



HAL
open science

Population dynamics of RNA arthropod-borne viruses in experimental evolution

Quang-Dinh Tran

► **To cite this version:**

Quang-Dinh Tran. Population dynamics of RNA arthropod-borne viruses in experimental evolution. Microbiology and Parasitology. Université Paris Cité, 2022. English. NNT : 2022UNIP5207 . tel-04677750

HAL Id: tel-04677750

<https://theses.hal.science/tel-04677750v1>

Submitted on 26 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris Cité
Institut PASTEUR
Ecole doctorale 562 – Bio Sorbonne Paris Cité (BioSPC)

Population dynamics of RNA arthropod-borne viruses in experimental evolution

Unité des Populations Virales et Pathogénèse – UMR 3569

Par **Quang Dinh TRAN**

Thèse de doctorat de Microbiologie, spécialité Virologie

Dirigée par **Dr. Marco VIGNUZZI**

Présentée et soutenue publiquement le 16 décembre 2022

Devant un jury composé de :

Marco VIGNUZZI	DR	Université Paris Cité	Directeur de thèse
Ron GELLER	DR	University of Valencia	Rapporteur
Dimitri LAVILLETTE	PU	University of Chinese Academy of Sciences	Rapporteur
Gisa GEROLD	PU	University of Veterinary Medicine Hannover	Examinatrice
Sandie MUNIER	MCU	Université Paris Cité	Examinatrice

RÉSUMÉ

Les virus à ARN sont l'une des entités biologiques évoluant le plus rapidement. À l'aide de leur taux de mutation, recombinaison et replication élevés, ainsi que leurs petits génomes, ils s'adaptent aisément à de nouveaux hôtes, causant régulièrement des épidémies. Malgré leur simplicité, leur capacité à interagir entre eux par complémentation et d'autres formes de comportements collectifs fait émerger des dynamiques populationnelles complexes. Dans mon travail de thèse sur deux arbovirus, j'explore deux aspects de ces interactions sociales et leur influence sur l'évolution virale.

Une première partie de ce manuscrit s'intéresse au rôle des génomes viraux défectifs (DVG) généralement décrits comme des sous-produits délétaires de la polymérase ARN-dépendante des virus à ARN. Dans ces travaux, nous avons testé si les DVG étaient capables d'interagir positivement avec les génomes viraux standards. Premièrement, nous avons identifié une mutation de résistance à la ribavirine située dans la polymérase virale qui restaure les titres viraux en conditions mutagènes. Nous avons par la suite introduit cette mutation dans un DVG du virus Zika précédemment identifié, et démontré que ces DVGs accélèrent l'adaptation virale au traitement par recombinaison génétique. Nous avons montré que l'absence d'observation de complémentation était causée par l'impossibilité du domaine polymérase de NS5 à être *trans*-complémenté. Enfin, nous avons montré que non seulement les DVGs pouvaient être empaquetés en *trans*, mais qu'ils pouvaient également être co-empaquetés avec des génomes standards afin d'assurer leur maintien.

Dans une seconde partie, j'ai ré-étudié l'hypothèse du compromis évolutif (trade-off) des arbovirus qui expliquerait leurs faibles taux de substitution dans la nature. En utilisant des clones du virus chikungunya barcodés et des techniques de séquençage à haut débit, j'ai proposé un design expérimental capable de résoudre les contradictions entre précédentes études. J'ai identifié que la diversité génétique variait selon les régions du génome et le régime de sélection. Les virus soumis à l'alternation entre hôtes montrent une diversité restreinte à chaque locus au niveau de l'hôte le moins permissif, ainsi que pour l'accumulation des DVG spécifiques. L'analyse du nuage de variant a montré qu'ils avaient acquis en partie des mutations associées à l'adaptation chez le moustique, ainsi que des mutations spécifiques au régime alterné. Ces observations s'expliquent par l'asymétrie du trade-off: l'adaptation initiale aux cellules de mammifères a un coût adaptatif en cellules d'insectes, ce qui empêche ces mutations d'apparaître dans le régime alterné. De plus, nous n'avons pas observé de coût du généralisme, et nous proposons que cela s'explique par un paysage adaptatif différent avoisinant le chemin évolutif d'adaptation aux cellules d'insectes.

Mots-clés: virus chikungunya, virus Zika, virus à ARN, évolution expérimentale, sociovirologie, génomes viraux défectifs, complémentation, recombinaison, trade-off, arbovirus, séquençage à haut débit

ABSTRACT

RNA viruses are one of the fastest evolving biological entities. With their high mutation and recombination rates, small genomes, and fast replication rates, they easily adapt to new hosts, causing outbreaks regularly. Despite their simplicity, their ability to interact with each other through complementation and other forms of collective behaviors gives rise to complex population dynamics. In this work, I explore two aspects of these social interactions and their influence on virus evolution, using two arthropod-borne viruses.

A first part of this manuscript investigates the role of defective viral genomes (DVGs), which were typically described as ubiquitous deleterious by-products of the erroneous RNA-dependent RNA-polymerase of RNA viruses. In this work, we tested if these DVGs could positively interact with standard viral genomes. We first identified a ribavirin-resistance mutation in the viral polymerase that rescued viral titers in mutagenic conditions. We then introduced it in a previously described DVG of Zika virus, and showed that DVGs carrying this mutation accelerated viral population adaptation to ribavirin treatment through genetic recombination. We then showed that absence of observed complementation was due to a *trans*-complementation incompetent NS5 polymerase domain. Finally, we showed that DVGs could not only be *trans*-packaged, but they could also be co-packaged with standard genomes, ensuring their maintenance across new infectious cycles.

In the second half, I re-explored the question of the arbovirus trade-off hypothesis which is hypothesized to explain the low substitution rates of arboviruses in nature. Using barcoded chikungunya virus clones and high-throughput sequencing techniques, I proposed an extended experimental design to resolve conclusion discrepancies among previous studies. In a first phase of experimental evolution, I identified that genetic diversity was accumulated differentially in-between genomic regions and among passing regimes. Alternating passages were restricted in genetic diversity by the most stringent host at each locus, and did not tolerate host-specific DVGs. Mutant cloud analysis showed that they in part acquired insect-associated adaptive mutations, but also alternation-specific ones. This pattern was in turn explained by the observation of an asymmetric trade-off: initial adaptation to mammalian cells has a fitness cost in insect cells, blocking such mutations to appear in alternation. Further, no cost of generalism was observed, which we propose is explained by a different fitness landscape neighborhood around the insect-cell adaptive pathway.

Keywords: chikungunya virus, Zika virus, RNA virus, experimental evolution, sociovirology, defective viral genomes, complementation, recombination, trade-off, arbovirus, high-throughput sequencing

Résumé substantiel

Les virus à ARN sont des entités biologiques évoluant rapidement. Ils se distinguent par leurs petites tailles de génome, leurs forts taux de mutation et de recombinaison dus aux erreurs commises par leur ARN polymérase ARN-dépendante, et des taux de réplication typiquement très élevés. Ces caractéristiques en font des modèles biologiques idéaux pour l'étude des processus évolutifs en laboratoire. Néanmoins, la virologie évolutive est encore une discipline relativement jeune, ayant pris son essor depuis les années 1960 avec la découverte des taux élevés de mutation et de substitution des virus à ARN. Ces propriétés l'ont au départ éloigné de la génétique des populations issue de la synthèse moderne de la théorie de l'évolution, au profit du modèle des quasi-espèces développé par Manfred Eigen qui a dominé le cadre théorique de la discipline jusqu'à récemment. Cette théorie des quasi-espèces a popularisé l'idée que la sélection naturelle chez ces virus n'opérait pas à l'échelle des génomes individuels, mais plutôt au niveau d'un réseau de génomes interconnectés par des mutations abondantes. Malgré les controverses et les limitations du modèle, l'idée d'une sélection de groupe à l'échelle de la population est aujourd'hui portée par la discipline de la sociovirologie reposant sur les théories de la sociobiologie.

La théorie des quasi-espèces prédit que le taux de mutation des virus à ARN est tel que la fréquence de chaque haplotype ne dépend pas majoritairement de son taux de réplication et de sa valeur adaptative propres, mais également de sa génération par d'autres haplotypes par des processus mutationnels. Ainsi, l'effet de la sélection naturelle sur les fréquences alléliques ne pourrait agir directement sur les génomes isolés, mais opérerait plutôt sur le réseau de variants. Certains auteurs ont vivement critiqué l'application de la théorie des quasi-espèces aux virus à ARN, argumentant principalement que le taux de mutation chez ceux-ci étaient en-dessous de celui pré-requis pour permettre une telle dynamique populationnelle. Dans mon introduction générale, je présente d'autres mécanismes permettant des dynamiques de sélection de groupe chez ces virus et ne reposant pas sur des critères d'application débattus.

Tout d'abord, la capacité des virus à se compléter en *trans* au sein d'une même cellule infectée permet l'émergence d'interactions génome-génome influençant la valeur adaptative de chacune des molécules interagissantes. Ce phénomène est notamment observé lorsqu'un génome « aidant » vient compléter une fonction endommagée d'un génome dit défectif, permettant à ce dernier d'achever son cycle de réplication et de se propager. Par opposition, certains génomes défectifs interfèrent avec la réplication des génomes viraux standards, réduisant ainsi leur valeur sélective. Par ailleurs, la complémentation permet également à des mutations bénéfiques présentes sur des génomes différents d'interagir au niveau protéique et faire émerger de nouveaux phénotypes. Ceci permet notamment de contourner de potentielles contraintes épistatiques entre mutations bénéfiques. À noter, certains variants minoritaires sont capables d'influencer de manière dominante la valeur adaptative de la population virale à des fréquences alléliques faibles.

De plus, certains virus ont tendance à se disperser en groupe, formant des unités d'infection collective capable d'augmenter localement la multiplicité d'infection, et donc la fréquence de co-infection. Les événements d'infection « en masse » permettent non seulement à la complémentation entre génomes d'opérer, mais font également émerger des dynamiques

coopératives intrinsèques. En effet, un nombre initial plus important de virions infectant une même cellule permet d'une part d'accélérer la réplication des génomes, d'autre part d'antagoniser de manière collective les réponses immunitaires de l'hôte.

Les travaux de ma thèse portent sur ces aspects populationnels de l'évolution des virus à ARN, et utilisent comme modèle deux virus transmis par les moustiques, les virus Zika et virus du chikungunya. Dans un premier chapitre, je m'intéresse aux génomes viraux défectifs (*defective viral genomes*, DVG) et leur capacité ou non à interagir positivement avec des génomes viraux standards. Dans un second chapitre, je revisite la question de l'adaptation des arbovirus à alterner dans leur cycle de vie entre hôtes mammifères et insectes, et caractérise par séquençage profond la composition en variants contribuant à l'adaptation à chaque hôte, ou l'alternance entre les deux.

Dans mon premier chapitre, je me suis intéressé aux rôles des DVG dans l'évolution des populations virales. Ces DVG sont des génomes viraux incapables de compléter un cycle de réplication en l'absence de génomes standards venant compléter leurs fonctions endommagées. Pour la plupart, ces DVG possèdent de larges délétions ou sont des duplications des régions non-traduites à l'extrémité du génome (les « copy-backs »). De ce fait, ces génomes sont typiquement de petites tailles, se répliquent plus rapidement et entrent alors en compétition avec les génomes viraux standards pour les facteurs cellulaires et viraux nécessaires à leur réplication. De plus, ces DVG, et plus particulièrement les copy-backs, sont capables de stimuler les voies de l'immunité, induisant ainsi une réponse antivirale de l'hôte et une réponse inflammatoire. Ainsi, les DVG sont très souvent associés à un rôle d'interférence avec la réplication virale. Néanmoins, à l'exception d'un rôle potentiel dans la mise en place d'infection persistante, aucune interaction positive des DVG n'a encore été décrite. Nous avons alors imaginé deux scénarios possibles par lesquels les DVG pourraient contribuer à l'adaptation des populations virales : par complémentation ou recombinaison de mutations bénéfiques.

Dans un premier temps, nous avons identifié deux mutations de résistance à la ribavirine (un analogue de nucléoside mutagène) par séquençage profond d'expériences d'évolution expérimentale du virus Zika. Après validation du phénotype de résistance, nous avons introduit l'une de ces mutations dans un DVG précédemment caractérisé dans le laboratoire afin de pouvoir tester son effet sur le reste de la population virale. Nous avons également introduit différentes mutations synonymes additionnelles afin de pouvoir distinguer par séquençage à haut débit notre mutation de résistance introduite dans le DVG de mutations apparues *de novo*. Après quatre passages dans des cellules préalablement traitées à la ribavirine du virus Zika standards en présence de DVG contenant la mutation de résistance, nous avons démontré que la mutation avait permis de restaurer en partie les titres viraux lors d'un traitement à la ribavirine. Par séquençage à haut débit, nous avons pu démontrer que ce phénotype était dû à un événement de recombinaison entre génomes viraux standards et DVG, suggérant que les mutations acquises par les DVG pouvaient contribuer à l'évolution des virus.

Nous avons également montré par une technique de RNA FISH que le DVG identifié était capable d'être co-encapsidé avec des génomes standards, assurant ainsi son maintien au sein de la population virale. Par ailleurs, nous avons montré que la mutation de résistance se

trouvait sur un domaine non-complémentable du génome du virus Zika, ne nous permettant pas de conclure sur la possibilité de complémentation de génomes standards par des DVG.

Dans mon deuxième chapitre, j'ai ré-exploré la question du compromis évolutif des arbovirus à circuler dans des cycles entre hôtes insecte et vertébrés. En effet, les arbovirus ont des taux de substitutions environ 10 fois inférieurs aux virus à ARN ayant des taux de mutations similaires. L'une des hypothèses avancées pour expliquer cette différence est que les deux hôtes imposent des pressions de sélections très différentes, imposant un compromis évolutif sur ces virus : une mutation bénéfique dans un hôte a de grandes chances d'être délétère dans l'autre. Bien que ce sujet a été étudié à de nombreuses reprises au début du siècle, aucun consensus sur l'existence d'un tel compromis n'a pu être établi. Par ailleurs, la caractérisation génétique des mutations adaptatives au cours de ces expériences d'évolution expérimentale dans ces précédents travaux n'ont été réalisés qu'en séquençage Sanger, et n'ont permis d'identifier que des variants majoritaires. Néanmoins, nous savons désormais que les variants minoritaires sont capables d'influencer la dynamique des populations virales, et leur identification permettrait de mieux comprendre les contraintes qu'impose le cycle de vie des arbovirus.

Durant ma thèse, j'ai ainsi réalisé vingt passages de virus chikungunya barcodés dans des cellules de U4.4 de moustique *Aedes albopictus* et dans des cellules BHK-21 de hamster de manière alternée, ou dans un seul des types cellulaires. J'ai par la suite réalisé un séquençage profond de mes virus dérivés à intervalle de passages régulier et mesuré leurs valeurs sélectives relatives respectives. L'analyse des résultats de séquençage profond a permis de montrer que la diversité génétique au cours des passages variait selon la cellule hôte, mais également selon la région génétique observée. Ainsi, un type cellulaire peut être plus permissive à l'accumulation de mutations dans certains gènes, et inversement plus réfractaire dans d'autres régions. Le régime de sélection alterné quant à lui impose une restriction de la diversité génétique suivant celle de l'hôte le plus réfractaire tout le long du génome. De plus, l'accumulation de DVG est également restreinte par les deux hôtes dans un régime alternant. Cela conforte l'hypothèse d'un compromis évolutif restreignant le nombre de mutations pouvant être positivement sélectionnées chez les deux hôtes. Une analyse en composante principale de l'ensemble des variants identifiés lors du séquençage nous a permis de déterminer des variants minoritaires responsables de l'adaptation à chacun des hôtes. Notamment, nous avons observé une abondante accumulation de mutations non-synonymes dans le gène d'enveloppe E2 lorsque le virus a été passé en cellules de hamster. En outre, les virus passés en régime alterné tendent à suivre la trajectoire évolutive d'adaptation aux cellules de moustiques, mais accumulent également des mutations spécifiques à l'alternance entre les deux types cellulaires.

La mesure des valeurs sélectives relatives des populations virales dérivées par mise en compétition avec un clone barcodé ancestral du virus chikungunya et mesure d'abondance relative des code-barres génétiques a permis de mettre en évidence un compromis évolutif asymétrique : l'adaptation aux cellules de hamster a induit un coût adaptatif chez les cellules de moustiques, tandis que l'adaptation à ces dernières n'a pas de coût sélectif apparent. De plus, nous n'avons pas observé de coût du généralisme lorsque les virus ont été soumis à un régime alterné : l'augmentation des valeurs sélectives dans les deux types cellulaires ne sont pas significativement différents des deux autres régimes de sélection. Mis en relation avec les

données de séquençage précédemment décrites, ces résultats suggèrent que le régime alterné a poussé les populations virales à s'adapter dans un premier temps aux cellules de moustiques car ne provoquant pas de coût sélectif, et permis l'acquisition de mutations bénéfiques en cellules de mammifères sans coût adaptatif chez les cellules de moustiques.

Ainsi, au cours de mes travaux de thèse, j'ai pu montrer qu'une approche populationnelle de la virologie évolutive permet une meilleure compréhension de l'évolution des arbovirus. En outre, j'ai pu démontrer que les DVG ne sont pas forcément des génomes affectant négativement les populations virales, mais qu'ils peuvent dans certains cas accélérer l'adaptation de ces populations. J'ai également montré que ces DVG étaient capables de se disperser collectivement avec des génomes viraux standards, leur permettant ainsi de se maintenir dans la population. Cela suggère qu'il pourrait exister des mécanismes promouvant la co-encapsulation de ces génomes défectifs. Enfin, par analyse par séquençage profond d'expériences d'évolution expérimentale visant à étudier le compromis évolutif du cycle alterné des arbovirus, j'ai pu démontrer que les deux hôtes restreignaient l'accumulation de mutations et de DVG dans un schéma alterné. Par ailleurs, j'ai également pu également identifier des variants spécifiques à l'adaptation à chaque type cellulaire, et déterminé que l'adaptation au régime alterné suivait préférentiellement la trajectoire évolutive d'adaptation aux cellules de moustiques. Cela s'explique par l'absence de coût adaptatif initial à l'adaptation à ces cellules, contrairement aux cellules de mammifères. Finalement, l'identification de mutations spécifiques au régime alterné suggère que ce cycle imposerait des contraintes sélectives additionnelles.

ACKNOWLEDGMENTS

I want to first thank Marco for welcoming me in his lab when I was just a Master student with no prior knowledge in virology, and then trusting me enough to push me to do my PhD. Thank you for the pep talks whenever I was feeling not worthy enough to be in academia. You really know how to boost people's self-confidence.

Thank you to all the past members of the PVP lab, the ones that welcomed me five years ago with all your crazy energy, and those that stayed until the end. I learned and grew up a lot as a human being and a scientist thanks to you.

Special thanks to Vero for being my supervisor during my Master, and an unofficial one for the first years of my PhD. You taught me so much about science, and believed in my ability to pull off this degree. I would have never finished my PhD on time without your help! I miss our late night chats/rants. You are truly an amazing person!

Thank you Steph for being my office mate for all these years. Sorry you had to deal with my daily rants. Thank you Thomas for all your help, being here when I was breaking down, and all the gossips. Alice and Alex for all the fun when the lab was slowly coming to its end.

My sincere thanks to Nolwenn with whom we collaborated, and who helped a lot writing our common manuscript. Thanks also to the entire V2S team for all the fun and parties, you guys are craaaazy!

Also a big thank you to all my substitute teams that welcomed me when I became a labless PhD student. The VIA lab for giving me a bench space and inviting me to their lab meetings. It felt good to be shortly part of a growing and vibrant lab. The UVI lab for letting me keep my office and populating a space that was starting to feel empty and lonely. The Vartanian group for many shared lunches.

A special thank to all the colleagues and friends that I made along the way: Cass for your mental support and last minute corrections; Kyle for following me on an insane bike trip, our regular chats and more generally for making what could have been depressing months a lot more enjoyable; Mauro for being the best roommate at the virology department I could have asked for; Anvita and Ségo for all your positive energy; Sophie for offering your help the day you submitted your manuscript; Elodie for dealing with my chaotic energy and rants in the office; Jill-Léa for creating the best bike team on the campus; people from the Stapa bureau for feeding my alcoholism every Friday nights.

Thanks also to all my friends outside of Pasteur that supported me. Diem and Colas for being the best friends one could ask for and acting as my substitute therapists whenever I was feeling low. My friends from med school: Amina, Alex, Natacha, Débo, Mélanie, Julie and all those that helped me go through all these years.

TABLE OF CONTENTS

RÉSUMÉ.....	3
ABSTRACT.....	5
ACKNOWLEDGMENTS.....	7
TABLE OF CONTENTS.....	8
LIST OF FIGURES.....	12
LIST OF TABLES.....	14
LIST OF BOXES.....	14
LIST OF ABBREVIATIONS.....	15
PREAMBLE.....	19
GENERAL INTRODUCTION.....	23
1. MECHANISMS OF RNA VIRUS EVOLUTION.....	25
1.1. <i>High mutation rates in RNA viruses</i>	25
1.2. <i>Recombination and reassortment: forms of sex?</i>	27
1.3. <i>Natural selection and distribution of fitness effects</i>	30
1.4. <i>Population sizes and genetic drift: the importance of bottlenecks</i>	32
1.5. <i>Genome organization, replication and mode of transmission</i>	33
1.6. <i>Epistasis and pleiotropy</i>	35
1.7. <i>Mutational robustness</i>	37
2. RNA VIRUS POPULATIONS AS A UNIT OF SELECTION: FROM QUASISPECIES TO SOCIAL EVOLUTION THEORIES.....	40
2.1. <i>Quasispecies theory: a controversial debate among virologists</i>	41
2.1.1. <i>Definition of quasispecies terminology</i>	41
2.1.2. <i>The error threshold</i>	43
2.1.3. <i>Survival of the flattest quasispecies</i>	44
2.1.4. <i>Limitations and controversies of quasispecies theory</i>	45
2.2. <i>Collective behaviors of viral populations</i>	46
2.2.1. <i>Complementation and minority variants</i>	46
2.2.2. <i>Cheater genomes</i>	48
2.2.3. <i>Collective infections</i>	48
2.2.4. <i>Strength in numbers: collaboration through ‘mass effect’</i>	51
2.2.5. <i>Viral communication</i>	51
2.3. <i>Social evolution theories: alternatives to quasispecies to account for virus-virus interactions</i>	52
2.3.1. <i>Brief introduction to social evolution</i>	53
2.3.2. <i>Applications to virus-virus interactions</i>	55
2.4. <i>Defective viral genomes are an integral part of viral populations</i>	56

CHAPTER 1: POSITIVE INTERACTIONS OF DEFECTIVE VIRAL GENOMES IN ZIKA VIRUS.....	61
1. INTRODUCTION.....	61
1.1. Definition and types of defective viral genomes.....	61
1.2. Factors of DVG generation and maintenance.....	62
1.3. Defective Interfering Particles (DIPs): DVGs as social cheaters.....	62
1.3.1. Interference of viral production.....	63
1.3.2. Stimulation of host immunity.....	64
1.3.3. Therapeutic Interfering Particles (TIPs): using viral parasites for control, prevention and treatment.....	65
1.4. Positive interactions of DVGs?.....	66
1.4.1. Role in viral persistence.....	66
1.4.2. Adjusting protein stoichiometry.....	67
1.4.3. Can DVGs complement wild-type genomes?.....	67
1.4.4. Can DVGs enhance population evolvability?.....	68
1.5. Zika virus.....	68
1.5.1. Genome and virus structure.....	68
1.5.2. Replication cycle.....	69
1.5.3. A ZIKV DVG to explore positive interactions.....	72
1.6. Aims of Chapter 1.....	75
2. ARTICLE: DEFECTIVE VIRAL GENOMES CAN ACCELERATE VIRAL POPULATION ADAPTATION.....	76
3. COMPLEMENTARY RESULTS.....	102
3.1. Material and methods of complementary results.....	102
3.1.1. Cells.....	102
3.1.2. Plasmids.....	102
3.1.3. Viruses.....	103
3.1.4. Site-directed mutagenesis by <i>in vivo</i> assembly.....	103
3.1.5. Viral titers.....	104
3.1.6. Complementation assays.....	104
3.1.7. DVG mutant distribution analysis.....	105
3.2. Results.....	106
3.2.1. Mutations in NS5 RdRp domain cannot be <i>trans</i> -complemented.....	106
3.2.2. DVGs show signs of relaxed purifying selection.....	107
4. DISCUSSION AND PERSPECTIVES.....	109
4.1. Summary of results.....	109
4.2. Recombination between DVGs and standard genomes.....	109
4.3. Complementation: an unresolved question.....	111
4.4. Co-dispersal of DVGs and WT genomes.....	112
4.5. DVG genetic diversity and relaxed purifying selection.....	113
4.6. Investigations in less artificial set-ups.....	114
4.7. Concluding remarks.....	114

CHAPTER 2: POPULATION DYNAMICS OF CHIKUNGUNYA VIRUS IN EXPERIMENTAL ALTERNATING PASSAGES.....	119
1. INTRODUCTION.....	119
1.1. <i>Arboviruses life cycles and the trade-off hypothesis</i>	119
1.2. <i>A review on previous experimental studies</i>	122
1.3. <i>Chikungunya virus</i>	125
1.3.1. <i>Genome and virus structure</i>	125
1.3.2. <i>Replication cycle</i>	126
1.3.3. <i>Epidemiology</i>	130
1.4.4. <i>Experimental design and Aims of Chapter 2</i>	131
2. MATERIAL AND METHODS.....	134
2.1. <i>Cells, plasmids and viruses</i>	134
2.1.1. <i>Cells</i>	134
2.1.2. <i>Plasmids</i>	134
2.1.3. <i>Viruses</i>	135
2.2. <i>Site-directed mutagenesis</i>	136
2.3. <i>Viral titers</i>	136
2.4. <i>Viral experimental evolution passages</i>	137
2.5. <i>Competition fitness assays</i>	137
2.5.1. <i>Competition passages</i>	137
2.5.2. <i>Multiplex RT-qPCR</i>	137
2.5.3. <i>Relative fitness</i>	139
2.6. <i>Viral RNA extraction</i>	140
2.7. <i>Viral RNA sequencing</i>	140
2.8. <i>Sequencing data alignment and variant calling</i>	141
2.9. <i>Defective viral genomes calling</i>	142
2.10. <i>Shannon entropy</i>	142
2.11. <i>Principal Component Analysis</i>	142
2.12. <i>Statistical analysis</i>	143
3. RESULTS.....	144
3.1. <i>Alternating passages have restricted genetic diversity determined by permissivity of both hosts</i>	144
3.2. <i>Alternation between insects and mammals limits accumulation of defective viral genomes</i>	146
3.3. <i>Evolution of alternating passages follows insect-cell adaptive pathway while also selecting for alternation-specific mutations</i>	147
3.4. <i>Initial phase of experimental evolution does not detect cost of generalism and shows an asymmetric trade-off cost</i>	150
4. DISCUSSION AND PERSPECTIVES.....	153
4.1. <i>Summary of results</i>	153
4.2. <i>Genetic diversity varies along the virus genome</i>	153
4.3. <i>The insect host partially drives the evolution of the alternating passages</i>	155

4.4. <i>Alternation-specific mutations</i>	156
4.5. <i>The trade-off hypothesis: a pending question & future investigations</i>	156
REFERENCES	161
ANNEXES	183
1. DEFECTIVE VIRAL GENOMES AS THERAPEUTIC INTERFERING PARTICLES AGAINST FLAVIVIRUS INFECTION IN MAMMALIAN AND MOSQUITO HOSTS.....	185
2. BRD2 INHIBITION BLOCKS SARS-CoV-2 INFECTION BY REDUCING TRANSCRIPTION OF THE HOST CELL RECEPTOR ACE2.....	200
3. IDENTIFICATION OF DAXX AS A RESTRICTION FACTOR OF SARS-CoV-2 THROUGH A CRISPR/Cas9 SCREEN.....	215
4. A SARS-CoV-2 PROTEIN INTERACTION MAP REVEALS TARGETS FOR DRUG REPURPOSING.....	229
5. THE GLOBAL PHOSPHORYLATION LANDSCAPE OF SARS-CoV-2 INFECTION.....	245
6. HOST PDZ-CONTAINING PROTEINS TARGETED BY SARS-CoV-2.....	289

LIST OF FIGURES

Figure 1 Comparison of per-site mutation rates of different RNA viruses (orange) and other biological entities.....	26
Figure 2 Types of copy-choice recombination products.....	28
Figure 3 Distribution of lethal, deleterious, neutral and beneficial effects of random mutations in vesicular stomatitis virus.....	31
Figure 4 Adaptive pathways on a fitness landscape.....	32
Figure 5 Different types of epistasis and environmental pleiotropy.....	37
Figure 6 Simplified representation of a quasispecies.....	42
Figure 7 Survival of the flattest at high mutation rates or in quasispecies.....	44
Figure 8 Types of collective infectious units.....	50
Figure 9 Types of viral collective behaviors.....	52
Figure 10 Types of defective viral genomes.....	62
Figure 11 Modes of interference of defective viral genomes.....	65
Figure 12 Genome and polyprotein organizations and viral particle structure of Zika virus..	69
Figure 13 Zika virus replication cycle.....	71
Figure 14 The candidate ZIKV DVG requires a conserved ORF and WT-encoded NS1 in trans to be replicated and show potent interfering properties in vitro.....	74
Figure 15 NS5 polymerase activity cannot be trans-complemented to rescue DVG reporter replication.....	107
Figure 16 Mutations in DVGs can reach higher frequencies than in standard genomes.....	108
Figure 17 Arboviruses' sylvatic and urban life cycles.....	119
Figure 18 Model of population fitness in alternating environments.....	121
Figure 19 Genome organization and viral particle structure of chikungunya virus.....	126
Figure 20 Chikungunya virus replication cycle in mammalian cells.....	129
Figure 21 Two-phase experimental evolution design in constant or alternated passages of barcoded CHIKV in BHK-21 or U4.4 cells.....	133
Figure 22 Alternating passages accumulate less genetic diversity than single-host ones across passages.....	145

Figure 23 | DVG accumulation is limited by the two host tolerance in alternating passages...147

Figure 24 | Adaptive regimes show specific patterns of mutation acquisition.....149

Figure 25 | Competition assays in single-host challenges show linearity of fitness effects over passages..... 151

Figure 26 | Competition assays reveal an asymmetric trade-off in serial passages and no significant cost of generalism in alternating ones..... 152

LIST OF TABLES

Table 1 Zika virus proteins and functions.....	72
Table 2 Primers for ZIKV IVA site-directed mutagenesis.....	104
Table 3 Experimental evolution designs and results investigating the trade-off hypothesis in arboviruses.....	124
Table 4 Chikungunya virus proteins and functions.....	130
Table 5 Site-directed mutagenesis primers for barcoded CHIKV clones.....	135
Table 6 Probes and primers for CHIKV multiplex RT-qPCR.....	139

LIST OF BOXES

Box 1 Major points of the general introduction.....	58
---	----

LIST OF ABBREVIATIONS

C	Capsid protein (ZIKV and CHIKV)
CHIKV	Chikungunya virus
CIU	Collective infectious unit
COVID-19	Coronavirus disease 2019
C _T	Cycle threshold
CRE	<i>Cis</i> -regulatory element
DENV-2	Dengue virus serotype 2
DFE	Distribution of fitness effects
DIP	Defective interfering particle
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
DelVG	Deletion-containing viral genome
DVG	Defective viral genome
E	Envelope protein (ZIKV)
E1	Envelope protein 1 (CHIKV)
ECSA	East-Central-South-African
ER	Endoplasmic reticulum
ESS	Evolutionary stable strategy
GAG	Glycosaminoglycan
Hsp	Heat-shock protein
IFN	Interferon
IL	Interleukin
IOL	Indian Ocean lineage
IVA	<i>In vivo</i> assembly
M	Membrane protein (ZIKV)
MAVS	Mitochondria antiviral-signaling protein
MOI	Multiplicity of infection
mRNA	Messenger RNA
Mxra8	Matrix-remodeling associated protein 8
NS1	Non-structural protein 1 (ZIKV)

nsP1	Non-structural protein 1 (CHIKV)
ORF	Open reading frame
PCR	Polymerase chain reaction
PFU	Plaque-forming unit
poly-A	Poly-adenyl
pr	Precursor peptide (ZIKV)
prM	Precursor membrane protein (ZIKV)
PRR	Pattern recognition receptor
RBD	Receptor-binding domain
RdRp	RNA-dependent RNA polymerase
RLR	RIG-I-like receptor
RNAi	RNA interference
RT-qPCR	Quantitative reverse transcription PCR
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
sfRNA	Subgenomic flavivirus RNA
sgRNA	Subgenomic RNA
siRNA	Small interfering RNA
smFISH	Single-molecule fluorescence <i>in situ</i> hybridization
SNV	Single-nucleotide variant
ssDNA	Single-stranded DNA
+ssRNA	Positive-sense single-stranded RNA
-ssRNA	Negative-sense single-stranded RNA
TIM-1	T-cell immunoglobulin and mucin domain 1
TIP	Therapeutic interfering particle
TLR	Toll-like receptor
TNF	Tumor necrosis factor
TNFR1	TNF receptor 1
TNFR2	TNF receptor 2
TRAF1	TNF receptor associated factor 1
UTR	Untranslated region
vDNA	Viral DNA
VP	Vesicle pocket
vRNA	Viral RNA

ZIKV

Zika virus

PREAMBLE

This manuscript is divided into two independent chapters.

The first one argues for the possibility for defective viral genomes to play positive roles in virus populations, notably through genetic recombination. This chapter is mainly written and formatted as a preprint scientific article and has been submitted once to Nature Microbiology, but was desk-rejected. We aim to resubmit the manuscript to another journal such as PLOS Pathogens or PNAS after adding more controls to some of the experiments.

The second chapter describes chikungunya virus population compositions and fitness phenotypes in experimental evolution in mammalian-only, insect-only or alternated passages. It was the initial project of my Ph.D. but has been put aside in favor of the other one during my third year of PhD in order to be able to submit and defend in the context of the Viral Populations and Pathogenesis Unit (my laboratory) closing down. It is an incomplete story in regards to the initial objectives, but will nonetheless be reformatted to be submitted to a specialist journal such as Virus Evolution.

In addition, the general introduction has been written with the aim to be reformatted and submitted as a review for Nature Reviews Microbiology on the topic of “Mechanisms of RNA virus evolution”. An initial outline has already been submitted and accepted by the editor, and the initial submission of the manuscript proof is planned by the end of December 2022.

Finally, I have been involved at the beginning of my PhD in different scientific collaborations to respond to the COVID-19 pandemic crisis, which yielded several co-author publications that I’ve added in the Annexes section. This work does not, however, find an integral place in my thesis as they do not integrate with the main topic of my dissertation.

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Viruses are the most abundant biological entities on the planet^{1,2}, infecting virtually every branch of cellular life³⁻⁷ and even other viruses⁸. They are obligate parasites of their hosts, occasionally causing diseases, but their evolutionary history is independent due to their ability to cross species barriers^{1,4,9} on regular occasions throughout life's history. They play key roles in horizontal gene flows^{10,11}, the regulation of ecosystems⁶ and nutrient cycles⁶. In a more anthropocentric view, they are responsible for major outbreaks throughout human history, with a recent rise in public interest with the coronavirus disease 2019 (COVID-19) pandemic caused by SARS-CoV-2.

Despite viruses' importance in shaping the Tree of Life¹¹, evolutionary virology – the study of virus evolution – is a relatively young field¹², inaugurated by papers on RNA virus high mutation^{13,14} and evolutionary rates^{15,16} in the 1960s and 1970s. These findings highlighted the key characteristic of RNA viruses that set them apart from other evolving entities: they have one of the highest mutation rates⁹, and subsequently evolve rapidly. This property led to the application of quasispecies theory – an alternative model to classical population genetics – as the main theoretical framework for RNA virus evolution for several decades¹⁷⁻¹⁹. Despite its controversies and limitations^{9,20}, the theory popularized a fundamental idea that is now carried by social evolution theories applied to virology^{21,22}: that natural selection may act at the population level.

This work focuses on this population approach of evolutionary virology using two arthropod-borne viruses (arboviruses) with the goals to (1) explore the potential beneficial role that defective viral genomes play in virus evolution using a Zika virus defective genome previously described in the laboratory²³, and (2) characterize the population dynamics of chikungunya virus in experimental alternating cycles between insect and mammalian cell lines.

The following introduction aims at providing the reader with general knowledge on the mechanisms of RNA virus evolution, and the evolution of the idea of virus population as a unit of selection. I then briefly review in the introduction of Chapter 1 the generation mechanisms and roles of defective viral genomes, and highlight the lack of literature on positive interactions with wild-type genomes. Finally, in Chapter 2, I present the specificity of the

arbovirus life cycle and review the current literature on the experimental assessment of the trade-off hypothesis associated with generalism.

1. Mechanisms of RNA virus evolution

RNA viruses are fast-evolving biological entities, continuously engaged in an arms race against the cellular organisms they infect. Their evolutionary (or substitution) rates – the numbers of substitutions that reach fixation per nucleotide site and per year – range between 10^{-5} to 10^{-2} substitutions/nucleotide/year (sub/nt/year), with a median rate of 10^{-3} sub/nt/year^{12,24}. This number is many orders of magnitude above those of double-stranded DNA (dsDNA) viruses (typically between 10^{-9} to 10^{-7} sub/nt/year), one or more orders above single-stranded DNA (ssDNA) ones²⁵. Hence, they are ideal model systems to study evolutionary processes at the human-life timescale and have been used to test evolutionary theories such as Muller's ratchet²⁶ or the prisoner's dilemma²⁷.

In the modern synthesis of evolution, population genetics is a powerful theoretical framework used to decipher and explain molecular polymorphism observed in nature. It simply defines evolution as the change of allele frequencies in a population and defines four major forces that influence said frequencies: mutation, selection, genetic drift, and gene flow (or migration). In this section, I will review the principal mechanisms that modulate RNA virus micro-evolution through this lens, intentionally omitting quasispecies theory, which will be discussed in the following section.

1.1. High mutation rates in RNA viruses

If natural selection is by far the most popular force of 'Darwinian' evolution, to degrees such that evolution, as a whole, is commonly reduced to it, mutation is probably the most important of them. Indeed, it is the only force that creates genetic novelty (alleles) *de novo* in a population and takes the form of punctual substitutions, deletions or insertions. It may not be a coincidence that the main characteristic of RNA viruses that sets them apart from other organisms is their high mutation rates. With a range from 10^{-6} to 10^{-3} substitutions/nucleotide/cycle (sub/nt/cycle)²⁴, their error rates highly surpass those of DNA viruses (10^{-8} to 10^{-6} sub/nt/cycle)²⁴ (**Figure 1**). On average, these error rates translate to one mutation per genome replication cycle.

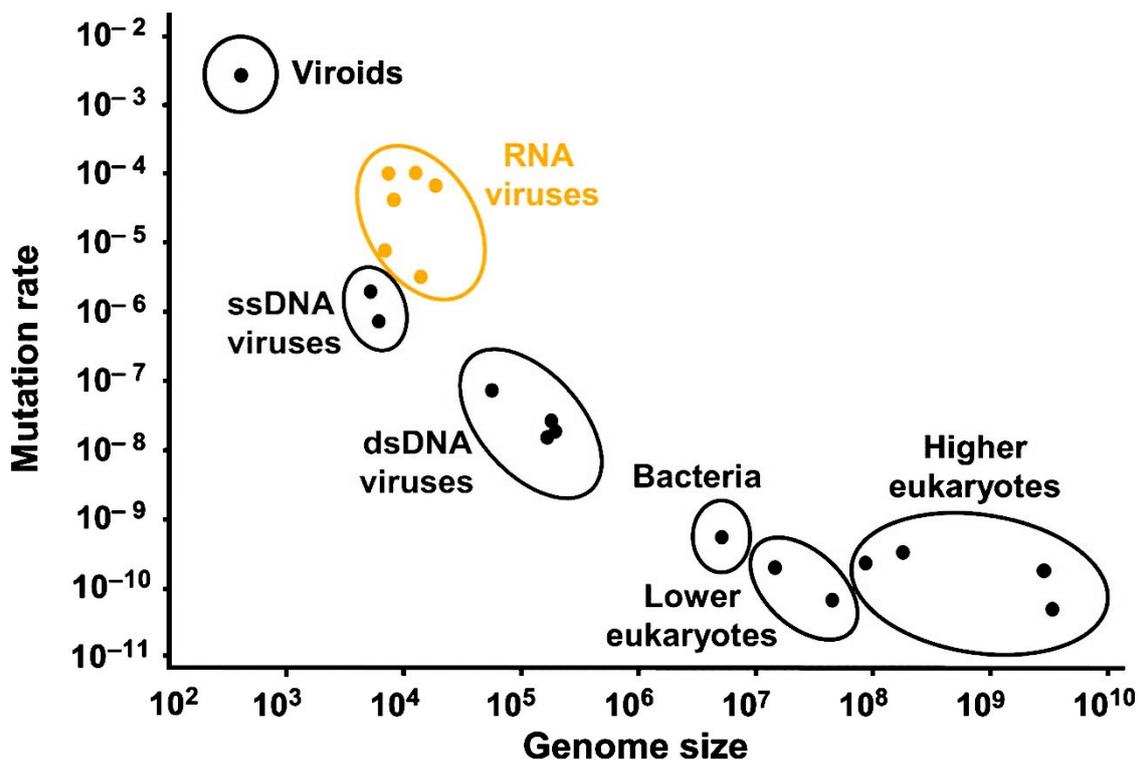


Figure 1 | Comparison of per-site mutation rates of different RNA viruses (orange) and other biological entities. Mutation rates are expressed in sub/nt/cycle and genome sizes in nucleotides. Adapted from Gago *et al.*, 2009²⁸.

A major determinant of these extremely high rates is the error-prone RNA-dependent RNA-polymerase (RdRp) of RNA viruses. While replication of genetic information is an erroneous process *per se*, mechanisms of mismatch repair and proofreading usually buffer the amount of mutations during synthesis in DNA organisms. In opposition, most RNA virus RdRps, with the exceptions of viruses in the *Coronaviridae* and *Roniviridae* families, lack 3'-to-5' exonuclease proofreading activity⁹. Moreover, mutations that affect the RdRp fidelity itself have been identified in many viruses²⁹⁻³³, allowing evolution of this heritable trait. Indeed, fitness cost of both low and high-fidelity variants have led many to believe that mutation rates in RNA viruses may be finely tuned to produce high genetic diversity while staying below a lethal threshold^{30,32-36}. However, biochemical assays have linked such mutations in poliovirus to be associated with modified replication speed³⁷. This evolutionary trade-off rather suggests that high mutation rates in RNA viruses are a rather tolerated than an optimized by-product of selection for faster replication^{37,38}.

Additionally, besides intrinsic polymerase error rates, other factors contribute to mutate viral genomes such as host cytosine deaminating enzymes from the APOBEC3³⁹⁻⁴¹ or

ADAR^{42,43} protein families. These enzymes are primarily used as a defense mechanism against viral infections, leading to hyper-mutated sterile genomes, but occasionally leave footprints on viral evolution³⁹. Of note, while they induce a bias in mutation types from cytosine to uracil, with sometimes specific 3' and 5' context bias such as with the APOBEC3 superfamily⁴⁴, this only influences the distribution of the probability of mutations but not their random nature.

1.2. Recombination and reassortment: forms of sex?

Recombination is often approached as a form of mutational process. This probably has to do with the fact that its main manifestation in RNA viruses, copy-choice recombination, occurs during replication by the RdRp. However, its effects on genetic diversity are of a different nature and can rather be assimilated with its sister process in segmented viruses, reassortment. Indeed, one can define recombination as the process that leads to the formation of a chimeric molecule from two or more parental genomes⁴⁵. As such, recombination does not really create new alleles, but rather shuffles them into new haplotypes.

The principal mechanism of recombination is copy-choice recombination. In this process, the RdRp starts the synthesis of the complementary strand of one parental RNA molecule (the donor template), then dissociates from the donor, switches to a different acceptor template and resumes synthesis, resulting in a chimeric molecule of mixed origins. The exchange can occur between the same genomic region in the two parental templates, in which case the resulting molecule is a fully complete genome and the recombination is termed 'homologous'. Conversely, if it occurs at different sites or between dissimilar parental molecules, the process is termed respectively 'non-homologous' or 'illegitimate'⁴⁵ (**Figure 2**). In the latter cases, the resulting product is often deleterious and is one of the main mechanisms for the production of defective viral genomes, which will be discussed in the introduction of Chapter 1.

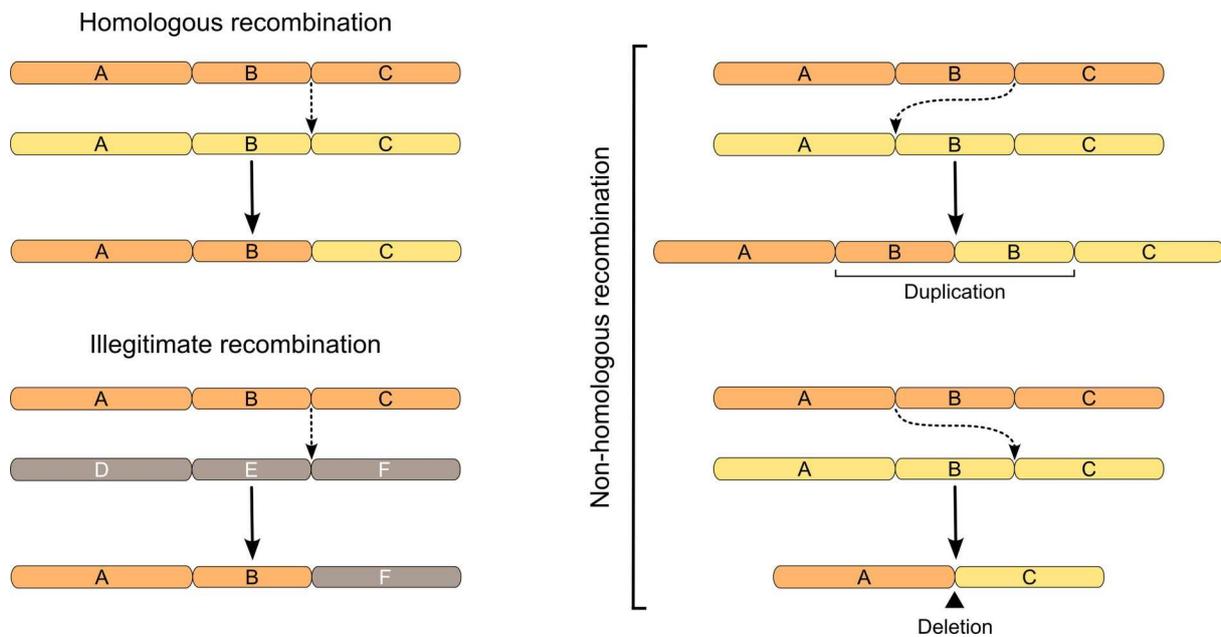


Figure 2 | Types of copy-choice recombination products. Homologous recombination products have the same genomic organization as their parental genomes. Non-homologous recombinations produce duplications or deletions depending if the RdRp resumes extension on the acceptor template respectively in 5' or in 3' of the dissociation site on the donor template. Illegitimate recombinations result in chimeric progeny with different set of genes than their parental genomes. Dashed arrows represent dissociation and reassociation events of the RdRp from the donor to the acceptor template. Adapted from Mac Kain *et al.*, 2021⁴⁶.

Alternative mechanisms of recombination involving the polymerase have also been theorized, such as the 'primer alignment-and-extension' model⁴⁷. This process involves the hybridization of the 3' end of a truncated RNA template with a complementary sequence. The truncated RNA then serves as a primer and is extended by the viral polymerase using the opposite polarity genome as a template^{46,47}. Non-replicative RNA recombination events have also been demonstrated experimentally^{48,49}. In these processes, truncated RNA templates (through cleavage, physical breakage, or aborted synthesis) spontaneously recombine through transesterification reactions^{46,49} or are enzymatically ligated by cellular proteins⁴⁶. The biological relevance of such mechanisms is still under debate⁴⁶.

Similar to mutations, replicative recombination is highly influenced by the erroneous replication of viral RdRps, and mutations affecting its rate are usually linked to altered fidelity and replication speed⁵⁰. However, while RNA viruses all typically display high rates of mutations, recombination rates vary immensely between RNA virus families, from almost clonal⁵¹ to highly recombinogenic⁵². However, these differences could be linked to differences induced by selection of recombinant products and should be validated through biochemical

assays. Factors such as sequence homology⁵² and RNA secondary structure⁵³ play major roles in the distribution of recombination breakpoints.

Reassortment, on the other hand, is specific to segmented viruses, whose genomes are divided into different RNA or DNA molecules. In this specific case, no chimeric molecule is produced, but individual segments are shuffled to produce progeny virions containing molecules of mixed parental origin. Reassortment is very common in segmented RNA viruses, with about 70% influenza A virus progeny being the result of reassortment after an *in vivo* infection^{55,56}. This process is highly constrained by RNA-protein and RNA-RNA signal compatibility^{57,58}, resulting in a skewed distribution of segments that participate in it⁵⁷⁻⁵⁹.

Both reassortment and recombination have been experimentally associated with increased genetic diversity and virulence^{16,60-62}, and are involved in virus emergence and re-emergence. For example, recombination between two distinct alphavirus species is responsible for the generation of western equine encephalitis virus⁶³, while multiple events of vaccine-derived poliovirus outbreaks are the result of recombination between live-attenuated vaccines and common enteroviruses⁶⁴. Moreover, compelling evidence shows high numbers of recombination events inside the sarbecovirus subgenus, which could be responsible for the recent SARS-CoV-2-related pandemic⁶⁵. Finally, reassortment in influenza A virus is one of the main drivers of the fast antigenic drift of seasonal flu⁶⁶.

These results have led many authors to hypothesize that recombination and reassortment are favored and fine-tuned by natural selection. Indeed, because both processes involve an exchange of parental genetic material resulting in chimeric progeny, recombination is often seen as a form of sex. As such, it falls into the prediction that sex can alleviate the accumulation of deleterious mutations in asexual populations predicted by Muller's ratchet^{26,67-69}, or accelerate the rate of adaptive evolution by combining advantageous mutations, thus reducing the impact of clonal interference between beneficial mutations at two different loci^{70,71}. However, other authors argue that non-adaptive explanations are more parsimonious to explain differences in observed recombination rates. Simon-Loriere and Holmes suggest that RNA virus large population sizes and high mutation rates are sufficient enough to counter Muller's ratchet and generate advantageous combinations of mutations⁴⁵. Rather, they propose that recombination rates are by-products of viral genome organization and life cycles: negative-sense single-stranded RNA (-ssRNA) viruses experience low rates of recombination because their RNA molecules are quickly bound and stabilized by

nucleoproteins; coronaviruses' high rates are encouraged by the necessity to produce subgenomic mRNAs through template switching⁴⁵.

1.3. Natural selection and distribution of fitness effects

Evidence of natural selection in RNA viruses is countless. From antiviral resistance, innate and adaptive immune escape, to adaptation to new hosts or tissues, selection gives a direction to evolution through adaptation. However, it is important to keep in mind that selection only acts on standing genetic variation already present in the population, and *de novo* mutations are produced randomly. It does not produce adaptive alleles but rather fixes among the ones present, those that are better adapted.

To model natural selection, population genetics uses the concept of 'fitness', which is defined as the expected number of viable progeny produced by one allele. While absolute fitness is difficult to measure experimentally in an unbiased manner, relative fitness w – the change in frequency of one allele in regard to the total population – is easier to assess and can be denoted as:

Equation 1
$$p(t+1) = \frac{w}{\bar{w}} p(t),$$

with $p(t)$ the frequency of the allele at generation t , and \bar{w} the mean relative fitness of the population, usually set to 1 for convenience. It is then possible to define a selection coefficient s as the relative advantage or disadvantage of one mutant allele against the reference wild-type as $w = 1 + s$. Of note, fitness and selection coefficient are not absolute concepts and are measured for specific environments and genetic backgrounds. Alleles can be advantageous in one condition and detrimental in another one, which will be further discussed in section 1.6.

A fundamental question in evolutionary biology concerns the proportion of mutants that are lethal, deleterious, neutral, or beneficial. Indeed, the pertinence of the neutral theory of molecular evolution theorized by Motoo Kimura^{72,73} as a null hypothesis and explanation of observed molecular polymorphism⁷⁴⁻⁷⁶ is still debated more than 50 years after its introduction. The underlying problem of the distribution of fitness effects (DFE) of mutations is even more relevant to RNA virus evolution given that high mutation and replication rates and small genome sizes allow RNA viruses to virtually explore all possible mutations. Initial assessment in vesicular stomatitis virus of the DFE in RNA viruses led by Sanjuán *et al.*

(**Figure 3**) confirmed the expectations that almost 70% of mutations were deleterious, including almost 40% lethal ones, while beneficial mutations were rare (less than 5%)⁷⁷. These observations were then further confirmed in other model bacteriophages^{78,79} and more recently with ultra-deep sequencing approaches in dengue virus⁸⁰. While the high proportions of deleterious mutations were expected for non-synonymous changes because of the compactness and lack of redundancy in RNA virus genomes, evidence of parallel fixation of non-coding or synonymous mutations⁸¹ suggests that these types of mutations may also not be strictly nor behave neutral⁸².

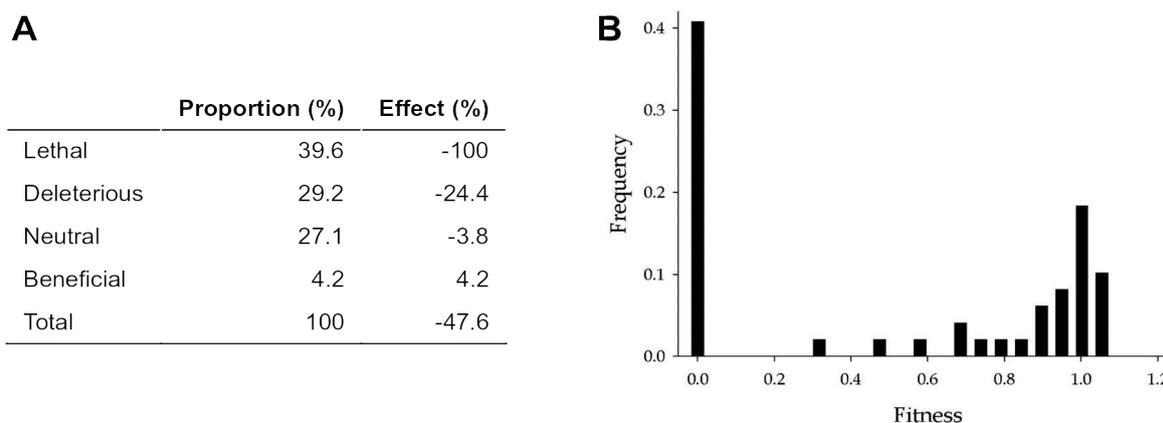


Figure 3 | Distribution of lethal, deleterious, neutral and beneficial effects of random mutations in vesicular stomatitis virus. (A) Table of proportions and mean fitness effects (in percentages of change) of each class of point mutations. **(B)** Histogram of the distribution of fitness effects. Adapted from Sanjuán *et al.*, 2004⁷⁷.

A common metaphor to visualize the directionality of evolution through natural selection is the adaptive fitness landscape. In this simplified model, genomes and/or populations explore a multi-dimensional sequence space comprising all possible combinations of mutations in which topology is defined by the fitness values of corresponding haplotypes. If selection is strong enough to dominate substitution rates, populations tend to climb fitness ‘hills’, increasing their mean fitness. The topography of this fitness landscape is therefore essential in defining the direction of adaptation: if the landscape only comprises a single hill, molecular evolution is highly parallel and the population eventually reaches the theoretical ‘highest peak’, while if multiple peaks exist and are distant enough, populations can climb local hills and reach local optimum fitness, thus getting trapped as they cannot go down

in fitness values to reach the global highest peak (**Figure 4**). Landscape ruggedness⁸³ and steepness of hills⁸⁴ are also predicted to play a role in molecular evolution, particularly under the quasispecies theory and the subsequent ‘survival of the flattest’ model which will be discussed in section 2. Nevertheless, this metaphor has caveats: populations that reach fitness peaks are predicted to enter stasis in rates of adaptation, but long-term experimental evolution on bacteria shows that while these rates slow down, they do not reach zero and better fit a power law dynamic⁸⁵.

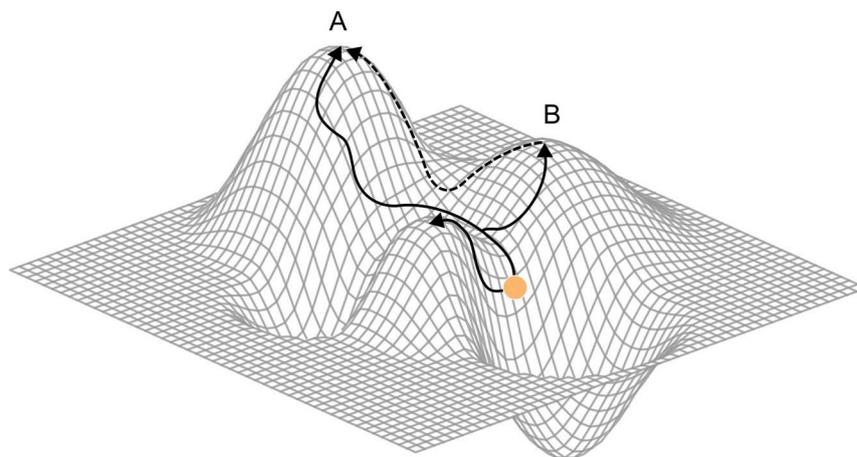


Figure 4 | Adaptive pathways on a fitness landscape. An ancestral genotype can climb fitness peaks by using adaptive routes which increases its fitness along the way. Multiple paths (solid lines) are possible, and the peak it reaches depends on stochastic contingency of evolution. A genotype located on a local sub-optimal peak (B) can no longer reach the highest peak (A) as it requires to decrease its fitness first (dashed line).

1.4. Population sizes and genetic drift: the importance of bottlenecks

In opposition to natural selection, genetic drift is a non-adaptive random process of evolution. It is the stochastic fluctuation of allele frequencies in finite populations and is usually modeled as the result of random sampling of alleles participating in the following generation. Consequently, the importance of drift is inversely proportional to the effective population size N_e – the number of genomes actively participating in the production of the next generation. The effects of genetic drift in small populations can be dramatic: it can eliminate rare beneficial alleles before they reach fixation, and conversely fix detrimental

ones, hindering population mean fitness. For *de novo* neutral mutations, their rate of fixation in asexual populations is estimated to be of the order of μ the mutation rate, and more importantly independent of population size. This substitution rate at neutral sites is often used as a molecular clock to reconstruct phylogenies.

One prediction of genetic drift in asexual populations is the stochastic accumulation of detrimental mutations over cycles of replication, leading to the collapse in mean population fitness values, termed ‘Muller’s ratchet’⁶⁷. This prediction has been verified in RNA viruses through experimental plaque-to-plaque passages^{68,69}. However, whether this phenomenon happens in natural viral populations has yet to be shown and might be dissipated by high recombination rates²⁶ or compensatory mutations⁴⁵.

More importantly, the estimation of effective population sizes N_e of viral populations is fundamental to deciphering which of genetic drift or selection plays the major role in substitution rates in RNA virus evolution. Indeed, population genetics predicts that when $N_e s \gg 1$, selection dominates the dynamic of substitution rates. A naive approach to the question would assume that because viral population sizes are so large during infection, even near-neutral low selective effect mutations can be selected. However, viral populations undergo successive transmission bottlenecks during their life cycles at within-⁸⁶⁻⁸⁸ and inter-host⁸⁷⁻⁹¹ levels, and studies have shown that such bottlenecks affect genetic diversity through founder effects^{91,92}.

These drops in genetic diversity and potential random fixation of detrimental mutations (which we previously stated are the majority) have been associated with decrease in viral population fitness experimentally^{93,94}. Surprisingly, bottlenecks can also have positive effects on fitness⁹⁵: in theory, they can displace populations away from a local fitness optimum in the adaptive landscape, allowing them to explore and potentially reach higher peaks⁸⁷. Bottlenecks can also prevent the emergence of social cheaters by limiting co-infection events^{21,87} and improve selection on *trans*-complementing alleles⁸⁷.

1.5. Genome organization, replication and mode of transmission

While mutation, selection and genetic drift are direct forces acting on allele frequencies, other factors constrain or influence the outcome of these forces.

Generally speaking, RNA viruses have very short genomes with an average size of about 10,000 nt⁹. Therefore, they have very low to no redundancy in gene functions and extremely high proportions of coding sequences. The remaining non-coding regions are often further constrained by bearing RNA structures essential to the virus life cycle. Moreover, some RNA viruses have overlapping open-reading frames (ORFs) usually involving a +1 or -1 frameshift, such as in the *Coronaviridae* family. These highly constrained genomes may explain the majority of deleterious mutations found in RNA viruses: mutations cannot be compensated by redundant genes and may affect multiple functions at once.

Genome organization itself also plays a major role. First, RNA polarity of the packaged genome influence how important RNA secondary structure is: positive-sense single-stranded RNA (+ssRNA) virus genomes must serve as templates for translation, replication and packaging, therefore *cis*-acting RNA elements (CREs) are fundamental for RNA modularity^{96,97}. On the other hand, I have discussed in section 1.2 how -ssRNA virus genomes rapidly form ribonucleoprotein complexes which limit the extent of recombination. Further, segmentation of RNA virus genomes limits the extent of hitch-hiking mutations – neutral or slightly deleterious mutations that reach fixation by being physically linked to positively selected loci – across the genome through reassortment⁴⁵. To an extreme, multipartite viruses – segmented viruses that package their individual molecules in separate virions – allow almost independent evolution of genome segments^{98,99}. Finally, translation strategies and how viruses regulate relative protein abundance constrain genetic regions that serve as signals; some viruses transcribe multiple mRNAs or subgenomic RNAs; others simply cleave the translated polyproteins, and some others use discontinuous transcription, frameshift or read-through stop codons⁴⁵.

In addition, replication mode has been theorized to affect mutation accumulation in the population. Evidently, replication rates, and more specifically the speed of a single infectious cycle from one cell to another, significantly modulates substitution rates. Fast-replicating viruses are predicted to accumulate more mutations over time than latent viruses. More dramatically, for retroviruses in which replication is dominated by clonal expansion of the infected cells they integrated their genome in, substitution rates tend to be closer to their host evolutionary rates¹⁰⁰. Of significance, which RNA molecules can be used as templates for replication in a single cell determines the distribution of the number of mutations per genome. If a single or few molecules are used to produce all progeny, mutations accumulate

linearly following a so-called ‘stamping machine replication’ mode and the number of mutations is expected to fit a Poisson distribution¹⁰¹. In opposition, if all progeny RNAs can serve as templates, mutations accumulate geometrically (‘geometric replication’) and therefore at a faster rate¹⁰¹.

Finally, transmission mode also affects evolutionary rates. Collective dispersal of virions¹⁰²⁻¹⁰⁴ locally increases the multiplicity of infection and therefore co-infection events, which favor recombination. Vector-borne viruses tend to have slower substitution rates due to having to find compromises to adapt to different hosts and environments (this will be discussed in detail in the introduction of Chapter 2). And even surprisingly, host alternation can promote genome integrity in Rift Valley fever virus¹⁰⁵.

1.6. Epistasis and pleiotropy

As stated previously, mutations do not possess fitness effects (or even phenotypic effects) *per se*, rather their fitness is dependent on their genetic background and the environment. To better comprehend the effect of genetic background on a mutation at a given locus, population genetics uses the concept of ‘epistasis’ which can be defined as the deviation from purely additive phenotypic effects of mutations at two or more loci. What that means is that epistasis is the result of the interaction of mutations at different loci. We can identify different types of epistatic interactions: magnitude epistasis and sign epistasis (**Figure 5A**).

Magnitude epistasis is defined when two mutations, both either beneficial or detrimental, interact in such a way that their absolute combined fitness effects is stronger (synergistic epistasis) or fainter (antagonistic epistasis) than the expected addition of the two. Theoretically, synergistic epistasis can happen through positive feedback loops when mutations affect the same metabolic pathway. However, studies in RNA viruses have shown that the widespread deleterious mutations interact mostly in an antagonistic (positive) way¹⁰⁶⁻¹⁰⁸; they tend to compensate each other. These compensatory mutations are thought to be mainly linked to protein¹⁰⁹ and RNA secondary structures¹¹⁰. In the latter case, this can explain the positive selection patterns of non-coding or non-synonymous mutations⁸¹ observed and argues against selective neutrality of these sites¹¹⁰. Analysis of interactions between beneficial mutations also shows a trend for antagonistic (negative) epistasis¹⁰⁸. This tendency for mostly

antagonistic epistasis in RNA viruses can be explained by their small, lack of redundancy genomes: mutations will often damage the same function^{111,112}.

Sign epistasis on the other hand happens when the deleterious or beneficial effect of a mutation is reversed in presence of another mutation. To an extreme, reciprocal sign epistasis occurs when two beneficial or deleterious mutations combined result in the opposite fitness effect. These interactions are of particular interest as they can lead viruses into fitness traps: one beneficial mutation could cause other otherwise beneficial mutations to become deleterious, blocking further adaptation¹¹³. Conversely, stabilizing mutations can permit the fixation of otherwise destabilizing mutations with advantageous phenotypes¹⁰⁹. More generally, sign epistasis highly constrains the trajectory of molecular evolution, and has been linked to the divergence of Asian and Indian Ocean lineages of chikungunya virus in their capacity to adapt to the *Aedes albopictus* mosquito species^{113,114}. Widespread sign epistasis is predicted to result in more rugged fitness landscapes¹¹⁵, increasing the role of contingency in evolution.

While the study of epistasis is interested in the interaction of two or more mutations on one phenotype, pleiotropy is the phenomenon of one mutation affecting different functions at once. Consequently, trade-offs are the result of pleiotropy.

First, pleiotropy can be considered at the genome level: because viral proteins usually have multiple functions, or change functions after post-translational modifications, single mutations are expected to affect different phenotypes at once. Moreover, for RNA viruses with overlapping ORFs, mutations affect different proteins at once. Interestingly, because they are usually the result of +1 or -1 frameshifts, synonymous mutations in one ORF usually results in a non-synonymous change in the other one, highly constraining the space of non-detrimental mutations. Additionally, RNA secondary structure and presence of CREs in coding regions means that structural stability of mutations must also be taken into consideration in addition to residue change in resulting proteins.

Alternatively, pleiotropy also applies to the difference in fitness effects of mutations in different environments. Similar to epistasis, pleiotropy can affect the strength of one mutational effect across different environments (magnitude pleiotropy) or the detrimental or beneficial nature of said mutation (sign or antagonistic pleiotropy) (**Figure 5B**). Environmental pleiotropic effects of mutations have been widely studied in RNA viruses for

adaptation to different temperatures^{116,117}, and in the case of vector-borne viruses for the adaptation to multiple hosts. Because antagonistic pleiotropy tends to be widespread in the latter case, its study is of significant importance to understand the evolution of generalist organisms and will be the topic of Chapter 2.

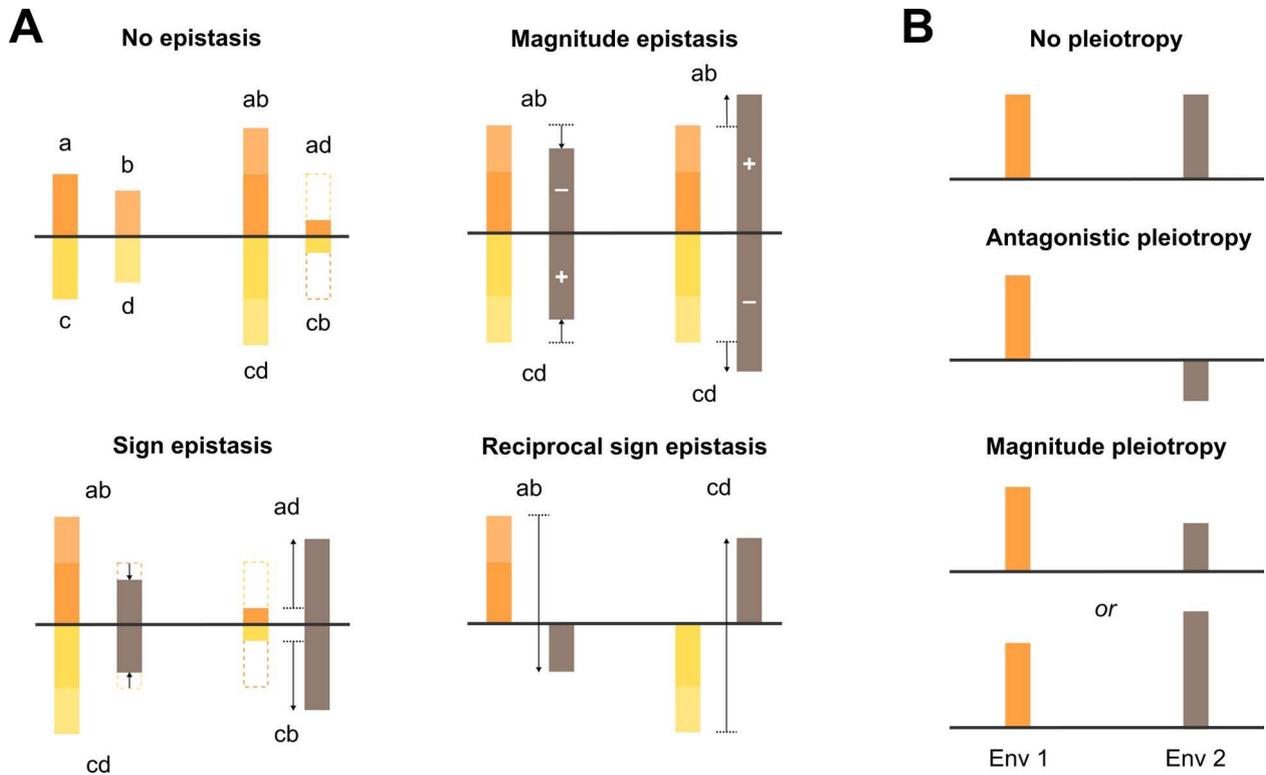


Figure 5 | Different types of epistasis and environmental pleiotropy. (A) Epistatic interactions result in fitness deviations from the expected linear sum of gene fitness effects. Stacked color bars correspond to the expected fitness effects of double-mutants without epistasis, while brown ones are the observed fitness effects. The left part of magnitude epistasis corresponds to antagonistic epistasis, the right part to synergistic epistasis. Plus (+) and minus (-) symbols indicate what are typically called respectively ‘positive’ and ‘negative’ epistasis depending on the initial fitness effects of mutants of interest. (B) Environmental pleiotropy corresponds to differences in fitness effect of the same mutation in different environments. Inspired by Bergh *et al.*, 2018¹¹⁸.

1.7. Mutational robustness

Because RNA viruses have elevated mutation rates and a majority of them are deleterious or lethal, they have developed different mechanisms to buffer the mutational

burden. This buffering of mutation effects, termed ‘mutational robustness’ occurs at the genomic, population and cellular levels.

Genetic mutational robustness often takes the form of codon optimization: while different codons can give rise to the same amino acid, they differ in their position in the fitness landscape, and consequently in their accessible mutational neighbourhood^{119,120}. For example, the use of codons that have higher chances of resulting in synonymous or similar amino acid changes through random mutations is more robust. Consequently, codon-deoptimized viruses have been shown to have reduced fitness and pathogenicity^{121,122} and can be used to make safer live-attenuated vaccines^{122,123}. These studies could be criticized because the cost of deoptimizing codons can be confounded with the effects of discrepancies in codon bias between the virus and its host^{122,124}; however mutational robustness has been elegantly demonstrated in coxsackie B3¹²⁵, influenza A¹²⁵ and chikungunya virus¹²³ by attenuating them through codon substitution with synonymous ones closer to non-sense mutations. Interestingly, mutational robustness through codon optimization reveals a property of the geometry of the fitness landscape: at high mutation rates, genomes that replicate the fastest (the fittest) do not necessarily outcompete more robust slower replicator^{84,119,126}. This can happen because the mean expected fitness of one progeny genome does not depend solely on the fitness of its progenitor, but is the integration of fitness values of all possible neighboring mutations. This phenomenon is thus termed ‘survival of the flattest’ in reference to the gentler gradient in fitness effects surrounding local peaks. Survival of the flattest has first been demonstrated in digital organisms⁸⁴, and *in vitro* evidence has shown that in artificially elevated mutation rate conditions, more robust and less fit populations outcompete otherwise fitter ones^{126,127}. Nevertheless, evidence of such dynamics at natural mutation rates has yet to be found⁹.

At the population level, robustness can occur first through population size. Because RNA virus populations are generally large, unmutated progeny numbers can be sufficiently large to survive and propagate to the following replication cycle. When combined with large amounts of lethal and highly deleterious mutations, population size allows the quick and efficient purge of unwanted alleles. This property has been coined ‘population robustness’¹²⁸. Of note, this form of robustness relies on individual genome hypersensitivity¹²⁹, therefore opposing itself against genetic robustness. Additionally, population-level robustness can emerge from networks of complementation between mutated genomes. Because random

mutations are expected to affect different functions in progeny, *trans*-complementation can buffer the effect of detrimental mutations through shared viral resources. This mechanism can extend itself to include co-infection events at high multiplicity of infection and has been demonstrated in RNA phage $\Phi 6$ ¹³⁰. Interestingly, in the latter study, Montville *et al.* showed that high levels of co-infection relax the need for genetic mutational robustness¹³⁰.

Finally, extrinsic factors also provide robustness to mutational changes. Molecular chaperones of the heat shock protein (Hsp) family are expressed during cellular stress to prevent misfolding of cellular proteins during biogenesis or facilitate the degradation of misfolded ones. As such, they can preserve protein functionality regardless of destabilizing detrimental mutations. Hsp proteins are known to often interact with viral proteins¹²⁸ for trafficking or assembly¹³¹, and play essential roles in viral entry^{132,133} and RNA synthesis¹³⁴ of some viruses. Moreover, over-expression of chaperones is common during viral infections¹³⁵, leading to the hypothesis that they can also buffer detrimental mutations in viral genomes. Indeed, inhibition of Hsp90 cellular expression affects poliovirus capsid variants' fitness and influences synonymous mutant distribution¹³⁶.

Because mutational robustness buffers the effects of mutations, it hereby reduces the heritability of traits. A naive approach would lead to the conclusion that evolvability – the capacity of a system for adaptive evolution through the selection of heritable traits – and robustness are antagonistic. However, contradictory results rather show increased evolvability in robust systems¹³⁷. This apparent contradiction can be resolved because robust populations can explore a wider neutral space of mutations, therefore having a wider panel of accessible alleles¹³⁸. In addition, robust populations with cryptic genetic diversity in one environment can show potentially higher phenotypic diversity through environmental pleiotropy¹³⁸. However, demonstrations of such phenomena have yet to be described in RNA viruses.

2. RNA virus populations as a unit of selection: from quasispecies to social evolution theories

Quasispecies theory was first formulated by Manfred Eigen in 1971 to describe and predict the evolutionary dynamics of self-replicating macromolecules (presumably RNAs) at the origins of life¹³⁹, although the term *quasispecies* itself was only coined in 1977¹⁴⁰. The theory was developed to study replicating entities at extremely high error rates, which was argued could not be done using classical population genetics as it was developed under the assumption of low mutation rates. At its core, it describes the dynamic of such populations as a distribution of mutant macromolecules closely linked by mutational networks and dominated by one or several master copies¹⁴⁰. Because these molecules are genetically coupled, frequencies of variant genomes do not depend on their own replication rates or fitness, but also on their production by erroneous replications of linked genomes. As a consequence, quasispecies theory defines the unit on which selection operates as the mutant distribution, and not individual genomes¹⁴¹.

After the findings of the high error rates of RNA bacteriophage Q β by Domingo *et al.* in 1978¹³, quasispecies theory was broadly adopted by virologists to describe and explain experimental evolution results, and served as the main theoretical framework for several decades¹⁷⁻¹⁹. Of note, it guided successful antiviral strategies such as lethal mutagenesis by nucleotide analog treatments^{142,143}, and increased interest in evolutionary concepts in virology⁹. Nevertheless, many authors criticized the relevance of quasispecies theory to RNA virus (and even DNA virus) evolution, arguing mainly that mutation rates were not sufficiently high to observe such mutation coupling dynamics^{9,20}. In recent years, alternative theories that also predict that selection acts on groups of viral genomes have gained popularity among virologists^{21,22}. Among them, social evolution theories rely on kin and/or group selection and is compatible with population genetics (although not without its own controversies). Additionally, it does not rely on the extreme mutation rate assumptions of quasispecies theory.

In this section, I briefly present quasispecies theory and its main concepts and predictions, and confront them with the main arguments against its application to RNA viruses. I then proceed to review different types of collective behaviors observed in RNA viruses and argue that a population-wide approach is still necessary to comprehend viral evolution, then present social evolution theories. Finally, I argue that defective viral genomes

and other non-standard genomes should be considered as part of viral populations, therefore contributing to their evolution.

2.1. Quasispecies theory: a controversial debate among virologists

2.1.1. DEFINITION OF QUASISPECIES TERMINOLOGY

A substantial amount of criticism towards quasispecies theory targets the misuse of its terminology in virology literature. Indeed, many authors use the term ‘quasispecies’ solely to refer to genetic polymorphism, or as a synonym for viral populations, without any consideration for its underlying mathematical model^{9,141}. As stated earlier, a quasispecies is a specific network of mutants bound under mutation coupling dynamics. Here, the terminology (molecular) *species* refer to a single replicative unit¹⁴⁰ and has no link with the naturalist species concept. Because one species abundance relies not only on its own replication but also on that of other mutationally linked ones, the network behaves as (*quasi-*) a single replicative unit (*-species*), and evolutionary forces act on this second-order *species* rather than on individual molecules. In the following sub-section, I will introduce the fundamental equations in quasispecies theory and the popular concepts associated with it.

First, quasispecies theory defines the rate of frequency changes \dot{x}_i of the mutant i as:

Equation 2

$$\dot{x}_i(t) = (A_i Q_{ii} - D_i) x_i(t) + \sum_{k=1, k \neq i}^n w_{ik} x_k(t) + \phi_i,$$

with A_i the replication rate, Q_{ii} the fraction of non-erroneous replication from sequence i to i , D_i the rate of degradation of i , w_{ik} the rate of production of sequence i from sequence k such as $w_{ik} = A_k Q_{ik}$, and ϕ_i the flow of removal of sequence i outside of chemical reactions¹⁴⁰. To note, A_i integrates the concept of fitness (the rate of reproduction), while Q_{ii} is related to the mutation rate μ such as $Q_{ii} = 1 - \mu$. The term $W_{ii} = A_i Q_{ii} - D_i$, called the ‘intrinsic selective value’ by Eigen¹⁴⁰, therefore represents a mutation-selection balance coefficient¹⁴⁴ which tends to resemble population genetics relative fitness w when mutation rates μ are close to zero¹⁴⁴. The quasispecies is then defined as the linear combination of x_i , and the fittest distribution corresponds to the quasispecies composition defined by the eigenvector associated with the maximum eigenvalue of the matrix W :

Equation 3

$$W = \begin{bmatrix} W_{11} & w_{21} & \cdots & w_{k1} \\ w_{12} & W_{22} & \cdots & w_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1k} & w_{2k} & \cdots & W_{kk} \end{bmatrix}$$

From the previous formula, one can then define the ‘master sequence’ as the one with the highest intrinsic selective value W_{ii} , which dominates the mutant distribution (sometimes called mutant ‘swarm’, ‘cloud’ or ‘spectrum’) and is critical for the maintenance of the information in the distribution¹⁴⁰. Importantly, the master sequence is not necessarily the most abundant in the distribution: this is rather determined by the values of the eigenvector, and its sequence does not necessarily match the consensus sequence – the sequence composed of the most dominant nucleotide at each position in the population. Even more surprisingly, the consensus sequence may not be represented at all in the viral swarm. This latter consequence is of particular importance, as a substantial part of virology literature bases its conclusions on the consensus sequence and admits it as the ‘wild-type’ (**Figure 6**).

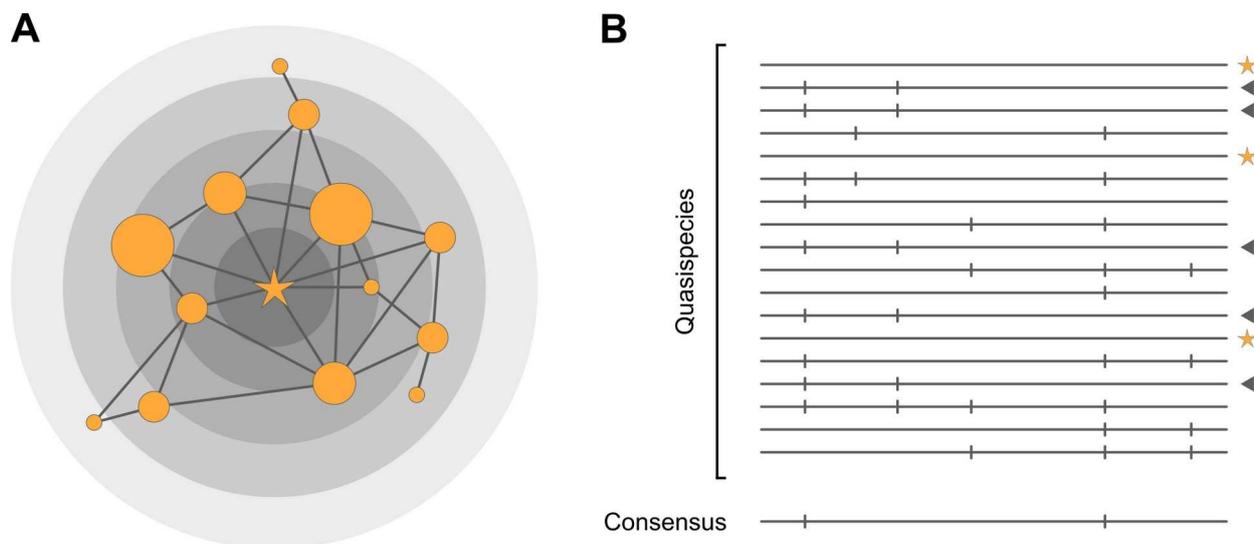


Figure 6 | Simplified representation of a quasispecies. (A) Mutant network of the quasispecies on a fitness landscape (gray circle gradient). The master sequence (star) is the fittest of the distribution. Circle sizes represent variant abundances. Solid lines correspond to mutational coupling between variants. (B) Composition of the quasispecies. The master sequence is not necessarily the most abundant sequence, which is indicated by the gray arrowhead. The consensus sequence corresponds to the ‘average’ sequence of the population and is not necessarily represented in the quasispecies.

2.1.2. THE ERROR THRESHOLD

A second aspect of quasispecies theory focuses more on the stability of the quasispecies itself. Briefly, because the master sequence is assumed to dominate the distribution and hold its information content, its disappearance through excessive error would lead to the collapse of the quasispecies itself, resulting in a random distribution of mutations (no information). This predicts an ‘error threshold’ above which the information is lost. It is defined when the average error rate p_{max}^- is

$$\text{Equation 4} \quad p_{max}^- = \frac{\ln(\sigma_m)}{v_m} \quad \text{with} \quad \sigma_m = \frac{A_m}{D_m + E_{k \neq m}^-}.$$

σ_m is the superiority parameter of the master sequence m , v_m the amount of information in the master sequence (which is proportional to its length), and $E_{k \neq m}^-$ the average productivity of all competitors k such as $E_{k \neq m}^- = \sum (A_k - D_k) x_k / \sum x_k$. Consequently, the more information v_m in a sequence (the longer it is), the lower the error threshold is at similar superiority values of the master sequence. Moreover, the master sequence must maintain a superior replication rate A_m (independent of its accurate replication) which compensates for its degradation D_m and the overall production of minor sequences $E_{k \neq m}^-$. Of note, as pointed out by Bull *et al.*, the error threshold of one quasispecies only signifies that the master sequence holding its structure gets lost, but another master sequence of lower selective value W_{ii} can replace it and generate a new distribution¹⁴⁵. Only when no other sequence can sustain its information content does the population reach the ‘extinction threshold’¹⁴⁵ where the viral population goes extinct.

The error threshold is a fundamental prediction of quasispecies theory and has been at the origin of the concept of lethal mutagenesis as an antiviral strategy^{9,143}. The major idea coming from studies around viral polymerase fidelity is that viral populations tend to have mutation rates situated just below the error threshold to maximize genomic diversity and thus adaptability^{30,32-36}. As such, using mutagens such as 5-fluorouracil or ribavirin can push the mutation rate passed the threshold and lead the population to an ‘error catastrophe’. The effects of lethal mutagenesis have shown successful attenuation of viruses *in vitro*¹⁴⁶⁻¹⁴⁹ and *in vivo*¹⁵⁰. However, whether this attenuation is due to a true ‘error catastrophe’ is still debated⁹. I will discuss in section 2.1.4 alternative explanations for the observation of lethal mutagenesis.

2.1.3. SURVIVAL OF THE FLATTEST QUASISPECIES

Another prediction of quasispecies theory is the ‘survival of the flattest’ which was introduced in section 1.7. Particularly for quasispecies, the evolutionary dynamic is not linked to genetic mutational robustness of single genomes, rather it is a property of the mutant network as a whole acting as a single unit. Because the distribution spans across a surface of sequence space, the topology of the landscape plays an important role: the fitness of the quasispecies is determined as the linear combination of the fitness of each mutant. As a result, highly diverse and spread-out distributions on lower but flatter peaks tend to have better fitness overall than ones situated at higher sharper peaks. In contrast to mutational robustness at the genome level, where more robust genomes outcompete lesser ones at elevated mutation rates, and is the result of contingent random mutations, this survival of the flattest dynamic in quasispecies is completely deterministic as the population is already composed of a mutant spectrum (**Figure 7**).

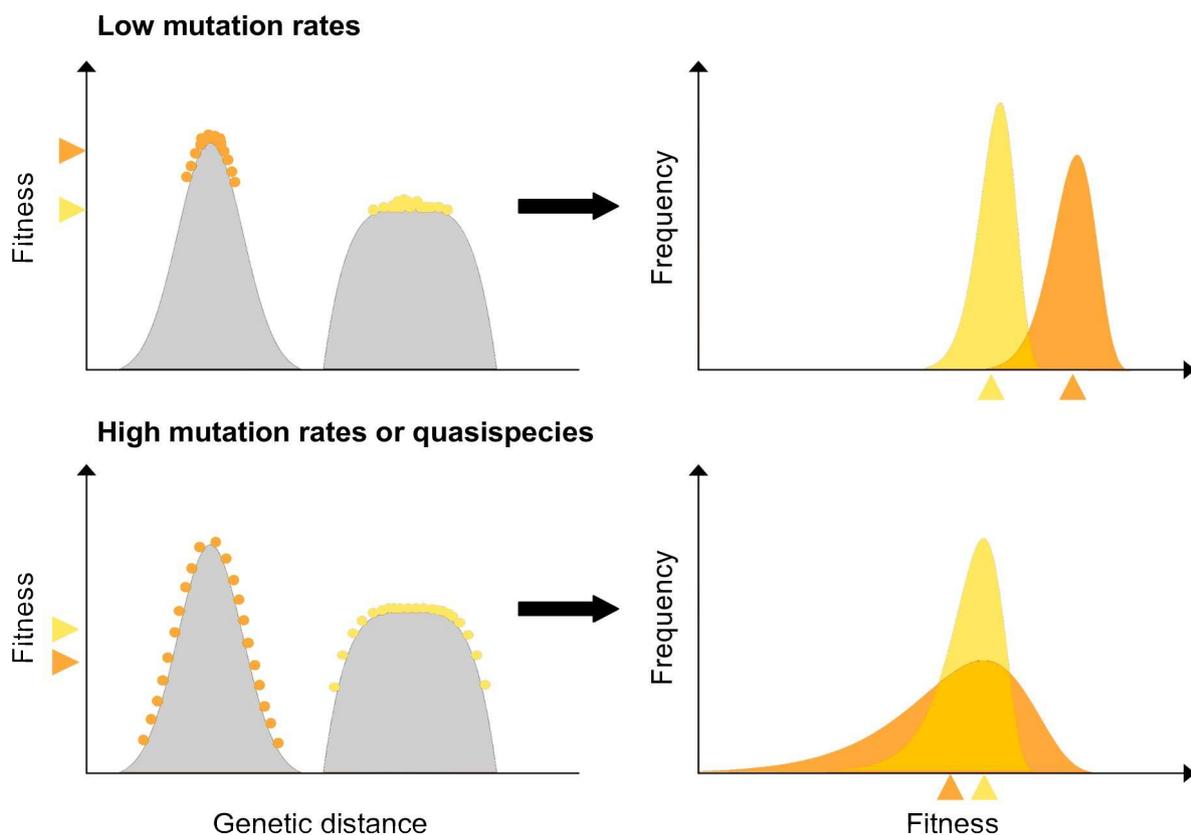


Figure 7 | Survival of the flattest at high mutation rates or in quasispecies. At low mutation rates the fittest variant (orange) produces on average a fitter progeny than the more robust but less fit variant (yellow). At high mutation rates, progeny genomes accumulate more mutations and are more

genetically distant than their progenitor. Because the more robust variant lies on a flatter peak, fitness values decrease less dramatically with genetic distance, resulting on average in a fitter progeny than the fittest variant. In quasispecies, the respective mutant networks already span the high mutation rate distribution, and the average fitness correspond to the fitness of their populations. Arrows indicate average fitness of progeny (mutational robustness) or mutant networks (quasispecies). Adapted from Sanjuán *et al.*, 2007¹²⁷.

2.1.4. LIMITATIONS AND CONTROVERSIES OF QUASISPECIES THEORY

As stated before, the main controversy around quasispecies theory is its applicability to RNA virus evolution. Indeed, many authors argue that mutation rates in RNA viruses, and even more so in DNA viruses, are too low to allow mutational coupling between mutants^{9,20}. Indeed, validation of quasispecies dynamics *in silico* with digital organisms by Wilke *et al.* has shown that survival of the flattest only emerges at mutation rates ranging from 1.13-3.5 mutations per genome per replication cycle (mut/genome/rep)⁸⁴, or for Comas *et al.* above 2 mut/genome/rep¹⁵¹. This is typically higher than the 1 mut/genome/rep usually associated with RNA viruses. In opposition, evidences of survival of the flattest *in vitro*^{126,127} push other authors to affirm that these tested viruses do indeed follow a quasispecies dynamic, although it is only observed at artificially elevated mutation rates.

Moreover, other critics target the use of quasispecies theory as an alternative to classical population genetics and argue that all current observations are more parsimoniously explained by an already robust and validated theoretical framework. For example, I developed in section 1.7 how survival of the flattest is also predicted in population genetics while only involving genetic mutational robustness and the majority of deleterious mutations in the DFE, without the need for mutation coupling. Additionally, the underlying mechanism of lethal mutagenesis is not necessarily the error catastrophe theorized under quasispecies theory. Indeed, the error catastrophe only predicts the loss of the master sequence but not necessarily the collapse in fitness of the population. First, mutagen treatments can lead to the selection of high-fidelity and mutagen-resistant variants^{29,149,152,153} or more robust genomes¹²⁷, thus actually increasing their fitness value in the fitness landscape under mutagenic conditions. Further, loss of viral infectivity can still happen while the viral population continues to replicate because of the action of defective genomes¹⁴⁸. Finally, the approach of viral extinction through loss of sequence information is also too stringent: mutational robustness allows the maintenance of phenotype despite the deformation of the initial information. Rather, lethal

mutagenesis can simply be explained by the increase in the expected number of lethal or deleterious mutations per replicated genomes. This last approach allows for a more relaxed definition of a different lethal mutation threshold that could be typically lower than the error threshold predicted by quasispecies theory.

In addition, other phenomena have been associated with quasispecies theory, although they are not directly predicted by it. Among them, the concept of ‘memory genomes’ in quasispecies refers to genomes that were once dominant (master sequences) in the past evolutionary history of the population, but declined and remained as minority variants through constant replenishing by mutational coupling. In theory, they can accelerate the rate of re-adaptation to former selective pressures through simple selection forces and become dominant again. The existence of such memory genomes has been reported in viral systems *in vivo*^{150,154}, and sheds light on the importance of minority variants and the historical contingency of viral evolution⁹. Nevertheless, the maintenance of sequences at minority levels is not specific to quasispecies and can be achieved if these sequences do not accumulate deleterious mutations over time and are lost through genetic drift.

In conclusion, I argue that whether or not quasispecies dynamics are possible in RNA virus populations, they do not provide better evolutionary explanations to experimentally or naturally observed viral evolution patterns. Nevertheless, quasispecies theory popularized among virologists the idea that the unit of selection is not necessarily the individual genome, but could be analyzed at a second-order level. In the next sub-sections, I will illustrate how viral genomes interact in complex manners to replicate, then argue for another type of group selection theory.

2.2. Collective behaviors of viral populations

2.2.1. COMPLEMENTATION AND MINORITY VARIANTS

One form of interaction between different viral genomes is complementation: one genome produces proteins that can then help in *trans* other genomes in their infectious cycle. Most of the examples of complementation described in the literature show how wild-type genomes can rescue functions of other genomes that have depleted functions¹⁵⁵⁻¹⁵⁷. However, the extent of complementation does not stop at helping otherwise defective genomes, but can also happen between strains containing beneficial mutations. In their paper, Xue *et al.* show

how two influenza A virus variants at residue 151 in the neuraminidase (NA) protein are capable of cooperating to increase viral yields superior to pure populations of one or the other¹⁵⁸. The proposed mechanism is that one of the variants (G151) is more proficient at cell entry while the other (D151) improves cell exit, and assembly would result in chimeric virions containing both NA protein variants, regardless of the genome variant packaged. This complementation between strains leads to stable mixed populations after multiple passages, and even becomes obligate when the hemagglutinin protein lacks receptor-binding activity¹⁵⁸.

Interestingly, complementation between genomes is not completely random, and some genomes are required to produce some proteins in *cis* to be able to be complemented in *trans* for other functions. In their paper, Rezelj *et al.* show that a defective viral genome (DVG) of Zika virus can only be complemented by a helper full-length genome if the DVG retains its ORF without frameshifts²³. More generally, it appears possible that not all proteins can be complemented in *trans*, and spatial distribution of replication units and viral factories in a single cell could affect protein flow. For example, structural proteins are thought to be typically more easily shared, to the extreme that envelope glycoproteins from other virus species can be used for delivery of genetic material of other species such as in virus pseudotypes¹⁵⁹. Further research should be conducted to resolve a complete map of possible complementation between viral genomes.

In segmented viruses, complementation is almost essential for successful infection cycles: Brooke *et al.* showed that at low multiplicity of infection with influenza A virus, up to 90% of infected cells failed to express at least one essential viral protein, leading to mostly abortive infections¹⁶⁰. They conclude that most virions are semi-infectious and complementation-dependent. To a certain extreme, this requirement for multiple infections to recapitulate a fully functional infectious cycle is at the core of multipartite viruses' life cycle.

In non-segmented viruses, such requirements for complementation do not seem to be the rule¹⁶⁰, but effects on population fitness and intra-host dissemination are nonetheless important. Ciota *et al.* demonstrated that fitness values of diverse West Nile virus populations are significantly higher than the mean or median fitness of individual clones from the same populations, and that this discrepancy increased with multiplicity of infection¹⁶¹. Even more striking, diverse population fitness can only be recapitulated when reconstructed mutant haplotypes are mixed with some at frequencies of only 10% in the population¹⁶². These effects of so-called 'minority variants' have emphasized viral population diversity as a main driver of

pathogenicity^{32,92,163}, which has been linked for example to neurotropism in poliovirus¹⁶³. However, genetic diversity by itself – the existence of beneficial haplotypes that can be positively selected – obscures the role of minority variant complementation found by Bordería *et al.*: some minority genomes can exert dominant phenotypes at very low frequencies, which could explain balanced selection of variants at intermediate frequencies in experimental evolution set-ups¹⁶².

An interesting property of complementation is that it alleviates epistatic interactions in the same genome, allowing interactions of beneficial mutations at the protein level, for example during oligomer assembly^{158,164,165}, while avoiding antagonistic negative epistasis. Moreover, as shown in the paper from Xue *et al.*, cooperation can happen between different alleles at the same locus¹⁵⁸, which could otherwise only be possible through gene duplication without complementation. Of note, cooperation between co-infecting genomes can produce new phenotypes^{158,165}.

2.2.2. CHEATER GENOMES

A direct consequence of complementation is the emergence of ‘cheater genomes’. Indeed, because some viral resources are shared, genomes can evolve to participate unequally in costly protein production while still benefiting from these common resources. Most cheater genomes described are defective viral genomes (DVGs): genomes that are incapable of completing a full replication cycle on their own and require ‘helper’ genomes to do so¹⁶⁶. DVGs are therefore complementation-dependent. However, some cheaters do not require complementation and can function on their own, albeit when they co-infect cells with wild-type genomes, they outcompete the latter ones by better exploiting common resources or producing more proteins that function preferentially for their progenitor genomes¹⁶⁷. Nevertheless, co-infections and complementation weaken purifying selection of deleterious mutations, therefore promoting the accumulation and maintenance of ‘inferior’ cheater genomes¹⁶⁸.

2.2.3. COLLECTIVE INFECTIONS

A determining factor for complementation and the emergence of cheater genomes is the probability of co-infections. The simplest way to achieve co-infection of cells is through

elevated multiplicities of infection (MOI): the distribution of the number of infectious units infecting a single cell follows a Poisson distribution. However, evidence of several forms of collective infectious units (CIUs) that locally increases the MOI have been reported and seems to represent a substantial proportion of infection events¹⁰³. Consequently, in typical titration by plaque assays, single plaques initially thought to represent single infectious genomes are often in reality ‘chimeric plaques’ composed of different variants^{103,169,170}. These CIUs can be formed from clustered single virions, or readily released as multi-virion structures¹⁰⁴. In addition, multiple genomes can be packaged in a single virion, resulting in polyploid viral particles^{171–173}. The ability for virions to co-package different genomes depends on encapsidation selectivity¹⁷⁴ and virion structure modularity¹⁷³.

Clustering of individual virions through virion aggregates can happen spontaneously depending on pH and saline properties of their containing fluids^{169,175}. This phenomenon is probably the underlying mechanism of virion aggregates in saliva^{102,176}. Additionally, interactions with extracellular matrix compounds can also promote aggregates, such as amyloid fibrils found in semen which interact with human immunodeficiency virus 1 (HIV-1) virions¹⁷⁷. Other forms of clustering also occur through concentration at the surface of ‘third-party cells’: HIV-1 particles are captured and accumulated in membrane invaginations of dendritic cells before being transmitted to target CD4+ lymphocytes¹⁷⁸. For enteric viruses, the accumulation of virions at the surface of commensal bacteria facilitates co-infections of host cells¹⁷⁹.

Moreover, viral particles can be delivered readily as multi-virion structures. Human T-cell leukemia viruses 1 were described to aggregate after egress in virus-induced extracellular matrix networks at infected cell surfaces, which subsequently rapidly adhere to other cells upon contact¹⁸⁰. This type of structure has been compared to bacterial biofilms¹⁸⁰. In addition, membrane-mediated structures have also been described. These extracellular vesicles originate either from the exosomal secretory pathway^{181–183} or autophagy bodies^{183–186} and enable *en bloc* infections without cell-to-cell contacts.

Finally, direct co-infections through cell-to-cell spread are common for many viruses. These can involve preexisting junctions such as plant plasmodesmata¹⁸⁷ or neural synapses¹⁸⁸, or virus-induced structures such as filopodial bridges¹⁸⁹. More dramatically, the formation of syncytia through membrane fusion of infected and non-infected cells^{190,191} leads to the delivery of the entire intracellular viral population to the new targeted cells.

Altogether, these mechanisms are thought to increase bottleneck sizes^{103,104}, especially at inter-host transmission events^{176,177}, therefore increasing initial genetic diversity. Moreover, co-dispersal structures can increase virion stability¹⁹² and protect from neutralizing antibodies¹⁸⁰. Complete avoidance of circulating antibodies can also be achieved by cell-to-cell spread. Experimental infectivity assessments of virion clusters against independent ones usually show more potent infections when viral particles disperse in CIUs^{180,182,183}, suggesting that these mechanisms might have been the result of positive selection. However, no genetic determinant for these mechanisms has been identified in viral genomes, shedding doubts on their heritability and evolvability *per se*¹⁰⁴. Finally, mechanisms that prevent co-infection events also exist. Among them, superinfection exclusion blocks the secondary infection of already infected cells at either the entry¹⁹³ or genome replication^{194,195} step. This phenomenon can dramatically modify cell-to-cell transmission events¹⁹⁶ and induce spatial segregation of virus populations *in situ*^{196,197}.

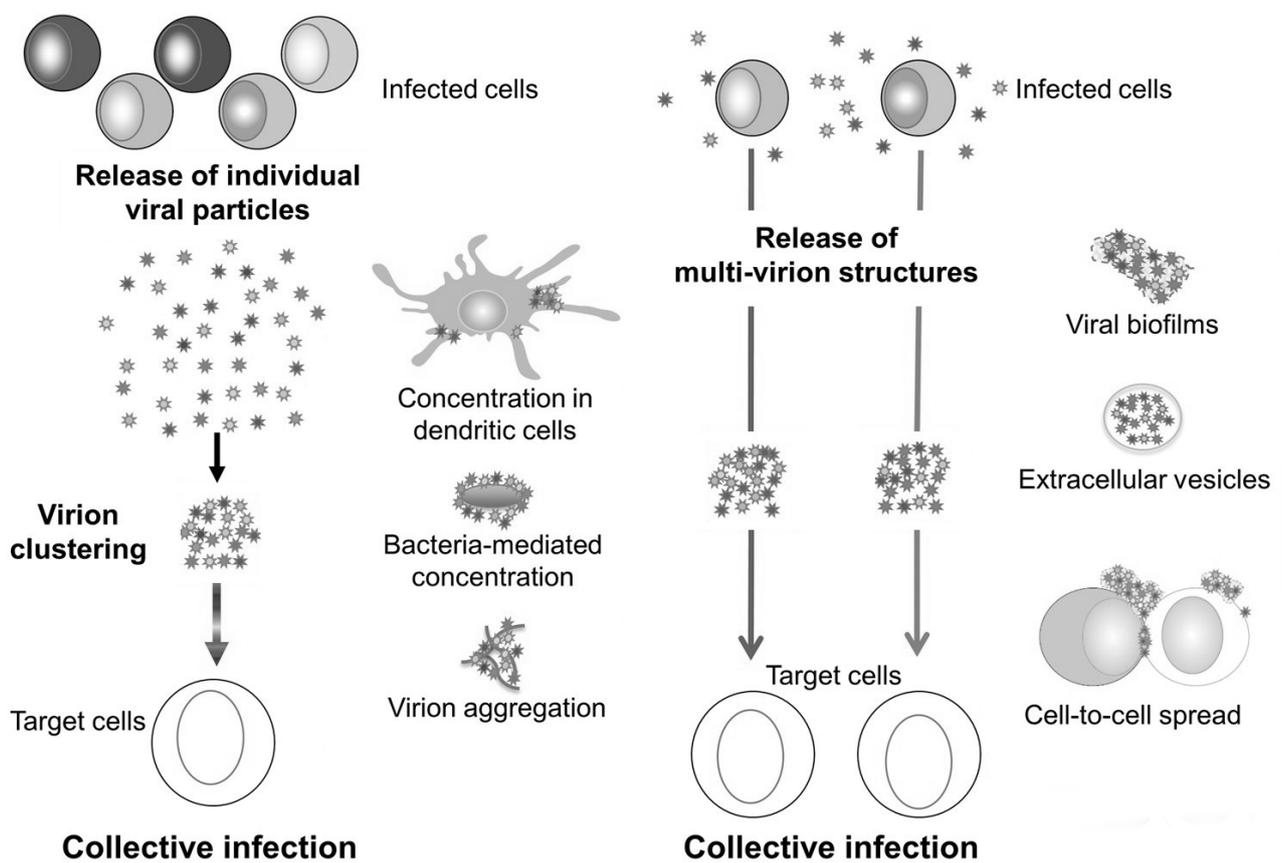


Figure 8 | Types of collective infectious units. Adapted from Sanjuán and Thoulouze, 2019¹⁰⁴.

2.2.4. STRENGTH IN NUMBERS: COLLABORATION THROUGH 'MASS EFFECT'

I have discussed how collective infections and complementation networks draw an interdependent picture of viral genomes, where selection cannot act solely on individual viral variants because of their interactions. In addition to this view, viral population size itself has emerging collaborating properties.

First, co-infection events can lead to cooperative replication, where early synthesis of viral proteins from multiple templates accelerates the initial speed at which progeny are produced. This is due to the positive feedback of viral genomes producing their own polymerase which then replicate their numbers. This cooperative process has been validated experimentally in different mammalian viruses¹⁹⁸.

Additionally, this 'mass effect' in single cells also allows cooperative suppression of the host immune response. Because host restriction factors are limited, early higher numbers of viral genomes and restriction factor antagonists can overwhelm the antiviral response, leading to reduced abortive infections¹⁹⁹⁻²⁰¹. This cooperative antagonism of antiviral immunity is well described in bacteriophages that must overcome the CRISPR-Cas defense system of host bacteria. They evolved anti-CRISPR (Acr) proteins that antagonize this immune response, however, the amount of Acr proteins delivered by a single phage is often insufficient, leading to a critical population number threshold under which infection fails^{200,201}. More surprisingly, previous abortive infections initiated by insufficient numbers of phages still temporarily hinder the CRISPR-Cas system of infected bacteria, therefore predisposing them to successful infections by other phages^{201,202}. In vertebrates, successful antagonism of the interferon signaling pathway not only helps other viral particles to infect the already infected cells, but also prevents neighboring uninfected cells from entering into an antiviral state^{21,203}.

2.2.5. VIRAL COMMUNICATION

The discovery of the arbitrium communication system regulating the lysis-lysogeny cycle of bacteriophages must be one of the most fascinating forms of collective behavior yet identified in viral populations. Briefly, the system consists in the production and sensing of a small peptide AimP by infected bacteria: AimP is secreted and internalized by host bacteria, where it then downregulates the transcription of an inhibitor of lysogeny AimX by binding to a phage-encoded receptor AimR, which upregulates AimX expression in its unbound form²⁰⁴.

The concentration of AimP in the medium therefore reflects the density of already infected bacteria. When AimP concentrations are low, lysogeny is inhibited by expressed AimX and infected bacteria undergo lysis which releases new virions during burst. Conversely, when AimP concentrations are high, AimR cannot activate the expression of AimX and lysogeny is promoted, allowing prudent exploitation of hosts when their uninfected concentration is expected to be low²⁰². Of note, while this communication system highly resembles the quorum-sensing system of the bacteria they infect, AimP, AimX and AimR are all produced from the phage genome itself. Therefore the arbitrium system is not simply a hijacking of their host communication system.

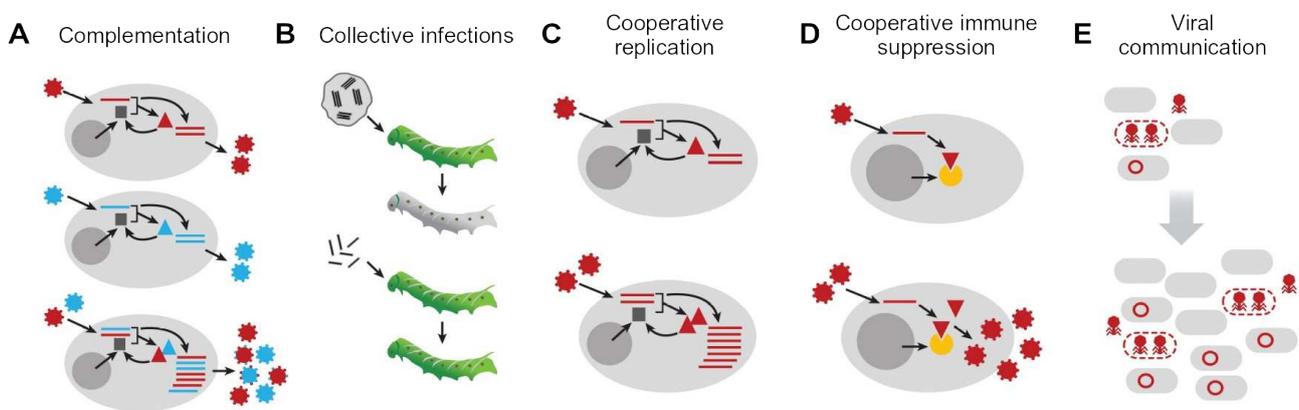


Figure 9 | Types of viral collective behaviors. (A) Complementation. Synergistic interactions between co-infecting variants can improve total population fitness and lead to new phenotypes (e.g. oligomer assembly). Complementation is sensible to the emergence of cheater viruses. (B) Collective infections. Virus genomes often disperse in groups, increasing the multiplicity of infection locally. (C) Cooperative replication. Positive feedbacks between the initial number of genome templates and polymerases being synthesized result in a short term replication advantage of co-infections over single infections. (D) Cooperative immune suppression. High multiplicity of infection can limit the number of abortive infections through higher concentrations of restriction factor antagonists. (E) Viral communication. The bacteriophage arbitrium system promotes prudent host exploitation by inhibiting lysis when available uninfected bacteria numbers are low. Adapted from Sanjuán, 2021²¹.

2.3. Social evolution theories: alternatives to quasispecies to account for virus-virus interactions

In the previous sub-section, I depicted RNA virus populations as complex pools of interacting genomes which sometimes disperse in groups. Because of the widespread network

of complementation between co-infecting genomes and *de novo* variants being produced by the erroneous RdRp, I argue that considering single genomes as a unit of selection is insufficient to describe viral infections and evolution. Nevertheless, quasispecies theory which has been previously used to account for selection at the mutant network (*quasispecies*) level has teleological flaws: if selection acts on groups of genomes, it is not mainly because they are linked through mutational coupling, but because at different levels (intracellular replication units, cells, tissues) the proteins they produce interact and can be shared between other genomes.

Under this particular light, social evolution theories can help widen the theoretical framework of population genetics without having to account for yet non-validated prerequisites such as mutational coupling. Indeed, social evolution combines kin selection, group selection and game theories to account for interactions between individuals when they share resources, usually referred to as ‘public goods’, which impact their fitness. While the theory has been first developed to account for social interactions in higher organisms, I will present how it has already been applied to virus populations to explain some observations.

2.3.1. BRIEF INTRODUCTION TO SOCIAL EVOLUTION

Social evolution theories were first formulated to explain traits that were identified as altruistic – organisms helping the reproduction of other organisms at the cost of their own fitness. An extreme form of altruistic behavior is the example of sterile eusocial insects, such as honeybees or ants, where large numbers within the colonies work to help the reproduction of a few queens. At first look, these observations defied the selection at the individual level thought to result from Darwinian evolution. Indeed, the emergence of social cheaters was thought to block any form of altruism, as they would benefit from cooperative traits without suffering their fitness costs, therefore having higher fitness than cooperators. Three main mechanisms were then theorized to explain such traits: kin selection, game theory and group selection.

First, kin selection focuses on a gene-centered approach to evolution. It theorizes that for altruistic behaviors to emerge, individuals should only cooperate with other genetically related individuals. William Donald Hamilton used the concept of ‘inclusive fitness’, in addition to the classical ‘personal fitness’, to describe the number of offspring one individual supports

regardless of their progenitor. Using this, he showed that not only does a gene's evolutionary success in a given population depend on its own replication, but success can also increase if it promotes the reproduction or survival of other individuals carrying the same gene. He derived a formula called Hamilton's rule which determines if a gene coding for altruistic traits can stably spread in a population:

Equation 5
$$rb - c > 0$$

where r is the probability that a cooperating trait benefits individuals carrying the altruistic gene and is usually referred to as the 'degree of relatedness', b the reproductive benefits of the recipient of the altruistic trait, and c the reproductive cost of the trait for the altruist. To increase the probability of cooperator assortment, kin recognition – the ability to selectively recognize related individuals – and “viscous” population – the spatial proximity of close kin dispersal – are different mechanisms thought to increase values of r .

Second, evolutionary game theory is the application of classical game theory used in economic sciences to explain stable evolutionary strategies. It was developed by John Maynard Smith to understand how different heritable strategies and their interactions in cost or gain of fitness for players could result in altruistic behaviors while maintaining the assumption that selection occurred upon individuals which act for self-interest. In evolutionary game theory, the evolutionary stable strategies (ESSs) are strategies or mixes of strategies that are stable once adopted by a population. One particularly interesting game is the Hawk-Dove game (also known as 'Snowdrift'), in which two types of strategies exist: hawk which wins the public goods (V) when it meets a dove, but suffers a cost (C) when it meets another hawk; and dove which shares the resources when it meets another dove, but leaves the public goods to the competitor if it is a hawk. When the costs suffered by hawks are higher than the gains ($C > V$), the ESS is a stable mix of hawks and doves in the population. If the hawk and dove strategies are respectively associated with cheaters and cooperators in social interactions, then cooperators can resist the invasion of cheaters if $C > V$. More complex models also exist with the addition of kin selection.

Last, in opposition to the previous theoretical frameworks, group selection rather argues for multi-level selection. It postulates that when competition between groups is more important than competition between individuals in said groups, then selection acts to maximize group fitness. Briefly, it argues that if cheaters tend to beat cooperators in a group,

groups of cooperators outcompete groups of cheaters. Group selection has been widely criticized by gene-centric evolutionary biologists, however, it allows the consideration of division of labor in populations, which can account for certain traits observed in microbial populations²⁰⁵. Division of labor allows for functional specialization and relieves organisms from certain trade-offs. Interestingly, division of labor can result in individual fitness loss for some individuals in mixed populations, while increasing population fitness overall²⁰⁵. The latter case is observed in the case of specialized replicators and maintainers, and is often compared to the division of labor between germinal and somatic cells in multicellular organisms, therefore reinforcing the argument that selection acts upon the group.

2.3.2. APPLICATIONS TO VIRUS-VIRUS INTERACTIONS

While all of these social evolution theories have been first formulated for higher organisms, they also have been applied to virus populations and form an emergent field of studies called sociovirology²².

For example, the study of defective viral genomes (DVGs) and their ubiquitous presence in virtually all viral families have attracted a lot of attention from evolutionary virologists. Because DVGs are social cheaters, their ubiquity should drive cooperative traits in viruses to extinction. However, as I have previously described, complementation and cooperative immune escape are observed in many viruses. An analysis of this problem through evolutionary game theory's Hawk-Dove game, where DVGs are hawks that preferentially hoard public goods when in presence of wild-type dove genomes, but suffer an absolute fitness cost when they are only met with other DVGs that cannot complement them, allows us to predict that cooperators and cheaters can still co-exist in an evolutionary stable state. Works from Chao and Elena found that it is the non-linear advantage of DVGs both not participating in public goods production and replicating faster than wild-type genomes because of their smaller sizes that promote the Hawk-Dove dynamic. Further, they predict that the system evolves towards the co-existence of ultra-cheaters and ultra-cooperators and high-level production of public goods²⁰⁶. This not only recapitulates current observations but can serve as a neutral evolutionary explanation for the existence of DVGs without having to find in them selective advantages.

In addition, kin selection has been used to explain the evolution of interferon suppression as a social trait^{21,203}. Because interferon suppression benefits other viruses infecting neighboring cells but has an inherent fitness cost, it acts as an altruistic trait and is sensible to cheater invasion. Using *in vitro* and *in vivo* experiments with vesicular stomatitis virus, Domingo-Calap *et al.* showed that the spatial structure of infections allows higher assortments between cooperating variants, which promotes higher yields in virus titers and prevents invasion of fitter cheater genomes²⁰³. The importance of spatial structure for social traits could explain the evolution of superinfection exclusion as a mean to increase kin assortments, although other cooperative mechanisms such as increased viral spread have also been proposed²⁰⁷.

Furthermore, mathematical models predict that collective infectious units produced from the same infected cells increase genetic relatedness between co-infecting genomes and limit cheater invasion risk even when complementation is high¹⁵⁶.

2.4. Defective viral genomes are an integral part of viral populations

I have described how viral populations function as complex interacting units, and how social evolution theories can help model their evolution. However, what should be counted as part of viral populations can be a subject of debate. I argue, similarly to Aparicio *et al.* in their review²⁰⁸, that DVGs are integral parts of viral populations, and could play specific roles in viral infections.

Indeed, as I have developed before, DVGs can be considered social cheaters. As such, their existence is not just a meaningless by-product of viral replication, but an expected part of viral populations predicted by social evolution, with conflict dynamics in shared viral and cellular resources. Moreover, as I will develop further in the introduction of Chapter 1, some DVGs have clear interfering properties, stimulating the host immunity^{23,166,209,210} or having enhanced competitive advantages over their wild-type counterparts¹⁶⁶. As a consequence, they can drive standard genomes to evolve resistance to their interfering properties^{211,212}. DVGs that persist over passages are also not random, exhibiting particular features such as preserved ORF^{23,213}, specific mutations¹⁶⁷ or deletions^{23,214,215}. They can compete with one another¹⁶⁷ and accumulate specific mutations not observed in standard genomes^{216,217}. Interestingly, contrary to their commonly accepted portrayal as purely detrimental members of virus populations, they were shown to promote persistent infections^{166,210,218}. A more complete exploration of

possible positive interactions of DVGs with the rest of viral populations has yet to be researched and is the topic of Chapter 1 of this thesis.

Box 1 | MAJOR POINTS OF THE GENERAL INTRODUCTION

- RNA viruses have high mutation and recombination rates due to their error-prone RNA-dependent RNA polymerase. Combined with their small sizes and high replication rates, they can virtually explore all possible mutations.
- Most mutations are deleterious or lethal. Severe bottlenecks at transmission or host-barrier crossing events can fix those deleterious mutations.
- Intra-genomic epistatic interactions between mutations are usually antagonistic, buffering additive deleterious effects and resulting in a rugged fitness landscape with multiple adaptive pathways.
- Viral population interactions can increase robustness to deleterious mutations through mass-effect population robustness and complementation.
- Complementation in viral populations is ubiquitous and can act dominantly (minority variants), with individual genomes being either semi-infectious or having significantly lower fitness than the population one.
- The widespread network of complementation (shared resources), collective infectious units and other forms of virus-virus interactions argue that natural selection acts rather on the population level than the individual genome one.
- Social evolution theories are more parsimonious than quasispecies theory to model and explain virus evolution.
- Defective viral genomes are integral parts of virus populations, driving standard genome evolution and harboring specific mutations.

CHAPTER ONE

CHAPTER 1: POSITIVE INTERACTIONS OF DEFECTIVE VIRAL GENOMES IN ZIKA VIRUS

1. Introduction

1.1. Definition and types of defective viral genomes

Defective viral genomes (DVGs) were first identified by Preben Von Magnus in 1947 in *in vitro* passages of influenza A virus, where he found incomplete viruses that could interfere with and reduce viral yields of wild-type viruses²¹⁹. This discovery raised interest in viral particles with interfering properties, and led to the coining of the term 'defective interfering particles (DIPs) two decades later by Huang and Baltimore²²⁰. However, in their paper, Huang and Baltimore also considered other types of defective viral particles which they excluded from the DIP definition because they lacked interfering properties. In my manuscript, following the definition from Vignuzzi and López¹⁶⁶, I will refer to DVGs simply as viral genomes incapable of completing a full replication cycle in the absence of co-infecting standard viral genomes. This definition includes genomes that do not necessarily interfere, but also genomes incapable of being packaged contrary to defective viral particles considered by Huang and Baltimore. Of note, while all DVGs are by definition cheater genomes, not all cheater genomes are defective on their own.

Because this definition only involves complementation-dependance, DVGs can be generated by diverse genomic alterations: single lethal mutation or hypermutation, frameshift, insertion, deletion, reassortment and copy-back or snap-back¹⁶⁶ (**Figure 10**). Among them, DVGs comprising deletions are the most abundant species observed *in vitro* and *in vivo* in virtually all viral families¹⁶⁶, and are sometimes referred specifically as deletion-containing viral genomes (DelVGs)²²¹. Moreover, copy-back and snap-back DVGs are specifically found in -ssRNA viruses and are generated when the polymerase detaches from its donor template and reattaches to the nascent RNA strand, forming viral RNAs with complementary 3' and 5' ends and typical hairpin structures²²².

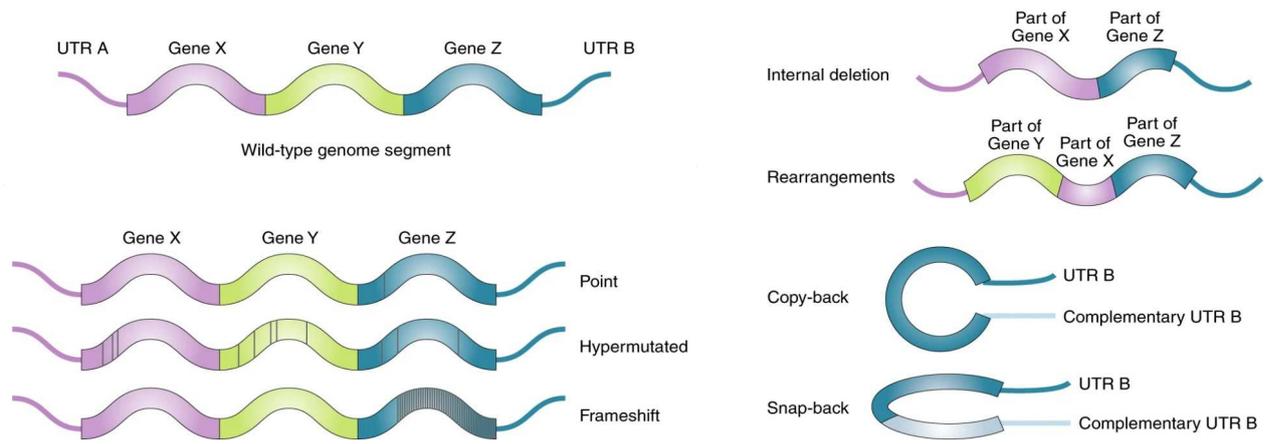


Figure 10 | Types of defective viral genomes. Extracted from Vignuzzi and López, 2019¹⁶⁶.

1.2. Factors of DVG generation and maintenance

Since DVG generation is mainly caused by mutational or recombinational events, random errors generated by viral polymerases are thought to be the main factor of their formation. Indeed, experimental results in RNA viruses showed that mutations in the RdRp affecting recombination rates were linked to increased amounts of DVGs^{50,223,224}. Additionally, mutations outside of the RdRp but affecting proteins interacting with replication complexes were also reported to affect DVG abundance²²⁵⁻²²⁷. Consequently, mutation and recombination hotspots are believed to affect the distribution of types of DVGs¹⁶⁶. The particularity of copy-back and snap-back DVGs in -ssRNA viruses but not in +ssRNA suggests that RNA polarity or yet undiscovered genomic signals might also drive DVG generation.

Finally, as discussed in the general introduction, mechanisms promoting co-infections consequently promote cheater genome (including DVGs) maintenance. Even more, because high multiplicities of infection have been associated with increased DVG generation, these CIUs play a double role: they increase DVG generation and promote their maintenance in new infectious cycles. Surprisingly, some viral protein domains have been directly associated to increased DVG packaging while not affecting standard genome packaging²²⁸.

1.3. Defective Interfering Particles (DIPs): DVGs as social cheaters

When DVGs were identified by Von Magnus, he noticed that infectious viral titers were fluctuating in an oscillating manner across passages²¹⁹. This phenomenon was attached to the interfering properties of the defective interfering particles he identified: when infectious titers

are high, high levels of co-infection occur and DIPs increase in frequency, resulting in following passages in a drop in viral titers; conversely, when titers become too low, the MOI becomes insufficient to sustain DIPs and titers recover. This phenomenon was observed in other viruses^{229,230} and mathematically modeled based on a modified predator-prey model²³¹.

In the following section, I review different mechanisms by which DIPs interfere with standard genomes, and discuss how DIPs can be used for novel antiviral therapies or prevention.

1.3.1. INTERFERENCE OF VIRAL PRODUCTION

First, an already discussed mechanism for DIP interference is their competition against standard genomes for public goods, namely viral and cellular resources¹⁶⁶, which has been modeled using social evolution theories^{156,203,206}.

Briefly, DVGs are thought to replicate faster because of their usual smaller sizes²²⁹, or by encoding signals that enhance their ability to replicate²³². The latter case is well documented for copy-back species which have highly efficient flanking trailer promoters that enhance their replication^{233,234}. The relationship between DVG size and replication kinetics on the other hand appears not to be completely linear, with intermediate optimal deletion sizes found across different viruses^{23,215,229}. Whether this non-linearity is due to necessary *cis*-acting genes²³ or structures that DVGs must preserve, or packaging limitations to ensure proper assembly of the nucleocapsid²³⁵ has still to be resolved (**Figure 11A**).

Moreover, DVGs can compete for other viral resources, notably structural proteins, while keeping their replication rates at similar levels than standard genomes. Enhanced packaging signals have been found in some DIP species^{167,236}, while others showed on the contrary decreased packaging efficiency²³⁵.

Finally, DIPs can interfere with standard genomes by affecting their infectious cycle. A fascinating study by Fonville *et al.* revealed that while semi-infectious particles in influenza A virus promoted reassortment between viral segments, DIPs on the other hand suppressed reassortment²³⁷. This phenomenon could affect population genetic diversity, and therefore viral population fitness. Additionally, DVGs in Sendai virus were reported to affect cellular trafficking, thus affecting released viral particles' composition²³⁸. Defective protein production

from DIPs can also negatively affect viral population through dominant phenotypes such as with oligomeric structural proteins^{239,240}.

Of note, DVGs do not necessarily need to outcompete standard genomes for replication or packaging to harbor interfering properties, as their defection in viral protein production readily reduces available resources for standard genomes.

1.3.2. STIMULATION OF HOST IMMUNITY

Other well characterized mechanisms by which DIPs interfere is through stimulation of their host immunity.

In vertebrates, DIPs are recognized by pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs) and RIG-I-like receptors (RLRs). This in turn stimulates the secretion of type I and III interferons (IFNs) and pro-inflammatory cytokines, including tumor necrosis factor (TNF), interleukin (IL)-6 and IL-1 β ^{209,227,241-244}. More particularly, copy-back DVGs were reported to be extremely potent at immunostimulation with higher magnitudes than DelVG²⁴⁵. This potency is mediated by copy-back-specific motifs enhancing RIG-I sensing of vRNA²⁴⁶. In addition, DIPs stimulate the adaptative immunity by promoting maturation of dendritic cells and enhancing antigen presentation capabilities^{242,247} (**Figure 11B**).

In plants, DIPs have been shown to activate the RNA interference (RNAi) response through enhancement of post-transcriptional gene silencing²⁴⁸. In insects, the RNAi response is also involved through the implication of Dicer-2. DVGs can be processed classically through their dsRNA form by the RNase III domain of Dicer-2 to produce small interfering RNAs (siRNAs), but they were shown to also serve as templates for the formation of linear and circular viral DNAs (vDNAs) which amplify the siRNA silencing pathway²⁴⁹.

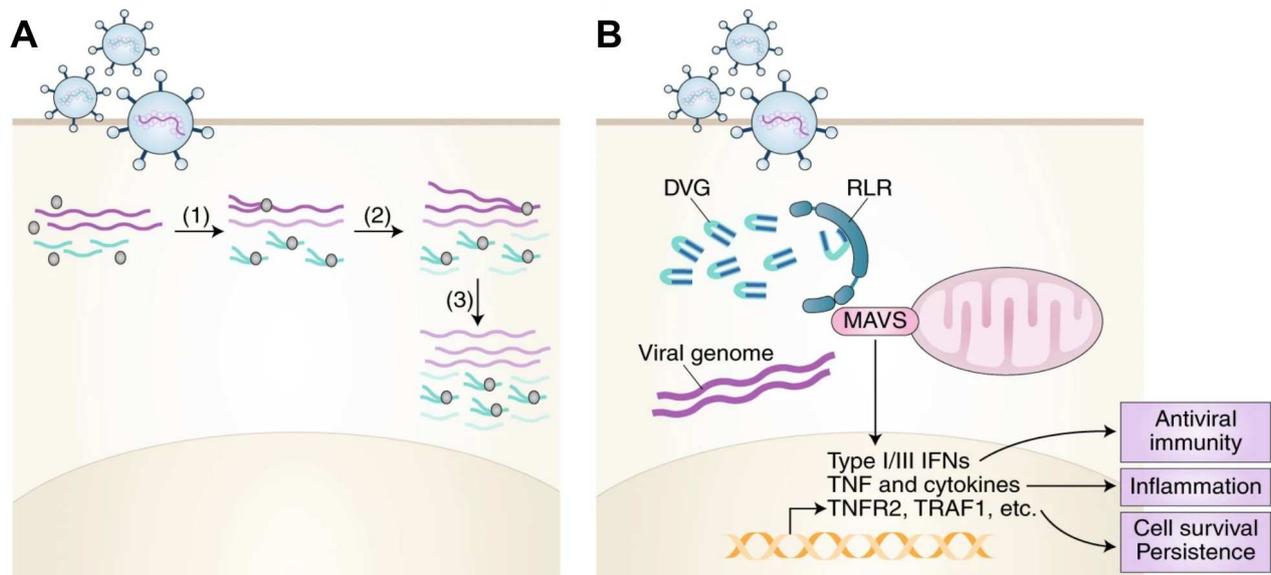


Figure 11 | Modes of interference of defective viral genomes. (A) Interference of standard viral production. Here is depicted competition for polymerase replication, where smaller size DVGs replicate faster than full-length genomes. (B) Interference through immunostimulation. Here is depicted copy-back DVGs being recognized by PRRs, inducing innate immunity and inflammation signaling. MAVS: mitochondria antiviral-signaling protein. Extracted from Vignuzzi and López, 2019¹⁶⁶.

1.3.3. THERAPEUTIC INTERFERING PARTICLES (TIP): USING VIRAL PARASITES FOR CONTROL, PREVENTION AND TREATMENT

The interfering properties of DIPs have initiated several studies to use them as antiviral therapies for mainly human infections. These specific DIPs termed ‘therapeutic interfering particles’ (TIPs) were shown to induce potent immune response *in vivo* and reduce viral loads when administered to animals^{23,214,250,251}. Specifically, some TIPs were shown to work as prophylactics to prevent against cognate standard viral infection challenges^{250,251}, but also to cross-protect against other RNA viruses from different viral families²⁵⁰. However, this latter broad-spectrum protection could also just be a result of boosted innate immune response through the interferon pathway, rather than directly interfering with viral replication.

In addition, a TIP of Zika virus identified in our laboratory was shown not only to work as a therapeutic in mice, but to also drastically limit viral dissemination and transmission when injected in mosquito vectors²³. This opens the way to the use of DIPs as a way to mitigate arbovirus epidemics by controlling viral populations directly in their vectors, therefore limiting spill-overs or breaking their urban life cycles.

Finally, DVGs were identified in different live-attenuated vaccines²⁵²⁻²⁵⁴. However, it remains unclear if these DVGs provide additional immunization, for example by acting as adjuvants through increased immune stimulation²⁵⁵.

1.4. Positive interactions of DVGs?

I have depicted in the previous section a strictly negative portrait of DVGs for their viral populations. However, while some of them first attracted interest because of their interfering properties, potential roles in promoting persistent infections and chronic diseases have been associated to DVGs^{166,210}. This opens new research questions concerning alternative positive roles of DVGs in viral populations. More precisely, investigations of DVGs that not only do not interfere but actively help increase viral population fitness would drastically change the current vision of DVGs as simple viral cheaters. In the following paragraphs, I first present current results in DVGs' implication in viral persistence, then explore alternative positive roles they could play for standard genomes.

1.4.1. ROLE IN VIRAL PERSISTENCE

It has been observed on several occasions that DIPs lead to persistently infected cell cultures *in vitro*^{218,230,256,257}. This phenomenon has been attached to the ability of DIPs to stimulate the interferon response, which correlates with the establishment of persistent infections²⁵⁸. In their work, Xu *et al.* showed that cells containing high levels of DVGs were capable of surviving infection²⁵⁹, therefore promoting persistence. They attached this phenotype to a differential activation of the TNF-induced immune pathway: high levels of DVGs induced the expression of TNF receptor associated factor 1 (TRAF1) through activation of mitochondria antiviral-sensing proteins (MAVS), which promoted cellular survival through the TNF receptor 2 (TNFR2) activation pathway; conversely, cells with low levels of DVGs were stimulated by endocrine TNF through TNFR1, promoting cell apoptosis^{210,259}.

While these observations of persistent infections induced by DVGs were validated *in vitro*, the presence of DVGs in infections *in vivo*²⁶⁰ suggests they could play a role in viral infections that typically result in chronic diseases. Notably, Von Magnus oscillations of viral titers could be linked to Ebola virus reactivation in convalescent patients²⁵⁷. Further

experimental investigations of DVGs *in vivo* should elucidate if they indeed play a role in persistent infections, or rather if other mechanisms are involved.

1.4.2. ADJUSTING PROTEIN STOICHIOMETRY

DelVGs that selectively preserve their ORF are believed to be able to translate their genomes into functional deleted polyproteins^{23,213}. While unbalanced protein production can lead to interfering properties, such DVGs could alternatively play similar roles than subgenomic RNAs in regulating protein stoichiometry. Indeed, a hepatitis C virus DVG has been reported to increase core protein accumulation without translation of envelope glycoproteins, resulting in an upregulation of viral genome replication and release²¹³. Moreover, DVGs observed in Flock House virus contain multiple deletions without accumulation of single-deletion intermediates, which could be the result of regulation signals that promote specific types of DelVGs, and consequently translation products²¹⁵. One of this DelVG in Zika virus has been however associated with interfering properties²³, but this only occurred at DelVG proportions probably higher than what occurs in natural infections, therefore having negative effects through artificially skewed protein stoichiometry.

If validated, this hypothesis would strengthen Simon-Lorier and Holmes's claim that recombination rates have evolved to regulate viral protein expression⁴⁵.

1.4.3. CAN DVGs COMPLEMENT WILD-TYPE GENOMES?

Another interesting consequence of DelVGs being capable of protein production is that their own mutations can in theory *trans*-complement standard viral genomes. Specific mutations were shown to covariate with deletion breakpoints, while others anti-correlated with the latter²¹⁷. Specifically, in Flock House virus, one single-nucleotide variant (SNV) found in a DelVG was found to be lethal in standard full-length genome²¹⁷, indicating that rearrangements in DVGs alleviate antagonistic or negative sign epistasis. This could signify that some SNVs play specific roles in DVGs, perhaps for the benefit of the viral population. Moreover, reports of a dengue virus type 1 DVG showing long-term co-transmission with full-length genomes over several years²⁶¹ suggest that maintenance of some DVGs could promote their evolution and allow division of labor if co-transmission is favored.

1.4.4. CAN DVGs ENHANCE POPULATION EVOLVABILITY?

Different studies have reported that DVGs tend to accumulate more mutations than standard genomes^{216,226,244,262}, usually with RNA editing signatures from the adenosine deaminase acting on RNA (ADAR) enzyme family^{226,262}. Whether these hypermutation patterns are linked to relaxed purifying selection due to complementation by standard genomes or preferential editing of DVGs has still to be determined¹⁶⁶. Nevertheless, this implies that DVGs can increase genetic diversity at particular loci, which could in turn benefit the adaptability of viral populations through complementation or genetic exchange through recombination.

1.5. Zika virus

1.5.1. GENOME AND VIRUS STRUCTURE

ZIKV is an enveloped, single-stranded positive-sense RNA virus from the flavivirus genus, *Flaviridae* family. Its genome consists of a linear non-segmented +ssRNA molecule of around 10.8 kb with a single ORF, a 5' a 7-methylguanosine cap and a 3' loop structure which leads to the formation of a non-coding subgenomic flavivirus RNA (sfRNA) upon degradation of the rest of the genome. The ORF encodes at its 5'-part three structural proteins, namely a capsid (C), a precursor membrane (prM) and an envelope protein (E). The remaining majority of the ORF codes for seven non-structural proteins (NS), namely NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5, and a 2 kDa peptide (2K) (**Figure 12A**).

The genome is enclosed in an icosahedral nucleocapsid, which is enveloped in a endoplasmic-reticulum-derived membrane decorated in its mature form by 90 antiparallel E dimers arranged in an icosahedral-like symmetry. The overall viral particle is a smooth spherical structure of around 50 nm (**Figure 12B**).

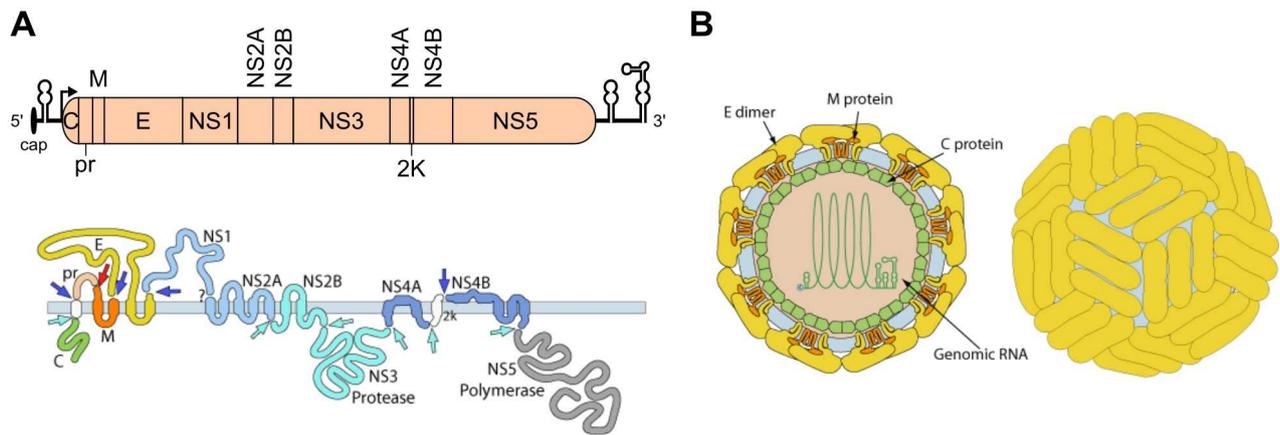


Figure 12 | Genome and polyprotein organizations and viral particle structure of Zika virus. (A) ZIKV genome viral RNA and its polyprotein membrane anchoring. Arrows show cleavage site during the polyprotein processing: dark blue arrows for host proteases, light blue arrows for NS3 protease activity and red arrow for Golgi protease. (B) Schematic representation of the mature viral particle. Extracted from ViralZone.

1.5.2. REPLICATION CYCLE

Similar to other flaviviruses, viral entry is mediated by the interaction of E proteins with a variety of host factors expressed at the host-cell surface such as heparan sulfate or phosphatidylserine receptors. However cellular receptors are still largely uncharacterized²⁶³. The virion is internalized through a clathrin-mediated endocytosis, and is then trafficked to late endosomal compartments. As the pH becomes more acidic, the E protein dimers disassociate and form E trimeric structures, exposing their fusion loops which interact with the endosomal membrane and trigger the nucleocapsid release in the cytosol.

The nucleocapsid disassembles and the positive-sense vRNA is readily translated by host ribosomes at the endoplasmic reticulum (ER) membrane. The genomic RNA is translated into a single polyprotein which anchors itself in the ER membrane via the transmembrane domains of prM, E, NS2A, NS2B, NS4A and NS5B. This polyprotein exposes C, NS3 and NS5 to the cytosol, while prM, E, NS1 and parts of NS2A, NS4A and NS4B are located in the ER lumen. It is then cleaved into individual proteins by host proteases on the ER lumen side, and through NS2B-NS3 protease activity for the cytosol-facing cleavage sites (**Figure 12A**).

NS1, NS2A, NS2B, NS4A and NS4B transmembrane proteins associate to provide a molecular scaffold for NS5 RdRp and NS3 activities, resulting in ER membrane

rearrangements into small membrane invaginations called vesicle pockets (VPs) where the replication complex is located. There, the positive-sense single-stranded vRNA is transcribed by the NS5 RdRp activity into dsRNA. The NS3 helicase then unwinds the generated dsRNA into positive- and negative-sense ssRNAs. The -ssRNA is then processed again by the RdRp to serve as a replication template, while produced +ssRNAs combine with structural membrane-associated proteins C, prM and E in assembly complexes.

The structural protein complexes bud into the ER lumen to form immature virus particles and are trafficked to the trans-Golgi network where they are processed. Immature virions are characterized by trimeric structures of prM-E heterodimers which form spikes at the particle surface, unlike the smooth mature virus particle. The acidic environment of the trans-Golgi network provokes a conformational change in E proteins which rearrange to form antiparallel dimers which lie flat against the viral membrane, exposing prM proteins at the surface. This exposes a furin cleavage site on prM proteins which is then cleaved by host proteases in the Golgi lumen into precursor peptides pr and the membrane proteins M. Due to the acidic conditions of the Golgi, pr peptides stay associated with the E-M complexes. Finally, mature virions are released into the neutral extracellular milieu through exocytosis, which releases the pr peptides from the final mature virion surface²⁶⁴ (**Figure 13B**).

Additionally, some +ssRNAs are partially degraded by host 5'-to-3' exoribonuclease 1 (Xrn1) proteins, resulting in sfRNAs from the degradation-resistant 3' loop structure at the end of the genome. These sfRNAs are known to antagonize host immune response²⁶⁵.

For a graphical summary of Zika virus replication cycle, see **Figure 13A**. **Table 1** summarizes known protein functions described in Lindenbach *et al.*, 2021²⁶⁶.

