A molecular mechanism for the quantitativemodel of the cell cycle : CDK-mediatedphosphorylation of intrinsically disordered regions

Geronimo Dubra

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale de Sciences Chimiques et Biologiques pour la Santé

Unité de recherche IGMM – Institut de Génétique Moléculaire de Montpellier

Un mécanisme moléculaire pour le modèle quantitatif du cycle cellulaire : La phosphorylation par les CDKs des régions intrinsèquement non structurées.

A molecular mechanism for the quantitative model of the cell cycle: CDK-mediated phosphorylation of intrinsically disordered regions.

Présentée par Geronimo Dubra

Le 23 septembre 2022

Sous la direction de Daniel Fisher
et la supervision de Liliana Krasinska

Devant le jury composé de

Dre. Ariane Abrieu, CRBM - Montpellier	Examinatrice
Dr. David Santamaria, CIC - Universidad de Salmanca	Rapporteur
Dr. Helfrid Hochegger, University of Sussex	Rapporteur
Dr. Daniel Fisher, IGMM - Montpellier	Directeur de thèse

UNIVERSITÉ DE MONTPELLIER

“All models are wrong, but some are useful”

George E. P. Box, British statistician.

Abstract

Eukaryotic cell cycle progression is controlled by Cyclin-Dependent Kinase family (CDK). These enzymes in complex with their mandatory binding partners, the cyclins, phosphorylate substrates to progress through the different phases of the cell cycle. The canonical view of the cell cycle states that different combinations between the members of CDK and cyclin families confer these complexes with the specificity needed to select their substrates and, therefore, promote specific reactions in each phase. An alternative model proposes that most CDK-cyclin function is redundant and it is the global CDK activity levels that drive the cell cycle. This so-called "quantitative model", implies that there exist low and high overall CDK activity thresholds for entry into S-phase and mitosis, respectively, determined by the CDK regulatory network.

Genetic systems that allow scientists to manipulate individual CDK levels are pivotal to addressing critical questions that help to elucidate which model explains better the current body of evidence. Here I present my advances in the meticulous design of such systems to be introduced into cells with different genetic contexts and into model organisms, as well.

One of the main dilemmas to be solved when considering the quantitative model is how a global activity can precisely control all the biochemical states of the cell during the cell cycle. Here, I suggest a mechanism of action by which phosphorylation of intrinsically disordered regions of proteins control the formation and dissolution of protein condensates acting as biochemical hubs in the cell. Not only CDKs but most cell cycle kinases share the tendency of phosphorylating disordered regions and their substrates contain more of these regions that the rest of the phosphorylated proteins. Moreover, a striking proportion of proteins in the protein condensates are CDK targets. We collaborated with a phosphoproteomics team and we obtained a high-resolution phosphorylation map through the first cell division of single embryos of Xenopus laevis. We confirmed that before mitosis, a rapid increase of global phosphorylation occurs, where most of those appear to be CDK-mediated. We detected a high number of interphase phosphorylations that were enriched in CDK phosphorylation motifs and other cell cycle kinases such as Aurora.
We selected the proliferation marker Ki-67 as a case study to investigate how phosphorylation can regulate the process of phase separation, responsible for the formation of the molecular condensates. Ki67 appears to present competing modes of regulation of phase separation by phosphorylation of its repeat domain, which will depend on the cellular and the

molecular context. It appears that different levels of phosphorylation will differentially localize Ki-67 to the perinucleolar heterochromatin during interphase and to the perichromosomal layer in mitosis, both described as being phase-separated.

The theory of global action of the CDK phosphorylation in controlling the formation of biochemical centers where cell cycle-specific reaction happens lacks, however, a detailed explanation of the downstream effects of this regulation. Taking Ki-67 as an effector of CDK-mediated regulation of phase separation, we investigated what are the changes in chromatin organization in cells lacking Ki-67. We found that knockout of Ki-67 produces massive changes in the transcriptome, partially due to changes in chromatin histone marks, especially the inhibitory H3K27 trimethylation. No visible alteration was detected in cells, except for the striking decrease of tumorigenic properties of cancer cells when injected into mice.
Another mechanism of action of CDK-mediated phosphorylation of disordered regions of proteins might be the regulation of the activity of condensates in which they participate actively. Members of the CDK8/19 subfamily in complex with cyclin C constitute the kinase domain of the Mediator complex, which phosphorylates the disordered C-terminal domain of Pol II. Several other subunits of the Mediator are also reported to be disordered and this complex appears to undergo phase separation and form the so-called “super-enhancers”. The deletion of both CDK8 and CDK19 produces transcriptional alterations in multiple genes, although those changes are rather low in magnitude. No strong phenotype is observed in any of the models tested, and the transcriptional changes observed seem to diverge considerably depending on the cell type observed, although they seem to be related to cell identity transcriptional programs. In summary, the loss of both kinases affects the proliferation of intestinal organoids and brings about a cystic fibrosis-like phenotype; it protects Hepatic cells against malignant transformation; and regulates stress responses in mouse embryonic fibroblasts.

Résumé (français)

La progression du cycle cellulaire eucaryote dépend de la sous-famille CDK1 des CDK et est postulée de résulter du comportement collectif d'états de phosphorylation de protéines modifiés. À l'exception notable de CDK1, la plupart des gènes de CDK et de cycline ne sont pas nécessaires pour la prolifération cellulaire dans la majorité des types de cellules chez la souris, tandis que chez la levure de fission, l'activité oscillante de CDK1 seule peut entraîner l'ensemble du cycle cellulaire. Cela suggère qu'un réseau assez limité de CDKs peut diriger le cycle cellulaire eucaryote et que les changements dans l'activité globale des CDK déterminent en la séquence des processus complexes nécessaires pour dupliquer le génome et distribuer les composants cellulaires pendant la division cellulaire. Ce "modèle quantitatif" implique qu'il existe des seuils d'activité globale des CDK bas et élevés pour l'entrée en phase S et en mitose, respectivement, déterminés par le réseau de régulation des CDK. Ce réseau implique des boucles de rétroaction positive et double-négative, ainsi que des cycles futiles de l'activité des CDK et des phosphatases s'opposant aux CDK. La modélisation mathématique montre que ces caractéristiques de l'organisation du réseau peuvent générer une ultra-sensibilité et une hystérésis dans l'activation de CDK1, tandis que la bistabilité de l'activité de CDK1 qui en résulte conduit à une transition G2/M de type interrupteur ou “*switch-like*”. Ces concepts théoriques sont étayés par des évidences expérimentales dans des extraits d'œufs de *Xenopus* et des cellules de mammifères. La dynamique présumée du réseau de régulation de CDK1 est cohérente avec la réorganisation morphologique abrupte de la cellule lors de la mitose. Chez les métazoaires, l'enveloppe et la lamelle nucléaires se décomposent et de nombreuses structures cellulaires sont rapidement dissociées. Il s'agit notamment des complexes de pores nucléaires, des nucléoles, du matériel péricentriolaire, des tâches d'épissage, des corps de Cajal, des corps nucléaires de la leucémie promyélocytaire (PML) et des granules de stress, qui ont été désignés collectivement comme des organites sans membrane (MLO). Ainsi, l'assemblage et le désassemblage des MLO se produisent de manière dépendante du cycle cellulaire. On pense que les MLO s'assemblent par des mécanismes impliquant des interactions multivalentes entre les régions intrinsèquement désordonnées (IDR) des protéines, et que ce processus peut être régulé par des protéines kinases, notamment les CDK. La phosphorylation des protéines, en général, est enrichie en régions intrinsèquement désordonnées et cela semble également vrai pour les CDK. En tant que tel, un modèle attrayant est que la phosphorylation des IDRs médiée par les CDKs pourrait déclencher la dissolution rapide de nombreux MLOs lors de la mitose. Cela serait cohérent avec le fait que les CDK de la famille CDK1 peuvent phosphoryler des centaines de sites sur diverses

protéines, et réguler la réplication de l'ADN, la mitose, la transcription, le remodelage de la chromatine, la réparation de l'ADN, le cytosquelette, le transport nucléaire, la traduction des protéines, la formation du fuseau mitotique et même la ciliogenèse.

L'objectif principal de cette thèse est de développer davantage le concept de "modèle quantitatif" pour le cycle cellulaire. Inspiré par des travaux antérieurs sur la levure, qui présentaient un système génétique-chimique permettant de contrôler de manière externe le cycle cellulaire, j'ai cherché à concevoir un outil similaire pour commander de manière externe l'activité endogène de CDK1 dans les cellules humaines. Je présente ici la conception d'une construction génétique à insérer silencieusement comme un intron artificiel dans un exon du locus du gène *CDK1*. Ce système peut passer de l'activité endogène de CDK1 à une version sensible aux analogues de CDK1 fusionnée soit avec la cycline B1, soit avec la cycline A2, par recombinaison médiée par l'activité Cre. Sa conception et sa construction ont constitué un véritable “tour de force”*,* où les concepts de base du génie génétique ont été appliqués pour générer un outil complexe de contrôle du cycle cellulaire. En dépit du fait que je n'ai pas été en mesure de conclure ce projet au cours de mon doctorat, des progrès importants ont été accomplis et plusieurs questions essentielles concernant le modèle quantitatif pourraient être résolues grâce à cette stratégie.

Des travaux antérieurs ont rapporté que CDK1 est le seul membre essentiel de la famille et qu'il suffit à diriger l'ensemble du cycle cellulaire chez la souris. En outre, CDK2 ne peut pas compenser son activité même lorsqu'elle est située dans le locus de *CDK1*, ce qui indique que c'est la protéine plutôt que sa régulation transcriptionnelle qui rend CDK1 essentielle. Néanmoins, il n'a jamais été démontré que le cycle cellulaire des mammifères pouvait être piloté par un seul complexe CDK-cycline comme cela a été montré chez la levure. Ce système chimique-génétique nous permettrait de répondre à cette question, ainsi qu'à d'autres, qui sont d'une grande importance pour la validation de la théorie du "modèle quantitatif", selon laquelle c'est le niveau de phosphorylation global des CDK, donné par le rapport d'activité CDK/Phosphatase, plutôt que des combinaisons spécifiques de CDK et de cyclines, qui entraîne les différents événements du cycle cellulaire. Des travaux récents dans ce domaine ont présenté des preuves soutenant l'idée que les différents complexes CDK-cyclines ne sont pas strictement essentiels et qu'ils pourraient n'être importants que pour réguler les niveaux globaux des activités des CDK. Il a été signalé qu'en l'absence de CDK1, CDK2 peut assumer le rôle de la kinase mitotique lorsqu'elle est surexprimée pour générer des niveaux de phosphorylation similaires à ceux observés pour CDK1. De plus, il a été récemment démontré que l'initiation de la phase S, historiquement attribuée au complexe CDK2-cycline E, se produit en l'absence de CDK2, et nécessite CDC7 ou CDK1, très probablement en complexe avec la cycline B, pour phosphoryler les protéines du complexe MCM.

Avec l'introduction de notre système génétique dans des cellules humaines, nous serons en mesure de réguler de manière externe l'activité cellulaire de CDK1 de façon dose-dépendante par l'ajout d'un inhibiteur spécifique. Avec des knockouts supplémentaires d'autres gènes CDK du cycle cellulaire, nous espérons pouvoir contrôler le cycle cellulaire en modulant l'activité d'un seul complexe CDK-cycline. Il est également intéressant de noter que ce système nous permettra d'étudier les conditions dans la cellule à un degré intermédiaire d'activité CDK. Si l'on considère les preuves précédentes, la phosphorylation médiée par les CDK pourrait fonctionner de manière bistable et les niveaux intermédiaires de CDK pourraient être une situation artificielle, on pourrait donc s'interroger sur la pertinence d'un système de modulation dépendant de la dose. Cependant, cela nous permettrait de mieux comprendre comment différents niveaux de phosphorylation de CDK1 déclenchent différentes phases du cycle cellulaire et d'observer les changements cellulaires pendant les transitions entre ces niveaux qui, autrement, se produisent trop rapidement pour être détectés. La modulation de l'activité de CDK1 pourrait également être cruciale dans un contexte où un seul complexe de fusion CDK1-cycline dirige le cycle cellulaire si son expression est régulée par le locus du gène *CDK1*. Les protéines CDK1-cyclines de fusion peuvent être dégradées par l'APC/C, ce qui permet la transition métaphase-anaphase. Cependant, il n'est pas clair si les cellules seraient capables d'effectuer un cycle sans la contribution de la régulation de l'expression des cyclines à l'oscillation de l'activité des CDK. Dans ce cas, l'activité des CDK peut être contrôlée de manière externe par des concentrations alternées d'un inhibiteur spécifique.

Au cours de mes études doctorales, un article a été publié dans lequel on a réussi à remplacer le locus *CDK1* par une construction génétique exprimant AS-CDK1 chez la souris. Malheureusement, cela est létal pour les embryons et il est impossible d'obtenir des animaux homozygotes portant cette construction. Néanmoins, ils ont obtenu des cellules souches embryonnaires (ESC) homozygotes AS-CDK1 qui ont été utilisées pour étudier les rôles de CDK1 dans la régulation épigénétique du développement. Cela m'a incité à réexaminer notre commutateur génétique pour l'adapter au fonctionnement des souris. De cette manière, les animaux présentant des allèles homozygotes avec le commutateur CDK1 peuvent être amenés à terme, puis la recombinase Cre peut être exprimée dans un tissu particulier pour étudier les effets dose-dépendants de l'inhibition spécifique de CDK1. Le fait de travailler avec des animaux nous donne la possibilité de croiser des individus hétérozygotes, ce qui simplifie la stratégie d'insertion : nous n'avons plus besoin de cibler les exons pour nous assurer que les allèles seront nuls s'ils ne sont pas mutés. Une approche similaire au système génétique original proposé ici a été soumise à un appel à propositions pour développer des modèles de souris génétiquement modifiées et a été acceptée pour être financée. Cela deviendra un outil essentiel pour le

laboratoire afin de développer un modèle de cycle cellulaire actualisé qui tienne compte de l'ensemble des preuves récentes.

Nous avons supposé qu'un mécanisme d'action possible pour le modèle quantitatif de contrôle du cycle cellulaire pourrait être qu'une augmentation coordonnée des niveaux globaux de phosphorylation peut générer des changements physico-chimiques dans les condensats de protéines en régulant leur formation et leur dissolution. Plusieurs processus biochimiques se produisant pendant l'interphase sont signalés comme se déroulant dans des organites sans membrane (MLOs) qui se produisent via la séparation des phases. La plupart de ces MLO sont dissous à l'entrée en mitose, tandis que la formation d'autres condensats est observée, notamment les centrosomes, les kinétochores et la couche péri chromosomique, et dont les composants subiraient une séparation de phase. Non seulement ces condensats des protéines sont régulés selon le cycle cellulaire, mais dans la plupart des cas, les protéines clés conduisant leur formation sont phosphorylées principalement par les CDK, et peut-être par d'autres kinases liées au cycle cellulaire. Nous avons donc émis l'hypothèse que les changements globaux de phosphorylation des CDK régulent la réorganisation des compartiments non membranaires de la cellule pour effectuer différentes réactions biochimiques au cours des différentes phases du cycle cellulaire.

Les deux principales caractéristiques moléculaires qui déterminent si une protéine peut être soumise à une séparation de phases sont la multivalence et le désordre intrinsèque. Les régions des protéines présentant ces caractéristiques sont capables de former des condensats macromoléculaires avec d'autres molécules par le biais d'interactions homotypiques et hétérotypiques. Il a été signalé que la phosphorylation se produit de préférence dans les régions désordonnées et que, de fait, l'activité kinase peut moduler le processus de séparation de phase. Pour vérifier si c'était le cas pour la phosphorylation médiée par les CDK, nous avons d'abord compilé une base de données de substrats de la famille CDK1 de haute confiance chez l'homme et la levure. J'ai observé dans les bases de données et les publications un biais de découverte clair en faveur des phosphorylations dirigées par les prolines dans la position +1 pour les CDK, et les sites non dirigés par les prolines sont presque toujours filtrés de l'analyse des substrats des CDK.

J'ai ensuite calculé les régions désordonnées de toutes les protéines détectées comme étant phosphorylées dans nos ensembles de données sur la levure et l'homme en utilisant de multiples prédicteurs de désordre. Je montre que la phosphorylation, en général, est significativement enrichie dans les régions désordonnées des protéines, conformément aux rapports précédents, bien que nos résultats aient été obtenus en normalisant les données de manière à prendre en

compte la composition d'acides aminés du désordre et à éliminer ainsi le biais que d'autres études pourraient présenter en raison de la composition différentielle des IDR. Il existe également un risque accru de phosphorylation médiée par la CDK dans les IDR, tant dans les ensembles de données sur la levure que sur l'homme. Étonnamment, cette tendance était analogue à celle observée pour d'autres kinases liées au cycle cellulaire et pour les membres de la famille des MAPK kinases qui, bien que n'étant pas des kinases du cycle cellulaire, partagent la tendance à phosphoryler des sites dirigés par des prolines. Ensuite, j'ai estimé la distribution du pourcentage de désordre pour chaque ensemble de données des CDK sur l'humain et la levure et je les ai comparés au reste de leurs protéines phosphorylées. J'ai constaté une augmentation d'environ 2 fois de la proportion de désordre dans les cibles CDK par rapport au reste du phosphoprotéome. C'était également le cas pour les autres kinases du cycle cellulaire, et bien que les cibles des MAPK aient montré un pourcentage de désordre plus élevé que le reste du phosphoprotéome humain, elles étaient nettement moins désordonnées que les cibles des CDK, PLK, AURK et DYRK. Même si les MAP kinases et les CDK partagent le même motif de reconnaissance dirigé par les prolines et présentent toutes deux un risque accru comparable de phosphoryler des régions désordonnées, les substrats des MAPK présentent notamment une proportion de résidus désordonnés inférieure à celle des cibles des CDK. Cela pourrait signifier que l'évolution façonne les protéines régulées par le cycle cellulaire pour qu'elles aient une plus grande proportion de régions désordonnées, présentant ainsi une proportion accrue de leur chaîne d'acides aminés qui peut être accessible aux kinases spécialisées et par conséquent phosphorylée.

Nous avons établi une ligne de travail en collaboration avec Maarten Altelaar et Juan Manuel Valverde, qui nous ont fourni un ensemble de données supplémentaires à inclure dans cette analyse. Ils ont obtenu des données de spectrométrie de masse de la phosphorylation sur les protéines qui proviennent d'un seul embryon de *Xenopus,* dans ses premiers stades de l'embryogenèse. Ce système naturel hautement synchrone nous a permis de recueillir des données de phosphorylation avec une résolution de temps suffisamment large pour les étudier pendant le cycle cellulaire. Ces données ont pu être comparées à notre analyse précédente des données publiques disponibles sur l'homme et la levure. Ils ont détecté une fraction des phosphorylations qui ont été modifiées pendant le parcours temporel de l'expérience ("phosphorylations dynamiques") et qui ont pu être regroupées selon quatre comportements différents. Nous trouvons des preuves que les CDK sont responsables de la majorité des phosphorylations régulées par le cycle cellulaire. Nous avons croisé nos substrats humains de la famille CDK1 et constaté que près d'un quart d'entre eux sont représentés parmi les protéines dynamiquement phosphorylées dans les embryons de *Xenopus*, tandis que plus de la moitié des

phosphosites sont conformes au motif consensus minimal des CDK. Nous nous sommes ensuite concentrés sur les phosphorylations dynamiques qui présentaient un comportement oscillatoire évident. En utilisant la spectrométrie de masse quantitative à haute résolution temporelle, nous avons démontré que les oscillations des niveaux de phosphorylation étaient, en effet, de type interrupteur. Ceci est cohérent avec la modélisation mathématique du réseau de régulation des CDK en mitose, qui prédit une activation des CDK1 sous forme de "*switch-like*". Cependant, il est surprenant de constater que cette phosphorylation abrupte de nombreux substrats CDK s'est produite malgré la diminution progressive de la phosphorylation de CDK1 sur la tyrosine-15, qui est considérée comme un facteur clé de l'ultra sensibilité de l'activation de CDK1. Nos données suggèrent que la boucle de rétroaction positive CDC25 est active et augmente avec le temps, tandis que la boucle doublement négative WEE1 reste constante, expliquant la légère oscillation mais la dérégulation générale progressive de CDK1-Y15 avec le temps.

Ensuite, en réalisant une phosphoprotéomique quantitative sur des extraits d'œufs de *Xenopus* en réplication, nous avons montré la dynamique de la phosphorylation de facteurs de réplication de l'ADN au cours de la phase S. Des protéines connues pour ses effets sur la réplication comment MCM4 et RIF1, qui sont impliqués dans l'initiation de la réplication de l'ADN médiée par les DDK et dans la stabilité du réplisome, sont parmi ceux qui présentent le plus grand nombre de phosphosites régulés dynamiquement, ce qui suggère un rôle pour l'hyper- phosphorylation d'un sous-ensemble de protéines de réplication. L'analyse des motifs de phosphorylation a révélé que les phosphosites potentiels de la DDK sont régulés à la hausse pendant la phase S, à la fois *in vivo* et *in vitro*. Nous avons observé que les phosphosites des motifs de phosphorylation des kinases aurorales étaient enrichis dans les groupes d'interphase. Les kinases Aurora sont mieux connues comme régulateurs de l'attachement microtubule-kinétochore à la prométaphase et de la cytokinèse, ce qui suggère donc que la phosphorylation de beaucoup de leurs substrats se produit en amont de la phosphorylation CDK mitotique, ou que les kinases Aurora jouent un rôle méconnu pendant la réplication de l'ADN. Les phosphoprotéines dynamiques de *Xenopus* ont reproduit tous les résultats obtenus pour les cibles CDK de levure et les kinases du cycle cellulaire humain : une tendance générale des phosphosites à se localiser dans les IDRs avec un risque accru de phosphorylation dans ces régions pour les phosphosites dynamiques spécifiquement; de plus, les protéines avec au moins un phosphosite dynamique étaient significativement plus désordonnées que le reste des phosphoprotéines détectées. Nous avons émis l'hypothèse que des mécanismes de régulation du cycle cellulaire ont été sélectionnés pour contrôler les protéines intrinsèquement désordonnées, vraisemblablement pour moduler la formation et la dissolution des MLOs agissant comme des centres biochimiques dans des phases spécifiques du cycle

cellulaire. Étant donné que les sites CDK ont tendance à se regrouper au sein des IDR et que nos données montrent que de nombreux substrats CDK régulés par le cycle cellulaire sont des composants clés des MLO, il est probable que la phosphorylation médiée par les CDK affecte les propriétés physicochimiques des MLO. Ainsi, la phosphorylation de type interrupteur des IDRs peut réguler la réorganisation rapide des MLOs en mitose, lorsque la plupart de ces MLOs sont dissous.

Je présente ici une approche collaborative multi-échelle *in vitro*, *in cellulo* et *in silico* pour étudier le marqueur de prolifération Ki-67, montrant une régulation de sa condensation clairement dépendante de la CDK1. Cependant, aucune conclusion simple ne peut être tirée concernant l'effet net de la phosphorylation de Ki-67 sur le processus de séparation de phase, puisqu'il semble exister des modes concurrents de régulation de la séparation de phase par la phosphorylation de son domaine répété. Je suppose que différents niveaux de phosphorylation et le contexte cellulaire localisent Ki-67 à l'hétérochromatine périnucléolaire pendant l'interphase et à la couche péri chromosomique en mitose.

Enfin, j'ai présenté les résultats de deux projets différents auxquels j'ai contribué, visant à comprendre la fonction de deux systèmes biologiques intimement liés à la régulation des MLOs par les CDK.

La première est la régulation de l'organisation de la chromatine par le Ki-67 et les changements consécutifs dans l'expression des gènes. Nous avons constaté que le knock-out de Ki-67 entraîne des changements massifs dans le transcriptome, bien que sans aucun phénotype cellulaire évident. Nous avons remarqué des altérations importantes dans les marques d'histones de la chromatine, en particulier la triméthylation inhibitrice H3K27 associée à la répression de l'expression des gènes par les polycombes. Néanmoins, ces changements ne semblent pas expliquer entièrement la reprogrammation globale de la transcription observée, et des modifications dans d'autres mécanismes de régulation de la chromatine pourraient également être responsables de cet effet.

Nous avons montré que toutes les étapes de la tumorigenèse étaient affectées par l'ablation du Ki-67. Les données d'analyse RNAseq ont révélé des altérations de l'expression des gènes impliqués dans la transition épithélio-mésenchymateuse (TEM), la présentation des antigènes, le métabolisme des médicaments et d'autres caractéristiques associées au cancer. Ces résultats ont ensuite été confirmés par l'incapacité à générer des tumeurs intestinales chez des souris mutantes Ki-67, la capacité réduite de certaines cellules cancéreuses knock-out Ki-67 à induire l'angiogenèse, la présence de caractéristiques épithéliales et mésenchymateuses altérées dans les cellules cancéreuses Ki-67-négatives, la capacité réduite de ces dernières à coloniser d'autres

tissus et à donner naissance à des métastases, leur sensibilité accrue aux médicaments et leurs interactions réduites avec le système immunitaire.

Nous avons détecté des changements transcriptionnels dans les régulateurs clés de la voie EMT qui est régulée transcriptionnellement par le complexe PRC2. Cependant, l'inactivation de Suz12 n'a pas restauré la tumorigénicité des cellules 4T1 knockout Ki-67. Il est important de noter que des voies similaires associées au cancer sont dérégulées dans les cellules 4T1 et MDA-MB-231, ce qui suggère une conservation générale des mécanismes de régulation.

Ki-67 pourrait agir comme un effecteur de la régulation médiée par CDK1 en se localisant dans l'hétérochromatine périnucléolaire et en contrôlant son organisation et, par conséquent, les programmes transcriptionnels généraux dépendant de cette organisation. Néanmoins, l'absence complète de Ki-67 ne génère aucun changement détectable dans la prolifération d'aucun des modèles cellulaires testés, ce qui pourrait indiquer que ces programmes ne sont importants que dans des conditions spécifiques ou que tout ce système est fortement étayé.

Le deuxième système biologique est la régulation transcriptionnelle dépendante du Mediator par la sous-famille CDK8/19. Cette sous-famille de CDK hautement conservée fait partie, avec la cycline C, du module kinase du Mediator. Une particularité de ces protéines est qu'elles présentent un domaine C-terminal désordonné qui est prédit avoir une composition de type prion. Les CDK8/19 sont capables de phosphoryler le CTD de Pol II, qui est également intrinsèquement désordonné. En tenant compte du fait que l'interaction entre Mediator et les amplificateurs est médiée par la séparation de phase, nous avons supposé que le module kinase pourrait agir comme un contrôle interne de ce mécanisme pour réguler l'expression génétique. Cependant, nos résultats ne soutiennent pas un rôle essentiel des kinases de Mediator dans la prolifération cellulaire ni la régulation globale de la transcription, contrairement à Mediator lui-même. Nous avons constaté que seulement un nombre relativement faible de gènes étaient fortement dérégulés pour les doubles knockouts des organoïdes intestinaux de souris et des MEFs. Nous avons remarqué une tendance selon laquelle un peu plus de gènes étaient régulés à la hausse qu'à la baisse lors de la perte de CDK8 seule ou des deux kinases, tandis que les effets de la délétion combinée étaient plus que cumulatifs des effets des délétions simples, ce qui indique une redondance fonctionnelle. Dans les organoïdes intestinaux, les doubles knockouts n'ont montré qu'un effet discret sur la prolifération et la croissance, bien qu'après quelques semaines de croissance, le génotype se soit inversé, peut-être en raison de la sélection négative des organoïdes *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$. De plus, un phénotype semblable à celui de la fibrose kystique a été observé pour les organoïdes dépourvus des deux kinases, avec une signature transcriptionnelle similaire à celle montrée lors de l'ablation du gène *CFTR* dans les cellules humaines. Dans les

cellules BMEL, seules des délétions uniques ont été réalisées et la délétion de CDK8 a protégé ces cellules contre la transformation maligne et l'initiation de la tumorigenèse, indiquant qu'elle est requise pour ces processus, peut-être en contrecarrant les effets de p53. En effet, les gènes dérégulés dans les doubles knockouts CDK8/19 chez les MEFs étaient clairement enrichis en cibles de p53 et d'autres voies de réponse au stress. De plus, ces cellules présentaient également un taux de croissance plus lent, avec une accumulation de p21 et davantage de cellules en interphase.

En résumé, j'ai présenté ici les résultats de plusieurs projets collaboratifs visant à contribuer au développement d'un modèle pour le cycle cellulaire qui peut intégrer de nouvelles preuves qui ne peuvent être expliquées de manière satisfaisante par le modèle classique. Je propose ici un mécanisme qui pourrait expliquer comment les niveaux de phosphorylation totale des CDK peuvent réguler le cycle cellulaire de manière précise: Des changements brusques dans l'activité autorégulée des CDK phosphorylent des protéines dans les MLO, les centres biochimiques de la cellule, régulant leur formation et leur dissolution de manière dépendante du cycle cellulaire.

Acknowledgments

I would like to thank, first, Lila and Dan for giving me the opportunity of accomplishing a long-overdue goal of mine. They trusted a person who did not align with the canonical profile of a scientist and they guide me through the entire process. I am grateful that they were able to recognize my passion for science, and I truly hope that they are proud of the work they have done for transforming me into a proper scientist.

I want to also acknowledge Krzysztof Rogowski whose influence was critical for me to realize that science is a beautiful discipline, despite how foreign I felt (and still feel) in an academic context. Kris was the first person who said to me “you have what it takes to be a scientist”. These are words that I will always take with me and I hope everyone in the world can have a Kris that reminds them that they are good enough.

I would like to thank Juan Manuel Valverde and Maarten Altelaar for their huge contribution to our collaborative project. In particular, I want to thank Juan for reminding me how working as a team is much more fun than working on a project by myself.
Thanks to all the other collaborators mentioned in this thesis, as well.

I would like to thank Dan’s laboratory members (former and current) that I have shared time with. Special thanks to Susana, Ana-Bella, Lucie, Nuria, Denisa, Mehdi and Karim for all the help you provided to a lost bioinformatician doing bench work. Thanks also for those great "lunch talks". Thank you, Remy, for being a great example of resilience.

Thanks to my family and friends. It is hard to imagine how we can go through the situations that one has to experience during a PhD without the unconditional support of those we love.
A special mention goes to my friends in La Plata (Argentine) who, a long time ago, put together my first computer with spare parts as a birthday present. I would like to thank, especially, Juan “Pepo” Guzman and Ger “Umpacu” Riera for introducing me to Linux and programming, skills that have become critical in my professional life.

Ana, a million words would not do justice to what you meant to me during this process. Considering this thesis as my biggest achievement to the date, I want to dedicate this to you, and only you.

Contents

Abbreviations

AF mutant: T14A and Y15F CDK mutations
APC/C: anaphase promoting complex / Cyclosome
AS or as: Analogue-sensitive
AURK: Aurora kinase
CAK: CDK activating kinase
CB: Cajal body
CDK: Cyclin Dependent kinase
CDS: coding sequence
CF: cystic fibrosis
CG: Coarse grained
CKM: CDK kinase module (Mediator)
CLK: CDC-like kinase
Cryo-EM: Cryogenic electron microscopy
Csat: Saturation concentration
CTD: C-terminal domain
DDK: Dbf4-dependent kinase
DDR: DNA damage response
DEG: Differentially expressed gene
DFC: Dense fibrillar component (Nucleolus)
DSB: DNA double-strand break
DYRK: Dual-specificity tyrosine-regulated kinase
EdU: 5-Ethynyl-2′-deoxyuridine
EMT: Epithelial-to-mesenchymal transition
FACS: Fluorescence-activated cell sorting
FC: Fibrillar center (Nucleolus)
FDR: False discovery rate
FRAP: Fluorescence recovery after photobleaching
GC: Granular component (Nucleolus)
GO: Gene ontology
gRNA: Guide RNA (CRISPR-Cas9)
GST: Glutathione S-transferase
HAL: left homology arm
HAR: right homology arm
HCC: hepatocellular carcinoma
HDR: Homology- directed repair
HRR: Homologous recombinational repair
IDR: Intrinsically disordered region
IF: Immunofluorescence
INM: Inner nuclear membrane
LC: Low complexity
LFC: Log fold change: logarithm of the fold change
LLPS: Liquid-liquid phase separation
MAPK: Mitogen-activated protein kinase
MD: Molecular dynamics
MEF: Mouse embryonic fibroblast
MLE: Maximum likelihood estimation
MLO: Membraneless organelle
MPF: M-phase Promoting Facto
MS: Mass spectrometry
NE: Nuclear envelope
NEB: Nuclear envelop breakdown
NEK: NIMA-related kinase
NHEJ: Non-homologous end joining
NMR: Nuclear magnetic resonance
NPC: Nuclear pore complex
OA: Okadaic acid
Nups: Nucleoporins

PA: Phosphatase A
PCA: Principal components analysis
PDB: Protein databank
PLK: Polo-like kinase
PP1: Protein phosphatase 1
PP2A: Protein phosphatase 2A
pre-IC: pre-initiation complex
pre-RC: pre-replicative complex
PML: promyelocytic leukemia body
PS: Phase separation
PTM: Post-translational modification
puroR: puromycin resistance cassette
RB: Retinoblastoma protein
RNP: Ribonucleoprotein
ROS: Reactive oxygen species
SAC: Spindle-associated checkpoint
SCDM: Sequence Charge Decoration matrices
SCF: Skp, Cullin, F-box containing complex
SILAC: Stable isotope labeling by amino acids in cell culture
SIM: SUMO Interacting Motif
SLiMs: short linear motifs
SUMO: Small ubiquitin-like modifier
TES: Transcription ending site
TSS: Transcription starting site
UV: Ultra violet
WT: Wild-type

Names
JMV: Juan Manuel Valverde
GD: Geronimo Dubra
LK: Liliana Krasinska
DF: Daniel Fisher
MA: Maarten Altelaar

1. Introduction

1.1 The cell cycle

From its origin, over 3.5 billion years ago, life is characterized by the transferring of chemical information from one generation to the next one. The first self-replicating molecules in the primordial soup were already subject to selective pressures allowing for Darwinian selection and resulting in more complex systems. These “replicators”, as Richard Dawkins named them in his book *The selfish gene*, found their way into membrane-enclosed organizations having their own metabolic networks that evolved to give rise to modern cells.

This increase in the organization had, then, to be accompanied by more sophisticated mechanisms for the replication of biological information. Cells, as we observe them nowadays, have to not only carry over their genetic information to the new generation but also divide all their components. This process, known as the Cell Cycle (or Cell Division Cycle) is extremely conserved and requires seemingly complicated molecular mechanisms to control and coordinate the different steps involved. The main scope of this thesis is to try to unveil how such complex and detailed control systems can be, at the same time, extremely robust and reliable.

1.1.1 Cell cycle phases

In 1951, Alma Howard and Stephen Pelc performed a remarkable experiment in which they managed to mark nascent DNA with radioactive P^{32}, demonstrating that replication only happens in a limited time window. This, together with further investigations led to the currently accepted model of the Cell Cycle, which essentially consists of two temporally delimited main events. First, a cell should replicate its hereditary material and, only then, can divide all its components into two daughter cells. Each one of the resulting progeny will go through this exact same process. DNA replication and cell division are two of the cell cycle phases, called S-phase (from synthesis) and M-phase (from Mitosis), respectively. In between those stages, there can exist two gap phases called G1 (originally, pre-S phase) and G2 (pre-M phase) which have been thought to be critical for the correct chronological separation of S and M phases. Jointly, G1, S, and G2 phases are referred to as *Interphase,* while M phase is subdivided into various stages depending on the state of separation of the chromosomes.

The generic view of the cell cycle considers that phase should finish in order for the next one to start. The cell cycle is inexorably progressive, meaning that it cannot be reversed and its phases cannot be skipped. However, the cell cycle does not escape the generally chaotic biological principles and these hardly result in such rigid and structured systems. Evidence showing numerous exceptions to most statements listed in the current paragraph will be detailed further on.

Nevertheless, all eukaryotic cells require a system to precisely regulate the phases of the cell cycle. To decipher the molecular mechanisms underlying this control system, researchers opted for the most instinctive approach to solving complex puzzles: they started with the simplest version of it and observed how it evolved in intricacy from there.

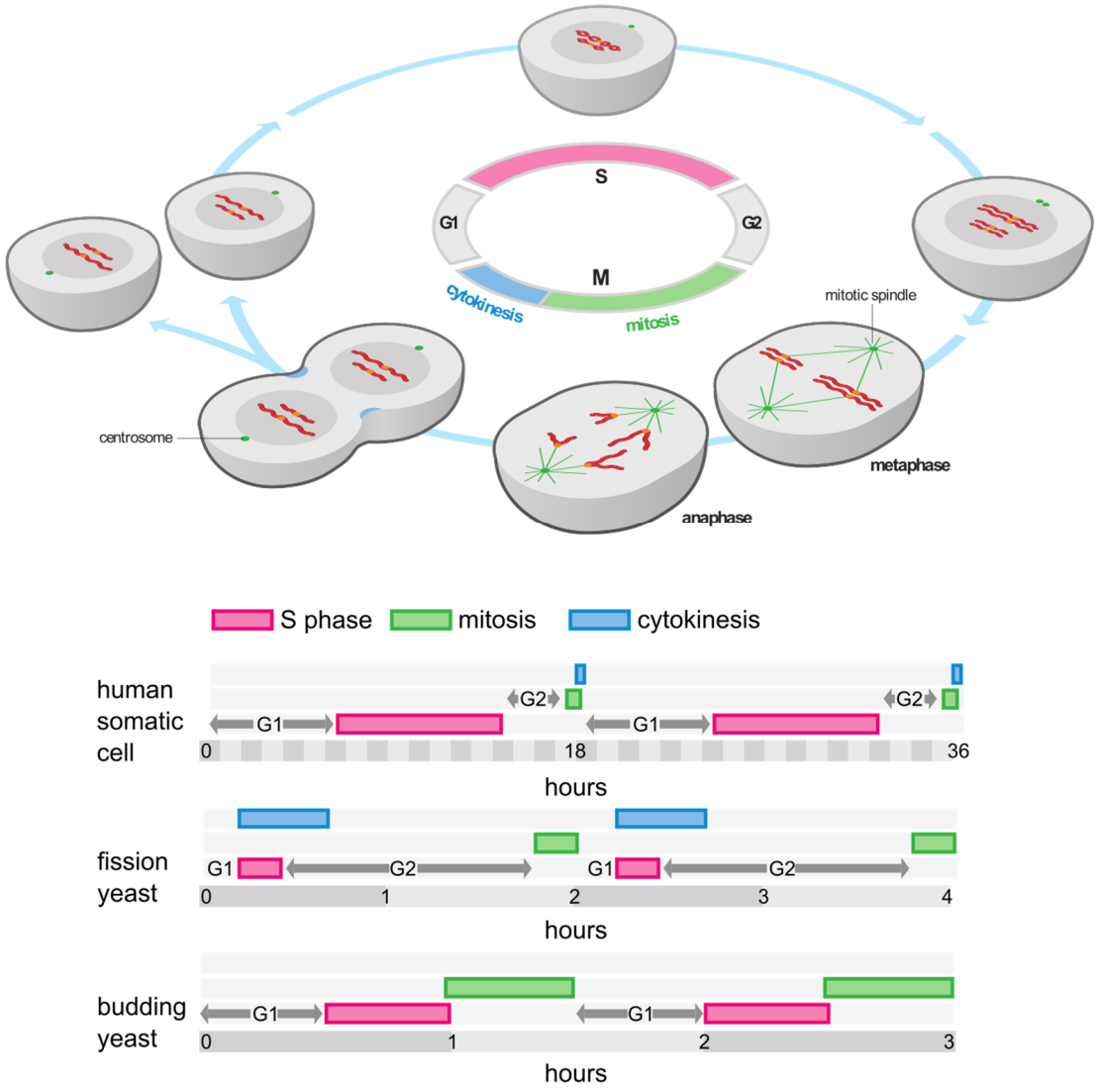


Figure 1.1 | **The eukaryotic cell cycle.** Adapted from *Morgan, 2007.*

1.1.2 Molecular control of the cell cycle

In the 1970s, screens of mutant strains of budding and fission yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively) revealed the existence of a family of genes required for the control of the cell cycle progression called *cdc*, for cell division cycle (Hartwell et al., 1970; Hartwell et al., 1974; Nurse, 1975; Nurse et al., 1976). The genes cdc28 and cdc2, from *S. cerevisiae* and *S. pombe* respectively, were highlighted as the most interesting candidates for the control of the cell cycle, given that they are required for both the start and the mitosis control points (Hartwell et al., 1974; Nurse and Bissett, 1981; Piggott et al., 1982).

During the mid-1980s, both genes were sequenced showing high identity between them, as well as some other relevant characteristics such as a conserved ATP binding site and a putative phosphorylation site (Lörincz and Reed, 1984; Hindley and Phear, 1984). It was later confirmed that cdc2/28 were protein kinases regulated by phosphorylation (Simanis and Nurse, 1986). This raised the interest of scientists in protein phosphorylation which had long been described to occur specifically in mitosis and had been proposed to drive mitosis itself (Bradbury et al., 1974a; Bradbury et al., 1974b; Guerrier et al.,1975).

Cdc2/28 homologs were shortly identified as the drivers of mitotic entry in all eukaryotes (Lee and Nurse, 1987; Gautier et al, 1988; Labbé et al., 1988). CDK phosphorylation was also defined as the responsible activity of the elusive Maturation Promoting Factor (also, M-phase Promoting Factor or MPF) which triggers Oocytes maturation by inducing cell division (Masui and Markert, 1971; Wasserman and Smith, 1978; Gerhart et al., 1984; Dunphy et al., 1988; Labbe et al., 1988; Lohka and Maller 1988; Labbe et al., 1989).

Interestingly, at that time when autoregulation of cdc2 had not been described, it was shown that neither the cdc2 expression levels nor its phosphorylation state could account for the mitotic cell cycle regulation *in vivo* since both remain relatively constant throughout its entire duration (Simanis and Nurse, 1986). In 1983, a group of proteins whose levels oscillated during the first divisions of sea urchin embryos was described (Evans et al., 1983). These proteins, called Cyclins, were shown to be synthesized upon embryo fertilization and then degraded after cell division. It was later demonstrated that Cyclins are the binding partners of the cdc2/28 family and that this association is mandatory for its kinase activity (Labbé et al., 1989). Scientific tradition led to the conservation of the original cdc2 and cdc28 names for yeast but, for most organisms, this protein is now referred to as Cyclin-Dependent Kinase or CDK and it is the founding member of the CDK family (Meyerson et al., 1992)

The contributions of Leland Hartwell, Paul Nurse, and Timothy Hunt to the discovery and description of the molecular control of the cell cycle were awarded the 2001 Nobel Prize in

Physiology and medicine. Yet how the activity of different members of the CDK family is organized in an orderly manner to generate a functional cell cycle is still subject to some controversy in the field. Questions persist to which answers remain elusive to scientists even in current times, some of which I will try to address in this thesis.

1.1.3 Evolution and diversity of CDKs and Cyclins

CDKs and Cyclins are extremely conserved across all eukaryotic life, to such an extent that *cdc2* yeast mutants could be complemented with the human homolog CDK1 (Lee and Nurse, 1987). In higher multicellular organisms, the expansion of CDK and Cyclin genes, mostly during Metazoa/Eumetazoa emergence (Cao et al., 2014), contributed to an increase in the complexity of the cell cycle regulatory network and functional diversification of some of these kinases. Humans, for instance, have 20 different CDKs, grouped in 8 subfamilies, and 29 different Cyclins (Malumbres, 2014), some of which do not appear to have direct roles in the cell cycle. Three CDK subfamilies -CDK1, CDK5, and CDK4/6 subfamilies- comprising 11 different proteins have been associated with the regulation of the cell cycle. In contrast, in fission yeast only 6 CDKs have been described with 2 of them, Cdc28 and Pho85 -homologs of CDK1 and CDK5 subfamilies, respectively- described as directly regulating cell cycle processes.

As expected, Cyclins also add to the evolutionary increasing complexity of the cell cycle regulation, and maybe in a more substantial fashion than their catalytic counterparts. There are 15 human subfamilies containing 29 individual cyclins (Gunbin et al., 2011). Only 4 of those subfamilies -A, B, D, and E-type cyclins-, containing 10 different members, have been proposed as direct regulators of the cell cycle (Morgan, 2007).

CDK2 and CDK1, which have been proposed as the S-phase and G2/M-phase main CDKs, respectively, are two of the members of the CDK1 subfamily. The third member, CDK3, presents a homozygous missense mutation in the *Cdk3* gene (T187X), generating two null alleles, in most of the commonly utilized mouse strains employed in the laboratory (Ye et al., 2001). In contrast, the vast majority of the wild mice analyzed don't present this mutation. Altogether, this implies that CDK3 is not essential and that, probably, there exists a functional redundancy with other CDKs. The multiple CDK1 subfamily members appear in metazoans by gene duplication-mediated divergence from a common ancestor of amoebas and fungi (Cao et al., 2014). As mentioned before, the CDK1 yeast homologs, and the sole members of their subfamilies, are Cdc2 and Cdc28 for *S. cerevisiae* and *S. pombe*, respectively.

In metazoans, three different subfamilies of cyclins bind CDK1 and CDK2 in different phases of the cell cycle: cyclin E during S-phase; cyclin A at the end of S-phase and until early mitosis; and cyclin B during mitosis. In contrast, yeast cyclins are categorized in two subgroups. Cln1-3 in budding yeast and Puc1 in fission yeast belong to the CLN subfamily. Not much is known about the evolutionary history of this subfamily, although common ancestors of these organisms seem to drive the cell cycle only with a single B-type cyclin, which has also been proven to be possible in the modern S. pombe strains (Coudreuse and Nurse, 2010). It was hypothesized that a single primordial B-type cyclin was the starting point from which the rest of cyclins evolved as a consequence of gene duplication followed by function specification (Nasmyth, 1995). S. pombe genome encodes for 3 different B-type cyclins (Cig1 and Cig2, G1 and S-phase; Cdc13 mitosis), while S. cerevisiae expresses 6 (Clb5-6, S-phase; Clb1-4, mitosis). In mammals, cyclin A and B are the mitotic cyclins, while cyclin E has been associated with S phase-related functions. Furthermore, adult mammals express one subtype of A cyclin, A2, and two subtypes of B cyclins, B1 and B2. The evolutionary mechanisms for the amplification in the number of genes in each cyclin subfamily appears to be quite divergent between higher eukaryotes and yeasts (Gunbin et al, 2011)

The CDK4/6 subfamily does not have a clear homolog in fungi and it is only present in eumetazoans (Cao et al., 2014). Indeed, these CDKs are considered mainly as specific regulators of the animal cell cycle. Historically, CDK4/6 have been described to associate with type-D cyclins to phosphorylate the retinoblastoma protein (RB) which, in turn, will activate specific expression programs for progressing through the G1 phase of the cell cycle and start the DNA replication (Narasimha et al., 2014). However, the ablation of both CDK4 and CDK6 or all D-type cyclins does not have any impact on cell proliferation nor on early stages of mouse development, indicating that CDK4/6-cyclin D complexes are not essential for the mammalian core cell cycle (Malumbres et al., 2004; Kozar et al., 2004). Nevertheless, none of these embryos can reach advanced embryogenesis stages, probably due to defects in hematopoiesis (Malumbres et al., 2004; Kozar et al., 2004). Even if the minimal cell cycle does not dramatically increase in complexity in mammals when compared with single cell eukaryotes, positive selection of additional CDK and cyclins might be due to their critical roles in the regulation of multicellularity and the development of complex organisms. Indeed, mice having CDK1 as the sole cell cycle CDK, are able to reach mid-gestation (Santamaria et al., 2007), in a situation reminiscent of that observed in yeasts where a single CDK drives the entire cell cycle.

The CDK5 subfamily is composed by 6 CDKs and they are considered to be orthologs of S. cerevisiae Pho85. CDK5 is mainly associated with cell cycle-related functions and molecular

signaling of specific cell types such as neurons, cells from the immune system and hepatocytes (Shupp et al., 2017). It can bind cyclin Y to reach their active version, but they can also be associated with cyclin E and D, which reduces its activity (Shupp et al., 2017).

CDK7, in complex with cyclin H, is a dual-function CDK: it indirectly modulates the cell cycle by catalyzing the activating phosphorylation of CDK1 and CDK2 (CDK activating kinase, or CAK), but it also regulates gene expression through the phosphorylation of the C-terminal domain of Pol II (Harper and Elledge, 1998). In budding yeast, the ortholog Kin28 regulates transcription but is incapable of CAK activity, function that is mainly achieved by the non-CDK-related Cak1 kinase (Espinoza et al., 1996). Fission yeast encode for two proteins, Mcs6 and Mcs2, which are the orthologs for CDK7 and cyclin H, respectively. It also has a second CAK, Csk1, more closely related to S. cerevisiae Cak1. Deletion of CDK7 in mice is lethal in a prenatal stage, although it allows first stages of embryogenesis, indicating that it is not essential for the minimal cell cycle, but it is required for normal organism development (Ganuza et al., 2012).

There is another subfamily of extremely conserved CDKs that can regulate gene expression by phosphorylating the C-terminal domain of Pol II: CDK8/19 (Rickert et al, 1999). These kinases have the particularity of forming the kinase module of the Mediatorcomplex together with the MED12 and MED13 subunits. This function is extremely conserved and the budding yeast homologs of the CDK8-cyclin C complex (Srb10-Srb11) are associated with the homologs of Med12 (Srb8) and Med13 (Srb9) to form the kinase module of S cerevisiae Mediator.

The rest of the CDK and Cyclin subfamilies are less relevant to the discussion proposed in this thesis and, therefore, not within the scope of the current manuscript. A fine summary of the entire CDK family has been published in 2014 by Marcos Malumbres (Malumbres, 2014), to which I refer the reader for further information.

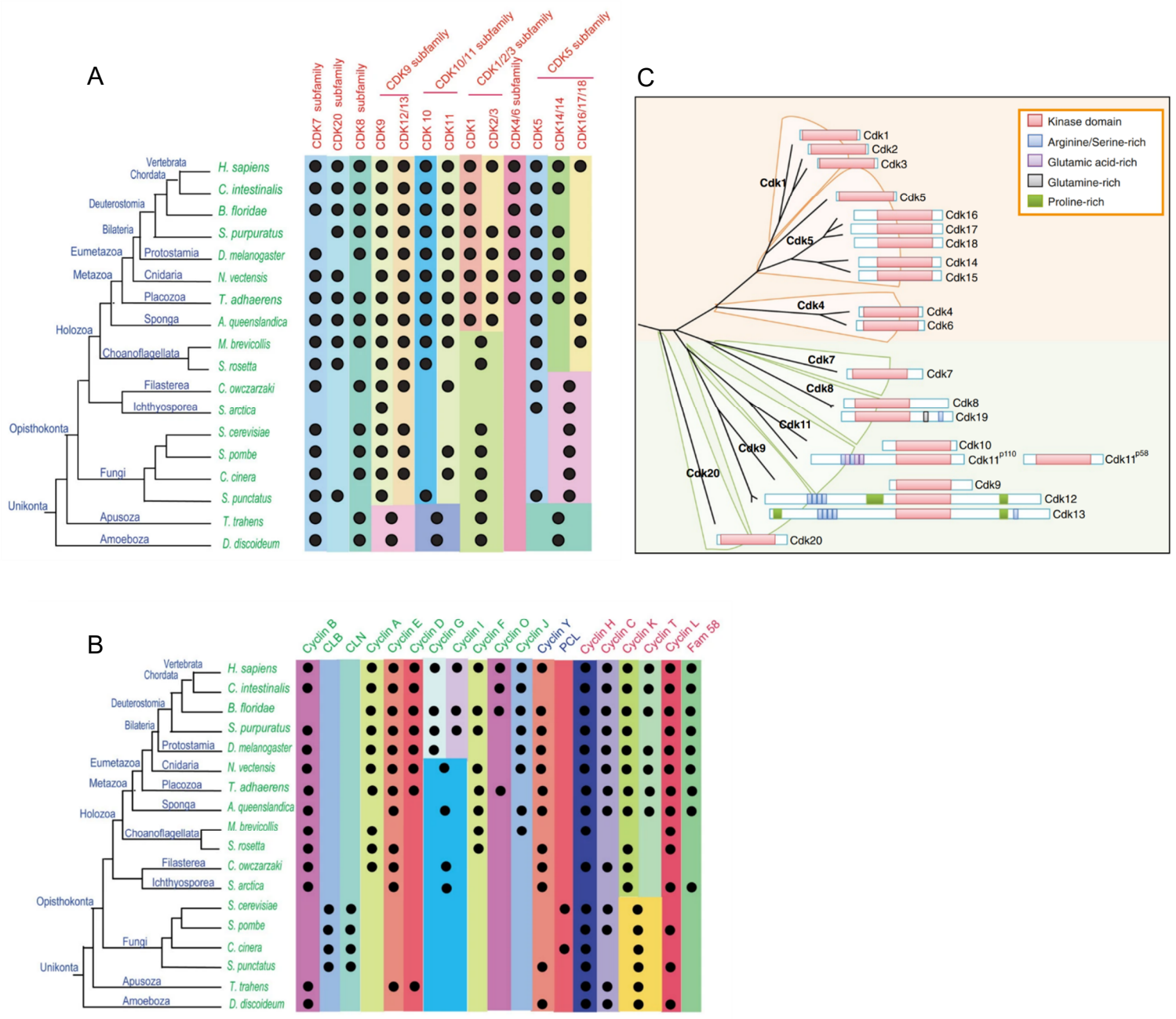


Figure 1.2 | Evolution of CDKs and cyclins. From *Cao et al., 2014 and Malumbres 2014.* Phylogenetic trees schematizing the tree of life where the presence of the different CDKs (A) and cyclins (B) subfamilies are marked with black dots. (C) phylogenetic tree schematizing the evolutionary distances between the different subfamilies of mammalian CDKs.

1.1.4 The CDK-Cyclin complex

The Cyclin-CDK binding as a requirement for the CDK kinase activity has raised the interest in the study of their molecular structure.

Cyclins are composed of two major domains, and both consist of 5 short helices. The N-terminal domain, named the cyclin box, has been shown to play an important role in the association with CDK (Kobayashi et al., 1992). The first crystal structure of a Cyclin-CDK complex obtained was that of human CyclinA2-CDK2 (Jeffrey et al., 1995). This structure sheds light on the fundamental requirements for the binding of these two proteins. At the Cyclin-CDK interface, CDK2 has a helix (PSTAIRE) and a loop (T-loop) that can be subject to conformational changes when it comes in contact with the more rigid cyclin box of Cyclin A2. It was postulated that this might be the explanation for the fact that a given CDK can normally bind more than one cyclin (Petri et al. 2007). This depiction is accurate to such a degree that CDK2 and CDK6 can accept viral cyclin-like proteins as functional binding partners, constituting one elegant evolutionary mechanism by which herpesviruses can take over the control of a host cell cycle (Swanton et al. 1997; Card et al., 2000).

For the most part, CDKs lack the kinase activity in absence of cyclin but they can, nonetheless, bind ATP which results in residual basal activity. The structural changes induced by the formation of the Cyclin-CDK heterodimers and phosphorylation of a conserved threonine residue at the T-loop (T160 in human CDK2, T161 in human CDK1, T167 in yeast Cdc2/28) rearrange the substrate-binding site stabilizing and activating the complex. The T-loop phosphorylation is catalyzed by the CDKs activating kinases or CAKs. The main CAKs are CDK7-CyclinH-Mat1 complex in humans, the ortholog Mcs6-Mcs2 complex in *S. pombe,* and the single CDK-unrelated Cak1 protein in *S. cerevisiae*. Other kinases were also described to be able to phosphorylate this residue, but CDK7 is essential for the mammalian cell cycle.

There exists an important distinction between human CyclinA2-CDK2 and CyclinB1-CDK1 complexes regarding the order in which T-loop phosphorylation and Cyclin binding happen. CDK2, unlike CDK1, allows this phosphorylation in the absence of Cyclin (Fisher and Morgan, 1994). That is concordant with recent studies showing a differential behavior of CDK1 and CDK2. *In vivo*, the absence of T160 phosphorylation turnover and the preference for the “CAK-first pathway” was observed only for CyclinA-CDK2 (Merrick et al., 2008). Besides, once assembled this particular complex is refractory to dephosphorylation of the T-loop by λ-phosphatase (Brown et al., 2015). The phosphate in the T160 of CyclinA-CDK2- might be buried in a milieu of positive charges while the homolog T161 in the CyclinB-CDK1 remains solvent accessible. Some researchers, notably R.P. Fisher and M.E.M. Noble, have accredited this divergence in the CDK7-

mediated T-loop phosphorylation, the mechanism by which CDK1 and CDK2 activities can be ordered during the cell cycle of higher eukaryotes.

Certainly, caution is advised when extrapolating conclusions on the general structural behavior of CDKs obtained, almost uniquely, from the examination of a single CDK-Cyclin complex even if others have been crystalized thereafter. This is especially true when considering different subfamilies, and it has been reviewed in detail by Wood and Endicott (Wood and Endicott, 2018).

Another conserved residue, the tyrosine residue in position 15 (Y15), was early described to be phosphorylated, and its phosphorylation state correlated with CDK activity (Gould and Nurse, 1989). Indeed, wee1 (Wee1 in humans and swe1 in budding yeast), the first “dose-dependent inhibitor of mitosis”, was discovered to be the kinase mainly responsible for the phosphorylation of Y15 (Russel and Nurse 1987). Another protein from the Wee family, Mik1 (Myt1 for higher eukaryotes) has been pointed to contribute to Y15 phosphorylation and to be the main driver kinase for the phosphorylation of the adjacent threonine 14 (T14), whose role is less well described (Lundgren et al., 1991; Mueller et al., 1995). Both residues are located on top of the ATP-binding pocket, and early studies have hypothesized about possible steric and/or electrostatic clashes between the γ-phosphate moiety of the phopho-Y15 and the ATP molecule. A recent work comparing the structures of phosphorylated and non-phosphorylated Y15 showed that the phosphorylation state does not induce any considerable conformational change. It would rather hinder the affinity for the substrate, reducing the CDK kinase activity (Welburn et al., 2007).

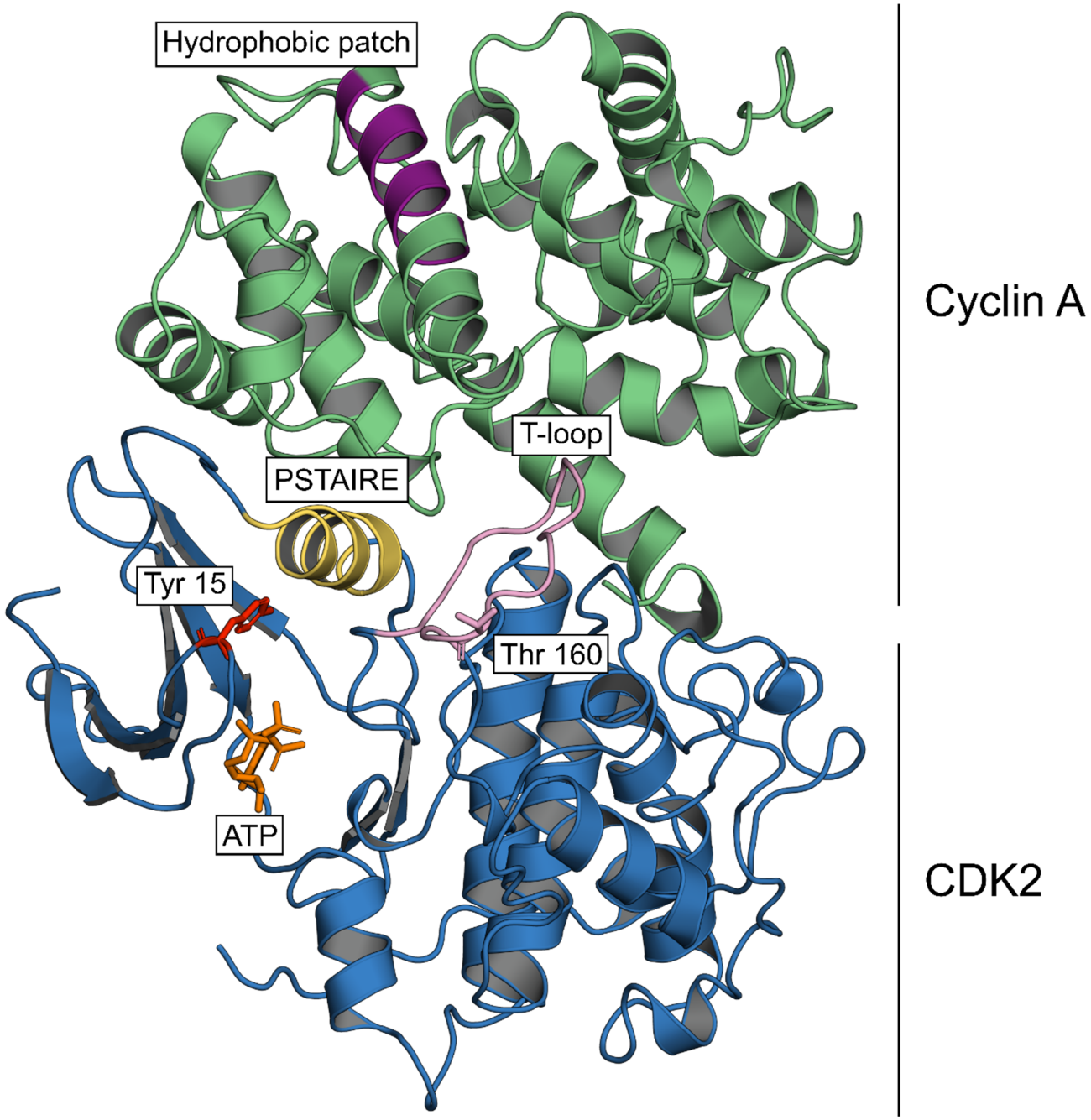


Figure 1.3 | Crystal structure of the CDK2–cyclin A complex (RCSB PDB ID: **1FIN**). Cartoon representation of CDK2 (blue) in complex with cyclin A (green). ATP, orange; Tyrosine 15, red; PSTAIRE helix, yellow; T-loop and the Threonine 160, pink; hydrophobic patch, purple.

1.2 CDK activity

Hundreds of proteins have been identified as substrates for cyclin-dependent kinases. How the cell brings about order in the phosphorylation of its targets is still in debate among experts in the field.

The dynamics of CDK-mediated phosphorylation have been, nonetheless, extensively studied and some sophisticated control mechanisms have been elucidated. Expression, degradation, and positive and negative feedback loops, are all working together to generate an extremely non-linear dynamic of CDK activity. This means that a positive increment in the level of CDK complexes does not generate a proportional number of phosphorylations. Terms borrowed from the mathematical modeling field, like ultrasensitivity and hysteresis, will be introduced in the next section to explain how such dynamics are being carried out within the cell.

1.2.1 Cyclin expression and degradation through the cell cycle

The amount of CDK protein remains relatively constant throughout the cell cycle. The levels of Cyclins, however, undergo successive fluctuations due to the coordinated regulation of their expression and degradation.

Different cyclins will peak during different phases of the cell cycle, and it is the cyclin-CDK kinase activity that will activate transcription factors for the expression of cyclins that govern the next phase. CDK activity will also initiate the ubiquitin-mediated proteolysis machinery that will degrade the currently operating cyclins.

Some overlap in the levels of some cyclins is observed during the cell cycle, and this is generally interpreted as an indication of redundancy in their functions. This redundancy is not conserved and, even for closely related species, the number of cyclin genes expressed in each cell cycle phase, as well as the total number of cyclins, differs considerably from one organism to the other. *S. cerevisiae*, for instance, has nine different cell cycle cyclins whereas *S. pombe* only expresses four, suggesting a simpler regulation of CDK kinase activity (Morgan, 1997).

Budding yeast can be selected to illustrate the complexity of the cyclin expression/degradation network. Of the nine cell cycle cyclins, three (Cln1-3) are considered to be active exclusively during the G1-phase (Cross 1988; Nash et al., 1988; Hadwiger et al., 1989; Richardson et al., 1989). Cln3 has been proposed as the initial cyclin and it is upstream-regulated by the size and metabolic state of the cell (Dirick et al., 1995; Shi and Tu 2013). It was originally proposed that Cln3-Cdc28 phosphorylates the transcription inhibitor Whi5, preventing its binding with the Swi4/Swi6 (SBF) transcription factor (de Bruin et al., 2004). SBF-mediated transcription would

then promote the expression of Cln1-2, further activating SBF and Mbp1/Swi6 (MBF) transcription factors. This positive feedback loop leads to the transcription of hundreds of genes among which are the S-phase cyclins required to initiate DNA replication (Skotheim et al. 2008; Ferrezuelo et al., 2010; Travesa et al., 2012). This model has, however, been recently discredited by evidence indicating that Whi5 represents a poor substrate for Cln3-Cdc28, and that this complex phosphorylates the C-terminal domain (CTD) of the Rbp1 subunit of RNA polymerase II. Cln3-Cdc28 binds to SBF-regulated promoters and stimulates the transcription of the genes under their control by direct *in situ* phosphorylation of RNA polymerase II.

During G1/S transition, Cln1-3 are marked for ubiquitin-mediated degradation. A protein complex formed by the subunits Skp1, Cdc53, Rbx1, and an interchangeable F-box containing adapter act as the ubiquitin ligase of phosphorylated Cln-Cdc28. The F-box protein drives the substrate recognition of the so-called Skp1-cullin-F-box complex (SCF) that will tag G1 cyclins for proteasomal degradation.

The first S-phase cyclins, Clb5-6 are expressed in an MBF-dependent manner. Later on, both MBF- and SBF-regulated transcription are repressed by increasing levels of Clb-CDC28 activity. A negative feedback loop occurs as cells transition from G1 to S, where the repressor Nrm1, encoded by an MBF-target gene, binds MBF to inhibit its gene expression activity (Amon et al, 1993; de Bruin et al, 2006). Rising levels of Clb-CDC28 also catalyze the phosphorylation and activation of Fkh2-Mcm1 and Ndd1 transcription factors, resulting in the expression of the rest of the cyclins, Clb1-4.

With the sole exception of Clb6, which is degraded by an SCF-dependent mechanism (Jackson et al., 2006), all B-type cyclins are ubiquitinated and targeted for proteasomal degradation by an E3 ubiquitin ligase called anaphase-promoting complex (also Cyclosome or simply APC/C). Similar to SCF, APC/C requires the association with an activator subunit, Cdc20 or Cdh1, that binds 3 major substrate recognition short linear motifs: the D box (Glotzer et al., 1991), the KEN box (Pfleger and Kirschner, 2000), and the ABBA motif (Burton et al., 2011; Lu et al., 2014; Di Fiore et al., 2015).

APC/C-Cdc20 is activated in early M-phase by Cdc28-mediated phosphorylation (Rudner and Murray, 2000) and its main function is to promote metaphase-to-anaphase transition by marking cyclin B and Securin for degradation. This will activate the protease separase that releases the cohesion between sister-chromatids, initiating anaphase. During M/G1 transition, Cdc20 is replaced by Cdh1, which redirects the substrate specificity toward mitotic cyclins. Cln- and Clb-dependent Cdh1 phosphorylation represses APC/C-Cdh1 until the mitotic exit and the early-G1 phase of a new cycle (Zachariae et al., 1998). The decreased affinity of APC/C-Cdh1 for S-phase

Clbs allows for the buildup of early Clb3-5, whose activity inhibits Cdh1 thus promoting the expression of mitotic Clb2 (Yeong et al., 2001).

In fission yeast, only 4 cyclins govern the cell cycle: Puc1, Cig1, Cig2, and Cdc13 (Forsburg and Nurse, 1991; Bueno et al., 1991; Bueno and Russell, 1993; Connolly and Beach, 1994; Obara-Ishihara and Okayama, 1994; Booher and Beach, 1988; Hagan et al., 1988). Nonetheless, an analogy can be found between the regulation of expression and degradation cycles of *S. pombe* and *S. cerevisiae*.

Puc1 is the only cyclin of this group that possesses homology to the Cln family and it has been shown to regulate the length of G1 phase and cell size in the absence of Cig1 and Cig2, most likely by negatively regulating the Cdc2 stoichiometric inhibitor rum1 (Martin-Castellanos et al., 2000).

Two B-type cyclins, Cig1 and Cig2, attain peak expression levels during G1/S and while Cig1 is maintained at constant levels until is degraded by APC/C at mitosis (Blanco et al., 2000), Cig2 levels diminish during G2 due to SCF-mediated degradation (Yamano et al., 2000; Yamano et al., 2004). Cig2 is transcriptionally regulated by MBF that in turn can be repressed by Cig2-mediated phosphorylation. This negative feedback loop influences the tight expression regulation of Cig2 expression throughout the cell cycle.

Cdc13 is the main mitotic cyclin. This B-type cyclin reaches its maximum level in G2/M and is degraded by the APC/C during anaphase (Moreno et al., 1989; Yamano et al., 1996). APC/C can be bound to the Cdc20 homolog (Slp1) or Cdh1 homolog (Ste9/Swr1), in an analogous way to what is observed in budding yeast. Ste9/Swr1 can promote Cig1 and Cdc13 degradation during mitotic exit and early-G1, after which CDK-mediated phosphorylation drives its inactivation and dissociation from APC/C (Blanco et al., 2000; Yamaguchi et al.,2000).

In higher eukaryotes, the principles of auto-regulation of cyclin expression and degradation are conserved. A review of the animal regulatory G1-S regulatory network shows remarkable topology conservation when compared with those observed in yeast, even if the sequence homology of its components is low (Cross et al., 2011). Basically, before the commitment point, which in animals has been named the “restriction point” (R), the protein Rb represses the E2F transcription factor similar to what occurs in budding yeast with Whi5 regulation of SBF. Type D cyclins (D1, D2, D3) increase their levels during early G1 acting as mitogen sensors to form complexes with CDK4 and CDK6 that will then mono-phosphorylate Rb (Planas-Silva and Weinberg, 1997). A more complex mechanisms for the commitment to the cell cycle seems to take place in higher eukaryotes, where the mono-phosphorylation of Rb partially activates the E2F-mediated expression allowing the accumulation of cyclin E (Narasimha et al., 2014). CDK2-cyclin E

complexes will hyperphosphorylate Rb, thus completing its inactivation and releasing E2F and driving the expression of genes needed to start S-phase (Johnson et al., 1994; Spencer et al., 2013). Cyclin E gets degraded at mid S-phase by ubiquitin-mediated proteolysis (Siu et al., 2012). Degradation of D-type cyclins is less clear, partially due to the fact that the multiple subtypes contain different degron tags for their recognition by the proteasome and might be removed at different phases of the cell cycle, but also as a result of disagreements among different authors regarding which are the ubiquitin ligases involved in this process (Qie and Diehl, 2020). The expression of the only isoform of A-type cyclins expressed in mammals, cyclin A2, peaks shortly after S-phase onset and it maintains its level until mitotic prometaphase, where it gets removed by APC/C-mediated proteasome degradation in a spindle checkpoint-independent manner (Elzen and Pines, 2001). Both cyclin A and E can thus be regarded as the S-phase cyclins in mammals. Indeed, microinjections of anti-cyclin A antibodies in human cells block DNA replication (Pagano et al., 1992). Nonetheless, cyclin A has important mitotic functions in adult cells, while in embryonic cells of Drosophila and Xenopus it is not required for DNA replication and cyclin E is the major S-phase cyclin (Morgan, 2007). Of the two subtypes of cyclin B expressed in mammalian cells, only cyclin B1 is essential, while cyclin B2 functions can be replaced by the former (Brandeis et al., 1998). Cyclin B starts to be transcribed during S-phase and it reaches its maximum level during mitosis at metaphase, and although it is degraded by an APC/C-mediated mechanism, similar to cyclin A, this process depends on the spindle-checkpoint. This assures a differential timing for the degradation of cyclins A and B, thus controlling specific functions of CDK1 mediated phosphorylation during the different stages of mitosis. A more detailed description of this critical mechanism will be addressed in the section 1.3.3

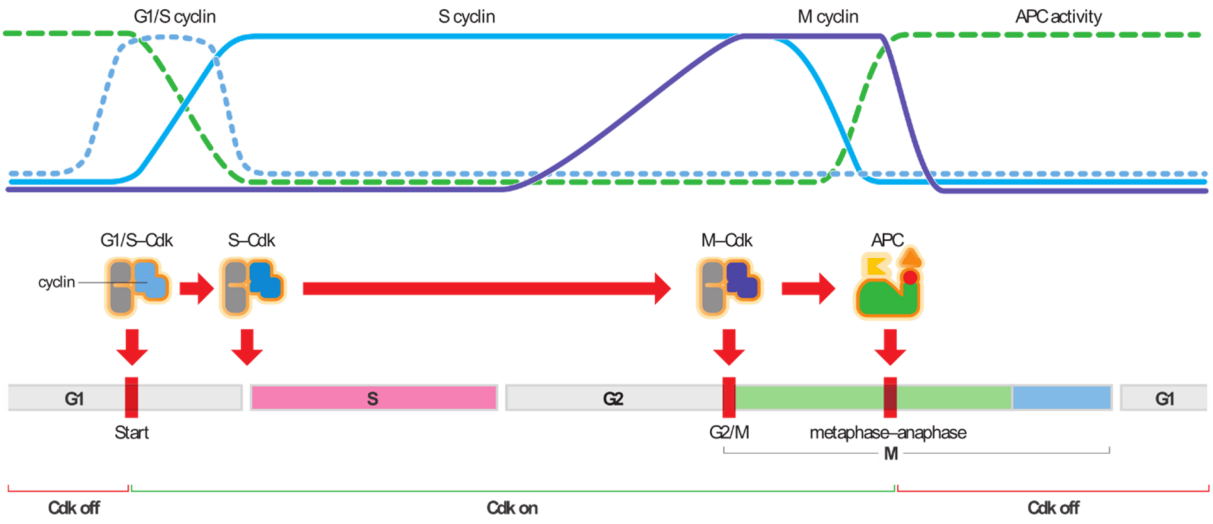


Figure 1.4 | Cyclin levels trough the cell cycle. From *Morgan, 2007.*

1.2.3 CDK auto-regulation: Y15 phosphorylation

While working in Mitchinson's laboratory, in Edinburgh, Paul Nurse was interested in fission yeast mutants presenting an increased size. His idea, largely inspired by Hartwell's previous work, was to find genes that once deleted could prevent cell division. The cell cycle of *S. pombe* is tied to its size, and they divide at an approximately constant volume. Therefore, the discovery of these higher-sized mutants would imply the finding of cell division controlling genes. He, instead, found smaller cells suspected to be some sort of contamination at first. These mutants were able to divide at roughly half the size of normal cells, the reason why they were called wee from the Scottish dialect word meaning "small". The gene responsible for the phenotype was named wee1 and was then shown to encode for a protein kinase acting antagonistically to the mitotic inducer cdc25, by controlling the activity of cdc2 (Russel and Nurse 1987).

It was later conclusively demonstrated that Wee1 is the kinase that catalyzes the phosphorylation of cdc2/CDK1 Y15 residue inhibiting its activity (Parker et al., 1992; Parker and Piwnica-Worms, 1992; McGowan and Russel, 1993). It was also shown that the product of the gene cdc25 is the phosphatase that dephosphorylates this same amino acid (Dunphy and Kumagai, 1991; Gautier et al., 1991; Kumagai and Dunphy, 1991; Strausfeld et al., 1991).

Moreover, Wee1 and Cdc25 are regulated by CDK phosphorylation. Wee1 can be directly phosphorylated by CDK eliciting its inactivation (Mueller et al., 1995; Okamoto and Sagata, 2007). Cdc25, on the other hand, is activated by CDK-mediated phosphorylation (Hoffman et al.,1993). This network of activations and inhibitions generated by a double-negative feedback loop between CDK and Wee1 and a positive feedback loop between Cdc25 have as a consequence a limited range of CDK activity in which Wee1 and Cdc25 are phosphorylated. Such kind of acute response to a continuously fluctuating input is called ultrasensitivity and was first made acquainted in the cell cycle field by Novak and Tyson (Novak and Tyson, 1993). Honoring the fruitful mathematics tradition of describing events with differential equations, they predicted the mechanisms responsible for the quick oscillations in the early mitotic cycles of *Xenopus* embryos. This somehow simplified nine-dimensional model, which accounted also for CAK-mediated activation of CDKs and Cyclin degradation, has been proven to be accurate in the description of several of the concepts introduced by it. Further on, increasingly sophisticated models were proposed adding the additional complexity of the molecular control of CDK activity learned throughout the years.

A tradeoff must be accepted when modeling intricate biological processes and, typically, simplification for the sake of interpretability is required if one wants to mathematically describe what happens within a complex system like a cell. In this case, the most obvious casualty is

Mik/Myt1 mediated T14 and Y15 phosphorylation, which tends to be considered to complement the Wee1 activity. Mutants for both of the genes encoding these kinases present a more severe phenotype than single mutants (Lundgren et al., 1991).

Indeed, replacing both T14 and Y15 residues of CDK for non-phosphorylatable counterparts (T14AY15F, also called AF mutant) results in an overall more linear dynamic giving place to low amplitude oscillations of CDK activity (Pomerening et al., 2005; Pomerening et al.,2008; Gavet and Pines, 2010). Surprisingly, in a simplified cell cycle context, the AF mutants appear to be remarkably healthy in fission yeast, contrasting with the low cell viability of Cdc2 Y15F mutants (Coudreuse and Nurse, 2010). This strain of *S. pombe* can, therefore, not only drive the entire cell cycle with a single cdk-cyclin complex, but it is also able to escape the tight Cdc2-Wee1-Cdc25 regulation. The resulting more linear CDK activation can be used to better understand its dynamics (Swaffer et al., 2016).

In mice, heterozygous Cdk1 AF mutants are embryonic lethal, and the number of blastocysts cells generated is greatly reduced, probably caused by DNA damage due to the premature onset of mitotic events during S-phase (Satyanarayana and Kaldis, 2009; Szmyd, 2018). Remarkably, this genotype produces a more severe embryonic phenotype than the concurrent complete ablation of CDK2, CDK4, and CDK6 (Santamaria et al., 2007). These results highlight that it is not only CDK1 that is essential for the normal progression of the cell cycle, but also its regulation.

1.2.4 Reversibility of CDK-mediated phosphorylation

Phosphorylation is the molecular mechanism by which the cell controls its division cycle and it is brought about by CDK-cyclin complexes. A huge body of scientific literature addresses the question of how this CDK-dependent control system functions, but much less is published regarding the contribution of the antagonistic phosphatase activity. The reversibility of CDK-mediated phosphorylation is critical to restore non-phosphorylated states of the cell cycle, but also to control the dynamics of simultaneous and opposing action of CDK and phosphatases to precisely control the phosphorylation output during phase transitions (Krasisnka et al., 2011). Another vital contribution of phosphatase in the cell cycle regulation is the mitotic activation of CDK1 by the removal of Y15 Wee1-mediated phosphorylation, which it has been previously detailed in the section 1.2.3. In budding yeast, the phosphatase Cdc14 was shown to be essential for exiting mitosis by counteracting CDK phosphorylation (Visintin et al., 1998). Nevertheless, this does not hold true for fission yeast, where it has shown not to be essential for mitotic exit but it is required for cytokinesis (Cueille et al., 2001; Trautmann et al., 2014). in S. cerevisiae Cdc14 acts through the intricate FEAR (Cdc fourteen early anaphase release) and MEN (the mitotic exit

network) regulatory networks to ultimately coordinate process of chromosome segregation and exit from mitosis (Stegmeier and Amon, 2004; Bloom et al., 2011; Kao et al., 2014). In humans, the role of CDC14 is not well defined, although its activity is not essential for mitosis and appears to alter alternative features, such as ciliogenesis (Berdougo et al., 2008; Partscht et al., 2021). The lack of functional conservation of Cdc14 suggest that multiple CDK-opposing phosphatases might be present in the eukaryotic cell, allowing for the diversification of the functions of some of them. Out of the 117 protein phosphatases identified in the Drosophila proteome, 22 have been shown as to have roles in the cell cycle (Chen et al., 2007). Two of them, PP1 and PP2A, have been long known to regulate cell cycle processes by directly opposing CDK activity (Cyert and Thorner, 1989). Both of these phosphatases are members of the same Ser/Thr phosphatase family, although they are not functionally redundant and can be differentiated by their response to Okadaic acid treatment, which is a potent inhibitor of PP2A but has a lesser effect over PP1 activity (Kinoshita et al.,1990; Bialojan and Takai, 1988).

PP1 has been associated with several different functions during the cell cycle, such as dephosphorylation of the Rb proteins in G1 (Ludlow et al., 1993), reversing CDC7-mediated phosphorylation of the MCM complex during S-phase (Hiraga et al., 2014), and regulation of the kinetochore disassembly during mitosis by counteracting AURB activity (Emanuele et al., 2008). PP1 was also shown to be partly responsible for the regulation of CDK1 mitotic activity by the modulation of the CDK1-CDC25 feedback loop, and therefore contributing to the mitotic entry (Margolis et al., 2006). PP1 has been identified as the phosphatase required for mitotic CDK1 substrate dephosphorylation needed for exit from mitosis in early stages of Xenopus embryos (Wu et al., 2009). PP1 is a holoenzyme that forms heterodimeric or heterotrimeric complexes with the so-called PP1-interacting proteins, which grant the substrate specificity to those complexes. One of these interacting proteins is the nuclear inhibitor of PP1 (NIPP1), which blocks dephosphorylation by PP1 with exception of substrates that are targeted by the FHA domain of NIPP1. This was exploited by researchers that overexpressed a mutated version of NIPP1 carrying a non-functional FHA domain and therefore acting as an extremely selective and potent inhibitor of PP1 (Winkler et al, 2015). Specific inhibition of PP1 in vertebrate cells causes defects in chromosome condensation and in spindle assembly, similar to what is observed in invertebrates (Axton et al., 1990; Hisamoto et al., 1994).

PP2A was originally described as the main phosphatase activity commanding the regulation of mitosis, by opposing Xenopus MPF activity (Cyert and Kirschner, 1988; Lee et al., 1991). During the early 1990s the first pieces of evidence showed that PP2A negatively regulates CDK1 activity and, at the same time, is capable of reversing CDK1-mediated phosphorylation of Histone H1 at

multiple sites (Yamashita et al., 1990; Félix et al., 1990; Sola et al., 1991). Considering that PP1 and PP2A might account for about 90% of all cellular phosphatase activity, both phosphatases have been deemed responsible for reversing cell cycle phosphorylation (Remmerie and Janssens, 2019). Nevertheless, specific phosphatase activities are hard to assess given that there exist multiple isoforms of the PP2A which bind regulatory subunits to form holoenzymes with different target specificities. These holoenzymes consist of 3 subunits, two of which A (scaffold) and C (catalytic) constitute the core, and a third B subunit that, much like in PP1, defines the substrate recognition. In humans, both A and C subunits are present in two isoforms (α and β) that can bind different isoforms of the 4 families of regulatory B subunits, resulting in a large number of possible PP2A holoenzyme configurations (Seshacharyulu et al., 2014). The most relevant complex for the cell cycle regulation is the PP2A-B55 complex that is associated with the regulation of CDK1 activity by dephosphorylating CDC25 and Wee1 to modulate the mitotic onset switch (Krasinska et al., 2011; Rata et al 2018). PP2A-B55 is also considered the main phosphatase activity opposing CDK1 in mitosis, and many mitotic substrates have been inferred by observing premature dephosphorylation of proteins upon releasing PP2A-B55 from its ENSA-Gwl control system, although only a few have been conclusively established (Amin et al., 2022).

1.2.5 Ultrasensitivity, hysteresis and bistability

The cell cycle could be conceived as an irreversible series of distinct biochemical states, where in each one of them a set of reactions will take place to advance to the next one. The transition between phases should be then governed by a control mechanism that acts as some kind of switch. Using concepts from dynamical system theory, Béla Novák and John Tyson first mathematically modeled the regulation of mitosis by MPF in Xenopus egg extracts (Novák and Tyson, 1993). they were able to reproduce observed experimental results employing equations that considered the intricate network that generates positive and negative feedback loops. Some years later, James Ferrell and colleagues applied concepts introduced by Goldbeter and Koshland in the early 1980s (Goldbeter and Koshland, 1981) to explain how a continuous input of phosphorylation could result in a switch between two states (Huang and Ferrel, 1996; Ferrell, 1996). If multiple phosphorylation events in a protein occur by a distributive mechanism, rather than processively, opposing activities of kinases and phosphatases result in a sigmoidal signal–response curve (Ferrell, 1996). Different from hyperbolic curves described by Michaelis-Menten for non-cooperative enzymes, sigmoidal curves have the characteristic of responding abruptly to small changes in the input around the inflexion point. This type of sharp response to a stimulus is called ultrasensitivity.

Another common behavior is also observed in complex systems with entangled negative and feedback loops: the response not only depends on the input signal, but also on the history. If we consider a change in the stimulus, some systems can have different responses depending on the initial state of the system. This concept is called hysteresis (Tyson et al., 1996), and the G2/M transition constitutes an elegant example. Mitosis onset requires a high number of phosphorylated substrates arising from a gradual increment in CDK activity. Because the system is ultrasensitive all substrates are phosphorylated quickly at a given threshold of CDK activity. Any disturbance generating a minimal drop in CDK activity at this point could send the system to low levels of phosphorylated substrates very quickly. In reality, this instability is fixed by having different thresholds of CDK activity for substrate phosphorylation and dephosphorylation depending on whether the initial state is a high or low level of phosphorylation (Thron, 1996). Simultaneous phosphorylation of mitotic substrates will require a given level of CDK activity considerably higher than the level required to reverse that phosphorylation. Hysteresis leads to another common property of biochemical switches that is multistability, and, more specifically for cell cycle switches, bistability. That is, the system only has two stable states similar to an on-off switch, and no transitional states can be observed.

An extra layer of complexity should be taken into account: similar to what was described for mitotic onset, most cell cycle transitions are showing bistable behavior (Yang and Ferrell, 2013; Yao et al., 2008; Barr et al., 2016) and they are all interconnected. This series of toggle switches are coordinated in such a way that the entire system has a memory slot to store information about its status and, thus, "'remembers' which stages it has already completed" which controls the unidirectionality of the process (Novák and Tyson, 2021). The details concerning the integration of the different cell cycle transitions are outside the scope of this thesis, although the reader can be referred to recent publications by Tyson and Novak which address these points (Novák and Tyson, 2021; Novák and Tyson, 2022; Tyson and Novák 2022).

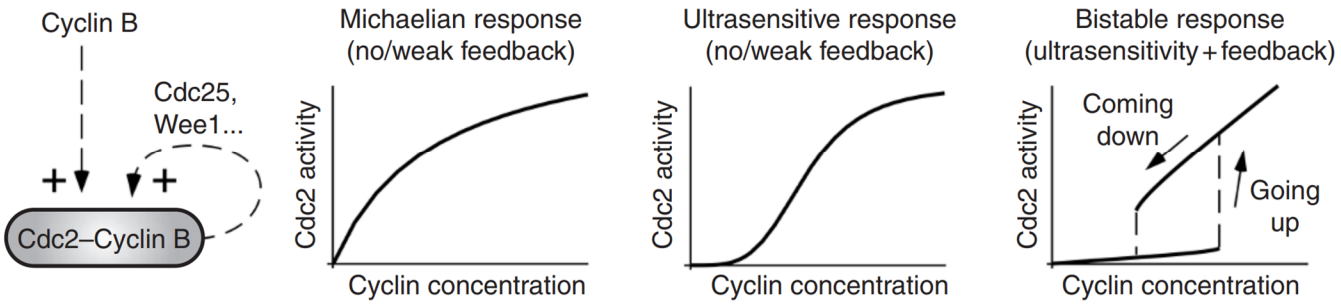


Figure 1.4 | Contribution of utransensitivity and feedback to CDK response. From *Pomerening et al., 2003.*

1.3 Models for the cell cycle control by CDKs

CDK-mediated phosphorylation is the force that propels the advance of the cell cycle. Until this point, I have described these molecular machineries and the precise self-regulation of their activities. However, comprehensive models explaining how the phosphorylation of multiple substrates can bring about the ordered succession of events of the cell cycle have not yet been presented. In this section, I will address the basis for CDK substrate recognition and the accepted models for S-phase and M-phase CDK regulation. Taking into account all this information, I will then introduce two general models describing the CDK-mediated control of the cell cycle, the evidence supporting them as well as their limitations.

1.3.1 CDK substrate specificity

In the 90s, biochemical and crystallographic studies were performed to understand how CDKs could recognize their substrates. Experiments using CDK1/2 determined their affinity for different sequences and revealed what is nowadays considered the CDK phosphorylation full consensus motif: S/T-P-X-K/R (Holmes and Solomon, 1996; Minshull et al., 1990; Songyang et al., 1994). Thus, a Serine or a Threonine (position 0) are candidates to be phosphorylated by CDK if there is a proline in the position +1, any amino acid in the position +2, and a basic residue in the position +3. Analysis of x-ray diffraction of the crystallization of CDK-cyclin complexes with target peptides revealed the structural basis of the importance of these residues. The proline at the +1 position helps to reshape and stabilize the binding pocket, directing the +2 amino acid towards the outside and contacting the solvent, while orienting the +3 basic residue to form a hydrogen bond with the phosphorylated T160 of CDK (Brown et al., 1999). Nonetheless, these strict requirements for CDK phosphorylation would later be subjected to further evaluation. Indeed, it seems that the proline residue in the position +1 (also called minimum consensus motif) accounts for most of the phosphorylation attributed to CDK in high-throughput proteomic analysis (Holt et al., 2009; Michowski et al., 2020; Chi et al., 2020; Blethrow et al., 2008). However, in budding yeast, more than six hundred targets have been described (Holt et al., 2009) and more than 1000 different substrates have been identified for human ES cells (Michowski et al., 2020), with 51% and 47% of them not being phosphorylated in the minimal consensus motif, respectively. A major conclusion can be drawn from these data: although preponderant, the consensus motif does not seem to be a strict requirement for CDK recognition of the substrates, implying that CDK might not be a particularly specific enzyme.

Cyclins have also been described to have a region, called “the hydrophobic patch”, that drives the interaction of CDK-cyclin complexes and their substrates and, therefore, contributes to the specificity of CDK phosphorylation (Schulman et al., 1998; Brown et al 1999). This mechanism of cyclin-substrate docking has prompted the interest of some researchers in the field, most notably David Morgan and Mart Loog. In 2005, they published a paper where they showed that, despite a general higher CDK1-Clb2 activity over CDK1-Clb5 for common substrates, specific CDK1-Clb5 substrates do not get phosphorylated when Clb5 hydrophobic patch gets mutated (Loog and Morgan, 2005). This highlights the importance of this recognition system for at least some CDK substrates, but it also provided them with a mechanism to support the existence of qualitative differences between CDK-cyclin complexes that could explain the expansion of cyclin genes and why they are differentially expressed during the cell cycle. The most recent work of Loog’s team will be further discussed in the context of the models proposed to describe the cell cycle control system.

Nevertheless, neither the consensus motif, the hydrophobic patch nor the combination of both seem to account entirely for CDK substrate recognition. The question then arises of how CDKs select their targets. The answer might come from an emergent concept in structural biology: intrinsically disordered regions of proteins (IDRs). These unstructured regions of proteins are exposed to the solvent and contain an increased proportion of serines and threonines, which can be phosphorylated by CDKs (Uversky, 2019; Quaglia et al., 2021). Moreover, IDRs seem to be far more frequent than previously thought and they are predicted to account for about 20% of the proteome for eukaryotes (Peng et al., 2014). The formal definition of this molecular feature, the structural consequences derived from it and its relationship with CDK-mediated phosphorylation will be detailed in section 1.4.

1.3.2 CDK control of G1 and S-phase

After mitosis, each daughter cell has the choice of following one of two paths: it can continue in a replicative state and restart the cell cycle to divide itself or it can exit the cell cycle.This decision should be made before the so-called restriction point (or “start” in yeast), after which cells are committed to continue with the cell cycle in its totality (Pardee, 1974). The current consensus for G1 progression states that upon growth factor signaling (Planas-Silva and Weinberg, 1997; Sherr, 2000), D-type cyclin levels increase and they form complexes with CDK4/6 that mono-phosphorylate the Retinoblastoma (Rb) tumor suppressor protein (Narasimha et al., 2014). Unphosphorylated Rb can bind and sequester the E2F transcription factors. CDK4/6-mediated mono-phosphorylation partially inactivates Rb, releasing E2F transcription factors that will

promote the expression of cyclin E, which in complex with CDK2 will hyper-phosphorylate Rb to complete its inactivation. This positive feedback loop assures that E2F and cyclin E activate their own expression and other genes needed to progress into S-phase (Johnson et al., 1994; Spencer et al., 2013). Once critical levels of Cyclin E and E2F are obtained, the system becomes insensitive to variations in growth factor signaling (Yao et al., 2008). Initial CDK4/6 phosphorylation of Rb has, therefore, become synonymous with the restriction point. Nevertheless, single cell analysis revealed that the restriction point and Rb phosphorylation do not coincide in time and there might exist more than one checkpoint before S-phase onset (Martinsson et al., 2005). This double restriction point theory was further supported by evidence showing that the commitment to the next cell cycle is made, at least in part, at the end of the preceding cycle and it is dependent on CDK2 activity levels at the end of mitosis (Spencer et al., 2013). Some authors have recently presented a unified model for the G1/S phase transition suggesting three different "commitment points" that cells have to go through to start the DNA replication (Hume et al., 2020). This model can more satisfactorily explain the evidence available, which is a desirable feature of a newly-introduced model, but it nevertheless relies exclusively on *post-hoc* reasoning and no prediction was proposed. The existence of a restriction point in G1 and even the existence of G1 itself, as well as the non-replicative phase G0, is not free of controversy and some researchers do not believe that the current evidence supports them. Stephen Cooper, an outspoken representative of this point of view, argues that there is no clear evidence that justifies the need of a control point in G1 and G0 occurrence, and that those models reflect, more likely, "anthropomorphic constructs" rather than reality (Cooper, 2003; Cooper, 2020). Indeed, the canonical model of 4 phases in the cell cycle (5, if we consider G0) was defined before obtaining any evidence that those are distinctive biochemical states of the cell. A recent review from Dr. Fisher and Dr. Krasinska also advocate for this idea which it is well illustrated by the following extract: "If, as suggested , the restriction point does not exist, and G1 is not a discrete phase of the cell cycle, but a variable period before exponential DNA replication origin activation occurs, then the so-called G1 cyclins might not be required for the cell cycle" (Fisher and Krasinska, 2022).

Whether or not G0 and G1, together with their proposed regulatory mechanism, occur as distinct phases of the cell cycle, the increase in CDK activity levels (CDK2, in the canonical model) is required for S-phase onset. Two individual steps are needed to initiate the DNA replication in eukaryotes. First, in late mitosis and during G1, the six subunits of the origin-recognition complex (ORC) are loaded onto the DNA at sites called "replication origins". This complex will then recruit the CDC6 and CDT1 proteins to form a complex that will then direct the loading of the

minichromosome maintenance (MCM2-7) complex, that performs the helicase activity for unwinding the DNA during replication (Donovan et al., 1997; Rowles et al., 1999; Seki and Diffley 2000; Evrin et al., 2009; Remus et al., 2009; Gambus et al., 2011). Altogether, this protein complex is known as the pre-replicative complex (pre-RC), which stays inactive until the beginning of S-phase, and this process is recognized as origin licensing. The second step is the firing of the replication origins, in which the helicase is activated allowing the unwinding of the double-stranded DNA and making each strand of the DNA accessible to polymerases. The activation of the helicase requires the formation the assembly of CDC45, MCM2-7, and GINS proteins, to form the the CMG complex (also known as pre-initiation complex or pre-IC) (Aparicio et al., 2009), that triggers the recruitment of DNA polymerases Pol ε, in the leading strand, and Pol α/primase and Pol δ, in the lagging strand (Garg and Burgers, 2005; Pursell et al., 2007; McElhinny et al., 2007). Both steps are cell cycle-coordinated and controlled by CDK activity. For instance, in budding yeast the ORC remains loaded for the entire duration of the cell cycle (Bell and Stillman, 1992; Diffley et al., 1994; Liang and Stillman, 1997; Tanaka et al., 1997). In contrast, Cdc6 is removed in S phase upon Cln-Cdc28 phosphorylation avoiding re-licensing of origins (Perkins et al., 2001; Piatti et al., 1996). In metazoan, particularly in humans, some details still remain unclear, such as the timing for the assembly of the pre-RC that has not been precisely determined. Differently from yeast, ORC1 is degraded at the onset of S phase and then re-synthesized in late G2 (DePamphilis, 2003; Kara et al., 2015; Kreitz et al., 2001; Méndez et al., 2002). CDC6 is phosphorylated by CDK2-cyclin E at some point during mid-G1, stabilizing it and promoting the formation of the pre-RC (Cook et al., 2002; Coverley et al., 2002). During S-phase CDC6 undergoes CDK2-cyclin A-mediated phosphorylation driving its re-localization to the cytoplasm (Delmolino et al., 2001; Jiang et al., 1999; Petersen et al., 2000). The activation of the helicase by the formation of the CMG complex is also mediated by phosphorylation. MCM subunits are phosphorylated by the Dbf4-dependent kinase (also DDK or CDC7) (Lei et al., 1997; Sheu and Stillman, 2010; Tsuji et al., 2006; Yeeles et al., 2015; Labib, 2010) and although DDK was suggested as essential for the replication process (Labib, 2010), a recent paper demonstrated the existence of redundancy between CDC7 and CDK1 in MCM phosphorylation (Suski et al., 2022). Indeed, this work shows that “at least one of these kinases must be present to allow S-phase entry”. Apart from the CMG complex, the replisome is also composed of other proteins that are not strictly part of the replication machinery but are nevertheless essential for DNA replication, such as RecQ4, Treslin/Ticrr, and TopBP1 (Sld2, Sld3 and Dpb11, respectively, in budding yeast). Those factors appear to be important for the recruitment of GINS to the complex and for origin firing (Yeeles et al., 2015). The scaffold protein TopBP1 contains BRCT repeated motifs that have

a high affinity for phosphorylated proteins (Manke et al. 2003; Yu et al. 2003). CDK-mediated phosphorylation of RecQ4 and Treslin regulates the localization of their interaction with TopBP1, required for the assembly and activation of the CMG helicase (Kumagai et al., 2010; Pagliuca et al., 2011; Kumagai et al., 2011; Boos et al., 2011; Gaggioli et al., 2014; Lu et al., 2017). Additionally, TopBP1 is shown to play key roles in DNA repair pathways and it might be an important factor to coordinate DNA replication with the DNA damage checkpoint at G1 and the end of S-phase (Cescutti et al., 2010; Liu et al., 2017; Kim et al., 2015). Although the DNA replication pathway is conserved across all eukaryotes, there exist many mechanistic differences in various organisms and sequence conservation of some key proteins is not as stringent as one would expect. Such is the case of Sld2 and Sld3 whose metazoan homologs (RecQ4 and Treslin, respectively) were only recently described (Kumagai et al, 2010; Labib, 2010; Zegerman, 2015). Whether these reported differences truly reflect a biological reality or they are due overestimation of functions of each protein is not yet clear. Throughout the extent of this manuscript, I will present an alternative model that could explain such inconsistencies, in which the formation and dissolution of protein condensates with biochemical activity is regulated by non-site specific CDK-mediated phosphorylation.