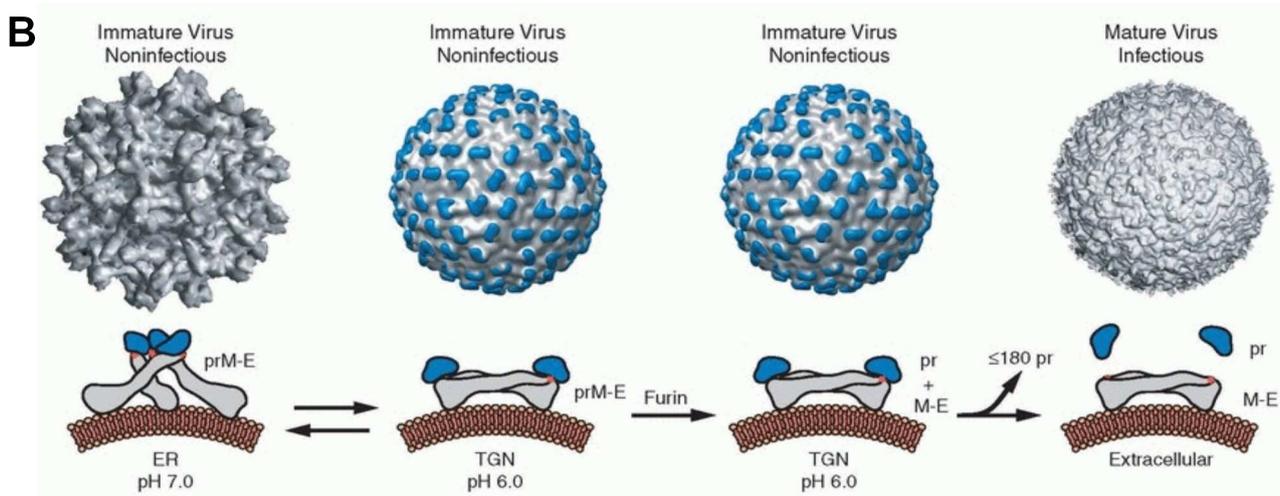
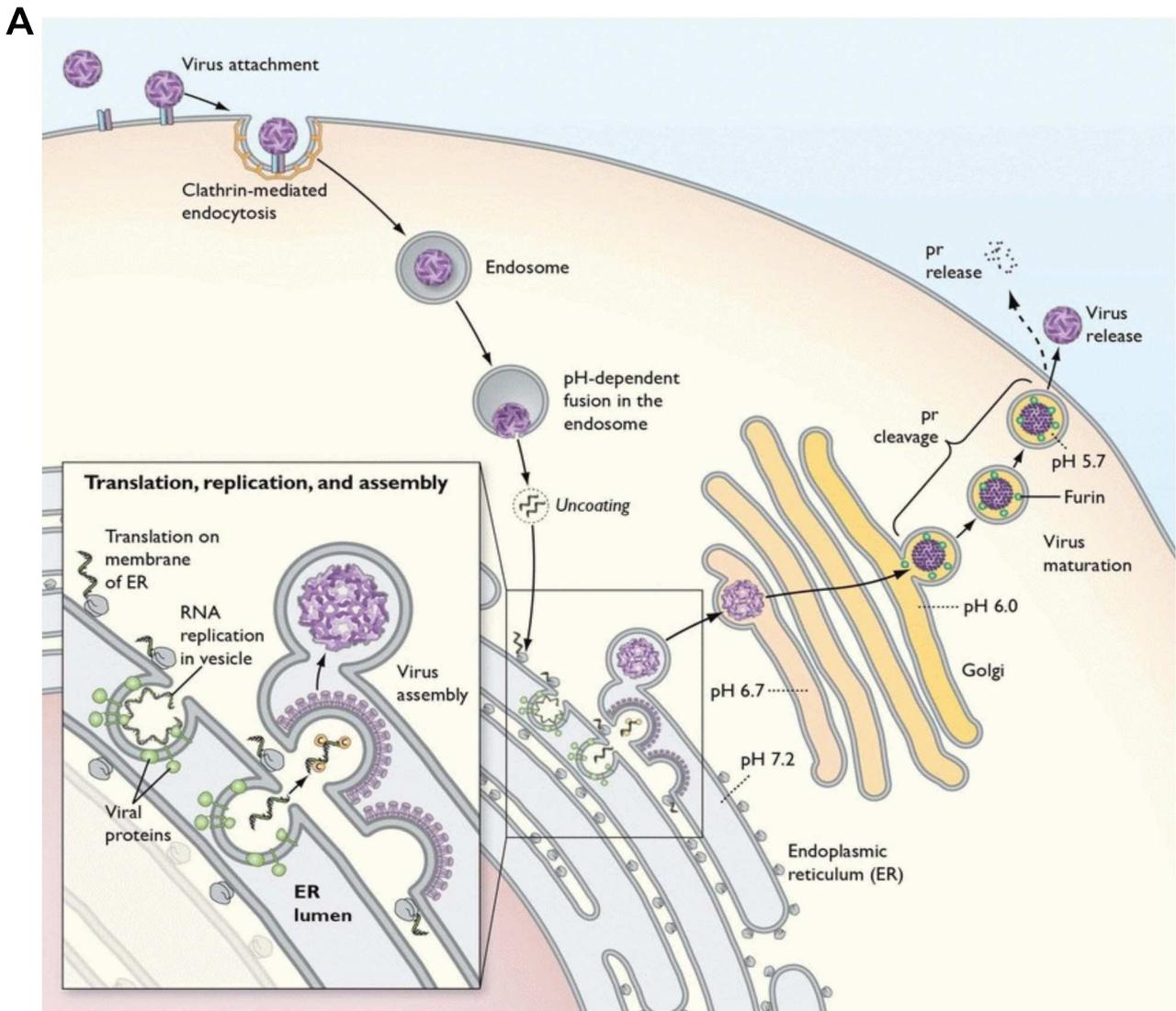


Figure 13 | Zika virus replication cycle. (A) The complete ZIKV replication cycle. **(B)** Maturation of ZIKV virions through pH-dependent conformational E proteins changes and prM furin cleavage in the trans-Golgi network (TGN). Extracted from Pierson *et al.*, 2021²⁶⁷ and Lindenbach *et al.*, 2021²⁶⁶.

Protein	Size (aa)	Function
C	104	Encapsidates viral genome.
prM	168	Assists proper folding of E and prevents fusion during maturation; anchors the prM-E complex through a C-terminal ER-retention domain.
E	500	Mediates receptor binding and membrane fusion through fusion loop; major neutralizing antigen.
NS1	352	Essential for formation of VPs and RNA replication; contributes to virus particle assembly; major antigen; may enhance transmission to mosquitoes ²⁶⁸ .
NS2A	226	Part of the replication complex molecular scaffold; inhibits interferon response.
NS2B	130	Part of the replication complex molecular scaffold; co-factor of the NS2B-NS3 serine protease activity; inhibits innate immunity.
NS3	617	Cleaves polyprotein and inhibits interferon response via N-terminal serine protease catalytic domain; RNA helicase.
NS4A	127	Part of the replication complex molecular scaffold; co-factor of NS3 helicase activity; induces membrane alterations via 2K cleavage; antagonizes innate immunity.
NS4B	251	Part of the replication complex molecular scaffold; antagonizes innate immunity.
NS5	903	N-terminal guanylyltransferase and methyltransferase activities for 5' RNA capping; C-terminal RdRp; inhibits interferon response; may dysregulate host gene expression.

Table 1 | Zika virus proteins and functions. Summarized from Lindenbach *et al.*, 2021²⁶⁶.

1.5.3. A ZIKV DVG TO EXPLORE POSITIVE INTERACTIONS

Previous work in the laboratory by Rezelj *et al.*²³ identified a TIP candidate ZIKV DVG comprising an in-frame deletion of nucleotides 581 to 3250, resulting in a conserved ORF which encodes for a polyprotein lacking the C-terminal 130 amino acids of prM, the complete envelope protein E, and the N-terminal 254 amino acids of NS1 (**Figure 14A**). This DVG was shown to be incapable of packaging its own genome due to the lack of most of M and E glycoproteins, but is also incapable of replicating its genome in absence of wild-type virus.

Complementation of this DVG by wild-type was shown to require an intact ORF (**Figure 14B**), suggesting that DVG RNA replication requires some of its own non-structural proteins encoded in 3' of the deletion to act in *cis*. Moreover, replication of DVG RNAs was restored by solely providing in *trans* the NS1 protein to the DVG (**Figure 14C**).

This DVG showed interfering properties both in mammalian and insect cells when co-transfected at 1:1 or 10:1 (DVG:WT) molar ratios. In mammalian cells, this interfering property was only observed when the DVG preserved its ORF (**Figure 14D**), but whether the high number of DVG RNAs replicating, stoichiometry imbalance in protein production, NS1 snatching, or other mechanisms are involved in this interference has still to be explored. In contrast, in insect cells the ORF is not required, and interference was shown to be mediated by the RNA interference pathway through Ago2 knock-downs. *In vivo* experiments in mice and mosquitoes have shown reduced viral replication of WT viruses when the animal models were treated with this DVG. In mice, DVG treatment reduced viral symptom severity, while treatment in mosquitoes showed drastic reductions of viral dissemination and transmission. Altogether, this DVG was proposed as a candidate TIP for treatment or prevention of ZIKV infections. Importantly, this specific type of DVG has been observed in West Nile virus and yellow fever virus high MOI passages, suggesting that it may be constitutive of flavivirus viral populations in general.

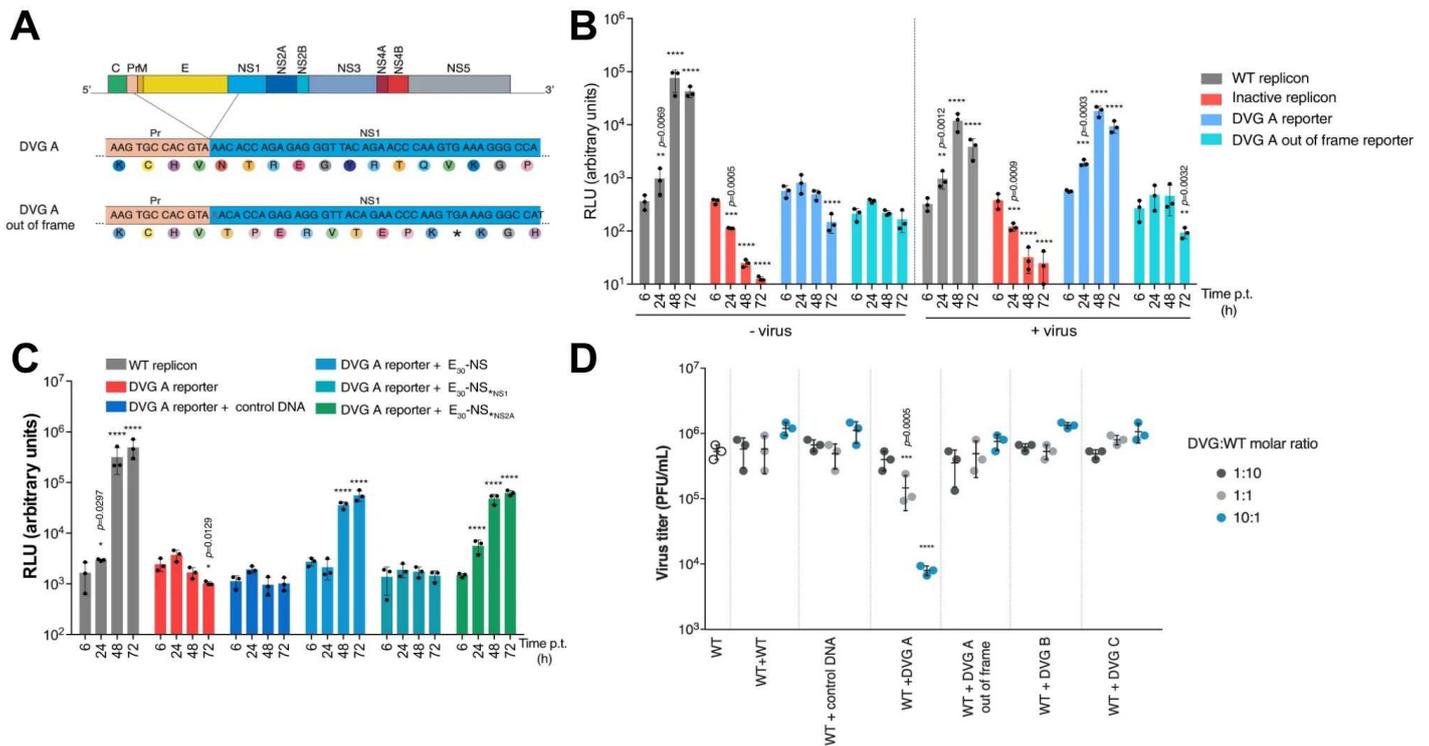


Figure 14 | The candidate ZIKV DVG requires a conserved ORF and WT-encoded NS1 in *trans* to be replicated and show potent interfering properties *in vitro*. (A) The candidate DVG comprises a deletion that conserves the ORF. A +1 framshift (DVG out of frame) results in a truncated polyprotein. (B) Complementation assay of nanoluciferase reporter DVGs or replicons in infected or uninfected cells. Replication of the DVG is only restored in presence of infectious virus and a conserved ORF. (C) Complementation of the DVG only requires NS1 in *trans*. (D) Co-transfection of DVG and WT genomes results in decreased infectious titers in HEK-293T only when DVGs retain the ORF. Extracted from Rezelj *et al.*, 2021²³.

Finally, this DVG was shown to persist and increase in frequency relative to other deleted genomes over successive passages, suggesting that it may be packaged and transmitted to new uninfected cells. However, apart from experimental virus-like particles (VLPs), no experimental evidence of constitutive viral packaging of these genomes has yet been performed, and their presence in the population could alternatively be the result of *de novo* generation. In case of constitutive DVG packaging, questions relative to their collective dispersal with WT genomes to be complemented and maintained need to be addressed.

Besides its promising role for antiviral therapy and control, this DVG is an ideal candidate to study possible positive interactions with full-length genomes. First, because it retains a functional ORF, positive non-synonymous mutations affecting coding sequences could be shared with WT through complementation. Second, since we showed that this DVG had potent interfering properties, showing positive interactions that alleviate this cost in

population fitness would suggest that viral populations could evolve to mitigate the cost of ubiquitous cheater genomes. Moreover, because it appears to be maintained across passages, mutations arising in this DVG could be transmitted to uninfected cells, selected for their population advantages, and allow division of labor or memory genomes dynamics. Interestingly, DVG genetic diversity could therefore play a role in population evolvability, either through dominant complementation like minority variants, or genetic recombination with full-length genomes. Finally, since this type of DVG is observed in mammalian and insect cells and in at least two different flaviviruses, experimental observations in the following study could lead to investigations of such dynamics in natural populations.

1.6. Aims of Chapter 1

To explore if DVGs could play positive roles in viral population infections and their evolution, we first aimed at identifying a beneficial mutation that could be used in controlled *in vitro* experiments to test our hypothesis. We realized classical experimental evolution passages of wild-type ZIKV in mutagenic conditions to select and identify resistance mutations to ribavirin treatments.

Then, by introducing the identified mutation in DVGs but not in WT genomes, we tested if co-infections of beneficial mutation-carrying DVGs and WT could benefit overall population survival in mutagenic conditions, therefore alleviating the DVG interfering cost. We then aimed at identifying if the underlying mechanism of this positive interaction was the result of complementation or genetic recombination.

Finally, we tried to validate that DVGs could be naturally packaged in virions to be maintained across multiple rounds of infections, and looked for collective infectious units which could promote DVG maintenance through DVG and WT co-infections.

2. Article: Defective viral genomes can accelerate viral population adaptation

The inserted manuscript corresponds to the version we submitted to Nature Microbiology in March 2022. It was desk-rejected by the editor. Since then, more investigations have been performed and were added in the following 'Complementary results' section.

1 **Defective viral genomes can accelerate viral population adaptation.**

2 Quang Dinh Tran^{1,2}, Veronica V. Rezelj^{1,*}, Maxime Chazal³, Lucía Carrau¹, Florian Mueller⁴,
3 Djoshkun Shengjuler¹, Thomas Vallet¹, Nolwenn Jouvenet^{3,*} & Marco Vignuzzi^{1,*}

4

5 **Author Affiliations:**

6 ¹ Institut Pasteur, Université de Paris, Viral Populations and Pathogenesis Unit, Centre National de
7 la Recherche Scientifique UMR 3569, F-75015 Paris, France

8 ² École Doctorale BioSPC, Université de Paris, Sorbonne Paris Cité, F-75006 Paris, France

9 ³ Institut Pasteur, Université de Paris, Virus Sensing and Signaling Unit, Centre National de la
10 Recherche Scientifique UMR 3569, F-75015 Paris, France.

11 ⁴ Institut Pasteur, Université de Paris, Imaging and Modeling Unit, CNRS UMR 3691, C3BI USR
12 3756 IP CNRS, F-75015 Paris, France

13

14 * corresponding authors

15

16

17 **Abstract**

18 Defective viral genomes (DVGs) are non-replicative genomes generated through non-
19 homologous recombination during viral replication. They are known to negatively impact viral
20 replication by triggering potent innate immune responses or competing with wild-type viruses for
21 viral and host resources required for replication. Positive interactions between DVGs and wild-type
22 viruses have not been described so far. By combining experimental evolution and barcoded viral
23 genomes, we demonstrate that DVGs are capable of exchanging genetic material with wild-type
24 genomes through recombination and can provide advantageous alleles in challenging environments.
25 RNA fluorescence *in situ* hybridization experiments on released virions showed that DVGs can be
26 co-packaged with wild-type genomes, which may contribute to the maintenance of DVGs in the
27 population. We challenge the current understanding of DVGs as solely deleterious for viral
28 replication and propose that, when present at a low frequency, they can contribute to viral
29 population adaptation by expanding the exploration of the fitness landscape.

30 **Introduction**

31 RNA virus populations exist as swarms of genotypes with vast variability, due to their high
32 mutation rate and large population sizes, which allow them to evolve rapidly and adapt to new
33 environments¹. Due to the lack of proofreading mechanism by viral RNA polymerases, at the
34 exception of coronaviruses, it is estimated that RNA viruses incorporate on average one mutation in
35 each progeny genome^{2,3}. While it was initially thought that recombination was a rare event in RNA
36 viruses, it is now established that exchange of genetic material between viral genomes has
37 considerable evolutionary significance. Indeed, RNA recombination has been implicated in
38 expanding viral host range^{4,5}, enhancing virulence⁶⁻⁸, adaptability^{6,9}, and resistance to antiviral
39 treatment^{10,11}. Thus, RNA recombination can also increase genetic diversity in a virus population.

40 During replication of the viral genome, recombination takes place when the RNA synthesis
41 along a genome template is disrupted and the RNA polymerase resumes synthesis elsewhere on the
42 template or along another genome, resulting in the formation of a chimeric molecule.
43 Recombination may be regarded as a form of viral sex that can accelerate the creation of
44 advantageous or the purging of deleterious genotypes in viruses¹². RNA recombination is also
45 responsible for the production of defective viral genomes (DVGs). DVGs are degenerate forms of
46 the viral genome that are generated as a result of ‘imperfect’ recombination leading to deletions,
47 insertions, genome rearrangements, or a combination of these. Because they lack genetic
48 information encoding key functions, DVGs cannot replicate or propagate on their own, and thus
49 require the presence of wild-type (WT) virus.

50 While initially regarded as an *in vitro* by-product of virus propagation, next-generation
51 sequencing (NGS) has revealed that DVGs are present in virtually any RNA virus population and
52 that their diversity is immense¹³⁻¹⁵. DVGs have recently been identified in patients infected with
53 influenza A virus¹⁶, respiratory syncytial virus¹⁷, hepatitis C virus^{18,19} or dengue virus^{20,21}, as well as
54 in birds infected with West Nile virus²². Experimental evolution combined with NGS approaches
55 have enabled the genotypic characterization of virus populations to further describe the nature of
56 DVGs. Jaworski and Routh have previously shown that, while a large pool of Flock House virus
57 DVGs are generated, only a small number of these can accumulate over passages *in vitro*¹⁴. Our
58 previous work has shown a similar trend for Zika virus (ZIKV) and other flaviviruses, as well as for
59 chikungunya virus: while many deletions were observed throughout the genome, only a handful of
60 DVGs was propagated over passages in mammalian and mosquito cells^{23,24}. These DVGs exerted
61 strong interfering activities against WT virus in mice or in mosquitoes, strengthening the
62 assumptions that DVGs mainly act as parasites^{25,26}.

63 Apart from a role in facilitating persistence of viral infections *in vitro*^{18,20,21,27-30}, positive
64 interactions with WT virus have not been extensively investigated. Here, using a ZIKV DVG that
65 we previously described²³ combined with a ribavirin resistance mutation identified by experimental
66 evolution, we demonstrate that even potent interfering DVGs can provide positive fitness to the
67 overall viral population by exchanging beneficial alleles through recombination in challenging
68 environments. Further, we show using RNA fluorescence *in situ* hybridization (FISH) that DVGs
69 can be co-packaged with WT viral genomes, albeit at low frequencies, suggesting that DVGs can
70 share viral resources. We propose that DVG-WT genome interactions are more complex than
71 simple parasitism and could play a significant role in shaping viral evolution.

72 **Results**

73 **Identification of mutations conferring ribavirin resistance in Zika virus**

74 Our goal was to understand if beneficial mutations carried by co-infecting DVGs could
75 positively increase viral population fitness and counteract their interfering properties. First, we
76 sought to identify mutations that could be beneficial under controlled conditions *in vitro*, and which
77 could be used in downstream experiments. Selection of ribavirin (a mutagenic guanosine analogue)
78 resistance has been achieved for several RNA viruses³¹⁻³⁷. To select for a ZIKV variant that is
79 ribavirin resistant, we serially passaged the virus fourteen times in Vero cells either in vehicle
80 (DMSO) alone, or in the presence of 400 μ M ribavirin (Fig 1a). While ribavirin treatment reduced
81 virus yields in early passages, this phenotype was lost in two replicates out of three (replicate A and
82 B) by the eighth passage, suggesting that mutagen resistance had been achieved (Fig 1b).
83 Sequencing of the virus population in each passage identified two mutations in the viral polymerase
84 that did not arise in control conditions, and for which frequencies increased throughout passaging.
85 These mutations were found in the replicates A and B, which exhibited ribavirin resistance in late
86 passages. No such mutations were identified in replicate C, which showed lower overall titres in the
87 presence of ribavirin in late passages, thus suggesting no ribavirin resistance arose after fourteen
88 passages. The mutation T9203G in replicate B (Fig 1b, right panel) emerged in passage number
89 four and became rapidly fixed in the population, reaching frequencies of 1 by passage number
90 seven. The mutation A9207C, found in replicate A, emerged in passage number six and reached
91 frequencies of 0.85 by passage number twelve (Fig 1c, left panel). Both T9203G and A9207C
92 mutations resulted in a non-synonymous change in neighbouring amino acids in the viral
93 polymerase: L513V and Q514P, respectively. Mapping of the amino acids on the crystal structure
94 of the ZIKV polymerase³⁸ (PDB entry 5U05) revealed that these residues are found in the
95 polymerase palm, in close proximity to the active site, and project towards the priming loop (Fig
96 1d). The appearance of independent non-synonymous mutations in two replicates that exhibited
97 ribavirin resistance, coupled with their proximity to the polymerase active site suggest that these
98 mutations may be mediators of ribavirin resistance.

99

100 **Zika virus polymerase L513V confers ribavirin resistance without reducing viral fitness *in*** 101 ***vitro*.**

102 To test if the mutations confer ribavirin resistance, we introduced the T9203G or the A9207C
103 mutation in a ZIKV infectious clone and rescued both viruses in 293T cells, hereafter referred to as
104 Zika virus L513V and Q514P, respectively. Next, we tested their ability to resist ribavirin treatment
105 by assessing the percentage of surviving infectious viruses in the supernatant of ribavirin-treated

106 Vero cells, compared to vehicle (DMSO)-treated cells. Both L513V and Q514P variants resisted
107 ribavirin treatment significantly better than WT virus when using 200 μ M ribavirin, whereas only
108 L513V resisted significantly better the 400 μ M ribavirin treatment (Fig 2a). Since the L513V
109 variant showed better resistance than Q514P, it was selected for further experiments.

110 L513V and WT viruses exhibited similar growth phenotypes in Vero cells (Fig 2b). To assess
111 whether L513V incurs a fitness cost *in vitro*, we performed a competition assay by infecting cells
112 with a population containing a 1:1 ratio of WT:L513V variant and changes in ratios were monitored
113 by Sanger sequencing the viral populations in three successive passages. WT and L513V alone
114 remained stable throughout passaging and L513V did not show signs of reversion (Fig 2c). In direct
115 competition, the 1:1 WT:L513V ratio was maintained throughout the passages, suggesting that the
116 L513V variant does not incur a significant fitness cost under normal conditions.

117

118 **Beneficial mutations in interfering DVGs can increase viral population fitness**

119 To assess if DVGs can increase viral population fitness in adverse environments, we
120 compared the evolution of WT virus on its own, or when it is complemented with a DVG carrying
121 the L513V mutation that confers ribavirin resistance, under normal or mutagenic conditions in Vero
122 cells. We chose a DVG which we have previously identified and characterized in Vero and C6/36
123 cells²³. The aforementioned DVG lacks the region encoding two viral structural proteins (pre-
124 membrane (PrM) and envelope (Env)) and the N-terminal part of the non-structural protein NS1 but
125 conserves the open reading frame following the deletion to maintain the translation of the remaining
126 NS proteins, which are essential for genome replication (Fig 3a). This DVG can replicate and
127 propagate with the virus population in the presence of WT virus; yet on its own, it is replication-
128 incompetent and functionally inert. Furthermore, we demonstrated that this DVG has high
129 interfering properties *in vitro* and *in vivo* against WT virus when present at a 10-fold excess ratio²³.

130 To differentiate between WT virus and DVG genomes in NGS experiments, we barcoded the
131 WT virus genome by introducing five silent mutations into its Env-coding region spanning a 40 nt
132 region (WT tag) (Fig 3a). Similarly, in order to discern between ribavirin resistance mutations
133 arising by *de novo* mutations or positive selection of the ones already present in the DVG genome
134 (namely T9202G and G9204A), we barcoded the DVG genome with four silent mutations spanning
135 a 30 nt region containing the L513V mutation (DVG tag). We generated virus populations derived
136 from infectious clones containing barcoded WT virus only, or a mixed population of barcoded WT
137 virus and barcoded L513V-carrying DVG. We next passaged these populations four times in mock-
138 or ribavirin- treated cells at low or high MOI (in triplicate). Ribavirin resistance of the derived
139 populations was assessed and the frequencies of the WT or DVG tags were determined by

140 sequencing each population (Fig. 3b). Of note, a lower concentration of ribavirin was used for these
141 passages (100 μ M) to prevent population extinction.

142 To test if the complementation of WT virus with L513V-encoding DVGs resulted in an
143 advantageous phenotype for competent viruses, we challenged each of the unpassaged (p0) or
144 passaged (p4) viral populations with higher amounts of ribavirin (200 μ M or 400 μ M) in Vero cells
145 and measured the percentage of surviving infectious progeny virus relative to control DMSO-
146 treated cells by plaque assays. Unpassaged mixes of WT and DVG did not show significant
147 differences with WT virus alone, suggesting the phenotype of L513V is either not available in *trans*
148 for competent viruses, or that its frequency is not sufficiently high in the population. WT virus
149 passaged alone also did not show significant differences in resistance to ribavirin compared to
150 unpassaged virus. On the other hand, when DVG was present, WT virus demonstrated significantly
151 higher survival in the presence of ribavirin at both concentrations (Fig 3c). These results
152 demonstrate that DVGs can accelerate the evolution of viral population phenotypes faster than *de*
153 *novo* mutations provided they carry beneficial alleles that WT genomes do not readily possess, thus
154 compensating their fitness cost due to interference.

155

156 **WT genomes can acquire beneficial mutations from DVGs through recombination**

157 To further understand how mutations present in DVGs can affect population phenotypes, we
158 monitored DVG and WT tags across these passages through RNA sequencing using the barcodes
159 described in Fig 3a. Due to recombination across passages, both of these tags could in principle
160 form part of the same genome, thus limiting our ability to monitor the true DVG population. To this
161 end, we additionally monitored DVG abundance, by following the DVG deletion junction (deletion
162 of nucleotides 581 to 3250). The WT tag was generally fixed in the population through the
163 passaging series, indicating that the barcode mutations did not revert (Fig 4a, left panels). However,
164 when WT was passaged in the presence of DVGs at high MOI, the frequency of the WT tag was
165 generally lower than WT only-passaged populations. This may be attributed to the use of high MOI,
166 which favours the propagation and maintenance of DVGs (Fig 4a, right upper panel), thus
167 preventing the WT population from reaching 100%. DVG tag frequencies rapidly increased through
168 the passage series and became almost fixed in the population both at low and high MOI passaging
169 conditions in the presence of ribavirin, but not in DMSO-treated cells (Fig 4a, middle panels). The
170 same kinetics were observed for the two mutations that lead to the L513V amino acid change and
171 which were introduced in the DVG tag (Fig 4b). This suggests that, under conditions in which
172 L513V is beneficial, the readily existing mutation carried by the DVG is positively selected and
173 reaches high frequencies in the overall viral population instead of appearing *de novo*. Importantly,
174 the frequencies of the DVG donor backbone did not show any increase as compared to the L513V-

175 containing tag (Fig 4a, right panels). This result indicates that WT virus must have acquired the
176 mutation through recombination and carried it to near fixation in the population.

177 Together, these data show that mutations found in DVGs could serve as a source of genetic
178 material that may be exploited through recombination by the WT virus under adverse conditions, to
179 quickly acquire beneficial alleles rather than relying on random *de novo* mutations. Indeed, in WT-
180 only passaged populations, a mutation in WT genomes that increased in frequency by passage
181 number four, namely T9202G (which results in L513V), was identified only in one out of three
182 ribavirin-treated samples (Fig 4a, bottom left panel). This simply mimics the original passaging
183 conditions that led to the emergence of ribavirin resistance mutations in two out of three replicates
184 in the first experiment at passage number four and six (Fig 1).

185

186 **DVG and WT genomes can be co-packaged in virions**

187 Ensuring DVG maintenance in the viral population can be regarded as a costly process for
188 WT virus, as structural proteins otherwise required for WT genome packaging and virion assembly
189 would be used to generate DVG-containing virions. Co-packaging of both WT and DVG RNA
190 molecules could ensure delivery of both molecules into the same target cell and avoid competition
191 of packaging resources. To assess whether DVG and WT RNAs can be co-packaged in virions
192 released from cells, we analysed the content of thousands of virions by RNA fluorescence *in situ*
193 hybridization (FISH) coupled with immunofluorescence and confocal microscopy. We used a FISH
194 technique that is sensitive enough to detect single RNA molecules.

195 To differentiate between WT and DVG RNAs by FISH, we introduced 164 silent nucleotide
196 substitutions into the DVG sequence that confer specificity for the FISH probe binding site. Virions
197 produced from 293T cells transfected with the WT infectious clone were immobilized in negatively
198 charged glass, incubated with the WT and DVG RNA FISH probes and stained for the viral Env
199 protein (Fig 5a). No DVG signal was detected in WT virions, validating the specificity of the WT
200 RNA probe (Fig 5a). More than three thousand virions were analysed for content. Around 57.9% of
201 released Env molecules were associated with WT viral RNA (Fig 5b, top panel). Env molecules that
202 are not associated with RNA likely stain subviral particles. Indeed, prM and Env proteins from
203 numerous flaviviruses, including yellow fever virus, Japanese encephalitis virus, West Nile virus
204 and tick-bone encephalitis virus, can assemble into subviral particles in the absence of any other
205 viral components³⁹⁻⁴². Around 61.3% of released WT RNAs were associated with Env molecules
206 (Fig 5a and 5b, bottom panel). The remaining WT RNAs may be associated with a number of Env
207 molecules that are under the detection limit of the assays or may be released in an Env-free form,
208 for instance in exosomes, as described for Hepatitis C virus RNA^{43,44}.

209 We then analysed the content of around five thousand virions produced from 293T cells
210 transfected with a 1:1 ratio of WT virus and DVG clones (Fig. 5c). Around 5.6% of Env-positive
211 virions contained both the WT and DVG RNA (Fig. 5d, top panel). The majority (around 61.3%) of
212 WT RNA positive-particles were associated with Env only (Fig. 5d, middle panel) and around 7.5%
213 of them were positive for both Env and DVG RNA (Fig. 5d, middle panel). The majority of DVG
214 RNA positive-particles (around 58.3%) were associated with Env only (Fig. 5d, bottom panel),
215 indicating that defective genome can be packaged by themselves. Finally, around 13.9% of DVG
216 RNA positive particles were associated with both Env and WT RNA signal (Fig. 5d, bottom panel).
217 These data indicate that WT and DVG RNAs can be co-packaged into virions, albeit at a low
218 frequency.

219 **Discussion**

220 Previous work has shown that the genetic diversity of DVGs is vast at multiple levels,
221 including type (deletions, insertions, copybacks, rearrangements), size and copy number^{13,14,23,24,45}. It
222 is well established that the high mutation rates of RNA viruses enable the generation of extensive
223 genetic diversity, and that this is fundamental for the survival and rapid adaptation of the virus
224 population. DVGs, which are non-replicative, rely on WT genomes for their maintenance. Since
225 DVGs are replicated by the error-prone RNA-dependent RNA polymerase, one can envisage that
226 DVGs incorporate mutations at the same rate as full-length WT genomes. Whether point mutations
227 in DVGs could play a biological role in virus populations remained to be investigated. Here we
228 explored whether mutations in DVGs could contribute to the viral population fitness. Our work
229 revealed that DVGs can represent a pool of genetic information that WT virus RNAs can acquire
230 via recombination, enabling the rapid acquisition of advantageous genotypes and phenotypes, and
231 thus facilitating rapid virus adaptation.

232 Since arboviruses alternate between vertebrate and invertebrate hosts, they face continuous
233 selective pressures from changing environments and extensive population bottlenecks. In such
234 cycling conditions, it is possible that mutations are purged or lost by genetic drift from WT
235 genomes if they do not confer a fitness advantage in one environment. Because DVGs are
236 functionally inert on their own but nonetheless propagated by WT virus, they could have the
237 capacity of carrying mutations that would normally negatively affect WT genomes^{46,47}. In principle,
238 this property could allow these genomes to explore a broader spectrum of mutations in the sequence
239 space or store mutations that display antagonistic pleiotropy between different hosts. DVGs have
240 been classically regarded as molecules that negatively interfere with virus replication by triggering
241 innate immune responses or hijacking viral or host resources required by WT virus^{17,48-51}. However,
242 provided that their frequency is low enough that it marginally interferes with WT virus replication,
243 DVGs could, in principle, store valuable genetic information that can be used in new environments
244 or help cross fitness valleys for the broader viral population.

245 By using RNA FISH we showed that DVGs can be either packaged individually in virions
246 or co-packaged with WT genomes. This finding poses additional questions in relation to the role of
247 defective viral particles in the population and their maintenance and survival. Co-packaging of WT
248 and DVG RNAs into the same virion can be regarded as an efficient strategy for enabling
249 propagation and survival of a DVG within the viral population. However, structural constraints
250 imposed by the size of the viral particle may limit such co-packaging from happening. Nonetheless,
251 an increasing number of studies have now shown that viruses have exploited mechanisms to
252 facilitate co-infections, by delivering multiple viral genomes via extracellular lipid vesicles,

253 exosomes, virological synapses, occlusion bodies, virion aggregates and even by binding to enteric
254 bacteria^{13,43,52-60}. In addition, recent work has shown that collective virus spread by the formation of
255 virion aggregates in saliva promotes the emergence of defective interfering particles⁵⁵. The
256 existence of these collective behaviours which maintain and promote the emergence of mediocre
257 genomes poses a serious paradox, even under the ‘survival of the flattest’ model⁶¹, if DVGs are
258 solely regarded as deleterious for viral replication.

259 The work presented here aimed at building a more in-depth comprehension of viral
260 population composition and interactions. By demonstrating that even highly interfering DVGs can
261 be utilized for the advantage of WT viruses, we propose that DVGs can also display positive
262 interactions, mainly by exploring at little to no cost the fitness landscape for the viral population.
263 Therefore, WT-DVG interactions is probably more situated at the delimitation between parasitism
264 and mutualism: when DVG frequencies are high, such as in antiviral approaches^{23,24}, the frequency
265 of cheaters exceeds the lethal threshold and the viral population crashes; however, when frequencies
266 are below this threshold, WT viruses can benefit from the genetic diversity of DVGs and the
267 population becomes more robust at a moderate cost.

268 Advancements in NGS technologies, such as long-read sequencing with low error rates or
269 single-cell sequencing methods, which will respectively enable the genotypic characterization of all
270 RNA types present in a virus population, including DVG RNAs, and the identification of specific
271 virus-derived RNAs in individual cells should allow further investigations on the relationship
272 between defective and WT viral genomes.

273 **METHODS**

274 **Cells and viruses**

275 Vero and Vero-E6 cells were maintained in Dulbecco's modified Eagle's medium (DMEM)
276 containing 10% (v/v) fetal bovine serum (FBS, Invitrogen). HEK-293T cells were maintained in the
277 same medium supplemented with 1% (v/v) non-essential amino acids. All cell lines were
278 supplemented with 5 units/mL penicillin and 5 µg/mL streptomycin (Life Technologies) and
279 maintained in a humidified atmosphere at 37 °C with 5% CO₂.

280 The Zika virus strain used in this study is the prototype African MR-766 strain, derived from a
281 previously described infectious clone⁶². This clone was used as a template for site-directed
282 mutagenesis for the generation of mutant viruses containing the mutations L513V (codon TTG
283 changed to GTA) and Q514P (codon CAA changed to CCG) in the NS5 polymerase. The DVG
284 clone used has previously been described²³. WT and DVG clones were barcoded by introducing the
285 following mutations for the WT or DVG clones, respectively; G2298C, C2307A, T2319C, A2328T
286 and A2337G (WT) or C9189C, G9195A, T9202G, G9204A, A9284C and A9216T (DVG).

287 Rescue of Zika virus was carried out by transfection of the infectious clone in HEK-293T cells
288 (TransIT-LT1, Mirus Bio) according to the manufacturer's instructions. 3-4 days post-transfection
289 (p.t.), the cell culture supernatant was clarified by centrifugation. Virus stocks (passage 1) were then
290 prepared in Vero cells, by infecting the cells with a multiplicity of infection (MOI) of 0.01
291 PFU/cell. 5-6 days post infection (p.i.) the cell culture supernatant was clarified by centrifugation
292 and virus titres determined by plaque assays in Vero-E6 cells.

293 **Ribavirin passages**

294 Isolation of variants by ribavirin passages was carried out as follows. Vero cells were treated with
295 vehicle (DMSO)- or ribavirin- (400 µM) containing media 3 hours prior to infection. The cells were
296 then inoculated with Zika virus in triplicate at an MOI of 0.1 PFU/cell in 12-well plates. Following
297 a one-hour absorption period, the inoculum was removed, the cells washed with PBS three times,
298 and vehicle- or ribavirin- containing media replaced. 3 days p.i. the cell culture supernatant was
299 clarified by centrifugation, virus titre quantified by plaque assay in Vero-E6 cells and used for the
300 next passage.

301 **Viral RNA sequencing**

302 RNA from clarified cell culture supernatants was extracted using TRI reagent (Sigma). Purified
303 viral RNA was quantified with a Qubit fluorometer (Invitrogen) and used for sequencing by next-
304 generation RNA sequencing. Libraries were prepared using the NEBNext Ultra II RNA Library
305 preparation kit for Illumina (New England Biolabs) and loaded in a NextSeq500/550 Mid output kit
306 v2.5 (Illumina), for sequencing in a NextSeq500 (151 cycles, 8 nucleotides of index).

307 **Sequence data processing and analysis**

308 Reads were first demultiplexed using the bcl2fastq conversion software (Illumina). Next, the reads
309 were adapter- and quality-trimmed using BBDuk and optical duplicates removed using Clumpify.
310 Both BBDuk and Clumpify are part of the BBTools and BBDuk suite
311 (<https://sourceforge.net/projects/bbmap/>). Reads were then aligned to the reference genome using
312 BBMap and deletions called with the BBMap variant caller CallVariants.

313 **Competition fitness assays**

314 WT virus or L513V variants were mixed in a 1:1 ratio and used to inoculate Vero cells at 0.1
315 PFU/cell over three passages, in triplicate. RNA in the supernatant at each passage was extracted
316 using TRI reagent (Sigma), and used to amplify by RT-PCR a fragment flanking the position 513 in
317 the viral polymerase NS5. The PCR products were sequenced by Sanger sequencing and the
318 abundance of either WT or L513V estimated as the height of the peaks in sequencing
319 chromatograms, corresponding to the third nucleotide of the polymerase 513 codon encoding WT
320 (TTG) or L513V (GTA), using the QSVanalyser software (<http://dna.leeds.ac.uk/qsv/>).

321 **Passaging of virus populations enriched with barcoded DVG**

322 Three independent populations of barcoded WT virus only or distinctly barcoded WT virus and
323 DVGs were generated as follows. HEK-293T cells were transfected with the barcoded WT virus
324 clone alone, or in a 1:1 molar ratio with the barcoded DVG (TransIT-LT1, Mirus Bio) according to
325 the manufacturer's instructions. Three days p.t., the cell culture supernatant was clarified by
326 centrifugation and viral titres determined by plaque assay. Each independent population (p0) was
327 then used to inoculate Vero cells (p1) pre-treated for three hours with vehicle (DMSO)- or ribavirin-
328 (100 µM) containing media. Infections were carried out at a low (0.01 PFU/cell) or high (5
329 PFU/cell) MOI and media containing DMSO or ribavirin replaced following a one-hour absorption
330 period. 3 days p.i. the cell culture supernatant was clarified by centrifugation, and used for blind
331 infections in three subsequent passages using a low volume (3 µL) or high volume (300 µL) in cells
332 seeded in 12-well plates.

333 **Ribavirin sensitivity assays**

334 Vero cells were pre-treated for three hours with media containing vehicle DMSO, or 200 µM or 400
335 µM ribavirin (Sigma). Following this treatment, the media was removed and cells inoculated with
336 the indicated virus population (p0 or p4 virus, or engineered virus) at an MOI of 0.1 PFU/cell for 1h
337 at 37 °C. 24h post infection, virus was harvested, and virus titres determined by plaque assay and
338 the percentage of infectious virus surviving treatment used to assess virus sensitivity to ribavirin
339 treatment.

340 **RNA fluorescence *in situ* hybridization (FISH) virion assays**

341 To distinguish between WT and DVG RNA, the DVG genome carrying the L513V mutation was
342 engineered to carry 160 synonymous mutations within an 800 nt region in the viral polymerase.
343 Rescue of a WT virus population or a mixed WT and DVG population was carried out by
344 transfection of the WT infectious clone alone, or in a 1:1 ratio with the barcoded DVG clone in
345 HEK-293T cells (TransIT-LT1, Mirus Bio) according to the manufacturer's instructions. 4 days
346 post-transfection (p.t.), the cell culture supernatant was clarified by centrifugation, and used to
347 generate virus stocks in Vero cells. The virus stocks were then concentrated 10x using Amicon
348 Ultra-15 Centrifugal units 30K (Merck Millipore), and incubated for 6h in the CultureWell 16
349 chambered Coverglass wells at 37°C to allow virion binding in the negatively charged glass surface.
350 After the incubation period, bound virions were fixed in 4% PFA, washed with PBS and
351 permeabilized with PBS, 0,5% Triton X-100 for 10 min at RT, blocked with PBSTA for 30 min and
352 stained with anti-Env Mab 4G2 in PBSTA for 1 hour and secondary antibodies Alexa-Fluor 647
353 (Life Technologies) for 30 min. After immunostaining, the WT and DVG RNA were detected
354 following the manufacturer's protocol (ViewRNA ISH Cells Assays) using probe sets designed by
355 Affymetrix. The WT RNA probe set targets a region between position nt 109 and 983 of the ZIKV
356 genome, lacking in the DVG RNA due to a deletion. The DVG RNA probe set recognizes the
357 mutagenized region in the viral polymerase (approximately in the 8800-9600 nt region equivalent to
358 the WT sequence).

359 **Image acquisition and analysis**

360 Images of immobilized virions and transfected cells were acquired with a Zeiss LSM 720 laser
361 scanning confocal microscope equipped with a X100 objective. smFISH and IF images were
362 analyzed with our Python analysis package FISH-quant⁶³. In short, images were first filtered with a
363 LoG filter to enhance spots and reduce background. Individual spots were then identified with a
364 user-defined intensity threshold with a local maximum detection. Co-localization analysis between
365 spots in the three channels was performed with a custom-written Python script. Channels were first
366 analyzed pairwise with the Hungarian algorithm (SciPy function `optimize.linear_sum_assignment`).
367 To consider spherical aberration, two spots were considered to be co-localized when they are closer
368 than 700nm. In a post-processing step, the results from the three pair-wise co-localization analysis
369 were fused and the percentage of matched pairs and triplets reported.

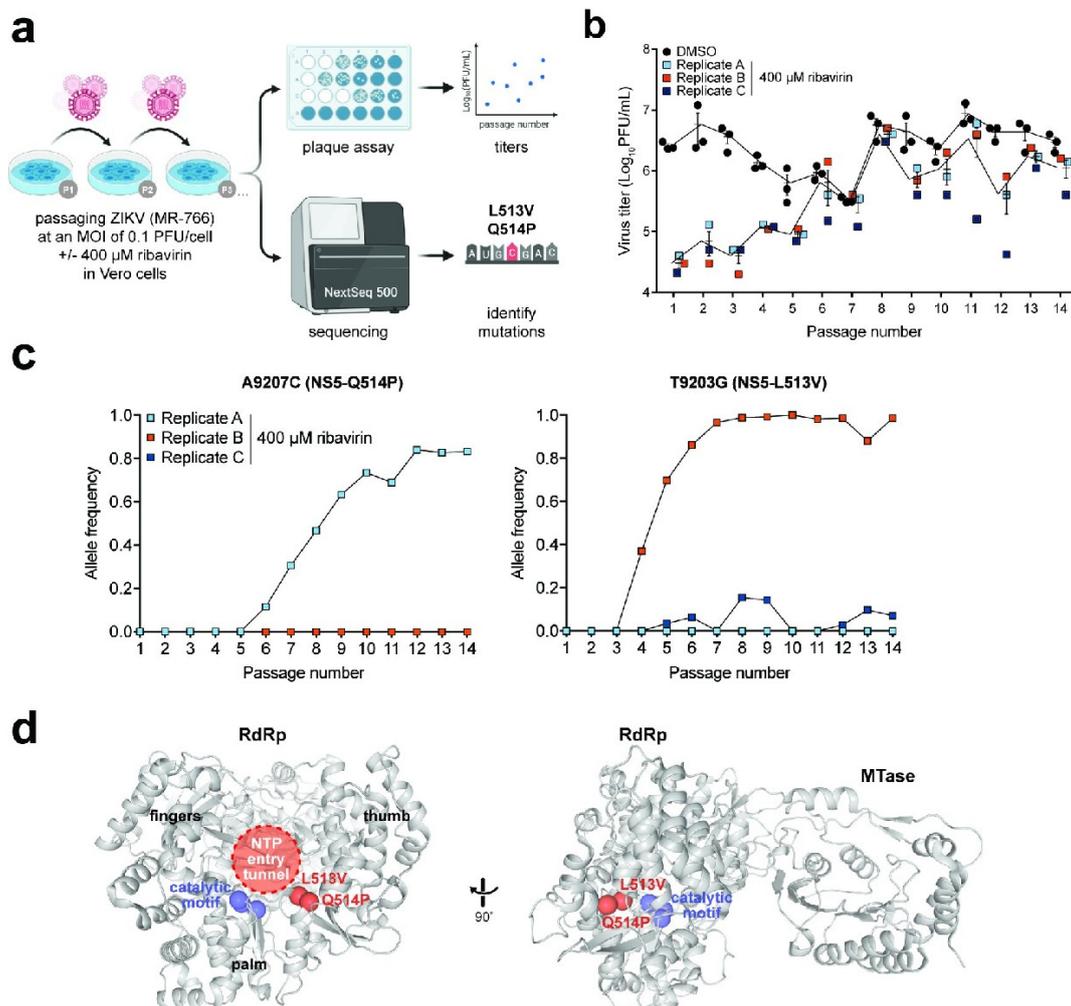
370 **Statistical analyses**

371 All data were analyzed using Prism 7 software (GraphPad) and are presented as means \pm standard
372 deviation (SD) unless indicated otherwise. For highly skewed data, statistics were carried out using

373 log-transformed data to improve the homoscedasticity of the data. The chosen method of statistical
 374 analysis is described in each figure legend. No statistical significance is indicated if p is >0.05 .

375

376 **FIGURES**



377

378 **Figure 1. Identification of ribavirin resistant mutations in Zika virus genome. (a)** Experimental

379 design for the isolation and identification of Zika virus ribavirin resistance mutations in Vero cells.

380 **(b)** Viral titres in fourteen passages performed with or without ribavirin (n=3).

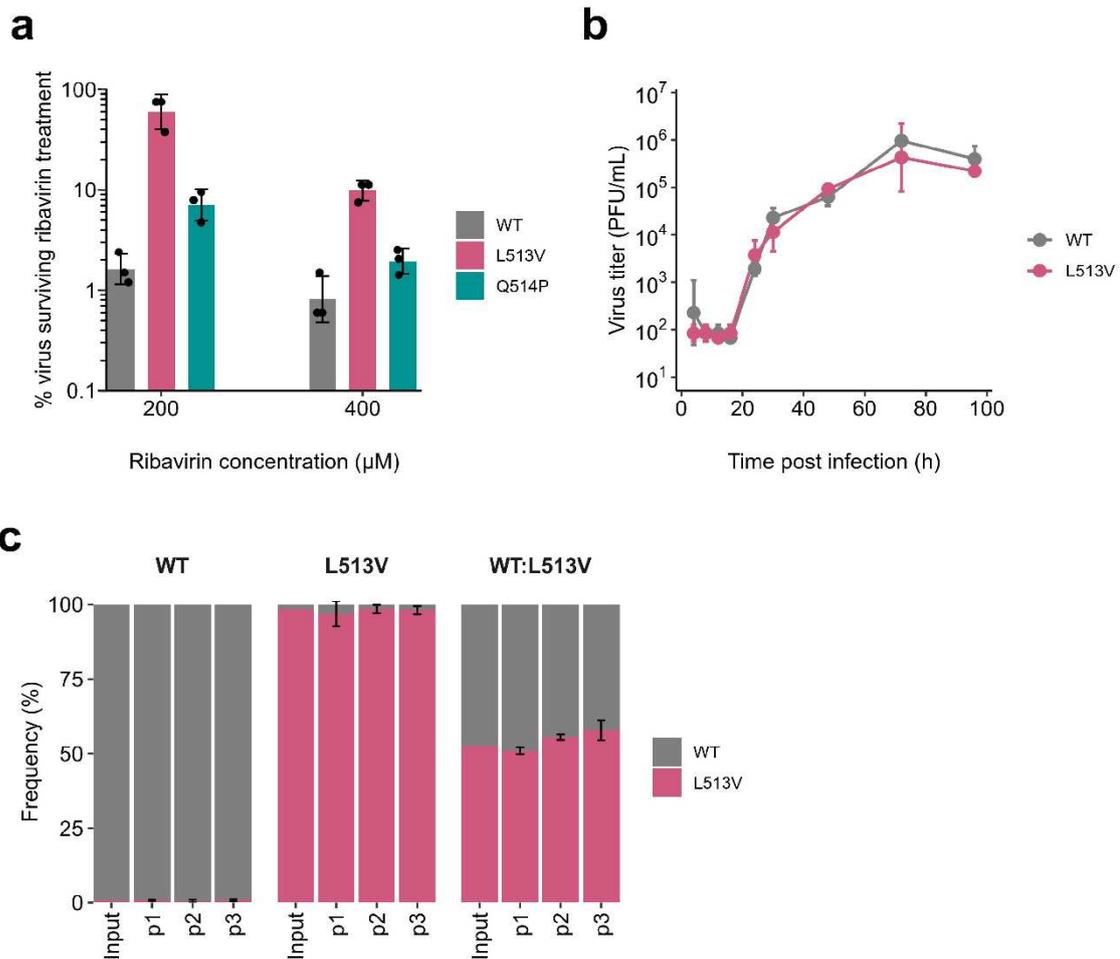
381 **(c)** Allele frequencies of two identified mutations for which frequencies throughout the passage series

382 increased; A9207C (resulting in NS5-Q514P, left panel) and T9203G (resulting in NS5-L513V,

383 right panel).

384 **(d)** Location of residues 514 and 513 in the Zika virus polymerase NS5 structure (PDB entry 5U05), relative to the priming loop and the catalytic motif in the active site.

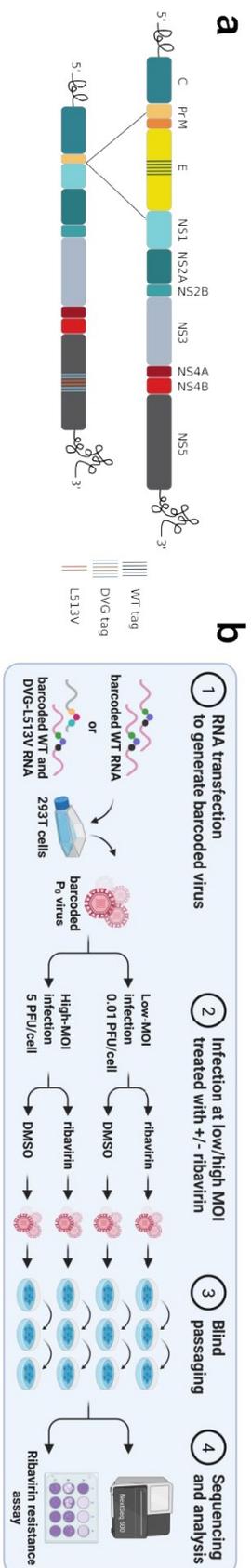
385



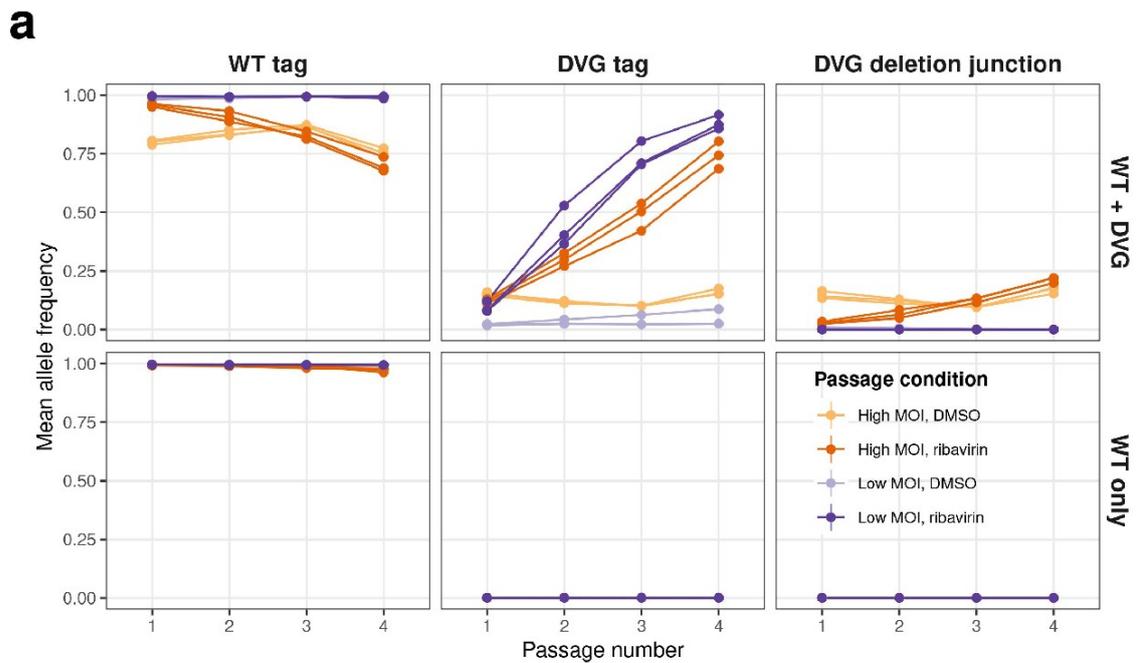
386

387 **Figure 2. Zika virus NS5 L513V confers ribavirin resistance.** (a) Ribavirin resistance assays
 388 with viruses encoding L513V or Q514P NS5 protein variants. Vero cells, which were pre-treated
 389 with media containing vehicle DMSO or 200 μM or 400 μM ribavirin, were infected with Zika
 390 virus (MOI 0.1 PFU/cell). 24h post infection, virus was harvested and virus titres determined by
 391 plaque assay. The percentage of infectious progeny virus surviving treatment, relative to titres in
 392 control DMSO-treated cells is shown (mean values (n=3) ± SD, *p at least <0.05, two-way ANOVA
 393 with Dunnet's post-test). (b) Replication kinetics of WT and L513V mutant virus in Vero cells
 394 (MOI 0.1 PFU/cell). Mean values (n=3) ± SD are shown at each indicated time point. (c)
 395 Competition assays comparing relative fitness of WT virus and L513V. The variants were mixed in
 396 a 1:1 ratio and used to inoculate Vero cells at 0.1 PFU/cell over three passages. RNA in the
 397 supernatant at each passage was used to amplify a fragment flanking the position 513 in the viral
 398 polymerase NS5. The abundance of either WT or L513V was estimated by Sanger sequencing, as
 399 the height of the peaks in sequencing chromatograms). Mean values (n=3) ± SD are shown.

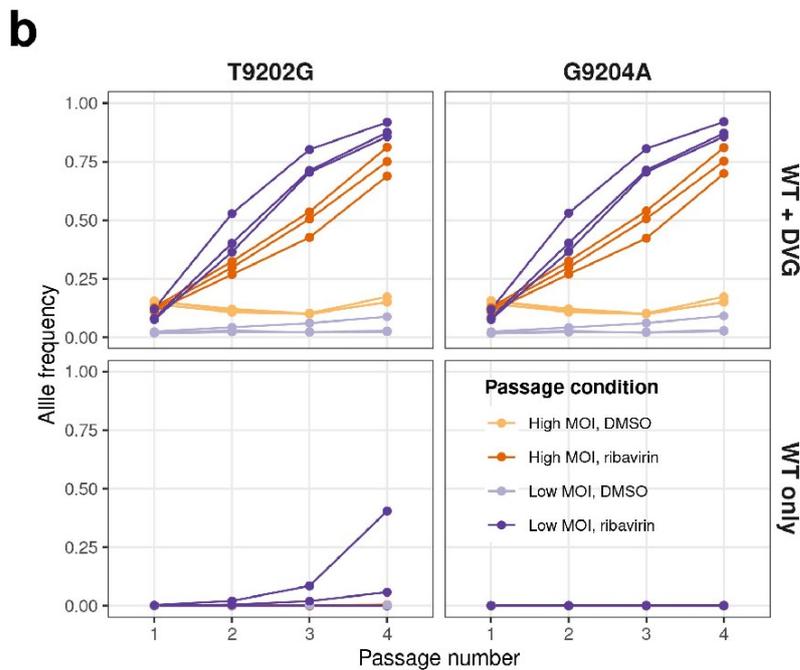
15



401 **Figure 3. Introducing the NS5 L513V mutation in DVGs results in rapid acquisition of**
402 **ribavirin resistance for viral populations under mutagenic conditions. (a)** Illustration of
403 mutations introduced in WT or DVG clones to differentiate between both molecules. The WT tag
404 included five silent mutations in the Env (E) gene, whereas the DVG tag included four silent
405 mutations and two mutations that result in the L513V amino acid change in the NS5 polymerase.
406 **(b)** Experimental design to follow the evolution of barcoded WT and DVG populations under
407 normal or mutagenic conditions. Populations containing barcoded WT virus only or a mix of
408 barcoded WT and barcoded, L513V-carrying DVG, were passaged four times at low or high MOI,
409 with or without ribavirin (100 μ M). Each virus population was phenotypically assayed for ribavirin
410 resistance and sequenced by NGS to follow the frequency of each tag. **(c)** Ribavirin resistance
411 assays with non-passaged viruses (p0) and the respective virus population passaged four times (p4)
412 with (+R) or without ribavirin (-R). p0 and p4 viruses were used to infect (MOI 0.1 PFU/cell) Vero
413 cells, which were pre-treated with media containing vehicle DMSO or 200 μ M or 400 μ M ribavirin.
414 At 24h post infection, viruses were harvested and virus titres determined by plaque assays. The
415 percentage of infectious progeny virus surviving treatment, relative to titres in control DMSO-
416 treated cells is shown on a square-root-transformed y-axis (mean values (n=3) \pm SD, *P at
417 least<0.05, two-way ANOVA with Dunnet's post-test).

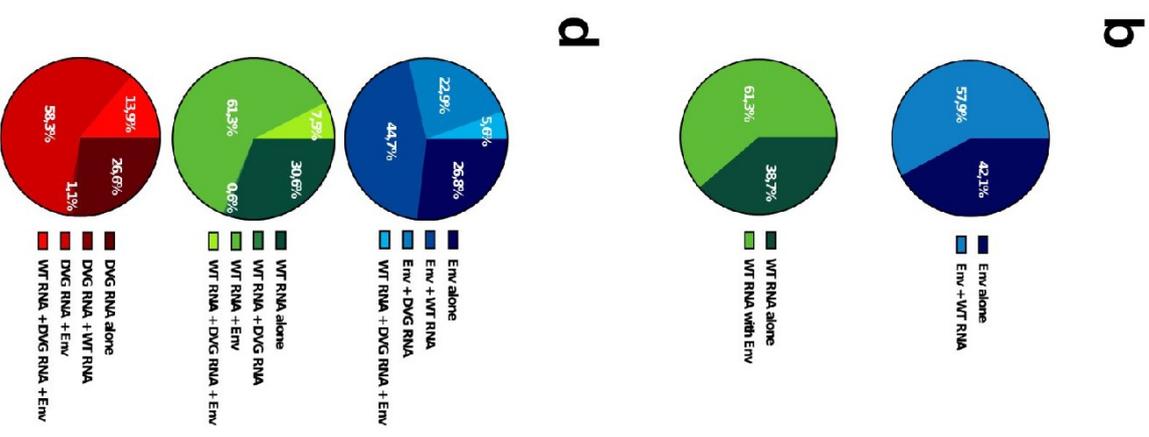
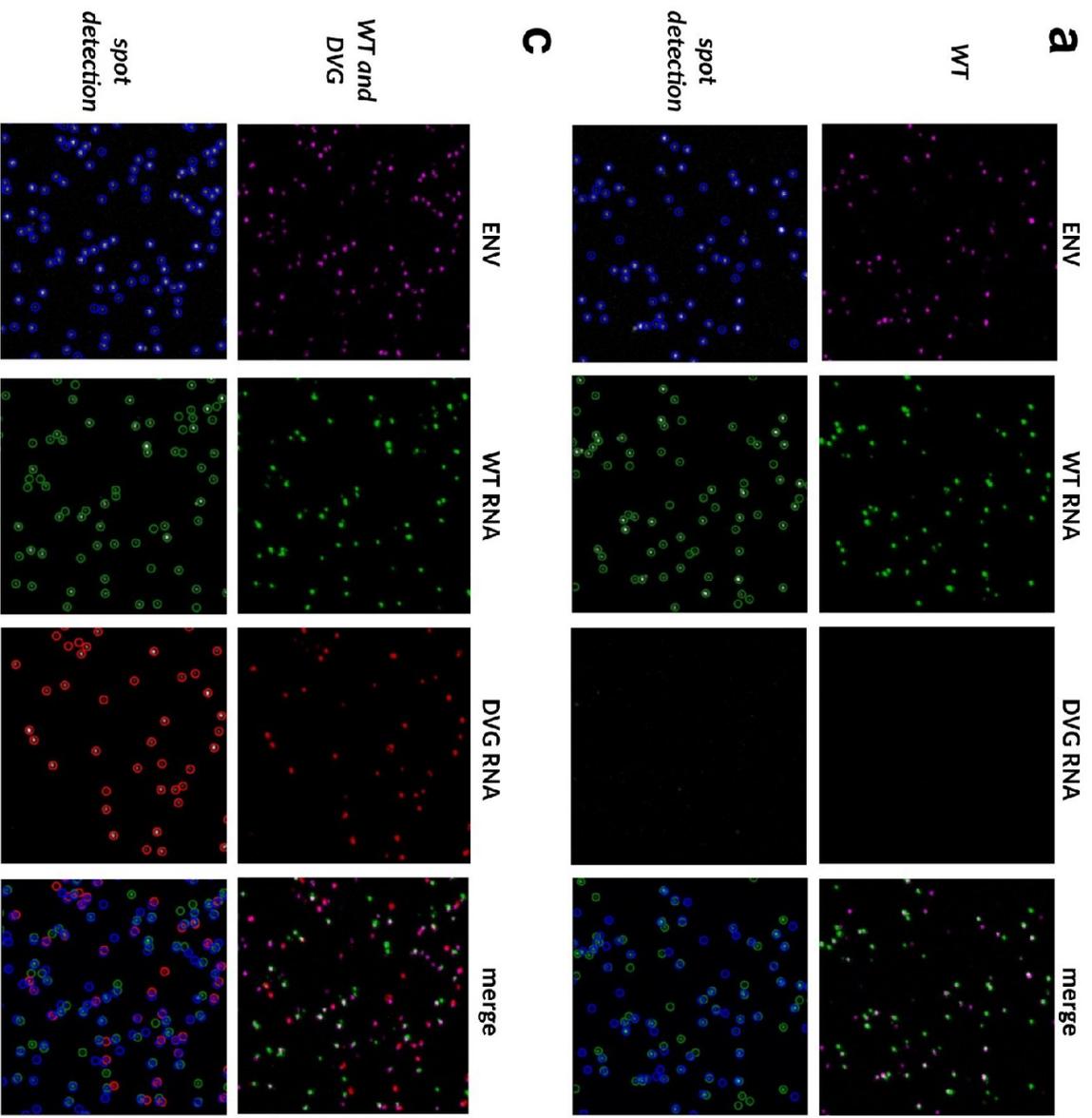


418



419

420 **Figure 4. L513V is rapidly acquired by WT genomes through recombination.** (a) Mean allele
 421 frequencies of the WT tag (left panels), DVG tag (middle panels) or the DVG deletion junction
 422 (right panels), across the passage series in the indicated passaging conditions. The mean allele
 423 frequency of each tag was derived from the frequency of each mutation in the corresponding tag
 424 (mean \pm SD, n=4 for WT tag, and n=6 for DVG tag). The frequency of three replicate populations is
 425 shown for each condition. (b) Allele frequencies of mutations that lead to the L513V change
 426 (T9201G and G9204A) throughout the passage series.



428 **Figure 5. WT and DVG genomes can be co-packaged in virions released from cells.**

429 Immobilized WT ZIKV virions (**a**) or WT and mutated DVG virions (**c**) were stained with
430 antibodies specific for viral Env protein (magenta) and processed for FISH using probes specific for
431 ZIKV WT RNA genome (green) and for mutated DVG RNA genome (red). Detected spots are
432 depicted with circles. (**b, d**) Quantification of the content of the virions. For every channel,
433 colocalization with the other two signals was evaluated. For WT ZIKV virions, 2486 Env and 2349
434 WT RNA spots were analyzed. For WT and DVG virions, 3834 Env, 2795 WT RNA and 1505
435 DVG were analyzed.

436

437 **ACKNOWLEDGMENTS**

438 We thank Matthew J. Evans (Icahn School of Medicine at Mount Sinai) for the MR766 Zika virus
439 clone, P. Desprès (Université de la Réunion, PIMIT) for 4G2 hybridoma cells, and the Région Île-
440 de-France (program DIM1Health) for the use of the Institut Pasteur imaging facility.

441 This work was funded by the DARPA INTERCEPT program managed by Dr. Jim Gimlett, Dr.
442 Brad Ringiesen, and Dr. Seth Cohen and administered through DARPA Cooperative Agreement
443 #HR0011-17-2-0023 (the content of the information does not necessarily reflect the position or the
444 policy of the U.S. government, and no official endorsement should be inferred). This work also
445 received funding from the Fondation pour la Recherche Médicale grant FRM EQU201903007777,
446 and the Agence Nationale de Recherche Laboratoire d'Excellence grant ANR-10-LABX-62-IBEID.

447

448

449 **Disclosure of Potential Competing Interests.** The Institut Pasteur has filed a U.S. Provisional
450 Application No. 63/000,998, Filed: March 27, 2020, entitled DEFECTIVE INTERFERING VIRAL
451 GENOMES, with V.V.R., D.S., T.V. and M.V. listed as inventors. The patent application covers the
452 experimental and computational methodology for identifying the best candidate TIPs for a given
453 virus, including the Zika virus TIP (DVG) described in this paper. All other authors declare no
454 competing interests.

455

456 **Correspondence** and requests for materials should be addressed to VVR (verorezelj@gmail.com),
457 NJ (nolwenn.jouvenet@pasteur.fr) or MV (marco.vignuzzi@pasteur.fr).

458

1. Holmes EC. *The Evolution and Emergence of RNA Viruses*. Oxford University Press; 2009.
2. Duffy S. Why are RNA virus mutation rates so damn high? *PLOS Biol*. 2018;16(8):e3000003. doi:10.1371/journal.pbio.3000003
3. Vignuzzi M, Andino R. Closing the gap: the challenges in converging theoretical, computational, experimental and real-life studies in virus evolution. *Curr Opin Virol*. 2012;2(5):515-518. doi:10.1016/j.coviro.2012.09.008
4. Cui J, Han N, Streicker D, et al. Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis*. 2007;13(10):1526-1532. doi:10.3201/eid1310.070448
5. Wang L, Su S, Bi Y, Wong G, Gao GF. Bat-origin coronaviruses expand their host range to pigs. *Trends Microbiol*. 2018;26(6):466-470. doi:10.1016/j.tim.2018.03.001
6. Xiao Y, Rouzine IM, Bianco S, et al. RNA recombination enhances adaptability and is required for virus spread and virulence. *Cell Host Microbe*. 2017;22(3):420. doi:10.1016/j.chom.2017.08.006
7. Suarez DL, Senne DA, Banks J, et al. Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerg Infect Dis*. 2004;10(4):693-699. doi:10.3201/eid1004.030396
8. Khatchikian D, Orlich M, Rott R. Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the haemagglutinin gene of an influenza virus. *Nature*. 1989;340(6229):156-157. doi:10.1038/340156a0
9. Filomatori CV, Bardossy ES, Merwaiss F, et al. RNA recombination at Chikungunya virus 3'UTR as an evolutionary mechanism that provides adaptability. Ebel GD, ed. *PLOS Pathog*. 2019;15(4):e1007706. doi:10.1371/journal.ppat.1007706
10. Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol*. 2007;81(14):7620-7628. doi:10.1128/jvi.00083-07
11. Gu Z, Gao Q, Faust EA, Wainberg MA. Possible involvement of cell fusion and viral recombination in generation of human immunodeficiency virus variants that display dual resistance to AZT and 3TC. *J Gen Virol*. 1995;76(10):2601-2605. doi:10.1099/0022-1317-76-10-2601
12. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol*. 2011;9(8):617-626. doi:10.1038/nrmicro2614
13. Langsjoen RM, Muruato AE, Kunkel SR, Jaworski E, Routh A. Differential alphavirus defective RNA diversity between intracellular and extracellular compartments is driven by subgenomic recombination events. Hardy RW, Griffin DE, eds. *mBio*. 2020;11(4). doi:10.1128/mbio.00731-20
14. Jaworski E, Routh A. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. Lauring AS, ed. *PLOS Pathog*. 2017;13(5):e1006365. doi:10.1371/journal.ppat.1006365
15. Martin MA, Kaul D, Tan GS, Woods CW, Koelle K. The dynamics of influenza A H3N2 defective viral genomes from a human challenge study. Published online October 2019. doi:10.1101/814673
16. Vasilijevic J, Zamarreño N, Oliveros JC, et al. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. Thomas PG, ed. *PLOS Pathog*. 2017;13(10):e1006650. doi:10.1371/journal.ppat.1006650

17. Sun Y, Jain D, Koziol-White CJ, et al. Immunostimulatory defective viral genomes from respiratory syncytial virus promote a strong innate antiviral response during infection in mice and humans. Thomas PG, ed. *PLoS Pathog*. 2015;11(9):e1005122. doi:10.1371/journal.ppat.1005122
18. Karamichali E, Chihab H, Kakkanas A, et al. HCV defective genomes promote persistent infection by modulating the viral life cycle. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.02942
19. Noppornpanth S, Smits SL, Lien TX, Poovorawan Y, Osterhaus ADME, Haagmans BL. Characterization of hepatitis c virus deletion mutants circulating in chronically infected patients. *J Virol*. 2007;81(22):12496-12503. doi:10.1128/jvi.01059-07
20. Li D, Lott WB, Lowry K, Jones A, Thu HM, Aaskov J. Defective interfering viral particles in acute dengue infections. Coffey LL, ed. *PLoS ONE*. 2011;6(4):e19447. doi:10.1371/journal.pone.0019447
21. Aaskov J, Buzacott K, Thu HM, Lowry K, Holmes EC. Long-term transmission of defective RNA viruses in humans and aedes mosquitoes. *Science*. 2006;311(5758):236-238. doi:10.1126/science.1115030
22. Pesko KN, Fitzpatrick KA, Ryan EM, et al. Internally deleted WNV genomes isolated from exotic birds in New Mexico: Function in cells, mosquitoes, and mice. *Virology*. 2012;427(1):10-17. doi:10.1016/j.virol.2012.01.028
23. Rezelj VV, Carrau L, Merwaiss F, et al. Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-22341-7
24. Levi LI, Rezelj VV, Henrion-Lacritick A, et al. Defective viral genomes from chikungunya virus are broad-spectrum antivirals and prevent virus dissemination in mosquitoes. Lauring AS, ed. *PLoS Pathog*. 2021;17(2):e1009110. doi:10.1371/journal.ppat.1009110
25. Meir M, Harel N, Miller D, et al. Competition between social cheater viruses is driven by mechanistically different cheating strategies. *Sci Adv*. 2020;6(34):eabb7990. doi:10.1126/sciadv.abb7990
26. Taylor DR, Jarosz AM, Fulbright DW, Lenski RE. The acquisition of hypovirulence in host-pathogen systems with three trophic levels. *Am Nat*. 1998;151(4):343-355. doi:10.1086/286123
27. Yoon SW, Lee SY, Won SY, Park SH, Park SY, Jeong YS. Characterization of homologous defective interfering RNA during persistent infection of Vero cells with Japanese encephalitis virus. *Mol Cells*. 2006;21(1):112-120.
28. Manzoni TB, López CB. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. *Future Virol*. 2018;13(7):493-503. doi:10.2217/fvl-2018-0021
29. De BK, Nayak DP. Defective interfering influenza viruses and host cells: establishment and maintenance of persistent influenza virus infection in MDBK and HeLa cells. *J Virol*. 1980;36(3):847-859. doi:10.1128/JVI.36.3.847-859.1980
30. Schmaljohn C, Blair CD. Persistent infection of cultured mammalian cells by Japanese encephalitis virus. *J Virol*. 1977;24(2):580-589. doi:10.1128/JVI.24.2.580-589.1977
31. Pfeiffer JK, Kirkegaard K. Ribavirin resistance in hepatitis c virus replicon-containing cell lines conferred by changes in the cell line or mutations in the replicon RNA. *J Virol*. 2005;79(4):2346-2355. doi:10.1128/jvi.79.4.2346-2355.2005
32. Pfeiffer JK, Kirkegaard K. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc Natl Acad Sci*. 2003;100(12):7289-7294. doi:10.1073/pnas.1232294100

33. Scheidel LM, Durbin RK, Stollar V. Sindbis virus mutants resistant to mycophenolic acid and ribavirin. *Virology*. 1987;158(1):1-7. doi:10.1016/0042-6822(87)90230-3
34. Young KC, Lindsay KL, Lee KJ, et al. Identification of a Ribavirin-resistant NS5B mutation of hepatitis C virus during ribavirin monotherapy. *Hepatol Baltim Md*. 2003;38(4):869-878. doi:10.1002/hep.1840380413
35. Zeng J, Wang H, Xie X, Yang D, Zhou G, Yu L. An increased replication fidelity mutant of foot-and-mouth disease virus retains fitness in vitro and virulence in vivo. *Antiviral Res*. 2013;100(1):1-7. doi:10.1016/j.antiviral.2013.07.008
36. Arnold JJ, Vignuzzi M, Stone JK, Andino R, Cameron CE. Remote site control of an active site fidelity checkpoint in a ViralRNA-dependent RNAPolymerase. *J Biol Chem*. 2005;280(27):25706-25716. doi:10.1074/jbc.m503444200
37. Coffey LL, Beeharry Y, Borderia AV, Blanc H, Vignuzzi M. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc Natl Acad Sci*. 2011;108(38):16038-16043. doi:10.1073/pnas.1111650108
38. Zhao B, Yi G, Du F, et al. Structure and function of the Zika virus full-length NS5 protein. *Nat Commun*. 2017;8(1). doi:10.1038/ncomms14762
39. Pincus S, Mason PW, Konishi E, et al. Recombinant vaccinia virus producing the prM and E proteins of yellow fever virus protects mice from lethal yellow fever encephalitis. *Virology*. 1992;187(1):290-297. doi:10.1016/0042-6822(92)90317-i
40. Mason PW, Pincus S, Fournier MJ, Mason TL, Shope RE, Paoletti E. Japanese encephalitis virus-vaccinia recombinants produce particulate forms of the structural membrane proteins and induce high levels of protection against lethal JEV infection. *Virology*. 1991;180(1):294-305. doi:10.1016/0042-6822(91)90034-9
41. Hunt AR, Cropp CB, Chang GJJ. A recombinant particulate antigen of Japanese encephalitis virus produced in stably-transformed cells is an effective noninfectious antigen and subunit immunogen. *J Virol Methods*. 2001;97(1-2):133-149. doi:10.1016/s0166-0934(01)00346-9
42. Davis BS, Chang GJJ, Cropp B, et al. West Nile virus recombinant DNA vaccine protects mouse and horse from virus challenge and expresses in vitro a noninfectious recombinant antigen that can be used in enzyme-linked immunosorbent assays. *J Virol*. 2001;75(9):4040-4047. doi:10.1128/jvi.75.9.4040-4047.2001
43. Ramakrishnaiah V, Thumann C, Fofana I, et al. Exosome-mediated transmission of hepatitis C virus between human hepatoma Huh7.5 cells. *Proc Natl Acad Sci*. 2013;110(32):13109-13113. doi:10.1073/pnas.1221899110
44. Dreux M, Garaigorta U, Boyd B, et al. Short-range exosomal transfer of viral RNA from infected cells to plasmacytoid dendritic cells triggers innate immunity. *Cell Host Microbe*. 2012;12(4):558-570. doi:10.1016/j.chom.2012.08.010
45. Lui WY, Yuen CK, Li C, et al. SMRT sequencing revealed the diversity and characteristics of defective interfering RNAs in influenza A (H7N9) virus infection. *Emerg Microbes Infect*. 2019;8(1):662-674. doi:10.1080/22221751.2019.1611346
46. Cattaneo R, Schmid A, Eschle D, Baczko K, ter Meulen V, Billeter MA. Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell*. 1988;55(2):255-265. doi:10.1016/0092-8674(88)90048-7
47. O'Hara PJ, Nichol ST, Horodyski FM, Holland JJ. Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell*. 1984;36(4):915-924. doi:10.1016/0092-8674(84)90041-2

48. Fonville JM, Marshall N, Tao H, Steel J, Lowen AC. Influenza virus reassortment is enhanced by semi-infectious particles but can be suppressed by defective interfering particles. Lauring A, ed. *PLoS Pathog.* 2015;11(10):e1005204. doi:10.1371/journal.ppat.1005204
49. Yount JS, Gitlin L, Moran TM, López CB. MDA5 participates in the detection of paramyxovirus infection and is essential for the early activation of dendritic cells in response to sendai virus defective interfering particles. *J Immunol.* 2008;180(7):4910-4918. doi:10.4049/jimmunol.180.7.4910
50. Tuffereau C, Roux L. Direct adverse effects of sendai virus DI particles on virus budding and on M protein fate and stability. *Virology.* 1988;162(2):417-426. doi:10.1016/0042-6822(88)90482-5
51. Tapia K, Kim W keun, Sun Y, et al. Defective viral genomes arising in vivo provide critical danger signals for the triggering of lung antiviral immunity. Iwasaki A, ed. *PLoS Pathog.* 2013;9(10):e1003703. doi:10.1371/journal.ppat.1003703
52. Erickson AK, Jesudhasan PR, Mayer MJ, Narbad A, Winter SE, Pfeiffer JK. Bacteria facilitate enteric virus co-infection of mammalian cells and promote genetic recombination. *Cell Host Microbe.* 2018;23(1):77-88.e5. doi:10.1016/j.chom.2017.11.007
53. Nagashima S, Jirintai S, Takahashi M, et al. Hepatitis E virus egress depends on the exosomal pathway, with secretory exosomes derived from multivesicular bodies. *J Gen Virol.* 2014;95(10):2166-2175. doi:10.1099/vir.0.066910-0
54. Gutiérrez S, Pirolles E, Yvon M, Baecker V, Michalakis Y, Blanc S. The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. Simon A, ed. *J Virol.* 2015;89(18):9665-9675. doi:10.1128/jvi.00537-15
55. Andreu-Moreno I, Sanjuán R. Collective viral spread mediated by virion aggregates promotes the evolution of defective interfering particles. Racaniello VR, ed. *mBio.* 2020;11(1). doi:10.1128/mbio.02156-19
56. Lawrence DMP, Patterson CE, Gales TL, D'Orazio JL, Vaughn MM, Rall GF. Measles virus spread between neurons requires cell contact but not CD46 expression, syncytium formation, or extracellular virus production. *J Virol.* 2000;74(4):1908-1918. doi:10.1128/jvi.74.4.1908-1918.2000
57. Sherer NM, Lehmann MJ, Jimenez-Soto LF, Horensavitz C, Pypaert M, Mothes W. Retroviruses can establish filopodial bridges for efficient cell-to-cell transmission. *Nat Cell Biol.* 2007;9(3):310-315. doi:10.1038/ncb1544
58. Cuevas JM, Durán-Moreno M, Sanjuán R. Multi-virion infectious units arise from free viral particles in an enveloped virus. *Nat Microbiol.* 2017;2(7). doi:10.1038/nmicrobiol.2017.78
59. Simón O, Williams T, Caballero P, López-Ferber M. Dynamics of deletion genotypes in an experimental insect virus population. *Proc R Soc B Biol Sci.* 2005;273(1588):783-790. doi:10.1098/rspb.2005.3394
60. Sexton NR, Bellis ED, Murrieta RA, et al. Genome number and size polymorphism in zika virus infectious units. Pfeiffer JK, ed. *J Virol.* 2021;95(6). doi:10.1128/jvi.00787-20
61. Codoñer FM, Darós JA, Solé RV, Elena SF. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2006;2(12):e136. doi:10.1371/journal.ppat.0020136
62. Schwarz MC, Sourisseau M, Espino MM, et al. Rescue of the 1947 zika virus prototype strain with a cytomegalovirus promoter-driven cDNA clone. Coyne CB, ed. *mSphere.* 2016;1(5). doi:10.1128/msphere.00246-16

63. Imbert A, Ouyang W, Safieddine A, et al. FISH-quant v2: a scalable and modular analysis tool for smFISH image analysis. Published online July 2021. doi:10.1101/2021.07.20.453024

3. Complementary results

3.1. Material and methods of complementary results

3.1.1. CELLS

Vero and Vero-E6 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) containing 10% (v/v) FBS and 1% (v/v) P/S in a humidified atmosphere at 37°C with 5% CO₂. HEK-293T cells were grown in similar conditions with the addition of 1% (v/v) non-essential amino acids in the culture medium.

3.1.2. PLASMIDS

The Zika virus (ZIKV) strain used is the prototype African MR-766 strain infectious clone under a cytomegalovirus (CMV) promoter (GenBank accession number KX830960)²⁶⁹. This clone was used as a template for site-directed mutagenesis for the generation of mutant viruses containing the L513V (codon TTG changed to GTA) and Q514P (codon CAA changed to CCG) in the non-structural protein 5 (NS5) polymerase. The DVG clone used has been previously described in Rezelj *et al.*²³ and comprises a deletion of nucleotides 581 to 3250 from the wild-type (WT) ZIKV clone. Additionally, the ZIKV replicon was described in Rezelj *et al.*²³ and Xie *et al.*²⁷⁰ and consists of the replacement of the structural protein genes (capsid C, precursor membrane prM and envelope E) from the N-terminal 38 amino acids of C protein to the C-terminal 30 amino acids of protein E with a NanoLuc reporter gene followed by foot-and-mouth disease virus protease 2A (NanoLuc-2A). The DVG reporter clone was similarly designed to contain the NanoLuc-2A cassette at the deletion site. Altered DVG or replicon (NS5_{AAA}) have their NS5 polymerase catalytic motif G₆₆₄D₆₆₅D₆₆₆ modified by three successive alanine residues. Additionally, a DVG with a deleted NS5 protein (DVG-ΔNS5) was constructed by removing the entire gene at the exception of its translation-termination codon (TAA). Barcoded WT and DVG clones were obtained through site-directed mutagenesis by introducing the following silent mutations, respectively: G2298C, C2307A, T2319C, A2328T and A2337G (WT) and C9189C, G9195A, A9284C and A9216T (DVG).

All plasmid stocks were obtained through classic transformation protocols of XL-10 Gold Ultracompetent cells (Agilent) or Top10 Chemically competent *E. coli* (ThermoFisher Scientific), followed by Mirapreps²⁷¹ of colonies grown in 100 mL LB Miller using NucleoSpin Plasmids kits (Macherey Nagel). Purity and concentrations of extracted DNAs were evaluated

on a NanoDrop (ThermoFisher Scientific Scientific). Full plasmid sequences were checked by Sanger sequencing (Eurofins Genomics).

2.1.3. VIRUSES

Rescue of ZIKV was carried out using the TransIT-LT1 kit (Mirus Bio) to transfect the infectious clone in HEK-293T cells following the manufacturer's instructions. Three days post-transfection (dpt), the cell culture supernatant was clarified by centrifugation at 11 000×g for 10 minutes (min). Virus stocks (passage 1) were then obtained by infecting Vero cells at a multiplicity of infection (MOI) of 0.01 plaque-forming unit per cell (PFU/cell). Supernatants were collected and clarified at 5 days post-infection (dpi), and titers determined by plaque assays in Vero-E6 cells in triplicates.

3.1.4. SITE-DIRECTED MUTAGENESIS BY *IN VIVO* ASSEMBLY

Site-directed mutagenesis was performed by *in vivo* assembly (IVA)²⁷². Briefly, overlapping primers containing the mutations of interest (Integrative DNA Technologies) were used to amplify target DNA products through polymerase-chain reaction (PCR) using the Phusion High-Fidelity DNA Polymerase kit (ThermoFisher Scientific) for 18 cycles at annealing temperatures of 60°C. Reaction products were then digested with DpnI (ThermoFisher Scientific) for 30 min at 37°C to eliminate plasmid DNA and used without purification to transform XL-10 Gold Ultracompetent cells (Agilent) or Top10 Chemically competent *E. coli* (ThermoFisher Scientific). Each transformation was performed by adding 2 µL PCR product to 25 µL suspended bacteria, incubated 30 min on ice, heat-shocked at 42°C for 40 seconds (s), supplemented with 100 µL SOC media (ThermoFisher Scientific), then incubated in a shaking incubator at 37°C, 120 rotations per minute (rpm) for 1 hour (h). After incubation, total volume was plated on a LB agar plate with 100 µg/mL ampicillin and incubated at 37°C overnight. Successful mutagenesis were validated by randomly picking colonies to grow in 5 mL LB Miller overnight followed by plasmid DNA extraction minipreps (NucleoSpin Plasmids, Macherey Nagel) and Sanger sequencing of the target mutation site (Eurofins Genomics).

Mutation	Sense	Primer
Δ NS5	Forward	5'-CTGGTTAAGAGACGTTAAGCACCAATTTTAGTGTTGTCAGG-3'
Δ NS5	Reverse	5'-ACGTCTCTTAACCAGGCCAG-3'
NS5 _{AAA}	Forward	5'-GGCGGTCAGTGCAGCTGCCTGCGTTGTGAAGCCAATC-3'
NS5 _{AAA}	Reverse	5'-CTGCACTGACCGCCATTCGTTTGAG-3'

Table 2 | Primers for ZIKV IVA site-directed mutagenesis.

3.1.5. VIRAL TITERS

Infectious viral titers were determined by plaque assays. Briefly, monolayers of Vero-E6 cells in 24-well plates were inoculated with 150 μ L serial 10-fold dilutions of viruses in unsupplemented DMEM and incubated at 37°C and 5% CO₂ for 1h. Then, wells were gently covered with approximately 1 mL of a solid overlay comprising DMEM supplemented with 2% FBS, 1% P/S and 0.8% agarose (Invitrogen) and incubated again for 5dpi until fixation with 4% formaldehyde (Sigma). Plaques were revealed by crystal violet 0.8% staining and manually counted.

3.1.6. COMPLEMENTATION ASSAYS

DVG and replicon reporter RNAs were generated by amplifying by PCR their respective plasmid constructs with a forward primer replacing the CMV promoter with an SP6 promoter sequence at its 5'-end (5'-CCGATTTAGGTGACACTATAGAGTTGTTGATCTGTGTGAGTCAGACT-3') and a reverse primer which anneals downstream of the ZIKV 3'-UTR and the SV40 poly-A tail signal of the molecular clone (5'-CCCCTGAACCTGAAACATAAAAATGAATGCAATTGTTGTTG-3'). PCR products were purified (NucleoSpin Gel and PCR Clean-Up, Macherey Nagel) and used for *in vitro* transcription of messenger RNA (mRNA) using the SP6 mMessage mMachine kit (Invitrogen), followed by mRNA purification using the RNeasy MinElute Cleanup kit (Qiagen). Purified RNAs were quantified using the Qubit RNA BR assay kit (Invitrogen) on a Qubit 2.0 fluorometer (Invitrogen).

Sub-confluent Vero cells in 24-well plates were infected with ZIKV at MOI 1 PFU/cell, or left uninfected. After 1 h incubation at 37°C, naive and infected cells were transfected with 50 ng reporter RNAs using the TransIT-mRNA transfection kit (Mirus Bio). At 6 h, 24 h, 48 h and 72 h post-transfection (hpt), supernatants were removed and cell lysed in 1X Luciferase Cell Culture Lysis Buffer (Promega) for 5 min, and prepared Nano-Glo Luciferase Assay reagents (Promega) added to the wells and incubated for 10 min on a shaking plate. 10-fold dilutions of the reaction in 1X PBS were finally loaded onto black 96-well plate and luminescence was measured in a Tecan Infinite M200 Pro plate reader (Tecan) with an integration time of 10 seconds. Relative light units (RLUs) were normalized against total protein concentration by performing in parallel a Bradford assay (Zymo) on cell lysates.

3.1.7. DVG MUTANT DISTRIBUTION ANALYSIS

Variant frequency distributions were investigated in sequencing data of DVG-WT co-infections at passage 3, which had the highest number of aligned reads to the ZIKV MR766 reference genome (GenBank accession number KX830960).

Reads containing the 581-3250 deletion breakpoint were identified using the ViReMa (Viral Recombination Mapper) algorithm²⁷³. Briefly, reads were clipped and trimmed (phred quality score < 30 and minimum length = 50) using BBDuk, and PCR duplicates removed with Clumpify from the BBMap v38.91 toolkit (<https://sourceforge.net/projects/bbmap/>). Then, alignments against the ZIKV MR766 reference sequence and recombination sites were determined using ViReMa v0.25 (<https://sourceforge.net/projects/virema/>) with seeds of 20 nucleotides and maximum 2 mismatches allowed. Reads containing the deletion breakpoint were then filtered from the original fastq files using seqtk v.1.3 (<https://github.com/lh3/seqtk>).

Variant calling on either total viral reads or filtered breakpoint-containing reads was performed with ViVan v0.43 (Viral Variant Analysis)²⁷⁴ against respectively the ZIKV MR766 WT or DVG reference sequence. Insufficiently covered bases (< 50X) were discarded prior variant calling and a generalized likelihood-ratio test was performed to identify statistically significant variants ($p_{\text{adj}} < 0.05$ after a Benjamini-Hochberg false-discovery post-hoc test). Only SNVs with base coverage > 100 and frequencies > 0.15% (the lowest frequency successfully called in DVG filtered reads) were considered.

3.2. Results

3.2.1. MUTATIONS IN NS5 RdRp DOMAIN CANNOT BE *TRANS*-COMPLEMENTED

The ribavirin-resistance mutation introduced in the DVG did not show complementation effects at initial passages (p1) in ribavirin-treated cells (**Article Figure 3c**). However, sequencing results showed that DVGs were at relative frequencies of only 10-20% in the viral population (**Article Figure 4a**), suggesting that the mutation might have been at too low frequency to rescue viral titers. To evaluate if *trans*-complementation by the ribavirin-resistance mutation was possible, we first aimed at verifying that NS5 was itself capable of complementation. We designed DVG reporters carrying a NanoLuc gene at the deletion junction with either no additional modification (DVG-NS5_{WT}), an altered polymerase GDD catalytic motif consisting of three successive alanine residues (DVG-NS5_{AAA}), or a deleted NS5 gene (DVG- Δ NS5). Additionally, we used similar replicon reporters as described previously in Rezelj *et al.*²³ for controls, which consist of replication-capable genomes with an identical NanoLuc reporter gene which replaces deleted structural genes. The active replicon (Replicon-NS5_{WT}) serves as a positive control for genome replication, while the inactive replicon carries the same alteration to its GDD motif (Replicon-NS5_{AAA}) and has been shown to be incapable of replication even in presence of infectious virus²³ (**Figure 15A**).

While the DVG-NS5_{WT} reporter showed restored replication when transfected in infected cells, neither DVG-NS5_{AAA} nor DVG- Δ NS5 were able to replicate through complementation by full-length infectious viruses (**Figure 15B**). This indicates that the RdRp activity of NS5 cannot be *trans*-complemented, and is required in *cis* for DVGs to replicate. As a consequence, because the L513V mutation affects the polymerase, absence of viral titer rescue in ribavirin assays at early passages was due to the impossibility for this mutation to act in *trans*.

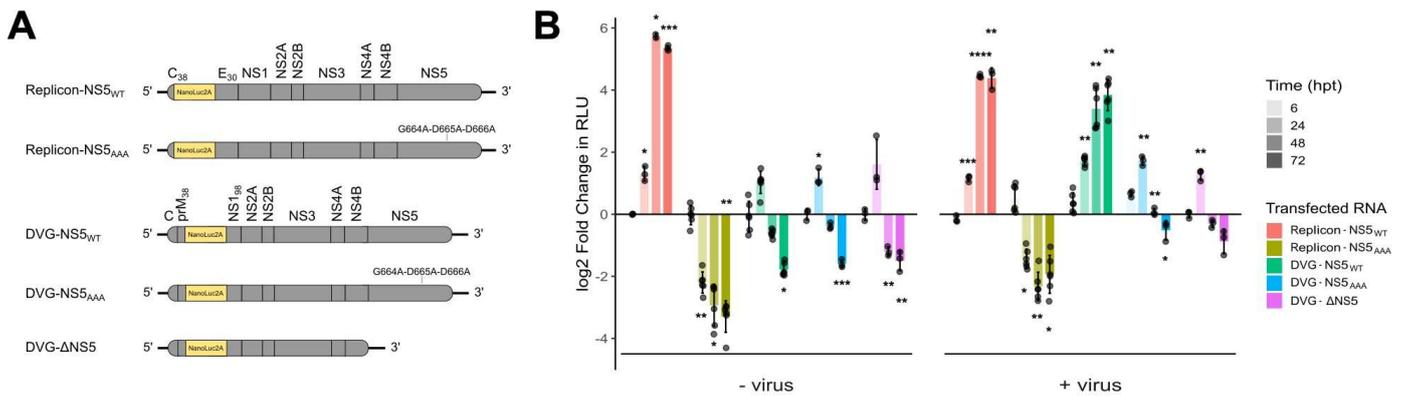


Figure 15 | NS5 polymerase activity cannot be *trans*-complemented to rescue DVG reporter replication. (A) Schematic diagram of reporter constructs. The NanoLuc-2A reporter gene (yellow) is inserted at the deletion breakpoints of the replicon or the DVG. (B) Complementation assays of DVG and replicon reporters in infected and uninfected cells. Equal molar amounts of replicon or DVG RNAs were transfected in naive or infected Vero cells (MOI 1 PFU/cell), and relative light units (RLUs) were measured and normalized at the 6, 24, 28 and 72 hours post-transfection (hpt). The y-axis shows log₂ fold changes of normalized RLUs against the mean values of the Replicon-NS5_{WT} positive control at 6 hpt in uninfected cells. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; Student *t*-test with Bonferroni post-hoc multiple comparison correction. Means and standard deviations (SDs) are shown; $n = 3$ per group.

3.2.2. DVGs SHOW SIGNS OF RELAXED PURIFYING SELECTION

Previous studies have shown that some DVGs showed extensive mutagenic RNA editing, resulting in hypermutation profiles^{216,226,244,262}. This increased genetic diversity could in theory benefit standard WT genomes through recombination similarly to the L513V we introduced. However, it is still unclear if these hypermutated DVGs are the result of preferential RNA editing of DVGs, or if DVGs can tolerate deleterious or lethal mutations through complementation by standard genomes. To provide insights on the question, we looked at the distribution of SNV frequencies in DVGs, and compared it to the distribution in total viral genomes. Because we only had short-read Illumina sequencing data, only reads containing the 581-3250 deletion could be attributed with certainty to DVGs. We therefore used ViReMa²⁷³ to identify reads containing the 581-3250 deletion breakpoint and ran ViVan²⁷⁴ to call variants on those selected reads, which typically covered with sufficient depth (> 100X) a 300-nucleotide window around the breakpoint. Conversely, we called SNVs on total reads and selected variants which corresponded to the same genomic region as the DVG window. While total viral genomes showed higher amounts (275) of called SNVs than DVGs (51), they are in their majority (89%) situated at frequencies below 0.5% (**Figure 16A**). In

DVGs, mutation frequencies span a wider distribution, with 59% of them located between 0.5% and 2.5% frequencies (**Figure 16B**). These results indicate that while our DVG do not necessarily mutate more than standard genomes, mutations seems less subject to purifying selection, allowing them to reach higher frequencies on average (0.75% against 0.31%; $p=6.74e-8$) through genetic drift. Finally, inspection of relative frequencies of RNA deamination editing substitutions caused by ADAR1 (namely A>G and U>C) in DVG or total genome reads show relative increases, but requires bigger datasets to support preferential editing of this specific species of DVG (respectively, $p=0.558$ and $p=0.4339$ with Pearson's χ^2 -test) (**Figure 16C**).

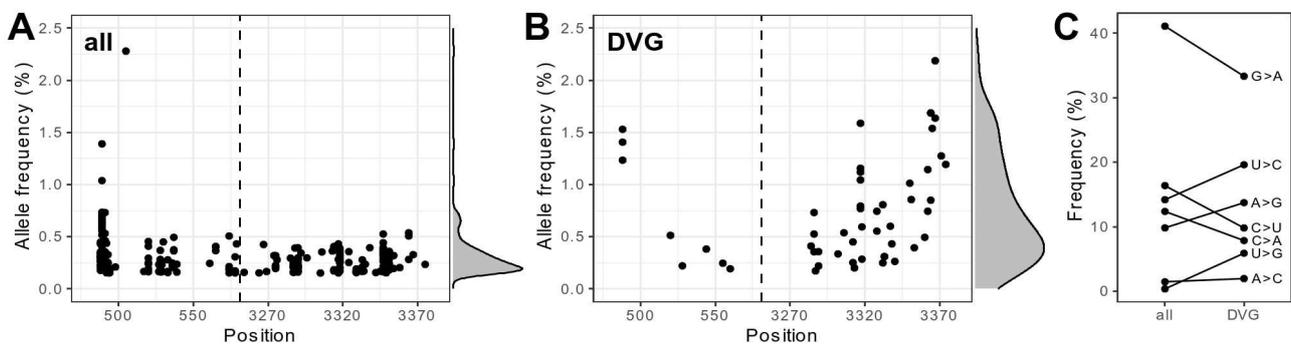


Figure 16 | Mutations in DVGs can reach higher frequencies than in standard genomes. Allele frequencies of called variants in all reads (**A**) or in DVG reads (**B**) spanning a 300-nucleotide window around the deletion breakpoints (dashed vertical line). Each point represent a SNV. Marginal curves on the right of each panel represent density curves of variant counts in regards to allele frequencies. $p=6.74e-8$; Student t -test. (**C**) Relative frequencies of nucleotide substitution.

4. Discussion and perspectives

4.1. Summary of results

We successfully identified a ribavirin-resistance mutation resulting in a L513V residue substitution in the NS5 polymerase through a classical experimental evolution approach. After introduction of this mutation in a barcoded DVG, we challenged viral populations of either WT genomes alone or a mix of WT genomes and mutated DVGs through successive passages in presence of ribavirin. We demonstrated that the presence of the resistance mutation in DVGs allowed faster adaptation of the entire viral population to the mutagenic conditions. We showed that this phenotype was the result of genetic recombination between WT genomes and DVGs, suggesting that DVGs can serve as a pool of genetic diversity for standard genomes. While we did not observe higher numbers of mutations in DVG-associated reads than in total viral reads, mutations present in DVGs arose to significantly higher frequencies, suggesting relaxed purifying selection on these genomes. Moreover, while we first hypothesized that mutations in coding regions of DVGs could benefit standard genomes through complementation, we were unable to test this hypothesis. Indeed, the identified mutation was incapable of *trans*-complementation as it is located on the NS5 polymerase domain, which we demonstrated cannot be complemented. Finally, we uncovered through smFISH that not only our DVG was capable of being *trans*-packaged, a low yet significant proportion of DVGs were co-packaged with standard genomes, therefore facilitating their maintenance through co-dispersal.

4.2. Recombination between DVGs and standard genomes

In our study, we showed that beneficial mutations readily present in DVGs could be exchanged to standard genomes through recombination events, therefore accelerating adaptation to mutagenic conditions. However, this is an artificial situation as it is very unlikely that a beneficial mutation would be present in all DVGs before being transferred to standard genomes. In more realistic scenarios, DVGs could randomly acquire mutations, possibly increase their frequencies to intermediate levels if they provide selective advantages in *cis* or in *trans*, and through random events provide through recombination these beneficial mutations to WT genomes. This raises at least two fundamental questions: (1) Are mutation

rates in DVGs higher than in WT genomes for them to significantly accelerate the rate of adaptation? (2) How frequently does recombination occur between DVGs and WT genomes?

Evidences of hypermutation and RNA editing have already been found in several viruses^{216,226,244,262}. However, these types of mutations provoke biased substitutions with 3'- and 5'-nucleotide context dependence, as they rely on deaminating enzymes from the ADAR or APOBEC protein families^{41,43,44}. Moreover, these proteins originally act as restriction factors, and induce lethal mutagenesis in standard genomes^{41,43}. It is therefore probable that mutation patterns from such processes would mostly induce deleterious mutations to DVGs, and subsequently result in non-functional recombination products with standard genomes. Indeed, hypermutation in DVGs has rather been associated with increased interfering properties of DIPs²⁴⁴. Nevertheless, some ADAR1 editing has also been shown to have proviral effects^{42,216} and could potentially be carried and exchanged by DVGs.

In addition, if DVGs replicate at faster rate than WT genomes, at similar mutation rates per replication cycle, DVGs could still in theory harbor more genetic diversity through geometric replication kinetics. Our results were however inconclusive as our sequencing data were not of sufficient quality to have sufficient reads mapping to the deletion breakpoint. Moreover, our deep-sequencing technique has limitations in terms of technical error rates: it does not allow us to explore mutations below 0.1%. This could be remedied by using high-accuracy library preparation protocols such as CirSeq²⁷⁵, that can even detect lethal mutations as they are generated⁸⁰. However, such techniques result in shorter read lengths (85-100 nt), therefore reducing the probability of them spanning the deletion breakpoint. Development of algorithms such as CoVaMa which associate long-read sequencing haplotype reconstruction abilities and short-read sequencing high-accuracy²¹⁷ could theoretically break this trade-off and offer a wider overview of mutation frequency distributions in DVGs.

Concerning the second question, a genome-wide barcoding approach of either DVG or standard genomes and estimation of their linkage disequilibrium when co-transfected together could help estimate the frequency of recombination events between these two types, and even identify recombination hotspots if linkage disequilibrium is not homogeneous. Moreover, because our results showed that NS5 polymerase activity could only act in *cis* for DVG replication, this raises new questions on the biochemical mechanisms of copy-choice recombination. Indeed, if NS5 preferentially replicates the template genome it was translated from, mechanisms and factors of template-switching should be investigated. A possible

explanation to resolve this apparent contradiction could be that initiation of replication has to occur in *cis*, which is then followed by template-switching that is no longer restricted to certain templates afterwards.

Finally, a possible approach to resolve if DVGs can accelerate adaptive evolution would be to co-infect WT genomes and functional (DVG-NS5_{WT}) or dysfunctional (DVG-NS5_{AAA}) DVGs without prior adaptive mutations at very high scale, and perform experimental evolution passages of such preparations in selective environments, such as mutagenic treatments. If the rate of phenotypic response is higher with functional DVGs, it would suggest that their ability to replicate and thus accumulate mutations helped viral populations to adapt faster to new selective pressures.

4.3. Complementation: an unresolved question

One of the goals of our work was to test if beneficial mutations in coding regions of DVGs could complement standard genomes. This was motivated by the observations that some reported DVGs had accumulated specific mutations, while counter-selecting for others observed in full-length genomes²¹⁷. Unfortunately, the ribavirin-resistance mutation we identified is located on NS5 which is not complementation-competent for its polymerase domain, preventing us from testing this hypothesis. It is therefore necessary to identify new mutations that confer fitness advantages, and that are capable of *trans*-complementation to pursue our investigation.

A first possible approach would be to select for mutations that confer resistance to drugs targeting structural protein functions or assembly. Indeed, structural proteins are believed to be readily complementation-competent as they were shown to *trans*-package DVGs in our study. Moreover, they are one source of competition for public goods between defective and standard genomes¹⁶⁷. Such molecules targeting oligomeric structural protein assembly²⁴⁰ and viral entry²⁷⁶ have already shown efficacy in dengue virus, and resistance mutations have been identified^{240,276}. However, as shown with the resistance against ST-148, the oligomeric properties of their target could lead to susceptibility-dominance to the drug, rendering complementation ineffective and preventing selection of such alleles²⁴⁰. Moreover, efficacy of these antivirals on ZIKV have not yet been characterized, and flaviviruses have shown differential susceptibilities to antivirals²⁷⁷.

One promising antiviral and its associated resistance mutations is the nucleoside inhibitor MK-0608 described in Mateo *et al.*²⁴⁰ which showed potent inhibition of hepatitis C virus²⁷⁸ and dengue virus serotype 2 (DENV-2)²⁷⁹ replication. In their study, they showed that evolved populations of DENV-2 treated with this drug selected independently for two mutations in the methyltransferase domain of NS5, namely NS5-A60T and NS5-Y201H, that partially restored viral titers upon treatment. Moreover, they showed that not only were these mutations complementation-competent, but their phenotypes were also dominant even when present at 10% in the inocula²⁴⁰. Nevertheless, translation of these results to ZIKV might prove to be unsuccessful: while the 60A residue is conserved in ZIKV, the 202 residue (ZIKV's homologue of DENV's 201 residue) is already a histidine, suggesting that the virus might already be resistant to the drug. Additive resistance phenotype of double-mutant viruses have not been explored by the authors. An alternative way to continue our investigations would be to create a structurally similar DVG in DENV-2 and to introduce in its genome the described mutations. Such DENV-2 DVG should have similar replicating and interfering properties, as Rezelj *et al.* have shown that this specific in-frame DelVG species seems to consistently appear in other flaviviruses²³, and have been reported in other studies^{280,281}.

Additionally, Mateo *et al.* showed that the methyltransferase domain of NS5 of dengue virus (and probably other flaviviruses) is capable of *trans*-complementation, in opposition to the strictly *cis*-acting RdRp activity that we described. This suggests that protein localization is probably not involved in limitations for complementation in this specific case. Future investigations of factors affecting complementation-competence of protein functions would not only inform what type of DVGs are structurally capable of being replicated, but as Mateo *et al.* suggest, identify targets for either susceptibility or resistance dominance during drug development.

4.4. Co-dispersal of DVGs and WT genomes

We have shown that not only DVGs could be *trans*-packaged and transmitted to new uninfected cells, but a minor but still significant proportion of them (13.9%) was capable of being co-packaged with WT genomes, ensuring their maintenance in following infection cycles, even at lower MOIs. Whether this ability for co-packaging is specific of this species of DVG, or rather a more common phenomenon of polyploid virions in flaviviruses has still to be investigated. Interestingly, this ability for DVG co-packaging could explain the report of a

transmissible DVG of DENV-1 over several years in Myanmar²⁸². Indeed, such phenomenon could enhance co-transmission even through tight cross-species bottlenecks that are believed to occur multiple times during the full arbovirus life cycle. In addition, Ke *et al.* have shown that such co-transmission events increased transmission rates compared to standard genome-only transmission chains²⁶¹. An interesting hypothesis would be that co-packaging of DVGs could be an evolvable trait, either encoded by the DVG itself or its helper functional genome. To go further, investigations on other type of CIUs DVGs could exploit to co-disperse should be performed. For example, their ability to affect cellular trafficking and the composition of viral particles released has already been shown in Sendai virus²³⁸. Finally, a hepatitis C virus DVG has been shown to exploit the exosomal transport system²¹³, which is a known form of CIU.

4.5. DVG genetic diversity and relaxed purifying selection

I have already discussed how increased genetic diversity in DVGs could benefit standard genomes through genetic recombination and complementation. Because of their reliance on complementation for some parts of their replication cycle, one could envisage that DVGs are subject to more relaxed purifying selection on complemented genes. We have found in our data that mutations in DVGs could reach higher frequencies, probably through random drift. However, these results only indicated a distribution spanning below a 2% allele frequency, and was not attached to any fitness effects of SNVs. On a macro-evolution level, relaxed selective pressure has been observed in the long-term transmitted DENV-1 DVG, where d_N/d_S ratios were significantly higher in the region in 3' of the non-sense mutation²⁸². More strikingly, some DVGs were shown to tolerate a normally lethal mutation in flock house virus²¹⁷, suggesting they reside in a different, more robust fitness landscape than their standard genome counterparts. These results suggest that, in theory, DVGs could explore fitness valleys that standard genomes could not because of their fitness costs. I propose a model in which, through random exploration of fitness valleys by DVGs, viral populations stuck in local fitness peaks could transition to higher peaks through genetic recombination. In order to explore this hypothesis, identification of a beneficial phenotype that requires two successive mutations linked by reciprocal sign epistasis is necessary. In a theoretical experimental design, a wild-type standard genome could be co-passaged in presence of barcoded functional and/or defective genomes containing one of these deleterious mutations.

If standard genomes acquire the second epistatically beneficial mutation preferentially from the DVG, this would validate the proposed model.

4.6. Investigations in less artificial set-ups

While our results serve as a proof-of-concept for beneficial interactions between DVGs and standard genomes, our current experimental design is still very artificial. Nevertheless, we hope that it will inspire future investigations of such interactions in more natural set-ups, or in natural infections. To our knowledge, besides promoting persistent infection, only two studies reported advantages associated with DVGs. The first paper corresponds to the DENV-1 DVG presented earlier, which appeared to have conferred higher transmission rates to its co-infecting viral population lineage through phylodynamic analysis²⁶¹. However, experimental validation of this effect has not yet been performed, and a molecular dissection of the underlying mechanism could reinforce the view that DVGs are integral part of virus populations. A second study reported that a hepatitis C virus DVG enhanced viral replication and release, and suggested that the increased production of core proteins could be responsible for this phenotype²¹³. Of importance, both DVGs were identified in natural infections. Interestingly, different reports of in-frame DelVGs have been observed in chronic patients infected by *Flaviviridae* viruses^{283,284}, suggesting they might play important roles in their virus life cycles, similar to subgenomes²⁸⁴.

4.7. Concluding remarks

Further investigations on positive interactions of DVGs should be pursued to finally remove their current association with interfering properties. Indeed, while DIPs probably represent a substantial proportion of DVGs, they have probably benefited from a biased exposure in the published literature for their potential as therapeutics. For example, our DVG was not reported to decrease viral titers in the experimental passages it was first identified, suggesting it was present at low frequencies²⁵⁵. Only when ratios were of 1:1 (DVG:WT) did we observe significant interference. However, these high levels of DVGs probably do not occur in natural infections. This would suggest that even potent identified interfering DVGs could have relative neutral effects in natural infections, therefore suggesting that a majority of DVGs could be neutral products of viral replication.

Finally, as my goal was to demonstrate that DVGs were capable of positive interactions, I want to highlight the importance of avoiding adaptationist narratives. If DVGs can improve fitness and evolvability, it does not necessarily mean that they evolved for this reason. More generally, even if some DVGs have been associated with life cycle functions, more parsimonious neutral explanations for their ubiquity in viral populations exist. Among them, the error-prone RdRp and the widespread complementation network in viral populations naturally promote social cheaters, and game theory has established that mixed populations of cheaters and cooperators could be stable. If DVGs have evolved to serve specific functions in virus life cycles, it is a secondary evolutionary process, as their existence does not require them. As Gould and Lewontin wrote in their critique of the adaptationist program²⁸⁵, it is critical to “distinguish current utility from reasons of origin”.

CHAPTER TWO

CHAPTER 2: POPULATION DYNAMICS OF CHIKUNGUNYA VIRUS IN EXPERIMENTAL ALTERNATING PASSAGES

1. Introduction

1.1. Arboviruses life cycles and the trade-off hypothesis

Arboviruses are RNA viruses that circulate between an arthropod vector and a vertebrate host. In most cases, this circulation is referred as an endozootic cycle, as it does not involve humans. The virus can exit this cycle and be transmitted by spill-over to a secondary host, which often is a dead-end host incapable of sustaining a transmission cycle. On some occasions, arboviruses successfully escape their primary endozootic cycle and circulate in a secondary epizootic one. For arboviruses of human concern, this creates a so-called urban cycle that is usually sustained by other species of competent vectors, which feed preferentially on humans (**Figure 17**).

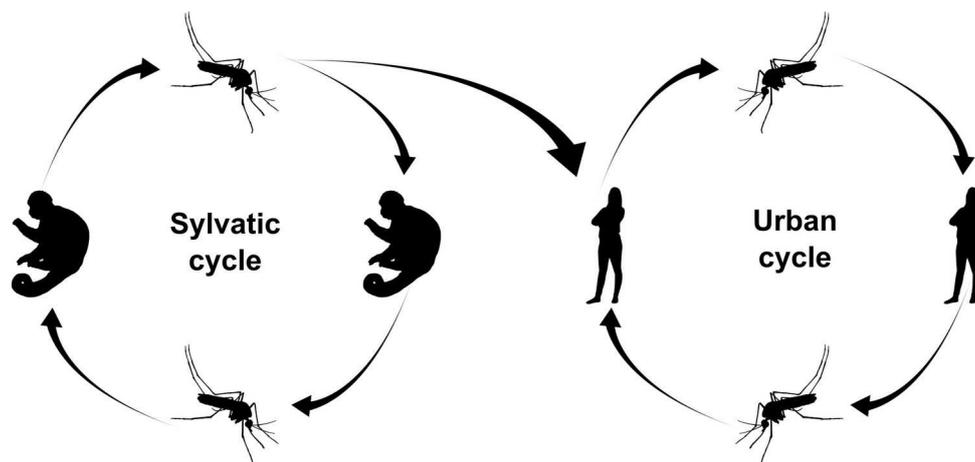


Figure 17 | Arboviruses' sylvatic and urban life cycles. Example of a mosquito-borne arbovirus that circulates in non-human primates in its sylvatic cycle, and is capable of establishing an urban cycle in humans.

During their life cycle, arboviruses undergo very different selective pressures between their two dissimilar hosts. Body temperature, cellular factors, codon usage, microbiotes and more importantly immune responses are a few examples of dissimilar factors to which these viruses need to concomitantly adapt. For the latter factor, vertebrate hosts rely on the

interferon response and their adaptive immunity to overcome viral infections, while insect hosts use mainly the RNA interference response to antagonize infection.

These constraints have been used to explain why arboviruses tend to have around 10-fold lower substitution rates ($1.4-5 \times 10^{-4}$ mut/nt/year) than other RNA viruses²⁸⁶. Indeed, in what is called the arbovirus trade-off hypothesis, it is hypothesized that arboviruses undergo higher levels of purifying selection. Because of antagonistic pleiotropy and the majority of single mutations being deleterious or lethal in one environment⁷⁷, beneficial or neutral mutations in one host are expected to be deleterious in the other one. This highly reduces the pool of truly neutral or beneficial mutations available, therefore constraining arboviruses adaptive pathways and neutral mutation accumulation through genetic drift and founder effects. In addition, it theoretically limits optimal fitness values of generalists in both environments compared to specialists^{118,287}. This 'cost of generalism' is quantifiable through the geometry of the Pareto front of the fitness trade-off, and allows predictions on overall fitness values over multiple environments¹¹⁸. Conversely, specialists are expected to outcompete generalists in constant environments to which they are adapted, and suffer a cost of adaptation when they adapt to new niches¹¹⁸ (**Figure 18A**).

However, evolutionary stasis is not necessarily predicted for generalism. First, while Pareto fronts are useful models for simple phenotype trade-offs, they fail to model multi-variable complex traits such as fitness. Indeed, because of widespread epistatic effects in RNA virus genomes, fitness landscapes are rugged and adaptive pathways taken by viral populations change the shape of the trade-off front depending on their position in sequence space. Consequently, dynamics of adaptation depends on fitness landscape congruence between the two environments:

- if landscapes are congruent on restricted numbers of low fitness peaks, populations suffer the cost of generalism and stasis occurs through back-and-forth movements in fitness landscapes;
- if congruent peaks are more densely scattered, generalists can wander through fitness landscape local peaks and substitution rates are not diminished;
- if congruent peaks correspond to high optimal fitness peaks in their respective environments, then stasis occur but generalists do not suffer the cost of generalism;
- if there is no congruence, adaptation can be dominated by one environment and generalists behave similarly to specialists of that given environment²⁸⁷ (**Figure 18B**).

Moreover, as I have introduced in the general introduction, viral populations are cooperative groups of viral genomes that can complement each other. This allows the possibility for division of labor inside populations, and therefore the existence of sub-populations specialized in each environment^{205,288}. In mixed populations without interactions, this already predicts higher geometric mean fitness than generalists under a fitness trade-off through expansion of the specialist sub-population^{118,288} (**Figure 18A**). If interactions through complementation are added, this could completely alleviate the trade-off. However, the possibility of evolution of specialized sub-populations in alternating conditions is still debated, as complementation puts a break on adaptation.

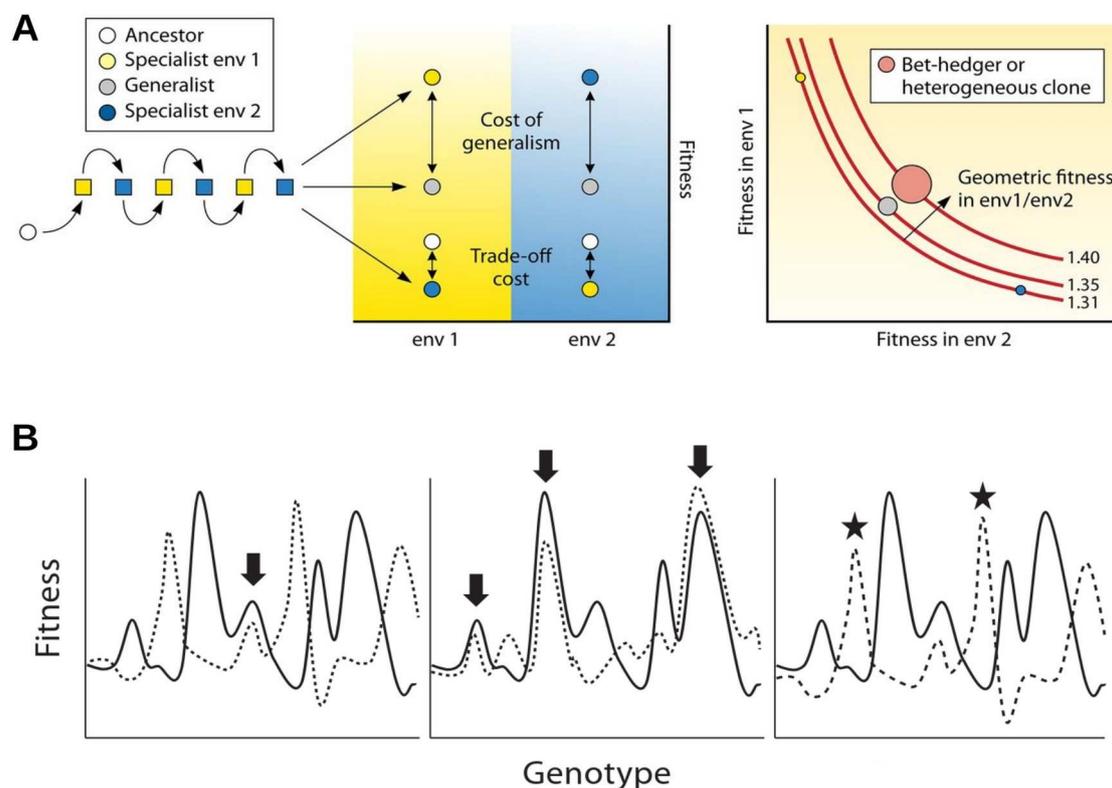


Figure 18 | Model of population fitness in alternating environments. (A) Cost of generalism and trade-off cost of adaptation (left). Schematic representation of the Pareto front of a fitness trade-off (right). Heterogeneous populations can partially alleviate the trade-off, pushing the Pareto front. (B) Host fitness landscapes' congruence determines available fitness peaks of arboviruses. The fitness landscapes can be congruent on a single low fitness peak (left), multiple peaks or the highest peak (middle). Alternatively, one host can dominate the adaptation (right). Extracted from Bergh *et al.*, 2018¹¹⁸ and Novella *et al.*, 2011²⁸⁷.

1.2. A review on previous experimental studies

Many authors have tried to test the trade-off hypothesis and derive conclusions on the topology of the fitness landscapes by submitting arboviruses (usually alphaviruses or flaviviruses) to experimental evolution passages *in vitro* or *in vivo*. However, studies from different teams yielded contradicting results, leaving the questions partially unresolved. In the following section, I summarize their results and discuss their technical and experimental design limitations.

Most works on the trade-off hypothesis used a similar approach: investigators experimentally passaged viruses in either a single host (insect or vertebrate cells, or live arthropods and rodents) or in an alternating pattern for a variable number of experimental passages, then they submitted the evolved viruses to a variety of assays to estimate fitness changes in both environments. Additionally, some researchers also performed challenges on cell types or animals which were not used for the experimental evolution. Sequencing of consensus changes and measures of genetic diversity were often also performed.

As a proxy for fitness, authors were divided on two different approaches to measure it. A first category measured the replication rate of viable viruses through classical growth curves in the competition cell lines or hosts^{105,289-293}. They usually derived the estimated replicative fitness from the slope of the log-transformed titer curve during the exponential growth, or the area under the curve to account for higher titer plateaus. This approach does not require any pre-required manipulation of the virus genome, and can easily be performed in viral isolates. However, because there is no direct competition between evolved lineages and ancestral stocks, conclusions on relative fitness are subject to errors: in cases where initial replication is faster yet yields fewer progeny during infection, it is unclear which lineage is more fit. Moreover, the assay is less sensitive to small differences in fitness and requires more biological replicates to increase statistical power. The second method to estimate relative fitness is through competition assays, in which evolved lineages are co-inoculated with an ancestor surrogate. Relative abundance changes can then be used to calculate the population genetics relative fitness w through neutral molecular markers such as restriction enzyme sites²⁹⁴⁻²⁹⁷, neutralizing antibody resistance^{155,288,298-302} or molecular barcoding^{303,304}. This method can be very sensitive to small fitness differences if carried over multiple competition passages, but requires either pre-introduction or pre-identification of target neutral markers in order to be carried out. If competition results are measured by antibody resistance, only

live viruses are included, while restriction sites or sequencing of molecular barcodes give information on genome abundance regardless of infectivity. Moreover, complementation or recombination between ancestral and derived lineages can occur, usually reducing the magnitude of the estimated relative fitness.

While all studies reported viral adaptation of serially passaged viruses to their respective single host, conclusions on the existence of a trade-off and therefore a cost of adaptation in the by-passed host are contradictory, even with the same virus model. Some authors found a fitness cost of adaptation in both single-host evolved lineages^{290,292-294,296,297}, whereas others found no fitness change in the by-passed host^{105,291,295}, or even increases in fitness²⁹⁸. Asymmetric cost of adaptation was also reported when viruses adapted to the mosquito host, but not when passaged in mammalian hosts^{289,295,303}. A particularly interesting case is the study from Coffey *et al.* where cost of adaptation was not observed when viruses were passaged *in vivo*, but was found when passaged *in vitro*²⁹⁵. Moreover, they showed that while *in vivo* passages did not result in fitness cost in by-passed hosts, viruses passed in BHK-21 cells lost fitness when challenged in hamsters²⁹⁵. As suggested by Novella *et al.*, the lack of expected trade-off could be linked to cell culture adaptation and released selective pressure²⁸⁷, which would indicate that viruses used are not already on the Pareto front and still have margins for adaptation in both hosts. This could be even true for *in vivo* studies where the virus model was usually isolated through successive cell passages²⁹⁵.

As a consequence of the absence of observable trade-offs in some experimental studies, the cost of generalism was not always reported^{289,293,294,296,298}. In instances where generalists suffered a cost, in most cases they still gained in fitness value, although at magnitudes lower than specialists^{105,290,291,295,297,303}. In one study with vesicular stomatitis virus, alternated lineages even suffered fitness loss in BHK-21 cells, which was linked to persistent infection passages of LL-5 sand fly cells³⁰². More generally, it appears that generalists tend to maximize their fitness gain in the vector, suggesting that they play a more determinant role in evolutionary trajectories^{290,293}.

See **Table 3** for a summary of experimental designs and conclusions.

Article	Virus	Vertebrate	Invertebrate	Nb passages	Nb replicates	MOI	Fitness proxy	Cost of generalism	Cost of adaptation
Arias-Goeta, 2014 ²⁸⁹	CHIKV	BHK-21	Aag-2	30	1	0.1	Growth curve	No	Yes in Aag-2
Coffey, 2011 ²⁹⁴	CHIKV	BHK-21, HeLa	C6/36	7	1	0.1	Competition assay (restriction site)	No	Yes
Cooper, 2001 ²⁹¹	EEEV	PDE	C7-10	10	10	0.01	Growth curve	Yes	No
Weaver, 1999 ²⁹⁶	EEEV	BHK-21	C6/36	100 (20-40)	1-4	0.01	Competition assay (clonal) (restriction site)	No	Yes
Greene, 2005 ²⁹⁷	SINV	BHK-21	C6/36	20-50	2	0.01	Competition assay (restriction site)	Yes	Yes
Coffey, 2008 ²⁹⁵	VEEV	Hamsters, mice	<i>Ae. aegypti</i>	10	1	10 ⁶ PFU	Competition assay (restriction site)	Yes	No
Coffey, 2008 ²⁹⁵	VEEV	BHK-21	RML12	10	1	0.1	Competition assay (restriction site)	Yes	Yes in RML12
Chen, 2003 ²⁹³	DENV	Vero	C6/36	30	1	10	Growth curve	No	Yes
Vasilakis, 2009 ²⁹⁰	DENV	Huh-7	C6/36	10	2	0.01	Growth curve	Yes	Yes
Jerzak, 2008 ³⁰⁵	WNV	Chickens	<i>Cx. spp.</i>	20	3	100 ID ₅₀	Competition assay (Sanger)	Yes in mosquito	Yes in mosquito
Deardorff, 2011 ³⁰³	WNV	DF-1	C6/36	20-40	2	0.1	Growth curve	N/D	Yes
Zarate & Novella, 2004 ³⁰²	VSV	BHK-21	LL-5	25	4	0.1	Competition assay (MARMU)	Yes in BHK	Yes
Novella, 1999 ²⁹⁸	VSV	BHK-21	LL-5	80	4	0.1	Competition assay (MARMU)	No	Opposite in BHK
Moutailler, 2011 ¹⁰⁵	RVFV	BHK-21	Aag-2	30	1	0.1	Growth curve	Yes in BHK	No

Table 3 | Experimental evolution designs and results investigating the trade-off hypothesis in arboviruses. Cost of generalism is observed if the fitness gain in generalist is lower than in specialist. Cost of adaptation is observed if adaptation to a serial condition leads to fitness cost in the by-passed condition. CHIKV: chikungunya virus; EEEV: eastern equine encephalitis virus; SINV: Sindbis virus; VEEV: Venezuelan equine encephalitis virus; DENV: dengue virus; WNV: West Nile virus; VSV: vesicular stomatitis virus; RVFV: Rift Valley fever virus; ID₅₀: infectious dose 50%; MARM U: VSV clone resistant to neutralizing antibody.

Some of these studies not only assayed the fitness phenotype of derived viral lineages, but also tried to identify causal mutations for the population fitness changes. Most of them used Sanger sequencing of either amplified extracted total vRNA^{293,296–298,302}, isolated plaques²⁹⁰ or subcloned vRNA after RT-PCR^{105,288,289,292,294,295,305}, therefore restricting themselves to near-consensus SNV identification. Methods using plaque isolation and subcloning allowed for

determination of genomic diversity of lineages. Generally, the alternated lineages did not undergo evolutionary stasis and accumulated substitutions^{290,295,298,302,305}. However, conclusions on stronger purifying selection in alternating conditions were contradictory as well. Coffey *et al.* found that alternation reduced mutation frequencies, and associated it with higher proportions of viable genomes than in serially passaged populations, suggesting higher selective constraints²⁹⁴. On the other hand, other authors rather found that while single-host passaging restricted genetic diversity, whether the vector^{289,290,294,295} or the host^{302,305} was responsible for reduced genetic diversity was not consensual. Additionally, when substitutions were cloned and assayed for replication efficiency, they did not necessarily recapitulate the observed phenotype in evolved lineages²⁸⁹, suggesting that undetected minority variants played a role in population adaptive values. This is even more striking in the study from Coffey *et al.* where no consensus change was observed in viruses passaged in mosquitoes, despite significant fitness increase of the population²⁹⁵.

Altogether, these investigations on the existence of a trade-off hypothesis do not draw a definite conclusion. These inconsistencies can be attributed to differences in virus and cellular or animal models used; however, limitations in experimental design also spurred hasty conclusions. First, on the cost of adaptation and generalism, it is impossible to conclude if ancestor populations are not readily situated on the Pareto front of the trade-off. Second, estimation of nucleotide diversity is only partial when using low-throughput sequencing techniques which cannot identify low frequency variants. Finally, in order to link genotype and phenotype, minority variants should be identified. In this chapter, my aim is to re-explore this evolutionary question using an updated experimental design and high-throughput sequencing which has since become more affordable. In the following section, I briefly present chikungunya virus (CHIKV) which is my virus model, and then present the initial experimental design I had planned.

1.3. Chikungunya virus

1.3.1. GENOME AND VIRUS STRUCTURE

CHIKV is a small enveloped, positive-sense single-stranded RNA virus belonging to the alphavirus genus within the *Togaviridae* family. It is part of the Semliki Forest virus complex together with closely related viruses such as o'nyong-nyong, Semliki Forest, Ross River and Mayaro viruses³⁰⁶. Its genome consists of a non-segmented linear RNA molecule of

approximately 11.8kb, with a 7-methylguanosine cap at its 5'-end and a 3' polyadenyl (poly-A) tail. It contains two ORFs separated by a non-coding region containing a subgenomic promoter, which allows the generation of a 26S subgenomic RNA. The first ORF encodes for four non-structural proteins (nsPs, namely nsP1 to nsP4), which constitute the replication complex. Similarly to most alphaviruses, CHIKV *nsp3* gene contains a read-through opal termination codon (UGA) near the 3'-end of its sequence. The second ORF codes for six structural proteins, namely the capsid (C), 3 envelope glycoproteins (E1, E2 and E3), a small 6 kDa peptide (6K) and a transframe peptide (TF) produced from a ribosomal frameshift sharing the same N-terminal region with 6K³⁰⁷ (**Figure 19A**).

The genome is surrounded by an icosahedral nucleocapsid formed by 240 capsid monomers, which is enclosed in a plasma-membrane-derived lipid bilayer decorated with 80 trimers of E1-E2 heterodimers. These trimers constitute the spikes on the virus surface and mediate virus-cell contact and viral entry. Overall, the mature virion has a diameter of approximately 70 nm³⁰⁷ (**Figure 19B**).

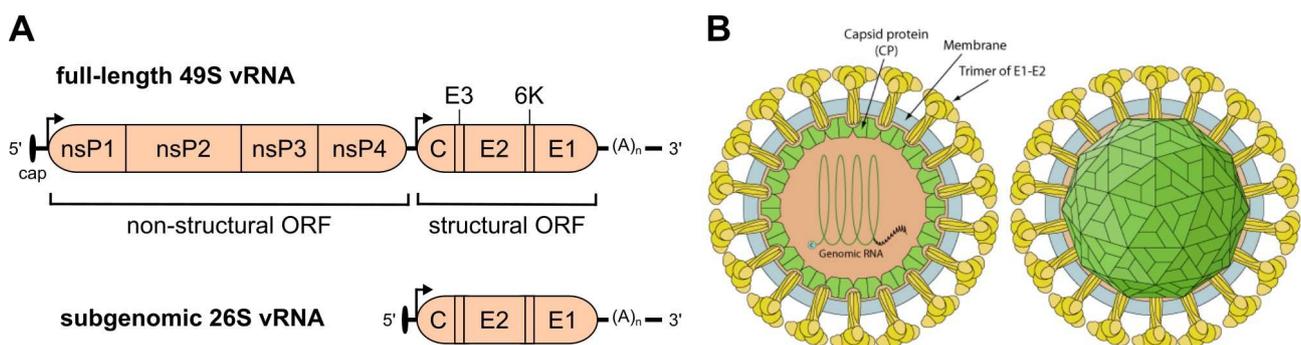


Figure 19 | Genome organization and viral particle structure of chikungunya virus. (A) Full-length genomic (49S) and subgenomic (26S) viral RNA of CHIKV. (B) Schematic representation of the viral particle. Extracted from ViralZone.

1.3.2. REPLICATION CYCLE

CHIKV has a broad cellular tropism, replicating in many mammalian and insect cells. In mammalian cells, matrix-remodeling associated protein 8 (Mxra8) has been identified as an entry receptor for the virus^{308,309}. However, partial protection against CHIKV infection from knock-outs of Mxra8 in some cell lines suggest that other receptors could also be used³⁰⁸. Interactions with phosphatidylserine receptors T-cell immunoglobulin and mucin domain 1 (TIM-1)³¹⁰ and glycosaminoglycans (GAGs)³¹¹ promote viral binding and entry. On the virus

side, E2 mediates the attachment to cells, and the viral particle is internalized by a clathrin-mediated endocytosis.

Lower pH in endosomes induces conformational changes in the viral glycoproteins, exposing the E1 fusion loop that inserts itself in the endosomal membrane, leading to the fusion of the viral and endosomal membranes and the release of the nucleocapsid in the cytoplasm. Upon release, the nucleocapsid disassembles and the full-length 49S genomic viral RNA (vRNA) is delivered into the cytosol for translation.

There, the host ribosomal machinery translates and processes the first ORF into two non-structural precursor polyproteins P123 and P1234. P123 represents the majority of products, while P1234 is synthesized through a low-frequency ribosomal read-through of the opal stop codon at the end of *nsp3*, yielding around 10% of full-length polyproteins³¹². After translation, nsP4 is cleaved autoproteolytically through the cysteine protease activity residing at the C-terminus of nsP2. The complex formed by P123 and nsP4 is the early polymerase complex capable of synthesizing the full-length 49S minus-strand vRNA which serves as a template for the synthesis of subgenomic 26S and *de novo* 49S plus-strand vRNAs. Minus-strand vRNA synthesis takes place at the plasma membrane, through anchoring of the replication complex mediated by nsP1 amphipathic domain, in membrane invaginations called 'spherules'. These viral replication compartments are thought to protect double-stranded RNAs (dsRNAs) from degradation and detection by the host immune system.

The P123 polyprotein is rapidly cleaved into nsP1 and P23, which in turn is cleaved *trans* into nsP2 and nsP3. Concomitantly, a fraction of spherules are internalized to form type I cytopathic vacuoles (CPV-I) containing numerous spherules at the membrane interface. All fully processed nsPs form the late replication complex which can only synthesize plus-strand 49S and 26S vRNAs. This induces a switch to strictly plus-strand vRNA synthesis in mammalian cells after 4 hours post infection.

Translation of the subgenomic 26S vRNAs, which are produced in excess comparatively to the 49S vRNAs, leads to the production of the structural polyprotein. During its synthesis, the capsid protein is cleaved and released early through its autocatalytic serine protease domain, and forms oligomers with newly synthesized 49S plus-strand vRNAs to form nucleocapsids each containing a full-length vRNA. The remaining polyprotein synthesis produces a majority of E3-E2-6K-E1, and about 10-18% of E3-E2-TF due to a -1 ribosomal

frameshift. The signal peptide at the N-terminus of E3 translocates the two polypeptides to the endoplasmic reticulum (ER) where they are cleaved by host proteases into pE2 (E3-E2), 6K or TF, and E1. As the E1 and pE2 complex matures through the secretory pathway as a non-covalent hetero-oligomer, it undergoes post-translational modifications such as palmitoylation and glycosylation to form mature glycoproteins. Next, E3 is released through cleavage, leading to trimerisation of E1-E2 heterodimers to form mature spikes which are transported to the plasma membrane. Finally, recruitment of nucleocapsids at the plasma membrane results in mature virion budding. Alternatively, at a late stage of infection, type II cytopathic vacuoles (CPV-II) are formed from the trans-Golgi network. They typically contain oligomers of E1-E2 dimers and are studded with nucleocapsids, which allow assembly and egress of mature virions.

For a graphical summary of CHIKV replication cycle, see **Figure 20. Table 4** recapitulates known functions of CHIKV proteins.

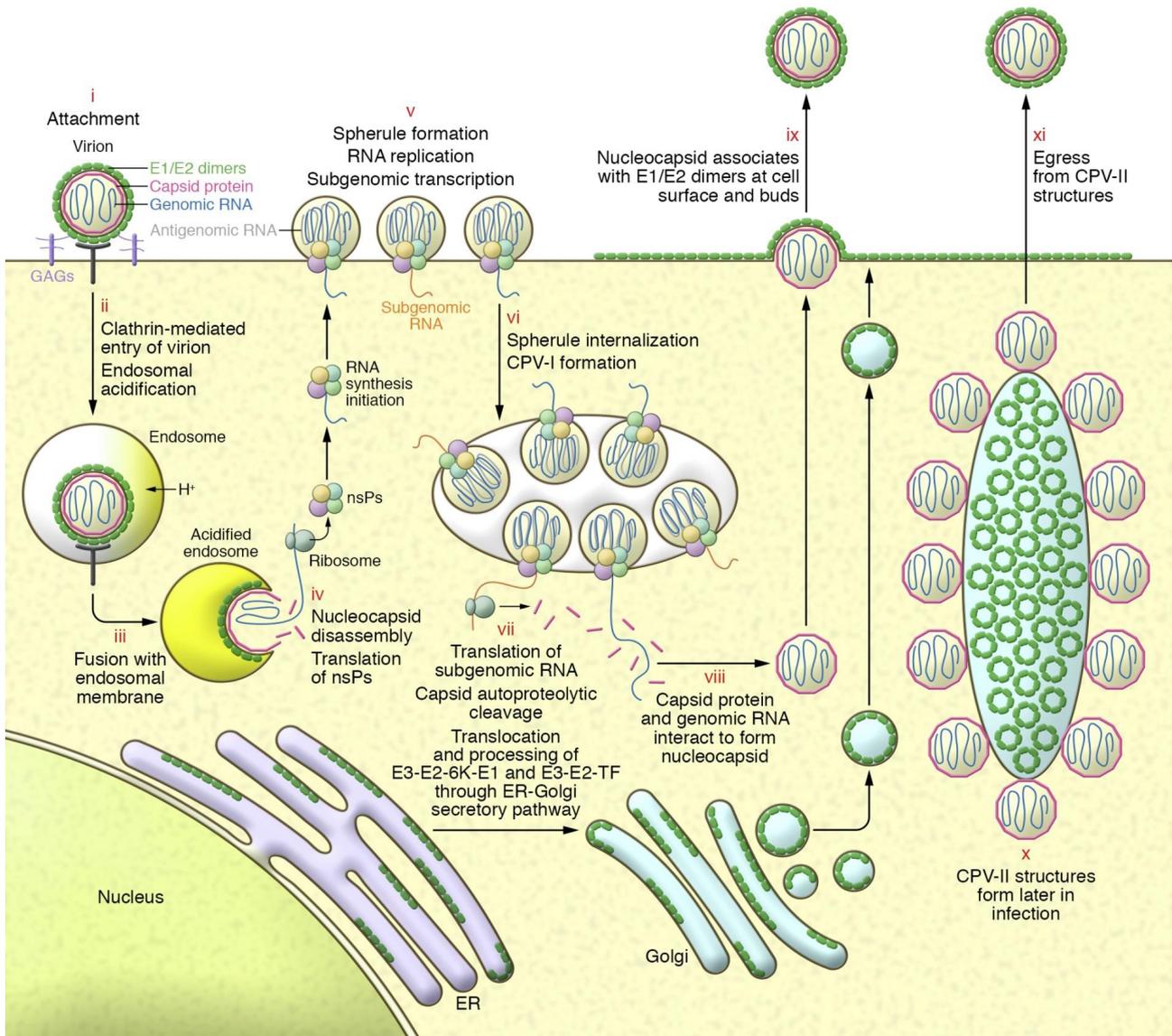


Figure 20 | Chikungunya virus replication cycle in mammalian cells. Extracted from Silva and Dermody, 2017³⁰⁷.

Protein	Size (aa)	Functions
nsP1	535	Methyltransferase and guanylyltransferase for vRNA capping; membrane anchor through amphiphatic domain.
nsP2	798	N-terminal NTPase, helicase and RNA triphosphatase activities; non-structural polyprotein cleavage through C-terminal cysteine protease.
nsP3	530	Phosphoprotein with unknown functions, but required for minus-stran synthesis; likely interacts with host proteins.
nsP4	611	RNA-dependent RNA polymerase; putative terminal transferase activity.
C	261	Encapsidates vRNA to form nucleocapsid; autocatalytic serine protease.
E3	64	N-terminus directs structural polyprotein to the ER-Golgi pathway.
E2	423	Mediates attachment to cell receptors; major target of neutralizing antibodies.
6K/TF	61/76	Putative ion channels; may enhance viral egress.
E1	437	Mediates fusion of viral envelope and cellular membrane via fusion loop.

Table 4 | Chikungunya virus proteins and functions. Adapted from Silva and Dermody, 2017³⁰⁷.

1.3.3. EPIDEMIOLOGY

CHIKV is an arbovirus which was first identified in an 1952-1953 outbreak in Southern Tanganyika (current Tanzania)³¹³. The different strains are classified into different lineages according to their genetic diversity, namely the West African, Asian, East-Central-South-African (ECSA)³¹⁴, and more recently the Indian Ocean lineage (IOL) which emerged and evolved from the ECSA genotype³¹⁵.

The virus circulates in a sylvatic cycle between non-human primates and sylvatic *Aedes* spp. mosquitoes such as *Aedes furcifer*, *Aedes luteocephalus* and *Aedes africanus*³¹⁶. The ECSA lineage has been detected in multiple non-human primate species in Africa³¹⁷, and causes outbreaks in human populations through spillovers from its sylvatic cycle to its urban life cycle where *Aedes aegypti* becomes the main vector³¹⁸.

The first major outbreak in modern times occurred in 2005, when a strain from the ECSA lineage spread to the Indian Ocean islands, India and South-East Asia, giving rise to the IOL lineage. This outbreak caused around 6.5 million cases, including imported cases in

Europe and North America and small autochthonous outbreaks in Italy³¹⁹ and southern France³²⁰. The importance of this epidemic was linked to a change in vector species: while ECSA strains in Africa are typically transmitted by *Aedes aegypti* mosquitoes in their urban cycle, most human IOL infections were linked to *Aedes albopictus* ones, which distribution has expanded to tropical and temperate regions worldwide³²¹. In particular, a specific mutation in the E1 glycoprotein inducing a A226V substitution was linked to the outbreak³¹⁵, and was later confirmed to increase infectivity, dissemination and transmission rates in *Aedes albopictus*, while having no effect in *Aedes aegypti*^{322,323}. Additional substitutions, namely E2-L210Q and E2-K252Q, were shown to increase virus fitness in *Aedes albopictus* in the E1-A226V genetic background^{324,325}.

In Asia, introduction of the Asian lineage is thought to have happened through shipping at the beginning of the 20th century³¹⁸, and has caused sporadic outbreaks mediated by *Aedes aegypti*³¹⁸, until the 2013 epidemic where it spread to the Caribbean Sea and the American continent, causing the world second major CHIKV outbreak^{326,327}. Interestingly, while *Aedes albopictus* is abundant in South-East Asia, there is no evidence of its implication in Asian lineage outbreaks. This phenomenon was attributed to an epistatic interaction with the E1-98T residue present in all Asian genotypes, which prevents the acquisition of the E1-A226V substitution¹¹⁴, and therefore adaptation to this new vector.

More recently, outbreaks in India, Pakistan, Bangladesh, Thailand, Myanmar and Cambodia between 2010 and 2020 were linked to another sub-lineage of the ECSA clade, which is characterized by E1-K211E and E2-V264A substitutions, and the absence of the IOL E1-A226V one³²⁸.

1.4.4. EXPERIMENTAL DESIGN AND AIMS OF CHAPTER 2

In order to re-explore and test the arbovirus trade-off hypothesis, we will use an IOL lineage CHIKV clone previously used in the laboratory^{123,214} to serially or alternately passage it in BHK-21 or U4.4 cells. Both of these cell lines were chosen as they are widely used in virology literature, yield sufficient viral titers for subsequent assays and high-throughput deep-sequencing, and relatively preserve their functional innate immune pathways. Fitness estimations will be performed by competition assays as they are more sensible to small fitness differences, and relative abundance of derived populations measured with an accuracy-

improved multiplex RT-qPCR approach described in Carrasco *et al.*³⁰⁴. To identify evolved populations from the ancestral surrogate competitor, all experimental evolution lineage replicates and the surrogate contain a unique 6-nucleotide molecular barcode inserted at the 3' end of the subgenomic promoter. While exogenous, this small barcode is non-coding and has been shown to be relatively neutral¹²³, and it does not affect the initial position of each clone in fitness landscape. In addition to its use for competition assay read-outs, it allows identification of cross-contamination between samples, and the possibility to mix evolved populations to evaluate their contributions in complementation. Sequencing of lineages across passages will be performed with Illumina short-read RNA-sequencing, which allows the identification of minority variants down to 0.1% in frequencies.

Briefly, similarly to previous studies, barcoded clones are first passaged serially or alternately for 20 experimental evolution passages in BHK-21 and U4.4, and variant composition and population fitness assayed across passages. This allows initial adaptation to each passaging condition, and should move viral populations towards the theoretical Pareto front. In a second phase of experimental evolution, each lineage switches regime: alternated passages are released from alternation and serially passaged in each cell line; serial passages are either switched to the by-passed host, or alternately passaged (**Figure 21**). In theory, if a trade-off exists, new adaptation to their former by-passed host should result in fitness cost to the formerly adapted one, as the virus population moves along the Pareto front. Combined with high-throughput sequencing, loci under antagonistic epistasis could be identified and mapped onto the viral genome. Moreover, other conditions aim to explore the ruggedness and congruence of the two fitness landscapes, allowing us to conclude if evolutionary stasis is expected. Finally, to test if division of labor and collaborative complementation between specialist populations could be maintained in alternated regimes, we will mix serially-passaged populations over several passages.

To account for evolutionary contingency and measure parallel evolution of each condition, six independent biological replicates for each regime, with their own distinctive molecular barcode, will be passaged.

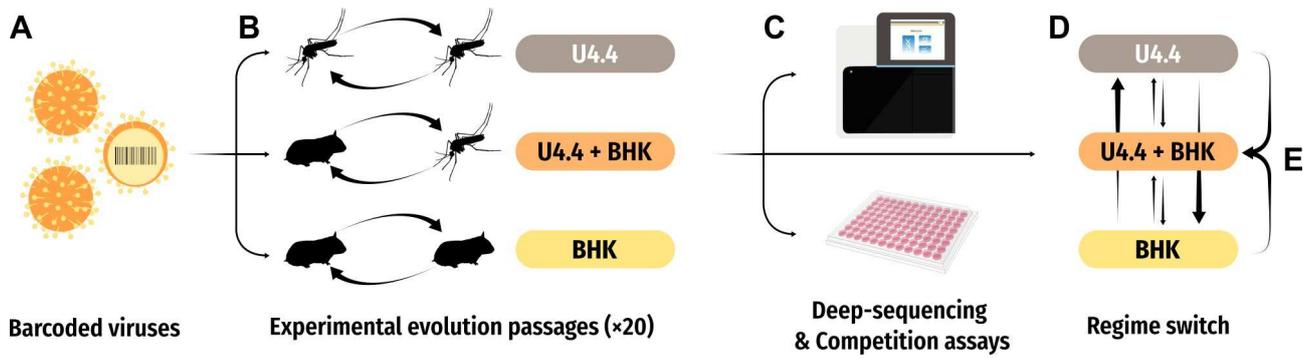


Figure 21 | Two-phase experimental evolution design in constant or alternated passages of barcoded CHIKV in BHK-21 or U4.4 cells. (A) Each lineage replicate is barcoded with a unique 6-nucleotide insert in 3' of the subgenomic promoter. (B) Virus populations are initially passaged either serially in U4.4 or BHK-21 cells, or alternately between the two cell lines for 20 passages. (C) Viral population fitness and variant compositions are determined through Illumina RNA deep-sequencing and a competition assays followed by a multiplex RT-qPCR read-out. (D) Terminal passages of the initial phase of experimental evolution are switched to new regimes to map the Pareto front geometry and determine fitness landscape congruence and loci of antagonistic pleiotropy. (E) Serially-passaged populations are mixed together to test if division of labor yield higher population fitness in alternated passages.

2. Material and methods

2.1. Cells, plasmids and viruses

2.1.1. CELLS

U4.4 cells were maintained in Leibovitz's L-15 medium (Gibco) complemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v) penicillin-streptomycin (P/S; Gibco), 1% (v/v) non-essential amino acids (Gibco) and 2% (v/v) tryptose phosphate broth (Sigma) at 28°C without CO₂.

Vero-E6 and BHK-21 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) containing 10% (v/v) FBS and 1% (v/v) P/S in a humidified atmosphere at 37°C with 5% CO₂.

2.1.2. PLASMIDS

The chikungunya virus (CHIKV) infectious clone used as the marked ancestor is the Indian Ocean Lineage (IOL) La Réunion 06-049 strain (GenBank accession number AM258994) under a SP6 promoter, and modified with an AvrII restriction site (CCTAGG) inserted at position 7612 after the subgenomic promoter as previously described in Carrau *et al*¹²³. Subsequent tagged lineages were generated by site-directed mutagenesis modifying the AvrII restriction site with a unique 6-nucleotide barcode (**Table 5**). Each barcode is unique and has a minimum of 3 nucleotide differences with all the other ones in order to distinguish them after numerous passages and potential fixation of random mutations.

All plasmid stocks were obtained through classic transformation protocols of XL-10 Gold Ultracompetent cells (Agilent) or Top10 Chemically competent *E. coli* (ThermoFisher Scientific), followed by Mirapreps²⁷¹ of colonies grown in 100 mL LB Miller using NucleoSpin Plasmids kits (Macherey Nagel). Purity and concentrations of extracted DNAs were evaluated on a NanoDrop (ThermoFisher Scientific). Full plasmid sequences were checked by Sanger sequencing (Eurofins Genomics).

EE regime	ID	Barcode	Sense	Primer sequence
U4.4	I1	GAGTAC	Forward	5'-GAGTACATGGAGTTCATCCCAACCC-3'
	I2	CATCTC	Forward	5'-CATCTCATGGAGTTCATCCCAACCC-3'
	I3	CTAGCT	Forward	5'-CTAGCTATGGAGTTCATCCCAACCC-3'
	I4	AACTGT	Forward	5'-AACTGTATGGAGTTCATCCCAACCC-3'
	I5	TTCGAG	Forward	5'-TTCGAGATGGAGTTCATCCCAACCC-3'
	I6	TCACTC	Forward	5'-TCACTCATGGAGTTCATCCCAACCC-3'
BHK-21	V1	ATCCGG	Forward	5'-ATCCGGATGGAGTTCATCCCAACCC-3'
	V2	ACTAAC	Forward	5'-ACTAACATGGAGTTCATCCCAACCC-3'
	V3	CGTACC	Forward	5'-CGTACCATGGAGTTCATCCCAACCC-3'
	V4	GCTACT	Forward	5'-GCTACTATGGAGTTCATCCCAACCC-3'
	V5	CTGCAG	Forward	5'-CTGCAGATGGAGTTCATCCCAACCC-3'
	V6	AAGCCT	Forward	5'-AAGCCTATGGAGTTCATCCCAACCC-3'
Alternated	A1	GCCATC	Forward	5'-GCCATCATGGAGTTCATCCCAACCC-3'
	A2	TGGATC	Forward	5'-TGGATCATGGAGTTCATCCCAACCC-3'
	A3	TGTTCA	Forward	5'-TGTTCAATGGAGTTCATCCCAACCC-3'
	A4	TACGGC	Forward	5'-TACGGCATGGAGTTCATCCCAACCC-3'
	A5	TATCAG	Forward	5'-TATCAGATGGAGTTCATCCCAACCC-3'
	A6	ATTCAT	Forward	5'-ATTCATATGGAGTTCATCCCAACCC-3'
All	All	-	Reverse	5'-TGTAGCTGATTAGTGTTTAGATACTTG-3'

Table 5 | Site-directed mutagenesis primers for barcoded CHIKV clones. EE = Experimental evolution. ID = Lineage identifier.

2.1.3. VIRUSES

CHIKV rescues were carried out by first linearizing the infectious clone plasmids with the NotI restriction enzyme (Anza) for 30 min at 37°C, followed by DNA purification using the NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel), *in vitro* transcription of messenger RNA (mRNA) using the SP6 mMessage mMachine kit (Invitrogen), and finally mRNA purification using the RNeasy MinElute Cleanup kit (Qiagen). Purified RNAs were quantified using the Qubit RNA BR assay kit (Invitrogen) on a Qubit 2.0 fluorometer (Invitrogen). RNA was subsequently transfected into BHK-21 cells using the TransIT-mRNA kit (Mirus Bio)

following the manufacturer's instructions, and clarified supernatant collected at 2 dpt. Viral stocks were obtained after an initial passage (passage 1) in either BHK-21 or U4.4 cells at MOI 0.1 and collected at 2 dpi. Stocks were titered in Vero-E6 cells by plaque assays in triplicates.

2.2. Site-directed mutagenesis

Site-directed mutagenesis was performed by classical amplification and ligation protocol. Briefly, primers containing the mutations of interest (Integrative DNA Technologies) were used to amplify target DNA products through polymerase-chain reaction (PCR) using the Phusion High-Fidelity DNA Polymerase kit (ThermoFisher Scientific) for 18 cycles at annealing temperatures of 60°C. Reaction products were then digested with DpnI (ThermoFisher Scientific) for 30 min at 37°C to eliminate plasmid DNA, and subsequently purified (NucleoSpin Gel and PCR Clean-up, Macherey Nagel). Linear DNAs were circularized by adding 4 µL Ligase Buffer (New England BioLabs) and 1 µL T4 polynucleotide kinase (PNK; New England BioLabs) for 30 min at 37°C, followed by 1 µL T4 DNA ligase (New England BioLabs) and incubating at 4°C overnight. Circular DNA was then used to transform XL-10 Gold Ultracompetent cells (Agilent) or Top10 Chemically competent *E. coli* (ThermoFisher Scientific). Each transformation was performed by adding 2 µL circular DNA product to 25 µL suspended bacteria, incubated 30 min on ice, heat-shocked at 42°C for 40 seconds (s), supplemented with 100 µL SOC media (ThermoFisher Scientific), then incubated in a shaking incubator at 37°C, 120 rotations per minute (rpm) for 1 hour (h). After incubation, total volume was plated on a LB agar plate with 100 µg/mL ampicillin and incubated at 37°C overnight. Successful mutagenesis were validated by randomly picking colonies to grow in 5 mL LB Miller overnight followed by plasmid DNA extraction minipreps (NucleoSpin Plasmids, Macherey Nagel) and Sanger sequencing of the target mutation site (Eurofins Genomics).

2.3. Viral titers

Infectious viral titers were determined by plaque assays. Briefly, monolayers of Vero-E6 cells in 24-well plates were inoculated with 150 µL serial 10-fold dilutions of viruses in unsupplemented DMEM and incubated at 37°C and 5% CO₂ for 1h. Then, wells were gently covered with approximately 1 mL of a solid overlay comprising DMEM supplemented with 2%

FBS, 1% P/S and 0.8% agarose (Invitrogen) and incubated again for 2-3 dpi until fixation with 4% formaldehyde (Sigma). Plaque-forming units (PFUs) were manually enumerated after crystal violet 0.2% staining .

2.4. Viral experimental evolution passages

Barcoded virus populations were blindly passaged either in BHK-21 cells only, U4.4 cells only or alternately between the two different cell lines in 6-well plates or T25 flasks at approximately MOI 0.1 PFU/cell following regular volume adjustments after viral titration every 5 passages. Briefly, inoculum was added on 80% confluent cells in their respective media complemented with 2% (v/v) FBS and incubated for 2 dpi at 37°C (BHK-21) or 28°C (U4.4). Infectious supernatants were collected and clarified by centrifugation for 10 min at 4,000×g, then aliquoted for subsequent viral assays or RNA extractions. Initial passages were carried until passage 20, and infections in T25 flasks were done for all passages that were sequenced (passage 1, 5, 10, 15, 19 and 20).

2.5. Competition fitness assays

2.5.1. COMPETITION PASSAGES

Relative fitness of passaged virus populations in either BHK-21 or U4.4 were determined through three subsequent competition passages in presence of an unpassaged marked clone in three biological replicates. Inocula (passage 0) were prepared by mixing assayed populations and the marked competitor (AvrII) at 1:1 ratio following their infectious titers, and competition passages carried out blindly in 96-well plates at approximately MOI 0.1 PFU/cell for 2 dpi. Clarified supernatants were heat-inactivated at 95°C for 5 min and used directly for multiplex real-time quantitative reverse transcription polymerase chain reaction (RT-qPCR) to determine relative abundance of each barcoded populations.

2.5.2. MULTIPLEX RT-qPCR

Relative abundance of assayed virus populations were determined through an improved multiplex RT-qPCR version of the assay described in Carrasco *et al.*³⁰⁴ A dual-probe TaqMan set-up was used to improve accuracy and limit sensibility to pipetting errors and measurement stochasticity. Briefly, a first probe TaqMan 6-FAM™ probe (Integrative DNA

Technologies, **Table 6**) binds to a 125-nucleotide amplicon downstream of the engineered molecular barcode introduced in each virus populations. Two different forward primers specific at their 3' ends to the barcode of either the assayed population or the marked competitor are used along a common reverse primer into two separated reaction mixes using the same RNA preparation, allowing discriminant absolute quantification of each population. A second TaqMan JOE NHS probe (Integrative DNA Technologies; equivalent to VIC™; **Table 6**) binds to a 98-nucleotide amplicon in the *nsp3* gene region using common forward and reverse primers in the two reaction mixes, and is used to quantify total viral RNAs in order to correct for pipetting errors in-between the two mixes. To account for potential amplification-specificity differences between marked competitor and assayed population RNAs, serial 10-fold dilutions of *in vitro* transcription stocks (250 ng/μL) of each clone were used as standards, and standard curve values of the assayed population and associated cycle threshold (C_T) values were linearly transformed to match the dynamic range of the marked clone standard curve.

Individual RT-qPCR reactions were set up using 1 μL of 5-fold diluted inactivated supernatant in DNase- and RNase-free water, mixed with their respective forward primer for the 6-FAM™ amplicon, common primers and probes (Integrative DNA Technologies; **Table 6**) and Luna® Universal Probe One-Step RT-qPCR kit reagents (New England BioLabs) up to a final volume of 5.5 μL following ratios recommended by the manufacturer. RT-qPCRs were carried out in 384-well plates in an Applied Biosystems QuantStudio 6 or 7 Flex thermocycler with the following cycling conditions: 55°C for 10 min, 95°C for 1 min and 40 cycles of 95°C for 10 seconds followed by 60°C for 1 min.

Region	Target	Type	Sequence
nsP3	All	Probe	5'-[6JOEN]TAAAACGCATGGACATCGCGAAGAACGATG[IABkFQ]-3'
	All	Reverse	5'-ACAGGTCACCCGAGCAG-3'
	All	Forward	5'-GTATACTGCCTTGCAAACACC-3'
Barcode	All	Probe	5'-[6FAM]TTTACAATAGGAGGTACCAGCCTCGACCCT[IABkFQ]-3'
	All	Reverse	5'-TAACTGCTGAGATCAGCTGG-3'
	AvrII	Forward	5'-GTATCTAAACACTAATCAGCTACACCTAGG-3'
	V1	Forward	5'-GTATCTAAACACTAATCAGCTACAATCCGG-3'
	V2	Forward	5'-GTATCTAAACACTAATCAGCTACAATAAC-3'
	V3	Forward	5'-GTATCTAAACACTAATCAGCTACACGTACC-3'
	V4	Forward	5'-GTATCTAAACACTAATCAGCTACAGCTACT-3'
	V5	Forward	5'-GTATCTAAACACTAATCAGCTACACTGCAG-3'
	V6	Forward	5'-GTATCTAAACACTAATCAGCTACAAAGCCT-3'
	A1	Forward	5'-GTATCTAAACACTAATCAGCTACAGCCATC-3'
	A2	Forward	5'-GTATCTAAACACTAATCAGCTACAAGGATC-3'
	A3	Forward	5'-GTATCTAAACACTAATCAGCTACATCTTCA-3'
	A4	Forward	5'-GTATCTAAACACTAATCAGCTACATACGGC-3'
	A5	Forward	5'-GTATCTAAACACTAATCAGCTACATATCAG-3'
	A6	Forward	5'-GTATCTAAACACTAATCAGCTACAATTCAT-3'
	I1	Forward	5'-GTATCTAAACACTAATCAGCTACAGAGTAC-3'
	I2	Forward	5'-GTATCTAAACACTAATCAGCTACACATCTC-3'
I3	Forward	5'-GTATCTAAACACTAATCAGCTACACTAGCT-3'	
I4	Forward	5'-GTATCTAAACACTAATCAGCTACAAACTGT-3'	
I5	Forward	5'-GTATCTAAACACTAATCAGCTACATTCGAG-3'	
I6	Forward	5'-GTATCTAAACACTAATCAGCTACATCACTC-3'	

Table 6 | Probes and primers for CHIKV multiplex RT-qPCR.

2.5.3. RELATIVE FITNESS

Relative fitness values w of assayed evolved populations can be derived after rearranging formula 1, such as:

Equation 6³⁰⁴
$$w = \left(\frac{p(t)/q(t)}{p(0)/q(0)} \right)^{1/t},$$

with $p(t)$ and $q(t)=1-p(t)$ the frequencies of respectively the assayed population and the marked competitor at competition passage t . In our assay, frequency ratios $R(t)=p(t)/q(t)$ at each passage t are determined for each biological replicates, and \log_2 -transformed relative fitness values $\log_2(w)$ estimated by the slope of the linear regression

Equation 7
$$\log_2(R(t)/R(1)) \sim \log_2(w) \times t.$$

Competition passage 1 frequency ratios $R(1)$ were used to normalize relative frequency ratios $R(t)/R(1)$ as they have consistently less missing values due to undetermined C_T values than inocula (passage 0). Linear regressions were performed using base R 3.6.3. Standard errors of the relative fitness estimates are obtained from the linear model.

2.6. Viral RNA extraction

Viral RNAs were obtained through classic trizol extractions of clarified supernatants after incubating supernatants with 15 μ L RNase T1/A mix (ThermoFisher Scientific), 3 μ L Ambion™ DNase I (Invitrogen) and 3 μ L RQ1 DNase (Promega) at 37°C for 1 hour. Briefly, 1 ml trizol reagent (Sigma) was added to 1 mL thawed supernatants and allowed to incubate 5 min at room temperature. Then, 400 μ L chloroform was mixed to the suspension, incubated for 3 min and centrifuged 15 min at 12,000 \times g at 4°C for phase separation. Upper phases were collected and nucleic acids were precipitated by mixing 1 mL isopropanol (Sigma) and 1.5 μ L glycoblue (15 mg/mL; Invitrogen) and incubating at -20°C overnight. RNA pellets were recovered by centrifuging at 12,000 \times g at 4°C for 20 min, followed by a 1 mL 75% ethanol wash. RNAs were finally recovered in 30 μ L DNase- and RNase-free water. The quality and quantity of extracted RNAs were analyzed using the Quant-iT RNA BR assay kit (ThermoFisher Scientific) on a Tecan infinite M200 PRO plate reader (Tecan).

2.7. Viral RNA sequencing

Viral RNA preparations were first purified using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs) to get rid of abundant ribosomal RNAs (rRNAs) found particularly in U4.4 passages, and quantified afterwards using the Quant-iT RNA BR

assay kit (ThermoFisher Scientific). Libraries were prepared using the NEBNext Ultra II RNA Library kit (New England BioLabs) and multiplexed using the 96-index primers from the NEBNext Multiplex Oligos for Illumina kit (New England BioLabs). Quality of the libraries was checked using a High Sensitivity DNA Chip (Agilent) and quantity determined using the Quant-iT dsDNA assay kit (ThermoFisher Scientific). Libraries were diluted to 1 nM and pooled together, then quantified again using the Qubit dsDNA HS assay kit (Invitrogen) on a Qubit 2.0 fluorometer (Invitrogen). Sequencing of the libraries was performed by single-end on a NextSeq 500/550 MidOutput v2.5 (150 cycles) flowcell (Illumina) using a NextSeq 500 sequencing machine (Illumina).

2.8. Sequencing data alignment and variant calling

Sequencing reads were demultiplexed using bcl2fastq v2.20.0.422 (Illumina). Sample quality control was performed using FastQC v0.11.9 (<https://github.com/s-andrews/FastQC>), and number of reads aligning to the chikungunya virus reference sequence (GenBank accession number AM258994) were determined using Bowtie 2³²⁹ v2.3.5.1 (<https://github.com/BenLangmead/bowtie2>). Samples with number of aligned reads below 200,000 (approximately 18X genome coverage) were re-extracted and re-sequenced after a fresh library preparation. In case of low viral RNA yields in different independent sequencing runs, fastq files were concatenated and quality control re-performed on the new files.

Variant calling was performed with ViVan v0.43 (Viral Variant Analysis)²⁷⁴, an in-house pipeline which performs multiple rounds of alignments and consensus reference sequence modifications to increase alignment and variant calling accuracies. Briefly, reads are clipped and trimmed using fastq-mcf from ea-utils v1.1.2 (<https://expressionanalysis.github.io/ea-utils/>) to remove Illumina adapters and low quality ends with phred scores below 30. Reads are then aligned using the Burrows-Wheeler aligner^{330,331} (BWA v0.7.4, <https://github.com/lh3/bwa>) against their corresponding barcoded reference sequence, allowing a maximum of 2 mismatches and no gap. Alignments are then processed using the SAMtools v0.1.19 toolkit³³² (<https://github.com/samtools/samtools>) to generate pileups of called bases at each position. Insufficiently covered bases (< 200X) are discarded and a generalized likelihood-ratio test is performed to identify statistically significant variants ($p_{\text{adj}} < 0.05$ after a Benjamini-Hochberg false-discovery post-hoc test). Majority variants

identified are then modified in the reference sequence, and alignment and variant calling are re-performed.

2.9. Defective viral genomes calling

Demultiplexed fastq reads from variant calling were used to call recombination points. First, reads were clipped and trimmed (phred quality score < 30 and minimum length = 50) using BBDuk, and PCR duplicates removed with Clumpify from the BBMap v38.91 toolkit (<https://sourceforge.net/projects/bbmap/>). Then, alignments against corresponding barcoded reference sequence and recombination sites were determined using ViReMa v0.25 (Viral Recombination Mapper) algorithm²⁷³ (<https://sourceforge.net/projects/virema/>) with seeds of 20 nucleotides and maximum 2 mismatches allowed. ViReMa can identify recombination points in-between positive or negative sense RNAs, and between opposite polarity RNAs. For simplicity, we only filtered recombination points resulting in deletions (5'-positive to 3'-positive; and 3'-negative to 5'-negative) to be considered as defective viral genomes, and discarded possible duplications or copy-backs.

2.10. Shannon entropy

Average Shannon entropies \bar{S}_n of viral genes were manually computed using variants with frequencies over 0.1% on R v3.6.3. The following formula was used:

Equation 8

$$\bar{S}_n = \frac{-\sum p_i \log_2(p_i) + (1 - p_i) \log_2(1 - p_i)}{N}$$

with p_i the allele frequency of SNVs at each position i , and N the length of the gene.

2.11. Principal Component Analysis

Principal component analysis was performed on R v3.6.3 using the FactoMineR package available on CRAN. Briefly, variants with frequencies > 0.5% were filtered, and only non-synonymous mutations which were reported in at least two different passages in the same replicate were used to create a matrix of all SNV frequencies across replicates and passages. Allele frequencies were \log_{10} -transformed to reduce heteroscedasticity of the data. PCA was performed on all passages but P1, and passage 1 components predicted *a posteriori*.

2.12. Statistical analysis

R v3.6.3 was used to perform statistical analysis and generate graphs. Error bars correspond to standard deviations (SD) unless otherwise indicated.

3. Results

3.1. Alternating passages have restricted genetic diversity determined by permissivity of both hosts

To assess the effect of host alternation on the accumulation of genetic diversity across the genome, we deep-sequenced evolved populations from the initial phase of experimental evolution at passage 1, 5, 10, 15, 19 and 20, and analyzed variant composition using our in-house pipeline ViVan²⁷⁴. We observed at terminal passages (19 and 20) signs of positive selection with mainly non-synonymous mutations reaching frequencies over 1%. While nsP1 and E2 genes appear to accumulate mutations in all regime types, the serially-passaged replicates in U.4.4 showed signs of positive selection in nsP2, E1 and in the two UTRs at the extremities of the viral genome. Populations passaged in BHK only showed additional clear selection patterns at a specific site at the 5' half of nsP4. Interestingly, the alternated lineages acquired less mutations at moderate to high frequencies overall, and selection patterns only matched regions common in the two serially-passaged lineages (**Figure 23A-B**).

In order to observe changes in genetic diversity over passages at the gene-level, we computed the average Shannon entropy of each lineage at every passages. We observed overall that in most cases, genetic diversity increased over passages, although at different levels and rates depending on the gene. This is at the exception of nsP1 and the 5' UTR, which showed a decrease in genetic diversity with passages. Given nsP1 showed signs of positive selection, this diminution could be the result of a selective sweep on the 5' region which occurred in all regimes. Alternatively, sign epistasis of fixed mutations could limit the accumulation of new ones. Of interest, E2 which also had signs of positive selection showed on the opposite an increase in entropies, suggesting that fixation of mutations did not displace other variants, or that genetic diversity reconstituted faster than the rate of variant loss. Moreover, as already suggested by **Figure 22A-B**, viruses passaged in U4.4 increased in genetic diversity faster in nsP2. Lineages passaged in BHK had significantly higher SNV diversity in the E2 gene across all passages, which suggests that mutations are readily more tolerated by mammalian cells. Finally, at all genomic regions, the Shannon entropies of viruses passaged alternately consistently followed the dynamic of the most restrictive host (**Figure 22C**). This observation agrees with the prediction of the trade-off hypothesis that in alternation, arboviruses are limited by their most stringent host. Of importance here, this pattern is observable at the gene-level.

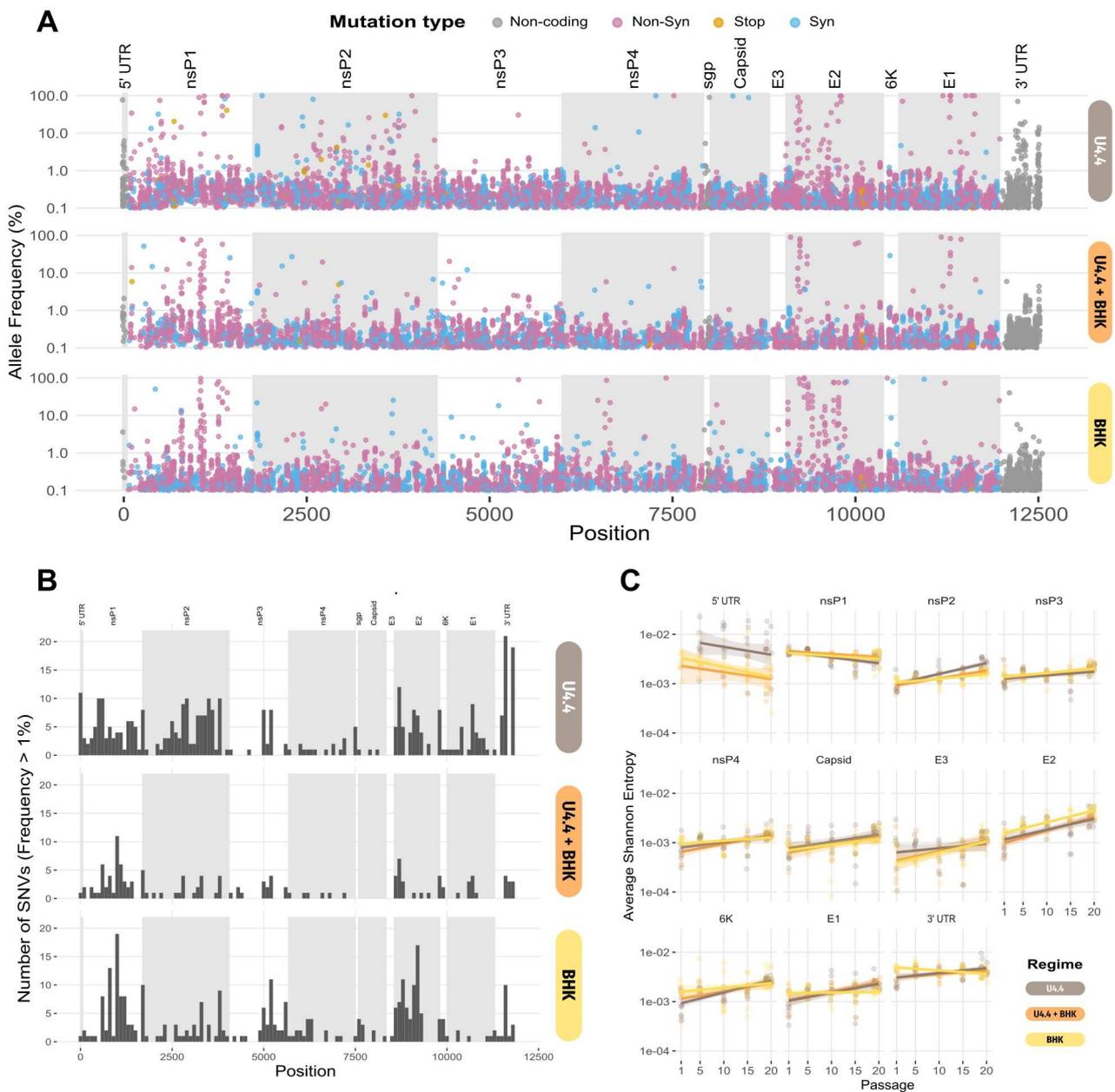


Figure 22 | Alternating passages accumulate less genetic diversity than single-host ones across passages. (A) Allele frequencies of SNVs with frequency over 0.1%. Shaded rectangles indicate gene positions. Colors correspond to mutation types: non-coding (gray), non-synonymous (pink), non-sense (orange) and synonymous (blue). The y-axis is \log_{10} -transformed for better visualization of minority SNVs. (B) Distribution of SNVs with frequency > 1% across the genome. Bins correspond to 100-nucleotide windows. (C) Genetic diversity across passages for each gene. Diversity is measured by the average Shannon entropy of individual positions for SNVs with frequency > 0.1%. In A and B, all replicates are shown.

3.2. Alternation between insects and mammals limits accumulation of defective viral genomes

Previous works in CHIKV have identified specific hotspots of deletion emerging in viral populations depending on the cell types they were generated²¹⁴. This is indicative of differences in host selective pressures on DelVG formation and maintenance. To test if alternation affected DVG abundance and species types, we compared deletion points between our different selective regimes using ViReMa²⁷³. While at initial passages (P1, top panels), mostly small deletions were observed across all regime types, at terminal passages (P19, bottom panels), larger deletions accumulated at high frequencies, suggestive of DVG replication. Strikingly, mosquito-passaged viruses accumulated more DVGs than other regimes (**Figure 23B**), and had three specific neighborhoods of deletion: neighborhood A spanning from the 5' UTR to nsP2, neighborhood B with internal deletions of nsP2, and neighborhood C which spans from the 5' to the 3' UTR (**Figure 23A**). Mammalian-passaged viruses on the other hand showed significantly less DVG amounts, and had a specific species of DVG (neighborhood D) which spanned from E2 or downstream to the 3' UTR, resulting in deleted structural protein genes (**Figure 23A**). Quantification of overall DVG abundance showed that alternately-passaged populations had significantly less DVGs. In addition, only marginal numbers of DVGs in host-specific neighborhoods were observed (**Figure 23B**). Overall, these observations indicate that mosquito cells are more permissive to DelVGs, but reciprocal intolerance of each cell type to the other's DVG species limits their accumulation in the alternating regime.

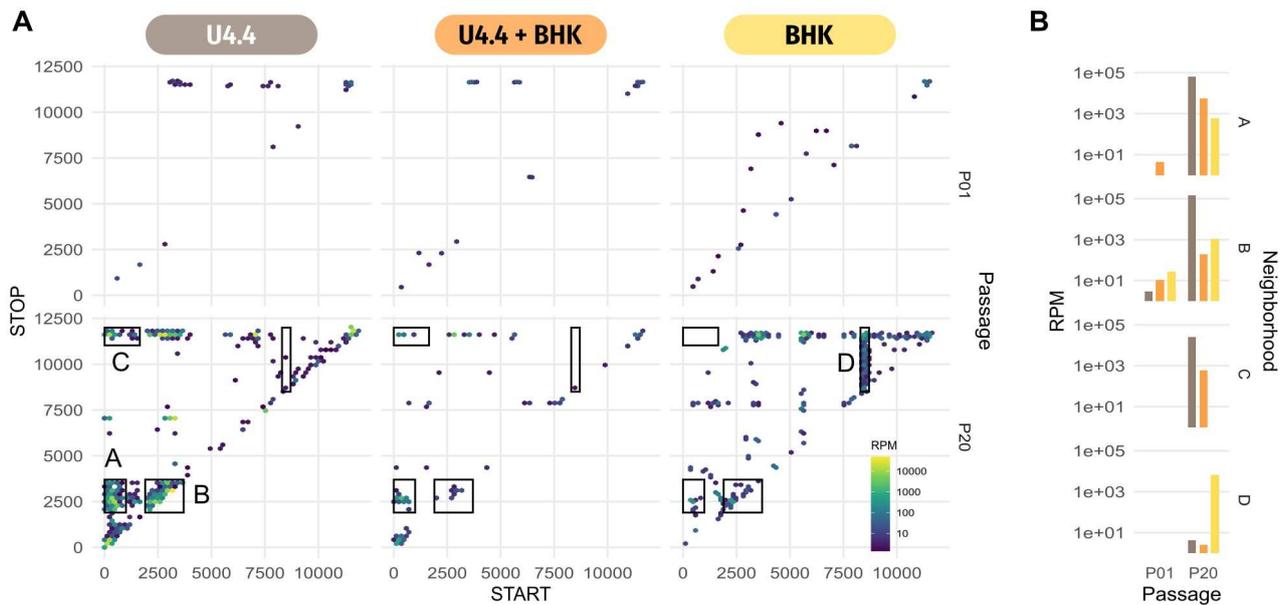


Figure 23 | DVG accumulation is limited by the two host tolerance in alternating passages. (A) Deletion hotspots in serially- or alternately-passaged viruses. Host-specific neighborhoods are delimited by rectangles. (B) Quantification of deletion breakpoints genome-wide or in host-specific neighborhoods. RPM: reads per million total aligned reads, represented on a log₁₀ scale.

3.3. Evolution of alternating passages follows insect-cell adaptive pathway while also selecting for alternation-specific mutations

Previous studies reported individual mutations to assess if common variants were shared among different regimes. However, because we used deep-sequencing approaches, reports of the entire mutant cloud of each replicate is difficult to comprehend. Rather, similarly to Dolan *et al.*⁸⁰, we reduced the dimensions of our data by performing a principal component analysis (PCA) on the frequency matrix of all SNVs that are at least shared by two different replicates regardless of regime type, and over all sequenced passages but the initial P1. The analysis resulted in separation of the three selective regimes across passages, indicating parallel evolution between replicates of the same regime. Moreover, principal component 1 (PC1) separated insect- and mammalian-only selective regimes, accounting for 19.5% of the total variation. Of importance, the alternating regime evolved towards values corresponding to insect-specific mutations on PC1, indicating it shared more common variants with insect-adapted lineages. Surprisingly, it also separated from the other regimes on PC2, which was driven by few non-synonymous mutations, notably E2-W64R (**Figure 24A**). This suggests that the replicates in the alternating regime accumulated specific

mutations, specifically in E2 and nsP1, which do not appear at similar levels or parallelism in other regimes.

Inspections of mutation contributions to each PC allows us to also identify mutations specific to U4.4 cells, including the E2-V226A mutation which confers adaptation to *Aedes albopictus*³²², and a few in nsP1 and E2. Mutations specific of BHK-21 cells mainly cluster in E2 and nsP1 at high densities (**Figure 24B**). While these high numbers of mutations clustered in the E2 genes are consistent with previous observations of genetic diversity (**Figure 22C**), this is nonetheless surprising and suggests extensive conformational changes of the E2 protein.

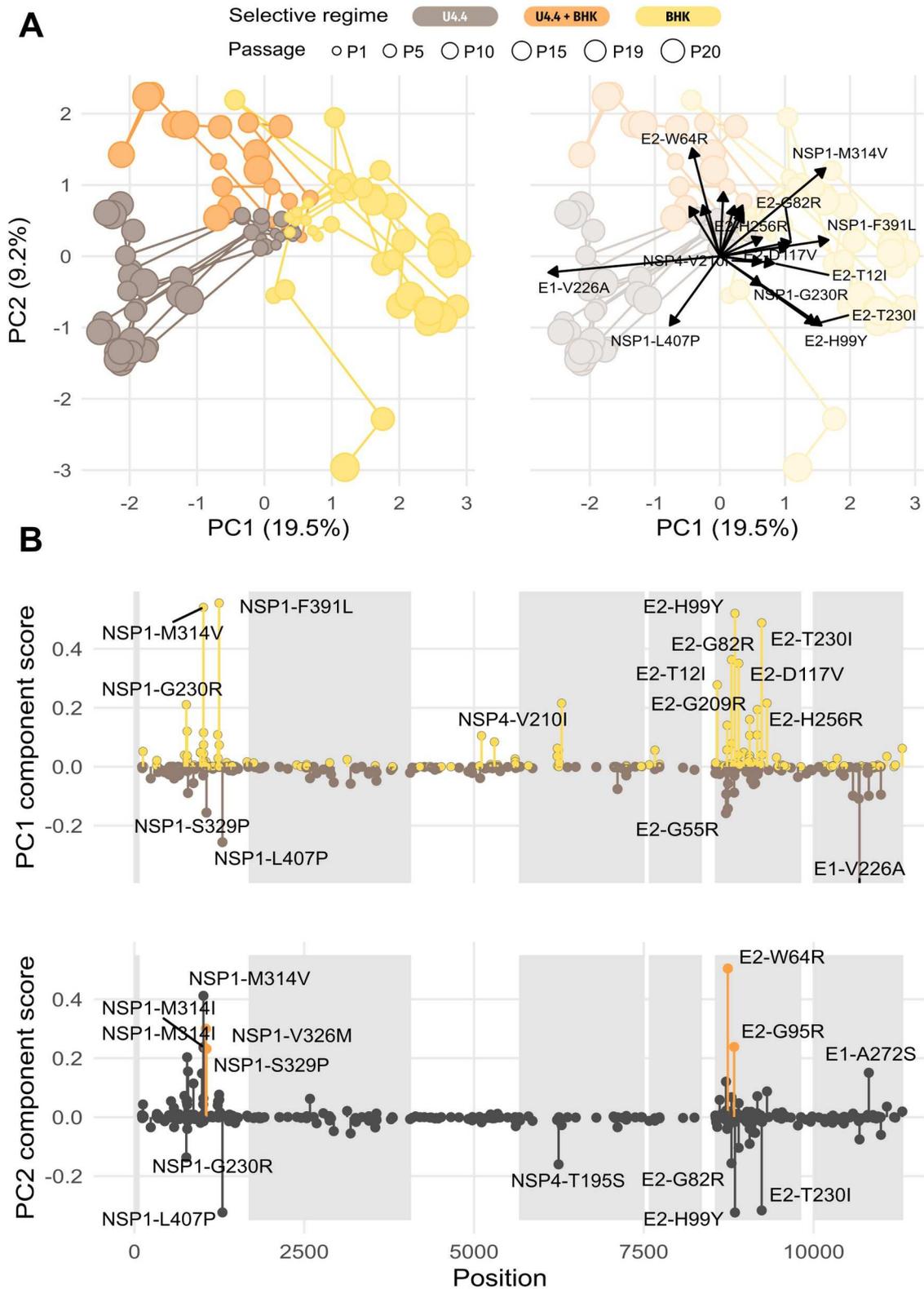


Figure 24 | Adaptive regimes show specific patterns of mutation acquisition. (A) Principal component analysis of non-synonymous mutations with frequencies superior to 0.5%. Colors indicate selective regime, dot sizes extracted passages. Lines represent adaptive paths of individual biological replicates. Black arrows indicate the contribution of each variant to the principal components (PC). (B) Component scores of SNVs. Colors indicate putative specific mutations to the corresponding selective regime.

3.4. Initial phase of experimental evolution does not detect cost of generalism and shows an asymmetric trade-off cost

We further submitted all passages that were sequenced to competition assays against a non-passaged marked surrogate competitor virus in either BHK-21 or U4.4 cells in three sequential competition passages. Relative abundance of each competitor was measured through a high-accuracy multiplex RT-qPCR approach, and changes in relative frequencies over the three passages determined. The slope of the linear regression on the relative frequency ratios is then used as our relative fitness proxy. In the majority of cases, linearity was observed, confirming the additivity of log-transformed fitness across passages (**Figure 25**).

Viruses passaged in BHK-21 cells consistently displayed a cost of adaptation in the bypassed U4.4 cells. On the other hand, populations passaged in insect cells consistently increased their fitness in the adapted host, but individual replicates showed all possible scenarios: minor cost of adaptation, no-cost neutrality, or even increase in fitness for one replicate (**Figure 26A**). On average, no cost of adaptation was reported (**Figure 26B**). These results indicate that while adaptation to mammalian cells necessarily incur a fitness cost in insect cells, the opposite is not true. Adaptation to U4.4 can randomly accumulate mutations that have antagonistic pleiotropic effects, but the possible adaptive mutations have a broader range of effects in mammals, resulting in no clear cost of adaptation. Interestingly, viruses passaged alternately between mammalian and insect cells showed consistent fitness increases in both hosts, although to lower magnitudes in BHK cells. Overall, no significant cost of generalism was reported in U4.4 and BHK cells (**Figure 26A-B**). This is consistent with the sequencing observations in which alternating passages followed to some degrees the adaptive pathway of insect cells, which offers the possibility for fitness increase in the other host.

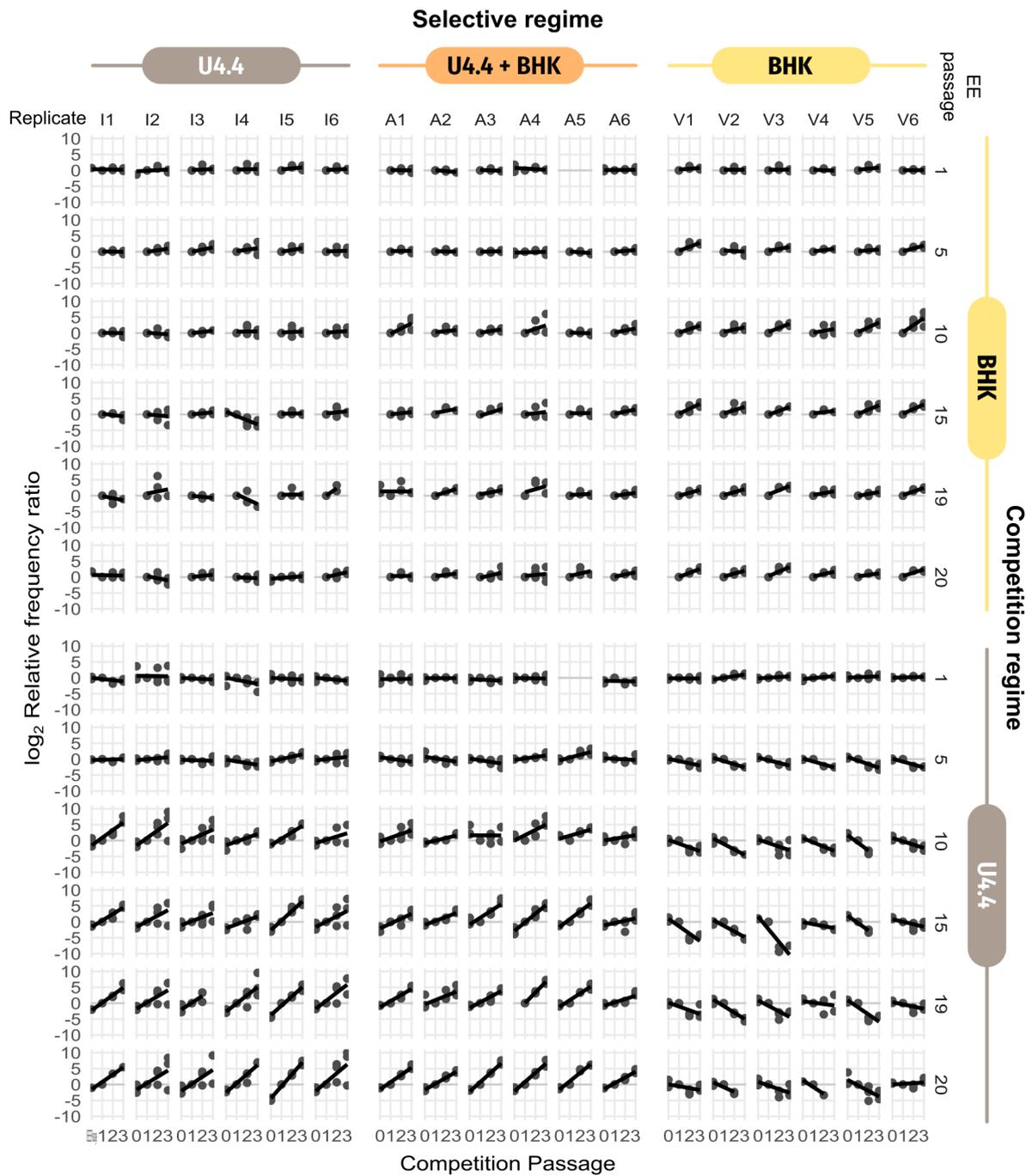


Figure 25 | Competition assays in single-host challenges show linearity of fitness effects over passages. The y-axis represent the \log_2 relative frequency ratios of assayed populations against the surrogate competitor. The frequency ratios were normalized against the first competitive passage (p1) as some inocula (p0) had degraded vRNA and therefore undetermined frequencies. Each competition experiment was run in biological triplicates. Slopes of regression lines give the average relative fitness of the viral population over one passage (used in **Figure 26**).

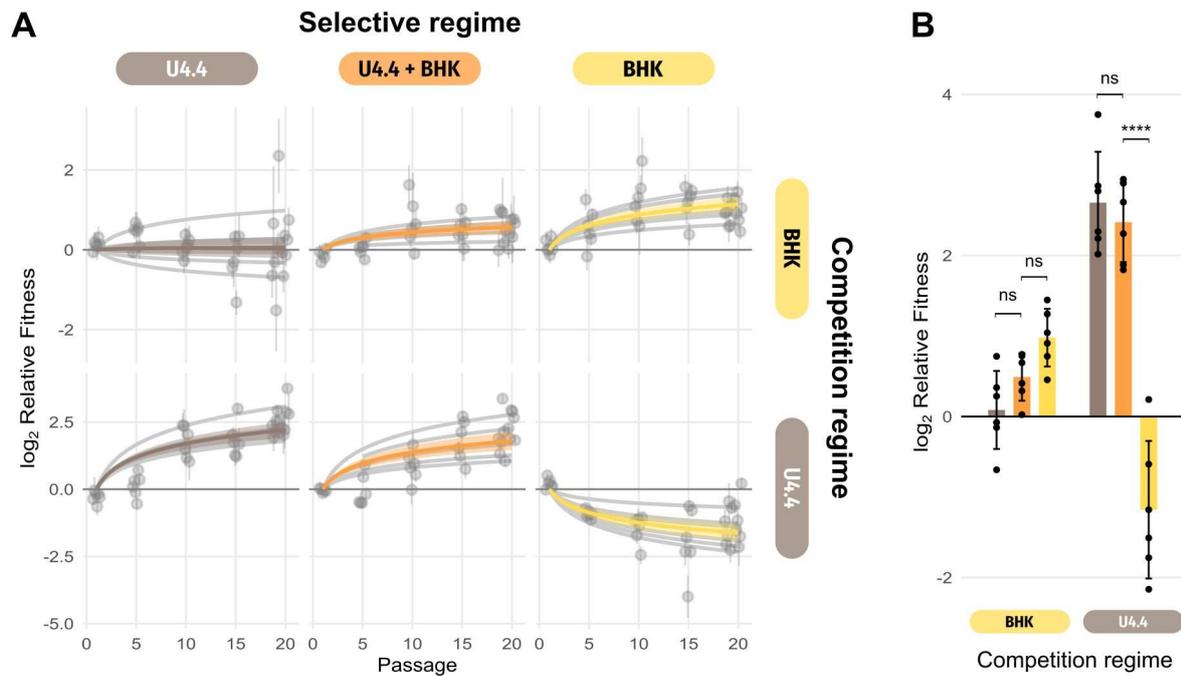


Figure 26 | Competition assays reveal an asymmetric trade-off in serial passages and no significant cost of generalism in alternating ones. (A) Relative fitness changes across experimental evolution passages. Estimated relative fitness \pm standard error of each passaged population is derived from the linear regression performed in **Figure 25**, and represented in light gray. Individual replicate lineage fitness trajectories are fitted in light gray. Average fitness trajectory of each selective regime is fitted on average fitness values at each passage and represented in their respective colors. (B) Comparisons of relative fitness values at passage 20. **** $p < 0.001$; ns: non-significant; Student *t*-tests with Holm's post-hoc correction for multiple tests.

Unfortunately, despite our best efforts, results from the second phase of experimental evolution showed signs of cross-contamination between samples, rendering the results unexploitable. These experiments will be repeated with caution in the future.

4. Discussion and perspectives

4.1. Summary of results

To re-investigate the arbovirus trade-off hypothesis, we proposed an experimental design in two experimental evolution phases to first push viral populations to the trade-off Pareto front, then in a second time identify fitness costs after the regime switch. To that aim, we used barcoded CHIKV clones and passaged it serially or alternately in BHK-21 and U4.4 cells. Sequencing was performed using deep sequencing, which allowed us to uncover that genetic diversity accumulated differentially along the virus genome and depending on the selective regime. Overall, all regimes accumulated diversity mainly in nsP1 and E2, although at different rates. At every region of the genome, the alternating regime diversity was restricted to the level of the most stringent host, suggesting antagonistic pleiotropy predicted in the trade-off hypothesis. Similar observations were made by looking at DelVG accumulation over passages: overall numbers of DVGs and inside host-specific deletion neighborhoods were significantly lower, suggesting higher constrains in alternation. Surprisingly, the insect cells showed relaxed purifying selection against DVGs, with very high accumulation of some species. Analysis of mutant cloud composition through PCA showed that although the alternating regime followed to some degrees the molecular evolution imposed by the insect host, independent accumulation of specific mutations was also identified, notably the E2-W64R substitution. High molecular parallelism between replicates was reported. Finally, the first phase of experimental evolution reported an asymmetric cost of adaptation for viruses adapted to BHK-21 cells, but not in insect cells. The alternating passages did not show signs of cost of generalism and gained fitness in both cell lines at similar levels to the specialist regimes. Unfortunately, the second phase of the experimental design was unexploitable, rendering us incapable of concluding on the existence of the trade-off.

4.2. Genetic diversity varies along the virus genome

In our study, we showed that genetic diversity accumulated differentially over the viral genome, with especially nsP1 and E2 genes accumulating more mutations than other sites. Moreover, diversity patterns were different among selective regimes: while the insect cells showed more diversity in nsP2 and the terminal UTRs, mammalian cells showed a localized increase of diversity in nsP4. This indicates that in our experimental design, viruses passaged

in insect cells tend to have more relaxed purifying selection on a bigger part of their genomes. Additionally, we found that insect-passaged viruses also tolerated higher amounts of DVGs overall, suggesting again that they are more permissive to mutations. An alternative explanation would be that the genetic diversity found in U4.4 in nsP2 is localized on DelVGs which have been shown to have more relaxed selection in chapter 1 or in other studies²⁸², therefore biasing our diversity measurements. Nevertheless, this contradicts previous studies that found that the mosquito host puts a break on diversity accumulation in arboviruses^{289,290,294,295}, and rather align with opposite observations^{302,305}.

Because genetic diversity varies along the genome, one possible explanation of this discrepancies could be that some studies only calculated genetic diversity in some particular regions. For example, Coffey *et al.* only amplified E1 sequences in their samples for diversity measurements because it was previously reported that it tolerated more mutations^{294,295}. Moreover, because all these previous reports used low-throughput sequencing methods, only mutations with more than 5-10% frequencies were used in their analysis. This incorporates another bias as mutations at such frequencies are for the majority either beneficial mutations under selection, or hitchhiking neutral mutations. Our high-throughput sequencing approach allowed us to go down as far as 0.1% SNV frequencies, which allowed a more accurate depiction of neutral variation in the viral genome. Ultra-deep sequencing techniques such as CirSeq²⁷⁵ can even go to 10^{-6} mutation frequencies, therefore being able to accurately identify *de novo* mutations being produced and draw an accurate distribution of fitness effects along the genome. This approach has been realized by Dolan *et al.* with dengue virus in serial passages only, and showed that adaptation to human cells had higher amounts of sites under positive selection⁸⁰. Consequently, because mammalian cells select positively for more sites, which was also suggested in the non-synonymous cluster in E2 in our results, previous measure of diversity could have been biased to report higher diversity in such models.

Nonetheless, we clearly showed a pattern in which the alternated-passaged viruses were restricted in genetic diversity accumulation depending on the stringency of both hosts at each locus.

4.3. The insect host partially drives the evolution of the alternating passages

In our results, we also showed that the mutant cloud and substitution pattern of the alternating regime followed at least partially the trajectory of viruses adapted to mosquito cells only. This observation could be linked to the selection of the E2-A226V substitution in this condition, as it contributes to the majority of the negative component score of PC1, and has been selected on multiple occasions in natural outbreaks^{315,322,323}, suggesting it does not incur a fitness cost to mammalian hosts. However, we re-performed the PCA by excluding the E2-A226V substitution, and did not find significant changes in the PCA topology apart from reduced explained variance on both PCs (data not shown), therefore suggesting that other mutations contributed to this similar adaptive pathway. By associating these results with subsequent fitness assays, we theorize that the main driver that pushes alternating passages to follow the insect-specific adaptive pathway is the absence of initial fitness cost to first adapt to U4.4. Indeed, because they must adapt to both hosts, the evolutionary pathway taken by mammalian-adapted lineages would be too costly, while fitness gain in one host without adaptation cost still benefits the alternating population.

Moreover, specific mutations in alternating passages associated with fitness gain in mammalian cells were observed. We propose that adaptation to insect cells pushed the alternating regime in a new neighborhood where it could acquire beneficial mutations to mammalian cells without paying a fitness cost in insect cells. Report of one lineage of serially-passaged virus in U4.4 cells stochastically gaining fitness in the by-passed host suggests that such pathways exist and that alternation drives the population towards it. This would signify that while insect-specific mutations do not display antagonistic pleiotropy, they induce antagonistic or sign epistasis on loci responsible for mammalian adaptation.

Overall, we propose that the reason for similar adaptive pathways between insect- and alternately-passaged populations rely solely on the absence of cost of adaptation towards insect-specific adaptation. Further, this drives the alternated population to a new neighborhood in sequence space where further adaptations to mammalian cells do not induce adaptation cost. This conclusion opposes the hypothesis that the reason the vector host tend to drive arbovirus evolution is due to more restricted genetic diversity^{290,295}, stringent purifying selection⁸⁰ or more severe bottlenecks^{89,91}. Additionally, this suggests that punctual asymmetric trade-offs in the fitness landscape allow to avoid evolutionary stasis^{288,298}.

4.4. Alternation-specific mutations

The identification of alternation specific mutations raises questions on their specific fitness effects. Indeed, if they provide beneficial fitness effects to both hosts, why do they not also emerge at similar levels in the serially-passaged ones? One mutation in particular, E2-W64R, drove most of the component score on PC2 towards the alternated-passage trajectory. Interestingly, this mutation has been reported in multiple studies. Because the residue 64 in E2 resides at the interaction between two E2 proteins in their homodimer form³³³, mutations are expected to affect CHIKV spike assembly and conformation. Moreover, the residue is exposed at the spike surface and interacts with the Mxra8 mammalian receptor³⁰⁹, explaining why it has been reported in antibody escape³³³ and in patient samples³³⁴. The acquisition of the W64R substitution has been reported to decrease viral dissemination in *Aedes aegypti* but not in *Aedes albopictus* and to interact epistatically with the E2-A226V substitution³¹⁸. The latter observation could have explained the emergence of the 64R residue, but we did not observe concomitant fixation of A226V and W64R in our samples (data not shown). Nevertheless, extensive epistatic interactions between mutations affecting CHIKV glycoproteins have been demonstrated^{114,289,324,325}. Combined with our results showing extensive non-synonymous mutation accumulation in E2 in mammalian-passaged lineages, emergence of W64R could be inhibited by the accumulation of more beneficial initial mutations in E2, which in turns interact through negative epistasis, blocking further acquisition of the mutation. This follows our previous hypothesis that initial adaptation to mammalian cells could not only lead viruses to an adaptation cost for insect cells, but also could trap viral populations in evolutionary dead-ends. Future investigations should thrive to test if such epistatic interactions with mammalian-selected E2 mutations could inhibit W64R fitness effects.

Alternatively, W64R could not necessarily confer fitness benefits in either host on their own, but act as a specific beneficial mutation when viruses shed by one cell type infects the other. This hypothesis could be tested through entry assays.

4.5. The trade-off hypothesis: a pending question & future investigations

While we were able to draw some conclusions on the fitness landscape of alternating-regime adaptation from the initial phase of our experimental evolution design, these results are still subject to the same criticisms and limitations than previous studies. Repetition of the

second phase of experimental evolution and associated experiments described in the introduction are still lacking to conclude on the nature of the trade-off hypothesis, and the possibility that it slows down the substitution rate of arboviruses. Nevertheless, some questions can still be resolved without extensive passaging and will be performed in the future.

First, we want to test the hypothesis that mixes of specialist populations can interact through complementation and limit the cost of adaptation and generalism. To that end, we will co-infect either initial or late passages of our serially-passaged viruses with our surrogate ancestor clone at different multiplicity of infection, and perform early-replication viral titration and total or surrogate-genome specific RT-qPCR quantification to assess if, when at sufficient MOIs to allow co-infection, replication of the surrogate is improved when co-infected with adapted lineages. This would first inform us if evolved passages are capable of complementation. Moreover, sequencing of successive alternating passages of mixed populations would inform us if one population takes the advantage over the other, rejecting sub-population structures, or rather if complementation allows mixed populations to remain stable over time. In addition, sequencing of two successive passages of our evolved lineages in all cell types should allow us to identify sites subject to antagonistic pleiotropy.

Finally, a study from Novella *et al.* showed that infecting a previously mammalian-adapted lineage that was then adapted to persistent insect cell infection back to mammalian cells resulted in fast adaptation to the mammalian host²⁸⁸. This was mediated by a minority variant that was previously selected for mammalian adaptation, but was subsequently relegated to low frequencies during persistent infection of insect cells²⁸⁸. Existence of such memory genomes could also accelerate adaptation to both hosts even under trade-offs, and result in stasis by back-and-forth adaptation through expansion of previously acquired substitutions. This could be explored in a medium- to long-term extension of our experimental evolution design.

BIBLIOGRAPHY

REFERENCES

1. Wasik, B. R. & Turner, P. E. On the Biological Success of Viruses. *Annu. Rev. Microbiol.* **67**, 519–541 (2013).
2. Bergh, Ø., Børshem, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
3. Harvey, E. & Holmes, E. C. Diversity and evolution of the animal virome. *Nat. Rev. Microbiol.* **20**, 321–334 (2022).
4. Shi, M. *et al.* The evolutionary history of vertebrate RNA viruses. *Nature* **556**, 197–202 (2018).
5. Koskella, B., Hernandez, C. A. & Wheatley, R. M. Understanding the Impacts of Bacteriophage Viruses: From Laboratory Evolution to Natural Ecosystems. *Annu. Rev. Virol.* **9**, null (2022).
6. Proctor, L. M. & Fuhrman, J. A. Viral mortality of marine bacteria and cyanobacteria. *Nature* **343**, 60–62 (1990).
7. Dávila-Ramos, S. *et al.* A Review on Viral Metagenomics in Extreme Environments. *Front. Microbiol.* **10**, (2019).
8. Desnues, C., Boyer, M. & Raoult, D. Chapter 3 - Sputnik, a Virophage Infecting the Viral Domain of Life. in *Advances in Virus Research* (eds. Łobocka, M. & Szybalski, W. T.) vol. 82 63–89 (Academic Press, 2012).
9. Holmes, E. C. *The evolution and emergence of RNA viruses*. (Oxford University Press, 2009).
10. Koonin, E. V. & Krupovic, M. The depths of virus exaptation. *Curr. Opin. Virol.* **31**, 1–8 (2018).
11. Forterre, P. & Prangishvili, D. The origin of viruses. *Res. Microbiol.* **160**, 466–472 (2009).
12. Geoghegan, J. L. & Holmes, E. C. Evolutionary Virology at 40. *Genetics* **210**, 1151–1162 (2018).
13. Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735–744 (1978).
14. Eggers, H. J. & Tamm, I. Cocksackie A9 Virus: Mutation from Drug Dependence to Drug Independence. *Science* **148**, 97–98 (1965).
15. Young, J. F., Desselberger, U. & Palese, P. Evolution of human influenza A viruses in nature: Sequential mutations in the genomes of new H1N1 isolates. *Cell* **18**, 73–83 (1979).
16. Young, J. F. & Palese, P. Evolution of human influenza A viruses in nature: recombination contributes to genetic variation of H1N1 strains. *Proc. Natl. Acad. Sci.* **76**, 6547–6551 (1979).

17. Domingo, E. *et al.* Viruses as Quasispecies: Biological Implications. in *Quasispecies: Concept and Implications for Virology* (ed. Domingo, E.) 51–82 (Springer, 2006). doi:10.1007/3-540-26397-7_3.
18. Domingo, E., Sheldon, J. & Perales, C. Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* **76**, 159–216 (2012).
19. Domingo, E. & Perales, C. Viral quasispecies. *PLOS Genet.* **15**, e1008271 (2019).
20. Holmes, E. C. & Moya, A. Is the Quasispecies Concept Relevant to RNA Viruses? *J. Virol.* **76**, 460–462 (2002).
21. Sanjuán, R. The Social Life of Viruses. *Annu. Rev. Virol.* (2021) doi:10.1146/annurev-virology-091919-071712.
22. Díaz-Muñoz, S. L., Sanjuán, R. & West, S. Sociovirology: Conflict, Cooperation, and Communication among Viruses. *Cell Host Microbe* **22**, 437–441 (2017).
23. Rezelj, V. V. *et al.* Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. *Nat. Commun.* **12**, 1–14 (2021).
24. Sanjuán, R. From Molecular Genetics to Phylodynamics: Evolutionary Relevance of Mutation Rates Across Viruses. *PLOS Pathog.* **8**, e1002685 (2012).
25. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).
26. Chao, L., Tran, T. & Matthews, C. Muller's ratchet and the advantage of sex in the RNA virus ϕ 6. *Evolution* **46**, 289–299 (1992).
27. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. *Nature* **398**, 441–443 (1999).
28. Gago, S., Elena, S. F., Flores, R. & Sanjuán, R. Extremely High Mutation Rate of a Hammerhead Viroid. *Science* **323**, 1308–1308 (2009).
29. Pfeiffer, J. K. & Kirkegaard, K. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc. Natl. Acad. Sci.* **100**, 7289–7294 (2003).
30. Vignuzzi, M., Wendt, E. & Andino, R. Engineering attenuated virus vaccines by controlling replication fidelity. *Nat. Med.* **14**, 154–161 (2008).
31. Beaucourt, S. *et al.* Isolation of Fidelity Variants of RNA Viruses and Characterization of Virus Mutation Frequency. *JoVE J. Vis. Exp.* e2953 (2011) doi:10.3791/2953.
32. Coffey, L. L., Beeharry, Y., Bordería, A. V., Blanc, H. & Vignuzzi, M. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc. Natl. Acad. Sci.* **108**, 16038–16043 (2011).
33. Gnädig, N. F. *et al.* Cocksackievirus B3 mutator strains are attenuated in vivo. *Proc. Natl. Acad. Sci.* **109**, E2294–E2303 (2012).
34. Warmbrod, K. L. *et al.* Viral RNA-dependent RNA polymerase mutants display an altered mutation spectrum resulting in attenuation in both mosquito and vertebrate hosts. *PLoS Pathog.* **15**, e1007610 (2019).

35. Caldwell, H. S., Ngo, K., Pata, J. D., Kramer, L. D. & Ciota, A. T. Y. 2020. West Nile Virus fidelity modulates the capacity for host cycling and adaptation. *J. Gen. Virol.* **101**, 410–419 (2020).
36. Rozen-Gagnon, K. *et al.* Alphavirus Mutator Variants Present Host-Specific Defects and Attenuation in Mammalian and Insect Models. *PLoS Pathog.* **10**, e1003877 (2014).
37. Fitzsimmons, W. J. *et al.* A speed–fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol.* **16**, e2006459 (2018).
38. Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* **16**, e3000003 (2018).
39. Poulain, F., Lejeune, N., Willemart, K. & Gillet, N. A. Footprint of the host restriction factors APOBEC3 on the genome of human viruses. *PLoS Pathog.* **16**, e1008718 (2020).
40. Vieira, V. C. & Soares, M. A. The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections. *BioMed Res. Int.* **2013**, 683095 (2013).
41. Milewska, A. *et al.* APOBEC3-mediated restriction of RNA virus replication. *Sci. Rep.* **8**, 5960 (2018).
42. Gélinas, J.-F., Clerzius, G., Shaw, E. & Gatignol, A. Enhancement of Replication of RNA Viruses by ADAR1 via RNA Editing and Inhibition of RNA-Activated Protein Kinase γ . *J. Virol.* **85**, 8460–8466 (2011).
43. Weiden, M. D. *et al.* Adenosine Deaminase Acting on RNA-1 (ADAR1) Inhibits HIV-1 Replication in Human Alveolar Macrophages. *PLoS ONE* **9**, e108476 (2014).
44. Khalfi, P. *et al.* APOBEC3C S188I Polymorphism Enhances Context-Specific Editing of Hepatitis B Virus Genome. *J. Infect. Dis.* jiac003 (2022) doi:10.1093/infdis/jiac003.
45. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626 (2011).
46. Mac Kain, A., Joffret, M.-L., Delpeyroux, F., Vignuzzi, M. & Bessaud, M. A cold case: non-replicative recombination in positive-strand RNA viruses. *Virol. Montrouge Fr.* **25**, 62–73 (2021).
47. Pierangeli, A. *et al.* ‘Primer alignment-and-extension’: a novel mechanism of viral RNA recombination responsible for the rescue of inactivated poliovirus cDNA clones. *J. Gen. Virol.* **80**, 1889–1897.
48. Gallei, A., Pankraz, A., Thiel, H.-J. & Becher, P. RNA Recombination In Vivo in the Absence of Viral Replication. *J. Virol.* **78**, 6271–6281 (2004).
49. Chetverin, A. B., Chetverina, H. V., Demidenko, A. A. & Ugarov, V. I. Nonhomologous RNA Recombination in a Cell-Free System: Evidence for a Transesterification Mechanism Guided by Secondary Structure. *Cell* **88**, 503–513 (1997).
50. Poirier, E. Z. *et al.* Low-Fidelity Polymerases of Alphaviruses Recombine at Higher Rates To Overproduce Defective Interfering Particles. *J. Virol.* **90**, 2446–2454 (2016).

51. Taucher, C., Berger, A. & Mandl, C. W. A trans-Complementing Recombination Trap Demonstrates a Low Propensity of Flaviviruses for Intermolecular Recombination. *J. Virol.* **84**, 599–611 (2010).
52. Muslin, C., Mac Kain, A., Bessaud, M., Blondel, B. & Delpeyroux, F. Recombination in Enteroviruses, a Multi-Step Modular Evolutionary Process. *Viruses* **11**, 859 (2019).
53. Baird, H. A. *et al.* Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res.* **34**, 5203–5216 (2006).
54. Staroseletz, Y. *et al.* Non-enzymatic recombination of RNA: Ligation in loops. *Biochim. Biophys. Acta Gen. Subj.* **1862**, 705–725 (2018).
55. Tao, H., Steel, J. & Lowen, A. C. Intrahost Dynamics of Influenza Virus Reassortment. *J. Virol.* **88**, 7485–7492 (2014).
56. Tao, H., Li, L., White, M. C., Steel, J. & Lowen, A. C. Influenza A Virus Coinfection through Transmission Can Support High Levels of Reassortment. *J. Virol.* **89**, 8453–8461 (2015).
57. Lowen, A. C. Constraints, Drivers, and Implications of Influenza A Virus Reassortment. *Annu. Rev. Virol.* **4**, 105–121 (2017).
58. Gerber, M., Isel, C., Moules, V. & Marquet, R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* **22**, 446–455 (2014).
59. Nibert, M. L., Margraf, R. L. & Coombs, K. M. Nonrandom segregation of parental alleles in reovirus reassortants. *J. Virol.* **70**, 7295–7300 (1996).
60. Xiao, Y. *et al.* RNA Recombination Enhances Adaptability and Is Required for Virus Spread and Virulence. *Cell Host Microbe* **22**, 420 (2017).
61. Filomatori, C. V. *et al.* RNA recombination at Chikungunya virus 3UTR as an evolutionary mechanism that provides adaptability. *PLOS Pathog.* **15**, e1007706 (2019).
62. Acevedo, A. *et al.* Genetic recombination of poliovirus facilitates subversion of host barriers to infection. *bioRxiv* 273060 (2018) doi:10.1101/273060.
63. Hahn, C. S., Lustig, S., Strauss, E. G. & Strauss, J. H. Western equine encephalitis virus is a recombinant virus. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5997–6001 (1988).
64. Combelas, N., Holmblat, B., Joffret, M.-L., Colbère-Garapin, F. & Delpeyroux, F. Recombination between Poliovirus and Coxsackie A Viruses of Species C: A Model of Viral Genetic Plasticity and Emergence. *Viruses* **3**, 1460–1484 (2011).
65. Lytras, S. *et al.* Exploring the natural origins of SARS-CoV-2 in the light of recombination. 2021.01.22.427830 Preprint at <https://doi.org/10.1101/2021.01.22.427830> (2021).
66. Steel, J. & Lowen, A. C. Influenza A Virus Reassortment. in *Influenza Pathogenesis and Control - Volume I* (eds. Compans, R. W. & Oldstone, M. B. A.) 377–401 (Springer International Publishing, 2014). doi:10.1007/82_2014_395.
67. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res. Mol. Mech. Mutagen.* **1**, 2–9 (1964).
68. Chao, L. Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455 (1990).

69. Duarte, E., Clarke, D., Moya, A., Domingo, E. & Holland, J. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci.* **89**, 6015–6019 (1992).
70. Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal Interference and the Evolution of RNA Viruses. *Science* **285**, 1745–1747 (1999).
71. Poon, A. & Chao, L. Drift Increases the Advantage of Sex in RNA Bacteriophage $\Phi 6$. *Genetics* **166**, 19–24 (2004).
72. Kimura, M. Evolutionary Rate at the Molecular Level. *Nature* **217**, 624–626 (1968).
73. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, 1983).
74. Kern, A. D. & Hahn, M. W. The Neutral Theory in Light of Natural Selection. *Mol. Biol. Evol.* **35**, 1366–1371 (2018).
75. Jensen, J. D. *et al.* The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution* **73**, 111–114 (2019).
76. Hahn, M. W. Toward a Selection Theory of Molecular Evolution. *Evolution* **62**, 255–265 (2008).
77. Sanjuán, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci.* **101**, 8396–8401 (2004).
78. Peris, J. B., Davis, P., Cuevas, J. M., Nebot, M. R. & Sanjuán, R. Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage $\phi 1$. *Genetics* **185**, 603–609 (2010).
79. Vale, P. F., Choisy, M., Froissart, R., Sanjuán, R. & Gandon, S. THE DISTRIBUTION OF MUTATIONAL FITNESS EFFECTS OF PHAGE $\phi X174$ ON DIFFERENT HOSTS. *Evolution* **66**, 3495–3507 (2012).
80. Dolan, P. T. *et al.* Principles of dengue virus evolvability derived from genotype-fitness maps in human and mosquito cells. *eLife* **10**, e61921 (2021).
81. Novella, I. S., Zárata, S., Metzgar, D. & Ebendick-Corpus, B. E. Positive Selection of Synonymous Mutations in Vesicular Stomatitis Virus. *J. Mol. Biol.* **342**, 1415–1421 (2004).
82. Novella, I. S. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Curr. Opin. Microbiol.* **6**, 399–405 (2003).
83. Cervera, H., Lalić, J. & Elena, S. F. Effect of Host Species on Topography of the Fitness Landscape for a Plant RNA Virus. *J. Virol.* **90**, 10160–10169 (2016).
84. Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001).
85. Wisner, M. J., Ribbeck, N. & Lenski, R. E. Long-Term Dynamics of Adaptation in Asexual Populations. *Science* **342**, 1364–1367 (2013).

86. Patterson, E. I. *et al.* Mosquito bottlenecks alter viral mutant swarm in a tissue and time-dependent manner with contraction and expansion of variant positions and diversity. *Virus Evol.* **4**, vey001 (2018).
87. Zwart, M. P. & Elena, S. F. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2**, 161–179 (2015).
88. Gutiérrez, S., Michalakakis, Y. & Blanc, S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr. Opin. Virol.* **2**, 546–555 (2012).
89. Forrester, N. L., Guerbois, M., Seymour, R. L., Spratt, H. & Weaver, S. C. Vector-borne transmission imposes a severe bottleneck on an RNA virus population. *PLoS Pathog.* **8**, e1002897 (2012).
90. McCrone, J. T. & Lauring, A. S. Genetic bottlenecks in intraspecies virus transmission. *Curr. Opin. Virol.* **28**, 20–25 (2018).
91. Weaver, S. C., Forrester, N. L., Liu, J. & Vasilakis, N. Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. *Nat. Rev. Microbiol.* **19**, 184–195 (2021).
92. Kadoya, S. *et al.* Bottleneck Size-Dependent Changes in the Genetic Diversity and Specific Growth Rate of a Rotavirus A Strain. *bioRxiv* 702233 (2019) doi:10.1101/702233.
93. Clarke, D. K. *et al.* Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* **67**, 222–228 (1993).
94. Novella, I. S., Elena, S. F., Moya, A., Domingo, E. & Holland, J. J. Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *J. Virol.* **69**, 2869–2872 (1995).
95. Cases-González, C., Arribas, M., Domingo, E. & Lázaro, E. Beneficial effects of population bottlenecks in an RNA virus evolving at increased error rate. *J. Mol. Biol.* **384**, 1120–1129 (2008).
96. Sagan, S. M., Chahal, J. & Sarnow, P. cis-Acting RNA elements in the hepatitis C virus RNA genome. *Virus Res.* **206**, 90–98 (2015).
97. Steil, B. P. & Barton, D. J. Cis-active RNA elements (CREs) and picornavirus RNA replication. *Virus Res.* **139**, 240–252 (2009).
98. Sicard, A. *et al.* A multicellular way of life for a multipartite virus. *eLife* **8**, e43599 (2019).
99. Moreau, Y. *et al.* The Genome Segments of Bluetongue Virus Differ in Copy Number in a Host-Specific Manner. *J. Virol.* **95**, e01834-20 (2020).
100. Salemi, M., Vandamme, A.-M., Desmyter, J., Casoli, C. & Bertazzoni, U. The origin and evolution of human T-cell lymphotropic virus type II (HTLV-II) and the relationship with its replication strategy. *Gene* **234**, 11–21 (1999).
101. Sardanyés, J., Solé, R. V. & Elena, S. F. Replication Mode and Landscape Topology Differentially Affect RNA Virus Mutational Load and Robustness. *J. Virol.* **83**, 12579–12589 (2009).

102. Andreu-Moreno, I. & Sanjuán, R. Collective viral spread mediated by virion aggregates promotes the evolution of defective interfering particles. *mBio* **11**, (2020).
103. Sanjuán, R. Collective Infectious Units in Viruses. *Trends Microbiol.* **25**, 402–412 (2017).
104. Sanjuán, R. & Thoulouze, M.-I. Why viruses sometimes disperse in groups†. *Virus Evol.* **5**, vez014 (2019).
105. Moutailler, S. *et al.* Host Alternation Is Necessary to Maintain the Genome Stability of Rift Valley Fever Virus. *PLoS Negl. Trop. Dis.* **5**, e1156 (2011).
106. Elena, S. F. Little Evidence for Synergism Among Deleterious Mutations in a Nonsegmented RNA Virus. *J. Mol. Evol.* **49**, 703–707 (1999).
107. Bonhoeffer, S., Chappey, C., Parkin, N. T., Whitcomb, J. M. & Petropoulos, C. J. Evidence for Positive Epistasis in HIV-1. *Science* **306**, 1547–1550 (2004).
108. Sanjuán, R., Moya, A. & Elena, S. F. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci.* **101**, 15376–15379 (2004).
109. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
110. Simmonds, P. & Smith, D. B. Structural constraints on RNA virus evolution. *J. Virol.* **73**, 5787–5794 (1999).
111. Sanjuán, R. & Nebot, M. R. A Network Model for the Correlation between Epistasis and Genomic Complexity. *PLOS ONE* **3**, e2663 (2008).
112. Sanjuán, R. & Elena, S. F. Epistasis correlates to genomic complexity. *Proc. Natl. Acad. Sci.* **103**, 14402–14405 (2006).
113. Chen, R. *et al.* Lineage Divergence and Vector-Specific Adaptation Have Driven Chikungunya Virus onto Multiple Adaptive Landscapes. *mBio* **12**, e0273821 (2021).
114. Tsetsarkin, K. A. *et al.* Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7872–7877 (2011).
115. Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J. & Tans, S. J. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144 (2011).
116. Arribas, M., Kubota, K., Cabanillas, L. & Lázaro, E. Adaptation to Fluctuating Temperatures in an RNA Virus Is Driven by the Most Stringent Selective Pressure. *PLOS ONE* **9**, e100940 (2014).
117. Alto, B. W., Wasik, B. R., Morales, N. M. & Turner, P. E. Stochastic Temperatures Impede Rna Virus Adaptation. *Evolution* **67**, 969–979 (2013).
118. Bergh, B. V. den, Swings, T., Fauvart, M. & Michiels, J. Experimental Design, Population Dynamics, and Diversity in Microbial Experimental Evolution. *Microbiol. Mol. Biol. Rev.* **82**, (2018).
119. Lauring, A. S., Frydman, J. & Andino, R. The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.* **11**, 327–336 (2013).

120. Burch, C. L. & Chao, L. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**, 625–628 (2000).
121. Lauring, A. S., Acevedo, A., Cooper, S. B. & Andino, R. Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus. *Cell Host Microbe* **12**, 623–632 (2012).
122. Coleman, J. R. *et al.* Virus Attenuation by Genome-Scale Changes in Codon Pair Bias. *Science* **320**, 1784–1787 (2008).
123. Carrau, L. *et al.* Chikungunya Virus Vaccine Candidates with Decreased Mutational Robustness Are Attenuated In Vivo and Have Compromised Transmissibility. *J. Virol.* **93**, e00775-19.
124. Jenkins, G. M., Pagel, M., Gould, E. A., Zanutto, P. M. de A. & Holmes, E. C. Evolution of Base Composition and Codon Usage Bias in the Genus Flavivirus. *J. Mol. Evol.* **52**, 383–390 (2001).
125. Moratorio, G. *et al.* Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol.* **2**, 17088 (2017).
126. Codoñer, F. M., Darós, J.-A., Solé, R. V. & Elena, S. F. The Fittest versus the Flattest: Experimental Confirmation of the Quasispecies Effect with Subviral Pathogens. *PLOS Pathog.* **2**, e136 (2006).
127. Sanjuán, R., Cuevas, J. M., Furió, V., Holmes, E. C. & Moya, A. Selection for Robustness in Mutagenized RNA Viruses. *PLOS Genet.* **3**, e93 (2007).
128. Elena, S. F., Carrasco, P., Daròs, J. A. & Sanjuán, R. Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* **7**, 168–173 (2006).
129. Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1405–9 (2002).
130. Montville, R., Froissart, R., Remold, S. K., Tenaillon, O. & Turner, P. E. Evolution of Mutational Robustness in an RNA Virus. *PLOS Biol.* **3**, e381 (2005).
131. Mayer, M. P. Recruitment of Hsp70 chaperones: a crucial part of viral survival strategies. in *Reviews of Physiology, Biochemistry and Pharmacology* 1–46 (Springer, 2005). doi:10.1007/s10254-004-0025-5.
132. Alfano, C. & McMacken, R. Heat Shock Protein-mediated Disassembly of Nucleoprotein Structures Is Required for the Initiation of Bacteriophage λ DNA Replication. *J. Biol. Chem.* **264**, 10709–10718 (1989).
133. Guerrero, C. A. *et al.* Heat Shock Cognate Protein 70 Is Involved in Rotavirus Cell Entry. *J. Virol.* **76**, 4096–4102 (2002).
134. Kampmueller, K. M. & Miller, D. J. The Cellular Chaperone Heat Shock Protein 90 Facilitates Flock House Virus RNA Replication in *Drosophila* Cells. *J. Virol.* **79**, 6827–6837 (2005).

135. Jockusch, H., Wiegand, C., Mersch, B. & Rajes, D. Mutants of Tobacco mosaic virus with Temperature-Sensitive Coat Proteins Induce Heat Shock Response in Tobacco Leaves. *Mol. Plant-Microbe Interactions*® **14**, 914–917 (2001).
136. Geller, R., Pechmann, S., Acevedo, A., Andino, R. & Frydman, J. Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat. Commun.* **9**, 1781 (2018).
137. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
138. Wagner, A. Robustness and evolvability: a paradox resolved. *Proc. R. Soc. B Biol. Sci.* **275**, 91–100 (2008).
139. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
140. Eigen, M. & Schuster, P. A principle of natural self-organization. *Naturwissenschaften* **64**, 541–565 (1977).
141. Eigen, M. On the nature of virus quasispecies. *Trends Microbiol.* **4**, 216–8 (1996).
142. Holland, J. J., Domingo, E., de la Torre, J. C. & Steinhauer, D. A. Mutation Frequencies at Defined Single Codon Sites in Vesicular Stomatitis Virus and Poliovirus Can Be Increased Only Slightly by Chemical Mutagenesis. *J. Virol.* **64**, 3960–3962 (1990).
143. Domingo, E., García-Crespo, C. & Perales, C. Historical Perspective on the Discovery of the Quasispecies Concept. *Annu. Rev. Virol.* **8**, (2021).
144. Wilke, C. O. Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* **5**, 44 (2005).
145. Bull, J. J., Meyers, L. A. & Lachmann, M. Quasispecies Made Simple. *PLOS Comput. Biol.* **1**, e61 (2005).
146. Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci.* **98**, 6895–6900 (2001).
147. Pariente, N., Sierra, S., Lowenstein, P. R. & Domingo, E. Efficient Virus Extinction by Combinations of a Mutagen and Antiviral Inhibitors. *J. Virol.* **75**, 9723–9730 (2001).
148. Grande-Pérez, A., Lázaro, E., Lowenstein, P., Domingo, E. & Manrubia, S. C. Suppression of viral infectivity through lethal defection. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4448–4452 (2005).
149. Vignuzzi, M., Stone, J. K. & Andino, R. Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *Virus Res.* **107**, 173–181 (2005).
150. Arias, A., Ruiz-Jarabo, C. M., Escarmís, C. & Domingo, E. Fitness Increase of Memory Genomes in a Viral Quasispecies. *J. Mol. Biol.* **339**, 405–412 (2004).
151. Comas, I., Moya, A. & González-Candelas, F. Validating viral quasispecies with digital organisms: a re-examination of the critical mutation rate. *BMC Evol. Biol.* **5**, 1–10 (2005).

152. Pfeiffer, J. K. & Kirkegaard, K. Ribavirin resistance in hepatitis c virus replicon-containing cell lines conferred by changes in the cell line or mutations in the replicon RNA. *J. Virol.* **79**, 2346–2355 (2005).
153. Scheidel, L. M., Durbin, R. K. & Stollar, V. Sindbis virus mutants resistant to mycophenolic acid and ribavirin. *Virology* **158**, 1–7 (1987).
154. Briones, C., de Vicente, A., Molina-Paris, C. & Domingo, E. Minority memory genomes can influence the evolution of HIV-1 quasispecies in vivo. *Gene* **384**, 129–138 (2006).
155. Novella, I. S., Reissig, D. D. & Wilke, C. O. Density-dependent selection in vesicular stomatitis virus. *J. Virol.* **78**, 5799–5804 (2004).
156. Segredo-Otero, E. & Sanjuán, R. The effect of genetic complementation on the fitness and diversity of viruses spreading as collective infectious units. *Virus Res.* **267**, 41–48 (2019).
157. Xu, H. *et al.* Coinfection in the host can result in functional complementation between live vaccines and virulent virus. *Virulence* **13**, 980–989 (2022).
158. Xue, K. S., Hooper, K. A., Ollodart, A. R., Dingens, A. S. & Bloom, J. D. Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. *eLife* **5**, e13974 (2016).
159. Munis, A. M., Bentley, E. M. & Takeuchi, Y. A tool with many applications: vesicular stomatitis virus in research and medicine. *Expert Opin. Biol. Ther.* **20**, 1187–1201 (2020).
160. Brooke, C. B. *et al.* Most Influenza A Virions Fail To Express at Least One Essential Viral Protein. *J. Virol.* **87**, 3155–3162 (2013).
161. Ciota, A. T., Ehrbar, D. J., Van Slyke, G. A., Willsey, G. G. & Kramer, L. D. Cooperative interactions in the West Nile virus mutant swarm. *BMC Evol. Biol.* **12**, 58 (2012).
162. Bordería, A. V. *et al.* Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLoS Pathog.* **11**, 1–20 (2015).
163. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348 (2006).
164. Tanner, E. J. *et al.* Dominant drug targets suppress the emergence of antiviral resistance. *eLife* **3**, e03830 (2014).
165. Shirogane, Y., Watanabe, S. & Yanagi, Y. Cooperation between different RNA virus genomes produces a new phenotype. *Nat. Commun.* **3**, 1235 (2012).
166. Vignuzzi, M. & López, C. B. Defective viral genomes are key drivers of the virus–host interaction. *Nat. Microbiol.* **4**, 1075–1087 (2019).
167. Meir, M. *et al.* Competition between social cheater viruses is driven by mechanistically different cheating strategies. *Sci. Adv.* **6**, eabb7990 (2020).
168. Froissart, R. *et al.* Co-infection weakens selection against epistatic mutations in RNA viruses. *Genetics* **168**, 9–19 (2004).

169. Aguilera, E. R., Erickson, A. K., Jesudhasan, P. R., Robinson, C. M. & Pfeiffer, J. K. Plaques Formed by Mutagenized Viral Populations Have Elevated Coinfection Frequencies. *mBio* **8**, (2017).
170. Aguilera, E. R. & Pfeiffer, J. K. Strength in numbers: Mechanisms of viral co-infection. *Virus Res.* **265**, 43–46 (2019).
171. Luque, D. *et al.* Infectious bursal disease virus is an icosahedral polyplloid dsRNA virus. *Proc. Natl. Acad. Sci.* **106**, 2148–2152 (2009).
172. Rager, M., Vongpunsawad, S., Duprex, W. P. & Cattaneo, R. Polyplloid measles virus with hexameric genome length. *EMBO J.* **21**, 2364–2372 (2002).
173. Beniac, D. R. *et al.* The Organisation of Ebola Virus Reveals a Capacity for Extensive, Modular Polyplloidy. *PLOS ONE* **7**, e29608 (2012).
174. Schreur, P. J. W. & Kortekaas, J. Single-Molecule FISH Reveals Non-selective Packaging of Rift Valley Fever Virus Genome Segments. *PLOS Pathog.* **12**, e1005800 (2016).
175. Floyd, R. & Sharp, D. G. Aggregation of poliovirus and reovirus by dilution in water. *Appl. Environ. Microbiol.* **33**, 159–167 (1977).
176. Cuevas, J. M., Durán-Moreno, M. & Sanjuán, R. Multi-virion infectious units arise from free viral particles in an enveloped virus. *Nat. Microbiol.* **2**, (2017).
177. Usmani, S. M. *et al.* Direct visualization of HIV-enhancing endogenous amyloid fibrils in human semen. *Nat. Commun.* **5**, 3508 (2014).
178. McDonald, D. Dendritic Cells and HIV-1 Trans-Infection. *Viruses* **2**, 1704–1717 (2010).
179. Erickson, A. K. *et al.* Bacteria facilitate enteric virus co-infection of mammalian cells and promote genetic recombination. *Cell Host Microbe* **23**, 77-88.e5 (2018).
180. Pais-Correia, A.-M. *et al.* Biofilm-like extracellular viral assemblies mediate HTLV-1 cell-to-cell transmission at virological synapses. *Nat. Med.* **16**, 83–89 (2010).
181. Nagashima, S. *et al.* Hepatitis E virus egress depends on the exosomal pathway, with secretory exosomes derived from multivesicular bodies. *J. Gen. Virol.* **95**, 2166–2175 (2014).
182. Ramakrishnaiah, V. *et al.* Exosome-mediated transmission of hepatitis C virus between human hepatoma Huh7.5 cells. *Proc. Natl. Acad. Sci.* **110**, 13109–13113 (2013).
183. Santiana, M. *et al.* Vesicle-Cloaked Virus Clusters Are Optimal Units for Inter-organismal Viral Transmission. *Cell Host Microbe* **24**, 208-220.e8 (2018).
184. Bird, S. W., Maynard, N. D., Covert, M. W. & Kirkegaard, K. Nonlytic viral spread enhanced by autophagy components. *Proc. Natl. Acad. Sci.* **111**, 13081–13086 (2014).
185. Robinson, S. M. *et al.* Coxsackievirus B Exits the Host Cell in Shed Microvesicles Displaying Autophagosomal Markers. *PLOS Pathog.* **10**, e1004045 (2014).
186. Chen, Y.-H. *et al.* Phosphatidylserine Vesicles Enable Efficient En Bloc Transmission of Enteroviruses. *Cell* **160**, 619–630 (2015).

187. Kumar, D., Kumar, R., Hyun, T. K. & Kim, J.-Y. Cell-to-cell movement of viruses via plasmodesmata. *J. Plant Res.* **128**, 37–47 (2015).
188. Lawrence, D. M. P. *et al.* Measles virus spread between neurons requires cell contact but not CD46 expression, syncytium formation, or extracellular virus production. *J. Virol.* **74**, 1908–1918 (2000).
189. Sherer, N. M. *et al.* Retroviruses can establish filopodial bridges for efficient cell-to-cell transmission. *Nat. Cell Biol.* **9**, 310–315 (2007).
190. Leroy, H. *et al.* Virus-Mediated Cell-Cell Fusion. *Int. J. Mol. Sci.* **21**, 9644 (2020).
191. Symeonides, M. *et al.* HIV-1-Induced Small T Cell Syncytia Can Transfer Virus Particles to Target Cells through Transient Contacts. *Viruses* **7**, 6590–6603 (2015).
192. Robinson, C. M., Jesudhasan, P. R. & Pfeiffer, J. K. Bacterial Lipopolysaccharide Binding Enhances Virion Stability and Promotes Environmental Fitness of an Enteric Virus. *Cell Host Microbe* **15**, 36–46 (2014).
193. Laliberte, J. P. & Moss, B. A Novel Mode of Poxvirus Superinfection Exclusion That Prevents Fusion of the Lipid Bilayers of Viral and Cellular Membranes. *J. Virol.* **88**, 9751–9768 (2014).
194. Boussier, J. *et al.* Chikungunya virus superinfection exclusion is mediated by a block in viral replication and does not rely on non-structural protein 2. *PLOS ONE* **15**, e0241592 (2020).
195. Zou, G. *et al.* Exclusion of West Nile Virus Superinfection through RNA Replication. *J. Virol.* **83**, 11765–11776 (2009).
196. Gutiérrez, S. *et al.* The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *J. Virol.* **89**, 9665–9675 (2015).
197. Sims, A. *et al.* Superinfection exclusion creates spatially distinct influenza virus populations. 2022.06.06.494939 Preprint at <https://doi.org/10.1101/2022.06.06.494939> (2022).
198. Andreu-Moreno, I., Bou, J.-V. & Sanjuán, R. Cooperative nature of viral replication. *Sci. Adv.* **6**, eabd4942 (2020).
199. Stiefel, P. *et al.* Cooperative Vaccinia Infection Demonstrated at the Single-Cell Level Using FluidFM. *Nano Lett.* **12**, 4219–4227 (2012).
200. Landsberger, M. *et al.* Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell* **174**, 908–916.e12 (2018).
201. Borges, A. L. *et al.* Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. *Cell* **174**, 917–925.e10 (2018).
202. Domingo-Calap, P., Mora-Quilis, L. & Sanjuán, R. Social Bacteriophages. *Microorganisms* **8**, 533 (2020).
203. Domingo-Calap, P., Segredo-Otero, E., Durán-Moreno, M. & Sanjuán, R. Social evolution of innate immunity evasion in a virus. *Nat. Microbiol.* **4**, 1006–1013 (2019).

204. Erez, Z. *et al.* Communication between viruses guides lysis–lysogeny decisions. *Nature* **541**, 488–493 (2017).
205. Giri, S., Waschina, S., Kaleta, C. & Kost, C. Defining Division of Labor in Microbial Communities. *J. Mol. Biol.* **431**, 4712–4731 (2019).
206. Chao, L. & Elena, S. F. Nonlinear trade-offs allow the cooperation game to evolve from Prisoner’s Dilemma to Snowdrift. *Proc. R. Soc. B Biol. Sci.* **284**, 20170228 (2017).
207. Doceul, V., Hollinshead, M., van der Linden, L. & Smith, G. L. Repulsion of Superinfecting Virions: A Mechanism for Rapid Virus Spread. *Science* **327**, 873–876 (2010).
208. Aparicio, L. J. G., López, C. B. & Felt, S. A. A Virus Is a Community: Diversity within Negative-Sense RNA Virus Populations. *Microbiol. Mol. Biol. Rev.* **0**, e00086-21.
209. Tapia, K. *et al.* Defective viral genomes arising in vivo provide critical danger signals for the triggering of lung antiviral immunity. *PLoS Pathog.* **9**, e1003703 (2013).
210. Manzoni, T. B. & López, C. B. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. *Future Virol.* **13**, 493–503 (2018).
211. Brinton, M. A. & Fernandez, A. V. A replication-efficient mutant of West Nile virus is insensitive to DI particle interference. *Virology* **129**, 107–115 (1983).
212. DePolo, N. J., Giachetti, C. & Holland, J. J. Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles. *J. Virol.* **61**, 454–464 (1987).
213. Karamichali, E. *et al.* HCV defective genomes promote persistent infection by modulating the viral life cycle. *Front. Microbiol.* **9**, (2018).
214. Levi, L. I. *et al.* Defective viral genomes from chikungunya virus are broad-spectrum antivirals and prevent virus dissemination in mosquitoes. *PLOS Pathog.* **17**, e1009110 (2021).
215. Jaworski, E. & Routh, A. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLOS Pathog.* **13**, e1006365 (2017).
216. Cattaneo, R. *et al.* Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55**, 255–265 (1988).
217. Wang, S. *et al.* Covariation of viral recombination with single nucleotide variants during virus evolution revealed by CoVaMa. *Nucleic Acids Res.* **50**, e41 (2022).
218. De, B. K. & Nayak, D. P. Defective interfering influenza viruses and host cells: establishment and maintenance of persistent influenza virus infection in MDBK and HeLa cells. *J. Virol.* **36**, 847–859 (1980).
219. Von Magnus, P. Studies on Interference in Experimental Influenza. I. Biological Observations. *Ark. Kemi Mineral. Och Geol.* (1947).

220. Huang, A. S. & Baltimore, D. Defective Viral Particles and Viral Disease Processes. *Nature* **226**, 325–327 (1970).
221. Alnaji, F. G. & Brooke, C. B. Influenza virus DI particles: Defective interfering or delightfully interesting? *PLOS Pathog.* **16**, e1008436 (2020).
222. Lazzarini, R. A., Keene, J. D. & Schubert, M. The origins of defective interfering particles of the negative-strand RNA viruses. *Cell* **26**, 145–154 (1981).
223. Fodor, E., Mingay, L. J., Crow, M., Deng, T. & Brownlee, G. G. A Single Amino Acid Mutation in the PA Subunit of the Influenza Virus RNA Polymerase Promotes the Generation of Defective Interfering RNAs. *J. Virol.* **77**, 5017–5020 (2003).
224. Panavienė, Ž. & Nagy, P. D. Mutations in the RNA-binding domains of tombusvirus replicase proteins affect RNA recombination in vivo. *Virology* **317**, 359–372 (2003).
225. Odagiri, T. & Tobita, K. Mutation in NS2, a nonstructural protein of influenza A virus, extragenetically causes aberrant replication and expression of the PA gene and leads to generation of defective interfering particles. *Proc. Natl. Acad. Sci.* **87**, 5988–5992 (1990).
226. Pfaller, C. K. *et al.* Measles Virus Defective Interfering RNAs Are Generated Frequently and Early in the Absence of C Protein and Can Be Destabilized by Adenosine Deaminase Acting on RNA-1-Like Hypermutations. *J. Virol.* **89**, 7735–7747 (2015).
227. Sánchez-Aparicio, M. T. *et al.* Loss of Sendai virus C protein leads to accumulation of RIG-I immunostimulatory defective interfering RNA. *J. Gen. Virol.* **98**, 1282–1293.
228. Ziegler, C. M. *et al.* The Lymphocytic Choriomeningitis Virus Matrix Protein PPTY Late Domain Drives the Production of Defective Interfering Particles. *PLOS Pathog.* **12**, e1005501 (2016).
229. Pelz, L. *et al.* Semi-continuous Propagation of Influenza A Virus and Its Defective Interfering Particles: Analyzing the Dynamic Competition To Select Candidates for Antiviral Therapy. *J. Virol.* **95**, e01174-21 (2021).
230. Kawai, A., Matsumoto, S. & Tanabe, K. Characterization of rabies viruses recovered from persistently infected BHK cells. *Virology* **67**, 520–533 (1975).
231. Frensing, T. *et al.* Continuous Influenza Virus Production in Cell Culture Shows a Periodic Accumulation of Defective Interfering Particles. *PLOS ONE* **8**, e72288 (2013).
232. Oster, S. K., Wu, B. & White, K. A. Uncoupled Expression of p33 and p92 Permits Amplification of Tomato Bushy Stunt Virus RNAs. *J. Virol.* **72**, 5845–5851 (1998).
233. Li, T. & Pattnaik, A. K. Replication Signals in the Genome of Vesicular Stomatitis Virus and Its Defective Interfering Particles: Identification of a Sequence Element That Enhances DI RNA Replication. *Virology* **232**, 248–259 (1997).
234. Calain, P. & Roux, L. Functional Characterisation of the Genomic and Antigenomic Promoters of Sendai Virus. *Virology* **212**, 163–173 (1995).
235. Alnaji, F. G., Reiser, W. K., Rivera-Cardona, J., te Velhuis, A. J. W. & Brooke, C. B. Influenza A Virus Defective Viral Genomes Are Inefficiently Packaged into Virions Relative to Wild-Type Genomic RNAs. *mBio* **12**, e02959-21 (2021).

236. Easton, A. J. *et al.* A novel broad-spectrum treatment for respiratory virus infections: Influenza-based defective interfering virus provides protection against pneumovirus infection in vivo. *Vaccine* **29**, 2777–2784 (2011).
237. Fonville, J. M., Marshall, N., Tao, H., Steel, J. & Lowen, A. C. Influenza virus reassortment is enhanced by semi-infectious particles but can be suppressed by defective interfering particles. *PLoS Pathog.* **11**, e1005204 (2015).
238. Genoyer, E. & López, C. B. Defective Viral Genomes Alter How Sendai Virus Interacts with Cellular Trafficking Machinery, Leading to Heterogeneity in the Production of Viral Particles among Infected Cells. *J. Virol.* **93**, e01579-18 (2019).
239. Tuffereau, C. & Roux, L. Direct adverse effects of sendai virus DI particles on virus budding and on M protein fate and stability. *Virology* **162**, 417–426 (1988).
240. Mateo, R., Nagamine, C. M. & Kirkegaard, K. Suppression of Drug Resistance in Dengue Virus. *mBio* **6**, e01960-15.
241. Yount, J. S., Gitlin, L., Moran, T. M. & López, C. B. MDA5 participates in the detection of paramyxovirus infection and is essential for the early activation of dendritic cells in response to sendai virus defective interfering particles. *J. Immunol.* **180**, 4910–4918 (2008).
242. Yount, J. S., Kraus, T. A., Horvath, C. M., Moran, T. M. & López, C. B. A Novel Role for Viral-Defective Interfering Particles in Enhancing Dendritic Cell Maturation. *J. Immunol.* **177**, 4503–4513 (2006).
243. Sun, Y. *et al.* Immunostimulatory defective viral genomes from respiratory syncytial virus promote a strong innate antiviral response during infection in mice and humans. *PLoS Pathog.* **11**, e1005122 (2015).
244. van den Hoogen, B. G. *et al.* Excessive production and extreme editing of human metapneumovirus defective interfering RNA is associated with type I IFN induction. *J. Gen. Virol.* **95**, 1625–1633.
245. Strahle, L., Garcin, D. & Kolakofsky, D. Sendai virus defective-interfering genomes and the activation of interferon-beta. *Virology* **351**, 101–111 (2006).
246. Xu, J. *et al.* Identification of a Natural Viral RNA Motif That Optimizes Sensing of Viral RNA by RIG-I. *mBio* **6**, e01265-15 (2015).
247. Mercado-López, X. *et al.* Highly immunostimulatory RNA derived from a Sendai virus defective viral genome. *Vaccine* **31**, 5713–5721 (2013).
248. Simon, A. E., Roossinck, M. J. & Havelda, Z. Plant virus satellite and defective interfering RNAs: New paradigms for a new century. *Annu. Rev. Phytopathol.* **42**, 415–437 (2004).
249. Poirier, E. Z. *et al.* Dicer-2-Dependent Generation of Viral DNA from Defective Genomes of RNA Viruses Modulates Antiviral Immunity in Insects. *Cell Host Microbe* **23**, 353-365.e8 (2018).
250. Xiao, Y. *et al.* A defective viral genome strategy elicits broad protective immunity against respiratory viruses. *Cell* **184**, 6037-6051.e14 (2021).

251. Dimmock, N. J., Rainsford, E. W., Scott, P. D. & Marriott, A. C. Influenza Virus Protecting RNA: an Effective Prophylactic and Therapeutic Antiviral. *J. Virol.* **82**, 8570–8578 (2008).
252. McLaren, L. C. & Holland, J. J. Defective interfering particles from poliovirus vaccine and vaccine reference strains. *Virology* **60**, 579–583 (1974).
253. Shingai, M. *et al.* Differential Type I IFN-Inducing Abilities of Wild-Type versus Vaccine Strains of Measles Virus. *J. Immunol.* **179**, 6123–6133 (2007).
254. Gould, P. S., Easton, A. J. & Dimmock, N. J. Live Attenuated Influenza Vaccine contains Substantial and Unexpected Amounts of Defective Viral Genomic RNA. *Viruses* **9**, 269 (2017).
255. Rezelj, V. V., Levi, L. I. & Vignuzzi, M. The defective component of viral populations. *Curr. Opin. Virol.* **33**, 74–80 (2018).
256. Schmaljohn, C. & Blair, C. D. Persistent infection of cultured mammalian cells by Japanese encephalitis virus. *J. Virol.* **24**, 580–589 (1977).
257. Calain, P., Monroe, M. C. & Nichol, S. T. Ebola Virus Defective Interfering Particles and Persistent Infection. *Virology* **262**, 114–128 (1999).
258. Sekellick, M. J. & Marcus, P. I. Persistent infection I. Interferon-inducing defective-interfering particles as mediators of cell sparing: Possible role in persistent infection by vesicular stomatitis virus. *Virology* **85**, 175–186 (1978).
259. Xu, J. *et al.* Replication defective viral genomes exploit a cellular pro-survival mechanism to establish paramyxovirus persistence. *Nat. Commun.* **8**, 799 (2017).
260. Barrett, A. D. T. *et al.* Subclinical Infections in Mice Resulting from the Modulation of a Lethal Dose of Semliki Forest Virus with Defective Interfering Viruses: Neurochemical Abnormalities in the Central Nervous System. *J. Gen. Virol.* **67**, 1727–1732.
261. Ke, R., Aaskov, J., Holmes, E. C. & Lloyd-Smith, J. O. Phylodynamic Analysis of the Emergence and Epidemiological Impact of Transmissible Defective Dengue Viruses. *PLOS Pathog.* **9**, e1003193 (2013).
262. O'Hara, P. J., Nichol, S. T., Horodyski, F. M. & Holland, J. J. Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell* **36**, 915–924 (1984).
263. Pierson, T. C. & Kielian, M. Flaviviruses: braking the entering. *Curr. Opin. Virol.* **3**, 3–12 (2013).
264. Sirohi, D. & Kuhn, R. J. Zika Virus Structure, Maturation, and Receptors. *J. Infect. Dis.* **216**, S935–S944 (2017).
265. Slonchak, A. & Khromykh, A. A. Subgenomic flaviviral RNAs: What do we know after the first decade of research. *Antiviral Res.* **159**, 13–25 (2018).
266. Lindenbach, B. D., Randall, G., Bartenschlager, R. & Rice, C. M. Flaviviridae: The Viruses and Their Replication. in *Fields Virology: Emerging Viruses - Volume 1* (eds. Howley, P. M. & Knipe, D. M.) 246–301 (Lippincott Williams & Wilkins, 2021).

267. Pierson, T. C., Lazear, H. M. & Diamond, M. S. Flaviviruses: Dengue, Zika, West Nile, Yellow Fever and Other Flaviviruses. in *Fields Virology: Emerging Viruses - Volume 1* (eds. Howley, P. M. & Knipe, D. M.) 345–409 (Lippincott Williams & Wilkins, 2021).
268. Liu, Y. *et al.* Evolutionary enhancement of Zika virus infectivity in *Aedes aegypti* mosquitoes. *Nature* **545**, 482–486 (2017).
269. Schwarz, M. C. *et al.* Rescue of the 1947 zika virus prototype strain with a cytomegalovirus promoter-driven cDNA clone. *mSphere* **1**, (2016).
270. Xie, X. *et al.* Zika Virus Replicons for Drug Discovery. *EBioMedicine* **12**, 156–160 (2016).
271. Pronobis, M. I., Deutch, N. & Peifer, M. The Miraprep: A Protocol that Uses a Miniprep Kit and Provides Maxiprep Yields. *PLOS ONE* **11**, e0160509 (2016).
272. García-Nafría, J., Watson, J. F. & Greger, I. H. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci. Rep.* **6**, 27459 (2016).
273. Routh, A. & Johnson, J. E. Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. *Nucleic Acids Res.* **42**, e11 (2014).
274. Isakov, O. *et al.* Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. *Bioinformatics* **31**, 2141–2150 (2015).
275. Acevedo, A. & Andino, R. Library preparation for highly accurate population sequencing of RNA viruses. *Nat. Protoc.* **9**, 1760–1769 (2014).
276. Talarico, L. B. & Damonte, E. B. Characterization of in vitro dengue virus resistance to carrageenan. *J. Med. Virol.* **88**, 1120–1129 (2016).
277. Bassi, M. R., Sempere, R. N., Meyn, P., Polacek, C. & Arias, A. Extinction of Zika Virus and Usutu Virus by Lethal Mutagenesis Reveals Different Patterns of Sensitivity to Three Mutagenic Drugs. *Antimicrob. Agents Chemother.* **62**, e00380-18 (2018).
278. Olsen, D. B. *et al.* A 7-Deaza-Adenosine Analog Is a Potent and Selective Inhibitor of Hepatitis C Virus Replication with Excellent Pharmacokinetic Properties. *Antimicrob. Agents Chemother.* **48**, 3944–3953 (2004).
279. Schul, W., Liu, W., Xu, H.-Y., Flamand, M. & Vasudevan, S. G. A Dengue Fever Viremia Model in Mice Shows Reduction in Viral Replication and Suppression of the Inflammatory Response after Treatment with Antiviral Drugs. *J. Infect. Dis.* **195**, 665–674 (2007).
280. Yoon, S. W. *et al.* Characterization of homologous defective interfering RNA during persistent infection of Vero cells with Japanese encephalitis virus. *Mol. Cells* **21**, 112–120 (2006).
281. Lancaster, M. U., Hodgetts, S. I., Mackenzie, J. S. & Urosevic, N. Characterization of Defective Viral RNA Produced during Persistent Infection of Vero Cells with Murray Valley Encephalitis Virus. *J. Virol.* **72**, 2474–2482 (1998).
282. Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K. & Holmes, E. C. Long-term transmission of defective RNA viruses in humans and aedes mosquitoes. *Science* **311**, 236–238 (2006).

283. Noppornpanth, S. *et al.* Characterization of hepatitis c virus deletion mutants circulating in chronically infected patients. *J. Virol.* **81**, 12496–12503 (2007).
284. Yagi, S. *et al.* Identification of novel HCV subgenome replicating persistently in chronic active hepatitis C patients. *J. Med. Virol.* **77**, 399–413 (2005).
285. Gould, S. J., Lewontin, R. C., Maynard Smith, J. & Holliday, R. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B Biol. Sci.* **205**, 581–598 (1979).
286. Weaver, S. C., Rico-Hesse, R. & Scott, T. W. Genetic Diversity and Slow Rates of Evolution in New World Alphaviruses. in *Genetic Diversity of RNA Viruses* (ed. Holland, J. J.) 99–117 (Springer, 1992). doi:10.1007/978-3-642-77011-1_7.
287. Novella, I. S., Presloid, J. B., Smith, S. D. & Wilke, C. O. Specific and nonspecific host adaptation during arboviral experimental evolution. *J. Mol. Microbiol. Biotechnol.* **21**, 71–81 (2011).
288. Novella, I. S., Ebendick-Corpus, B. E., Zárate, S. & Miller, E. L. Emergence of Mammalian Cell-Adapted Vesicular Stomatitis Virus from Persistent Infections of Insect Vector Cells. *J. Virol.* **81**, 6664–6668 (2007).
289. Arias-Goeta, C. *et al.* Chikungunya virus adaptation to a mosquito vector correlates with only few point mutations in the viral envelope glycoprotein. *Infect. Genet. Evol.* **24**, 116–126 (2014).
290. Vasilakis, N. *et al.* Mosquitoes Put the Brake on Arbovirus Evolution: Experimental Evolution Reveals Slower Mutation Accumulation in Mosquito Than Vertebrate Cells. *PLoS Pathog.* **5**, e1000467 (2009).
291. Cooper, L. A. & Scott, T. W. Differential Evolution of Eastern Equine Encephalitis Virus Populations in Response to Host Cell Type. *Genetics* **157**, 1403–1412 (2001).
292. Ciota, A. T. *et al.* Role of the mutant spectrum in adaptation and replication of West Nile virus. *J. Gen. Virol.* **88**, 865–874 (2007).
293. Chen, W.-J., Wu, H.-R. & Chiou, S.-S. E/NS1 Modifications of Dengue 2 Virus after Serial Passages in Mammalian and/or Mosquito Cells. *Intervirology* **46**, 289–295 (2003).
294. Coffey, L. L. & Vignuzzi, M. Host Alternation of Chikungunya Virus Increases Fitness while Restricting Population Diversity and Adaptability to Novel Selective Pressures. *J. Virol.* **85**, 1025–1035 (2011).
295. Coffey, L. L. *et al.* Arbovirus evolution in vivo is constrained by host alternation. *Proc. Natl. Acad. Sci.* **105**, 6970–6975 (2008).
296. Weaver, S. C., Brault, A. C., Kang, W. & Holland, J. J. Genetic and Fitness Changes Accompanying Adaptation of an Arbovirus to Vertebrate and Invertebrate Cells. *J. Virol.* **73**, 4316–4326 (1999).
297. Greene, I. P. *et al.* Effect of Alternating Passage on Adaptation of Sindbis Virus to Vertebrate and Invertebrate Cells. *J. Virol.* **79**, 14253–14260 (2005).

298. Novella, I. S., Hershey, C. L., Escarmis, C., Domingo, E. & Holland, J. J. Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. Edited by J. Karn. *J. Mol. Biol.* **287**, 459–465 (1999).
299. Turner, P. E. & Elena, S. F. Cost of Host Radiation in an RNA Virus. *Genetics* **156**, 1465–1470 (2000).
300. Turner, P. E., Morales, N. M., Alto, B. W. & Remold, S. K. Role of evolved host breadth in the initial emergence of an RNA virus. *Evol. Int. J. Org. Evol.* **64**, 3273–86 (2010).
301. Smith-Tsurkan, S. D., Wilke, C. O. & Novella, I. S. Incongruent fitness landscapes, not tradeoffs, dominate the adaptation of vesicular stomatitis virus to novel host types. *J. Gen. Virol.* **91**, 1484–1493 (2010).
302. Zarate, S. & Novella, I. S. Vesicular Stomatitis Virus Evolution during Alternation between Persistent Infection in Insect Cells and Acute Infection in Mammalian Cells Is Dominated by the Persistence Phase. *J. Virol.* **78**, 12236–12242 (2004).
303. Deardorff, E. R. *et al.* West Nile Virus Experimental Evolution in vivo and the Trade-off Hypothesis. *PLOS Pathog.* **7**, e1002335 (2011).
304. Carrasco, P., Daròs, J. A., Agudelo-Romero, P. & Elena, S. F. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *J. Virol. Methods* **139**, 181–188 (2007).
305. Jerzak, G. V. S., Brown, I., Shi, P.-Y., Kramer, L. D. & Ebel, G. D. Genetic diversity and purifying selection in West Nile virus populations are maintained during host switching. *Virology* **374**, 256–260 (2008).
306. Powers, A. M. *et al.* Evolutionary Relationships and Systematics of the Alphaviruses. *J. Virol.* **75**, 10118–10131 (2001).
307. Silva, L. A. & Dermody, T. S. Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. *J. Clin. Invest.* **127**, 737–749 (2017).
308. Zhang, R. *et al.* Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* **557**, 570–574 (2018).
309. Song, H. *et al.* Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein. *Cell* **177**, 1714–1724.e12 (2019).
310. Kirui, J. *et al.* The Phosphatidylserine Receptor TIM-1 Enhances Authentic Chikungunya Virus Cell Entry. *Cells* **10**, 1828 (2021).
311. Silva, L. A. *et al.* A Single-Amino-Acid Polymorphism in Chikungunya Virus E2 Glycoprotein Influences Glycosaminoglycan Utilization. *J. Virol.* **88**, 2385–2397 (2014).
312. Pietilä, M. K., Hellström, K. & Ahola, T. Alphavirus polymerase and RNA replication. *Virus Res.* **234**, 44–57 (2017).
313. Ross, R. W. The Newala epidemic: III. The virus: isolation, pathogenic properties and relationship to the epidemic. *Epidemiol. Infect.* **54**, 177–191 (1956).

314. Powers, A. M., Brault, A. C., Tesh, R. B. & Weaver, S. C. Y. 2000. Re-emergence of chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *J. Gen. Virol.* **81**, 471–479.
315. Schuffenecker, I. *et al.* Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. *PLOS Med.* **3**, e263 (2006).
316. Diallo, M., Thonnon, J., Traore-Lamizana, M. & Fontenille, D. Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *Am. J. Trop. Med. Hyg.* **60**, 281–286 (1999).
317. Valentine, M. J., Murdock, C. C. & Kelly, P. J. Sylvatic cycles of arboviruses in non-human primates. *Parasit. Vectors* **12**, 1–18 (2019).
318. Volk, S. M. *et al.* Genome-Scale Phylogenetic Analyses of Chikungunya Virus Reveal Independent Emergences of Recent Epidemics and Various Evolutionary Rates. *J. Virol.* **84**, 6497–6504 (2010).
319. Rezza, G. *et al.* Infection with chikungunya virus in Italy: an outbreak in a temperate region. *The Lancet* **370**, 1840–1846 (2007).
320. Delisle, E. *et al.* Chikungunya outbreak in Montpellier, France, September to October 2014. *Eurosurveillance* **20**, 21108 (2015).
321. Gratz, N. G. Critical review of the vector status of *Aedes albopictus*. *Med. Vet. Entomol.* **18**, 215–227 (2004).
322. Vazeille, M. *et al.* Two Chikungunya Isolates from the Outbreak of La Reunion (Indian Ocean) Exhibit Different Patterns of Infection in the Mosquito, *Aedes albopictus*. *PLOS ONE* **2**, e1168 (2007).
323. Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. *PLOS Pathog.* **3**, e201 (2007).
324. Tsetsarkin, K. A. & Weaver, S. C. Sequential Adaptive Mutations Enhance Efficient Vector Switching by Chikungunya Virus and Its Epidemic Emergence. *PLOS Pathog.* **7**, e1002412 (2011).
325. Tsetsarkin, K. A. *et al.* Multi-peaked adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in *Aedes albopictus* mosquitoes. *Nat. Commun.* **5**, 4084 (2014).
326. Burt, F. J. *et al.* Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. *Lancet Infect. Dis.* **17**, e107–e117 (2017).
327. Leparc-Goffart, I., Nougairede, A., Cassadou, S., Prat, C. & de Lamballerie, X. Chikungunya in the Americas. *The Lancet* **383**, 514 (2014).
328. Khongwichit, S., Chansaenroj, J., Chirathaworn, C. & Poovorawan, Y. Chikungunya virus infection: molecular biology, clinical characteristics, and epidemiology in Asian countries. *J. Biomed. Sci.* **28**, 1–17 (2021).

329. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
330. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
331. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
332. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
333. Jin, J. *et al.* Neutralizing Monoclonal Antibodies Block Chikungunya Virus Entry and Release by Targeting an Epitope Critical to Viral Pathogenesis. *Cell Rep.* **13**, 2553–2564 (2015).
334. Zou, M. *et al.* Genetic Characterization of Chikungunya Virus Among Febrile Dengue Fever–Like Patients in Xishuangbanna, Southwestern Part of China. *Front. Cell. Infect. Microbiol.* **12**, (2022).

ANNEXES

1. Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts

Veronica V. Rezelj, Lucía Carrau, Fernando Merwaiss, Laura I. Levi, Diana Erazo, **Quang Dinh Tran**, Annabelle Henrion-Lacritick, Valérie Gausson, Yasutsugu Suzuki, Djoshkun Shengjuler, Bjoern Meyer, Thomas Vallet, James Weger-Lucarelli, Veronika Bernhauerová, Avi Titievsky, Vadim Sharov, Stefano Pietropaoli, Marco A. Diaz-Salinas, Vincent Legros, Nathalie Pardigon, Giovanna Barba-Spaeth, Leonid Brodsky, Maria-Carla Saleh & Marco Vignuzzi

DOI: <https://doi.org/10.1038/s41467-021-22341-7>

Contributions: I have helped with ZIKV *in vitro* experiments of Figure 2 and 3. Moreover, I have contributed to initial bioinformatics analysis presented in Figure 1, S2 and S6.

Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts

Veronica V. Rezelj ¹, Lucía Carrau ¹, Fernando Merwaiss ², Laura I. Levi ^{1,3}, Diana Erazo ¹, Quang Dinh Tran ^{1,3}, Annabelle Henrion-Lacritick², Valérie Gausson², Yasutsugu Suzuki ², Djoshkun Shengjuler ¹, Bjoern Meyer ¹, Thomas Vallet ¹, James Weger-Lucarelli¹, Veronika Bernhauerová¹, Avi Titievsky⁴, Vadim Sharov ⁴, Stefano Pietropaoli⁵, Marco A. Diaz-Salinas ⁶, Vincent Legros⁵, Nathalie Pardigon⁶, Giovanna Barba-Spaeth ⁵, Leonid Brodsky⁴, Maria-Carla Saleh ^{2✉} & Marco Vignuzzi ^{1✉}

Arthropod-borne viruses pose a major threat to global public health. Thus, innovative strategies for their control and prevention are urgently needed. Here, we exploit the natural capacity of viruses to generate defective viral genomes (DVGs) to their detriment. While DVGs have been described for most viruses, identifying which, if any, can be used as therapeutic agents remains a challenge. We present a combined experimental evolution and computational approach to triage DVG sequence space and pinpoint the fittest deletions, using Zika virus as an arbovirus model. This approach identifies fit DVGs that optimally interfere with wild-type virus infection. We show that the most fit DVGs conserve the open reading frame to maintain the translation of the remaining non-structural proteins, a characteristic that is fundamental across the flavivirus genus. Finally, we demonstrate that the high fitness DVG is antiviral *in vivo* both in the mammalian host and the mosquito vector, reducing transmission in the latter by up to 90%. Our approach establishes the method to interrogate the DVG fitness landscape, and enables the systematic identification of DVGs that show promise as human therapeutics and vector control strategies to mitigate arbovirus transmission and disease.

¹Institut Pasteur, Viral Populations and Pathogenesis Unit, Centre National de la Recherche Scientifique UMR 3569, Paris, France. ²Institut Pasteur, Viruses and RNA Interference Unit, Centre National de la Recherche Scientifique UMR 3569, Paris, France. ³École doctorale BioSPC, Université Paris Diderot, Sorbonne Paris Cité, Paris, France. ⁴Tauber Bioinformatics Research Center, University of Haifa, Haifa, Israel. ⁵Institut Pasteur, Unité de Virologie Structurale, Centre National de la Recherche Scientifique UMR 3569, Paris, France. ⁶Institut Pasteur, Unité de Recherche et d'Expertise Environnement et Risques Infectieux, Groupe Arbovirus, Paris, Cedex 15, France. ✉email: carla.saleh@pasteur.fr; marco.vignuzzi@pasteur.fr

Viral infections give rise to degenerate forms of the viral genome, known as defective viral genomes (DVGs), as a by-product of virus replication. These DVGs manifest as genomes with complementary ends, deleterious point mutations, deletions, insertions, mosaic rearrangements, or a combination of these. As a result, DVGs cannot complete a full replication cycle. Following their discovery after serial passaging of influenza virus at high multiplicity of infection (MOI) in 1947¹, DVGs have been described for nearly every virus family.

Although DVGs cannot self-replicate, complementation by wild-type (WT) virus enables their propagation within a virus population, by providing functions that they have lost. Certain DVGs interfere with virus replication by competing for viral and/or host resources or by enhancing immunostimulation, among other means. These are known as defective interfering particles (DIPs). While DIPs are known to be effective against some viruses in cell culture, they are largely considered a nuisance and artifact of cell culture passaging conditions. However, a rising number of recent reports describe DVGs in clinical and natural isolates: in patients infected with influenza A virus², respiratory syncytial virus³, hepatitis C virus^{4,5}, and dengue virus⁶; in birds infected with West Nile virus⁷; and even in plants^{8,9}, suggesting that DVGs are not strictly an *in vitro* phenomenon. A major obstacle in exploring their potential for application as therapeutic agents for viral diseases is the difficulty in isolating DVGs with antiviral capacity from the much larger spectrum of defective genomes generated during WT virus replication.

Arthropod-borne viruses (arboviruses) pose a major global public health threat, given their continuous and rapid (re-) emergence. These viruses are maintained in nature in a cycle between invertebrate vectors and vertebrate amplification hosts. The effect of DVGs on either or both hosts has not been investigated extensively. Here, we describe a method to triage DVG sequence space to identify fit DVGs generated during virus replication in both vertebrate and invertebrate environments. Further, we demonstrate that these DVGs can exert antiviral activities in both hosts, highlighting their potential as an intervention to perturb or control arbovirus infection and transmission.

Results

Experimental evolution to identify Zika virus defective viral genomes (DVGs) in vertebrate and invertebrate host environments. Our primary goal was to develop a method to identify DVGs with the best potential to inhibit WT virus. While hundreds to thousands of different DVGs are generated during virus replication, we hypothesized that DVGs capable of competing and interfering with WT virus should satisfy certain criteria: occur more frequently and with greater abundance than other DVGs in conditions favoring co-infection; be observed in several replicates and in different host cell types; and be packaged into virions to infect new cells over several passages. We thus passaged Zika virus in vertebrate (Vero) and invertebrate (C6/36) cell lines at high multiplicity of infection (MOI), a condition that favors co-infection of cells by multiple genomes and thus, complementation by WT virus. We additionally performed low MOI passages to compare DVGs that arose in conditions with reduced probability of complementation with WT virus. After 12 serial passages (Supplementary Fig. 1), we sequenced encapsidated RNA genomes to identify all deletions generated in the passage series. In total, we identified 6303 and 6184 different deletions in the Vero and C6/36 high MOI passage series, respectively (Supplementary Data 1 and 2). While small deletions (1–10 nucleotides) were common in both low and high MOI conditions, larger deletions were abundant only in high MOI conditions (Supplementary

Fig. 2), suggesting that larger deletions constitute competitive DVGs that rely on WT virus for propagation. It is worth noting that we did not observe a clear reduction in WT infectious virus in high MOI compared to low MOI conditions (Supplementary Fig. 1), suggesting that inhibitory DVGs did not reach high enough levels to outcompete WT virus under these growth conditions in this cell type.

A deletion cluster with higher relative fitness is identified across multiple flaviviruses. To down-select DVGs with the highest potential to outcompete WT virus, we identified deletions of higher relative fitness with respect to the overall population of DVGs presenting random deletions (that is, deletions occurring with significantly increasing frequency over passages). To do this, we computationally defined regions along the Zika virus genome where deletions were most predominant using a nested neighborhood algorithm. Three neighborhoods enriched in deletions were detected in Vero cell passages (Fig. 1a, left panel) encompassing: (i) deletions from nucleotide 500 to 3500, approximately (DVG-A) (ii) large deletions stretching from the beginning of the genome to nearly the end (DVG-B); and (iii) a small deletion around nucleotide position 5500 (DVG-C). In mosquito cells, only the DVG-A neighborhood was identified as significantly enriched with deletions (Fig. 1a, right panel). Deletions were uniformly distributed within each neighborhood (Supplementary Fig. 3). Enrichment z-scores then allowed us to estimate the significance of deletions within the neighborhood, across passages. The DVG-A neighborhood exhibited increasing enrichment z-scores over the passage series in Vero and C6/36 cells, whereas the other two showed the opposite effect (Fig. 1b). Thus, deletions in the DVG-A neighborhood were predicted as fit, whereas those in DVG-B and DVG-C neighborhoods were less fit. Importantly, the deletion hotspot corresponding to the DVG-A neighborhood was exclusively found in replicates performed at high, but not low, MOI conditions, with increased frequency compared to other deletions (Supplementary Figs. 4 and 5). Once this deletion hotspot appeared, it was maintained in subsequent passages, confirming propagation from one passage to another. By contrast, deletions corresponding to neighborhoods B and C were transient, found to appear in some passages and disappear in later ones. We next asked whether this type of DVG was unique to Zika virus, or was shared among the flaviviruses. We thus performed similar passaging experiments with two other flaviviruses, yellow fever virus (Fig. 1c) and West Nile virus (Fig. 1d) and found this same neighborhood of deletions in both cases.

High fitness DVG-A conserves the open reading frame following the deletion site. We evaluated the effect that DVG-A-type deletions inflicted on the remainder of the polyprotein still encoded in its sequences. We further delineated the deletion hotspot for both Vero and C6/36 cell lines with boundaries encompassing deletions with the highest frequencies (suggesting higher biological relevance): where the breakpoint was between nucleotide positions 500 and 900, and the rejoin was between nucleotide positions 2800 and 3400 (Supplementary Fig. 6a, d). Deletions outside the delineated region were randomly generated with respect to maintaining the open reading frame (ORF) following the deletion or frame-shifting. Strikingly, nearly 100% of deletions within the high-fitness hotspot conserved the ORF (Supplementary Fig. 6g). This observation was further supported by the yellow fever and West Nile virus data, for which 99.5% and 98.2% of deletions also conserved the ORF downstream of the deletion (Supplementary Fig. 6b, c, e, f, g).

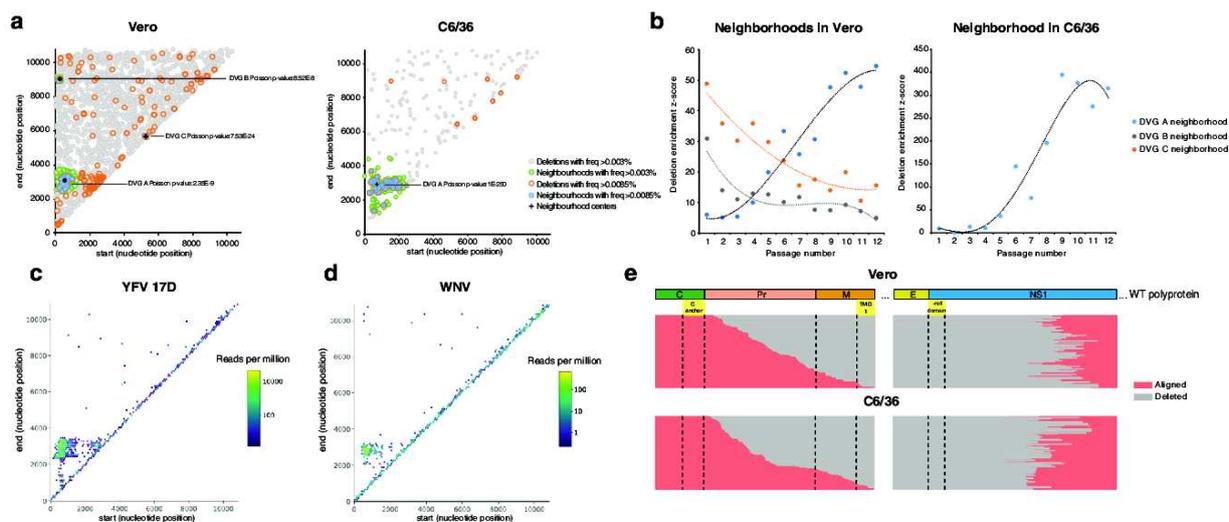


Fig. 1 Identification and genetic characterization of Zika virus DVGs with high fitness. Virus populations enriched in DVGs, passaged in Vero or C6/36 cells, were sequenced and deletions identified. **a** Computationally-determined neighborhoods enriched with deletions in Vero or C6/36 cells are shown in a 2-dimensional plane representation of the Zika virus genome, with start and end positions of the identified deletions denoted as dots on the plane. Poisson p values for each neighborhood are shown. **b** Dynamics of deletion enrichment within DVG neighborhoods identified in Vero and C6/36 cells. Enrichment z -scores indicate total enrichment significance of deletions within each neighborhood from all replicates per passage. **c, d** Hexbin plot of deletions identified following Yellow fever virus 17D (**c**) or West Nile virus high MOI passaging in SW-13 or C6/36 cells, respectively. Identified deletions with defined start and end positions within each hexagon (>10 nt in length, bin = 100) are shown. The color of each hexagon is relative to the sum of the reads per million of deletions in all passages and replicates. For YFV, passages number 1, 8, 14, and 19 ($n = 12$) were sequenced and shown. For WNV, data from all passages and replicates ($n = 3$) are shown. **e** Putative Zika virus DVG-A genomes identified in Vero or C6/36 cells that conserve the open reading frame aligned against the Zika virus polyprotein. Aligned regions are shown in pink and deleted regions in gray. Dashed lines denote limits of proteins or relevant protein domains. Source data are provided in the Supplementary Data 1 and 2 for Vero and C6/36 data, respectively, and in Supplementary Data 3 and 4 for Yellow fever virus and West Nile virus, respectively.

To understand if the resulting DVG-A polyprotein retained relevant signal peptides or domains required for correct processing, translocation, and topology in the ER membrane, we aligned each putative DVG within this hotspot with the viral polyprotein (Fig. 1e). For the 244 or 418 deletions identified in Vero or C6/36 cells, respectively, every deletion followed the capsid anchor that directs the translocation of Pr into the ER lumen. Concomitantly, these deletions ended at the N-terminus of NS1, which is also found in the luminal side of the ER. Thus, although the complete E protein, part of PrM, and NS1 are missing, the topology of the rest of the polyprotein is presumably maintained in these DVGs. Moreover, all deletions lack the β -roll domain of NS1, essential for its dimerization and function^{10,11}. Taken together, our analysis suggests that while other non-structural proteins would be properly translated and processed, the truncated NS1 protein of these DVGs would be non-functional.

DVG-A is non-replicative per se and relies on WT virus NS1 protein for genome replication. The ubiquity and high fitness of this DVG neighborhood led us to select DVG-A as the optimal candidate to be used as a viral competitor in downstream experiments. While some genomes containing deletions can be defective in producing their own structural proteins, they may retain their ability to self-replicate, and would thus be classified as replicons. Replicons are often used as surrogates to study replication of positive-sense RNA viruses, including Zika virus, because they retain all non-structural functions required to copy themselves and need WT virus only to encapsidate their progeny genomes^{12,13}. We thus evaluated if DVG-A was either a true replicon or a truly defective genome. We genetically engineered DVG-A to represent one of the most abundant deletions

identified in the DVG-A high fitness neighborhood. We also engineered a cognate DVG in which the ORF following the deletion was disrupted by removing a nucleotide at the deletion junction (hereafter referred to as DVG-A_{out-of-frame}). A NanoLuc reporter gene was inserted into each construct to measure genome replication by luminescence. As a positive control, we employed a WT replicon that can self-replicate and as a negative control, an inactive version of this replicon in which the catalytic motif at the polymerase active site was mutated (Fig. 2a). While the WT replicon was replication-competent on its own, the inactive replicon, DVG-A, and DVG-A_{out-of-frame} did not replicate (Fig. 2b, left side of panel). Importantly, DVG-A replication could be rescued in infected cells (Fig. 2b, right side of panel). Thus, DVG-A is not a true replicon, as it requires the presence of WT virus to replicate. DVG-A_{out-of-frame} did not replicate even in infected cells, indicating that preserving the ORF after the deletion is an absolute requirement for DVG-A genome replication.

Since DVG-A carries intact non-structural sequences except for NS1, we investigated whether this protein rescues DVG-A replication through trans-complementation. We assessed DVG-A genome replication in cells pre-transfected with plasmids encoding the non-structural proteins: namely a plasmid encoding all of the non-structural proteins, from NS1 through NS5 (E_{30-NS}); a control plasmid in which a stop codon was placed in lieu of the first amino acid of NS1, such that no non-structural proteins are expressed ($E_{30-NS*NS1}$); and a third plasmid in which a stop codon replaced the first amino acid of NS2A, so that only NS1 protein is expressed ($E_{30-NS*NS2A}$). Only cells expressing the entire non-structural proteins or the single NS1 protein could support DVG-A replication (Fig. 2c). Thus, DVG-A relies solely on WT virus-encoded NS1 function to facilitate its genome

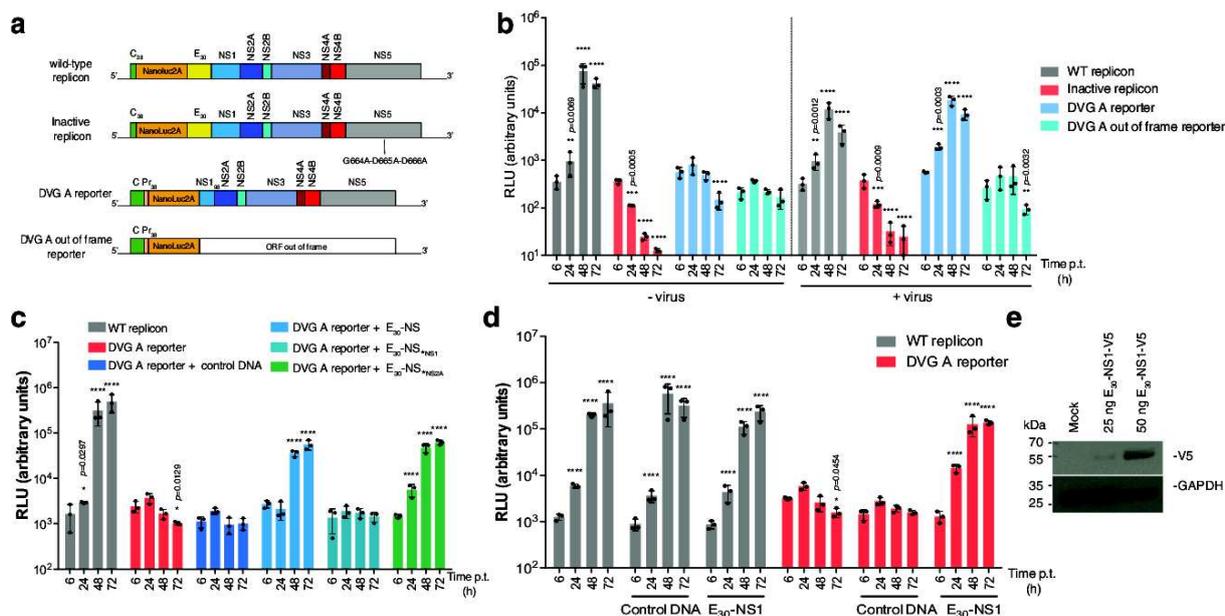


Fig. 2 Zika virus DVG-A relies on WT-encoded NS1 for its genome replication. **a** Schematic diagram of reporter constructs. For the wild-type replicon, C₃₈ and E₃₀ represent the N-terminal 38 aminoacids of C and the C-terminal 30 aminoacids of E proteins, respectively. For the DVG-A replicon, Pr₃₈ and NS1₉₈ represent the N-terminal 38 aminoacids and the C-terminal 98 aminoacids of Pr and NS1 proteins, respectively. NanoLuc2A represents the NanoLuc reporter sequence followed by the foot-and-mouth disease virus 2A protease. An inactive replicon (with a mutated polymerase active site) was also constructed. **b** Replicon assays in infected or uninfected cells. Equal amounts of replicon or reporter RNA was transfected in naïve or infected Vero cells (MOI 1 PFU/cell), and relative light units (RLU) measured at the indicated times post-transfection. WT and inactive replicons were used as controls. **** $p \leq 0.0001$. **c** DVG-A RNA reporter activity in HEK-293T cells pre-transfected with control DNA, E₃₀-NS1, E₃₀-NS1_{stopNS1} or E₃₀-NS1_{stopNS2A}. WT replicon was used as a control. RLU were measured at the indicated times p.t. **** $p \leq 0.0001$. **d** WT and DVG-A reporter assays in HEK-293T cells pre-transfected with control DNA or E₃₀-NS1. RLU were measured at the indicated times p.t. **** $p \leq 0.0001$. **e** Cells transfected with a V5-tagged version of the E₃₀-NS1 were harvested 24 h p.t. for western blotting, a representative blot of two is shown. Corresponding uncropped images of the western blot are shown in Supplementary Fig. 12. All graphs show the mean and SD; $n = 3$ per group for a representative experiment of three. Control DNA: GFP-expressing plasmid. Source data are provided as a Source Data file.

replication. Indeed, DVG-A replication was rescued in cells transfected with a plasmid encoding only the NS1 sequence (E₃₀-NS1), devoid of any downstream NS sequences that were present in the previous experiment (Fig. 2d). The expression of NS1 protein was confirmed by western blot with an antibody against a V5 epitope used to tag NS1 in the plasmid construct (Fig. 2e).