1.3.3 G2 and M-phase control

Cyclin A-associated CDK1 and CDK2 are important to phosphorylate multiple substrates involved in DNA replication and DNA damage repair systems. No clear evidence is available to demonstrate differential functions of these kinases during this point of the cell cycle. Cyclin A expression begins at S-phase onset and increases steadily until late G2 where it reaches its maximum levels, promoting the initial steps of chromosome condensation and presumably nuclear envelope breakdown (Pagano and Draetta, 1991; Furuno et al., 1999; Gong et al., 2007). It is then ubiquitinated by APC/C during prometaphase and degraded by the proteasome in a spindle checkpoint-independent manner (Elzen and Pines, 2001). At this moment cyclin B accumulation reaches its peak and in complex with CDK1 it promotes the continuation of M-phase. Indeed, high levels of Cdk1-cyclin B kinase activity drives the complete cellular re-organization observed during mitosis (Morgan, 2007). Some details of the regulation of the cyclin B degradation by APC/C remain unclear, although the contribution of the spindle assembly checkpoint (SAC) and autoregulation by CDK1-cyclin B phosphorylation seem to be critical (Morgan, 2016). In contrast to cyclin A, cyclin B degradation was reported to be dependent on the spindle checkpoint by a mechanism in which Mad1-Mad2 loaded onto kinetochores promote the

formation of the Mad2-Cdc20 complex that, in turn, will bind BubR1-Bub3 to produce the mitotic checkpoint complex (MCC), inhibiting APC/C^{Cdc20}-mediated degradation of cyclin B (Musacchio, 2015; Corbett, 2017). Furthermore, a dual mechanism for controlling the APC/C^{Cdc20} activity during G2/M transition has been recently reported in *C. elegans* and human cells, by which soluble Mad1-Mad2 complexes and CDK mediated phosphorylation block specific Cdc20 motifs interfering with its binding to APC/C and thus lowering APC/C^{Cdc20} activity.

The SAC is, therefore, a complex molecular machinery that controls the proper attachment of chromosomes to the mitotic spindle through a precise modulation of APC/C activity (Musacchio, 2015). APC/C ubiquitylation and the consequent degradation of proteins, most notably cyclin B, seem to be the key effector of the spindle checkpoint signaling and mitotic progression (Pines, 2006). Silencing of SAC and the advancement of mitosis depend on signaling pathways capable of sensing cell size and the mitotic spindle geometry (Bloomfield et al., 2021; Chen and Liu, 2014). Sensing of kinetochore attachment is also crucial for SAC silencing timing and phosphorylation has been proposed as the main signaling for this process. First, the phosphorylation of SAC members by the chromosomal passenger complex CPC, composed of the inner centromere protein (INCENP), survivin (BIRC5), borealin (CDCA8) and aurora kinase B (AURKB) as the catalytic subunit, has been associated with the state of the attachment between microtubules and kinetochores (Tanaka et al., 2002; Welburn et al., 2010; Liu et al., 2009). AURKB also promotes the recruitment of MPS1 to kinetochores, a kinase that has been reported to be responsible for sensing microtubule attachment by being displaced from kinetochores due either to competition with microtubules for the binding to the major microtubule receptor complex Ndc80C (Hiruma et al., 2015; Ji et al., 2015) or to a mechanical switch that increases the distance between MPS1 and its substrates (Aravamudhan et al., 2015). In the absence of MPS1 activity, PP1 phosphatase dephosphorylate sites in kinetochore proteins which are critical for the assembly of the MCC and thus releasing the APC/C^{Cdc20} inhibition (Emanuele et al., 2008; Corbett, 2017). High levels of CDK1-cyclin B activities promote APC/C^{Cdc20} activation through phosphorylation of APC (Golan et al., 2002; Kraft et al., 2003; Qiao et al., 2016) that, in turn, targets cyclin B1 for degradation by the proteasome. It has been reported that this negative feedback loop generates a “time-delayed, hypersensitive switch” making CDK1-APC/C an oscillator that controls the timing of the anaphase-metaphase transition onset (Yang and Ferrell, 2013). At anaphase APC/C switches its substrate adapter proteins CDC20 for CDH1 (Hagting et al., 2002), another WD repeat protein with a broader substrate recognition range. Opposite to what is observed for CDC20, CDH1 binding and activation of APC/C is inhibited by CDK1-mediated phosphorylation (Zachariae et al., 1998; Jaspersen et al., 1999; Kramer et al., 2017).

Degradation of cyclin B and securin promotes the metaphase-to-anaphase transition by releasing their inhibition over the endopeptidase separase (Holland and Taylor, 2008; Kumada et al., 1998; Ciosk et al., 1998; Stemmann et al., 2001; Gorr et al., 2005). Recent structural studies shed light on the mechanisms of separase inhibition by securin which acts as a pseudo-substrate and block the active site, and by CDK1-cyclin B that phosphorylates and binds separase to activate autoinhibitory motifs of the protein (Yu et al., 2021). Another inhibitor, the shugoshin 2 (SGO2) protein, has been recently described to block separase in a SAC-dependent manner but independent from securin (Hellmuth et al., 2020). Indeed, SGO2 contains a pseudosubstrate motif but no structural evidence has yet been presented. When active, separase can catalyze the cleavage of the kleisin subunit (Rad21/Scc1/Mcd1/Rec8) of the cohesin complex (Uhlmann et al., 2000). The initial step of removal of the cohesin complex requires the action of CDK1, AURKB and polo-like kinase 1 (PLK1) that phosphorylate specific subunits driving their dissociation and the consequent unloading from most chromosomal regions (Tedeschi et al., 2013; Nishiyama et al., 2013). Centromeric cohesin, however, reminds bound to chromosomes protected against kinase activity by the action of a complex formed by shugoshin 1 (SGO1) and the protein phosphatase 2A (PP2A) that antagonizes the phosphorylation of cohesin subunits (Liu et al., 2013). The presence of cohesin at centromeres allows for the chromosome alignment in the metaphase plate. At anaphase onset, APC/C degradation of inhibitors of separase releases its endopeptidase activity to cleave the RAD21 subunit of the centromeric cohesin thereby disassembling the complex and allowing separation of the sister chromatids by mechanical forces carried out by the mitotic spindle.

Another way in which CDK activity can control the timing of mitotic events is by modulating its own activity beyond CDK and cyclin expression and degradation. Indeed, mitotic phosphorylation levels are determined by the ratio in the activity of CDK1 and the opposing phosphatase PP2A. Analytical resolution of mathematical models that consider futile cycles of opposing enzymes predict that those systems will present ultrasensitivity in the absence of allosteric regulation, generating a switch-like network output (Goldbeter and Koshland, 1981). Additionally, the CDK1-PP2A phosphorylation network presents auxiliary feedback loops that contribute tp the switch-like and irreversible mitotic onset. It was previously introduced that CDK1 itself can phosphorylate Wee1 and Cdc25, the kinase and the phosphatase that regulate the inhibitory Y15 phosphorylation. This CDK-mediated phosphorylation inhibits the kinase activity of Wee1 (Mueller et al., 1995; Okamoto and Sagata, 2007) and at the same time promotes Cdc25 phosphatase activity (Hoffman et al.,1993), generating a positive feedback loop. A second interconnected feedback loop occurs through the regulatory pathway Greatwall (Gwl)-endosulfine

(ENSA)/ARPP19, capable of phosphorylating and inhibiting the PP2A phosphatase in complex with its regulatory subunit B55 (Vigneron et al., 2009; Gharbi-Ayachi et al., 2010). PP2A:B55 can reverse the activatory CDK-mediated phosphorylation of Gwl to increase its own activity (Hégarat et al., 2014). PP2A:B55 is also responsible for dephosphorylating Wee1 and Cdc25 to oppose CDK autoactivation (Kinoshita et al., 1993; Chica et al., 2016; Lucena et al., 2017). An additional regulatory point lies in the switch from cyclin A to cyclin B as binding partner of CDK1 to modulate its mitotic activity levels. Cyclin A appears to be required for mitotic entry, while cyclin B is essential for sister chromatids separation and exiting mitosis (Hégarat et al., 2020). This complex network results in a series of CDK1 bi-stable switches that are responsible for the timely ordering of the different mitotic events, such as chromosome condensation, nuclear envelope breakdown (NEB), chromosome segregation and cytokinesis (Hégarat et al., 2020; Hégarat et al., 2016).

In order to proceed correctly with the separation of the sister chromatids, cellular DNA has to be condensed to form mitotic chromosomes. Estimates by different experimental methods indicate that mitotic chromosomes are 2- or 3-fold more compacted than chromatin during interphase (Vagnarelli, 2012). During early mitosis, before the breakdown of the nuclear envelope, the activation and relocation of CDK1-cyclinB to the nucleus drives the molecular processes that mediate chromosome condensation (Gavet and Pines, 2010a; Gavet and Pines, 2010b). Premature onset of CDK1 phosphorylation during interphase by okadaic acid-mediated inhibition of PP2A results in premature chromosome condensation (Gotoh and Durante, 2006).). The mechanism of the condensation of chromosomes has not been completely deciphered yet due in part to the fact that more than 4000 different proteins have been identified as to be associated with mitotic chromosomes, apart from histone proteins that constitute around 40% of the total chromosomal mass (Ohta et al., 2010). Nonetheless, there are some well characterized proteins known to contribute to this process such as condensin I and condensin II complexes, topoisomerase IIα and the KIF4A kinesin (Hirano, 2005; Lewis and Laemmli, 1982; Maeshima and Laemmli, 2003; Mazumdar et al., 2004). Their established roles in chromosomal condensation have been reviewed in more detail by Vagnarelli (Vagnarelli, 2012). Not surprisingly, those roles seem to be regulated by CDK1 and, most probably, by other cell cycle-related kinases such as AURKB, PLK1 and casein kinase II (Bazile et al., 2010; Dong et al., 2018; Wei-Shan et al., 2019; Maeshima and Laemmli, 2003).

After chromosome condensation, another process takes place in the cell to restructure it: the nuclear membrane gets dissociated. Activation of CDK1-cyclin B in early mitosis relocalizes this complex to the nucleus and enhances its own nuclear import (Gavet and Pines, 2010b). CDK1, together with other cell cycle kinases such as the NIMA-related kinase (Nek) family and PLK, can

specifically phosphorylate important members of the nuclear pore complex (NPC) and the nuclear lamina driving its dissolution (Blethrow et al., 2008; Laurell et al., 2011). Most notably, the soluble nucleoporin NUP98 is hyperphosphorylated and it appears to be the first member to be released from the NPC (Dultz et al., 2008; Hase and Cordes, 2003; Lenart et al., 2003). This phosphorylation burst occurs simultaneously with the loss of the nuclear permeability barrier (Dultz et al., 2008). To complete the nuclear envelope breakdown (NEB), CDK1-mediated phosphorylation drives the polymerization of the nuclear lamina (Heald and McKeon, 1990; Peter et al., 1990). CDK1 also phosphorylates multiple components of the inner nuclear membrane (INM), such as LAP2α and lamin B receptor (Courvalin et al., 1992; Macaulay et al., 1995; Dechat et al., 1998; Tseng and Chen, 2011). The breakdown of NE promotes the search-and-capture process by which spindle microtubules attach to kinetochores in sister chromatids and separate them by pulling toward the opposite spindle poles. Upon the inactivation of CDK1-cyclinB in late anaphase, the reassembly of the nuclear envelope begins and by the moment the new cell cycle initiates, all chromatin of the daughter cells is contained within integral nuclear envelopes (Ungricht and Kutay, 2017).

1.3.4 The qualitative model vs. the quantitative model

The eukaryotic cell cycle, as we understand it, is the biochemical organization of a cell in a timely ordered manner to perform the reactions that will allow the cell to duplicate and to transfer the hereditary information to its progeny. CDK-mediated phosphorylation was identified as the molecular control that coordinates the progression of different cellular states through the cell cycle. In yeast, a single CDK associates with different cyclins to drive the different phases of the cell cycle. In contrast, higher eukaryotes have a considerably bigger *repertoire* of genes encoding CDKs and cyclins which scientist intuitively attributed to an increase in the complexity of the cell cycle control mechanism and the coordination of multicellularity to develop an organism. Borrowing control systems concepts from other fields such as clocks, oscillators, switches and checkpoints, researchers modeled what is today considered the canonical model to explain the cell cycle,to which I will refer as the “qualitative model” hereinafter. The qualitative model assumes that intrinsic biochemical properties of the different CDK-cyclin complexes can phosphorylate specific substrates and autoregulate the shift to the next cell cycle phase. In humans, for instance, CDK4 and CDK6 in complex with cyclin D will be active during G1 and their main target is the retinoblastoma protein. These two kinases together with CDK2-cyclin E will hyperphosphorylate Rb releasing it from E2F, thus promoting the expression of genes essential for DNA replication and the G1-S transition. CDK2-cyclin A activity is critical to drive S-phase until G2, where CDK1

binds to cyclin A for the mitotic onset, followed by the activation and nuclear accumulation of CDK1-cyclin B that drive sister chromatids segregation and its subsequent degradation that allow the exit from mitosis.

This model that proposed the cell cycle as an ordered series of events with intuitive control systems seemed to be compelling to the scientific community, being helpful to explain some observations, and was broadly adopted. However, early evidence in fission yeast indicated that in reality the cell cycle might be controlled in a more robust and flexible manner. Indeed, the specificity provided by different cyclins might not be essential (Stern and Nurse, 1996), and yeast strains bearing only the mitotic cyclin Cdc13 can adequately coordinate all phases of the cell cycle (Fisher and Nurse, 1996). Years later it was confirmed that a yeast strain containing only one CDK-cyclin complex was viable (Coudreuse and Nurse, 2010). Furthermore, this complex was encoded by a genetic construction allowing for the external modulation of total CDK activity. Low and high CDK activity thresholds could trigger S-phase and M-phase, respectively, independently of the cell cycle phase that the cell was traversing. This provided strong evidence to theorize that the cell cycle does not present an “intrinsic directionality” and, ultimately, it only depends on the quantities of global CDK activities (Coudreuse and Nurse, 2010). This model that considered that the cell cycle is driven by the global activity of all CDKs, rather than the contribution of different CDK-cyclin complexes for specific processes, was titled “the quantitative model” (Fisher and Nurse, 1996; Stern and Nurse, 1996). More definitive evidence came with genetic manipulation in animals: CDK6 knockout mice are viable while double knockouts of CDK4 and CDK6 do not impair the first stages of embryonic development (Malumbres et al., 2004). Moreover, while studying the kinetics of the cell cycle, researchers did not observe any difference between double mutant fibroblasts and wild-type cells (Malumbres et al., 2004). Furthermore, mice lacking all three D-type cyclins showed similar results (Kozar et al., 2004). Additionally, mice where all interphase CDKs have been deleted except for CDK1 are able to reach mid-gestation and mouse embryonic fibroblast (MEFs) produced from these embryos can proliferate *in vitro* (Santamaria et al., 2007). In that case, CDK1 is capable of binding all cell cycle-related cyclins and proceed with functions normally associated with other CDK-cyclin complexes, such as Rb phosphorylation. Altogether, this evidence illustrates a high degree of functional overlap between different CDKs and cyclins, supporting the hypothesis that levels of CDK-mediated phosphorylation are what determines the onset of different phases of the cell cycle. A literature review of cell cycle CDK substrates revealed a striking overlap between pairs of CDK-cyclin complexes that are consecutive in the cell cycle chronology, indicating that specificity might be due to timing rather than intrinsic biochemical properties of cyclins (Errico et al., 2010). Interactomes analized by mass spectrometry for cyclin

E, A and B in HeLa cells have shown a considerable amount of overlap between substrates bound by different cyclins, even when the determination for each cyclin was performed in the cell cycle phase when they are more enriched. A recent work in fission yeast showed a remarkable lack of specificity between S-CDK and M-CDK phosphorylation *in vivo*, with 65% of the phosphorylation detected presenting an identical behavior in both conditions (Basu et al., 2022). In engineered human cells where CDK1 and CDK2 were under the control of a chemical-genetic system that allows a selective and quick degradation of these kinases, CDK1 is enough for driving the entire cell cycle (Lau et al., 2021). CDK2 has less intrinsic phosphorylation activity than CDK1 but, in its absence, CDK2 can take over mitotic functions by binding cyclin B and phosphorylating mitosis-specific substrates. However, it can only drive the entire cell cycle if its activity is brought to levels equivalent to those of CDK1 (Lau et al., 2021).

Evidence appears to be accumulating to support a model where the overall phosphorylation levels, resulting from opposite activities of CDKs and phosphatases, is the main driver of the cell cycle. This coarse regulation explains very well the robustness of the process, although it is hard to interpret how a single input can precisely coordinate all the different biochemical states observed along the cell cycle. Additionally, research previously introduced still identified CDK-specific or cyclin-specific substrates (Basu et al., 2022; Errico et al., 2010) and cell cycles driven with alterations in CDKs and/or cyclin expression do not occur seamlessly. Mice models lacking CDK4/6 or D-type cyclins are not able to fully develop, and a similar phenotype is observed for those having only CDK1 (Malumbres et al., 2004; Kozar et al., 2004; Santamaria et al., 2007). MEFs generated from mice only having CDK1 showed an extended cell cycle which authors attributed to an inefficient phosphorylation of Rb (Santamaria et al., 2007). If embryos can be detected in post implantation stages, the “core cell cycle” is not considered to be affected because some cell proliferation can take place. Nevertheless, different organisms, cells, or developmental stages present variations in the way their cell cycles are organized and might require particular sets of cell cycle regulators. Currently the best way to conciliate all the evidence available is, most probably, to consider that the mechanisms described in both, the qualitative model and the quantitative model, might contribute to the regulation of the cell cycle.

1.4 The common feature of CDK substrates: disorder

The full consensus substrate recognition motif for CDK (S/T-P-X-R/K) and, to a lesser extent, the minimal consensus motif (S/T-P) promote phosphorylation mediated by CDKs. Nevertheless, modern proteome-wide studies have shown that a considerable proportion of these phosphorylations might occur at serines and threonines that do not conform with these consensus motifs (Michowski et al., 2020; Holt et al., 2009). This prompts the question of whether there may be another molecular feature determining phosphorylation sites, specifically, structural features of the proteins that might favor phosphorylation by CDKs. Indeed, high-throughput studies for assessing its targets showed that CDK phosphorylation appears to be enriched in regions that are not folded, also called intrinsically unstructured regions (IUR) or intrinsically disordered regions (IDR) (Michowski et al., 2020; Holt et al., 2009; Moses et al., 2007). Indeed, the importance of general phosphorylation in IDRs was already discussed, to such a degree that bioinformatic tools based on protein disorder were developed for predicting phosphorylation sites (Iakoucheva et al., 2004).

This section will explore the concept of protein disorder and its relevance in the field of cell cycle for understanding CDK-mediated phosphorylation as a single global force rather than the addition of multiple minor specific events.

1.4.1 Protein disorder: a change of paradigm

The genetic code is the one biological information transfer system that we understand the best. This rather simple cryptographic riddle was almost completely solved by a set of rules proposed by Crick, Barnett, Brenner, and Watts-Tobin in 1961 (Crick et al.,1961). Based mostly on indirect evidence, they stated that each amino acid of a given protein is determined by a continuous non-overlapping set of three nucleotides in the DNA that are read from a fixed starting point.

Curiously enough, in the same year, Anfinsen and Haber showed that proteins could regain their enzymatic activity after being denatured by the reduction of disulfide bonds (Anfinsen and Haber, 1961). This occurs by removing the denaturing agents alone and in the absence of any other catalytic reaction. Moreover, also in 1961, Frederik White confirmed that the native tertiary and secondary structures are being regenerated and the disulfide bonds are spontaneously formed only by air oxidation (White, 1961).

Both of these discoveries set the conceptual basis for the development of Structural Biology applied to proteins. If the DNA holds the information of every amino acid sequence, and these

sequences encode the information of their own tridimensional structures that define the protein function, evolution should be also understood from the structural point of view. Indeed, as a general rule of thumb, protein structures are more conserved than their amino acid sequences, and amino acid sequences are more conserved than the DNA sequences that encode them. Ultimately, a fair amount of evidence exists for concluding that the three-dimensional organization of an amino acid chain is what defines the protein function and, hence, the need for a method for the determination of protein structures.

The answer came from X-ray crystallography (also macromolecular crystallography or MX) that, in fact, had achieved the first protein structure at the atomic resolution some years before the discoveries of Anfinsen, Haber, and White (Kendrew and Parrish,1957; Kendrew et al.,1958; Perutz et al., 1959).

In 1976, a database for depositing the resolved crystal structures was introduced: The Protein Data Bank or PDB. Due to several technical advances in the crystallography field, together with the optimization of heterologous protein expression systems, the number of solved proteins deposited in the PDB saw an explosion during the 1990s. It was only in the year 2008 that depositing MX data was declared mandatory (Burkley, 2021).

The progress in X-ray crystallography, and more recently the breakthrough in nuclear magnetic resonance (NMR) and cryogenic electron microscopy (Cryo-EM) for the determination of macromolecular structures, led to overly optimistic predictions stating that the human structural proteome was going to be completed in the following years.

Regardless, to the present day, only around 20% of the proteome amino acids are covered by the PDB (SWISS MODEL, H. sapiens, 2021). Considering 30% of sequence identity as a threshold to classify proteins to the same folding, the coverage number increases up to 46% of the human ‘structurome’. A number that is relatively small given how permissive the threshold employed is. Thus, the question emerges: what are the limiting factors that scientists are facing to solve all the structures of the human proteome?

Firstly, several technical hurdles need to be sorted in order to obtain the amount of protein required in its native state. Then, each one of the methods used for solving the three-dimensional conformations of proteins has its own pitfalls. While X-ray crystallography has good overall coverage of the amino acid chains, not all proteins can be crystallized, and even if it is possible, a structure obtained under non-physiological conditions might render biologically irrelevant results (Harkey et al., 2019). Cryo-EM does not require a crystallization step and has notoriously grown in usage during the last decade due to a great improvement in the resolution of the acquisitions. However, this resolution is still suboptimal for its use in drug discovery/design (Renaud et al.,

2018). NMR has one desirable advantage over the other methods. It can provide a three-dimensional structure of a protein in solution. However, the amount of pure protein needed is extremely high, it is considerably more expensive, and its interpretation increases exponentially with the size of the protein.

Taking into account these historical and methodological aspects, it is not surprising that about 80% of the structures deposited in the PDB are coming from X-ray crystallography experiments. And that spawned yet another limitation into the structure determination process. Those amino acids located in high-mobility regions of the proteins are not resolved by the X-ray diffraction, and they are cataloged as “missing electronic densities”. Indeed, most of the X-ray structures in the PDB present at least one stretch of 30 or more unstructured residues, also called intrinsically disordered regions of the protein.

We could formally define IDRs as regions in the protein with a lack of a fixed three-dimensional organization in solution. Rather than presenting a funnel-like energy landscape, as folded proteins do, IDPs show a “hilly plateau” geography where multiple local minimums are favored (Fisher and Stultz, 2011; Turoverov et al.,2010; Chebaro et al., 2015). Is like this that these proteins can transition through a continuum of several semi-stable conformations, instead of having a single folded native conformation.

The structure-centered perspective is so strongly embedded in the scientific community that the evolutionary implications of protein disorder have been long neglected. It was originally thought that the proportion of disorder residues of a proteome would reflect the increase of organism complexity throughout evolution and that disorder is a positively selected evolutionary trait. However, apart from the step-like increase between prokaryotes and eukaryotes, the proportion of disordered residues is not increasing steadily with proteome size nor the number of cells of a given organism (Xue et al., 2012; Schad et al., 2014).

Evolutionary rates and constraints change significantly between structured and disorder regions. Studies using different evolutionary models consistently showed that disordered proteins accept a higher number of point mutations (Brown et al., 2010; Schaefer et al., 2010), but also insertions and deletions (Brown et al., 2011). When comparing amino acid change rates, scientists observed that disordered regions evolve more rapidly than structured regions (Brown et al., 2002). This increase in the evolutionary rates of IDRs might be, at least, partially explained by repeat expansions that seem to be located predominantly in those regions (Tompa, 2003; Darling and Uversky, 2017).

This general greater rate of amino acid changes does not imply the lack of function conservation. The existence of predicted disordered region domains with a high level of conservation has been

shown to be associated with a number of biological functions, notably binding to nucleic acids (Chen et al., 2006a; Chen et al., 2006b; Brown et al., 2002). In some other cases, as regions acting as flexible linkers, functional protein disorder seems to be maintained despite the absence of sequence conservation (Daughdrill et al., 2007).

One of the reported functions of disorder is to serve as flexible and accessible display sites for protein modifications, most notably phosphorylation (Iakoucheva et al., 2004). Conservation of this function would have, as a consequence, not only an enrichment of phosphorylatable residues in IDRs, but also their maintenance throughout evolution. Indeed, Serines and, to a lesser extent, Threonines are enriched in disorder but for the mutation rate of specific phosphorylation sites, studies have reported that results change depending on the evolutionary scale used. Landry and collaborators found that, in vertebrates, evolutionary rates are higher for phosphosites located to disordered regions while inferring the phosphorylation status of common ancestors using only human information (Landry et al., 2009). Contrastingly, Chen and collaborators reported functional conservation of phosphorylated sites, regardless of where they are located, for the human-mouse split when using experimental information of phosphorylation for both organisms (Chen et al., 2010). Using experimental phosphoproteomic information of *S. cerevisia*e and complete genome sequence of 32 fungal species, it was found that phosphosites in structured domains are conserved in a position-specific manner while phosphorylation sites in disorder regions seem to be clustered within a given IDR but not conserved in the exact same location (Holt et al., 2009).

Functions and even the existence of intrinsic disorder have been the subject of some controversy. Janin and Sternberg, two notable biochemists and structural biologists, argue that disorder is an artifact arising from *in vitro* experimental setups rather than a property of proteins under physiological conditions (Janin and Sternberg, 2013). They claimed that flexibility in proteins is an established concept that can fully explain most, if not all, the functions attributed to disorder, and that, in fact, IDPs are just “protein waiting for partners” (PWP). Furthermore, experiments and theory both support the idea that cellular crowding should promote the formation of more compact native protein conformations (Minton, 2000). It was also believed that IDPs could not persist in the cell due to their increased solvent exposure that will render them more susceptible to protease degradation (Dunker et al., 2002). This was mostly based on studies showing that local unfolding is required for proteolytic cleavage (Hubbard, 1994). Indeed, proteins presenting IDRs have shorter half-lives than completely globular proteins but this is probably due to cellular homeostasis (van der Lee et al., 2014b)

In 2013, Dunker and Uversky, the two most well-known advocates of protein disorder, agreed that evidence for protein-protein interaction without, at a minimum, some kind of local folding had never been shown (Uversky and Dunker, 2013). Eventually, evidence for proteins retaining their intrinsic disorder while experiencing picomolar affinity protein-protein interaction was shown (Borgia et al., 2018). They also contest the idea that these are just proteins waiting for partners and that their flexibility is also key for performing some of their inherent functions, as in the case of flexible linkers or entropic chains. They also dismiss the PWP concept on the basis of some examples of proteins interacting with several partners and forming distinct conformations. In those cases, “waiting for a partner” seems to imply a much less dynamic structural landscape than the term “disorder”.

There is accumulating evidence pointing towards the increasing prevalence of IDRs in Eukaryotic proteomes defying some traditional concepts of biochemistry for substrate recognition and protein interaction. As is the case with any paradigm’s change, it takes a long time for the scientific community to acquire new concepts and practices and it often requires a generational turnover to reach a wide acceptance.

1.4.2 Bioinformatic prediction of IDRs

The increasing interest in unstructured regions of the proteins during the decades of 1990 and 2000 was accompanied by the aforementioned technical advances in X-Ray crystallography, NMR, and Cryo-EM. Those advances, however, were not sufficient to lay the foundations of the experimental determination of IDRs due to their inherent limitations. This, and the simultaneous infiltration of computer science into structural biology, acted as a catalyst for the development of computational algorithms capable of predicting protein disorder taking advantage of the already available information.

The first-ever reported bioinformatic method for the prediction of intrinsically disordered regions was reported in the 1997 Proceedings of International Conference on Neural Networks by Pedro Romero, who was working in Keith Dunker’s group (Romero et al., 1997). From this moment a plethora of different predictors, and their successive improved versions, were made available to the scientific community. A subset of the most widely used predictors was recently benchmarked at the Critical Assessment of protein Intrinsic Disorder prediction (CAID), a biennial community-wide double-blinded experiment for the determination of IDRs (Necci et al., 2021). Even in the light of these results, the selection of the most suitable method for the detection of disordered regions is not a trivial choice. It might be, in part, a subjective decision depending if one wants to prioritize long (>30 amino acids) or short regions, the sensitivity and specificity desired, and the

ease of use of the method. There exists, however, objective evidence for comparing the performance of different predictors over the same datasets (Liu et al., 2017; Nielsen and Mulder, 2019; Necci et al., 2021). How well these predictors will reproduce experimental results depends on the way they have been conceived. The main algorithm chosen for establishing a model, the selection of relevant predictor variables, and the gold standard used for optimizing the results are all critical factors that define the overall success of a particular method.

There are two main categories of disorder predicting algorithms. The first group corresponds to predictors that are based on physicochemical principles and the disorder propensity is calculated using mathematical equations that model those principles. As such, they are characterized by being extremely cost-effective in terms of computational times and highly interpretable. For example, IUPred assumes that a large number of inter-residue interactions are required for the stability of protein structures, and the method is developed as an estimation of the total pairwise interaction energy based on the amino acid composition (Dosztányi et al., 2005). Considering that structured regions have a higher energy computation than disordered regions, IUPred is able to generate a residue-wise scoring for disorder propensity.

The second group consists of predictors using machine-learning (also referred to as statistical learning or, more recently, artificial intelligence) algorithms. These methods use computational power to train a predictive model by feeding them with an array of variables, or features, from known positive and negative cases, called the training set. If the training set is accurate and sizable, and the features selected are the most relevant to the prediction desired, the model will "learn" to recognize the patterns that define disordered or structured regions. More computationally intensive than the first group, machine learning methods have a better performance in general, yet at the expense of interpretability. As a general rule, the more complex the algorithm used, the more difficult it will be to interpret which features are relevant and how they impact the prediction.

Within the category of machine-learning-based methods, there exists yet another subdivision defined by the information used for training the models. Multiple predictors rely strongly on sequence properties such as amino acid composition or sequence complexity, like PONDR (Linding et al., 2003) and VSL2 (Peng et al., 2006) predictors. Others are based on evolutionary information coming from sequence alignment data, and which is the case of DISpro (Cheng et al., 2005) and ESpritz (Walsh et al., 2012). There is a third subcategory that uses structural information, like secondary structure prediction of solvent accessibility. As an example of predictors in this group, the less popular predictors like DisPredict (Iqbal and Hoque, 2015) or SPINE-D (Zhang et al., 2012) can be listed.

Currently, modern methods or newer versions of older predictors take advantage of more than one of this type of information. This improves overall performance by taking into account relevant data coming from different and unrelated sources. Such is the case of SPOT-disorder (Hanson et al., 2017) that utilizes evolutionary and predicted structural information in combination with the physicochemical properties of the amino acid sequences for training a neural network with a complex architecture. The result is a method that is very well ranked in a number of benchmarks (Nielsen and Mulder, 2019; Necci et al., 2021; Katuwawala et al., 2020), and that is able to perform a good prediction of short and long IDRs simultaneously.

Lastly, there is another set of predictors that aim to reach a consensus prediction from other well-established methods in the community, rather than focusing on a selection of experimental information or the optimization of a statistical model. These are called meta-predictors, and the main difference between them is the strategy used to combine the input information. Some examples of this type of predictors are MobiDB-lite (Necci et al., 2017) and PONDR FIT (Xue et al., 2010).

Recent reviews have extensively described and benchmarked the functioning of the most widely used predictors (Meng et al., 2017; Liu et al., 2019). These publications are recommended reads if one wants to delve into the details of protein disorder prediction and its historical evolution.

A noteworthy aspect of IDR determination is the fact that the information of the primary protein sequence alone is enough for a satisfactory prediction. Indeed, the amino acid array seems to encode for the instructions to determine which parts of the protein will adopt a folded conformation and which parts will remain unstructured. Consistently, a specific amino acid compositional bias is observed in disorder regions, in which charged and polar amino acids are enriched while the proportion of hydrophobic and aromatic species is decreased, compared to the entire proteome (Uversky, 2019; Quaglia et al., 2021).

Overall, protein disorder prediction is consistent with experimental data, regardless of the method used, and different predictors render similar results for a vast majority of proteins or regions analyzed. However, dataset-level analysis has been shown to underperform when compared to protein-level analysis of a subset of the same dataset (Katuwawala et al., 2020). This implies the presence of two subsets of proteins: one, whose disorder is being properly predicted, and another one, where it is harder to predict. Since most of the users will use the predictor at the protein level, unaware of how difficult the prediction is for their target proteins, resources having easily accessible experimental and predicted data are highly desirable. Recently, different online databases containing diverse categories of information about protein disorder have been established. Altogether, these efforts constitute a powerful tool for the scientific community. The

centralization and documentation of data that, otherwise, is scattered or requires a certain level of bioinformatics skills to obtain, opened the possibility for researchers to inspect proteins being the subject of their studies in the context of intrinsic disorder. The first attempt to collect manually curated experimental data of IDRs can be attributed to DisProt (Vucetic et al, 2005). With the successive revisions and reformatting of its website, it has become the reference deposit of functional annotations for protein disorder (Quaglia et al., 2021). The DisProt data are obtained from the literature by a group of curators, a process that guarantees a high level of accuracy for this database. For this same motif, and totaling 2366 proteins across all the kingdoms of life, the coverage over entire proteomes is relatively narrow. Another less widely-known database, and which is based on the same principles, IDEAL, has claimed to have complementary information to that found on DisProt (Fukuchi et al., 2012). In order to address the need for a wider proteome coverage, two other databases, D^2P^2 (Oates et al., 2012) and MobiDB (Piovesan et al., 2021), have also compiled information from different disorder predictors. Available as user-friendly websites, both of these resources make present data allowing for more comprehensive proteome-wide analysis, while also giving access to results coming from disorder predictors that, otherwise, might be difficult to obtain for the regular user.

1.4.3 Biological functions of IDRs

Since the discovery of proteins as the main catalytic and structural forces in the cell, biologists have tried to understand the relevance of protein function by using natural selection as their magnifying glass. Proteins, or sections of them, that have any evolutionary advantages would be strictly conserved even for distant homologs. This basic principle applies almost universally when considering structured proteins, where the changes in the amino acid sequence is constrained by the function of the resulting tri-dimensional conformation. For disorder proteins this same principle cannot be applied generally. However, disordered regions are maintained throughout evolution, even when the primary amino acid sequence is not conserved to the same extent as in structured regions, suggesting some degree of function conservation.

The early proponents of disorder as a functional conformational state of proteins, promptly linked the advantageous characteristic features of IDRs to known biological functions. Moreover, they stated that they are a fundamental requirement for these functions to be carried out. In 1999, only two years after the publication of the first disorder predictor, Peter Wright and Jane Dyson, reviewed the data on a few well-reported case studies of unstructured domains being a necessary condition for key cellular mechanisms such as transcriptional regulation, translation of proteins, and membrane dynamics, among others (Wright and Dyson, 1999). Some years later, Dunker

and collaborators systematically reviewed the scientific literature to obtain 115 experimentally-determined disordered regions and for 98 of them, they could identify 28 different molecular functions, with the majority of these functions being related to macromolecule binding or post-translational modifications (Dunker et al., 2002). More recently, Peng and collaborators analyzed the disordered regions of 965 complete proteomes (Peng et al., 2014). For eukaryotes, where Gene Ontology (GO) (Ashburner et al., 2000) annotations are more comprehensive, they found that between 4 and 10 % of cellular processes, molecular functions, and cell components present in GO are significantly enriched in disorder. Similarly to Dunker, they observed that most of these annotations were related to interactions between proteins and nucleic acids. Interestingly, they also reported differential enrichment of diverse PTMs depending on the domain of life considered, although calculations were not corrected for compositional bias of IDRs.

A group of the most renowned scientists in the field emphasized the need for comprehensive classification and functional annotations of IDRs (van der Lee et al., 2014a). Eight different classification systems were proposed, one of which is based on their functions. It consists of 6 categories divided into 3 classes depending on the binding requirements. The first class comprises the IDPs requiring permanent binding to other molecules to perform their functions which contain 3 categories: effectors, assemblers, and scavengers. **Effectors** are proteins that fold upon binding and modifying the catalytic activity of their binding partners, such as p21 and p27 acting as stoichiometric inhibitors of CDKs (Galea et al., 2008; Yoon et al., 2012). The category **assemblers** refer to IDRs with interfaces sufficiently vast to act as a scaffold for several binding partners and facilitate the formation of complexes. Differently from structured complexes, these interactions rely on the dynamic properties conferred by the lack of a fixed conformation. The expression “fuzziness” has been proposed as a general term to extend the reach of protein disorder to protein complexes (Tompa and Fuxreiter, 2008). A fuzzy complex is, therefore, a model to explain the “continuous spectrum of structural states”, from a group of discrete rigid and static conformations to complete random protein interactions where the majority of the molecule persist disordered even in a bound state. The mediator complex is one of the classic examples selected to illustrate how large multi-protein complexes handle interactions through IDRs. In humans, 75% of the 19 mediator subunits present a region, longer than 30 residues, predicted to be disordered and 32% of them have a disordered region exceeding 100 amino acids in length (Tóth-Petróczy et al., 2008). Structural experiments not only show conformational changes of the mediator complex upon binding with RNA polymerase II, specific transcriptions factors, or the CDK8 kinase module but also provide evidence of the inherent conformational flexibility of their subunits (Fuxreiter et al., 2014; Koschubs et al., 2009). At last, in this category are the

scavengers whose function is to bind small molecules and act as a *"reservoir"* until the moment they are needed by the organism. Some examples of this behavior are Chromogranin A storing ATP and adrenaline (Yoo and Albanesi, 1990; Daniels and Wright, 1978), calcium-binding phosphoproteins (SCPPs) solubilizing calcium phosphate in biological fluids (Holt, 2013), or salivary glycoproteins that bind tannin in the digestive tract (Asquith et al., 1987; Boze et al., 2010).

The second functional class of IDRs comprises those that only require a transient binding with their partners. It comprises two different categories: proteins acting as chaperones for the folding of other macromolecules and display sites for post-translational modifications of proteins. A big proportion of protein and RNA **chaperones** are predicted to be disordered (van deer Lee et al., 2014; Ivangy-Nagy et al., 2005; Tompa and Csermely, 2004), and even if structural evidence showing the flexibility of certain chaperonins was reported in the mid-90s (Saibil, 2000), some mechanistic insights were still missing. More recently it was proposed that, given the multivalent nature of IDRs and their lack of structures, they can provide initial binding and recognition, flexibility to accommodate several protein-folding intermediates, and an increase of the solubility of the client-chaperone complex (Voth and Jackob, 2017). The acid-inducible intrinsically disordered chaperone Asr, from *E.coli,* constitutes a nice example of how charge can affect function. Replacing the positively-charged repeat sequence by a negatively-charged analogous sequence does not change its unbound conformation but it abolishes its effects on protein aggregation (Ren et al., 2022). This also seems to be the case for RNA folding mediated by the nucleocapsid domain of the hepatitis C virus core protein where molecular simulations show that "the chaperone acts as a flexible macromolecular counterion that screens the repulsive negative charges along the phosphate backbone and allows the nucleic acid to more frequently adopt compact conformations" (Holmstrom et al., 2019).

Intrinsically disordered regions as **display sites** for protein modification: this is one of the most extensively studied functional correlations in the field. Statistically significant enrichments in IDRs of various PTMs in disorder has been reported, such as phosphorylation, methylation, ubiquitination, S-palmitoylation (Narasumani and Harrison, 2018; Iakoucheva et al., 2004; Daily et al., 2005 ; Radivojac et al., 2010), among others (Gao and Xu, 2012). Moreover, proteins that are subject to multiple PTMs tend to have a greater number of long IDRs, and sites sharing multiple modifications have even a stronger preference for disorder (Pejaver et al., 2014). Interestingly, the enrichment of modifications in IDRs seems to be a feature only globally shared by eukaryotes, and not by bacteria or archaea (Peng et al., 2014), reflecting the step-like difference in proteomic disorder content between eukaryotes and prokaryotes (Xue et al., 2012).

Two special cases of IDRs that are heavily modified are worth mentioning: core histones N-terminal tails and Tubulin C-terminal tails (Peng et al., 2012; Roll-Mecak, 2015). Histone and Tubulin tails have a code in which a set of enzymes, the writers, modify them and there are proteins, the readers, that can decode those signals to promote different functions accordingly (Jenuwein and Allis, 2001; Janke et al., 2020). Interestingly, some of the PTMs present in tubulin tails are extremely rare and seem to be unique to these proteins like detyrosination, deglutamylation, tyrosination, polyglutamylation, and polyglycylation. Some understanding of the mechanism underlying post-translational modification of IDRs could be inferred from the co-crystallization of modifying enzymes with peptides derived from their substrates. In most cases, we can observe the accommodation of extended peptides over the groove where the catalytic site is located (Darling and Uversky, 2018). Considering enzymes to be folded and rigid, their substrates should account for the flexibility needed in order to obtain a system where a single protein can modify several targets.

The last class is for disorder functions that do not require binding with other molecules and corresponds with the category called **entropic chains**. For this special type of IDRs their function is directly related to the high mobility granted by the lack of a fixed conformation. Classic examples of this are sections of the proteins that work as flexible linkers between two globular domains, as in the case of the 70 Kb subunit of the replication protein A (RPA70). The globular domains contain the binding sites for single-stranded DNA and protein partners that regulate replication, and the unstructured linker seems to be required for those interactions even without showing strong sequence conservation (Daughdrill et al., 2007). Another well-known example is the microtubule-associated protein 2 (MAP), which consists of two domains: the positively charged C-terminal repeated domain binds directly to the negatively charged surface of microtubules, and the disordered N-terminal domain that protrudes to the exterior forming a repelling protective interface (Guharoy et al., 2013). More specialized types of entropic changes are the entropic springs, such as the “PEVK” repeated domain of Tintin that restore the length of muscle cells after stretching (Cheng et al., 2010), and entropic clocks of spacers that repel unsought molecules from “sticky surfaces”.

In contrast, proteins with IDRs that act as chaperones for the folding of macromolecules or as display sites for post-translational modifications of proteins (PTM) only require a transient binding with their partners. The third class, where no binding is required, is composed only of the so-called entropic chains. For this group, their function is due to disorder, as can be observed in linkers/spacers, entropic springs, or entropic clocks where the conformational shifts in IDRs

provide a mechanism for keeping the time such in the ball and “chain activation” of ion channels (Zandany et al., 2015).

There exists evidence that assigns a novel, yet indirect, function of disordered regions: IDRs seem to make good targets for immune system recognition. Some groups of researchers have been working on this idea for some years, and recently, immunogenic properties of IDRs were confirmed to the point that vaccines are being developed using unstructured regions of proteins (Olugbile et al., 2009; Ramamurthy et al., 2019; Ameri et al., 2021)

For all the functional classes of IDRs described previously, only entropic chains, whose function is directly controlled by their mobility and flexibility, all require some kind of binding with other molecules. Also, none of them seem to be exclusive to disorder, inasmuch as structured proteins are described to be able to perform the same functions. There is a vast amount of documentation about how folded proteins engage in interactions with their partners, but how does this work for disordered proteins? What is the evolutive advantage of those types of IDR-mediated interactions?

For some systems where interactions between partners induce the folding of a disordered region, such as HIF1α - TAZ1, p21 - Cyclin1 - CDK2 or Cadherin - β-catenin, the measured binding affinities tend to be very high. The enthalpic contribution associated with the interactions is then countered by the entropic loss due to the disorder-to-order transition, resulting in an overall reduction in the magnitude of the negative free energy. This produces low affinity - high specificity protein-protein interactions that allow for a single protein to recognize many others, and it is a “characteristic feature of signaling interactions” (Darling and Uversky, 2018; Dyson and Wright, 2005). This kind of one-to-many versatility in their interactions gives certain proteins the possibility of switching functions depending on the partner they bind, a mechanism called “moonlighting”, that has the advantage of drastically increasing the network complexity of an organism at a constant number of encoded proteins (Tompa et al., 2005).

Entropy changes mediated by disorder regions of proteins are also reported to modulate enzyme activities by modifying their affinity for specific binding partners. The disordered C-terminal tail of the human UDP-α-D-glucose-6-dehydrogenase increases the affinity of the enzyme for an allosteric inhibitor (Keul et al., 2018). Amazingly, this regulation only seems to depend on the presence -or not- and the length of the IDR rather than on its sequence.

An emerging concept is rapidly spreading to different fields in biology: phase separation. There exist phase-separated molecular condensates in the cell and observations indicate that IDRs might be their *“raison d’etre”* (Banani et al., 2017). This provides scientists with a model for explaining how unfolded proteins could form compact and delimited hubs carrying biological

functions. How these condensates are formed, regulated and which is their relation with the cell cycle is the subject of the next sections.

1.4.4 Biomolecular condensates

The uttermost extreme change in the evolution of life is, probably, the emergence of the eukaryotic cell. The vast majority of the hallmarks of eukaryotic life have been found also in prokaryotes as short-lived traits, and it is theorized that the energetic edge given by the acquisition of mitochondria is the reason why a cell could evolve to bear them all at once (Lane and Martin, 2010): a sophisticated genetic architecture that allows cells with identical genomes to have different phenotypes; an internal membrane system for physically separating linear chromosomes, delimiting other cellular organelles, and providing the means for an import-export system; an intricate and dynamic cytoskeleton that grants structural strength and a route map for internal transport. All these examples illustrate how substantial the increase in complexity from prokaryotes to eukaryotes is.

Nevertheless, all cells could be simplistically defined as containers for a bundle of biochemical reactions, regardless of their sophistication levels. And even in the simplest configuration, those reactions shall be spatiotemporally organized. There is evidence demonstrating the existence of different types of bacterial organelles at the nano-scale (Greening and Lithgow, 2020; Dobro et al., 2017). Although not fully ubiquitous across all bacterial phyla, the magnetosomes constitute an example of organelles that are surrounded by a lipidic membrane. These organelles contain nano-crystals of magnetic iron species that govern the magnetotactic bacteria motility depending on external magnetic fields (Uebe and Schuler, 2016). Another noteworthy example is the carboxysome, a rigid quasi-icosahedral proteinaceous shell whose function is to cluster enzymes needed for the fixation of carbon dioxide (Shively et al., 1973; Sutter et al., 2017).

A more widespread mechanism for the organization of cell biochemistry is the so-called “biomolecular condensates". This model, first introduced for eukaryotic cells (Banani et al., 2017; Shin and Brangwynne, 2017), describes that weak and unspecific multivalent interactions between molecules can drive their concentration into a liquid condensate within a liquid milieu by separation of phases. Analogous to what happens with oil in water, the de-mixing of the molecules of the condensates and the cytoplasm (or nucleoplasm) results in a droplet-like structure with distinct surface tension and viscosity. Some examples of bacterial bodies described as being mediated by liquid-liquid phase separation (LLPS) are the RNA degradosomes (bacterial ribonucleoprotein bodies or BR-bodies) (Al-Husini et al., 2018; Al-Husini et al., 2020), RNA polymerase clusters (Ladouceur et al., 2020), the FtsZ-SlmA condensates (Monterroso et al.,

2019), and others (Azaldegui et al, 2021). Interestingly, the less-acknowledged bacterial condensates might be the evolutionary origin of the better described eukaryotic membraneless organelles (MLOs). Muthunayake and collaborators reviewed the literature on the mRNA decay ribonucleoprotein bodies and they found sufficient experimental evidence for proposing that “The RNA decay machinery within the mitochondria and chloroplasts were derived from α-proteobacterial and cyanobacterial ancestors" (Muthunayake et al., 2020). Moreover, they highlighted similarities between BR-bodies and some eukaryotic cytoplasmic organelles, namely P-bodies and stress granules (SGs).