DVG-A inhibits WT Zika virus replication in vertebrate cells through competition for cellular resources. Our primary goal was to identify DVGs of high fitness that could best compete with WT virus. To evaluate this, we measured WT virus production in the presence of DVG. In addition to DVG-A and DVG-A_out-of-frame, we included two other DVGs representing deletion neighborhoods with lower fitness: DVG-B and DVG-C (Fig. 3a). DVG-A significantly decreased WT virus titers in a dose-dependent manner by up to 2.2-log (Fig. 3b). Interestingly, both DVG-B and DVG-C failed to inhibit WT virus production even at the highest DVG:WT molar ratio and no inhibition was observed when DVG-A_out-of-frame was used. Since DVG-A is missing the structural proteins, we further examined whether inhibition was related to competition for structural proteins provided by WT virus. To address this, we evaluated whether DVG-A could also inhibit a WT replicon lacking structural proteins. Indeed, DVG-A RNA significantly decreased WT replicon activity (Fig. 3c) and this effect was dose-dependent, with pronounced inhibition at the highest DVG:WT replicon RNA molar ratios (Supplementary Fig. 7a). Finally, since DVGs

of other viruses, such as the copyback DVGs of negative-sense RNA were shown to inhibit WT virus replication by stimulating innate immune responses^{3,14}, we measured the activation of three antiviral genes in cells harboring DVG-A and WT virus at a 10:1 molar ratio. No significant differences in the expression of interferon-alpha, interferon-lambda, or RIG-I were found (Supplementary Fig. 7a-c), indicating that DVG-A is no more immunostimulatory than WT Zika virus.

Thus, in mammalian cells, DVG-A seemingly inhibits WT virus production by entering in competition with WT virus for cellular machinery and resources required for genome replication, rather than by activation of innate immunity. These results also suggest that the observed inhibition requires the expression of the replicase and non-structural proteins directly by the DVG itself, once NS1 has been provided *in trans* by WT virus.

DVG-A inhibits WT Zika virus in invertebrate cells in an RNAi-dependent manner, regardless of whether the remaining ORF is in frame. To determine if DVG-A could also inhibit Zika virus in the invertebrate host environment, we infected C6/36 mosquito cells harboring either DVG-A, DVG-A_out-of-frame, or control RNA. We observed no reduction of virus titer for any treatment condition (Fig. 3d, left side of panel). Since C6/36 cells are known to have a defective Dicer-2 activity and they cannot initiate the RNA interference (RNAi) antiviral response^{15,16}, we repeated the experiment in Dicer-2 competent U4.4 mosquito cells^{17,18}. In this case, both DVG-A and DVG-A_out-of-frame

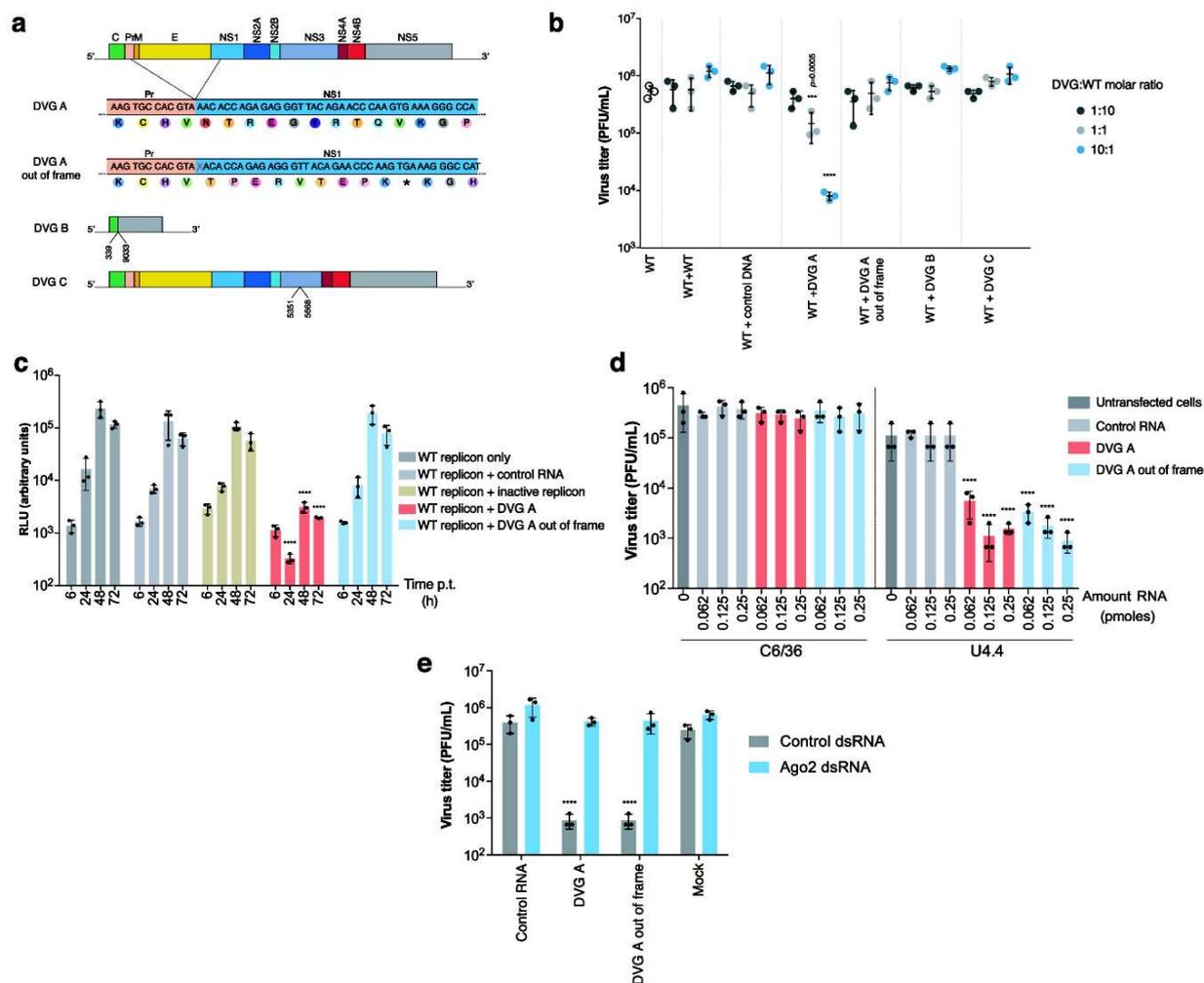


Fig. 3 DVG-A inhibits WT virus production in vertebrate and invertebrate cells. **a** Schematic diagram of the engineered DVGs. For DVG-A and DVG-A out-of-frame, the respective nucleotide and amino acid sequences are shown. Symbol (*) refers to a stop codon. The deleted nucleotide in DVG-A out-of-frame is crossed in red. For DVG-B and DVG-C, the nucleotides at the deletion junction are indicated. **b** Virus titers 72 h following co-transfection of DVG and WT-encoding plasmids in HEK-293T cells at different DVG:WT molar ratios (1:10, 1:1, and 10:1) are shown. More WT-encoding plasmid or control DNA were used as controls. **** $p \leq 0.0001$ (compared to WT + WT conditions by two-way ANOVA with Dunnett's multiple comparison). **c** Vero cells (commonly used for replicon assays) were transfected with WT replicon RNA complemented with control RNA, inactive replicon, DVG-A or DVG-A out-of-frame. Replicon activity was measured at the indicated time points p.t. **** $p \leq 0.0001$ (compared to RLU values in WT replicon only-transfected cells, by two-way ANOVA with Dunnett's multiple comparison). **d** Inhibition assays in mosquito cells. C6/36 (left) or U4.4 (right) cells were transfected with the indicated amount of control RNA, DVG-A or DVG-A out-of-frame. 18 h later, cells were infected with a MOI of 0.1 PFU/cell and virus titers measured 5 days p.i. **** $p \leq 0.0001$ (compared to untransfected cells by two-way ANOVA with Dunnett's multiple comparison). **e** Inhibition assays in cells knocked down for Ago-2 expression. U4.4 cells were transfected with Ago-2 or control dsRNA and the indicated RNA. 18 h later, cells were infected at an MOI of 0.1 PFU/cell and virus titers measured at 5 days p.i. **** $p \leq 0.0001$ (compared to control RNA-transfected cells by two-way ANOVA with Dunnett's multiple comparison). All graphs show the mean and SD; $n = 3$ per group for a representative experiment of three. Control RNA: pTRI-Xef RNA; control DNA: GFP-expressing plasmid. Source data are provided as a Source Data file.

elicited a significant decrease in virus titers (up to 2-log) compared to control conditions (Fig. 3d, right side of panel). Since RNAi is a sequence-dependent silencing mechanism, both DVG-A and DVG-A_{out-of-frame} inhibited virus replication, as a single nucleotide change in the DVG sequence should not prevent the mounting of an efficient RNAi response. Knockdown of Argonaute-2 (Ago-2, an essential component of the RNAi response) in U4.4 cells (up to 90% knock-down efficiency, Supplementary Fig. 7e) abrogated the inhibition of Zika virus infection (Fig. 3e). Therefore, DVG-A-induced inhibition of virus replication in the invertebrate host environment is governed by the RNAi response.

DVG-A can be packaged into virus-like particles (VLPs) for in vivo delivery. To continue exploring DVG-A as a potential therapeutic agent, we required an appropriate delivery method. Thus, we established a packaging system to encapsidate DVG-A into virus-like particles (VLPs). We first needed to confirm that DVG-A could be encapsidated at all. To this end, we transfected uninfected or infected Vero cells with DVG-A reporter or control WT replicon RNA. These cells, named 'producer' cells, should (if infected) produce WT virus progeny, as well as virions containing genomes that can be packaged. Transfer of the supernatant derived from infected producer cells to naïve (recipient) cells enabled us to assess packaging of reporter-expressing

genomes. Significantly higher WT replicon and DVG-A reporter activities were observed in recipient cells in the presence of WT virus, compared to background activity in the absence of virus, confirming that DVG-A was packaged by WT virus (Fig. 4a, producer cell reporter activity is shown in Supplementary Fig. 8a). Comparatively, no increase in reporter activity was detected for the inactive replicon or DVG-A_out-of-frame, suggesting that genome replication is a requirement for genome packaging.

Next, we developed a VLP-producing system to encapsidate DVG-A in the absence of WT virus. Briefly, we transfected HEK293T cells (chosen due to their high transfection efficiency) with plasmids encoding DVG-A, the structural proteins (CPrME), and the non-structural protein NS1, to enable packaging and replication of DVG-A, respectively (Fig. 4b). Packaging of DVG-A into VLPs was then assessed following transfer of supernatants onto naïve recipient Vero cells. Using reporter DVG-A, we confirmed that the reporter activity in recipient cells for packaged DVG-A was comparable to that of packaged WT replicon (Fig. 4c, producer cell reporter activity shown in Supplementary Fig. 8b). Finally, we confirmed that native DVG-A was packaged into VLPs, as shown by RT-qPCR of transfected cell supernatants (Fig. 4d). Importantly, we confirmed by plaque assay that no infectious virus was produced with the VLPs (Supplementary Fig. 9).

Thus, DVG-A can be actively packaged into VLPs, which could serve as therapeutic interfering particles (TIPs) for in vivo evaluation.

Zika virus TIPs reduce infection and virulence in mice. We next assessed the efficacy of these TIPs as an antiviral measure in mice. We used mice deficient in α/β and γ receptors (AG129 mice), as these are highly susceptible to Zika virus infection and disease progression^{19,20}. Mice were mock-infected (vehicle alone), infected with 10^4 PFU of WT virus alone, or with a mixture of WT virus and DVG-A TIPs. Approximately 10^4 DVG-A TIP genome copies per Zika virus infectious virion were used. As an additional control, mice received a mix of WT virus and supernatant from mock conditions of TIP production (“control TIPs,” generated in the absence of CPrME or NS1) (Fig. 5a). All mice lost weight by day 6, except mock-infected mice. At this time, TIP-treated (but not control TIP-treated) mice presented significantly lower weight loss than WT virus-infected mice (Fig. 5b). Further, unlike mice receiving WT virus alone or control treatment, TIP-treated mice presented significantly lower viremia at 2, 4, and 6 days p.i., with up to 2-log differences in virus titer (Fig. 5c). Viral loads in the footpads, spleens, ovaries, and brains were also 1–2 log lower in TIP-treated mice on day 6, underscoring the protective effect of this TIP (Fig. 5d). While DVG-A RNA in circulating blood was below detection levels, we confirmed the presence of DVG-A RNA in the footpads of TIP-treated mice, and at lower quantities in the spleens, ovaries, and brains (Fig. 5e). These results confirm that DVG-A TIPs successfully disseminate from the injection site to distal organs. Importantly, the safety profile of DVG-A TIPs was demonstrated in mice. In the absence of WT virus, DVG-A TIPs did not persist and did not disseminate to distal organs. The TIPs were degraded and undetectable within a day of administration (Supplementary Fig. 10).

In further support, we assessed TIP efficacy in a more immunologically competent mouse model²⁰. C57BL/6 mice were treated with IFNARI-blocking monoclonal antibody before infection with Zika virus, with or without TIPs. Weight loss was not monitored, as mice do not lose weight nor succumb to infection under these conditions²⁰. Viremia was significantly lower in TIP-treated mice at 3 and 6 days p.i. (Fig. 5f). While we did not observe

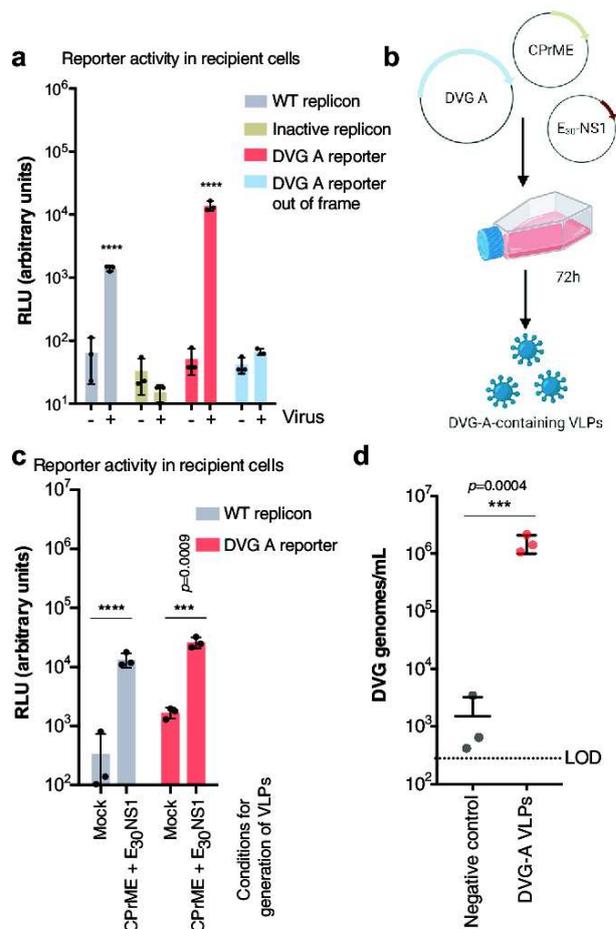


Fig. 4 Generation of DVG-A-carrying virus-like particles. **a** Assessment of DVG-A packaging ability by WT virus. Infected or uninfected Vero cells were transfected with WT replicon, inactive replicon, DVG-A reporter, or DVG-A out-of-frame reporter RNA. 48 h post-transfection, the cell supernatant was collected from donor cells, depleted of naked RNA, and used to infect naïve recipient cells, which were harvested 24 h later for measuring reporter activity. Luciferase activity in recipient cells is shown. **** $p \leq 0.0001$ (by two-way ANOVA with Dunnett’s multiple comparison, as compared to uninfected cells). **b** Schematic illustration of VLP assays: HEK-293T cells were transfected with a CPrME, E₃₀-NS1 and DVG-A encoding plasmids. 72 h post-transfection the cell culture supernatant with DVG-containing VLPs was harvested. **c** VLP assays using WT replicon and DVG-A reporter. Donor HEK-293T cells were transfected with WT replicon or DVG-A reporter plasmid, with or without (mock) CPrME and E₃₀-NS1. The cell culture supernatant was treated with nuclease and used to infect naïve recipient Vero cells. Reporter activity in recipient cells is shown. **** $p \leq 0.0001$ (by two-way ANOVA with Dunnett’s multiple comparison, as compared to mock conditions). **d** Quantification of DVG-A genomes in VLPs generated using native DVG-A. Clarified and nuclease-treated supernatant from donor cells were subjected to DVG-A quantification using RT-qPCR. Negative control refers to supernatant derived from cells transfected with the same amount of DVG-A-containing plasmid in the absence of CPrME or E₃₀-NS1. *** $p = 0.0004$ (by two-tailed t-test). All graphs show the mean and SD; $n = 3$ per group of a representative experiment out of three. Source data are provided as a Source Data file.

differences in viral loads in the footpads or spleens, viral loads in ovaries and brains were significantly lower in TIP-treated mice compared to control conditions, with up to 2-log difference in viral loads (Fig. 5g). As noted in AG129 mice, the high quantities of

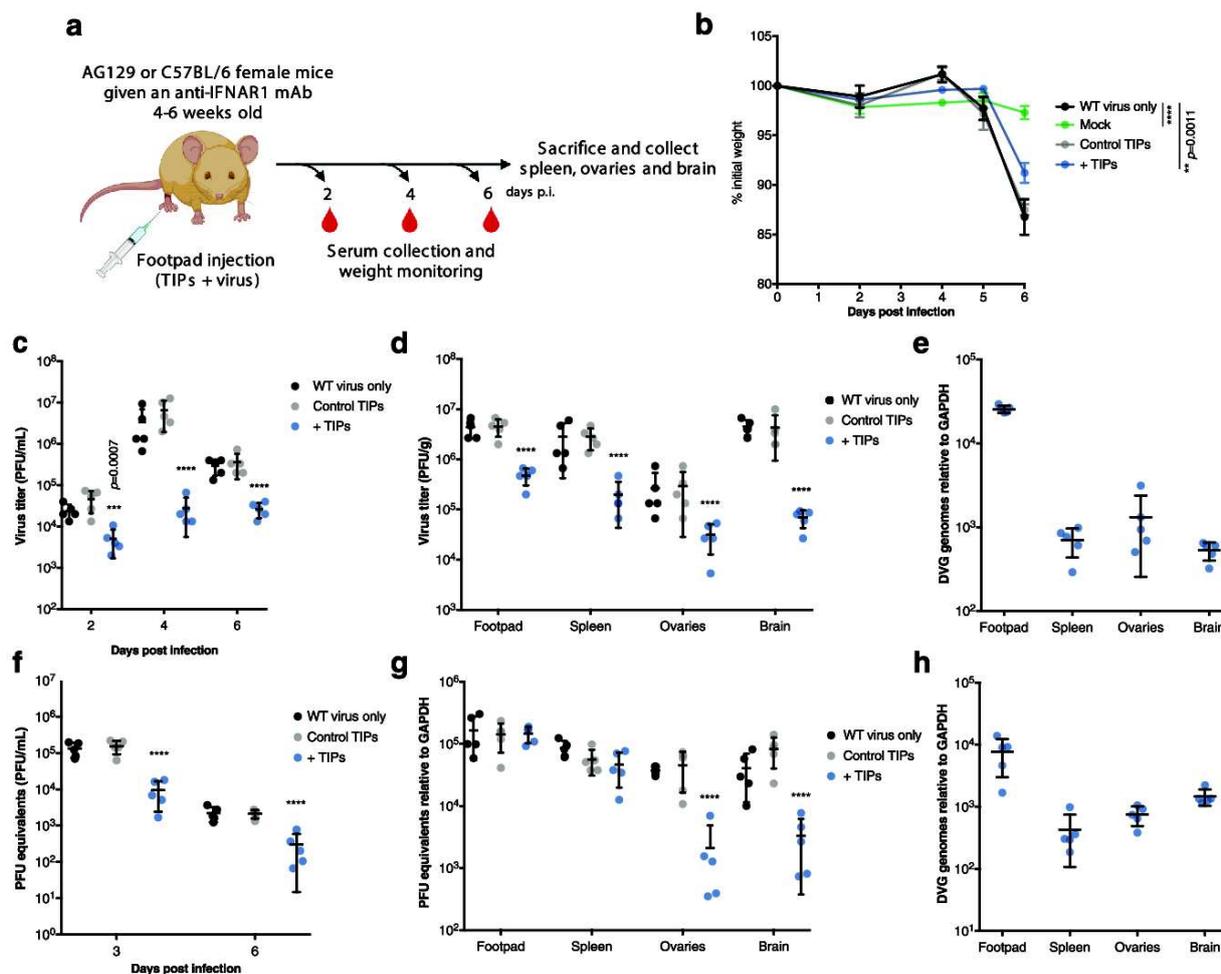


Fig. 5 Zika virus DVG-A-carrying TIPs inhibit virus replication in mice. **a** Experimental design. 4–6-week-old AG129 or C57BL/6 female mice were inoculated with WT virus diluted in DMEM, mock TIP supernatant (from transfections performed with no structural proteins), or with supernatant confirmed to contain DVG-packaging VLPs (TIPs). C57BL/6 mice were treated with 2 mg of IFNAR1-blocking monoclonal antibody 24 h prior to infection. At different times post-infection, sera were collected. 6 days p.i. mice were euthanized and organs harvested. **b–e** AG129 mice: **(b)** Weight loss of infected mice is shown as a % of weight at day 0 ($n = 5$). $****p \leq 0.0001$. **(c)** Viremia on days 2, 4, and 6 post infection ($n = 5$). $****p \leq 0.0001$. **(d)** Virus load in the injected footpad, spleen, ovaries and brain, 6 days p.i. ($n = 5$) $****p \leq 0.0001$. **(e)** DVG amounts in all organs collected measured by RT-qPCR ($n = 5$). **f–h** C57BL/6 female mice that were given an IFNAR1-blocking monoclonal antibody 24 h prior to infection. **(f)** 3 and 6 days p.i. viremia was measured by RT-qPCR and is depicted as PFU equivalents/mL ($n = 5$). $****p \leq 0.0001$. **(g)** 6 days p.i. mice were euthanized and virus load in organs measured by RT-qPCR ($n = 5$). $****p \leq 0.0001$. **(h)** DVG-A amounts in all organs measured by RT-qPCR ($n = 5$). All graphs show the mean and SD; $n = 5$. In all cases, two-way ANOVA with Dunnett’s multiple comparison was carried out comparing each condition to mice infected with WT virus only. Source data are provided as a Source Data file.

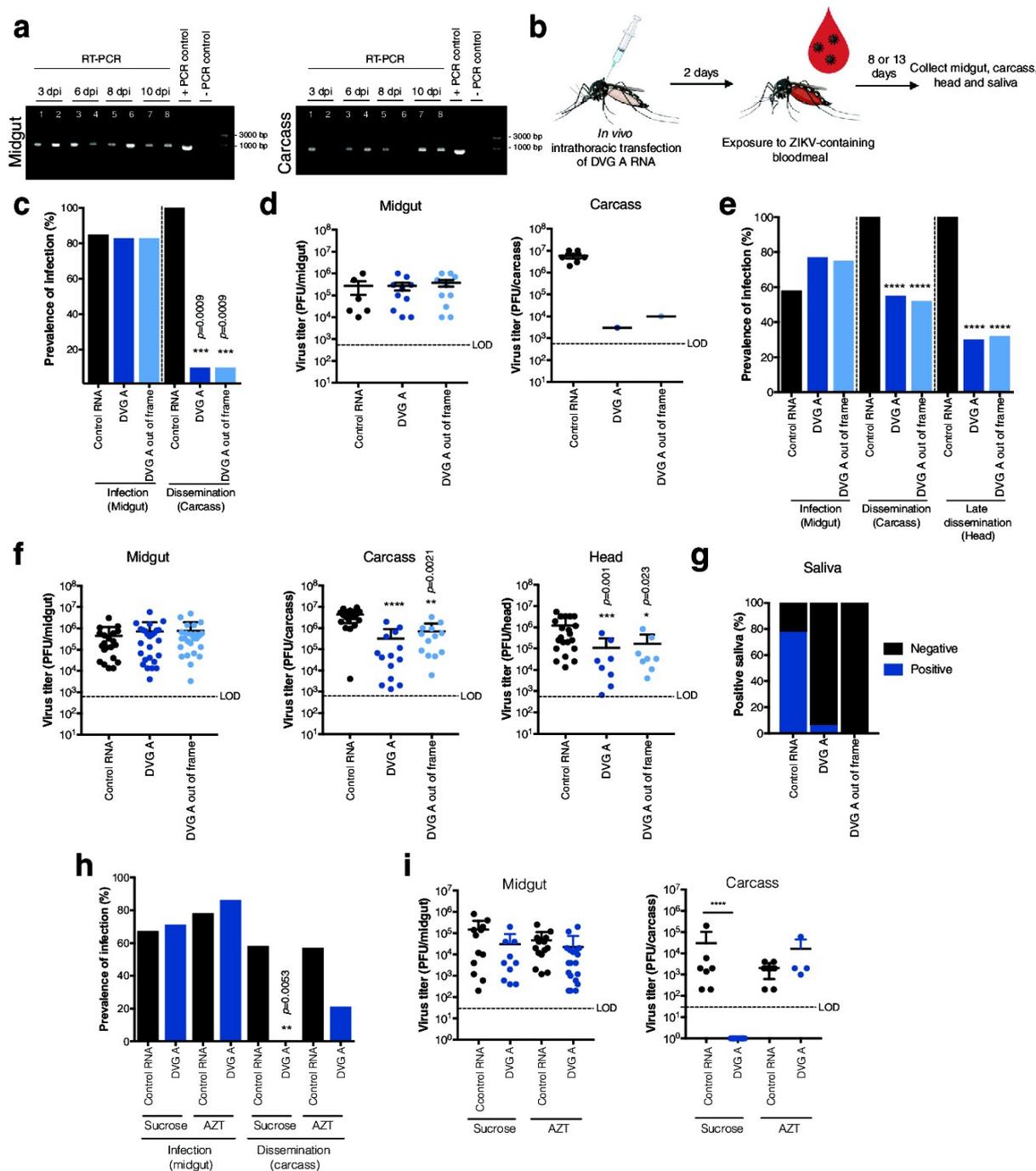
DVG-A at the injection site disseminated to the spleens, brains, and ovaries of TIP-treated mice (Fig. 5h).

DVG-A impairs Zika virus dissemination and transmission in mosquitoes. Finally, we investigated DVG-A inhibitory capacity in the mosquito vector. First, *Ae. aegypti* mosquitoes were injected with a mix containing equimolar amounts of RNA (DVG-A, DVG-A_out-of-frame or control RNA), anatomical compartments were dissected, and the longevity of DVGs in vivo was determined by RT-PCR. DVG-A and DVG-A_out-of-frame RNA was detectable in the midgut and carcass of transfected mosquitoes for at least 10 days post transfection, even in the absence of virus (Fig. 6a), while no RNA could be detected in head or saliva samples (Supplementary Fig. 11a).

Next, we fed mosquitoes an infectious blood meal containing Zika virus two days after injection of DVG or control RNA, and

examined viral loads in mosquito compartments 8 days after infection (Fig. 6b). The midguts and carcasses of mosquitoes were scored for the presence of virus to determine the infection and dissemination rates, respectively. Similar infection rates were observed, regardless of the pre-treatment mosquitoes received. While virus disseminated in 100% of mosquitoes that underwent control RNA pre-treatment, only 10% of mosquitoes that received either version of DVG-A had disseminated virus (Fig. 6c). In both cases, the viral titer in the single positive mosquito carcass from each group was 2–3-logs lower than the mean viral load in control RNA-transfected mosquito carcasses, whereas no differences were found in midgut viral loads (Fig. 6d).

To determine if dissemination was either completely blocked or delayed by DVG-A treatment, we repeated the experiment at a later time point (13 days post oral exposure to Zika virus). DVG-A RNA was still present in mosquitoes at this time point



(Supplementary Fig. 11b). It should be noted that a lower percentage of mosquitoes were productively infected compared to the previous experiment (58%, 77%, and 75% infection rates for control, DVG-A, or DVG-A_out-of-frame treated mosquitoes, respectively) (Fig. 6e). As in the previous experiment, 100% of control RNA-treated mosquitoes presented disseminated virus, whereas virus had disseminated in only 55% and 52% of DVG-A or DVG-A_out-of-frame treated mosquitoes (Fig. 6e). While 100% of heads (a proxy for late dissemination) of control RNA-treated mosquitoes were also virus-positive, only 30% or 32% DVG-A or DVG-A out-of-frame treated mosquitoes manifested late virus dissemination (Fig. 6e). Further, there were no

differences in viral load in the midgut, but the viral load was significantly lower in the carcass and heads of both DVG-treated groups than in control mosquitoes (Fig. 6f). Thus, both DVG-A and DVG-A_out-of-frame block or significantly delay Zika virus dissemination, and for mosquitoes in which dissemination is not blocked, disseminated viral loads are significantly lower than in control mosquitoes.

These observations prompted us to question if lower amounts of disseminated virus could result in impaired transmission. We thus analyzed saliva that was collected from a subset of mosquitoes in the last experiment. While infectious virus was present in the saliva of 78% of carcass-positive mosquitoes that

Fig. 6 Zika virus DVG inhibits virus dissemination and blocks transmission in experimentally-infected mosquitoes. **a** Longevity of DVG-A in mosquitoes. *Ae. aegypti* mosquitoes were in vivo transfected with equimolar solutions of DVG-A or DVG-A out-of-frame RNA. At the indicated times post transfection, 5 mosquitoes were salivated and midgut, carcass, and head dissected. RNA was extracted from homogenates of a pool of the 5 mosquitoes and RT-PCR specific for the DVG was performed. Results from a representative experiment out of two is shown. Lanes 1, 3, 5, and 7 = DVG-A out-of-frame, Lanes 2, 4, 6, and 8 = DVG-A. Results for head and saliva samples are shown in Supplementary Fig. 11a. **b** Schematic representation of the experimental design. *Ae. aegypti* mosquitoes were injected with a transfection mix 0.02 pmoles of RNA. 2 days post-transfection, mosquitoes were fed a bloodmeal containing 2×10^6 PFU/mL of Zika virus, and dissected at 8 or 13 d p.i. **c** Prevalence of Zika virus infection in the midgut or carcass of mosquitoes pre-transfected with control ($n = 7$), DVG-A ($n = 12$) or DVG-A out-of-frame ($n = 12$) RNA 8 d p.i. $***p = 0.0009$ (by Fisher's exact test, two-sided). **d** Viral load in infected mosquitoes from the midgut (control: $n = 6$; DVG-A: $n = 10$; DVG-A out-of-frame: $n = 10$) or carcass (control: $n = 6$; DVG-A: $n = 1$; DVG-A out-of-frame: $n = 1$) of the mosquitoes from (c). Mean and SD are shown. For midgut data, a one-way ANOVA with Dunnett's multiple comparison was performed. For carcass, no statistical test was performed because due to $n = 1$. **e** Prevalence of Zika virus infection in the midgut, carcass or heads of mosquitoes pre-transfected with control ($n = 36$), DVG-A ($n = 35$) or DVG-A out-of-frame ($n = 33$) RNA 13 d p.i. $****p \leq 0.0001$ (by Fisher's exact test, two-sided). **f** Viral load in the organs of the same infected mosquitoes from (e): midgut (control: $n = 21$; DVG-A: $n = 26$; DVG-A out-of-frame: $n = 25$), carcass (control: $n = 21$; DVG-A: $n = 13$; DVG-A out-of-frame: $n = 13$), or heads (control: $n = 21$; DVG-A: $n = 8$; DVG-A out-of-frame: $n = 8$). Mean and SD are shown. $****p \leq 0.0001$ (by one-way ANOVA with Dunnett's multiple comparison). **g** Percentage Zika virus-positive saliva in mosquitoes pre-transfected with control ($n = 9$), DVG-A ($n = 16$) or DVG-A out-of-frame ($n = 13$) RNA 13 d p.i. **h** Prevalence of Zika virus infection in the midgut or carcass of sucrose- or AZT-fed mosquitoes that were injected with control or DVG-A RNA. Prevalence was assessed at 7 d p.i. Control RNA, sucrose fed ($n = 18$); DVG-A RNA, sucrose fed ($n = 14$); control RNA, AZT fed ($n = 18$); DVG-A RNA, AZT fed ($n = 22$). $**p = 0.0053$ (by Fisher's exact test, two-sided). **i** Viral load in the midgut (control, sucrose fed: $n = 12$; DVG-A, sucrose fed: $n = 10$; control, AZT fed: $n = 14$; DVG-A, sucrose fed: $n = 19$) or carcass (control: $n = 7$; DVG-A: $n = 10$ <LOD; control, AZT fed: $n = 8$; DVG-A, sucrose fed: $n = 4$) of the same mosquitoes from (h). Mean and SD are shown. $****p \leq 0.0001$ (by one-way ANOVA with Dunnett's multiple comparison). Data from two independent experiments is shown. LOD = limit of detection. Control RNA: pTRI-Xef RNA. Source data are provided as a Source Data file.

received control RNA treatment ($n = 9$), only one carcass-positive mosquito (6.25%) contained infectious virus for DVG-A treated mosquitoes ($n = 16$), and no saliva of DVG-A out-of-frame treated mosquitoes ($n = 13$) had infectious virus (Fig. 6g).

We have previously shown that arboviruses establish persistent infections in the mosquito host via the production of a viral-derived DNA (vDNA) and that DVGs can serve as templates for the synthesis of this vDNA^{21,22}. To assess whether the Zika virus DVG-induced protection in mosquitoes was mediated by the presence of vDNA, DVG-A, or control RNA-injected mosquitoes were treated with azidothymidine (AZT), a compound that inhibits reverse transcription and thus the synthesis of the vDNA^{22,23}. AZT treatment rescued dissemination rates (Fig. 6h) and viral load of disseminated virus (Fig. 6i) at 7 days post infection, confirming that in the mosquito host, Zika virus DVGs mediate an RNAi-dependent inhibition of WT virus via vDNA production.

Discussion

The potential to use DVGs as antivirals or prophylactics has previously been proposed in studies showing that virus stocks rich in DVGs can interfere with disease outcomes in mice or ferrets^{24–27}. To date, synthetically engineered DVGs with interfering potential were shown to be effective as prophylactic (vaccine adjuvants) or therapeutic antivirals for two respiratory viruses, RSV and influenza A, in mouse models^{28–30}. DVG diversity has been largely overlooked in the field, with only few DVGs per virus family described. This is because early reports relied on classic methods of isolation, such as RT-PCR amplification, that bias toward the shortest, most readily amplified, and abundant DVGs. Here, using a combined approach based on experimental evolution, NGS, and computational methods to measure fitness, we explored the vast DVG sequence space in three different flaviviruses to identify a top candidate DVG with potential to interfere with WT virus infection. Indeed our NGS data show that deletions occur randomly throughout the viral genome with no apparent bias (Supplementary Fig. 6a–c) and suggest that the subset observed at higher frequency is dictated by biological constraints, such as the retention of coding sequences, replication elements, and packaging signals. Our results

demonstrate that DVG-A meets all these conditions, and could explain why the DVG-A cluster exists.

One key element to success was relying on high replicate numbers in the passage series that is customary in evolutionary sciences, rather than the minimal triplicate that is the default in molecular biology. Triplicate passage simply would not have been enough to observe convergent evolution in experiments, as the variation among replicates illustrated (Supplementary Fig. 4). The methodology described here to identify DVG therapeutic candidates could be applied to any virus species. While we focused solely on deletions, other types of DVGs such as rearrangements, insertions or a combination of these that cannot be easily characterized by short-read sequencing may also constitute therapeutic candidates. Emerging sequencing technologies such as long-read sequencing or ClickSeq³¹ and analysis tools such as ViReMa³² will allow us to identify more complex genomes in a high-throughput manner.

As expected, our list included DVGs that resemble those identified by more classic methods, similar to the dengue virus DVGs with large deletions containing only key regulatory elements at the ends of the genome³³. However, our analysis predicted and we experimentally confirmed that such DVGs did not have sustained fitness and showed no antiviral activity. This observation questions the classical notion that competition between DVGs and WT genomes is mostly a matter of genome size, where the smallest DVGs would be replicated at much faster rates and outcompete WT virus. Instead, we found that a DVG with a deletion of the structural proteins, that retains non-structural protein sequences except NS1, was most fit. This same DVG was identified as high fitness in other flaviviruses and found to be most capable of inhibiting WT virus. DVGs similar to DVG-A have been reported in flavivirus persistent infections in vitro and in natural isolates^{7,34–38}, although their naturally occurring frequency was seemingly not high enough to successfully interfere with WT virus.

By packaging the fittest DVG candidate into VLPs, we show that these TIPs are not only active at the site of injection in the mammalian host, but disseminate to distal organs that are primary targets of Zika virus infection, an important feature for combatting most arbovirus infections. Importantly, the reduced viral loads in brains and ovaries of TIP-treated mice suggest that

DVGs may help avert the neurological complications of Zika virus infection in severe cases of the disease^{39,40}, or vertical and/or sexual transmission of the virus^{41–45}. Thus, our results provide a proof of concept that DVGs can target infection against flaviviruses such as Zika virus in the vertebrate acute infection cycle. A limitation of our study, in terms of antiviral efficacy, is that we did not evaluate to what extent TIPs can be administered after virus exposure, since we only administered TIPs at the same time as virus infection. The success of this DVG in the vertebrate host may lie in its ability to compete and interfere with WT on more than one level. First, it competes with WT virus in the most classic parasitic sense by ‘stealing’ key viral proteins away from virus, which it requires for genome replication and packaging (e.g. NS1 or structural proteins). Second, while it may interfere with WT replication by taking viral replicase proteins away from the virus, it may not solely rely on WT-derived non-structural proteins. In fact, while DVG-A is not a true replicon, and it behaves like an inert replicon with an ‘On’ switch that only becomes active once NS1 is supplied in *trans*, whereby it then produces its own non-structural proteins. These DVG-encoded proteins can then enter into competition with WT virus proteins for cellular resources as well.

While studies on DVGs from other viruses showed them to be strong triggers of innate immune responses affecting viral replication^{3,14,46–50}, DVG-A did not upregulate these interferon-related genes. Further, the mouse models used here are either deficient in interferon α/β and γ receptors or have temporally diminished IFN α/β signaling. Thus, the protection conferred by DVG-A in these mice provides further evidence that the mechanism of action by DVG-A is independent of the type I and II interferon responses. Presumably, DVGs such as DVG-A with internal deletions are weaker activators of the interferon response than those used in previous studies, such as copy-back DVGs, which can form dsRNA known to be a critical factor in immunostimulation⁴⁶.

Contrary to what was observed for acute infections in the vertebrate host, DVG-A was antiviral in mosquito cells regardless of whether the ORF was maintained, suggesting that the antiviral activity of the DVG is independent of the proteins or functions it encodes. Indeed, RNAi is the major antiviral pathway in insects, and our work shows that inhibition by DVG-A relies on an intact RNAi pathway. Furthermore, we show that the major mechanism by which DVGs inhibit viral infection in insects is through the generation of a DNA form of the DVG RNA which boosts the RNAi response. Indeed, our previous work suggested that DVGs are a preferential template from which to for the viral DNA form²¹. Our results provide a proof of concept that DVGs could constitute a new vector control strategy for any arbovirus. Previous work showed that West Nile virus stocks containing DVGs correlated with decreased infection rates in *Culex* mosquitoes⁷. Here, we show that synthetic DVG-A RNA introduced into mosquitoes can block virus dissemination in mosquito bodies by 50% and most importantly, reduce transmission by 90%. Since inhibition occurred regardless of the retention of the reading frame, it is plausible that a large array of DVGs presenting random deletions could be as effective at reducing transmission in mosquitoes if expressed at high enough amounts. Given the recent advances in mosquito genome editing techniques^{51–53}, it would be interesting to assess the use of knock-in mosquitoes carrying DVGs or DVG-derived sequences in endemic settings and to apply this approach to other arboviruses to reduce global disease burden.

Arboviral diseases constitute a public health threat worldwide. To tackle the (re-)emergence of arboviruses, novel strategies for therapeutic and prophylactic treatments, as well as vector control are required. Our study provides an in-depth approach to identify

fit, competitive DVG molecules naturally generated by a virus and most suitable for interfering with infection, and underscores the potential of DVGs to address these needs; not only as therapeutics by reducing virulence in the vertebrate host, but also as a vector control by limiting virus dissemination and transmission during persistent infection of the natural mosquito vector. Given that virtually all RNA viruses can generate DVGs, our experimental and computational method to select high fitness DVG may be useful as a global intervention strategy against RNA viruses.

Methods

Cells and viruses. Vero and Vero-E6 cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% (v/v) fetal bovine serum (FBS, Invitrogen). HEK-293T cells were maintained in the same medium supplemented with 1% (v/v) non-essential amino acids. SW-13 cells were maintained in Leibovitz’s L-15 medium (Life Technologies) with 2mM L-glutamine and 10% (v/v) FBS. C6/36 and U4.4 were maintained in Leibovitz’s L-15 medium (Life Technologies) supplemented with 10% FBS (v/v), 1% (v/v) non-essential amino acids and 1% (v/v) tryptose phosphate broth (Sigma). All cell lines were supplemented with 5 units/mL penicillin and 5 μ g/mL streptomycin (Life Technologies). Mammalian cell lines were maintained in a humidified atmosphere at 37°C with 5% CO₂. C6/36 and U4.4 cells were maintained at 28°C without CO₂.

The Zika virus strain used in this study is the prototype African MR-766 strain, derived from a previously described infectious clone⁵⁴. Rescue of WT Zika virus was carried out by transfection of the infectious clone in HEK-293T cells (TransIT-LT1, Mirus Bio) according to the manufacturer’s instructions. 3–4 days post-transfection (p.t.), the cell culture supernatant was clarified by centrifugation. Virus stocks were then prepared in Vero cells, by infecting cells with a multiplicity of infection (MOI) of 0.01 plaque-forming units (PFU)/cell. 5–6 days post-infection (p.i.) the cell culture supernatant was clarified by centrifugation, and virus titers determined by plaque assay in Vero-E6 cells.

The yellow fever virus 17D strain was derived from pACNR/FLYF plasmid containing the full-length infectious YF17D genome under a SP6 promoter⁵⁵.

West Nile virus (Israeli strain IS-98) was generated from the “two-plasmid” system infectious clone previously described⁵⁶.

Serial passaging of Zika virus to generate virus populations enriched in DVGs.

Zika virus stocks were used to inoculate Vero or C6/36 cells at a low (0.01 PFU/cell) or high (20 PFU/cell) MOI in 12-well plates. 3 or 5 days p.i. (for Vero or C6/36 cells, respectively) the cell culture supernatant was clarified by centrifugation and used for the next passage. After passage 1 (p1), blind passaging was performed, using 5 μ l or 300 μ l of the virus from the previous passage per well for low and high MOI conditions, respectively. All samples were titrated and we confirmed that indeed the high and low MOI conditions were respected (Supplementary Fig. 1). Yellow fever virus was passaged at high (5 FFU/cell) and low (0.1 FFU/cell) MOI in SW13 cells for 2 days. West Nile virus was passaged similarly, in C6/36 cells.

Viral RNA sequencing. RNA from clarified cell culture supernatants was extracted using TRI reagent (Sigma). Purified viral RNA was quantified with a Qubit fluorometer (Invitrogen) and used for sequencing by next-generation RNA sequencing. Libraries were prepared using the NEBNext Ultra II RNA Library preparation kit for Illumina (New England Biolabs) and loaded in a NextSeq500/550 Mid output kit v2.5 (Illumina), for sequencing in a NextSeq500 (151 cycles, 8 nucleotides of index).

Sequence data processing and analysis. Reads were first demultiplexed using the bcl2fastq conversion software (Illumina). Next, the reads were adapter- and quality-trimmed using BBDuk and optical duplicates removed using Clumpify. Both BBDuk and Clumpify are part of the BBTools and BBMap suite version 38.06 (<https://sourceforge.net/projects/bbmap/>). Reads were then aligned to the reference genome using BBMap and deletions called with the BBMap variant caller CallVariants. The number of reads mapping to a deletion was normalized relative to the total number of reads that aligned to the reference genome. This value was multiplied by 10⁶ to obtain deletion event reads per million reads (RPM), as described previously³.

Computational determination of neighborhoods enriched with deletions.

“Junction” reads (reads that align to the reference virus genome but not as a continuous alignment) were grouped into clusters of fragments with similar start and end deletion positions. Specifically, deleted fragments with less than 10 positions in divergence in their start and end positions were grouped in a cluster (“elementary deletion cluster”). The size of the elementary deletion cluster is a “junction coverage” of the deletions’ narrow start and end position interval by reads with the corresponding junctions. The ratio of “junction coverage” to the sum of “junction coverage” and “continuous coverage” (derived from reads with

continuous alignment that cover these deletions' start and end intervals) is a frequency of the elementary deletion cluster.

The most abundant deletions (with frequencies higher than 0.0085%) from all high MOI replicates were selected to determine regions in the viral genome in which deletions were more predominant. Since deletions with this threshold are not uniformly distributed, enriched areas can be found unequivocally by the applied nested neighborhood algorithm in a 2-dimensional plane composed of start and end positions of deletions as X and Y coordinates. This method allows detecting the area (neighborhood) enriched by points around a certain center by sequentially extending the neighborhood's border (distance to the center from the next closest point) and calculating the neighborhood's fractal dimension to get more accurate *p* values of enrichments on each step. For methods describing the calculation of fractal dimension, refer to the "Fractal dimensions of neighborhoods" section.

Putative centers of the neighborhoods enriched by deletions were detected on the plane using a grid method. Namely, a number of deletions ("points") were randomly selected as the grid references. For all points, distances to these reference points were calculated. For every point, a product of all-rounded logs of its Euclidean distances to the reference points was used as a hashing index. The hashing indexes of all points were sorted, and big enough islands of points (threshold ≥ 7 elementary deletion points) in the sorting that have the same hash index were considered as containing putative centers of significant enrichment. Any point of the island can be used as a center for the subsequent determination of the center's neighborhood most significantly enriched by points/deletions. Let the null hypothesis assumption be that all points/deletions are uniformly distributed on the start/end plane (i.e., no enrichments assumption). Then probability for a number of points to be in a neighborhood of radius *r* from the center with volume (or an area if the space is a two-dimensional plane) *V_r* can be calculated from Poisson distribution. Indeed, if points are uniformly distributed in the neighborhood with radius *R*, *R* > *r* of volume *V_R*, then the number of points in a neighborhood with the same center and radius *r* (therefore, with volume *V_r*) will be a random variable, having a Poisson distribution with the parameter $\lambda = \alpha \cdot r^k$, where *k* is the dimension of the space, and α is the density of the uniform distribution of *n* points in *V_R* ($\alpha = \frac{n}{V_R}$). Thus, *n* is proportional to *V_r*. The probability (*P*) of finding more or equal to *m* points in a neighborhood with radius *r* (*p*value) will be:

$$P_m(r) = \sum_{i=m}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = \sum_{i=m}^{\infty} \frac{(\alpha \cdot r^k)^i}{i!} e^{-(\alpha \cdot r^k)} \quad (1)$$

All deletions/points were sorted according to closeness to the selected central point. Those regions that are most enriched by points were determined as follows. In the sorting, let us consider a transition from a neighborhood with radius *r_p* which is equal to a distance from the center to point *t* in the sorting, to a neighborhood with radius *r_{t+1}*. The Poisson *p* value of the enrichment of the neighborhood of radius *r_t* containing *t* points is calculated from the perspective of the extended neighborhood with the radius *r_{t+1}* and assumptions that its *t* + 1 points are uniformly distributed in this volume of space of fractal dimension: the fractal dimension is defined by the sequence of distances from the *t* + 1 points to the center. For non-uniform dense areas, the Hausdorff fractal dimension is higher than the geometrical dimension of the plane equal to two. This higher dimension makes a drop of *p* value sharper on a transition from *t* to *t* + 1 than in two-dimensional space. The *t*-neighborhood with the most significant Poisson *p* value was selected as the best neighborhood, i.e., the one that is most enriched by deletions/points.

Statistical significance of deletion frequency within enriched neighborhoods.

We assumed a binomial distribution for a number of "junction coverage" events in trials numbered as the sum of numbers of all "junction coverage" and all "continuous coverage" events. Since the "continuous coverage" was high, the 95% confidence interval for the "junction coverage" frequency was evaluated via a chi-square distribution for binomial log-likelihood ratio (Wilks theorem). In this way, the significance of the deletion, i.e., its deviation from zero, can be measured in standard deviation (SD) units of the normal distribution (z-score of the elementary deletion). This SD-unit measurement is based on the consideration that the 95% confidence interval of the normal distribution nearly equals to four SDs.

Significance of a cluster of deletions in a neighborhood was inferred by calculating the z-score of the neighborhood. For each passage, the neighborhood z-score was calculated as the sum of z-scores of its deletions divided by the square root of the total number of deletions in the neighborhood. Since significance of each deletion is measured in SD and therefore distributed normally *N*(0,1), the calculated z-score for total significance of all deletions in the neighborhood is also distributed normally *N*(0,1) and therefore easily translated to a *p* value. Thus, the dynamics of the neighborhood z-scores across passages show how the total significance of the neighborhood's set of deletion frequencies increases or decreases.

Fractal dimensions of neighborhoods. Fractal dimensions of neighborhoods were calculated for a more sensitive determination of the most enriched neighborhoods. A fractal dimension for a neighborhood of a center of radius *r_{t+1}* with *t* + 1 closest distinct points in it can be calculated as follows. Let each point in the neighborhood be

exchanged with the same neighborhood compressed with the $\frac{\text{minimumDistance}}{\text{maximumDistance}} \cong \frac{r_1}{2 \cdot r_{t+1}}$ ratio. Next, each point in the compressed neighborhood *N₁* will be exchanged with the double compressed original neighborhood: compressed with the $\left(\frac{r_1}{2 \cdot r_{t+1}}\right)^2$ ratio, next each point of the double compressed neighborhood will be exchanged with the triple compressed neighborhood: $\left(\frac{r_1}{2 \cdot r_{t+1}}\right)^3$ ratio, and so on.

In this nested construction, balls of diminishing radiuses cover the infinite sequence of the nested compressed self-similar (fractal) ball-neighborhoods. Hence, the Hausdorff fractal dimension of the ball-neighborhood can be calculated as a minimal dimension *d* of a ball (volume without compression is *r_{t+1}^d*) in order to get the lowest estimation for sum of volumes of infinite series of all diminishing balls that cover all points in this nested fractal construction.

The Hausdorff content of this series is

$$H^d = \lim_{n \rightarrow \infty} \left\{ (t+1)^n \cdot \left[r_{t+1} \cdot \left(\frac{r_1}{2 \cdot r_{t+1}} \right)^n \right]^d \right\} \quad (2)$$

Dimension *d* of this Hausdorff content of the ball-neighborhood is defined as:

$$\dim(H^d) = \inf \left\langle d > 0 : \lim_{n \rightarrow \infty} \left\{ (t+1)^n \cdot \left[r_{t+1} \cdot \left(\frac{r_1}{2 \cdot r_{t+1}} \right)^n \right]^d \right\} = 0 \right\rangle \quad (3)$$

Alternatively, the equivalent formula will be as follows:

$$\dim(H^d) = \inf \left\langle d > 0 : \lim_{n \rightarrow \infty} \left\{ (r_{t+1})^d \cdot \left[(t+1) \cdot \left(\frac{r_1}{2 \cdot r_{t+1}} \right)^d \right]^n \right\} = 0 \right\rangle \quad (4)$$

This infimum for dimension is calculated as:

$$d = \frac{\ln(t+1)}{\ln(2r_{t+1}) - \ln(r_1)} \quad (5)$$

Indeed, for the above limit under *n* → ∞ to be equal to zero the inequality as follows has to be true:

$$(t+1) \cdot \left(\frac{r_1}{2 \cdot r_{t+1}} \right)^d < 1 \quad (6)$$

Taking the logarithm from both parts, the inequality will be transformed into equivalent:

$$\ln(t+1) + d \cdot \ln\left(\frac{r_1}{2 \cdot r_{t+1}}\right) < 0 \quad (7)$$

$$d > \frac{\ln(t+1)}{-\ln\left(\frac{r_1}{2 \cdot r_{t+1}}\right)} \quad (8)$$

Taking infimum we will get:

$$d = \frac{\ln(t+1)}{\ln(2 \cdot r_{t+1}) - \ln(r_1)} \quad (9)$$

Generation of DVG clones. The Zika virus infectious clone was used as a template for the generation of DVG-encoding clones, using primers to excise deleted region of interest. The deleted nucleotides in each DVG clone were: 581-3250 for DVG-A, 339-9033 for DVG-B, and 5351-5668 for DVG-C. Zika virus replicons were generated as previously described¹². Briefly, the region encoding the structural proteins (C, Pr, M, and E) was replaced by a cassette containing C38-NanoLuc-2A-E30 (the N-terminal 38 amino acids of C protein, NanoLuc reporter, foot-and-mouth disease virus 2 A protease, and the C-terminal 30 amino acids of the E protein). The N-terminal 38 amino acids of the C protein contains the *cis*-acting element of nucleotides complementary to the 3' cyclization sequence⁵⁷. The C-terminal 30 amino acids of E correspond to the transmembrane domain required to maintain the topology of NS1 in the ER. In addition, the intron present in NS1 in the Zika virus clone was removed. A mutant replicon plasmid was generated for use as a negative control, by replacing the catalytic motif G₆₆₄D₆₆₅D₆₆₆ in the viral polymerase NS5 by three alanine residues (inactive replicon). The DVG-A reporter clone was generated similarly; with NanoLuc-2A introduced at the deletion site, namely between the N-terminal 38 amino acids of Pr and the C-terminal 98 amino acids of NS1.

Replicon assays. To generate replicon RNA, the full replicon cassette from CMV-driven clones was amplified using primers annealing to the backbone of the plasmid, and in which the forward primer contained an SP6 sequence. RNA was in vitro transcribed from the PCR product (mMESSAGE mMACHINE SP6 Transcription kit, Thermo Fisher), and purified using the RNeasy mini kit (QIAGEN). RNA was quantified using a Qubit fluorometer (Invitrogen) and used for transfection of HEK-293T or Vero cells. Briefly, sub-confluent cells in 24-well plates were transfected with 100 ng of replicon RNA using TransIT-mRNA transfection kit (Mirus Bio). Over a 72 h time course, the supernatant was removed and cells lysed in passive lysis buffer (Promega). Luciferase expression was

measured using the Nano-Glo luciferase assay system (Promega) in a Tecan infinite M200 pro plate reader (Tecan group).

Assays in which the effect of non-structural proteins on DVG replicon activity was studied were performed as described above, except HEK-293T cells were transfected with E₃₀-NS, E₃₀-NS^{NS1}, E₃₀-NS^{NS2A}, E₃₀-NS1 plasmids using LT-1 transfection reagent (Mirus Bio) 24 h prior to reporter RNA transfection. E₃₀-NS is a CMV-driven clone encoding the Zika virus C-terminal 30 amino acids of E and all nonstructural proteins. E₃₀-NS^{NS1}, E₃₀-NS^{NS2A} plasmids are identical to E₃₀-NS, except a stop codon introduced as the first amino acid of NS1 or NS2A, respectively. E₃₀-NS1 encodes only E₃₀ and NS1. We performed these assays in HEK-293T cells because of their high transfection efficiency, which proved to be important for the successful transfection of both DNA and RNA molecules. To confirm successful expression from the clones, we generated a C-terminal, V5-tagged version of E₃₀-NS1. Cell lysates of transfected cells were collected at 24 h p.t. in RIPA buffer (Sigma). V5-tagged NS1 was detected by conventional western blotting, using an anti-V5 antibody (Abcam; ab27671) and anti-mouse HRP linked antibody (GE Healthcare Life Sciences).

DVG inhibitory activity in vitro. HEK-293T cells were transfected with the DVG-encoding clones along with the WT Zika virus clone at different molar ratios (1:10, 1:1 and 10:1 DVG:WT) using LT-1 transfection reagent (Mirus Bio). As controls, additional WT virus clone or a GFP-encoding clone (control clone unrelated to Zika virus) were used. 72 h p.t. WT virus titers were determined by plaque assay.

Due to the low efficacy of the CMV promoter in invertebrate cells⁵⁸, we tested DVG-induced inhibition in mosquito cells by transfection of DVG RNA. Briefly, PCR amplicons of the full DVG sequences were generated using primers with an SP6 promoter at the 5' end. SP6 in vitro transcription was then performed using the PCR product as a template (mMessage mMachine SP6 Transcription kit, ThermoFisher). Following DNase treatment, the reaction was cleaned up (RNeasy MinElute Cleanup kit, QIAGEN) and RNA quantified with a Qubit fluorimeter (Invitrogen). 0.062, 0.125, or 0.25 pmoles DVG or pTRI Xef control RNA were transfected into sub-confluent C6/36 or U4.4 cells seeded in a 48-well plate (Lipofectamine LTX, ThermoFisher). The next day, cells were infected with Zika virus at an MOI 0.1 PFU/cell and the medium replaced. 5 days p.i., infectious virus in the cell culture medium was quantified by plaque assay. For assessing the effect of DVG RNA in Ago-2 knockdown cells, co-transfection of the DVG or control RNA was carried out with Ago-2 or control dsRNA.

Induction of antiviral genes. For quantitative reverse transcription PCR (RT-qPCR) of antiviral genes, total cellular RNA was reverse transcribed using the Maxima H Minus First Strand cDNA Synthesis Kit with random primers (ThermoFisher). cDNA was used for qPCR using specific primers (Supplementary Table 3) and the Power SYBR Green PCR Master Mixture (Applied Biosystems) on a Step-One-Plus Real-Time PCR thermocycler (Applied Biosystems). The relative fold gene expression of genes was calculated relative to GAPDH using the delta-delta Ct method, comparing to mock conditions.

Knockdown of Ago-2 in U4.4 cells. RNA from U4.4 cells was extracted (TRIzol, Thermo Fisher) and cDNA generated using the Maxima First Strand cDNA Synthesis Kit with random primers (Thermo Fisher). A 686 bp Ago-2 region was amplified from the cDNA using specific T7-promoter flanked primers (T7-Ago2-F and T7-Ago2-R, Supplementary Table 3). PCR products were cleaned up (NucleoSpin Gel and PCR Clean-up, Macherey Nagel) and used for dsRNA production (MEGAScript RNAi kit, Thermo Fisher). Knock-down of Ago-2 was performed by transfection of 250 ng dsRNA per well in sub-confluent U4.4 cells seeded in a 48 well plate (Lipofectamine LTX, Thermo Fisher). Control knock-downs were carried out using the control dsRNA provided in the MEGAScript RNAi kit.

The efficiency of Ago-2 knockdown was assessed by mRNA quantification. Briefly, 24 h following dsRNA transfection, cells were lysed in RNA lysis buffer and RNA extracted using the *Quick-RNA* 96 kit (Zymo), including the in-column DNase digestion step to remove genomic DNA. RT-qPCR was carried out using the Luna Universal One-Step RT-qPCR kit (New England Biolabs) following the manufacturer's instructions, with Ago-2 and actin specific primers (Supplementary Table 3), on a Step-One-Plus Real-Time PCR thermocycler (Applied Biosystems). Quantification of Ago-2 mRNA expression was calculated relative to actin mRNA using the delta-delta Ct method.

Determination of DVG packaging capacity. Uninfected or infected Vero cells (producer cells, MOI 1 PFU/cell at 24 h) in 24-well plates were transfected with 100 ng of DVG replicon RNA, WT replicon (positive control), or inactive replicon (negative control), using TransIT-mRNA transfection reagent (Mirus Bio). 48 h p.t. the supernatant was endonuclease-treated (25 Units/μl BaseMuncher, Expedeon) and used to infect naive Vero cells in 24-well plates (recipient cells). Replicon activity in recipient cells was assessed 24 h p.i. Donor and recipient cells were lysed in passive lysis buffer (Promega) and luciferase activity measured using the Nano-Glo luciferase assay system (Promega) in a Tecan infinite M200 pro plate reader (Tecan group).

Generation of DVG-containing VLPs. DVG-containing VLPs were generated as previously described¹³. Briefly, HEK-293T cells seeded in a 24 well plate were transfected with a mix of 200 ng of a CPrME-encoding plasmid, 200 ng of E₃₀-NS1, and 200 ng of the DVG-encoding plasmid (LT-1 transfection reagent, Mirus Bio). 72 h p.t., the medium was clarified by centrifugation. Naïve recipient cells were infected with producer cell supernatant following endonuclease (Basemuncher, Expedeon) treatment, and successful packaging determined with luciferase measurement when using reporter DVG clones, or by RT-qPCR when using native DVG.

RT-qPCR for WT or DVG RNA. Detection of WT and DVG RNA was performed by RT-qPCR, using WT and DVG specific primers and probes. For WT RNA, the probe and primers were designed to bind to the deleted region in the DVG. DVG-specific primers were designed such that a 113 nt region containing the deleted part of the genome is amplified by primers binding to the Pr gene and the NS1 gene. Amplification of this region in WT virus RNA would result in a significantly larger product (2783 nt). In addition, the DVG recognizing probe was designed such that it would bind to the breakpoint of the deletion. The primer sequences are shown in Supplementary Table 3. DVG or WT RT-qPCR was carried out using the TaqMan RNA-to-Ct One-step RT-PCR kit (Applied Biosystems) or the Luna Universal One-Step RT-qPCR kit (Abcam) on a Step-One-Plus Real-Time PCR thermocycler (Applied Biosystems). The number of RNA genomes was derived from a standard curve produced in each RT-qPCR run using in vitro generated DVG or WT RNA.

Mouse experiments. AG129 mice⁵⁹ were bred in a specific-pathogen-free facility at Institut Pasteur's animal facility, and C57BL/6 mice were obtained from Charles River Laboratories. Infected adult mice were housed in BSL-3 level isolators and handled in compliance with the Animal Committee regulations and guidelines of Institut Pasteur Paris France, under the 2010/63 European Union Council directive. Animal protocols were approved by the Ethics Committee on Animal Experimentation (CETE) under dossier number dap160116/CHCST 18.176 and the USAMRMC Animal Care and Use Review Office (ACURO), under the protocol number DARPA-5417.04. Mice were monitored daily, with food and water supplied ad libitum. Endpoints were defined and mice humanely euthanized when these were reached.

4-6 week old AG129 or C57BL/6 female mice were given a mix of Ketamine (10 mg/mL)/Xylazine (1 mg/mL) by intraperitoneal injection. Once anesthetized, mice were inoculated by a subcutaneous (footpad) route with vehicle (DMEM media), 10⁴ PFU of Zika virus alone, or complemented with DVG-containing VLPs (TIPs). C57BL/6 mice were treated with 2 mg of an IFNAR1 blocking mouse MAb (MAR-5A3, Euromedex) by intraperitoneal injection one day prior to infection. Weight loss and viremia were monitored at the indicated times after infection. Blood was collected from the facial vein in Microtainer blood collection tubes (BD) and allowed to clot at room temperature. Serum was separated by centrifugation and stored at -80 °C. Viremia was quantified either by plaque assay or RT-qPCR on unextracted and diluted serum samples, as described previously⁶⁰. Infected mice were euthanized by cervical dislocation. Spleen, ovaries, brain, and injected footpads were harvested and homogenized with 600 μl DMEM supplemented with 2% FBS in Precellys tubes containing ceramic beads, using a Precellys 24 homogenizer (Bertin Technologies) at 5000 rpm for 2 cycles of 20 seconds. Homogenates were cleared by centrifugation, and the supernatant stored at -80 °C until virus titration or RT-qPCR. Viral burden in organs was measured either by plaque assay (expressed as PFU/g for organs) or by RT-qPCR (expressed as PFU equivalents relative to GAPDH) following RNA extraction using Direct-zol-96 (Zymo). The number of DVG or WT RNA genomes in mouse organs was normalized to GAPDH genomes, derived from a standard curve using mouse GAPDH primers (Supplementary Table 3) and RNA extracted from the respective organ of uninfected mice.

Mosquito experiments. 6-8 days old female *Ae. aegypti* mosquitoes (1 colony, F7 generation, collected originally in Kamphaeng Phet Province, Thailand) were cold-anesthetized and injected with a transfection mix of CellFectin II reagent (Thermo Fisher) containing Liebovitz's L-15 medium, 0.02 moles of pTRI Xef control or DVG RNA in a total volume of 200 nL. The injection was performed intrathoracically using a nanoinjector (Nanoject III, Drummond Scientific) and a glass capillary needle. 2 days after transfection and following starvation for one day, mosquitoes were allowed to feed for 30 min on an artificial blood meal at 37 °C containing previously washed human blood (2 mL), Zika virus (1 mL, 6 × 10⁶ PFU) and 5 mM ATP. Engorged mosquitoes were counted, selected and maintained at 28 °C in a humidified atmosphere (70%), supplemented with 10% sucrose *ad libitum*. For experiments where mosquitoes were fed with azidothymidine (AZT), mosquitoes were supplemented daily with sucrose-containing 37 mM AZT (Adenosine 5'-triphosphate disodium salt solution A6559, Sigma-Aldrich). 8 days after the blood meal, mosquitoes were cold-anesthetized and dissected. Midguts and carcasses were collected in microtubes containing steel beads (5 mm diameter) and 200 μL DMEM supplemented with 2% FBS. 13 days after the blood meal, mosquitoes were salivated in a tip containing 5 μL FBS, followed by the collection of heads, midgut, and carcass. Body parts were homogenized using a TissueLysor II (QIAGEN) at 30 shakes/second for 2 min. Homogenates were clarified by

centrifugation and virus titer in the supernatant determined by plaque assay. Mosquito saliva was amplified in Vero cells and cytopathic effect assessed 5 days p.i. following fixation and staining with crystal violet.

For detection of DVG or WT RNA from whole mosquitoes or mosquito parts, RNA extracted from homogenates was used for reverse transcription using random primers (Maxima first-strand cDNA synthesis kit, Thermo Fisher). PCRs were then carried out (DreamTaq [Thermo Fisher] or Phusion [NEB]) using the PCR-WT/DVG set of primers (Supplementary Table 3) that detect both WT and DVG RNA, but with different amplicon sizes.

Human blood and ethics statement. Human blood used to feed mosquitoes was obtained from healthy volunteer donors. Healthy donor recruitment was organized by the local investigator assessment using medical history, laboratory results, and clinical examinations. Biological samples were supplied through participation of healthy volunteers at the ICAREB biobanking platform (BB-0033-00062/ICAREB platform/Institut Pasteur, Paris/BBMRI AO203/[BIORESOURCE]) of the Institut Pasteur to the CoSImmGen and Diagnostics protocols, which have been approved by the French Ethical Committee (CPP) Ile-de-France I. The Diagnostics protocol was declared to the French Research Ministry under the reference: DC 2008–68 COL 1.

Statistical analyses. All data were analyzed using Prism 7 software (GraphPad) and are presented as means \pm standard deviation (SD) unless indicated otherwise. For highly skewed data, statistics were carried out using log-transformed data to improve the homoscedasticity of the data. The chosen method of statistical analysis is described in each figure legend. No statistical significance is indicated if p is >0.05 .

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The NGS datasets generated and analyzed during the current study have been deposited in the NCBI Sequence Read Archive under accession number [PRJNA703982](https://doi.org/10.1038/s41467-021-22341-7). Source data are provided with this paper.

Code availability

The code used for the computational determination of neighborhoods enriched with deletions is available at <https://github.com/sharvadim07/DeletionsIslands>⁶¹.

Received: 13 October 2020; Accepted: 12 March 2021;

Published online: 16 April 2021

References

- von Magnus P. Studies on Interference in experimental Influenza. II. Purification and Centrifugation Experiments. *Arkiv for Kemi, Mineralogi och Geologi*. 24. Stockholm, Sweden: Almqvist & Wiksell; 1947.
- Vasilijevic, J. et al. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLoS Pathog.* **13**, e1006650 (2017).
- Sun, Y. et al. Immunostimulatory defective viral genomes from respiratory syncytial virus promote a strong innate antiviral response during infection in mice and humans. *PLoS Pathog.* **11**, e1005122 (2015).
- Karamichali, E. et al. HCV defective genomes promote persistent infection by modulating the viral life cycle. *Front Microbiol* **9**, 2942 (2018).
- Noppompanth, S. et al. Characterization of hepatitis C virus deletion mutants circulating in chronically infected patients. *J. Virol.* **81**, 12496–12503 (2007).
- Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K. & Holmes, E. C. Long-term transmission of defective RNA viruses in humans and Aedes mosquitoes. *Science* **311**, 236–238 (2006).
- Pesko, K. N. et al. Internally deleted WNV genomes isolated from exotic birds in New Mexico: Function in cells, mosquitoes, and mice. *Virology* **427**, 10–17 (2012).
- Hillman, B. I., Carrington, J. C. & Morris, T. J. A defective interfering RNA that contains a mosaic of a plant-virus genome. *Cell* **51**, 427–433 (1987).
- Li, X. H., Heaton, L. A., Morris, T. J. & Simon, A. E. Turnip crinkle virus defective interfering RNAs intensify viral symptoms and are generated de novo. *Proc. Natl Acad. Sci. USA* **86**, 9173–9177 (1989).
- Akey, D. L. et al. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science* **343**, 881–885 (2014).
- Akey, D. L., Brown, W. C., Jose, J., Kuhn, R. J. & Smith, J. L. Structure-guided insights on the role of NS1 in flavivirus infection. *Bioessays* **37**, 489–494 (2015).
- Xie, X. et al. Zika virus replicons for drug discovery. *EBioMed.* **12**, 156–160 (2016).
- Garg, H., Sedano, M., Plata, G., Punke, E. B. & Joshi, A. Development of Virus-Like-Particle Vaccine and Reporter Assay for Zika Virus. *J. Virol.* **91**, e00834–17 (2017).
- Yount, J. S., Gitlin, L., Moran, T. M. & Lopez, C. B. MDA5 participates in the detection of paramyxovirus infection and is essential for the early activation of dendritic cells in response to Sendai virus defective interfering particles. *J. Immunol.* **180**, 4910–4918 (2008).
- Brackney, D. E. et al. C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. *PLoS Negl. Trop. Dis.* **4**, e856 (2010).
- Chotkowski, H. L. et al. West Nile virus infection of Drosophila melanogaster induces a protective RNAi response. *Virology* **377**, 197–206 (2008).
- Vodovar, N. et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. *PLoS One* **7**, e30861 (2012).
- Schnettler, E. et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *J. Gen. Virol.* **94**, 1680–1689 (2013).
- Aliota, M. T. et al. Characterization of Lethal Zika Virus Infection in AG129 Mice. *PLoS Negl. Trop. Dis.* **10**, e0004682 (2016).
- Lazear, H. M. et al. A mouse model of zika virus pathogenesis. *Cell Host Microbe* **19**, 720–730 (2016).
- Poirier, E. Z. et al. Dicer-2-dependent generation of viral DNA from defective genomes of RNA viruses modulates antiviral immunity in insects. *Cell Host Microbe* **23**, 353–65 e8 (2018).
- Goic, B. et al. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. *Nat. Commun.* **7**, 12410 (2016).
- Tassetto, M., Kunitomi, M. & Andino, R. Circulating immune cells mediate a systemic RNAi-based adaptive antiviral response in drosophila. *Cell* **169**, 314–25 e13 (2017).
- Santak, M. et al. Accumulation of defective interfering viral particles in only a few passages in Vero cells attenuates mumps virus neurovirulence. *Microbes Infect.* **17**, 228–236 (2015).
- Dimmock, N. J. & Marriott, A. C. In vivo antiviral activity: defective interfering virus protects better against virulent Influenza A virus than avirulent virus. *J. Gen. Virol.* **87**, 1259–1265 (2006).
- Barrett, A. D. & Dimmock, N. J. Modulation of semliki forest virus-induced infection of mice by defective-interfering virus. *J. Infect. Dis.* **150**, 98–104 (1984).
- Tapia, K. et al. Defective viral genomes arising in vivo provide critical danger signals for the triggering of lung antiviral immunity. *PLoS Pathog.* **9**, e1003703 (2013).
- Dimmock, N. J. & Easton, A. J. Cloned defective interfering influenza RNA and a possible pan-specific treatment of respiratory virus diseases. *Viruses* **7**, 3768–3788 (2015).
- Dimmock, N. J., Rainsford, E. W., Scott, P. D. & Marriott, A. C. Influenza virus protecting RNA: an effective prophylactic and therapeutic antiviral. *J. Virol.* **82**, 8570–8578 (2008).
- Mercado-Lopez, X. et al. Highly immunostimulatory RNA derived from a Sendai virus defective viral genome. *Vaccine* **31**, 5713–5721 (2013).
- Jaworski, E. & Routh, A. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS Pathog.* **13**, e1006365 (2017).
- Routh, A. & Johnson, J. E. Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. *Nucleic Acids Res* **42**, e11 (2014).
- Li, D. et al. Defective interfering viral particles in acute dengue infections. *PLOS One*. **6**, e19447 (2011).
- Lancaster, M. U., Hodgetts, S. I., Mackenzie, J. S. & Urosevic, N. Characterization of defective viral RNA produced during persistent infection of Vero cells with Murray Valley encephalitis virus. *J. Virol.* **72**, 2474–2482 (1998).
- Debnath, N. C., Tiernery, R., Sil, B. K., Wills, M. R. & Barrett, A. D. In vitro homotypic and heterotypic interference by defective interfering particles of West Nile virus. *J. Gen. Virol.* **72**, 2705–2711 (1991).
- Yoon, S. W. et al. Characterization of homologous defective interfering RNA during persistent infection of Vero cells with Japanese encephalitis virus. *Mol. Cells* **21**, 112–120 (2006).
- Tsai, K.-N., Tsang, S.-F., Huang, C.-H. & Chang, R.-Y. Defective interfering RNAs of Japanese encephalitis virus found in mosquito cells and correlation with persistent infection. *Virus Res.* **124**, 139–150 (2007).
- Sempere, R. N. & Arias, A. Establishment of a cell culture model of persistent flaviviral infection: usutu virus shows sustained replication during passages and resistance to extinction by antiviral nucleosides. *Viruses* **11**, 560 (2019).
- Mlakar, J. et al. Zika virus associated with microcephaly. *N. Engl. J. Med.* **374**, 951–958 (2016).
- Cao-Lormeau, V. M. et al. Guillain-Barre syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* **387**, 1531–1539 (2016).

41. Musso, D. et al. Potential sexual transmission of Zika virus. *Emerg. Infect. Dis.* **21**, 359–361 (2015).
42. Yockey, L. J. et al. Vaginal exposure to zika virus during pregnancy leads to fetal brain infection. *Cell* **166**, 1247–56 e4 (2016).
43. Vermillion, M. S. et al. Intrauterine Zika virus infection of pregnant immunocompetent mice models transplacental transmission and adverse perinatal outcomes. *Nat. Commun.* **8**, 14575 (2017).
44. Shi, Y. et al. Vertical transmission of the zika virus causes neurological disorders in mouse offspring. *Sci. Rep.* **8**, 3541 (2018).
45. Miner, J. J. et al. Zika virus infection during pregnancy in mice causes placental damage and fetal demise. *Cell* **165**, 1081–1091 (2016).
46. Strahle, L., Garcin, D. & Kolakofsky, D. Sendai virus defective-interfering genomes and the activation of interferon-beta. *Virology* **351**, 101–111 (2006).
47. Shivakoti, R., Siwek, M., Hauer, D., Schultz, K. L. & Griffin, D. E. Induction of dendritic cell production of type I and type III interferons by wild-type and vaccine strains of measles virus: role of defective interfering RNAs. *J. Virol.* **87**, 7816–7827 (2013).
48. Fuller, F. J. & Marcus, P. I. Interferon induction by viruses. IV. Sindbis virus: early passage defective-interfering particles induce interferon. *J. Gen. Virol.* **48**, 63–73 (1980).
49. Marcus, P. I. & Sekellick, M. J. Defective interfering particles with covalently linked [+/–]RNA induce interferon. *Nature* **266**, 815–819 (1977).
50. van den Hoogen, B. G. et al. Excessive production and extreme editing of human metapneumovirus defective interfering RNA is associated with type I IFN induction. *J. Gen. Virol.* **95**, 1625–1633 (2014).
51. Kistler, K. E., Voss hall, L. B. & Matthews, B. J. Genome engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*. *Cell Rep.* **11**, 51–60 (2015).
52. Li, M. et al. Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA.* **114**, E10540–E10549 (2017).
53. Chaverra-Rodriguez, D. et al. Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. *Nat. Commun.* **9**, 3008 (2018).
54. Schwarz, M. C. et al. Rescue of the 1947 zika virus prototype strain with a cytomegalovirus promoter-driven cDNA clone. *mSphere* **1**, e00246-16 (2016).
55. Bredenbeek, P. J. et al. A stable full-length yellow fever virus cDNA clone and the role of conserved RNA elements in flavivirus replication. *J. Gen. Virol.* **84**, 1261–1268 (2003).
56. Alsaleh, K. et al. The E glycoprotein plays an essential role in the high pathogenicity of European-Mediterranean IS98 strain of West Nile virus. *Virology* **492**, 53–65 (2016).
57. Khromykh, A. A., Meka, H., Guyatt, K. J. & Westaway, E. G. Essential role of cyclization sequences in flavivirus RNA replication. *J. Virol.* **75**, 6719–6728 (2001).
58. Atieh, T. et al. New reverse genetics and transfection methods to rescue arboviruses in mosquito cells. *Sci. Rep.* **7**, 13983 (2017).
59. van den Broek, M. F., Muller, U., Huang, S., Aguet, M. & Zinkernagel, R. M. Antiviral defense in mice lacking both alpha/beta and gamma interferon receptors. *J. Virol.* **69**, 4792–4796 (1995).
60. Pastorino, B. et al. Development of a TaqMan RT-PCR assay without RNA extraction step for the detection and quantification of African Chikungunya viruses. *J. Virol. Methods* **124**, 65–71 (2005).
61. Titievski A., Sharov V., Brodsky L. High fitness deletion hotspots in defective viral genomes enable the development of therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. *GitHub* <https://doi.org/10.5281/zenodo.4553175>, 2021.

Acknowledgements

The authors thank Cassandra Koh, Alessandro Torri, and Margus Varjak for helpful discussions. We thank Germinate Science Consulting for professional editorial services.

Illustrative figures in this manuscript were created with BioRender.com. This work was funded by the DARPA INTERCEPT program managed by Dr. Jim Gimlett, Dr. Brad Ringesen, and Dr. Seth Cohen and administered through DARPA Cooperative Agreement #HR0011-17-2-0023 (the content of the information does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred). This work also received funding from the Fondation de recherche médicale, FRM EQU201903007777, and the Agence Nationale de Recherche Laboratoire d'Excellence grant ANR-10-LABX-62-IBEID.

Author contributions

The study was conceived by V.V.R., M.C.S., and M.V. In vivo Zika virus experiments were carried out and processed by V.V.R., L.C., F.M., and L.L., with support from A.H.L., V.G., and Y.S. In vitro experiments Zika virus were performed and analysed by V.V.R. with support from L.C., L.L., D.T., D.S., B.M., T.V., and J.W.-L. Work with other flaviviruses was performed by S.P., M.A.D.S., and V.L. under the supervision of N.P. and G.B. S. Library preparation was carried out by V.R. and T.V. Initial bioinformatics analysis and interpretation was done by D.E., V.B. with support from Q.D.T. Further bioinformatics analysis and computational determination of deletion-enriched neighborhoods were conceived and performed by A.T. and V.S. under the supervision of L.B. The manuscript was written by V.V.R., L.B., M.C.S., and M.V. The work was supervised by N.P., G.B.S., L.B., M.C.S., and M.V.

Competing interests

The Institut Pasteur has filed a U.S. Provisional Application No. 63/000,998, Filed: March 27, 2020, entitled DEFECTIVE INTERFERING VIRAL GENOMES, with V.V.R., L.L., D.S., B.M., T.V., V.B., N.P., G.B.S., M.C.S. and M.V. listed as inventors. The patent application covers the experimental and computational methodology for identifying the best candidate TIPs for a given virus, including the flavivirus TIPs described in this paper. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22341-7>.

Correspondence and requests for materials should be addressed to M.-C.S. or M.V.

Peer review information *Nature Communications* thanks Alistair Russell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

2. BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2

Avi J. Samelson, **Quang Dinh Tran**, Rémy Robinot, Lucia Carrau, Veronica V. Rezelj, Alice Mac Kain, Merissa Chen, Gokul N. Ramadoss, Xiaoyan Guo, Shion A. Lim, Irene Lui, James K. Nuñez, Sarah J. Rockwood, Jianhui Wang, Na Liu, Jared Carlson-Stevermer, Jennifer Oki, Travis Maures, Kevin Holden, Jonathan S. Weissman, James A. Wells, Bruce R. Conklin, Benjamin R. TenOever, Lisa A. Chakrabarti, Marco Vignuzzi, Ruilin Tian & Martin Kampmann

DOI: <https://doi.org/10.1038/s41556-021-00821-8>

Contributions: I have contributed to SARS-CoV-2 *in vitro* infection experiments in Figure 2c-d,f and 3b-f and Extended Data Figure 4.



BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2

Avi J. Samelson^{1,2}, Quang Dinh Tran^{3,4}, Rémy Robinot⁵, Lucia Carrau⁶, Veronica V. Rezelj³, Alice Mac Kain^{3,4}, Merissa Chen^{1,2}, Gokul N. Ramadoss^{7,8}, Xiaoyan Guo^{1,2}, Shion A. Lim^{9,19}, Irene Lui⁹, James K. Nuñez^{10,11}, Sarah J. Rockwood⁷, Jianhui Wang¹², Na Liu¹², Jared Carlson-Stevermer¹³, Jennifer Okí¹³, Travis Maures¹³, Kevin Holden¹³, Jonathan S. Weissman^{10,11,14,15}, James A. Wells^{12,9,10}, Bruce R. Conklin^{7,10,16,17}, Benjamin R. TenOever¹⁶, Lisa A. Chakrabarti¹⁵, Marco Vignuzzi¹³, Ruilin Tian^{1,2,12}✉ and Martin Kampmann^{1,2,18}✉

SARS-CoV-2 infection of human cells is initiated by the binding of the viral Spike protein to its cell-surface receptor ACE2. We conducted a targeted CRISPRi screen to uncover druggable pathways controlling Spike protein binding to human cells. Here we show that the protein BRD2 is required for ACE2 transcription in human lung epithelial cells and cardiomyocytes, and BRD2 inhibitors currently evaluated in clinical trials potently block endogenous ACE2 expression and SARS-CoV-2 infection of human cells, including those of human nasal epithelia. Moreover, pharmacological BRD2 inhibition with the drug ABBV-744 inhibited SARS-CoV-2 replication in Syrian hamsters. We also found that BRD2 controls transcription of several other genes induced upon SARS-CoV-2 infection, including the interferon response, which in turn regulates the antiviral response. Together, our results pinpoint BRD2 as a potent and essential regulator of the host response to SARS-CoV-2 infection and highlight the potential of BRD2 as a therapeutic target for COVID-19.

The ongoing COVID-19 pandemic is a public health emergency. As of September 2021, SARS-CoV-2, the novel coronavirus causing this disease, has infected over 200 million people worldwide, causing at least four and a half million deaths (<https://covid19.who.int>). New infections are still rapidly increasing, despite current vaccination programmes. The emergence of novel viral variants with the potential to partially overcome vaccine-elicited immunity highlights the need to elucidate the molecular mechanisms that underlie SARS-CoV-2 interactions with host cells to enable the development of therapeutics to treat and prevent COVID-19, complementing ongoing vaccination efforts.

SARS-CoV-2 entry into human cells is initiated by the interaction of the viral Spike protein with its receptor on the cell surface, angiotensin-converting enzyme 2 (ACE2). To uncover new therapeutic targets targeting this step of SARS-CoV-2 infection, we have conducted a focused CRISPR interference (CRISPRi)-based screen for modifiers of Spike binding to human cells. We expected that ACE2 and factors regulating ACE2 expression would be major hit genes in

this screen. A second motivation for identifying regulators of ACE2 was the fact that ACE2 affects inflammatory responses and is itself regulated in the context of inflammation^{1–3}. Inflammatory signalling, in particular the type I interferon response, is known to be misregulated in the most severely affected patients with COVID-19^{4–7}. Therefore, regulators of ACE2 expression would probably be relevant for COVID-19 in human patients, as suggested by clinical data⁸.

Previous CRISPR screens have been performed in cell-based models of SARS-CoV-2 infection that overexpressed an ACE2 transgene^{9,10}, represented cell types not primarily targeted by SARS-CoV-2¹¹ or were non-human cells¹². Although these studies elucidated major features of SARS-CoV-2 biology, we reasoned that the cell lines used would not have enabled the discovery of regulators of ACE2 expression in relevant human cell types.

We selected a lung epithelial cancer cell line, Calu-3, which endogenously expresses ACE2, to perform a targeted CRISPRi screen to find regulators of Spike protein binding. We found that the strongest hit genes are potent regulators of ACE2 levels. Knockdown

¹Institute for Neurodegenerative Diseases, University of California San Francisco, San Francisco, CA, USA. ²Chan-Zuckerberg Biohub, San Francisco, CA, USA. ³Viral Populations and Pathogenesis Unit, Institut Pasteur, Paris, France. ⁴École Doctorale BioSPC, Université de Paris, Sorbonne Paris Cité, Paris, France. ⁵Institut Pasteur, CIVIC Group, Virus and Immunity Unit, Université de Paris, Paris, France. ⁶Microbiology Department, NYU-Langone, New York, NY, USA. ⁷Gladstone Institutes, San Francisco, CA, USA. ⁸Biomedical Sciences PhD Program, University of California San Francisco, San Francisco, CA, USA. ⁹Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA. ¹⁰Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA. ¹¹Howard Hughes Medical Institute, University of California San Francisco, San Francisco, CA, USA. ¹²School of Medicine, Southern University of Science and Technology, Shenzhen, China. ¹³Synthego Corporation, Redwood City, CA, USA. ¹⁴Whitehead Institute for Biomedical Research, Cambridge, MA, USA. ¹⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, USA. ¹⁶Department of Ophthalmology, University of California San Francisco, San Francisco, CA, USA. ¹⁷Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ¹⁸Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA. ¹⁹Present address: Department of Antibody Engineering, Genentech Inc., San Francisco, CA, USA. ✉e-mail: tianrl@sustech.edu.cn; martin.kampmann@ucsf.edu

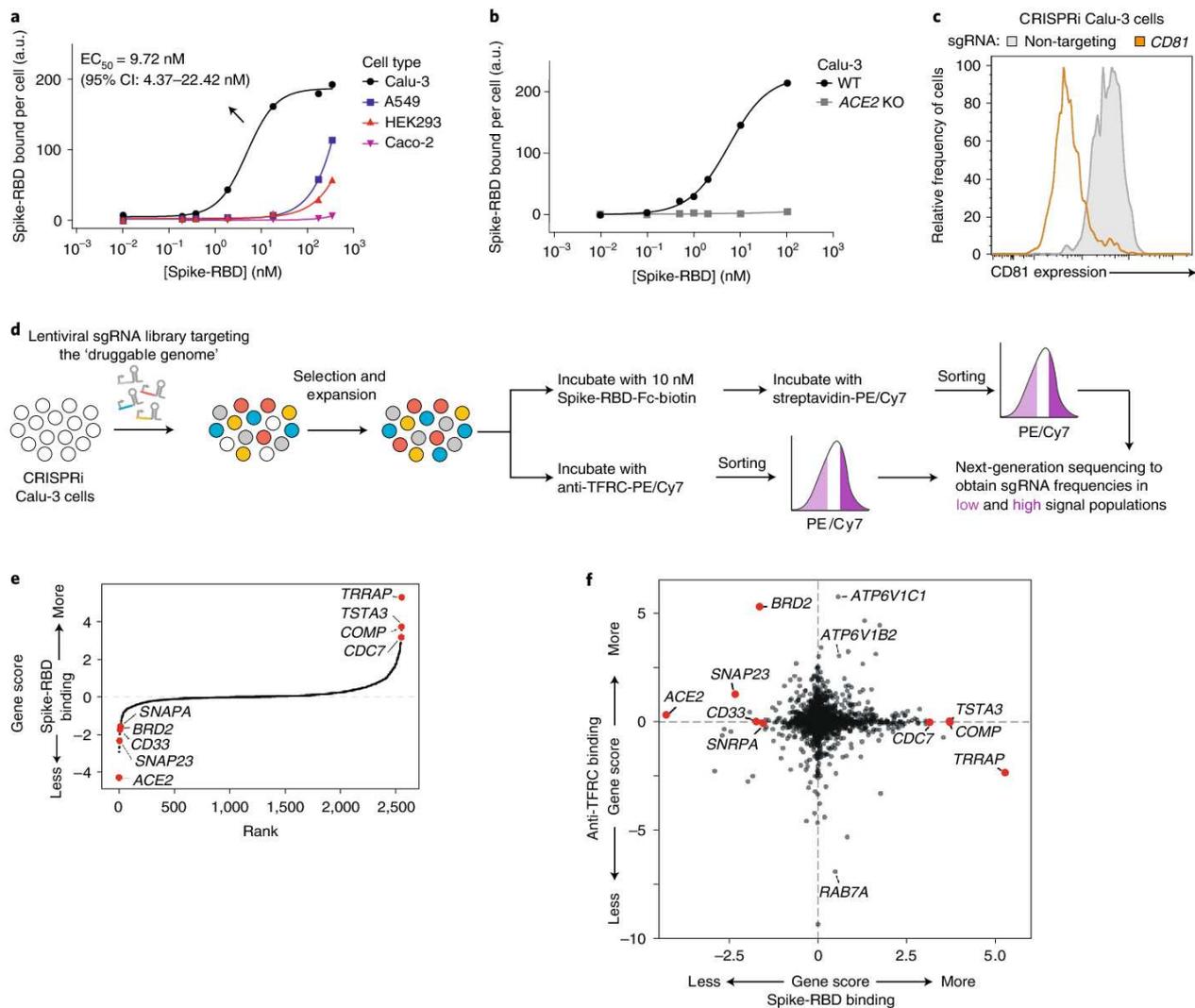


Fig. 1 | CRISPRi screen reveals cellular factors controlling Spike protein binding. **a**, Human cell lines were incubated with different concentrations of recombinant SARS-CoV-2 Spike-RBD, and the amount of Spike-RBD bound per cell was quantified by flow cytometry. Only Calu-3 cells (black) display saturable binding, for which an EC_{50} value was fit. The error of EC_{50} is the 95% confidence interval. **b**, Spike-RBD binding is eliminated in Calu-3 cells with ACE2 knockout (KO, grey). **c**, Validation of CRISPRi activity of a Calu-3 CRISPRi line. Calu-3 cells stably expressing CRISPRi machinery were transduced with an sgRNA targeting CD81 (orange) or a non-targeting sgRNA (grey), and CD81 levels were determined by flow cytometry. The experiments in **a–c** were performed once. **d**, CRISPRi screen strategy. CRISPRi Calu-3 cells transduced with an sgRNA library targeting the 'druggable genome' were stained either with Spike-RBD or an anti-TFRC antibody. Cells were then FACS-sorted into bins (top and bottom 30%) based on Spike-RBD or anti-TFRC binding signal, and frequencies of cells expressing each sgRNA were determined for each bin by targeted next-generation sequencing. The CRISPRi screen was performed once. **e**, Rank-order plot of Spike-RBD hit genes. Genes selected for follow-up experiments are highlighted as red dots. **f**, Scatter plot of gene scores for the Spike-RBD screen (x axis) and the anti-TFRC screen (y axis). Genes selected for follow-up experiments are highlighted as red dots.

of these genes reduced or increased ACE2 levels transcriptionally, and prevented or enhanced, respectively, SARS-CoV-2 infection in cell culture.

We identified the transcriptional regulator bromodomain-containing protein 2 (BRD2) as a major node for host–SARS-CoV-2 interaction. BRD2 is part of the bromodomain and extra-terminal domain (BET) family of proteins that includes Brd3, Brd4 and BrdT. BETs are being explored as targets for a number of cancers¹³. These proteins are known to be master transcriptional regulators and serve to bridge chromatin marks (mostly acetyl-lysines) to the transcriptional machinery¹⁴. We found BRD2 inhibition to

downregulate ACE2 expression in Calu-3 cells, induced pluripotent stem cell (iPSC)-derived cardiomyocytes, primary human lung epithelial cells and reconstructed human nasal epithelia. Inhibition of BRD2 with small molecules, some of which are in phase I clinical trials, inhibited SARS-CoV-2 infection in primary human nasal epithelia and in Syrian hamsters. We propose BRD2 as a key regulator and potential therapeutic target for COVID-19.

Results

CRISPRi screen for Spike-RBD binding to cells. To identify cellular mechanisms controlling the binding of SARS-CoV-2 to human

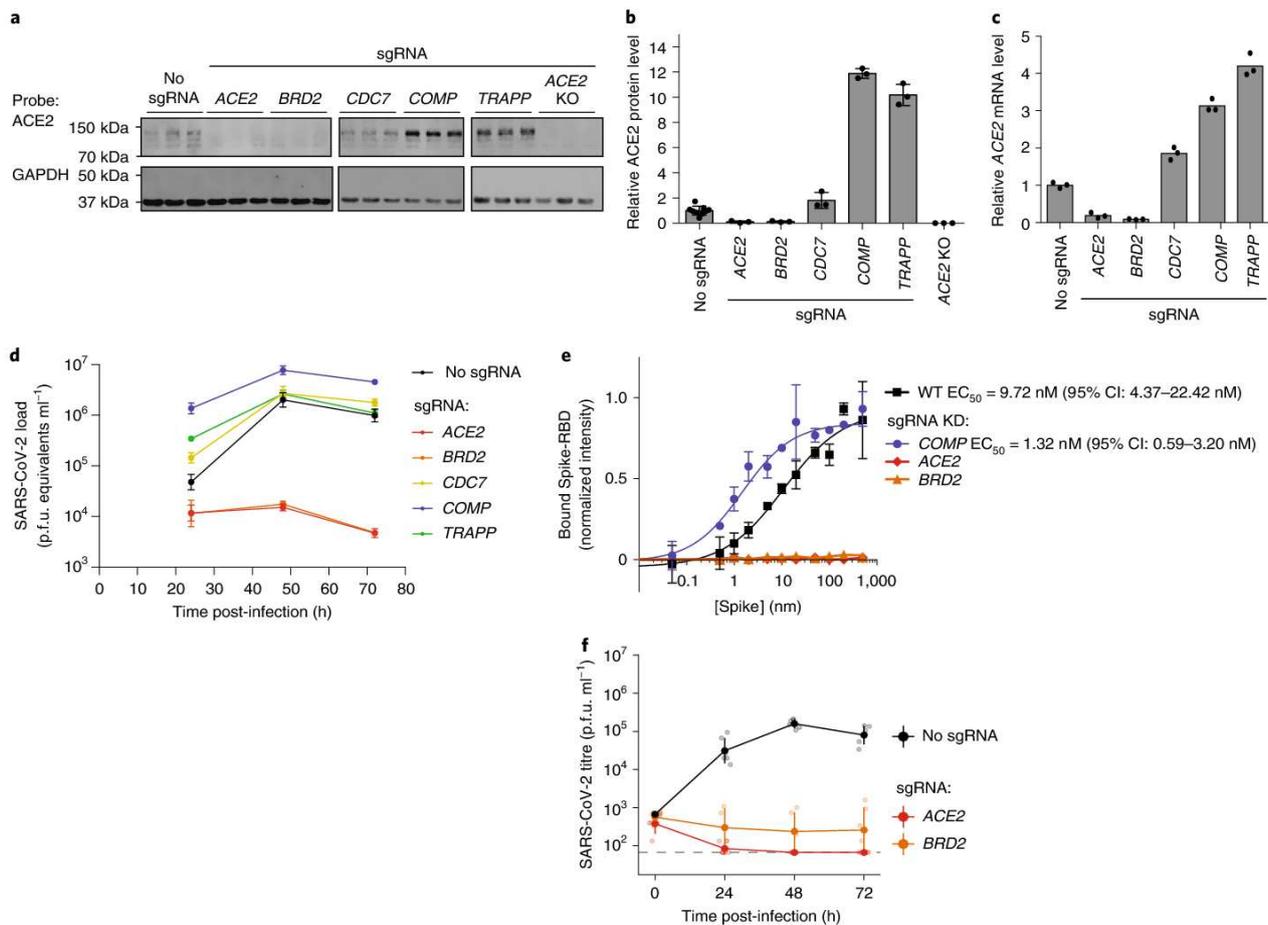


Fig. 2 | Hit genes modulate ACE2 levels and SARS-CoV-2 infection. **a**, Western blotting for ACE2 and GAPDH in Calu-3 CRISPRi cells expressing no sgRNA or sgRNAs targeting different hit genes, or ACE2 knockout Calu-3 cells. The three lanes represent biological triplicates for each cell line. **b**, Quantification of ACE2 protein levels relative to GAPDH based on the data in **a**. Average normalized ACE2 intensity as a fraction of no sgRNA and the standard deviation for three biological replicates are shown. **c**, Relative amounts of ACE2 transcript levels measured by qPCR in Calu-3 CRISPRi cells expressing sgRNAs targeting different hit genes, compared to cells without sgRNA. Average relative gene expression of three technical replicates is shown from a single experiment. **d**, Calu-3 CRISPRi cells expressing different sgRNAs targeting hit genes were infected with SARS-CoV-2 and viral RNA in the supernatant was measured by RT-qPCR as a function of time post-infection. The average SARS-CoV-2 load and the standard deviation of three biological replicates are shown. **e**, Spike-RBD binding to Calu-3 cells was quantified by flow cytometry of Calu-3 cells expressing sgRNAs targeting individual hit genes after incubation with increasing concentrations of Spike-RBD. For genes for which data could be fitted with a binding curve, the EC₅₀ was determined along with the 95% confidence intervals. Data points are average values from three biological replicates for each gene knockdown, with error bars representing the standard deviation, except for ACE2 and BRD2, where only one experiment at each concentration was performed. **f**, Plaque assays in Calu-3 CRISPRi cells expressing different sgRNAs targeting hit genes were infected with SARS-CoV-2 as a function of time post-infection. Average SARS-CoV-2 titre and the standard deviation of six biological replicates are shown.

cells, we identified a cell line that would robustly bind the viral Spike protein (S). We measured binding of a previously described recombinant protein construct encompassing the SARS-CoV-2 Spike protein receptor-binding domain (RBD) with a C-terminal human immunoglobulin-G (IgG) Fc-domain fusion¹⁵, referred to hereafter as Spike-RBD, to several commonly used human cell lines (Fig. 1a and Extended Data Fig. 1a). Within this cell line panel, only Calu-3 cells displayed a binding curve consistent with specific binding of Spike-RBD, with a half-maximum effective concentration (EC₅₀) of 9.72 nM (95% CI: 4.37–22.42 nM). This value agrees with the dissociation constant of Spike-RBD–ACE2 binding determined *in vitro* (4.7 nM)¹⁶, within measurement error. Spike-RBD binding is dependent on ACE2 expression, as binding is abrogated in Calu-3 ACE2 knockout cells (Fig. 1b). Calu-3 cells are challenging to culture

compared to most cell lines—they proliferate slowly, and their adherence properties pose challenges for flow cytometry. Nevertheless, Calu-3 cells are a particularly attractive cell culture model for studying SARS-CoV-2 from the biological point of view, because they are derived from lung epithelia, which is selectively infected by SARS-CoV-2 in patients¹⁷, are known to support infection of SARS-CoV and SARS-CoV-2¹⁸ and have recently been reported to closely recapitulate gene-expression changes that occur in patients¹⁹.

We generated a polyclonal Calu-3 line constitutively expressing machinery to enable CRISPRi-based genetic screens^{20,21} and validated its CRISPRi activity (Fig. 1c and Extended Data Fig. 1b). CRISPRi uses a catalytically dead Cas9 (dCas9) fused to a transcriptional repressor, KRAB, to knock down genes at specific sites programmed by the loading of the dCas9-KRAB with a single-guide

RNA (sgRNA). Using this line, we then performed a focused CRISPRi screen for factors controlling Spike-RBD binding (Fig. 1d). To maximize our chances of identifying potential undiscovered therapeutic targets for COVID-19, we screened a sgRNA library targeting the ‘druggable genome’, comprising ~2,300 genes, with ~16,000 total sgRNAs including non-targeting control sgRNAs²². In parallel, we screened the same library using a fluorophore-conjugated antibody against the transferrin receptor (TFRC, also known as CD71) to control for factors that generally affect protein trafficking or protein binding to the cell surface (Fig. 1d–f). Owing to the limitations of the Spike-RBD binding assay and Calu-3 cells, this screen was conducted at a lower average representation (~100 sorted cells per sgRNA) than ideal, resulting in relatively high noise and therefore fewer hits crossing a false discovery rate cutoff of 0.1 than in typical CRISPR screens (Extended Data Fig. 1c and Methods).

Despite the increased noise, *ACE2*, as expected, was the strongest hit gene, knockdown of which decreased binding of Spike-RBD, while having no effect on TFRC levels (Fig. 1e,f). Conversely, *RAB7A*, which was recently reported to be essential for the trafficking of TFRC to the cell surface²³, was the strongest hit that decreased TFRC levels, with no effect on Spike-RBD binding (Fig. 1f). Generally, hits were not correlated between the two screens (Fig. 1f), demonstrating the specificity of each screen. Although the screens did not result in a large number of strong hits (Supplementary Table 1), we decided to validate the top 15 genes for which knockdown decreased Spike-RBD binding and the top five genes for which knockdown increased Spike-RBD binding. We cloned individual sgRNAs targeting each of these genes and evaluated their effect on Spike-RBD binding (Extended Data Fig. 2). Based on these experiments, we selected hits that robustly recapitulated their phenotypes from the primary screen for further characterization: two genes for which knockdown decreased Spike-RBD binding (*ACE2* and *BRD2*) and three genes for which knockdown increased Spike-RBD-binding (*CDC7*, *COMP* and *TRRAP*).

Hit genes modulate ACE2 levels and infection with SARS-CoV-2.

Because Spike-RBD binding is dependent on ACE2 expression, we hypothesized that other hit genes might act by modulating ACE2 levels. Western blots for ACE2 levels in Calu-3 cell lines expressing sgRNAs against validated target genes (hereafter referred to as knockdown lines) indeed revealed marked changes in ACE2 protein levels. For hits associated with lower levels of Spike-RBD binding in the primary screen, we observed lower levels of ACE2 protein, and vice versa for those hits associated with higher levels of Spike-RBD binding (Fig. 2a,b). To distinguish whether hit genes affected ACE2 protein levels via transcriptional or post-transcriptional mechanisms, we performed quantitative polymerase chain reaction (qPCR) to measure *ACE2* transcript levels in these same

knockdown lines. For all tested genes, we observed changes in *ACE2* transcript levels that were concordant with the changes in ACE2 protein levels (Fig. 2c), indicating that they acted at the transcriptional level. Some genes, such as *COMP* and *TRAPP*, showed relatively modest effects on *ACE2* transcript levels, but quite large effects on ACE2 protein levels, suggesting that these hit genes additionally affect post-transcriptional regulation of ACE2 expression.

We next determined the effect of hit gene knockdown on susceptibility to SARS-CoV-2 infection. We infected cells expressing sgRNAs against hit genes with SARS-CoV-2 and measured virus replication 24, 48 and 72 h post-infection using qPCR with reverse transcription (RT-qPCR; Fig. 2d). Already at 24 h post-infection, viral genome copies diverged concordantly with changes in ACE2 levels and Spike-RBD binding: sgRNAs that lowered Spike-RBD binding reduced virus replication, and sgRNAs that increased Spike-RBD binding resulted in higher virus replication. *BRD2* knockdown abrogated viral replication in these cells to levels similar to *ACE2* knockdown, even at 72 h post-infection, and *COMP* knockdown supported an order of magnitude increase in viral titres. Focusing on these three hit genes—*ACE2*, *BRD2* and *COMP*—we then quantified how gene knockdown modulates Spike-RBD binding to cells (Fig. 2e). Knockdown of *ACE2* and *BRD2* abolished spike binding, while *COMP* decreased the EC₅₀ compared to wild type (WT) from 9.72 nM (95% CI: 4.37–22.42 nM) to 1.32 nM (95% CI: 0.59–3.20 nM), by almost a full order of magnitude (Fig. 2e). We also confirmed that gene knockdown decreased SARS-CoV-2 replication by performing plaque assays on WT, *BRD2* KD and *ACE2* knockdown cells (Fig. 2f).

BRD2 inhibitors prevent SARS-CoV-2 infection of human cells.

Given the stringent inhibition of SARS-CoV-2 infection achieved by *BRD2* knockdown, and the fact that BRD2 is currently being evaluated as a therapeutic target in cancer^{13,24}, with several small-molecule inhibitors in clinical trials²⁵, we decided to focus on this hit gene.

We validated that CRISPRi knockdown of *BRD2* robustly reduced BRD2 protein levels (Extended Data Fig. 3). Transgenic expression of full-length BRD2 restored *ACE2* transcript levels (Fig. 3a), validating that the reduction in ACE2 expression triggered by CRISPRi targeting of *BRD2* was indeed due to *BRD2* knockdown. Transgenic expression of truncation mutants of *BRD2* did not rescue ACE2 expression (Fig. 3a), indicating that full-length BRD2 is required for ACE2 expression.

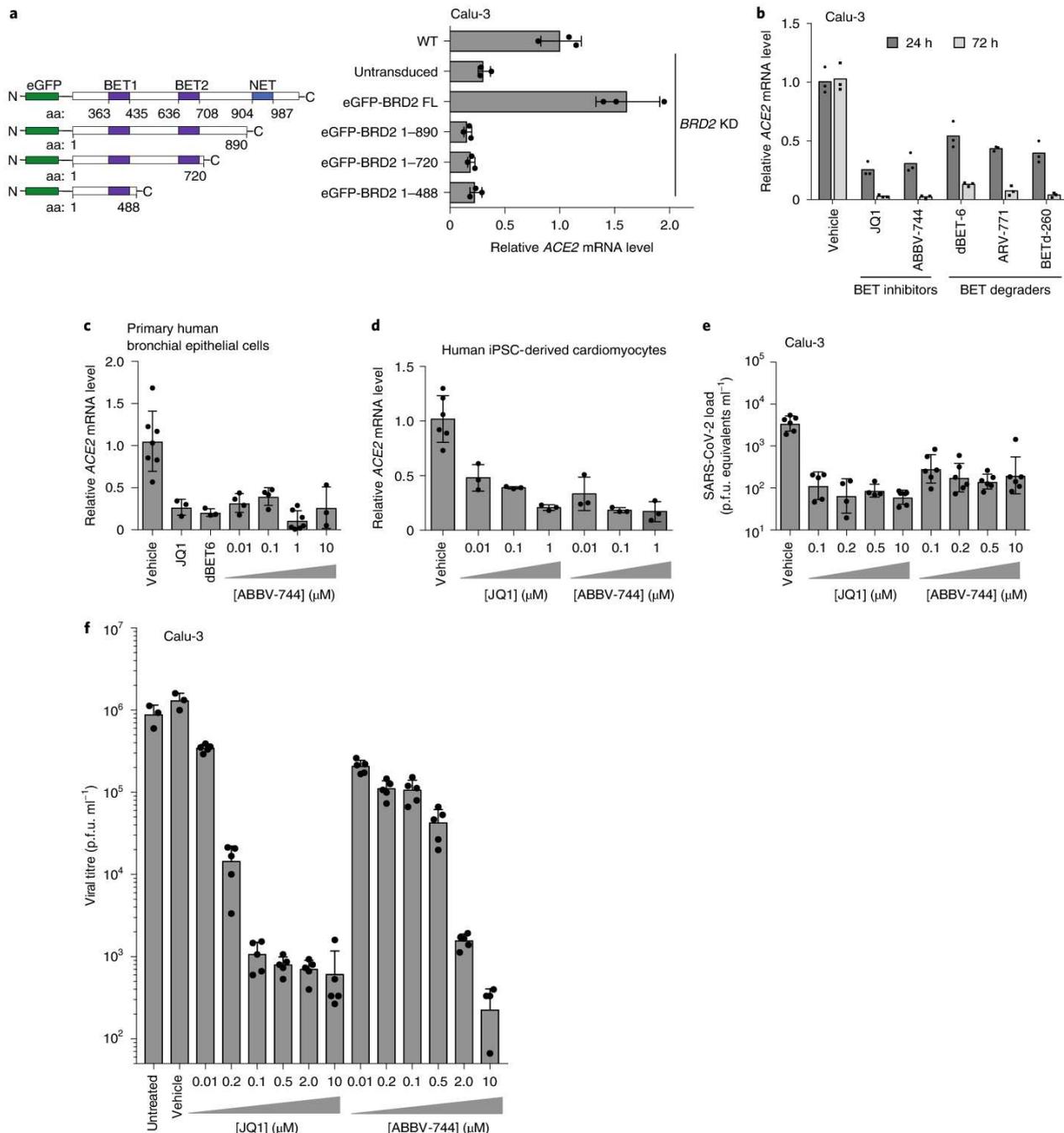
To test the potential of BRD2 as a therapeutic target for COVID-19, we treated cells with a panel of compounds targeting BRD2: two BET-domain inhibitors (JQ1²⁶ and ABBV-744²⁷; the latter is currently in clinical trials NCT03360006 and NCT04454658) and three proteolysis targeting chimaeric (PROTAC) compounds that lead to the degradation of BRD2 (dBET-6²⁸, ARV-771²⁹ and BETd-260²⁹).

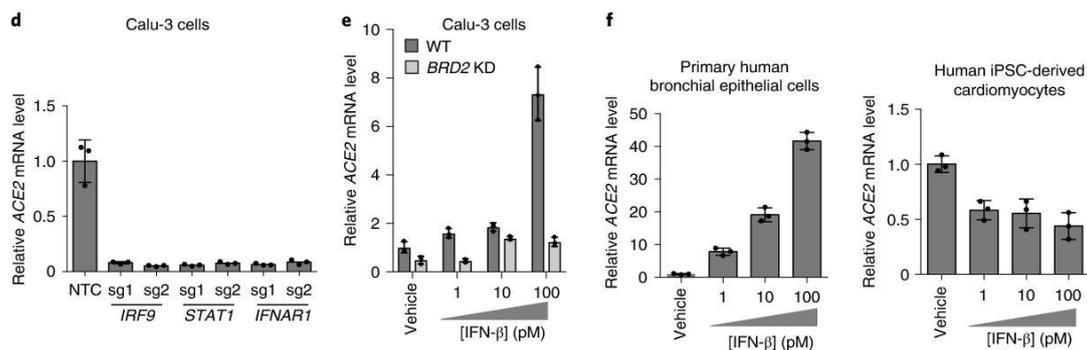
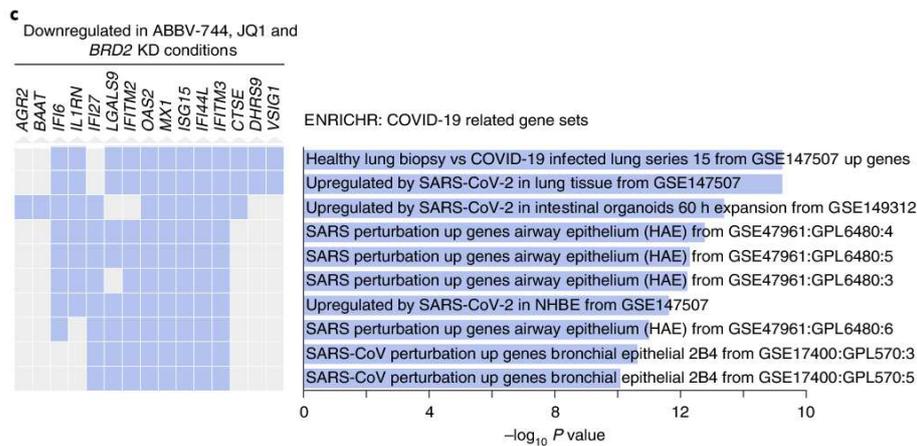
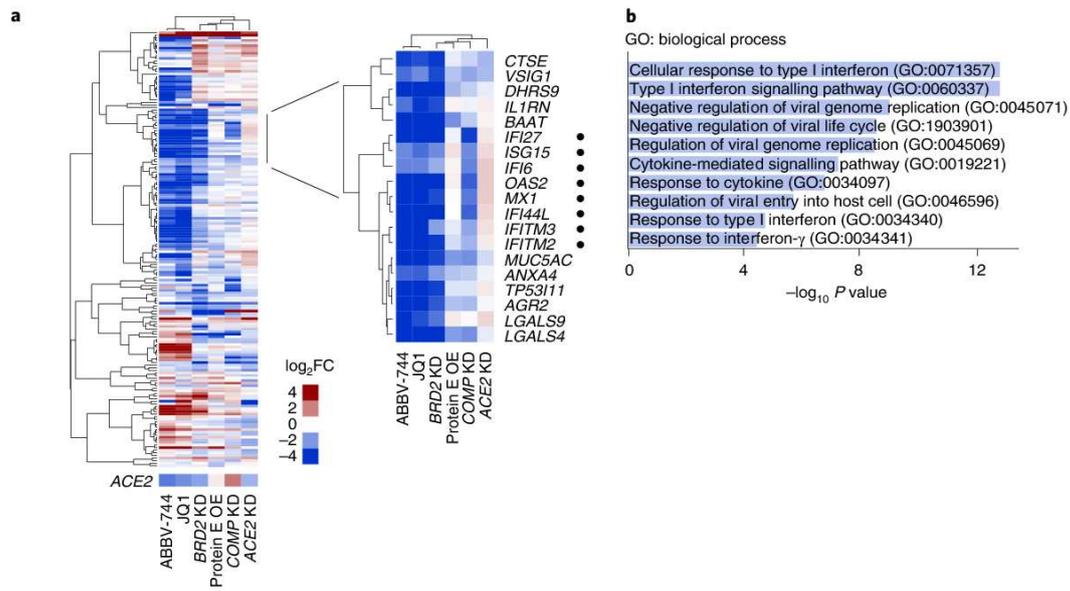
Fig. 3 | BRD2 inhibitors potently reduce ACE2 levels and SARS-CoV-2 infection. **a**, Transgenic constructs expressed in Calu-3 cells (left), and transcript levels of *ACE2* relative to *ACTB* in Calu-3 cells transduced with enhanced green fluorescent protein–BRD2 truncations in a *BRD2* knockdown background (right). Average relative *ACE2* gene expression compared to non-transduced cells and the standard deviation of $n=3$ biological replicates are shown. aa, amino acid. **b**, Transcript levels of *ACE2* relative to *ACTB* in Calu-3 cells treated with BRD2 inhibitors (JQ1 at 10 μ M, ABBV-744 at 10 μ M and dBET-6 at 200 nM) were quantified at 24 h (dark grey bars) and 72 h (light grey bars) post-treatment. Average *ACE2* mRNA levels relative to vehicle-treated of three technical replicates are shown from a single experiment. **c**, Transcript levels of *ACE2* relative to *ACTB* in primary human bronchial epithelial cells treated (NHBE) with BRD2 inhibitors (JQ1 at 10 μ M, dBET-6 at 20 nM, ABBV-744 at 0.01–10 μ M) were quantified at 72 h post-treatment. Average *ACE2* mRNA levels relative to vehicle and the standard deviation of $n=3$ biological replicates are shown, except for vehicle-treated, for which $n=6$ replicates are shown. **d**, Transcript levels of *ACE2* relative to 18S ribosomal RNA (rRNA) in human iPSC-derived cardiomyocytes treated with the indicated concentrations of BRD2 inhibitors were quantified at 72 h post-treatment. Average *ACE2* mRNA levels relative to vehicle-treated and the standard deviation of $n=3$ biological replicates for each condition are shown, except for the vehicle-treated sample, for which $n=6$ biological replicates are shown. **e**, SARS-CoV-2 viral RNA in supernatant measured by RT-qPCR at 24 h post-infection of Calu-3 cells. Infection took place 72 h after treatment with the indicated concentrations of BRD2 inhibitors. Measurements are the average of $n=5$ biological replicates. Error bars show the standard deviation. **f**, Plaque assays in Calu-3 CRISPRi cells treated with increasing concentrations of the BET inhibitors JQ1 or ABBV-744 infected with SARS-CoV-2 as a function of time post-infection. Average viral titre and the standard deviation of six biological replicates are shown, except for vehicle and untreated, for which three biological replicates are shown.

After only 24 h of treatment with these drugs, *ACE2* messenger RNA (mRNA) levels measured by qPCR decreased roughly twofold (Fig. 3b). This effect was magnified after treatment for 72 h, when almost no *ACE2* mRNA was detectable for any of the BRD2-targeting compounds tested, phenocopying *BRD2* knockdown (Fig. 3b). Similarly, we found that BET inhibitors led to substantial decreases in *ACE2* mRNA levels in primary human bronchial epithelial cells (Fig. 3c) and human iPSC-derived cardiomyocytes (Fig. 3d)—two non-transformed cell types that are susceptible to SARS-CoV-2 infection^{30,31}. Importantly, BET inhibitors were non-toxic to Calu-3

cells, primary human bronchial epithelial cells and cardiomyocytes at effective concentrations (Extended Data Fig. 4).

Because pharmacological inhibition of BRD2 phenocopied *BRD2* knockdown, we hypothesized that these same compounds might prevent infection of cells exposed to SARS-CoV-2. To test this, we treated Calu-3 cells for 72 h with the BET inhibitors JQ1 and ABBV-744, and measured SARS-CoV-2 replication at 48 h post-infection. Strikingly, we found that treated cells displayed 100-fold decreased viral replication versus untreated cells (Fig. 3e), a similar effect size to *BRD2* or *ACE2* knockdown (Fig. 2d,e).





BRD2 regulates ACE2 and SARS-CoV-2-induced host genes. We next asked whether BRD2 controls transcription of additional genes beyond ACE2. We performed RNA-sequencing (RNA-seq) of Calu-3 cells after treatment with the BET-domain inhibitors JQ1 and ABBV-744 as well as BRD2 CRISPRi knockdown (Supplementary Table 2). We also included CRISPRi knockdown of two other validated hit genes from our screen, COMP and ACE2, as well as over-expression of the viral protein E, which has been reported to interact

with BRD2³². RNA-seq of BRD2 knockdown and BET-domain inhibitor treated cells recapitulated downregulation of ACE2 (Fig. 4a). *TMPRSS2*, the gene encoding a protease important for viral entry in many cell types, was not a differentially expressed gene in any condition (Supplementary Table 2). Surprisingly, BRD2 knockdown or pharmacological inhibition also resulted in marked downregulation of genes involved in the type I interferon response, while ACE2 knockdown slightly increased expression of those same genes

Fig. 4 | BRD2 controls genes induced by interferon and SARS-CoV-2 infection. **a**, Differentially expressed genes (DEGs) from RNA-sequencing of Calu-3 cells under different treatment conditions compared to control cells: 72-h treatment with 10 μ M JQ1 or 10 μ M ABBV-744, *BRD2* knockdown (KD), SARS-CoV-2 protein E overexpression (OE), *COMP* KD and *ACE2* KD. The quasi-likelihood method glmQLFTest was implemented in edgeR. *P* values were adjusted for multiple comparisons using the Benjamini-Hochberg method. The heatmap shows the log₂-fold change (log₂FC) relative to untreated controls for each condition (columns) for genes (rows) among the top 50 DEGs (by *P* values) in at least one condition. *ACE2* was not among these genes and is shown as a separate row. Inset: a cluster of genes downregulated upon both *BRD2* inhibition and *BRD2* knockdown. Among these, genes associated with the Gene Ontology (GO) term ‘Cellular Response to Type I interferon’ are marked by black circles. **b**, Significantly enriched (FDR < 0.05) GO biological process terms for the genes shown in the inset in **a**. **c**, Enrichment analysis for genes in the inset in **a** reveals COVID-19-related gene sets. Genes associated with a gene set are marked in blue. **d**, Calu-3 cells expressing sgRNAs knocking down genes essential for interferon signalling assayed for *ACE2* gene expression relative to *ACTB* by qPCR. Average *ACE2* mRNA level relative to non-targeting control (NTC) sgRNA and the standard deviation of *n* = 3 biological replicates are shown for each condition. **e**, WT (dark grey) or *BRD2* KD (light grey) Calu-3 cells were treated with IFN- β , and transcript levels of *ACE2* relative to *ACTB* were quantified at 72 h post-treatment by qPCR. Average *ACE2* levels relative to vehicle-treated and the standard deviation of *n* = 3 biological replicates are shown for each condition. **f**, Primary human bronchial epithelial cells (left) and human iPSC-derived cardiomyocytes (right) were treated with the indicated concentrations of IFN- β , and transcript levels of *ACE2* relative to *ACTB* (for Calu-3 and primary human bronchial epithelial cells) or 18S rRNA (cardiomyocytes) were quantified at 72 h post-treatment by qPCR. Average *ACE2* mRNA levels relative to WT and the standard deviation of *n* = 3 biological replicates are shown for each condition.

(Fig. 4a,b). Furthermore, the genes downregulated by both *BRD2* knockdown and inhibition were strongly enriched in genes induced by SARS-CoV-2 infection in patient and cultured cells (Fig. 4c).

These findings are compatible with two distinct mechanisms: *BRD2* could independently regulate *ACE2* and SARS-CoV-2-induced interferon response genes, or *BRD2* could mediate the response to interferon, which in turn regulates *ACE2* transcription. *ACE2* expression has been reported to be induced by interferons in some studies^{1,2}. Other studies, however, suggest that interferon suppresses *ACE2* expression³.

In Calu-3 cells, disruption of basal interferon signalling, via knockdown of the genes essential for interferon signal transduction—*IRF9*, *STAT1* or *IFNAR1*—abrogated *ACE2* expression (Fig. 4d and Extended Data Fig. 5). Conversely, treatment with exogenous interferon- β (IFN- β) stimulates *ACE2* expression in a concentration-dependent manner. Upon *BRD2* knockdown, however, this concentration-dependent increase in *ACE2* expression is inhibited (Fig. 4e). Thus, *BRD2* is required for interferon-induced *ACE2* expression. Treatment with IFN- β similarly strongly increased *ACE2* mRNA levels in primary human bronchial epithelial cells, but reduced *ACE2* mRNA levels in human iPSC-derived cardiomyocytes (Fig. 4f), suggesting that the effect of interferons on *ACE2* can be context-dependent.

To test whether *BRD2* is a direct transcriptional regulator of *ACE2*, we performed CUT&RUN³³ to comprehensively map genomic loci bound by *BRD2* in Calu-3 cells (Supplementary Table 3). CUT&RUN is similar to chromatin immunoprecipitation sequencing (ChIP-seq), as it measures the occupancy of factors bound to DNA, but has the advantage of higher sensitivity and a lower requirement for cell numbers³³. Genes adjacent to *BRD2*-bound sites detected in our experiment showed a highly significant overlap with *BRD2*-bound sites previously mapped by ChIP-seq in NCI-H23³⁴ cells, another lung epithelium-derived

cancer cell line (Fig. 5a). To further validate our CUT&RUN analysis, we performed binding and expression target analysis³⁵ (BETA) to uncover direct *BRD2* targets that were differentially expressed upon *BRD2* knockdown, and identified several interferon response genes as direct *BRD2* targets that were downregulated (Fig. 5b). We verified that a previously described³⁴ *BRD2* binding site upstream of *PVT1* was also detected in our experiment (Fig. 5c). We also mapped a *BRD2* binding site upstream of the several interferon-stimulated genes (ISGs), including *IRF9*, *STAT1* and *MX1* (Fig. 5d). Although there was some signal in the WT background at the *ACE2* locus that is decreased in *BRD2* knockdown cells, there were no peaks as determined by the peak calling algorithm (Fig. 5e), suggesting that *BRD2* is not a direct transcriptional regulator of *ACE2* expression.

We also performed CUT&RUN for histone H2A.Z, which was previously reported to modulate the magnitude of ISG expression and thus connect *BRD2* activity to interferon stimulation³⁶. We found decreased H2A.Z occupancy at ISGs in *BRD2* knockdown cells (Fig. 5d), recapitulating the role of *BRD2* as a potential chaperone of H2A.Z. These results support a model in which *BRD2* controls the transcription of key interferon response genes, which can in turn induce *ACE2* transcription in some cell types (Fig. 5f). Alternatively, *ACE2* expression may be controlled by other genes that are expressed in a *BRD2*-dependent, interferon-stimulated manner (Fig. 5f).

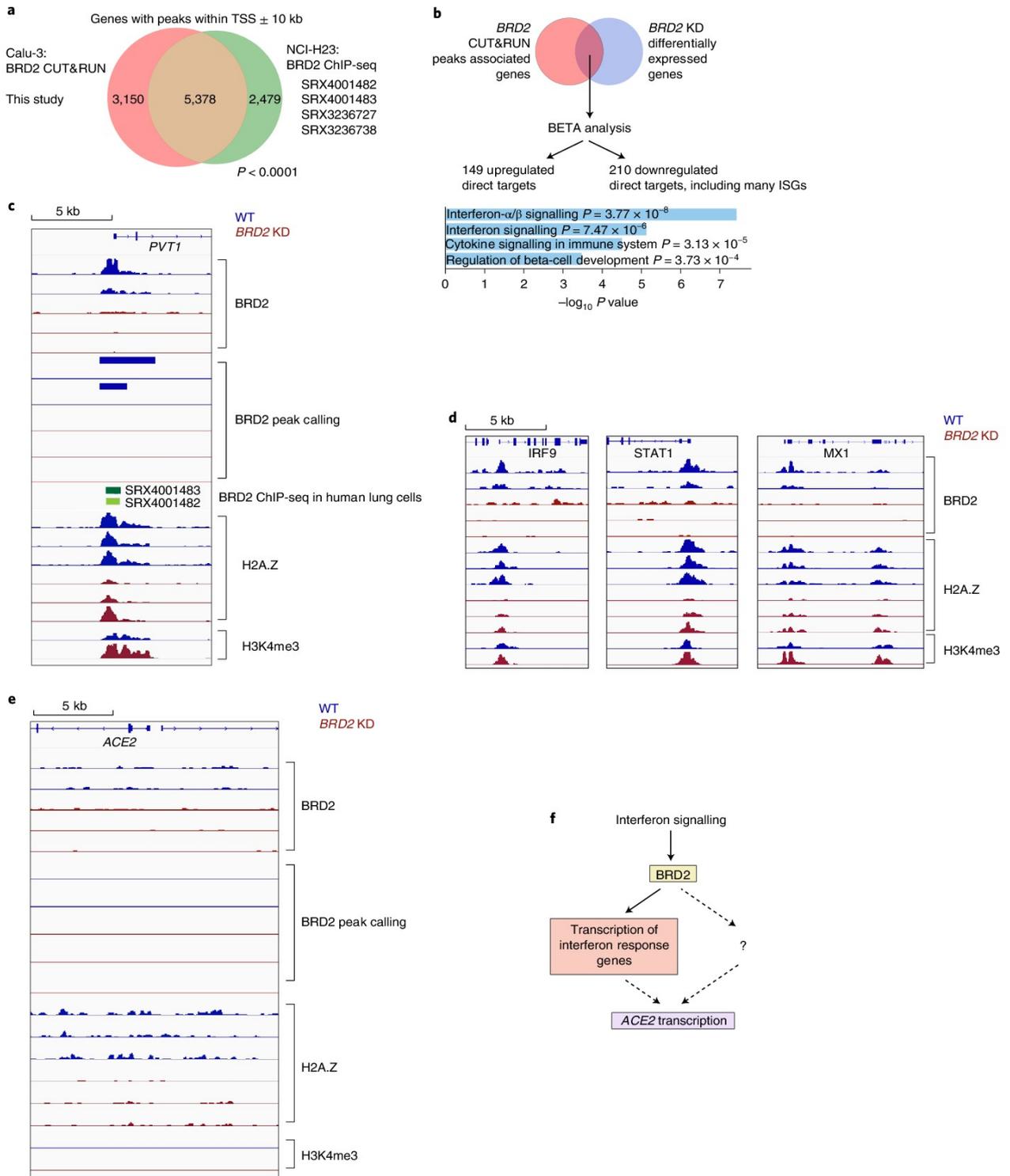
ABBV-744 reduces infection in primary cells and in vivo. We then tested whether ABBV-744, a bromodomain inhibitor currently in clinical trials, could reduce SARS-CoV-2 infection and infection-associated phenotypes in more physiological models.

First, we investigated a human nasal epithelial model³⁷. We treated reconstituted nasal epithelia maintained in air/liquid interphase conditions with 100 nM and 300 nM ABBV-744 (above the reported half-maximum inhibitory concentration (IC₅₀) of 4–18 nM)

Fig. 5 | BRD2 directly regulates transcription of interferon-induced genes. **a**, Genes associated with *BRD2* CUT&RUN peaks within 10 kb of a transcription start site determined in this study in Calu-3 cells overlap significantly with published *BRD2* ChIP-seq peaks from the indicated datasets (*P* < 0.0001, two-sided Fisher’s exact test). **b**, BETA was performed to identify direct *BRD2* targets that were differentially expressed upon *BRD2* knockdown using the one-tailed Kolmogorov-Smirnov test. Many interferon response genes were identified as direct *BRD2* targets. Direct *BRD2* targets that were downregulated upon *BRD2* knockdown were analysed by ENRICH for enriched Reactome pathways. Pathways with adjusted *P* values less than 0.05 are displayed. **c–e**, CUT&RUN experiments were conducted to map *BRD2*, H2A.Z and H3K4me3 genomic localization in WT (blue traces) and *BRD2* knockdown (red) Calu-3 cells. Each trace represents an independent biological replicate. **c**, Known *BRD2* regulatory sites are recapitulated. Raw signal tracks for WT and *BRD2* knockdown cells are shown at the known *BRD2* locus *PVT1*. *BRD2* ChIP-seq tracks from human lung cells are shown. *BRD2* peaks were called over IgG using SEACR at FDR < 0.05. **d**, Identified *BRD2* and histone H2A.Z occupancy and peak calling at ISGs in *BRD2* knockdown and WT Calu-3 cells. Raw signal tracks for *BRD2*, histone H2A.Z and H3K4me are shown. *BRD2* peaks were called over IgG using SEACR at FDR < 0.05. *BRD2* ChIP-seq tracks from human lung cells are also shown. **e**, Raw *BRD2* signal tracks and peak calling as for **c**, at the *ACE2* locus. **f**, Proposed model for *BRD2* control of *ACE2* expression.

and performed SARS-CoV-2 or mock infections (Fig. 6a). First, we found that ABBV-744 treatment reduced *ACE2* levels in these conditions (Fig. 6b). Apical supernatants did not show significant changes in viral RNA concentrations at two or four days post-infection (Extended Data Fig. 6a). Intracellular viral RNA concentrations, however, were significantly decreased in the

ABBV-744 conditions (Fig. 6c). Furthermore, epithelial barrier integrity, as measured by transepithelial electrical resistance (Fig. 6d) and cytotoxicity (Fig. 6e), were rescued in infected cells treated with ABBV-744. Thus, ABBV-744 partially inhibited SARS-CoV-2 replication and fully rescued epithelial barrier integrity in a primary human nasal epithelial model.



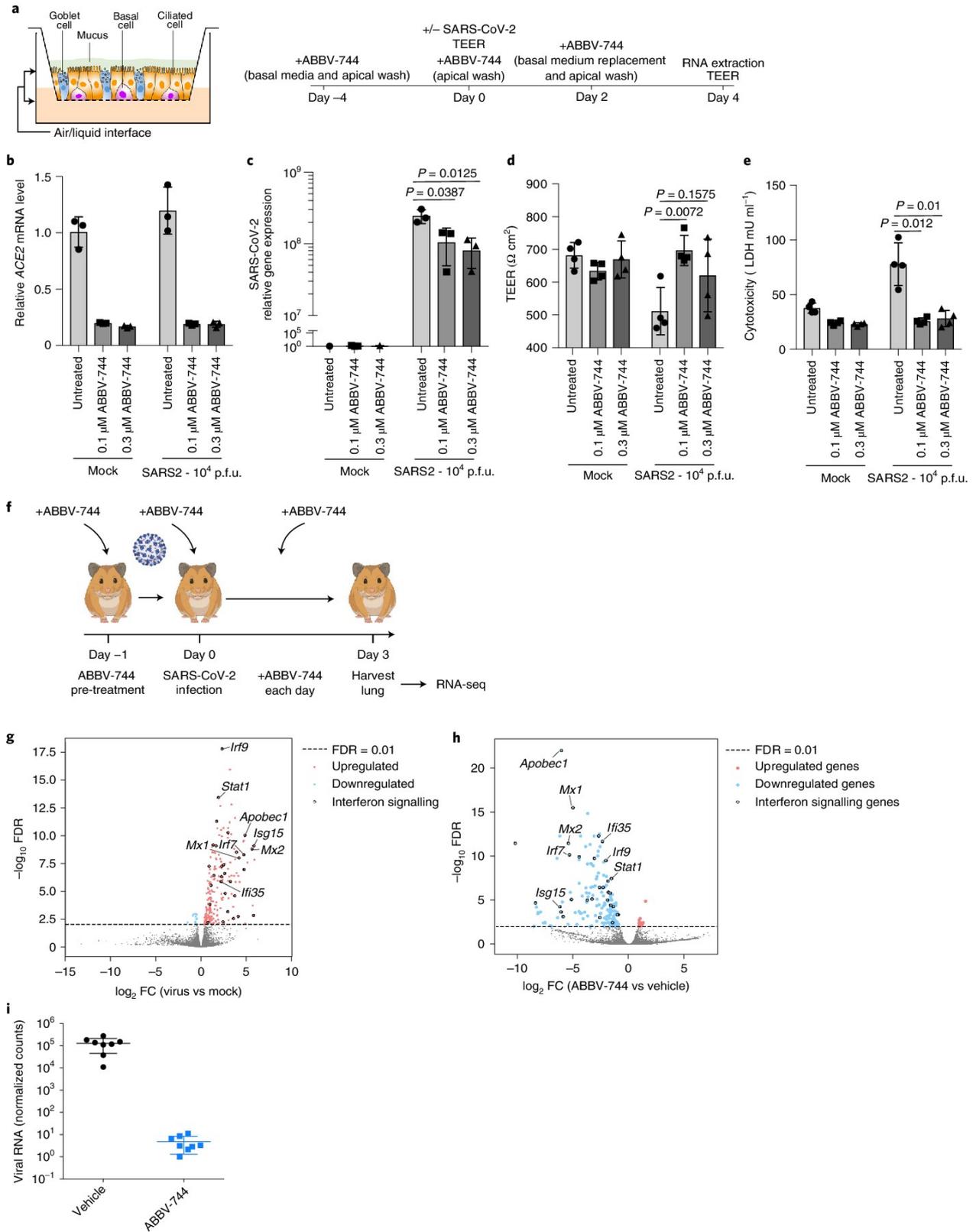


Fig. 6 | BRD2 inhibitors prevent cytotoxicity and reduce SARS-CoV-2 infection in human primary nasal epithelia and inhibit SARS-CoV-2 infection in Syrian hamsters. **a**, Design for reconstructed human nasal epithelia experiments. **b**, *ACE2* transcript levels relative to the average of *GAPDH*, *TFRC*, *RPL13* and *ACTB* as a function of ABBV-744 concentration and/or SARS-CoV-2 infection (right). The average *ACE2* mRNA level relative to mock infection and vehicle-treated of $n=3$ biological replicates, with error bars representing the standard deviation, are shown. **c**, Intracellular SARS-CoV-2 gene expression relative to the average of *GAPDH*, *TFRC*, *RPL13* and *ACTB* as a function of ABBV-744 concentration and/or SARS-CoV-2 infection (right). The average of $n=3$ biological replicates and error bars representing the standard deviation are shown. *P* values were determined using Student's unpaired two-tailed *t*-test. SARS-CoV-2 gene expression relative to mock infection and vehicle-treated are shown. **d**, Transepithelial electrical resistance (TEER), evaluated as a function of ABBV-744 concentration and/or SARS-CoV-2 infection. The average of $n=4$ biological replicates and error bars representing the standard deviation are shown. *P* values were determined using Student's unpaired two-tailed *t*-test. **e**, Cytotoxicity, as measured by LDH release, evaluated as a function of ABBV-744 concentration and/or SARS-CoV-2 infection. The average of $n=4$ biological replicates and error bars representing the standard deviation are shown. *P* values were determined using Student's unpaired two-tailed *t*-test. **f**, Experimental design for the Syrian hamster experiments. **g**, Volcano plot showing differentially expressed genes for Syrian hamster lungs infected or not with SARS-CoV-2. A subset of ISGs for which BRD2 peaks were identified by CUT&RUN are labelled. **h**, Volcano plot showing differentially expressed genes for Syrian hamster lungs infected with SARS-CoV-2 and treated with vehicle or ABBV-744 at 100 nM. A subset of ISGs for which BRD2 peaks were identified by CUT&RUN are labelled. **i**, Normalized viral RNA counts for Syrian hamsters infected with SARS-CoV-2 and treated with vehicle or ABBV-744 at 100 nM. The average of $n=8$ biological replicates and error bars representing the standard deviation are shown.

Next, we tested whether ABBV-744 could reduce SARS-CoV-2 infection in golden Syrian hamsters. Syrian hamsters provide a physiologically relevant model for SARS-CoV-2 infection, with high viral replication and signs of lung involvement^{38–41}. In our hands, hamsters did not show substantial weight loss following SARS-CoV-2 infection (Extended Data Fig. 6b). After 24 h of treatment with ABBV-744 or vehicle, hamsters were infected with SARS-CoV-2 (Fig. 6f) and treated daily with ABBV-744 or vehicle. Three days post-infection, the lungs of hamsters were harvested and subjected to RNA-seq. Infected, but untreated, hamsters showed marked upregulation of a number of genes including ISGs when compared to uninfected controls (Fig. 6g). By contrast, infected hamsters treated with ABBV-744 showed a downregulation of ISG (Fig. 6h) levels relative to vehicle-treated infected hamsters, confirming ABBV-744 activity. Remarkably, viral RNA counts were reduced by about five orders of magnitude in the ABBV-744-treated hamsters versus those treated with vehicle controls (Fig. 6i). Thus, Brd2 inhibition can dramatically decrease SARS-CoV-2 infection in Syrian hamsters.

Discussion

We have demonstrated that BRD2 is necessary for *ACE2* expression in a number of different SARS-CoV-2-relevant systems. We also found that treatment with ABBV-744, a bromodomain inhibitor, can reduce SARS-CoV-2 viral RNA concentrations in primary human nasal epithelial cells and Syrian hamsters. These findings suggest that pharmacological BRD2 inhibitors may be of therapeutic benefit to prevent or reduce the impact of SARS-CoV-2 infection.

Our data suggest that BRD2 is an indirect regulator of *ACE2* transcription and thus viral entry in COVID-19-relevant cell types. This is consistent with two other recent studies that show that BET-domain inhibitors can prevent SARS-CoV-2 infection in culture and reduce *ACE2* gene expression^{42,43}. Our data suggest that BRD2 is the most likely target of the BET-domain inhibitors used in these studies, as opposed to other BET-domain containing proteins, such as BRD4.

We also show that BRD2 is required for interferon-mediated stimulation of *ACE2* expression, as both exogenous interferon stimulation and basal interferon stimulation of *ACE2* expression are blocked upon BRD2 knockdown or pharmacological inhibition (Fig. 6i). Indeed, inhibition of the Janus kinase (required for transduction of interferon signalling) with ruxolitinib has been shown to downregulate *ACE2* levels in cell culture⁴⁴. This does not, however, preclude a more direct, and interferon-independent, regulatory mode.

Our data also show that BRD2 activity is essential for the transcription of ISGs in cell culture and in Syrian hamsters. Based on

our findings and the previous literature³⁶, BRD2 regulation of ISG transcription is probably mediated by a reduction in histone H2A.Z occupancy at these promoters. Intriguingly, these BRD2-dependent effects might be cell-type specific, as one study⁴⁵ has shown that BET inhibition, but not the BET inhibitor ABBV-744, can block cardiac dysfunction due to inflammation caused by SARS-CoV-2 infection. Either way, our data indicate that BRD2 could be a key regulator of the host response to SARS-CoV-2 infection.

The previously described³² interaction between the SARS-CoV-2 E protein and BRD2 might have evolved to manipulate gene expression during infection, including the expression of *ACE2*. In isolation, however, protein E overexpression in Calu-3 cells did not recapitulate expression changes resulting from BRD2 knockdown or inhibition (Fig. 4a). These data suggest that there is no direct effect of protein E on BRD2 function, or that other viral or host factors expressed during SARS-CoV-2 infection are required to modulate BRD2 function. Further studies are needed to define the function of the protein E–BRD2 interaction.

Several previous CRISPR screens aiming to uncover strategies to inhibit SARS-CoV-2 infection were carried out in cell lines in which an *ACE2* transgene was overexpressed^{9,10}. Those screens therefore failed to uncover BRD2 as a regulator of endogenous *ACE2* expression. BRD2 did show a phenotype, however, in a CRISPR screen carried out in Vero-E6 cells (which express *ACE2* endogenously)¹², although it was not further characterized in that study. These differences highlight the importance of conducting CRISPR-based screens in disease-relevant cell types.

There is a growing literature about the relationship between COVID-19 disease severity, *ACE2* expression and interferon regulation^{1–6}. Given that *ACE2* is known to promote recovery after lung injury and that SARS-CoV-2 manipulates the host interferon response^{46–48}, the misregulation of these two pathways may play a major role in enhancing the severity of COVID-19. Our data suggest that BRD2 is central to this regulatory network and therefore pharmacological targeting of BRD2 may be a promising therapeutic strategy for the treatment of COVID-19: BRD2 inhibition could both block viral entry, through *ACE2* downregulation, and act as an 'emergency brake' for misregulated patient immune responses to COVID-19, via downregulation of ISGs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-021-00821-8>.

Received: 18 January 2021; Accepted: 24 November 2021;
Published online: 13 January 2022

References

- Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**, 1016–1035 (2020).
- Chu, H. et al. Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an ex vivo study with implications for the pathogenesis of COVID-19. *Clin. Infect. Dis.* **71**, 1400–1409 (2020).
- Gutiérrez-Chamorro, L. et al. SARS-CoV-2 infection suppresses ACE2 function and antiviral immune response in the upper respiratory tract of infected patients. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.11.18.388850> (2020).
- Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045 (2020).
- Bastard, P. et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
- Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
- Zhang, Q. et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
- Samuel, R. M. et al. Androgen signaling regulates SARS-CoV-2 receptor levels and is associated with severe COVID-19 symptoms in men. *Cell Stem Cell* **27**, 876–889 (2020).
- Daniloski, Z. et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105 (2020).
- Wang, R. et al. Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. *Cell* **184**, 106–119.e14 (2020).
- Schneider, W. M. et al. Genome-scale identification of SARS-CoV-2 and pan-coronavirus host factor networks. *Cell* **184**, 120–132 (2021).
- Wei, J. et al. Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cell* **184**, 76–91 (2020).
- Shi, J. & Vakoc, C. R. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol. Cell* **54**, 728–736 (2014).
- Fujisawa, T. & Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.* **18**, 246–262 (2017).
- Lui, I. et al. Trimeric SARS-CoV-2 Spike interacts with dimeric ACE2 with limited intra-Spike avidity. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.05.21.109157> (2020).
- Lan, J. et al. Structure of the SARS-CoV-2 Spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
- Chua, R. L. et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).
- Tseng, C.-T. K. et al. Apical entry and release of severe acute respiratory syndrome-associated coronavirus in polarized Calu-3 lung epithelial cells. *J. Virol.* **79**, 9470–9479 (2005).
- Kuchi, S., Gu, Q., Palmarini, M., Wilson, S. J. & Robertson, D. L. Meta-analysis of virus-induced host gene expression reveals unique signatures of immune dysregulation induced by SARS-CoV-2. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.12.29.424739> (2020).
- Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
- Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
- Horlbeck, M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).
- Deffieu, M. S. et al. Rab7-harboring vesicles are carriers of the transferrin receptor through the biosynthetic secretory pathway. *Sci. Adv.* **7**, 1–17 (2021).
- Doroshov, D. B., Eder, J. P. & LoRusso, P. M. BET inhibitors: a novel epigenetic approach. *Ann. Oncol.* **28**, 1776–1787 (2017).
- Xu, Y. & Vakoc, C. R. Targeting cancer cells with BET bromodomain inhibitors. *Cold Spring Harb. Perspect. Med.* **7**, a026674 (2017).
- Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
- Faivre, E. J. et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **578**, 306–310 (2020).
- Winter, G. E. et al. BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol. Cell* **67**, 5–18 (2017).
- Shi, C. et al. PROTAC induced-BET protein degradation exhibits potent anti-osteosarcoma activity by triggering apoptosis. *Cell Death Dis.* **10**, 815 (2019).
- Bermejo, J. A. P. et al. SARS-CoV-2 infection of human iPSC derived cardiac cells reflects cytopathic features in hearts of patients with COVID-19. *Sci. Transl. Med.* **13**, 1–15 (2021).
- Mulay, A. et al. SARS-CoV-2 infection of primary human lung epithelium for COVID-19 modeling and drug discovery. *Cell Rep.* **35**, 109055 (2021).
- Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
- Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
- Handoko, L. et al. JQ1 affects BRD2-dependent and independent transcription regulation without disrupting H4-hyperacetylated chromatin states. *Epigenetics* **13**, 410–431 (2018).
- Wang, S. et al. Target analysis by integration of transcriptome and ChIP-seq data with BET. *Nat. Protoc.* **8**, 2502–2515 (2013).
- Au-Yeung, N. & Horvath, C. M. Histone H2A.Z suppression of interferon-stimulated transcription and antiviral immunity is modulated by GCN5 and BRD2. *iScience* **6**, 68–82 (2018).
- Robinet, R. et al. SARS-CoV-2 infection induces the dedifferentiation of multiciliated cells and impairs mucociliary clearance. *Nat. Commun.* **12**, 4354 (2021).
- Osterrieder, N. et al. Age-dependent progression of SARS-CoV-2 infection in Syrian hamsters. *Virus* **12**, 94301 (2020).
- Imai, M. et al. Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. *Proc. Natl Acad. Sci. USA* **117**, 16587–16595 (2020).
- Sia, S. F. et al. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* **583**, 834–838 (2020).
- Rosenke, K. et al. Defining the Syrian hamster as a highly susceptible preclinical model for SARS-CoV-2 infection. *Emerg. Microbes Infect.* **9**, 2673–2684 (2020).
- Qiao, Y. et al. Targeting transcriptional regulation of SARS-CoV-2 entry factors ACE2 and TMPRSS2. *Proc. Natl Acad. Sci. USA* **118**, e2021450118 (2020).
- Gilham, D. et al. Bromodomain and extraterminal protein inhibitor, apabetalone (RVX-208), reduces ACE2 expression and attenuates SARS-CoV-2 infection in vitro. *Biomedicines* **9**, 437 (2021).
- Lee, H. K., Jung, O. & Hennighausen, L. JAK inhibitors dampen activation of interferon-stimulated transcription of ACE2 isoforms in human airway epithelial cells. *Commun. Biol.* **4**, 654 (2021).
- Mills, R. J. et al. BET inhibition blocks inflammation-induced cardiac dysfunction and SARS-CoV-2 infection. *Cell* **184**, 2167–2182 (2021).
- Ribero, M. S., Jouvenet, N., Dreux, M. & Nisole, S. Interplay between SARS-CoV-2 and the type I interferon response. *PLoS Pathog.* **16**, e1008737 (2020).
- Lei, X. et al. Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat. Commun.* **11**, 3810 (2020).
- Xia, H. et al. Evasion of type I interferon by SARS-CoV-2. *Cell Rep.* **33**, 108234 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Cell culture. Calu-3 cells were cultured in RPMI 1640 (Life Technologies, 22400-105) with 10% fetal bovine serum (FBS; VWR, 89510-186), 1% pen/strep (Life Technologies, 15140122) and 5 mM glutamine (Life Technologies, 25030081) at 37 °C and 5% CO₂. Cells were split by treatment with TrypLE (Life Technologies, 12604013) for 15 min, quenching with media and spinning down at 200g for 5 min.

At Institut Pasteur, where virus infections were carried out, Calu-3 cells were cultured in minimum essential medium (MEM; Gibco, 11095-080) with 20% FBS (Gibco, A3160801), 1% pen/strep (Gibco, 15140-122), 1% non-essential amino acids (NEAA; Sigma-Aldrich, M7145) and 1 mM sodium pyruvate (Sigma-Aldrich, S8636). These were split in trypsin-EDTA 0.05% (Gibco, 11580626).

HEK293 cell culture and the production of lentivirus were performed as previously described¹⁹.

A vial of short tandem repeat (STR) authenticated Caco-2 cells was obtained from the UCSF Cell and Genome Engineering Core (CGEC). Caco-2 cells were cultured in Eagle's minimum essential medium (EMEM; ATCC, 30-2003) with 20% FBS (VWR, 89510-186), 1% pen/strep (Life Technologies, 15140122) and 5 mM glutamine (Life Technologies, 25030081) at 37 °C and 5% CO₂.

A vial of A549 cells was obtained from D. Ruggero's laboratory as a gift. A549 cells were cultured in Dulbecco's modified Eagle medium (DMEM; Thermo Fisher Scientific, 10313-039) with 10% FBS (VWR, 89510-186), 1% pen/strep (Life Technologies, 15140122) and 5 mM glutamine (Life Technologies, 25030081) at 37 °C and 5% CO₂.

Human iPSC-derived cardiomyocytes were generated and cultured, as previously described³⁰, from AIC590 iPSCs (Allen Institute Cell catalogue). Drugs were added on day 69 of differentiation, and cardiomyocytes were collected for analysis on day 72.

Normal human bronchial epithelia (Mattek NHBE-CRY) were cultured following the supplier's instructions.

Generation of the Calu-3 ACE2 knockout line. The polyclonal ACE2 knockout Calu-3 cell line was generated using the Gene KO kit V2 from Synthego, using three sgRNAs targeting ACE2 with the following protospacer sequences: sRNA1, 5'-GACAUUCUCUUCAGUAAUUAU-3'; sgRNA2, 5'-AAACUUGUCCAAA AUGUCU-3'; sgRNA3, 5'-UUACAGCAACAAGGUCUGAGA-3'. The sgRNAs were designed according to Synthego's multiguide gene knockout kit³⁰. Briefly, two or three sgRNAs are bioinformatically designed to work in a cooperative manner to generate small, knockout-causing fragment deletions in early exons. These fragment deletions are larger than standard indels generated from single guides. The genomic repair patterns from a multiguide approach are highly predictable on the basis of the guide spacing and design constraints to limit off-targets, resulting in a higher-probability protein knockout phenotype.

The ribonucleoprotein complex with a ratio of 4.5:1 of sgRNA and Cas9 was delivered following the protocol of the SE Cell Line 4D-Nucleofector™ X Kit (Lonza, V4XC-1012), using the nucleofection program DS-130 on the Lonza 4D X unit. At 72 h post-transfection, genomic DNA was extracted to serve as the template for PCR amplification of the region that covers the sites targeted by the sgRNAs with the following two primers: ACE2-F, 5'-CTGGGACTCCAAAATCA GGA-3'; ACE2-R, 5'-CGCCCAACCAAGTTCAAAG-3'. Sanger sequencing reactions using the sequencing primer ACE2-seq 5'-CAAAATCAGGGATATGG AGCCAAACATC-3' were then performed, and the knockout efficiency was determined to be 80% with ICE software from Synthego³¹ (v2, <https://ice.synthego.com/#/>).

Generation of the Calu-3 CRISPRi line. The parental Calu-3 line was obtained from the UCSF Cell and Genome Engineering Core. Calu-3 cells were cultured at 37 °C with 5% CO₂ in EMEM medium containing 10% FBS, 100 U ml⁻¹ streptomycin, 100 µg ml⁻¹ penicillin and 2 mM glutamine. To generate the CRISPRi lines, ~3 × 10⁸ cells were seeded into medium containing lentiviral particles packaging dCas9-BFP-KRAB under a UCOE (universal chromatin opening element)-SPFV (spleen focus-forming virus) promoter³². Five days post-infection, blue fluorescent protein (BFP)-positive cells were sorted using BD Fusion. To validate the CRISPRi line, Calu-3-CRISPRi cells were transduced with lentiviral particles expressing non-targeting sgRNA (protospacer 5'-GCTCCAGTCGGC ACCACAG-3') or CD81-targeting sgRNA (protospacer 5'-GGCCTGGCAGGATG CGCGG-3'). CD81 expression was measured seven days post-transduction by dislodging cells with TrypLE and live cells were stained with Allophycocyanin (APC)-conjugated anti-human CD81 antibody (Biolegend, 349509). CD81 expression was assessed on a BD LSRiI system with >90% of Calu-3 CRISPRi cells with CD81 knocked down compared to a non-targeting sgRNA control.

Spike-RBD binding assay. Recombinant biotinylated SARS-CoV-2 spike Spike-receptor-binding domain with a C-terminal human IgG Fc domain fusion (referred to as Spike-RBD) was prepared as previously described³³. Calu-3 cells were grown in 96-well flat-bottom plates until >50% confluence. Medium was aspirated and cells were washed once with phosphate-buffered saline (PBS). Cells were then treated with TrypLE to release them from the plate, RPMI 1640 medium was added to dilute TrypLE, and cells were pelleted by centrifugation at 200g for 5 min. From

this point on, all steps were carried out on ice. Cells were incubated in 3% bovine serum albumin (BSA; Sigma-Aldrich, A7030) in Dulbecco's phosphate-buffered saline (DPBS; Sigma-Aldrich, D8537) for 15 min to block, then washed twice in 3% BSA in DPBS by centrifugation at 200g for 5 min in V-bottom plates, followed by resuspension. Spike-RBD was diluted in 3% BSA to the appropriate concentrations and incubated with cells for 30 min on ice. Cells were then washed twice with 3% BSA in DPBS and incubated with anti-strep PE-Cy7 (Thermo Fisher, SA1012) at 5 µg ml⁻¹. Cells were washed twice and subjected to flow cytometry on a FACSCelesta in HTS mode. Cells were gated to exclude doublets and the median PE-Cy7 signal was calculated for each sample. The gating strategy is shown in Supplementary Fig. 2. BD FACSDiva (version 8.01.1) was used to collect flow cytometry data and perform fluorescence activated cell sorting (FACS). EC₅₀ values and their 95% confidence intervals were calculated by fitting the RBD binding data into a Sigmoidal 4PL model in Prism 6.

CRISPRi screen. Calu-3 cells were infected with the H1 CRISPRi sgRNA library²² as described in ref. ⁴⁹ and selected using treatment with 1 µg ml⁻¹ puromycin for 3 days. After selection, cells were stained with 10 nM Spike-RBD as described above or for TFRC as previously described⁴⁹ and subjected to FACS, where cells were sorted into the top 30% and bottom 30% based on high and low expression of TFRC or Spike-RBD. Because of the viability and stickiness known for Calu-3 cells, coverage was lower than optimal, at 200-fold over the library diversity. Sorted populations were spun down at 200g for 5 min, and genomic DNA was isolated as described in ref. ⁴⁹. sgRNA cassettes were amplified by PCR and sequencing and analysis was performed as described in ref. ⁴⁹ but with FDR of 0.1 rather than 0.05 or 0.01, due to noise.

Validation of screening hits. Individual sgRNAs were selected based on phenotypes in the primary screens and cloned into a lentiviral expression vector as described⁴⁹. Protospacer sequences of these sgRNAs are provided in cells expressing sgRNAs were selected using treatment with 1 µg ml⁻¹ puromycin for 3–7 days.

Drug treatments. Drugs (ABBV-744, Selleckchem, S8723; JQ1, Sigma-Aldrich, SML1524; dBET-6, Selleckchem, S8762) were dissolved in dimethyl sulfoxide (DMSO) or water according to the manufacturers' instructions. Cells were treated with drugs for 72 h with medium changes performed every 24 h with medium containing fresh drug.

Interferon treatments. IFN-β (R&D Systems, 8499-IF) was dissolved according to the manufacturer's instructions. Cells were treated with IFN-β for 72 h, with medium changes performed every 24 h with medium containing fresh IFN-β.

Quantitative PCR. qPCR was performed and analysed as described in ref. ⁴⁹. Primers: ACE2 forward, GGCTTCTGTCCACCGATT; ACE2 reverse, CACT CACCTCCACTTCTCTAAC; ACTB forward, ACCTTCTACAATGAGCTGCG; ACTB reverse, CTGGATAGCAACGTCATG; IRF9 forward, GCCCTACAA GGTGTATCAGTTG; IRF9 reverse, TGCTGTCGCTTTGATGGTACT; IFNAR1 forward, AACAGGAGCGATGAGTCTGTC; IFNAR1 reverse, TGCGAAATGG TGTAATGAGTCA; STAT1 forward, CAGCTTGACTCAAATTCCTGGA; STAT1 reverse, TGAAGATTACGCTTCTTTCTT.

Western blotting. Cells from one confluent well of a six-well plate were lysed in RIPA buffer plus cOmplete EDTA-free protease inhibitor tablets (Roche, 11873580001) and spun for 10 min at 21,000g at 4 °C. The pellet was removed and a bicinchoninic acid (BCA) assay (Thermo Fisher, 23225) was performed on the remaining supernatant. Lysate volumes with equivalent protein content were diluted with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) loading dye and subjected to gel electrophoresis on 4–12% BisTris SDS-PAGE gels (Life Technologies, NP0322). Gels were then transferred and blocked in 5% non-fat dry milk (NFD) for 1 h at room temperature. Antibodies in fresh 5% NFD were added (mouse monoclonal GAPDH 1:10,000; goat polyclonal ACE2, R&D Tech AF933, 1:200; rabbit monoclonal BRD2, Abcam 197865, 1:5,000) and incubated at 4 °C for at least 16 h. Membranes were washed four times with TBS + 0.1% Tween-20 and incubated with secondary antibodies (1:10,000 donkey anti-goat-800 (LI-COR, 926-32214), donkey anti-mouse-680 (LI-COR, 926-68072) and horse radish peroxidase (HRP) donkey anti-rabbit (CST, 7074P2)). Membranes were visualized using a LI-COR or Fento HRP kit (Thermo Fisher, 34094). Uncropped images of western blots are provided as Supplementary Fig. 1.

Virus. The SARS-CoV-2 strain used (BetaCoV/France/IDF0372/2020 strain) was propagated once in Vero-E6 cells and was a gift from the National Reference Centre for Respiratory Viruses at Institut Pasteur, Paris, originally supplied through the European Virus Archive goes global platform.

Cytotoxicity measurements of Calu-3 cells. Thirty-thousand Calu-3 cells per well were seeded into Greiner 96-well white-bottom plates and incubated for 48 h at 37 °C and 5% CO₂. The cells were then treated with identical drug concentrations as in the infection assays, for five days, by refreshing the medium with 100 µl per

well fresh drug-containing medium every 24 h. Cell viability was then assayed by adding 100 μ l per well of CellTiter-Glo 2.0 (Promega) and incubated for 10 min at room temperature. Luminescence was recorded with an Infinite 200 Pro plate reader (Tecan) using an integration time of 1 s.

Virus infection assays. Thirty-thousand Calu-3 cells per well were seeded into 96-well plates and incubated for 48 h at 37°C and 5% CO₂. At the time of infection, the medium was replaced with virus inoculum (multiplicity of infection of 0.1 plaque-forming units (p.f.u.) per cell) and incubated for 1 h at 37°C and 5% CO₂. Following the 1-h adsorption period, the inoculum was removed, replaced with fresh medium, and the cells incubated at 37°C and 5% CO₂. At 24 h, 48 h and 72 h post-infection, the cell-culture supernatant was collected, and the viral load was assessed by RT-qPCR as described previously³². Briefly, the cell-culture supernatant was collected, heat-inactivated at 95°C for 5 min, and used for RT-qPCR analysis. SARS-CoV-2-specific primers targeting the N gene region—5'-TAATCAGACAAGGAAGTATGATTA-3' (forward) and 5'-CGAAGGTGTGACTTCCATG-3' (reverse)—were used with the Luna Universal One-Step RT-qPCR Kit (New England Biolabs) in an Applied Biosystems QuantStudio 6 thermocycler or an Applied Biosystems StepOnePlus system, with the following cycling conditions: 55°C for 10 min, 95°C for 1 min, and 40 cycles of 95°C for 10 s, followed by 60°C for 1 min. The number of viral genomes is expressed as p.f.u. equivalents per ml, and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titre.

Plaque assays. Viruses were quantified by plaque-forming assays. For this, Vero-E6 cells were seeded in 24-well plates at a concentration of 1×10^5 cells per well. The following day, tenfold serial dilutions of individual virus samples in serum-free DMEM medium were added to infect the cells at 37°C for 1 h. After the adsorption time, a solid agarose overlay (DMEM, 10% (vol/vol) PBS and 0.8% agarose) was added. The cells were incubated for a further three days before fixation with 4% formalin and visualization using crystal violet solution.

CUT&RUN. CUT&RUN was performed with one million Calu-3 cells. Cells were removed from the plate by treatment with Versene (Life Technologies 15040066) for 20 min and resuspended in fresh medium. They were then spun down and washed twice with DPBS before proceeding with the CUTANA CUT&RUN kit (EpiCypher, 14-0050). The experiment was performed with the included IgG and H3K4Me control antibodies and the BRD2 antibody (Abcam 197865) as well as *Escherichia coli* spike-in DNA according to the kit protocol.

QuantSeq analysis. Raw sequencing reads from QuantSeq were trimmed using Trimmomatic³³ (v0.39, PMID 24695404) and mapped to the human reference transcriptome (GRCh38, GENCODE Release 36) using Salmon³⁴ (v1.3.0) to obtain transcript abundance counts. Gene-level count estimates were obtained using tximport³⁴ (v1.18.0) with default settings. Subsequently, differential gene-expression analyses were performed using the glmQLFTest method implemented in the edgeR package³⁵ (v3.28.1). Cluster³⁶ (v3.0) was used for hierarchical clustering and Java TreeView³⁶ (v1.1.6r4) for visualization.

CUT&RUN analysis. CUT&RUN analysis was performed as previously described³⁷. Briefly, paired-end reads were mapped to the human genome GRCh38 using Bowtie2 (v2.3.4.1) with the options --end-to-end --very-sensitive --no-unal --no-mixed --no-discordant --phred33 -1 10 -X 1000. Sparse enrichment analysis for CUT&RUN (SEACR³⁸, <https://seacr.fredhutch.org/>) was used for peak calling. H3K4me3 and BRD2 peaks were normalized to IgG control. Published BRD2 ChIP-seq data in human lung cells were obtained from ChIP-Atlas (<https://chip-atlas.org/>). The Integrative Genomics Viewer (IGV, igv.org) was used for visualization.

SARS-CoV-2 infection of reconstructed human nasal epithelia. MucilAir (<https://www.epithelix.com/products/mucilair>), corresponding to reconstructed human nasal epithelium cultures differentiated in vitro for at least four weeks, was purchased from Epithelix. Epithelix obtains donor consent and this study does not involve any human participants. The cultures were generated from pooled nasal tissues obtained from 14 human adult donors. Cultures were maintained in air/liquid interface (ALI) conditions in transwells with 700 μ l of MucilAir medium (Epithelix) in the basal compartment, and kept at 37°C under a 5% CO₂ atmosphere. SARS-CoV-2 infection was performed as previously described³⁷. Briefly, the apical side of ALI cultures was washed for 20 min at 37°C in Mucilair medium (+/- drug) to remove mucus. Cells were then incubated with 10^4 p.f.u. of the isolate BetaCoV/France/IDF00372/2020 (EVAg collection, Ref-SKU 014V-03890; provided by S. Van der Werf). The viral input was diluted in DMEM medium (+/- drug) to a final volume 100 μ l, and left on the apical side for 4 h at 37°C. Control wells were mock-treated with DMEM (Gibco) for the same duration. Viral inputs were removed by washing twice with 200 μ l of PBS (5 min at 37°C) and once with 200 μ l of Mucilair medium (20 min at 37°C). The basal medium was replaced every 2–3 days. Apical supernatants were collected every 2–3 days by adding 200 μ l of Mucilair medium on the apical side, with an incubation of 20 min at 37°C before collection.

For ABBV-774 treatment, cultures were pretreated for four days with 100 nM or 300 nM of the drug. For this pretreatment, ABBV-774 was added to an apical wash at day -4 and to the basal compartment from day -4 to day 0. The drug was then added on the apical side during viral adsorption at day 0, and then every 2–3 days to both the apical wash and the basal compartment throughout the infection.

Transepithelial electrical resistance measurements. The apical side of transwell cultures was washed for 20 min at 37°C in Mucilair medium. The transwells were then transferred in a new 24-well plate and DMEM was added to both the apical (200 μ l) and basal (700 μ l) sides. The transepithelial electrical resistance was then measured using an Evom3 ohm meter (World Precision Instruments).

Lactate dehydrogenase cytotoxicity assay. Diluted culture supernatants (1:25) were pretreated with Triton-X100 1% for 2 h at room temperature for viral inactivation. Lactate dehydrogenase (LDH) dosage was performed using the LDH-Glo cytotoxicity assay kit (Promega) following the manufacturer's instructions. Luminescence was measured using an EnSpire luminometer (Perkin Elmer).

Viral RNA quantification. Apical supernatants were stored at -80°C until thawing and were diluted fourfold in PBS for quantification in a 96-well PCR plate. Supernatants were then inactivated for 20 min at 80°C. One microlitre of supernatant was added directly to 4 μ l of PCR reaction mix for SARS-CoV-2 RNA quantification. PCR was carried out in a final volume of 5 μ l per reaction in 384-well plates using the Luna Universal Probe One-Step RT-qPCR kit (New England Biolabs) with SARS-CoV-2 N-specific primers (forward, 5'-TAA TCA GAC AAG GAA CTG ATT A-3'; reverse, 5'-CGA AGG TGT GAC TTC CAT G-3') on a QuantStudio 6 Flex thermocycler (Applied Biosystems). Standard curves were established in parallel using purified SARS-CoV-2 viral RNA.

Tissue RNA quantification. ACE2 and SARS-CoV-2 expression levels were quantified in epithelial cells by real-time qPCR. The epithelial cultures were washed in ice-cold PBS and then lysed in 150 μ l of Trizol reagent (Thermo Fisher Scientific) added to the apical side of the insert for 5 min. RNA was purified using the Direct-zol miniprep kit (ZR2080, Zymo Research). Transcripts of genes of interest (ACE2, SARS-CoV-2 N gene) were amplified in a final volume of 5 μ l per reaction in 384-well plates using the Luna Universal Probe One-Step RT-qPCR kit (New England Biolabs) on a QuantStudio 6 Flex thermocycler. The RT-qPCR results were normalized to the mean expression of four reference genes (GAPDH, TFR3, ALAS1 and RPL13) to compute relative gene expression, as described previously³⁷. The ACE2 primers used were ACE2-For 5'-TGG GAC TCT GCC ATT TAC TTA C-3' and ACE2-Rev 5'-CCA GAG CCT CTC ATT GTA GTC T-3'.

In vivo infections. All animal infections were conducted at the Icahn School of Medicine at Mount Sinai, in the biosecurity level 3 (BSL-3) facility of the Global Health and Emerging Pathogen Institute approved by the Institutional Animal Care and Use Committee at Icahn School of Medicine at Mount Sinai under protocol number IACUC#20-0743. Six- to eight-week-old male golden Syrian hamsters (*Mesocricetus auratus*) were purchased from Jackson Laboratories, housed in pairs and fed ad libitum. On the day of infection, animals were anesthetized by administration of 100 μ l of a ketamine HCl/xylazine (4:1) mix by intraperitoneal injection and infected intranasally with 1,000 p.f.u. of SARS-CoV-2 USA-WA1/2020 diluted in 100 μ l of PBS. At the indicated time points, animals were treated with 1 ml of ABBV-774 by oral gavage. The drug was prepared fresh daily to a final concentration of 20 mg kg⁻¹ in 0.5% hydroxypropyl methylcellulose/0.5% Tween 80 in water. On day 3 after infection, animals received 100 μ l of a mix of pentobarbital/PBS (1:4) intraperitoneally and, once anaesthetized they were cervically dislocated and lung lobes collected in 1 ml of Trizol reagent. Tissues were homogenized in a tissue lyser for two cycles of 40 s, spun down for 5 min at 8,000g, and the supernatants were stored at -80°C for plaque assay or RNA extraction.

RNA extraction. RNA extraction was performed following the instructions from the manufacturer of TRIzol reagent (Invitrogen). Briefly, a 1/5 volume of chloroform was added to the lung supernatants in TRIzol, phases were separated by centrifugation and RNA was precipitated by overnight incubation with isopropanol at -20°C. The RNA pellet was washed with ethanol 70% and resuspended in RNase-free water. RNA was quantified by NanoDrop and resuspended to a final concentration of 100 ng μ l⁻¹ in water.

RNA-seq. One microgram of RNA was used as the starting material for library preparation. The kit employed was a TruSeq RNA Library Prep kit v2 (Illumina) over polyadenylated RNA and the manufacturer's instructions were followed. The sequencing was performed on an Illumina NextSeq 500 instrument. The raw reads obtained from the run were aligned against the Syrian golden hamster genome (MesAur1.0) in the BaseSpace platform by Illumina, with the tool 'RNA-Seq Alignment'.

Statistics and reproducibility. All experiments were carried out using standard cell biological/biochemical techniques with pre-validated reagents including

cell lines. Accordingly, the minimum standard of at least three replicates per experiment was used for all experiments, with the following exceptions. In Fig. 1a,b we report the dose responses of different cell types to Spike protein where each dose response has data points for seven doses. These experiments were performed once. In Extended Data Fig. 1 we report a dose response with data points for two doses. These experiments were performed once. In Extended Data Fig. 2 we report measured dose responses where each dose response was performed in duplicate and with data points for seven to eight doses. Experiments that were performed with $n = 1$ are reported in Figs. 1c,e,f, 2c and 3b. No statistical method was used to predetermine sample size. No data were excluded from analyses. Experiments were not randomized. Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

Sequencing data are available from the NCBI Gene Expression Omnibus (GEO) with the following accession numbers: GSE165025 (RNA-sequencing data associated with Fig. 4), GSE182993 (CUT&RUN data associated with Fig. 5) and GSE182994 (RNA-sequencing data associated with Fig. 6f–h). Previously published BRD2 ChIP-seq data that were re-analysed here are available under accession codes GSE113714 and GSE104481. All other data supporting the findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

Code availability

Analysis of the CRISPRi screen results was carried out using custom code (MAGeCK-iNC) developed in the Kampmann laboratory. This has been described previously⁴⁹ and is freely available at <https://kampmannlab.ucsf.edu/mageck-inc>.

References

- Tian, R. et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* **104**, 239–255 (2019).
- Stoner, R., Maures, T. & Conant, D. Methods and systems for guide RNA design and use US patent application 16/418,893 (2019).
- Hsiao, T. et al. Inference of CRISPR edits from Sanger trace data. Preprint at *bioRxiv* <https://doi.org/10.1101/251082> (2019).
- Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
- Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
- Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
- Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by sparse enrichment analysis for CUT&RUN chromatin profiling. *Epigenet. Chromatin* **12**, 42 (2019).

Acknowledgements

We thank members of the Kampmann, Vignuzzi and Conklin laboratories, as well as V. Ramani, D. Ruggero, M. Ott and her laboratory and other members of the UCSF QBI Coronavirus Research Group (QCRG) for helpful discussions. We thank K. Leng for feedback on the manuscript. We acknowledge the Gladstone Stem Cell Core for help with cardiomyocyte production. A.J.S. is supported by NIH grant F32AG063487. G.N.R. is supported by the NSF Graduate Research Fellowship Program (GRFP) and UCSF Discovery Fellowship. S.A.L. was a Merck Fellow of the Helen Hay Whitney Foundation. I.L. was supported by an NSF GRFP award. M.K. is a Chan Zuckerberg Biohub Investigator.

Author contributions

R.T., A.J.S. and M.K. conceptualized the overall project, analysed the results and prepared the manuscript, with input from all co-authors. V.V.R., A.M.K. and Q.D.T. performed and analysed live-virus experiments in Calu-3 cells, with guidance from M.V. R.R. performed and analysed human nasal epithelia experiments with guidance from L.A.C. L.C. performed and analysed the Syrian hamster experiments, with guidance from B.R.T. G.N.R. and S.J.R. performed and analysed the experiments with cardiomyocytes, with guidance from B.R.C. R.T., A.J.S., M.C. and X.G. performed and analysed all other experiments, with guidance from M.K. J.W. performed and analysed basal interferon signalling knockdown experiments in Calu-3 cells, with guidance from R.T. N.L. analysed the QuantSeq data, with guidance from R.T. S.A.L., I.L. and J.A.W. generated Spike-RBD. J.K.N. and J.S.W. generated the Calu-3 CRISPRi cell line. J.C.-S., J.O., T.M. and K.H. designed and provided sgRNAs to generate the ACE2 knockout cell line.

Competing interests

J.C.-S., J.O., T.M. and K.H. are employees and shareholders of Synthego Corporation. All other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41556-021-00821-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-021-00821-8>.

Correspondence and requests for materials should be addressed to Ruilin Tian or Martin Kampmann.

Peer review information *Nature Cell Biology* thanks Andrew Bowie, Ke Lan and the other, anonymous, reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

3. Identification of DAXX as a restriction factor of SARS-CoV-2 through a CRISPR/Cas9 screen

Alice Mac Kain, Ghizlane Maarifi, Sophie-Marie Aicher, Nathalie Arhel, Artem Baidaliuk, Sandie Munier, Flora Donati, Thomas Vallet, **Quang Dinh Tran**, Alexandra Hardy, Maxime Chazal, Françoise Porrot, Molly OhAinle, Jared Carlson-Stevermer, Jennifer Oki, Kevin Holden, Gert Zimmer, Etienne Simon-Lorière, Timothée Bruel, Olivier Schwartz, Sylvie van der Werf, Nolwenn Jouvenet, Sébastien Nisole, Marco Vignuzzi & Ferdinand Roesch

DOI: <https://doi.org/10.1038/s41467-022-30134-9>

Contributions: I have helped with experiments and analysis of Figure 2a-c and 3a-c.

Identification of DAXX as a restriction factor of SARS-CoV-2 through a CRISPR/Cas9 screen

Alice Mac Kain ^{1,12}, Ghizlane Maarifi^{2,12}, Sophie-Marie Aicher³, Nathalie Arhel ², Artem Baidaliuk ⁴, Sandie Munier^{5,6}, Flora Donati ^{5,6}, Thomas Vallet ¹, Quang Dinh Tran ¹, Alexandra Hardy ¹, Maxime Chazal ³, Françoise Porrot ⁷, Molly OhAinle ⁸, Jared Carlson-Stevermer ⁹, Jennifer Oki⁹, Kevin Holden ⁹, Gert Zimmer ¹⁰, Etienne Simon-Lorière ⁴, Timothée Bruel ⁷, Olivier Schwartz ⁷, Sylvie van der Werf ^{5,6}, Nolwenn Jouvenet ³ , Sébastien Nisole ² , Marco Vignuzzi ¹  & Ferdinand Roesch ^{1,11} 

Interferon restricts SARS-CoV-2 replication in cell culture, but only a handful of Interferon Stimulated Genes with antiviral activity against SARS-CoV-2 have been identified. Here, we describe a functional CRISPR/Cas9 screen aiming at identifying SARS-CoV-2 restriction factors. We identify DAXX, a scaffold protein residing in PML nuclear bodies known to limit the replication of DNA viruses and retroviruses, as a potent inhibitor of SARS-CoV-2 and SARS-CoV replication in human cells. Basal expression of DAXX is sufficient to limit the replication of SARS-CoV-2, and DAXX over-expression further restricts infection. DAXX restricts an early, post-entry step of the SARS-CoV-2 life cycle. DAXX-mediated restriction of SARS-CoV-2 is independent of the SUMOylation pathway but dependent on its D/E domain, also necessary for its protein-folding activity. SARS-CoV-2 infection triggers the re-localization of DAXX to cytoplasmic sites and promotes its degradation. Mechanistically, this process is mediated by the viral papain-like protease (PLpro) and the proteasome. Together, these results demonstrate that DAXX restricts SARS-CoV-2, which in turn has evolved a mechanism to counteract its action.

¹Institut Pasteur, Université de Paris Cité, CNRS UMR 3569, Viral populations and pathogenesis Unit, F-75015 Paris, France. ²Institut de Recherche en Infectiologie de Montpellier (IRIM), Université de Montpellier, CNRS, 34090 Montpellier, France. ³Institut Pasteur, Université de Paris Cité, CNRS UMR 3569, Virus sensing and signaling Unit, F-75015 Paris, France. ⁴Institut Pasteur, G5 Evolutionary genomics of RNA viruses, F-75015 Paris, France. ⁵Institut Pasteur, Université de Paris Cité, CNRS UMR 3569, Molecular Genetics of RNA Viruses Unit, F-75015 Paris, France. ⁶Institut Pasteur, CNR Virus des infections respiratoires, F-75015 Paris, France. ⁷Institut Pasteur, Université de Paris Cité, CNRS UMR 3569, Virus and Immunity, F-75015 Paris, France. ⁸Divisions of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁹Synthego Corporation, 3565 Haven Avenue, Menlo Park, CA 94025, USA. ¹⁰Institute of Virology and Immunology, Bern & Mithelhäusern, Switzerland, and Department of Infectious Diseases and Pathobiology, Vetsuisse Faculty, University of Bern, Bern, Switzerland. ¹¹UMR 1282 ISP, INRAE Centre Val de Loire, Nouzilly, France. ¹²These authors contributed equally: Alice Mac Kain, Ghizlane Maarifi. email: nolwenn.jouvenet@pasteur.fr; sebastien.nisole@inserm.fr; marco.vignuzzi@pasteur.fr; ferdinand.roesch@inrae.fr

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19 and the third coronavirus to cause severe disease in humans after the emergence of SARS-CoV in 2002 and Middle East Respiratory Syndrome-related Coronavirus (MERS-CoV) in 2012. Since the beginning of the pandemic, SARS-CoV-2 has infected more than 500 million people and claimed at least 6 million lives. While the majority of infected individuals experience mild (or no) symptoms, severe forms of COVID-19 are associated with respiratory failure, shock and pneumonia. Innate immune responses play a key role in COVID-19 pathogenesis: immune exhaustion¹ and reduced levels of type-I and type-III interferons (IFN) have been observed in the plasma of severe COVID-19 patients^{2,3}. Imbalanced immune responses to SARS-CoV-2, with a low and delayed IFN response coupled to early and elevated levels of inflammation, have been proposed to be a major driver of COVID-19^{4,5}. Neutralizing auto-antibodies against type-I IFN⁶ and genetic alterations in several IFN pathway genes⁷ have also been detected in critically ill COVID-19 patients. These studies highlight the crucial need to characterize the molecular mechanisms by which IFN effectors may succeed, or fail, to control SARS-CoV-2 infection.

Although SARS-CoV-2 has been described to antagonize the IFN pathway by different mechanisms involving the viral proteins ORF3b, ORF9b, ORF6, and Nsp15⁸, detection of SARS-CoV-2 by the innate immune sensor MDA5^{9,10} leads to the synthesis of IFN and expression of IFN Stimulated Genes (ISGs) in human airway epithelial cells⁴. IFN strongly inhibits SARS-CoV-2 replication when added in cell culture prior to infection^{11,12} or when administered intranasally in hamsters¹³, suggesting that some ISGs might have antiviral activity¹⁴. Relatively few ISGs with antiviral activity against SARS-CoV-2, however, have been identified so far. For instance, spike-mediated viral entry and fusion is restricted by LY6E^{15,16} and IFITMs^{17,18}. Mucins have also been suggested to restrict viral entry¹⁹. ZAP, which targets CpG dinucleotides in RNA viruses, also restricts SARS-CoV-2, albeit moderately²⁰. OAS1 has been recently identified in an ISG overexpression screen to restrict SARS-CoV-2 replication, through the action of RNaseL, both in cell lines and in patients²¹. Another overexpression screen identified 65 ISGs as potential inhibitors of SARS-CoV-2²², and found that BST-2 is able to restrict viral budding, although this activity is counteracted by the viral protein ORF7a. We hypothesize that additional ISGs with antiviral activity against SARS-CoV-2 remain to be discovered. Other antiviral factors that are not induced by IFN may also inhibit SARS-CoV-2: for instance, the RNA helicase DDX42 restricts several RNA viruses, including SARS-CoV-2²³. While several whole-genome CRISPR/Cas9 screens identified host factors required for SARS-CoV-2 replication^{24–29}, none focused on antiviral genes.

In this work, we performed a CRISPR/Cas9 screen designed to identify restriction factors for SARS-CoV-2, assessing the ability of 1905 ISGs to modulate SARS-CoV-2 replication in human epithelial lung cells. We report that the Death domain-associated protein 6 (DAXX), a scaffold protein residing in PML nuclear bodies³⁰ and restricting DNA viruses³¹ as well as retroviruses^{32,33}, is a potent inhibitor of SARS-CoV-2 replication. SARS-CoV-2 restriction by DAXX is largely independent of the action of IFN, and unlike most of its other known activities, of the SUMOylation pathway. Within hours of infection, DAXX relocalizes to sites of viral replication in the cytoplasm, targeting an early, post-entry step of the viral life cycle such as viral transcription or uncoating. We show that the SARS-CoV-2 papain-like protease (PLpro) induces the proteasomal degradation of DAXX, demonstrating that SARS-CoV-2 developed a mechanism to evade, at least partially, the restriction imposed by DAXX.

Table 1 Gene editing efficiency.

Gene	% of alleles edited
LY6E	96 ± 1.73
DAXX	79,67 ± 2.52
APOL6	99 ± 0
HERC5	97 ± 0
CTSL	91 ± 1
IFI6	88,33 ± 0.58
IFNAR1	76,67 ± 3.21

The frequency of editing was determined using Sanger sequencing and ICE analysis. Values are represented as mean ± SD (n = 3).

Results

A restriction factor-focused CRISPR/Cas9 screen identifies genes potentially involved in SARS-CoV-2 inhibition. To identify restriction factors limiting SARS-CoV-2 replication, we generated a pool of A549-ACE2 cells knocked-out (KO) for 1905 potential ISGs, using the sgRNA library we previously developed to screen HIV-1 restriction factors³⁴. This library includes more ISGs than most published libraries, as the inclusion criteria was less stringent (fold-change in gene expression in THP1 cells, primary CD4 + T cells or PBMCs ≥ 2). Therefore, some genes present in the library may not be ISGs per se in A549 cells. Transduced cells were selected by puromycin treatment, treated with IFN α and infected with SARS-CoV-2. Infected cells were immuno-labelled with a spike (S)-specific antibody and analyzed by flow cytometry. As expected^{11,12}, IFN α inhibited infection by sevenfold (Fig. S1). Infected cells were sorted based on S expression (Fig. 1a), and DNA was extracted from infected and non-infected control cells. Integrated sgRNA sequences in each cell fraction were amplified by PCR and sequenced by Next Generation Sequencing (NGS). Statistical analyses using the MAGeCK package³⁵ led to the identification of sgRNAs significantly enriched or depleted in infected cells representing antiviral and proviral genes, respectively (Fig. 1b). Although our screen was not designed to study proviral genes, we did successfully identify the well-described SARS-CoV-2 co-factor cathepsin L (CTSL)³⁶, validating our approach. *USP18*, a gene encoding a negative regulator of the IFN signaling pathway³⁷, and *ISG15*, which favors Hepatitis C Virus replication³⁸, were also identified as proviral ISGs. Core IFN pathway genes such as those encoding for the IFN receptor (*IFNAR1*), *STAT1*, and *STAT2*, were detected as antiviral factors, further validating our screening strategy. *LY6E*, a gene previously described to encode an inhibitor of SARS-CoV-2 entry^{15,16}, was also a significant hit. Moreover, our screen identified *APOL6*, *IFI6*, *DAXX* and *HERC5*, genes that are known to encode proteins with antiviral activity against other viruses^{39–42}, but had not previously been studied in the context of SARS-CoV-2 infection. For all these genes except *APOL6*, individual sgRNAs were consistently enriched (for antiviral factors) or depleted (for proviral factors) in the sorted population of infected cells, while non-targeting sgRNAs were not (Fig. 1c).

LY6E and DAXX display antiviral activity against SARS-CoV-2.

To validate the ability of the identified hits to modulate SARS-CoV-2 replication in human cells, we generated pools of A549-ACE2 knocked-out (KO) cells for different genes of interest by electroporating a mix of 3 sgRNA/Cas9 ribonucleoprotein (RNP) complexes per gene target. Levels of gene editing were above 80% in all of the A549-ACE2 KO cell lines, as assessed by sequencing of the edited *loci* (Table 1). As controls, we used cells KO for *IFNAR1*, for the proviral gene *CTSL* or for the antiviral gene *LY6E*, as well as cells

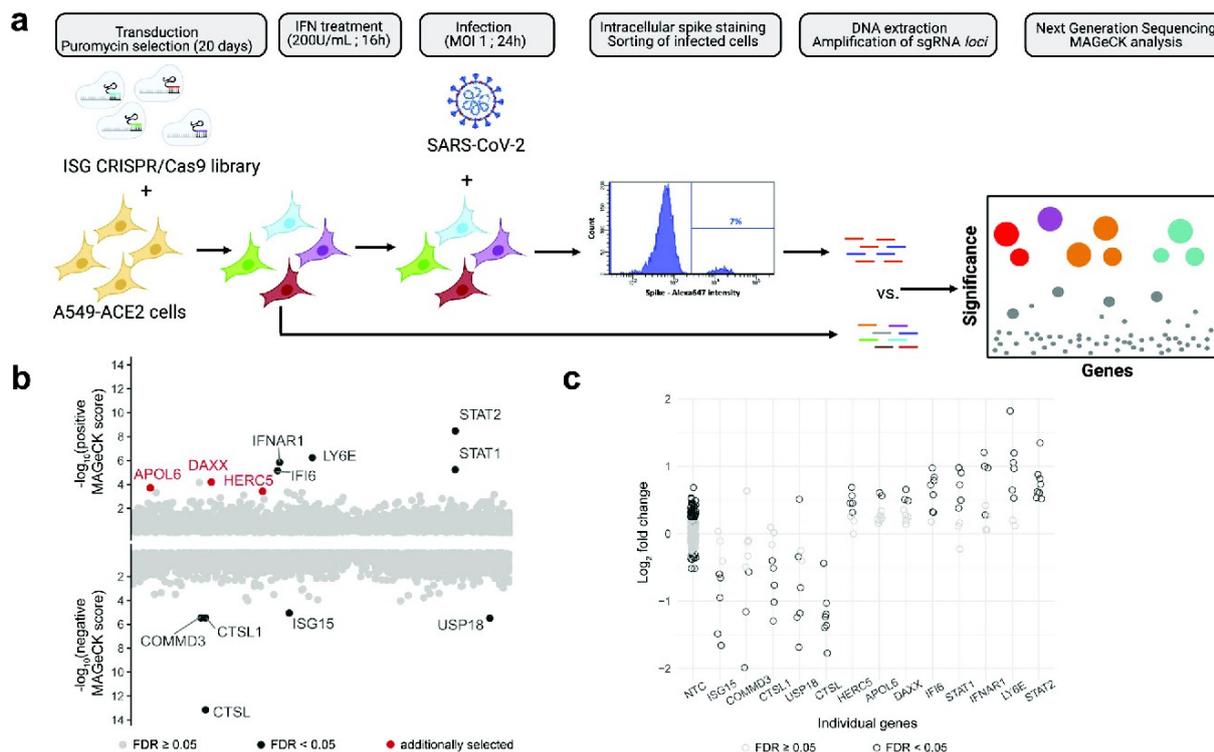


Fig. 1 ISG-focused CRISPR/Cas9 screening approach to identify restriction factors for SARS-CoV-2. **a** CRISPR/Cas9 screen outline. A549-ACE2 cells were transduced with lentivectors encoding the ISG CRISPR/Cas9 library and selected by puromycin treatment for 20 days. Library cells were then pre-treated with 200 U/mL of IFN α for 16 h, and infection with SARS-CoV-2 at an MOI of 1. At 24 h p.i., infected cells were fixed with formalin treatment, permeabilized by saponin treatment and stained with a monoclonal anti-spike antibody. After secondary staining, infected cells were sorted and harvested. Non-infected, non-IFN α treated cells were harvested as a control. DNA was extracted from both cellular fractions and sgRNA *loci* amplification was carried out by PCR. Following NGS, bio-informatic analysis using the MAGECK package was conducted. This figure was created with BioRender.com. **b** Screen results. By taking into account the enrichment ratios of each of the 8 different sgRNAs for every gene, the MAGECK analysis provides a positive score for KO enriched in infected cells (i.e. restriction factor, represented in the top fraction of the graph) and a negative score for KO depleted in infected cells (i.e. proviral factors, represented in the bottom portion of the graph). Genes with an FDR < 0.05 are represented in black. 3 genes with a FDR > 0.05, but with a *p* value < 0.005 were additionally selected and are represented in red. **c** Individual sgRNA enrichment. For the indicated genes, the enrichment ratio of the 8 sgRNAs present in the library was calculated as the MAGECK normalized read counts in infected cells divided by those in the original pool of cells and is represented in log₂ fold change. As a control, the enrichment ratio of the 200 non-targeting control (NTCs) is also represented.

electroporated with non-targeting sgRNAs/Cas9 RNPs (indicated here as WT). These different cell lines were then treated with IFN α and infected with SARS-CoV-2. Viral replication was assessed by measuring the levels of viral RNA in the supernatant of infected cells using RT-qPCR (Fig. 2a). In parallel, we titrated the levels of infectious viral particles released into the supernatant of infected cells (Fig. 2b). As expected, infection was significantly reduced in *CTSL* KO cells, confirming the proviral effect of this gene³⁶. Among the selected antiviral candidate genes, only 2 had a significant impact on SARS-CoV-2 replication: *LY6E* (as expected), and to an even greater degree, *DAXX*. Both genes restricted replication in absence of IFN α , an effect which was detectable at the level of viral RNA (8-fold and 42-fold reduction of infection, respectively, Fig. 2a) and of infectious virus (15-fold and 62-fold reduction, Fig. 2b). Based on available single-cell RNAseq datasets,⁴³ *DAXX* is expected to be expressed in cell types physiologically relevant for SARS-CoV-2 infection such as lung epithelial cells and macrophages (Fig. S2).

In IFN α -treated cells, *DAXX* and *LY6E* KO led to a modest, but significant rescue of viral replication, which was particularly visible when measuring the levels of infectious virus by plaque assay titration (Fig. 2b), while the antiviral effect of IFN α treatment was completely abrogated in *IFNAR1* KO cells, as expected (Fig. 2c). However, IFN α still had robust antiviral effect on SARS-CoV-2 replication in both *DAXX* KO and *LY6E* KO cells (Fig. 2c). This

suggests that other ISGs likely contribute to the anti-SARS-CoV-2 IFN response. *DAXX* is sometimes referred to as an ISG, and was originally included in our ISG library, although its expression is only weakly induced by IFN in some human cell types^{32,44}. Consistent with this, we found little to no increase in *DAXX* gene expression in IFN α -treated A549-ACE2 cells (Fig. S3). In addition, we tested the antiviral effect of *DAXX* on several SARS-CoV-2 variants that have been suggested to be partially resistant to the antiviral effect of IFN in A549-ACE2 cells⁴⁵. Our results confirmed that Lineage B.1.1.7. (Alpha) and Lineage P1 (Gamma) SARS-CoV-2 variants were indeed less sensitive to IFN (Fig. 2d). *DAXX*, however, restricted all variants to a similar level than the original Lineage B strain of SARS-CoV-2 (Fig. 2d), suggesting that while some variants may have evolved towards IFN-resistance, they are still efficiently restricted by *DAXX*. To determine whether *DAXX* restriction is specific to SARS-CoV-2 or also inhibits other RNA viruses, including coronaviruses, A549-ACE2 WT and *DAXX* KO cells were infected with SARS-CoV, MERS-CoV, and 2 RNA viruses belonging to unrelated families: Yellow Fever Virus (YFV) and Measles Virus (MeV), which are positive and negative strand RNA viruses, respectively. *DAXX* restricted SARS-CoV, but had no effect on the replication of YFV, MeV or MERS-CoV (Fig. 2e, f). Thus, our data suggests that *DAXX* restriction may exhibit some level of specificity.

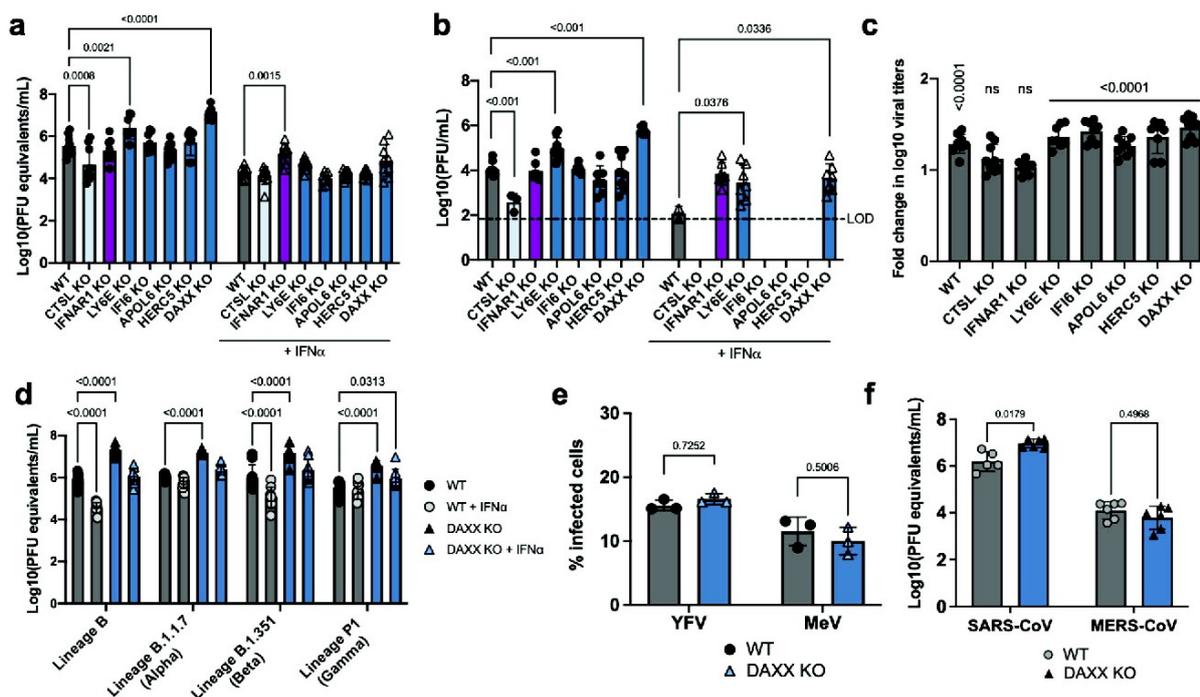


Fig. 2 DAXX is a restriction factor for SARS-CoV-2. **a–c** Antiviral activity of ISGs against SARS-CoV-2. A549-ACE2 knocked-out for the indicated genes were generated using a multi-guide approach, leading to pools of KO cells with a high frequency of indels. KO cells were pre-treated with 0 (circles) or 200 (triangles) U/mL of IFN α 24 h prior to infection with SARS-CoV-2 (at an MOI of 0.1). Supernatants were harvested at 72 h p.i. The mean \pm SD of three independent experiments, each performed in three biological replicates, is shown. **a** For the titration of RNA levels, supernatants were heat-inactivated prior to quantification by qRT-PCR. Genome copies/mL were calculated by performing serial dilutions of a synthetic RNA with a known concentration. Statistics: 2-way ANOVA using Dunnett’s test. Significant *p* values (below 0.05) are indicated on the graph. **b** For the titration of infectious virus levels by plaque assay, supernatants were serially diluted and used to infect VeroE6 cells. Plaques formed after 3 days of infection were quantified using crystal violet coloration. The limit of detection (LOD) is indicated as a dotted line. Statistics: Dunnett’s test on a linear model, (two-sided). Significant *p* values (below 0.05) are indicated on the graph. **c** For each of the indicated KO, the data shown in (**a**) is represented as fold change in log₁₀ titers (i.e. the log₁₀ titers of the non-treated condition divided by the mean of the triplicate log₁₀ titers IFN α -treated condition, *n* = 3). Statistics: 2-way ANOVA using Sidak’s test. *P* values are indicated on the graph (ns: *p* value > 0.05). **d–f** Antiviral activity of DAXX against SARS-CoV-2 variants and other viruses. **d** A549-ACE2 WT or DAXX KO cells were infected at an MOI of 0.1 with the following SARS-CoV-2 strains: Lineage B (original strain); Lineage B.1.1.7. (Alpha variant); Lineage B.1.351.1 (Beta variant); Lineage P1 (Gamma variant). Supernatants were harvested at 72 h p.i. Supernatants were heat-inactivated prior to quantification by qRT-PCR. Genome copies/mL were calculated by performing serial dilutions of a synthetic RNA with a known concentration. The mean \pm SD of three independent experiments, with infections carried out in three biological replicates, is shown. Statistics: 2-way ANOVA using Dunnett’s test. Significant *p* values (below 0.05) are indicated on the graph. **e** A549-ACE2 WT or DAXX KO cells were infected with Yellow Fever Virus (YFV, Asibi strain, MOI of 0.3) or with Measles Virus (MeV, Schwarz strain expressing GFP, MOI of 0.2). At 24 h p.i., the percentages of cells positive for viral protein E (YFV) or GFP (MeV) was assessed by flow cytometry. The mean \pm SD of 3 independent experiments is represented. Statistics: 2-way ANOVA using Sidak’s test. *P* values are indicated on the graph. **f** WT or DAXX KO cells were infected at an MOI of 0.1 with SARS-CoV or MERS-CoV. Supernatants were harvested at 72 h p.i. Supernatants were heat inactivated prior to quantification by qRT-PCR. Serial dilutions of a stock of known infectious titer was used as a standard. The mean \pm SD of 2 independent experiments, with infections carried out in three biological replicates, is represented. Statistics: 2-way ANOVA using Dunnett’s test. *P* values are indicated on the graph. Source data are provided as a Source Data file.

DAXX targets an early post-entry step. Next, we investigated which steps of the SARS-CoV-2 replication cycle were targeted by DAXX. To assess whether DAXX affects SARS-CoV-2 Spike-mediated entry, we took advantage of a replication-competent Vesicular Stomatitis Virus expressing GFP (VSV*) and carrying the Spike protein instead of its G envelope (VSV* Δ G-S). This approach allows to study Spike-mediated viral entry without relying on lentiviral pseudotypes, which are likely to be targeted by DAXX³³. We first ensured that DAXX was not affecting VSV replication in A549-ACE2 cells using the VSV* control virus (Fig. 3a). Flow cytometric analysis revealed that VSV* Δ G-S replicated at similar levels in WT and DAXX-KO cells (Fig. 3a), suggesting that DAXX does not inhibit the entry steps that are mediated by SARS-CoV-2 Spike.

To determine whether DAXX targets viral transcription, A549-ACE2 WT or DAXX KO cells were infected with SARS-CoV-2, and the intracellular levels of viral RNA were assessed at different time post-infection (Fig. 3c, d). At early time points (from 2 h to 6 h p.i.), the levels of viral RNA transcripts were similar in WT and DAXX KO cells, further suggesting that comparable amounts of SARS-CoV-2 RNA were entering cells in both cell lines. The levels of viral transcripts significantly increased starting at 8 h p.i., representing the initiation of viral transcription. The levels of viral RNA as detected by amplification of the 5’ UTR (Fig. 3c) were 6.4-fold higher at 8 h; 4.1-fold higher at 10 h; and eightfold higher at 24 h post-infection in DAXX KO cells compared to WT cells. We observed a similar effect when using primers amplifying the RdRp region (Fig. 3d) with levels of viral transcripts 1.7-fold and 3.5-fold

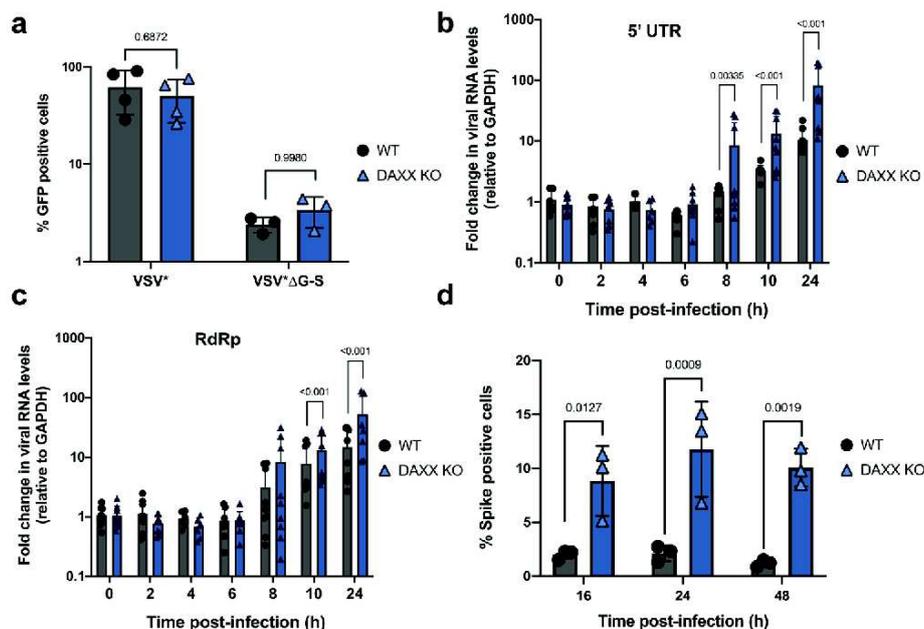


Fig. 3 DAXX inhibits SARS-CoV-2 at an early post-entry step. **a** DAXX has no effect on Spike-mediated entry. A549-ACE2 WT or DAXX KO were infected with the GFP reporter viruses VSV* or VSV*ΔG-S. Cell monolayers were harvested at 16 h p.i. and analyzed by flow cytometry. The mean of 3 independent experiments \pm SD is shown. Statistics: 2-way ANOVA using Sidak's test. *P* values are indicated on the graph. **b, c** DAXX inhibits viral RNA synthesis. A549-ACE2 WT or DAXX KO were infected at an MOI of 1. Cell monolayers were harvested at the indicated time points, and total RNA was extracted. The levels of viral RNA (**c**: 5' UTR; **d**: RdRp) were determined by qRT-PCR and normalized against GAPDH levels. The mean \pm SD of 3 independent experiments, with infections carried out in three biological replicates, is represented. Statistics: Dunnett's test on a linear model (two-sided). Significant *p* values (below 0.05) are indicated on the graph. **d** DAXX inhibits viral protein synthesis. A549-ACE2 WT or DAXX KO were infected at an MOI of 2. Cell monolayers were harvested at the indicated time points and Spike levels were quantified by flow cytometry. The mean of 3 independent experiments \pm SD is represented. Statistics: 2-way ANOVA using Sidak's test. *P* values are indicated on the graph. Source data are provided as a Source Data file.

higher in DAXX KO cells compared to WT cells at 10 h and 24 h post-infection, respectively.

Finally, we measured the levels of Spike protein synthesis at different times post-infection by flow cytometry in A549-ACE2 WT or DAXX KO cells. In agreement with the observed effect of DAXX KO on viral RNA synthesis (Fig. 3c, d), we observed a 4- to 8-fold increase in the intracellular levels of Spike (Fig. 3e). Together, these results suggest that while DAXX has no effect on viral entry, it restricts a post-entry step of the viral life cycle such as viral transcription or uncoating.

DAXX restriction is mediated by its D/E domain but is SUMOylation-independent. DAXX encodes a small scaffold protein that acts by recruiting other SUMOylated proteins in nuclear bodies through its C-terminal SUMO-Interacting Motif (SIM) domain⁴⁶. The recruitment of these factors is required for the effect of DAXX on various cellular processes such as transcription and apoptosis, and on its antiviral activities^{32,47–49}. DAXX can also be SUMOylated itself⁵⁰, which may be important for some of its functions. To investigate the role of SUMOylation in DAXX-mediated SARS-CoV-2 restriction, overexpression assays using WT and two previously described mutated versions of DAXX⁵¹ were performed (Fig. 4a). Fifteen lysine residues have been mutated to arginine in the first mutant (DAXX 15KR), which is unable to be SUMOylated. The second mutant is a truncated version of DAXX that lacks its C-terminal SIM domain (DAXXΔSIM)⁴⁸ and is thus unable to interact with its SUMOylated partners. A549-ACE2 were refractory to SARS-CoV-2 infection upon transfection with any plasmid, precluding us from using this cell line. The experiments were performed in

293T-ACE2 cells, which are permissive to SARS-CoV-2¹⁸ and easy to transfect. In order to quantify the antiviral effect of overexpressed DAXX WT and mutants, we assessed the number of transfected cells that were positive for the Spike protein by flow cytometry. Western blot (Fig. S4a) and flow cytometry (Fig. S4b) analyses showed that DAXX WT and mutants were expressed at similar levels, with around 40 to 50% cells expressing the HA-tagged constructs. DAXX WT, 15KR and ΔSIM efficiently restricted SARS-CoV-2 replication. Indeed, at 24 h p.i., the proportion of infected cells (among HA-positive cells) was reduced by 2 to 3-fold as compared to control transfected cells (Fig. 4b). This effect was less pronounced but still significant at 48 h p.i. (Fig. 4c). Moreover, DAXX overexpression led to a significant reduction of the levels of intracellular viral RNA (Fig. S5), in line with our earlier results showing that DAXX targets viral transcription (Fig. 3c, d). Together, these results show that DAXX overexpression restricts SARS-CoV-2 replication in a SUMOylation-independent mechanism.

DAXX was recently described as a protein chaperone able to solubilize protein aggregates and unfold misfolded proteins⁵². We investigated whether this activity was required for SARS-CoV-2 restriction using a DAXX mutant lacking the D/E domain, critical for this chaperone activity⁵². The DAXXΔD/E mutant, while expressed at similar levels than WT DAXX in transfected cells (Fig. S4c), was unable to restrict SARS-CoV-2, as assessed by Western Blot analysis on Spike levels (Fig. 4d) and RT-qPCR analysis detecting intracellular SARS-CoV-2 RNA (Fig. 4e). Taken together, these results suggest that DAXX dampens SARS-CoV-2 replication through a SUMOylation-independent mechanism that likely involves its chaperone activity.

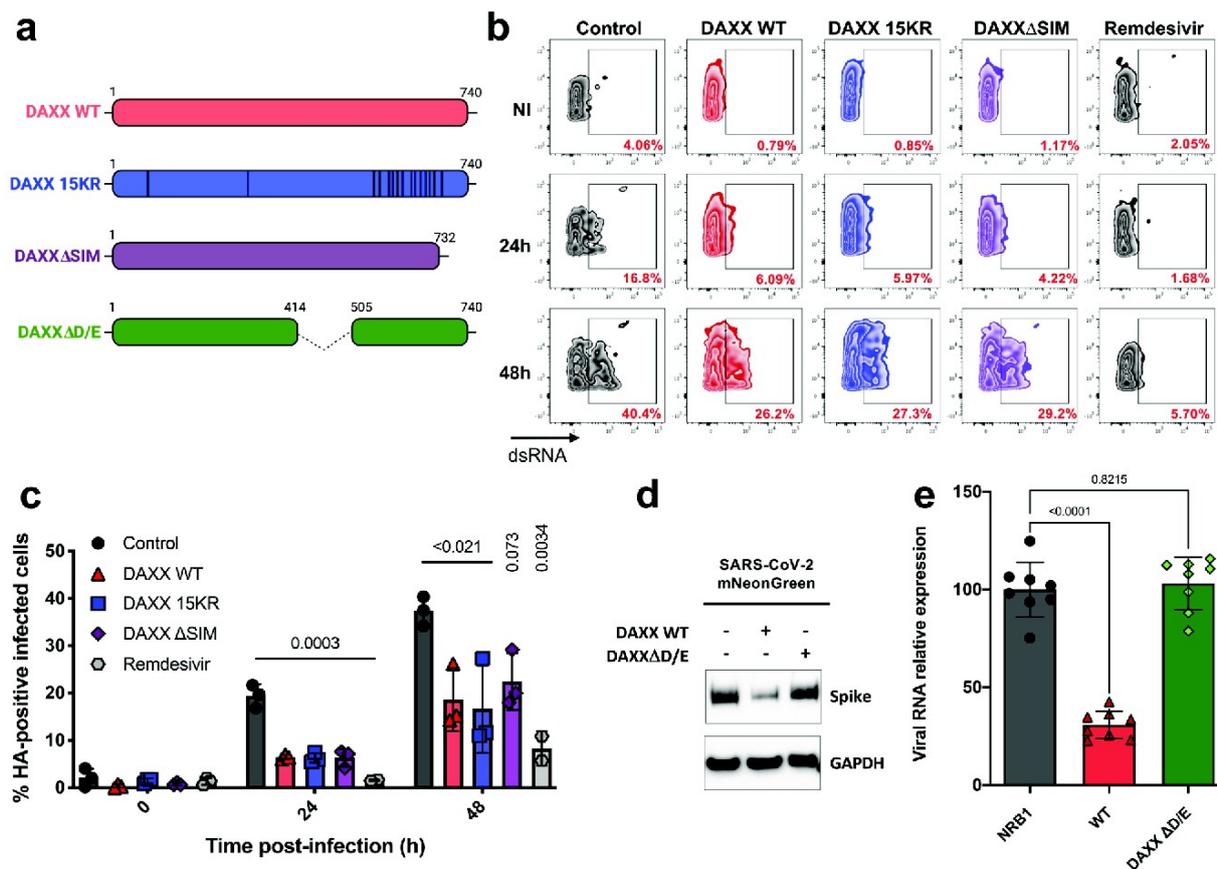


Fig. 4 DAXX restriction of SARS-CoV-2 is dependent on its chaperone activity but SUMOylation-independent. **a** Schematic of the DAXX mutants used. The fifteen lysine residues of DAXX 15KR have been mutated to arginine. DAXX Δ SIM lacks the 732–740 C-terminal residues. Both mutants were described in⁴⁸ DAXX Δ D/E is lacking its 414–505 domain and has been described in⁵² **b**, **c** SUMOylation-deficient DAXX mutants are still able to restrict SARS-CoV-2. 293T-ACE2 cells were transfected with HA-DAXX WT; HA-DAXX 15KR; HA-DAXX Δ SIM; or with HA-NBR1 as negative control plasmid. 24 h after transfection, cells were infected with SARS-CoV-2 at an MOI of 0.1. When indicated, cells were treated with remdesivir at the time of infection. After 24 h or 48 h, infected cells were double-stained for dsRNA (to read out infection) and HA (to read out transfection efficiency) and acquired by flow cytometry. The percentage of infected cells among HA-positive (transfected) cells for one representative experiment is shown in **(b)**, for the mean \pm SD of 3 independent experiments in **(c)**. Statistics: one-way ANOVA using Dunnett's test, Holm corrected. *P* values are indicated on the graph. **d**, **e**: The chaperone activity of DAXX is required for SARS-CoV-2 restriction. 293T-ACE2 cells were transfected with DAXX WT or with the DAXX Δ D/E mutant. 24 h after transfection, cells were infected with SARS-CoV-2-mNeonGreen at an MOI of 1. After 24 h, the levels of Spike and GAPDH levels were analyzed by Western Blot **(d)**. A Western Blot representative of 3 independent experiments is shown. In parallel, SARS-CoV-2 replication levels were measured by RT-qPCR targeting the 5' UTR and normalized against RPL13A transcript levels **(e)**. The mean \pm SD of 4 independent experiments, with infections carried out in two biological replicates, is represented. Statistics: 1-way ANOVA using Dunnett's test. *P* values are indicated on the graph. Source data are provided as a Source Data file.

SARS-CoV-2 infection triggers DAXX re-localization. DAXX mostly localizes in PML nuclear bodies³⁰, whereas SARS-CoV-2 replication occurs in the cytoplasm. We reasoned that DAXX may re-localize during the course of infection in order to exert its antiviral effect. We first examined the effect of DAXX overexpression on the replication of SARS-CoV-2-mNeonGreen⁵³ by microscopy. DAXX overexpression starkly reduced the number of infected cells (Fig. 5a, b), confirming our flow cytometry data (Fig. 4). Using double staining for HA-tagged DAXX and SARS-CoV-2, we found that most of the DAXX-transfected cells were negative for infection, and conversely, that most of the infected cells did not express transfected DAXX (Fig. 5c), confirming that DAXX imposes a major block to SARS-CoV-2 infection. Next, we infected 293T-ACE2 cells with SARS-CoV-2 and used high-resolution confocal microscopy to study the localization of endogenous DAXX (Fig. 5d). As expected³⁰, DAXX localized in discrete nuclear foci in non-infected cells. Strikingly, SARS-CoV-2 infection induced the re-localization of DAXX in the cytoplasm,

as early as 6 h post-infection, although some nuclear foci were still detected. At 24 h post-infection however, DAXX was completely absent from nuclear bodies, and was found almost exclusively in the cytoplasm, in close association with dsRNAs, likely representing SARS-CoV-2 replication sites. These results suggest that early events following SARS-CoV-2 infection trigger the cytoplasmic translocation of DAXX.

SARS-CoV-2 PLpro induces proteasomal degradation of DAXX. Next, we asked whether this relocalization of DAXX following infection destabilizes the protein. Western blot analysis revealed that SARS-CoV-2 infection induces a marked decrease of total DAXX expression levels in infected cells (Fig. 6a). In contrast, SARS-CoV-2 infection had no effect on DAXX mRNA levels (Fig. S6). Importantly, the decrease in DAXX protein levels is likely not attributed to a global host protein expression shut down, as the levels of Lamin B, HSP90, Actin, GAPDH, Tubulin, TRIM22 and RIG-I were unchanged upon infection (Fig. 6a).

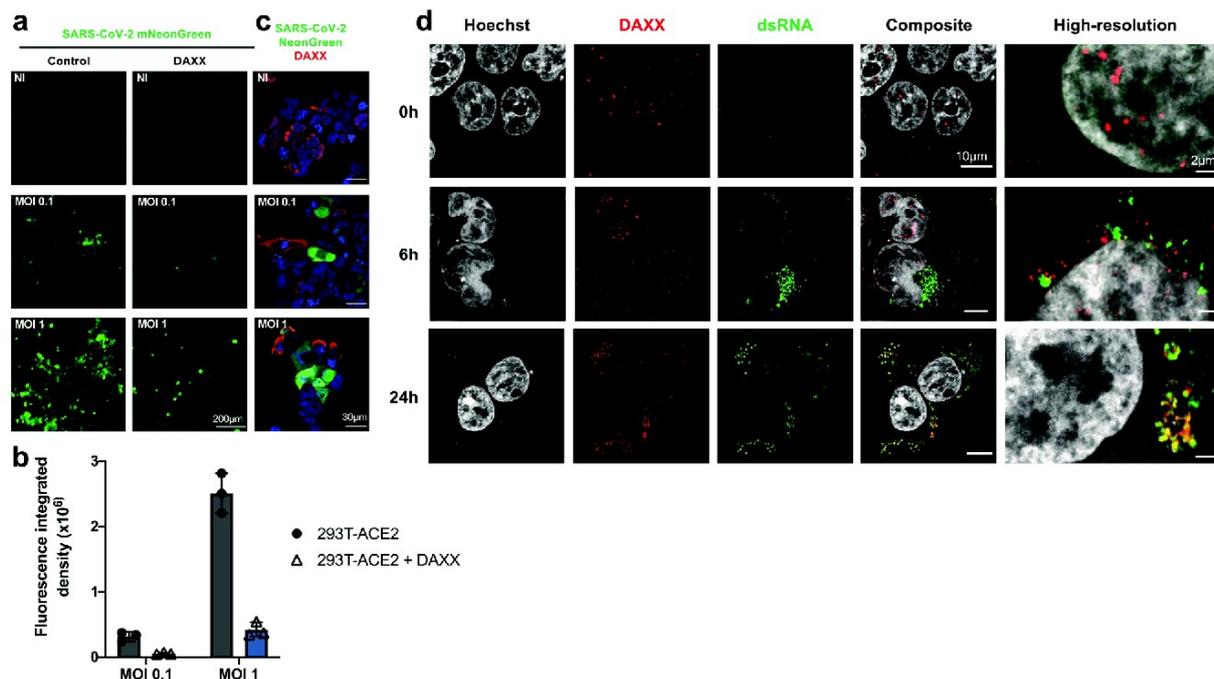


Fig. 5 SARS-CoV-2 infection induces DAXX cytoplasmic re-localization to sites of viral replication. DAXX overexpression restricts SARS-CoV-2. 293T-ACE2 cells were transfected with DAXX WT. 24 h after transfection, cells were infected with the mNeonGreen fluorescent reporter SARS-CoV-2 at the indicated MOI. Cells were either visualized with an EVOS fluorescence microscope (**a**, **b**) or stained with an HA-antibody detecting DAXX and imaged by confocal microscopy (**c**). Scale bars correspond to 200 μm (**a**) and 30 μm (**c**). Images shown in (**a**) were quantified using ImageJ software (**b**). Data shows the mean \pm SD of fluorescence integrated densities. The analysis was performed on around 200 cells from 3 different fields. Images are representative of 3–6 different fields from 2 independent experiments. **d** Relocalization of endogenous DAXX during SARS-CoV-2 infection. 293T-ACE2 cells were infected with SARS-CoV-2 at an MOI of 1. 24 h post-infection, cells were labelled with Hoescht and with antibodies against dsRNA (detecting viral RNA, in green) and HA (detecting DAXX, in red). When indicated, the high-resolution Airyscan mode was used. Scale bars correspond to 10 μm for confocal images, and 2 μm for the high-resolution images. Images are representative of 3–6 different fields from 2 independent experiments. Source data are provided as a Source Data file.

These results suggest that DAXX may be specifically targeted by SARS-CoV-2 for degradation. SARS-CoV-2 papain-like protease (PLpro) is a possible candidate for this activity, as it cleaves other cellular proteins such as ISG15^{54,55}, and ULK1⁵⁶. Moreover, PLpro of foot-and-mouth disease virus (FDMV) degrades DAXX⁵⁷. To investigate this possibility, we treated cells with GRL0617, an inhibitor of SARS-CoV-2 PLpro⁵⁵; MG132, a well-described proteasome inhibitor; or Masitinib, an inhibitor of SARS-CoV-2 3CL protease⁵⁸. These inhibitors had minimal effects on cell viability at the selected concentrations (Fig. S7). Strikingly, GRL0617 treatment partially restored DAXX expression (Fig. 6b), especially at the highest concentration. Similarly, MG132 also prevented DAXX degradation in SARS-CoV-2 infected cells. In contrast, Masitinib treatment had no effect on DAXX levels. These results suggest that PLpro, but not 3CL, targets DAXX for proteasomal degradation. Consistently, GRL0617 treatment also restored DAXX subcellular localization to nuclear bodies (Fig. 6c). As expected, GRL0617 treatment also inhibited the production of SARS-CoV-2 proteins, such as Spike (Fig. 6b), and may thus have an indirect effect on DAXX levels by inhibiting SARS-CoV-2 replication itself. However, the fact that Masitinib also inhibits Spike production but does not restore DAXX expression suggested that DAXX degradation is not an unspecific consequence of a reduced viral replication but rather a specific activity of PLpro. To investigate further the direct contribution of PLpro to DAXX degradation, we assessed the impact of overexpressing a panel of individual SARS-CoV-2 proteins in 293T-ACE2 cells on DAXX levels. We included in the analysis

mCherry-tagged SARS-CoV-2 Non-structural proteins (Nsp⁵⁹, which were not expressed from a lentiviral vector that may be targeted by DAXX antiviral activity³³. This included Nsp3 (which encodes PLro), Nsp4, Nsp6, Nsp7, Nsp10, Nsp13 and Nsp14. All proteins were expressed at similar levels (Fig. S8a). Only the overexpression of Nsp3 led to DAXX reduced expression (Fig. 6d, Fig. S8b). This effect was dose-dependent (Fig. 6e, Fig. S8c, d), and was abrogated when cells were treated with GRL0617 (Fig. 6f, Fig. S8e). Taken together, these results indicate that PLpro directly induces the proteasomal degradation of DAXX.

Discussion

The whole-genome CRISPR/Cas9 screens conducted to date on SARS-CoV-2 infected cells mostly identified host factors necessary for viral replication^{24–29} and did not focus on antiviral genes, as did our screen. Three overexpression screens, however, identified ISGs with antiviral activity against SARS-CoV-2^{16,21,22}. In the first one, Pfaender et al. screened 386 ISGs for their antiviral activity against the endemic human coronavirus 229E, and identified *LY6E* as a restriction factor inhibiting both 229E and SARS-CoV-2. Our screen also identified *LY6E* as a top hit (Fig. 1), further validating the findings of both studies. Four additional genes had significant *p*-values in both Pfaender et al. and our work: *IFI6*, *HERC5*, *OAS2* and *SPSB1* (Supplementary Data 1 and 2). We showed that knocking-out *LY6E* and *DAXX* only partially rescued SARS-CoV-2 replication in IFN-treated cells (Fig. 2), suggesting that other IFN effectors active against SARS-CoV-2 remain to be identified. For instance, other proteins, such as

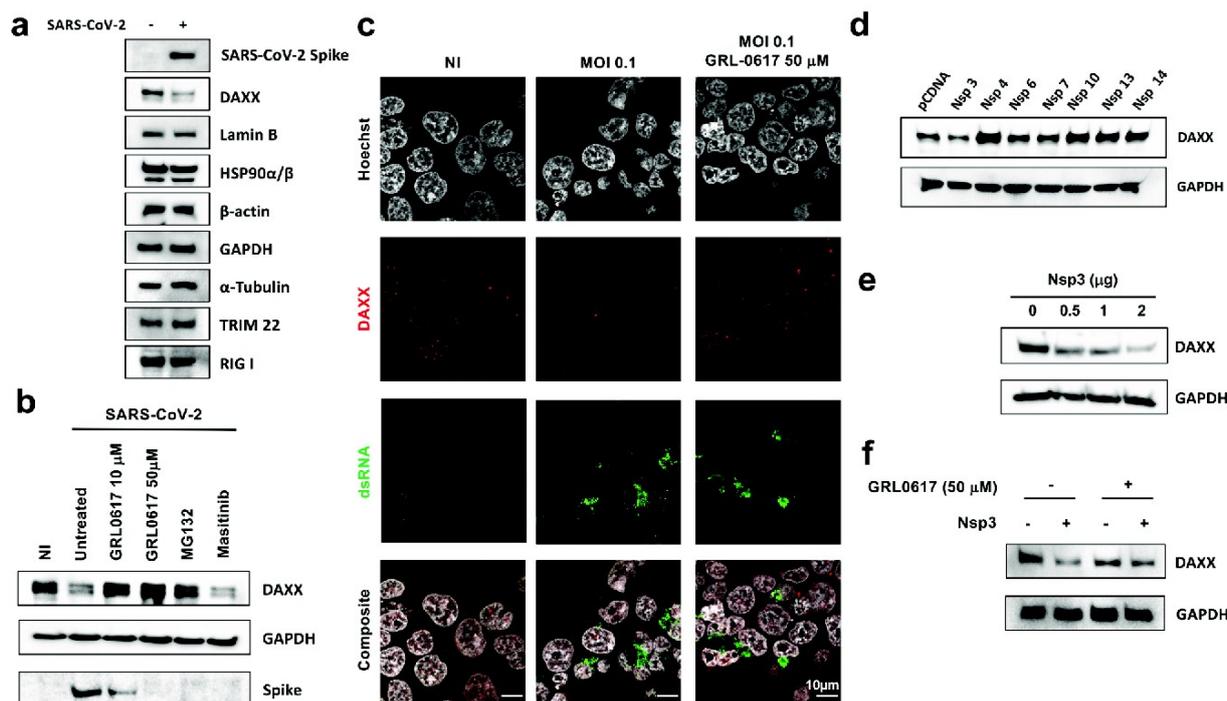


Fig. 6 SARS-CoV-2 PLpro induces the proteasomal degradation of DAXX. **a** DAXX degradation after infection. 293T-ACE2 cells were infected with SARS-CoV-2 at MOI 0.1. After 24 h, cells were harvested and levels of DAXX, Lamin B, HSP90, Actin, Tubulin, TRIM22, RIG-I and of the viral protein spike were analyzed by Western Blot. A Western Blot representative of 3 independent experiments is shown. **b** GRL0617 and MG132 treatments restore DAXX expression. 293T-ACE2 cells were infected with SARS-CoV-2 at an MOI of 0.1. When indicated, cells were pretreated 2 h before infection with GRL0617 (at the indicated concentrations), or with MG132 (10 μ M), a proteasome inhibitor, or Masitinib (10 μ M) a 3CL inhibitor. After 24 h, cells were harvested and levels of DAXX, GAPDH and of the viral protein spike were analyzed by Western Blot. A Western Blot representative of 3 independent experiments is shown. **c** GRL0617 treatment restores DAXX localization. 293T-ACE2 cells were infected with SARS-CoV-2 at an MOI of 0.1. 24 h post-infection, cells were labelled with Hoechst and with antibodies against dsRNA (detecting viral RNA, in green) and HA (detecting DAXX, in red). When indicated, cells were treated with 50 μ M of GRL0617 at the time of infection. Scale bars correspond to 10 μ m. Images are representative of 3–6 different fields from 2 independent experiments. **d–f**: Nsp3 induces DAXX degradation. **d** 293T-ACE2 cells were transfected with 1 μ g of the indicated viral proteins. After 24 h, the levels of DAXX and GAPDH were analyzed by Western Blot. **e** 293T-ACE2 cells were transfected with the indicated amounts of Nsp3. After 24 h, the levels of DAXX and GAPDH were analyzed by Western Blot. **f** 293T-ACE2 cells were transfected with 1 μ g of Nsp3 or of pcDNA. 6 h post transfection, cells were also, when indicated, treated with 50 μ M of GRL0617. Of, 24 h after transfection, the levels of DAXX and GAPDH were analyzed by Western Blot. Western Blots representative from 2 independent experiments are shown. The quantification of band intensity for Fig. 6d–f is shown in Fig. S8b,d, e. Source data are provided as a Source Data file.

IFITMs, inhibit SARS-CoV-2 viral entry^{17–19}. In the second screen, Martin Sancho et al. tested 399 ISGs against SARS-CoV-2. Among the 65 antiviral ISGs identified, they focused on *BST-2*, which encodes a protein targeting viral budding. *BST-2* was not a significant hit in our screen (Supplementary Data 1 and 2). This discrepancy is likely due to the fact that our screen relies on the sorting of S-positive cells, and is therefore unable to detect factors restricting the late stages of the viral replication cycle. The most recent overexpression screen assessed the contribution of 539 human and 444 macaque ISGs in SARS-CoV-2 restriction, and further characterized the role of *OAS1* in sensing SARS-CoV-2 and restricting its replication through *RNASEL*. While we did not identify *OAS1* or *RNASEL* in our screen (Supplementary Data 1 and 2), we did identify hits in common with this screen, including *IFI6* and *OAS2* (that were also identified by Pfaender et al.). Of note, *DAXX* was absent from the ISG libraries used by these overexpression screens, which explains why it was not previously identified as an antiviral gene for SARS-CoV-2. Our sgRNA library, by including 1905 genes, targeted a wider set of ISGs and “ISG-like” genes, including genes like *DAXX* that are not (or only weakly) induced by IFN in some cell types^{32,44}. One potential caveat to our screen is that it compared IFN-treated infected cells

to non-infected untreated cells (rather than IFN-treated non-infected cells). Although we validated this approach in previous studies^{34,60}, it may cause enrichment of genes induced by IFN, but not antiviral against SARS-CoV-2 per se. Interestingly, IFN has a stronger effect on *DAXX* gene expression levels in cells from other mammals such as bats.⁶¹ Future studies may investigate whether *DAXX* orthologs of different species are also able to restrict SARS-CoV-2 and whether *DAXX* participates in IFN-mediated viral restriction in these species.

We identify *DAXX* as a potent antiviral factor restricting the replication of SARS-CoV-2, acting independently of IFN (Fig. 3). *DAXX* fulfills all of the criteria defining a bona fide SARS-CoV-2 restriction factor: knocking-out endogenous *DAXX* leads to enhanced viral replication (Fig. 2), while over-expression of *DAXX* restricts infection (Figs. 4–5). While *DAXX* had no effect on Spike-mediated viral entry (Fig. 3), it led to a stark reduction in the levels of viral transcripts at 8 h post-infection, in the levels of Spike protein at 16 h post-infection (Fig. 3). This suggests that *DAXX* blocks a post-entry step of the viral life cycle such as viral transcription. *DAXX* co-localizes with viral replication sites (Fig. 5) and SARS-CoV-2 antagonizes *DAXX* to some extent, as evidenced by the proteasomal degradation of *DAXX* induced by

PLpro (Fig. 6). Although DAXX expression is not upregulated by IFN α in A549 cells (Fig. S3), basal levels of expression are sufficient for its antiviral activity, as has been shown for other potent restriction factors. Publicly available single-cell RNAseq analyses (Fig. S2) indicated that DAXX is expressed in cell types targeted by the virus in vivo, such as lung epithelial cells and macrophages. Interestingly, DAXX exhibited some degree of specificity in its antiviral activity, as unrelated viruses such as YFV and MeV, as well as the closely related MERS-CoV were not sensitive to its action, in contrast to SARS-CoV and SARS-CoV-2 (Fig. 2). Future work will determine which viral determinants are responsible for the specific antiviral activity of DAXX.

DAXX is mostly known for its antiviral activity against DNA viruses replicating in the nucleus, such as adenovirus 5 (AdV5)⁶² and human papillomavirus (HPV)⁶³. Most of these viruses antagonize PML and/or DAXX, which interacts with PML in nuclear bodies³⁰. We show here that DAXX is also able to restrict SARS-CoV-2, a positive sense RNA virus that replicates in the cytoplasm. Recent studies have shown that DAXX inhibits the reverse transcription of HIV-1 in the cytoplasm^{32,33}. Within hours of infection, DAXX subcellular localization was altered, with DAXX accumulating in the cytoplasm and colocalizing with incoming HIV-1 capsids³³. Here, we observed a similar phenomenon, with a rapid re-localization of DAXX from the nucleus to cytoplasmic viral replication sites (Fig. 5), where it likely exerts its antiviral effect. Early events in the replication cycle of both HIV-1 and SARS-CoV-2, such as fusion between viral and cellular membranes, or virus-induced stress, may thus trigger DAXX re-localization to the cytoplasm. DAXX seems to inhibit SARS-CoV-2 by a distinct mechanism than HIV-1: whereas the recruitment of interaction partners through the SIM-domain is required for the effect of DAXX on HIV-1 reverse transcription³², it was not the case in the context of SARS-CoV-2 restriction. This result was unexpected, since DAXX has no enzymatic activity and rather acts as a scaffold protein recruiting SUMOylated partners through its SIM domain⁵¹. Other DAXX functions, such as interaction with the chromatin remodeler ATRX³⁰ or its role as a chaperone protein⁵² are, however, SIM-independent. This last activity was recently shown⁵² to be dependent on the D/E domain (amino acids 414 to 505). We found that DAXX anti-SARS-CoV-2 activity also depends on this domain (Fig. 4), suggesting that it restricts SARS-CoV-2 replication through an original mechanism involving protein refolding. Future work will determine whether DAXX binds and refolds viral proteins to hamper viral replication, or acts through binding and folding another host factor.

Our results suggest that SARS-CoV-2 evolved a mechanism to antagonize DAXX restriction, with PLpro inducing its degradation via the proteasome (Fig. 6). This antagonism, however, is not complete, since knocking-out DAXX expression enhanced SARS-CoV-2 replication (Fig. 2). Another possibility is that DAXX, by acting early in the viral life cycle (i.e. as soon as 8 h p.i., Fig. 3) exert its antiviral effect before the expression of PLpro. Proteins expressed by other viruses are also able to degrade DAXX: for instance, the AdV5 viral factor E1B-55K targets DAXX for proteasomal degradation⁶², and FDMV PLpro cleaves DAXX⁵⁷. We showed in Fig. 2 that several SARS-CoV-2 variants and SARS-CoV were sensitive to DAXX, but MERS-CoV was not. Thus, it will be interesting to test whether PLpro from these different coronaviruses differ in their ability to degrade DAXX, and whether this has an impact on their sensitivity to DAXX restriction. Future research may also establish whether PLpro induces the degradation of DAXX through direct cleavage, or whether it acts in a more indirect way, such as cleaving or recruiting cellular co-factors. Such investigations may be relevant for the development of PLpro inhibitors⁶⁴: indeed, in addition to directly blocking

SARS-CoV-2 replication, PLpro inhibitors may also sensitize the virus to existing antiviral mechanisms such as DAXX restriction.

Methods

Cells, viruses & plasmids. HEK 293T (ATCC #CRL-11268) were cultured in MEM (Gibco #11095080) complemented with 10% FBS (Gibco #A3160801) and 2 mM L-Glutamine (Gibco # 25030081). VeroE6 (ATCC #CRL-1586), A549 (ATCC #CCL-185) and HEK 293T, both overexpressing the ACE2 receptor (A549-ACE2 and HEK 293T-ACE2, respectively), were grown in DMEM (Gibco #31966021) supplemented with 10% FBS (Gibco #A3160801), and penicillin/streptomycin (100 U/mL and 100 μ g/mL, Gibco # 15140122). Blasticidin (10 μ g/mL, Sigma-Aldrich #SBR00022-10ML) was added for selection of A549-ACE2 and HEK 293T-ACE2. All cells were maintained at 37 °C in a 5% CO₂ atmosphere. Universal Type I Interferon Alpha (PBL Assay Science #11200-2) was diluted in sterile-filtered PBS 1% BSA according to the activity reported by the manufacturer. The strains BetaCoV/France/IDF0372/2020 (Lineage B); hCoV-19/France/IDF-IPP11324/2020 (Lineage B.1.1.7); and hCoV-19/France/PDL-IPP01065/2021 (Lineage B.1.351) were supplied by the National Reference Centre for Respiratory Viruses hosted by Institut Pasteur and headed by Pr. Sylvie van der Werf. The human samples from which the lineage B, B.1.1.7 and B.1.351 strains were isolated were provided by Dr. X. Lescure and Pr. Y. Yazdanpanah from the Bichat Hospital, Paris, France; Dr. Besson J., Bioliance Laboratory, saint-Herblain France; Dr. Vincent Foissaud, HIA Percy, Clamart, France, respectively. These strains were supplied through the European Virus Archive goes Global (Evag) platform, a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement #653316. The hCoV-19/Japan/TY7-501/2021 strain (Lineage P1) was kindly provided by Jessica Vanhongen (Cellule d'Intervention Biologique d'Urgence; Institut Pasteur). The mNeonGreen reporter SARS-CoV-2 was provided by Pei-Yong Shi⁵³. SARS-CoV FFM-1 strain⁶⁵ was kindly provided by H.W. Doerr (Institute of Medical Virology, Frankfurt University Medical School, Germany). The Middle East respiratory syndrome (MERS) Coronavirus, strain IP/COV/MERS/Hu/France/FRA2 (Genbank reference KJ361503) isolated from one of the French cases⁶⁶ was kindly provided by Jean-Claude Manuguerra (Cellule d'Intervention Biologique d'Urgence; Institut Pasteur). SARS-CoV-2 viral stocks were generated by infecting VeroE6 cells (MOI 0.01, harvesting at 3 dpi) using DMEM supplemented with 2% FBS and 1 μ g/mL TPCK-trypsin (Sigma-Aldrich #1426-100MG). SARS-CoV and MERS-CoV viral stocks were generated by infecting VeroE6 cells (MOI 0.0001) using DMEM supplemented with 5% FCS and harvesting at 3 dpi or 6 dpi, respectively. The Yellow Fever Virus (YFV) Asibi strain was provided by the Biological Resource Center of the Institut Pasteur. The Measles Schwarz strain expressing GFP (MeV-GFP) was described previously.⁶⁷ Both viral stocks were produced on Vero NK cells. The Human Interferon-Stimulated Gene CRISPR Knockout Library was a gift from Michael Emerman and is available on Addgene (Pooled Library #125753). The plentiCRISPRv2 backbone was ordered through Addgene (Plasmid #52961). pMD2.G and psPAX2 were gifts from Didier Trono (Addgene #12259; #12260). pcDNA3.1 was purchased from Invitrogen. Plasmids constructs expressing WT and mutant HA-tagged DAXX constructs were kindly provided by Hsiu-Ming Shih⁵¹. The plasmids encoding mCherry-tagged viral proteins were a gift from Bruno Antony and ordered through Addgene: Nsp3-mCherry (#165131); Nsp4-mCherry (#165132); Nsp6-mCherry (#165133); Nsp7-mCherry (#165134); Nsp10-mCherry (#165135); Nsp13-mCherry (#165136); Nsp14-mCherry (#165137). Vesicular stomatitis virus (VSV) encoding green fluorescent protein (GFP) has been previously described as VSV*⁶⁸. The chimeric virus VSV* Δ G-SARS-CoV-2-S₂₂₁ (VSV* Δ G-S), which lacks the homotypic glycoprotein G but rather encodes the spike protein of SARS-CoV-2 (Wuhan-Hu-1 strain) along with GFP has recently been described⁶⁹. The DAXX Δ D/E mutant is a kind gift from Pr. Xiaolu Yang.

Antibodies. For Western Blot, we used mouse anti-DAXX (diluted 1:1000, Abnova #7A11), rat anti-HA clone 3F10 (diluted 1:3000, Sigma #2158167001), mouse anti-GAPDH clone 6C5 (diluted 1:3000, Millipore #FCMAB252F), Goat anti-Lamin B clone M-20 (diluted 1:500, Santa Cruz sc-6217), mouse monoclonal HSP90 α / β clone F-8 (diluted 1:500, Santa Cruz sc-13119), mouse monoclonal α -actin clone AC-15 (1:3000 Sigma #A1978), mouse monoclonal α -Tubulin clone DMA1 (diluted 1:1000, Sigma #T9026), rabbit anti-TRIM22 (diluted 1:1000, Proteintech #13744-1-AP) and mouse Monoclonal RIG-I clone Alme-1 (diluted 1:1000, adipoGen #AG-20B-0009). To detect SARS-CoV-2 Spike protein, we used mouse anti-spike clone 1A9 (diluted 1:1000, GeneTex GTX632604). Secondary antibodies were goat anti-mouse and anti-rabbit HRP-conjugates (diluted 1:5000, ThermoFisher #31430 and #31460) and horse anti-goat HRP (diluted 1:1000, Vector # PI-9500). For immunofluorescence, we used rabbit anti-DAXX (diluted 1:50, Proteintech #20489-1-AP) and mouse anti-dsRNA J2 (diluted 1:50, Scicons #10010200). Secondary antibodies were goat anti-rabbit AF555 and anti-mouse AF488 (diluted 1:1000, ThermoFisher #A-21428 and #A-28175). For flow sorting of infected cells, we used the anti-S2 H2 162 antibody (diluted 1:150), a kind gift from Dr. Hugo Mouquet, (Institut Pasteur, Paris, France). Secondary antibody was donkey anti-mouse AF647 (diluted 1:1000, Invitrogen #A31571). For FACS analysis, we used rat anti-HA clone 3F10 (diluted 1:100, Sigma #2158167001) and mouse anti-dsRNA J2 (diluted 1:500, Scicons #10010200).

Secondary antibodies were goat anti-rat AF647 and anti-mouse AF488 (diluted 1:1000, ThermoFisher #A-21247 #A-28175). The pan-flavivirus anti-Env 4G2 antibody was a kind gift from Philippe Desprès.

Generation of CRISPR/Cas9 library cells. HEK 293T cells were transfected with the sgRNA plasmid library together with plasmids coding for Gag/Pol (R8.2) and for the VSVg envelope (pVSVg) using a ratio of 5:5:1 and calcium phosphate transfection. Supernatants were harvested at 36 h and 48 h, concentrated 80-fold by ultracentrifugation (22,000 g, 4 °C for 1 h) and pooled. To generate the ISG KO library cells, 36×10^6 A549-ACE2 cells were seeded in 6-well plates (10^6 cells per well) 24 h before transduction. For each well, 100 μ L of concentrated lentivector was diluted in 500 μ L of serum-free DMEM, supplemented with 10 μ g/mL of DEAE dextran (Sigma #D9885). After 48 h, transduced cells were selected by puromycin treatment for 20 days (1 μ g/mL; Sigma #P8833).

CRISPR/Cas9 screen. In total, 4×10^7 A549-ACE2 cells were treated with IFN α (200 U/mL). 16 h later, cells were infected at an MOI of 1 in serum-free media complemented with TPCK-trypsin and IFN α (200 U/mL). After 90 min, the viral inoculum was removed, and cells were maintained in DMEM containing 5% FBS and IFN α (200 U/mL). After 24 h, cells were harvested and fixed for 15 min in Formalin 1%. Fixed cells were washed in cold FACS buffer containing PBS, 2% Bovine Serum Albumin (Sigma-Aldrich #A2153-100G), 2 mM EDTA (Invitrogen #15575-038) and 0.1% Saponin (Sigma-Aldrich #S7900-100G). Cells were incubated for 30 min at 4 °C under rotation with primary antibody diluted in FACS buffer. Incubation with the secondary antibody was performed during 30 min at 4 °C under rotation. Stained cells were resuspended in cold sorting buffer containing PBS, 2% FBS, 25 mM HEPES (Sigma-Aldrich #H0887-100ML) and 5 mM EDTA. Infected cells were sorted on a BD FACS Aria Fusion. Sorted and control (non-infected, not IFN-treated) cells were centrifuged (20 min, 2,000 g) and resuspended in lysis buffer (NaCl 300 mM, SDS 0.1%, EDTA 10 mM, EGTA 20 mM, Tris 10 mM) supplemented with 1% Proteinase K (Qiagen #19133) and 1% RNase A/T1 (ThermoFisher #EN0551) and incubated overnight at 65 °C. Two consecutive phenol-chloroform (Sigma #P3803-100ML) extractions were performed and DNA was recovered by ethanol precipitation. Nested PCR was performed using the Herculase II Fusion DNA Polymerase (Agilent, #600679) and the DNA oligos indicated in Supplementary Table 1. PCR1 products were purified using QIAquick PCR Purification kit (Qiagen #28104). PCR2 products were purified using Agencourt AMPure XP Beads (Beckman Coulter Life Sciences #A63880). DNA concentration was determined using Qubit dsDNA HS Assay Kit (Thermo Fisher #Q32854) and adjusted to 2 nM prior to sequencing. NGS was performed using the NextSeq 500/550 High Output Kit v2.5 75 cycles (Illumina #20024906).

Screen analysis. Reads were demultiplexed using bcl2fastq Conversion Software v2.20 (Illumina) and fastx_toolkit v0.0.13. Sequencing adapters were removed using cutadapt v1.9.1.⁷⁰ The reference library was built using bowtie2 v2.2.9.⁷¹ Read mapping was performed with bowtie2 allowing 1 seed mismatch in -local mode and samtools v1.9.⁷² Mapping analysis and gene selection were performed using MAGeCK v0.5.6, normalizing the data with default parameters. sgRNA and gene enrichment analyses are available in Supplementary Data 1 and 2, respectively and full MAGeCK output at https://github.com/Simon-LorriereLab/crispr_isg_sarscov2.

Generation of multi-guide gene knockout cells. 3 sgRNAs per gene were designed (Supplementary Table 2). 10 pmol of NLS-Sp.Cas9-NLS (SpCas9) nuclease (Aldevron #9212) was combined with 30 pmol total synthetic sgRNA (10 pmol for each sgRNA) (Synthego) to form RNPs in 20 μ L total volume with SE Buffer (Lonza #V55C-1002). The reaction was incubated at room temperature for 10 min. In total, 2×10^5 cells per condition were pelleted by centrifugation at $100 \times g$ for 3 min, resuspended in SE buffer and diluted to 2×10^4 cells/ μ L. 5 μ L of cell solution was added to the pre-formed RNP solution and gently mixed. Nucleofections were performed on a Lonza HT 384-well nucleofector system (Lonza #AAU-1001) using program CM-120. Immediately following nucleofection, each reaction was transferred to a 96-well plate containing 200 μ L of DMEM 10% FBS (5×10^4 cells per well). Two days post-nucleofection, DNA was extracted using DNA QuickExtract (Lucigen #QE09050). Cells were lysed in 50 μ L of QuickExtract solution and incubated at 68 °C for 15 min followed by 95 °C for 10 min. Amplicons were generated by PCR amplification using NEBNext polymerase (NEB #M0541) or AmpliTaq Gold 360 polymerase (ThermoFisher #4398881) and the primers indicated in Supplementary Table 3. PCR products were cleaned-up and analyzed by Sanger sequencing. Sanger data files and sgRNA target sequences were input into Inference of CRISPR Edits (ICE) analysis <https://ice.synthego.com/#/> to determine editing efficiency and to quantify generated indels.⁷³ Percentage of alleles edited is shown in Table 1 ($n = 3$).

Hit validation. In total, 2.5×10^4 A549-ACE2 KO cells were seeded in 96-well plates 18 h before the experiment. Cells were treated with IFN α and infected as described above. At 72 h post-infection, supernatants and cellular monolayers were harvested in order to perform qRT-PCR and plaque assay titration. Infectious

supernatants were heat-inactivated at 80 °C for 10 min. For intracellular RNA, cells were lysed in a mixture of Trizol Reagent (Invitrogen #15596018) and PBS at a ratio of 3:1. Total RNA was extracted using the Direct-zol 96 RNA kit (Zymo Research #R2056) or the Direct-zol RNA Miniprep kit (Zymo Research #R2050). For SARS-CoV-2 detection, qRT-PCR was performed either directly on the inactivated supernatants or on extracted RNA using the Luna Universal One-Step RT-qPCR Kit (NEB #E3005E) in a QuantStudio 6 thermocycler (Applied Biosystems) or in a StepOne Plus thermocycler (Applied Biosystems). The primers used are described in Supplementary Table 4. Cycling conditions were the following: 10 min at 55 °C, 1 min at 95 °C and 40 cycles of 95 °C for 10 s and 60 °C for 1 min. Results are expressed as genome copies/mL as the standard curve was performed by diluting a commercially available synthetic RNA with a known concentration (EURM-019, JRC). For SARS-CoV and MERS-CoV, qRT-PCR were performed using FAM-labelled probes (Eurogentec) and the Superscript III Platinum One-Step qRT-PCR System (Thermo Fisher Scientific, #11732020). The cycling conditions were the following: 20 min at 55 °C, 3 min at 95 °C and 50 cycles of 95 °C for 15 s and 58 °C for 30 s. The primers used are described in Supplementary Table 4. Standard curves were performed using serial dilutions of RNA extracted from and SARS-CoV and MERS-CoV viral culture supernatants of known infectious titer. For plaque assay titration, VeroE6 cells were seeded in 24-well plates (10^5 cells per well) and infected with serial dilutions of infectious supernatant diluted in DMEM during 1 h at 37 °C. After infection, 0.1% agarose semi-solid overlays were added. At 72 h post-infection, cells were fixed with Formalin 4% (Sigma #HT501128-4L) and plaques were visualized using crystal violet coloration. Time-course experiments were performed the same way, except that supernatants and cellular monolayers were harvested at 0 h, 2 h, 4 h, 6 h, 8 h, 10 h and 24 h post-infection.

SARS-CoV-2, SARS-CoV and MERS-CoV infection assays. A549-ACE2 cells were infected by incubating the virus for 1 h with the cells maintained in DMEM supplemented with 1 μ g/ml TPCK-trypsin (Sigma #4370285). The viral input was then removed and cells were kept in DMEM supplemented with 2% FBS. For 293T-ACE2 cells, infections were performed without TPCK-trypsin. MERS-CoV and SARS-CoV infections were performed in DMEM supplemented with 2% FBS and cells were incubated 1 h at 37 °C 5% CO₂. Viral inoculum was then removed and replaced by fresh DMEM supplemented with 2% FBS. All experiments involving infectious material were performed in Biosafety Level 3 facilities in compliance with Institut Pasteur's guidelines and procedures. When indicated, SARS-CoV-2 infected cells were stained for intracellular Spike levels as described below.

Yellow fever virus and measles virus infection assays. Cells were infected with YFV (at an MOI of 0.3) or MeV-GFP (at an MOI of 0.2) in DMEM without FBS for 2 h in small volume of medium to enhance contacts with the inoculum and the cells. After 2 h, the viral inoculum was replaced with fresh DMEM 10% FBS 1% P/S. FACS analysis was performed at 24 h p.i. Cells were fixed and permeabilized using BD Cytofix/Cytoperm (Fisher Scientific, # 15747847) for 30 min on ice (all the following steps were performed on ice and centrifuged at 4 °C) and then washed three times with wash buffer. Cells infected with YFV were incubated with the pan-flavivirus anti-Env 4G2 antibody for 1 h at 4 °C and then with Alexa 488 anti-mouse IgG secondary antibodies (Thermo Fisher, #A28175) for 45 min at 4 °C in the dark. Non-infected, antibody-stained samples served as controls for signal background. The number of cells infected with MeV-GFP were assessed with the GFP signal, using non-infected cells as controls. Data were acquired with an Attune NxT Acoustic Focusing Cytometer (Life technologies) and analyzed using FlowJo software.

Entry assays. Cells were seeded at 1×10^5 cells per well in 24-well plates in DMEM with 1% FBS. The next day, cells were infected with VSV* (MOI 0.001) or VSV* Δ G-S (MOI 7) in DMEM without FBS. The virus suspension was removed after 2 h and replaced with DMEM with 1% FBS. 16 h p.i., cells were washed once with PBS, trypsinized and subsequently fixed in 4% PFA. Fixed cells were washed once with PBS and analyzed by flow cytometry. The percentage of infected cells was identified based on GFP expression.

Overexpression assay. In total, 2×10^5 293T-ACE2 cells were seeded in a 24-well plate 18 h before the experiment. Cells were transfected with 500 ng of plasmids expressing HA-DAXX WT, HA-DAXX 15KR and HA-DAXX Δ SIM plasmids, using Fugene 6 (Promega # E2691), following the manufacturer's instructions. HA-NBR1 was used as negative control. After 24 h cells were infected at the indicated MOI in DMEM 2% FBS. When indicated, cells were treated with 10 mM of remdesivir (MedChemExpress #HY-104077) at the time of infection. For flow cytometry analysis, cells were fixed with 4% formaldehyde and permeabilized in a PBS 1% BSA 0.025% saponin solution for 30 min prior to staining with corresponding antibodies for 1 h at 4 °C diluted in the permeabilization solution. Samples were acquired on a BD LSR Fortessa and analyzed using FlowJo. Total RNA was extracted using a RNeasy Mini kit and submitted to DNase treatment (Qiagen). RNA concentration and purity were evaluated by spectrophotometry (NanoDrop 2000c, ThermoFisher). In addition, 500 ng of RNA were reverse transcribed with both oligo dT and random primers, using a PrimeScript RT

Reagent Kit (Takara Bio) in a 10 mL reaction. Real-time PCR reactions were performed in duplicate using Takyon ROX SYBR MasterMix blue dTTP (Eurogentec) on an Applied Biosystems QuantStudio 5 (ThermoFisher). Transcript levels were quantified using the following program: 3 min at 95 °C followed by 35 cycles of 15 s at 95 °C, 20 s at 60 °C, and 20 s at 72 °C. Values for each transcript were normalized to expression levels of *RPL13A*. The primers used are indicated in Supplementary Table 4.

Microscopy immunolabeling and imaging. 293T-ACE2 cells were cultured and infected with SARS-CoV-2 as described above. When indicated, cells were treated with 50 μM of GRL0617 (MedChemExpress #HY-117043) at the time of infection. Cultures were rinsed with PBS and fixed with 4% paraformaldehyde (electronic microscopy grade; Alfa Aesar) in PBS for 10 min at room temperature, treated with 50 mM NH₄Cl for 10 min, permeabilized with 0.5% Triton X-100 for 15 min, and blocked with 0.3% BSA for 10 min. Cells were incubated with primary and secondary antibodies for 1 h and 30 min, respectively, in a moist chamber. Nuclei were labeled with Hoechst dye (Molecular Probes). Images were acquired using a LSM700 (Zeiss) confocal microscope equipped with a 63X objective or by Airyscan LSM800 (Zeiss). Image analysis and quantification was performed using ImageJ (Fiji) v2.1.

Western blot. 293T-ACE2 cells were transfected with the indicated plasmids or treated with the indicated concentrations of GRL0617; with 10 μM of Masitinib (MedChemExpress #HY-10209); or with 10 μM of MG132 (SIGMA #M7449), an inhibitor of the proteasome and infected with SARS-CoV-2. Cell lysates were prepared using RIPA lysis and extraction buffer (ThermoFisher #89901). Protein concentration was determined using Bradford quantification. Proteins were denatured using 4X Bolt LDS Sample Buffer (Invitrogen) and 10X Bolt Sample Reducing Agent (Invitrogen). 40 μg of proteins were denatured and loaded on 12% ProSieve gel and then subjected to electrophoresis. Gels were then transferred (1 h, 90 V) to Western blotting membranes, nitrocellulose (GE Healthcare #GE10600002) using Mini Trans-Blot Electrophoretic Transfer Cell (Biorad #1703930EDU). Membranes were blocked with 5% BSA in PBS (blocking buffer) and incubated with primary antibodies diluted in blocking buffer. Membranes were washed and incubated with secondary antibodies diluted in blocking buffer. Chemiluminescent acquisitions were performed on a Chemidoc™ MP Imager and analyzed using Image Lab v6 software (Bio-Rad Laboratories).

Flow cytometry. For flow cytometry analysis, all cells were fixed with 4% formaldehyde. For intracellular staining, cells were permeabilized in a PBS 1% BSA 0.025% saponin solution for 30 min prior to staining with corresponding primary antibodies for 1 h at 4 °C and then secondary antibodies for 45 min at 4 °C, diluted in the permeabilization solution. Acquisition was done with BD Fortessa and Attune NxT cytometers. Data was analyzed with the FlowJo software (Treestar Inc., Oregon, USA v10.8.1).

Single-cell RNAseq analysis. Single cell RNAseq analysis were performed in the BioTuring Browser software v2.8.42 developed by BioTuring, using a dataset made available by Liao et al.⁴³ (GSE145926). All processing steps were done by BioTuring Browser.⁷⁴ Cells with less than 200 genes and mitochondrial genes higher than 10% were excluded from the analysis.

Statistical analysis. Graphpad Prism v9.1.0 was used for statistical analyses. Linear models were computed using Rstudio v1.2.1335.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

CRISPR/Cas9 screen NGS raw (fastq) and preprocessed data (read counts) is available in NCBI GEO (GSE173418); sgRNA and gene enrichment analyses are available in Supplementary Data 1 and 2, respectively and full MAGECK output at https://github.com/Simon-LorieriLab/crispr_isg_sarscov2. The single cell data from Liao et al. 2020 Nature is available in NCBI GEO (GSE145926). Source data are provided with this paper.

Code availability

All code used in this study is available at https://github.com/Simon-LorieriLab/crispr_isg_sarscov2.

Received: 27 April 2021; Accepted: 11 April 2022;

Published online: 04 May 2022

References

1. Ruetsch, C. et al. Functional exhaustion of type I and II interferons production in severe COVID-19 patients. *Front Med.* **7**, 603961 (2021).
2. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
3. Combes, A. J. et al. Global absence and targeting of protective immune states in severe COVID-19. *Nature* **591**, 124–130 (2021).
4. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).
5. Galani, I.-E. et al. Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat. Immunol.* **22**, 32–40 (2021).
6. Bastard, P. et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
7. Zhang, Q. et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
8. Sa Ribero, M., Jovenet, N., Dreux, M. & Nisole, S. Interplay between SARS-CoV-2 and the type I interferon response. *PLOS Pathog.* **16**, e1008737 (2020).
9. Rebendenne, A. et al. SARS-CoV-2 triggers an MDA-5-dependent interferon response which is unable to control replication in lung epithelial cells. *J. Virol.* **95**, e02415–e02420 (2021).
10. Yin, X. et al. MDA5 governs the innate immune response to SARS-CoV-2 in lung epithelial cells. *Cell Rep.* **34**, 108628 (2021).
11. Felgenhauer, U. et al. Inhibition of SARS-CoV-2 by type I and type III interferons. *J. Biol. Chem.* **295**, 13958–13964 (2020).
12. Lokugamage, K. G. et al. Type I interferon susceptibility distinguishes SARS-CoV-2 from SARS-CoV. *J. Virol.* **94**, e01410–e01420 (2020).
13. Hoagland D. A et al. Leveraging the antiviral type I interferon system as a first line of defense against SARS-CoV-2 pathogenicity. *Immunity*. 2021; S1074761321000406.
14. Busnadiego, I. et al. Antiviral activity of Type I, II, and III interferons counterbalances ACE2 inducibility and restricts SARS-CoV-2. *mBio* **11**, e01928–20 (2020).
15. Zhao, X. et al. LY6E restricts entry of human Coronaviruses, including currently pandemic SARS-CoV-2. *J. Virol.* **94**, e00562–20 (2020).
16. Pfander, S. et al. LY6E impairs coronavirus fusion and confers immune control of viral disease. *Nat. Microbiol.* **5**, 1330–1339 (2020).
17. Shi, G. et al. Opposing activities of IFITM proteins in SARS-CoV-2 infection. *EMBO J.* **40**, e106501 (2021).
18. Buchrieser J, et al. Syncytia formation by SARS-CoV-2-infected cells. *EMBO J.* **39**, e106267 (2020).
19. Biering, S. B., et al. Genome-wide, bidirectional CRISPR screens identify mucins as critical host factors modulating SARS-CoV-2 infection. *bioRxiv*. 2021;2021.04.22.440848.
20. Nchioua, R. et al. SARS-CoV-2 is restricted by zinc finger antiviral protein despite preadaptation to the Low-CpG environment in humans. *mBio* **11**, e01930–20 (2020).
21. Wickenhagen A, et al. A prenylated dsRNA sensor protects against severe COVID-19. *Science* 2021;eabj3624.
22. Martin-Sancho L, et al. Functional landscape of SARS-CoV-2 cellular restriction. *Mol. Cell.* **81**, 2656–2668.e8 (2021).
23. Bonaventure B, et al. A genome-wide CRISPR/Cas9 knock-out screen identifies the DEAD box RNA helicase DDX42 as a broad antiviral inhibitor. *bioRxiv*. 2020;2020.10.28.359356.
24. Hoffmann, H.-H. et al. Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. *Cell Host Microbe* **29**, 267–280.e5 (2021).
25. Daniloski, Z. et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105.e16 (2021).
26. Wei, J. et al. Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cell* **184**, 76–91.e13 (2021).
27. Schneider, W. M. et al. Genome-scale identification of SARS-CoV-2 and pan-coronavirus host factor networks. *Cell* **184**, 120–132.e14 (2021).
28. Wang, R. et al. Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. *Cell* **184**, 106–119.e14 (2021).
29. Baggen, J. et al. Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2. *Nat. Genet.* **53**, 435–444 (2021).
30. Tang, J. et al. A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. *J. Biol. Chem.* **279**, 20369–20377 (2004).
31. Porter, S. S., Stepp, W. H., Stamos, J. D. & McBride, A. A. Host cell restriction factors that limit transcription and replication of human papillomavirus. *Virus Res* **231**, 10–20 (2017).
32. Maillet S, et al. Daxx inhibits HIV-1 reverse transcription and uncoating in a SUMO-dependent manner. *Viruses*. **12**, 636–655 (2020).
33. Dutrieux, J. et al. PML/TRIM19-dependent inhibition of retroviral reverse-transcription by Daxx. *PLoS Pathog.* **11**, e1005280 (2015).

34. OhAinle, M. et al. A virus-packagable CRISPR screen identifies host factors mediating interferon inhibition of HIV. *eLife* **7**, e39823 (2018).
35. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
36. Katopodis, P. et al. Pan-cancer analysis of transmembrane protease serine 2 and cathepsin L that mediate cellular SARS-CoV-2 infection leading to COVID-19. *Int J. Oncol.* **57**, 533–539 (2020).
37. Malakhova, O. A. et al. Protein ISGylation modulates the JAK-STAT signaling pathway. *Genes Dev.* **17**, 455–460 (2003).
38. Broering, R. et al. The interferon stimulated gene 15 functions as a proviral factor for the hepatitis C virus and as a regulator of the IFN response. *Gut* **59**, 1111–1119 (2010).
39. McLaren, P. J. et al. Identification of potential HIV restriction factors by combining evolutionary genomic signatures with functional analyses. *Retrovirology* **12**, 41 (2015).
40. Richardson, R. B. et al. A CRISPR screen identifies IFI6 as an ER-resident interferon effector that blocks flavivirus replication. *Nat. Microbiol.* **3**, 1214–1223 (2018).
41. Merkl P. E., Orzalli M. H., Knipe D. M. Mechanisms of Host IFI6, PML, and Daxx protein restriction of herpes simplex Virus 1 replication. *J. Virol.* **92**, e00057–18 (2018).
42. Jacquet, S., Pontier, D. & Etienne, L. Rapid evolution of HERC6 and duplication of a chimeric HERC5/6 gene in rodents and bats suggest an overlooked role of HERCs in Mammalian immunity. *Front Immunol.* **11**, 605270 (2020).
43. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
44. Haugh, K. A. et al. The interferon-inducible antiviral protein Daxx is not essential for interferon-mediated protection against avian sarcoma virus. *Virol. J.* **11**, 100 (2014).
45. Guo, K., Barrett, B. S., Mickens K. L., Hasenkrug K. J., Santiago M. L. Interferon Resistance of Emerging SARS-CoV-2 Variants. *BioRxiv Prepr Serv Biol.* 2021.
46. Mailliet, S. & Nisole, S. Daxx, a broad-spectrum viral restriction factor. *Virol. Montrouge Fr.* **20**, 261–272 (2016).
47. Kuo, H.-Y. et al. SUMO modification negatively modulates the transcriptional activity of CREB-binding protein via the recruitment of Daxx. *Proc. Natl. Acad. Sci. USA* **102**, 16973–16978 (2005).
48. Shih, H.-M., Chang, C.-C., Kuo, H.-Y. & Lin, D.-Y. Daxx mediates SUMO-dependent transcriptional control and subnuclear compartmentalization. *Biochem Soc. Trans.* **35**, 1397–1400 (2007).
49. Sudharsan, R. & Azuma, Y. The SUMO ligase PIAS1 regulates UV-induced apoptosis by recruiting Daxx to SUMOylated foci. *J. Cell Sci.* **125**, 5819–5829 (2012).
50. Jang, M.-S., Ryu, S.-W. & Kim, E. Modification of Daxx by small ubiquitin-related modifier-1. *Biochem Biophys. Res Commun.* **295**, 495–500 (2002).
51. Lin, D.-Y. et al. Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol. Cell* **24**, 341–354 (2006).
52. Huang, L. et al. DAXX represents a new type of protein-folding enabler. *Nature* **597**, 132–137 (2021).
53. Xie, X. et al. An Infectious cDNA Clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848.e3 (2020).
54. Liu, G. et al. ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat. Microbiol.* **6**, 467–478 (2021).
55. Shin, D. et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).
56. Mohamud, Y. et al. The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem Biophys. Res Commun.* **540**, 75–82 (2021).
57. Saiz M, Martinez-Salas E. Uncovering targets of the Leader protease: Linking RNA-mediated pathways and antiviral defense. *Wiley Interdiscip. Rev. RNA.* **12**, e1645 (2021).
58. Drayman, N. et al. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science* **373**, 931–936 (2021).
59. Miserey-Lenkei, S. et al. A comprehensive library of fluorescent constructs of SARS-CoV-2 proteins and their initial characterisation in different cell types. *Biol. Cell* **113**, 311–328 (2021).
60. Roesch, F., OhAinle, M. & Emerman, M. A CRISPR screen for factors regulating SAMHD1 degradation identifies IFITMs as potent inhibitors of lentiviral particle delivery. *Retrovirology* **15**, 26 (2018).
61. Shaw, A. E. et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol.* **15**, e2004086–e2004086 (2017).
62. Schreiner, S. et al. Control of human adenovirus type 5 gene expression by cellular Daxx/ATRAX chromatin-associated complexes. *Nucleic Acids Res.* **41**, 3532–3550 (2013).
63. Guion, L. G. & Sapp, M. The role of promyelocytic leukemia nuclear bodies during HPV infection. *Front Cell Infect. Microbiol.* **10**, 35 (2020).
64. Mahmoudvand, S. & Shokri, S. Interactions between SARS coronavirus 2 papain-like protease and immune system: A potential drug target for the treatment of COVID-19. *Scand. J. Immunol.* **94**, e13044 (2021).
65. Drosten, C. et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
66. Guery, B. et al. Clinical features and viral diagnosis of two cases of infection with Middle East Respiratory Syndrome coronavirus: a report of nosocomial transmission. *Lancet* **381**, 2265–2272 (2013).
67. Combredet, C. et al. A molecularly cloned Schwarz strain of measles virus vaccine induces strong immune responses in macaques and transgenic mice. *J. Virol.* **77**, 11546–11554 (2003).
68. Hoffmann, M. et al. Fusion-active glycoprotein G mediates the cytotoxicity of vesicular stomatitis virus M mutants lacking host shut-off activity. *J. Gen. Virol.* **91**, 2782–2793 (2010).
69. Cramer, J. et al. Sweet drugs for bad bugs: a glycomimetic strategy against the DC-SIGN-mediated dissemination of SARS-CoV-2. *J. Am. Chem. Soc.* **143**, 17465–17478 (2021).
70. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10 (2011).
71. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
72. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
73. Hsiao T, et al. Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv.* 2019;251082.
74. Le T, et al. BBrowse: Making single-cell data easily accessible. *bioRxiv.* 2020;2020.12.11.414136.

Acknowledgements

We thank the Cytometry Platform, Center for Technological Resources and Research, Institut Pasteur, for cell sorting experiments. This work was funded by the Institut Pasteur Coronavirus Task Force, CNRS (UMR 3569), the Labex IBEID (ANR-10-LABX-62-IBEID), the Occitanie region and by the ANR (ANR-20-COVI-000, projects IDIS-COVR to M.V. and Alpha-COV to S.N.). A.M.K. is supported by a grant of the French Ministry of Higher Education, Research and Innovation. G.M. is supported by a grant from the Agence nationale de recherches sur le sida et les hépatites virales (ANRS). S.M.A. is supported by the Pasteur-Paris University (PPU) International PhD Program. We thank Michael Emerman, Daniel Marc and Ignacio Caballero-Posadas for helpful comments on the manuscript. Figure 1 was created with BioRender.com.

Author contributions

F.R. designed the research project. F.R., N.J., S.N. and M.V. secured the funding for the study. A.M.K., S.M.A., A.H., N.A., S.N., G.M., D.Q.T., M.C., T.V. and F.R. performed and analyzed the in vitro experiments. F.P. produced the stocks of lentiviruses. J.C.S., J.O. and K.H. generated and validated KO cell lines. T.B. performed the single-cell RNAseq data analysis. S.M. and F.D. performed the SARS-CoV and MERS-CoV experiments. A.B. and E.S.L. performed the bio-informatic analyses of the CRISPR/Cas9 screen. G.Z. supplied the VSV* and VSV*ΔG-S viruses used for entry assays and share protocols. M.O., T.B., O.S., N.J., S.N., S.V.D.W. and M.V. analyzed the data and supervised the project. A.M.K., N.J., S.N. and F.R. wrote the manuscript. All authors edited the manuscript. Correspondence and requests for materials should be addressed to either M.V., N.J., S.N. or F.R.

Competing interests

J.C.S., J.O. and K.H. are employees and shareholders from Synthego Corporation. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30134-9>.

Correspondence and requests for materials should be addressed to Nolwenn Jouvenet, Sébastien Nisole, Marco Vignuzzi or Ferdinand Roesch.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

4. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing

David E. Gordon, Gwendolyn M. Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M. White, Matthew J. O'Meara, Veronica V. Rezelj, Jeffrey Z. Guo, Danielle L. Swaney, Tia A. Tummino, Ruth Hüttenhain, Robyn M. Kaake, Alicia L. Richards, Beril Tutuncuoglu, Helene Foussard, Jyoti Batra, Kelsey Haas, Maya Modak, Minkyu Kim, Paige Haas, Benjamin J. Polacco, Hannes Braberg, Jacqueline M. Fabius, Manon Eckhardt, Margaret Soucheray, Melanie J. Bennett, Merve Cakir, Michael J. McGregor, Qiongyu Li, Bjoern Meyer, Ferdinand Roesch, Thomas Vallet, Alice Mac Kain, Lisa Miorin, Elena Moreno, Zun Zar Chi Naing, Yuan Zhou, Shiming Peng, Ying Shi, Ziyang Zhang, Wenqi Shen, Ilsa T. Kirby, James E. Melnyk, John S. Chorba, Kevin Lou, Shizhong A. Dai, Inigo Barrio-Hernandez, Danish Memon, Claudia Hernandez-Armenta, Jiankun Lyu, Christopher J. P. Mathy, Tina Perica, Kala Bharath Pilla, Sai J. Ganesan, Daniel J. Saltzberg, Ramachandran Rakesh, Xi Liu, Sara B. Rosenthal, Lorenzo Calviello, Srivats Venkataramanan, Jose Liboy-Lugo, Yizhu Lin, Xi-Ping Huang, YongFeng Liu, Stephanie A. Wankowicz, Markus Bohn, Maliheh Safari, Fatima S. Ugur, Cassandra Koh, Nastaran Sadat Savar, **Quang Dinh Tran**, Djoshkun Shengjuler, Sabrina J. Fletcher, Michael C. O'Neal, Yiming Cai, Jason C. J. Chang, David J. Broadhurst, Saker Klippsten, Phillip P. Sharp, Nicole A. Wenzell, Duygu Kuzuoglu-Ozturk, Hao-Yuan Wang, Raphael Trenker, Janet M. Young, Devin A. Cavero, Joseph Hiatt, Theodore L. Roth, Ujjwal Rathore, Advait Subramanian, Julia Noack, Mathieu Hubert, Robert M. Stroud, Alan D. Frankel, Oren S. Rosenberg, Kliment A. Verba, David A. Agard, Melanie Ott, Michael Emerman, Natalia Jura, Mark von Zastrow, Eric Verdin, Alan Ashworth, Olivier Schwartz, Christophe d'Enfert, Shaeri Mukherjee, Matt Jacobson, Harmit S. Malik, Danica G. Fujimori, Trey Ideker, Charles S. Craik, Stephen N. Floor, James S. Fraser, John D. Gross, Andrej Sali, Bryan L. Roth, Davide Ruggero, Jack Taunton, Tanja Kortemme, Pedro Beltrao, Marco Vignuzzi, Adolfo García-Sastre, Kevan M. Shokat, Brian K. Shoichet & Nevan J. Krogan

DOI: <https://doi.org/10.1038/s41586-020-2286-9>

Contributions: I have helped with infection experiments presented in Figure 6a-d and Extended Data Figure 9.

A SARS-CoV-2 protein interaction map reveals targets for drug repurposing

<https://doi.org/10.1038/s41586-020-2286-9>

Received: 23 March 2020

Accepted: 22 April 2020

Published online: 30 April 2020

 Check for updates

A list of authors and affiliations appears at the end of the paper

A newly described coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is the causative agent of coronavirus disease 2019 (COVID-19), has infected over 2.3 million people, led to the death of more than 160,000 individuals and caused worldwide social and economic disruption^{1,2}. There are no antiviral drugs with proven clinical efficacy for the treatment of COVID-19, nor are there any vaccines that prevent infection with SARS-CoV-2, and efforts to develop drugs and vaccines are hampered by the limited knowledge of the molecular details of how SARS-CoV-2 infects cells. Here we cloned, tagged and expressed 26 of the 29 SARS-CoV-2 proteins in human cells and identified the human proteins that physically associated with each of the SARS-CoV-2 proteins using affinity-purification mass spectrometry, identifying 332 high-confidence protein–protein interactions between SARS-CoV-2 and human proteins. Among these, we identify 66 druggable human proteins or host factors targeted by 69 compounds (of which, 29 drugs are approved by the US Food and Drug Administration, 12 are in clinical trials and 28 are preclinical compounds). We screened a subset of these in multiple viral assays and found two sets of pharmacological agents that displayed antiviral activity: inhibitors of mRNA translation and predicted regulators of the sigma-1 and sigma-2 receptors. Further studies of these host-factor-targeting agents, including their combination with drugs that directly target viral enzymes, could lead to a therapeutic regimen to treat COVID-19.

SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA beta-coronavirus of the family *Coronaviridae*^{3,4}. Coronaviruses that infect humans historically included several common cold viruses, including hCoV-OC43, HKU and 229E⁵. However, over the past two decades, highly pathogenic human coronaviruses have emerged, including SARS-CoV in 2002, which is associated with 8,000 cases worldwide and a death rate of around 10%, and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012, which caused 2,500 confirmed cases and had a death rate of 36%. Infection with these highly pathogenic coronaviruses can result in acute respiratory distress syndrome, which may lead to a long-term reduction in lung function, arrhythmia or death. In comparison to MERS-CoV or SARS-CoV, SARS-CoV-2 has a lower case-fatality rate but spreads more efficiently⁶, making it difficult to contain. To devise therapeutic strategies to counteract SARS-CoV-2 infection and the associated COVID-19 pathology, it is crucial to understand how this coronavirus hijacks the host during infection, and to apply this knowledge to develop new drugs and repurpose existing ones.

Thus far, no clinically available antiviral drugs have been developed for SARS-CoV, SARS-CoV-2 or MERS-CoV. Clinical trials are ongoing for treatment of COVID-19 with the nucleoside-analogue RNA-dependent RNA polymerase (RdRP) inhibitor remdesivir⁷, and recent data suggest that a new nucleoside analogue may be effective against SARS-CoV-2 infection in laboratory animals⁸. Clinical trials using several vaccine candidates are also underway⁹, as are trials of repurposed compounds that inhibit the human protease TMPRSS2¹⁰. We believe that there is great potential in systematically exploring the host dependencies of

the SARS-CoV-2 virus to identify other host proteins that are already targeted by existing drugs. Therapies that target the host–virus interface, where the emergence of mutational resistance is arguably less likely, could potentially present durable, broad-spectrum treatment modalities¹¹. Unfortunately, limited knowledge of the molecular details of SARS-CoV-2 precludes a comprehensive evaluation of small-molecule candidates for host-directed therapies. We sought to address this gap by systematically mapping the interaction landscape between SARS-CoV-2 proteins and human proteins.

Cloning and expression of SARS-CoV-2 proteins

Sequence analysis of SARS-CoV-2 isolates suggests that the 30-kb genome encodes as many as 14 open-reading frames (ORFs). ORF1a and ORF1ab encode polyproteins, which are auto-proteolytically processed into 16 non-structural proteins (NSP1–NSP16) that form the replicase–transcriptase complex (Fig. 1a). The replicase–transcriptase complex consists of multiple enzymes, including the papain-like protease (NSP3), the main protease (NSP5), the NSP7–NSP8 primase complex, the primary RNA-dependent RNA polymerase (NSP12), a helicase–triphosphatase (NSP13), an exoribonuclease (NSP14), an endonuclease (NSP15) and N7- and 2′-O-methyltransferases (NSP10 and NSP16)^{11,12,13}. At the 3′ end of the viral genome, as many as 13 ORFs are expressed from 9 predicted sub-genomic RNAs. These include four structural proteins: spike (S), envelope (E), membrane (M) and nucleocapsid (N)¹³, and nine putative accessory factors^{1,12} (Fig. 1a). The SARS-CoV-2 genome is very similar to SARS-CoV. Although both viruses have an

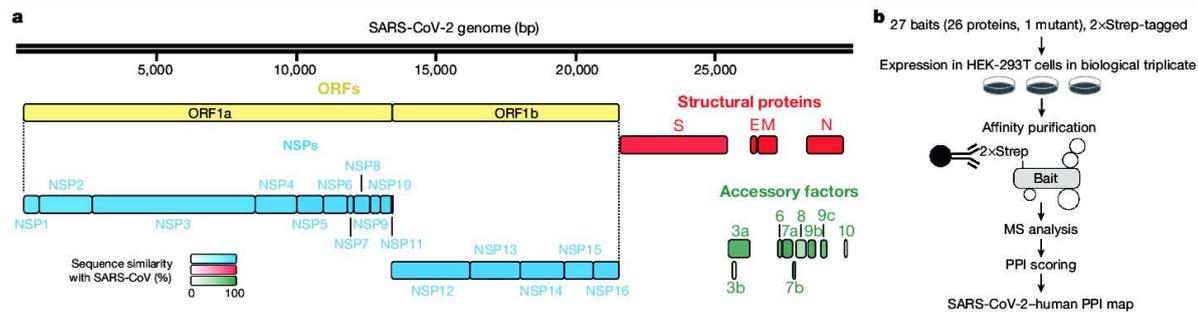


Fig. 1 | AP-MS workflow for the identification of SARS-CoV-2–host protein–protein interactions. **a**, SARS-CoV-2 genome annotation. The colour intensity is proportional to the protein sequence similarity with SARS-CoV homologues

(when homologues exist). $n = 4$ structural proteins; $n = 16$ NSPs; $n = 9$ accessory factors. **b**, Experimental workflow for AP-MS studies. MS, mass spectrometry; PPI, protein–protein interaction.

ORF1ab that encodes the 16 predicted NSPs as well as the four typical structural proteins of coronaviruses, they differ in their complement of 3' ORFs: SARS-CoV-2 possesses an ORF3b and ORF10, which have limited detectable homology to SARS-CoV proteins^{1,12} (Extended Data Fig. 1a).

Mature NSPs, with the exception of NSP3 and NSP16, and all predicted proteins expressed from other SARS-CoV-2 ORFs (27 proteins and one mutant) were codon-optimized and cloned into a mammalian expression vector that contained a 2×Strep-tag II affinity tag that can be used for affinity-purification–mass spectrometry (AP-MS)-based proteomics when expressed in HEK-293T/17 cells. High-confidence interacting proteins were identified using SAINTexpress and MiST scoring algorithms^{14,15}.

To verify the expression of viral proteins, we performed western blot using an anti-Strep antibody on the input cell lysate, and with the exception of NSP4, NSP6, NSP11, and ORF3b, we observed bands consistent with the predicted protein sizes (24 out of 28 constructs) (Extended Data Fig. 1b). Despite the lack of detection by western blot, we detected expression of viral peptides NSP4, NSP6 and ORF3b in the proteomic analysis. The fourth construct not confirmed by western blot, the small peptide NSP11, had a predicted molecular mass of 4.8 kDa (including tag) but an apparent mass of approximately 30 kDa (Extended Data Fig. 1b).

Alignment of 2,784 SARS-CoV-2 sequences revealed a premature stop codon at position 14 of ORF3b in 17.6% of isolates (Extended Data Fig. 1c), and two mutations were also observed that resulted in premature stop codons in ORF9c (Extended Data Fig. 1d). These data suggest that ORF3b and ORF9c might not be bona fide SARS-CoV-2 reading frames, or are dispensable for replication. Pending a comprehensive evaluation of viral protein expression, we nevertheless proceeded with the analysis for all possible viral proteins. Out of the 27 bait proteins (Fig. 1b), the affinity purification of ORF7b showed an unusually high number of background proteins and was therefore excluded from protein interaction analysis. We have thus far sent these plasmids to almost 300 laboratories in 35 countries.

Analysis of SARS-CoV-2–host protein interactions

Our AP-MS analysis identified 332 high-confidence protein interactions between SARS-CoV-2 proteins and human proteins, observing correlations between replicate experiments of each viral bait (Pearson's $R = 0.46–0.72$) (Extended Data Fig. 2 and Supplementary Tables 1, 2). We studied the interacting human proteins with regards to their biological functions, anatomical expression patterns, expression changes during SARS-CoV-2 infection¹⁶ and in relation to other maps of host-pathogen interacting proteins^{15,17} (Fig. 2a). We analysed each viral protein for Gene Ontology enrichment (Fig. 2b and Extended Data Fig. 3) and identified the major cell processes of the interacting proteins,

including lipoprotein metabolism (S), nuclear transport (NSP7) and ribonucleoprotein complex biogenesis (NSP8). To discover potential binding interfaces, we enriched for domain families within the interacting proteins of each viral bait (Extended Data Fig. 4). For instance, DNA polymerase domains are enriched among proteins that interact with NSP1, and bromodomains and extra-terminal domain (BET) family domains are enriched among proteins that interact with E (Supplementary Discussion and Supplementary Methods).

Although the cell line used for these AP-MS experiments, HEK-293T/17, can be infected with the SARS-CoV-2 virus¹⁸, it does not represent the primary physiological site of infection—lung tissue. From 29 human tissues¹⁹, we identified the lung as the tissue with the highest expression of the prey proteins relative to the average proteome (Fig. 2c). Consistent with this, the interacting proteins were enriched in the lung relative to other tissues (Extended Data Fig. 5a), and compared to overall RefSeq gene expression in the lung (median transcripts per million (TPM) = 3.198), proteins that interacted with SARS-CoV-2 proteins were expressed at a higher level (median TPM = 25.52, $P = 0.0007$; Student's t -test) (Extended Data Fig. 5b), supporting the hypothesis that SARS-CoV-2 preferentially hijacks proteins that are expressed in lung tissue.

We also studied the evolutionary properties of the host proteins bound by SARS-CoV-2 (Supplementary Table 3, Supplementary Methods and Supplementary Discussion). In addition, we analysed changes in protein abundance during SARS-CoV-2 infection¹⁶. We calculated, when possible, the correlation between changes in the abundance of viral proteins and their human interaction partners across four time points. Interacting pairs typically had stronger correlated changes than other pairs of viral–human proteins (Fig. 2d) (Kolmogorov–Smirnov test $P = 4.8 \times 10^{-5}$), indicating that the AP-MS-derived interactions are relevant for the target tissue and the infection context. We compared our SARS-CoV-2 interaction map with those of ten other pathogens (Fig. 2e) and found that West Nile virus²⁰ and *Mycobacterium tuberculosis*²¹ had the most similar host-protein interaction partners. The association with *M. tuberculosis* is of particular interest as it also infects lung tissue.