The first MLO ever described was the nucleolus, in the 1830s, as merely a cytology curiosity (Pederson, 2011). It was not until the 1960s were several scientific contributions linked nucleolus with their currently accepted function: the biogenesis of ribosomal RNA (rRNA). Rather than being a piece of uniform machinery for the synthesis of rRNA, nucleoli seem to present a complex substructure consisting of three layers working similarly to an assembly line (Harmon and Jülicher, 2022). The first steps of synthesis and processing of pre-rRNA occur at the fibrillar center (FC) and the surrounding dense fibrillar component (DFC), while the assembly of the pre-ribosomal ribonucleoproteins takes place most probably at the outer layers, namely the granular component (GC) and the DFC-GC interface (Krüger et al., 2007). Both the nucleolus as a whole and the de-mixed layers that form its components, have been shown to be regulated by phase separation (PS) (Brangwynne et al., 2011; Feric et al., 2016; Yao et al., 2019). A recent work shows that biophysical properties of the different sub-compartments of the nucleolus might be defined by the radial diffusion of rRNA from the center, where entangled nascent rRNA forms a highly-viscoelastic gel-like phase, to the outer layer, where the folded pre -ribonucleoproteins allow for the existence of a liquid-like phase (Riback et al., 2022). Nevertheless, more than 180 years had to pass from the first observation of nucleoli in order for scientists to acknowledge their liquid-like properties, and this only happened after this characteristic was reported for other membraneless organelles (Brangwynne et al., 2009).

The belated discovery of nucleolus being mediated by LLPS was probably due to technical advances in experimental setups, most notably those enhancing microscopy observations. These seem to be the same experimental limitations that stalled advances on the detailed description of MLOs in the smaller bacterial cells. Consequently, one could ask what are the experimental requirements for establishing that a given cellular body is mediated by PS. In order to define experiments for validating this hypothesis, the properties of a phase-separated MLO should be first defined. Primarily, liquid-like condensates within the cell should behave as de-mixed liquids at the macro scale. That is, the difference in surface tension should impose geometrical

constraints rendering the droplets approximately spherical. Other events, such as fission and fusion should be also noticeable, together with wetting and dripping effects. Whereas observing the roundness of phase-separated condensates does not require any sophisticated microscopy technique and immunofluorescence is often enough, the dynamics of MLOs are only evident through cell-live imaging of fluorescent-tagged proteins. One special case is the fluorescence recovery after photobleaching (or FRAP) technique, which can follow, in a time-resolved manner, how fast and complete the recovery of the fluorescence region is after it has been bleached by laser exposure. For a liquid-liquid interface, the rates of diffusion are high and the expected times of full-recovery is in the seconds or minutes scale (Brangwynne et al 2009; Feric et al., 2016; Handwerger et al., 2003). Some authors called for caution when FRAP is used as a "gold standard" *in vivo* evidence of LLPS (McSwiggen et al., 2019). They stated that FRAP results might vary greatly depending on very sensitive variables, like the microscopes or detectors being used. This was indeed observed in the literature where differences in the recovery times of orders of magnitudes were reported for the same protein. Moreover, they affirmed that deeming the fast recovery times as sufficient evidence for LLPS is a gross oversimplification and that, indeed, these results are not incompatible with other models of macromolecular interaction.

Another characteristic feature of liquid phase-separated condensates is that their formation and dissolution is concentration dependent. A phase diagram can be envisioned, where, for a given set of conditions, a concentration of the components saturates the system and two phases start to coexist (show image). For *in cellulo* observations of the concentration dependency, the aforementioned microscopy techniques can be employed with a system that allows for the regulation of protein expression (Pak et al., 2016). It was shown recently, however, that the complexity of the cell might render the results hard to interpret. Riback and collaborators reported that increasing concentrations of Nucleophosmin (NPM1), a key driver of nucleolar phase separation, indeed resulted in bigger nucleoli (Riback et al., 2020). Surprisingly, the concentration of NPM1 in the nucleoplasm also increased substantially breaking the principle of equilibrium between the two phases, i.e., the saturation concentration (C_{sat}) was not a fixed value. They concluded that multicomponent MLOs governed by heterotypic interactions cannot be explained by the simplistic model of a single saturation concentration. To circumvent these limitations, they set up the optoDroplet system, in which the *Arabidopsis* Cry2 tag is fused with the protein of interest and upon blue light exposure the tag dimerizes promoting phase separation (Shin et al., 2017). Cry2-fused DDX4, a system reported to form phase-separated exogenous homotypic condensates (Shin et al., 2017; Nott et al., 2015), showed intracellular droplet formation at constant C_{sat}. For testing their hypothesis, they calculated, at different concentrations of NPM1,

the free energy of transfer nucleoli as $\Delta Gtr = -RTlnK$, where K is the ratio between the concentration in the condensates and the concentration of the nucleoplasm. As expected, the free energy increased as the total concentration added increased, indicating a non-equilibrium state where Csat is not constant. To further explore these results, they simulate nucleoli formation in vitro by mixing recombinant NPM1, SURF6 (a nucleolar multivalent protein), and rRNA, and they obtained similar NPM1 concentration-dependent increase of ΔGtr values to those observed in the cell. When only NPM1 was present in the test tube, ΔGtr were constant reflecting the fixed Csat value. Clearly, in vitro phase separation experiments with recombinant proteins constitute another powerful tool used in the field. Despite not closely representing physiological conditions, these experimental setups allow scientists to investigate LLPS while controlling other variables such as temperature, pH, salt concentration, presence of crowding agents (such as Dextran) or dissolving agents (1-6, Hexanediol).

Apart from the already mentioned MLOs of eukaryotic cells, such as nucleoli, stress granules, or P-bodies, there are several other well-established phase-separated condensates worth mentioning, such as Cajal bodies, nuclear speckles, PML bodies, 53BP1 bodies, superenhancers-mediator clusters and nuclear pore complexes. Other subcellular systems like histone locus bodies, chromatin, the perichromosomal layer, centrosomes, different DNA damage foci, among many others, are also recognized as being phase-separated condensates (Banani et al., 2017; Uversky, 2017; Zaslavsky and Uversky, 2018; Gomes and Shorter, 2019).

Nuclear pore complexes (NPC) are, as their name clearly indicates, molecular machines that form pores in the nuclear envelope (NE) (Watson, 1954). They are composed of multiple copies of a series of more than 30 different proteins called Nucleoporins -or Nups for short- (Rout et al., 2000; Cronshaw et al., 2002) that form a channel-like structure anchored to the NE. Rather than having a specific catalytic activity, NPCs regulate the diffusion through the center of the channel by the material properties of a mesh arrangement of Nups with FG-rich domains (FG Nups) (Schmidt and Görlich, 2016). The highly conserved Nup98 is often signaled as the main FG Nup in charge of the permeability properties of NPC in vertebrates (Hülsmann et al., 2012), and it has been shown to form phase-separated hydrogels by its own (Labokha et al., 2013). More recently, researchers used a microfluidics approach to demonstrate that condensates of yeast FG Nup49, a putative analog of Nup98, rather present liquid-like properties (Celetti et al., 2020).

Another well characterized example are Cajal bodies, “small round structures inside the nucleus” first described in 1903 by the notorious researcher Santiago Ramon y Cajal (Gall, 2000). This discovery was awarded the 1906 Nobel prize, but it was not until 1999, that these organelles were named after the Spanish neurobiologist. The main function of the Cajal bodies described is that

they serve as spliceosomal and telomerase ribonucleoprotein (RNP) maturation sites (Sawyer et al., 2016) and, recently, they have been linked to the genome structural organization (Wang et al., 2016). The recruitment and sequestration of their components to increase their concentration while maintaining a high exchange rate with the nucleoplasm are all properties conferred by its LLPS nature, and they are key for carrying out their functions. Following GFP-tagged Coilin, an essential protein for the formation of Cajal bodies and their principal molecular marker, showed that these nuclear bodies undergo events of fission and fusion *in cellulo* (Platani et al., 2000), in accordance with what is expected from phase-separated MLOs.

As it happens with the rest of the DNA damage response (DDR) systems, 53BP1 bodies are cell cycle-regulated. At the site of a double-strand breaks (DSB), the tumor suppressor protein 53BP1 gets recruited via a γ-H2AX-mediated mechanism and forms observable foci (Anderson et al., 2001; Fernandez-Capetillo et al., 2002; Wang et al., 2002). It is proposed that 53BP1 blocks DNA end resection promoting repair systems based on end-joining, and impeding BRCA1 homologous recombination repair (Chapman et al., 2012; Bothmer et al., 2010; Bothmer et al., 2011). Recently, researchers have averted that 53BP1 foci, that form during G1 and early S-phase upon exposure to exogenous clastogens, are different from the 53BP1 bodies present (Lukas et al., 2011). These bodies seem to be formed in mid-to-late S-phase in response to damage induced by inherited unreplicated DNA in the previous cell cycle, and they coordinate repair mechanisms mediated by RAD52 (Spies et al., 2019). Kilic and collaborators showed that 53BP1 condensates, required for the downstream activation of the p53 response, present all the hallmarks of a phase-separated MLO: liquid-like behavior with fusion and fission events, optodroplet formation of Cry2-53BP1 fusion upon blue light, *in vitro* droplet formation, and fast recovery times in FRAP experiments. Control of 53BP1 multimerization by the scaffolding protein AHNAK has been reported as an important mechanism for avoiding an enhanced phase separation that will lead to an exacerbated p53 response (Ghodke et al., 2021). A recent work claimed that, in fact, 53BP1 phase separation is not directly related to DNA repair and might be controlling genome stability by regulating heterochromatin in a mutual dependency with HP1α (Zhang et al., 2022).

Heterochromatin was also recently proposed to be a liquid-like nuclear phase mediated by HP1 (Strom et al., 2017; Larson et al., 2017; Sanulli et al., 2019), although other data suggest that the chromatin itself behaves like a solid but can act as a scaffold that support formation of liquid-like protein assemblies (Erdel et al., 2020; Strickfaden et al., 2020). HP1 proteins appear not to possess the intrinsic capacity to undergo LLPS in cells (Erdel et al., 2020). However, HP1 colocalises with the methyl-CpG binding protein, MeCP2, which was found as a novel Ki-67-interacting protein (Sobecki et al., 2016), and which can undergo LLPS in cells (Li et al., 2020).

Ki-67 is historically recognized as a proliferation marker that, surprisingly, is dispensable for cell proliferation but essential for heterochromatin organization (Sobecki et al., 2016). Cells lacking Ki-67 have reduced heterochromatin compaction and disrupted long-range interactions between pericentromeric and perinucleolar heterochromatin, resulting in nonspecific wide-ranging transcriptional effects (Sobecki et al., 2016; Mrouj et al., 2021).

Another interesting type of nuclear body is the PML body. These condensates, however, appear to be less specific and they have been linked to several different functions, which is also reflected by their composition. They concentrate a great variety of different proteins with no more apparent functional relationship between them than being sumoylated (Bernardi and Pandolfi, 2007; Lallemand-Breitenbach and de Thé, 2010). PML bodies are stress-sensitive bodies and upon different stress conditions, they show distinct phenotypes as, for instance, an increase in size and number upon interferon exposure, or changes in localization after exposure to UVc radiation (Stadler et al., 1995; Seker et al., 2003; Bøe et al., 2006). These differences, together with the evidence showing that there exist different isotopes of PML that interact with different partners, led the scientist to consider the possibility that rather than being a single type of condensates, “PML body” is a broad definition for PML-mediated MLOs that respond to stress (Lallemand-Breitenbach and de Thé, 2010). In this way, different types of cellular stress could recruit different partners to PML bodies to activate specific pathways (Lallemand-Breitenbach and de Thé, 2018). If oxidative stress is identified, PML bodies will regulate p53 PTMs that, in turn, will promote the expression of specific genes that mediate the response to the reactive oxygen species (ROS) (Niwa-Kawakita et al., 2017). They can also control the epigenetic status of the cell by regulating the histone pool and chromatin assembly via the DAXX/ATRX histone chaperone complex (Delbarre et al., 2012; Delbarre et al., 2017). It has also been reported that PML nuclear bodies might be involved in the alternative lengthening of telomeres (ALT) pathway in cancer cells lacking telomerase. Even if the presence of nucleic acids within the bodies has not been clearly demonstrated, they associate with telomeres, sheltering proteins and DNA repair proteins to establish ALT-associated PML bodies (APB) (Chung et al., 2012; Lallemand-Breitenbach and de Thé, 2018). There is increasing evidence showing the LLPS characteristics of PML bodies such as roundness, high recovery rates in FRAP experiments, fusion, and PML concentration-dependent formation (Corpet et al., 2020; Hoischen et al., 2018; Lang et al., 2010; Dellaire et al., 2006; Hancock 2004). Ultimately, definitive evidence with the optodroplet system or *in vitro* phase separation of PML is lacking. Interestingly, that is not the case for recombinant repeats of SUMO and the SUMO interaction motif (SIM), whose ability to *in vitro* phase separate has been shown (Banani et al., 2016). Furthermore, it has recently been proposed that, in fact, it is the SUMO-SIM

interaction which drives the phase separation of APB and, therefore, controls the ALT activity of telomere-associated PML bodies (Min et al., 2019; Zhang et al., 2020).

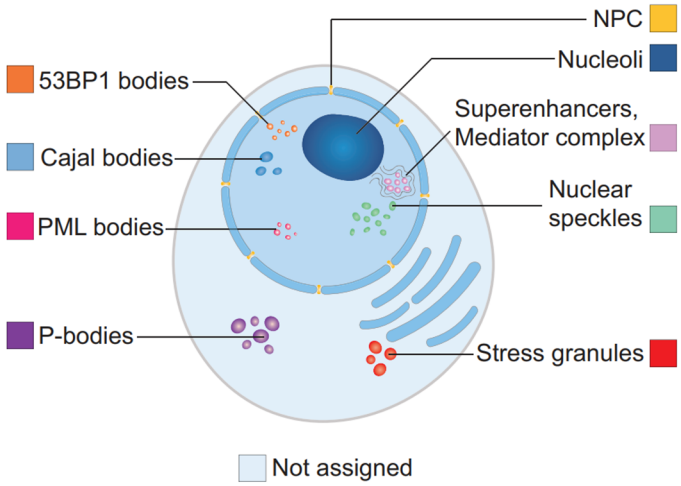


Figure 1.5 | Membraneless organelles in the cell. *Author: Juan Manuel Valverde.*

There exists a broad consensus regarding which are the molecular features driving phase separation and the mechanisms underlying this process. In 2017, two seminal reviews from the laboratories of Michael Rosen and Clifford Brangwynnehave have addressed the principles of phase separation (Banani et al., 2017; Shin and Brangwynne, 2017). In both cases, they concluded that the most important requirement for biomolecules to phase separate is the presence of multivalency, i.e., molecules bearing multiple discrete interaction domains. Both, based in “classic concepts of polymer science” (Rubinstein and Colby, 2003), which state that multivalent molecules tend to inherently form oligomers - or polymers - that will decrease the entropy of the system decreasing the solubility of those molecules and causing them to form a distinct phase. The properties of this phase will largely depend on its composition, concentration and the interactions by which the phase separation of these condensates is driven. For instance, several proteins that phase separate seem to present low complexity (LC) regions that participate in weak low-affinity interactions driving the formation of condensates. On the other hand, the LC domains of the FUS are able to stack and to form amyloid-like fibrils mediated by stronger cross-β structures upon an increase of their concentration (Murray et al., 2017; Kato et al., 2012). As a consequence, FUS has been shown to present a liquid-like behavior *in vivo*, like fast FRAP recovery rates and the formation of optoDroplets (Patel et al. 2015; Shin et al., 2017). However, increasing concentrations of the proteins make the condensates lose their liquid-like properties and start presenting hydrogel-like properties or even to start behaving like solids. Curiously, different from what is observed on pathogenic fibrils, if the fibrils are not aged, they are reversible

and a decrease in their concentration or changes in temperature will disassemble these structures (Kato et al., 2012). The discrepancies observed are most probably due to the nature of their side chains interactions: while pathogenic fibrils are stabilized by a core of hydrophobic interactions, FUS-LC side chains are not enriched in hydrophobic residues (Kato et al., 2012; Hughes et al., 2018). Indeed, the compositional signature of the low-complexity domain of FUS is not comparable to that in proteins involved in the formation of pathogenic fibrils, but rather similar to what is observed for intrinsically disordered regions.

That brings us to the second molecular feature greatly related with phase separation: MLOs and phase separation are systematically linked with protein disorder. Most of these associations are commonly observational and anecdotal and there is not a great body of evidence based on systematic and quantitative analysis. In 2017, Darling and collaborators fetched 4796 human proteins with Gene Ontology terms (Ashburner et al., 2000) placing them as components of MLOs. They perform disorder prediction using 3 methods from the PONDR family of predictors and they found, in all cases, that the proteome in MLOs is substantially more enriched in disorder than the entire human proteome (Darling et al., 2018). Thus, FUS liquid-gel-solid phase transitions may not be unique and the general dynamics of IDR-mediated phase separation might be considerably more complex than the two-way phase diagram. A conceptual free energy landscape can be envisioned where different IDR dynamics could lead to different types of ensembles, ranging from free solved proteins all the way through amyloid-like aggregates or glassy-like solids, which retain the unstructured nature of disorder proteins but their mobility is almost completely lost (Tsang et al., 2020).

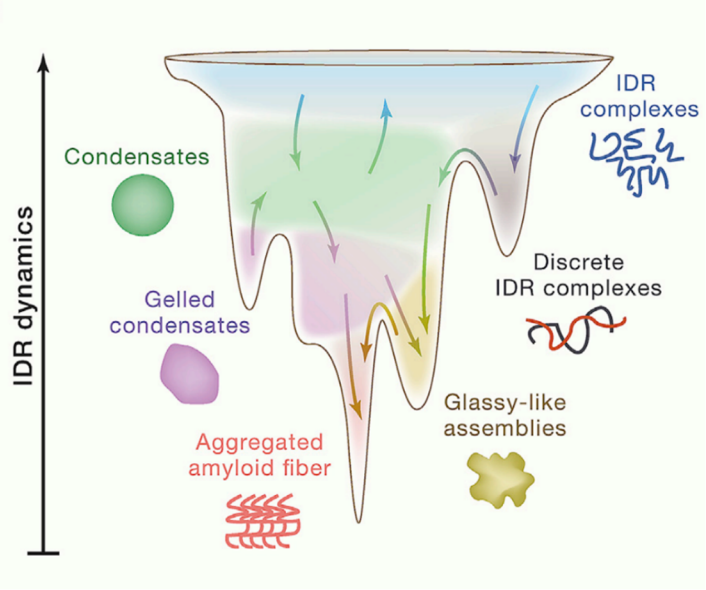


Figure 1. | **Free energy landscape of the dynamics of disordered proteins.** From *Tsang et al., 2020*

Protein disorder certainly conforms to the intuitive conceptual mechanism for phase separation, in which the lack of structure of a polypeptide chain may render it highly mobile, and intra- or inter-molecular interactions are more easily occurring in this dynamic context. Recently, researchers have proposed a model based on the “stickers and spacers” framework, derived from associative polymers biophysics (Choi et al., 2020). Stickers are regions that can interact one with another while spacers act as flexible chains whose interactions neutralize one another. This rather simple model is able to generalize important concepts, such as the saturation concentration, for both homotypic and heterotypic systems. Another interesting notion emerging from this framework, is the fact that it does not assume any restrictions for the nature of the stickers and they could be structured interactions domains, or simply disordered small linear interaction motifs (SLIMs). In this last case the concepts of “intrinsically disordered regions” and “multivalent proteins” might become analogous, where a given stretch of disordered residues presents multiple regions of sticky linear interaction motifs.

The existence of *in vivo* phase separation and its relevance for organizing the cell biochemistry is, of course, not free of controversy. More conservative biochemists and structural biologists have been reluctant to these new concepts, especially given how predominant it is becoming for explaining a wide variety of different molecular mechanisms. Notable cases are Andrea Musacchio, who argues that membraneless cellular compartments are driven by highly specific interactions instead of by low affinity unspecific ones (Musacchio 2022), or McSwiggen and collaborators who advise against the use of high rates of recovery in FRAP experiments as a gold standard for phase separation (McSwiggen et al., 2020). In both cases, these authors emphasize that simplistically categorizing interactions as “phase separation” without a strict experimental definition does not add any value and it is irrelevant from a biophysics standpoint. They called for a redefinition of the experimental framework and for more stringency and consistency in the nomenclature. Nevertheless, the burden of evidence they propose for claims of phase separation-mediated condensates is quite excessive and such severity is surely not universally applied to all biology-related disciplines.

Regardless of the biophysical minutiae of molecular interactions within biomolecular condensates, the basic evidence for their existence, composition, and their formation and dissolution are rather compelling. But generalized models applicable to all -or most- of the membraneless organelles tend to fail when considering all the experimental data. This might well mean that fairly different condensates with substantial variances in composition, physical behavior and biological function are not generalizable. Is there then a reason why the cell might benefit from a common control system for all MLOs?

1.4.5 CDK-mediated regulation of membraneless organelles

In general, the different identified membraneless compartments in the cell vary greatly from one another. As we look into details, we can find a wide divergence between them. Some examples of properties that can be unique in each case are composition, localization and biological functions. Nonetheless, some other properties are general such as the already mentioned liquid-like behavior or their enrichment in IDRs and multivalency.

Not surprisingly, multiple MLOs are reported to be composed of proteins that are post-translational modified. This observation might seem trivial considering the predominance of IDRs in proteins composing biomolecular condensates, and the increased tendency of PTMs to occur in disorder regions (Pejaver et al., 2014). However, a functional aspect lies therein: post-translational modifications modulate phase separation (Owen and Shewmaker, 2019; Bah and Forman-Kay, 2016; Hofweber and Dormann, 2019). How the covalent attachment of different chemical species affects proteinaceous condensates differ vastly from one case to the other depending on the characteristics of the modifications, the targets and the condensates. The same PTM can act as a driver for phase transitions or as a dissolving agent for different proteins in different contexts (Keiten-Schmitz et al., 2021; cite). Two generic mechanisms have been recently proposed to model how post-translational modifications of proteins can regulate MLOs, using phosphorylation as an example (Söding et al., 2019). In the Enrichment-Inhibition model the modifying enzyme is concentrated within the limits of the condensate and the modification will generate repulsive interactions causing a diffusive flux of the modified proteins towards the condensate. In this way the modification helps to maintain constant the C_{sat} of unmodified proteins inside the condensate. In contrast, the Localization-Induction model will modify the proteins causing stronger interaction forces between them and the rest of the components of the condensate. This will require an initial fixed localization of the modifying enzyme that will generate a local increase in the concentration of the modified protein, forming a nucleation site. The mesh of interacting modified protein will diffuse out from the reaction site, increasing the size of the condensate. In both models, a counteracting activity for the modification should happen in the exterior phase, increasing the concentration of unmodified protein outside the condensate to cause diffusive influx.

Seemingly neutral modifications as arginine methylation, for example, can alter critical inter- and intra-molecular interactions by diminishing the hydrogen bond potential or impeding π-cation interactions with neighboring aromatic residues (Owen and Shewmaker, 2019). *In vitro* phase separated condensates of the human DEAD-box helicase DDX4 (a key component of the nuage

organelle, a.k.a. chromatoid body or perinuclear P granules) is greatly destabilized when its RGG motif gets dimethylated (Nott et al., 2015).

More complex modifications such as the covalent binding of Ubiquitin or the Small Ubiquitin-like Modifier (SUMO) have also been associated with phase transition processes. Ubiquitination and SUMOylation consist of the attachment of small proteins of around 76 residues and 100 residues, respectively, presenting a discrete level of identity between them (~20%) but a similar fold (Gill, 2004). Both modifications are also related in terms of the molecular machineries that regulate them and functions associated with them, that in most cases are mediated by some kind of stress response (Flick and Kaiser, 2012; Maxwell et al., 2021; Enserink, 2015), and it is thus not surprising that key proteins of stress-driven MLOs present these modifications (Dao et al., 2018; Keiten-Schmitz et al., 2021). In the case of PML nuclear bodies, for instance, the PML protein presents both SUMOylation and SUMO-interacting motifs and it is the SUMO-SIM-mediated oligomerization of PML that act as the nucleation step for the formation of the bodies (Müller et al., 1998; Shen et al., 2006). As mentioned in the previous section, the SUMO-SIM interaction is not only able to promote phase separation, but also modulates the function of PML bodies (Banani et al., 2016; Min et al., 2019; Zhang et al., 2020). Interestingly, SUMO-SIM interaction is heavily regulated by another post-translational modification: phosphorylation. Both SUMO enzymes and targets are substrates for different kinases that allow for both positive and negative regulation of the pathway (Gareau and Lima, 2010).

Indeed, protein phosphorylation is not just another example of PTM enriched in MLOs but a seminal one. Being as simple as 4 added atoms, the phosphoryl group also brings notable changes in the charge and steric hindrance to the proteins modified. It is also the most abundant modification in the cell, with an estimation of at least 75 percent of the proteome being phosphorylated (Sharma et al., 2014). Just as important, protein phosphorylation is the chemical modification selected by evolution to modulate the cell cycle via the kinase activity of CDKs. And it is this particular fact that highlights the importance of yet another common feature among all MLOs that have not been mentioned: they are cell cycle regulated. All the previously described biocondensates are formed at some point during interphase, depending on their biological functions, and get dissolved concurrently during mitosis. By the time chromosomes start to segregate, most phase-separated compartments of biomolecules are no longer observed (Misteli, 2001; Hernandez-Verdun, 2011; Lång et al., 2019), although some mitotic protein arrangements, such as centromeres, kinetochores and the perichromosomal layer, have been speculated as to be phase separated (Liu et al., 2020; Booth and Earnshaw, 2017)

A reductionist view could propose the cell cycle merely as a cycle of cellular reorganizations. In each cycle DNA gets copied and packed, internal membranes get disassembled (including the nuclear envelope), sister chromatids are mechanically pulled to opposite sides of the cell and, lastly, the cell membrane gets contracted until two different and complete cells are formed. Given that all these processes are regulated by CDK phosphorylation, it is reasonable to think that the cell cycle-dependent reorganization of MLOs should be governed by CDK activity as well. For instance, Cajal bodies are regulated throughout the cell cycle, they get assembled during early-to-mid G1 and get dissolved in mitosis. Coilin levels, however, remain constant throughout the different cell cycle phases (Andrade et al., 1993), but it is its phosphorylation which is affected, peaking precisely in mitosis (Carmo-Fonseca et al., 1993). The hypothesis of cell cycle phosphorylation-dependent regulation of CB is supported also by the fact that human Coilin has been shown to be phosphorylated by CDK2 *in vitro* and that the CDK2-CyclinE complex is located to Cajal bodies (Liu et al., 2000). A more recent study shows that these differential phosphorylation states of Coilin regulate the formation, localization and mobility of Cajal bodies (Hearst et al., 2009). Furthermore, human cells presenting hyper-phosphorylated versions of Coilin lacked observable CBs (Hearst et al., 2009).

A critical step for mitotic entry is the nuclear envelope breakdown, that is accompanied by disassembly of nuclear pore complexes. Multiple Nups are phosphorylated during mitosis and CDK1 kinase activity have been shown to be essential for dismantling NPCs (Onischenko et al., 2005; Mühlhäusser and Kutay, 2007; Linder et al., 2017; Laurell et al., 2011). Observations in different organisms indicate that Nup98, the FG repeat-containing protein that drives the formation of a phase-separated mesh that controls the permeability of the pores, is the first to be leave the NPC (Dultz et al., 2008; Hase and Cordes, 2003; Lenart et al., 2003). Other cell cycle-related kinases, like NEK and PLK, might contribute to Nup98 hyperphosphorylation driving nuclear pore disassembly (Laurell et al., 2011).

Another interesting case is DYRK3, a member of the dual-specificity DYRK family that shares the CMGC group of kinases with CDK, MAPK, Cdc2-like kinases (CLKs) and members of the RCK family (Manning et al., 2002). A screen of chemical kinase inhibitors that would block dissolution of SGs rendered DYRK3 inhibitor as a primary hit (Wippich et al., 2013). The disordered N-terminal region of DYRK3 was shown responsible for its localization to stress granules, under stress conditions, sequestering the kinase and impeding its upstream action on the mTORC1 pathway. The same inhibitor was used in a more recent work from the same laboratory, in which they showed that DYRK3 acts as “dissolvase” (Rai et al., 2018). When the kinase was inhibited, markers of splicing speckles, stress granules and pericentriolar matrix appeared to form aberrant

co-condensates, in which also poly-A RNAs could be found. Authors also claimed that the formation of those atypical granules from the amalgamation of components of different MLOs delayed mitotic progression, which fit perfectly into the mitotic MLO reorganization hypothesis. However, some caution is required as nonspecific inhibition of other kinases, including CDK1, cannot be ruled out, and CDK1 was not included in the DYRK3 inhibitor specificity screen.

There are two important common observations in all examples described above. First these MLOs are cell cycle-regulated and, second, a being the key driver for the formation of the aforesaid MLO is a CDK target, or a target for another cell cycle-related kinase such as DYRK3. Moreover, these characteristics are common for other MLOs such as Nucleoli (Sirri et al., 2002; Dranovsky et al., 2001), the histone locus bodies (Hur et al., 2020) or 53BP1 bodies (Zhang et al., 2022; Kilic et al.,2019). A notable exception to this, seemingly, general behavior are PML nuclear bodies which do not get dissolved during mitosis (Everett et al., 1999), even if PML protein is phosphorylated by CDKs. It has been speculated that this unique property might be mediated by a liquid-to-solid phase transition (Lång et al., 2019). Regardless of the phase transition involved, there is plenty of evidence for establishing a clear correlation between CDK phosphorylation and the regulation of MLOs throughout the cell cycle. A direct causation however, is not yet conclusively proven and, rather than from a single extraordinary experiment, the definitive answer will be probably delivered as accumulation of evidence emerging from multidisciplinary approaches.

The fact that the discovery of new CDK targets vastly outweighs the description of their functions, may indicate that only a few specific CDK phosphorylations are functional while the vast majority of them occur in a nonspecific manner without any functional consequence. This idea is supported by a study from Landry and collaborators in which they found that phosphorylations of known functions have more sequence conservations than those where a function has not been described, “indicating that the apparent rapid evolution of phosphoproteomes results from a large fraction of phosphosites being non-functional” (Landry et al., 2009). However, an important defect of this hypothesis resides in the fact that authors did not consider that, even if protein phosphorylation is reversible, phosphatases that counteract kinase activity do not restore the cellular ATP pool. That would imply that a great amount of chemical energy is spent in trivial biochemical reactions. Energetic efficiency might be, arguably, one of the stronger evolutionary constraints and it would probably outweigh any of the advantages that a rapidly-evolving phosphoproteome could offer.

One way to conciliate the evidence indicating that most phosphorylation sites are located into fast-evolving unstructured regions of proteins with their functional character is to consider them as a general regulatory system rather than punctual individual regulators. In the context of CDK

regulation of mitotic reorganization of MLOs, CDK can be thought of as the global regulator of condensates by non-site-specific hyperphosphorylation of their key drivers. In this case, conservation of specific phosphosites does not grant any advantage if the evolutionary rate at which phosphorylatable residues are lost is compensated by a comparable rate of spawning new sites in other positions. Even the net loss of a site of phosphorylation -or all of them- for a given protein might not have great evolutionary consequences over the overall fitness, since phosphorylations might still happen in other proteins of the same complex with an equivalent outcome.

2. A genetic tool for commanding the cell cycle

Early work in the fission yeast *S. pombe*, which has a small genetic complement of CDKs and cyclins, suggested that the cell cycle can be driven by oscillations in the activity of a single CDK1-cyclin complex alone (Fisher and Nurse, 1996). In this “quantitative model”, the different combinations of CDKs and cyclins are not a requirement for cell proliferation and, therefore, neither for the cell cycle progression. Instead, oscillations of total CDK activity drive sequential entry into S phase and mitosis. Indeed, this principle was proven in a fission yeast strain where all CDKs and cyclins were deleted except for the CDK1 analog Cdc2 fused with the Cdc13 cyclin. These cells could operate with a single CDK-cyclin fusion complex where CDK was, additionally, converted to an analog-sensitive (as)-CDK (Shokat) version (Bishop et al., 2000), rendering it sensitive to ATP analogs that do not affect wild-type kinases (Bishop et al., 2000). A mutation in a phenylalanine in the ATP-binding pocket of CDK enlarges the groove and bulky analogs of ATP can be engineered to act as specific inhibitors of (as)-CDKs. Remarkably, by using these specific inhibitors, the activity of the complex was experimentally modulated and the different phases of the cell cycle could be prompted by different activity levels of the same CDK-Cyclin complex (Coudreuse and Nurse, 2010).

Similar principles may apply to mammals. Multiple simultaneous CDK gene knockouts in mice demonstrated that CDK1 is the only CDK that is essential for the core cell cycle (Santamaría et al., 2007), while mammalian cyclins also show a high degree of functional redundancy (Malumbres et al., 2004; Kozar et al., 2004). We, therefore, propose that the proliferation of mammalian cells, like fission yeast, may fundamentally rely on quantitative oscillations of overall CDK activity, while the additional complexity of the CDK network might increase mechanistic coupling between the cell cycle and developmental programs. For example, deletion of type-D cyclins in mice does not impair the progression of the cell cycle, although deficient hematopoiesis results in embryonic lethality (Malumbres et al., 2004; Kozar et al., 2004). A similar outcome was observed for mice embryos lacking CDK2, CDK3, CDK4 and CDK6, which can reach midgestation where developmental defects render them not viable (Santamaría et al., 2007). This is in stark contrast to mice knockout of CDK1 that fail to progress through early developmental stages. Moreover, specifically modulating CDK1 activity levels in mouse embryonic stem cells by genetic replacement of CDK1 with as-CDK1 (F80G) showed that CDKs couple pluripotency to cell cycle regulation (Michowski et al., 2020). Nevertheless, this allele is hypomorphic, and viable

homozygous mutant mice could not be produced, suggesting that the precise regulation of CDK1 activity is critical for development.

It has been reported that the Cyclin fusion to CDK can help to stabilize Shokat alleles that, otherwise, present an impaired association with cyclins due to conformational defects (Merrick et al., 2011). As a result, by generating the CDK-Cyclin fusion complex in mammals the issue of a hypomorphic allele might be circumvented. Besides, there is one additional theoretical implication emerging from this approach if it is successful: the core cell cycle of higher eukaryotes can be driven by one CDK-Cyclin complex.

To test this hypothesis, we will employ two different strategies and they will be presented in this section. The first, in human cell lines, consists of CRISPR-Cas9-mediated homologous recombination for the insertion of an artificial flipping intron, within an exon of the endogenous human CDK1 gene. Alleles where recombination has occurred ($CDK1^{rec}$), will present normal expression of CDK1, while non-homologous end-joining (NHEJ) repair pathway will probably generate a null allele (CDK^{-}). Recombinant clones will present antibiotic resistance and the other allele should be genetically screened and only those presenting $CDK1^{rec/rec}$ or $CDK1^{rec/-}$ genotypes will be selected. Upon Cre recombinase expression, the $CDK1^{rec}$ alleles will switch to express the (as)-CDK1-Cyclin fusion protein, whose activity can be externally modulated using specific inhibitors. Diverse cell models with different genetic backgrounds will be used for answering specific questions.

The second strategy is to develop a similar genetic switch in mice. Here, all the added complexity of experimentation with animals is partially compensated by the simplification of the genetic system. In this case, targeting exons of the endogenous CDK1 gene is not a requirement and the flipping cassette can be inserted within an intron. Increasing the recombination efficiency by CRISPR-Cas9-mediated double-strand breaks (DSB) is not always an option, and the best approach might be to microinject zygotes with the donor DNA. The recombinant alleles will switch to the (as)-CDK1-Cyclin fusion version of CDK1 upon action of Cre recombinase, but the other allele will most probably not have the switch. The advantage of the mouse model lies in the possibility of *a posteriori* crossing the heterozygous individuals to obtain homozygous mice.

As mentioned before, a mammalian cell cycle running on a tunable single CDK-Cyclin complex is not only a result in itself, but will also allow answering key questions regarding the so-called quantitative model. Namely, a dose-dependent control of CDK activity could be greatly useful to observe partial phenotypes compared with more extreme approaches. Complete disruption of CDK1 activity blocks the progression of the cell cycle and, therefore, cell proliferation. These extreme phenotypes highlight the importance of this kinase, but they don't reveal any information

about the mechanism. External modulation of CDK1 activity will be crucial to detect the precise CDK levels that trigger which S phase or M phase. Ultimately, I aim to replicate the results obtained in yeast by Coudreuse and Nurse showing that the cell cycle could be completely manipulated by the activity of a single kinase, to such a degree that a lower CDK activity after G2 phase triggered re-replication instead of mitosis (Coudreuse and Nurse, 2010). If the same holds true for mammalian cells, this will allow us to synchronize a cell culture in a less artifactual manner than with the currently available methods. Using precise modulation of the CDK activity combined with SILAC experiments, in different cell lines, will allow me to characterize specific S phase and M phase CDK targets required for the core cell cycle by phosphoproteomics. Analysis of CDK targets in different cellular models will serve to discriminate between direct cell cycle functions of CDKs and tissue/cell-specific functions, required for the normal development of an organism.

Although some authors are optimistic regarding the imminent emergence of single-cell phosphoproteomics, currently this approach is not viable. For instance, if we consider the hypothesis in which CDK act as the main reorganizer of proteinaceous membraneless cellular organelles during the cell cycle stochasticity might be a main characteristic of CDK phosphorylation and effects on specific proteins or complexes might be concealed by bulk proteomic experiments, due to cellular heterogeneity or absence of synchronization. By experimentally setting CDK1 activity at different levels, and using immuno-fluorescence (IF) or live imaging techniques, specific CDK targets that are located to those MLOs can be followed. Focusing on CDK targets that are key drivers of MLOs, we could assess if there exists direct and general association between the degree of formation/dissolution of these protein condensates and the level of CDK1 activity, regardless of the specific CDK1 phosphorylations.

2.1 Genetic switch for tuning CDK1 activity in human cell lines.

In this project, I wished to explore the reasons for the complexity of cell cycle regulation in human cell lines using relevant cell models. To do this, I sought to adapt the FLIP-mediated KO approach to generate human cells carrying an inducible genetic switch in its essential regulator, CDK1, to a synthetic and chemically controllable CDK1 fusion protein. With this objective in mind, I adapted a genetic system reported by Andersson-Rolf and colleagues in which they developed a tool for generating conditional and reversible gene knockouts (KO) (Andersson-Rolf et al., 2017). By the means of CRISPR-Cas9, an exon of a protein-coding gene is targeted with a guide RNA (gRNA) for the generation of double-strand breaks in the genomic sequence. A plasmid containing a designed “flipping cassette” flanked by left and right homology arms (HAL and HAR, respectively)

is co-transfected and serves as donor DNA for homology-directed repair (HDR). If indeed this is the case, the flipping cassette is inserted within the targeted exon in such a way that an artificial splicing site is formed rendering the allele fully functional and the endogenous protein is expressed (Fig. 1A). In contrast, if the NHEJ pathway is selected for the repair of the DNA, insertions or deletions of nucleotides in the coding sequence are the most probable consequence, generating mutations that will result in a null allele (Fig. 1A).

The flipping cassette contains a puromycin resistance marker for selecting clones with one recombined allele, bearing the flipping cassette. The other allele should be screened and those clones containing an additional null or recombined allele are selected. Fig. 1B illustrates the details of the flipping intron once inserted in the gene of interest. The construct was carefully designed in order to generate a split on the targeted exon, in the exact point where a consensus sequence for mammalian splicing sites -MAGR, where M is adenine or cytosine and R is a purine- is formed (Burset et al., 2000). A sequence motif for the branching point (BP1), required for intron splicing, is also present in the construct. Under these conditions, the phosphoglycerate kinase (PGK) promoter will drive the expression of the puromycin resistance cassette ($puro^R$) independently of the expression of the endogenous gene targeting.

The flipping intron contains two sets of Lox recombination sites: LoxP1 sites are represented as orange triangles and Lox5171 sites as purple triangles. Upon expression of the Cre recombinase enzyme, a first round of recombination using Lox sites in opposite directions will result in the flipping of the sequence contained between those sites. Indistinctively of which pair of Lox sites is first targeted, the result is a pair of complementary Lox sites in the same direction enclosing one of the other non-complementary Lox sites, the BP1 element, and the PGK promoter. A second round of recombination in the Lox sites with the same orientation will result in the excision of all elements flanked by those sites. The result of the succession of recombination events is that the puromycin resistance cassette, together with its transcriptional termination signal, gets oriented in the same direction as the gene of interest. A new branching point (BP2) is also inverted and oriented in the same sense between the first section of the exon split and the $puro^R$, which contains a splicing acceptor (SA2). In cells, the spliceosome will recognize the puromycin resistance cassette with the transcription termination elements as the contiguous exon, which will encode for an aberrant truncated protein, generating a null allele.

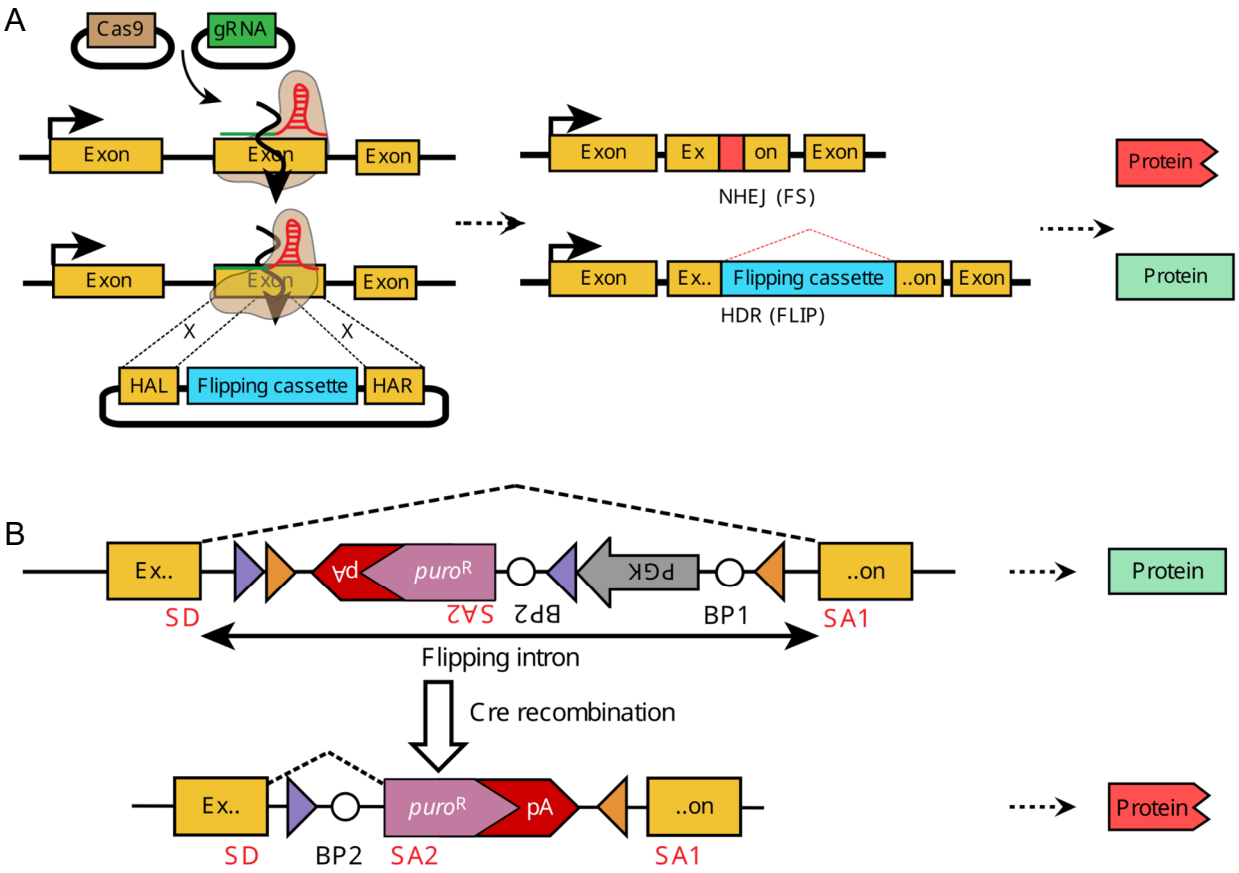


Figure 2.1 | Genetic system for conditional and reversible knockouts. Adapted from *Andersson-Rolf et al., 2017.* (A) Scheme for the strategy of the CRISPR-Cas9 mediated insertion of the flipping cassette. (B) The design of the FLIP cassette. HAL, homologous arm left; HAR, homologous arm right; PGK, phosphoglycerate kinase promoter. SD, splice donor; SA1, SA2, splice acceptor; orange triangles, LoxP1 sites; purple triangles, Lox5171 sites; BP1, BP2 (blue circles), branching point; pA, polyadenylation signal.

To make this system reversible, the authors flanked the puromycin resistance cassette and the termination signal with FRT recombination sites, sensitive to the Flippase recombinase. Upon expression of this enzyme, those elements get excised and the splicing occurs again between the splicing donor at the end of the first section of the split exon (SD), and the splicing acceptor at the start of the second section of the split exon (SA1). The gene regains its endogenous expression splicing isoform.

Andersson-Rolf and colleagues did not only design the genetic tools for the generation of the conditional and reversible knockouts but also used bioinformatics tools to examine all protein-encoding genes in the genome of mice and humans for insertion sites amenable for this strategy. In order for this approach to become viable, several criteria have to be met. First, the insertion point should overlap with a CRISPR site. Second, the exon in which it is located must be present

in all transcripts of the gene and be located within the first half of the protein-coding sequence. Third, the size of the split exons should be at least 60 bp, to conform to the minimum length of mammalian exons (Dominski and Kole, 1991). Finally, insertion points should match the consensus sequence for mammalian splice junctions (MAGR). Using these rules they mapped all the possible FLIP insertion sites obtaining over 1.7 million for both organisms, covering 16,460 and 15,177 genes in the mouse and human genomes, respectively.

1.1.1 Design of the genetic switch for the endogenous CDK1 locus

I designed a scheme on how to adapt the plasmid used by Andersson-Rolf and colleagues (Fig. 2A) and generated a FLIP intron, containing the coding DNA sequence (CDS) of as-CDK1-Cyclin fusion protein with a GFP CDS as a reporter gene, as shown in Fig. 2B.

The design of the homology arms (HA) should consider the requirements for defining a viable insert point defined by Andersson-Rolf and colleagues, but should also be located before the exon 4 of the endogenous CDK1 locus, which contains the sequence for the phenylalanine in the position 80 of the protein. Then, the CDK1* element in Fig. 2B will encode for the downstream CDS of CDK1 starting from the insertion site and will contain the F80G mutation. Two CDK-Cyclin fusion proteins are designed, with CyclinB1 and Cyclin A2, respectively. The CDS of the Cyclins is joint in the same reading frame of the CDK1* element by a sequence encoding for an S3 flexible linker, composed of three repeats of the Ser-Gly-Gly-Gly peptide. This linker should grant enough mobility freedom for the CDK-Cyclin complex to be formed correctly, but it also helps to stabilize the CDK1 allele otherwise hindered by the F80G mutation (Merrick et al., 2011). Both CDK and Cyclin genes are devoid of their stop codons. The fusion protein is joined to a GFP reporter gene by a sequence encoding for a P2A ‘self-cleavable peptide’. The GFP coding sequence contains its stop codon and the polyadenylation signal for the transcription termination.