The interactome reveals SARS-CoV-2 biology

Our study highlights interactions between SARS-CoV-2 proteins and human proteins that are involved in several complexes and biological processes (Fig. 3). These included DNA replication (NSP1), epigenetic and gene-expression regulators (NSP5, NSP8, NSP13 and E), vesicle trafficking (NSP2, NSP6, NSP7, NSP10, NSP13, NSP15, ORF3a, E, M and ORF8), lipid modification (S), RNA processing and regulation (NSP8 and N), ubiquitin ligases (ORF10), signalling (NSP7, NSP8, NSP13, N and ORF9b), nuclear transport machinery (NSP9, NSP15 and ORF6),

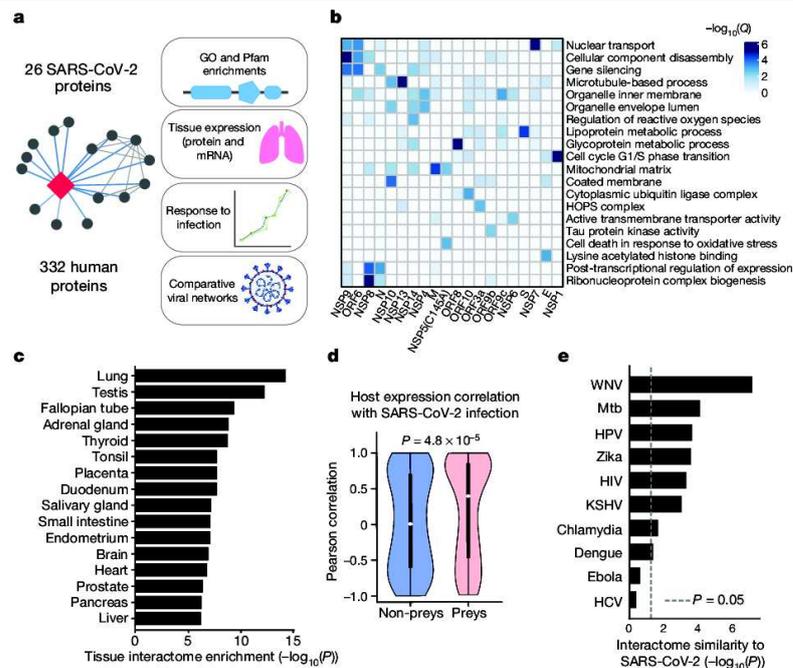


Fig. 2 | Global analysis of SARS-CoV-2 protein interactions. **a**, Overview of global analyses performed. **b**, Gene Ontology (GO) enrichment analysis was performed on the human interacting proteins of each viral protein. *P* values were calculated by hypergeometric test and a false-discovery rate was used to account for multiple hypothesis testing (Methods). The top GO term of each viral protein was selected for visualization. **c**, Degree of differential protein expression for the human interacting proteins ($n = 332$) across human tissues. We obtained protein abundance values for the proteome in 29 human tissues and calculated the median level of abundance for the human interacting proteins (top 16 tissues shown). This was then compared with the abundance values for the full proteome in each tissue and summarized as a Z-score from which a *P* value was calculated. A false-discovery rate was used to account for multiple hypothesis testing. **d**, The distribution of the correlation of protein level changes during SARS-CoV-2 infection for pairs of viral-human proteins

(median, white circles; interquartile range, black bars) is higher than non-interacting pairs of viral-human proteins ($P = 4.8 \times 10^{-5}$; Kolmogorov-Smirnov test). The violin plots show each viral-human protein correlation for preys ($n = 210$, minimum = -0.986 , maximum = 0.999 , quartile (Q)₁ = -0.468 , Q₂ = 0.396 , Q₃ = 0.850) and non-preys ($n = 54765$, minimum = -0.999 , maximum = 0.999 , Q₁ = -0.599 , Q₂ = 0.006 , Q₃ = 0.700). **e**, Significance of the overlap of human interacting proteins between SARS-CoV-2 and other pathogens using a hypergeometric test (unadjusted for multiple testing). The background gene set for the test consisted of all unique proteins detected by mass spectrometry across all pathogens ($n = 10,181$ proteins). HCV, hepatitis C virus; HIV, human immunodeficiency virus; HPV, human papillomavirus; KSHV, Kaposi's sarcoma-associated herpesvirus; Mtb, *M. tuberculosis*; WNV, West Nile virus.

cytoskeleton (NSP1 and NSP13), mitochondria (NSP4, NSP8 and ORF9c) and the extracellular matrix (NSP9).

Approximately 40% of SARS-CoV-2-interacting proteins were associated with endomembrane compartments or vesicle trafficking pathways. Host interactions with NSP8 (signal recognition particle (SRP)), ORF8 (protein quality control in the endoplasmic reticulum), M (morphology of the endoplasmic reticulum) and NSP13 (organization of the centrosome and Golgi) may facilitate the marked reconfiguration of endoplasmic reticulum and Golgi trafficking during coronavirus infection, and interactions in peripheral compartments with NSP2 (WASH), NSP6 and M (vacuolar ATPase), NSP7 (Rab proteins), NSP10 (AP2), E (AP3) and ORF3a (HOPS) may also modify endomembrane compartments to favour coronavirus replication. NSP6 and ORF9c interact with Sigma receptors that have been implicated in lipid remodelling and the stress response of the endoplasmic reticulum; these proteins interact with many human drugs (see 'Antiviral activity of host-directed compounds').

Trafficking into the endoplasmic reticulum and mitochondria may also be affected by the main protease of SARS-CoV-2, NSP5. We identified one high-confidence interaction between wild-type NSP5 and the epigenetic regulator histone deacetylase 2 (HDAC2), and predicted a cleavage site between the HDAC domain and the nuclear localization

sequence of HDAC2 (Extended Data Fig. 6a–d), suggesting that NSP5 may inhibit the transport of HDAC2 into the nucleus and could potentially affect the ability of HDAC2 to mediate the inflammation and interferon response^{22,23}. We also identified an interaction between catalytically dead NSP5(C145A) and tRNA methyltransferase 1 (TRMT1), which is responsible for the dimethylguanosine base modification (m₂2G) in both nuclear and mitochondrial tRNAs²⁴. We predict that TRMT1 is also cleaved by NSP5 (Extended Data Fig. 6a–d), leading to the removal of its zinc finger and nuclear localization signal and probably resulting in an exclusively mitochondrial localization of TRMT1.

SARS-CoV-2 interacts with innate immune pathways

Several innate immune signalling proteins are targeted by SARS-CoV-2 viral proteins. The interferon pathway is targeted by NSP13 (TBK1 and TBKBP1), NSP15 (RNF41 (also known as NRDPI)) and ORF9b (TOMM70); and the NF- κ B pathway is targeted by NSP13 (TLE1, TLE3 and TLE5) and ORF9c (NLRX1, F2RL1 and NDFIP2). Furthermore, two other E3 ubiquitin ligases that regulate antiviral innate immune signalling, TRIM59 and MIB1, are bound by ORF3a and NSP9, respectively^{25,26}.

We also identified interactions between SARS-CoV-2 ORF6 and the NUP98-RAE1 complex (Fig. 4a), an interferon-inducible mRNA

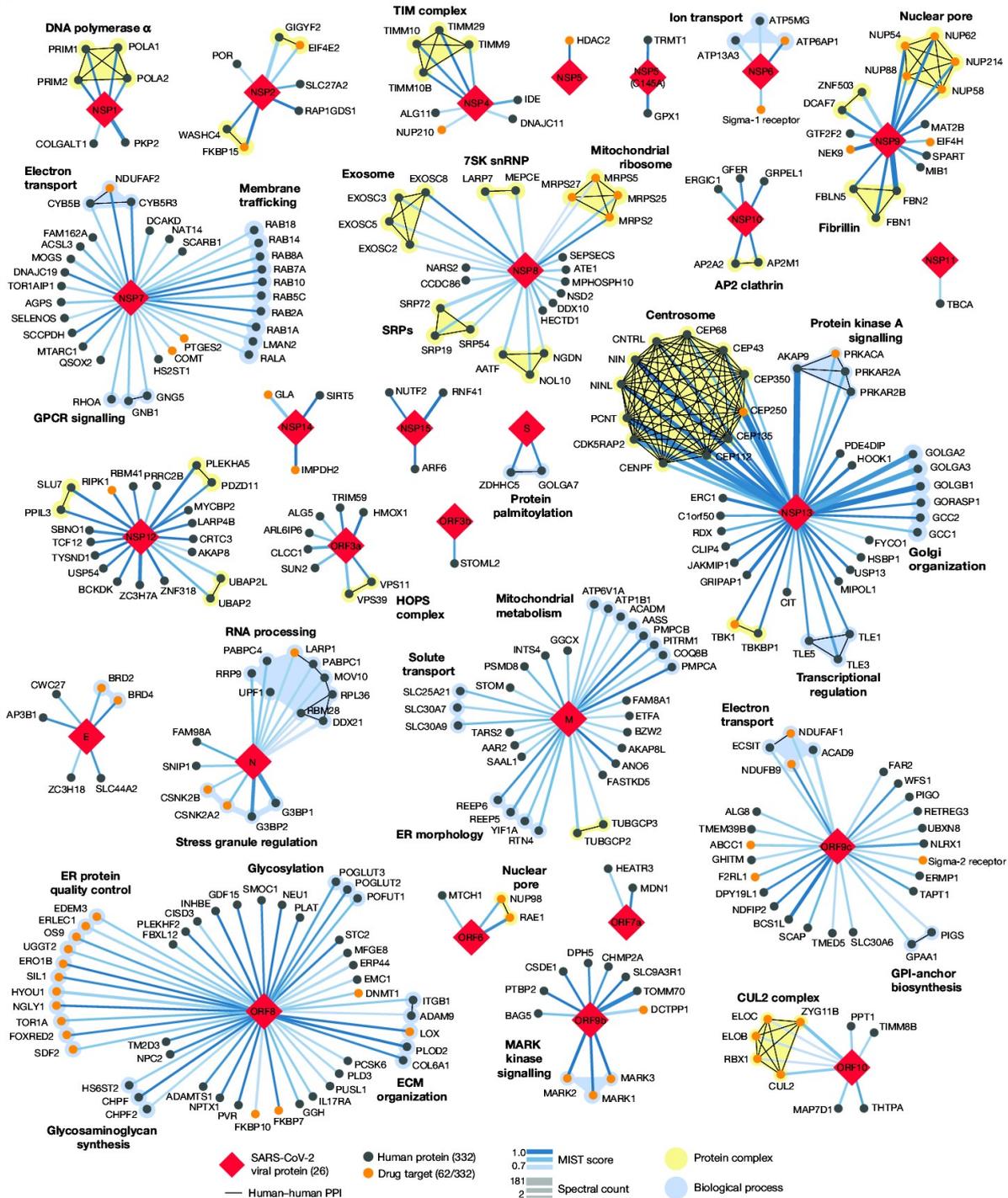


Fig. 3 | SARS-CoV-2 protein-protein interaction network. 332 high-confidence interactions between 26 SARS-CoV-2 proteins (red diamonds) and human proteins (circles; drug targets: orange; protein complexes: yellow; proteins in the same biological process: blue). Edge colour proportional to MIST score; edge thickness proportional to spectral counts. Physical interactions among

host proteins (thin black lines) were curated from CORUM, IntAct, and Reactome. An interactive protein-protein interaction map can be found at kroganlab.ucsf.edu/network-maps. ECM, extracellular matrix; ER, endoplasmic reticulum; snRNP, small nuclear ribonucleoprotein. $n = 3$ biologically independent samples.

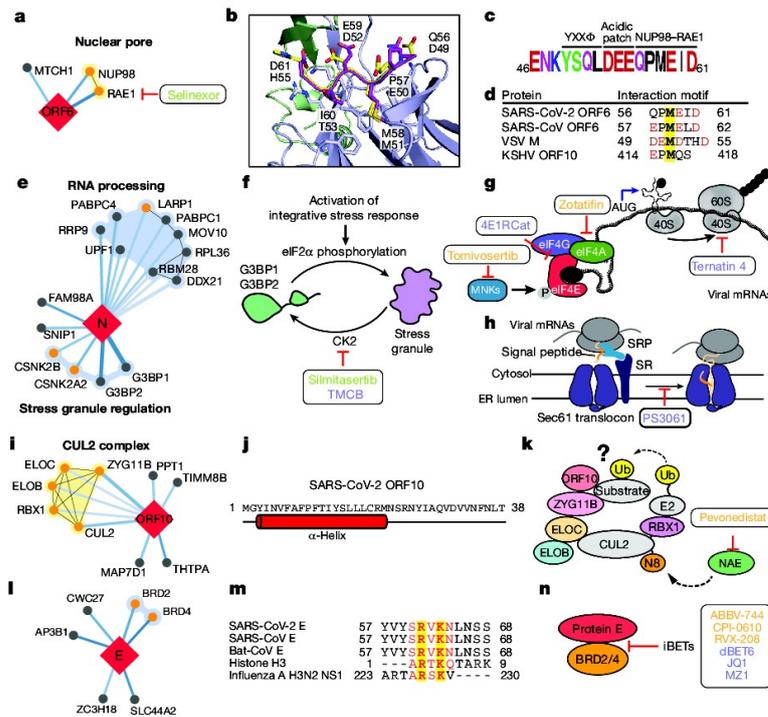


Fig. 4 | The SARS-CoV-2 interactome reveals novel aspects of SARS-CoV-2 biology and pharmacological targets. **a**, ORF6 interacts with an mRNA nuclear export complex that can be targeted by selinexor. **b**, The C-terminal peptide of SARS-CoV-2 ORF6 (dark purple) is modelled into the binding site of the VSV M protein (yellow)-NUP98 (green)-RAE1 (light purple) complex (Protein Data Bank (PDB) ID: 4OWR). ORF6 and M protein residues are labelled. **c**, The C-terminal sequence of SARS-CoV-2 ORF6, highlighting the described trafficking motifs and putative NUP98-RAE1 binding sequence. The chemical properties of amino acids are shown as follows: polar, green; neutral, purple; basic, blue; acidic, red; and hydrophobic, black. **d**, Putative NUP98-RAE1 interaction motifs. Negatively charged residues (red) surround a conserved methionine (yellow) in several virus species. **e**, N targets stress granule proteins. **f**, Inhibition of casein kinase II (by sinitasertib or TMCB) disrupts stress granules. **g**, Inhibition of translation initiation. The MNK inhibitor tomivosertib prevents phosphorylation of eIF4E. 4E1Rcat blocks the

interaction of eIF4E with eIF4G. Inhibition of eIF4A (zotatifin) may prevent the unwinding of the viral 5' untranslated region to prevent its translation. **h**, Targeting the translation elongation factor-1A ternary complex (ternatin-4) or the Sec61 translocon (PS3061) may prevent viral protein production and membrane insertion, respectively. **i**, ORF10 interacts with the CUL2^{ZYG11B} complex. **j**, Predicted secondary structure of ORF10. **k**, ORF10 possibly hijacks CUL2^{ZYG11B} for the ubiquitination (Ub) of host proteins, which can be inhibited by pevonedistat. **l**, The protein E interacts with bromodomain proteins. **j**, Alignment of E proteins from SARS-CoV-2, SARS-CoV and bat-CoV with histone H3 and NS1 protein of influenza A H3N2. Identical and similar amino acids are highlighted. **m**, Bromodomain inhibitors (iBETs) may disrupt the interaction between E and BRDs. **a**, **f**-**h**, **k**, **n**, FDA-approved drugs are shown in green, clinical candidates are shown in orange and preclinical candidates are shown in purple.

nuclear export complex²⁷ that is hijacked or degraded by multiple viruses including vesicular stomatitis virus (VSV), influenza A, Kaposi's sarcoma-associated herpesvirus and poliovirus, and is a restriction factor for influenza A infection²⁸⁻³¹. The X-ray structure of the VSV M protein complexed with NUP98-RAE1³² reveals key binding interactions, including a buried methionine residue on the M protein that packs into a hydrophobic pocket in RAE1, and neighbouring acidic residues that interact with a basic patch on the NUP98-RAE1 complex³². These features are also present in a conserved motif in the C-terminal region of SARS-CoV-2 ORF6 (Fig. 4b-d and Extended Data Fig. 7a, b), providing a structural hypothesis for the observed interaction. ORF6 of SARS-CoV antagonizes host interferon signalling by perturbing nuclear transport³³, and the NUP98-RAE1 interaction with ORF6 may perform the same function for SARS-CoV-2.

SARS-CoV-2 interacts with host translation machinery

Nucleocapsid (N) of SARS-CoV-2 binds to the stress granule proteins G3BP1 and G3BP2, and to other host mRNA-binding proteins including the mTOR-regulated translational repressor LARP1, two subunits

of casein kinase 2 (CK2), and mRNA decay factors UPF1 and MOV10 (Fig. 4e). Manipulation of the stress granule and related RNA biology is common among *Coronaviridae*³⁴⁻³⁶ and stress granule formation is thought to be a primarily antiviral response. The promotion of G3BP aggregation by eIF4A inhibitors^{28,37} may partially explain their antiviral activity (see 'Antiviral activity of host-directed compounds').

All coronavirus mRNAs rely on cap-dependent translation to produce their proteins, a process enhanced *in trans* by the SARS-CoV N protein³⁸. Key eIF4-cap binding complex constituents—the cap binding protein eIF4E, scaffold protein eIF4G and the DEAD-box helicase eIF4A—are candidates for therapeutic targeting of coronaviruses^{39,40}. Therapeutic targeting (Fig. 4f, g) of viral translation by interfering with the eIF4F complex formation or the interactions between viral proteins N, NSP2 or NSP8 and the translational machinery may have therapeutic benefits (see 'Antiviral activity of host-directed compounds').

Cotranslational entry into the secretory pathway is a potential target for SARS-CoV-2 inhibition. Up to ten SARS-CoV-2 proteins are predicted to undergo insertion into the membrane of the endoplasmic reticulum mediated by the Sec61 translocon, which localizes to SARS-CoV replication complexes⁴¹. Furthermore, high-confidence interactions

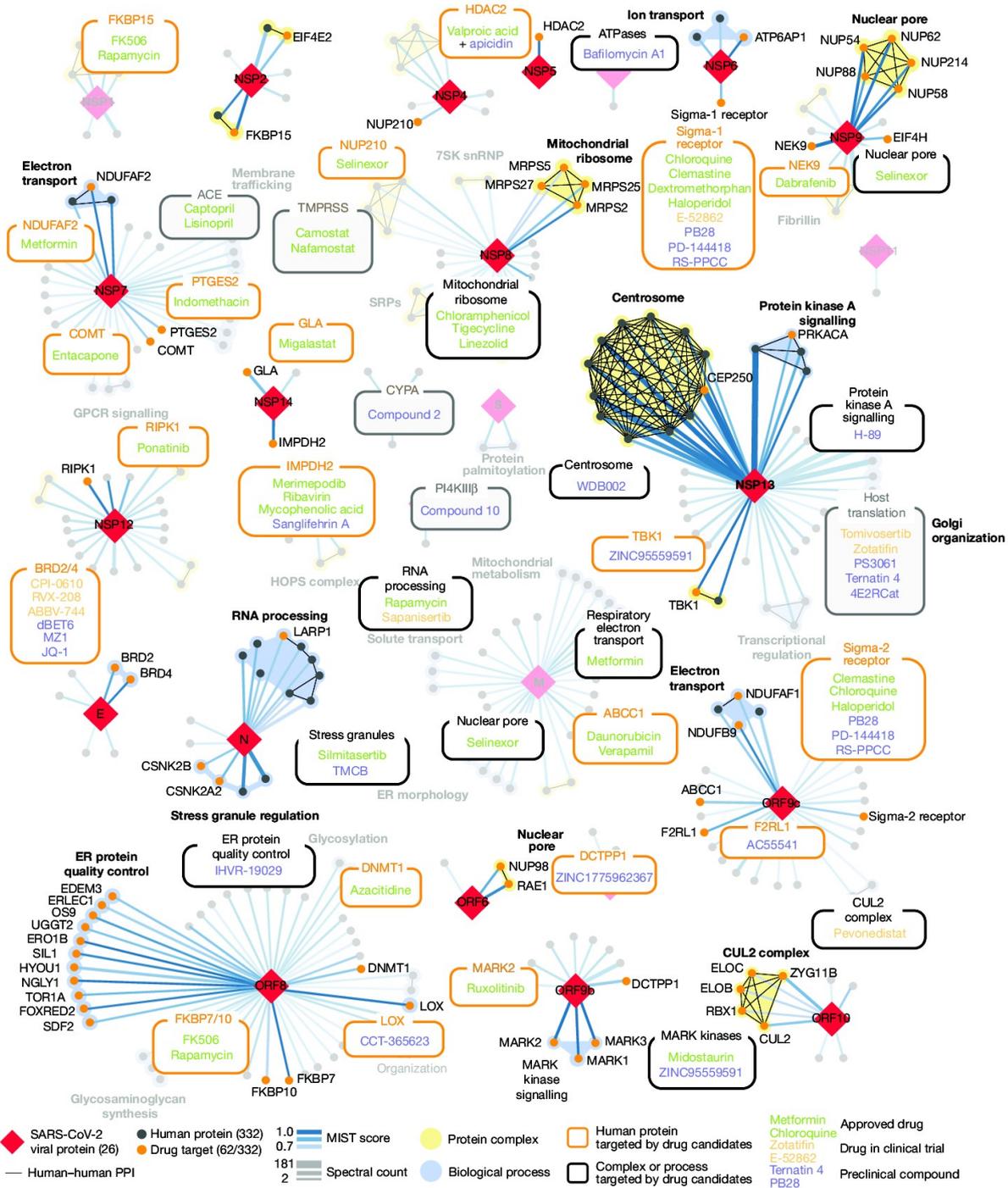


Fig. 5 | Drug-human target network. Protein-protein interactions of SARS-CoV-2 baits with approved drugs (green), clinical candidates (orange) and preclinical candidates (purple) with experimental activities against the host proteins (white background) or previously known host factors (grey background) are shown.

between NSP8 and three SRP components suggest that the virus hijacks the Sec61-mediated protein translocation pathway for entry into the endoplasmic reticulum. Sec61 inhibitors of protein biogenesis such

as PS3061 (Fig. 4h), which has previously been shown to inhibit other enveloped RNA viruses^{42,43}, may also block SARS-CoV-2 replication and assembly.

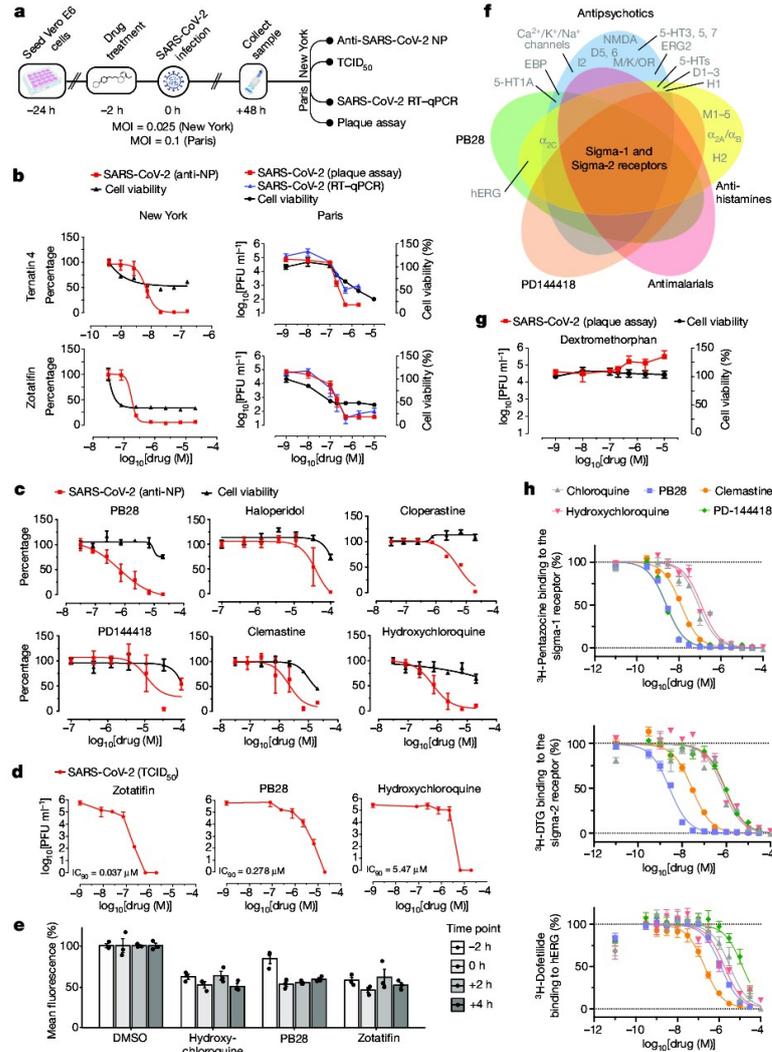


Fig. 6 | The antiviral activity of the translation inhibitors and Sigma receptor ligands. **a**, Schematic of viral infectivity assays. **b**, The mRNA translation inhibitors (zotatifin and ternatin-4) reduce viral infectivity in a concentration-dependent manner. Red, viral infectivity (anti-NP or plaque assays); blue, RT-qPCR; black, cell viability. The initial decrease in cell viability probably reflects cytostatic and not cytotoxic effects. Data are mean \pm s.d.; $n = 6$ biologically independent samples for cell viability data and DMSO controls from Paris; all other data, $n = 3$. PFU, plaque-forming unit. **c**, Sigma receptor drugs and preclinical molecules inhibit viral infectivity (coloured as in **b**). Data are mean \pm s.d.; $n = 3$ biologically independent samples. **d**, IC_{50} assays using zotatifin, PB28 and hydroxychloroquine. **e**, Drugs added before or after high-titre virus ($MOI = 2$) addition had similar antiviral effects. Data are mean \pm s.d.; $n = 3$ biologically independent samples. **f**, Sigma-1 and sigma-2 receptors are the common targets of the sigma ligands at the 1- μ M activity

threshold⁵⁵. **g**, Dextromethorphan increases viral titres. Red, viral titre (plaque assay); black, cell viability. Data are mean \pm s.d.; $n = 3$ biologically independent samples. **h**, On-target K_d values for the sigma-1 and sigma-2 receptor compared with the K_d values for the hERG ion channel. PB28 and PD-144418 show 500- to 5,000-fold selectivity, whereas chloroquine and hydroxychloroquine have about 30-fold selectivity for these targets. pK_i values for hERG compared with sigma-1 receptor compared with sigma-2 receptor are: chloroquine (5.5 ± 0.1 ; 7.1 ± 0.1 ; 6.3 ± 0.1); hydroxychloroquine (5.6 ± 0.2 ; 6.9 ± 0.2 ; 6.0 ± 0.1); PB28 (6.0 ± 0.1 ; 8.7 ± 0.1 ; 8.6 ± 0.1); PD-144418 (5.0 ± 0.2 ; 8.7 ± 0.1 ; 6.1 ± 0.1); clemastine (6.8 ± 0.2 ; 8.0 ± 0.1 ; 7.6 ± 0.1). Data are mean \pm s.d.; PB28, clemastine, PD-144418, $n = 9$ biologically independent samples for sigma-1 and sigma-2 receptors and hERG; chloroquine, hydroxychloroquine, $n = 6$ for sigma-1 and sigma-2 receptors and $n = 4$ for hERG.

SARS-CoV-2 interacts with a Cullin ubiquitin ligase

Viruses commonly hijack ubiquitination pathways for replication and pathogenesis⁴⁴. The ORF10 of SARS-CoV-2 interacts with members of a cullin-2 (CUL2) RING E3 ligase complex (Fig. 4i), specifically the CUL2^{ZYG11B} complex. ZYG11B is the highest scoring protein in the ORF10

interactome, suggesting that there is a direct interaction between ORF10 and ZYG11B. Despite its small size (38 amino acids), ORF10 appears to contain an α -helical region (Fig. 4j) that may be adopted in complex with CUL2^{ZYG11B}. The ubiquitin transfer to a substrate requires neddylation of CUL2 by NEDD8-activating enzyme (NAE), which is a druggable target⁴⁵ (Fig. 4k). ORF10 may bind to the CUL2^{ZYG11B} complex

Article

and hijack it for ubiquitination and degradation of restriction factors, or alternatively, ZYG11B may bind to the N-terminal glycine in ORF10 to target it for degradation³¹.

SARS-CoV-2 interacts with bromodomain proteins

We found that the transmembrane E protein, which is probably resident on the endoplasmic reticulum–Golgi intermediate compartment and Golgi membranes, binds to BRD2 and BRD4 (Fig. 4l), members of the bromodomain and extra-terminal (BET) domain family of epigenetic readers that bind to acetylated histones to regulate gene transcription⁴⁶. The C-terminal region of E mimics the N-terminal segment of histone H3, which is a known interacting partner of bromodomains⁴⁷. Notably, this region of E is highly conserved in SARS and bat coronaviruses, which suggests that it has a conserved function (Fig. 4m). A similar short peptide motif has also been identified in the NS1 protein of the influenza A H3N2 strain, in which it interferes with transcriptional processes that support an antiviral response^{47,48}. Bromodomain inhibitors might disrupt the interaction between protein E and BRDs (Fig. 4n).

For a more comprehensive overview of virus–host interactions, see Supplementary Discussion and Supplementary Methods.

Identification of drugs that target host factors

To disrupt the SARS-CoV-2 interactome, we sought ligands of human proteins that interact with viral proteins (Methods). Molecules were prioritized by the MiST score of the interaction between the human and viral proteins; by their status as approved drugs, investigational drugs (drugs in clinical trials) or as preclinical candidates; by their selectivity; and by their availability (Supplementary Tables 4, 5). Chemoinformatics searches from the IUPHAR/BPS Guide to Pharmacology (2020-3-12) and the ChEMBL25 database on the human interactors yielded 16 approved drugs, 3 investigational drugs and 18 pre-clinical candidates (Supplementary Table 4); and target- and pathway-specific literature search revealed 13 approved drugs, 9 investigational drugs and 10 pre-clinical candidates (Supplementary Table 5). Of the 332 human targets that interact with the viral bait proteins with a high-confidence score (Fig. 3), 62 have 69 drugs, investigational drugs or preclinical molecules that modulate them and can be overlaid on our protein–interaction network (Fig. 5).

Antiviral activity of host-directed compounds

We next investigated the antiviral activity of these drugs and compounds, using two viral assays (Fig. 6a). First, at Mount Sinai Hospital in New York, we developed a medium-throughput immunofluorescence-based assay (which detects the viral NP protein) to screen 37 compounds for inhibition of SARS-CoV-2 infection in the Vero E6 cell line. Second, at the Institut Pasteur in Paris, viral RNA was monitored using quantitative PCR with reverse transcription (RT–qPCR) after treatment with 44 drugs and compounds. Together, both locations tested 47 of the 69 compounds that we identified, plus 13 to expand testing of the sigma-1 and sigma-2 receptors and mRNA translation targets, and 15 additional molecules that had been prioritized by other methods (Methods and Supplementary Table 6). Viral growth and cytotoxicity were monitored at both institutions (Extended Data Figs. 8, 9 and Supplementary Table 6). Two classes of molecules emerged as effectively reducing viral infectivity: protein biogenesis inhibitors (zotatifin, ternatin-4 and PS3061) (Fig. 6b and Extended Data Fig. 9) and ligands of the sigma-1 and sigma-2 receptors (haloperidol, PB28, PD-144418 and hydroxychloroquine, which is undergoing clinical trials in patients with COVID-19 (ClinicalTrials.gov, trial number NCT04332991)). We also subsequently found that the sigma-1- and sigma-2-receptor active drugs clemastine, cloperastine and progesterone, and the clinical molecule siramesine, were antiviral drugs (Fig. 6c and Extended Data Fig. 9). Median tissue

culture infectious dose (TCID₅₀) assays on supernatants from infected cells treated with PB28 (90% inhibitory concentration (IC₉₀) = 0.278 μM) and zotatifin (IC₉₀ = 0.037 μM) revealed a more potent inhibition than was observed in the NP-staining assay (Fig. 6d). Notably, in this assay, PB28 was around 20 times more potent than hydroxychloroquine (IC₉₀ = 5.78 μM).

To better understand the mechanism by which these inhibitors exert their antiviral effects, we performed a time course assay in which the drugs were added at different times before or after infection (Fig. 6e). Cells were infected during a single cycle of infection at high multiplicity of infection (MOI = 2) over the course of 8 h, and the drugs were added either 2 h before infection or at 0, 2 or 4 h after infection. PB28, zotatifin and hydroxychloroquine all decreased the detection of the viral NP protein even in this single cycle assay, indicating that the antiviral effect occurs before viral egress from the cell (Fig. 6e). Furthermore, all three molecules inhibited NP expression when added up to 4 h after infection, after viral entry has occurred. Thus, these molecules seem to exert their antiviral effect during viral replication.

Coronaviruses rely on cap-dependent mRNA translation through the translation machinery of the host. eIF4H, which interacts with NSP9, is a partner of eIF4A, and we observed a strong antiviral effect after treatment with the eIF4A inhibitor zotatifin (Fig. 6b), which is currently in a phase-I clinical trial for the treatment of cancer. We also observed potent antiviral effects of the elongation factor-1A (eEF1A) inhibitor ternatin-4⁴⁹ (Fig. 6b), which may suggest that the rate of translation elongation is critical for obtaining optimal levels of viral proteins. Of note, the eEF1A inhibitor aplidin/plitidepsin is used clinically in patients with multiple myeloma⁵⁰. Multiple SARS-CoV-2 proteins are predicted to undergo SRP- and Sec61-mediated co-translational insertion into the endoplasmic reticulum, and SRP19, SRP54 and SRP72 were identified as NSP8-interacting proteins (Fig. 3). Consistent with previous studies of flaviviruses⁴², the Sec61 inhibitor PS3061 also blocked SARS-CoV-2 replication (Extended Data Fig. 9). The two translation inhibitors had cytostatic effects in uninfected Vero cells, which are immortalized cell lines with indefinite proliferative capacity that have mutations in key cell cycle inhibitors. These cells are more sensitive to anti-cancer compounds, which affect the cell cycle state of immortalized cells more strongly than non-immortalized cells. A critical question going forward is whether these or related inhibitors of viral protein biogenesis will show therapeutic benefits in patients with COVID-19. Plitidepsin is currently under consideration by the Spanish Medicines Agency for a phase-II trial in hospitalized patients with COVID-19.

Molecules that target the sigma-1 and sigma-2 receptors perturb the virus through different mechanisms than the translation inhibitors, which could include the cell stress response⁵¹. These molecules are also active against other aminergic receptors; however, the only targets shared among all of the tested molecules are the sigma-1 and sigma-2 receptors (Fig. 6f), into which these drugs can be readily modelled (Extended Data Fig. 10a). For instance, the antipsychotic drug haloperidol inhibits the dopamine D2 and histamine H1 receptors, whereas clemastine and cloperastine are antihistamines; each of these drugs is also a Sigma receptor ligand with antiviral activity (Fig. 6c). Conversely, the antipsychotic drug olanzapine, which also inhibits histamine H1 and dopamine D2 receptors, has little Sigma receptor activity and does not show antiviral activity (Extended Data Fig. 10b). Which of the Sigma receptors is responsible for the activity remains uncertain, as does the role of pharmacologically related targets, such as EBP and related sterol isomerases, the ligand recognition of which resembles those of the Sigma receptors. Notably, the sigma-1-receptor benzomorphan agonist dextromethorphan has proviral activity (Fig. 6g), further supporting the role of these receptors in viral infection. Overall, two features should be emphasized. First, several of the molecules that target the Sigma receptors, such as clemastine, cloperastine and progesterone, are approved drugs with a long history in human therapy. Many other widely used drugs, which show activity against the Sigma receptors,

remain to be tested; and indeed, several drugs such as astemizole, which we show is a sigma-2 receptor ligand (with an K_i of 95 nM) (Extended Data Fig. 11), verapamil and amiodarone, have been reported by others to be active in viral replication assays, although this has not been linked to their Sigma receptor activity^{52,53}. Second, the Sigma receptor ligands have a clear separation between antiviral and cytotoxic effects (Fig. 6b, c), and ligands such as PB28 have substantial selectivity for the Sigma receptors compared with side-effect targets, such as the hERG ion channel. Indeed, the lack of selectivity of chloroquine and hydroxychloroquine for hERG (Fig. 6h) and other off-targets (Extended Data Fig. 12) may be related to the adverse cardiac drug reactions⁵⁴ that have limited their use.

Discussion

In this study, we have identified 332 high-confidence SARS-CoV-2 protein–human protein interactions that are connected with multiple biological processes, including protein trafficking, translation, transcription and regulation of ubiquitination. We found 69 ligands, including FDA-approved drugs, compounds in clinical trials and pre-clinical compounds, that target these interactions. Antiviral tests in two different laboratories reveal two broad sets of active drugs and compounds; those that affect translation and those modulate the sigma-1 and sigma-2 receptors. Within these sets are at least five targets and more than ten different chemotypes, providing a rich landscape for optimization.

The chemo-proteomic analysis that emerges from this study not only highlights clinically actionable drugs that target human proteins in the interactome, but also provides a context for interpreting their mechanism of action. The potent efficacy of the translation inhibitors on viral infectivity—in the 10 to 100 nM range—makes these molecules attractive as candidate antiviral agents, and also highlights this pathway as a point of intervention. Although the mechanism of action of the drugs that target the sigma-1 and sigma-2 receptors remains less defined, their activity as both anti- and proviral agents is mechanistically suggestive. The relatively strong efficacy of PB28, at an IC_{50} of 280 nM in the viral titre assay, and its high selectivity against off-target proteins, suggests that molecules of this class may be optimized as therapeutic agents. Although it is unclear whether approved drugs such as clemastine and cloperastine, which are used as antihistamines and antitussive drugs, have pharmacokinetics that are suitable for antiviral therapy, and although they are not free of binding to targets that cause side effects (Fig. 6f and Extended Data Fig. 12), these drugs have been used for decades. We caution against their use outside of controlled studies, because of their side-effect liabilities. By the same standard, we find that the widely used antitussive drug dextromethorphan has proviral activity and that therefore its use should merit caution and further study in the context of the treatment of COVID-19. More positively, there are dozens of approved drugs that show activity against the Sigma receptors, not all of which are generally recognized as Sigma receptor ligands. Many of these drugs remain to be tested as a treatment for COVID-19; although some have begun to appear in other studies^{52,53}. This area of pharmacology has great promise for the repurposing and optimization of new agents in the fight against COVID-19.

Our approach of host-directed intervention as an antiviral strategy overcomes problems associated with drug resistance and may also provide panviral therapies as we prepare for the next pandemic. Furthermore, the possibilities for cotherapies are expanded—for example, with drugs that directly target the virus, including remdesivir—and, as we demonstrate in this study, there are numerous opportunities for the repurposing of FDA-approved drugs. More broadly, the pipeline described here represents an approach for drug discovery not only for panviral strategies, but also for the research of many diseases, and illustrates the speed with which science can be moved forward using a multi-disciplinary and collaborative approach.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2286-9>.

1. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
2. WHO. Coronavirus disease (COVID-2019) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (2020).
3. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473 (2020).
4. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
5. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).
6. Gates, B. Responding to Covid-19 — a once-in-a-century pandemic? *N. Engl. J. Med.* **382**, 1677–1679 (2020).
7. Sheahan, T. P. et al. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nat. Commun.* **11**, 222 (2020).
8. Sheahan, T. P. et al. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronavirus in mice. *Sci. Transl. Med.* **12**, eabb5883 (2020).
9. Paton, J. Moderna's coronavirus vaccine trial set to begin this month. *Bloomberg News* (6 March 2020).
10. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).
11. Prussia, A., Thepchatri, P., Snyder, J. P. & Plemper, R. K. Systematic approaches towards the development of host-directed antiviral therapeutics. *Int. J. Mol. Sci.* **12**, 4027–4052 (2011).
12. Chan, J. F.-W. et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **9**, 221–236 (2020).
13. Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **1282**, 1–23 (2015).
14. Teo, G. et al. SAINTEpress: improvements and additional features in Significance Analysis of INteractome software. *J. Proteomics* **100**, 37–43 (2014).
15. Jäger, S. et al. Global landscape of HIV–human protein complexes. *Nature* **481**, 365–370 (2012).
16. Bojkova, D. et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* <https://doi.org/10.1038/s41586-020-2332-7> (2020).
17. Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R. & Krogan, N. J. A systems approach to infectious disease. *Nat. Rev. Genet.* **21**, 339–354 (2020).
18. Harcourt, J. et al. Severe acute respiratory syndrome coronavirus 2 from patient with coronavirus disease, United States. *Emerg. Infect. Dis.* **26**, 1266–1273 (2020).
19. Wang, D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
20. Li, M. et al. Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection. *Nat. Microbiol.* **4**, 985–995 (2019).
21. Penn, B. H. et al. An *Mtb*-human protein–protein interaction map identifies a switch between host antiviral and antibacterial responses. *Mol. Cell* **71**, 637–648 (2018).
22. Barnes, P. J. Role of HDAC2 in the pathophysiology of COPD. *Annu. Rev. Physiol.* **71**, 451–464 (2009).
23. Xu, P. et al. NOST1 inhibits the interferon response of cancer cells by S-nitrosylation of HDAC2. *J. Exp. Clin. Cancer Res.* **38**, 483 (2019).
24. Dewe, J. M., Fuller, B. L., Lentini, J. M., Kellner, S. M. & Fu, D. TRMT1-catalyzed tRNA modifications are required for redox homeostasis to ensure proper cellular proliferation and oxidative stress survival. *Mol. Cell. Biol.* **37**, e00214-17 (2017).
25. Kondo, T., Watanabe, M. & Hatakeyama, S. TRIM59 interacts with ECSIT and negatively regulates NF- κ B and IRF-3/7-mediated signal pathways. *Biochem. Biophys. Res. Commun.* **422**, 501–507 (2012).
26. Li, S., Wang, L., Berman, M., Kong, Y.-Y. & Dorf, M. E. Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. *Immunity* **35**, 426–440 (2011).
27. Faria, P. A. et al. VSV disrupts the Rae1/mrnp41 mRNA nuclear export pathway. *Mol. Cell* **17**, 93–102 (2005).
28. Slaine, P. D., Kleer, M., Smith, N. K., Khapersky, D. A. & McCormick, C. Stress granule-inducing eukaryotic translation initiation factor 4A inhibitors block influenza A virus replication. *Viruses* **9**, 388 (2017).
29. Reineke, L. C. et al. Casein kinase 2 is linked to stress granule dynamics through phosphorylation of the stress granule nucleating protein G3BP1. *Mol. Cell. Biol.* **37**, e00596-16 (2017).
30. Kindrachuk, J. et al. Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinase analysis. *Antimicrob. Agents Chemother.* **59**, 1088–1099 (2015).
31. Timms, R. T. et al. A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* **365**, eaaw4912 (2019).
32. Quan, B., Seo, H.-S., Blobel, G. & Ren, Y. Vesiculoviral matrix (M) protein occupies nucleic acid binding site at nucleoporin pair (Rae1-Nup98). *Proc. Natl Acad. Sci. USA* **111**, 9127–9132 (2014).

33. Frieman, M. et al. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J. Virol.* **81**, 9812–9824 (2007).

34. Nakagawa, K., Narayanan, K., Wada, M. & Makino, S. Inhibition of stress granule formation by Middle East respiratory syndrome coronavirus 4a accessory protein facilitates viral translation, leading to efficient virus replication. *J. Virol.* **92**, e00902-18 (2018).

35. Raaben, M., Groot Koerkamp, M. J. A., Rottier, P. J. M. & de Haan, C. A. M. Mouse hepatitis coronavirus replication induces host translational shutoff and mRNA decay, with concomitant formation of stress granules and processing bodies. *Cell. Microbiol.* **9**, 2218–2229 (2007).

36. Ivanov, P., Kedersha, N. & Anderson, P. Stress granules and processing bodies in translational control. *Cold Spring Harb. Perspect. Biol.* **11**, a032813 (2019).

37. Thompson, P. A. et al. Abstract 2698: eFT226, a potent and selective inhibitor of eIF4A, is efficacious in preclinical models of lymphoma. *Cancer Res.* **79**, 2698 (2019).

38. Nakagawa, K., Lokugamage, K. G. & Makino, S. Viral and cellular mRNA translation in coronavirus-infected cells. *Adv. Virus Res.* **96**, 165–192 (2016).

39. Müller, C. et al. Broad-spectrum antiviral activity of the eIF4A inhibitor silvestrol against corona- and picornaviruses. *Antiviral Res.* **150**, 123–129 (2018).

40. Cencic, R. et al. Blocking eIF4E–eIF4G interaction as a strategy to impair coronavirus replication. *J. Virol.* **85**, 6381–6389 (2011).

41. Knoops, K. et al. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* **6**, e226 (2008).

42. Shah, P. S. et al. Comparative flavivirus–host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. *Cell* **175**, 1931–1945 (2018).

43. Heaton, N. S. et al. Targeting viral proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* **44**, 46–58 (2016).

44. Mahon, C., Krogan, N. J., Craik, C. S. & Pick, E. Cullin E3 ligases and their rewiring by viral factors. *Biomolecules* **4**, 897–930 (2014).

45. Soucy, T. A. et al. An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **458**, 732–736 (2009).

46. Faivre, E. J. et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **578**, 306–310 (2020).

47. Filippakopoulos, P. et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).

48. Marazzi, I. et al. Suppression of the antiviral response by an influenza histone mimic. *Nature* **483**, 428–433 (2012).

49. Carelli, J. D. et al. Ternatin and improved synthetic variants kill cancer cells by targeting the elongation factor-1A ternary complex. *eLife* **4**, e10222 (2015).

50. Spicka, I. et al. Randomized phase III study (ADMYRE) of plitidepsin in combination with dexamethasone vs. dexamethasone alone in patients with relapsed/refractory multiple myeloma. *Ann. Hematol.* **98**, 2139–2150 (2019).

51. Mitsuda, T. et al. Sigma-1Rs are upregulated via PERK/eIF2α/ATF4 pathway and execute protective function in ER stress. *Biochem. Biophys. Res. Commun.* **415**, 519–525 (2011).

52. Si, L. et al. Human organs-on-chips as tools for repurposing approved drugs as potential influenza and COVID19 therapeutics in viral pandemics. Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.13.039917v1> (2020).

53. Riva, L. et al. A large-scale drug repositioning survey for SARS-CoV-2 antivirals. Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.16.044016v1> (2020).

54. White, N. J. Cardiotoxicity of antimalarial drugs. *Lancet Infect. Dis.* **7**, 549–558 (2007).

55. Armstrong, J. F. et al. The IUPHAR/BPS Guide to pharmacology in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to malaria pharmacology. *Nucleic Acids Res.* **48**, D1006–D1021 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

David E. Gordon^{1,2,3,4,25}, Gwendolyn M. Jang^{1,2,3,4,35}, Mehdi Bouhaddou^{1,2,3,4,35}, Jiewei Xu^{1,2,3,4,35}, Kirsten Obernier^{1,2,3,4,35}, Kris M. White^{5,6,25}, Matthew J. O'Meara^{7,35}, Veronica V. Rezelj^{8,35}, Jeffrey Z. Guo^{1,2,3,4}, Danielle L. Swaney^{1,2,3,4}, Tia A. Tummino^{1,2,9}, Ruth Hüttenhain^{1,2,3,4}, Robyn M. Kaake^{1,2,3,4}, Alicia L. Richards^{1,2,3,4}, Beril Tutuncuoglu^{1,2,3,4}, Helene Foussard^{1,2,3,4}, Jyoti Batra^{1,2,3,4}, Kelsey Haas^{1,2,3,4}, Maya Modak^{1,2,3,4}, Minkyu Kim^{1,2,3,4}, Paige Haas^{1,2,3,4}, Benjamin J. Polacco^{1,2,3,4}, Hannes Braberg^{1,2,3,4}, Jacqueline M. Fabius^{1,2,3,4}, Manon Eckhardt^{1,2,3,4}, Margaret Souchery^{1,2,3,4}, Melanie J. Bennett^{1,2,3,4}, Merve Cakir^{1,2,3,4}, Michael J. McGregor^{1,2,3,4}, Qiongyu Li^{1,2,3,4}, Björn Meyer⁹, Ferdinand Roesch⁹, Thomas Vallet⁹, Alice Mac Kain⁹, Lisa Miorin^{5,6}, Elena Moreno^{5,6}, Zun Zar Chi Naing^{1,2,3,4}, Yuan Zhou^{1,2,3,4}, Shiming Peng^{1,2,9}, Ying Shi^{1,2,4,30}, Ziyang Zhang^{1,2,4,30}, Wenqi Shen^{1,2,4,30}, Ilsa T. Kirby^{1,2,4,30}, James E. Melnyk^{1,2,4,30}, John S. Chorba^{1,2,4,30}, Kevin Lou^{1,2,4,30}, Shizhong A. Dai^{1,2,4,30}, Inigo Barrio-Hernandez¹⁰, Danish Memon¹¹,

Claudia Hernandez-Armenta¹¹, Jiankun Lyu^{1,2,9}, Christopher J. P. Mathy^{1,2,12,33}, Tina Perica^{1,2,12}, Kala Bharath Pilla^{1,2,12}, Sai J. Ganesan^{1,2,12}, Daniel J. Saltzberg^{1,2,12}, Ramachandran Rakesh^{1,2,12}, Xi Liu^{1,2,9}, Sara B. Rosenthal¹⁴, Lorenzo Calviello¹⁵, Srivats Venkataramanan¹⁵, Jose Liboy-Lugo¹⁵, Yizhu Lin¹⁵, Xi-Ping Huang¹⁶, YongFeng Liu¹⁶, Stephanie A. Wankowicz^{1,2,12,17}, Markus Bohn^{1,2,9}, Malihah Safari^{1,2,18}, Fatima S. Ugur^{1,2,4,9}, Cassandra Koh⁸, Nastaran Sadat Sava⁸, Quang Dinh Tran⁸, Djoshkun Shengjuler⁸, Sabrina J. Fletcher⁸, Michael C. O'Neal¹⁹, Yiming Cai¹⁹, Jason C. J. Chang¹⁹, David J. Broadhurst¹⁹, Saker Klippsten¹⁹, Phillip P. Sharp⁴, Nicole A. Wenzell^{1,2,4}, Duygu Kuzuoglu-Ozturk^{1,20,21}, Hao-Yuan Wang^{1,2,4}, Raphael Trenke^{1,2,22}, Janet M. Young²³, Devin A. Caverso^{1,3,24}, Joseph Hiatt^{1,3,25}, Theodore L. Roth^{1,3,24,25}, Ujjwal Rathore^{1,3,24}, Advait Subramanian^{1,2,24}, Julia Noack^{1,2,24}, Mathieu Hubert²⁶, Robert M. Stroud^{1,2,38}, Alan D. Frankel^{1,2,38}, Oren S. Rosenberg^{1,2,38,27}, Kliment A. Verba^{1,2,9}, David A. Agard^{1,2,30}, Melanie Ott^{1,2,3,27}, Michael Emerman²⁸, Natalia Jura^{1,2,4,22}, Mark von Zastrow^{1,2,4,29}, Eric Verdin^{1,2,7,30}, Alan Ashworth^{1,2,30}, Olivier Schwartz²⁹, Christophe d'Enfert³⁰, Shaeri Mukherjee^{1,2,34}, Matt Jacobson^{1,2,9}, Harmit S. Malik²³, Danica G. Fujimori^{1,2,4,9}, Trey Ideker^{1,32}, Charles S. Craik^{1,2,9,20}, Stephen N. Floor^{1,15,20}, James S. Fraser^{1,2,32}, John D. Gross^{1,2,9}, Andrej Sali^{1,2,9,32}, Bryan L. Roth¹⁸, Davide Ruggero^{1,2,4,20,21}, Jack Taunton^{1,2,4}, Tanja Kortemme^{1,1,2,12,33}, Pedro Beltrao¹³, Marco Vignuzzi¹³, Adolfo Garcia-Sastre^{5,6,33,34}, Kevan M. Shokat^{1,2,4,30}, Brian K. Shoichet^{1,2,33} & Nevan J. Krogan^{1,2,3,4,5,33}

¹QBI COVID-19 Research Group (QCRG), San Francisco, CA, USA. ²Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, CA, USA. ³J. David Gladstone Institutes, San Francisco, CA, USA. ⁴Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA. ⁵Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁶Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁷Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. ⁸Viral Populations and Pathogenesis Unit, CNRS UMR 3569, Institut Pasteur, Paris, France. ⁹Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA. ¹⁰Howard Hughes Medical Institute, University of California San Francisco, San Francisco, CA, USA. ¹¹European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK. ¹²Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. ¹³The UC Berkeley-UCSF Graduate Program in Bioengineering, University of California San Francisco, San Francisco, CA, USA. ¹⁴Center for Computational Biology and Bioinformatics, Department of Medicine, University of California San Diego, San Diego, CA, USA. ¹⁵Department of Cell and Tissue Biology, University of California San Francisco, San Francisco, CA, USA. ¹⁶Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. ¹⁷Biophysics Graduate Program, University of California San Francisco, San Francisco, CA, USA. ¹⁸Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA. ¹⁹Zoic Labs, Culver City, CA, USA. ²⁰Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA. ²¹Department of Urology, University of California San Francisco, San Francisco, CA, USA. ²²Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA. ²³Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ²⁴George William Hooper Foundation, Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA, USA. ²⁵Medical Scientist Training Program, University of California San Francisco, San Francisco, CA, USA. ²⁶Virus and Immunity Unit, Institut Pasteur, Paris, France. ²⁷Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ²⁸Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ²⁹Department of Psychiatry, University of California San Francisco, San Francisco, CA, USA. ³⁰Buck Institute for Research on Aging, Novato, CA, USA. ³¹Direction Scientifique, Institut Pasteur, Paris, France. ³²Division of Genetics, Department of Medicine, University of California San Diego, San Diego, CA, USA. ³³Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³⁴The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³⁵These authors contributed equally: David E. Gordon, Gwendolyn M. Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M. White, Matthew J. O'Meara, Veronica V. Rezelj. ✉e-mail: marco.vignuzzi@pasteur.fr; Adolfo.Garcia-Sastre@mssm.edu; Kevan.Shokat@ucsf.edu; shoichet@cgl.ucsf.edu; nevan.krogan@ucsf.edu

Methods

Genome annotation

The GenBank sequence for SARS-CoV-2 isolate 2019-nCoV/USA-WA1/2020, accession MN985325, was downloaded on 24 January 2020. In total, we annotated 29 possible ORFs and proteolytically mature proteins encoded by SARS-CoV-2¹¹². Proteolytic products that result from NSP3- and NSP5-mediated cleavage of the ORF1a/ORF1ab polyprotein were predicted on the basis of the protease specificity of SARS-CoV proteases⁵⁶, and 16 predicted nonstructural proteins (NSPs) were subsequently cloned (NSP1–NSP16). For the NSP5 protease (3CLike/3CLpro), we also designed the catalytically dead mutant NSP5(C145A)^{57,58}. ORFs at the 3' end of the viral genome annotated in the original GenBank file included 4 structural proteins: S, E, M, N and the additional ORFs ORF3a, ORF6, ORF7a, ORF8 and ORF10. On the basis of the analysis of ORFs in the genome and comparisons with other annotated SARS-CoV ORFs, we annotated a further four ORFs: ORF3b, ORF7b, ORF9b and ORF9c.

Cloning

ORFs and proteolytically mature NSPs annotated in the SARS-CoV-2 genome were human codon-optimized using the IDT codon-optimization tool (<https://www.idtdna.com/codonopt>) and internal EcoRI and BamHI sites were eliminated. Start and stop codons were added as necessary to NSPs 1–16, a Kozak sequence was added before each start codon, and a 2×Strep tag with linker was added to either the N or C terminus. To guide our tagging strategy, we used GPS-Lipid to predict protein lipid modification of the termini (<http://lipid.biocuckoo.org/webserver.php>)^{59,60}, TMHMM Server v.2.0 to predict transmembrane/hydrophobic regions (<http://www.cbs.dtu.dk/services/TMHMM/>)⁶¹ and SignalP v.5.0 to predict signal peptides (<http://www.cbs.dtu.dk/services/SignalP/>)⁶². IDT gBlocks were ordered for all reading frames with 15-bp overlaps that corresponded to flanking sequences of the EcoRI and BamHI restriction sites in the lentiviral constitutive expression vector pLVX-EF1alpha-IRES-Puro (Takara). Vectors were digested and gel-purified, and gene fragments were cloned using InFusion (Takara). The S protein was synthesized and cloned into pTwist-EF1alpha-IRES-Puro (Twist Biosciences). NSP16 showed multiple mutations that could not be repaired before the time-sensitive preparation of this manuscript, and NSP3 was too large to be synthesized in time to be included in this study. Strep-tagged constructs encoding NSP3, NSP3(C857A) (catalytically dead mutant) and NSP16 will be used in future AP-MS experiments.

Cell culture

HEK-293T/17 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Corning) supplemented with 10% fetal bovine serum (FBS; Gibco, Life Technologies) and 1% penicillin–streptomycin (Corning) and maintained at 37 °C in a humidified atmosphere of 5% CO₂. HEK-293T/17 cells were procured from the UCSF Cell Culture Facility, now available through UCSF's Cell and Genome Engineering Core (<https://cgec.ucsf.edu/cell-culture-and-banking-services>); cell line collection listed here: <https://ucsf.app.box.com/s/6xkydeqhr8a2xes0mbo2333i3k1ndqv> (CCLZR076). STR analysis by the Berkeley Cell Culture Facility on 8 August 2017 authenticated HEK-293T/17 cells with 94% probability. Cells were tested on 3 July 2019 using the MycoAlert Mycoplasma Detection Kit (Lonza LT07-318) and were negative: B/A ratio < 1 (no detected mycoplasma).

Transfection

For each affinity purification (26 wild-type baits and one catalytically dead SARS-CoV-2 bait, one GFP control and one empty vector control), ten million HEK-293T/17 cells were plated per 15-cm dish and transfected with up to 15 µg of individual Strep-tagged expression constructs after 20–24 h. Total plasmid was normalized to 15 µg with

empty vector and complexed with PolyJet Transfection Reagent (SigmaGen Laboratories) at a 1:3 µg:µl ratio of plasmid:transfection reagent based on the manufacturer's recommendations. After more than 38 h, cells were dissociated at room temperature using 10 ml Dulbecco's phosphate-buffered saline without calcium and magnesium (DPBS) supplemented with 10 mM EDTA for at least 5 min and subsequently washed with 10 ml DPBS. Each step was followed by centrifugation at 200g, 4 °C for 5 min. Cell pellets were frozen on dry ice and stored at –80 °C. For each bait, *n* = 3 independent biological replicates were prepared for affinity purification.

Affinity purification

Frozen cell pellets were thawed on ice for 15–20 min and resuspended in 1 ml lysis buffer (IP buffer (50 mM Tris-HCl, pH 7.4 at 4 °C, 150 mM NaCl, 1 mM EDTA) supplemented with 0.5% Non-idet P40 substitute (NP40; Fluka Analytical) and cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Roche)). Samples were then frozen on dry ice for 10–20 min and partially thawed at 37 °C before incubation on a tube rotator for 30 min at 4 °C and centrifugation at 13,000g, 4 °C for 15 min to pellet debris. After reserving 50 µl lysate, up to 48 samples were arrayed into a 96-well Deepwell plate for affinity purification on the KingFisher Flex Purification System (Thermo Scientific) as follows: MagStrep 'type3' beads (30 µl; IBA Lifesciences) were equilibrated twice with 1 ml wash buffer (IP buffer supplemented with 0.05% NP40) and incubated with 0.95 ml lysate for 2 h. Beads were washed three times with 1 ml wash buffer and then once with 1 ml IP buffer. To directly digest bead-bound proteins as well as elute proteins with biotin, beads were manually suspended in IP buffer and divided in half before transferring to 50 µl denaturation–reduction buffer (2 M urea, 50 mM Tris-HCl pH 8.0, 1 mM DTT) and 50 µl 1× buffer BXT (IBA Lifesciences) dispensed into a single 96-well KF microtitre plate. Purified proteins were first eluted at room temperature for 30 min with constant shaking at 1,100 rpm on a ThermoMixer C incubator. After removing eluates, on-bead digestion proceeded (see 'On-bead digestion'). Strep-tagged protein expression in lysates and enrichment in eluates were assessed by western blot and silver stain, respectively. The KingFisher Flex Purification System was placed in the cold room and allowed to equilibrate to 4 °C overnight before use. All automated protocol steps were performed using the slow mix speed and the following mix times: 30 s for equilibration and wash steps, 2 h for binding and 1 min for final bead release. Three 10-s bead collection times were used between all steps.

On-bead digestion

Bead-bound proteins were denatured and reduced at 37 °C for 30 min and after being brought to room temperature, alkylated in the dark with 3 mM iodoacetamide for 45 min and quenched with 3 mM DTT for 10 min. Proteins were then incubated at 37 °C, initially for 4 h with 1.5 µl trypsin (0.5 µg/µl; Promega) and then another 1–2 h with 0.5 µl additional trypsin. To offset evaporation, 15 µl 50 mM Tris-HCl, pH 8.0 were added before trypsin digestion. All steps were performed with constant shaking at 1,100 rpm on a ThermoMixer C incubator. Resulting peptides were combined with 50 µl 50 mM Tris-HCl, pH 8.0 used to rinse beads and acidified with trifluoroacetic acid (0.5% final, pH < 2.0). Acidified peptides were desalted for MS analysis using a BioPureSPE Mini 96-Well Plate (20 mg PROTO 300 C18; The Nest Group) according to standard protocols.

MS data acquisition and analysis

Samples were resuspended in 4% formic acid, 2% acetonitrile solution, and separated by a reversed-phase gradient over a Nanoflow C18 column (Dr Maisch). Each sample was directly injected via an Easy-nLC 1200 (Thermo Fisher Scientific) into a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific) and analysed with a 75 min acquisition, with all MS1 and MS2 spectra collected in the orbitrap; data were

Article

acquired using the Thermo software Xcalibur (4.2.47) and Tune (2.11 QF1 Build 3006). For all acquisitions, QCloud was used to control instrument longitudinal performance during the project⁶². All proteomic data were searched against the human proteome (Uniprot-reviewed sequences downloaded 28 February 2020), the eGFP sequence and the SARS-CoV-2 protein sequences using the default settings for MaxQuant (v.1.6.11.0)^{64,65}. Detected peptides and proteins were filtered to 1% false-discovery rate in MaxQuant, and identified proteins were then subjected to protein–protein interaction scoring with both SAINTexpress (v.3.6.3)¹⁴ and MiST (<https://github.com/kroganlab/mist>)^{15,66}. We applied a two-step filtering strategy to determine the final list of reported interactors, which relied on two different scoring stringency cut-offs. In the first step, we chose all protein interactions that had a MiST score ≥ 0.7 , a SAINTexpress Bayesian false-discovery rate (BFDR) ≤ 0.05 and an average spectral count ≥ 2 . For all proteins that fulfilled these criteria, we extracted information about the stable protein complexes that they participated in from the CORUM⁶⁷ database of known protein complexes. In the second step, we then relaxed the stringency and recovered additional interactors that (1) formed complexes with interactors determined in filtering step 1 and (2) fulfilled the following criteria: MiST score ≥ 0.6 , SAINTexpress BFDR ≤ 0.05 and average spectral counts ≥ 2 . Proteins that fulfilled filtering criteria in either step 1 or step 2 were considered to be high-confidence protein–protein interactions (HC-PPIs) and visualized with Cytoscape (v.3.7.1)⁶⁸. Using this filtering criteria, nearly all of our baits recovered a number of HC-PPIs in close alignment with previous datasets reporting an average of around 6 PPIs per bait⁶⁹. However, for a subset of baits (ORF8, NSP8, NSP13 and ORF9c), we observed a much higher number of PPIs that passed these filtering criteria. For these four baits, the MiST scoring was instead performed using a larger in-house database of 87 baits that were prepared and processed in an analogous manner to this SARS-CoV-2 dataset. This was done to provide a more comprehensive collection of baits for comparison, to minimize the classification of non-specifically binding background proteins as HC-PPIs. All MS raw data and search results files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018117^{70,71}. PPI networks have also been uploaded to NDEX.

GO overrepresentation analysis

The targets of each bait were tested for enrichment of GO biological process terms. The overrepresentation analysis was based on the hypergeometric distribution and performed using the enricher function of clusterProfiler package in R with default parameters. The GO terms were obtained from the c5 category of Molecular Signature Database (MSigDBv6.1). Significant GO terms (1% false-discovery rate) were identified and further refined to select non-redundant terms. To select non-redundant gene sets, we first constructed a GO term tree based on distances (1 – Jaccard similarity coefficients of shared genes) between the significant terms. The GO term tree was cut at a specific level ($h = 0.99$) to identify clusters of non-redundant gene sets. For a bait with multiple significant terms belonging to the same cluster, we selected the broadest term that is, largest gene set size.

Virus interactome similarity analysis

Interactome similarity was assessed by comparing the number of shared human-interacting proteins between pathogen pairs, using a hypergeometric test to calculate significance. The background gene set for the test consisted of all unique proteins detected by MS across all pathogens ($n = 10,181$ genes).

ORF6 peptide modelling

The proposed interaction between ORF6 and the NUP98–RAE1 complex was modelled in PyRosetta 4 (release v.2020.02-dev61090)⁷² using the crystal structure of VSV matrix (M) protein bound to NUP98–RAE1 as

a template³² (PDB 4OWR; downloaded from the PDB-REDO server⁷³). The M protein chain (C) was truncated after residue 54 to restrict the model to the putative interaction motif in ORF6 (M protein residues 49–54, sequence DEMDTH). These residues were mutated to the ORF6 sequence, QPMEID, using the mutate_residue function in the module pyrosetta.toolbox, without repacking at this initial step. After all six residues were mutated, the full model was relaxed to a low-energy conformation using the FastRelax protocol in the module pyrosetta.rosetta.protocols.relax. FastRelax was run with constraints to starting coordinates and scored with the ref2015 score function. The resulting model was inspected for any large energetic penalties associated with the modelled peptide residues or those NUP98 and RAE1 residues interacting with the peptide, and was found to have none. The model was visualized in PyMOL (The PyMOL Molecular Graphics System, v.3.4, Schrödinger).

ORF10 secondary structure prediction

The secondary structure of ORF10 was predicted using JPRED (<https://www.compbio.dundee.ac.uk/jpred/index.html>)⁷⁴.

Protein E alignment

Protein E sequences from SARS-CoV-2 (YP_009724392.1), SARS-CoV (NP_828854.1) and bat SARS-like CoV (AGZ48809.1) were aligned using Clustal Omega⁷⁵, and then manually aligned to the sequences of histone H3 (P68431) and influenza A H3N2 NS1 (YP_308845.1).

Chemoinformatic analysis of SARS-CoV-2-interacting partners

To identify drugs and reagents that modulate the 332 host factors that interact with SARS-CoV-2 and HEK-293T/17 cells (MiST ≥ 0.70), we used two approaches: (1) a chemoinformatic analysis of open-source chemical databases and (2) a target- and pathway-specific literature search, drawing on specialist knowledge within our group. Chemoinformatically, we retrieved 2,472 molecules from the IUPHAR/BPS Guide to Pharmacology (2020-3-12)⁵⁵ (Supplementary Table 7) that interacted with 30 human ‘prey’ proteins (38 approved, 71 in clinical trials), and found 10,883 molecules (95 approved, 369 in clinical trials) from the ChEMBL25 database⁷⁶ (Supplementary Table 8). For both approaches, molecules were prioritized on their FDA approval status, activity at the target of interest better than 1 μM and commercial availability, drawing on the ZINC database⁷⁷. FDA-approved molecules were prioritized except when clinical candidates or preclinical research molecules had substantially better selectivity or potency on-target. In some cases, we considered molecules with indirect mechanisms of action on the general pathway of interest based solely on literature evidence (for example, captopril modulates ACE2 indirectly via its direct interaction with angiotensin-converting enzyme, ACE). Finally, we predicted 6 additional molecules (2 approved, 1 in clinical trials) for proteins with MiST scores between 0.7 and 0.6 to viral baits (Supplementary Tables 4, 5). Complete methods can be found at <https://github.com/momeara/BioChemPantry/tree/master/vignette/COVID19>.

Molecular docking

After their chemoinformatic assignment to the sigma-1 receptor, cloperastine and clemastine were docked into the agonist-bound state structure of the receptor (6DK1)⁷⁸ using DOCK3.7⁷⁹. The best scoring configurations that ion pair with Glu172 are shown; both l-cloperastine and clemastine receive solvation-corrected docking scores between -42 and -43 kcal/mol, indicating high complementarity.

Viral growth and cytotoxicity assays in the presence of inhibitors

For studies carried out at Mount Sinai, SARS-CoV-2 (isolate USA-WA1/2020 from BEI RESOURCES NR-52281) was propagated in Vero E6 cells. Two thousand Vero E6 cells were seeded into 96-well plates in DMEM (10% FBS) and incubated for 24 h at 37 °C, 5% CO₂.

Vero E6 cells used were purchased from ATCC and thus authenticated (VERO C1008 (Vero 76, clone E6, Vero E6) (ATCC CRL-1586); tested negative for mycoplasma contamination before commencement). Then, 2 h before infection, the medium was replaced with 100 μ l of DMEM (2% FBS) containing the compound of interest at concentrations 50% greater than those indicated, including a DMSO control. The Vero E6 cell line used in this study is a kidney cell line; therefore, we cannot exclude that lung cells yield different results for some inhibitors (see also 'Cells and viruses', 'Antiviral activity assays', 'Cell viability assays' and 'Plaque-forming assays' for studies carried out at Institut Pasteur). Plates were then transferred into the Biosafety Level 3 (BSL3) facility and 100 PFU (MOI = 0.025) was added in 50 μ l of DMEM (2% FBS), bringing the final compound concentration to those indicated. Plates were then incubated for 48 h at 37 °C. After infection, supernatants were removed and cells were fixed with 4% formaldehyde for 24 h before being removed from the BSL3 facility. The cells were then immunostained for the viral NP protein (anti-sera produced in the Garcia-Sastre laboratory; 1:10,000) with a DAPI counterstain. Infected cells (488 nm) and total cells (DAPI) were quantified using the Celigo (Nexcelcom) imaging cytometer. Infectivity is measured by the accumulation of viral NP protein in the nucleus of the Vero E6 cells (fluorescence accumulation). Percentage infection was quantified as ((infected cells/total cells) – background) \times 100 and the DMSO control was then set to 100% infection for analysis. The IC₅₀ and IC₉₀ for each experiment were determined using the Prism (GraphPad) software. For some inhibitors, infected supernatants were assayed for infectious viral titres using the TCID₅₀ method. For this, infectious supernatants were collected at 48 h after infection and frozen at –80 °C until later use. Infectious titres were quantified by limiting dilution titration on Vero E6 cells. In brief, Vero E6 cells were seeded in 96-well plates at 20,000 cells per well. The next day, SARS-CoV-2-containing supernatant was applied at serial tenfold dilutions ranging from 10⁻¹ to 10⁻⁶ and, after 5 days, viral cytopathogenic effect was detected by staining cell monolayers with crystal violet. The TCID₅₀/ml values were calculated using the previously described method⁸⁰. Cytotoxicity was also performed using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected Vero E6 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates.

Cells and viruses

For studies at the Institut Pasteur, African green monkey kidney epithelial Vero E6 cells (ATCC, CRL-1586, authenticated by ATCC and tested negative for mycoplasma contamination before commencement (Vero 76, clone E6, Vero E6) (ATCC CRL-1586)) were maintained in a humidified atmosphere at 37 °C with 5% CO₂, in DMEM containing 10% (v/v) FBS (Invitrogen) and 5 units/ml penicillin and 5 μ g/ml streptomycin (Life Technologies). The Vero E6 cell line used in this study is a kidney cell line; therefore, we cannot exclude that lung cells yield different results for some inhibitors (see also 'Viral growth and cytotoxicity assays in the presence of inhibitors' for studies performed at Mount Sinai). SARS-CoV-2, isolate France/IDF0372/2020, was supplied by the National Reference Centre for Respiratory Viruses hosted by the Institut Pasteur and headed by S. van der Werf. The human sample from which strain BetaCoV/France/IDF0372/2020 was isolated has been provided by X. Lescure and Y. Yazdanpanah from the Bichat Hospital, Paris, France. The BetaCoV/France/IDF0372/2020 strain was supplied through the European Virus Archive goes Global (Evag) platform, a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no. 653316. Viral stocks were prepared by propagation in Vero E6 cells in DMEM supplemented with 2% FBS and 1 μ g/ml TPCK-trypsin (Sigma-Aldrich). Viral titres were determined by plaque-forming assay in minimum essential medium supplemented with 2% (v/v) FBS (Invitrogen) and 0.05% agarose. All experiments involving live SARS-CoV-2

were performed at the Institut Pasteur in compliance with the guidelines of the Institut Pasteur following BSL3 containment procedures in approved laboratories. All experiments were performed in at least three biologically independent samples.

Antiviral activity assays

Vero E6 cells were seeded at 1.5×10^4 cells per well in 96-well plates 18 h before the experiment. Then, 2 h before infection, the cell-culture supernatant of triplicate wells was replaced with medium containing 10 μ M, 2 μ M, 500 nM, 200 nM, 100 nM or 10 nM of each compound or the equivalent volume of maximum DMSO vehicle used as a control. At the time of infection, the drug-containing medium was removed, and replaced with virus inoculum (MOI of 0.1 PFU per cell) containing TPCK-trypsin (Sigma-Aldrich). Following a 1-h adsorption incubation at 37 °C, the virus inoculum was removed and 200 μ l of drug- or vehicle-containing medium was added. Then, 48 h after infection, the cell-culture supernatant was used to extract RNA using the Direct-zol-96 RNA extraction kit (Zymo) following the manufacturer's instructions. Detection of viral genomes in the extracted RNA was performed by RT-qPCR, using previously published SARS-CoV-2-specific primers⁸¹. Specifically, the primers target the *N* gene region: 5'-TAATCAGACAAGGAAGTATTA-3' (forward) and 5'-CGAAGGTGTGACTTCCATG-3' (reverse). RT-qPCR was performed using the Luna Universal One-Step RT-qPCR Kit (NEB) in an Applied Biosystems QuantStudio 6 thermocycler, using the following cycling conditions: 55 °C for 10 min, 95 °C for 1 min, and 40 cycles of 95 °C for 10 s, followed by 60 °C for 1 min. The quantity of viral genomes is expressed as PFU equivalents, and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titre. In addition to measuring viral RNA in the supernatant derived from drug-treated cells, infectious virus was quantified by plaque-forming assay.

Cell viability assays

Cell viability in drug-treated cells was measured using Alamar blue reagent (ThermoFisher). In brief, 48 h after treatment, the drug-containing medium was removed and replaced with Alamar blue and incubated for 1 h at 37 °C and fluorescence measured in a Tecan Infinity 2000 plate reader. Percentage viability was calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol (0% viability), included in each plate.

Plaque-forming assays

Viruses were quantified by plaque-forming assays. For this, Vero E6 cells were seeded in 24-well plates at a concentration of 7.5×10^4 cells per well. The following day, tenfold serial dilutions of individual virus samples in serum-free MEM medium were added to infect the cells at 37 °C for 1 h. After the adsorption time, the overlay medium was added at final concentration of 2% FBS/MEM medium and 0.05% agarose to achieve a semi-solid overlay. Plaque-forming assays were incubated at 37 °C for 3 days before fixation with 4% formalin and visualization using crystal violet solution.

Off-target assays for Sigma receptor drugs and ligands

hERG binding assays were carried out as previously described⁸². In brief, compounds were incubated with hERG membranes, prepared from HEK-293 cells stably expressing hERG channels, and [³H]dofetilide (5 nM final) in a total of 150 μ l for 90 min at room temperature in the dark. Reactions were stopped by filtering the mixture onto a glass fibre and were quickly washed three times to remove unbound [³H]dofetilide. The filter was dried in a microwave, melted with a scintillant cocktail and wrapped in a plastic film. Radioactivity was counted on a MicroBeta counter and results were analysed in Prism by fitting to the built-in one binding function to obtain affinity *K*_d. Radioligand binding assays for the muscarinic and alpha-adrenergic receptors were performed as previously described⁸³. Detailed protocols are available

Article

on the NIMH PDSP website at <https://pdspdb.unc.edu/html/tutorials/UNC-CH%20Protocol%20Book.pdf>.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The AP-MS raw data and search results files generated during the current study are available in the ProteomeXchange Consortium via the PRIDE partner repository with dataset identifier PXD018117 (<https://www.ebi.ac.uk/pride/archive/projects/PXD018117>) and PPI networks have also been uploaded to NDEX (<https://public.ndexbio.org/#/network/43803262-6d69-11ea-bfcd-0ac135e8bacf>). An interactive version of these networks, including relevant drug and functional information, can be found at <http://kroganlab.ucsf.edu/network-maps>. All data generated or analysed during this study are included in the article and its Supplementary Information. Expression vectors used in this study are readily available from the authors to biomedical researchers and educators in the non-profit sector. Source data are provided with this paper.

Code availability

Complete methods for chemoinformatic analysis can be found on GitHub (<https://github.com/momeara/BioChemPantry/tree/master/vignette/COVID19>); details on MIST scoring can be found on GitHub (<https://github.com/kroganlab/mist>).

- Yang, D. & Leibowitz, J. L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **206**, 120–133 (2015).
- Yang, H. et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl Acad. Sci. USA* **100**, 13190–13195 (2003).
- Thiel, V. et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315 (2003).
- Xie, Y. et al. GPS-Lipid: a robust tool for the prediction of multiple lipid modification sites. *Sci. Rep.* **6**, 28249 (2016).
- Ren, J. et al. CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng. Des. Sel.* **21**, 639–644 (2008).
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
- Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
- Chiva, C. et al. QCloud: a cloud-based quality control system for mass spectrometry-based proteomics laboratories. *PLoS ONE* **13**, e0189209 (2018).
- Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
- Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
- Verschueren, E. et al. Scoring large-scale affinity purification mass spectrometry datasets with MiST. *Curr. Protoc. Bioinformatics* **49**, 8.19.1–8.19.16 (2015).
- Giurgiu, M. et al. CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).
- Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Huttlin, E. L. et al. The BioPlex Network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
- Vizcaino, J. A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).
- Deutsch, E. W. et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).
- Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
- Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCr* **1**, 213–220 (2014).
- Drozdzetkiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
- Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
- Gaulton, A. et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).

- Sterling, T. & Irwin, J. J. ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
- Schmidt, H. R., Betz, R. M., Dror, R. O. & Kruse, A. C. Structural basis for σ_1 receptor ligand recognition. *Nat. Struct. Mol. Biol.* **25**, 981–987 (2018).
- Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **50**, 1561–1573 (2010).
- Reed, L. J. & Muench, H. A simple method of estimating fifty per cent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).
- Chu, D. K. W. et al. Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. *Clin. Chem.* **66**, 549–555 (2020).
- Huang, X.-P., Mangano, T., Hufeisen, S., Setola, V. & Roth, B. L. Identification of human Ether-à-go-go related gene modulators by three screening platforms in an academic drug-discovery setting. *Assay Drug Dev. Technol.* **8**, 727–742 (2010).
- Besnard, J. et al. Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215–220 (2012).
- Jimenez-Morales, D., Rosa Campos, A., Von Dollen, J., Krogan, N. J. & Swaney, D. L. artMS: analytical R tools for mass spectrometry. R package version 1.6.5 <https://bioconductor.org/packages/release/bioc/html/artMS.html> (2020).
- El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

Acknowledgements We thank J. Sarlo for his artistic contributions to the manuscript; A. Kruse for his advice on Sigma receptor functional activity and pharmacologically related enzymes; S. Worland and eFFECTOR Therapeutics for providing zotatifin; T. Haley from FedEx for his effort helping us to ship out drugs and SARS-CoV-2 expression constructs across the globe; R. Albrecht for support with BSL3 procedures and their partners and families for support in child care and other matters during this time. This research was funded by grants from the National Institutes of Health (P50AI150476, U19AI135990, U19AI135972, R01AI143292, R01AI120694, P01A1063302 and R01AI122747 to N.J.K.; R35GM122481 to B.K.S.; R01CA221969 and R01CA244550 to K.M.S.; K08HL124068 to J.S.C.; 1F32CA236347-01 to J.E.M., NIMH Psychoactive Drug Screening Contract to B.L.R.); from the Defense Advance Research Projects Agency HRO011-19-2-0020 (to B.K.S., N.J.K., A.G.-S., K.M.S., D.A.A. and K.M.V.); by CRIP (Center for Research for Influenza Pathogenesis), a NIAID supported Center of Excellence for Influenza Research and Surveillance (CEIRS, contract no. HHSN27201400008C) and by supplements to NIAID grant U19AI135972 and DoD grant W81XWH-19-PRMRP-FPA to A.G.-S.; by the Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' (grant ANR-10-LABX-62-IBED) to M.V., O.S. and C.d'E.; funding from F. Hoffmann-La Roche and Vir Biotechnology and gifts from The Ron Conway Family. K.M.S. is an investigator of the Howard Hughes Medical Institute. The views, opinions, and findings contained in this study are those of the authors and do not represent the official views, policies, or endorsement of the Department of Defense or the US Government.

Author contributions The study was conceived by N.J.K. and D.E.G. Genome annotation was carried out by D.E.G., G.M.J. and B.J.P. Molecular cloning was performed by D.E.G., G.M.J. and J.Z.G. Cell culture, affinity purifications and peptide digestion was carried out by G.M.J. and J.X. Mass spectrometer operation and peptide search were performed by D.L.S. and proteomics data were processed by M. Bouhaddou, Y.Z., B.J.P., D.L.S. and R.H. Network annotation was led by M. Bouhaddou with support from D.E.G., N.J.K., R.M.K. and the Supplementary Discussion literature review team. Interactome meta-analysis was performed by P.B., M. Bouhaddou, H.B., M.C., Z.Z.C.N., I.B.-H., D.M., C.H.-A., T.P., S.B.R., M.C.O., Y.C., J.C.J.C., D.J.B. and S.K. Drug selection and annotation was done by M.J.O., T.A.T., S.P., Y.S., Z.Z., W.S., I.T.K., J.E.M., J.S.C., K.L., S.A.D., J.L., L.C., S.V., J.L.-L., Y. Lin, X.-P.H., Y. Liu, P.P.S., N.A.W., D.K., H.-Y.W., K.M.S. and B.K.S. Structural modelling was performed by C.J.P.M., K.B.P., S.J.G., D.J.S., R.R., X.L., S.A.W., M. Bohn, F.S.U. and T.K. Live SARS-CoV-2 virus assays were led by K.M.W. and V.V.R. with support from F.R., T.V., A.M.K., L.M., E.M., C.K., N.S.S., Q.D.T., D.S., S.J.F. and M.H. Analysis of SARS-CoV-2 genomic diversity was done by M. Safari. Analysis of human gene positive selection was performed by J.M.Y., B.M. and H.S.M. NSP5 cleavage prediction and analysis was carried out by M. Bohn and C.S.C. with support from D.E.G. Figures were prepared by D.E.G., M. Bouhaddou, K.O., K.M.W., M.J.O., D.L.S., T.A.T., R.H., R.M.K., M.K., H.B., Y.Z., Z.Z., C.J.P.M., T.P., S.A.W., M. Bohn, M. Safari, F.S.U., N.A.W., D.G.F., S.N.F., J.D.G., D.R., T.K., P.B., K.M.S., B.K.S. and N.J.K. The Supplementary Information was assembled by R.M.K. with support from the literature review team: K.O., R.H., R.M.K., A.L.R., B.T., H.F., J.B., K.H., M.M., M.K., P.H., J.M.F., M. Eckhardt, M. Soucheray, M.J.B., M.C., M.J.M., Q.L., C.J.P.M., T.P., X.L., L.C., S.V., J.L.-L., Y. Lin, M. Bohn, R.T., D.A.C., J.H., T.L.R., U.R., A. Subramanian, J.N., N.J., S.M., S.N.F. and J.D.G. The manuscript was prepared by D.E.G., M. Bouhaddou, K.O., M.J.O., D.L.S., T.A.T., R.H., R.M.K., M. Eckhardt, M. Soucheray, M.J.B., P.B., K.M.S., B.K.S. and N.J.K. The work was supervised by R.M.S., A.D.F., O.S.R., K.A.V., D.A.A., M.O., M. Emerman, N.J., M.v.Z., E.V., A.A., O.S., C.d'E., S.M., M.J., H.S.M., D.G.F., T.I., C.S.C., S.N.F., J.S.F., J.D.G., A. Sali, B.L.R., D.R., J.T., T.K., P.B., M.V., A.G.-S., K.M.S., B.K.S. and N.J.K.

Competing interests The Krogan laboratory has received research support from Vir Biotechnology and F. Hoffmann-La Roche. K.M.S. has consulting agreements for the following companies, which involve monetary and/or stock compensation: Black Diamond Therapeutics, BridGene Biosciences, Denali Therapeutics, Dice Molecules, eFFECTOR Therapeutics (zotatifin and tomivosertib), Erasca, Genentech/Roche, Janssen Pharmaceuticals, Kumquat Biosciences, Kura Oncology, Merck, Mitokinin, Petra Pharma, Qalab, Revolution Medicines (WDB002), Type6 Therapeutics, Venthera, Wellspring Biosciences (Araxes Pharma). J.T. is a cofounder and shareholder of Global Blood Therapeutics, Principia Biopharma, Kezar Life Sciences and Cedilla Therapeutics. J.T. and P.P.S. are listed as inventors on a patent application that describes the cyclic depsipeptide, PS3061 (PCT/US2019/024731) by the Regents of the University of California and Kezar Life Sciences. D.R. is a shareholder of eFFECTOR Therapeutics and a member of its scientific advisory board (zotatifin and tomivosertib). C.S.C. is a co-founder of Catalyt Biosciences and Alauus Biosciences, and receives grant/research support from

ShangPharma/ChemPartners. A. A. is a co-founder of Tango Therapeutics, Azkarrá Therapeutics, Ovibio Corporation; a consultant for SPARC, Bluestar, ProLynx, Earli, Cura, GenVivo and GSK; a member of the SAB of Genentech and GLAdiator; receives grant/research support from SPARC and AstraZeneca; holds patents on the use of PARP inhibitors held jointly with AstraZeneca, from which he has benefitted financially (and may do so in the future). The authors have not filed for patent protection on the SARS-CoV-2 host interactions or the use of predicted drugs for the treatment of COVID-19 to ensure all of the information is freely available to accelerate the discovery of a treatment.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2286-9>.

Correspondence and requests for materials should be addressed to M.V., A.G.-S., K.M.S., B.K.S. or N.J.K.

Peer review information Nature thanks Michael P. Weekes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

5. The global phosphorylation landscape of SARS-CoV-2 infection

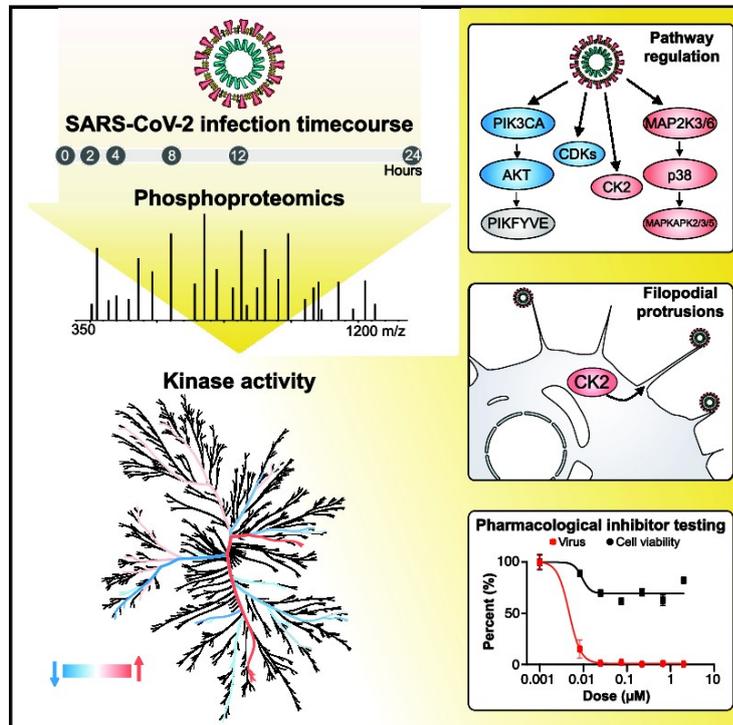
Mehdi Bouhaddou, Danish Memon, Bjoern Meyer, Kris M. White, Veronica V. Rezelj, Miguel Correa Marrero, Benjamin J. Polacco, James E. Melnyk, Svenja Ulferts, Robyn M. Kaake, Jyoti Batra, Alicia L. Richards, Erica Stevenson, David E. Gordon, Ajda Rojc, Kirsten Obernier, Jacqueline M. Fabius, Margaret Soucheray, Lisa Miorin, Elena Moreno, Cassandra Koh, **Quang Dinh Tran**, Alexandra Hardy, Rémy Robinot, Thomas Vallet, Benjamin E. Nilsson-Payant, Claudia Hernandez-Armenta, Alistair Dunham, Sebastian Weigang, Julian Knerr, Maya Modak, Diego Quintero, Yuan Zhou, Aurelien Dugourd, Alberto Valdeolivas, Trupti Patil, Qiongyu Li, Ruth Hüttenhain, Merve Cakir, Monita Muralidharan, Minkyu Kim, Gwendolyn Jang, Beril Tutuncuoglu, Joseph Hiatt, Jeffrey Z. Guo, Jiewei Xu, Sophia Bouhaddou, Christopher J.P. Mathy, Anna Gaulton, Emma J. Manners, Eloy Félix, Ying Shi, Marisa Goff, Jean K. Lim, Timothy McBride, Michael C. O'Neal, Yiming Cai, Jason C.J. Chang, David J. Broadhurst, Saker Klippsten, Emmie De wit, Andrew R. Leach, Tanja Kortemme, Brian Shoichet, Melanie Ott, Julio Saez-Rodriguez, Benjamin R. tenOever, R. Dyche Mullins, Elizabeth R. Fischer, Georg Kochs, Robert Grosse, Adolfo García-Sastre, Marco Vignuzzi, Jeffery R. Johnson, Kevan M. Shokat, Danielle L. Swaney, Pedro Beltrao & Nevan J. Krogan

DOI: <https://doi.org/10.1016/j.cell.2020.06.034>

Contributions: I have helped with infection experiments presented in Figure 7B-K and S5.

The Global Phosphorylation Landscape of SARS-CoV-2 Infection

Graphical Abstract



Highlights

- Phosphoproteomics analysis of SARS-CoV-2-infected cells uncovers signaling rewiring
- Infection promotes host p38 MAPK cascade activity and shutdown of mitotic kinases
- Infection stimulates CK2-containing filopodial protrusions with budding virus
- Kinase activity analysis identifies potent antiviral drugs and compounds

Authors

Mehdi Bouhaddou, Danish Memon, Bjoern Meyer, ..., Danielle L. Swaney, Pedro Beltrao, Nevan J. Krogan

Correspondence

robert.grosse@pharmakol.uni-freiburg.de (R.G.),
adolfo.garcia-sastre@mssm.edu (A.G.-S.),
marco.vignuzzi@pasteur.fr (M.V.),
jeffrey.johnson@mssm.edu (J.R.J.),
kevan.shokat@ucsf.edu (K.M.S.),
danielle.swaney@ucsf.edu (D.L.S.),
pbeltrao@ebi.ac.uk (P.B.),
nevan.krogan@ucsf.edu (N.J.K.)

In Brief

Phosphoproteomics analysis of SARS-CoV-2-infected Vero E6 cells reveals host cellular pathways hijacked by viral infection, leading to the identification of small molecules that target dysregulated pathways and elicit potent antiviral efficacy.



Bouhaddou et al., 2020, Cell 182, 685–712
August 6, 2020 © 2020 Elsevier Inc.
<https://doi.org/10.1016/j.cell.2020.06.034>



Article

The Global Phosphorylation Landscape of SARS-CoV-2 Infection

Mehdi Bouhaddou,^{1,2,3,4,21} Danish Memon,^{5,21} Bjoern Meyer,^{6,21} Kris M. White,^{7,8,21} Veronica V. Rezelj,^{6,21} Miguel Correa Marrero,^{5,21} Benjamin J. Polacco,^{1,2,3,4,21} James E. Melnyk,^{1,2,4,9,21} Svenja Ulferts,^{10,21} Robyn M. Kaake,^{1,2,3,4,21} Jyoti Batra,^{1,2,3,4,21} Alicia L. Richards,^{1,2,3,4} Erica Stevenson,^{1,2,3,4} David E. Gordon,^{1,2,3,4} Ajda Rojc,^{1,2,3,4} Kirsten Obernier,^{1,2,3,4} Jacqueline M. Fabius,^{1,2,3,4} Margaret Soucheray,^{1,2,3,4} Lisa Miorin,^{7,8} Elena Moreno,^{7,8} Cassandra Koh,⁵ Quang Dinh Tran,⁶ Alexandra Hardy,⁵ Rémy Robinot,^{11,12} Thomas Vallet,⁶ Benjamin E. Nilsson-Payant,⁷ Claudia Hernandez-Armenta,⁵ Alistair Dunham,⁵ Sebastian Weigang,¹³ Julian Knerr,¹⁰ Maya Modak,^{1,2,3,4} Diego Quintero,^{1,2,3,4} Yuan Zhou,^{1,2,3,4} Aurelien Dugourd,¹⁴ Alberto Valdeolivas,¹⁴ Trupti Patil,^{1,2,3,4} Qiongyu Li,^{1,2,3,4} Ruth Hüttenhain,^{1,2,3,4} Merve Cakir,^{1,2,3,4} Monita Muralidharan,^{1,2,3,4} Minkyu Kim,^{1,2,3,4} Gwendolyn Jang,^{1,2,3,4} Beril Tutuncuoglu,^{1,2,3,4} Joseph Hiatt,^{1,2,3,4} Jeffrey Z. Guo,^{1,2,3,4} Jiewei Xu,^{1,2,3,4} Sophia Bouhaddou,¹ Christopher J.P. Mathy,^{1,2,19} Anna Gaulton,⁵ Emma J. Manners,⁵ Eloy Félix,⁵ Ying Shi,^{1,2,4,9} Marisa Goff,⁷ Jean K. Lim,⁷ Timothy McBride,¹⁵ Michael C. O'Neal,¹⁵ Yiming Cai,¹⁵ Jason C.J. Chang,¹⁵ David J. Broadhurst,¹⁵ Saker Klippsten,¹⁵ Emmie De wit,¹⁶ Andrew R. Leach,⁵ Tanja Kortemme,^{1,2,19} Brian Shoichet,^{1,2} Melanie Ott,³ Julio Saez-Rodriguez,¹³ Benjamin R. tenOever,⁷ R. Dyche Mullins,^{1,2,4,8} Elizabeth R. Fischer,¹⁶ Georg Kochs,^{12,17} Robert Grosse,^{9,17,18,*} Adolfo Garcia-Sastre,^{7,8,20,*} Marco Vignuzzi,^{6,*} Jeffery R. Johnson,^{7,*} Kevan M. Shokat,^{1,2,4,9,*} Danielle L. Swaney,^{1,2,3,4,*} Pedro Beltrao,^{1,5,*} and Nevan J. Krogan^{1,2,3,4,7,22,*}

¹QBI COVID-19 Research Group (QCRG), San Francisco, CA 94158, USA

²Quantitative Biosciences Institute (QBI), University of California, San Francisco, San Francisco, CA 94158, USA

³J. David Gladstone Institutes, San Francisco, CA 94158, USA

⁴Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA

⁵European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

⁶Viral Populations and Pathogenesis Unit, CNRS UMR 3569, Institut Pasteur, 75724 Paris, Cedex 15, France

⁷Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

⁸Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA

⁹Howard Hughes Medical Institute

¹⁰Institute for Clinical and Experimental Pharmacology and Toxicology, University of Freiburg, Freiburg 79104, Germany

¹¹Virus & Immunity Unit, Department of Virology, CNRS UMR 3569, Institut Pasteur, 75724 Paris, Cedex 15, France

¹²Vaccine Research Institute, 94000 Creteil, France

¹³Institute of Virology, Medical Center – University of Freiburg, Freiburg 79104, Germany

¹⁴Institute for Computational Biomedicine, Bioquant, Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Heidelberg 69120, Germany

¹⁵Zoic Labs, Culver City, CA 90232, USA

¹⁶NIH/NIAID/Rocky Mountain Laboratories, Hamilton, MT 59840, USA

¹⁷Faculty of Medicine, University of Freiburg, Freiburg 79008, Germany

¹⁸Centre for Integrative Biological Signalling Studies (CIBSS), Freiburg 79104, Germany

¹⁹Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, USA

²⁰The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA

²¹These authors contributed equally

²²Lead Contact

*Correspondence: robert.grosse@pharmakol.uni-freiburg.de (R.G.), adolfo.garcia-sastre@mssm.edu (A.G.-S.), marco.vignuzzi@pasteur.fr (M.V.), jeffrey.johnson@mssm.edu (J.R.J.), kevan.shokat@ucsf.edu (K.M.S.), danielle.swaney@ucsf.edu (D.L.S.), pbeltrao@ebi.ac.uk (P.B.), nevan.krogan@ucsf.edu (N.J.K.)

<https://doi.org/10.1016/j.cell.2020.06.034>

SUMMARY

The causative agent of the coronavirus disease 2019 (COVID-19) pandemic, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has infected millions and killed hundreds of thousands of people worldwide, highlighting an urgent need to develop antiviral therapies. Here we present a quantitative mass spectrometry-based phosphoproteomics survey of SARS-CoV-2 infection in Vero E6 cells, revealing dramatic rewiring of phosphorylation on host and viral proteins. SARS-CoV-2 infection promoted casein kinase II (CK2) and p38 MAPK activation, production of diverse cytokines, and shutdown of mitotic kinases, resulting in cell cycle arrest. Infection also stimulated a marked induction of CK2-containing filopodial protrusions possessing budding viral particles. Eighty-seven drugs and compounds were identified by mapping global phosphorylation profiles to dysregulated kinases and pathways. We found pharmacologic inhibition of the p38, CK2, CDK, AXL, and PIKfyve kinases to possess antiviral efficacy, representing potential COVID-19 therapies.



INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped positive-sense RNA virus that belongs to the lineage B Betacoronavirus family. It is closely related to SARS-CoV, the causative agent of SARS, which emerged in the human population in 2002 (79% genetic similarity), and several SARS-related coronaviruses that circulate in bats (up to 98% genetic similarity) (Lai et al., 2020; Zhou et al., 2020). The pathophysiology of severe coronavirus disease 2019 (COVID-19) is similar to that of severe disease caused by SARS-CoV and is characterized by acute respiratory distress and excessive inflammation capable of inducing respiratory failure, multi-organ failure, and death (Wong et al., 2004; Zhang et al., 2020).

To enter host cells, the SARS-CoV-2 spike (S) protein binds to an ACE2 receptor on the target cell and is subsequently primed by a serine protease, TMPRSS2, that cleaves the S protein and allows fusion of viral and lysosomal membranes (Hoffmann et al., 2020). Following entry, viral genomic RNA is translated to produce the polyproteins ORF1a and ORF1ab, which are subsequently cleaved by viral proteases into non-structural proteins that form the viral replication/transcription complex (RTC). Extensive remodeling of the host endoplasmic reticulum leads to formation of double-membrane vesicles, within which viral RNA synthesis occurs. The viral RNA genome is replicated by transcription of the negative-strand genomic RNA template, whereas subgenomic mRNAs are transcribed and translated to produce structural and accessory proteins. Structural proteins and viral genomes assemble at the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) (de Haan and Rottier, 2005), followed by transport to the cell surface for release by exocytosis (Fehr and Perlman 2015).

Although much about SARS-CoV-2 biology can be inferred based on similarity to SARS-CoV, SARS-CoV-2 is a novel coronavirus with unique properties that contribute to its pandemic-scale spread. Unlike SARS-CoV, SARS-CoV-2 infection is commonly asymptomatic, particularly in the younger population (Guan et al., 2020), and contagious prior to symptom onset (Rothe et al., 2020; Peiris et al., 2003; Bai et al., 2020). These characteristics contribute to the difficulty of containing SARS-CoV-2 spread through public health strategies and amplify the need to develop vaccines and therapies to protect against and treat COVID-19. Clinical management of COVID-19 is largely limited to infection prevention and supportive care. So far, remdesivir, a broad-spectrum antiviral agent, is the only medication approved for emergency use to treat COVID-19 by the US Food and Drug Administration (FDA) (Grein et al., 2020). Although the evidence supporting remdesivir use in patients with advanced COVID-19 is promising, there remains an urgent need for potent SARS-CoV-2 therapeutic agents, especially those that could be given in an outpatient setting, to effectively combat the COVID-19 pandemic.

Proteomics approaches that globally quantify changes in protein abundance and phosphorylation are powerful tools to elucidate mechanisms of viral pathogenesis by providing a snapshot of how cellular pathways and processes are rewired upon infection (Johnson et al., 2020). Importantly, the functional outcomes of many phosphorylation events are well annotated, especially

for kinases where phosphorylation directly regulates their activity. State-of-the-art bioinformatics approaches can then be employed to readily identify regulated kinases from phosphorylation profiles, many of which are likely to be established drug targets with therapeutic potential (Ochoa et al., 2016, 2020). Here we present a quantitative survey of the global phosphorylation and protein abundance landscape of SARS-CoV-2 infection, map phosphorylation changes to disrupted kinases and pathways, and use these profiles to rapidly prioritize drugs and compounds with the potential to treat SARS-CoV-2 infection.

RESULTS

Phosphorylation Signaling Represents a Primary Host Response to SARS-CoV-2 Infection

To determine how SARS-CoV-2 hijacks host-protein signaling, a global phosphoproteomics experiment was performed in Vero E6 cells, a cell line originating from the kidney of a female African green monkey (*Chlorocebus sabaeus*) (Osada et al., 2014). This cell line was selected because of its high susceptibility to SARS-CoV-2 infection (Harcourt et al., 2020). Cells were harvested in biological triplicate at 6 time points after SARS-CoV-2 infection (0, 2, 4, 8, 12, or 24 h) or after mock infection at 0 or 24 h (Figure 1A). Using a data-independent acquisition (DIA) proteomics approach, each sample was then partitioned and analyzed for changes in global protein abundance or phosphorylation (data available in Table S1). *Chlorocebus sabaeus* and human protein sequences were aligned, and phosphorylation sites and protein identifiers were mapped to their respective human protein orthologs. Phosphorylation fold changes calculated using the 0- or 24-h mock control were highly comparable (correlation coefficient $r = 0.77$); therefore, the 0-h mock control was used for all subsequent comparisons.

Quality control filtering of the data was performed, and two samples from each of the phosphorylation and protein abundance datasets were removed because of poor correlation with their respective replicates (Figures S1A and S1B). Principal component analysis (PCA) of the remaining samples revealed good separation of mock and infected samples as well as high quantitative reproducibility between biological replicates (Figures 1B, 1C, and S1C). In total, high-quality quantification of 4,624 human-orthologous phosphorylation sites and 3,036 human-orthologous proteins was obtained (Figure S1D). Successful infection was confirmed by the observation of a dramatic increase in viral protein abundance over the course of a 24-h infection period (Figures 1D and S1E).

As expected, an increase was observed in the number of significantly regulated phosphorylation sites and proteins over the infection time course, with the majority of regulation occurring at the level of phosphorylation (Figures 1E and 1F) as opposed to protein abundance (Figures 1G and 1H). Of the few proteins that significantly increased in abundance upon infection, the vast majority were SARS-CoV-2 viral proteins (Figure 1H). In contrast, the majority of host proteins decreased in abundance. This finding is consistent with mechanisms of host mRNA nuclear export and/or host mRNA translation inhibition, which are common in viral infections (Kuss et al., 2013; Walsh

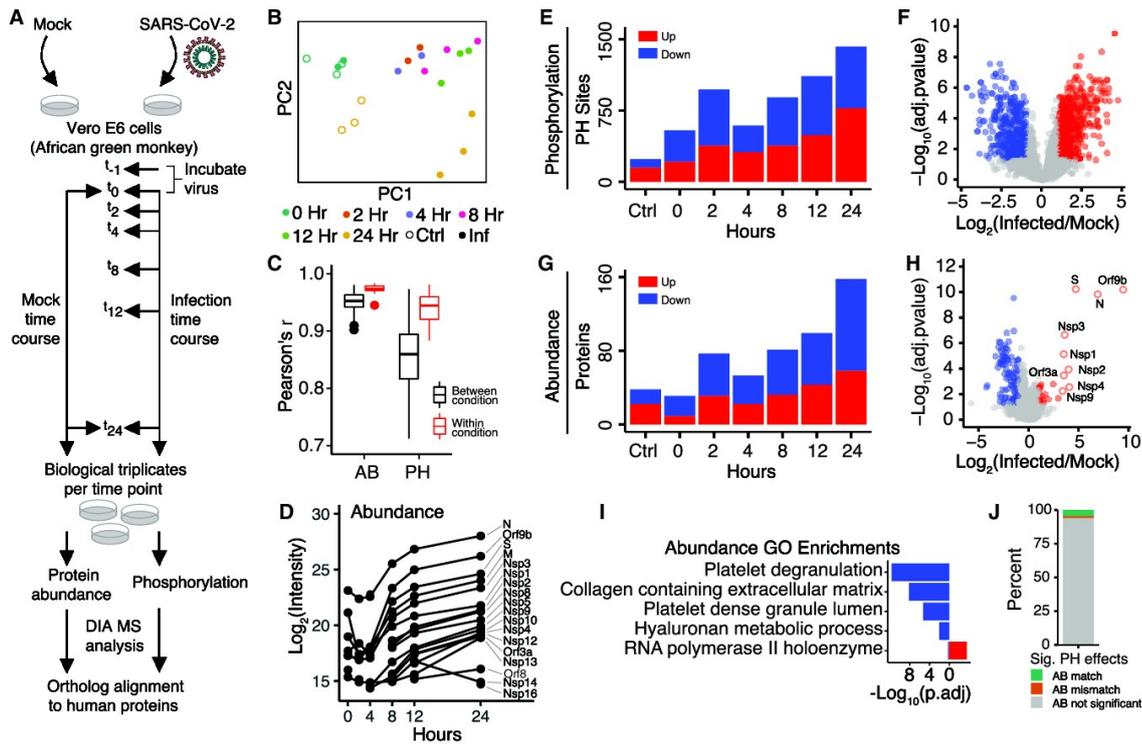


Figure 1. Global Proteomics of Phosphorylation and Abundance Changes upon SARS-CoV-2 Infection

(A) Vero E6 cells were infected with SARS-CoV-2 (MOI 1.0). After 1 h of viral uptake, cells were harvested (0 h) or, subsequently, after 2, 4, 8, 12, or 24 h. As a control, Vero E6 cells were also mock infected for 1 h and harvested immediately thereafter (0 h) or after 24 h of mock infection. All conditions were performed in biological triplicate. Following cell harvest, cells were lysed, and proteins were digested into peptides. Aliquots of all samples were analyzed by mass spectrometry (MS) to measure changes in protein abundance upon infection, whereas the remaining sample was enriched for phosphorylated peptides and subsequently analyzed to measure changes in phosphorylation signaling. A DIA approach was used for all MS acquisitions. Last, all phosphorylation sites and protein identifiers were mapped to their respective human protein orthologs.

(B) Principal-component analysis (PCA) of phosphorylation replicates after removing outliers. See also Figure S1.

(C) Correlation of protein abundance (AB) and phosphorylation sites (PHs) between replicates within a biological condition (red) and across biological conditions (black). Boxplots depict median (horizontal lines), interquartile range (boxes), maximum and minimum values (vertical lines), and outliers (solid circles).

(D) Median AB of individual SARS-CoV-2 proteins in the protein AB analysis.

(E) The number of significantly regulated PH groups across the infection time course.

(F) Volcano plot of PH group quantification 24 h after infection.

(G) The number of significantly regulated proteins across the infection time course.

(H) Volcano plot of protein AB quantification 24 h after infection.

(I) Gene Ontology enrichment analysis of all significantly changing proteins in terms of AB divided into two sets: downregulated (blue) and upregulated (red).

(J) Proportion of significantly regulated PH groups with a correlated (i.e., same direction, AB match) or anticorrelated (i.e., opposite direction, AB mismatch) significant or insignificant (gray) change in protein AB.

In (E)–(H), all infection time points are compared with the mock infection at 0 h, and significantly regulated proteins are defined as (absolute value of $\log_2(\text{infection}/\text{mock}) > 1$ and adjusted $p < 0.05$ or when only detected in infected or mock based on replicate and MS feature counts; STAR Methods).

See also Figure S1.

and Mohr 2011). Gene Ontology enrichment analysis of significantly downregulated proteins revealed several terms related to platelet regulation (Figures 1I and S1F). Several downregulated host proteins are known to be involved in platelet regulation, thrombosis, and prevention of blood coagulation, including APOH, CD9, TSPAN14, AHSG, SERPINA1, and A2M (Mather et al., 2016; Mangin et al., 2009; Taggart et al., 2000). The downregulation of these proteins suggests that they may mechanisti-

cally contribute to symptoms of blood coagulation and stroke in COVID-19 patients (Han et al., 2020).

Lastly, the contribution of protein abundance to phosphorylation level changes was evaluated. For nearly all cases of a significantly changed phosphorylation site, no corresponding significant change in protein abundance was observed (Figure 1J), further suggesting that phosphorylation signaling represents a primary host response over this time course of infection as

opposed to transcriptional regulation, which would influence protein abundance.

Phosphorylation of SARS-CoV-2 Viral Proteins by the Host Proteome

Viral protein phosphorylation within the host cell may play a role in sensing and responding to cell state. We detected 25 phosphorylation sites in SARS-CoV-2 viral proteins that we combined with another proteomics dataset (Davidson et al., 2020) to amass a total of 49 sites detected across seven viral proteins (Table S2). Of note, this analysis does not distinguish cleaved from uncleaved viral proteins in the assignment of viral phosphorylation sites. The degree of conservation, indicative of functional constraint, was estimated for each residue position (Figure 2A; Ng and Henikoff 2003), and the sites were mapped to positions within structured regions for five proteins, with the majority observed in accessible positions (i.e., loops) (Figure 2B). The top kinase families predicted by sequence to regulate these sites included casein kinase II (CK2), cyclin-dependent kinase (CDK), and protein kinase C (PKC), among others (Figure 2C), suggesting that these kinases may contribute to regulation of viral replication.

Although it is unlikely that all phosphorylation sites on viral proteins play important functional roles, several sites in membrane (M) protein, Nsp9, and nucleocapsid (N) protein (Figures 2D–2F) suggest potential functionality. Five phosphorylation sites were detected in the M protein cluster within a short C-terminal region of the protein (207–215; Figure 2D). Although these acceptor residues are not predicted to be conserved, several are negatively charged residues in M proteins of other related viruses (Figure 2E). This evolutionary pattern suggests that a negative charge in this region may play a functional role, reminiscent of other multi-site phosphorylation events (Serber and Ferrell 2007).

To identify phosphorylation sites that may regulate protein-protein interactions, all sites were mapped to 3D structures, and solvent accessibility based protein-protein interface identification and recognition (SPPIDER) was used to assess whether sites resided within interface regions (Porollo and Meller 2007; Figure 2A; Table S2). The single phosphorylation site in Nsp9 was predicted to be at an interface region (“True”), which was supported by inspection of the homodimer structure (PDB: 6W4B). Additional phosphorylation sites were predicted to be at interface residues within the S protein (Figure 2A). However, inspection of S in complex with the ACE2 receptor (Shang et al., 2020; Lan et al., 2020) reveals some of these phosphorylation sites to be near but not at the interface region.

Finally, phosphorylation sites in N protein, a structural protein that binds to and assists with packaging viral RNA, were investigated. Most sites occurred within the N-terminal portion of the protein, at or near the RNA binding region, but avoided the C-terminal dimerization domain. The cluster of phosphorylation sites within an arginine/serine (RS)-dipeptide rich region, C-terminal to the RNA binding region (Figure 2A), is conserved in other coronavirus N proteins. This region is phosphorylated in SARS-CoV by serine-arginine (SR) protein kinases, modulating the role of SARS-CoV N protein in host translation inhibition (Peng et al., 2008). It is likely that phosphorylation of this same region in SARS-CoV-2 plays a similar role. Interestingly, *in vitro* inhibition

of SARS-CoV N protein phosphorylation at the RS-rich region results in reduced viral load and cytopathic effects (Wu et al., 2009), highlighting its importance for viral fitness. In addition, sites spanning the sequence of the RNA binding domain, which forms a claw-like structure, have been observed (Kang et al., 2020). Several phosphorylation sites cluster in the structural model, predicted to affect the surface charge of the so-called acidic wrist region (Figure 2F) but not the positive surface charge of the RNA binding pocket. We hypothesize that this surface charge difference may modulate N protein function, potentially via allosteric regulation of RNA binding capacity.

Phosphorylation of SARS-CoV-2 Host-Interacting Proteins during Infection

The recently published SARS-CoV-2 virus-human protein-protein interaction map identified 332 human proteins interacting with 27 (26 wild-type and 1 mutant) viral proteins (Gordon et al., 2020). Here we found some of these host proteins (40 of 332) to be significantly differentially phosphorylated upon infection (Figure 3). Virus-host protein-protein interactions could drive changes in phosphorylation by affecting host protein subcellular localization or by sterically blocking kinase access. Furthermore, phosphorylation of these proteins upon infection may signify an additional mode of functional control over these potential dependency and restriction factors.

The SARS-CoV-2 N protein is known to interact with several RNA-processing proteins that are differentially phosphorylated during infection, including LARP1 and RRP9. Here LARP1 phosphorylation decreases on several sites, which is known to consequently increase LARP1 affinity for 3' untranslated regions (UTRs) of mRNAs encoding ribosomal proteins, driving inhibition of protein synthesis (Hong et al., 2017). This mechanism may be utilized by SARS-CoV-2 to prioritize synthesis of viral proteins over host proteins. In addition, ORF6 interacts with the NUP98/RAE complex, and NUP98 phosphorylation was observed to increase at S888, a site within its peptidase domain. NUP98 autocatalytic cleavage is required for localization to the nuclear pore; thus, it is possible that NUP98 interaction with ORF6 and/or its virus-induced phosphorylation prevents host mRNA export through the nuclear pore (Krull et al., 2010; Hodel et al., 2002). A similar mechanism is employed by vesicular stomatitis virus (VSV) matrix protein to block host mRNA export by targeting the NUP98/RAE complex, leading to exclusive translation of cytoplasmic VSV mRNAs (Quan et al., 2014).

For Nsp12, the majority of its protein interactors displayed decreased phosphorylation during infection. Because Nsp12 is known to encode the RNA-dependent RNA polymerase, responsible for replicating the viral genome, and several of these interacting proteins are related to RNA processing (LARP4B and CRT3), their regulation may possess functional implications for Nsp12 in viral RNA replication. In addition, Nsp8 interacts with several proteins whose phosphorylation increases (LARP7 and MPHOSPH10) and decreases (CCDC86) on several sites. Notably, LARP7 and MEPCE are important regulators of RNA polymerase II-mediated transcription elongation as part of the 7SK small nuclear ribonucleoprotein particle (snRNP) complex. Regulation of these phosphorylation sites may contribute to

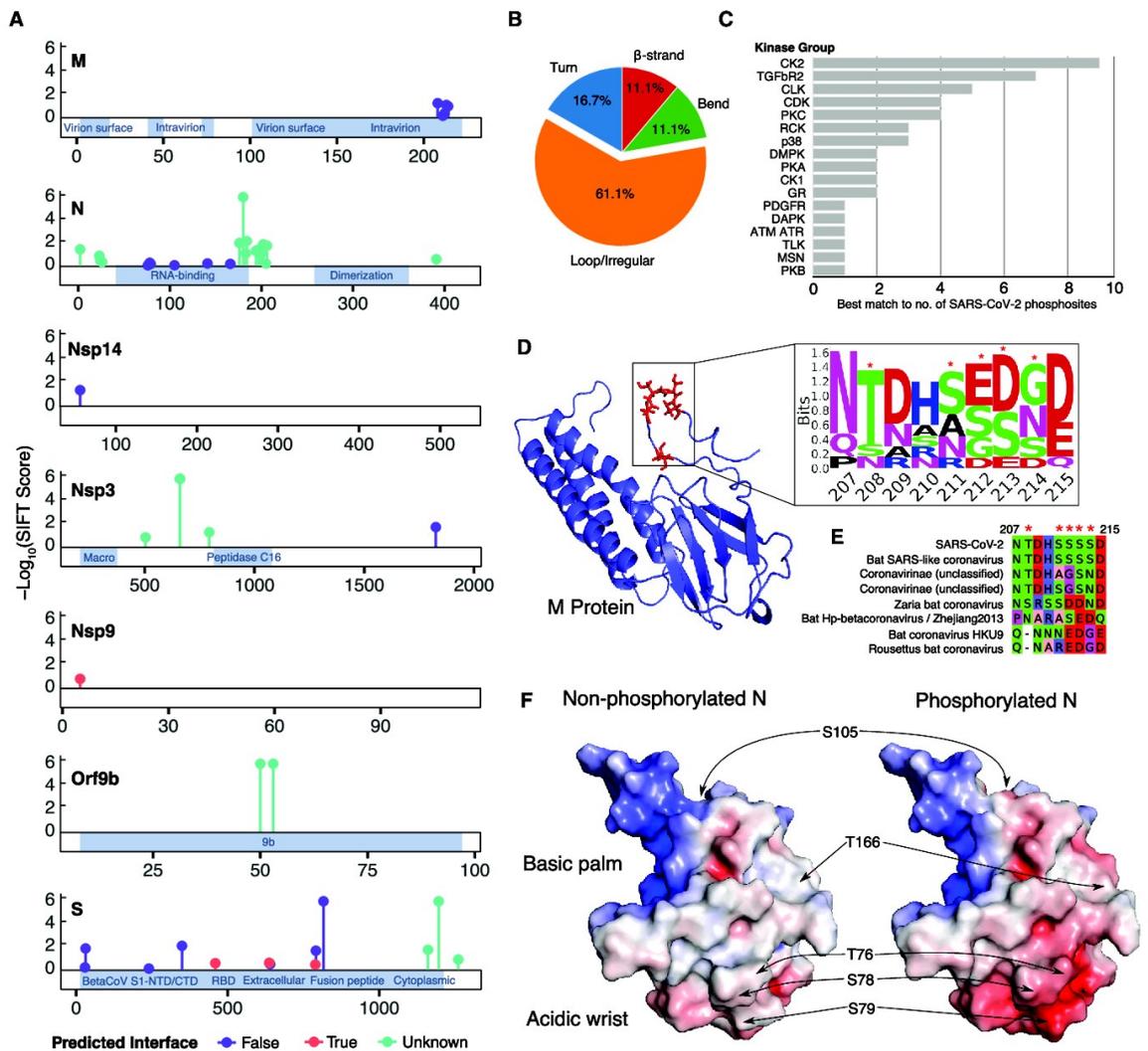


Figure 2. Overview of SARS-CoV-2 Viral Protein PHs in the Host Cell

(A) Localization of PHs across viral protein sequences from this study and a previous study (Davidson et al., 2020). Stem height indicates predicted deleteriousness of alanine substitutions. Dot color indicates whether the residue is (true) or is not (false) predicted to form part of an interaction interface based on SPPIDER analysis. Positions with no structural coverage are excluded from interface prediction.

(B) Distribution of secondary structure elements in which viral PHs were found, as classified by the define secondary structure of proteins (DSSP) tool.

(C) Distribution of top matching host kinases to viral PHs according to NetPhorest tool (Horn et al., 2014).

(D) Phosphorylation cluster in the C-terminal tail of the M protein (red residues) structure (Heo and Feig 2020) and associated sequence motif. Asterisks indicate PHs.

(E) Alignment of M protein phosphorylation clusters across different coronaviruses. Asterisks indicate PHs.

(F) Surface electrostatic potential of non-phosphorylated (left) and phosphorylated (right) RNA-binding domains of the N protein (PDB: 6M3M). Positions of PHs are indicated by arrows. Blue denotes a positive charge potential, and red indicates a negative charge potential. Electrostatic potential was computed with the Advanced Poisson-Boltzmann Solver (APBS) tool after preparation with the PDB2PQR tool.

the regulation of positive transcription elongation factor b (P-TEFb [CDK9]) and transcriptional regulation of the virus, similar to how these proteins are regulated during HIV infection (Mbonye et al., 2015).

SARS-CoV-2 Infection Regulates Host Kinase Signaling

To study global changes in kinase signaling and their effect on host protein phosphorylation, regulated phosphorylation sites were grouped in five clusters based on their dynamics using a

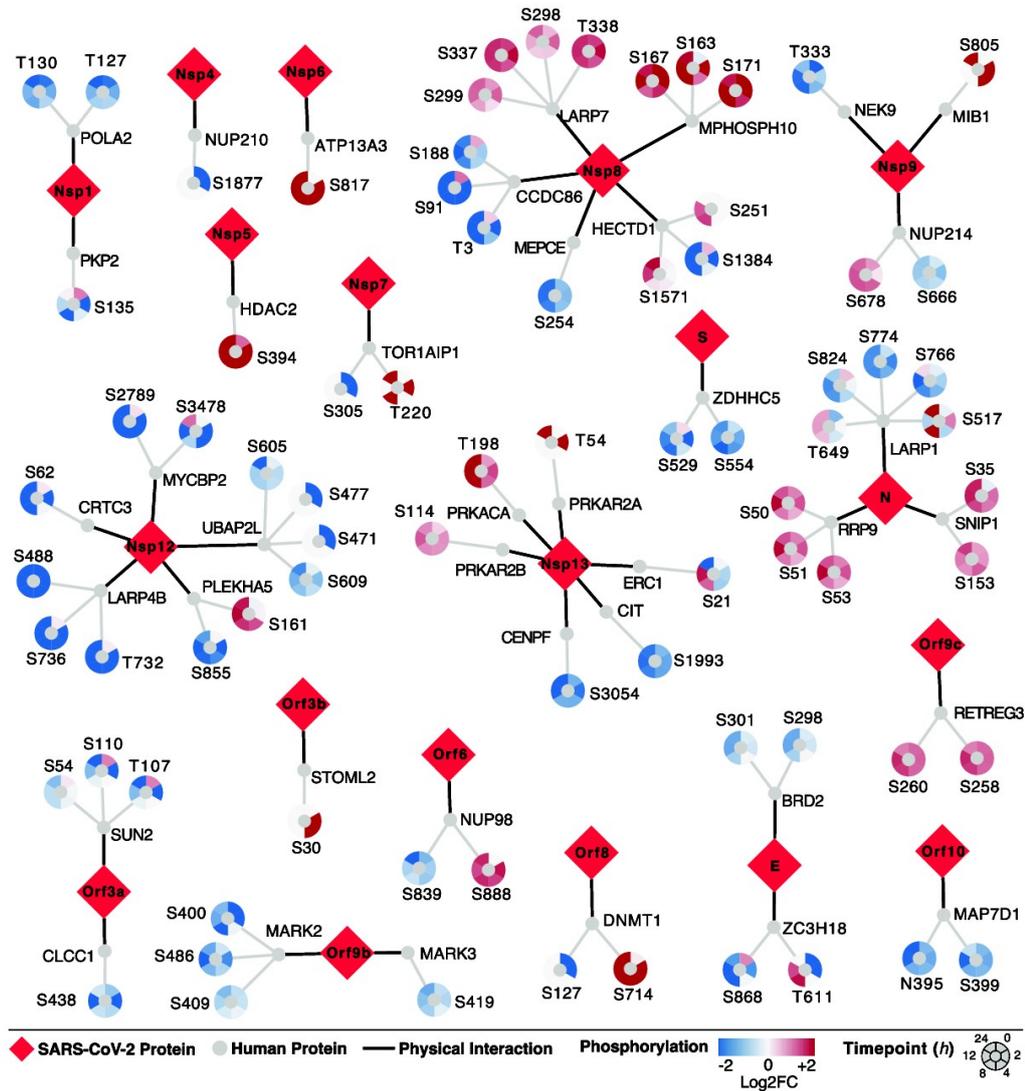


Figure 3. Phosphorylation on SARS-CoV-2 Virus-Human Interacting Proteins

The SARS-CoV-2 virus-host protein-protein interaction map (Gordon et al., 2020) found 332 human proteins interacting with 27 (26 wild-type and 1 mutant) viral proteins. Here we found 40 of 332 proteins significantly differentially phosphorylated across at least two time points (adjusted $p < 0.05$ and absolute value of \log_2 fold change $[\text{abs}(\log_2\text{FC})] > 1$). Viral proteins are shown as red diamonds. Interacting human proteins are shown as gray circles. PHs emanate from human proteins, colored by their \log_2 fold change compared with uninfected control samples (red, increase; blue, decrease) at each time point (0, 2, 4, 8, 12, and 24 h after infection) in a clockwise fashion. An interactive version of phosphorylation data can be found at <https://kroganlab.ucsf.edu/network-maps>.

data-driven clustering approach (Figure 4A; STAR Methods). For each of the groups, an enrichment analysis was performed for functions and pathways (Figure 4A; Table S3). The dynamics of these changes can be linked to the viral life cycle: entry (0–2 h), replication (4–12 h), and egress (24 h). Clusters 1 and 2 include phosphorylation sites that are, on average, upregulated during infection. Cluster 1 sites tended to be upregulated within 2 h

(i.e., linked to viral entry) and were enriched in mRNA processing, cell cycle, apoptosis, and proteins involved in HIV infection. Cluster 2 included apoptosis proteins with later onset of phosphorylation, associated with replication and/or egress. Phosphorylation sites in clusters 3 and 4 were downregulated and enriched in RNA-processing functions. Sites within cluster 5 possessed a dynamic response to infection, with immediate

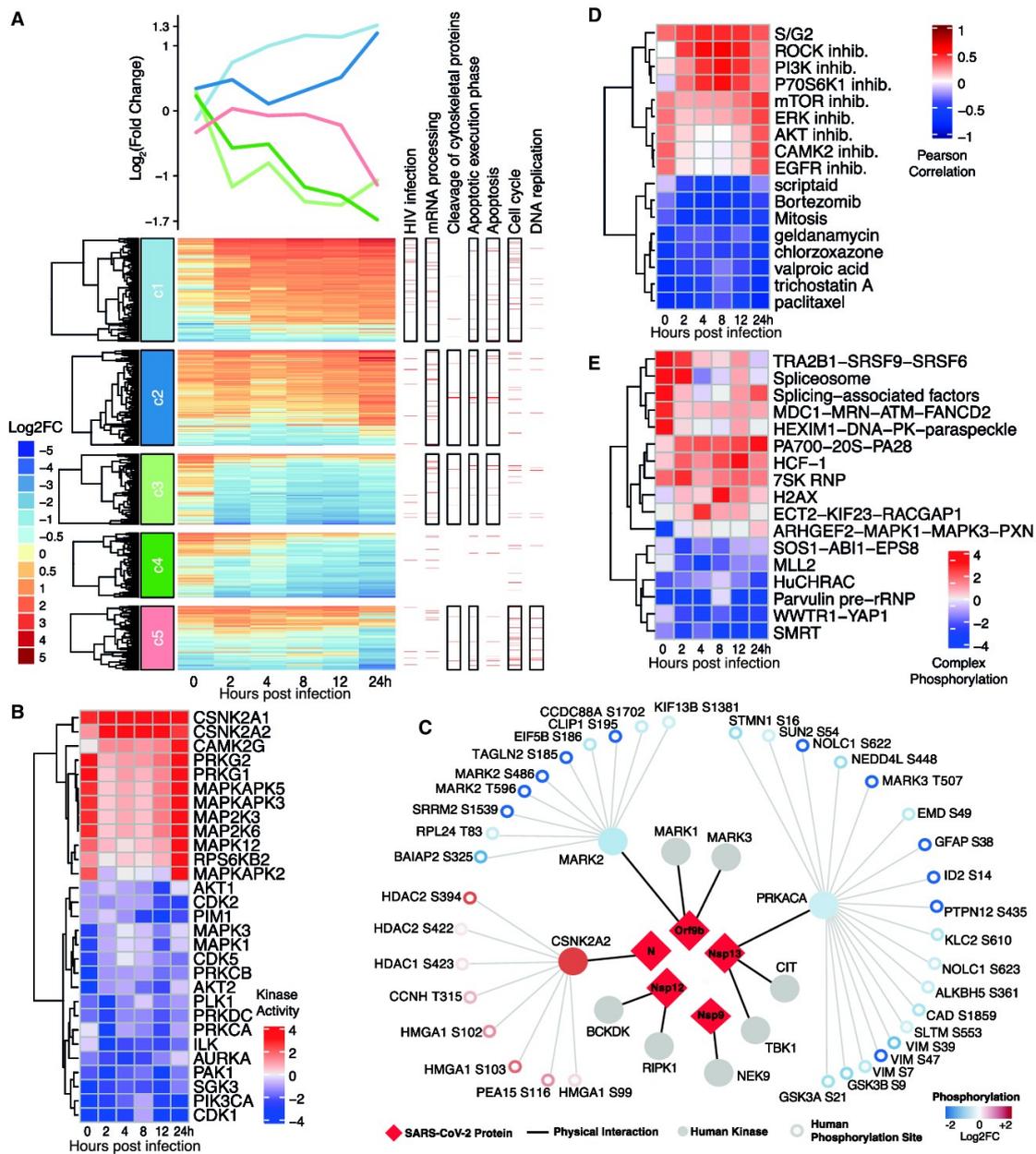


Figure 4. Signaling Changes in Host Cells in Response to SARS-CoV-2 Infection

(A) Clusters of significantly changing PHs ($abs(\log_2FC) > 1$ and adjusted $p < 0.05$) across the time course of infection with non-redundant enriched Reactome pathway terms (adjusted p value [q] < 0.01) for each cluster. Horizontal red lines below each pathway term correspond to phosphorylated proteins belonging to the pathway, and a black-bordered rectangle is indicative of a significantly enriched term.

(B) Kinases depicting a strong change in activity upon infection ($abs(\log_{10}(p)) > 2.5$) in at least one time point, with predicted activity in at least 5 of 6 time points.

(C) Schematic representation of interaction between host kinases and SARS-CoV-2 viral proteins from Gordon et al. (2020). Substrate PHs for each kinase are color-coded as blue (down) and red (up) based on the direction of change during infection. Only PHs corresponding to the kinase activity direction are shown.

(legend continued on next page)

downregulation followed by a rise during the middle and renewed downregulation at late time points. This cluster was enriched for DNA replication and the cell cycle, among others. These observations are corroborated by standard Gene Ontology (GO) enrichment analyses of biological processes regulated by phosphorylation (Figure S1G; Table S3; STAR Methods).

We estimated activity regulation for 97 kinases based on the regulation of their known substrates (Ochoa et al., 2016; Hernandez-Armenta et al., 2017; Table S4), with the strongest regulation linked to viral entry (0–2 h) and late replication/egress (24 h). The kinases predicted to be most strongly activated (Figures 4B and S2A) include several members of the p38 pathway, including p38 γ (MAPK12), CK2 (CSNK2A1/2), Ca²⁺/calmodulin-dependent protein kinase (CAMK2G), and the guanosine monophosphate (GMP)-dependent protein kinases PRKG1/2, which can inhibit Rho signaling. Kinases predicted to be downregulated include several cell cycle kinases (CDK1/2/5 and AURKA), cell growth-related signaling pathway kinases (PRKACA, AKT1/2, MAPK1/3, and PIM1), and the cytoskeleton regulators (PAK1), among others. Kinase activity estimates based on the 24-h mock control gave highly correlated results ($r = 0.81$), identifying the same set of highly regulated kinases (Figure S2D). Some of the changes in kinase activity can be directly linked to host-viral protein interactions (Figure 4C). Among the 10 interacting kinases detected in a virus-host protein-protein interaction map (Gordon et al., 2020), an increase in activity for CK2 and a decrease for MARK2 and PRKACA were observed (Figure 4C). Of note, although we predict decreased activity for PRKACA based on phosphorylation of its substrates, we simultaneously detected a significant increase in T198 phosphorylation (8, 12, and 24 h post infection) within its activation loop, suggesting an increase in PRKACA activity. It is possible that Nsp13 is sequestering active PRKACA away from its typical substrates.

To better understand the signaling states of cells over the course of infection, we compared our data with a compilation of public phosphoproteomics datasets of other conditions (Ochoa et al., 2016; Figures 4D and S2B). The first and last time point of infection resembled a kinase activation state induced by inhibition of mTOR, ERK, AKT, and EGFR, consistent with the estimated kinase activities of these growth-related pathways. Between 2 and 12 h after infection, kinase activity states resembling inhibition of phosphatidylinositol 3-kinase (PI3K), p70S6K, and Rho-associated protein kinases (ROCKs) were observed. Finally, several of the time points resembled S/G2 cell cycle state, suggestive of a cell cycle block. Conversely, some conditions were anticorrelated with kinase activity profiles (Figures 4D and S2B). In line with an S/G2 cell-cycle block, infection signaling appeared opposite to that of a mitotic cell. In addition, inhibitors of histone deacetylases (HDACs) (scriptaid and trichostatin A), the proteasome (bortezomib), Hsp90 (geldanamycin), and voltage-gated sodium channels (valproic acid)

were also anticorrelated. These drugs, or drugs targeting these protein activities, could induce a signaling state that inhibits viral replication.

To further link kinase activities to downstream protein complexes, enrichment of up- or downregulated phosphorylation sites was determined within a curated set of human protein complexes defined by CORUM (Giurgiu et al., 2019; Figures 4E and S2C). This analysis revealed significant changes in phosphorylation of splicing related complexes (spliceosome), the proteasome (PA700-20S-PA28), and chromatin remodeling complexes (HuCHRAC and MLL2). In addition, a subset of regulated phosphorylation sites were detected that have known regulatory functions or high predicted functional scores (Ochoa et al., 2020) that are linked to regulation of protein activities (Table S5). Consistent with the observed signaling changes described above, these regulatory phosphorylation sites are involved in activation of chaperones (including HSP90), proteasome activity, inhibition of the anaphase-promoting complex (APC), and regulation of HDACs and cytoskeleton proteins, among others.

CK2 and N Co-localize at Virus-Induced Filopodial Protrusions

The phosphoproteomics data indicated regulation of several kinases and effector proteins related to cytoskeleton organization upon SARS-CoV-2 infection. Kinases downstream of the Rho/Rac/Cdc42 GTPases (PAK1/2 and ROCK1/2) and several well-characterized phosphorylation site targets of PAK1/2 kinase in vimentin (VIM S39 and S56) and stathmin (STMN1 S16 and S25) were found to be downregulated during infection (Figures 5A and 5B). The interaction of Nsp7 with RHOA (Gordon et al., 2020) may contribute to this downregulation. In contrast, signaling via CK2 is strongly upregulated, as determined by the increase in phosphorylation of well-characterized target sites (Figures 5A and 5B). Among the many roles of this kinase, we noted increased phosphorylation of cytoskeleton protein targets such as α -Catenin (CTNNA1 S641) and the heavy chain of the motor protein Myosin IIa (MYH9 S1943). In addition to these kinase-mediated effects, the Nsp2 protein of SARS-CoV-2 also interacts directly with strumpellin (WASHC5), a subunit of the actin assembly-inducing WASH complex (Gordon et al., 2020), further implicating cytoskeleton regulation during infection. To study the relevance of these observations in a human infection model, high-resolution immunofluorescence imaging of fixed Caco-2 human colon epithelial cells was performed 24 h after infection (STAR Methods).

SARS-CoV-2 infected Caco-2 cells were imaged for filamentous actin and the SARS-CoV-2 M protein, revealing prominent M protein clusters, possibly marking assembled SARS-CoV-2 viral particles, localized along the shafts and at the tips of actin-rich filopodia (Figures 5B and S3B). SARS-CoV-2 infection induced a dramatic increase in filopodial protrusions, which were significantly longer and more branched than in uninfected

(D) Correlation of kinase activity profiles of each time point with other biological conditions with at least one significantly changing kinase ($|\text{abs}(\log_{10}(p)) > 2.5$) and having significant correlation with at least one time-point (false discovery rate [FDR] < 5%).

(E) Overall phosphorylation change ($-\log_{10}(p)$) of a protein complex, estimated as the change in phosphorylation on member proteins. Only non-redundant protein complexes with a significant change in phosphorylation ($|\text{abs}(\log_{10}(p)) > 2.5$) in at least one time point are shown. See also Figure S2.

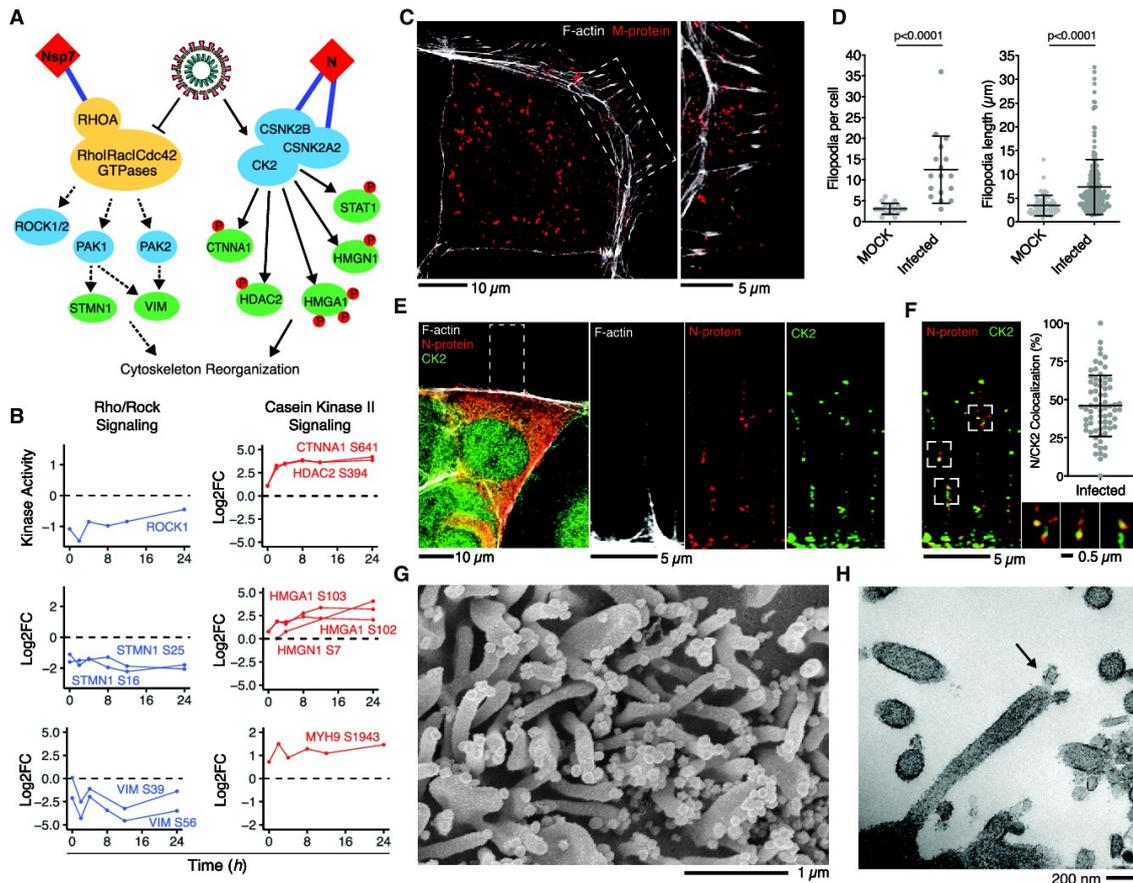


Figure 5. Colocalization of CK2 and Viral Proteins at Actin Protrusions

(A) Pathway of regulated PHs and SARS-CoV-2 interaction partners involved in cytoskeletal reorganization. Dashed lines indicate downregulation of activity, while solid lines indicate upregulation of activity.
 (B) Regulation of individual kinase activity or PHs depicted in (A).
 (C) Caco-2 cells infected with SARS-CoV-2 at an MOI of 0.1 for 24 h prior to immunostaining for F-actin and M protein, as indicated. Shown is a confocal section revealing M protein localization along and to the tip of filopodia (left) and magnification of the dashed box (right).
 (D) Dot plot quantification of the number and length of filopodia in untreated (mock) or infected Caco-2 cells for 24 h with SARS-CoV-2. Filopodium length was measured from the cortical actin to the tip of the filopodium. Error bars represent SD. Statistical testing by Mann-Whitney test.
 (E) Caco-2 cells infected with SARS-CoV-2 at an MOI of 0.01 for 24 h prior to immunostaining for F-actin, N protein, and casein kinase II (CK2) as indicated (left). Shown is magnification of the dashed box as single channels (right).
 (F) Magnification of the dashed box from (E) with quantification of colocalization between CK2 and N protein throughout infected Caco-2 cells. Displayed is the proportion of N protein-positive particles colocalizing with CK2. Error bars represent SD.
 (G and H) Scanning electron microscopy (G) and transmission electron microscopy (H) images of SARS-CoV-2 budding from Vero E6 cell filopodia (black arrow in H).
 See also Figure S3.

cells (Figure 5D). Uninfected cells also exhibited filopodial protrusions, but their frequency and shape were dramatically different (Figure S3A). Reorganization of the actin cytoskeleton is a common feature of many viral infections and is associated with different stages of the viral life cycle (Taylor et al., 2011).