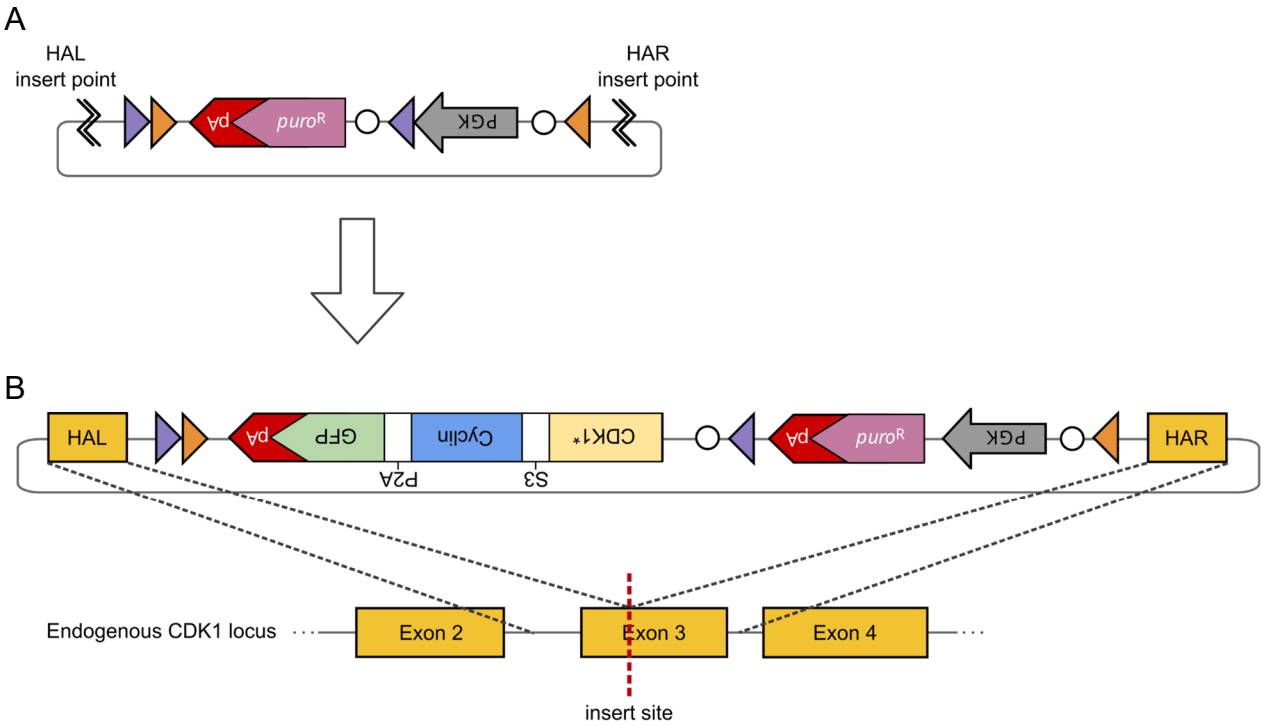


Figure 2.2 | Schematic design of the donor plasmid containing the FLIP intron. (A) The plasmid design was based on the system proposed by Andersson-Rolf et al. (Addgene number: #84539). The insert points for the homology arms (HA) use SapI-mediated Golden gate assembly. (B) The diagram of the final flipping cassette. The as-CDK CDS is fused by an S3 flexible linker to Cyclin CDS, which is joined by a P2A sequence encoding for an auto-cleavable peptide to the GFP CDS with the transcription termination signal. PGK, phosphoglycerate kinase promoter; SD, splice donor; orange triangles, LoxP1 sites; purple triangles, Lox5171 sites; S3, a 3X(Ser-Gly-Gly-Gly) flexible linker; P2A, an auto-cleavable peptide; white circles, branching points; pA, polyadenylation signal.

The flipping intron is designed in such a manner that upon Cre recombination, the PGK promoter and the puromycin resistance are lost and the endogenous CDK1 gene is switched to encode as-CDK1-Cyclin protein expressing GFP as a reporter (Fig 3).

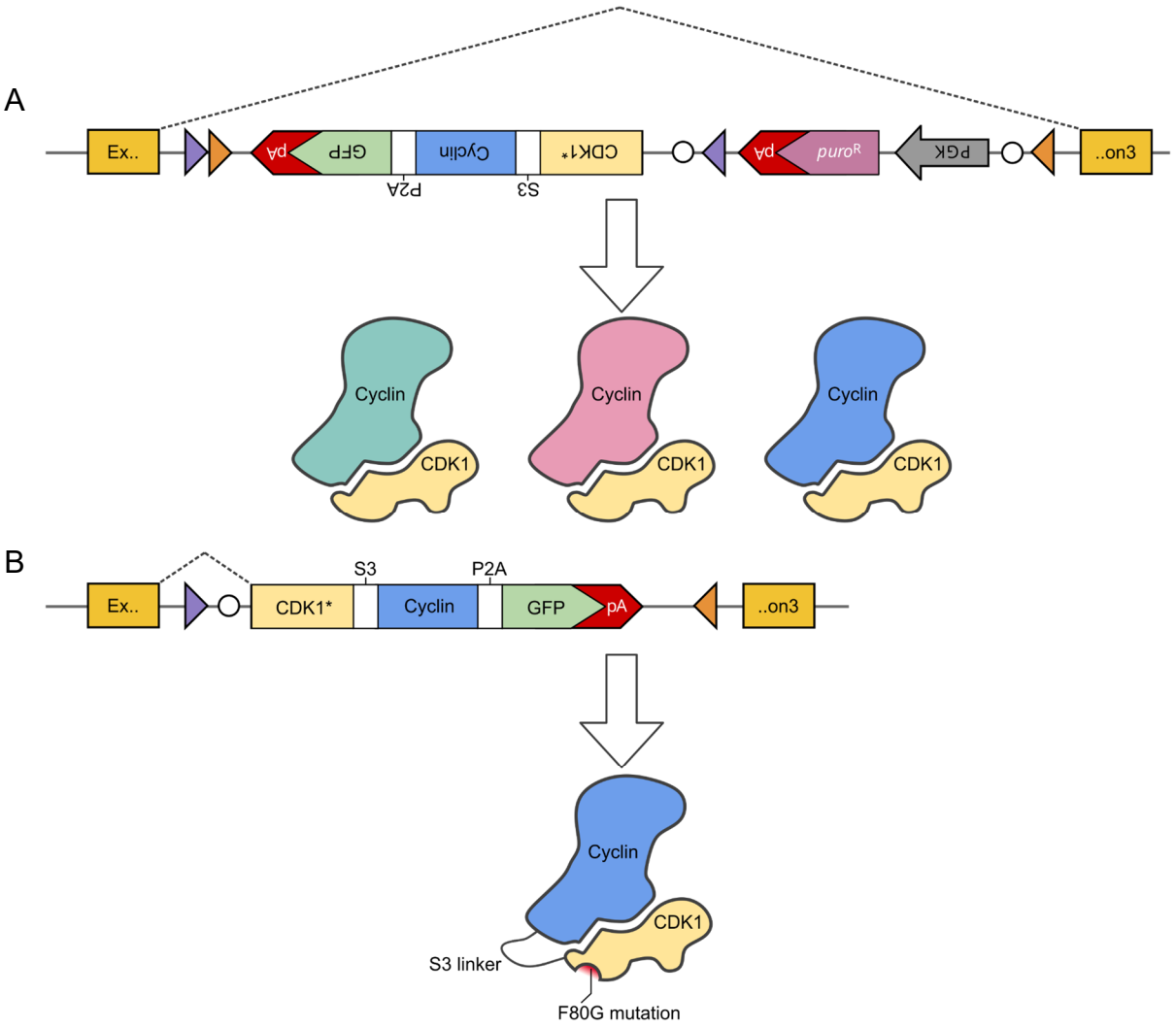


Figure 2.3 | Chemical-genetic switch for the expression of as-CDK1-Cyclin proteins. (A) In the absence of Cre recombinase expression, the flipping cassette is treated as an intron by the cellular splicing machinery. The endogenous CDK1 gene is expressed and CDK1 protein can interact freely with the corresponding Cyclins to form different complexes. (B) Upon Cre recombination, the as-CDK1-Cyclin is expressed instead.

2.1.2 Puromycin resistance cassette translocation

The first step for obtaining the donor plasmid DNA described in Fig. 2B is to translocate the puromycin resistance from where it is located in the backbone plasmid designed by Andersson-Rolf et al. (Fig. 2A, pUC118 - Addgene #84539) to a position between the PGK promoter and the neighboring Lox5171 site. The whole process consisted of two defined steps: removal and reinsertion of the puro^R cassette with the polyadenylation signal in a specific site, resulting in a translocation of this antibiotic selection genetic element (Fig. 4A). For the deletion of the puromycin resistance, I took advantage of the FRT sites flanking the resistance used in the

original design to reverse the KOs obtained. Recombination with the Flippase recombinase will recognize the FRT sites in the same orientation and will result in the loss of the DNA enclosed between these sites. Bacteria featuring arabinose-inducible Flippase recombinase expression (SW105 strain) were transformed with the pUC118 plasmid and were grown at two different concentrations of arabinose (0,1 and 0,2 %). The plasmid DNA of 10 clones was profiled by restriction reactions with *Hind*III to identify the recombination. *Hind*III digested DNA showed that, independently of the arabinose concentration, all the clones selected present a plasmid with a digestion pattern compatible with the loss of the puro^R cassette (pUC118-FLPrec) (Fig. 4B).

From the pUC118 plasmid, the puromycin resistance and the downstream termination signal are amplified by PCR using primers containing the *Cla*I restriction enzyme recognition site at both ends. An intense band of the expected size was observed in gel electrophoresis, together with multiple less strong unspecific bands. The band containing the fragment of interest was gel purified, and a new electrophoretic run of the purified DNA showed an almost complete absence of contaminant bands (Fig. 4C).

Using a *Cla*I unique restriction site as the insertion location, a two-coupled reaction protocol of digestion of both the puro^R amplicon and the pUC118-FLPrec plasmid followed by ligation was performed. In between the restriction and ligation steps, an extra step of phosphate removal at the ends of DNA is set by incubation of both the plasmid and the puromycin resistance cassette with Antarctic phosphatase. This helps to avoid plasmid recircularization and the polymerization of the insert. The products of the coupled reactions were used to transform competent bacteria, and a colony PCR screen was set up for detecting clones with the Puromycin resistance re-inserted.

Twenty-two PCR reaction tubes, each containing bacteria from 4 or 5 colonies, were set to amplify a 523bp fragment to verify if the puromycin resistance is present in some of the versions of the pUC118 plasmid in the correct orientation. The products were subjected to gel electrophoresis to observe the presence of a band of the correct size. Multiple tubes showed a band corresponding to the puro^R presence (Fig. 4D). I repeated the same procedure for individual clones from the tubes 12, 15, 16, 21, and 22 with, again, multiple reactions showing a band of the correct size. Clones 58, 60, 71, 72, 79, 104 were selected for XbaI-mediated restriction, where only clones 58 and 60 showed a restriction pattern compatible with what was expected if the translocation had occurred at the correct site. The translocation of the puromycin resistance for clone 58 was confirmed by DNA sequencing of the entire plasmid which will be referred to as pUC118-PUROtrans hereinafter.

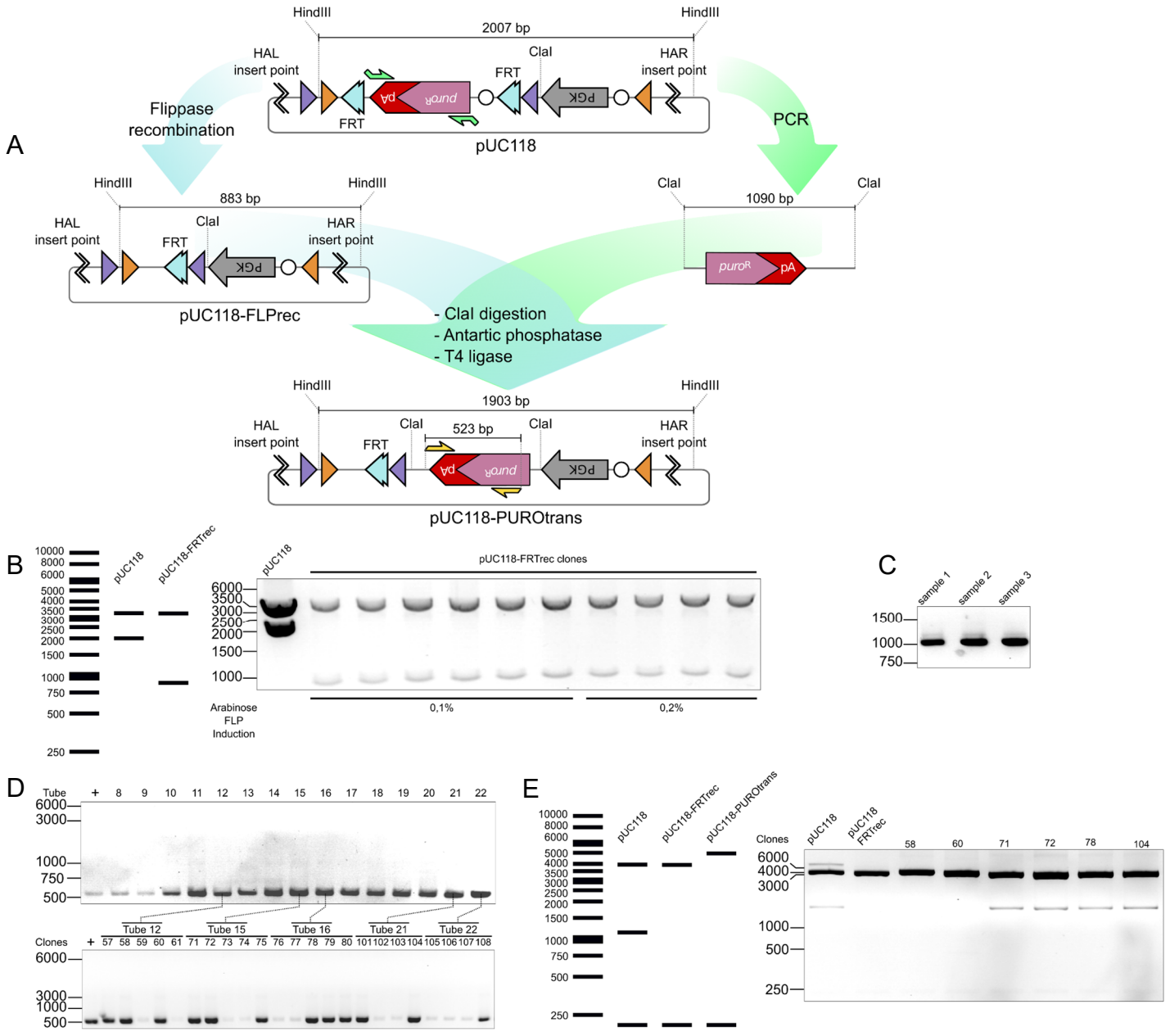


Figure 2.4 | Translocation of the puro^R cassette. (A) Schematic representation of the translocation process. Flippase recombination in bacteria was employed to generate the pUC118-FLPrec version of the plasmid and PCR for the amplification of the puromycin resistance cassette. Successive reactions with ClaI and antarctic phosphatase prepare both the plasmid and the puro^R cassette for the ligation with the T4 ligase to obtain the pUC118-PUROtrans plasmid. Light blue double triangles, FRT sites; green arrows, puro^R genotyping PCR primers; yellow arrows, puro^R PCR primers. (B) Gel electrophoresis of HindIII digested plasmid DNA from pUC118 transformed FLP-expressing bacteria. On the left, the expected pattern for each plasmid. (C) Gel purified PCR amplicons of the puro^R. PCR primers (green arrows Fig. 4A) were designed with ClaI overhangs. (D) Gel electrophoresis of colony PCR using genotyping primers (yellow arrows Fig. 4A). On the top, tubes 8 to 22 contain bacteria of 4 or 5 colonies each. Below, PCR of individual clones of a few positive tubes. (E) Gel electrophoresis of XbaI digested plasmid DNA from positive clones in Fig. 4D. On the left, the expected pattern for each plasmid.

2.1.3 Defining the insert site and design of homology arms

The data published by Andersson-Rolf et al. shows 4 potential insert sites in the CDK1 locus that fulfill all the requirements for implementing their tool, with 3 of them being located in the exon 3 and upstream of the Shokat mutation point (Fig. 5; Table 1).

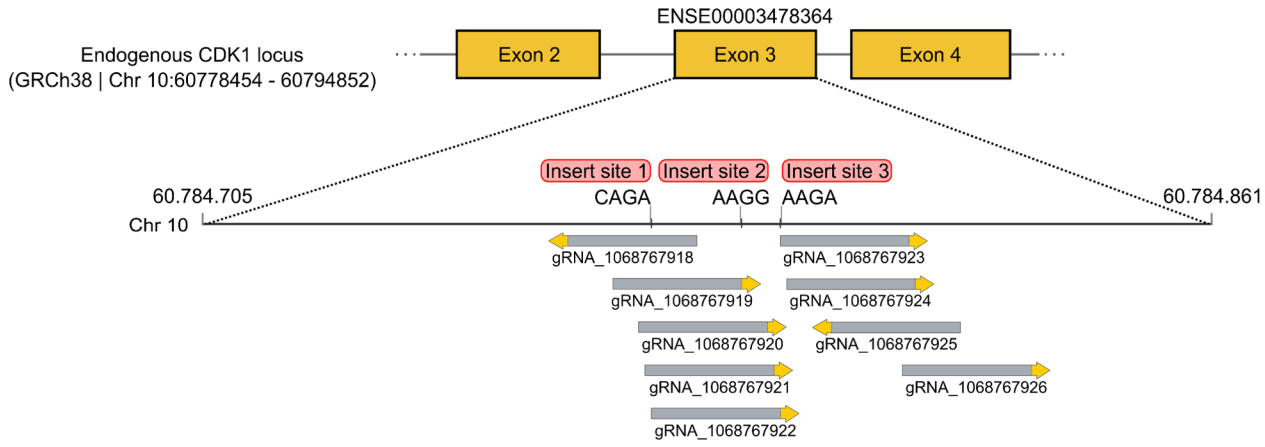


Figure 2.5 | Representation of the potential insert sites in the Exon 3 of the human CDK1 gene. The 157 bp of the exon 3 with the 3 potential insert sites are schematized in scale. The insert sites - with their sequences - are indicated with red tags and, below, the guide RNAs reported by Andersson-Rolf et al. with their corresponding IDs (Table 1). Yellow arrowheads indicate the PAM sequence of the gRNAs. All genomic coordinates reported are for the GRCh38 version of the human genome.

gRNA ID	Insertion point (chr position)	gRNA sequence
1068767918	Insert site 1 (60784772)	CCATGAAAAAAATCAGACTAGAA
1068767919	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	AATCAGACTAGAAAGTGAAGAGG
1068767920	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	AGACTAGAAAGTGAAGAGGAAGG
1068767921	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	GACTAGAAAGTGAAGAGGAAGGG
1068767922	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	ACTAGAAAGTGAAGAGGAAGGGG
1068767923	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	GGGTTCCTAGTACTGCAATTCGG
1068767924	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	GGTTCCTAGTACTGCAATTCGGG
1068767925	Insert site 1 (60784772) Insert site 2 (60784786) Insert site 3 (60784792)	CCTAGTACTGCAATTCGGGAAAT
1068767926	Insert site 2 (60784786) Insert site 3 (60784792)	TCGGGAAATTTCTCTATTAAAGG

Table 2.1 | **Guide RNAs targeting the exon 3 of the human CDK1 gene**. All the gRNAs presented in the table are informed by Andersson-Rolf and colleagues to comply with the requirements to target an insertion point for the flipping intron construct.

PCR primers were designed, *in silico*, for all 3 insert sites for the generation of homology arms ranging from 400 to 800 bp. Only insert site 1 allowed for the design of two couples of primers for the generation of the two homology arms at the specific splitting point. Subsequently, the homology arm fragments were obtained by PCR amplification of the genomic DNA of one of the cellular models to be used (SUM159).

Both homology arms shall be introduced in the backbone plasmid with the translocated Puromycin resistance by *Sap*I-mediated Golden Gate assembly (Engler et al., 2008). This type II restriction enzyme allows for DNA to cut one nucleotide away from the recognition site generating different single-strand overhangs while using the same enzyme. The restriction recognition sites were inserted at the extremes of the HA by adding their sequence into the PCR primers and followed by a specific sequence that will result in complementary overhangs with those in the backbone plasmid. This design allows for a single tube reaction in which *Sap*I digestion and T4 ligase-mediated ligation occur sequentially resulting in both homology arms inserted in the correct place and orientation (Fig. 6A).

The genomic DNA of the SUM159 cell line was isolated and purified to be used as a template for the PCR reactions with the designed oligos to obtain the homology arms. In both cases, the reactions rendered a clear band of the expected size (Fig. 6B). After a clean-up step, the DNA was used as the substrate for the Golden Gate reaction, which failed in the first instance. After testing the restriction and ligation steps separately, I found that the backbone plasmid had an extra recognition site for *Sap*I. A mutagenesis PCR with overlapping primers was designed to remove this site while maintaining the exact sequence of the original plasmid. The repaired plasmid was then used for the Golden Gate assembly protocol, this time obtaining a circular DNA of the expected size, that was used to transform competent bacteria. A miniprep was performed to purify the bacterial plasmid DNA that was, afterwards corroborated by Sanger sequencing of the entire plasmid.

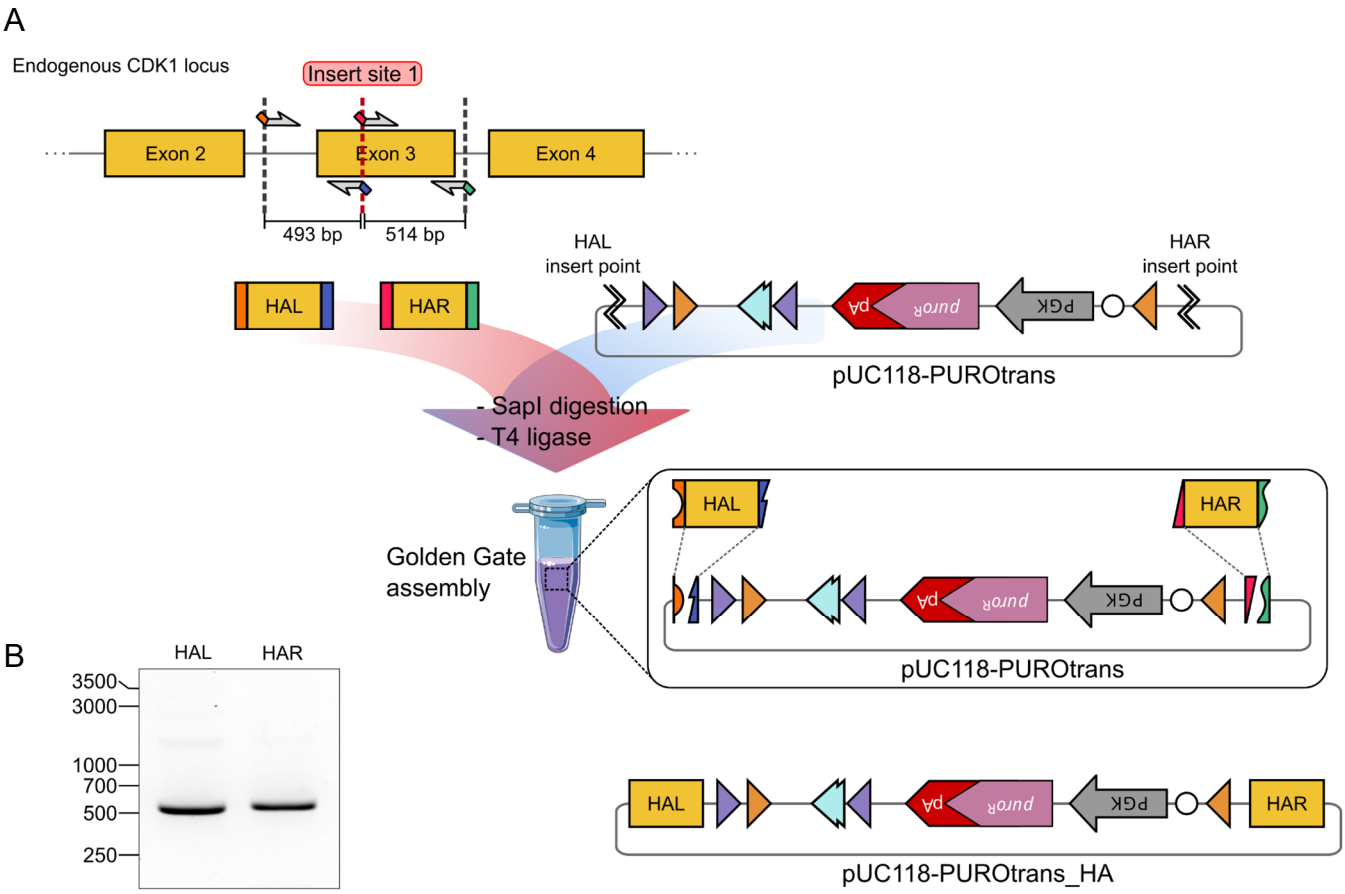


Figure 2.6 | **Insertion of the homology arms in the pUC118-PUROtrans plasmid**. (A) Schematic representation of the Golden Gate assembly reaction for the insertion of the homology arms. Different colors at both ends of the homology arm fragments represent the cut sites for SapI, and the different shapes indicate the specific overhangs formed in both HA and the plasmid. (B) Gel electrophoresis of the purified homology arm fragments.

2.1.4 Fusion proteins cassette synthesis and cloning

After the modification of the pUC118 plasmid to obtain the version of this plasmid where the puromycin resistance was translocated and the designed homology arms inserted (pUC118-PUROtrans), the next step consisted of inserting the gene encoding for the (as)-CDK1-Cyclin fusion proteins. We selected Cyclin B1 and Cyclin A2 for the generation of two different models with, presumably, different effects on the cell cycle when controlled only by their exclusive CDK1-Cyclin activity.

Given the inherent complexity and length (for the Cyclin B1 and A2 versions, 2908 bp and 2905 bp, respectively) of these constructs, we opted for a commercial solution. The company Synbio carried out the synthesis of both DNA fragments and cloned them into our pUC118-PUROtrans plasmid. They also performed a sequencing of the final plasmids.

2.1.5 Cellular models and transfection

The inducible genetic system for controlling the endogenous CDK1 activity was designed to be transfected into human cells and incorporated into their genomes by recombination. The two homology arms in the pUC118-PUROtrans were designed to target the CDK1 locus and to be inserted at a specific site.

If the flipping cassette is indeed added into the CDK1 endogenous gene, the puromycin resistance provides the selection marker. Puromycin-resistant cells will have at least one of the CDK1 alleles modified. In order to increase the recombination efficiency, DSBs are generated by the co-transfection of cells with the CAS9 recombinant protein coupled with a specific guide RNA targeting the insertion point as a ribonucleoprotein (RNP) complex. The first cellular model to be tested is HAP1, a nearly-haploid cell line derived from the chronic myeloid leukemia KBM7 line (Carette et al., 2011). In this case, only one allele of CDK1 is present, making this model a suitable option for establishing a proof of concept.

Two other diploid cell models are proposed for addressing different questions, but in these cases, genotyping should be performed to confirm homozygous recombination or the presence of one recombined allele and the other one carrying a null mutation promoted by CRISPR-Cas9. The first of these two models are a $CDK2^{-/-}$ human colon cancer HCT-116 cell line that was purchased from Horizon Discovery (HD R02-015). Clones presenting a $CDK1^{rec/rec}$ or $CDK1^{rec/-}$ genotypes will automatically demonstrate not only that the cell cycle can be driven by a single CDK1-Cyclin complex but also that this complex can substitute for CDK2 functions during S phase in human cells. Indeed, CDK1 has also been described as having overlapping functions with CDK2 during DNA replication in other organisms (Aleem et al., 2005; Hochegger et al., 2007). A new study shows that, in fact, the CDK1 functions in the G1/S transition overlaps with those of CDC7, demonstrated by the restricted S-phase entry of cells where both, CDK1 and CDC7, have been inhibited but the same effect is absent when a single kinase is targeted (Suski et al., 2022). CDK1 activity is not only relevant for the different stages of DNA replication, but it has also been proposed as to be required for the regulation of DNA damage repair systems (Ira et al., 2004; Xu et al., 2012). We have developed a dataset of *bona fide* human CDK1 subfamily targets (formally introduced in the Chapter 3) and we observed that the members of the CDK1 subfamily combined can phosphorylate a great proportion of the proteins shared by the DNA homologous recombination repair pathway (HRR) and DNA replication stress response pathway, as it has been suggested other authors (Cerqueira et al., 2009). The second model was then thought to be useful for providing mechanistic insights of CDK1 subfamily phosphorylations in DNA damage and replication stress responses: SUM159, a triple-negative breast cancer cell line for which we

have two versions, one being BRCA-deficient and the other one BRCA-proficient. BRCA1 and BRCA2 are critical for the HRR, and the loss of BRCA activity results in an increased rate of mutation and genomic instability due to suboptimal functioning of the DNA repair machinery during replication stress. Both BRCA1 and BRCA2, together with multiple other factors active in HRR are within the group of proteins observed to be phosphorylated by CDK1 subfamily members in our dataset, and we surmise that the inhibition of CDK1 activity during S phase could lead to a "BRCAness" phenotype, i.e., to emulate the phenotype observed in the absence of BRCA activity without any germline mutations in those genes.

Additionally, common effects observed after external CDK1 modulation in these different cellular models could help us understand more general functions of CDK-mediated phosphorylation and how it regulates the core processes of the cell cycle, such as mitotic cell reorganization.

The first pilot experiment in the HAP1 cell line was performed using the CRISPRMAX lipofection agent that is optimized for the administration of proteins into cells, notably Cas9. Three different guide RNAs were complexed with the recombinant Cas9 protein and were then mixed with the transfection reagent and both pUC118-PUROtrans plasmid containing the different CDK1-Cyclin versions of the flipping cassette. Three gRNAs (1068767919, 1068767920, and 1068767921 see Table 1) overlapping with the insert site 1 were tested in each condition, in addition to a negative control without the guides. An extra condition without a donor plasmid was also set to evaluate the efficiency of each guide. In this case, it is expected that frameshift mutations in CDK1 promoted by the CRISPR-Cas9 system would affect cell proliferation. In all cases, cells showed reduced growth after transfection with signs of recovery after 48 hours, a time point in which the selection marker Puromycin was added to the culture. No Puromycin-resistant cells were detected in any of the conditions, indicating that the insertion of the flipping intron did not take place. There was no indication of CDK1 Cas9-mediated loss of function since no reduction of proliferation was observed in any of the cases, including those conditions where specific CDK1-targeting guide RNAs were absent. Nevertheless, none of the conditions showed any transient mild puromycin resistance either. I attempted to troubleshoot the individual transfection steps separately by administering plasmids using a less aggressive lipofection protocol in the absence of Cas9-gRNA RNP complexes. Again, transient puromycin-resistant HAP1 cells were not observed in any of the conditions.

We surmised that HAP1 cells might not be receptive to lipofection reagents and another more general delivery method can be suitable for the simultaneous administration of Cas9, RNA, and DNA to cells. We aimed to set up an electroporation experimental protocol using the Neon transfection system, with which one of our cellular models (SUM159) was successfully transfected

by collaborators. Detailed information of the parameters used and the experimental conditions were also available for HCT-116, but not for HAP1 cells. We resolved, hence, to limit the first electroporation pilot experiment to SUM159 and HCT-116 cells. A colleague expert in genetics suggested that linearizing the donor DNA can effectively augment the recombination efficiency. By using specific PCR primers, I obtained a linear version of the flipping intron flanked by the homology arms. Both cell lines were subjected to short pulses of high voltages in the presence of the Cas9-gRNA complexes and the donor linear DNA and were left to recover during 24 hours, before the addition of puromycin. A vast majority of HCT-166 were observed dead 16 hours after electroporation whereas SUM159 cells were able to recover, although with a considerable rate of cell death. The addition of puromycin had a notorious effect on cell viability by 24 hours after the beginning of selection, and by 48 hours. Virtually all cells in the culture were showing morphologic signs of apoptosis or necrosis.

2.2 Genetic switch for tuning CDK1 activity in a mouse model.

The unsuccessful experiments in human cellular systems, even if preliminary, led us to consider the possibility that our approach might be overly complex. Principally, relying on the formation of an alternative splicing site can introduce additional failing points to a system that already is remarkably intricate. In spite of our meticulous design of each component of our chemical-genetic tool, low efficiency in the delivery and integration steps together with the possibility that the insertion might not be recognized as an intron, and disrupt the gene expression could greatly compromise our chances of success. Delivering macromolecules into cells and site-specific integration of DNA fragments are common experimental protocols in cell biology and easier to troubleshoot if any issue ensues. Generating an artificial splicing site, however, is not.

A simpler system, where the insertion of the flipping cassette takes place in the downstream intron of the exon 3, can be envisaged. In this case, the integrity of the coding sequence of the endogenous gene is not endangered, resulting in a system less prone to failures. Nevertheless, considering the low rate of homozygous insertions, even screening all possible clones will not assure to find cells with both alleles carrying our genetic system. Heterozygous clones can be selected for a second CRISPR-Cas9-mediated insertion or a knockout of the second allele, but again, in the absence of a specific marker for differentiating heterozygous clones from those bearing a $CDK1^{rec/rec}$ or $CDK1^{rec/-}$ genotype, a posterior genetic screen should be performed to detect an event with an extremely low prevalence.

One way of overcoming these limitations is to use an *in vivo* system that allows crossing heterozygous individuals to obtain homozygotes with a much higher chance, according to Mendelian inheritance. We therefore, designed a similar chemical-genetic system that was previously described for cells, based on the FLEx system (Schnütgen et al., 2003) and our previous construct. We submitted this proposal to a call for the development of novel genetically modified mouse models by PHENOMIN, a consortium of laboratories with the infrastructure and expertise required to conduct such a project, and it was accepted to be funded. This genetic switch also uses a combination of inwardly oriented Lox sites (LoxP1 and Lox5171 sites), such that, upon induction of Cre, recombination will result in flipping of the cassette, replacing the endogenous CDK1 with the synthetic fusion construct (Fig. 7). As I described above, the fusion of Cyclin helps to stabilize the (as)-CDK1 allele that, otherwise, is hypomorphic *in vivo* (Michowski et al., 2020). This is strengthened by an additional mutation to introduce a Methionine-to-Valine change in the position 32 (suppressor of glycine gatekeeper) of the CDK1 protein (Zhang et al., 2005). Besides, the presence of integrated antibiotic resistance cassettes can generate hypomorphic phenotypes and, therefore, the puromycin selection marker would not be controlled by the Cre recombination. Both, the puromycin resistance cassette and the PGK promoter should be flanked by FRT sites and excised by FLP recombination once the heterozygous CDK1$^{rec/+}$ mice are obtained, without altering the Cre switch system.

PHENOMIN informed us that they employ homologous recombination of embryonic stem cells (ES cells) as a strategy for the generation of the mice lines, which would require a set of larger homology arms than those of the previous design, in order to achieve a good insertion efficiency (Fig. 7). Mice carrying the genetic construct will be crossed to different inducible Cre driver lines, such as Villin-CreERT2, allowing inducible recombination in adults in the highly proliferative intestinal epithelium compartment.

Homozygous Cre-recombined (as)-CDK-Cyclin-expressing mice will allow us to evaluate whether in adult mice the cell cycle can be driven by a single CDK1-cyclin complex; if so, this will demonstrate that the core cell cycle architecture is conserved from protozoans to vertebrates. Additionally, by further crossing to available Cdk2$^{lox/lox}$ mice (from the Barbacid lab), we will assess whether this single complex suffices to drive the entire cell cycle.

This allele will further be used for specific manipulation of CDK1 activity in intestinal organoids and primary ES cells, allowing the study of the effects of specific modulation of the cell cycle in different cellular models.

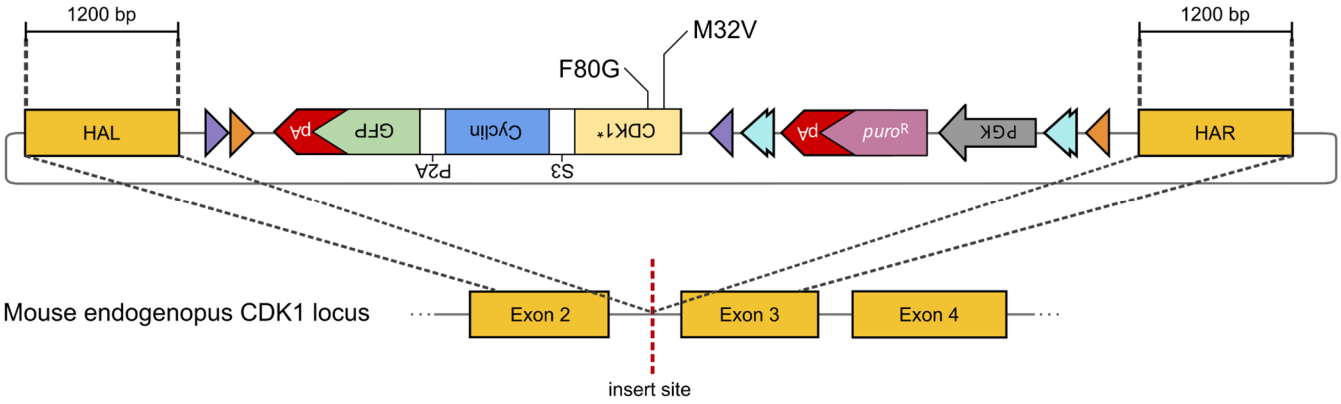


Figure 2.7 | Schematic design of the donor plasmid for the genetic switch of CDK1 in mice. The plasmid design was based on our previous construct and the FLEx system. The as-CDK CDS is fused to Cyclin CDS by a S3 flexible linker, and the Cyclin CDS is joined by a P2A sequence encoding for an auto-cleavable peptide to the GFP CDS with the transcription termination signal. PGK, phosphoglycerate kinase promoter; SD, splice donor; orange triangles, LoxP1 sites; purple triangles, Lox5171 sites; light blue double arrowheads, FRT recombination sites; S3, a 3X(Ser-Gly-Gly-Gly) flexible linker; P2A, an auto-cleavable peptide; pA, polyadenylation signal.

3. CDK regulation of protein condensation

Eukaryotic cell cycle progression depends on the CDK family of kinases, presumably arising from the collective behavior of altered protein phosphorylation states. It was long believed that the specific combinations of CDKs and cyclins were critical for the progression through the different phases of the cell cycle. However, most CDK and cyclin genes are dispensable for cell proliferation in the majority of cell types in the mouse and only CDK1 is essential for the cell cycle in mammals (Santamaria et al., 2007) in association with either cyclin A or E (Kalaszczynska et al, 2009), and cyclin B1 (Brandeis et al., 1998). Moreover, in fission yeast, the oscillating activity of CDK1 alone can drive the entire cell cycle (Coudreuse and Nurse, 2010; Fisher and Nurse, 1996). This proves that a rather limited core network of CDKs can drive the eukaryotic cell cycle, and that changes in overall CDK activity somehow determine the sequence of the complex processes required to duplicate the genome and reorganize the cellular components during cell division.

A major implication of this “quantitative model” (Fisher and Nurse, 1996) is that the cell cycle appears to be ruled by the kinetics of CDK activity rather than by CDK-Cyclin complex-specific phosphorylations. Indeed, when only CDK1 is present in mice, this kinase is able to compensate for other CDKs and bind all types of cyclins (Santamaria et al., 2007) to proceed with the phosphorylation required to progress through the cell cycle. This implies that there exist low and high overall CDK activity thresholds for entry into S-phase and mitosis, respectively, determined by the CDK-regulatory network. This network involves positive and double-negative feedback loops, as well as futile cycles of CDK and CDK-opposing phosphatase activity (Fisher et al., 2012). Such features of network organization can generate ultrasensitivity and hysteresis in CDK1 activation (Novak et al., 2010), while the resulting bistability of CDK1 activity leads to a switch-like G2/M transition (Tyson and Novak, 2001). These theoretical concepts are supported by experimental evidence in *Xenopus* egg extracts and mammalian cells (Rata et al., 2018; Novak et al., 2010; Trunnell et al., 2011; Kim and Ferrell, 2007; Pomerening et al., 2003; Sha et al., 2003).

The presumed switch-like dynamics of the CDK1 regulatory network are consistent with the observed abrupt reorganization of the cell at mitosis. In metazoans, the nuclear envelope and lamina break down and many cellular structures are rapidly disassembled, with most of MLOs (nuclear pore complexes, nucleoli, splicing speckles, Cajal bodies, PML-nuclear bodies, and stress granules) falling in this category (Banani et al., 2017; Hyman et al., 2014; Shin and

Brangwynne, 2017; Woodruff et al., 2018). This might indicate that the control of MLO assembly and disassembly is coordinated in a cell cycle-dependent manner. MLOs are thought to assemble by mechanisms of phase separation involving multivalent protein interactions and intrinsically-disordered regions of proteins (Banani et al., 2017). This process has been described to be regulated by protein kinases, including CDK (Berchtold et al., 2018; Hur et al., 2020; Rai et al., 2018; Yahya et al., 2021). Considering that protein phosphorylation is in general enriched in IDRs (Iakoucheva et al., 2004), and this also appears to be true for CDKs (Holt et al., 2009; Michowski et al., 2020; Moses et al., 2007), a compelling hypothesis is that CDK-mediated IDR phosphorylation regulates phase separation dynamics of MLOs in mitosis.

This would be consistent with the fact that CDK1-family CDKs can phosphorylate hundreds of sites on diverse proteins (Blethrow et al., 2008; Chi et al., 2008; Errico et al., 2010; Ubersax et al., 2003) following a switch-like dynamics model. Nevertheless, direct evidence for switch-like dynamics of cell cycle-regulated phosphorylation *in vivo* is currently lacking, mostly due to the lack of sensitivity of single-cell proteomics studies (Budnik et al., 2018; Lombard-Banek et al., 2019) for low stoichiometry and highly dynamic targets such as phosphosites. Therefore, studies analyzing cell cycle phosphorylation have generally used cells blocked at different stages of the cell cycle to generate “snapshots” of the phosphorylation landscape (Olsen et al., 2010). However, highly dynamic phosphorylation states cannot readily be determined from populations of cells (Purvis and Lahav, 2013). Moreover, whole-culture synchronization methods generate artifacts due to cell cycle perturbation (Cooper, 2019; Ly et al., 2015) and alternative phosphoproteomics approaches on unsynchronised cells selected with centrifugal elutriation (Ly et al., 2014) or FACS (Ly et al., 2017), lack the temporal resolution to determine the dynamics of protein phosphorylation throughout the cell cycle.

In this chapter I present my work in the context of a multidisciplinary collaboration aimed to study several of the aforesaid hypotheses. I wished to answer the question “Is CDK phosphorylation enriched in IDRs?”. A question that has been addressed in the past by multiple authors, however, not in sufficient detail to completely rule out random or indirect associations. This was done by the bioinformatic study of publicly-available protein phosphorylation data for yeast and humans, and was then extended to other cell cycle-related kinases. Next, by establishing a collaboration with the proteomic specialists J.M. Valverde and M. Altelaar from the University of Utrecht, we sought to understand the dynamics of the cell cycle phosphorylation in an untainted *in vivo* system. We used *Xenopus laevis* single-embryo phosphoproteomic approaches to describe how protein phosphorylation progresses through the time course of cell divisions. By combining these results with our previous bioinformatic approach we confirmed that cell cycle phosphorylation is

dominated by CDK, and we aimed to assess whether the enrichment of phosphorylation in disordered regions revealed a similar behavior to what was observed for human and yeast datasets. We collected conclusive evidence showing a categorical tendency of cell cycle phosphorylations, commanded by CDK1 subfamily activity, of being located to IDRs. Finally, these results led us to pursue a collaborative effort with experts in the fields of biochemistry, cell biology, biophysics and structural biology, for presenting a case report for the proliferation marker KI67, a CDK1, target for which we demonstrated that its phosphorylation state regulates its involvement in phase separation processes.

3.1 Bioinformatic analysis of CDK phosphorylation

Considering the vast amount of research published on protein phosphorylation, we judged the volume of publicly-available experimental data on CDK phosphorylation sufficient to explore the hypothesis that CDK phosphorylation is enriched in unstructured regions of the proteins. Using bioinformatics and biostatistics as the main tools, a large-scale phosphoproteome-wide analysis of CDK targets and their phosphosites can be proposed for multiple organisms. This analysis, however, must be based on bioinformatic prediction of protein disorder rather than experimental data of protein structure, due to limitations previously discussed (see introduction section 1.4.2). In order to avoid biases associated with particular prediction methods, all the results reported in this section were obtained using multiple predictors.

The combination of CDK phosphorylation and protein disorder data coming from different sources will help to address some shortcomings observed in previous publications. Most notably, to consider the fact that predicted disordered regions of proteins possess a compositional signature where the amino acids that constitute the consensus recognition motifs for CDK phosphorylation are enriched. Therefore, an analysis taking into account this compositional bias together with the replication using different predictors is needed to discern if phosphorylation is happening primarily in disordered regions or if disorder predictors detect regions enriched in phosphorylatable residues as disordered. To break this circularity, we also repeated the analysis in evolutionary distant organisms. If results are replicated this might imply that an evolutionary mechanism is acting to select IDRs as sites of phosphorylation by CDK, contributing to the body of evidence that supports our hypothesis.

3.1.1 CDK phosphorylation data collection

Mostly based on availability we selected two different organisms to collect data of CDK phosphorylation: humans and budding yeast. This process was principally done by the use of informatic tools, although the collaboration of Dr. Krasinska in the manual curation of the datasets was crucial to ensure high-fidelity data.

For humans we compiled a set of 1200 phosphosites for 656 human CDK1-subfamily targets, combining data on 450 CDK1, CDK2 and CDK3 substrates from PhosphoSite Plus (Hornbeck et al., 2015) with manually curated information on 206 targets from several human CDK substrate screens and other studies (Fig. 1A). From those phosphosites, a striking 89% (1076) corresponds with proline-directed phosphorylations, that is equivalent to the minimal consensus motif, and 34% (412) presented the full consensus motif for CDK. This highlights the preponderance of CDK phosphorylation in specific sequence motifs, although the discovery bias towards proline-directed sites should be also taken into account given that the presence of the consensus motifs has been considered a requirement for CDK-mediated phosphorylation by the the scientific community. The rest of the phosphorylations deposited in PhosphoSite Plus were used to define the phosphoproteome universe covering 82% of the entire human proteome as reported by UniProt (Fig. 1A).

We selected budding yeast as the evolutionary distant counterpart for humans. We took advantage of two phosphoproteome-wide studies from the laboratory of David Morgan in which they defined the targets for the CDK1 analog, Cdc2, by using *in vitro* and *in vivo* approaches, respectively (Ubersax et al., 2003; Holt et al., 2009). Ubersax and colleagues used an as-CDK mutant to perform *in vitro* phosphorylation reactions with a γ-P^{32} radiolabeled bulky analog of ATP of different strains of yeast extracts, each one containing a single ORF with an amino-terminal fusion to glutathione-S-transferase (GST). A total of 695 proteins (11% of the yeast proteome) were selected using different criteria in order to generate a representative sample of the CDK phosphorylation landscape, where 522 were specifically chosen for having at least one minimal CDK consensus motif in their sequence. After performing kinase assays, proteins are purified using the GST tag, and their phosphorylation level was estimated as a ratio of the phosphorylated protein and the total amount of protein, measured by quantification of autoradiography signal and gel silver staining signal, respectively. The logarithm of the amount of radioactivity incorporated divided by the nanograms of protein was then defined as the “P-score”, and proteins with values greater than 2 were considered to be CDK targets.

Holt and colleagues used Stable Isotope Labeling of Amino Acids in culture (SILAC) in an as-CDK1 mutant yeast strain that can be rapidly inhibited by the 1-NMPP1 ATP analog. This strain

is also lysine and arginine auxotroph and requires exogenous supplementation of these amino acids to survive. Two growth media were prepared, one containing normal lysine and arginine (the 'light' culture), and the other supplied with arginine and lysine labeled with stable heavy isotopes (the 'heavy' culture). This last culture was briefly treated (15 min) with ATP analog 1-NM-PP1 to specifically inhibit as-CDK1 activity. The cultures were then mixed together, lysed, and subjected to trypsinization and the phosphopeptides were purified to be analyzed by tandem mass spectrometry. In the original publication authors used two criteria to consider a phosphorylation site as a Cdk1 substrate: the phosphorylated residue should conform to the minimal Cdk1 consensus, and the phosphopeptide must have declined in abundance by at least half after Cdk1 inhibition. They found 547 phosphosites belonging to 308 different proteins that were identified as CDK1 targets.

This last approach also provided information about specific phosphorylation sites on CDK targets. We defined *bona fide* CDK1 sites by intersecting data for *in vitro* CDK1 substrates with all the targets that show a decrease in abundance by half upon CDK1 inhibition *in vivo,* and not only those which are proline-directed. Of these 604 CDK targets (1300 phosphosites), 100 intersected with yeast CDK1 substrates defined *in vitro* were also phosphorylated in a CDK1-dependent manner *in vivo* (Fig. 1B), with 19 of these not being proline-directed. These 352 phosphosites in 100 proteins constitute our high-confidence yeast CDK target dataset. The complete set of phosphoproteins detected in both studies constitute the phosphoproteome for yeast.

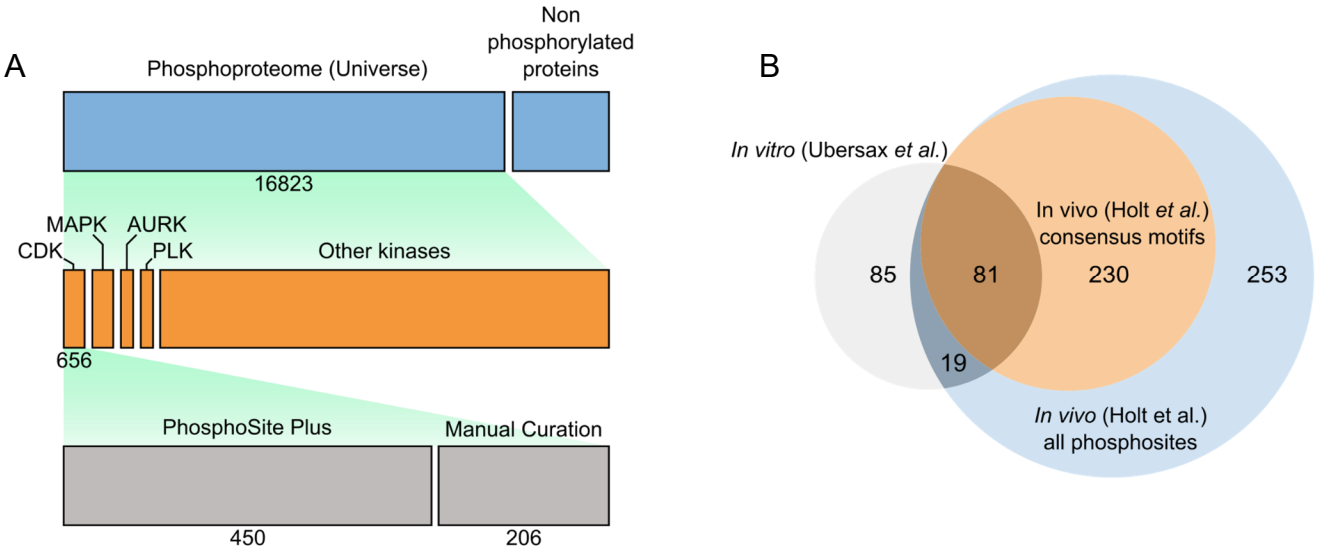


Figure 3.1 | CDK phosphorylation datasets. (A) Bar scheme illustrating the different subsets of the human phosphoproteome. Blue bars correspond to the entire proteome, orange bars to the phosphoproteome, and gray bars represent the CDK targets. MAPK, MAP kinase; AURK, aurora kinases; PLK, polo-like kinases. (B) Venn diagram showing the intersection of both studies used to define the high-confidence set of CDK targets for yeast.

3.1.2 Protein disorder data collection

Information of disordered regions in proteins was collected from multiple disorder predictors for the proteomes of both yeast and humans, rather than selecting a single method or a meta-predictor for reaching a consensus. I explored the database MobiDB, in which disorder scores are pre-calculated with several predictors for 17 different eukaryotic whole proteomes. I obtained the information of 12 predictors, detailed below.

- pFilt and SEG: predictors for low complexity regions in proteins.
- DisEMBL-465 and DisEMBL-HL: disorder detection by artificial neural networks trained with information of missing electronics densities in PDB structures and loops having a high temperature factor, respectively.
- ESpritz-DisProt, ESpritz-NMR and ESpritz-Xray: disorder detection by bidirectional recursive neural networks trained with manually curated data, NMR-solved structures and X-ray crystallography structures, respectively.
- GlobProt: analytical calculation of the propensity of whether a region is globular or not
- IUPred-long and IUPred-short: residue-level energy estimation method to detect values incompatibles with structured regions optimized for long and short disordered regions, respectively.
- JRONN: disorder detection by a regional order neural network trained with information from the Molecular Structure Database.
- VSL2b: a meta predictor for disorder detection by integrating linear support vector machines trained in short and long disorder regions of proteins

An additional novel predictor, SPOT2, was added to our catalog. This algorithm, based on long short-term memory networks, is trained with input data from evolutionary information, which allows the simultaneous detection of short and long disordered regions. It is not available as a pre-calculated score in MobiDB; we therefore used the distributable standalone software provided by their developers to run it through both yeast and human proteomes. SPOT2 is found among the top-ranked disorder predictors in recent benchmarks (Liu et al., 2019; Necci et al., 2021), which is also true for IUPred-long and VSL2b. These three predictors were chosen to perform most of the quantitative analysis reported in this section. This selection was based on the fact that they are methodologically different enough to be representative, and they also have a markedly different sensitivity. As a general rule-of-thumb, VSL2b is a more permissive, predictor detecting more and longer disorder stretches of IDRs, SPOT has an intermediate performance, and IUPred is the most restrictive. This is the reason why the latter was chosen as the method used for general

analysis and qualitative graphical representations. Nevertheless, all predictors seem to agree in the regions they predict as disordered, and while more permissive methods extend the prediction of more restrictive methods, they cover them almost completely. This was illustrated by the identification of all the predicted disordered residues in the yeast and human proteomes with the three predictors highlighted above. Subsequently, a venn diagram for each organism was produced, where the common prediction across the different methods can be observed as the overlapping area (Fig. 2A, B). In both cases, the more restrictive predictors have only a small proportion of predicted residues that are not recognized by the more permissive ones, confirming that there is a high consensus among the predictors on the regions recognized as disordered.

Next, I wanted to confirm if the reported characteristic amino acid composition of IDRs (cites) could be observed also in our datasets for the 3 disorder predictors. Multiple authors have claimed that disordered regions of proteins are enriched in polar and charged amino acids, while non-polar and aromatic residues are significantly less abundant. I calculated the differential composition in disordered regions as:

$$Differential\ composition = \frac{composition\ in\ disordered\ regions\ -\ composition\ in\ the\ phosphoproteome}{composition\ in\ the\ phosphoproteome}$$

Positive values indicate which are the amino acids that are enriched in IDRs, while amino acids depleted in disordered regions present negative values. For the phosphoproteomes of yeast (Fig. 2C) and humans (Fig. 2D), this tendency could be confirmed for the IUPred, SPOT and VLS2b predictors. One interesting observation is that in all cases Serine, Proline and to a lesser extent Threonine are considerably enriched in IDRs. This three amino acids are of special interest due to the fact that they conform with the minimal consensus motif for CDK phosphorylation ([S/T]*P). This should be taken into account when analyzing the enrichment of phosphorylation in IDRs, to avoid artifactual results due to the compositional bias.

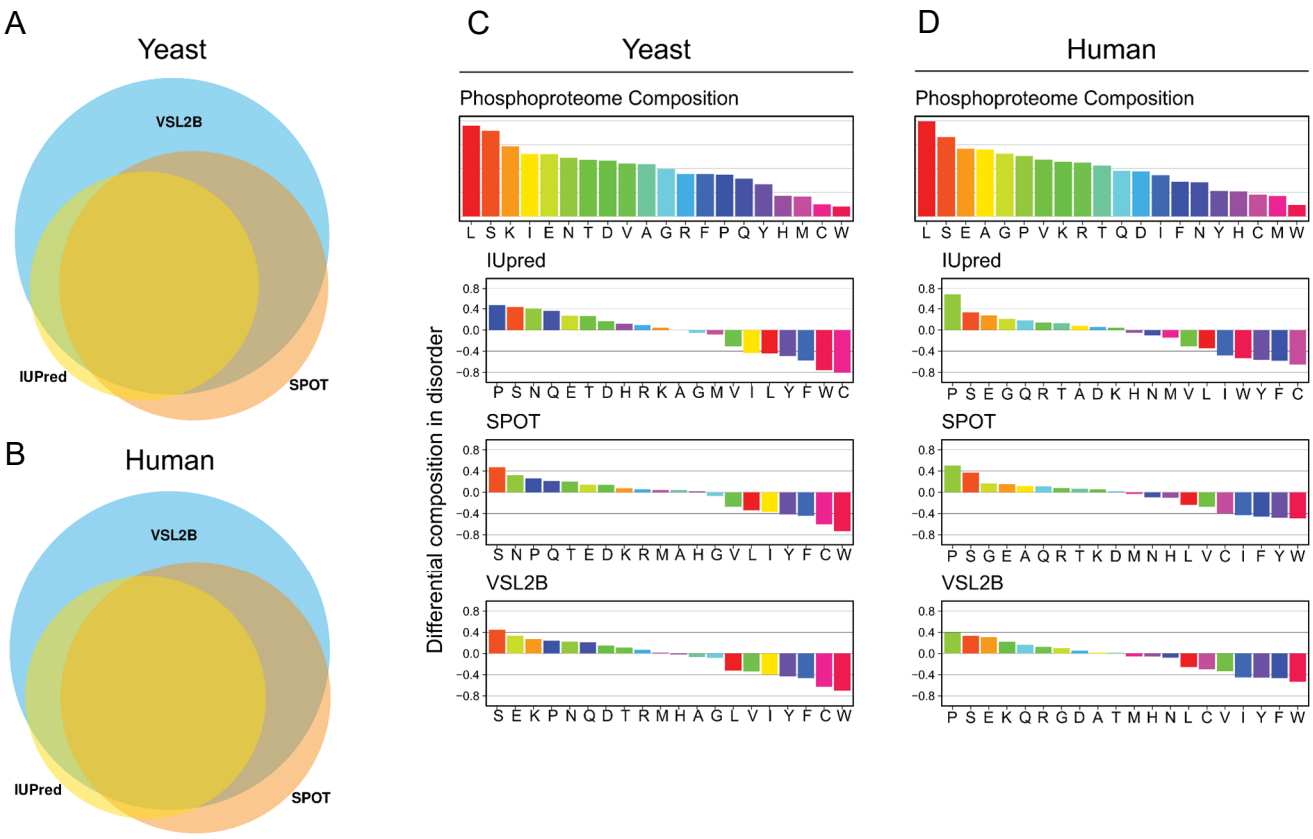


Figure 3.2 | Evaluation of disorder predictors. Left, Venn diagrams showing the overlapping predictions of disorder of IUPred, SPOT and VLS2b predictors for the entire proteomes of yeast (A) and humans (B). Right, the differential amino acid composition in IDRs. Positive values imply enrichment in IDRs while negative values represent amino acids depleted in IDRs. The analysis was done with the IUPred, SPOT and VLS2b predictors for yeast (C) and human (D). Amino acids are colored in a rainbow fashion based on their increasing abundance in the phosphoproteome. Disruption of the rainbow patterns in the differential composition plots indicates specific amino acid composition in IDRs.

3.1.3 Phosphorylation is enriched in IDRs

Phosphorylation has been reported to be enriched in disordered regions of proteins. However, in all the previous publications I examined, the compositional bias of disordered regions was not taken into account when analyzing the propensity of phosphorylations to localize in IDRs. We wondered if that observation was true for our two datasets of phosphorylations in yeast and humans, when correcting for this bias. A simple normalization was applied, in which the expected probability of phosphorylation being located in IDRs, rather than depending on the length of disordered or structured regions, was based on the amount of phosphorylatable residues (Ser and Thr; phospho-Tyr were not considered in this analysis) present in each region (Fig. 3A). Thus, if an IDR covers 10% of the protein length, but contains 80% of the phosphorylatable residues the expected probability of a given phosphorylation of being located in a disordered region is 80%.

By using this normalization, I modeled the distributions of phosphorylation between disordered and structured regions, as defined by UIPred (the version optimized for detecting long disordered regions), of each protein using the binomial distribution. Using the proportion of phosphorylatable amino acids in disorder as the expected probability, I assessed with a binomial test if the proportion of phosphorylations in disordered regions were greater than the expected value for each protein. The results for both organisms were plotted in a scatter plot where each point has as coordinates the expected and observed values of phosphorylations in IDRs, and its color indicates the adjusted p-value (Benjamini-Hochberg) of the binomial test (Fig. 3B). In addition, the proteins identified as CDK targets are highlighted with a different shape.

In both yeast and human phosphoproteomes, a clear tendency of phosphosites to be located to disordered regions is observed, indicated by the almost unanimously greater value of observed phosphorylated Ser and Thr than what would be expected if we considered the proportion of those amino acids in disordered regions. No obvious difference is detected for CDK targets, however, when compared with the rest of the phosphorylated proteins.

In conclusion, even when considering the natural bias of IDRs to contain a higher proportion of phosphorylatable residues, phosphorylation has a tendency to occur in disordered regions, and these results could be reproduced in two evolutionary distant organisms such as humans and yeast.

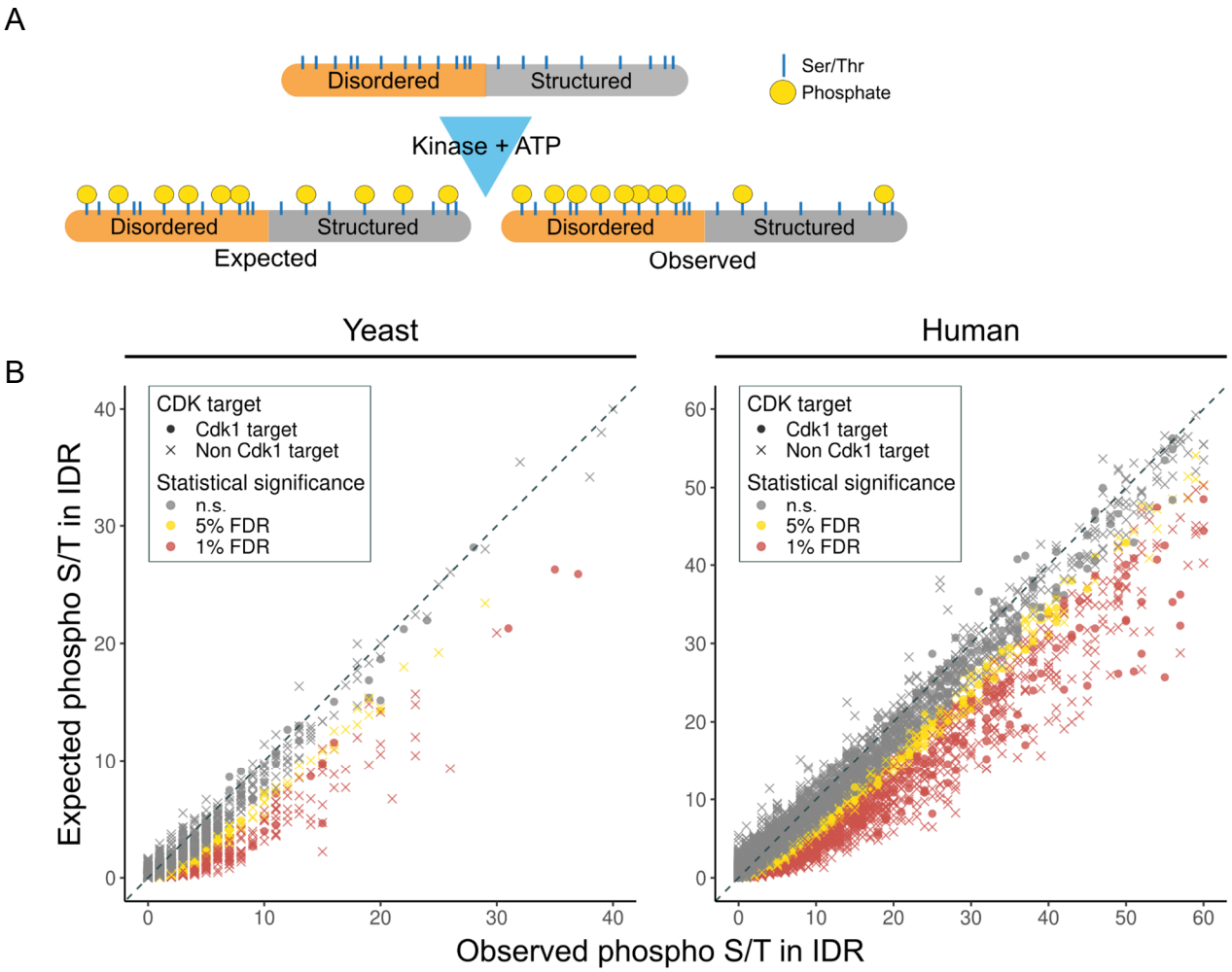
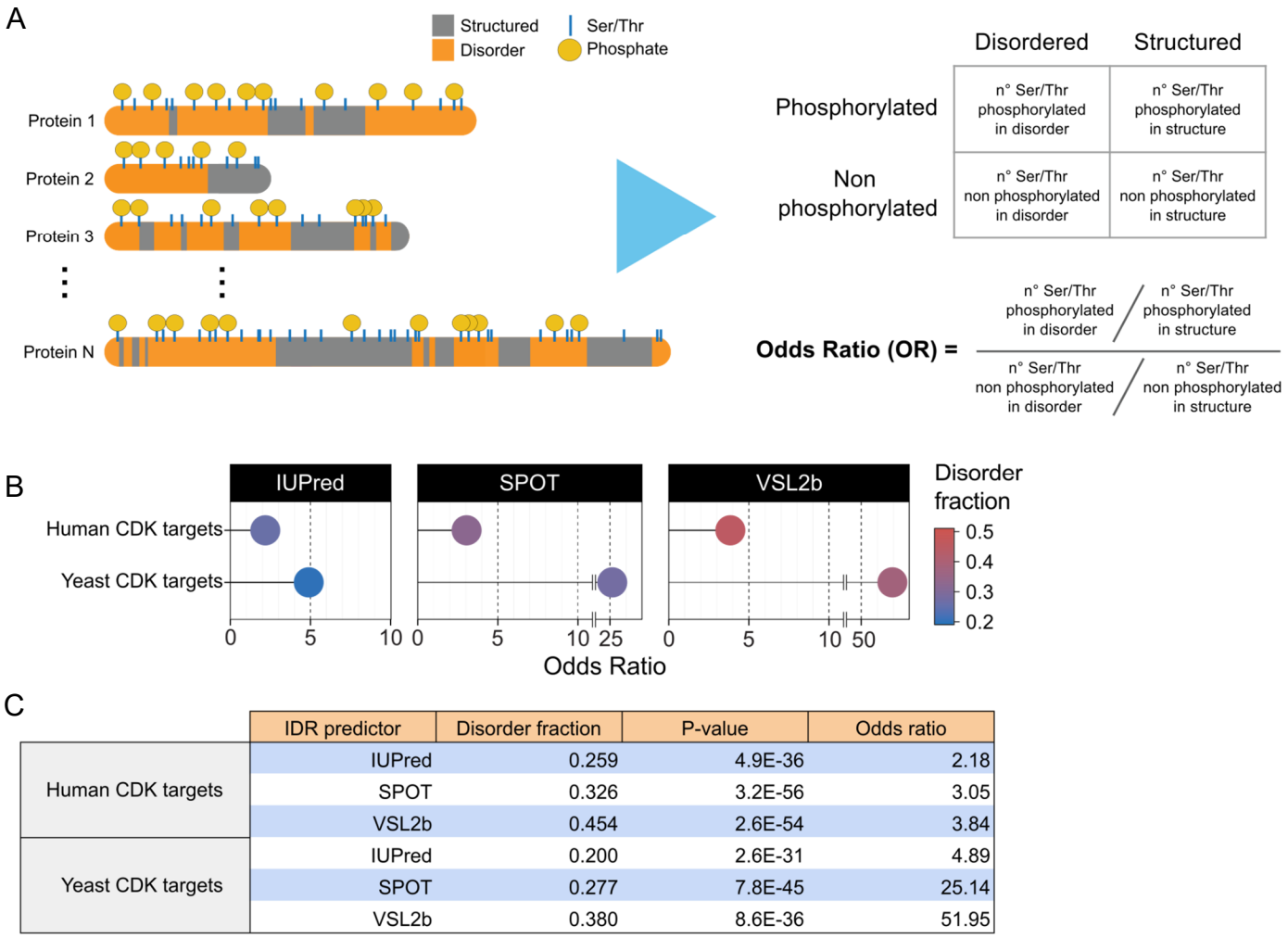


Figure 3.3 | Phosphorylation enrichment in IDRs. (A) Illustrative scheme of the distribution of phosphorylatable residues (blue lines) in structured (gray) and disordered (orange) regions. The expected values of phosphorylations for each type of region depend only on the abundance of phosphorylatable residues, rather than the length of those regions, symbolized here as equal lengths. (B) Scatter plots of the observed phosphorylations in Ser and Thr vs the expected value for each protein considering a binomial distribution. A binomial test was performed to establish statistical significance, and the resulting p-values were adjusted by the Benjamini-Hochberg method (FDR). Two thresholds of 5% and 1% false discovery rate were highlighted with yellow and red colors, respectively. CDK targets are shown with filled circles, while the rest of the phosphoproteins are represented by crosses.

3.1.4 Cell cycle phosphorylation is particularly enriched in IDRs

Next, I focused on CDK-mediated phosphorylation. I sought to answer if CDK phosphorylation specifically was enriched in disordered regions and how it can be quantified. For each CDK target identified in yeast and human datasets, I counted the number of phosphorylated and non-phosphorylated and Ser and Thr in both structured and disordered regions. Using these data, I constructed a two-by-two contingency table (Fig. 4A) employed to perform a Fisher exact test and obtain a p-value and an estimation of the sample odds ratio for evaluating the effect-size. For the three predictors in both organisms, p-values showed extreme statistical significance that the estimated odds ratio is greater than 1, which implies a tendency of phosphorylated Ser and Thr to be located in IDRs (Fig. 4 C; D). Unfortunately, the odds ratio is a mathematical expression which does not have an interpretation in colloquial language. However, for low-prevalence events, as in this case (most Ser and Thr are not phosphorylated), the odds ratio can be approximate to the risk ratio (also called, relative risk), which can be interpreted. This value indicates the fold-increase in the risk of an event happening under certain conditions. In this case, if we take the value 3.05 from the risk ratio of human CDK targets using SPOT predictor, that would mean that a phosphorylatable amino acid have 3 times the risk of being phosphorylated if it is located in a disordered region, compared with those in structured regions.

Yeast CDK targets present extreme odds ratio values for two predictors (Fig. 4 C; D). A closer inspection of this analysis did not show any error and the most probable cause of this abnormality might be the stricter selection of CDK targets. I conducted further exploratory analysis which indicated that when CDK substrates are defined by the *in vivo* datasets, instead of the intersection, the values of the odds ratios approximate those observed for humans (data not shown).



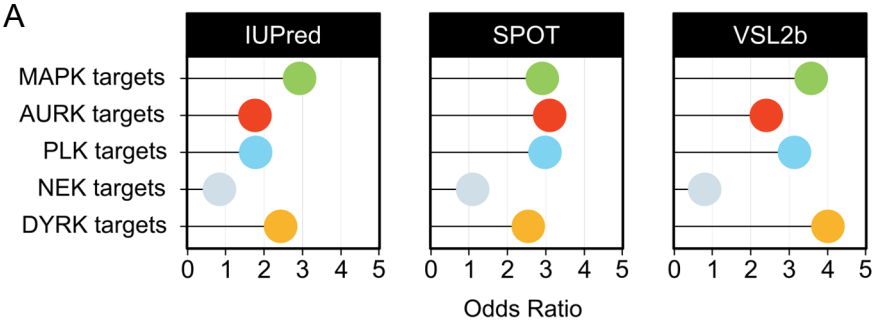
C

	IDR predictor	Disorder fraction	P-value	Odds ratio
Human CDK targets	IUPred	0.259	4.9E-36	2.18
	SPOT	0.326	3.2E-56	3.05
	VSL2b	0.454	2.6E-54	3.84
Yeast CDK targets	IUPred	0.200	2.6E-31	4.89
	SPOT	0.277	7.8E-45	25.14
	VSL2b	0.380	8.6E-36	51.95

Figure 3.4 | CDK-mediated phosphorylation in IDRs. (A) Scheme showing the process to generate the contingency table and the equation for calculating the odds ratio. (B) Lollipop plots showing the sample odds ratio, estimated by the Fisher test (conditional Maximum Likelihood Estimate -MLE-) for each predictor in both sets. The color illustrates the calculated disorder fraction of the entire proteome, highlighting different levels of permissiveness of different predictors for calling a region as disordered. (C) The table with the detailed results of the Fisher test for all the cases evaluated.

We wondered whether this enrichment in IDRs was a characteristic behavior of CDK-mediated phosphorylation or if, on the contrary, other kinases presented the same trend. I scrutinized the Phosphosite Plus database to retrieve data for other human cell cycle-related kinases and obtained the targets of Aurora kinases (AURK, 412 targets), Polo-like kinases (PLK, 459 targets), NimA-related kinases (NEK, 43 targets) and dual-specificity yak-related kinases (DYRK, 42 targets). I also extracted substrates for the mitogen-activated protein kinases (MAPK, 408 targets), as a control. This family of kinases are not directly associated with the control of the cell cycle processes but, like CDKs, they phosphorylate proline-directed sites.

A



B

	IDR predictor	Disorderf raction	P-value	Odds ratio
Human MAPK targets	IUPred	0.259	6.2E-50	2.92
	SPOT	0.326	1.6E-45	2.91
	VSL2b	0.454	1.1E-45	3.57
Human AURK targets	IUPred	0.259	7.7E-15	1.77
	SPOT	0.326	6.0E-40	3.10
	VSL2b	0.454	1.6E-20	2.41
Human PLK targets	IUPred	0.259	1.0E-19	1.78
	SPOT	0.326	2.3E-51	2.98
	VSL2b	0.454	7.1E-42	3.14
Human NEK targets	IUPred	0.259	5.0E-01	0.84
	SPOT	0.326	7.4E-01	1.10
	VSL2b	0.454	3.5E-01	0.80
Human DYRK targets	IUPred	0.259	3.7E-04	2.43
	SPOT	0.326	3.2E-04	2.54
	VSL2b	0.454	6.2E-05	4.01

Figure 3.5 | Cell cycle kinases phosphorylation in IDRs. (A) Lollipop plots showing the sample odds ratio, estimated by the Fisher test (conditional Maximum Likelihood Estimate) for each predictor in datasets of substrates of human kinases. The color illustrates the calculated disorder fraction of the entire proteome. (B) The table with detailed results of the Fisher test for all the cases evaluated in A.

All these datasets have not been manually curated, and therefore their accuracy in terms of representing the real set of targets might be lower than for CDK. Nevertheless, except for NEK, all kinases showed comparable results to those observed for CDKs (Fig. 5A; B). At these levels of statistical significance, it seems clear that all cell cycle-related kinases, as well as MAPK, are prone to phosphorylate sites located in disordered regions of the proteins. It is not evident, however, how relevant the difference between the effect-size (odds ratio) of these kinases is, considering the extreme dependency of these values in the curation of the input data, as it was reported for yeast.

In summary, these analyses showed a significant propensity of phosphorylation sites of cell cycle related kinases to be located in disordered regions. This behavior is probably shared by other kinases, as it was illustrated by the similar results obtained for MAPK.

3.1.5 CDK targets are more disordered than the average phosphoproteome

Disordered regions have several properties by which their enrichment in phosphorylations can be explained, as previously discussed. Compositional bias towards phosphorylatable residues, solvent exposure and structural flexibility are among the most relevant. Then, if our hypothesis that CDK activity, and probably other kinases as well, act as a global force for the cell reorganization throughout the cell cycle, evolution will probably result in the expansion of disordered regions enriched in accessible phosphorylatable residues. This will be in contrast with an evolutionary model that selects specific phosphorylation sites in the protein sequence. If this is true, there should exist a noticeable difference in the percentage of disordered amino acids between a CDK target - or another cell cycle-related kinase substrate - and the rest of the phosphorylated proteins.

To test this theory, I calculated the percentage of disordered amino acids for each protein of both phosphosproteomes - yeast and human - with SPOT, IUPred and VSL2b. I subdivided this set in two subsets: CDK substrates (CDK1 targets) and the rest of the phosphoproteome (non-CDK1 targets). I then used a non-parametric hypothesis test to assess the statistical significance of the difference between these distributions, and the results are represented by a combination of violin and box plots (Fig. 6A). In all cases evaluated, CDK targets have a significantly greater percentage of disordered amino acids than the rest of the phosphoproteome. For humans, the effect-size ranges from 18 to 21 percent approximately, while in yeast, this difference is accentuated ranging from 25 to 28. That means an around two-fold increase in the percentage of disordered amino acids for CDK targets when compared with the rest of the phosphorylated proteins.

A similar approach was used for all human kinases datasets, and an ANOVA analysis and Dunn post-hoc pairwise tests were applied to compare all the distributions (Fig. 6A).

The substrates of all the cell cycle kinases evaluated, except for the substrates of NEK, showed a statistically significant increase in disorder when compared with the rest of the phosphoproteome. Remarkably, MAPK targets did not show the same distribution as the cell-cycle related kinases and the median is clearly smaller. Even if MAPK-mediated phosphorylation seems to be enriched in IDRs in the same extent as kinases that directly control cell cycle processes, the proportion of disordered regions in their targets is seemingly smaller than for substrates of CDK, AURK, PLK or DYRK.

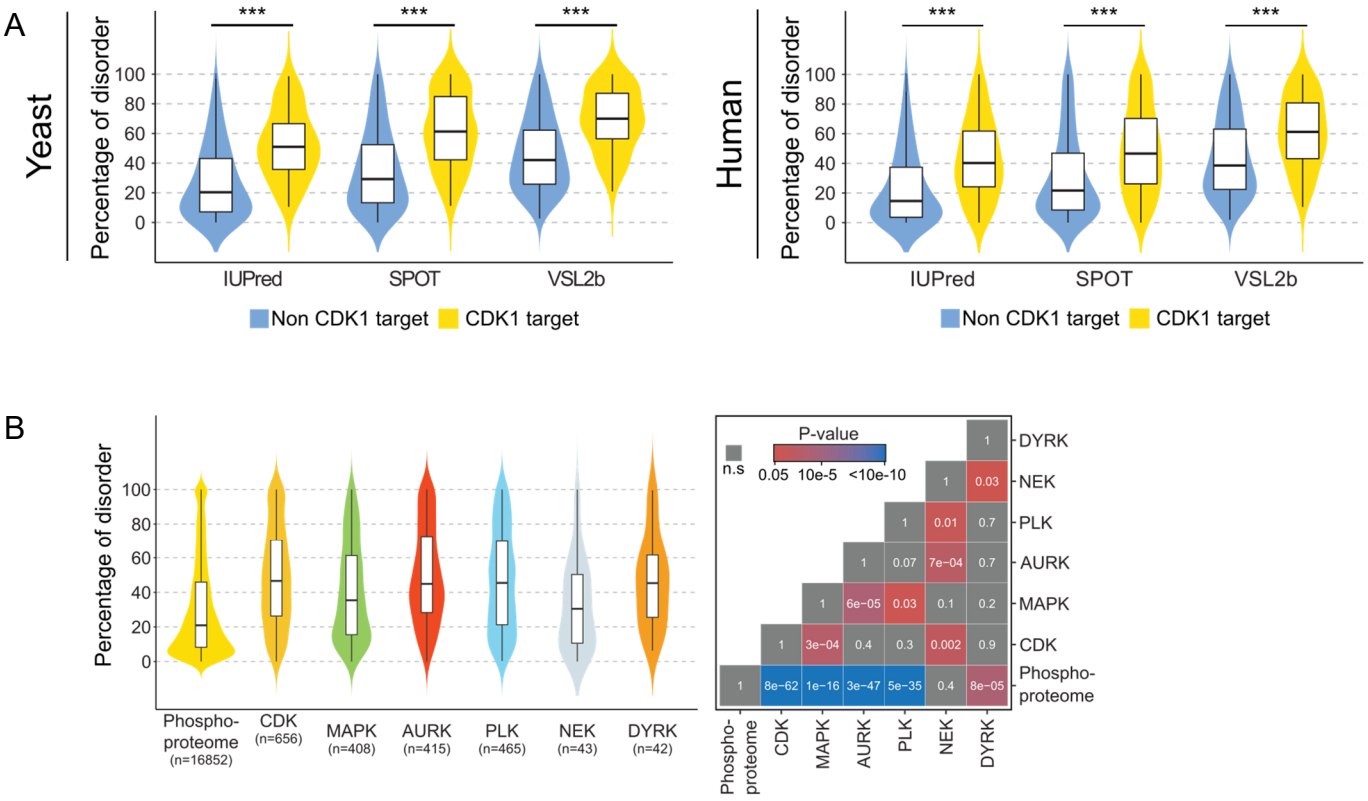


Figure 3.6 | Percentage of disorder in phosphorylated proteins. (A) Violin-Box plot illustrating the distributions of the protein-wide percentage of disordered amino acids. For yeast and human datasets, in yellow, the CDK1 targets and, in blue, the rest of the phosphoproteome. Hypothesis testing was performed with the Wilcoxon rank sum test; three stars indicate p-values smaller than the lower detection limit for the method (< 2.2e-16). (B) Left, Violin-Box plot with the distributions of percentage of disordered residues for all the datasets of targets of the indicated kinases and the phosphoproteome. Right, matrix showing the p-values obtained with the Dunn post-hoc pairwise tests. In gray, non-significant test; scale color from red to blue representing maximum and minimum values of p-values, respectively.

3.2 Cell cycle phosphorylations in *Xenopus*

Soon after discovering the first pieces of solid evidence showing that cell cycle phosphorylation was enriched in IDRs, our team got in contact with members of the team directed by Maarten Altelaar (MA) in the University of Utrecht through Puck Knipscheer. Dr. Altelaar himself and a PhD candidate in his team, J. M. Valverde (JMV), were performing advanced phosphoproteomics in *Xenopus* embryos. Dr. Knipscheer recognized the value of the data they were generating and recommended them to contact us for assistance in the interpretation of the results. They, being experts on proteomics of modified proteins, had generated a comprehensive dataset of

phosphorylations during the first cell divisions of *Xenopus* embryogenesis and Daniel Fisher (DF) and Liliana Krasinska (LK), two highly experienced cell cycle scientists, were appointed to help elucidate and interpret these data. The first recommendation from doctors Fisher and Krasinska, was to perform a replication of the first experiment, but using the system of *Xenopus* eggs extracts. In this technique, interphase extracts are spiked with sperm chromatin *in vitro,* triggering the molecular machinery for DNA replication and simulating the first S-phase of the fertilized egg in a controlled setup. In this way, the battery of phosphorylations related with DNA replication and repair systems could be detected and the results could be then compared with those obtained in embryos.

I was, at this moment, collaborating in a project aimed to understand CDK-mediated phosphorylation as a regulator of DNA repair mechanisms, and I was invited to participate in a meeting where J. M. Valverde presented us with some preliminary results. They had performed a time course of quantitative mass spectrometry (MS) on *Xenopus* single embryos during their first divisions, and by simple clustering methods they had observed patterns consistent with cell cycle dependent phosphorylations. The importance of these results should be properly acknowledged, since they not only achieved single-embryo phosphoproteomics measurements but their results were sensitive enough to detect oscillating patterns throughout the time course. Surprisingly, they had recognized a particular behavior for those phosphorylations: they seemed enriched in disordered regions of proteins.

We decided, then, to establish a collaborative project in which we aimed to merge all our expertise by integrating the *Xenopus* data into our bioinformatic pipeline and, together with J.M. Valverde, developing new tools to obtain an exhaustive interpretation of cell cycle phosphorylations as a global control system.

3.2.1 Dynamics of cell cycle-regulated phosphorylation in *Xenopus*

The system of fertilized Xenopus embryos was selected to perform a temporal map of cell cycle phosphorylation, initially, by technical reasons. Currently the methods for phosphopeptide enrichment are not sensitive enough to allow single-cell phosphoproteomic approaches to be feasible. Sufficient material can be retrieved by using an extremely sensitive phosphopeptide enrichment strategy (Post et al., 2017) to perform quantitative phosphoproteomics on the highly synchronous early cell cycles of *Xenopus* laevis embryos, which consist solely of S and M-phase (Newport and Kirschner, 1982; Newport and Kirschner, 1984).

This technical decision had major conceptual implications, since the chosen experimental setup allowed us to collect phosphorylation information with a sufficient temporal resolution to explore

the phosphorylations occurring within a cell cycle time frame. These results would not, otherwise, be possible to obtain with bulk phosphoproteomics, artificial synchronization methods or cell sorting approaches.

We investigated cell cycle-regulated phosphorylation in this unperturbed *in vivo* system by analyzing individual *Xenopus laevis* embryos undergoing highly synchronous cell cycles of early development. JMV collected single embryos at 18 time-points separated by 15-minute intervals, while recording visual cues of cortical rotation of fertilized eggs and subsequent cell divisions. He then purified the phosphopeptides from each embryo, separated them by nano-LC and analyzed them by high-resolution mass spectrometry (Fig. 7A). This procedure identified 4583 phosphosites with high localisation probability (>0.75) mapping to 1843 proteins, the majority being phosphoserines. Each time point was analyzed in triplicates and individual embryo phosphorylation states strongly agreed between replicates, demonstrating their synchrony and the robustness of our methodology. We thus generated a cell cycle map of protein phosphorylation from an unfertilized egg to a 16-cell embryo.

We focused on 1032 sites on 646 proteins showing at least one time point in which a variation of the signal was statistically significant (ANOVA, Benjamini-Hochberg correction, FDR 0.05) and we called them “dynamic phosphosites”. Proteins containing at least one dynamic phosphosite are hereafter denoted “dynamic phosphoproteins”.

Gene ontology (GO) and network analysis revealed high functional association and interconnectivity between groups of proteins involved in RNA binding and the nuclear pore complex, DNA replication and chromatin remodeling, and microtubule regulation (Fig. 7B).

Hierarchical clustering of the time-resolved phosphorylation signal for the dynamic phosphosites revealed four distinct groups that appear to reflect cell cycle-regulated behavior (Fig. 7C). The levels of cluster A and B phosphosites were highest in eggs and post-fertilization, and decreased during the first round of DNA replication, coincident with cortical rotation. This behavior reflects the transition from meiosis to mitosis and suggests that dephosphorylation of these sites may prepare the zygote for upcoming cell divisions (Clift and Schuh, 2013).

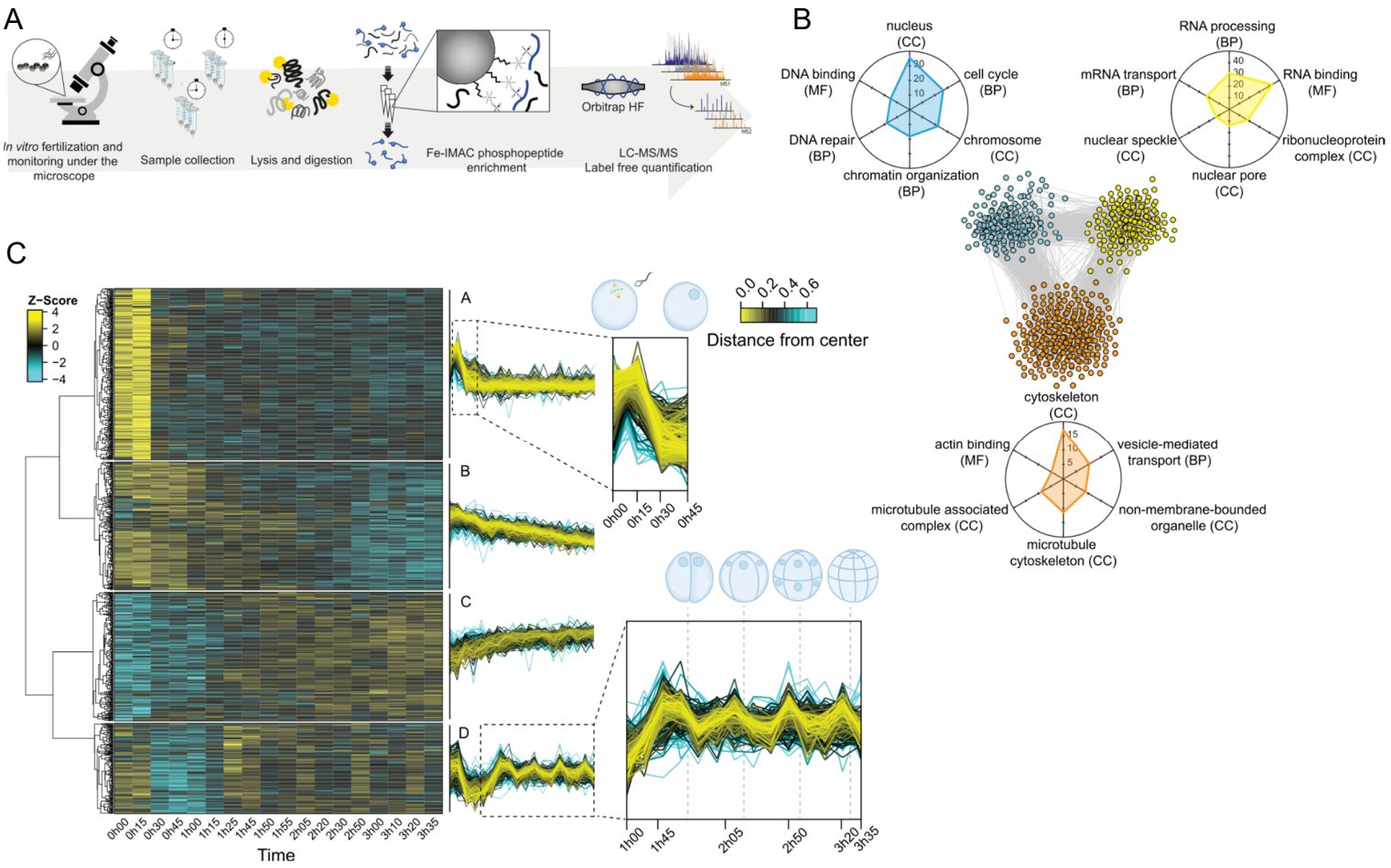


Figure 3.7 | *Xenopus* embryos phosphorylation dynamics. (*Author: JMV*) (A) Schematic representation of the workflow. Single *Xenopus* eggs and embryos were collected followed by cell lysis, protein digestion, phosphopeptide enrichment and high-resolution proteomics analysis. (B) STRING network of functionally associated proteins undergoing dynamic phosphorylation (each node represents a protein). Vicinity clustering reveals three main groups (yellow, blue and orange) with a high degree of association. Radar plots show the corresponding GO terms (adjusted p value <0.05) for each group (axes show -Log10(adj p-value) for each GO term). (C) Heatmap showing the normalized abundance (Z-scores) of each dynamic phosphosite (Y axis) during time (X axis). Hierarchical clustering of phosphosites reveals 4 clusters with distinct regulation (A-D). Dashed boxes in clusters A and D are zoomed-in to highlight dynamic phosphorylation patterns (dashed lines depict the time points of cell division).

GO analysis for group A highlighted proteins involved in RNA regulation and nuclear organization, including the NPC and nuclear transport, chromosomal structure and segregation (Fig. 8A), as also observed in a recent

study on meiosis exit in *Xenopus* eggs (Presler et al., 2017). Cluster B phosphosites were of lower initial intensity and dephosphorylation rate was gradual, and were enriched in regulators of RNA biosynthesis and stability, translation, actin, DNA replication and repair (Fig. 8A). Cluster C phosphosites progressively increased after meiotic exit, showing minor fluctuations over the time

course, while cluster D phosphosites had a clear oscillating signature with upregulation preceding each cell division. GO analysis of cluster C shows dominance of interphase cell cycle processes including DNA replication, RNA-related processes and chromosome organization (Fig. 8A), and included phosphosites displaying a reciprocal oscillating trend and a lower amplitude compared to cluster D sites.

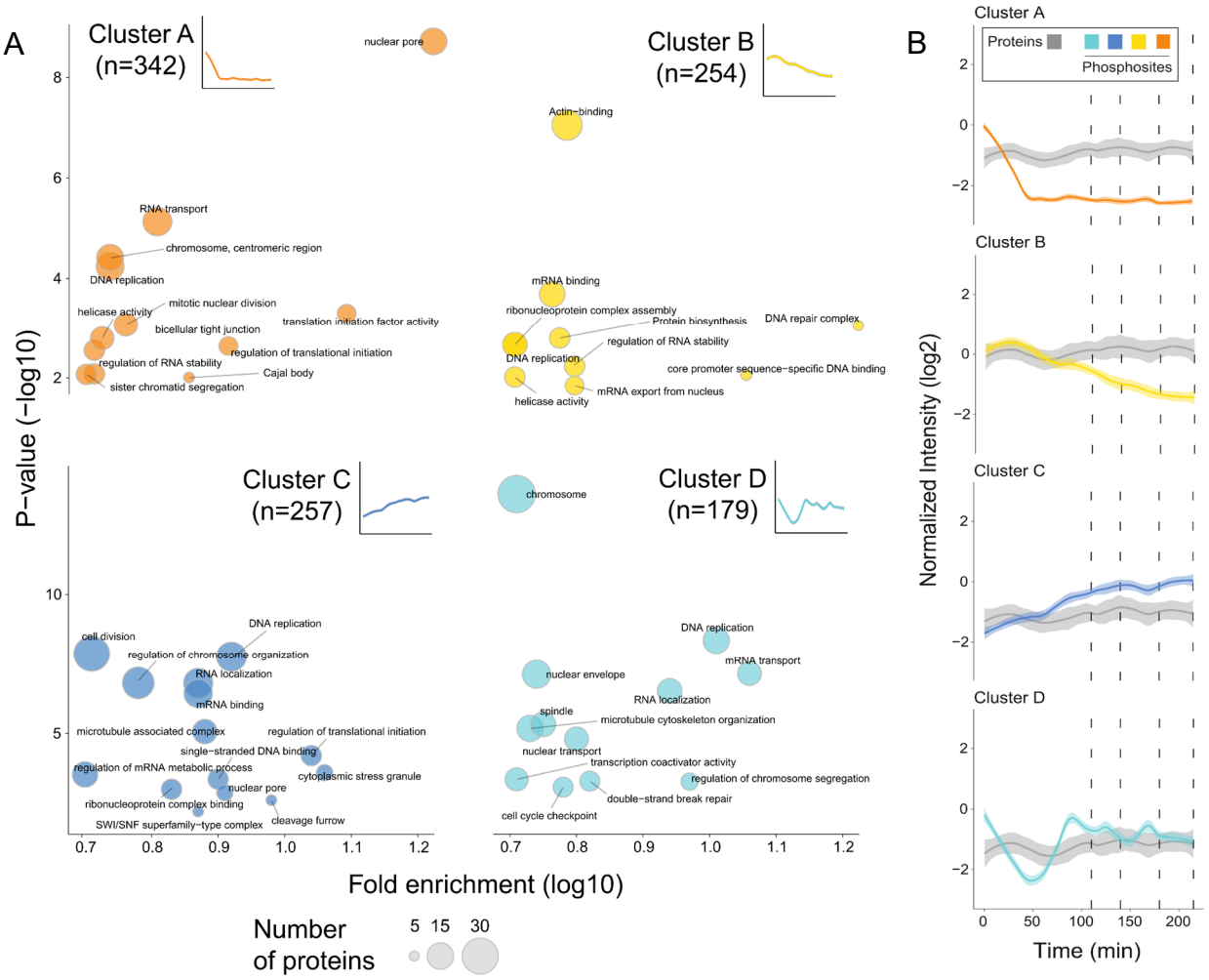


Figure 3.8 | *Xenopus* embryo phosphorylation dynamics. (*Author: JMV*) (A) Correlation coefficients for two randomly selected time points. Scatter plots of significantly enriched (Fisher's exact test with Bonferroni correction, p<0.05) GO (BP, MF, CC, Uniprot keywords) terms for all dynamic phosphosites per cluster in the *in vivo* experiment, presenting the fold-enrichment of specific terms vs statistical significance. The size of the circles correlates with the number of proteins associated with the specific term. (B) Comparison of dynamic variations in total protein compared to total phosphosites from the four clusters shown in Fig. 7C (dashed lines depict the time points of cell division).

In cluster D, coordinated phosphorylation of multiple members of protein complexes involved in diverse processes occurred, suggesting a common mechanism of regulation (Fig. 8A). Importantly, phosphoproteome changes were not simply a reflection of changes in abundance of

the corresponding proteins (Fig. 8B), which are comparatively negligible during *Xenopus* early development (Peuchen et al., 2017).

In cluster D, coordinated phosphorylation of multiple members of protein complexes involved in diverse processes occurred, suggesting a common mechanism of regulation (Fig. 8A). Importantly, phosphoproteome changes were not simply a reflection of changes in abundance of the corresponding proteins (Fig. 8B), which are comparatively negligible during *Xenopus* early development (Peuchen et al., 2017).

We assigned *in vivo* embryo phosphosites to different cell cycle stages by comparing with phosphorylation patterns of replicating or mitotic egg extracts (Fig. 9A). Replication was initiated by adding purified sperm chromatin to interphase egg extracts and quantified over time (Fig. 9B, top), while mitosis was triggered by adding recombinant cyclin B and verified microscopically. We also used egg extracts arrested at meiotic metaphase II (Cytostatic Factor, CSF-arrested). This experiment, performed jointly by LK and JMV, allowed us to identify 6937 phosphosites, which included 71% of the sites identified *in vivo* (Fig. 9C). 1728 sites presented at least one time point with a variation in the z-score (ANOVA, Benjamini-Hochberg correction, FDR 0.05) and using cluster analysis we could identify 6 different behaviors (Fig. 9B).

We grouped clusters 1, 2 and 3 as the M-phase phosphorylations (1035 sites), and clusters 4, 5 and 6 as S-phase phosphorylations (693 sites) (Fig. 9B). A high correlation can be observed between the mitotic and phosphorylations and the extracts arrested at metaphase, highlighting shared functional phosphorylation landscapes in those states, such as high CDK1 activity levels (Hörmanseder et al., 2103). Surprisingly, only 36% of *in vivo* dynamic phosphosites could be mapped to the *in vitro* determined M-phase or S-phase phosphorylations.