We hypothesize that induction of virus-containing filopodia could be important for SARS-CoV-2 egress and/or cell-to-cell spread within epithelial monolayers. Given that Rho/PAK/

ROCK signaling is downregulated, we next asked whether CK2 could play a role in this process. At 24 h, infected cells showed CK2 expression along the thin filopodial protrusions (Figure 5E), partially co-localized with SARS-CoV-2 N protein (Figure 5F). Scanning and transmission electron microscopy were used (Figures 5G, 5H, S3C, and S3D) to image the cellular protrusions at higher resolution. Assembled viral particles are clearly visible along these filopodia (Figure 5G), with instances where the viral

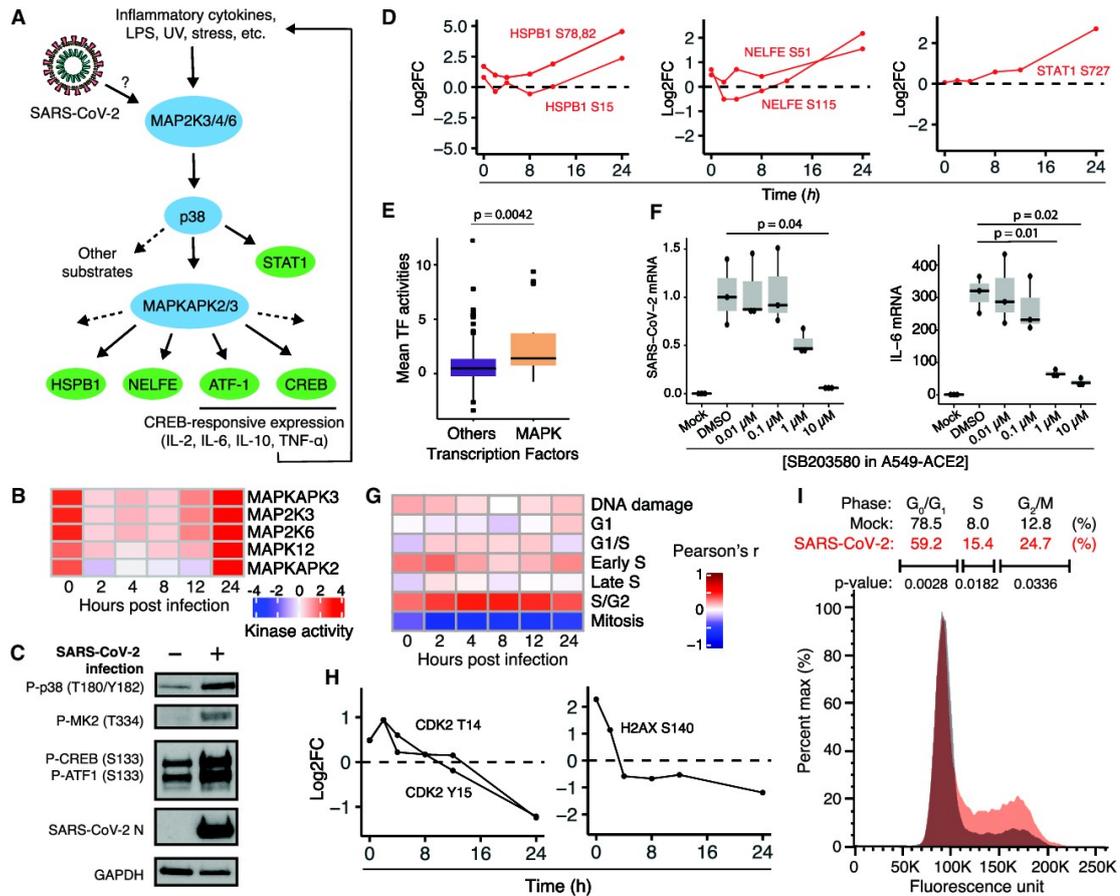


Figure 6. SARS-CoV-2 Activates the p38/MAPK Signaling Pathway and Causes Cell Cycle Arrest

(A) Diagram of the p38/MAPK signaling pathway.
 (B) Kinase activity analysis for kinases in the p38/MAPK pathway.
 (C) Western blot analysis of phosphorylated p38/MAPK signaling components in mock- and SARS-CoV-2-infected ACE2-A549 cells 24 h after infection.
 (D) Log₂ fold change profiles of indicated p38/MAPK substrates during SARS-CoV-2 infection in Vero E6 cells.
 (E) Transcription factor activity analysis of SARS-CoV-2-infected A549, Calu-3, and NHBE cells, comparing p38/MAPK transcription factors with transcription factors not associated with the p38/MAPK pathway. Statistical test: Mann-Whitney test.
 (F) qRT-PCR analysis of the indicated mRNA from ACE2-A549 cells pre-treated with the p38 inhibitor SB203580 at the indicated concentrations for 1 h prior to infection with SARS-CoV-2 for 24 h. Statistical test: Student's t test. See also Figures S4 and S5.
 (G) Heatmap of Pearson's correlation coefficients comparing SARS-CoV-2-infected Vero E6 phosphorylation profiles with profiles of cells with induced DNA damage and cells arrested at the indicated cell cycle stages.
 (H) Log₂ fold change profiles of the indicated cell cycle and DNA damage substrates during SARS-CoV-2 infection in Vero E6 cells.
 (I) DNA content analysis of cells infected with SARS-CoV-2 for 24 h compared with mock-infected cells.

particles appear to be budding from the protrusions (Figure 5H). Finally, we performed a global phosphoproteomics analysis of Vero E6 cells overexpressing N protein and observed CK2 activity to be significantly upregulated (Figure S3E; Tables S1 and S4). Because CK2 activity can promote actin polymerization (D'Amore et al., 2019), we hypothesize that N protein may allosterically control CK2 activity and regulate cytoskeleton organization.

SARS-CoV-2 Infection Promotes p38/MAPK Signaling Activity and Cell Cycle Arrest

Kinase activity analysis of SARS-CoV-2 phosphorylation profiles predicted upregulation of several components of the p38/mitogen-activated protein kinase (MAPK) signaling pathway, including MAP2K3, MAP2K6, MAPK12, MAPKAPK2 (MK2), and MAPKAPK3 (Figures 6A and 6B). Immunoblotting for activated phospho-p38 (T180/Y182), phospho-MK2 (T334), and

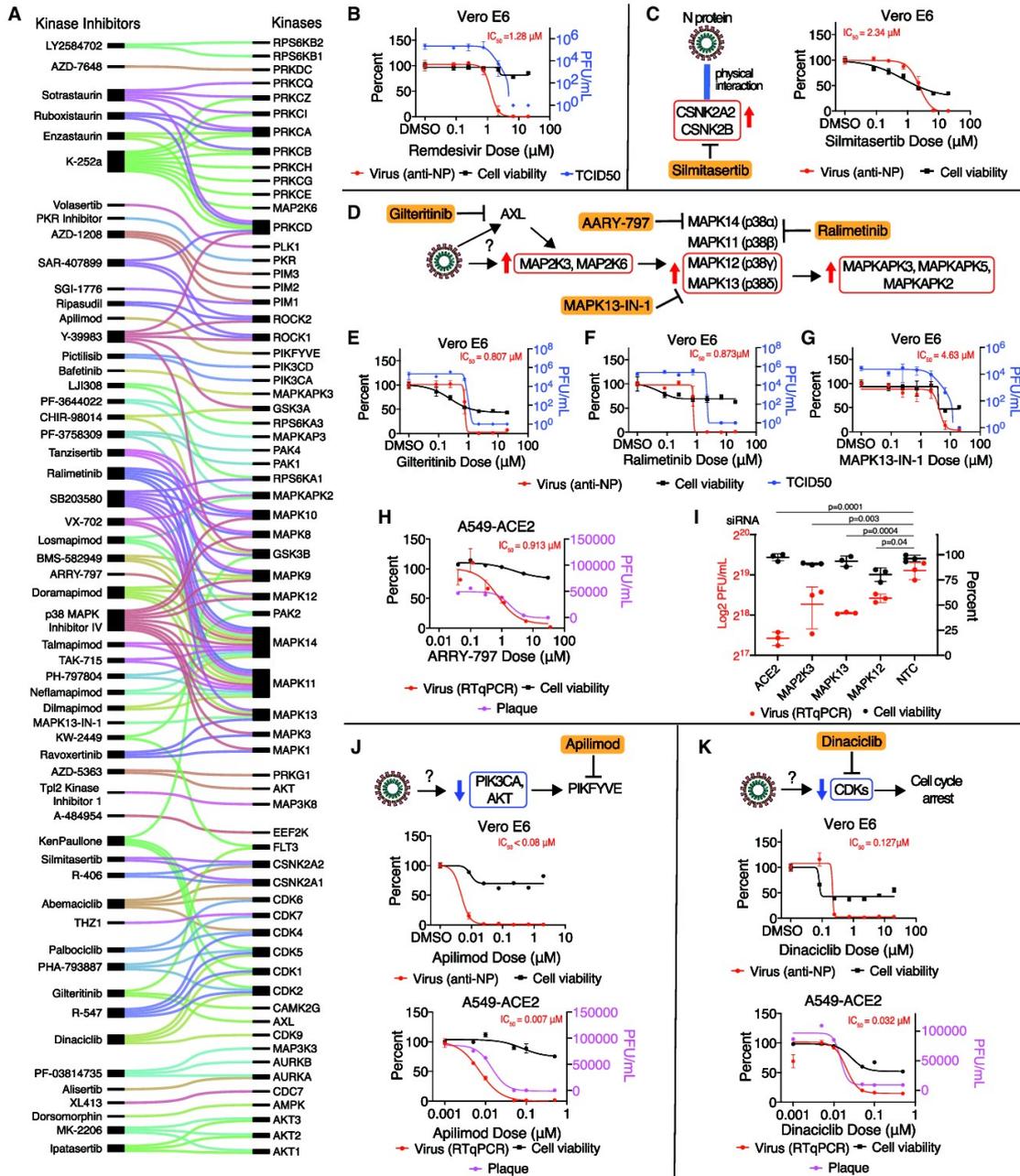


Figure 7. Mapping Regulated Kinases to Kinase Inhibitors Identifies SARS-CoV-2 Therapies

(A) Kinase inhibitors (left) mapped to kinases (right) whose activity was regulated by SARS-CoV-2 infection. Lines connecting them indicate known kinase targets for each drug/compound.

(B) Vero E6 cells pre-treated with remdesivir at the indicated doses, followed by SARS-CoV-2 infection for 48 h. Percent viral titer compared with mock drug treatment (anti-NP antibody; red line, dots, and text) and cell viability (black) is depicted. Error bars represent SD.

(C) As in (B). Vero E6 cells were treated with the CK2 inhibitor silmitasertib. Physical interactions between N protein and the CSNK2A2 and CSNK2B CK2 subunits were observed in a prior study (Gordon et al., 2020).

(legend continued on next page)

phospho-cAMP response element-binding protein (CREB) and phospho-ATF-1 at their respective MAPKAPK2 sites (S133 in both) confirmed activation of the p38/MAPK pathway during SARS-CoV-2 infection in ACE2-expressing A549 human lung carcinoma cells (ACE2-A549) (Figure 6C). Furthermore, phosphoproteomics data depict increased phosphorylation of p38 pathway substrates such as negative elongation factor E (NELFE), heat shock protein beta-1 (HSPB1), and signal transducer and activator of transcription 1-alpha/beta (STAT1), among others (Figure 6D). Regulation of these sites occurs late in the time course (24 h after infection), likely reflecting a more advanced stage of viral infection, replication, and egress.

The p38/MAPK pathway mediates the cellular response to environmental stress, pathogenic infection, and pro-inflammatory cytokine stimulation, whereas downstream effectors of the pathway include transcription factors and RNA binding proteins that promote inflammatory cytokine production (Cua-drado and Nebreda 2010; Wen et al., 2010). Analysis of estimated transcription factor activity from gene expression data (STAR Methods; Table S6) derived from the infection of a human lung carcinoma cell line (A549), a human epithelial lung cancer cell line (Calu3), and primary human bronchial epithelial (NHBE) cells demonstrated that transcription factors regulated by the p38/MAPK pathway were among the most highly activated upon infection (Figure 6E; Blanco-Melo et al., 2020).

To investigate the contribution of the p38/MAPK pathway to cytokine production, SARS-CoV-2-infected ACE2-A549 cells were treated with the p38 inhibitor SB203580. The mRNA of the inflammatory cytokines interleukin-6 (IL-6), tumor necrosis factor alpha (TNF- α), and others increased during infection and were inhibited by p38 inhibition in a dose-dependent manner (Figures 6F, right, and S4A). Interestingly, p38 inhibition also reduced SARS-CoV-2 subgenomic mRNA (Figure 6F, left) in the absence of major cellular toxicity (Figure S5), indicative of reduced viral replication. The SB203580-induced decrease in virus production was further confirmed using an anti-SARS-CoV-2 N protein (anti-NP) antibody-based assay (Figure S5, New York Vero E6). Multiplexed ELISA analysis of supernatants of cells from the same experiment demonstrated strong upregulation of inflammatory cytokines at the protein level, including IL-6, CXCL8, CCL20, and CCL2, which were decreased upon p38 inhibition (Figure S4B; Table S7). However, because SARS-CoV-2 replication is also inhibited by SB203580, we cannot deconvolve the contributions of p38/MAPK pathway activity and SARS-CoV-2 virus presence on cytokine production.

Comparing phosphoproteomics profiles of SARS-CoV-2-infected cells with a database of phosphorylation profiles collected at specific cell cycle stages, viral infection was most

highly correlated with cells arrested at the S/G2 transition and was negatively correlated with profiles of cells in mitosis (Figure 6G). We also observed SARS-CoV-2-dependent regulation of CDK2 T14/Y15 phosphorylation, initially increased in response to SARS-CoV-2 infection at 2 h, followed by a decrease over the remainder of the infection time course (Figure 6H, left). CDK2 activity promotes transition from the G2 phase of the cell cycle into mitosis and is inhibited by phosphorylation at positions T14 and Y15 by kinases WEE1 and MYT1, preventing premature entry into mitosis (Parker and Piwnicka-Worms 1992; Mueller et al., 1995). CDK2 can also become phosphorylated when the cell cycle is arrested because of checkpoint failure or DNA damage. In addition, H2AX S140 phosphorylation (i.e., γ -H2AX), a hallmark of the DNA damage response, exhibited a profile similar to CDK2, suggesting that the DNA damage response may become activated early during infection (Rogakou et al., 1998; Figure 6H, right).

To more directly test whether SARS-CoV-2 infection affects cell cycle progression, cells were infected with SARS-CoV-2 for 24 h, and their DNA content was measured using DAPI DNA staining and flow cytometry. A significant increase in the fraction of cells in S phase and at the G2/M transition and a decrease in the fraction of cells in G0/G1 phase were observed (Figures 6I and S4C). This observation is consistent with arrest between S and G2 phases of the cell cycle. A relationship between p38 activity and cell cycle arrest has been described previously, and the two could be linked mechanistically during SARS-CoV-2 infection (Lee et al., 2002; Yee et al., 2004).

Mapping Kinase Activities to Pharmacological Modulators Identifies SARS-CoV-2 Therapies

To identify effective therapies for SARS-CoV-2 infection, kinase inhibitors were mapped to the most differentially regulated kinase activities (Figure 7A) and to specific phosphorylation sites (Table S8; STAR Methods). This resulted in a list of 87 drugs and compounds: 10 FDA-approved, 53 undergoing clinical testing, and 24 pre-clinical. Many of the drugs and compounds identified were reported to target several host kinases in cell-free assays at a minimum, but many have been observed to hit targets in cellular assays as well (Figure 7A). We reasoned that testing molecules with both overlapping and unique targets would help specify the molecular targets of greatest importance for SARS-CoV-2. Here, 68 total drugs and compounds were tested for antiviral efficacy (via qRT-PCR, anti-NP antibody, plaque assay, and/or TCID₅₀) and cellular toxicity at two different institutions (in New York [Mount Sinai, 25 drugs/compounds] and Paris [Institut Pasteur, 62]) and in two cell lines (Vero E6 [68] and

(D) Predicted increased kinase activity for the p38 signaling pathway and drugs/compounds targeting pathway members (ralimetinib, MAPK13-IN-1, and ARRY-797) and upstream drivers (gilteritinib).

(E–G) As in (B). Vero E6 cells treated with the AXL inhibitor gilteritinib (E), the MAPK11/14 inhibitor ralimetinib (F), or the MAPK13 inhibitor MAPK13-IN-1 (G) prior to SARS-CoV-2 infection.

(H) A549-ACE2 lung epithelial cells were treated with the MAPK14 inhibitor ARRY-797 prior to SARS-CoV-2 infection.

(I) Small interfering RNA (siRNA) knockdown of p38 pathway genes in A549-ACE2 leads to a significant decrease in SARS-CoV-2 viral replication (red), as assessed by qRT-PCR in the absence of effects on cell viability (black). ACE2 and non-targeting siRNAs are included as positive and negative controls, respectively.

(J and K) Vero E6 or A549-ACE2 cells were treated with PIKFYVE inhibitor apilimod (J) or the CDK inhibitor dinaciclib (K) prior to SARS-CoV-2 infection. See also Figure S6.

A549-ACE2 [61]). All pharmacological profiling results can be found in [Figure S5](#) and [Table S8](#).

We found pharmacological inhibitors of CK2, p38 MAPK signaling, PIKFYVE, and CDKs to possess strong antiviral efficacy. Cells were pre-treated with inhibitor molecules, followed by SARS-CoV-2 infection ([STAR Methods](#)), and virus quantity (anti-NP antibody against SARS-CoV-2) and cell viability were quantified 48 h after infection. As a positive control and for comparison, remdesivir was tested, and the expected favorable antiviral activity was observed (half maximal inhibitory concentration [IC_{50}] = 1.28 μ M; [Figure 7B](#)). Silmitasertib, an inhibitor of CSNK2A1 and CSNK2A2, was found to possess antiviral activity (IC_{50} = 2.34 μ M; [Figures 7C](#) and [S5](#)). In conjunction with data supporting physical interaction ([Gordon et al., 2020](#)) and colocalization with N protein ([Figure 5F](#)), as well as a potential role in remodeling extracellular matrix upon infection ([Figures 5](#) and [S3](#)), CK2 signaling appears to be an important pathway hijacked by SARS-CoV-2. Furthermore, silmitasertib is currently being considered for human testing as a potential treatment for COVID-19.

To probe SARS-CoV-2 dependence on MAPK signaling, SARS-CoV-2 replication was measured in response to pharmacological and genetic perturbation of MAPK components that were upregulated during infection ([Figure 7D](#)). Potent antiviral activity was observed for gilteritinib ([Figure 7E](#); IC_{50} = 0.807 μ M), an inhibitor of AXL kinase, upstream of p38; ralimetinib ([Figure 7F](#); IC_{50} = 0.873 μ M), an inhibitor of MAPK11 (p38 α) and MAPK14 (p38 β); MAPK13-IN-1 ([Figure 7G](#); IC_{50} = 4.63 μ M), an inhibitor of MAPK13 (p38 δ); and ARRY-797 ([Figure 7H](#); IC_{50} = 0.913 μ M) in A549-ACE2 cells, a MAPK14 inhibitor. To further probe the dependence of SARS-CoV-2 on p38 pathway members, small interfering RNA (siRNA)-mediated knockdown of MAP2K3, p38 δ (MAPK13), and p38 γ (MAPK12) was performed in A549-ACE2 cells, and a significant decrease in SARS-CoV-2 replication was observed for all three, with little to no effect on cell viability ([Figure 7I](#)).

In addition, we noted marked regulation of phosphatidylinositol enzyme activities for PIK3CA, PLCB3, and PIKFYVE, suggesting a potential role for the appropriate balance of phosphatidylinositol species. To target this process, apilimod, a small-molecule inhibitor of PIKFYVE, was tested and found to possess strong antiviral activity in two cell lines (Vero E6, IC_{50} < 0.08 μ M; A549-ACE2, IC_{50} = 0.007 μ M), corroborated by a recent study ([Ou et al., 2020](#); [Figure 7J](#)). Lastly, we noted pronounced regulation of CDK signaling pathways ([Figure 4B](#)) and cell cycle stage ([Figure 6I](#)) during viral infection, suggesting that the virus may regulate the cell cycle to enhance viral replication. Accordingly, strong antiviral activity for the CDK inhibitor dinaciclib was observed across two cell lines (Vero E6, IC_{50} = 0.127 μ M; A549-ACE2, IC_{50} = 0.032 μ M) ([Figure 7K](#)).

DISCUSSION

We used a mass spectrometry-based approach to study perturbations in protein abundance and phosphorylation during SARS-CoV-2 infection. Viral proteins increased, starting 8 h after infection, indicative of viral replication, whereas only small changes in host protein abundance were observed within 24 h. In contrast,

large changes were observed in protein phosphorylation, highlighting the degree by which the virus makes use of the host post-translational regulatory systems to promote rapid changes in cellular signaling.

Changes in phosphorylation reflect altered activities of kinases that are hijacked during the infection. Based on changes in phosphorylation of their annotated substrates, we estimated changes in activity of 97 of the 518 human kinases. The changes in kinase activity offer insights into the biology of viral infection. Kinases represent ideal drug targets; here, we identified kinases and pathways altered by SARS-CoV-2 infection that can be targeted by 87 FDA-approved drugs and compounds in clinical trials or in preclinical development ([Table S8](#)).

The most strongly regulated kinases fall into a set of signaling pathways that include p38/MAPK signaling, AKT and ERK signaling, Rho GTPase and CK2 cytoskeleton signaling, and cell cycle regulation. The downregulation of ROCK and PAK kinase activity and upregulation of CK2 cytoskeleton-related targets suggest virus-induced changes in cytoskeleton organization. Imaging of infected cells revealed formation of actin-rich filopodia containing viral proteins. Higher-resolution electron microscopy data confirm the presence of assembled viral particles in these structures. Many viruses, including vaccinia, Ebola, and Marburg, hijack the host cell cytoskeleton to promote egress and rapid cell-to-cell spread across epithelial monolayers. Vaccinia promotes Arp2/3-dependent actin assembly, producing a filopodial protrusion with a virus at the tip ([Leite and Way 2015](#)). In contrast, Marburg virus hijacks the unconventional motor protein Myosin X, which promotes filopodium formation and trafficks the virus along the filopodium shaft. The SARS-CoV-2 protein clusters peppered throughout the length of filopodial protrusions more closely resemble Marburg than vaccinia, but additional work is required to understand whether SARS-CoV-2 makes use of either Myosin X motor activity or actin filament assembly to move along filopodia. CK2 is known to phosphorylate myosin proteins at endocytic sites to drive actin polymerization ([Fernández-Golbano et al., 2014](#)). Furthermore, CK2 has been found to regulate actin tail formation during vaccinia virus infection, enabling efficient cell-to-cell spread of the virus ([Alvarez and Agaisse 2012](#); [Smith and Law 2004](#)). Here the CK2 inhibitor silmitasertib displayed robust antiviral activity, suggesting a role of this kinase in regulating the SARS-CoV-2 life cycle.

In addition, kinase activity profiling analysis shows that CDK1/2 activities are significantly reduced by SARS-CoV-2 infection, leading to a S/G2 phase arrest that is similar to infectious bronchitis virus (IBV), a prototypical coronavirus ([Dove et al., 2006](#); [Li et al., 2007a](#)), and other RNA viruses ([Lilley et al., 2007](#); [Ariumi et al., 2008](#)). Arresting cells in S/G2 phase may provide benefits for viral replication and progeny production by ensuring an abundant supply of nucleotides and other essential host DNA repair/replication proteins ([Chaurushiya and Weitzman 2009](#)).

The predicted increase in p38/MAPK activity led us to investigate the effects of p38/MAPK inhibition on pro-inflammatory cytokine production and viral replication in SARS-CoV-2-infected cells. Recent immunological studies have indicated that increased IL-6, IL-10, and TNF- α and lymphopenia are associated with severe COVID-19 cases ([Pedersen and Ho 2020](#)).

The p38/MAPK pathway responds to and controls production of potentially harmful pro-inflammatory cytokines. Several pathogenic viral infections induce a p38/MAPK signaling state that exhibits uncontrolled positive feedback regulation, leading to excessive inflammation associated with severe disease. Inhibition of p38/MAPK signaling suppressed the overproduction of inflammatory cytokines induced by several viral infections, including SARS-CoV, Dengue virus, and influenza A virus, improving survival in mice (Fu et al., 2014; Growcott et al., 2018; Jimenez-Guardeño et al., 2014). However, p38/MAPK inhibition did not directly impair the virus in these cases but, instead, the host's immune response to the infection. In contrast, during SARS-CoV-2 infection, p38/MAPK inhibition suppressed cytokine production and impaired viral replication by a still unknown mechanism, suggesting that p38/MAPK inhibition may target multiple mechanisms related to COVID-19 pathogenesis.

We tested 68 drugs and compounds and found antiviral activity for several that are FDA approved, in clinical testing, or under preclinical development for various diseases, including silmitasertib (CK2, phase 2), gilteritinib (AXL, FDA approved), ARRY-797 (p38, phase 2/3), MAPK13-IN-1 (p38, preclinical), SB203580 (p38, preclinical), ralimetinib (p38, phase 2), apilimod (PIKFYVE, phase 1), and dinaciclib (CDK, phase 3), among others (Figure S5; Table S8). Silmitasertib, a small molecule undergoing clinical trials for various cancers, is now being considered for testing in humans to combat COVID-19. Although the effectiveness of CK2 inhibition may be attributed to its regulation of stress granules (Gordon et al., 2020), viral egress and dissemination could be facilitated by CK2-mediated remodeling of the extracellular matrix (Figure 5).

Ralimetinib is currently in phase 2 clinical trials for treatment of ovarian cancer (Patnaik et al., 2016), and ARRY-797 is in phase 3 clinical trials for treatment of cardiomyopathy. The antiviral activity observed for gilteritinib, an FDA-approved drug for treatment of acute myeloid leukemia, is supported by involvement of another AXL inhibitor, bencentininib, in the RECOVERY COVID-19 clinical trial in the United Kingdom. AXL is known to regulate various intracellular signaling pathways (Allen et al., 2002; Hafizi and Dahlbäck 2006), including Ras/ERK, PI3K, and p38 (Allen et al., 2002); AXL inhibition here may contribute to the downregulation of p38 signaling. Apilimod, a PIKFYVE inhibitor, has been described in a recent study to have antiviral capacity (Ou et al., 2020). Here we expand this into a mechanism of regulation by phosphorylation of PIKFYVE upon SARS-CoV-2 infection.

Similar to successful antiretroviral therapy for HIV, a combinatorial drug cocktail may be a viable treatment option for SARS-CoV-2 infection. Specifically, combining remdesivir with the kinase inhibitors identified in this study as well as with translation inhibitors and/or modulators of sigma-1 receptor (Gordon et al., 2020) warrants further testing. Furthermore, pairing genetic and pharmacological perturbations in a systematic fashion could identify new combination therapy approaches and illuminate disease mechanisms.

The unbiased, global phosphoproteomics approaches used here highlight cellular processes hijacked during SARS-CoV-2 infection. To address the need for improved therapeutic strategies to fight COVID-19, we employed a data-driven approach by mapping phosphorylation profiles of dysregulated signaling

pathways to drugs and compounds targeting those signaling pathways. We hope this paradigm can be employed in the future to find additional therapies for COVID-19 and other infectious diseases.

Limitations of Study

A limitation of the current study is the use of a non-human cell line for proteomics analysis upon SARS-CoV-2 infection; here, an African green monkey cell line (Vero E6) was used because it has been shown previously to be highly permissive to SARS-CoV-2 infection (Harcourt et al., 2020). However, pharmacological inhibition of SARS-CoV-2 was assessed in human lung A549-ACE2 cells in addition to Vero E6 cells. The majority of drug effects were found to be replicated between cell lines.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cells
 - Viruses (Institut Pasteur, Paris)
 - Viruses (Mount Sinai, New York)
 - Viruses (University of Freiburg, Germany)
- METHODS DETAILS
 - Vero E6 cell infection for proteomic analysis
 - Cell lysis and digestion
 - Phosphopeptide enrichment
 - Mass spectrometry data acquisition
 - Spectral library generation and raw data processing
 - Immunofluorescence microscopy
 - Electron microscopy
 - N overexpression in Vero E6 cells
 - Cell cycle analysis
 - SARS-CoV-2 infections in ACE2-A549 cells and lysis for cytokine analysis
 - Cytokine RT-qPCR analysis
 - Multiplexed cytokine ELISA
 - SARS-CoV-2 viral quantification via antibody staining in presence of inhibitors
 - Cytotoxicity cell viability assays
 - siRNA-based knockdown of host kinases
 - SARS-CoV-2 viral quantification via RT-qPCR in presence of inhibitors
 - SARS-CoV-2 titration by plaque assay
 - siRNA Cell viability assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Mass spectrometry data pre-processing
 - Identifying significantly regulated proteins
 - Mapping *C. sabaeus* and *H. sapiens* proteins
 - GO enrichment analysis
 - Clustering of phosphorylation changes

- Estimation of kinase activities in the time-course experiment
- Comparison of phosphorylation profiles with other conditions
- Phosphorylation of protein complexes
- Transcription factor activity after SARS-CoV-2 infection
- Pharmacological profiling dose response analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.06.034>.

ACKNOWLEDGMENTS

This research was funded by grants from the National Institutes of Health (P50AI150476, U19AI135990, U19AI135972, R01AI143292, R01AI120694, P01A1063302, and R01AI122747 to N.J.K.; 1R01CA221969 and 1R01CA244550 to K.M.S.; R01GM133981 to D.L.S.; 1F32CA236347-01 to J.E.M.; U19AI118610 to J.R.J.; and F32CA239333 to M.B.), Defense Advance Research Projects Agency HR0011-19-2-0020 (to N.J.K., A.G.S., and K.M.S.); by the Laboratory for Genomics Research (LGR) Excellence in Research Award (ERA) from the Innovative Genomics Institute at UC Berkeley (grant number 133122P); by CRIP (Center for Research for Influenza Pathogenesis), a NIAID-supported Center of Excellence for Influenza Research and Surveillance (CEIRS; contract HHSN272201400008C) (to A.G.S.); by supplements to NIAID grant U19AI135972 and DoD grant W81XWH-19-PRMRP-FPA (to A.G.S.); and by the generous support of the JPB Foundation, the Open Philanthropy Project (research grant 2020-215611 [5384]), and other philanthropic donations (to A.G.S.); by the Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" grant ANR-10-LABX-62-IBEID (to M.V.); by the DFG under Germany's Excellence Strategy (EXC-2189, project ID 390939984 to R.G.); by a Starting Grant Award from the European Research Council (ERC-2014-STG 638884 PhosFunc to P.B.); by the Federal Ministry of Education and Research (BMBF, Computational Life Sciences grant 031L0181B to J.S.R.); by the Intramural Research Program of the NIH, National Institute of Allergy and Infectious Diseases (to E.R.F. and E.D.W.); and by funding from F. Hoffmann-La Roche and Vir Biotechnology and gifts from The Ron Conway Family. K.M.S. is an investigator of the Howard Hughes Medical Institute. The views, opinions, and findings contained in this study are those of the authors and do not represent the official views, policies, or endorsement of the Department of Defense or the U.S. Government. We would like to acknowledge Desiree Ho, Innovative Genomics Institute, for the SARS-CoV-2 virus illustrations. We thank Randy Albert for support with the BSL3 facility and procedures at the Icahn School of Medicine at Mount Sinai, New York.

AUTHOR CONTRIBUTIONS

Conceptualization, M.B., D.L.S., and N.J.K.; Infection Experiments, B.M., V.V.R., B.E.N., L.M., C.K., Q.D.T., A.H., T.V., K.M.W., and E.M.; Proteomics Sample Coordination and Preparation, E.S., M.S., J.M.F., J.Z.G., and J.X.; Proteomics Data Acquisition, A.L.R. and D.L.S.; Data Analysis, M.C.M., B.J.P., D.M., C.H.-A., A. Dunham, M. Modak, D.Q., Y.Z., J.R.J., D.L.S., P.B., J.K.L., M.G., M.B., and C.J.P.M.; Figure Generation, M.C.M., B.J.P., D.M., C.H.-A., A. Dunham, M. Modak, D.Q., J.R.J., D.L.S., P.B., and M.B.; Cell Cycle Experiments, R.R. and B.M.; Infection Imaging, S.W., J.K., S.U., G.K., and R.G.; Manuscript Preparation, M.B., D.E.G., K.O., J.R.J., D.L.S., P.B., N.J.K., and R.D.M.; Literature Review, by R.M.K., M. Modak, J.B., A.R., T.P., Q.L., R.H., M.C., M. Muralidharan, M.K., G.J., B.T., J.H., D.L.S., and M.B.; RNA-seq Analysis, A. Dugourd, A.V., and J.S.-R.; Drug Curation, Y.S., J.E.M., K.M.S., A.R.L., E.J.M., E.F., and S.B.; Interactive Map, T.M., M.C.O., Y.C., J.C.J.C., D.J.B., S.K., M.B., and R.M.K.; Electron Microscopy, E.R.F. and E.D.W.; Work Supervision, T.K., J.K.L., A.G., B.S., M.O., J.S.-R., G.K., R.G., B.R.t., K.M.S., J.R.J., D.L.S., A.G.-S., M.V., P.B., and N.J.K.

DECLARATION OF INTERESTS

The Krogan laboratory has received research support from Vir Biotechnology and F. Hoffmann-La Roche. K.M.S. has consulting agreements for the following companies involving cash and/or stock compensation: Black Diamond Therapeutics, BridGene Biosciences, Denali Therapeutics, Dice Molecules, eFFECTOR Therapeutics (zotatifin and tomivosertib), Erasca, Genentech/Roche, Janssen Pharmaceuticals, Kumquat Biosciences, Kura Oncology, Merck, Mitokinin, Petra Pharma, Qulab Inc. Revolution Medicines (WDB002), Type6 Therapeutics, Venthera, and Wellspring Biosciences (Araxes Pharma).

Received: May 20, 2020
Revised: June 9, 2020
Accepted: June 23, 2020
Published: August 6, 2020

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Abraham et al. (2000); Alfonso et al. (2011); Ali et al. (2000); Amin et al. (2014); Aoh et al. (2009); Araujo et al. (1993); Archambault and Carmena (2012); Aubrey et al. (2016); Auld et al. (2005); Bai et al. (1996); Bakircioglu et al. (2011); Bandau et al. (2012); Barkley et al. (2012); Bartrons et al. (2018); Bassemann et al. (2005); Bayliss et al. (2012); Beeson (2013); Beli et al. (2012); Ben Djoudi Ouadda et al. (2018); Beurel et al. (2015); Biaoxue et al. (2016); Bieling et al. (2008); Bischoff et al. (1990); Bitko et al. (2008); Blethrow et al. (2008); Böck et al. (2010); Boeing et al. (2016); Bradshaw et al. (2013); Brattsand et al. (1994); Bromberg and Damell (2000); Bruinsma et al. (2014); Byon et al. (1998); Cahill et al. (2016a); Cahill et al. (2016b); Cai et al. (2006); Cai et al. (2007); Casado et al. (2014); Chang et al. (2004); Chen et al. (1999); Chen et al. (2002); Chen et al. (2004); Chen et al. (2009); Chen et al. (2013); Chen et al. (2017); Cheng et al. (1998); Cheng et al. (2000); Chiu et al. (2012); Cho et al. (2019); Chou et al. (1990); Chu et al. (2012); Corkery et al. (2015); Costa-Mattioli et al. (2004); Courcelles et al. (2013); Cox et al. (2003); Das et al. (2018); Das et al. (2019); Day et al. (2010); de Haan and Rodtger (2005); Decker and Kovarik (2000); DeDiego et al. (2011); DeDiego et al. (2014); Degrise et al. (2018); Deora et al. (2004); Dephore et al. (2008); Diaz-Moreno et al. (2009); Dou et al. (2015); Eguchi et al. (2009); Elia et al. (2015); Ellis et al. (2019); Eriksson et al. (2004); Escobar et al. (2015); Eto et al. (2012); Eulalio et al. (2007); Fallahi et al. (2016); Fan et al. (1997); Fan et al. (2018); Fang et al. (2000); Farquar et al. (2008); Feldman et al. (1997); Fleming et al. (2000); Folker et al. (2005); Foltz et al. (2009); Formstecher et al. (2001); Franchin et al. (2015); Frémin et al. (2016); Fujimoto et al. (2017); Gagnon and Delpire (2010); Gao and Luo (2006); Garcia-Moreno et al. (2018); Geuens et al. (2007); Gherzi et al. (2004); Goel et al. (2015); Goldstein et al. (2000); Grzenda et al. (2009); Guo et al. (2017); Hajjar (1991); Hajjar et al. (1994); Halder and Johnson (2011); Handschick et al. (2014); Harder et al. (2014); He et al. (2010); He et al. (2018); Heald and McKeon (1990); Heaton et al. (2016); Hedglin and Benkovic (2015); Hein and Nilsson (2016); Hensley and Kursula (2016); Holz et al. (2005); Hou et al. (2011); Hu et al. (2010); Hu et al. (2013); Hu et al. (2015); Huang and Lai (2001); Huang et al. (2006); Huang et al. (2014); Hwang et al. (2014); Ichimura et al. (2005); Inoki et al. (2002); Inoue et al. (2011); Intine et al. (2003); Ishii et al. (2013); Ishikawa et al. (2015); Izawa and Inagaki (2006); Jacque and Stevenson (2006); Jang et al. (2015); Jiang et al. (2014); Johnson et al. (2011); Kane et al. (2002); Kanzawa et al. (2006); Karin (1995); Kastenhuber and Lowe (2017); Katayama et al. (2004); Kennedy et al. (2017); Khurana and Bhattacharyya (2015); Killoran et al. (2015); Kim et al. (2006); Kindrachuk et al. (2015); Kishimoto et al. (2003); Kjolby et al. (2010); Koff et al. (1992); Kohl et al. (2019); Koseki et al. (1999); Kostenko et al. (2009); Kotula et al. (2013); Kovacina et al. (2003); Kraft et al. (2003); Kramer et al. (1998); Kremer et al. (2007); Kuang et al. (2016); Kumar et al. (2014); Kumar et al. (2018); Kunapuli et al. (2006); Labit et al. (2012); Lamarche-Vane and Hall (1998); Larance et al. (2010); Lau et al. (2006); Law et al. (2007); Le Sage et al. (2016); Lee and Myung (2008); Lee et al. (2011); Lee et al. (2012); Lee et al. (2013); Leite and Way (2015); Lekmine et al. (2003); Lekmine et al. (2004); Leukel and Jost (1995); Levy and Lee (2002); Li and Martinez (2011);

- Li and Nagy (2011); Li et al. (2007a); Li et al. (2007b); Li et al. (2008); Li et al. (2016); Li et al. (2017); Liang et al. (2016); Liao et al. (2005); Lieberman et al. (2017); Lim et al. (2013); Lin et al. (2009); Lin et al. (2013); Lingel et al. (2017); Lisnock et al. (2000); Liu and Schneider (2013); Liu et al. (2011); Liu et al. (2013); Liu et al. (2015); Lobry et al. (2007); Lokireddy et al. (2015); Long et al. (2014); Lu et al. (2014); Lu et al. (2017); Lubas et al. (2011); Luciani et al. (2000); Lukas et al. (2011); Luo et al. (2005); Ly et al. (2014); Ly et al. (2017); Machado-Neto et al. (2014); Machowska et al. (2015); Madureira et al. (2012); Magnuson et al. (2012); Magron et al. (2015); Mahony et al. (2017); Mancini et al. (2017); Mandl et al. (2005); Manning and Toker (2017); Marcos-Villar et al. (2016); Marklund et al. (1993); Marklund et al. (1996); Marquette et al. (2011); Matsuhashi et al. (2019); Matsumoto et al. (2015); Matsuo et al. (2003); McCamphill et al. (2015); McClain et al. (2002); Mehellou et al. (2018); Meier and Blobel (1990); Meier and Blobel (1992); Meier and Blobel (1992); Mendoza et al. (2011); Menon et al. (2014); Mizutani et al. (2004); Mo et al. (2012); Moldave (1985); Morris et al. (2017); Muda et al. (1996); Mueller et al. (2007); Müller et al. (2014); Na et al. (2015); Naim et al. (2001); Nandi et al. (2007); Needham et al. (2019); Nielsen et al. (2001); Ogiso et al. (2004); Ohtsubo et al. (1995); Old et al. (2009); Olsen et al. (2006); Olsen et al. (2010); Oshiro et al. (2007); Palamarchuk et al. (2005); Pan (2010); Paquette et al. (2018); Park et al. (2000); Parpys et al. (2015); Patel et al. (2015); Penzo et al. (2019); Pfeiffer et al. (2011); Pires et al. (2017); Ponnamm et al. (2019); Popov (2012); Poppe et al. (2017); Potter et al. (2002); Potthoff and Olson (2007); Preisinger et al. (2004); Proteau et al. (2005); Qiao et al. (2016); Qiu et al. (2016); Radic (2016); Radtke et al. (2006); Radulescu and Cleveland (2010); Rahim et al. (2018); Rai et al. (2017); Ramm et al. (2006); Raught et al. (2004); Ravichandran et al. (2001); Rebeaud et al. (2008); Renganathan et al. (2005); Revenko et al. (2010); Riad et al. (2018); Robinson et al. (2018); Robinson-White and Stratakis (2002); Robitaille et al. (2013); Rogalla et al. (1999); Rogers et al. (2015); Roy et al. (2014); Roy et al. (2017); Ryan et al. (2016); Salvi et al. (2009); Sanborn et al. (2011); Sánchez-Martín and Komatsu (2018); Santamaria et al. (2003); Saredi et al. (1997); Schmucker and Sumara (2014); Schweppe et al. (2013); Sengupta et al. (2013); Shah et al. (2018); Shahbazian et al. (2006); Shandilya et al. (2014); Sharma et al. (2011); Sharma et al. (2014); Shi and Sun (2015); Shi et al. (2008); Shih et al. (2012); Sinclair et al. (2006); Smith and Enquist (2002); Snyder et al. (2004); Stoilov et al. (2004); Stokoe et al. (1992); Su et al. (2001); Tacke et al. (1998); Tafforeau et al. (2011); Tan et al. (2004); Tan et al. (2012); Tanenbaum et al. (2006); Tantos et al. (2013); Tarakhovskiy and Prinjha (2018); Tavares et al. (2015); Tee et al. (2002); Thaker et al. (2019); Thejir et al. (2020); Tiedje et al. (2015); Tokuyama et al. (2001); Toughiri et al. (2013); Trencia et al. (2003); Trigg and Ferguson (2015); Tsai and Seto (2002); Tsai et al. (1993); Tsoi et al. (2014); Tsvetkov and Stern (2005); Ui et al. (2015); Ujike and Taguchi (2015); Valvezan and Klein (2012); Vind et al. (2011); Vishwanatha and Kumble (1993); Wada et al. (2001); Walter et al. (2002); Walter et al. (2016); Wang and Prives (1995); Wang and Zhang (1999); Wang et al. (2000); Wang et al. (2007); Wang et al. (2008); Wang et al. (2013); Wang et al. (2020); Ward et al. (2004); Wen et al. (1995); Wertz et al. (2015); Willems et al. (1996); Williams et al. (1997); Wohlbold et al. (2012); Wolin and Cedervall (2002); Wortzel and Seger (2011); Wu et al. (2014); Wu et al. (2018); Wurm et al. (2001); Xu et al. (2011); Yang et al. (2012a); Yang et al. (2012b); Yang et al. (2014); Yao et al. (2017); Ye et al. (2008); Yim et al. (2013); Yoo et al. (2019); Yu et al. (2009); Yu et al. (2011); Yu et al. (2012); Yui et al. (2001); Zhang et al. (1998); Zhang et al. (2009); Zhang et al. (2015); Zhang et al. (2019a); Zhang et al. (2019b); Zhao et al. (2015); Zhou et al. (2016); Zhu et al. (2001); Zhu et al. (2017); Zou et al. (2007); Zuo and Manley (1993).
- Ali, N., Pruijn, G.J., Kenan, D.J., Keene, J.D., and Siddiqui, A. (2000). Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation. *J. Biol. Chem.* 275, 27531–27540.
- Allen, M.P., Linseman, D.A., Udo, H., Xu, M., Schaack, J.B., Varnum, B., Kandel, E.R., Heidenreich, K.A., and Wierman, M.E. (2002). Novel mechanism for gonadotropin-releasing hormone neuronal migration involving Gas6/Ark signaling to p38 mitogen-activated protein kinase. *Mol. Cell. Biol.* 22, 599–613.
- Alvarez, D.E., and Agaisse, H. (2012). Casein kinase 2 regulates vaccinia virus actin tail formation. *Virology* 423, 143–151.
- Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 48, 838–847.
- Amin, M.A., Itoh, G., Iemura, K., Ikeda, M., and Tanaka, K. (2014). CLIP-170 recruits PLK1 to kinetochores during early mitosis for chromosome alignment. *J. Cell Sci.* 127, 2818–2824.
- Aoh, Q.L., Castle, A.M., Hubbard, C.H., Katsumata, O., and Castle, J.D. (2009). SCAMP3 negatively regulates epidermal growth factor receptor degradation and promotes receptor recycling. *Mol. Biol. Cell* 20, 1816–1832.
- Araujo, H., Danziger, N., Cordier, J., Glowinski, J., and Chneiweiss, H. (1993). Characterization of PEA-15, a major substrate for protein kinase C in astrocytes. *J. Biol. Chem.* 268, 5911–5920.
- Archambault, V., and Carmena, M. (2012). Polo-like kinase-activating kinases: Aurora A, Aurora B and what else? *Cell Cycle* 11, 1490–1495.
- Auriumi, Y., Kuroki, M., Dansako, H., Abe, K., Ikeda, M., Wakita, T., and Kato, N. (2008). The DNA damage sensors ataxia-telangiectasia mutated kinase and checkpoint kinase 2 are required for hepatitis C virus RNA replication. *J. Virol.* 82, 9639–9646.
- Aubrey, B.J., Strasser, A., and Kelly, G.L. (2016). Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harb. Perspect. Med.* 6, a026062.
- Auld, G.C., Campbell, D.G., Morrice, N., and Cohen, P. (2005). Identification of calcium-regulated heat-stable protein of 24 kDa (CRHSP24) as a physiological substrate for PKB and RSK using KESTREL. *Biochem. J.* 389, 775–783.
- Bachman, J.A., Györi, B.M., and Sorger, P.K. (2019). Assembling a phosphoproteomic knowledge base using ProtMapper to normalize phosphosite information from databases and text mining. *bioRxiv*. <https://doi.org/10.1101/822668>.
- Bai, C., Sen, P., Hofmann, K., Ma, L., Goebel, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* 86, 263–274.
- Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.-Y., Chen, L., and Wang, M. (2020). Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* 323, 1406–1407.
- Bakircioglu, M., Carvalho, O.P., Khurshid, M., Cox, J.J., Tuysuz, B., Barak, T., Yilmaz, S., Caglayan, O., Dincer, A., Nicholas, A.K., et al. (2011). The essential role of centrosomal NDE1 in human cerebral cortex neurogenesis. *Am. J. Hum. Genet.* 88, 523–535.
- Bandau, S., Knebel, A., Gage, Z.O., Wood, N.T., and Alexandru, G. (2012). UBXN7 docks on neddylated cullin complexes using its UIM motif and causes HIF1 α accumulation. *BMC Biol.* 10, 36.
- Barkley, L.R., Palle, K., Durando, M., Day, T.A., Gurkar, A., Kakusho, N., Li, J., Masai, H., and Vaziri, C. (2012). c-Jun N-terminal kinase-mediated Rad18 phosphorylation facilitates Pol η recruitment to stalled replication forks. *Mol. Biol. Cell* 23, 1943–1954.
- Bartrons, R., Simon-Molas, H., Rodríguez-García, A., Castaño, E., Navarro-Sabaté, Á., Manzano, A., and Martínez-Outschoom, U.E. (2018). Fructose 2,6-Bisphosphate in Cancer Cell Metabolism. *Front. Oncol.* 8, 331.
- Bassermann, F., von Klitzing, C., Münch, S., Bai, R.-Y., Kawaguchi, H., Morris, S.W., Peschel, C., and Duyster, J. (2005). NIPA defines an SCF-type mammalian E3 ligase that regulates mitotic entry. *Cell* 122, 45–57.
- Bayliss, R., Fry, A., Haq, T., and Yeoh, S. (2012). On the molecular mechanisms of mitotic kinase activation. *Open Biol.* 2, 120136.

REFERENCES

Abraham, J., Kelly, J., Thibault, P., and Benchimol, S. (2000). Post-translational modification of p53 protein in response to ionizing radiation analyzed by mass spectrometry. *J. Mol. Biol.* 295, 853–864.

Alfonso, R., Lutz, T., Rodriguez, A., Chavez, J.P., Rodriguez, P., Gutierrez, S., and Nieto, A. (2011). CHD6 chromatin remodeler is a negative modulator of influenza virus replication that relocates to inactive chromatin upon infection. *Cell. Microbiol.* 13, 1894–1906.

- Beeson, D. (2013). Hereditary and Autoimmune Myasthenias. In Emery and Rimoin's Principles and Practice of Medical Genetics, D. Rimoin, R. Peryitz, and B. Korf, eds. (Elsevier), pp. 1–17.
- Beli, P., Lukashchuk, N., Wagner, S.A., Weinert, B.T., Olsen, J.V., Baskcomb, L., Mann, M., Jackson, S.P., and Choudhary, C. (2012). Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol. Cell* **46**, 212–225.
- Ben Djoudi Ouadda, A., He, Y., Calabrese, V., Ishii, H., Chidiac, R., Gratton, J.-P., Roux, P.P., and Lamarche-Vane, N. (2018). CdGAP/ARHGAP31 is regulated by RSK phosphorylation and binding to 14-3-3 β adaptor protein. *Oncotarget* **9**, 11646–11664.
- Beurel, E., Grieco, S.F., and Jope, R.S. (2015). Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. *Pharmacol. Ther.* **148**, 114–131.
- Biaoxue, R., Hua, L., Wenlong, G., and Shuanying, Y. (2016). Overexpression of stathmin promotes metastasis and growth of malignant solid tumors: a systemic review and meta-analysis. *Oncotarget* **7**, 78994–79007.
- Bieling, P., Kandels-Lewis, S., Telley, I.A., van Dijk, J., Janke, C., and Surrey, T. (2008). CLIP-170 tracks growing microtubule ends by dynamically recognizing composite EB1/tubulin-binding sites. *J. Cell Biol.* **183**, 1223–1233.
- Bischoff, J.R., Friedman, P.N., Marshak, D.R., Prives, C., and Beach, D. (1990). Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2. *Proc. Natl. Acad. Sci. USA* **87**, 4766–4770.
- Bitko, V., Musiyenko, A., Bayfield, M.A., Maraja, R.J., and Barik, S. (2008). Cellular La protein shields nonsegmented negative-strand RNA viral leader RNA from RIG-I and enhances virus growth by diverse mechanisms. *J. Virol.* **82**, 7977–7987.
- Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Uhl, S., Hoagland, D., Moller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D., et al. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036–1045.e9.
- Blethrow, J.D., Glavy, J.S., Morgan, D.O., and Shokat, K.M. (2008). Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc. Natl. Acad. Sci. USA* **105**, 1442–1447.
- Böck, B.C., Tagscherer, K.E., Fassl, A., Krämer, A., Oehme, I., Zentgraf, H.-W., Keith, M., and Roth, W. (2010). The PEA-15 protein regulates autophagy via activation of JNK. *J. Biol. Chem.* **285**, 21644–21654.
- Boeing, S., Williamson, L., Encheva, V., Gori, I., Saunders, R.E., Instrell, R., Aygün, O., Rodríguez-Martínez, M., Weems, J.C., Kelly, G.P., et al. (2016). Multicentric Analysis of the UV-Induced DNA Damage Response. *Cell Rep.* **15**, 1597–1610.
- Bradshaw, N.J., Hennah, W., and Soares, D.C. (2013). NDE1 and NDEL1: twin neurodevelopmental proteins with similar 'nature' but different 'nurture'. *Biol. Concepts* **4**, 447–464.
- Brattsand, G., Marklund, U., Nylander, K., Roos, G., and Gullberg, M. (1994). Cell-cycle-regulated phosphorylation of oncoprotein 18 on Ser16, Ser25 and Ser38. *Eur. J. Biochem.* **220**, 359–368.
- Bromberg, J., and Darnell, J.E., Jr. (2000). The role of STATs in transcriptional control and their impact on cellular function. *Oncogene* **19**, 2468–2473.
- Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., et al. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410.
- Bruinsma, W., Macurek, L., Freire, R., Lindqvist, A., and Medema, R.H. (2014). Bora and Aurora-A continue to activate Plk1 in mitosis. *J. Cell Sci.* **127**, 801–811.
- Byon, J.C., Kusari, A.B., and Kusari, J. (1998). Protein-tyrosine phosphatase-1B acts as a negative regulator of insulin signal transduction. *Mol. Cell. Biochem.* **182**, 101–108.
- Cahill, M.A., Jazayeri, J.A., Kovacevic, Z., and Richardson, D.R. (2016a). PGRMC1 regulation by phosphorylation: potential new insights in controlling biological activity. *Oncotarget* **7**, 50822–50827.
- Cahill, M.A., Jazayeri, J.A., Catalano, S.M., Toyokuni, S., Kovacevic, Z., and Richardson, D.R. (2016b). The emerging role of progesterone receptor membrane component 1 (PGRMC1) in cancer biology. *Biochim. Biophys. Acta* **1866**, 339–349.
- Cai, S.-L., Tee, A.R., Short, J.D., Bergeron, J.M., Kim, J., Shen, J., Guo, R., Johnson, C.L., Kiguchi, K., and Walker, C.L. (2006). Activity of TSC2 is inhibited by AKT-mediated phosphorylation and membrane partitioning. *J. Cell Biol.* **173**, 279–289.
- Cai, Y., Liu, Y., and Zhang, X. (2007). Suppression of coronavirus replication by inhibition of the MEK signaling pathway. *J. Virol.* **81**, 446–456.
- Casado, P., Rodríguez-Prados, J.-C., Cosulich, S.C., Guichard, S., Vanhaesebroeck, B., Joel, S., and Cutillas, P.R. (2013). Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* **6**, rs6.
- Casado, P., Bilanges, B., Rajeeve, V., Vanhaesebroeck, B., and Cutillas, P.R. (2014). Environmental stress affects the activity of metabolic and growth factor signaling networks and induces autophagy markers in MCF7 breast cancer cells. *Mol. Cell. Proteomics* **13**, 836–848.
- Chang, Y.-J., Liu, C.Y., Chiang, B.L., Chao, Y.C., and Chen, C.C. (2004). Induction of IL-8 release in lung cells via activator protein-1 by recombinant baculovirus displaying severe acute respiratory syndrome-coronavirus spike proteins: identification of two functional regions. *J. Immunol.* **173**, 7602–7614.
- Chaurushiya, M.S., and Weitzman, M.D. (2009). Viral manipulation of DNA repair and cell cycle checkpoints. *DNA Repair (Amst.)* **8**, 1166–1176.
- Chen, H.K., Pai, C.Y., Huang, J.Y., and Yeh, N.H. (1999). Human Nopp140, which interacts with RNA polymerase I: implications for rRNA gene transcription and nucleolar structural organization. *Mol. Cell. Biol.* **19**, 8536–8546.
- Chen, H., Wurm, T., Britton, P., Brooks, G., and Hiscox, J.A. (2002). Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. *J. Virol.* **76**, 5233–5250.
- Chen, W., Yazicioglu, M., and Cobb, M.H. (2004). Characterization of OSR1, a member of the mammalian Ste20p/germinal center kinase subfamily. *J. Biol. Chem.* **279**, 11129–11136.
- Chen, C.-B., Ng, J.K.W., Choo, P.-H., Wu, W., and Porter, A.G. (2009). Mammalian sterile 20-like kinase 3 (MST3) mediates oxidative-stress-induced cell death by modulating JNK activation. *Biosci. Rep.* **29**, 405–415.
- Chen, J.-T., Ho, C.-W., Chi, L.-M., Chien, K.-Y., Hsieh, Y.-J., Lin, S.-J., and Yu, J.-S. (2013). Identification of the lamin A/C phosphopeptide recognized by the antibody P-STM in mitotic HeLa S3 cells. *BMC Biochem.* **14**, 18.
- Chen, J.-Y., Lin, C.-H., and Chen, B.-C. (2017). Hypoxia-induced ADAM 17 expression is mediated by RSK1-dependent C/EBP β activation in human lung fibroblasts. *Mol. Immunol.* **88**, 155–163.
- Cheng, X., Ma, Y., Moore, M., Hemmings, B.A., and Taylor, S.S. (1998). Phosphorylation and activation of cAMP-dependent protein kinase by phosphoinositide-dependent protein kinase. *Proc. Natl. Acad. Sci. USA* **95**, 9849–9854.
- Cheng, J., Yang, J., Xia, Y., Karin, M., and Su, B. (2000). Synergistic interaction of MEK kinase 2, c-Jun N-terminal kinase (JNK) kinase 2, and JNK1 results in efficient and specific JNK1 activation. *Mol. Cell. Biol.* **20**, 2334–2342.
- Chiu, M., Tardito, S., Barilli, A., Bianchi, M.G., Dall'Asta, V., and Bussolati, O. (2012). Glutamine stimulates mTORC1 independent of the cell content of essential amino acids. *Amino Acids* **43**, 2561–2567.
- Cho, H., Lee, W., Kim, G.-W., Lee, S.-H., Moon, J.-S., Kim, M., Kim, H.S., and Oh, J.-W. (2019). Regulation of La/SSB-dependent viral gene expression by pre-tRNA 3' trailer-derived tRNA fragments. *Nucleic Acids Res.* **47**, 9888–9901.
- Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526.
- Chou, Y.H., Bischoff, J.R., Beach, D., and Goldman, R.D. (1990). Intermediate filament reorganization during mitosis is mediated by p34cdc2 phosphorylation of vimentin. *Cell* **62**, 1063–1071.

- Chu, J., Loughlin, E.A., Gaur, N.A., SenBanerjee, S., Jacob, V., Monson, C., Kent, B., Oranu, A., Ding, Y., Ukumadu, C., and Sadler, K.C. (2012). UHRF1 phosphorylation by cyclin A2/cyclin-dependent kinase 2 is required for zebrafish embryogenesis. *Mol. Biol. Cell* **23**, 59–70.
- Chu, D.K.W., Pan, Y., Cheng, S.M.S., Hui, K.P.Y., Krishnan, P., Liu, Y., Ng, D.Y.M., Wan, C.K.C., Yang, P., Wang, Q., et al. (2020). Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. *Clin. Chem.* **66**, 549–555.
- Corkery, D.P., Holly, A.C., Lahsae, S., and Dellaire, G. (2015). Connecting the speckles: Splicing kinases and their role in tumorigenesis and treatment response. *Nucleus* **6**, 279–288.
- Costa-Mattioli, M., Svitkin, Y., and Sonenberg, N. (2004). La autoantigen is necessary for optimal function of the poliovirus and hepatitis C virus internal ribosome entry site in vivo and in vitro. *Mol. Cell. Biol.* **24**, 6861–6870.
- Courcelles, M., Frémin, C., Voisin, L., Lemieux, S., Meloche, S., and Thibault, P. (2013). Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions. *Mol. Syst. Biol.* **9**, 669.
- Cox, D.M., Du, M., Marback, M., Yang, E.C.C., Chan, J., Siu, K.W.M., and McDermott, J.C. (2003). Phosphorylation motifs regulating the stability and function of myocyte enhancer factor 2A. *J. Biol. Chem.* **278**, 15297–15303.
- Cuadrado, A., and Nebreda, A.R. (2010). Mechanisms and functions of p38 MAPK signalling. *Biochem. J.* **429**, 403–417.
- D'Amore, C., Salizzato, V., Borgo, C., Cesaro, L., Pinna, L.A., and Salvi, M. (2019). A Journey through the Cytoskeleton with Protein Kinase CK2. *Curr. Protein Pept. Sci.* **20**, 547–562.
- Das, T., Chen, Z., Hendriks, R.W., and Kool, M. (2018). A20/Tumor Necrosis Factor α -Induced Protein 3 in Immune Cells Controls Development of Autoinflammation and Autoimmunity: Lessons from Mouse Models. *Front. Immunol.* **9**, 104.
- Das, T., Kim, E.E., and Song, E.J. (2019). Phosphorylation of USP15 and USP4 Regulates Localization and Spliceosomal Deubiquitination. *J. Mol. Biol.* **437**, 3900–3912.
- Davidson, A.D., Williamson, M.K., Lewis, S., Shoemark, D., Carroll, M.W., Heesom, K., Zamboni, M., Ellis, J., Lewis, P.A., Hiscox, J.A., et al. (2020). Characterisation of the transcriptome and proteome of SARS-CoV-2 using direct RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. *bioRxiv*. <https://doi.org/10.1101/2020.03.22.002204>.
- Day, T.A., Palle, K., Barkley, L.R., Kakusho, N., Zou, Y., Tateishi, S., Verreault, A., Masai, H., and Vaziri, C. (2010). Phosphorylated Rad18 directs DNA polymerase η to sites of stalled replication. *J. Cell Biol.* **171**, 953–966.
- de Haan, C.A.M., and Rottier, P.J.M. (2005). Molecular interactions in the assembly of coronaviruses. *Adv. Virus Res.* **64**, 165–230.
- Decker, T., and Kovarik, P. (2000). Serine phosphorylation of STATs. *Oncogene* **19**, 2628–2637.
- DeDiego, M.L., Nieto-Torres, J.L., Jiménez-Guardeño, J.M., Regla-Nava, J.A., Álvarez, E., Oliveros, J.C., Zhao, J., Fett, C., Perlman, S., and Enjuanes, L. (2011). Severe acute respiratory syndrome coronavirus envelope protein regulates cell stress response and apoptosis. *PLoS Pathog.* **7**, e1002315.
- DeDiego, M.L., Nieto-Torres, J.L., Regla-Nava, J.A., Jiménez-Guardeño, J.M., Fernández-Delgado, R., Fett, C., Castaño-Rodríguez, C., Perlman, S., and Enjuanes, L. (2014). Inhibition of NF- κ B-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. *J. Virol.* **88**, 913–924.
- Degrise, S., de Bock, C.E., Demeyer, S., Govaerts, I., Borna-Schein, S., Verbeke, D., Jacobs, K., Binos, S., Skerrett-Byrne, D.A., Murray, H.C., et al. (2018). Mutant JAK3 phosphoproteomic profiling predicts synergism between JAK3 inhibitors and MEK/BCL2 inhibitors for the treatment of T-cell acute lymphoblastic leukemia. *Leukemia* **32**, 788–800.
- Deora, A.B., Kreitzer, G., Jacovina, A.T., and Hajjar, K.A. (2004). An annexin 2 phosphorylation switch mediates p11-dependent translocation of annexin 2 to the cell surface. *J. Biol. Chem.* **279**, 43411–43418.
- Dephore, N., Zhou, C., Villén, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. USA* **105**, 10762–10767.
- Díaz-Moreno, I., Hollingworth, D., Frenkiel, T.A., Kelly, G., Martin, S., Howell, S., Garcia-Mayoral, M., Gherzi, R., Briata, P., and Ramos, A. (2009). Phosphorylation-mediated unfolding of a KH domain regulates KSRP localization via 14-3-3 binding. *Nat. Struct. Mol. Biol.* **16**, 238–246.
- Dou, X., Wei, J., Sun, A., Shao, G., Childress, C., Yang, W., and Lin, Q. (2015). PBK/TOPK mediates geranylgeranylation signaling for breast cancer cell proliferation. *Cancer Cell Int.* **15**, 27.
- Dove, B., Brooks, G., Bicknell, K., Wurm, T., and Hiscox, J.A. (2006). Cell cycle perturbations induced by infection with the coronavirus infectious bronchitis virus and their effect on virus replication. *J. Virol.* **80**, 4147–4156.
- Eguchi, S., Oshiro, N., Miyamoto, T., Yoshino, K., Okamoto, S., Ono, T., Kikawa, U., and Yonezawa, K. (2009). AMP-activated protein kinase phosphorylates glutamine : fructose-6-phosphate amidotransferase 1 at Ser243 to modulate its enzymatic activity. *Genes Cells* **14**, 179–189.
- Elia, A.E.H., Boardman, A.P., Wang, D.C., Huttlin, E.L., Everley, R.A., Dephore, N., Zhou, C., Koren, I., Gygi, S.P., and Elledge, S.J. (2015). Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the DNA Damage Response. *Mol. Cell* **59**, 867–881.
- Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214.
- Ellis, R., Katerelos, M., Choy, S.W., Cook, N., Lee, M., Paizis, K., Pell, G., Walker, S., Power, D.A., and Mount, P.F. (2019). Increased expression and phosphorylation of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase isoforms in urinary exosomes in pre-eclampsia. *J. Transl. Med.* **17**, 60.
- Eriksson, J.E., He, T., Trejo-Skalli, A.V., Härmälä-Braskén, A.-S., Hellman, J., Chou, Y.-H., and Goldman, R.D. (2004). Specific in vivo phosphorylation sites determine the assembly dynamics of vimentin intermediate filaments. *J. Cell Sci.* **117**, 919–932.
- Escobar, D.J., Desai, R., Ishiyama, N., Folmsbee, S.S., Novak, M.N., Flozak, A.S., Daugherty, R.L., Mo, R., Nanavati, D., Sarpal, R., et al. (2015). α -Catenin phosphorylation promotes intercellular adhesion through a dual-kinase mechanism. *J. Cell Sci.* **128**, 1150–1165.
- Eto, K., Goto, S., Nakashima, W., Ura, Y., and Abe, S.-I. (2012). Loss of programmed cell death 4 induces apoptosis by promoting the translation of procaspase-3 mRNA. *Cell Death Differ.* **19**, 573–581.
- Eulalio, A., Behm-Ansmant, I., and Izaurralde, E. (2007). P bodies: at the crossroads of post-transcriptional pathways. *Nat. Rev. Mol. Cell Biol.* **8**, 9–22.
- Fallahi, E., O'Driscoll, N.A., and Matallanas, D. (2016). The MST1/Hippo Pathway and Cell Death: A Non-Canonical Affair. *Genes (Basel)* **7**, 28.
- Fan, H., Sakulich, A.L., Goodier, J.L., Zhang, X., Qin, J., and Maraja, R.J. (1997). Phosphorylation of the human La antigen on serine 366 can regulate recycling of RNA polymerase III transcription complexes. *Cell* **88**, 707–715.
- Fan, Y., Sanyal, S., and Bruzzone, R. (2018). Breaking Bad: How Viruses Subvert the Cell Cycle. *Front. Cell. Infect. Microbiol.* **8**, 396.
- Fang, X., Yu, S.X., Lu, Y., Bast, R.C., Jr., Woodgett, J.R., and Mills, G.B. (2000). Phosphorylation and inactivation of glycogen synthase kinase 3 by protein kinase A. *Proc. Natl. Acad. Sci. USA* **97**, 11960–11965.
- Farquhar, M.J., Harris, H.J., Diskar, M., Jones, S., Mee, C.J., Nielsen, S.U., Brimacombe, C.L., Molina, S., Toms, G.L., Maurel, P., et al. (2008). Protein kinase A-dependent step(s) in hepatitis C virus entry and infectivity. *J. Virol.* **82**, 8797–8811.
- Fehr, A.R., and Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **1282**, 1–23.
- Feldman, R.M., Correll, C.C., Kaplan, K.B., and Deshaies, R.J. (1997). A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **97**, 221–230.
- Fernández-Golbano, I.M., Idrissi, F.-Z., Giblin, J.P., Grosshans, B.L., Robles, V., Grötsch, H., Borrás, Mdel.M., and Geli, M.I. (2014). Crosstalk between

- PI(4,5)P₂ and CK2 modulates actin polymerization during endocytic uptake. *Dev. Cell* 30, 746–758.
- Fleming, Y., Armstrong, C.G., Morrice, N., Paterson, A., Goedert, M., and Cohen, P. (2000). Synergistic activation of stress-activated protein kinase 1/c-Jun N-terminal kinase (SAPK1/JNK) isoforms by mitogen-activated protein kinase kinase 4 (MKK4) and MKK7. *Biochem. J.* 352, 145–154.
- Folker, E.S., Baker, B.M., and Goodson, H.V. (2005). Interactions between CLIP-170, tubulin, and microtubules: implications for the mechanism of Clip-170 plus-end tracking behavior. *Mol. Biol. Cell* 16, 5373–5384.
- Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell* 137, 472–484.
- Formstecher, E., Ramos, J.W., Fauquet, M., Calderwood, D.A., Hsieh, J.C., Canton, B., Nguyen, X.T., Bamier, J.V., Camonis, J., Ginsberg, M.H., and Chneiweiss, H. (2001). PEA-15 mediates cytoplasmic sequestration of ERK MAP kinase. *Dev. Cell* 1, 239–250.
- Franchin, C., Cesaro, L., Salvi, M., Million, R., Iori, E., Cifani, P., James, P., Arignoni, G., and Pinna, L. (2015). Quantitative analysis of a phosphoproteome readily altered by the protein kinase CK2 inhibitor quinalizarin in HEK-293T cells. *Biochim. Biophys. Acta* 1854, 609–623.
- Frémin, C., Guégan, J.-P., Plutoni, C., Mahaffey, J., Philips, M.R., Emery, G., and Meloche, S. (2016). ERK1/2-induced phosphorylation of R-Ras GTPases stimulates their oncogenic potential. *Oncogene* 35, 5692–5698.
- Fu, Y., Yip, A., Seah, P.G., Blasco, F., Shi, P.-Y., and Hervé, M. (2014). Modulation of inflammation and pathology during dengue virus infection by p38 MAPK inhibitor SB203580. *Antiviral Res.* 110, 151–157.
- Fujimoto, M., Takii, R., Takaki, E., Katiyar, A., Nakato, R., Shirahige, K., and Nakai, A. (2017). The HSF1-PARP13-PARP1 complex facilitates DNA repair and promotes mammary tumorigenesis. *Nat. Commun.* 8, 1638.
- Gagnon, K.B., and Delpire, E. (2010). On the substrate recognition and negative regulation of SPAK, a kinase modulating Na⁺-K⁺-2Cl⁻ cotransport activity. *Am. J. Physiol. Cell Physiol.* 299, C614–C620.
- Gao, G., and Luo, H. (2006). The ubiquitin-proteasome pathway in viral infections. *Can. J. Physiol. Pharmacol.* 84, 5–14.
- García-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 29, 1363–1375.
- García-Moreno, M., Järvelin, A.I., and Castello, A. (2018). Unconventional RNA-binding proteins step into the virus-host battlefield. *Wiley Interdiscip. Rev. RNA* 9, e1498.
- Geuens, T., De Winter, V., Rajan, N., Achsel, T., Mateiu, L., Almeida-Souza, L., Asselbergh, B., Bouhy, D., Auer-Grumbach, M., Bagni, C., and Timmerman, V. (2017). Mutant HSPB1 causes loss of translational repression by binding to PCBP1, an RNA binding protein with a possible role in neurodegenerative disease. *Acta Neuropathol. Commun.* 5, 5.
- Gherzi, R., Lee, K.-Y., Briata, P., Wegmüller, D., Moroni, C., Karin, M., and Chen, C.-Y. (2004). A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. *Mol. Cell* 14, 571–583.
- Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* 47 (D1), D559–D563.
- Goel, P., Manning, J.A., and Kumar, S. (2015). NEDD4-2 (NEDD4L): the ubiquitin ligase for multiple membrane proteins. *Gene* 557, 1–10.
- Goldstein, B.J., Bittner-Kowalczyk, A., White, M.F., and Harbeck, M. (2000). Tyrosine dephosphorylation and deactivation of insulin receptor substrate-1 by protein-tyrosine phosphatase 1B. Possible facilitation by the formation of a ternary complex with the Grb2 adaptor protein. *J. Biol. Chem.* 275, 4283–4289.
- Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O’Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. Published online April 30, 2020. <https://doi.org/10.1038/s41586-020-2286-9>.
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., Feldt, T., Green, G., Green, M.L., Lescure, F.-X., et al. (2020). Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N. Engl. J. Med.* 382, 2327–2336.
- Growcott, E.J., Bamba, D., Galarneau, J.-R., Leonard, V.H.J., Schul, W., Stein, D., and Osborne, C.S. (2018). The effect of P38 MAP kinase inhibition in a mouse model of influenza. *J. Med. Microbiol.* 67, 452–462.
- Grzenda, A., Lomberk, G., Zhang, J.-S., and Urrutia, R. (2009). Sin3: master scaffold and transcriptional corepressor. *Biochim. Biophys. Acta* 1789, 443–450.
- Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., Liu, L., Shan, H., Lei, C.-L., Hui, D.S.C., et al.; China Medical Treatment Expert Group for Covid-19 (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720.
- Guo, X., Huang, X., and Chen, M.J. (2017). Reversible phosphorylation of the 26S proteasome. *Protein Cell* 8, 255–272.
- Hafizi, S., and Dahlbäck, B. (2006). Signalling and functional diversity within the Axl subfamily of receptor tyrosine kinases. *Cytokine Growth Factor Rev.* 17, 295–304.
- Hajjar, K.A. (1991). The endothelial cell tissue plasminogen activator receptor. Specific interaction with plasminogen. *J. Biol. Chem.* 266, 21962–21970.
- Hajjar, K.A., Jacovina, A.T., and Chacko, J. (1994). An endothelial cell receptor for plasminogen/tissue plasminogen activator. I. Identity with annexin II. *J. Biol. Chem.* 269, 21191–21197.
- Halder, G., and Johnson, R.L. (2011). Hippo signaling: growth control and beyond. *Development* 138, 9–22.
- Han, H., Yang, L., Liu, R., Liu, F., Wu, K.-L., Li, J., Liu, X.-H., and Zhu, C.-L. (2020). Prominent changes in blood coagulation of patients with SARS-CoV-2 infection. *Clin. Chem. Lab. Med.* 58, 1116–1120.
- Handschick, K., Beuerlein, K., Jurida, L., Bartkuhn, M., Müller, H., Soelch, J., Weber, A., Dittrich-Breiholz, O., Schneider, H., Scharfe, M., et al. (2014). Cyclin-dependent kinase 6 is a chromatin-bound cofactor for NF- κ B-dependent gene expression. *Mol. Cell* 53, 193–208.
- Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S.K., Murray, J., Queen, K., Tao, Y., Paden, C.R., Zhang, J., et al. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 From Patient With Coronavirus Disease, United States. *Emerg. Infect. Dis.* 26, 1266–1273.
- Harder, L.M., Bunkenborg, J., and Andersen, J.S. (2014). Inducing autophagy: a comparative phosphoproteomic study of the cellular response to ammonia and rapamycin. *Autophagy* 10, 339–355.
- He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N.J., Alber, T., and Zhou, Q. (2010). HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol. Cell* 38, 428–438.
- He, L., Gomes, A.P., Wang, X., Yoon, S.O., Lee, G., Nagiec, M.J., Cho, S., Chavez, A., Islam, T., Yu, Y., et al. (2018). mTORC1 Promotes Metabolic Reprogramming by the Suppression of GSK3-Dependent Foxk1 Phosphorylation. *Mol. Cell* 70, 949–960.e4.
- Heald, R., and McKeon, F. (1990). Mutations of phosphorylation sites in lamin A that prevent nuclear lamina disassembly in mitosis. *Cell* 61, 579–589.
- Heaton, N.S., Moshkina, N., Fenouil, R., Gardner, T.J., Aguirre, S., Shah, P.S., Zhao, N., Manganaro, L., Hultquist, J.F., Noel, J., et al. (2016). Targeting Viral Proteostasis Limits Influenza Virus, HIV, and Dengue Virus Infection. *Immunity* 44, 46–58.
- Hedglin, M., and Benkovic, S.J. (2015). Regulation of Rad6/Rad18 Activity During DNA Damage Tolerance. *Annu. Rev. Biophys.* 44, 207–228.
- Hein, J.B., and Nilsson, J. (2016). Interphase APC/C-Cdc20 inhibition by cyclin A2-Cdk2 ensures efficient mitotic entry. *Nat. Commun.* 7, 10975.
- Hensley, K., and Kursula, P. (2016). Collapsin Response Mediator Protein-2 (CRMP2) is a Plausible Etiological Factor and Potential Therapeutic Target in

- Alzheimer's Disease: Comparison and Contrast with Microtubule-Associated Protein Tau. *J. Alzheimers Dis.* 53, 1–14.
- Heo, L., and Feig, M. (2020). Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. *bioRxiv*. <https://doi.org/10.1101/2020.03.25.008904>.
- Hernandez-Armenta, C., Ochoa, D., Gonçalves, E., Saez-Rodríguez, J., and Beltrao, P. (2017). Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics* 33, 1845–1851.
- Hodel, A.E., Hodel, M.R., Griffis, E.R., Hennig, K.A., Ratner, G.A., Xu, S., and Powers, M.A. (2002). The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol. Cell* 10, 347–358.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271–280.e8.
- Holz, M.K., Ballif, B.A., Gygi, S.P., and Blenis, J. (2005). mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell* 123, 569–580.
- Hong, S., Freeberg, M.A., Han, T., Kamath, A., Yao, Y., Fukuda, T., Suzuki, T., Kim, J.K., and Inoki, K. (2017). LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. *eLife* 6, e25237.
- Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cēsareni, G., Jensen, L.J., and Lindling, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods* 11, 603–604.
- Hou, H., Wang, F., Zhang, W., Wang, D., Li, X., Bartlam, M., Yao, X., and Rao, Z. (2011). Structure-functional analyses of CRHSP-24 plasticity and dynamics in oxidative stress response. *J. Biol. Chem.* 286, 9623–9635.
- Hu, F., Gartenhaus, R.B., Eichberg, D., Liu, Z., Fang, H.-B., and Rapoport, A.P. (2010). PBK/TOPK interacts with the DBD domain of tumor suppressor p53 and modulates expression of transcriptional targets including p21. *Oncogene* 29, 5464–5474.
- Hu, J., Stites, E.C., Yu, H., Germino, E.A., Meharena, H.S., Stork, P.J.S., Kornev, A.P., Taylor, S.S., and Shaw, A.S. (2013). Allosteric activation of functionally asymmetric RAF kinase dimers. *Cell* 154, 1036–1046.
- Hu, C.-W., Hsu, C.-L., Wang, Y.-C., Ishihama, Y., Ku, W.-C., Huang, H.-C., and Juan, H.-F. (2015). Temporal Phosphoproteome Dynamics Induced by an ATP Synthase Inhibitor Citreoviridin. *Mol. Cell. Proteomics* 14, 3284–3298.
- Huang, P., and Lai, M.M. (2001). Heterogeneous nuclear ribonucleoprotein A1 binds to the 3'-untranslated region and mediates potential 5'-3'-end cross talks of mouse hepatitis virus RNA. *J. Virol.* 75, 5009–5017.
- Huang, Y.-W., Surka, M.C., Reynaud, D., Pace-Asciak, C., and Trimble, W.S. (2006). GTP binding and hydrolysis kinetics of human septin 2. *FEBS J.* 273, 3248–3260.
- Huang, G., Yan, H., Ye, S., Tong, C., and Ying, Q.-L. (2014). STAT3 phosphorylation at tyrosine 705 and serine 727 differentially regulates mouse ESC fates. *Stem Cells* 32, 1149–1160.
- Hwang, S.-K., Baker, A.R., Young, M.R., and Colburn, N.H. (2014). Tumor suppressor PDCD4 inhibits NF- κ B-dependent transcription in human glioblastoma cells by direct interaction with p65. *Carcinogenesis* 35, 1469–1480.
- Ichimura, T., Yamamura, H., Sasamoto, K., Tominaga, Y., Taoka, M., Kakiuchi, K., Shinkawa, T., Takahashi, N., Shimada, S., and Isobe, T. (2005). 14-3-3 proteins modulate the expression of epithelial Na⁺ channels by phosphorylation-dependent interaction with Nedd4-2 ubiquitin ligase. *J. Biol. Chem.* 280, 13187–13194.
- Inoki, K., Li, Y., Zhu, T., Wu, J., and Guan, K.-L. (2002). TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nat. Cell Biol.* 4, 648–657.
- Inoue, Y., Sato, H., Fujita, K., Tsukiyama-Kohara, K., Yoneda, M., and Kai, C. (2011). Selective translation of the measles virus nucleocapsid mRNA by a protein. *Front. Microbiol.* 2, 173.
- Intine, R.V., Tenenbaum, S.A., Sakulich, A.L., Keene, J.D., and Marais, R.J. (2003). Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. *Mol. Cell* 12, 1301–1307.
- Ishii, T., Warabi, E., Siow, R.C.M., and Mann, G.E. (2013). Sequestosome1/p62: a regulator of redox-sensitive voltage-activated potassium channels, arterial remodeling, inflammation, and neurite outgrowth. *Free Radic. Biol. Med.* 65, 102–116.
- Ishikawa, Y., Kawabata, S., and Sakurai, H. (2015). HSF1 transcriptional activity is modulated by IER5 and PP2A/B55. *FEBS Lett.* 589, 1150–1155.
- Izawa, I., and Inagaki, M. (2006). Regulatory mechanisms and functions of intermediate filaments: a study using site- and phosphorylation state-specific antibodies. *Cancer Sci.* 97, 167–174.
- Jacque, J.-M., and Stevenson, M. (2006). The inner-nuclear-envelope protein emerin regulates HIV-1 infectivity. *Nature* 441, 641–645.
- Jang, S.H., Jun, C.-D., and Park, Z.-Y. (2015). Label-free quantitative phosphorylation analysis of human transgelin2 in Jurkat T cells reveals distinct phosphorylation patterns under PKA and PKC activation conditions. *Proteome Sci.* 13, 14.
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48 (D1), D498–D503.
- Jiang, Y., Li, X., Yang, W., Hawke, D.H., Zheng, Y., Xia, Y., Aldape, K., Wei, C., Guo, F., Chen, Y., and Lu, Z. (2014). PKM2 regulates chromosome segregation and mitosis progression of tumor cells. *Mol. Cell* 53, 75–87.
- Jimenez-Guardeño, J.M., Nieto-Torres, J.L., DeDiego, M.L., Regla-Nava, J.A., Fernandez-Delgado, R., Castaño-Rodríguez, C., and Enjuanes, L. (2014). The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS Pathog.* 10, e1004320.
- Johnson, C., Tinti, M., Wood, N.T., Campbell, D.G., Toth, R., Dubois, F., Geraghty, K.M., Wong, B.H.C., Brown, L.J., Tyler, J., et al. (2011). Visualization and biochemical analyses of the emerging mammalian 14-3-3-phosphoproteome. *Mol. Cell. Proteomics* 10, M110.005751.
- Johnson, J.R., Crosby, D.C., Hultquist, J.F., Li, D., Marlett, J., Swann, J., Hüttenhain, R., Verschuere, E., Johnson, T.L., Newton, B.W., et al. (2020). Global post-translational modification profiling of HIV-1-infected cells reveals mechanisms of host cellular pathway remodeling. *bioRxiv*. <https://doi.org/10.1101/2020.01.06.896365>.
- Kane, S., Sano, H., Liu, S.C.H., Asara, J.M., Lane, W.S., Garner, C.C., and Lienhard, G.E. (2002). A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain. *J. Biol. Chem.* 277, 22115–22118.
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., Chen, Q., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B*. Published online April 20, 2020. <https://doi.org/10.1016/j.apsb.2020.04.009>.
- Kanzawa, N., Nishigaki, K., Hayashi, T., Ishii, Y., Furukawa, S., Niuro, A., Yasui, F., Kohara, M., Morita, K., Matsushima, K., et al. (2006). Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF- κ B activation. *FEBS Lett.* 580, 6807–6812.
- Karin, M. (1995). The regulation of AP-1 activity by mitogen-activated protein kinases. *J. Biol. Chem.* 270, 16483–16486.
- Kastenhuber, E.R., and Lowe, S.W. (2017). Putting p53 in Context. *Cell* 170, 1062–1078.
- Katayama, H., Sasai, K., Kawai, H., Yuan, Z.-M., Bondaruk, J., Suzuki, F., Fujii, S., Arlinghaus, R.B., Czerniak, B.A., and Sen, S. (2004). Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nat. Genet.* 36, 55–62.
- Kennedy, D., Mnich, K., Oommen, D., Chakravarthy, R., Almeida-Souza, L., Krols, M., Saveljeva, S., Doyle, K., Gupta, S., Timmerman, V., et al. (2017). HSPB1 facilitates ERK-mediated phosphorylation and degradation of BIM to

- attenuate endoplasmic reticulum stress-induced apoptosis. *Cell Death Dis.* 8, e3026.
- Khurana, N., and Bhattacharyya, S. (2015). Hsp90, the concertmaster: tuning transcription. *Front. Oncol.* 5, 100.
- Killoran, R.C., Fan, J., Yang, D., Shilton, B.H., and Choy, W.-Y. (2015). Structural Analysis of the 14-3-3/Chibby Interaction Involved in Wnt/ β -Catenin Signaling. *PLoS ONE* 10, e0123934.
- Kim, Y.-K., Lee, K.J., Jeon, H., and Yu, Y.G. (2006). Protein kinase CK2 is inhibited by human nucleolar phosphoprotein p140 in an inositol hexakisphosphate-dependent manner. *J. Biol. Chem.* 281, 36752–36757.
- Kindrachuk, J., Ork, B., Hart, B.J., Mazur, S., Holbrook, M.R., Frieman, M.B., Traynor, D., Johnson, R.F., Dyall, J., Kuhn, J.H., et al. (2015). Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis. *Antimicrob. Agents Chemother.* 59, 1088–1099.
- Kishimoto, H., Nakagawa, K., Watanabe, T., Kitagawa, D., Momose, H., Seo, J., Nishitai, G., Shimizu, N., Ohata, S., Tanemura, S., et al. (2003). Different properties of SEK1 and MKK7 in dual phosphorylation of stress-induced activated protein kinase SAPK/JNK in embryonic stem cells. *J. Biol. Chem.* 278, 16595–16601.
- Kjølby, M., Andersen, O.M., Breiderhoff, T., Fjorback, A.W., Pedersen, K.M., Madsen, P., Jansen, P., Heeren, J., Willnow, T.E., and Nykjaer, A. (2010). Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metab.* 12, 213–223.
- Koff, A., Giordano, A., Desai, D., Yamashita, K., Harper, J.W., Elledge, S., Nishimoto, T., Morgan, D.O., Franza, B.R., and Roberts, J.M. (1992). Formation and activation of a cyclin E-cdk2 complex during the G1 phase of the human cell cycle. *Science* 257, 1689–1694.
- Kohl, B., Zhong, X., Hermann, C., and Stoll, R. (2019). Phosphorylation orchestrates the structural ensemble of the intrinsically disordered protein HMGA1a and modulates its DNA binding to the NF κ B promoter. *Nucleic Acids Res.* 47, 11906–11920.
- Koseki, T., Inohara, N., Chen, S., Carrio, R., Merino, J., Hottiger, M.O., Nabel, G.J., and Núñez, G. (1999). CIPER, a novel NF kappaB-activating protein containing a caspase recruitment domain with homology to Herpesvirus-2 protein E10. *J. Biol. Chem.* 274, 9955–9961.
- Kostenko, S., Johannessen, M., and Moens, U. (2009). PKA-induced F-actin rearrangement requires phosphorylation of Hsp27 by the MAPKAP kinase MK5. *Cell. Signal.* 21, 712–718.
- Kotula, E., Faigle, W., Berthault, N., Dingli, F., Loew, D., Sun, J.-S., Dutreix, M., and Quanz, M. (2013). DNA-PK target identification reveals novel links between DNA repair signaling and cytoskeletal regulation. *PLoS ONE* 8, e80313.
- Kovacina, K.S., Park, G.Y., Bae, S.S., Guzzetta, A.W., Schaefer, E., Birnbaum, M.J., and Roth, R.A. (2003). Identification of a proline-rich Akt substrate as a 14-3-3 binding partner. *J. Biol. Chem.* 278, 10189–10194.
- Kraft, C., Herzog, F., Gieffers, C., Mechtler, K., Hagting, A., Pines, J., and Peters, J.M. (2003). Mitotic regulation of the human anaphase-promoting complex by phosphorylation. *EMBO J.* 22, 6598–6609.
- Kramer, E.R., Gieffers, C., Hölzl, G., Hengstschläger, M., and Peters, J.M. (1998). Activation of the human anaphase-promoting complex by proteins of the CDC20/Fizzy family. *Curr. Biol.* 8, 1207–1210.
- Kremer, B.E., Adang, L.A., and Macara, I.G. (2007). Septins regulate actin organization and cell-cycle arrest through nuclear accumulation of NCK mediated by SOCS7. *Cell* 130, 837–850.
- Krull, S., Dörries, J., Boysen, B., Reidenbach, S., Magnus, L., Norder, H., Thyberg, J., and Cordes, V.C. (2010). Protein Tpr is required for establishing nuclear pore-associated zones of heterochromatin exclusion. *EMBO J.* 29, 1659–1673.
- Kuang, X.-Y., Jiang, H.-S., Li, K., Zheng, Y.-Z., Liu, Y.-R., Qiao, F., Li, S., Hu, X., and Shao, Z.-M. (2016). The phosphorylation-specific association of STMN1 with GRP78 promotes breast cancer metastasis. *Cancer Lett.* 377, 87–96.
- Kumar, R., Whitehurst, C.B., and Pagano, J.S. (2014). The Rad6/18 ubiquitin complex interacts with the Epstein-Barr virus deubiquitinating enzyme, BPLF1, and contributes to virus infectivity. *J. Virol.* 88, 6411–6422.
- Kumar, R., Khandelwal, N., Thachamvally, R., Tripathi, B.N., Barua, S., Kashyap, S.K., Maherchandani, S., and Kumar, N. (2018). Role of MAPK/MNK1 signaling in virus replication. *Virus Res.* 253, 48–61.
- Kunapuli, P., Kasyapa, C.S., Chin, S.-F., Caldas, C., and Cowell, J.K. (2006). ZNF198, a zinc finger protein rearranged in myeloproliferative disease, localizes to the PML nuclear bodies and interacts with SUMO-1 and PML. *Exp. Cell Res.* 312, 3739–3751.
- Kuss, S.K., Mata, M.A., Zhang, L., and Fontoura, B.M.A. (2013). Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. *Viruses* 5, 1824–1849.
- Labit, H., Fujimitsu, K., Bayin, N.S., Takaki, T., Gannon, J., and Yamano, H. (2012). Dephosphorylation of Cdc20 is required for its C-box-dependent activation of the APC/C. *EMBO J.* 31, 3351–3362.
- Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., and Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents* 55, 105924.
- Lamarche-Vane, N., and Hall, A. (1998). Cdc42 GAP, a novel proline-rich GTPase-activating protein for Cdc42 and Rac. *J. Biol. Chem.* 273, 29172–29177.
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220.
- Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24, 719–720.
- Larance, M., Rowland, A.F., Hoehn, K.L., Humphreys, D.T., Preiss, T., Guilhaum, M., and James, D.E. (2010). Global phosphoproteomics identifies a major role for AKT and 14-3-3 in regulating EDC3. *Mol. Cell. Proteomics* 9, 682–694.
- Lau, E., Zhu, C., Abraham, R.T., and Jiang, W. (2006). The functional role of Cdc6 in S-G2/M in mammalian cells. *EMBO Rep.* 7, 425–430.
- Law, A.H.Y., Lee, D.C.W., Cheung, B.K.W., Yim, H.C.H., and Lau, A.S.Y. (2007). Role for nonstructural protein 1 of severe acute respiratory syndrome coronavirus in chemokine dysregulation. *J. Virol.* 81, 416–422.
- Le Sage, V., Cinti, A., Amorim, R., and Moulard, A.J. (2016). Adapting the Stress Response: Viral Subversion of the mTOR Signaling Pathway. *Viruses* 8, 152.
- Lee, K.-Y., and Myung, K. (2008). PCNA modifications for regulation of post-replication repair pathways. *Mol. Cells* 26, 5–11.
- Lee, J., Hong, F., Kwon, S., Kim, S.S., Kim, D.O., Kang, H.S., Lee, S.J., Ha, J., and Kim, S.S. (2002). Activation of p38 MAPK induces cell cycle arrest via inhibition of Raf/ERK pathway during muscle differentiation. *Biochem. Biophys. Res. Commun.* 298, 765–771.
- Lee, S., Salwinski, L., Zhang, C., Chu, D., Sampankpanich, C., Reyes, N.A., Vangeloff, A., Xing, F., Li, X., Wu, T.-T., et al. (2011). An integrated approach to elucidate the intra-viral and viral-cellular protein interaction networks of a gamma-herpesvirus. *PLoS Pathog.* 7, e1002297.
- Lee, K.-Y., Liu, L., Jin, Y., Fu, S.-B., and Rosales, J.L. (2012). Cdk5 mediates vimentin Ser56 phosphorylation during GTP-induced secretion by neutrophils. *J. Cell. Physiol.* 227, 739–750.
- Lee, W.-K., Son, S.H., Jin, B.-S., Na, J.-H., Kim, S.-Y., Kim, K.-H., Kim, E.E., Yu, Y.G., and Lee, H.H. (2013). Structural and functional insights into the regulation mechanism of CK2 by IP6 and the intrinsically disordered protein Nopp140. *Proc. Natl. Acad. Sci. USA* 110, 19360–19365.
- Leite, F., and Way, M. (2015). The role of signalling and the cytoskeleton during Vaccinia Virus egress. *Virus Res.* 209, 87–99.
- Lekmine, F., Uddin, S., Sassano, A., Parmar, S., Brachmann, S.M., Majchrzak, B., Sonenberg, N., Hay, N., Fish, E.N., and Platanius, L.C. (2003). Activation of the p70 S6 kinase and phosphorylation of the 4E-BP1 repressor of mRNA translation by type I interferons. *J. Biol. Chem.* 278, 27772–27780.

- Lekmine, F., Sassano, A., Uddin, S., Smith, J., Majchrzak, B., Brachmann, S.M., Hay, N., Fish, E.N., and Platanius, L.C. (2004). Interferon-gamma engages the p70 S6 kinase to regulate phosphorylation of the 40S S6 ribosomal protein. *Exp. Cell Res.* **295**, 173–182.
- Leukel, M., and Jost, E. (1995). Two conserved serines in the nuclear localization signal flanking region are involved in the nuclear targeting of human lamin A. *Eur. J. Cell Biol.* **68**, 133–142.
- Levy, D.E., and Lee, C.-K. (2002). What does Stat3 do? *J. Clin. Invest.* **109**, 1143–1148.
- Li, Q., and Martinez, J.D. (2011). Loss of HSF1 results in defective radiation-induced G(2) arrest and DNA repair. *Radiat. Res.* **176**, 17–24.
- Li, Z., and Nagy, P.D. (2011). Diverse roles of host RNA binding proteins in RNA virus replication. *RNA Biol.* **8**, 305–315.
- Li, F.-Q., Tam, J.P., and Liu, D.X. (2007a). Cell cycle arrest and apoptosis induced by the coronavirus infectious bronchitis virus in the absence of p53. *Virology* **365**, 435–445.
- Li, Y., Roux, C., Lazereg, S., LeCaer, J.-P., Laprévotte, O., Badet, B., and Badet-Denisot, M.-A. (2007b). Identification of a novel serine phosphorylation site in human glutamine:fructose-6-phosphate amidotransferase isoform 1. *Biochemistry* **46**, 13163–13169.
- Li, F.-Q., Mofunanya, A., Harris, K., and Takemaru, K. (2008). Chibby cooperates with 14-3-3 to regulate beta-catenin subcellular distribution and signaling activity. *J. Cell Biol.* **181**, 1141–1154.
- Li, G., Ma, D., and Chen, Y. (2016). Cellular functions of programmed cell death 5. *Biochim. Biophys. Acta* **1863**, 572–580.
- Li, J., Chen, C., Li, Y., Matye, D.J., Wang, Y., Ding, W.-X., and Li, T. (2017). Inhibition of insulin/PI3K/AKT signaling decreases adipose Sortilin 1 in mice and 3T3-L1 adipocytes. *Biochim Biophys Acta Mol Basis Dis* **1863**, 2924–2933.
- Liang, J., Cao, R., Zhang, Y., Xia, Y., Zheng, Y., Li, X., Wang, L., Yang, W., and Lu, Z. (2016). PKM2 dephosphorylation by Cdc25A promotes the Warburg effect and tumorigenesis. *Nat. Commun.* **7**, 12431.
- Liao, Q.-J., Ye, L.-B., Timani, K.A., Zeng, Y.-C., She, Y.-L., Ye, L., and Wu, Z.-H. (2005). Activation of NF-kappaB by the full-length nucleocapsid protein of the SARS coronavirus. *Acta Biochim. Biophys. Sin. (Shanghai)* **37**, 607–612.
- Lieberman, H.B., Panigrahi, S.K., Hopkins, K.M., Wang, L., and Broustas, C.G. (2017). p53 and RAD9, the DNA Damage Response, and Regulation of Transcription Networks. *Radiat. Res.* **187**, 424–432.
- Lilley, C.E., Schwartz, R.A., and Weitzman, M.D. (2007). Using or abusing: viruses and the cellular DNA damage response. *Trends Microbiol.* **15**, 119–126.
- Lim, H.C., Xie, L., Zhang, W., Li, R., Chen, Z.-C., Wu, G.-Z., Cui, S.-S., Tan, E.K., and Zeng, L. (2013). Ribosomal S6 Kinase 2 (RSK2) maintains genomic stability by activating the Atm/p53-dependent DNA damage pathway. *PLoS ONE* **8**, e74334.
- Lin, J.-Y., Li, M.-L., and Shih, S.-R. (2009). Far upstream element binding protein 2 interacts with enterovirus 71 internal ribosomal entry site and negatively regulates viral translation. *Nucleic Acids Res.* **37**, 47–59.
- Lin, X., Li, S., Zhao, Y., Ma, X., Zhang, K., He, X., and Wang, Z. (2013). Interaction domains of p62: a bridge between p62 and selective autophagy. *DNA Cell Biol.* **32**, 220–227.
- Lingel, H., Wissing, J., Arra, A., Schanze, D., Lienenklaus, S., Klawonn, F., Pierau, M., Zenker, M., Jänsch, L., and Brunner-Weinzierl, M.C. (2017). CTLA-4-mediated posttranslational modifications direct cytotoxic T-lymphocyte differentiation. *Cell Death Differ.* **24**, 1739–1749.
- Lisnock, J., Griffin, P., Calaycay, J., Frantz, B., Parsons, J., O'Keefe, S.J., and LoGrasso, P. (2000). Activation of JNK3 alpha 1 requires both MKK4 and MKK7: kinetic characterization of in vitro phosphorylated JNK3 alpha 1. *Biochemistry* **39**, 3141–3148.
- Lista, M.J., Page, R., Sertkaya, H., Matos, P., Ortiz-Zapater, E., Maguire, T.J.A., Poulton, K., O'Byrne, A., Bouton, C., Dickenson, R.E., et al. (2020). Resilient SARS-CoV-2 diagnostics workflows including viral heat inactivation. *Nature* **581**, 215–220.
- Liu, Y., and Schneider, M.F. (2013). Opposing HDAC4 nuclear fluxes due to phosphorylation by beta-adrenergic activated protein kinase A or by activity or Epac activated CaMKII in skeletal muscle fibres. *J. Physiol.* **597**, 3605–3623.
- Liu, Y., Tan, H., Tian, H., Liang, C., Chen, S., and Liu, Q. (2011). Autoantigen La promotes efficient RNAi, antiviral response, and transposon silencing by facilitating multiple-turnover RISC catalysis. *Mol. Cell* **44**, 502–508.
- Liu, R.-Y., Diao, C.-F., Zhang, Y., Wu, N., Wan, H.-Y., Nong, X.-Y., Liu, M., and Tang, H. (2013). miR-371-5p down-regulates pre mRNA processing factor 4 homolog B (PRPF4B) and facilitates the G1/S transition in human hepatocellular carcinoma cells. *Cancer Lett.* **335**, 351–360.
- Liu, A.-L., Li, Y.-F., Qi, W., Ma, X.-L., Yu, K.-X., Huang, B., Liao, M., Li, F., Pan, J., and Song, M.-X. (2015). Comparative analysis of selected innate immune-related genes following infection of immortal DF-1 cells with highly pathogenic (H5N1) and low pathogenic (H9N2) avian influenza viruses. *Virus Genes* **50**, 189–199.
- Lobry, C., Lopez, T., Israël, A., and Weil, R. (2007). Negative feedback loop in T cell activation through IkkappaB kinase-induced phosphorylation and degradation of Bcl10. *Proc. Natl. Acad. Sci. USA* **104**, 908–913.
- Lokireddy, S., Kukushkin, N.V., and Goldberg, A.L. (2015). cAMP-induced phosphorylation of 26S proteasomes on Rpn6/PSMD11 enhances their activity and the degradation of misfolded proteins. *Proc. Natl. Acad. Sci. USA* **112**, E7176–E7185.
- Long, L., Thelen, J.P., Furgason, M., Haj-Yahya, M., Brik, A., Cheng, D., Peng, J., and Yao, T. (2014). The U4/U6 recycling factor SART3 has histone chaperone activity and associates with USP15 to regulate H2B deubiquitination. *J. Biol. Chem.* **289**, 8916–8930.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Lu, Y., Liu, C., Xu, Y.-F., Cheng, H., Shi, S., Wu, C.-T., and Yu, X.-J. (2014). Stathmin destabilizing microtubule dynamics promotes malignant potential in cancer cells by epithelial-mesenchymal transition. *HBPD INT* **13**, 386–394.
- Lu, H., Shamanna, R.A., de Freitas, J.K., Okur, M., Khadka, P., Kulikowicz, T., Holland, P.P., Tian, J., Croteau, D.L., Davis, A.J., and Bohr, V.A. (2017). Cell cycle-dependent phosphorylation regulates RECQL4 pathway choice and ubiquitination in DNA double-strand break repair. *Nat. Commun.* **8**, 2039.
- Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **43**, 624–637.
- Luciani, M.G., Hutchins, J.R., Zheleva, D., and Hupp, T.R. (2000). The C-terminal regulatory domain of p53 contains a functional docking site for cyclin A. *J. Mol. Biol.* **300**, 503–518.
- Lukas, J., Lukas, C., and Bartek, J. (2011). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nat. Cell Biol.* **13**, 1161–1169.
- Luo, H., Chen, Q., Chen, J., Chen, K., Shen, X., and Jiang, H. (2005). The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS Lett.* **579**, 2623–2628.
- Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M.J., Stratton, M.R., and Lamond, A.I. (2014). A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *eLife* **3**, e01630.
- Ly, T., Whigham, A., Clarke, R., Brenes-Murillo, A.J., Estes, B., Madhessian, D., Lundberg, E., Wadsworth, P., and Lamond, A.I. (2017). Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic sub-phases, using PRIMMUS. *eLife* **6**, e27574.
- Machado-Neto, J.A., Saad, S.T.O., and Traina, F. (2014). Stathmin 1 in normal and malignant hematopoiesis. *BMB Rep.* **47**, 660–665.
- Machowska, M., Piekarczyk, K., and Rzepecki, R. (2015). Regulation of lamin properties and functions: does phosphorylation do it all? *Open Biol.* **5**, 150094.
- Madureira, P.A., Hill, R., Lee, P.W.K., and Waisman, D.M. (2012). Genotoxic agents promote the nuclear accumulation of annexin A2: role of annexin A2 in mitigating DNA damage. *PLoS ONE* **7**, e50591.

- Magnuson, B., Ekim, B., and Fingar, D.C. (2012). Regulation and function of ribosomal protein S6 kinase (S6K) within mTOR signalling networks. *Biochem. J.* **447**, 1–21.
- Magron, A., Elowe, S., and Carreau, M. (2015). The Fanconi Anemia C Protein Binds to and Regulates Stathmin-1 Phosphorylation. *PLoS ONE* **10**, e0140612.
- Mahony, R., Broadbent, L., Maier-Moore, J.S., Power, U.F., and Jefferies, C.A. (2017). The RNA binding protein La/SS-B promotes RIG-I-mediated type I and type III IFN responses following Sendai viral infection. *Sci. Rep.* **7**, 14537.
- Mancini, M., Soverini, S., Gugliotta, G., Santucci, M.A., Rosti, G., Cavo, M., Martinelli, G., and Castagnetti, F. (2017). Chibby 1: a new component of β -catenin-signaling in chronic myeloid leukemia. *Oncotarget* **8**, 88244–88250.
- Mandl, M., Slack, D.N., and Keyse, S.M. (2005). Specific inactivation and nuclear anchoring of extracellular signal-regulated kinase 2 by the inducible dual-specificity protein phosphatase DUSP5. *Mol. Cell Biol.* **25**, 1830–1845.
- Mangin, P.H., Kleitz, L., Boucheix, C., Gachet, C., and Lanza, F. (2009). CD9 negatively regulates integrin α IIb β 3 activation and could thus prevent excessive platelet recruitment at sites of vascular injury. *J. Thromb. Haemost.* **7**, 900–902.
- Manning, B.D., and Toker, A. (2017). AKT/PKB Signaling: Navigating the Network. *Cell* **169**, 381–405.
- Marcos-Villar, L., Pazo, A., and Nieto, A. (2016). Influenza Virus and Chromatin: Role of the CHD1 Chromatin Remodeler in the Virus Life Cycle. *J. Virol.* **90**, 3694–3707.
- Marklund, U., Brattsand, G., Osterman, O., Ohlsson, P.I., and Gullberg, M. (1993). Multiple signal transduction pathways induce phosphorylation of serines 16, 25, and 38 of oncoprotein 18 in T lymphocytes. *J. Biol. Chem.* **268**, 25671–25680.
- Marklund, U., Larsson, N., Gradin, H.M., Brattsand, G., and Gullberg, M. (1996). Oncoprotein 18 is a phosphorylation-responsive regulator of microtubule dynamics. *EMBO J.* **15**, 5290–5298.
- Marquette, A., André, J., Bagot, M., Bensussan, A., and Dumaz, N. (2011). ERK and PDE4 cooperate to induce RAF isoform switching in melanoma. *Nat. Struct. Mol. Biol.* **18**, 584–591.
- Mather, K.A., Thalamuthu, A., Oldmeadow, C., Song, F., Armstrong, N.J., Poljak, A., Holliday, E.G., McEvoy, M., Kwok, J.B., Assareh, A.A., et al. (2016). Genome-wide significant results identified for plasma apolipoprotein H levels in middle-aged and older adults. *Sci. Rep.* **6**, 23675.
- Matsushashi, S., Manirujjaman, M., Hamajima, H., and Ozaki, I. (2019). Control Mechanisms of the Tumor Suppressor PDCD4: Expression and Functions. *Int. J. Mol. Sci.* **20**, 2304.
- Matsumoto, T., Urushido, M., Ide, H., Ishihara, M., Hamada-Ode, K., Shimamura, Y., Ogata, K., Inoue, K., Taniguchi, Y., Taguchi, T., et al. (2015). Small Heat Shock Protein Beta-1 (HSPB1) Is Upregulated and Regulates Autophagy and Apoptosis of Renal Tubular Cells in Acute Kidney Injury. *PLoS ONE* **10**, e0126229.
- Matsuoka, Y., Nagahara, Y., Ikekita, M., and Shinomiya, T. (2003). A novel immunosuppressive agent FTY720 induced Akt dephosphorylation in leukemia cells. *Br. J. Pharmacol.* **138**, 1303–1312.
- Mbonye, U.R., Wang, B., Gokulrangan, G., Chance, M.R., and Karn, J. (2015). Phosphorylation of HEXIM1 at Tyr271 and Tyr274 Promotes Release of P-TEFb from the 7SK snRNP Complex and Enhances Proviral HIV Gene Expression. *Proteomics* **15**, 2078–2086.
- McCampbell, P.K., Farah, C.A., Anadol, M.N., Hoque, S., and Sossin, W.S. (2015). Bidirectional regulation of eEF2 phosphorylation controls synaptic plasticity by decoding neuronal activity patterns. *J. Neurosci.* **35**, 4403–4417.
- McClain, D.A., Lubas, W.A., Cooksey, R.C., Hazel, M., Parker, G.J., Love, D.C., and Hanover, J.A. (2002). Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *Proc. Natl. Acad. Sci. USA* **99**, 10695–10699.
- Mehellou, Y., Alamri, M.A., Dhiani, B.A., and Kadri, H. (2018). C-terminal phosphorylation of SPAK and OSR1 kinases promotes their binding and activation by the scaffolding protein MO25. *Biochem. Biophys. Res. Commun.* **503**, 1868–1873.
- Meier, U.T., and Blobel, G. (1990). A nuclear localization signal binding protein in the nucleolus. *J. Cell Biol.* **111**, 2235–2245.
- Meier, U.T., and Blobel, G. (1992). Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell* **70**, 127–138.
- Mendoza, M.C., Er, E.E., Zhang, W., Ballif, B.A., Elliott, H.L., Danuser, G., and Blenis, J. (2011). ERK-MAPK drives lamellipodia protrusion by activating the WAVE2 regulatory complex. *Mol. Cell* **41**, 661–671.
- Menon, S., Dibble, C.C., Talbott, G., Hoxhaj, G., Valvezan, A.J., Takahashi, H., Cantley, L.C., and Manning, B.D. (2014). Spatial control of the TSC complex integrates insulin and nutrient regulation of mTORC1 at the lysosome. *Cell* **156**, 771–785.
- Mizutani, T., Fukushi, S., Saijo, M., Kurane, I., and Morikawa, S. (2004). Phosphorylation of p38 MAPK and its downstream targets in SARS coronavirus-infected cells. *Biochem. Biophys. Res. Commun.* **319**, 1228–1234.
- Mo, M., Shahar, S., Fleming, S.B., and Mercer, A.A. (2012). How viruses affect the cell cycle through manipulation of the APC/C. *Trends Microbiol.* **20**, 440–448.
- Moldave, K. (1985). Eukaryotic protein synthesis. *Annu. Rev. Biochem.* **54**, 1109–1149.
- Morris, E.J., Kawamura, E., Gillespie, J.A., Balgi, A., Kannan, N., Muller, W.J., Roberge, M., and Dedhar, S. (2017). Stat3 regulates centrosome clustering in cancer cells via Stathmin/PLK1. *Nat. Commun.* **8**, 15289.
- Muda, M., Theodosiou, A., Rodrigues, N., Boschert, U., Camps, M., Gillieron, C., Davies, K., Ashworth, A., and Arkinstall, S. (1996). The dual specificity phosphatases M3/6 and MKP-3 are highly selective for inactivation of distinct mitogen-activated protein kinases. *J. Biol. Chem.* **271**, 27205–27208.
- Mueller, P.R., Coleman, T.R., Kumagai, A., and Dunphy, W.G. (1995). Myt1: a membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15. *Science* **270**, 86–90.
- Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L., and Slany, R.K. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood* **110**, 4445–4454.
- Müller, S., Montes de Oca, R., Lacoste, N., Dingli, F., Loew, D., and Almouzni, G. (2014). Phosphorylation and DNA binding of HJURP determine its centromeric recruitment and function in CenH3(CENP-A) loading. *Cell Rep.* **8**, 190–203.
- Na, B.-R., Kim, H.-R., Piragyte, I., Oh, H.-M., Kwon, M.-S., Akber, U., Lee, H.-S., Park, D.-S., Song, W.K., Park, Z.-Y., et al. (2015). TAGLN2 regulates T cell activation by stabilizing the actin cytoskeleton at the immunological synapse. *J. Cell Biol.* **209**, 143–162.
- Naim, A.C., Matsushita, M., Nastiuk, K., Horiuchi, A., Mitsui, K., Shimizu, Y., and Palfrey, H.C. (2001). Elongation factor-2 phosphorylation and the regulation of protein synthesis by calcium. *Prog. Mol. Subcell. Biol.* **27**, 91–129.
- Nandi, A.K., Ford, T., Fleksher, D., Neuman, B., and Rapoport, A.P. (2007). Attenuation of DNA damage checkpoint by PBK, a novel mitotic kinase, involves protein-protein interaction with tumor suppressor p53. *Biochem. Biophys. Res. Commun.* **358**, 181–188.
- Needham, E.J., Humphrey, S.J., Cooke, K.C., Fazakerley, D.J., Duan, X., Parker, B.L., and James, D.E. (2019). Phosphoproteomics of Acute Cell Stressors Targeting Exercise Signaling Networks Reveal Drug Interactions Regulating Protein Secretion. *Cell Rep.* **29**, 1524–1538.e6.
- Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814.
- Nielsen, M.S., Madsen, P., Christensen, E.I., Nykjaer, A., Gliemann, J., Kasper, D., Pohlmann, R., and Petersen, C.M. (2001). The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20**, 2180–2190.
- Ochoa, D., Jonikas, M., Lawrence, R.T., El Debs, B., Selkrig, J., Typas, A., Villén, J., Santos, S.D., and Beltrao, P. (2016). An atlas of human kinase regulation. *Mol. Syst. Biol.* **12**, 888.

- Ochoa, D., Jarnuczak, A.F., Viéitez, C., Gehre, M., Soucheray, M., Mateus, A., Kleefeldt, A.A., Hill, A., Garcia-Alonso, L., Stein, F., et al. (2020). The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **38**, 365–373.
- Ogiso, H., Kagi, N., Matsumoto, E., Nishimoto, M., Arai, R., Shirouzu, M., Mimura, J., Fujii-Kuriyama, Y., and Yokoyama, S. (2004). Phosphorylation analysis of 90 kDa heat shock protein within the cytosolic arylhydrocarbon receptor complex. *Biochemistry* **43**, 15510–15519.
- Ohtsubo, M., Theodoras, A.M., Schumacher, J., Roberts, J.M., and Pagano, M. (1995). Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol. Cell Biol.* **15**, 2612–2624.
- Old, W.M., Shabb, J.B., Houel, S., Wang, H., Coutts, K.L., Yen, C.-Y., Litman, E.S., Croy, C.H., Meyer-Arendt, K., Miranda, J.G., et al. (2009). Functional proteomics identifies targets of phosphorylation by B-Raf signaling in melanoma. *Mol. Cell* **34**, 115–131.
- Olsen, J.V., Blagoev, B., Gnäd, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648.
- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnäd, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3.
- Osada, N., Kohara, A., Yamaji, T., Hirayama, N., Kasai, F., Sekizuka, T., Kuroda, M., and Hanada, K. (2014). The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res.* **21**, 673–683.
- Oshiro, N., Takahashi, R., Yoshino, K., Tanimura, K., Nakashima, A., Eguchi, S., Miyamoto, T., Hara, K., Takehana, K., Avruch, J., et al. (2007). The proline-rich Akt substrate of 40 kDa (PRAS40) is a physiological substrate of mammalian target of rapamycin complex 1. *J. Biol. Chem.* **282**, 20329–20339.
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **11**, 1620.
- Palamarchuk, A., Efanov, A., Maximov, V., Aqeilan, R.I., Croce, C.M., and Pekarsky, Y. (2005). Akt phosphorylates and regulates Pcdcd4 tumor suppressor protein. *Cancer Res.* **65**, 11282–11286.
- Pan, D. (2010). The hippo signaling pathway in development and cancer. *Dev. Cell* **19**, 491–505.
- Paquette, M., El-Houjeiri, L., and Pause, A. (2018). mTOR Pathways in Cancer and Autophagy. *Cancers (Basel)* **10**, 18.
- Park, E.S., Kim, H., Suh, J.M., Park, S.J., You, S.H., Chung, H.K., Lee, K.W., Kwon, O.Y., Cho, B.Y., Kim, Y.K., et al. (2000). Involvement of JAK/STAT (Janus kinase/signal transducer and activator of transcription) in the thyrotropin signaling pathway. *Mol. Endocrinol.* **14**, 662–670.
- Parker, L.L., and Pivnicka-Worms, H. (1992). Inactivation of the p34cdc2-cyclin B complex by the human WEE1 tyrosine kinase. *Science* **257**, 1955–1957.
- Parplys, A.C., Zhao, W., Sharma, N., Groesser, T., Liang, F., Maranon, D.G., Leung, S.G., Grundt, K., Dray, E., Idate, R., et al. (2015). NUCKS1 is a novel RAD51AP1 paralog important for homologous recombination and genome stability. *Nucleic Acids Res.* **43**, 9817–9834.
- Patel, K.M., Strong, A., Tohyama, J., Jin, X., Morales, C.R., Billheimer, J., Millar, J., Kruth, H., and Rader, D.J. (2015). Macrophage sortilin promotes LDL uptake, foam cell formation, and atherosclerosis. *Circ. Res.* **116**, 789–796.
- Patnaik, A., Haluska, P., Tolcher, A.W., Erlichman, C., Papadopoulos, K.P., Lensing, J.L., Beeram, M., Molina, J.R., Rasco, D.W., Arcos, R.R., et al. (2016). A First-in-Human Phase I Study of the Oral p38 MAPK Inhibitor, Ralimetinib (LY2228820 Dimesylate), in Patients with Advanced Cancer. *Clin. Cancer Res.* **22**, 1095–1102.
- Pedersen, S.F., and Ho, Y.-C. (2020). SARS-CoV-2: a storm is raging. *J. Clin. Invest.* **130**, 2202–2205.
- Peiris, J.S.M., Yuen, K.Y., Osterhaus, A.D.M.E., and Stöhr, K. (2003). The severe acute respiratory syndrome. *N. Engl. J. Med.* **349**, 2431–2441.
- Peng, T.-Y., Lee, K.-R., and Tarn, W.-Y. (2008). Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J.* **275**, 4152–4163.
- Penzo, C., Arnoldo, L., Pegoraro, S., Petrosino, S., Ros, G., Zanin, R., Wiśniewski, J.R., Manfioletti, G., and Sgarra, R. (2019). HMGA1 Modulates Gene Transcription Sustaining a Tumor Signalling Pathway Acting on the Epigenetic Status of Triple-Negative Breast Cancer Cells. *Cancers (Basel)* **11**, 1105.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47** (D1), D442–D450.
- Pfeiffer, J.R., McAvoy, B.L., Fecteau, R.E., Deleault, K.M., and Brooks, S.A. (2011). CARHSP1 is required for effective tumor necrosis factor alpha mRNA stabilization and localizes to processing bodies and exosomes. *Mol. Cell Biol.* **31**, 277–286.
- Pires, E., Sung, P., and Wiese, C. (2017). Role of RAD51AP1 in homologous recombination DNA repair and carcinogenesis. *DNA Repair (Amst.)* **59**, 76–81.
- Ponnam, S., Sevrieva, I., Sun, Y.-B., Irving, M., and Kampourakis, T. (2019). Site-specific phosphorylation of myosin binding protein-C coordinates thin and thick filament activation in cardiac muscle. *Proc. Natl. Acad. Sci. USA* **116**, 15485–15494.
- Popov, D. (2012). Endoplasmic reticulum stress and the on site function of resident PTP1B. *Biochem. Biophys. Res. Commun.* **422**, 535–538.
- Poppe, M., Wittig, S., Jurida, L., Bartkuhn, M., Wilhelm, J., Müller, H., Beuerlein, K., Karl, N., Bhujju, S., Ziebuhr, J., et al. (2017). The NF- κ B-dependent and -independent transcriptome and chromatin landscapes of human coronavirus 229E-infected cells. *PLoS Pathog.* **13**, e1006286.
- Porollo, A., and Meller, J. (2007). Prediction-based fingerprints of protein-protein interactions. *Proteins* **66**, 630–645.
- Potter, C.J., Pedraza, L.G., and Xu, T. (2002). Akt regulates growth by directly phosphorylating Tsc2. *Nat. Cell Biol.* **4**, 658–665.
- Potthoff, M.J., and Olson, E.N. (2007). MEF2: a central regulator of diverse developmental programs. *Development* **134**, 4131–4140.
- Preisinger, C., Short, B., De Corte, V., Bruyneel, E., Haas, A., Kopajtic, R., Gettemans, J., and Barr, F.A. (2004). YSK1 is activated by the Golgi matrix protein GM130 and plays a role in cell migration through its substrate 14-3-3zeta. *J. Cell Biol.* **164**, 1009–1020.
- Proteau, A., Blier, S., Albert, A.L., Lavoie, S.B., Traish, A.M., and Vincent, M. (2005). The multifunctional nuclear protein p54nrb is multiphosphorylated in mitosis and interacts with the mitotic regulator Pin1. *J. Mol. Biol.* **346**, 1163–1172.
- Qiao, R., Weissmann, F., Yamaguchi, M., Brown, N.G., VanderLinden, R., Imre, R., Jarvis, M.A., Brunner, M.R., Davidson, I.F., Litos, G., et al. (2016). Mechanism of APC/CCDC20 activation by mitotic phosphorylation. *Proc. Natl. Acad. Sci. USA* **113**, E2570–E2578.
- Qiu, Q., Jiang, J., Lin, L., Cheng, S., Xin, D., Jiang, W., Shen, J., and Hu, Z. (2016). Downregulation of RSK2 influences the biological activities of human osteosarcoma cells through inactivating AKT/mTOR signaling pathways. *Int. J. Oncol.* **48**, 2508–2520.
- Quan, B., Seo, H.-S., Blobel, G., and Ren, Y. (2014). Vesiculoviral matrix (M) protein occupies nucleic acid binding site at nucleoporin pair (Rae1 • Nup98). *Proc. Natl. Acad. Sci. USA* **111**, 9127–9132.
- R Core Team. (2019). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
- Radic, V. (2016). A Novel Role for SCAMP3 as an Innate Immune Factor (UC Riverside).
- Radtke, K., Döhner, K., and Sodeik, B. (2006). Viral interactions with the cytoskeleton: a hitchhiker's guide to the cell. *Cell. Microbiol.* **8**, 387–400.
- Radulescu, A.E., and Cleveland, D.W. (2010). NuMA after 30 years: the matrix revisited. *Trends Cell Biol.* **20**, 214–222.

- Rahim, M.N., Klewes, L., Zahedi-Amiri, A., Mai, S., and Coombs, K.M. (2018). Global Interactomics Connect Nuclear Mitotic Apparatus Protein NUMA1 to Influenza Virus Maturation. *Viruses* 10, 731.
- Rai, R., Hu, C., Broton, C., Chen, Y., Lei, M., and Chang, S. (2017). NBS1 Phosphorylation Status Dictates Repair Choice of Dysfunctional Telomeres. *Mol. Cell* 65, 801–817.e4.
- Ramm, G., Larance, M., Guilhaus, M., and James, D.E. (2006). A role for 14-3-3 in insulin-stimulated GLUT4 translocation through its interaction with the Rab-GAP AS160. *J. Biol. Chem.* 281, 29174–29180.
- Raught, B., Peiretti, F., Gingras, A.-C., Livingstone, M., Shahbazian, D., Mayeur, G.L., Polakiewicz, R.D., Sonenberg, N., and Hershey, J.W.B. (2004). Phosphorylation of eucaryotic translation initiation factor 4B Ser422 is modulated by S6 kinases. *EMBO J.* 23, 1761–1769.
- Ravichandran, L.V., Chen, H., Li, Y., and Quon, M.J. (2001). Phosphorylation of PTP1B at Ser(50) by Akt impairs its ability to dephosphorylate the insulin receptor. *Mol. Endocrinol.* 15, 1768–1780.
- Rebeaud, F., Halfinger, S., Posevitz-Fejfar, A., Tapernoux, M., Moser, R., Rueda, D., Gaide, O., Guzzardi, M., Iancu, E.M., Rufer, N., et al. (2008). The proteolytic activity of the paracaspase MALT1 is key in T cell activation. *Nat. Immunol.* 9, 272–281.
- Renganathan, H., Vaidyanathan, H., Knapinska, A., and Ramos, J.W. (2005). Phosphorylation of PEA-15 switches its binding specificity from ERK/MAPK to FADD. *Biochem. J.* 390, 729–735.
- Revenko, A.S., Kalashnikova, E.V., Gemo, A.T., Zou, J.X., and Chen, H.-W. (2010). Chromatin loading of E2F-MLL complex by cancer-associated coregulator ANCCA via reading a specific histone mark. *Mol. Cell. Biol.* 30, 5260–5272.
- Riad, A., Zeng, C., Weng, C.-C., Winters, H., Xu, K., Makvandi, M., Metz, T., Carlin, S., and Mach, R.H. (2018). Sigma-2 Receptor/TMEM97 and PGRMC-1 Increase the Rate of Internalization of LDL by LDL Receptor through the Formation of a Ternary Complex. *Sci. Rep.* 8, 16845.
- Robinson, M., Schor, S., Barouch-Bentov, R., and Einav, S. (2018). Viral journeys on the intracellular highways. *Cell. Mol. Life Sci.* 75, 3693–3714.
- Robinson-White, A., and Stratakis, C.A. (2002). Protein kinase A signaling: “cross-talk” with other pathways in endocrine cells. *Ann. N Y Acad. Sci.* 968, 256–270.
- Robitaille, A.M., Christen, S., Shimobayashi, M., Cornu, M., Fava, L.L., Moes, S., Prescianotto-Baschong, C., Sauer, U., Jenoe, P., and Hall, M.N. (2013). Quantitative phosphoproteomics reveal mTORC1 activates de novo pyrimidine synthesis. *Science* 339, 1320–1323.
- Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* 273, 5858–5868.
- Rogalla, T., Ehmsperger, M., Preville, X., Kotlyarov, A., Lutsch, G., Ducasse, C., Paul, C., Wieske, M., Arrigo, A.P., Buchner, J., and Gaestel, M. (1999). Regulation of Hsp27 oligomerization, chaperone function, and protective activity against oxidative stress/tumor necrosis factor α by phosphorylation. *J. Biol. Chem.* 274, 18947–18956.
- Rogers, S., McCloy, R.A., Parker, B.L., Chaudhuri, R., Gayevskiy, V., Hoffman, N.J., Watkins, D.N., Daly, R.J., James, D.E., and Burgess, A. (2015). Dataset from the global phosphoproteomic mapping of early mitotic exit in human cells. *Data Brief* 5, 45–52.
- Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., et al. (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N. Engl. J. Med.* 382, 970–971.
- Roy, R., Durie, D., Li, H., Liu, B.-Q., Skehel, J.M., Mauri, F., Cuorvo, L.V., Barbaresi, M., Guo, L., Holcik, M., et al. (2014). hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2/S6K2 signaling. *Nucleic Acids Res.* 42, 12483–12497.
- Roy, R., Huang, Y., Seckl, M.J., and Pardo, O.E. (2017). Emerging roles of hnRNPA1 in modulating malignant transformation. *Wiley Interdiscip. Rev. RNA* 8.
- Ryan, E.L., Hollingworth, R., and Grand, R.J. (2016). Activation of the DNA Damage Response by RNA Viruses. *Biomolecules* 6, 2.
- Salvi, M., Xu, D., Chen, Y., Cabrelle, A., Sarno, S., and Pinna, L.A. (2009). Programmed cell death protein 5 (PDCD5) is phosphorylated by CK2 in vitro and in 293T cells. *Biochem. Biophys. Res. Commun.* 387, 606–610.
- Sanborn, K.B., Mace, E.M., Rak, G.D., Difeo, A., Martignetti, J.A., Pecci, A., Bussel, J.B., Favier, R., and Orange, J.S. (2011). Phosphorylation of the myosin IIA tailpiece regulates single myosin IIA molecule association with lytic granules to promote NK-cell cytotoxicity. *Blood* 118, 5862–5871.
- Sánchez-Martín, P., and Komatsu, M. (2018). p62/SQSTM1 - steering the cell through health and disease. *J. Cell Sci.* 131, jcs222836.
- Santamaria, P.G., Finley, D., Ballesta, J.P.G., and Remacha, M. (2003). Rpn6p, a proteasome subunit from *Saccharomyces cerevisiae*, is essential for the assembly and activity of the 26 S proteasome. *J. Biol. Chem.* 278, 6687–6695.
- Saredi, A., Howard, L., and Compton, D.A. (1997). Phosphorylation regulates the assembly of NuMA in a mammalian mitotic extract. *J. Cell Sci.* 110, 1287–1297.
- Schmucker, S., and Sumara, I. (2014). Molecular dynamics of PLK1 during mitosis. *Mol. Cell. Oncol.* 1, e954507.
- Schweppe, D.K., Rigas, J.R., and Gerber, S.A. (2013). Quantitative phosphoproteomic profiling of human non-small cell lung cancer tumors. *J. Proteomics* 91, 286–296.
- Sengupta, S., Lorente-Rodríguez, A., Earnest, S., Stippec, S., Guo, X., Trudgian, D.C., Mirzaei, H., and Cobb, M.H. (2013). Regulation of OSR1 and the sodium, potassium, two chloride cotransporter by convergent signals. *Proc. Natl. Acad. Sci. USA* 110, 18826–18831.
- Serber, Z., and Ferrell, J.E., Jr. (2007). Tuning bulk electrostatics to regulate protein function. *Cell* 128, 441–444.
- Shah, P.S., Link, N., Jang, G.M., Sharp, P.P., Zhu, T., Swaney, D.L., Johnson, J.R., Von Dollen, J., Ramage, H.R., Satkamp, L., et al. (2018). Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* 175, 1931–1945.e18.
- Shahbazian, D., Roux, P.P., Mieulet, V., Cohen, M.S., Raught, B., Taunton, J., Hershey, J.W.B., Blenis, J., Pende, M., and Sonenberg, N. (2006). The mTOR/P3K and MAPK pathways converge on eIF4B to control its phosphorylation and activity. *EMBO J.* 25, 2781–2791.
- Shandilya, J., Senapati, P., Dhanasekaran, K., Bangalore, S.S., Kumar, M., Kishore, A.H., Bhat, A., Kodaganur, G.S., and Kundu, T.K. (2014). Phosphorylation of multifunctional nucleolar protein nucleophosmin (NPM1) by aurora kinase B is critical for mitotic progression. *FEBS Lett.* 588, 2198–2205.
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224.
- Sharma, K., Åkerström, S., Sharma, A.K., Chow, V.T.K., Teow, S., Abrenica, B., Booth, S.A., Booth, T.F., Mirzami, A., and Lal, S.K. (2011). SARS-CoV 9b protein diffuses into nucleus, undergoes active Crm1 mediated nucleocytoplasmic export and triggers apoptosis when retained in the nucleus. *PLoS ONE* 6, e19436.
- Sharma, K., D’Souza, R.C.J., Tyanova, S., Schaab, C., Wiśniewski, J.R., Cox, J., and Mann, M. (2014). Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* 8, 1583–1594.
- Shi, J.-H., and Sun, S.-C. (2015). TCR signaling to NF- κ B and mTORC1: Expanding roles of the CARMA1 complex. *Mol. Immunol.* 68 (2 Pt C), 546–557.
- Shi, J., Zhang, T., Zhou, C., Chohan, M.O., Gu, X., Wegiel, J., Zhou, J., Hwang, Y.-W., Iqbal, K., Grundke-Iqbal, I., et al. (2008). Increased dosage of Dyrk1A alters alternative splicing factor (ASF)-regulated alternative splicing of tau in Down syndrome. *J. Biol. Chem.* 283, 28660–28669.
- Shih, M.-C., Chen, J.-Y., Wu, Y.-C., Jan, Y.-H., Yang, B.-M., Lu, P.-J., Cheng, H.-C., Huang, M.-S., Yang, C.-J., Hsiao, M., and Lai, J.M. (2012). TOPK/PBK promotes cell migration via modulation of the PI3K/PDEN/AKT pathway and is associated with poor prognosis in lung cancer. *Oncogene* 31, 2389–2400.
- Sinclair, A., Yarranton, S., and Schelcher, C. (2006). DNA-damage response pathways triggered by viral replication. *Expert Rev. Mol. Med.* 8, 1–11.

- Smith, G.A., and Enquist, L.W. (2002). Break ins and break outs: viral interactions with the cytoskeleton of Mammalian cells. *Annu. Rev. Cell Dev. Biol.* 18, 135–161.
- Smith, G.L., and Law, M. (2004). The exit of vaccinia virus from infected cells. *Virus Res.* 106, 189–197.
- Snyder, P.M., Olson, D.R., Kabra, R., Zhou, R., and Steines, J.C. (2004). cAMP and serum and glucocorticoid-inducible kinase (SGK) regulate the epithelial Na⁺ channel through convergent phosphorylation of Nedd4-2. *J. Biol. Chem.* 279, 45753–45758.
- Stoilov, P., Daoud, R., Nayler, O., and Stamm, S. (2004). Human tra2-beta 1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* 13, 509–524.
- Stokoe, D., Engel, K., Campbell, D.G., Cohen, P., and Gaestel, M. (1992). Identification of MAPKAP kinase 2 as a major enzyme responsible for the phosphorylation of the small mammalian heat shock proteins. *FEBS Lett.* 313, 307–313.
- Su, B., Cheng, J., Yang, J., and Guo, Z. (2001). MEKK2 is required for T-cell receptor signals in JNK activation and interleukin-2 gene expression. *J. Biol. Chem.* 276, 14784–14790.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Tacke, R., Tohyama, M., Ogawa, S., and Manley, J.L. (1998). Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93, 139–148.
- Tafforeau, L., Chantier, T., Pradezynski, F., Pellet, J., Mangeot, P.E., Vidalain, P.-O., Andre, P., Rabourdin-Combe, C., and Lotteau, V. (2011). Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network. *J. Virol.* 85, 13010–13018.
- Taggart, C., Cervantes-Laurean, D., Kim, G., McElvaney, N.G., Wehr, N., Moss, J., and Levine, R.L. (2000). Oxidation of either methionine 351 or methionine 358 in alpha 1-antitrypsin causes loss of anti-neutrophil elastase activity. *J. Biol. Chem.* 275, 27258–27265.
- Tan, Y.-J., Fielding, B.C., Goh, P.-Y., Shen, S., Tan, T.H.P., Lim, S.G., and Hong, W. (2004). Overexpression of 7a, a protein specifically encoded by the severe acute respiratory syndrome coronavirus, induces apoptosis via a caspase-dependent pathway. *J. Virol.* 78, 14043–14047.
- Tan, S.-X., Ng, Y., Burchfield, J.G., Ramm, G., Lambright, D.G., Stöckli, J., and James, D.E. (2012). The Rab GTPase-activating protein TBC1D4/AS160 contains an atypical phosphotyrosine-binding domain that interacts with plasma membrane phospholipids to facilitate GLUT4 trafficking in adipocytes. *Mol. Cell. Biol.* 32, 4946–4959.
- Tanenbaum, M.E., Gajjar, N., van Vugt, M.A.T.M., and Medema, R.H. (2006). CLIP-170 facilitates the formation of kinetochore-microtubule attachments. *EMBO J.* 25, 45–57.
- Tantos, A., Szrnka, K., Szabo, B., Bokor, M., Kamasa, P., Matus, P., Bekesi, A., Tompa, K., Han, K.-H., and Tompa, P. (2013). Structural disorder and local order of hNopp140. *Biochim. Biophys. Acta* 1834, 342–350.
- Tarakhovskiy, A., and Prinjha, R.K. (2018). Drawing on disorder: How viruses use histone mimicry to their advantage. *J. Exp. Med.* 215, 1777–1787.
- Tavares, M.R., Pavan, I.C.B., Amaral, C.L., Meneguello, L., Luchessi, A.D., and Simabuco, F.M. (2015). The S6K protein family in health and disease. *Life Sci.* 137, 1–10.
- Taylor, M.P., Koyuncu, O.O., and Enquist, L.W. (2011). Subversion of the actin cytoskeleton during viral infection. *Nat. Rev. Microbiol.* 9, 427–439.
- Tee, A.R., Fingar, D.C., Manning, B.D., Kwiatkowski, D.J., Cantley, L.C., and Blenis, J. (2002). Tuberous sclerosis complex-1 and -2 gene products function together to inhibit mammalian target of rapamycin (mTOR)-mediated downstream signaling. *Proc. Natl. Acad. Sci. USA* 99, 13571–13576.
- Thaker, S.K., Ch'ng, J., and Christofk, H.R. (2019). Viral hijacking of cellular metabolism. *BMC Biol.* 17, 59.
- Thejer, B.M., Adhikary, P.P., Kaur, A., Teakel, S.L., Van Oosterum, A., Seth, I., Pajic, M., Hannan, K.M., Pavy, M., Poh, P., et al. (2020). PGRMC1 phosphorylation affects cell shape, motility, glycolysis, mitochondrial form and function, and tumor growth. *BMC Mol. Cell Biol.* 21, 24.
- Tiedje, C., Lubas, M., Tehrani, M., Menon, M.B., Ronkina, N., Rousseau, S., Cohen, P., Kotlyarov, A., and Gaestel, M. (2015). p38MAPK/MK2-mediated phosphorylation of RBM7 regulates the human nuclear exosome targeting complex. *RNA* 21, 262–278.
- Tokuyama, Y., Hom, H.F., Kawamura, K., Tarapore, P., and Fukasawa, K. (2001). Specific phosphorylation of nucleophosmin on Thr(199) by cyclin-dependent kinase 2-cyclin E and its role in centrosome duplication. *J. Biol. Chem.* 276, 21529–21537.
- Toughiri, R., Li, X., Du, Q., and Bieberich, C.J. (2013). Phosphorylation of NuMA by Aurora-A kinase in PC-3 prostate cancer cells affects proliferation, survival, and interphase NuMA localization. *J. Cell. Biochem.* 114, 823–830.
- Trencia, A., Perletti, A., Cassese, A., Vigliotta, G., Miele, C., Oriente, F., Santopietro, S., Giacco, F., Condorelli, G., Formisano, P., and Beguinot, F. (2003). Protein kinase B/Akt binds and phosphorylates PED/PEA-15, stabilizing its antiapoptotic action. *Mol. Cell. Biol.* 23, 4511–4521.
- Trigg, B.J., and Ferguson, B.J. (2015). Functions of DNA damage machinery in the innate immune response to DNA virus infection. *Curr. Opin. Virol.* 15, 56–62.
- Tsai, S.-C., and Seto, E. (2002). Regulation of histone deacetylase 2 by protein kinase CK2. *J. Biol. Chem.* 277, 31826–31833.
- Tsai, L.-H., Lees, E., Faha, B., Harlow, E., and Riabowol, K. (1993). The cdk2 kinase is required for the G1-to-S transition in mammalian cells. *Oncogene* 8, 1593–1602.
- Tsoi, H., Li, L., Chen, Z.S., Lau, K.-F., Tsui, S.K.W., and Chan, H.Y.E. (2014). The SARS-coronavirus membrane protein induces apoptosis via interfering with PDK1-PKB/Akt signalling. *Biochem. J.* 464, 439–447.
- Tsvetkov, L., and Stern, D.F. (2005). Phosphorylation of Plk1 at S137 and T210 is inhibited in response to DNA damage. *Cell Cycle* 4, 166–171.
- Ui, A., Nagaura, Y., and Yasui, A. (2015). Transcriptional elongation factor ENL phosphorylated by ATM recruits polycomb and switches off transcription for DSB repair. *Mol. Cell* 58, 468–482.
- Ujike, M., and Taguchi, F. (2015). Incorporation of spike and membrane glycoproteins into coronavirus virions. *Viruses* 7, 1700–1725.
- Valvezan, A.J., and Klein, P.S. (2012). GSK-3 and Wnt Signaling in Neurogenesis and Bipolar Disorder. *Front. Mol. Neurosci.* 5, 1.
- Vind, B.F., Pehmoller, C., Trebak, J.T., Birk, J.B., Hey-Mogensen, M., Beck-Nielsen, H., Zierath, J.R., Wojtaszewski, J.F.P., and Hojlund, K. (2011). Impaired insulin-induced site-specific phosphorylation of TBC1 domain family, member 4 (TBC1D4) in skeletal muscle of type 2 diabetes patients is restored by endurance exercise-training. *Diabetologia* 54, 157–167.
- Vishwanatha, J.K., and Kumble, S. (1993). Involvement of annexin II in DNA replication: evidence from cell-free extracts of *Xenopus* eggs. *J. Cell Sci.* 105, 533–540.
- Wada, T., Nakagawa, K., Watanabe, T., Nishitai, G., Seo, J., Kishimoto, H., Kitagawa, D., Sasaki, T., Penninger, J.M., Nishina, H., and Katada, T. (2001). Impaired synergistic activation of stress-activated protein kinase SAPK/JNK in mouse embryonic stem cells lacking SEK1/MKK4: different contribution of SEK2/MKK7 isoforms to the synergistic activation. *J. Biol. Chem.* 276, 30892–30897.
- Walsh, D., and Mohr, I. (2011). Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microbiol.* 9, 860–875.
- Walter, B.L., Parsley, T.B., Ehrenfeld, E., and Semler, B.L. (2002). Distinct poly(rC) binding protein KH domain determinants for poliovirus translation initiation and viral RNA replication. *J. Virol.* 76, 12008–12022.
- Walter, D., Hoffmann, S., Komseli, E.-S., Rappsilber, J., Gorgoulis, V., and Sørensen, C.S. (2016). SCF(Cyclin F)-dependent degradation of CDC6 suppresses DNA re-replication. *Nat. Commun.* 7, 10530.
- Wang, Y., and Prives, C. (1995). Increased and altered DNA binding of human p53 by S and G2/M but not G1 cyclin-dependent kinases. *Nature* 376, 88–91.