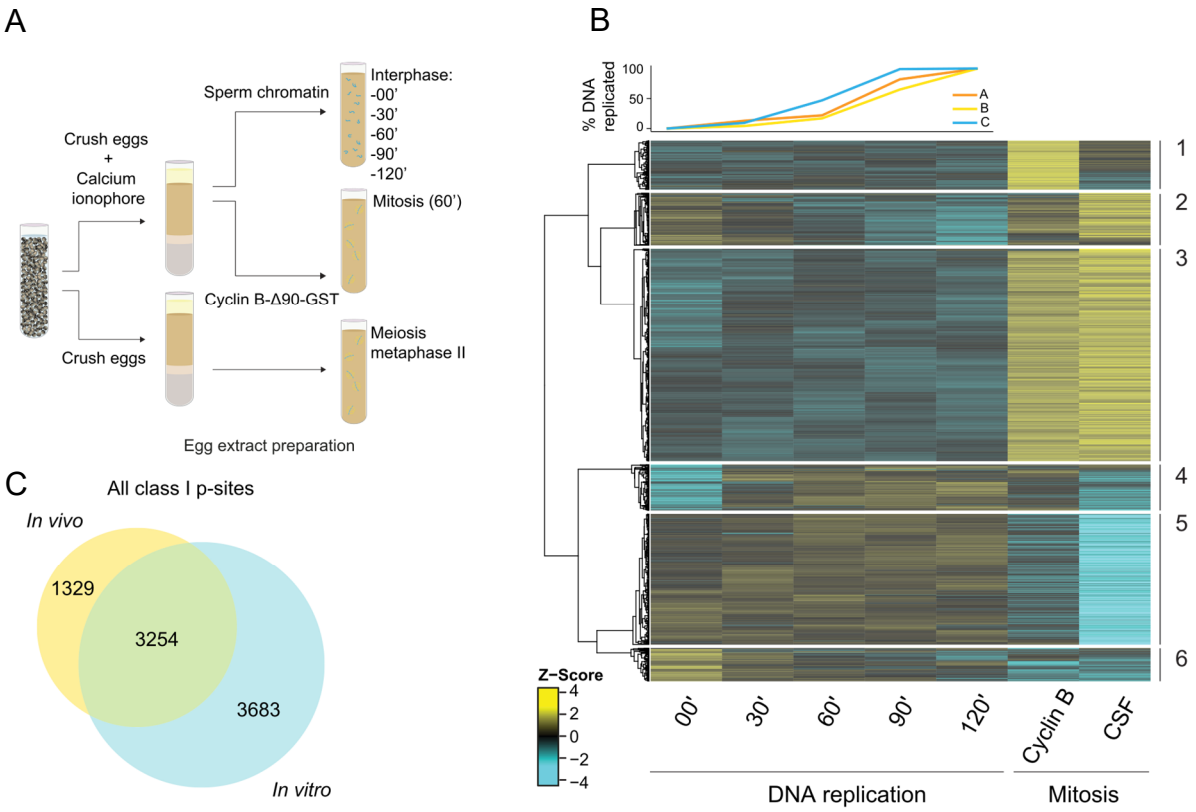


Figure 3.9 | *Xenopus* egg extracts phosphorylation dynamics. (*Author: JMV*) (A) Schematic representation of the experimental design. (B) Heatmap showing the normalized abundance (Z-scores) of each dynamic phosphosite (Y axis) in different time / condition (X axis). Hierarchical clustering of phosphosites reveals 6 clusters with distinct regulation divided in M-phase clusters (1-3) and S-phase clusters (4-6). (C) Venn diagram showing the overlapping phosphosites detected in the single-embryo (*in vivo*) and egg extracts (*in vitro*) phosphoproteomic experiments.

The different clusters from the *in vivo* experiment presented disparate levels of coverage (Fig. 10A), that seems to correlate with how well defined those clusters were. For example, Clusters A and D were well described by phosphorylations at the end of meiosis and cell cycle oscillating phosphorylations respectively. Clusters B and C however showed a less clear behavior. For sites in each cluster that could be assigned *in vitro* to mitosis or S-phase, we calculated the number of sites per 100 sites in the clusters and per 100 sites in each phase (mitosis or S-phase) (Fig. 10B). This double normalization allows us to directly compare this number across all clusters and for both M- and S-phase.

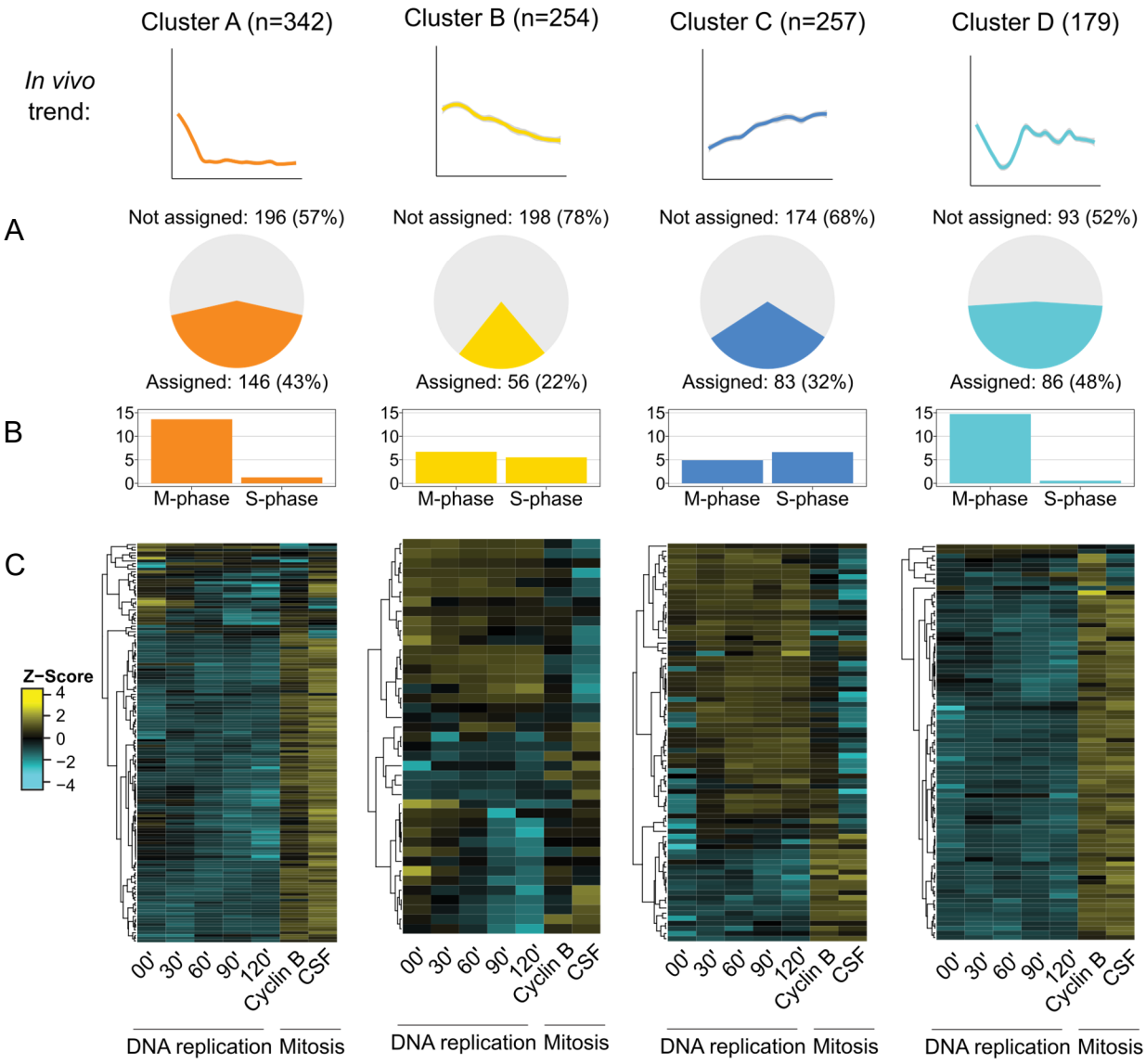


Figure 3.10 | Characterization of phosphorylation clusters. (*Authors: JMV & GD*) (A) Pie charts showing the proportions of each cluster from the *in vivo* experiment that can be assigned to M- or S-phase in the egg extracts experiment. (B) Barplot with the number of sites per 100 sites in the clusters and per 100 sites in each phase. (C) Heatmaps with the *in vitro* phosphorylation data for each of the single-embryo phosphorylation clusters.

Clusters A and D are clearly enriched in mitotic phosphorylations, and while B and C present a more attenuated difference, the latter is the only cluster where an S-phase enrichment can be observed. Finally, to understand in detail the composition of each cluster, we selected those dynamic phosphosites that could be detected *in vivo* and *in vitro* and, for each cluster, we reconstructed the heatmaps with the eggs extract data (Fig. 10C). This provided us with a clear

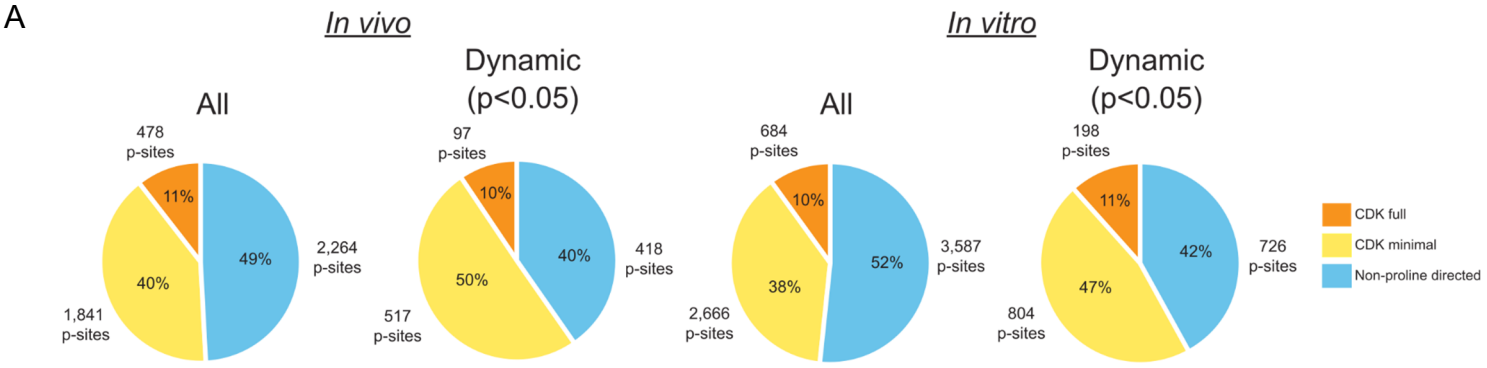
picture of how the clusters from the single-embryo phosphoproteomic data are composed. Egg extracts experiment showed that cluster B have a rather mixed behavior, and given the low coverage we can surmise that this cluster presents phosphorylations more related to developmental signaling processes than to cell cycle regulation. The *in vivo* behavior of clusters A and D was confirmed by the *in vitro* data, but did not allow us to discriminate between phosphorylations occurring at the meiotic exit, and those that oscillated with a maximum peak before each cell division, which had a clear distinction in the single-embryo data. Looking at the egg extracts, some of cluster C phosphorylations seem to have been miss-clustered, particularly those not active during the meiotic exit or presenting lower amplitude oscillations. Nevertheless, sites in cluster C were clearly defined by the *in vitro* data, showing an unequivocal tendency of having an enrichment of S-phase phosphorylations.

3.2.2 Predominance of CDK targets in *Xenopus* dynamic phosphorylation

To identify probable kinases responsible for phosphorylations occurring in these experiments, we analyzed kinase consensus motifs. Around 51% of all the detected phosphosites *in vivo* were proline-directed (S/T-P), thus, conforming to the minimal consensus for CDK sites (Fig. 11A). This proportion increased to 60% among dynamic sites, with around 10% of all phosphosites matching the full canonical CDK1-family sequence motif S/TPxK/R. Phosphosites in replicating and mitotic extracts displayed a similar trend for minimal and full CDK consensus motifs (Fig. 11A). Putative CDK targets dominated all clusters, with over 80% of sites in cluster D *in vivo* and mitotic clusters *in vitro* conforming to at least the minimal CDK motif (Fig. 11A).

While in meiosis, MAP kinases, which have the same consensus motif as CDKs, are likely responsible for a subset of these sites (i.e. those specific to embryo cluster A or CSF extracts), these kinases are inactivated upon meiotic exit and not reactivated during embryogenesis (Ferrell et al., 1991), suggesting that most of the dynamic proline-directed sites are due to CDKs. We extended the analysis to other cell cycle related kinases, and we counted how many sites complied with their consensus motifs (Fig. 11B). Some interesting observations arose from this data, such as the clear enrichment of AURK and DDK motifs for the *in vivo* cluster C and the clusters associated to S-phase in the *in vitro* experiment.

A



B

In vivo

	PLK	AURKA/B	CK1/2	DDK	CDK full	CDK min
Cluster A (n=342)	14	12	97	12	14	205
Cluster B (n=254)	11	27	67	12	16	124
Cluster C (n=257)	9	29	72	31	36	135
Cluster D (n=179)	7	5	51	4	31	150
Total (n=1032)	41	73	287	59	97	614

In vitro

		PLK	AURKA/B	CK1/2	DDK	CDK full	CDK min
M-phase	Cluster 1 (n=161)	23	7	56	5	12	67
	Cluster 2 (n=172)	5	6	62	6	23	114
	Cluster 3 (n=702)	16	23	224	5	123	565
Interphase	Cluster 4 (n=151)	13	52	48	8	6	20
	Cluster 5 (n=433)	15	34	136	67	25	196
	Cluster 6 (n=109)	6	9	36	14	9	40
	Total (n=1728)	78	131	562	105	198	1002

C

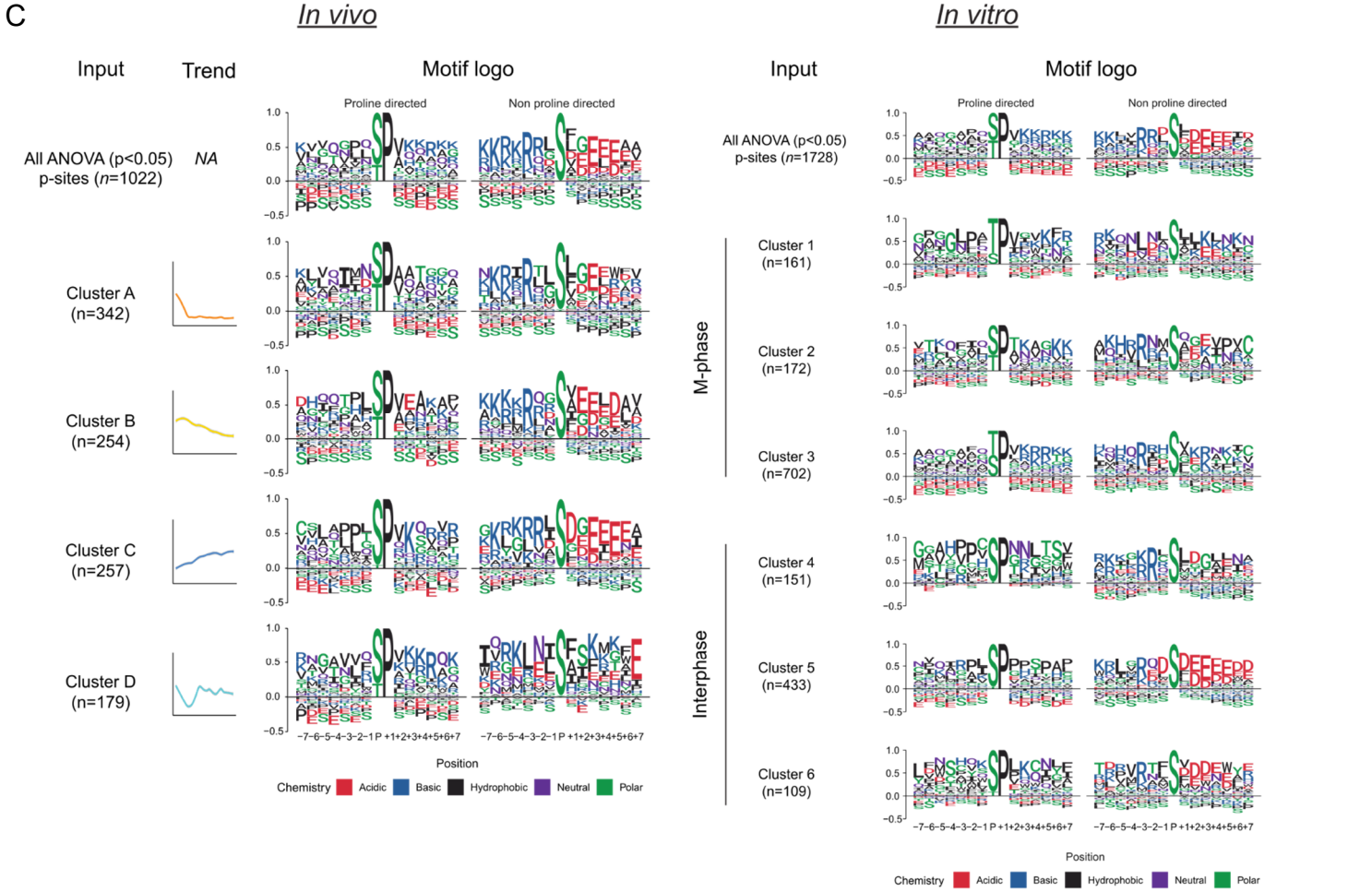


Figure 3.11 | CDK targets dominate *Xenopus* cell cycle phosphorylations. (*Authors: JMV & GD*) (A) Pie charts showing the proportions of phosphosites presenting the CDK full or minimal consensus motifs. (B) Table with the number of sites presenting consensus motifs for cell cycle related kinases in both *Xenopus* experiments. PLK, Polo-like kinase; AURK, Aurora kinase; CK, casein kinase; DDK, Dbf4-dependent kinase. (C) Motif logo analysis for the different clusters in both experiments. Proline-directed sites and non-proline-directed sites are shown separately to better assess the motif patterns in each cluster.

To assess different sequence patterns that might be found surrounding the phosphorylations of different clusters, we aligned the phosphopeptide sequences centered around the phosphosite (Fig. 11C). Proline-directed and non-proline-directed phosphorylations were analyzed separately to have a clearer picture of the sites that do not conform with the minimum CDK consensus motif. *In vivo* data show similar sequence patterns for proline-directed phosphorylations in all clusters, except for cluster A that seems depleted of the positively charged residue in the position +3, which is a feature that distinguishes the CDK full consensus motif from other proline-directed motifs (Fig. 11C). This is probably due to the presence of abundant MAPK-mediated phosphorylation at meiotic exit. Non proline-mediated phosphorylations seem to happen almost exclusively on Ser, and all clusters, except for D, display an enrichment of negatively and positively charged amino acids, downstream and upstream of the phosphorylation sites, respectively (Fig. 11C). This differential behavior of mitotic non proline-directed phosphorylations can be confirmed by observing the motif logos for the M-phase clusters in the *in vitro* data (Fig. 11C, right)

At last, to have a closer estimation of the predominance of CDK targets within the detected *Xenopus* cell cycle phosphorylations, I mapped our *bona fide* set of human CDK targets to the Xenopus proteome. The downstream analysis of the mass spectrometry data required a pipeline that aligns the peptide sequences obtained against a tailored version of the *Xenopus* proteome, to avoid multiple alignments due to its allotetraploid nature. This custom proteome does not have a clear ID correspondence with all the other databases used and, thus, the mapping had to be performed using the BLAST algorithm. An additional step of manual curation allowed me to map 628 of the 656 CDK1 subfamily substrates into the *Xenopus* proteome, representing 5% of the 13576 total proteins (Fig. 12). This proportion increased to 16% and 23% when looking at proteins in our *in vivo* data presenting phosphorylations in general or presenting phosphorylations that changed throughout the time course, respectively (Fig. 12).

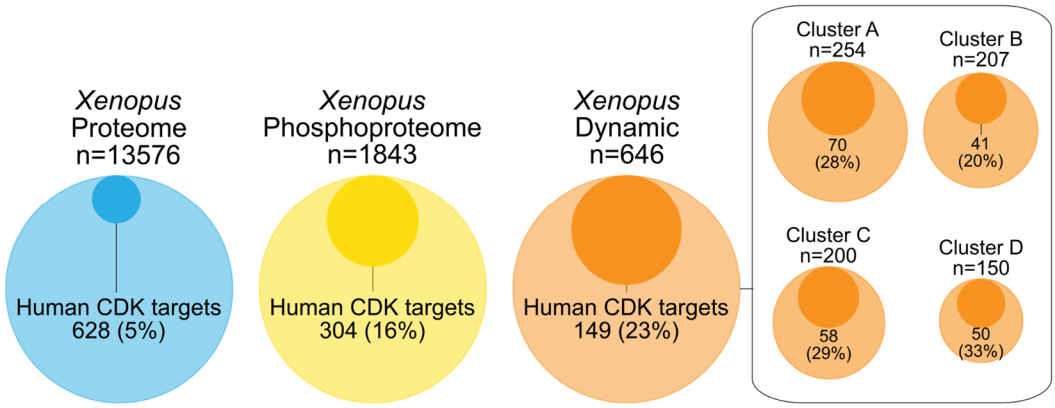


Figure 3.12 | *Xenopus* dynamic phosphoproteins are enriched in CDK targets. Circle plots illustrating the proportion of proteins that could be identified as homologs of CDK targets in humans throughout the different steps of the *in vivo* experiment.

None of the clusters presented a striking enrichment in CDK targets (Fig. 12) and, therefore, for further studies that benefit from an increased number of proteins we chose to use the entire dynamic phosphoprotein data set as a representative set.

3.2.3 Mitotic phosphorylation is switch-like *in vivo*

The dynamics of mitotic phosphorylation are not well understood. While theoretical modeling suggests that it should occur in a switch-like manner due to the bistable mitotic CDK control network (Krasinska et al., 2011), experimental data for CDK-dependent phosphorylations in synchronized cells shows a rather progressive increase throughout S-phase and G2 (Swaffer et al., 2016). We suspect that this behavior may be due to incomplete cell synchronization (Ly et al., 2017) and we estimated that our *Xenopus* single-embryo approach could help to settle this debate. Nevertheless, our first experimental design based on label-free quantification methods did not have the required resolution nor the sensitivity to provide the data needed for answering this question. To see whether mitotic phosphorylation of individual phosphosites is progressive or switch-like *in vivo*, JMV designed heavy labeled peptides as standards to quantify 64 phosphorylation sites selected from cluster D. He analyzed dynamics of these sites in single embryos every 180-seconds using quantitative targeted phosphoproteomics (Lawrence et al., 2016; Schmidlin et al., 2019) by parallel reaction monitoring (Peterson et al., 2012). We thus obtained an extremely high-time resolution quantitative description of mitotic phosphorylation *in vivo* (Fig. 12 A).

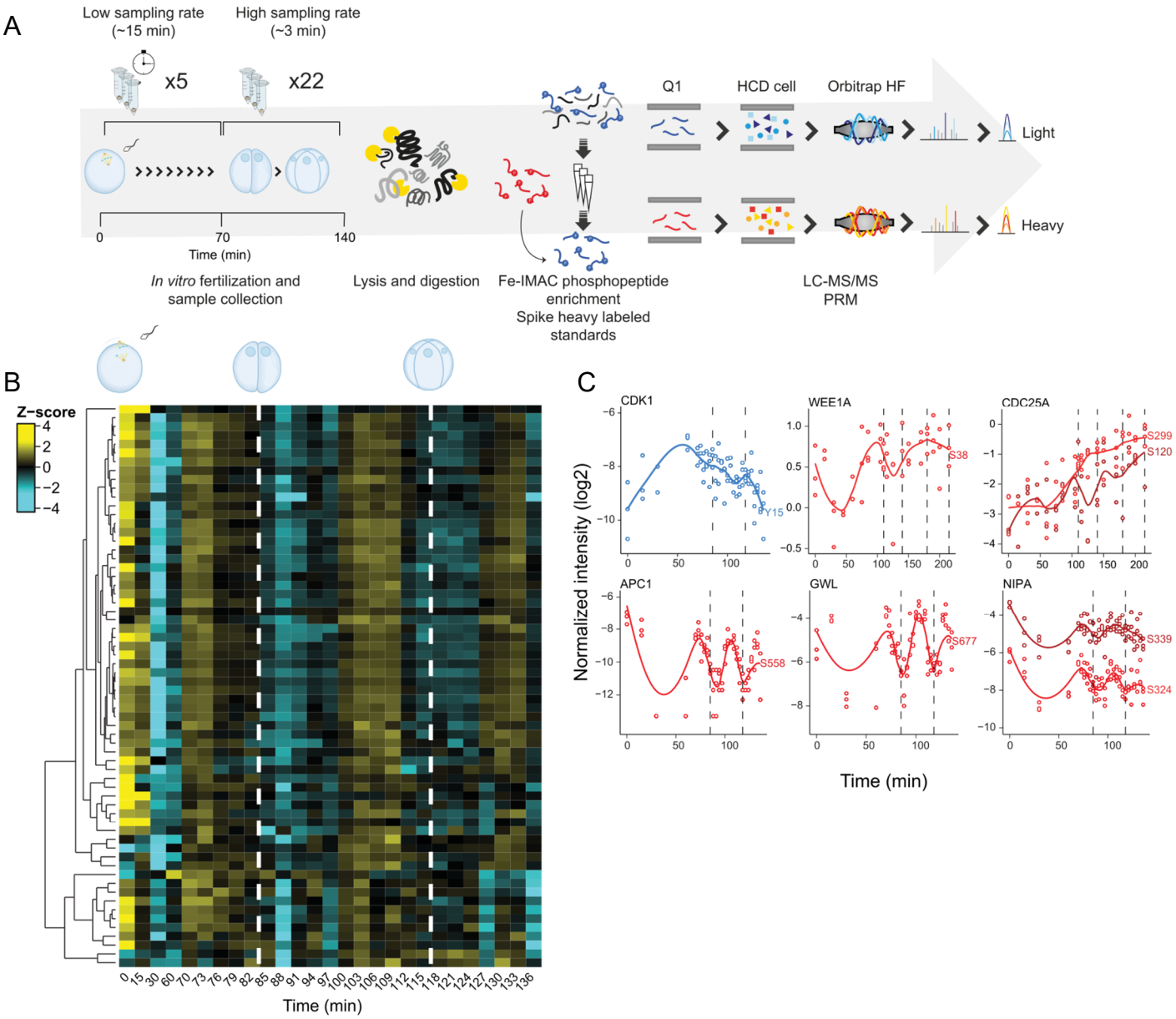


Figure 3.13 | Switch-like mitotic phosphorylation *in vivo*. (*Author: JMV*) (A) Schematic representation of the workflow. Single *Xenopus* eggs and embryos were collected followed by cell lysis, protein digestion, and phosphopeptide enrichment. An extra step of addition of the heavy labeled standards for the quantification took place before the high-resolution proteomics analysis. (B) Heatmap showing the normalized abundance (Z-scores) of each analyzed phosphosite (Y axis) during time (X axis). White dashed lines indicate cell division events. (C) Curves showing individual phosphorylation data of proteins involved in the mitotic entry regulation.

This revealed parallel and abrupt upregulation of all phosphosites preceding each cell division (Fig. 12B), indicating switch-like phosphorylation of diverse protein complexes at mitotic onset. This was not due to oscillation of CDK1-Y15 inhibitory phosphorylation, which was downregulated over time (Fig. 12C, top-left corner), as previously reported (Tsai et al., 2014), consistent with lack of corresponding phosphorylation of the CDK1-Y15-regulatory enzymes, CDC25 and WEE1. In contrast, oscillating phosphorylations on NIPA and the APC/C, which regulate mitotic cyclin

accumulation, as well as Greatwall kinase, which activates the PP2A inhibitors Arpp19/ENSA, were apparent (Fig. 12C). These data suggest that control of mitotic cyclin levels and PP2A activity, and therefore the overall CDK/phosphatase activity ratio, suffices for switch-like mitotic phosphorylation whereas regulated CDK1-Y15 phosphorylation is not essential. This is consistent with the self-sufficiency of futile cycles of opposing enzymes in generating switch-like network output in the absence of allosteric regulation (Goldbeter and Koshland, 1981).

3.2.3 *Xenopus* dynamic phosphorylation is enriched in IDRs

We wanted to understand if the dynamic set of *Xenopus* embryo phosphorylations presented a similar trend to that observed for CDK in yeast and humans of being enriched in IDRs. I repeated the analysis reported above for the CDK target datasets, now with the data obtained from the single-embryo phosphoproteomic experiment. First, I calculated the disorder with the three predictors used in previous analysis (IUPred, SPOT and VSL2B) for the entire *Xenopus* set of proteins. Since this proteome is a non-redundant version specifically designed for the proteomics analysis pipeline, scores for those predictors are not available and had to be calculated with the standalone software in all three cases. Using all detected phosphorylation in the *Xenopus in vivo* experiment I estimated the differential amino acid composition in disorder and obvious similarities to the other organisms could be observed for this dataset (Fig. 14A). As observed for humans, phosphorylation in general seems to be enriched in disordered regions of proteins, as shown by the unanimous trend of phosphoproteins to present more phosphorylations in disorder than what would be expected by chance (Fig. 14B). As we move from the set containing all phosphoproteins to consider only dynamic proteins and dynamic proteins that are also homologs of human CDK targets, the difference between observed and expected phosphorylations increase indicating that dynamic and, more specifically, CDK-mediated phosphorylations drive this tendency (Fig. 14C). The odds ratio values for dynamic phosphorylations in IDRs, when considering the phosphorylatable residues rendered by the Fisher test, were also comparable with what I reported for the human CDK target dataset (Fig. 14D). In an analogous manner, this might be due to an overall increase of disorder in the dynamic phosphoproteins, and, although is not as marked as for CDK targets, the disorder percentage of dynamic phosphoproteins is higher than for the rest of the phosphoproteins detected for all predictors used (Fig. 14E).

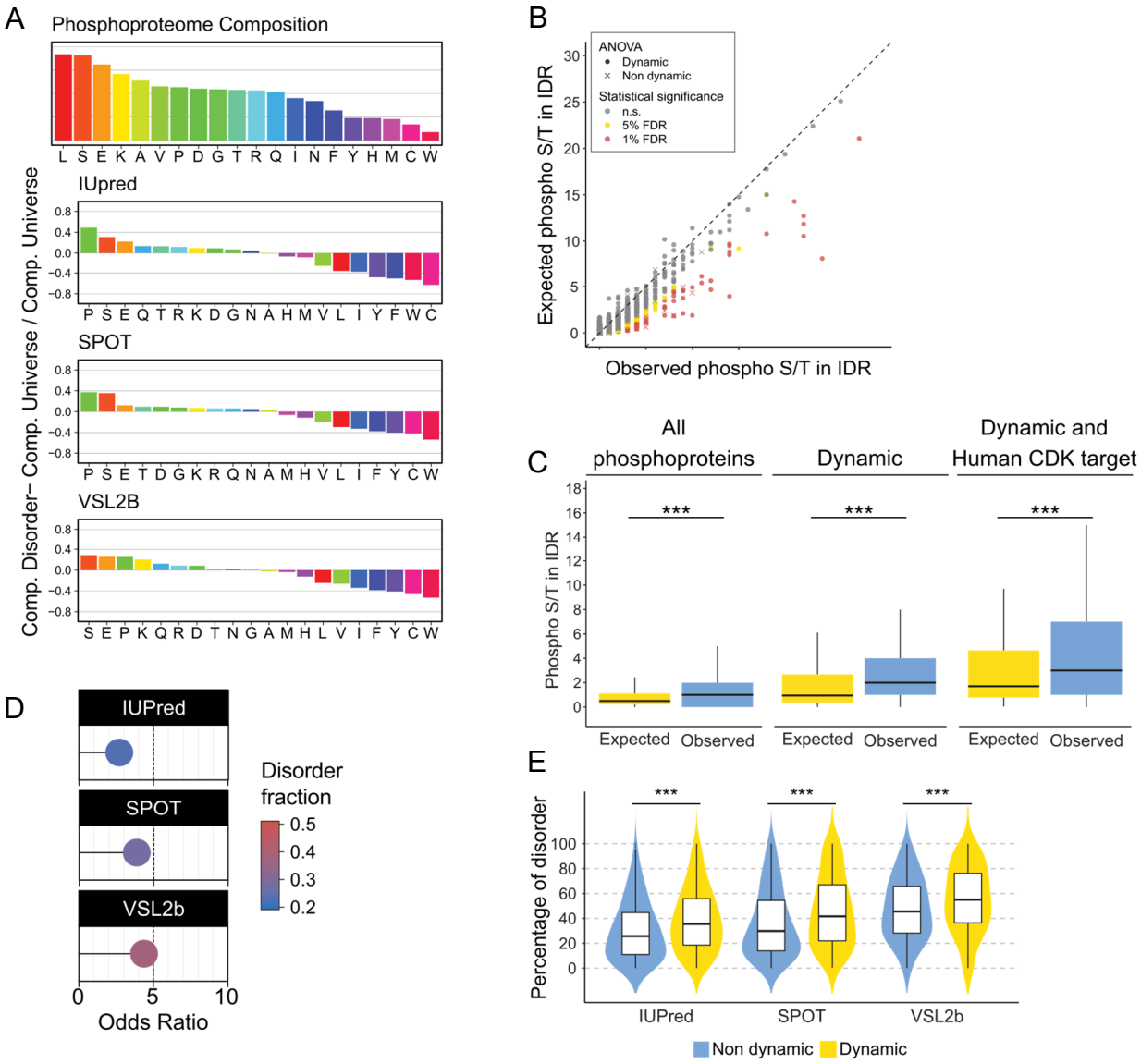


Figure 3.14 | *Xenopus* dynamic phosphorylation is enriched in disorder. (A) The differential amino acid composition in IDRs. Positive values imply enrichment in IDRs while negative values represent amino acids depleted in IDRs (B) Scatter plots of the observed phosphorylations in Ser and Thr vs the expected value for each protein considering a binomial distribution. Statistical significance was assessed by a binomial test (adjusted p-values by Benjamini-Hochberg). Two thresholds of 5% and 1% false discovery rate were highlighted with yellow and red colors, respectively. Dynamic phosphoproteins are shown with filled circles, while the rest of the phosphoproteins are represented with crosses. (C) Box plot with the distributions of the expected vs. observed phosphorylations in all phosphoproteins, dynamic phosphoproteins and dynamic phosphoproteins homologs to human CDK targets. Hypothesis testing was performed with the Wilcoxon signed-rank test; three stars indicate p-values smaller than 1e-10. (D) Lollipop plots showing the sample odds ratio, estimated by the Fisher test (MLE) for each predictor in *Xenopus* dynamic phosphoproteins. The color indicates disorder fraction of the entire proteome. (E) Violin-Box plot with the distributions of the protein-wide percentage of disordered amino acids for *Xenopus*. In yellow, dynamic phosphoproteins and, in blue, the rest of the phosphoproteome. Hypothesis testing was performed with the Wilcoxon rank sum test; three stars indicate p-values smaller than 1e-5.

3.3 Cell cycle phosphorylation regulate protein condensation

Recently, the emergent theory that there exist protein complexes that form condensates by separation of phases has revolutionized biology (Brangwynne et al., 2009). It is proposed that unspecific multivalent interaction between molecules forming those condensates drive their assembly (Banani et al., 2017; Shin and Brangwynne, 2017). One of the hallmarks of biological phase separation is intrinsic disorder (Darling et al., 2018; Uversky, 2017; Shapiro et al., 2021), that not only provides regions enriched in charged and polar amino acids that participate in those interactions, but also the structural flexibility required. It is reasonable then to think that general modifications in those unstructured residues that alter their physicochemical properties can regulate the formation or dissolution of protein condensates. This is particularly interesting if we consider that most membraneless organelles in the cell have now been described to be phase separated and in almost all cases their assembly and disassembly processes are regulated in a cell cycle-dependent manner. Altogether, this information led us to propose a model in which CDK-mediated phosphorylation of disordered regions in proteins acts as a switch-like regulator of the formation and disassembly of MLOs throughout the cell cycle.

We have provided, so far, evidence to conclusively state that cell cycle phosphorylation of disordered regions, specifically by CDKs, occurs in a switch-like manner. Despite the fact that these data suggest an association between this behavior and regulation of phase separation, we did not have direct evidence for supporting this claim. We, therefore, contacted experts in different areas of biology and biochemistry to help us address our theory. In this last section, I will detail how we obtained *in silico* proof for generalizing the idea of CDK phosphorylation regulating phase separation processes and by using the proliferation marker KI67 as a case report we collected experimental *in vitro* and *in vivo* data to confirm our hypothesis.

3.3.1 Enrichment of MLO components among CDK substrates

We first wanted to assess if components of some of the membrane-less organelles confirmed to be phase-separated presented phosphorylations patterns consistent with our theory. We selected a small set of key drivers of different MLOs (Coilin for Cajal bodies; Nucleophosmin and Nucleolin for nucleolus; NUP53 and NUP98 for nuclear pore complex; 53BP1 for 53BP1 bodies; PML for PML bodies) that are also confirmed CDK targets. We then qualitatively observed if CDK-mediated phosphorylation has a specific tendency of being located in IDRs, as calculated by IUPred. We repeated the analysis for the same set of proteins in *Xenopus*. In all cases, we noticed a clear preference of phosphorylation to be located in disordered regions but no obvious

conclusion can be drawn specifically for CDK or dynamic phosphorylation (Fig. 15A). We sought to understand how prevalent these phosphorylations are in proteins constituting different MLOs. By manual data curation, Liliana Krasinska, collected information of proteins forming condensates that have been unequivocally reported as being phase-separated. Thus, we established a high-confidence non-redundant database that can be crossed with phosphorylation data from our previous analysis. Out of the 656 human CDK targets 654 could be uniquely mapped to uniprot IDs compatible with both databases, and 39.2% (257) of those were also present in our MLO dataset, while homologs of the *Xenopus* dynamic phosphoproteins showed a 31.6% (204) overlap (Fig. 15B). Interestingly, when considering proteins that are both CDK targets in humans and dynamic phosphoproteins in *Xenopus* the intersections with the MLO database increase up to 50 % (73 of the 149 proteins) (Fig. 15B).

We then applied a recently-developed machine learning classifier (van Mierlo et al., 2021) to predict whether cell cycle-regulated phosphoproteins, or CDK substrates, show an increase in average propensity for phase separation (defined as PSAP score). Indeed, in the human dataset we observed that the propensity for phase separation is far higher amongst targets of the cell cycle (CDK, Aurora, PLK, again, with the exception of NEK) and DYRK kinases than the overall phosphoproteome, but less so for MAP kinase substrates. We then similarly analyzed *Xenopus* data, which showed a sharp increase in the PSAP score, from the proteome to the phosphoproteome, and a further increase for dynamic phosphoproteins, with the highest score for mitotic cluster D (Fig. 15C). We noted that, unlike the *Xenopus* early embryo phosphoproteome, the human phosphoproteome did not show a striking difference in PSAP score in comparison with the proteome. This suggests that the early embryonic phosphoproteome is already highly enriched in proteins with a propensity to phase separate and in targets of cell cycle kinases.

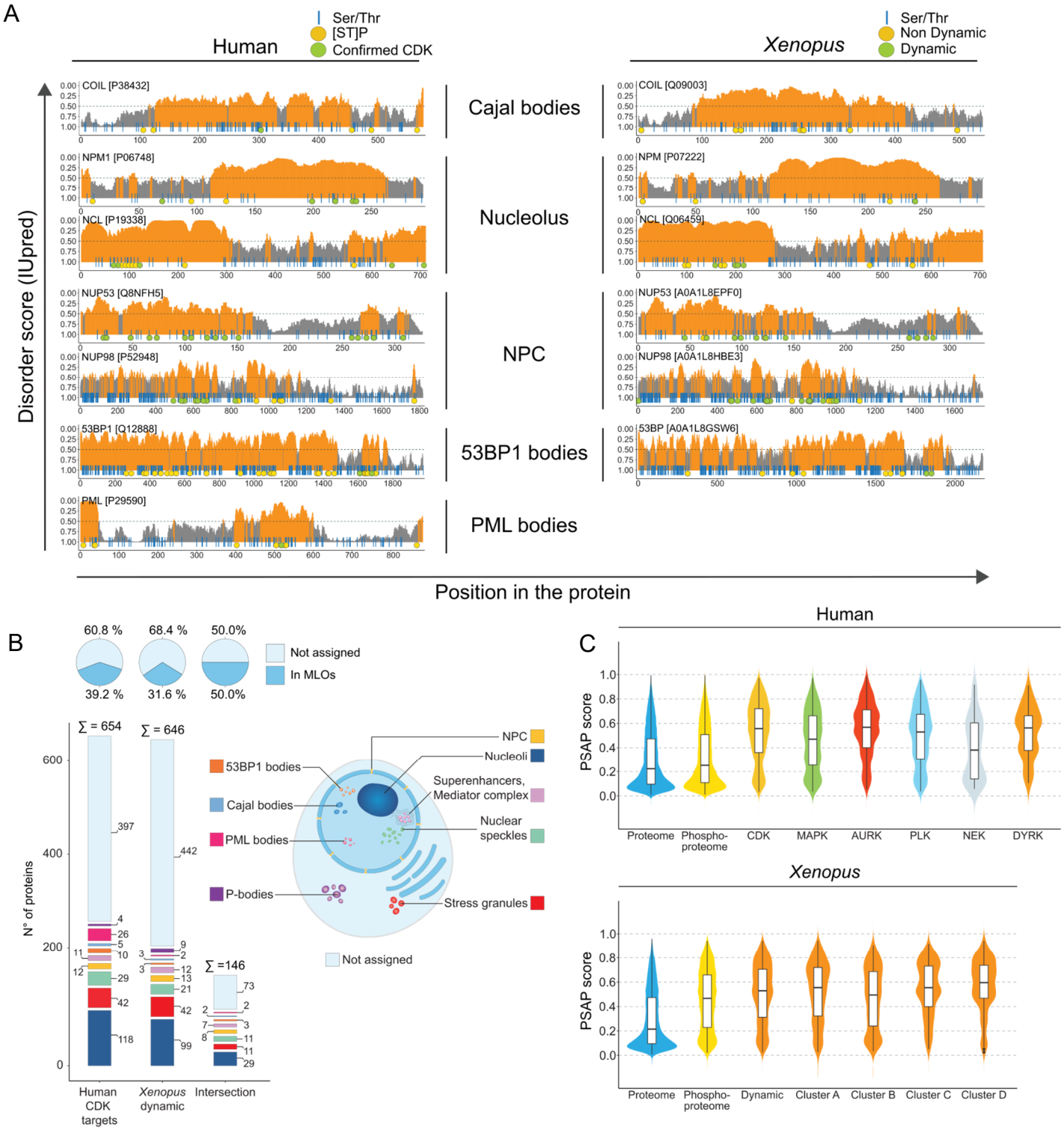


Figure 3.15 | CDK targets are located in MLOs. (A) Diagram of IUPred score over the length of different human proteins and key components of MLOs. Regions with scores >0.5 (orange) are considered to be disordered, and <0.5 (gray) structured. Blue vertical lines indicate Ser and Thr residues; yellow circles, phosphorylated [S/T]P sites in human, and non-dynamic sites in *Xenopus*; green circles, confirmed CDK1 phosphorylations in human, and dynamic phosphorylation in *Xenopus*. (B) Human CDK1 subfamily targets, *Xenopus* dynamic phosphoproteins, and the intersection of both sets, that are present in our manually curated proteome of membraneless organelles. (C) Violin plots presenting PSAP score for human kinase targets (top) and *Xenopus* dynamic phosphoproteins (bottom), in comparison with total proteome and phosphoproteome.

3.3.2 *In silico* proof of CDK-mediated regulation of protein condensation

The next step was to provide direct evidence of CDK-mediated phosphorylation directly changing the physicochemical properties of its targets. We set collaborations with two laboratories specialized in *in silico* modeling of intrinsically disordered proteins and the underlying principles of phase separation.

In the first instance, we used an analytical approach developed by the team directed by Kingshuk Gosh, in the University of Denver, to assess the propensity of phase separation of a selection of IDRs from CDK substrates. This method is based in a general heteropolymer theory that uses sequence charge decoration matrices (SCDM), based on electrostatic pairwise interactions only, to identify intra-chain interaction topology (Huihui and Ghosh, 2020; Huihui and Ghosh, 2021). Since this should correlate with inter-chain interactions that promote PS, SCDMs provide residue-level maps with information of the propensity to phase separate. These maps indicate in color, assigned scores to each combination of pairs of residues reflecting their tendency to interact. Higher levels of inter-chain interactions are indicated with negative scores and the color blue. Thus, the larger areas are colored in blue, the more prone these regions will be to form homotypic interactions. For each IDR selected, we plotted the matrices for the phosphorylated version in the top-right corner and the non-phosphorylated version in the bottom right corner. Of the 12 IDPs tested, 7 (nucleolin, nucleophosmin, NUP53, ELYS, MCM4, 53BP1 and the splicing factor SF3B1) had SCDM maps showing visibly decreased self-association propensity (increased red regions Fig. 16), implying reduced propensity to phase separate, upon CDK-site phosphorylation. Conversely, for SRRM2, CDK-mediated phosphorylation is predicted to increase intra-chain attraction (Fig. 16) and hence PS tendency.

To further analyze these trends, the team directed by Alessandro Barducci at CBS (Centre de Biologie Structurale, Montpellier) calculated the radius of gyration (Rg) of several IDRs using all-atom simulation. This measurement, which polymer physicists adapted from classic mechanics, indicates the average distance of all monomers to the center of mass. Thus, the radius of gyration will reflect how densely packed and entangled a polymer is, and smaller values indicate higher tendencies of phase separation. Effects of phosphorylation on CDT1 (28.4Å to 30.3Å), TICRR (56.2Å to 57.3Å) and coilin (39 Å to 37.9 Å) were minor, while MCM4 IDR expands upon phosphorylation (21.9Å to 26.3Å), consistent with SCDM analysis. Overall, these data suggest that phosphorylation is a key regulator of homotypic interactions, an important element of PS propensity, of most IDRs.

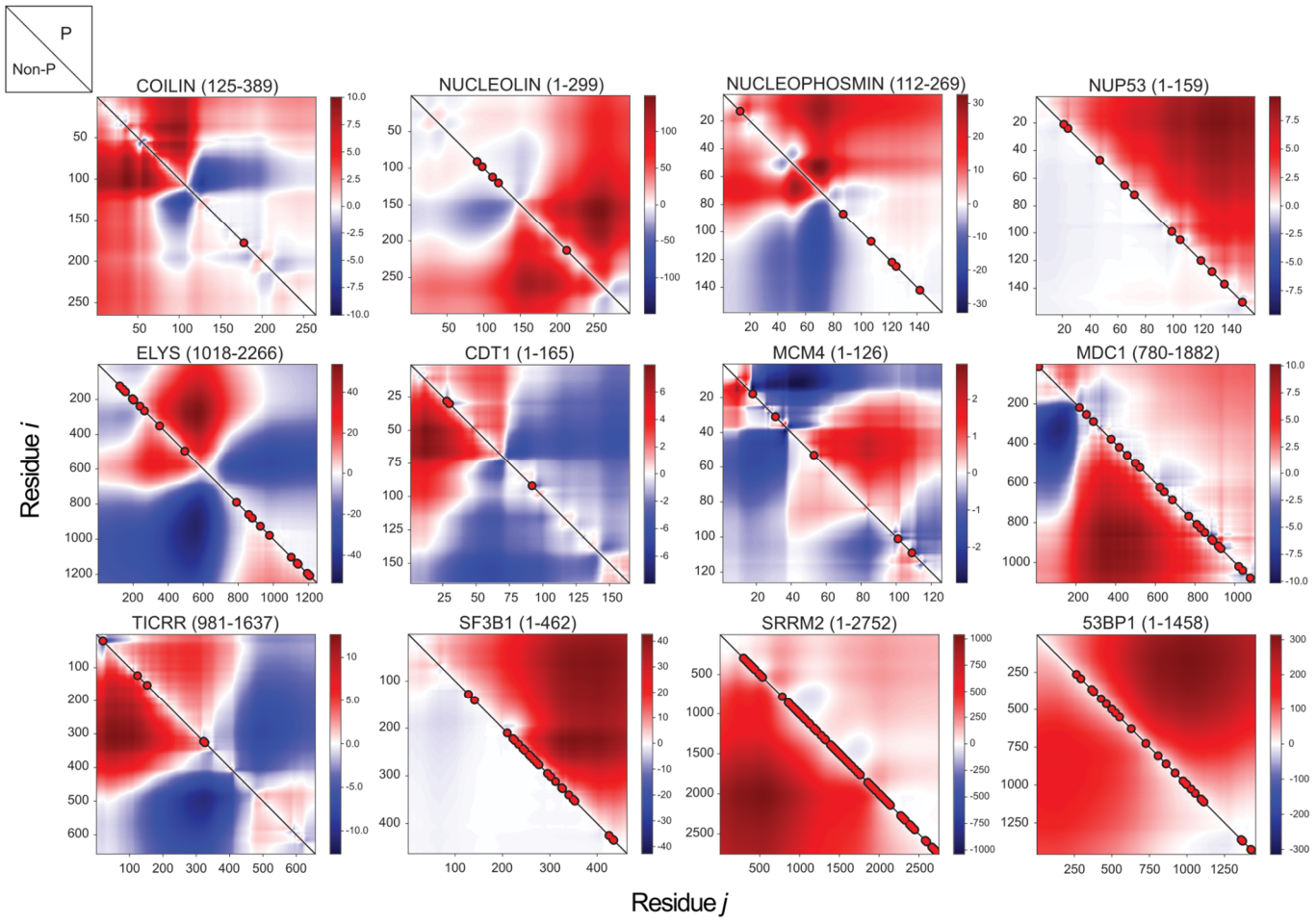


Figure 3.16 | CDK-mediated phosphorylation regulates IDR phase separation propensity (*Authors: Gosh team*). Sequence Charge Decoration Matrix (SCDM) maps for a selection of human CDK targets and major MLO components (IDRs analyzed are indicated in parentheses), depicting the contribution of electrostatic interaction dictating the distance between two amino acid residues i and j (shown in x and y axes). The values of SCDM for different residue pairs (i,j) are shown using color schemes with red and blue denoting positive (repulsive) and negative (attractive) values, respectively. The lower and upper triangles indicate SCDM map for the unphosphorylated (non-P) and phosphorylated (P) sequences, respectively. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are indicated with red circles.