- Wang, Y., and Zhang, X. (1999). The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 in vitro and in vivo. *Virology* 265, 96–109.
- Wang, A.H., Kruhlak, M.J., Wu, J., Bertos, N.R., Vezmar, M., Posner, B.I., Bazett-Jones, D.P., and Yang, X.J. (2000). Regulation of histone deacetylase 4 by binding of 14-3-3 proteins. *Mol. Cell. Biol.* 20, 6904–6912.
- Wang, W., Ye, L., Ye, L., Li, B., Gao, B., Zeng, Y., Kong, L., Fang, X., Zheng, H., Wu, Z., and She, Y. (2007). Up-regulation of IL-6 and TNF- α induced by SARS-coronavirus spike protein in murine macrophages via NF- κ B pathway. *Virus Res.* 128, 1–8.
- Wang, L., Harris, T.E., and Lawrence, J.C., Jr. (2008). Regulation of proline-rich Akt substrate of 40 kDa (PRAS40) function by mammalian target of rapamycin complex 1 (mTORC1)-mediated phosphorylation. *J. Biol. Chem.* 283, 15619–15627.
- Wang, T., Cleary, R.A., Wang, R., and Tang, D.D. (2013). Role of the adapter protein Abi1 in actin-associated signaling and smooth muscle contraction. *J. Biol. Chem.* 288, 20713–20722.
- Wang, Z., Lu, W., Zhang, Y., Zou, F., Jin, Z., and Zhao, T. (2020). The Hippo Pathway and Viral Infections. *Front. Microbiol.* 10, 3033.
- Ward, Y., Spinelli, B., Quon, M.J., Chen, H., Ikeda, S.R., and Kelly, K. (2004). Phosphorylation of critical serine residues in Gem separates cytoskeletal reorganization from down-regulation of calcium channel activity. *Mol. Cell. Biol.* 24, 651–661.
- Wen, Z., Zhong, Z., and Darnell, J.E., Jr. (1995). Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell* 82, 241–250.
- Wen, A.Y., Sakamoto, K.M., and Miller, L.S. (2010). The role of the transcription factor CREB in immune function. *J. Immunol.* 185, 6413–6419.
- Wertz, I.E., Newton, K., Seshasayee, D., Kusam, S., Lam, C., Zhang, J., Popovich, N., Helgason, E., Schoeffler, A., Jeet, S., et al. (2015). Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. *Nature* 528, 370–375.
- Willems, A.R., Lankester, S., Patton, E.E., Craig, K.L., Nason, T.F., Mathias, N., Kobayashi, R., Wittenberg, C., and Tyers, M. (1996). Cdc53 targets phosphorylated G1 cyclins for degradation by the ubiquitin proteolytic pathway. *Cell* 86, 453–463.
- Williams, R.S., Shohet, R.V., and Stillman, B. (1997). A human protein related to yeast Cdc6p. *Proc. Natl. Acad. Sci. USA* 94, 142–147.
- Wohlbold, L., Merrick, K.A., De, S., Amat, R., Kim, J.H., Larochelle, S., Allen, J.J., Zhang, C., Shokat, K.M., Petrini, J.H.J., and Fisher, R.P. (2012). Chemical genetics reveals a specific requirement for Cdk2 activity in the DNA damage response and identifies Nbs1 as a Cdk2 substrate in human cells. *PLoS Genet.* 8, e1002935.
- Wolin, S.L., and Cedervall, T. (2002). The La protein. *Annu. Rev. Biochem.* 71, 375–403.
- Wong, C.K., Lam, C.W.K., Wu, A.K.L., Ip, W.K., Lee, N.L.S., Chan, I.H.S., Lit, L.C.W., Hui, D.S.C., Chan, M.H.M., Chung, S.S.C., and Sung, J.J. (2004). Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. *Clin. Exp. Immunol.* 136, 95–103.
- Wortzel, I., and Seger, R. (2011). The ERK Cascade: Distinct Functions within Various Subcellular Organelles. *Genes Cancer* 2, 195–209.
- Wu, C.-H., Yeh, S.-H., Tsay, Y.-G., Shieh, Y.-H., Kao, C.-L., Chen, Y.-S., Wang, S.-H., Kuo, T.-J., Chen, D.-S., and Chen, P.-J. (2009). Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J. Biol. Chem.* 284, 5229–5239.
- Wu, X., Renuse, S., Sahasrabudhe, N.A., Zahari, M.S., Chaerkady, R., Kim, M.-S., Nirujogi, R.S., Mohseni, M., Kumar, P., Raju, R., et al. (2014). Activation of diverse signalling pathways by oncogenic PIK3CA mutations. *Nat. Commun.* 5, 4961.
- Wu, H., Deng, W.-W., Yang, L.-L., Zhang, W.-F., and Sun, Z.-J. (2018). Expression and phosphorylation of Stathmin 1 indicate poor survival in head and neck squamous cell carcinoma and associate with immune suppression. *Biomarkers Med.* 12, 759–769.
- Wurm, T., Chen, H., Hodgson, T., Britton, P., Brooks, G., and Hiscox, J.A. (2001). Localization of the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. *J. Virol.* 75, 9345–9356.
- Xu, L.H., Huang, M., Fang, S.G., and Liu, D.X. (2011). Coronavirus infection induces DNA replication stress partly through interaction of its nonstructural protein 13 with the p125 subunit of DNA polymerase δ . *J. Biol. Chem.* 286, 39546–39559.
- Yang, P., Guo, L., Duan, Z.J., Tepper, C.G., Xue, L., Chen, X., Kung, H.-J., Gao, A.C., Zou, J.X., and Chen, H.-W. (2012a). Histone methyltransferase NSD2/MMSET mediates constitutive NF- κ B signaling for cancer cell proliferation, survival, and tumor growth via a feed-forward loop. *Mol. Cell. Biol.* 32, 3121–3131.
- Yang, W., Zheng, Y., Xia, Y., Ji, H., Chen, X., Guo, F., Lyssiotis, C.A., Aldape, K., Cantley, L.C., and Lu, Z. (2012b). ERK1/2-dependent phosphorylation and nuclear translocation of PKM2 promotes the Warburg effect. *Nat. Cell Biol.* 14, 1295–1304.
- Yang, W., Xia, Y., Hawke, D., Li, X., Liang, J., Xing, D., Aldape, K., Hunter, T., Yung, W.K.A., and Lu, Z. (2014). PKM2 Phosphorylates Histone H3 and Promotes Gene Transcription and Tumorigenesis. *Cell* 158, 1210.
- Yao, K., Peng, C., Zhang, Y., Zykova, T.A., Lee, M.-H., Lee, S.-Y., Rao, E., Chen, H., Ryu, J., Wang, L., et al. (2017). RSK2 phosphorylates T-bet to attenuate colon cancer metastasis and growth. *Proc. Natl. Acad. Sci. USA* 114, 12791–12796.
- Ye, Z., Wong, C.K., Li, P., and Xie, Y. (2008). A SARS-CoV protein, ORF-6, induces caspase-3 mediated, ER stress and JNK-dependent apoptosis. *Biochim. Biophys. Acta* 1780, 1383–1387.
- Yee, A.S., Paulson, E.K., McDevitt, M.A., Rieger-Christ, K., Summerhayes, I., Berasi, S.P., Kim, J., Huang, C.-Y., and Zhang, X. (2004). The HBP1 transcriptional repressor and the p38 MAP kinase: unlikely partners in G1 regulation and tumor suppression. *Gene* 336, 1–13.
- Yim, H., Park, J.-W., Woo, S.U., Kim, S.-T., Liu, L., Lee, C.-H., and Lee, S.K. (2013). Phosphorylation of Cdc6 at serine 74, but not at serine 106, drives translocation of Cdc6 to the cytoplasm. *J. Cell. Physiol.* 228, 1221–1228.
- Yoo, S.-M., Lee, C.-J., An, H.-J., Lee, J.Y., Lee, H.S., Kang, H.C., Cho, S.-J., Kim, S.-M., Park, J., Kim, D.J., and Cho, Y.Y. (2019). RSK2-Mediated ELK3 Activation Enhances Cell Transformation and Breast Cancer Cell Growth by Regulation of c-fos Promoter Activity. *Int. J. Mol. Sci.* 20, 1994.
- Yu, W., Ding, X., Chen, F., Liu, M., Shen, S., Gu, X., and Yu, L. (2009). The phosphorylation of SEPT2 on Ser218 by casein kinase 2 is important to hepatoma carcinoma cell proliferation. *Mol. Cell. Biochem.* 325, 61–67.
- Yu, Y., Yoon, S.-O., Poulogiannis, G., Yang, Q., Ma, X.M., Villén, J., Kubica, N., Hoffman, G.R., Cantley, L.C., Gygi, S.P., and Blenis, J. (2011). Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science* 332, 1322–1326.
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.
- Yui, D., Yoneda, T., Oono, K., Katayama, T., Imaizumi, K., and Tohyama, M. (2001). Interchangeable binding of Bcl10 to TRAF2 and cIAPs regulates apoptosis signaling. *Oncogene* 20, 4317–4323.
- Zhang, Y., Sun, Z.W., Iratni, R., Erdjument-Bromage, H., Tempst, P., Hampsey, M., and Reinberg, D. (1998). SAP30, a novel protein conserved between human and yeast, is a component of a histone deacetylase complex. *Mol. Cell* 1, 1021–1031.
- Zhang, H.H., Huang, J., Düvel, K., Boback, B., Wu, S., Squillace, R.M., Wu, C.-L., and Manning, B.D. (2009). Insulin stimulates adipogenesis through the Akt-TSC2-mTORC1 pathway. *PLoS ONE* 4, e6189.
- Zhang, H., Diab, A., Fan, H., Mani, S.K.K., Hullinger, R., Merle, P., and Andriani, O. (2015). PLK1 and HOTAIR Accelerate Proteasomal Degradation of SUZ12 and ZNF198 during Hepatitis B Virus-Induced Liver Carcinogenesis. *Cancer Res.* 75, 2363–2374.



- Zhang, H., Wang, A., Tan, Y., Wang, S., Ma, Q., Chen, X., and He, Z. (2019a). NCBP1 promotes the development of lung adenocarcinoma through up-regulation of CUL4B. *J. Cell. Mol. Med.* **23**, 6965–6977.
- Zhang, Q., Sharma, N.R., Zheng, Z.-M., and Chen, M. (2019b). Viral Regulation of RNA Granules in Infected Cells. *Virology* **541**, 175–191.
- Zhang, B., Zhou, X., Qiu, Y., Feng, F., Feng, J., Jia, Y., Zhu, H., Hu, K., Liu, J., Liu, Z., et al. (2020). Clinical characteristics of 82 death cases with COVID-19. medRxiv. <https://doi.org/10.1101/2020.02.26.20028191>.
- Zhao, J., He, S., Minassian, A., Li, J., and Feng, P. (2015). Recent advances on viral manipulation of NF- κ B signaling pathway. *Curr. Opin. Virol.* **7**, 103–111.
- Zhou, Q., Wang, H., Schwartz, D.M., Stoffels, M., Park, Y.H., Zhang, Y., Yang, D., Demirkaya, E., Takeuchi, M., Tsai, W.L., et al. (2016). Loss-of-function mutations in TNFAIP3 leading to A20 haploinsufficiency cause an early-onset autoinflammatory disease. *Nat. Genet.* **48**, 67–73.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273.
- Zhu, Y., Doray, B., Poussu, A., Lehto, V.P., and Kornfeld, S. (2001). Binding of GGA2 to the lysosomal enzyme sorting motif of the mannose 6-phosphate receptor. *Science* **292**, 1716–1718.
- Zhu, G., Qiu, W., Li, Y., Zhao, C., He, F., Zhou, M., Wang, L., Zhao, D., Lu, Y., Zhang, J., et al. (2017). Sublytic C5b-9 Induces Glomerular Mesangial Cell Apoptosis through the Cascade Pathway of MEK2-p38 MAPK-IRF1-TRADD-Caspase 8 in Rat Thy-1 Nephritis. *J. Immunol.* **198**, 1104–1118.
- Zou, J.X., Revenko, A.S., Li, L.B., Gemo, A.T., and Chen, H.-W. (2007). ANCCA, an estrogen-regulated AAA+ ATPase coactivator for ERalpha, is required for coregulator occupancy and chromatin modification. *Proc. Natl. Acad. Sci. USA* **104**, 18067–18072.
- Zuo, P., and Manley, J.L. (1993). Functional domains of the human splicing factor ASF/SF2. *EMBO J.* **12**, 4727–4737.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Phospho-p38 MAPK (Thr180/Tyr182)	Cell Signaling	Cat#9211S
α-Rabbit IgG HRP Conjugate	Biorad	Cat#1706515
α-Mouse IgG HRP Conjugate	Biorad	Cat#1706516
SARS-CoV-2 N (recombinant mouse monoclonal)	Donated from the laboratory of Thomas Moran (Mount Sinai)	N/A
GAPDH	Cell Signaling	Cat#2118S
Phospho-MAPKAPK2 (T334)	Cell Signaling	Cat#3007T
Phospho-CREB/ATF-1 (S133)	Cell Signaling	Cat#9198S
CK2α	Abcam	Cat#ab70774
SARS-CoV membrane (M)	Rockland	Cat#100-401-A55
SARS-CoV nucleocapsid (N)	Rockland	Cat#200-401-A50
AF568-labeled goat-anti-rabbit	InvitrogenCat#	Cat#A11011
AF647-labeled goat-anti-mouse	Invitrogen	Cat#A21235
AF488-labeled Phalloidin	Hypermol	Cat#8813-01
viral NP protein anti-sera	produced in the Garcia-Sastre lab	N/A
Bacterial and Virus Strains		
SARS-CoV-2, isolate BetaCoV/France/IDF0372/2020	European Virus Archive goes Global (EVAg)	IDF0372/2020
SARS-CoV-2, isolate USA-WA1/2020	BEI Resources	Cat#NR-52281
SARS-CoV-2, Isolate Muc-IMB-1/2020	Bundeswehr Institute of Microbiology, Munich, Germany	N/A
Chemicals, Peptides, and Recombinant Proteins		
cOmplete protease inhibitor cocktail tablets mini, EDTA-free	Roche	Cat#11846170001
PhosStop phosphatase inhibitor cocktail tablets	Roche	Cat#4906837001
Sep-Pak C18 cartridge	Waters	Cat#WAT054955
Sequencing-grade modified trypsin	Promega	Cat#V5111
Fetal bovine serum (FBS)	Thermo Fisher Scientific	Cat#10082
DMEM	Thermo Fisher Scientific	Cat#MT10013CV
Water, HPLC grade	Sigma-Aldrich	Cat#270733-4 L; CAS#7732-18-5
Ni-NTA Superflow beads	QIAGEN	Cat#30210
Igepal	Sigma-Aldrich	Cat#I3021; CAS#9002-93-1
Minimal Essential Media (MEM)	Corning	Cat#10-009-CV
Opti-MEM	Thermo Fisher Scientific	Cat#31985062
Paraformaldehyde, 4% solution in PBS (PFA)	Thermo Scientific	Cat#MFCD00133991
Formalin	Fisher Scientific	Cat#SF100-20; CAS#50-00-0, 67-56-1, 7558-79-4, 10049-21-5, 7732-18-5
4',6-diamidino-2-phenylindole (DAPI)	Thermo Scientific	Cat#62247
ReprosilPur 1.9 μm particles	ESI SOURCE SOLUTIONS	Cat#R119
Crystal Violet Solution	Sigma Aldrich	Cat#HT90132-1L
Thermanox coverslips	Ted Pella	Cat#26028
Silicon Chips	Ted Pella	Cat#16007

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Aluminum specimen mounts	Ted Pella	Cat#16111
Double-sided carbon tape	Ted Pella	Cat#16084-1
Karnovsky's EM fixative	Electron Microscopy Sciences	Cat#15720
Sodium Cacodylate	Sigma	Cat#C4945-10G; CAS#6131-99-3
Osmium Tetroxide	Electron Microscopy Sciences	Cat#19190; CAS#20816-12-0
Potassium Ferrocyanide	Sigma	Cat#P-3289; CAS#14459-95-1
Uranyl Acetate	Ted Pella	Cat#19481; CAS#6159-44-0
Spurr's resin	Ted Pella	Cat#18300-4221
Iridium target	Electron Microscopy Sciences	Cat#3431
PolyJet	SignaGen	Cat#SL100688
SuperScript IV Reverse Transcriptase	Invitrogen	Cat#18090010
ARRY-371797 (ARRY-797)	MedKoo Biosciences	Cat#555466; CAS#1036404-17-7
Dilmapimod (SB-681323)	MedChemExpress	Cat#HY-10404; CAS#444606-18-2
Doramapimod (BIRB-796)	SelleckChem	Cat#S1574; CAS#285983-48-4
Losmapimod	SelleckChem	Cat#S7215; CAS#585543-15-3
MAPK13-IN-1	MedChemExpress	Cat#HY-18850; CAS#229002-10-2
Neflamapimod (VX-745)	SelleckChem	Cat#S1458; CAS#209410-46-8
P38 MAPK Inhibitor (4)	Cayman Chemical	Cat#22219; CAS#1638-41-1
PF-3644022	Tocris Bioscience	Cat#4483; CAS#142557-61-7
PH-797804	SelleckChem	Cat#S2726; CAS#586379-66-0
Ralimetinib (LY2228820)	SelleckChem	Cat#S1494; CAS#862507-23-1
SB203580	SelleckChem	Cat#S1076; CAS#152121-47-6
TAK-715	SelleckChem	Cat#S2928; CAS#303162-79-0
Talmapimod (SCIO-469)	MedChemExpress	Cat#HY-10406; CAS#309913-83-5
VX-702	SelleckChem	Cat#S6005; CAS#745833-23-2
4EGI-1	SelleckChem	Cat#S7369; CAS#315706-13-9
A-484954	Tocris Bioscience	Cat#4279; CAS#1276121-88-0
Abemaciclib mesylate	SelleckChem	Cat#S7158; CAS#1231930-82-7
Alisertib	SelleckChem	Cat#S1133; CAS#1028486-01-2
Apilimod	SelleckChem	Cat#S6414; CAS#541550-19-0
AT-13148	SelleckChem	Cat#S7563; CAS#1056901-62-2
AZD-1208	SelleckChem	Cat#S7104; CAS#1204144-28-4
AZD-5363 (Capivasertib)	SelleckChem	Cat#S8019; CAS#1143532-39-1
AZD-7648	SelleckChem	Cat#S8843; CAS#2230820-11-6
Bafetinib	SelleckChem	Cat#S1369; CAS#859212-16-1
BMS-582949	SelleckChem	Cat#S8124; CAS#623152-17-0
Bortezomib (PS-341)	SelleckChem	Cat#S1013; CAS#179324-69-7
CHIR-98014	SelleckChem	Cat#S2745; CAS#252935-94-7
Dabrafenib (GSK2118436)	SelleckChem	Cat#S2807; CAS#1195765-45-7
Dasatinib	SelleckChem	Cat#S1021; CAS#302962-49-8
Dinaciclib	SelleckChem	Cat#S2768; CAS#779353-01-4
Dorsomorphin 2HCl	SelleckChem	Cat#S7306; CAS#1219168-18-9
EHT 1864 2HCl	SelleckChem	Cat#S7482; CAS#754240-09-0
Enzastaurin	SelleckChem	Cat#S1055; CAS#170364-57-5
Fludarabine	SelleckChem	Cat#S1491; CAS#S1491
Fosbretabulin (Combretastatin A4 Phosphate (CA4P)) Disodium	SelleckChem	Cat#S7204; CAS#168555-66-6
Ganetespib	SelleckChem	Cat#S1159; CAS#888216-25-9

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Gilteritinib	SelleckChem	Cat#S7754; CAS#1254053-43-4
Idarubicin HCl	SelleckChem	Cat#S1228; CAS#57852-57-0
Ipatasertib	SelleckChem	Cat#S2808; CAS#1001264-89-6
Ispinesib (SB-715992)	SelleckChem	Cat#S1452; CAS#336113-53-2
Ixazomib (MLN2238)	SelleckChem	Cat#S2180; CAS#1072833-77-2
K-252a	Cayman Chemical	Cat#11338; CAS#99533-80-9
KenPaullone	SelleckChem	Cat#S7917; CAS#142273-20-9
KW-2449	SelleckChem	Cat#S2158; CAS#1000669-72-6
LJ1308	SelleckChem	Cat#S7871; CAS#1627709-94-7
LMK-235	SelleckChem	Cat#S7569; CAS#1418033-25-6
LY2584702 Tosylate	SelleckChem	Cat#S7704; CAS#1082949-68-5
MBQ-167	SelleckChem	Cat#S8749; CAS#2097938-73-1
Mdivi-1	SelleckChem	Cat#S7162; CAS#338967-87-6
Midostaurin	SelleckChem	Cat#S8064; CAS#120685-11-2
MK-2206 2HCl	SelleckChem	Cat#S1078; CAS#1032350-13-2
ML141	SelleckChem	Cat#S7686; CAS#71203-35-5
NSC319276	SelleckChem	Cat#S7149; CAS#71555-25-4
NVP-BEP800	SelleckChem	Cat#S1498; CAS#847559-80-2
Palbociclib (PD-0332991) HCl	SelleckChem	Cat#S1116; CAS#827022-32-2
PF-03814735	SelleckChem	Cat#S2725; CAS#942487-16-3
PF-3758309	SelleckChem	Cat#S7094; CAS#898044-15-0
PHA-793887	SelleckChem	Cat#S1487; CAS#718630-59-2
Pictilisib	SelleckChem	Cat#S1065; CAS#957054-30-7
PKR Inhibitor	Cayman Chemical	Cat#15323; CAS#608512-97-6
QNZ (EVP4593)	SelleckChem	Cat#S4902; CAS#545380-34-5
R-406	SelleckChem	Cat#S2194; CAS#841290-81-1
R-547	SelleckChem	Cat#S2688; CAS#741713-40-6
Rapamycin	SelleckChem	Cat#S1039; CAS#53123-88-9
Ravoxertinib	SelleckChem	Cat#S7554; CAS#1453848-26-4
Remdesivir	BioVision	Cat#B2997; CAS#1809249-37-3
Ripasudil (K-115) hydrochloride dihydrate	SelleckChem	Cat#S7995; CAS#887375-67-9
Romidepsin (FK228, Depsipeptide)	SelleckChem	Cat#S3020; CAS#128517-07-7
Ruboxistaurin HCl(LY333531)	SelleckChem	Cat#S7663; CAS#169939-93-9
Sapanisertib	SelleckChem	Cat#S2811; CAS#1224844-38-5
SAR-407899	Cayman Chemical	Cat#21717; CAS#923359-38-0
SB743921 HCl	SelleckChem	Cat#S2182; CAS#940929-33-9
SGL-1776 free base	SelleckChem	Cat#S2198; CAS#1025065-69-3
Silmitasertib	SelleckChem	Cat#S2248; CAS#1009820-21-6
Sorafenib Tosylate	SelleckChem	Cat#S1040; CAS#475207-59-1
Sotrastaurin	SelleckChem	Cat#S2791; CAS#425637-18-9
Stattic	SelleckChem	Cat#S7024; CAS#19983-44-9
Staurosporine	SelleckChem	Cat#S1421; CAS#62996-74-1
Tanzisertib	SelleckChem	Cat#S8490; CAS#899805-25-5(freebase)
Tasquinimod	SelleckChem	Cat#S7617; CAS#254964-60-8
THZ1 2HCl	SelleckChem	Cat#S7549; CAS#1604810-83-4(freebase)
Tp12 Kinase Inhibitor 1	Cayman Chemical	Cat#19710; CAS#871307-18-5
U73122	SelleckChem	Cat#S8011; CAS#112648-68-7
Volasertib	SelleckChem	Cat#S2235; CAS#755038-65-4

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
XL413 (BMS-863233)	SelleckChem	Cat#S7547; CAS#1169562-71-3
Y-39983 HCl	SelleckChem	Cat#S7935; CAS#173897-44-4
CCL2/MCP-1, Capture Antibody	R&D Systems	Cat#MAB679
CCL5/RANTES, Capture Antibody	R&D Systems	Cat#MAB678
CCL11/Eotaxin, Capture Antibody	R&D Systems	Cat#MAB320
GM-CSF, Capture Antibody	R&D Systems	Cat#MAB615
IL-1RA, Capture Antibody	Peprtech	Cat#500-P209
CCL13/MCP-4, Capture Antibody	R&D Systems	Cat#MAB327
IL-4, Capture Antibody	R&D Systems	Cat#MAB604
IL-10, Capture Antibody	Peprtech	Cat#500-P20
IL-15, Capture Antibody	R&D Systems	Cat#MAB647
IL-1beta, Capture Antibody	R&D Systems	Cat#MAB601
CCL8/MCP-2, Capture Antibody	Peprtech	Cat#500-P35
CXCL9/MIG, Capture Antibody	Peprtech	Cat#500-P50
CCL3/MIP-1alpha, Capture Antibody	R&D Systems	Cat#MAB670
CXCL2/GRO-Beta, Capture Antibody	Peprtech	Cat#500-P104
TGF-Beta1, Capture Antibody	R&D Systems	Cat#MAB240
CXCL16, Capture Antibody	Peprtech	Cat#500-P200
CXCL8/IL-8, Capture Antibody	R&D Systems	Cat#MAB208
MMP-9, Capture Antibody	R&D Systems	Cat#MAB936
CXCL1/GRO-alpha, Capture Antibody	Peprtech	Cat#500-P92
IL-7, Capture Antibody	R&D Systems	Cat#MAB207
CCL4/MIP-1beta, Capture Antibody	R&D Systems	Cat#MAB271
IL-1alpha, Capture Antibody	R&D Systems	Cat#MAB200
IL-12 (p70), Capture Antibody	Peprtech	Cat#500-P154G
IL-16, Capture Antibody	R&D Systems	Cat#MAB316
CCL20/MIP-3a, Capture Antibody	R&D Systems	Cat#MAB360
TNF-alpha, Capture Antibody	R&D Systems	Cat#MAB610
Trappin-2/Elafin, Capture Antibody	R&D Systems	Cat#AF1747
IL-17A, Capture Antibody	R&D Systems	Cat#MAB317
CXCL10/IP-10, Capture Antibody	R&D Systems	Cat#MAB266
S100A8, Capture Antibody	R&D Systems	Cat#DS8900
IL-22, Capture Antibody	Peprtech	Cat#500-P211
CXCL5/ENA-78, Capture Antibody	R&D Systems	Cat#MAB654
IL-6, Capture Antibody	R&D Systems	Cat#MAB206
CCL7/MCP-3, Capture Antibody	R&D Systems	Cat#MAB282
CCL2/MCP-1, Protein std	R&D Systems	Cat#279-MC
CCL5/RANTES, Protein std	R&D Systems	Cat#278-RN
CCL11/Eotaxin, Protein std	Peprtech	Cat#300-21
GM-CSF, Protein std	R&D Systems	Cat#215-GM
IL-1RA, Protein std	Peprtech	Cat#200-01RA
CCL13/MCP-4, Protein std	R&D Systems	Cat#327-P4
IL-4, Protein std	R&D Systems	Cat#204-IL
IL-10, Protein std	Peprtech	Cat#200-10
IL-15, Protein std	R&D Systems	Cat#247-ILB
IL-1beta, Protein std	R&D Systems	Cat#201-LB
CCL8/MCP-2, Protein std	Peprtech	Cat#300-15
CXCL9/MIG, Protein std	R&D Systems	Cat#392-MG

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CCL3/MIP-1alpha, Protein std	R&D Systems	Cat#270-LD
CXCL2/GRO-Beta, Protein std	Peprtech	Cat#300-39
TGF-Beta1, Protein std	Peprtech	Cat#100-21
CXCL16, Protein std	Peprtech	Cat#300-55
CXCL8/IL-8, Protein std	R&D Systems	Cat#208-IL
MMP-9, Protein std	R&D Systems	Cat#911-MP
CXCL1/GRO-alpha, Protein std	Peprtech	Cat#300-11
IL-7, Protein std	R&D Systems	Cat#207-IL
CCL4/MIP-1beta, Protein std	R&D Systems	Cat#271-BME
IL-1alpha, Protein std	Peprtech	Cat#200-01A
IL-12(p70), Protein std	R&D Systems	Cat#219-IL
IL-16, Protein std	Peprtech	Cat#200-16
CCL20/MIP-3a, Protein std	Peprtech	Cat#300-29A
TNF-alpha, Protein std	R&D Systems	Cat#210-TA
Trappin-2/Elafin, Protein std	R&D Systems	Cat#1747-PI
IL-17A, Protein std	Peprtech	Cat#200-17
CXCL10/IP-10, Protein std	R&D Systems	Cat#266-IP
S100A8, Protein std	R&D Systems	Cat#DS8900
IL-22, Protein std	Peprtech	Cat#200-22
CXCL5/ENA-78, Protein std	R&D Systems	Cat#254-XB
IL-6, Protein std	R&D Systems	Cat#206-IL
CCL7/MCP-3, Protein std	R&D Systems	Cat#282-P3
CCL2/MCP-1, Detection Antibody	R&D Systems	Cat#BAF279
CCL5/RANTES, Detection Antibody	R&D Systems	Cat#BAF278
CCL11/Eotaxin, Detection Antibody	R&D Systems	Cat#BAF320
GM-CSF, Detection Antibody	R&D Systems	Cat#BAM 215
IL-1RA, Detection Antibody	Peprtech	Cat#500-P209BT
CCL13/MCP-4, Detection Antibody	R&D Systems	Cat#BAF327
IL-4, Detection Antibody	R&D Systems	Cat#BAF204
IL-10, Detection Antibody	Peprtech	Cat#500-P20BT
IL-15, Detection Antibody	R&D Systems	Cat#BAM 247
IL-1beta, Detection Antibody	R&D Systems	Cat#BAF201
CCL8/MCP-2, Detection Antibody	Peprtech	Cat#500-P35BT
CXCL9/MIG, Detection Antibody	Peprtech	Cat#500-P50BT
CCL3/MIP-1alpha, Detection Antibody	R&D Systems	Cat#BAF270
CXCL2/GRO-Beta, Detection Antibody	Peprtech	Cat#500P104BT
TGF-Beta1, Detection Antibody	R&D Systems	Cat#BAM 2462
CXCL16, Detection Antibody	Peprtech	Cat#500-P200BT
CXCL8/IL-8, Detection Antibody	R&D Systems	Cat#BAF208
MMP-9, Detection Antibody	R&D Systems	Cat#BAF911
CXCL1/GRO-alpha, Detection Antibody	Peprtech	Cat#500P92BT
IL-7, Detection Antibody	R&D Systems	Cat#BAF207
CCL4/MIP-1beta, Detection Antibody	R&D Systems	Cat#BAF271
IL-1alpha, Detection Antibody	R&D Systems	Cat#BAF200
IL-12(p70), Detection Antibody	Peprtech	Cat#500-P154GBT
IL-16, Detection Antibody	R&D Systems	Cat#BAF316
CCL20/MIP-3a, Detection Antibody	R&D Systems	Cat#BAF360
TNF-alpha, Detection Antibody	R&D Systems	Cat#BAF210

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Trappin-2/Elafin, Detection Antibody	R&D Systems	Cat#BAF1747
IL-17A, Detection Antibody	R&D Systems	Cat#BAF317
CXCL10/IP-10, Detection Antibody	Peprtech	Cat#500-P93BT
S100A8, Detection Antibody	R&D Systems	Cat#DS8900
IL-22, Detection Antibody	Peprtech	Cat#500-P211BT
CXCL5/ENA-78, Detection Antibody	R&D Systems	Cat#BAF254
IL-6, Detection Antibody	R&D Systems	Cat#BAF206
CCL7/MCP-3, Detection Antibody	R&D Systems	Cat#BAF282
Critical Commercial Assays		
MTT assay	Roche	Cat#11465007001
Luna Universal One-Step RT-qPCR Kit	New England Biolabs	Cat#E3005L
CellTiter-Glo Luminescent Cell Viability Assay	Promega	Cat#G7570
Direct-zol RNA Miniprep Plus kit	Zymo Research	Cat#R2072
KAPA SYBR FAST qPCR Master Mix	Roche	Cat#07959389001
Bio-Rad II Protein Assay Kit	Bio-Rad	Cat#5000002
Deposited Data		
Phosphoproteomics data of SARS-CoV-2 infected Vero E6 cells	This paper	PRIDE Project ID: PXD019113
Supplemental Tables in Mendeley Data	This paper	https://dx.doi.org/10.17632/dpkbh2g9hy.1
CORUM v3.0	Giurgiu et al., 2019	http://mips.helmholtz-muenchen.de/corum/
RNA-seq dataset for transcription factor activities	Blanco-Melo et al., 2020	GSE147507
Phosphoproteomics datasets of biological conditions	Ochoa et al., 2016	https://www.embopress.org/doi/full/10.15252/msb.20167295
Functional scores for phosphorylation sites	Ochoa et al., 2020	https://dp.nature.com/authorize?response_type=cookie&client_id=grover&redirect_uri=https%3A%2F%2Fwww.nature.com%2Farticles%2F41587-019-0344-3
Gene identifier mapping database from BioMart	Ensembl	http://uswest.ensembl.org/biomart/martview/5ccd994c794265409f21c6194730ce5b
Molecular Signatures Database for gene set enrichment analysis	Subramanian et al., 2005	https://www.gsea-msigdb.org/gsea/msigdb/index.jsp
Kinase-substrate relationships	Bachman et al., 2019	https://www.biorxiv.org/content/10.1101/822668v3.full
<i>Chlorocephus</i> proteome sequences for peptide searches	Uniprot	https://www.uniprot.org/proteomes/UP000029965
Sars-CoV-2 genomic sequences (hCoV-19/France/IDF0372/2020)	GISAID	Accession number EPI_ISL_406596
Experimental Models: Cell Lines		
ACE2-A549 human lung carcinoma	Donated from the laboratory of Brad Rosenberg	N/A
Caco-2	ATCC	Cat#HTB-37
VERO C1008 [Vero 76, clone E6, Vero E6] (Vero E6)	ATCC	Cat#CRL-1586
HEK293T/17 cells	ATCC	Cat#CRL-11268

(Continued on next page)



Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
ON-TARGETplus Human MAPK12 (6300) siRNA - SMARTpool, 5 nmol	Horizon Discovery/Dharmacon	Cat#L-003590-00-0005
ON-TARGETplus Human MAPK13 (5603) siRNA - SMARTpool, 5 nmol	Horizon Discovery/Dharmacon	Cat#L-003591-00-0005
ON-TARGETplus Human MAP2K3 (5606) siRNA - SMARTpool, 5 nmol	Horizon Discovery/Dharmacon	Cat#L-003509-00-0005
ON-TARGETplus Human ACE2 (59272) siRNA - SMARTpool	Horizon Discovery/Dharmacon	Cat#L-005755-00-0005
N gene region (reverse) 5'-CGAAGGTGTGACTTCCATG-3'	Eurofins	N/A
N gene region (forward) 5'-TAATCAGACAAGGAACTGATTA-3'	Eurofins	N/A
Recombinant DNA		
pLVX-TetOne-Puro empty vector	Takara	Cat#631847
pLVX-TetOne-Puro-SARS-CoV-2-N-2xStrep	This paper	N/A
pMD2.G, Addgene	Gift from Didier Trono	Addgene Plasmid #12259
Gag-Pol-Tat-Rev packaging construct	Gift from Judd Hultquist	pJH045
Software and Algorithms		
artMS	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/artMS.html
MSstats	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/artMS.html
The R Project for Statistical Computing	R Core Team, 2019	http://www.r-project.org/index.html
Spectronaut	Biognosys	https://biognosys.com/shop/spectronaut
ZEN (Blue edition)	Carl Zeiss Microscopy	https://www.zeiss.com/zen
SPPIDER v2	Porollo and Meller, 2007	http://sppider.cchmc.org/
DoRothEA	Garcia-Alonso et al., 2019	https://saezlab.github.io/dorothea/
VIPER	Alvarez et al., 2016	https://www.bioconductor.org/packages/release/bioc/html/viper.html
Biostrings	Bioconductor	https://bioconductor.org/packages/release/bioc/html/Biostrings.html
clusterProfiler	Bioconductor	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
DESeq2 package	Love et al., 2014	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
Imaris 64x 9.5.1	Oxford Instruments	https://imaris.oxinst.com/
MetaMorph (Version 7.8)	Molecular Devices	https://www.moleculardevices.com/
Other		
Orbitrap Exploris 480 MS with internal calibration option	Thermo Fisher Scientific	Cat#BRE725533
ZEISS LSM 800 Confocal Laser Scanning Microscope	Carl Zeiss Microscopy	Model#ZEISS LSM 800
QuantStudio 6 Real-Time PCR System	Applied Biosystems	Cat#4485697
FACSymphony	BD Life Sciences	N/A
Infinity 2000 Plate Reader	Tecan	N/A
Luminex MAGPIX platform	Luminex Corp	N/A
LightCycler 480 Instrument II	Roche	N/A
Bal-Tec Drier	Balzers, Liechtenstein	Cat#CPD 030

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Quorum sputter coater	Electron Microscopy Sciences, Hatfield, PA	Cat#EMS300T D
Hitachi field emission scanning electron microscope	Hitachi, Tokyo, Japan	Model#SU-8000
Leica UC7 ultramicrotome	Leica Microsystems	N/A
FEI BT Tecnai transmission electron microscope	ThermoFisher/FEI	N/A
Gatan Rio camera	Gatan	N/A
Celigo imaging cytometer	Nexcelcom	N/A

RESOURCE AVAILABILITY**Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by Nevan J. Krogan (Nevan.Krogan@ucsf.edu).

Materials Availability

Plasmid pLVX-TetOne-Puro-SARS-CoV-2-N-2xStrep is available upon request from the Lead Contact.

Data and Code Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD019113](https://doi.org/10.1093/bioinformatics/btad001) (Perez-Riverol et al., 2019). An interactive version of phosphorylation data can be found at <https://kroganlab.ucsf.edu/network-maps>. Supplementary tables have been deposited to Mendeley Data: <https://dx.doi.org/10.17632/dpkbh2g9hy.1>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Cells**

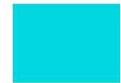
African green monkey kidney epithelial Vero E6 (Vero 76, clone E6, Vero E6, ATCC® CRL-1586) authenticated by ATCC and tested negative for mycoplasma contamination prior to commencement were maintained in a humidified atmosphere at 37°C with 5% CO₂, in Dulbecco's modified Eagle's medium (DMEM) containing 10% (v/v) fetal bovine serum (FBS, Invitrogen). ACE2-expressing A549 cells, a human lung epithelial cell line, were a kind gift from Brad Rosenberg. A549-ACE2 cells were cultured in DMEM supplemented with 10% FBS and maintained at 37°C with 5% CO₂. Caco-2 (ATCC® HTB-37) human colon epithelial cells were maintained in a humidified atmosphere at 37°C with 5% CO₂ in DMEM-containing 20% (v/v) FBS.

Viruses (Institut Pasteur, Paris)

The SARS-CoV-2 isolate BetaCoV/France/IDF0372/2020 was supplied by the National Reference Centre for Respiratory Viruses hosted by Institut Pasteur (Paris, France) and headed by Pr. Sylvie van der Werf. The isolate originated from a human sample provided by Dr. X. Lescure and Pr. Y. Yazdanpanah from the Bichat Hospital, Paris, France. The isolate was supplied through the European Virus Archive goes Global (EVAg) platform. Viral stocks were prepared by propagation in Vero E6 cells in DMEM supplemented with 2% FBS. Viral titers were determined by plaque assay in Minimum Essential Media (MEM) supplemented with 2% (v/v) FBS (Invitrogen) and 0.05% agarose. All experiments involving live SARS-CoV-2 were performed in compliance with Institut Pasteur Paris's guidelines for Biosafety Level 3 (BSL-3) containment procedures in approved laboratories. All experiments were performed in at least three biologically independent samples.

Viruses (Mount Sinai, New York)

For infection experiments in ACE2-A549 cells, SARS-CoV-2, isolate USA-WA1/2020 (NR-52281), which shares 99.983% sequence identity with the BetaCoV/France/IDF0372/2020 isolate, was deposited by the Center for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH. SARS-CoV-2 was propagated in Vero E6 cells in DMEM supplemented with 2% FBS, 4.5 g/L D-glucose, 4 mM L-glutamine, 10 mM Non-Essential Amino Acids, 1 mM Sodium Pyruvate and 10 mM HEPES as described previously (Blanco-Melo et al., 2020). All work involving live SARS-CoV-2 was performed in the CDC/USDA-approved BSL-3 facility of the Global Health and Emerging Pathogens Institute at the Icahn School of Medicine at Mount Sinai in accordance with institutional biosafety requirements.



Viruses (University of Freiburg, Germany)

For infection experiments in Caco-2 cells, SARS-CoV-2 isolate Muc-IMB-1/2020, kindly provided by the Bundeswehr Institute of Microbiology, Munich, Germany, was used. SARS-CoV-2 was propagated in Vero E6 cells in DMEM supplemented with 2% FBS. All work involving live SARS-CoV-2 was performed in the BSL-3 facility of the Institute of Virology, University Hospital Freiburg, and was approved according to the German Act of Genetic Engineering by the local authority (Regierungspraesidium Tuebingen, permit UNI.FRK.05.16/05).

METHODS DETAILS

Vero E6 cell infection for proteomic analysis

Vero E6 cells were seeded using 2×10^6 cells in T25 flasks. The following day cells were either mock infected or infected with SARS-CoV-2 at a MOI of 1 in serum-free DMEM at 37°C for 1 hour. After absorption the 0 hour samples were lysed immediately, while the media for other samples was replaced with 2% FBS / DMEM (Invitrogen) and incubated at 37°C for times indicated before lysis.

Cell lysis and digestion

Cells were lysed using 1% IGEPAL (Sigma) in PBS (Invitrogen) for 20 minutes at room temperature (RT) to inactivate the virus. These specific lysis conditions were used as this was the approved virus inactivation protocol. Proteins contained in the cell lysate were then immediately precipitated using 90% methanol (v/v) (Sigma) by centrifugation at 20,000x g for 10 minutes. The protein pellets were frozen at -80°C. Precipitated proteins were resuspended in lysis buffer (8 M urea, 100 mM ammonium bicarbonate (ABC), 150 mM NaCl, protease inhibitor (mini-cOmplete, Roche) and phosphatase inhibitors (phosSTOP, Roche). Tris-(2-carboxyethyl) phosphine (TCEP) was added to a final concentration of 4 mM. DNA was sheared via probe sonication, on ice, at 20% amplitude for 20 s., followed by 10 s of rest. This process was performed a total of three times. Following sonication, protein concentration was determined using Bradford assay. Iodoacetamide (IAA) was added to each sample to a final concentration of 10 mM, and samples were incubated in the dark at room temperature (RT) for 30 minutes. Excess IAA was quenched by the addition of dithiothreitol (DTT) to 10 mM, followed by incubation in the dark at RT for 30 minutes. Samples were then diluted with 0.1 M ABC (pH = 8.0) to a final urea concentration of 2 M. Trypsin (Promega) was added at a 1:100 (enzyme:protein w:w) ratio and digested overnight at 37°C with rotation. Following digestion, 10% trifluoroacetic acid (TFA) was added to each sample to a final pH ~2. Samples were desalted under vacuum using Sep Pak tC18 cartridges (Waters). Each cartridge was activated with 1 mL 80% acetonitrile (ACN)/0.1% TFA, then equilibrated with 3 × 1 mL of 0.1% TFA. Following sample loading, cartridges were washed with 4 × 1 mL of 0.1% TFA, and samples were eluted with 4 × 0.5 mL 50% ACN/0.25% formic acid (FA). 20 µg of each sample was kept for protein abundance measurements, and the remainder was used for phosphopeptide enrichment. Samples were dried by vacuum centrifugation.

Phosphopeptide enrichment

For each sample batch, 400 µL (30 µL per sample) of 50% Superflow bead slurry (QIAGEN) was added to a 2 mL bio-spin column. Beads were incubated with 4 × 500 µL of 100 mM EDTA for 30 s, washed with 2 × 500 µL H₂O, incubated 4 × 500 µL with 15 mM FeCl₃ for 1 minute, washed 3 × 500 µL H₂O, and washed once with 500 µL of 0.5% FA to remove residual Fe. Beads were resuspended in 600 µL of H₂O and 60 µL were aliquoted into a C18 NEST column that had been equilibrated with 150 µL of 80% ACN, 0.1% TFA. 1 mg of digested peptides were resuspended in 75% ACN/0.15% TFA and incubated with beads for 2 minutes, mixed by pipetting and incubated again for 2 minutes. Beads were washed 4 × 200 µL with 80% ACN, 0.1% TFA, then washed 3 × 200 µL with 0.5% FA, incubated 2 × 200 µL with 500 mM potassium phosphate buffer pH 7 and incubated 2 × 200 µL with 0.5% FA for 15 seconds. Phosphopeptides were eluted by centrifugation at 3000 RPM for 30 seconds with 2 × 75 µL of 50% ACN, 0.1% FA.

Mass spectrometry data acquisition

Digested samples were analyzed on an Orbitrap Exploris 480 mass spectrometry system (Thermo Fisher Scientific) equipped with an Easy nLC 1200 ultra-high pressure liquid chromatography system (Thermo Fisher Scientific) interfaced via a Nanospray Flex nano-electrospray source. For all analyses, samples were injected on a C18 reverse phase column (25 cm × 75 µm packed with ReprosilPur 1.9 µm particles). Mobile phase A consisted of 0.1% FA, and mobile phase B consisted of 0.1% FA/80% ACN. Peptides were separated by an organic gradient from 5% to 30% mobile phase B over 112 minutes followed by an increase to 58% B over 12 minutes, then held at 90% B for 16 minutes at a flow rate of 350 nL/minute. Analytical columns were equilibrated with 6 µL of mobile phase A. To build a spectral library, one sample from each set of biological replicates was acquired in a data dependent manner. Data dependent analysis (DDA) was performed by acquiring a full scan over a m/z range of 400-1000 in the Orbitrap at 60,000 resolving power (@200 m/z) with a normalized AGC target of 300%, an RF lens setting of 40%, and a maximum ion injection time of 60 ms. Dynamic exclusion was set to 60 seconds, with a 10 ppm exclusion width setting. Peptides with charge states 2-6 were selected for MS/MS interrogation using higher energy collisional dissociation (HCD), with 20 MS/MS scans per cycle. For phosphopeptide enriched samples, MS/MS scans were analyzed in the Orbitrap using isolation width of 1.3 m/z, normalized HCD collision energy of 30%, normalized AGC of 200% at a resolving power of 30,000 with a 54 ms maximum ion injection time. Similar settings were used for data dependent analysis of samples used to determine protein abundance, with an MS/MS resolving power of 15,000 and a 22 ms maximum ion injection time. Data-independent analysis (DIA) was performed on all samples. An MS scan at 60,000 resolving power over a scan

range of 390-1010 m/z, a normalized AGC target of 300%, an RF lens setting of 40%, and a maximum injection time of 60 ms was acquired, followed by DIA scans using 8 m/z isolation windows over 400-1000 m/z at a normalized HCD collision energy of 27%. Loop control was set to All. For phosphopeptide enriched samples, data were collected using a resolving power of 30,000 and a maximum ion injection time of 54 ms. Protein abundance samples were collected using a resolving power of 15,000 and a maximum ion injection time of 22 ms.

Spectral library generation and raw data processing

Raw mass spectrometry data from each DDA dataset were used to build separate libraries for DIA searches using the Pulsar search engine integrated into Spectronaut version 13.12.200217.43655 (Bruderer et al., 2015) by searching against a database of Uniprot *Chlorocephus* sequences (19,136 proteins, downloaded April 3, 2020) and 29 SARS-CoV-2 protein sequences translated from genomic sequence downloaded from GISAID (accession EPI_ISL_406596, downloaded April 7, 2020) with two mutations (G22661T Spike V367F and G26144T ORF3a G251V) detected by RNASeq analysis of virus stocks. For protein abundance samples, data were searched using the default BGS settings, variable modification of methionine oxidation, static modification of carbamidomethyl cysteine, and filtering to a final 1% false discovery rate (FDR) at the peptide, peptide spectrum match (PSM), and protein level (Elias and Gygi 2007). For phosphopeptide enriched samples, BGS settings were modified to include phosphorylation of S, T, and Y as a variable modification. The generated search libraries were used to search the DIA data. For protein abundance samples, default BGS settings were used, with no data normalization performed. For phosphopeptide enriched samples, the Significant PTM default settings were used, with no data normalization performed, and the DIA-specific PTM site localization score in Spectronaut was applied.

Immunofluorescence microscopy

Caco-2 cells seeded on glass coverslips were infected with SARS-CoV-2 isolate Muc-IMB-1/2020, second passage on Vero E6 cells (2×10^6 PFU/mL) at an MOI of 0.1 or 0.01. At 24 hours post-infection cells were washed with PBS and fixed in 4% paraformaldehyde in PBS for 20 minutes at RT, followed by permeabilization with 0.3% Triton X-100 in PBS for 10 minutes at RT and blocking in 5% fetal calf serum in PBS for 1 hour at RT. Incubation with primary antibodies against CK2 α (Abcam, ab70774, 1:500), SARS-CoV membrane (M) protein (Rockland, #100-401-A55, 1:500) and SARS-CoV nucleocapsid (N) protein (Rockland, #200-401-A50, 1:1000) was performed for 1 hour at RT. After washing with PBS, cells were incubated with AF568-labeled goat-anti-rabbit (Invitrogen, #A11011) and AF647-labeled goat-anti-mouse (Invitrogen, #A21235) secondary antibodies (1:200) as well as AF488-labeled Phalloidin (Hypermol, #8813-01, 1:250) for 2 hours at RT. Fluorescence images were generated using a LSM800 confocal laser-scanning microscope (Zeiss) equipped with a 63X, 1.4 NA oil objective and Airyscan detector and the Zen blue software (Zeiss) and processed with Zen blue software and ImageJ/Fiji. For 3D-reconstruction, cells were fixed and stained as indicated and imaged as z stack with 0.15 μ m sections. Z stack was processed using Imaris software 64x 9.5.1 and displayed as MIP. To quantify colocalization between casein kinase II and N protein colocalization events in filopodia of SARS-CoV-2 infected Caco-2 cells were counted. $42 \pm 19\%$ of the N protein particles detected in filopodia colocalize with CK2. Length of filopodia was measured using Metamorph (Version 7.8). Distance was measured starting at cortical actin to the tip of Filopodia.

Electron microscopy

Vero E6 cells were seeded overnight and then infected for 24 hours with SARS-CoV-2 isolate nCoV-WA1-2020 on silicon chips for scanning electron microscopy or Thermanox for transmission electron microscopy. For scanning electron microscopy (SEM), cells were fixed with Karnovsky's formulation of 2% paraformaldehyde/2.5% glutaraldehyde in 0.1 M Sorenson's phosphate buffer, and then post-fixed with 1.0% osmium tetroxide/0.8% potassium ferrocyanide in 0.1 M sodium cacodylate buffer washed with 0.1 M sodium cacodylate buffer then stained with 1% tannic acid in dH₂O. After additional buffer washes, the samples were further osmicated with 2% osmium tetroxide in 0.1M sodium cacodylate, then washed with dH₂O. Specimens were dehydrated with a graded ethanol series from 50%, 75%, 100% x 3 for 5 minutes each, critical point dried under CO₂ in a Bal-Tec model CPD 030 Drier (Balzers, Liechtenstein), mounted with double sided carbon tape on aluminum specimen mounts (Ted Pella), and sputter coated with 35 Å of iridium in a Quorum EMS300T D sputter coater (Electron Microscopy Sciences, Hatfield, PA) prior to viewing at 5 kV in a Hitachi SU-8000 field emission scanning electron microscope (Hitachi, Tokyo, Japan). For transmission electron microscopy (TEM), specimens were fixed as described above for scanning electron microscopy and additionally stained overnight with 1% uranyl acetate at 4°C after the second osmium staining and then dehydrated with the same graded ethanol series, and embedded in Spurr's resin. Thin sections were cut with a Leica UC7 ultramicrotome (Buffalo Grove, IL) prior to viewing at 120 kV on a FEI BT Tecnai transmission electron microscope (ThermoFisher/FEI, Hillsboro, OR). Digital images were acquired with a Gatan Rio camera (Gatan, Pleasanton, CA).

N overexpression in Vero E6 cells

The N-protein cassette was subcloned from pLVX-EF1a-SARS-CoV-2-N-2xStrep-IRES-Puro (Gordon et al., 2020) into pLVX-TetOne-Puro (Takara) using the restriction enzymes EcoRI and BamHI to create pLVX-TetOne-Puro-SARS-CoV-2-N-2xStrep. Sequence integrity was confirmed by Sanger sequencing (GeneWiz). To produce lentiviruses, HEK293T cells (50% confluent in a T175 flask) were transfected with 5 μ g of each of the pLVX-TetOne-Puro lentiviral plasmids, 3.33 μ g Gag-Pol packaging construct,

and 1.66 μg VSV-G envelope (pMD2.G, Addgene) using PolyJet (SigmaGen). Culture supernatant was precipitated with a final concentration of 8.5% PEG-6000 and 0.3M sodium chloride (NaCl), incubated at 4°C for 2-4 hours. Virions were pelleted at 3500 rpm for 20 minutes in a spin bucket rotor, suspended in 0.5 mL 1xPBS, and aliquoted for storage at 80°C. Vero E6 cells were seeded in T75 flasks at 50% confluence and transduced with 200 μL precipitated lentivirus derived from pLVX-TetOne-Puro-SARS-CoV-2-N-2xStrep or pLVX-TetOne-Puro empty vector. 48 hours post transduction, 10 $\mu\text{g}/\text{mL}$ Puromycin was added to cultures to select transduced cells. Polyclonal stable cell lines were seeded into 15 cm dishes in triplicate and treated with 1 $\mu\text{g}/\text{mL}$ doxycycline for 48 hours. Cells were washed in ice cold PBS and harvested by scraping in cold PBS. Cell pellets were lysed directly (8 M urea, 100 mM ammonium bicarbonate (ABC), 150 mM NaCl, protease inhibitor (mini-cOmplete, Roche) and phosphatase inhibitors (phosSTOP, Roche), and protein digestion and phosphopeptide enrichment was performed as described above for SARS-CoV-2 infected Vero E6 cells. During the analysis, one of three replicates of the N-overexpressed samples was found to be an outlier based on principal component analysis (as in [Figures S1A and S1B](#)) and was removed. N-overexpression ($n = 2$) was compared to empty vector (EV) controls ($n = 3$) during analysis (full data available in [Table S1](#)). Kinase activity predictions were also performed the same as for viral infection phosphoproteomics (full data available in [Table S4](#)).

Cell cycle analysis

1×10^5 Vero E6 cells were seeded per sample. The following day cells were either mock infected or infected with SARS-CoV-2 (isolate BetaCoV/France/IDF0372/2020) at an MOI of 1. The samples were incubated for 24 hours before detaching cells using Trypsin 0.05% (Thermo). Cells were pelleted at 500x g for 2 minutes, followed by fixation using 4% Formalin (Thermo) for 30 minutes. Cells were washed 2x with PBS (Thermo) before permeabilizing using 0.1% Triton X-100 (Thermo) for 20 minutes. Finally, cells were pelleted again and incubated with 1 mg/mL 4',6-diamidino-2-phenylindole (DAPI; Thermo) in PBS for 30 minutes before analyzing cell cycle using fluorescence by flow cytometry using the BD FACSymphony with a violet laser (405 nm) and 450/50 nm filter. Gating was performed by first selecting singlets by gating highly correlated cells in an SSC-A versus SSC-H plot. Next, typical cellular morphology was gated using a FSC-A versus SSC-A plot. These singles were then used to draw histograms which were gated for G0/G1, S, and G2/M phases.

SARS-CoV-2 infections in ACE2-A549 cells and lysis for cytokine analysis

Approximately 5×10^5 A549-ACE2 cells were pre-treated with either DMSO or SB203580 (0.01, 0.1, 1, 10 μM final concentration). After 1 hour of pre-treatment, cells were infected with SARS-CoV-2 (isolate USA-WA1/2020) at an MOI of 0.1 while maintaining inhibitor concentrations in the media for 30 hours post infection. The supernatants from infected cells were analyzed by multiplexed ELISA for the presence of secreted cytokines and chemokines. For RNA analysis, cells were lysed in TRIzol and total RNA was extracted and DNase treated using the Direct-zol RNA Miniprep Plus kit (Zymo Research) according to the manufacturer's instructions. For protein analysis, whole cell lysates were obtained by lysis in RIPA buffer and analyzed by SDS-PAGE and western blot.

Cytokine RT-qPCR analysis

Gene expression of selected cytokines as well as SARS-CoV-2 replication was quantified by RT-qPCR. Reverse transcription of extracted total RNA samples was performed using oligo (dT) primers and SuperScript IV Reverse Transcriptase according to the manufacturer's instructions. Quantitative real-time PCR analysis of cDNA samples was performed using KAPA SYBR FAST qPCR Master Mix (2X) Universal on a LightCycler 480 Instrument II (Roche). For viral quantification, primers specifically targeting the subgenomic viral N RNA were used. ΔCT values were determined relative to the ACTB and $\Delta\Delta\text{CT}$ values were normalized to the average of mock infected samples (for host genes) or to infected DMSO treated samples (for viral replication). Error bars indicate the standard deviation of the mean from three independent biological replicates.

Multiplexed cytokine ELISA

Supernatants from infected cells were evaluated for 34 cytokines/chemokines by multiplex ELISA for the following analytes: CCL2/MCP-1, CCL3/MIP-1 α , CCL4/MIP-1 β , CCL5/RANTES, CCL7/MCP-3, CCL8/MCP-2, CCL11/Eotaxin, CCL13/MCP-4, CCL20/MIP-3 α , CXCL1/GRO α , CXCL2/GRO β , CXCL5/ENA-78, CXCL8/IL-8, CXCL9/MIG, CXCL10/IP-10, CXCL16, IL-1 α , IL-1 β , IL-1RA, IL-4, IL-6, IL-7, IL-10, IL-12p70, IL-15, IL-16, IL-17A, IL-22, GM-CSF, MMP-9, S100A8, TNF α , TGF β , and Trappin-2/Elafin. All antibodies and cytokine standards were purchased as antibody pairs from R&D Systems (Minneapolis, MN) or Peprotech (Rocky Hill, NJ). Individual magnetic Luminex bead sets (Luminex Corp, CA) were coupled to cytokine-specific capture antibodies according to the manufacturer's recommendations. Samples were analyzed on a Luminex MAGPIX platform and quantified using a standard curve. For each bead region, > 50 beads were collected per analyte. The median fluorescence intensity of these beads was recorded for each bead and was used for analysis using a custom R script and a 5P regression algorithm.

SARS-CoV-2 viral quantification via antibody staining in presence of inhibitors

(Mount Sinai, New York). Two thousand (2,000) Vero E6 cells were seeded into 96-well plates and incubated for 24 hours. Two hours before infection, the medium was replaced with 100 μL of DMEM (2% FBS) containing the compound of interest at concentrations 50% greater than those indicated, including a DMSO control. Plates were then transferred into the BSL-3 facility and 100 PFU of SARS-CoV-2 (MOI 0.025) was added in 50 μL of DMEM (2% FBS), bringing the final compound concentration to those indicated.

Plates were then incubated for 48 hours at 37°C. After infection, supernatants were removed and cells were fixed with 4% formaldehyde for 24 hours prior to being removed from the BSL-3 facility. The cells were then immunostained for the viral NP protein (antiserum produced in the Garcia-Sastre lab; 1:10,000) with a DAPI counterstain. Infected cells (488 nM) and total cells (DAPI) were quantified using the Celigo (Nexcelcom) imaging cytometer. Infectivity was measured by the accumulation of viral NP protein in the nucleus of the Vero E6 cells (fluorescence accumulation). Percent infection was quantified as ((Infected cells/Total cells) - Background) *100 and the DMSO control was then set to 100% infection for analysis. The IC₅₀ for each experiment was determined using the Prism software (GraphPad). For select inhibitors, infected supernatants were assayed for infectious viral titer using the Median Tissue Culture Infectious Dose TCID₅₀ method. For this, infectious supernatants were collected at 48 hours post infection and frozen at -80°C until later use. Infectious titers were quantified by limiting dilution titration on Vero E6 cells. Briefly, Vero E6 cells were seeded in 96-well plates at 20,000 cells/well. The next day, SARS-CoV-2-containing supernatant was applied at serial 10-fold dilutions ranging from 10⁻¹ to 10⁻⁶ and, after 5 days, viral CPE was detected by staining cell monolayers with crystal violet. TCID₅₀/mL was calculated using the method of Reed and Muench.

Cytotoxicity cell viability assays

(Mount Sinai, New York). Cytotoxicity was also performed using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected Vero E6 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates.

siRNA-based knockdown of host kinases

(Institut Pasteur, Paris). Host kinases were knocked-down in A549 cells stably expressing ACE2 cells using OnTargetPlus siRNA SMARTpools (Horizon Discovery). Briefly, A549-ACE2 cells seeded at 2x10⁴ cells/well in 96-well plates 18 hours prior to the experiment. The next day, each well was transfected with 6 pmoles of siRNA SMARTpool, using Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. Twenty-four (24) hours post transfection, the cell culture supernatant was removed and replaced with virus inoculum (MOI of 0.1 PFU/cell). Following a 1 hour adsorption at 37°C, the virus inoculum was removed and replaced with fresh Opti-MEM media (Thermo Fisher Scientific) containing 2% FBS (v/v). Seventy-two (72) hours post-infection (p.i.), the cell culture supernatant was heat-inactivated at 95°C for 5 minutes, and detection of viral genomes performed by RT-qPCR.

SARS-CoV-2 viral quantification via RT-qPCR in presence of inhibitors

(Institut Pasteur, Paris). Detection of viral genomes was performed by RT-qPCR, directly from the inactivated supernatant (Lista et al., 2020). SARS-CoV-2 specific primers targeting the N gene region: 5'-TAATCAGACAAGGAAGTACTGATTA-3' (Forward) and 5'-CGAAGGTGTGACTTCCATG-3' (Reverse) (Chu et al., 2020) were used with the Luna Universal One-Step RT-qPCR Kit (NEB) in an Applied Biosystems QuantStudio 7 thermocycler, with the following cycling conditions: 55°C for 10 minutes, 95°C for 1 minute, and 40 cycles of 95°C for 10 s, followed by 60°C for 1 minute. The number of viral genomes is expressed as PFU equivalents/mL, and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titer.

SARS-CoV-2 titration by plaque assay

(Institut Pasteur, Paris). Vero E6 cells were seeded in 24-well plates at a concentration of 7.5x10⁴ cells/well. The following day, serial dilutions were performed in serum-free MEM media. After 1 hour adsorption at 37°C, 2x overlay media was added to the inoculum to give a final concentration of 2% (v/v) FBS / MEM media and 0.05% (w/v) Agarose (all Thermo Fisher Scientific) to achieve a semi-solid overlay. Plaque assays were incubated at 37°C for 3 days. Samples were fixed using 4% Formalin (Sigma Aldrich) and plaques were visualized using crystal Violet solution (Sigma Aldrich).

siRNA Cell viability assays

(Institut Pasteur, Paris). Cell viability in siRNA knocked-down cells was measured using the CellTiter Glo luminescent cell viability assay (Promega) following the manufacturer's instructions, and luminescence measured in a Tecan Infinity 2000 plate reader. Percent viability was calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol (0% viability), included in each experiment.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mass spectrometry data pre-processing

Quantitative analysis was performed in the R statistical programming language (version 3.6.1, 2019-07-05). Initial quality control analyses, including inter-run clusterings, correlations, principal components analysis, peptide and protein counts and intensities were completed with the R package artMS (version 1.5.3). Based on obvious outliers in intensities, correlations, and clusterings, 2 runs were discarded from the protein abundance data and 2 runs were discarded from the phosphopeptide data (Figure S1). Additionally, the phosphopeptide data were filtered based on feature (i.e., peptide ion) intensity, removing any single feature with intensity less than 2¹⁴—this decision was made based on apparent lack of correlation between runs for feature intensities below this intensity.

Thus, for both phosphopeptides and protein abundance, we had 2 control time points and 6 infected time points, each with 3 biologically distinct replicates, except for infected at 0 and 2 hours in the phosphopeptide data and control at 0 hours and infected at 0 hours in the protein abundance data which only had 2 replicates each.

Statistical analysis of protein abundance changes between control and infected runs were computed using peptide ion fragment data output from Spectronaut and processed using a pipeline of three functions from the R package MSstats (version 3.19.5) (Choi et al., 2014): function MSstats::SpectronautMSstatsFormat with default settings other than setting "removeProtein_with1Feature = TRUE"; function MSstats::dataProcess with default settings other than setting "censoredInt = 0," "featureSubset = highQuality," "remove_uninformative_feature_outlier = TRUE," "clusters=7"; and function MSstats::groupComparison with all default settings.

Phosphopeptide intensity data were summarized at the peptide ion level along with confident localization of phosphorylation as described in the previous section. Quantification of phosphorylation based on peptide ions were processed using artMS as a wrapper around MSstats, via functions artMS::doSiteConversion and artMS::artmsQuantification with default settings. All peptides containing the same set of phosphorylated sites were grouped and quantified together into phosphorylation site groups. For both phosphopeptide and protein abundance MSstats pipelines, MSstats performs normalization by median equalization, imputation of missing values and median smoothing to combine intensities for multiple peptide ions or fragments into a single intensity for their protein or phosphorylation site group, and statistical tests of differences in intensity between infected and control time points.

Identifying significantly regulated proteins

When not explicitly indicated, we used defaults for MSstats for adjusted p values, even in cases of $N = 2$. By default, MSstats uses Student's t test for p value calculation and Benjamini-Hochberg method of FDR estimation to adjust p values. Phosphorylation fold change data were filtered for quality based on consistency of observations between treatment and controls by requiring the MSstats reported value missingPercentage < 0.60. This value is the proportion of features (i.e., peptide ions) that are missing across the treatment and control replicates for a specific fold change computation. MSstats phosphorylation results had to be further simplified to effects at single sites. The results of artMS/MSstats are fold changes of specific phosphorylation site groups detected within peptides, so one phosphorylation site can have multiple measurements if it occurs in different phosphorylation site groups (see Table S1 for examples). This complex dataset was reduced to a single fold change per site and time point by choosing (per time point) the fold change with the lowest p value, favoring those detected in both treatment and control, i.e., non-infinite \log_2 fold change values. This single-site dataset, further reduced to those with human orthology (Table S1), was used as the input for kinase activity analysis.

Significance filters of adjusted p value < 0.05 and absolute \log_2 fold change (infected/mock) > 1 were applied to both phosphorylation and abundance datasets. Additionally, for protein abundance data, if a protein was detected in only infected but not mock, or vice versa, we required detection in 3 replicates or by 3 peptide ion fragments in at least 2 replicates. These significance-filtered data were used for enrichment analyses and counts of significant effects (Figure 1).

Mapping *C. sabaesus* and *H. sapiens* proteins

Orthologous pairs of gene identifiers between *C.sabaesus* and *H.sapiens* were downloaded from Ensembl using the BioMart web interface (<http://uswest.ensembl.org/biomart/martview/6c48d59058381e6b2b198a1f91ba5e50>) on April 6, 2020. Ensembl gene identifiers were mapped to UniProt identifiers and protein sequences using a table of ID mappings and reference proteome sequences for *H.sapiens* downloaded from Uniprot on April 6, 2020. Orthologous pairs of sequences were aligned using the Needleman-Wunsch global alignment algorithm implemented in the R package Biostrings (v. 2.52) function pairwiseAlignment with default parameters. The resulting alignments were used to convert the sequence positions of detected phosphorylations in *C.sabaesus* to positions in *H.sapiens* protein sequences.

GO enrichment analysis

Sets of genes with significant up and down effects were tested for enrichment of Gene Ontology (GO Biological Process, Molecular Function and Cellular Component) terms. Sets of genes were either combined across time points (Figure 1I) or collected separately per time point (Figures S1F and S1G). The over-representation analysis (ORA) was performed using the enricher function of clusterProfiler package (version 3.12.0) in R with default parameters. The gene ontology terms were obtained from the c5 category of Molecular Signature Database (MSigDBv7.1) (Subramanian et al., 2005). Significant GO terms (adjusted p value < 0.01) were identified and further refined to select non-redundant terms. In order to select non-redundant gene sets, we first constructed a GO term tree based on distances (1-Jaccard Similarity Coefficients of shared genes in MSigDB) between the significant terms. The GO term tree was cut at a specific level ($h = 0.99$) to identify clusters of non-redundant gene sets. For results with multiple significant terms belonging to the same cluster, we selected the most significant term (i.e., lowest adjusted p value).

Clustering of phosphorylation changes

Phosphorylation sites with one-to-one mapping to human S, T or Y amino acids and showing significant change in phosphorylation ($\text{abs}(\log_2\text{FC}) > 1$ and adjusted p value < 0.05) in one or more conditions were selected and clustered using hierarchical clustering (complete-linkage) with pearson correlation (1-r) as the distance measure. The cluster tree was cut into 5 clusters using a dynamic tree cutting method, cutreeHybrid function in dynamicTreeCut (Langfelder et al., 2008) package in R, with a minimum cluster size of 130 sites. Phosphorylated proteins in each cluster were tested for enrichment of Reactome pathways (Jassal et al., 2020). The over-

representation analysis (ORA) was based on the hypergeometric distribution and performed using the enricher function of clusterProfiler package in R (Yu et al., 2012). The pathway terms were obtained from the c2 category (Reactome) of Molecular Signature Database (MSigDBv6.1).

Estimation of kinase activities in the time-course experiment

Kinase activities were estimated using known kinase-substrate relationships in literature (Bachman et al., 2019). The resource comprises of a comprehensive collection of phosphosite annotations of direct substrates of kinases obtained from six databases, PhosphoSitePlus, SIGNOR, HPRD, NCI-PID, Reactome, and the BEL Large Corpus, and using three text-mining tools, REACH, Sparser, and RLIMS-P. Kinase activities were inferred as a $-\log_{10}(p \text{ value})$ of Z-test from the comparison of fold changes in phosphosite measurements of the known substrates against the overall distribution of fold changes across the sample. This statistical approach has been previously shown to perform well at estimating kinase activities (Hernandez-Armenta et al., 2017; Casado et al., 2013). We collected substrate annotations for 400 kinases with available data. Kinase activities for kinases with 3 or more measured substrates were considered leaving us with 97 kinases with activity estimates in at least one or more infection time points.

Comparison of phosphorylation profiles with other conditions

In addition to the time-course experiment, we also estimated kinase activities of cell line perturbation in a range of biological conditions using the K-test based approach, from a large resource of previously published phosphoproteomics datasets comprising of 435 biological conditions which include several drug treatments (chlorzoxazone, paclitaxel), inhibitors (EGFRi, AKTi) and cell cycle states (Ochoa et al., 2016). After selecting for kinase activities derived from 3 or more substrates, a total of 214 kinases had activity estimates for at least one biological condition including all 97 kinases which were estimated in the time-course experiment. Of the 435 conditions, 309 conditions with at least one kinase with significantly changing activity ($-\log_{10}(p \text{ value})$ from Z-test > 2.5) were selected. Individual time points of the infection experiment were correlated with each selected biological condition based on Pearson correlation of their kinase activity profiles. Conditions with significant correlation (5% FDR) with one or more time points were identified. Several conditions corresponded to the same biological treatment (i.e., drug perturbation) in different concentrations. In such cases, the conditions are represented together with strongest correlation.

Phosphorylation of protein complexes

Annotations for protein complexes were obtained from CORUM v3.0 database (Giurgiu et al., 2019). Overall significance of changes of phosphosites of a protein complex was inferred from the \log_2 fold change of individual phosphosites of protein members of the complex using the same Z-test based approach described for estimation of kinase activities. Among the 3512 complexes, 1070 complexes with phosphorylation changes obtained from 4 or more phosphosites were considered. Of these, 67 complexes had significant phosphorylation change ($-\log_{10}(p \text{ value})$ from Z-test > 2.5) in at least one time point. In order to select non-redundant complexes, we first constructed a complex term tree based on distances (1-Jaccard Similarity Coefficients of shared genes) between the significant complexes. The term tree was cut at a specific level ($h = 0.99$) to identify clusters of non-redundant complexes. The largest complex was selected from each cluster.

Transcription factor activity after SARS-CoV-2 infection

To evaluate the effect of SARS-CoV-2 infection at the Transcription Factor (TF) level, we applied DoRothEA (Garcia-Alonso et al., 2019) to RNA-seq datasets of different human lung cell lines from a recent study (GSE147507) (Blanco-Melo et al., 2020). DoRothEA is a comprehensive resource containing a curated collection of TF-target interactions. Each TF-target interaction is associated with a confidence level based on the number of supporting evidence. Here we selected the most reliable interactions (A, B, and C levels) and computed TF activities based on the normalized expression of their targets using the VIPER algorithm (Alvarez et al., 2016). For the TF activity enrichment analysis, VIPER was executed with the Wald statistic resulting from the differential expression analysis at the gene level between controls and SARS-CoV-2 infected cells using the DESeq2 package (Love et al., 2014). In VIPER, we set the eset.filter parameter to FALSE and consider five as the minimum number of targets allowed per regulon.

Pharmacological profiling dose response analysis

A hill function was fit to each dose response curve using the Isqcurvefit function in MATLAB (R2018a). IC_{50} (virus) and CC_{50} (cell viability) values were defined as the concentration at which the percent measure (virus or cell viability quantification) crossed the 50% mark. If the fit curve did not begin above 50% and cross to below 50% throughout the dose response, an IC_{50} or CC_{50} value was marked as greater than the maximum tested concentration. For $TCID_{50}$ and plaque assay results, which are not depicted in a percent scale, the IC_{50} values were extracted directly from the parameterized hill equation (see Figure S5 and Table S8).

6. Host PDZ-containing proteins targeted by SARS-CoV-2

Célia Caillet-Saguy, Fabien Durbesson, Veronica V. Rezelj, Gergö Gogl, **Quang Dinh Tran**, Jean-Claude Twizere, Marco Vignuzzi, Renaud Vincentelli & Nicolas Wolff

DOI: <https://doi.org/10.1111/febs.15881>

Contributions: I analyzed data from Figure 4.

Host PDZ-containing proteins targeted by SARS-CoV-2

Célia Caillet-Saguy¹ , Fabien Durbesson², Veronica V. Rezelj³, Gergő Gogl⁴, Quang Dinh Tran^{3,5} , Jean-Claude Twizere⁶, Marco Vignuzzi³, Renaud Vincentelli² and Nicolas Wolff¹ 

- 1 Institut Pasteur, Unité Récepteurs-Canaux, UMR CNRS 3571, Paris, France
 2 AFMB, UMR CNRS 7257, Marseille, France
 3 Institut Pasteur, Unité Populations Virales et Pathogénèse, UMR CNRS 3569, Paris, France
 4 IGBMC, INSERM U1258/UMR CNRS 7104, Illkirch, France
 5 École doctorale BioSPC, Université Paris Diderot, Sorbonne Paris Cité, France
 6 GIGA Institute, Molecular Biology of Diseases, Viral Interactomes laboratory, University of Liege, Belgium

Keywords

3A and N; human PDZ library; PDZ-binding motif; PDZ-containing protein; SARS-CoV-2; viral proteins E; viral replication

Correspondence

C. Caillet-Saguy, N. Wolff, Institut Pasteur, Unité Récepteurs-Canaux, UMR CNRS 3571, 75015 Paris, France
 Tel: +0033 145688872
 E-mails: celia.caillet-saguy@pasteur.fr and nicolas.wolff@pasteur.fr

Renaud Vincentelli and Nicolas Wolff contributed equally as last authors.

(Received 2 February 2021, revised 31 March 2021, accepted 13 April 2021)

doi:10.1111/febs.15881

Small linear motifs targeting protein interacting domains called PSD-95/Dlg/ZO-1 (PDZ) have been identified at the C terminus of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins E, 3a, and N. Using a high-throughput approach of affinity-profiling against the full human PDZome, we identified sixteen human PDZ binders of SARS-CoV-2 proteins E, 3A, and N showing significant interactions with dissociation constants values ranging from 3 to 82 μM . Six of them (TJP1, PTPN13, HTRA1, PARD3, MLLT4, LNX2) are also recognized by SARS-CoV while three (NHERF1, MAST2, RADIL) are specific to SARS-CoV-2 E protein. Most of these SARS-CoV-2 protein partners are involved in cellular junctions/polarity and could be also linked to evasion mechanisms of the immune responses during viral infection. Among the binders of the SARS-CoV-2 proteins E, 3a, or N, seven significantly affect viral replication under knock down gene expression in infected cells. This PDZ profiling identifying human proteins potentially targeted by SARS-CoV-2 can help to understand the multifactorial severity of COVID19 and to conceive effective anti-coronaviral agents for therapeutic purposes.

Introduction

Coronaviruses (CoV) are enveloped viruses that primarily infect birds and mammals. Some CoVs have evolved to establish zoonotic disease in humans, potentially causing severe respiratory illnesses [1]. The seriousness of CoV infection during the last outbreaks caused by the severe acute respiratory syndrome coronavirus 1 (SARS-CoV) in 2003, the Middle East Respiratory Syndrome-virus (MERS-CoV) in 2012 and the SARS-CoV-2 in 2020 [2], and the lack of effective treatment for CoV infection highlight the need of a

deeper understanding of coronavirus molecular biology to elaborate therapeutic strategies that counteract SARS-CoV-2 infection. However, limited knowledge of the molecular details of SARS-CoV-2 infection precludes a fast development of compounds targeting the host-virus interface. Several SARS-CoV-2 interactomes have been recently published using mass spectrometry [3–5]. Unfortunately, some transient and weak interactions with affinities in the 1–100 μM range can be hardly detected by such approaches. However, during

Abbreviations

BIs, binding intensities; CoV, coronaviruses; DV, Dengue virus; E, envelope; HPV, human papillomavirus; K_d , dissociation constant; M, membrane; MERS-CoV, Middle East Respiratory Syndrome virus; N, nucleocapsid; PBMs, PDZ-Binding Motifs; PDZ, PSD-95/Dlg/ZO-1; PDZome, PDZ domains identified in the human genome; PFU, plaque forming units; S, spike; SARS-CoV, severe acute respiratory syndrome coronavirus 1; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SLiMs, short linear motifs; TCR, T-cell receptor; TM, transmembrane; WNV, West Nile Virus.

infection, viruses target host signaling pathways that mainly involve protein–protein interactions of low affinity, as transient bindings are essential for the dynamics of cellular homeostasis.

Hijacking of cellular protein functions is a widely used strategy for viruses as they are obligate intracellular pathogens. Each step of the viral life cycle from entry to transmission is orchestrated through interactions with cellular proteins. All families of viruses manipulate the cell proteome by targeting key proteins involved in the control of cell defense and homeostasis. Interactions mediated by short linear motifs (SLiMs) are ubiquitous in the eukaryotic proteome [6], and the adaptation of viruses to their environment involves the extensive use of SLiM mimicry to subvert host functionality [7].

PSD-95/Dlg/ZO-1 (PDZ)-binding motifs (PBMs) are SLiMs that interact with a large family of protein interaction domains called PDZ. PBMs are usually located at the extreme carboxyl terminus of proteins [8]. The PDZ-PBM interactome is one of the most prominent instances of a SLiM-mediated protein interaction network serving key cell signaling purposes [9]. PBMs were identified in diverse proteins of many viruses responsible for acute to chronic infections [10,11]. Cellular PDZ-viral PBM interactions were shown to be directly involved in viral pathogenicity of SARS-CoV, as well as of rabies virus and in the oncogenicity of human papillomavirus (HPV) [10]. During infection, the viral proteins compete with the endogenous ligands through the binding to the PDZ domain of the host protein targets allowing a cellular relocation, sequestration, or degradation of protein [12] but can also affect the catalytic activity of signaling proteins [13]. Functional perturbations of cellular processes due to interactions between viral PBMs and cellular PDZ-containing proteins can facilitate the virus life cycle in the host and the dissemination to new hosts. Thus, targeting the PBM-PDZ interface could potentially lead to novel antiviral therapies.

The four major structural proteins encoded by the coronavirus genome, the spike (S) protein, the nucleocapsid (N) protein, the membrane (M) protein, and the envelope (E) protein, participate in different aspects of the replication cycle and ultimately form a structurally complete viral particle. While a minor fraction is incorporated into the virion envelope, the integral membrane E protein (~8–12 kDa) of CoVs including SARS-CoVs is a multifunctional protein that forms ion channels with its transmembrane (TM) domain. This channel activity is mediated by formation of pentameric oligomers and is only mildly selective for cations [14]. The channel would be active in

the secretory pathway, altering luminal environments, rearranging secretory organelles, and leading to efficient trafficking of virions [15]. In addition, E protein interacts with host proteins, related to its extramembrane domain particularly the C-terminal domain.

The C-terminal sequence in extra membrane position is predicted to be partially folded (helices and β -coil- β motif) which could affect the global conformation of the pentameric E channel suggesting that the channel activity of E and its protein interactions most likely interplay [14,16]. In addition, the E protein encodes a putative PBM sequence at its extreme C terminus (Fig. 1A) that could also interfere through protein–protein interactions with the oligomeric organization of E protein and consequently its channel activity. The PBM sequence of E protein has been identified as a virulence factor for SARS-CoV virus, influencing replication level, virus dissemination, and pathogenicity of SARS-CoV in animal models [17]. To date, E protein of SARS-CoV has only been reported to interact with five host proteins: Bcl-xL, PALS1/MPP5, syntenin, sodium/potassium ATPase α 1 subunit, and stomatin [16]. The relevance of these interactions is not yet fully understood. Among these five partners of SARS-CoV, PDZ domains of the PALS1/MPP5 tight junction protein and of Syntenin interact with the C-terminal PBM of E protein. PBM-PDZ interactions strongly impact the structure of mammalian epithelial cells and could contribute to the abrupt deterioration of the lung epithelium observed in patients infected by SARS-CoV [18]. Therefore, to better understand the influence of host PDZs and viral PBM motifs in the pathogenicity of SARS-CoV-2, we performed a targeted screen to identify cellular PDZ domains capable of interacting with PBM motifs CoV proteins.

In this study, we used the holdup assay [19], a high-throughput and quantitative approach that offers a high sensitivity for low-to-medium (1–100 μ M) affinity, typically found for PDZ/PBM interactions. We established the list of PDZ-containing partners potentially targeted by SARS-CoV-2 E protein through its PBM. This work identified 10 PDZ-containing proteins and determined the affinities of their PDZ to SARS-CoV-2 E protein PBM. Most of them are involved in cellular junction dynamics (PAR3, TJ1/ZO-1, LNX2, MLLT4, NHERF1). To further understand the potential implication of the PDZ-PBM in pathogenicity, we have also determined the PDZ-containing proteins recognized by the E protein of SARS-CoV and MERS-CoV viruses, as well as the PDZ binders of the SARS-CoV-2 3a and N proteins that also encode for a C-terminal PBM sequence. From this PDZ list, we have selected 20 PDZ-containing partners of SARS-

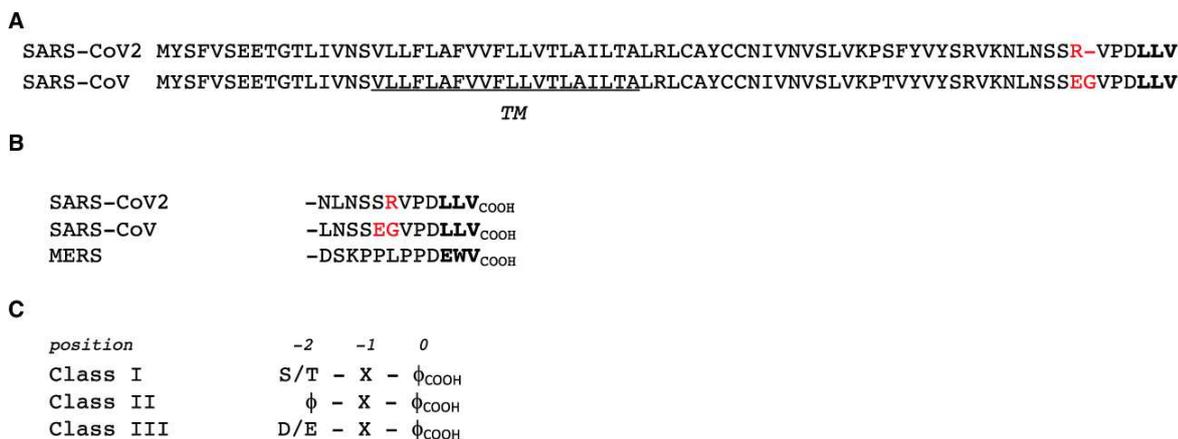


Fig. 1. E protein sequences. (A) Comparison of E protein sequences of SARS-CoV-2 and SARS-CoV viruses. The SARS-CoV-2 genome is very similar to SARS-CoV. The last three residues that form the PBM sequence (in bold) are strictly conserved. Deletion and substitution of amino acid close to the PBMs are colored in red. The TM helix is underlined. (B) Sequences of peptides used as baits in the holdup assays. PBM residues are in bold. (C) The three classes of PBM. X is any amino-acid (position -1), and ϕ is a hydrophobic amino-acid (position 0).

CoV-2 to assess their influence on SARS-CoV-2 replication in infected cells and found that the expression of seven of them is relevant for the efficiency of viral replication.

Results

SARS-CoV-2 E protein recognizes several cellular PDZ-containing proteins

The E protein of SARS-CoV virus is known to contain a C-terminal motif, a -LLV_{COOH} sequence corresponding to a PBM of class II (Fig. 1C). This motif is strictly conserved in SARS-CoV-2 E protein (Fig. 1A). To quantify the binding activity of the SARS-CoV-2 E PBM and its upstream sequence, we used an *in vitro* automated high-throughput chromatography assay called holdup [19,20]. A 12-mer peptide was synthesized encompassing the C-terminal PBM sequence of SARS-CoV-2 E protein linked to a biotinyl group (Fig. 1B). This peptide was used as bait to quantify the interaction between the SARS-CoV-2 E protein PBM peptide and the full human PDZ domain library expressed in *Escherichia coli*. The updated PDZ library called PDZome V.2 covers 98% of the human 'PDZome' (266 over 272 PDZ domains identified in the human genome) [20,21]. Each PDZ of the library is defined by the protein name followed by the PDZ number in the protein. The holdup approach displays a high sensitivity of detection for low-to-medium affinity PDZ/PBM pairs and provides an affinity-based

ranking of the identified binders corresponding to a specificity profile [22]. The holdup assay allows to measure binding intensities (BIs) for a large number of PDZ domain-PBM motif pairs. The datasets are plotted as specificity profiles (Fig. 2A) representing all the individual BIs measured for each PDZ domain for a peptide encompassing the PBM motif and its upstream sequence.

The specificity profiles for SARS-CoV, SARS-CoV-2, and MERS-CoV E PBM with the mean values of BI are reported in Fig. 2A with an affinity-based ranking. Steady-state holdup BIs can be converted to dissociation constants as described previously [23]. For this, we determined the dissociation constants of a few PBM-PDZ interactions using a competitive fluorescence polarization assay. The holdup assay detected several significant interactions for the three viruses. Using a stringent threshold previously defined for significant detectable binders in holdup assays at BI > 0.2 (~ 100 μM dissociation constant), 10 PDZ binders of SARS-CoV-2 E protein showed interactions with affinities values ranging from 2.6 μM (TJP1_2) to 82 μM (MLLT4; Fig. 2, zoomed-in view, Table 1). Thus, the PBM sequence of SARS-CoV-2 E protein has the capacity to bind *in vitro* several PDZ domains in the 1–100 μM affinity range typically found for PDZ/PBM interactions. As illustrated by the sharp profile of interaction of Fig. 2A, SARS-CoV-2 E PBM recognizes only 4% of PDZ domains from the human PDZ library reflecting the specificity of these interactions.

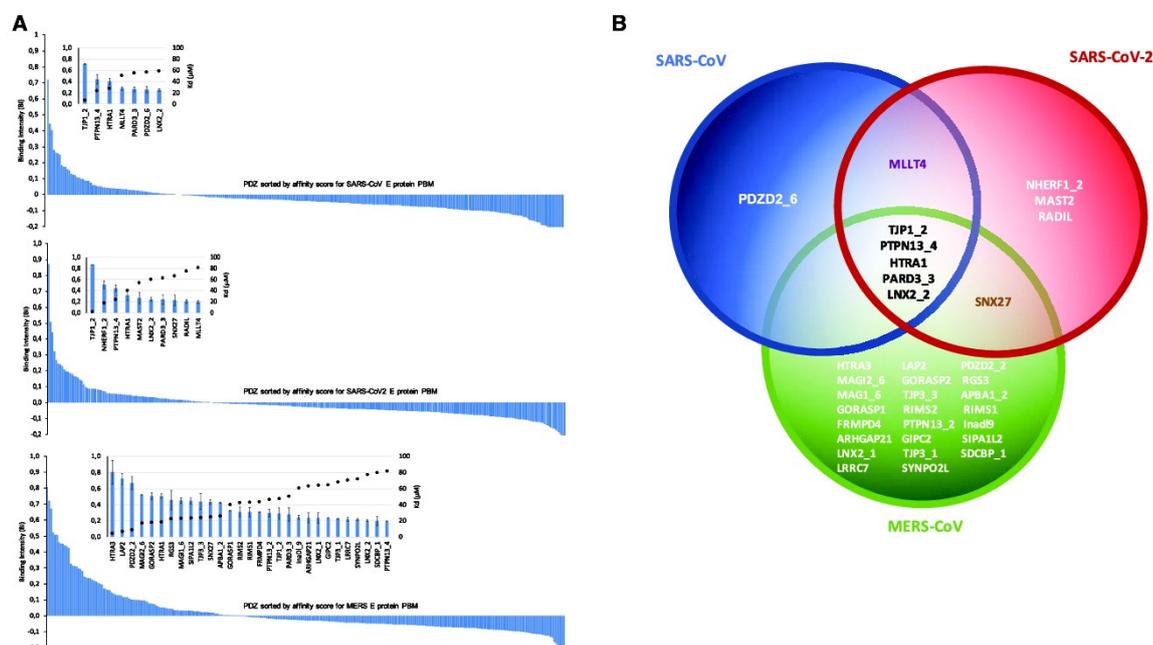


Fig. 2. PDZ binders of the E protein PBMs by holdup assay. (A) PDZome-binding profiles of the E protein PBMs by holdup assay of SARS-CoV (upper panel), SARS-CoV-2 (middle panel) and MERS-CoV (bottom panel). PDZ domains are ranked by decreased BI values. Zoomed-in views show the 7, 10 and 29 PDZ binders of SARS-CoV, SARS-CoV-2 and MERS-CoV E proteins, respectively, which displayed significant BI values higher than 0.2. Error bars are standard deviations of two independent experiments. The black circles correspond to affinity values (K_d) in μM (secondary Y axis). (B) Venn diagram of PDZ binders of the E protein PBMs of SARS-CoV, SARS-CoV-2 and MERS-CoV.

Patterns of human PDZ-containing proteins preferentially targeted by E proteins expressed by SARS-CoV, SARS-CoV-2, and MERS-CoV viruses

Two sequence differences between SARS-CoV and SARS-CoV-2 E proteins are observed upstream their identical PBM sequences (Fig. 1B).

Seven PDZ-containing proteins (3%) were identified as binders of SARS-CoV E PBM by high-throughput holdup (Table 1). Six of them are also SARS-CoV-2 E PBM binders, ranked by decreased affinity values in Fig. 2B; TJP1-2, PTPN13_4, HTRA1, MLLT4, PARD3_3 LNX2_2. The affinities of SARS-CoV and SARS-CoV-2 E PBMs for this set of proteins are similar, in agreement with the strict conservation of the last five C-terminal positions of the PBM (p0 to p-4) (Fig. 1A; Table 1). PDZ2_6 is only recognized by SARS-CoV E PBM while four additional PDZs are only recognized by SARS-CoV-2 E PBM (NHERF1_2, MAST2, RADIL, SNX27). Thus, the PBM upstream sequence differences increase the

number of binders for SARS-CoV-2 protein E with four additional host partners.

The MERS-CoV E protein encodes for a class III PBM (Fig. 1B,C). As compared to the SARS-CoV/SARS-CoV-2 E patterns, the profile of interaction of MERS-CoV E PBM is extended with a larger number of partners, that is, 29 (11%), within the same range of affinities between 5 and 82 μM (Table 1). Six out of seven PDZ-containing partners of SARS-CoV E PBM and six out of 10 PDZ-containing partners of SARS-CoV-2 E PBM are also recognized by MERS-CoV E PBM (Fig. 2B).

The C-terminal motifs of SARS-CoV-2 N and 3a proteins also bind to cellular PDZ-containing proteins

Apart from the E protein, following a systematic search of the sequences of the 29 proteins found in SARS-CoV-2, the 3a protein and the nucleocapsid protein N were also found to contain PBMs. The

Table 1. Measured BI and converted K_d (μM) values of PDZ domains for the SARS-CoV (A), SARS-CoV-2 (B) and MERS-CoV (C) E protein PBMs. Only significant BI values higher than 0.2 are reported. When multiple holdup experiments were performed, the averaged BI values and standard deviations are reported.

A				B				C			
SARS-CoV E	BI	St	K_d	SARS-CoV-2 E	BI	St	K_d	MERS E	BI	St	K_d
TJP1_2	0.72	0.01	7.0	TJP1_2	0.87	0.00	2.6	HTRA3	0.81	0.15	4.2
PTPN13_4	0.44	0.08	24	NHERF1_2	0.51	0.07	18	LAP2	0.72	0.07	6.9
HTRA1	0.40	0.06	28	PTPN13_4	0.44	0.06	24	PDZD2_2	0.67	0.08	8.9
MLLT4	0.28	0.03	51	HTRA1	0.32	0.10	41	MAGI2_6	0.52	0.01	17
PARD3_3	0.26	0.04	55	MAST2	0.27	0.10	54	GORASP2	0.51	0.04	18
PDZD2_6	0.26	0.06	57	LNX2_2	0.25	0.04	61	HTRA1	0.51	0.03	18
LNX2_2	0.25	0.02	59	PARD3_3	0.24	0.09	63	RGS3	0.46	0.12	22
				SNX27	0.23	0.10	67	MAGI1_6	0.45	0.03	23
				RADIL	0.21	0.03	75	TJP3_3	0.44	0.10	24
				MLLT4	0.20	0.02	82	SNX27	0.44	0.03	25
								APBA1_2	0.43	0.01	26
								GORASP1	0.33	0.00	40
								RIMS2	0.31	0.06	43
								RIMS1	0.31	0.06	43
								FRMPD4	0.31	0.00	43
								PTPN13_2	0.30	0.04	46
								TJP1_2	0.29	0.07	48
								PARD3_3	0.28	0.08	50
								InaDI_9	0.25	0.02	61
								ARHGAP21	0.24	0.06	63
								LNX2_1	0.24	0.06	64
								GIPC2	0.24	0.01	65
								TJP3_1	0.23	0.00	68
								SIPA1L2	0.22	0.22	69
								LRRC7	0.22	0.03	70
								SYNPO2L	0.22	0.01	72
								LNX2_2	0.21	0.01	77
								SDCBP_1	0.20	0.06	80
								PTPN13_4	0.20	0.01	82

PBM sequences of E and 3a proteins (Fig. 3A), two viroporins, belong to the same class of PBM, the class II (Fig. 1C). While the PBM of the E protein is involved in SARS-CoV virulence and replication [17], the motif of the protein 3a has no significant impact on viral growth [24]. However, it has been shown that this accessory 3a protein might exhibit compensatory function in the absence of E, on viral replication *in vitro* and *in vivo*. This homology suggests that the E and 3a proteins could interact with common host PDZ-containing proteins.

On the other hand, the potential PBM sequence at the C terminus of the structural nucleocapsid protein N is a class I motif with the atypical presence of a small hydrophobic residue, that is, an alanine, at the last position (Fig. 3A).

The binding profiles of PBM from SARS-CoV-2 3a and N proteins are reported in Fig. 3B. Five binders

(2%) displayed significant BI values higher than 0.2: TJP1_2, NHERF4_1, NHERF3_4, RGS3, and PARD3B_1, ranked by decreased affinity values. Interestingly, E proteins of SARS-CoV and SARS-CoV-2, and 3a protein of SARS-CoV-2 have the same best binder, TJP1_2, and both also recognized the PDZ domain of PARD3B. NHERF3 and NHERF4 are two new binders of SARS-CoV-2 while RGS3 is also recognized by MERS-CoV E protein. To note, the affinity range of 3a protein for its binders is lower than the one of E protein, from 25 to 80 μM . As for the N protein PBM binders, only two PDZ domains (1%) are identified: RHNP2 and GIPC1. These proteins have not been identified as binders of proteins E and 3a. GIPC1 is a PDZ-containing protein recognized by the PBM of the hepatitis B virus capsid protein and can accommodate another atypical small C-terminal residue, a cysteine [25].

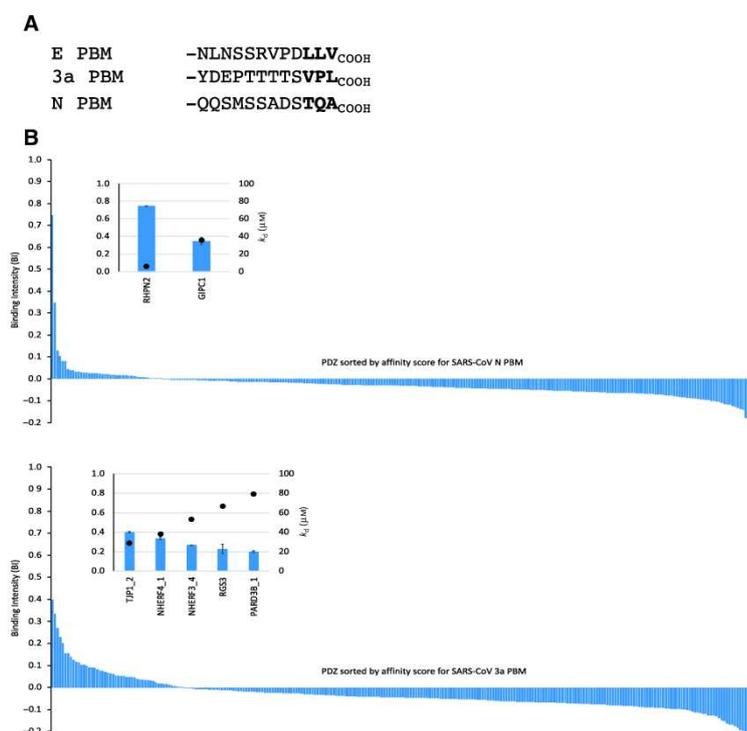


Fig. 3. PDZome-binding profiles of SARS-CoV-2 N protein and 3a protein PBMs by holdup assay. (A) sequences of E, N and 3a peptides used as baits in the holdup assays. (B) PDZ domains are ranked by decreased BI values. Zoomed-in views show the 2 and 5 PDZ binders of SARS-CoV-2 N and 3a proteins, respectively, which displayed significant BI values higher than 0.2. The black circles correspond to affinity values (K_d) in μM (secondary Y axis). The data are representative of two independent experiments and error bars correspond to the standard deviation.

Effects of depletion of cellular PDZ-containing proteins on SARS-CoV-2 replication

We next investigated the biological role of the PDZ-containing partners of SARS-CoV-2 proteins in the viral replication according to our holdup assay results.

We established a list of potential PDZ partners recognized by at least one of the PBM motif encoded by the SARS-CoV-2 proteins E, 3a or N (Table 2). We selected the 16 PDZ partners identified with BIs higher than 0.2 (Figs 2B and 3B) and added four additional PDZ partners with BI values comprised between 0.16 and 0.20; MPP5/PALS1, PDZD2, and DLG3 recognized by SARS-CoV-2 E protein and FRMPD4, a binder of both SARS-CoV-2 proteins E and 3A.

Twenty human PDZ-containing proteins were targeted for knockdown by siRNA transfection in human lung A549 cells. This cell line stably expresses the SARS-CoV-2 receptor ACE2 [26]. Our approach to knockdown gene expression used siRNA pools containing a mixture of four individual siRNAs, that are designed and modified for greater specificity, providing guaranteed gene silencing of human targets. We have validated these siRNA pools previously in a screen carried out in human A549 cells overexpressing ACE-

2. Our approach confirmed that knockdown of a number of different genes, including the receptor ACE-2, and other unexpected genes such as BZW2 or BRD4 reduces SARS-CoV-2 replication in these cells [3]. The effect of PDZ partner knockdown on virus replication was assessed in three independent experiments (each done in triplicate) and compared to virus replication in control siRNA-transfected cells (two-way ANOVA with Dunnett's test). As expected, knockdown of the SARS-CoV-2 receptor ACE2 resulted in significantly lower virus replication (Fig. 4). Virus replication was also significantly reduced in PARD3 and RSG3 knocked-down cells, suggesting a pro-viral role for these human proteins. PARD3 is found in our holdup assays as a PDZ partner of E protein of SARS-CoV, SARS-CoV-2, and MERS-CoV while RGS3 is recognized by SARS-CoV-2 3a protein. Comparatively, knockdown of MLLT4, PARD3B, MAST2, RHPN2, and HTRA generated an increased virus replication. MLLT4, MAST2, and HTRA1 interact with SARS-CoV-2 E protein, PARD3B with SARS-CoV-2 3a protein, and finally RHPN2 binds to SARS-CoV-2 N protein.

Altogether, our results suggested that the expression of these seven proteins is relevant for viral replication.

Table 2. Cellular PDZ-containing proteins recognized by at least one SARS-CoV-2 protein in holdup assays. List of the 20 PDZ partners of SARS-CoV-2 proteins selected for knockdown by siRNA transfection in human lung A549 cells. (+) indicated PDZ partners identified by assay with binding affinity values higher than 0.2; (+/–) represented PDZ binders with BI values comprised between 0.16 (105 μ M) and 0.20 (80 μ M).

Name prot	Name gene	Gene id	SARS-CoV-2		
			Protein E	Protein 3a	Protein N
TJP1	TJP1/ZO1	7082	+	+	
PTPN13	PTPN13	5783	+		
HTRA1	HTRA1	5654	+		
MLLT4	AFDN	4301	+		
PARD3	PARD3	56 288	+	+/–	
NHERF1	SLC9A3R1	9368	+		
LNX2	LNX2	222 484	+		
MAST2	MAST2	23 139	+		
SNX27	SNX27	81 609	+		
RADIL	RADIL	55 698	+		
NHERF4	PDZD3	79 849		+	
RGS3	RGS3	5998		+	
PARD3B	PARD3B	117 583		+	
RHPN2	RHPN2	85 415			+
GIPC1	GIPC1	10 755			+
MPP5	MPP5/ PALS1	64 398	+/–		
FRMPD4	FRMPD4	9758	+/–	+/–	
PDZD2	PDZD2	23 037	+/–		
DLG3	DLG3	1741	+/–		
NHERF3	PDZK1	5174		+	

Discussion

Only a few partners of SARS-CoV-2 E protein have been identified through extensive high-throughput approaches such as proteomics. As an example, among the structural SARS-CoV-2 proteins, only 6 binders of E protein are found as monitored in the Gordon *et al.* [3] study, while 30 partners for M protein and 15 binders to N protein are identified. Among them, no PDZ-containing protein has been identified in this study. Here, we have focused on specific interactions involving short C-terminal motifs of SARS-CoV-2 proteins, the PBMs, and PDZ domains of human partners. We mainly used the holdup high-throughput screening to identify the host PDZ partners of E, 3a, and N proteins and to quantify their PDZ-PBM affinities and specificities [19]. This chromatographic assay in solution is fully suitable to reliably rank low affinity interactions in the submillimolar range, which may escape detection by other high-throughput methods.

Conservation of the PBM sequence in Coronavirus E protein

The PBM motif at the C terminus of E protein is only expressed by coronaviral genera α and β , and is absent from the genus γ . The last 4 residues of E protein are strictly conserved in the SARS-CoV and SARS-CoV-2 viruses. A recent study has shown that, among the 3617 available complete genome sequences of SARS-CoV-2 in the NCBI database, only two variants (DFLV and YLLV) have been reported in the DLLV PBM motif of the E protein [27].

Both the E proteins of SARS-CoV-2 and MERS-CoV comprised putative PBM sequences at their C terminus but their nature is very different with a class II PBM for the genus α , and a class III PBM for the genus β . Our work shows that despite the absence of conservation between the PBM sequence of SARS-CoV/CoV-2 and MERS-CoV E protein, the two motifs interact with a common set of host partners; indeed, among the 10 host partners we identified for SARS-CoV-2, six are also binders of MERS-CoV.

Host-viral protein–protein interactions

Comparison of host partners of CoV and CoV-2

The affinity-based ranking of the identified binders of E protein provided a specificity profile. As comparison, the class I PBMs of NS5 protein from West Nile virus (WNV) and of capsid protein from hepatitis B virus interact with 29 and 28 PDZ binders [25,28], respectively, while about 50 PDZ proteins bind to the class I PBM of E6 viral oncoprotein from HPV16 [19]. By contrast, class II PBMs generally display lower affinities for their target and consequently a lowest number of binders in comparison with the class I PBMs. The hydrophobic character of the class II PBM sequences can explain the low affinities and sharp profiles of binding of SARS-CoV (seven binders) and SARS-CoV-2 (10 binders) E protein PBM with a reduced enthalpic contribution to the binding compared to the more hydrophilic class I sequences. Interestingly, the PBM of SARS-CoV-2 E protein can accommodate different classes of PDZ domains as illustrated in Fig. 5.

In the case of the class III PBM of E protein for MERS-CoV, the large number of targets (29) can be mainly explained by the presence of a tryptophan residue at position –1, a residue known to have a strong contribution to the binding affinity [29], potentially at the expense of the specificity. The PBM upstream sequence might also have a contribution to the

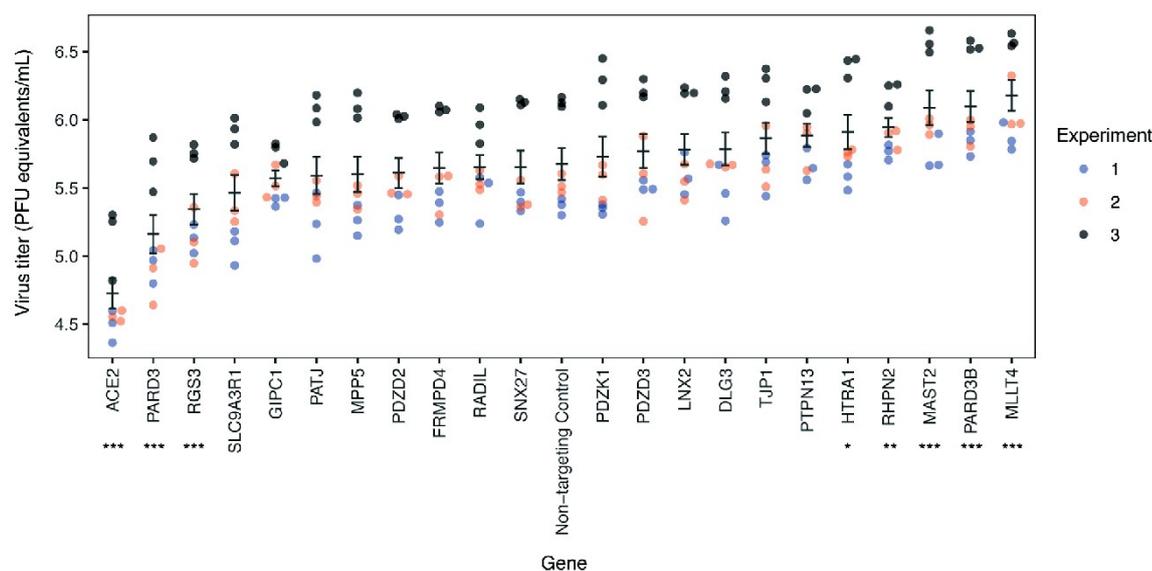


Fig. 4. Effect of knockdown of human PDZ-containing proteins on SARS-CoV-2 replication. The indicated genes were knocked down by siRNA transfection in human lung A549-ACE2 cells. ACE2 knockdown was used as a positive control. Virus load in knocked-down cells was assessed in three independent experiments (each done in triplicate) 72 h postinfection and is shown as PFU equivalents·mL⁻¹ supernatant. The virus load in each condition was compared to that in control siRNA-transfected cells (two-way ANOVA with Dunnett's test for multiple comparisons *** $P \leq 0.001$; ** $P = 0.00611$; * $P = 0.02672$). The data are representative of three independent experiments (each done in triplicate) and error bars correspond to the standard deviation of the three independent experiments.

extended profile of interaction of MERS-CoV E PBM peptide against the PDZome in the holdup assay. We have previously proposed that the large number of cell proteins targeted by WNV motif identified by holdup assay could be related to the broad host repertoire of flaviviruses replicating in very different cells and species necessary for its complex zoonotic transmission cycle. By contrast, SARS-CoV viruses infect very specialized cells, such as bronchial epithelial cells, pneumocytes, and upper respiratory tract cells in humans [30]. This observation can at least partially explain the narrow number of host proteins specifically targeted by SARS-CoV-2 during infection.

Six of the seven host PDZ partners of SARS-CoV E protein are also binders of SARS-CoV-2 E protein, in agreement with the strict conservation of their PBMs. However, the differences of sequence upstream the PBM enlarge the number of binders for SARS-CoV-2 protein E with four more host partners (Fig. 1A). Our results suggest that the SARS-CoV/SARS-CoV-2 PBM upstream sequence difference is responsible of the different binding profiles against the human PDZome. A few studies indicate that residues up to the upstream eleventh position (p-11) can also

be implicated in PDZ binding by modulating the affinity [29,31]. Thus, we can hypothesize that the substitution of Arg (p-6 in SARS-CoV sequence) by Glu (p-7 in SARS-CoV-2 sequence) may affect the potential electrostatic interaction of E protein with the β 2- β 3 loop of some PDZ domains. Interestingly, it has been reported that the substitution of a Glu at position -5 with Arg in the inward rectifier K(+) channel protein GIRK3 disrupts its interaction with the PDZ domain of SNX27 [32]. Here, we also observed that SNX27 binds to SARS-CoV-2 E protein exposing an Arg at p-7, while SARS-CoV E protein coding for Glu at p-6 does not. This example illustrates the fine tuning of the affinity/specificity profiles by the four residues of the PBM and the upstream sequence. Therefore, the charged residue substitution between SARS-CoV and SARS-CoV-2 PBM can modify the selective association of host PDZ proteins with the viral protein. The four host PDZ proteins, SNX27, RADIL, NHERF1, and MAST2 recognized by SARS-CoV-2, and not by SARS-CoV, can play a role in the differences observed between the two infections. The role of these PDZ proteins in viral pathogenesis remains to be explored.

PDZ domain	PDB code	$\alpha 2$ residue facing p-2
TJP1-2	2JWE	Leu
NHERF1-2	2OZF	Tyr
PTPN13-4	c2qg1A*	His
HTRA1	2YTW	Ala
MAST2	2KYL	His
LNK2-2	2VWR	Pro
PARD3-3 #	2K1Z	Asn
SNX27	6SAK	His
RADIL	1UM1	Tyr
MLLT4	1T2M	Gln

*Phyre2 model

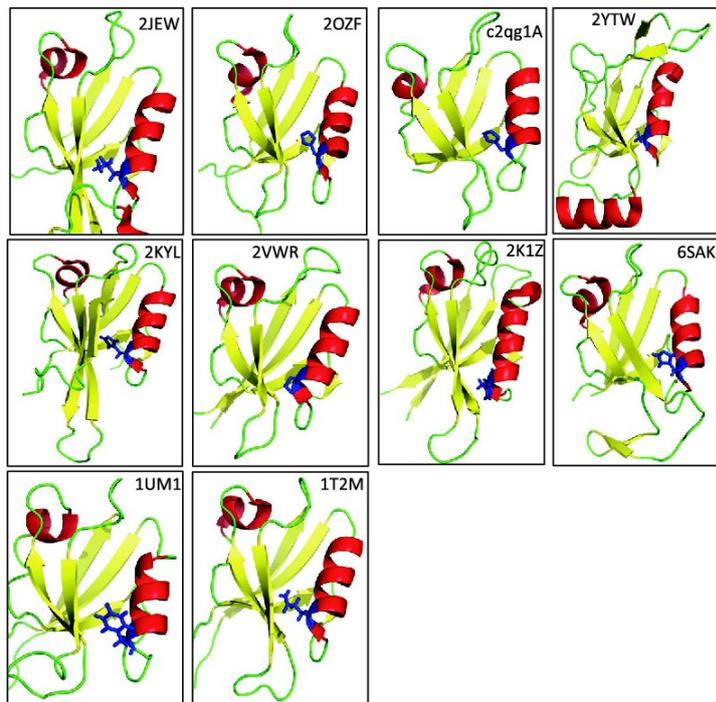
Human sequences excepted # from *Rattus norvegicus*

Fig. 5. Classes of PDZ domains targeted by the SARS-CoV-2 E protein PBM. Left panel: name of selected PDZ domain, PDB code of their structure and the residue of PDZ helix $\alpha 2$ in interaction with p-2 of PBM. Right panel: structural view of the PBM binding groove with the residue of helix $\alpha 2$ facing p-2 of PBM highlighted in blue sticks. *Phyre2 model. Human sequences excepted # from *Rattus norvegicus*

Functions of host partners of SARS E protein

Among the five host proteins already reported as protein E partners (Schoeman *et al.*, 2020) [33], syntenin/SCDP and MPP5/PALS1 encode for a PBM potentially involved in PDZ binding. We did not detect the interaction between syntenin PDZ domain and SARS-CoV-2 E protein. With the holdup assay, we measured a weak interaction between MPP5/PALS1 PDZ domain and SARS-CoV-2 E PBM [dissociation constant (K_d) of 102 μM], while we did not detect interaction between MPP5/PALS1 PDZ domain and SARS-CoV E PBM. Using equilibrium binding titration *in vitro* by Förster Resonance Energy Transfer (FRET), Toto and collaborators [34] have also calculated a better affinity of SARS-CoV-2 for MPP5/PALS1, illustrating the potential role of the upstream substitutions observed in the SARS-CoV-2 C-terminal E protein sequence in comparison with the SARS-CoV one. Teoh and collaborators [18] have proposed that this interaction between SARS-CoV E protein and MPP5/PALS1 alters the tight junctions in the lungs, MPP5/PALS1 being a key component of the Crumbs

complex that controls the apical–basal polarity and tight junction formation. We cannot exclude that the rest of MPP5 and E proteins are required to stabilize their complex, as well as for syntenin that is reported to bind the E protein by strategies using full-length proteins (yeast two-hybrid, reciprocal coimmunoprecipitation and confocal microscopy assays).

Remarkably, we pinpointed that most of binding proteins are involved in cellular junctions and polarity, cellular trafficking, and signaling. Indeed, our results highlighted that SARS-CoV-2 targets several proteins of the apical complex and tight/adherens junctions. Consequently, we expect that the SARS-CoV-2 proteins E, 3a, and N can contribute to globally affect the cellular junctions and cellular polarity. We identified TJP1 as the best binder for E and 3a proteins. TJP proteins, also called ZO proteins, provide a network of scaffold proteins that tether the TM tight junction proteins, comprising the Claudins, to the actin cytoskeleton [35]. The three TJP paralogous proteins, TJP1, TJP2, and TJP3, encode for three PDZ domains in their N-terminal region. In our *in vitro* assay, the

second PDZ of TJP1 displays the highest affinity for E and 3a proteins of SARS-CoV-2 in the micromolar range. This domain has been previously described as the binding site of TJP1 for the connexin43 [36], involved in the organization of the gap junction, but also in the ZO-2 and ZO-3 hetero-association. Dimerization of TJP1 occurs through the interaction between TJP1-PDZ2 domains and the swapped-dimerization of TJP1-PDZ2 regulates the connexin-43 binding [37]. Interestingly, TJP proteins also interact with Afadin/MLLT4 while the PBM sequence of TJP2 binds to the PDZ domain of SNX27 to regulate the trafficking of TJP2 to the tight junction [38,39]. PARD3 is also involved in the formation of adherens and tight junctions. Here, we showed that MLLT4, SNX27, PARD3, and MPP5/PALS1 to a lesser extent are also targeted by SARS-CoV-2 E PBM. By interacting by at least five proteins directly involved in the organization of the tight junction complexes, we expect a global perturbation of the epithelial/endothelial barriers by SARS-CoV-2 during infection. In this direction, SARS-CoV-2 replication causes a transient decrease in cellular barrier function and disruption of tight junctions [40] and a SARS-CoV-2 productive infection locally alters the distribution of TJP1, which associates to tight junctions. The characteristic TJP1 staining pattern at cell boundaries appeared disrupted in areas with viral expression, suggesting a possible impairment of epithelial barrier integrity [40].

Targeting host proteins from polarity and junctions complexes may have a consequence on the morphology of epithelium/endothelium but also on the immune response and cellular homeostasis [41]. In addition to their structural role in many tissues and organs, they can also regulate signaling mechanisms. Tight junctions are linked to many human diseases. Therefore, chronic inflammatory diseases and cancer have been related to dysfunction of tight junctions. MPP5/PALS1 of the Crumbs complex is expressed in T lymphocytes and is required for the T-cell receptor (TCR)-mediated activation [42]. TJP1 and TJP2 are also expressed in T lymphocytes and are relocated to the immune synapse after TCR stimulation. SNX27 regulates trafficking and recycling in T lymphocytes and controls TJP2 mobility at the immune synapse through a PDZ-dependent interaction [43]. LNX1/2 is involved in signaling with involvement in ubiquitination subsequent proteasomal degradation. The PDZ3 domain of PARD3 binds to the transcription factor YAP (Yes-associated protein) [44] regulating the activation of the Hippo pathway, a signaling pathway modulating cell proliferation and cell death. YAP protein can be found associated to the adherens and tight junctions and

several regulators of Hippo pathway are components of the adherens and tight junctions [45]. Finally, NHERF1 regulates the Wnt/b-catenin signaling, a pathway that controls cell growth and proliferation. NHERF1 has been shown to be targeted for degradation by E6 from HPV [46]. RHPN2 (Rho pathway; [47]), MAST2 (anti-survival pathway, PTEN/Akt signaling; [12]), RGS3 (G protein signaling; [48]), PTPN13 (tumor suppressor, Akt signaling; [49]), and HTRA1 (Signaling pathway TGF- β ; [50]) play also function in signaling pathways. All these proteins have been found to significantly interact with SARS-CoV-2 proteins in our holdup assay. Seven of the 14 host targets of SARS-CoV-2 proteins E, 3a, and N significantly impacted viral replication. We showed that knockdown of the PDZ-containing proteins PARD3 and RSG3 results in a decrease of SARS-CoV-2 replication in cells while the silencing of MLLT4, PARD3B, MAST2, RHPN2, and HTRA1 increases virus replication. These proteins are involved in the regulation of cellular junction and polarity, but also in cellular signaling. However, we cannot conclude on the PDZ/PBM dependence of this result, and further, in cell studies are needed to decipher the functional role of these proteins in the viral cycle of SARS-CoV-2.

PDZ proteins recognized by SARS-CoV-2 E protein are also host targets of other viruses

Many studies provide structural and biological evidence that the viral proteins act as competitors endowed with specificity and affinity in an essential cellular process by mimicking PBM of cellular partners. Disruption of critical endogenous protein-protein interactions by viral protein drastically alters intracellular protein trafficking and catalytic activity of cellular proteins that control cell homeostasis. Importantly, most of our selected host PDZ-containing proteins have been previously reported as host proteins targeted by viruses. Tight junctions are known to be targeted by multiple pathogenic viruses [51,52]. PARD3 and TJP1 are targets of flavivirus proteins (WNV, tick-borne encephalitis virus, and Dengue virus) [28,53,54,55]. Others cellular PDZ proteins were also identified as targets for other virus including DLG1 for hepatitis C virus and influenza virus [56,57]; NHERF1, HTRA1, PARD3, SNX27 for HPV [46,58,59,60].

Here, we showed that the three viral PBM-containing proteins expressed by SARS-CoV-2 target several cellular proteins involved in the maintenance and regulation of the tight junctions and immune response. By hijacking junctional proteins and signaling

mechanisms, viruses ensure their propagation through the manipulation of the cellular homeostasis, the immune system, and tissue barriers. The perturbation of cellular fitness promotes their replication and spread. Many PDZ-containing proteins are involved in cell–cell junctions, cell polarity, recycling, or trafficking. Mimicking short linear PBM motifs to target PDZ protein and subvert host functionality is a high adaptation capacity and a significant evolutionary advantage of many viruses [41]. We previously showed that the PBM from the glycoprotein encoded by the rabies virus competes with PTEN for binding to the anti-survival kinase MAST2 and drastically affects the PTEN/Akt signaling pathway. Consequently, the viral glycoprotein/MAST2 interaction promotes survival of the infected neurons to ensure viral propagation. As already observed for other viruses, potential pleiotropic perturbation of cellular functions can be induced by SARS-CoV-2 during infection comprising cell survival/death (MAST2, PARD3, NHERF1), immune response (MPP5/PALS1, syntenin), and cellular organization (TJP1, MLLT4).

Materials and methods

Peptide synthesis

The peptides were synthesized in solid phase using Fmoc strategy (Proteogenix, Schiltigheim, France) encompassing a biotinyl group, a (PEG)₃ or 6-aminohexanoic acid (Ahx) spacer, and the C-terminal PBM sequence of E, 3A, and N proteins (12 residues long). Peptides were resuspended in water or in a mix of DMSO and buffer and used for holdup and competitive fluorescence polarization assays.

Holdup assays

We used an updated version of our original PDZome library [19], that contains 266 soluble PDZs [20]. This PDZome v2 library was prepared as previously described [20].

The holdup assay was carried out with the biotinylated viral PBMs against the full human PDZ library in two independent experiments [19,20]. Error bars are standard deviations of the two experiments. In brief, we measured the fraction of PDZ depletion in the fluid phase during a pull-down experiment. For the detailed protocol, please look at [20]. All His6-MBP-PDZ constructs were expressed in *E. coli* following the previous protocol [19] with minimal modifications and cell lysate soluble fractions of all constructs were adjusted to approximately 4 μM concentration using a microcapillary electrophoretic system (Caliper; PerkinElmer, Waltham, MA, USA) and frozen in 96-well plates. Streptavidin resin was saturated with biotinylated

PBM peptides or with biotin used as a negative control. PDZ-containing cell lysates were incubated with the resins, and the supernatant was separated from the resin by a fast filtration step using filter plates. PDZ concentrations were evaluated using the microcapillary electrophoretic system and BI values were calculated using $\text{BI} = (\text{I}_{\text{tot}} - \text{I}_{\text{depl}}) / \text{I}_{\text{tot}}$, where I_{tot} is the total intensity of the PDZ peak (biotin control) and I_{depl} is the intensity of PDZ peak in the peptide depleted reaction. In the holdup buffer, lysozyme was used as internal standard for peak intensity normalization. $\text{BI} = 0.2$ was used as the minimal BI threshold value to define high-confidence PDZ-PBM pairs, as proposed previously [19]. For the detailed protocols of production and quality control, please look at [20].

Affinity values (Fig. 1, Table 1 and Table S1) were calculated from the BIs according to the method published in Gogl *et al.* [23]. Immobilized peptide concentrations were calculated using measured dissociation constants from competitive fluorescent polarization assays (see below). We have found that all determined $\text{BI}-K_{\text{d}}$ pairs resulted a mean peptide concentration of $\sim 20 \mu\text{M}$, a concentration that is coherent with other peptides in the same method [22,23,61]. The minimal BI threshold value is 0.2 to define a significant interaction, roughly corresponding to a 100 μM dissociation constant as previously reported [19].

Competitive fluorescence polarization assay

Competitive fluorescence polarization measurement and data evaluation were performed as described by Simon and collaborators [62]. Four different fluorescent peptides were used f16E6 (fluorescein-RTRRETQL), fRSK1 (fluorescein-KLPSTTL), fpRSK1 (fluorescein-KLPpSTTL), and MERS-CoV. The MERS-CoV peptide was labeled with sub-stoichiometric FITC (Sigma-Aldrich, St. Louis, MO, USA) in a basic HEPES buffer (pH 8.2), and the reaction was stopped with 100 mM TRIS. The peptide was buffer exchanged in order to remove fluorescent contaminants. Fluorescence polarization was measured with a PHERAstar (BMG Labtech, Offenburg, Germany) microplate reader by using 485 ± 20 and 528 ± 20 nm band-pass filters (for excitation and emission, respectively). In direct FP measurements, a dilution series of the Ni- and Amylose-affinity purified MBP-PDZ proteins was prepared in a 20 mM HEPES pH 7.5 buffer that contained 150 mM NaCl, 0.5 mM TCEP, 0.01% Tween 20, and 50 nM fluorescently labeled peptide. In total, the polarization of the probe was measured at 8 different protein concentrations. In competitive FP measurements, the same buffer was supplemented with the protein to achieve a complex formation of 60–80%, based on the direct titration. Then, this mixture was used for creating a dilution series of the competitor (i.e., the studied peptides) and the measurement was carried out identically as in the direct experiment. Analysis of FP experiments was carried out using ProFit.

Cells

A549 cells stably expressing ACE2 (A549-ACE2, kindly provided by O. Schwartz) were propagated at 37 °C, 5% CO₂ in DMEM with L-glutamine (Gibco, Co Dublin, Ireland) supplemented with 10% FBS, penicillin-streptomycin, and 20 µg·mL⁻¹ blasticidin S.

Virus

The SARS-CoV-2 strain used (BetaCoV/France/IDF0372/2020 strain) is a kind gift from the National Reference Centre for Respiratory Viruses at Institut Pasteur, Paris, originally supplied through the European Virus Archive goes Global platform. It was propagated once in Vero-E6 cells.

siRNA transfections

An OnTargetPlus siRNA SMARTpool library (Horizon Discovery, Waterbeach, UK) was purchased targeting 20 human PDZ-containing proteins (Table S2). This library was arrayed in 96-well format and also included nontargeting siRNAs and siRNA pools targeting ACE2. The siRNA library was transfected into A549-ACE2 cells seeded in a 384-well plate format (6250 cells per well), using Lipofectamine RNAiMAX reagent (Thermo Fisher, Waltham, MA, USA).

Briefly, 4 pmoles of each siRNA pool were mixed with 0.1 µL RNAiMAX transfection reagent and OptiMEM (Thermo Fisher) in a total volume of 10 µL. After a 5-min incubation period, the transfection mix was added to cells. Forty-eight hours post-transfection, the cells were either subjected to SARS-CoV-2 infection or incubated for 72 h to assess cell viability using the CellTiter-Glo luminescent viability assay, following the manufacturer's instructions (Promega, Madison, WI, USA). Luminescence was measured in a Tecan Infinity 2000 plate reader, and percentage viability calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol or 4% formalin (0% viability), included in each experiment.

Virus infections

For infections, the cell culture medium was replaced with virus inoculum [MOI 0.1 plaque forming units (PFU) per cell] and incubated for 1 h at 37 °C, 5% CO₂. After 1 h adsorption period, the inoculum was removed, replaced with fresh media, and cells incubated at 37 °C, 5% CO₂. 72 h postinfection, the cell culture supernatant was harvested, and viral load assessed by RT-qPCR. Briefly, the cell culture supernatant was collected, heat inactivated at 95 °C for 5 min, and used for RT-qPCR analysis. SARS-CoV-2 specific primers targeting the N gene region: 5'-TAATCAGACAAGGAACTGATTA-3' (Forward) and 5'-

CGAAGGTGTGACTTCCATG-3' (Reverse) were used with the Luna Universal One-Step RT-qPCR Kit (New England Biolabs, Courcouronnes, France) in an Applied Biosystems QuantStudio 6 thermocycler, with the following cycling conditions: 55 °C for 10 min, 95 °C for 1 min, and 40 cycles of 95 °C for 10 seconds, followed by 60 °C for 1 min. The number of viral genomes is expressed as PFU equivalents·mL⁻¹ and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titer. The effect of knockdowns on virus replication was compared to control siRNA-transfected cells in three independent experiments (each done in triplicate) with Dunnett's multiple comparison test based on a two-way ANOVA.

Acknowledgements

This work was supported by the URGENCE COVID-19 fundraising campaign of Institut Pasteur, the ANR Recherche Action Covid19-FRM PDZCov2 program and by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 for AFMB. GG is a recipient of the Post-doctorants en France program of the Fondation ARC fellowship. The authors acknowledge Gilles Travé for helpful discussion and for proofreading of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

CC-S, RV, and NW planned the experiments. CC-S, FD, VVR, and GG performed the experiments. CC-S, FD, VVR, GG, QDT, RV, and NW analyzed the data. J-CT and MV contributed reagents or other essential material. CC-S, RV, and NW wrote the paper.

Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/febs.15881>.

References

- 1 Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y & Gao GF (2016) Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* **24**, 490–502.
- 2 Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, Storgaard M, Al Khalili S &

- Simonsen L (2020) Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis* **20**, e238–e244.
- 3 Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL *et al.* (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468.
 - 4 Stukalov A, Girault V, Grass V, Bergant V, Karayel O, Urban C, Haas DA, Huang Y, Oubraham L, Wang A *et al.* (2020) Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *bioRxiv* [PREPRINT]
 - 5 Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, Cinatl J & Münch C (2020) Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472.
 - 6 Weatheritt RJ, Luck K, Petsalaki E, Davey NE & Gibson TJ (2012) The identification of short linear motif-mediated interfaces within the human interactome. *Bioinformatics* **28**, 976–982.
 - 7 Davey NE, Travé G & Gibson TJ (2011) How viruses hijack cell regulation. *Trends Biochem Sci* **36**, 159–169.
 - 8 Liu X & Fuentes EJ (2019) Emerging themes in PDZ domain signaling: structure, function, and inhibition. *Int Rev Cell Mol Biol* **343**, 129–218.
 - 9 Christensen NR, Čalyševa J, Fernandes EFA, Lüchow S, Clemmensen LS, Haugaard-Kedström LM & Strømgaard K (2019) PDZ domains as drug targets. *Adv Ther (Weinh)* **2**, 1800143.
 - 10 Javier RT & Rice AP (2011) Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J Virol* **85**, 11544–11556.
 - 11 Caillet-Saguy C, Maisonneuve P, Delhommel F, Terrien E, Babault N, Lafon M, Cordier F & Wolff N (2015) Strategies to interfere with PDZ-mediated interactions in neurons: what we can learn from the rabies virus. *Prog Biophys Mol Biol* **119**, 53–59.
 - 12 Terrien E, Chaffotte A, Lafage M, Khan Z, Préhaud C, Cordier F, Simenel C, Delepierre M, Buc H, Lafon M *et al.* (2012) Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci Signal* **5**, ra58.
 - 13 Maisonneuve P, Caillet-Saguy C, Vaney M-C, Edoe B-Z, Sawyer K, Raynal B, Delepierre M, Lafon M, Cordier F & Wolff N (2016) Molecular basis of the interaction of the human protein tyrosine phosphatase non-receptor type 4 (PTPN4) with the mitogen-activated protein kinase p38 γ . *J Biol Chem* **291**, 16699–16708.
 - 14 Li Y, Surya W, Claudine S & Torres J (2014) Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J Biol Chem* **289**, 12535–12549.
 - 15 Ruch TR & Machamer CE (2012) The coronavirus E protein: assembly and beyond. *Viruses* **4**, 363–382.
 - 16 Schoeman D & Fielding BC (2019) Coronavirus envelope protein: current knowledge. *Virol J* **16**, 69.
 - 17 Jimenez-Guardeño JM, Nieto-Torres JL, DeDiego ML, Regla-Nava JA, Fernandez-Delgado R, Castaño-Rodríguez C & Enjuanes L (2014) The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS Pathog* **10**, e1004320.
 - 18 Teoh K-T, Siu Y-L, Chan W-L, Schlüter MA, Liu C-J, Peiris JSM, Bruzzone R, Margolis B & Nal B (2010) The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Mol Biol Cell* **21**, 3838–3852.
 - 19 Vincentelli R, Luck K, Poirson J, Polanowska J, Abdal J, Blémont M, Turchetto J, Iv F, Ricquier K, Straub M-L *et al.* (2015) Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat Methods* **12**, 787–793.
 - 20 Duhoo Y, Girault V, Turchetto J, Ramond L, Durbesson F, Fourquet P, Nominé Y, Cardoso V, Sequeira AF, Brás JLA *et al.* (2019) High-throughput production of a new library of human single and tandem PDZ domains allows quantitative PDZ-peptide interaction screening through high-throughput holdup assay. *Methods Mol Biol* **2025**, 439–476.
 - 21 Amacher JF, Brooks L, Hampton TH & Madden DR (2020) Specificity in PDZ-peptide interaction networks: computational analysis and review. *J Struct Biol X* **4**, 100022.
 - 22 Jané P, Gógl G, Kostmann C, Bich G, Girault V, Caillet-Saguy C, Eberling P, Vincentelli R, Wolff N, Travé G *et al.* (2020) Interactomic affinity profiling by holdup assay: acetylation and distal residues impact the PDZome-binding specificity of PTEN phosphatase. *PLoS One* **15**, e0244613.
 - 23 Gogl G, Jane P, Caillet-Saguy C, Kostmann C, Bich G, Cousido-Siah A, Nyitray L, Vincentelli R, Wolff N, Nomine Y *et al.* (2020) Dual specificity PDZ- and 14-3-3-binding motifs: a structural and interactomics study. *Structure* **28**, 747–759.e3.
 - 24 Castaño-Rodríguez C, Honrubia JM, Gutiérrez-Álvarez J, DeDiego ML, Nieto-Torres JL, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, Verdía-Báguena C, Queralt-Martín M *et al.* (2018) Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis. *MBio* **9**, 1–23.
 - 25 Genera M, Quioc-Salomon B, Nourisson A, Colcombet-Cazenave B, Haouz A, Mechaly A, Matondo M, Duchateau M, König A, Windisch MP *et al.* (2021) Molecular basis of the interaction of the human tyrosine phosphatase PTPN3 with the hepatitis B virus core protein. *Sci Rep* **11**, 944.

- 26 Buchrieser J, Dufloo J, Hubert M, Monel B, Planas D, Rajah MM, Planchais C, Porrot F, Guivel-Benhassine F, Van der Werf S *et al.* (2020) Syncytia formation by SARS-CoV-2-infected cells. *EMBO J* **39**, e106267.
- 27 Hassan SS, Choudhury PP & Roy B (2020) Molecular phylogeny and missense mutations at envelope proteins across coronaviruses. *Genomics* **112**, 4993–5004.
- 28 Giraud E, Del Val CO, Caillet-Saguy C, Zehrouni N, Khou C, Caillet J, Jacob Y, Pardigon N & Wolff N (2021) Role of PDZ-binding motif from West Nile virus NS5 protein on viral replication. *Sci Rep* **11**, 3266.
- 29 Luck K, Charbonnier S & Travé G (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett* **586**, 2648–2661.
- 30 V'kovski P, Kratzel A, Steiner S, Stalder H & Thiel V (2020) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155–170.
- 31 Delhommel F, Chaffotte A, Terrien E, Raynal B, Buc H, Delepierre M, Cordier F & Wolff N (2015) Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem J* **469**, 159–168.
- 32 Balana B, Maslennikov I, Kwiatkowski W, Stern KM, Bahima L, Choe S & Slesinger PA (2011) Mechanism underlying selective regulation of G protein-gated inwardly rectifying potassium channels by the psychostimulant-sensitive sorting nexin 27. *Proc Natl Acad Sci USA* **108**, 5831–5836.
- 33 Schoeman D & Fielding BC (2020) Is There a Link Between the Pathogenic Human Coronavirus Envelope Protein and Immunopathology? A Review of the Literature. *Front Microbiol* **11**, 2086.
- 34 Toto A, Ma S, Malagrino F, Visconti L, Pagano L, Stromgaard K & Gianni S (2020) Comparing the binding properties of peptides mimicking the Envelope protein of SARS-CoV and SARS-CoV-2 to the PDZ domain of the tight junction-associated PALS1 protein. *Protein Sci* **29**, 2038–2042.
- 35 Otani T & Furuse M (2020) Tight junction structure and function revisited. *Trends Cell Biol* **30**, 805–817.
- 36 Giepmans BN & Moolenaar WH (1998) The gap junction protein connexin43 interacts with the second PDZ domain of the zona occludens-1 protein. *Curr Biol* **8**, 931–934.
- 37 Chen J, Pan L, Wei Z, Zhao Y & Zhang M (2008) Domain-swapped dimerization of ZO-1 PDZ2 generates specific and regulatory connexin43-binding sites. *EMBO J* **27**, 2113–2123.
- 38 Rouaud F, Sluysmans S, Flinois A, Shah J, Vasileva E & Citi S (2020) Scaffolding proteins of vertebrate apical junctions: structure, functions and biophysics. *Biochim Biophys Acta Biomembr* **1862**, 183399.
- 39 Zimmerman SP, Hueschen CL, Malide D, Milgram SL & Playford MP (2013) Sorting nexin 27 (SNX27) associates with zonula occludens-2 (ZO-2) and modulates the epithelial tight junction. *Biochem J* **455**, 95–106.
- 40 Robinot R, Hubert M, de Melo GD, Lazarini F, Bruel T, Smith N, Levallois S, Larrous F, Fernandes J, Gellenoncourt S *et al.* (2020) SARS-CoV-2 infection damages airway motile cilia and impairs mucociliary clearance. *bioRxiv* [PREPRINT].
- 41 Gutiérrez-González LH & Santos-Mendoza T (2019) Viral targeting of PDZ polarity proteins in the immune system as a potential evasion mechanism. *FASEB J* **33**, 10607–10617.
- 42 Carvalho G, Poalas K, Demian C, Hatchi E, Vazquez A & Bidère N (2011) Participation of the cell polarity protein PALS1 to T-cell receptor-mediated NF- κ B activation. *PLoS One* **6**, e18159.
- 43 Tello-Lafoz M, Martínez-Martínez G, Rodríguez-Rodríguez C, Albar JP, Huse M, Gharbi S & Merida I (2017) Sorting nexin 27 interactome in T-lymphocytes identifies zona occludens-2 dynamic redistribution at the immune synapse. *Traffic* **18**, 491–504.
- 44 Zhang P, Wang S, Wang S, Qiao J, Zhang L, Zhang Z & Chen Z (2016) Dual function of partitioning-defective 3 in the regulation of YAP phosphorylation and activation. *Cell Discov* **2**, 16021.
- 45 Yu F-X & Guan K-L (2013) The Hippo pathway: regulators and regulations. *Genes Dev* **27**, 355–371.
- 46 Drews CM, Case S & Pol SBV (2019) E6 proteins from high-risk HPV, low-risk HPV, and animal papillomaviruses activate the Wnt/ β -catenin pathway through E6AP-dependent degradation of NHERF1. *PLoS Pathog* **15**, e1007575.
- 47 Danussi C, Akavia UD, Niola F, Jovic A, Lasorella A, Pe'er D & Iavarone A (2013) RHPN2 drives mesenchymal transformation in malignant glioma by triggering RhoA activation. *Cancer Res* **73**, 5140–5150.
- 48 Dunn HA & Ferguson SSG (2015) PDZ protein regulation of G protein-coupled receptor trafficking and signaling pathways. *Mol Pharmacol* **88**, 624–639.
- 49 Long Q, Sun J, Lv J, Liang Y, Li H & Li X (2020) PTPN13 acts as a tumor suppressor in clear cell renal cell carcinoma by inactivating Akt signaling. *Exp Cell Res* **396**, 112286.
- 50 Li Y, Yuan J, Rothzerg E, Wu X, Xu H, Zhu S & Xu J (2020) Molecular structure and the role of high-temperature requirement protein 1 in skeletal disorders and cancers. *Cell Prolif* **53**, e12746.
- 51 Torres-Flores JM & Arias CF (2015) Tight junctions go viral!. *Viruses* **7**, 5145–5154.
- 52 Zihni C, Mills C, Matter K & Balda MS (2016) Tight junctions: from simple barriers to multifunctional molecular gates. *Nat Rev Mol Cell Biol* **17**, 564–580.
- 53 Ellencrona K, Syed A & Johansson M (2009) Flavivirus NS5 associates with host-cell proteins zonula occludens-1 (ZO-1) and regulating synaptic membrane exocytosis-

- 2 (RIMS2) via an internal PDZ binding mechanism. *Biol Chem* **390**, 319–323.
- 54 Melik W, Ellencrona K, Wigerius M, Hedström C, Elväng A & Johansson M (2012) Two PDZ binding motifs within NS5 have roles in Tick-borne encephalitis virus replication. *Virus Res* **169**, 54–62.
- 55 Werme K, Wigerius M & Johansson M (2008) Tick-borne encephalitis virus NS5 associates with membrane protein scribble and impairs interferon-stimulated JAK-STAT signalling. *Cell Microbiol* **10**, 696–712.
- 56 Awad A, Sar S, Barré R, Cariven C, Marin M, Salles JP, Erneux C, Samuel D & Gassama-Diagne A (2013) SHIP2 regulates epithelial cell polarity through its lipid product, which binds to Dlg1, a pathway subverted by hepatitis C virus core protein. *Mol Biol Cell* **24**, 2171–2185.
- 57 Golebiewski L, Liu H, Javier RT & Rice AP (2011) The avian influenza virus NS1 ESEV PDZ binding motif associates with Dlg1 and Scribble to disrupt cellular tight junctions. *J Virol* **85**, 10639–10648.
- 58 Stuqui B, Conceição ALG, Termini L, Sichero L, Villa LL, Rahal P, de Calmon M & Calmon MF (2016) The differential role of HTRA1 in HPV-positive and HPV-negative cervical cell line proliferation. *BMC Cancer* **16**, 840.
- 59 Facciuto F, Bugnon Valdano M, Marziali F, Massimi P, Banks L, Cavatorta AL & Gardiol D (2014) Human papillomavirus (HPV)-18 E6 oncoprotein interferes with the epithelial cell polarity Par3 protein. *Mol Oncol* **8**, 533–543.
- 60 Ganti K, Massimi P, Manzo-Merino J, Tomaić V, Pim D, Playford MP, Lizano M, Roberts S, Kranjec C, Doorbar J *et al.* (2016) Interaction of the human papillomavirus E6 oncoprotein with sorting nexin 27 modulates endocytic cargo transport pathways. *PLoS Pathog* **12**, e1005854.
- 61 Gógl G, Biri-Kovács B, Durbesson F, Jane P, Nomine Y, Kostmann C, Bilics V, Simon M, Reményi A, Vincentelli R *et al.* (2019) Rewiring of RSK-PDZ interactome by linear motif phosphorylation. *J Mol Biol* **431**, 1234–1249.
- 62 Simon MA, Ecsédi P, Kovács GM, Póti ÁL, Reményi A, Kardos J, Gógl G & Nyitray L (2020) High-throughput competitive fluorescence polarization assay reveals functional redundancy in the S100 protein family. *FEBS J* **287**, 2834–2846.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. All BI values obtained by holdup assay using PBM peptides from SARS-CoV, SARS-Cov-2, and MERS-CoV E proteins, and from SARS-CoV-2 N and 3a proteins.

Table S2. OnTargetPlus siRNA SMARTpool library (Horizon Discovery) was purchased targeting the 20 selected human PDZ-containing proteins.

Abstract

RNA viruses are one of the fastest evolving biological entities. With their high mutation and recombination rates, small genomes, and fast replication rates, they easily adapt to new hosts, causing outbreaks regularly. Despite their simplicity, their ability to interact with each other through complementation and other forms of collective behaviors gives rise to complex population dynamics. In this work, I explore two aspects of these social interactions and their influence on virus evolution, using two arthropod-borne viruses.

A first part of this manuscript investigates the role of defective viral genomes (DVGs), which were typically described as ubiquitous deleterious by-products of the erroneous RNA-dependent RNA-polymerase of RNA viruses. In this work, we tested if these DVGs could positively interact with standard viral genomes. We first identified a ribavirin-resistance mutation in the viral polymerase that rescued viral titers in mutagenic conditions. We then introduced it in a previously described DVG of Zika virus, and showed that DVGs carrying this mutation accelerated viral population adaptation to ribavirin treatment through genetic recombination. We then showed that absence of observed complementation was due to a *trans*-complementation incompetent NS5 polymerase domain. Finally, we showed that DVGs could not only be *trans*-packaged, but they could also be co-packaged with standard genomes, ensuring their maintenance across new infectious cycles.

In the second half, I re-explored the question of the arbovirus trade-off hypothesis which is hypothesized to explain the low substitution rates of arboviruses in nature. Using barcoded chikungunya virus clones and high-throughput sequencing techniques, I proposed an extended experimental design to resolve conclusion discrepancies among previous studies. In a first phase of experimental evolution, I identified that genetic diversity was accumulated differentially in-between genomic regions and among passaging regimes. Alternating passages were restricted in genetic diversity by the most stringent host at each locus, and did not tolerate host-specific DVGs. Mutant cloud analysis showed that they in part acquired insect-associated adaptive mutations, but also alternation-specific ones. This pattern was in turn explained by the observation of an asymmetric trade-off: initial adaptation to mammalian cells has a fitness cost in insect cells, blocking such mutations to appear in alternation. Further, no cost of generalism was observed, which we propose is explained by a different fitness landscape neighborhood around the insect-cell adaptive pathway.

Keywords: chikungunya virus, Zika virus, RNA virus, experimental evolution, sociovirology, defective viral genomes, complementation, recombination, trade-off, arbovirus, high-throughput sequencing