3.3.3 CDK phosphorylation regulates phase separation of KI67

Much of the body of evidence required to rigorously establish that a given protein is subject to phase separation in the cell escapes the proteome-wide level. Experiments with individual proteins, or sections of it, must be carefully designed and conducted specially if an additional mechanism, such as phosphorylation-mediated regulation shall be considered. We thus focused on a model CDK substrate, Ki-67, a highly disordered protein that organizes heterochromatin structure (Sobecki et al., 2016) and perichromosomal layer formation from nucleolar components in mitosis (Booth et al., 2014; Hayashi et al., 2017). Ki-67 contains a multivalent Ki-67 repeat domain that is highly phosphorylated in mitosis by CDKs (Fig. 17A), and which regulates its perichromosomal localization (Hégarat et al., 2020). SCDM analysis predicted that phosphorylation of full-length Ki-67 should promote self-interaction and thus PS, but this cannot be attributed to interactions within the repeat motif alone, since its phosphorylation is predicted to reduce homotypic interactions when performing the simulations in a consensus sequence for the repeats (Consensus repeat, CR) (Fig. 17B; Fig 18A). In agreement, coarse-grained (CG) molecular dynamics (MD) simulations performed by Dr. Barducci's team showed that the radius of gyration of full-length Ki-67 decreased upon phosphorylation (Fig. 17C, left) while that of a single consensus repeat motif increased (Fig. 17C, right). We also used MD simulations to model Ki-67 consensus repeat phase separation dependency on temperature (expressed in relative temperature) and protein concentration, and we identified the characteristic phase diagrams observed from experimental data. Consistent with SCDM analysis, the propensity of phase separation is counteracted by phosphorylation, as observed by the shift of the curve to lower relative temperatures (Fig. 17D). Interestingly, extending the sequence to two repetitions of the CR has an opposite effect suggesting that PS is enhanced by increasing repeat valency (Fig. 17D).

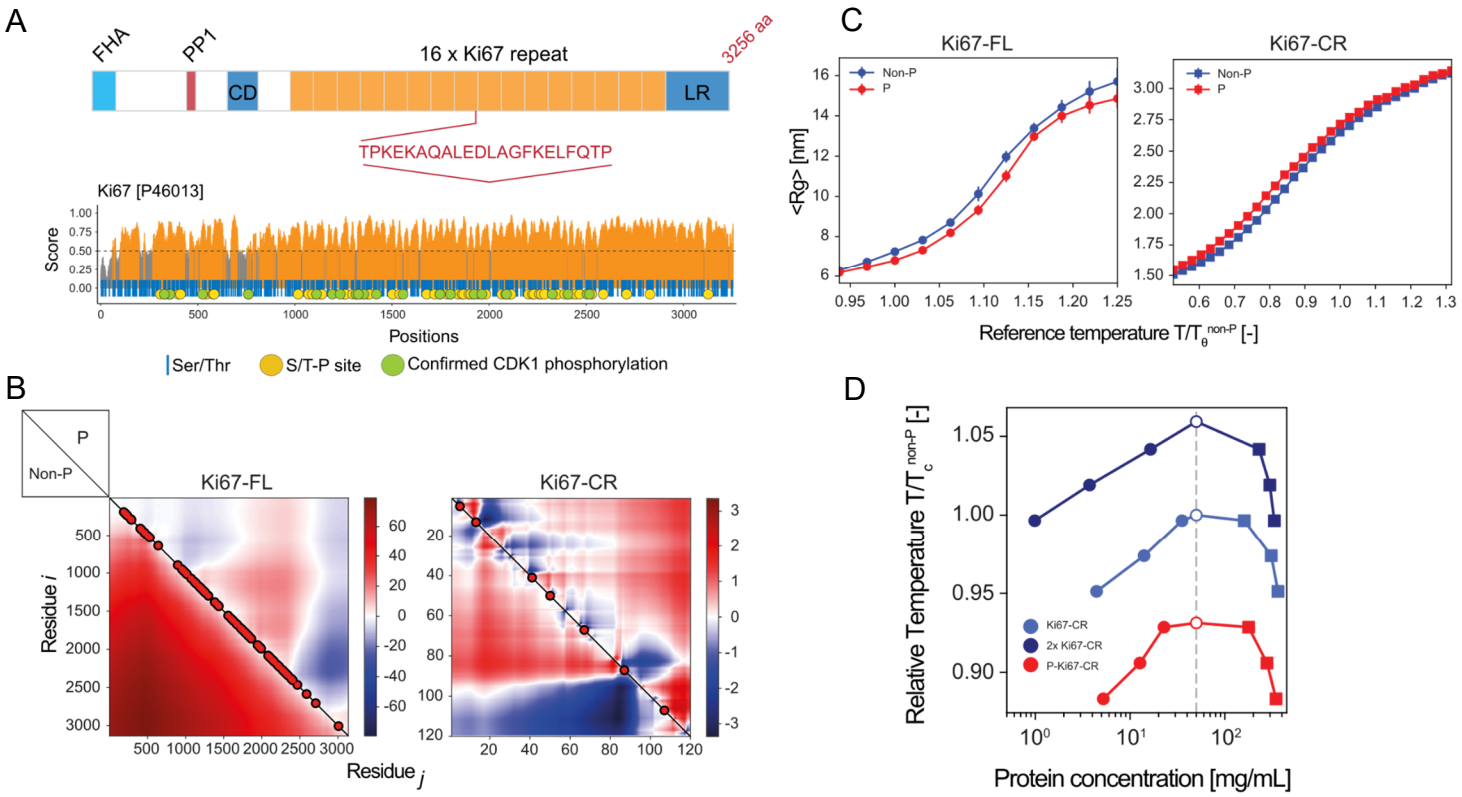


Figure 3.17 | CDK-mediated phosphorylation of KI67 repeat domain regulates its phase separation propensity (*Authors: GD & JMV; Gosh team; Barducci team*). (A) Top, scheme of the human Ki-67 protein (FHA, forkhead-associated domain; PP1, PP1 phosphatase-binding domain; CD, conserved domain; LR, leucine arginine-rich domain). Highlighted, Ki-67 repeat consensus motif. Bottom, diagram of IUPred score over the length of human Ki-67. Scores >0.5 (orange) are considered to be disordered, and <0.5 (gray) structured. Blue lines indicate Ser and Thr residues; yellow circles, known S/T-P sites; green circles, confirmed CDK1 phosphorylations. (B) Sequence Charge Decoration Matrix (SCDM) maps for full length Ki-67 (FL, left) and Ki-67 consensus repeat (CR, right), depicting the contribution of electrostatic interaction dictating the distance between two amino acid residues i and j (shown in x and y axes). The values of SCDM for different residue pairs (i,j) are shown using colour schemes with red and blue denoting positive (repulsive) and negative (attractive) values, respectively. The lower and upper triangles indicate SCDM map for the unphosphorylated (non-P) and phosphorylated (P) sequences, respectively. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are indicated with red circles. (C) Dependency of the radius of gyration (Rg) on the simulation temperature in single-chain MD simulations for full chain Ki-67 (left) and consensus repeat (right). The reference temperature is the θ temperature of the non-phosphorylated molecule for full chain and consensus repeat, respectively. (D) Binodal curves from phase coexistence simulations of the Ki-67 consensus repeat sequence. For each temperature, filled circles indicate the dilute phase density and squares indicate the coexisting dense phase density. Empty circles indicate the fitted critical temperature (Tc) of each system. The Tc of the non-phosphorylated monomer (light blue empty circle) was the reference for the normalization of the temperature values. The light gray dashed line indicates the total concentration used in the simulations.

To test these predictions experimentally, we first used the optogenetic Cry2 “optodroplet” system (Shin et al., 2017) with full length Ki-67 or a series of deletion mutants. Dhanvantri Chahar, a postdoc in our team, generated Flp-InT-Rex 293 opto-Ki-67 stable cell lines containing different inducible versions of KI67, and Emile AlGhoul, a former member of our lab currently working in the team directed by Angelos Constantinou, performed microscopy experiments. Full-length Ki-67 localized to the nucleolus, as expected, but exposure to blue light caused rapid appearance of small round foci in the nucleoplasm (Fig. 18A), which was dependent on the level of induced Ki-67 expression, consistent with PS. Importantly, promoting CDK-mediated phosphorylation by inhibiting PP2A with okadaic acid led to foci formation in the absence of blue light, while pan-CDK inhibition with purvalanol A prevented induction of foci upon light (Fig. 18A). These results indicate that, as predicted by SCDM and MD, phosphorylation of full-length Ki-67 promotes PS. Results were similar for constructs lacking the C-terminal LR domain, that binds chromatin, or the N-terminal domain, which is required for the nucleolar localization of Ki-67 (Fig. 18B, C).
Finally, we synthesized the CR polypeptide (Fig. 19A) and we used it as a substrate for *in vitro* kinase assays with recombinant CDK complexes.

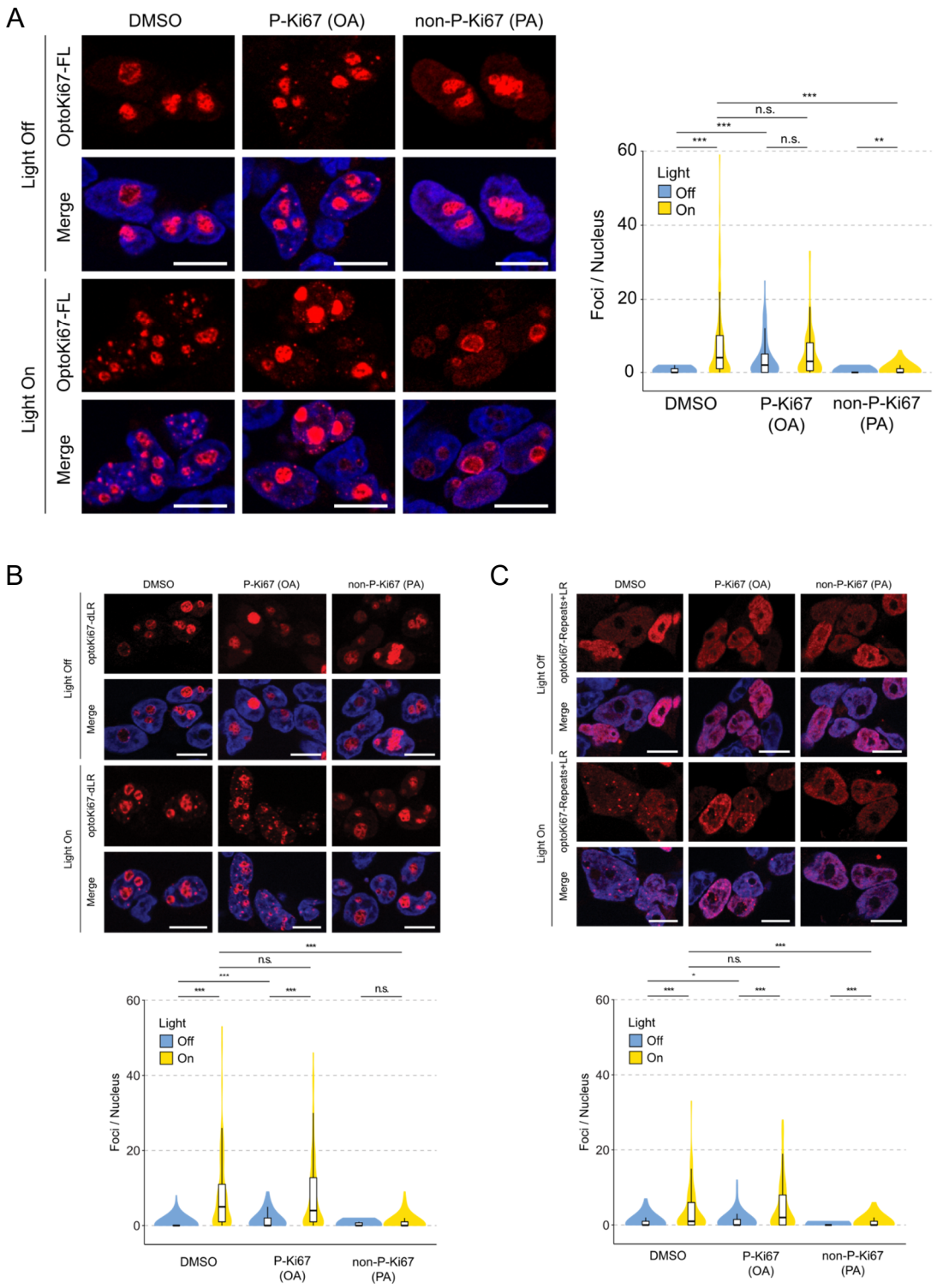


Figure 3.18 | Opto-droplet systems confirmed CDK-mediated regulation of Ki-67 phase separation (*Authors: GD and E. AlGhoul*). (A) Left, representative fluorescent images of HEK-293 cells expressing opto-Ki-67 (FL) construct before (Light Off) and after (Light On) exposure to blue light. Cells were pretreated for 1h with either vehicle (DMSO), 0.5 µM okadaic acid (OA), to inhibit protein phosphatase 2A, or 5 µM purvalanol A (PA), to inhibit CDKs. DNA was stained with Hoechst 33258; scale bars, 10m. Right, violin plot presenting quantification of the number of foci per nucleus. Statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise post-hoc comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method. B and C showed analogous results for the constructs lacking the LR domain and the N-ter domain, respectively.

The team directed by Pau Bernadó and Nathalie Sibille, at the CBS (Montpellier) performed Nuclear Magnetic Resonance spectroscopy of the consensus repeat peptide. It showed a reduced amide proton spectral dispersion of the peptide, a pattern typically observed for IDPs. The overlaid Heteronuclear Single Quantum Coherence (HSQC) spectrums for phosphorylated and non-phosphorylated Ki-67 CR confirmed the appearance of 7 new deshielded cross peaks appearing above 8.5 ppm in 1H phosphorylated residues upon incubation with purified CDKs and ATP (Fig. 18D). These peaks probably correspond to the seven putative phosphorylations sites defined in Fig. 18A. I then used Phos-Tag-SDS-PAGE to confirm stoichiometric phosphorylation. This western blot variation allows for the detection of phosphorylated species by different migration patterns in proteins depending on the number of phosphorylation states. Only two species are observed upon CDK phosphorylation, a faint band corresponding to the non-phosphorylated protein and the other stronger band, presumably corresponding to the fully phosphorylated version (Fig. 18C). Next, JMV mapped those phosphorylation sites using mass spectrometry approaches (Fig. 18 D). Five of the seven putative sites highlighted in Fig. 18A were identified plus 6 additional phosphorylations for a total of 11 different sites.

Purified GFP-tagged Ki-67 repeat motifs could instantly phase-separate *in vitro* when mixed with dextran as a crowding agent, and droplet size increased with time (Fig. 18 D, top). As predicted previously, PS was completely abolished when the peptide was fully phosphorylated by CDK prior to the PS assay (Fig. 18 E, bottom).

Taken together, these results confirm that CDK-mediated phosphorylation is able to both promote and inhibit homotypic interactions that contribute to PS, depending on the molecular context. This suggests that Ki-67 may have several competing modes of PS that are differentially regulated by phosphorylation. Our data supports a mechanism for Ki-67-mediated mitotic targeting of nucleolar components to the perichromosomal layer (Sobecki et al., 2016; Booth et al., 2014) via CDK-mediated phosphorylation, which reduces PS of several major nucleolar IDPs, thus triggering nucleolar disassembly, while simultaneously promoting PS of Ki-67 bound to chromatin to recruit nucleolar components.

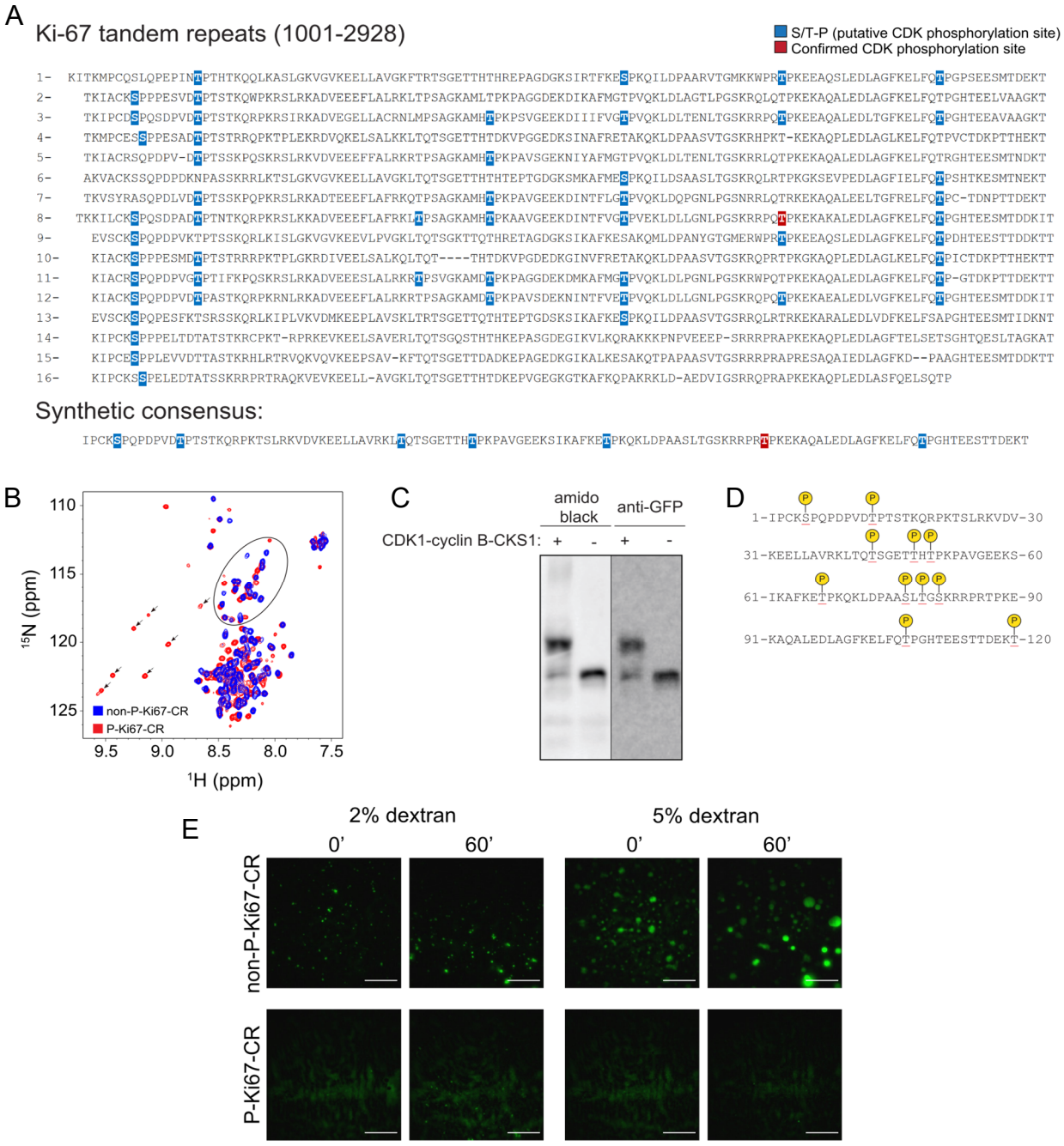


Figure 3.19 | Opto-droplet systems confirmed CDK-mediated regulation of KI67 phase separation (*Authors: GD, JMV; P. Bernadó and N. Sibille team)* (a) Alignment of human Ki-67 repeats, top, with the sequence of the consensus repeat depicted at the bottom. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are highlighted in red and blue, respectively. (B) Overlaid NMR 1H-15N HSQC of unphosphorylated (blue) and CDK-phosphorylated (red) GFP-tagged Ki-67 consensus repeat. Each peak corresponds to one residue. The new peaks (black arrows) correspond to phosphorylated serines or threonines. Non phosphorylated Ser/Thr residues are surrounded by a black oval. (C) WB after Phos-Tag SDS-PAGE indicating full stoichiometric of phosphorylation. (D) GFP-Ki-67 consensus repeat was phosphorylated in vitro using recombinant CDK1-cyclin B-CKS1 protein and the phosphosites were mapped by mass-spectrometry (E) Representative fluorescence images of in vitro phase separation assay with purified GFP-tagged Ki-67 consensus repeat (CR), non-phosphorylated (non-P) or in vitro phosphorylated with recombinant CDK1-cyclin B-CKS1 (P), at indicated dextran concentrations and time points; scale bars, 10µm.

In conclusion, this work reveals in vivo that CDK-dependent mitotic phosphorylation occurs in a switch-like manner on diverse proteins whose common denominators are a high level of disorder and localization to MLOs. Furthermore, our data show that CDK-mediated phosphorylation regulates homotypic interactions between IDPs, which may coordinate diverse cellular processes during the cell cycle. While this is not incompatible with models in which high-affinity interactions contribute to MLO formation by PS (Musacchio, 2022), it suggests that cell cycle control may be less specific than previously thought.

4. Mechanisms of action of CDK regulation of phase separation

Until this point, I have described the results obtained in the search of the understanding of the cell cycle control in the context of the quantitative model, where global activity of CDK controls the cell cycle rather than specific phosphorylations. The main hypothesis proposed in this thesis is that these coordinated peaks of phosphorylation act as regulators of phase-separated compartments throughout the cell cycle. The reorganization of cellular components seems to be essential not only for mitosis onset, but also to regulate other cellular processes such as DNA replication, transcription or stress responses. Simplifying the CDK-mediated phosphorylation functions in vertebrates to regulation of phase separation solely, might be a gross underestimation of the roles of these kinases. Nevertheless, essential site-specific CDK phosphorylations only seem to account for a miniscule fraction of their global kinase activity, and a common molecular mechanism for all CDKs, such as general regulation of protein interaction, constitute an enticing model.

The CDK8/19 subfamily is a good example of kinases that have been described as having defined molecular functions mediated by specific phosphorylations. Each of these transcriptional kinases bind Cyclin C and together with the proteins MED12 (or MED12L) and MED13 (or MED13L) form the kinase module of the Mediator (Daniels, 2013), a multiprotein complex that is of key importance for the transcription initiation. This kinase module is reported to modulate transcription by phosphorylating the C-terminal domain of the RNA polymerase II (Pol II CTD) (Nemet et al., 2014) or by directly phosphorylating transcription factors such as SREBP-1 (Zhao et al., 2012), NOTCH (Fryer et al., 2004; Li et al., 2014), SMAD (Alarcón et al., 2009) and STAT1 (Bancerek et al., 2013). The importance of CDK8/19-Cyclin C complexes is highlighted by the fact that they are essential for the correct embryonic development of vertebrates, which explain the extraordinary levels of conservation of both kinases and cyclin C in this clade. However, inducible depletion of both kinases or the cyclin in adult mice or cell cultures does not have any noticeable phenotype under no stress conditions. Thus, essentiality of CDK8/19 and cyclin C is not determined by their action over the cell cycle, as seen for CDK1, since cells can proliferate without them. It appears that CDK8/19-cyclin C activity controls the transcriptional regulation of gene expression programmes required for cell identity (Whyte et al., 2013; Zamudio et al., 2019), which could

explain the embryonic lethality observed for mice lacking CDK8 (Westerling et al., 2007) or cyclin C (Li et al., 2014).

Despite the glaring difference in cellular functions between the CDK1 and CDK8/19 subfamilies, their kinase domains responsible for their catalytic activities are closely conserved. Therefore, their molecular mechanism might be analogous. A closer look the CDK8/19 molecular context will show that, in fact, the Mediator has been reported as to be a fuzzy complex (Fuxreiter et al., 2014) with several of its subunits having some degree of intrinsic disorder, including CDK8/19 C-terminal prion-like domain. Also, the assembly between the Mediator complex and the enhancers/superhenhacer is described to be formed by phase separation processes (Zamudio et al., 2019). Moreover, CDK8/19 phosphorylation sites in Pol II CTD and the C-terminal region of STAT1 are located in intrinsically disordered regions and, in fact, most transcription factors are predicted to be highly unstructured (Liu et al., 2006), a feature that is pivotal for their functions (Garcia et al., 2021). A common mechanism of action can, thus, be hypothesized for the different CDK subfamilies: CDK-mediated phosphorylation of intrinsically disordered regions will modulate the assembly and disassembly of protein condensates. The main difference between the different subfamilies will reside in the scope of their action, and while CDK1 subfamily can control the overall state of MLOs during the cell cycle, other more specialized subfamilies will exclusively act over specific cellular processes, such as CDK8/19 regulating transcription.

The theory of global action of the CDK1 subfamily in controlling phase separation throughout the cell cycle, however, lacks a detailed explanation of the downstream effects of this regulation. Finding CDK targets that can act as effectors will help to understand the underlying specific mechanisms triggered by changing levels of CDK activity. The proliferating marker Ki-67 can be taken as an example of a highly disordered protein that is phosphorylated by CDK1, and we proposed that these different phosphorylation states modulate its participation in different phase separated compartments. During interphase, Ki-67 is located to the perinucleolar heterochromatin and binds bona fide chromatin interactors (Sobecki et al., 2016), while during mitosis, it situates to the perichromosomal layer (Starborg et al., 1996; Booth et al., 2014; Sobecki et al., 2016; Takagi et al., 2016; Cuylen et al., 2016), where it is important for the relocalization of the nucleolar components for their symmetric distribution in daughter cells (Booth et al., 2014). This interaction with the chromatin led us to propose that Ki-67 may act as an effector of the CDK1-mediated phosphorylation resulting in the regulation of specific gene expression patterns.

In this chapter, I will describe my collaborations in projects that sought to understand the transcriptional regulation resulting from CDK action in phase separated condensates both directly, as proposed for the CDK8/19 subfamily, and indirectly, as by the CDK1 phosphorylation of Ki-67.

Most of the studies described hereafter are submitted or published and available for the reader in the appendix section. Nevertheless, my contribution in these projects was not limited to the technical aspects of bioinformatics, and since I was committed to some degree of their conceptual development, I choose to present the results obtained, from my perspective, in the following section.

4.1 Ki-67 as a global transcriptional regulator

Ki-67 is a nuclear protein expressed exclusively in proliferating vertebrate cells, a feature that led to its widespread use in oncology as a biomarker (Endl and Gerdes, 2000). Consequently, for a long time Ki-67 was thought to be required for cell proliferation (Kausch et al., 2003; C Schlüter et al., 1993; Starborg et al., 1996; Zheng et al., 2006; Zheng et al., 2009), and early work suggested that it promotes ribosomal RNA transcription (Rahmanzadeh et al., 2007; Booth et al., 2014). However, recent genetic studies have shown that despite promoting formation of the peri chromosomal layer of mitotic chromosomes, it is not required for cell proliferation (Sobecki et al., 2016; Takagi et al., 2016; Cuylen et al., 2016; Cidado et al., 2016), and it is also dispensable for ribosomal RNA synthesis and processing. Although Ki-67 is expressed throughout the entirety of the cell cycle, its level is controlled by cell cycle regulators, including cyclin-dependent kinases (CDKs), the transcription factors E2F1/E2F2 and B-Myb, and APC/C-CDH1 complex for its degradation (Ishida et al., 2001; Ren et al., 2002; Miller et al., 2018; Sobecki et al., 2017). Nonetheless, it has been discarded as a relevant cell cycle regulator, due to the lack of homologs in invertebrate species. Moreover, mice lacking Ki-67 seem to develop and age normally, with no apparent phenotype. A closer inspection of cancer cells where Ki-67 had been knocked down showed alteration in their chromatin organization and affected gene expression (Sobecki et al., 2016). I collaborated on a project in which our laboratory studied the requirements of Ki-67 for the different stages of tumorigenesis. My particular contribution was to analyze the sequencing data generated to explore the transcriptional signatures associated with the lack of Ki-67 protein in cancer cells originating from cell culture and tumors grown in mice. In this section, I will present the results obtained.

4.1.1 Loss of Ki-67 causes global transcriptome changes in cultured cells.

Since we previously found that knockdown of Ki-67 in cancer cells altered their chromatin organization and affected gene expression (Sobecki et al., 2016), we hypothesized that the Ki-67 knockout-derived phenotypes might also result from gene expression changes. To test this

hypothesis, we first performed RNA-sequencing (RNA-seq) analysis of Ki-67 wild-type (WT) and knockout embryonic fibroblast NIH/3T3 cells to explore our theory and extend our conclusions to non-transformed cell lines. We observed surprisingly wide-ranging transcriptomic changes upon loss of KI67, with 2,558 genes significantly deregulated in independent clones of *Mki67*$^{-/-}$ cells (p-adjusted < 0.05) (Fig 1A, left). Given that Ki-67 was not recognized as a global regulator of transcriptional pathways nor a transcription factor, this level of transcriptome alteration suggested a more general effect on chromatin rather than a direct involvement in controlling multiple specific pathways. This hypothesis was also consistent with the previous finding of our team that Ki-67 interacts with many general chromatin regulators and transcription factors in the U2OS cancer cell line (Sobecki et al., 2016).

We therefore expected that Ki-67 knockout (KO) would also extensively affect the transcriptome of established cancer cells, with possible consequences for tumorigenicity. To investigate this, we used the syngeneic 4T1 mouse mammary carcinoma model, which is derived from BALB/c mice. This cell line mimics human triple-negative breast cancer, is highly invasive, and spontaneously metastasizes to distant organs (Heppner et al., 2000; Colnot et al., 2004).

Proliferation rates of the 4T1 cell line were unaffected by CRISPR-Cas9–mediated *Mki67* gene knockout in accordance with previous results. In these cells, *Mki67* knockout caused even more extensive gene expression alterations than those observed in NIH/3T3 cells: 4,979 genes were deregulated, of which 1,239 and 585 genes were more than two-fold down-regulated and up-regulated, respectively (Fig 1A, right). Considering the different nature of the NIH/3T3 (mesenchymal) lineage and 4T1 cells (more epithelial-like), the little overlap in the deregulated genes between those cell types upon *Mki67* KO was not surprising (Fig 1B). This supports our hypothesis that, by organizing chromatin, Ki-67 enables global gene regulation in different cell types rather than directly controlling specific genes. We investigated whether the extensive transcriptome changes seen in cancer cells upon Ki-67 knockout affected pathways involved in tumorigenesis. In 4T1 cells, bioinformatic gene enrichment analysis of the most up- and down-regulated genes revealed deregulation of various components of inflammation, apoptosis, p53, the epithelial-to-mesenchymal transition (EMT), estrogen response, K-Ras signaling, and hypoxia (Fig. 1C, left). Deregulated genes were enriched in targets of nuclear factor erythroid2-related factor 2, one of the major orchestrators of responses to oxidative stress; polycomb-repression complex 2 (PRC2), which mediates Histone H3 lysine-27 trimethylation (H3K27me3) and is a well-characterized regulator of the EMT (Tiwari et al., 2013; Chase and Cross, 2011); the pluripotency factors Nanog and Sox2; and interferon regulatory factor 8 (Fig. 1C, right). All of these pathways have previously been implicated in tumorigenesis. We also observed an up-

regulation of key genes of the Notch pathway, and down-regulation of main genes of the EMT, the Wnt pathway, antigen presentation, and aldehyde metabolism, which we validated by qRT-PCR (Fig. 1A, Fig. 1D).

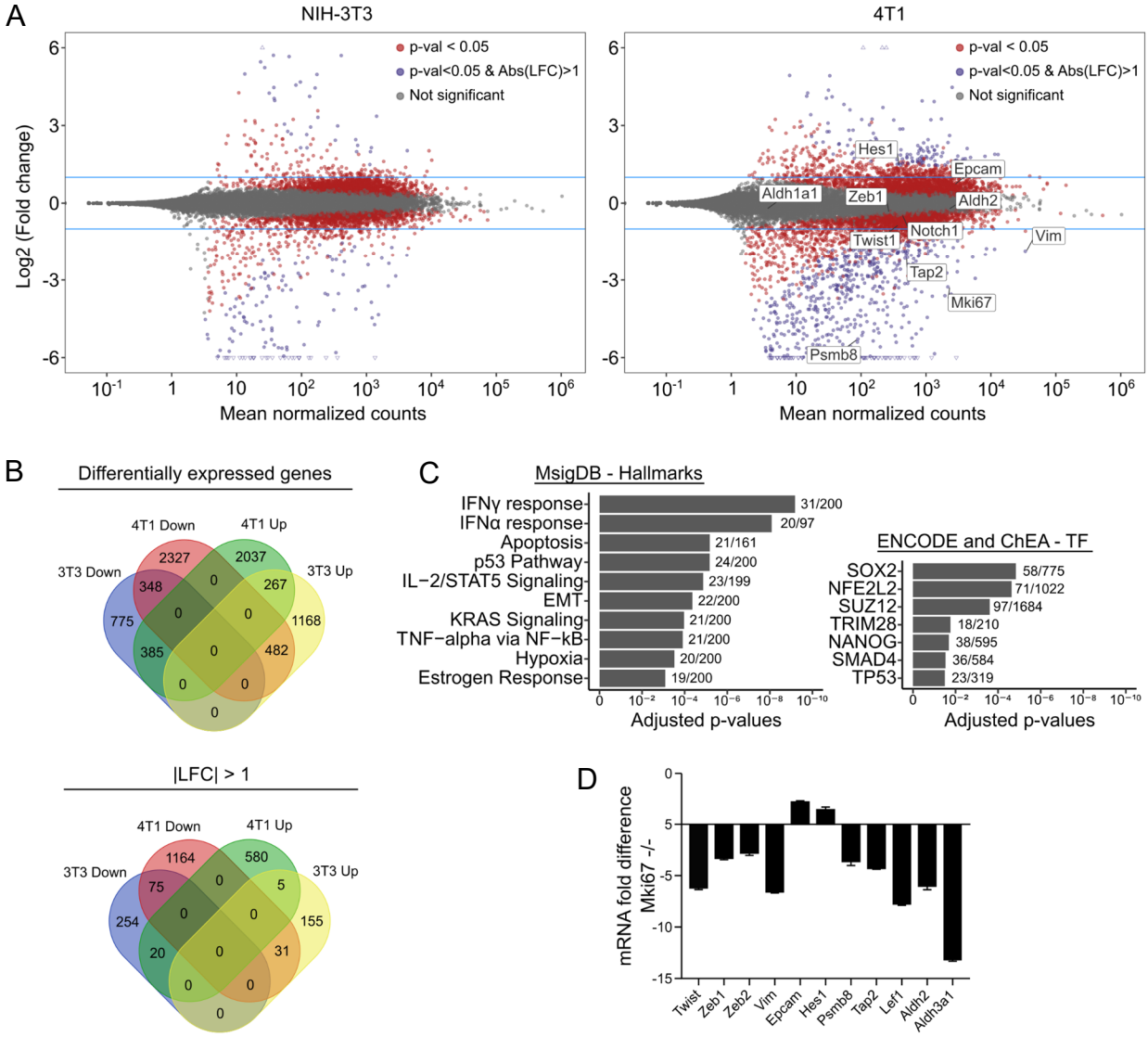


Figure 4.1 | Ki-67 ablation deregulates global gene expression programs in mouse cells. (A) MAplot analysis of differentially expressed genes (DEGs) in NIH/3T3 Mki67–/– cells (left) and 4T1 Mki67–/– cells (right). Red dots: DEGs with p-value<0.05; purple dots: log2fold change (LFC)>1 or <–1 and p-value<0.05; gray dots: not significant (NS). (B) Venn diagrams of DEGs in NIH/3T3 and 4T1 Mki6 –/– cells under condition of p-value<0.05 (top) and p-value < 0.05 and LFC>1 or <–1 (bottom). (C) Gene set enrichment analysis of deregulated genes in 4T1 Mki67–/– cells against MsigDB Hallmarks (left) and ENCODE and ChEA consensus transcription factors (right) databases. (D) qRT-PCR analysis of DEGs in 4T1 Mki67–/– cells; fold change in expression ± SD is shown.

Next, we sought to assess whether the global effect of Ki-67 knockout on gene expression was conserved across cancer cell types and species. We disrupted the *MKI67* gene by CRISPR-Cas9 in human MDA-MB-231 triple-negative breast cancer cells (Fig. 2A), which is a highly mesenchymal-like cell line due to an extensive epithelial–mesenchymal transition (EMT).

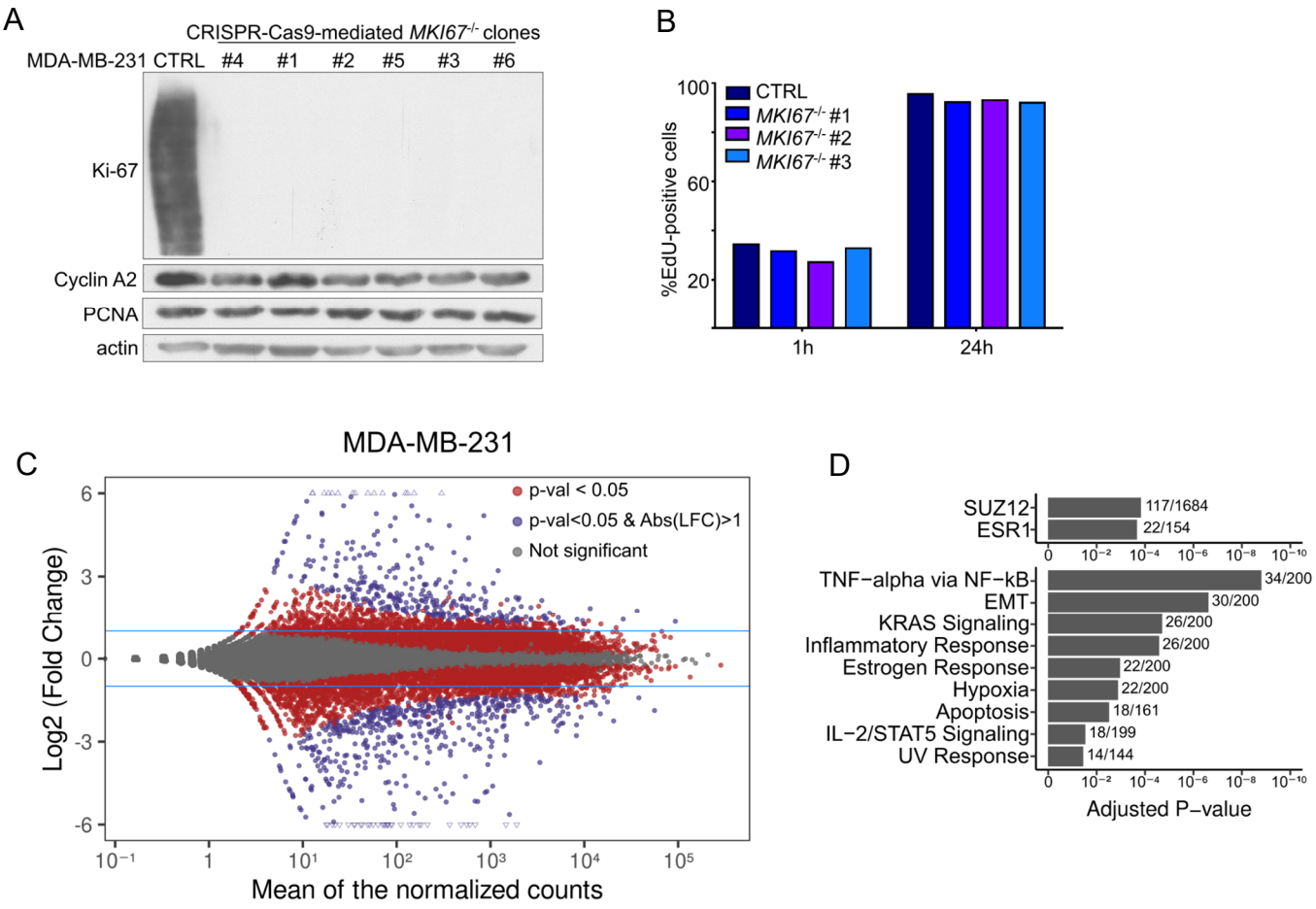


Figure 4.2 | Ki-67 ablation deregulates global gene expression programs in human cells. (*Authors: GD & N. Andrés*) (A) Immunoblotting for the indicated proteins of parental MDA-MB-231 cells and six MKI67–/– clones. (B) Quantification of the number of CTRL and MKI67-/- MDA-MB-231 EdU-positive cells, pulsed with EdU for either 1h or 24h. Cyclin A2 and PCNA proteins are indicators of proliferating cells. Actin serves as a loading control. (C) MAplot analysis of differentially expressed genes (DEGs) in MDA-MB-231 Mki67–/– cells. Red dots: DEGs with p-value<0.05; purple dots: log2 fold change (LFC)>1 or <–1 and p-value<0.05; gray dots: not significant (NS). (D) Gene set enrichment analysis of deregulated genes in 4T1 Mki67–/– cells against ENCODE and ChEA consensus transcription factors (top) and MsigDB Hallmarks (bottom) databases.

As expected, MKI67–/– MDA-MB-231 cells proliferated normally in vitro (Fig. 2A; B). Transcriptome analysis by RNA-Seq showed that Ki-67 knockout in this cell line also caused transcriptome-wide alterations in gene expression (Fig. 2C), with 9,127 genes deregulated, 914 of which were up- or down-regulated by a factor of more than two. In MDA-MB-231 cells, like 4T1

cells, pathway analysis revealed genes involved in the EMT, inflammatory response, early estrogen response, K-RAS signaling, and hypoxia, while a significant portion of the deregulated genes was under the control of PRC2 and estrogen receptor 1 (Fig. 2D). In summary, similar pathways involved in cancer are affected upon Ki-67 knockout in different cancer cell lines, although the specific genes with altered expression levels are not the same. In all cases there exists an imbalance between the number of genes that increase or decrease their RNA levels, with a larger number of down-regulated genes in Ki-67 knockout cells.

4.1.2 Ki-67 mediated expression changes are regulated by Histone marks

The prevalence of down-regulation of gene expression in Ki-67 knockout cells, the enrichment in PRC2 targets among these genes, and our previous observations that Ki-67 associates with the essential PRC2 component SUZ12 (Sobecki et al., 2016) prompted us to ask whether loss of Ki-67 affects genome-wide distribution of H3K27me3. To answer this question, we performed chromatin immunoprecipitation with high-throughput sequencing (ChIP-Seq) on WT and Ki-67 knockout 4T1 cells and compared them using qualitative and quantitative approaches. H3K27me3 presents a broad distribution across genes rather than having a specific localization in regulatory regions. We thus decided to explore this mark throughout the entire length of genes plus 5kb upstream of the transcription starting site (TSS) and 5kb downstream of transcription ending site (TES) for all mouse annotated genes (Fig 3A). Heatmaps with signal in each position (x axis) were generated for each gene (y axis) and the different patterns were clustered obtaining 3 groups, none of which showed major changes upon Ki-67 deletion. Although we did not observe genome-wide changes in H3K27me3 distribution, we sought to focus on the deregulated genes upon Ki-67 ablation, which constitute a small proportion of genes when compared with the entire set of annotated genes. We subsetted the data and we partitioned the heatmap between up- and down- regulated genes in control and *Mki67*$^{-/-}$ cells (Fig 3B). Deregulated genes seem to have an intermediate behavior between those observed in cluster 1 and 3 in Fig. 3A, with downregulated genes showing increased H3K27me3 upon Ki-67 KO (Fig 3B, left). Pathway analysis of these genes revealed a wide range of different cellular processes affected, such as autophagy, intracellular signaling, drug resistance, cancer and cell proliferation, cellular adhesion and drug resistance.

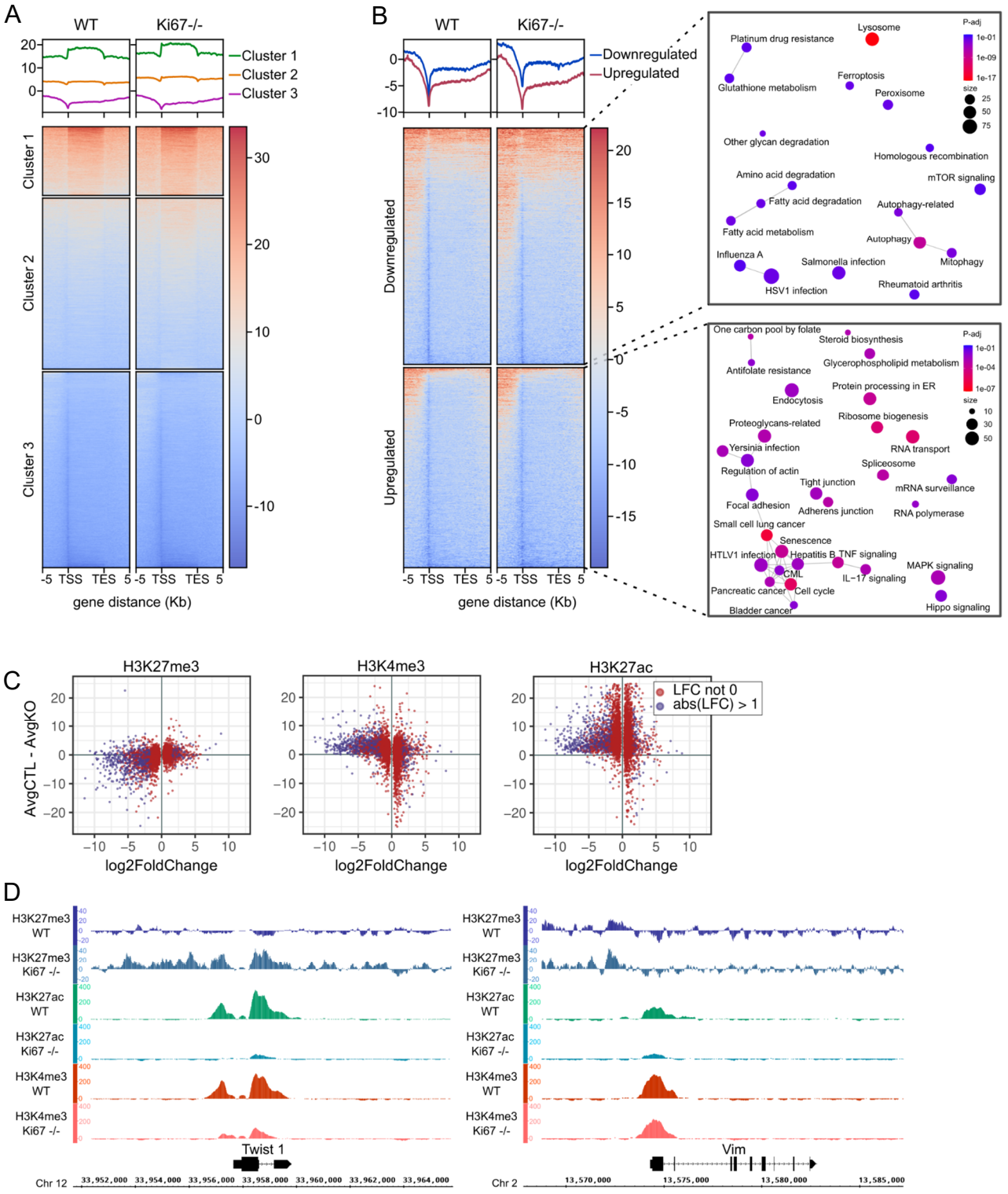


Figure 4.3 | Histone modifications contribute with Ki-67 KO gene expression changes. Heatmaps of ChIP-Seq analysis of H3K27me3 in all annotated mouse genes (A) and down- and up-regulated genes (B) in WT (CTRL) and Mki67–/– cells (TSS, transcription start site; TES, transcription end site). (B, Right) Gene set enrichment analysis associated with deregulated genes. (C) The average values of the H3K27me3, H3K27ac and H3K4me3 ChIP-seq reads over the 10kb region surrounding the gene in 4T1 Mki67–/– cells were subtracted from the values of wild-type cells, and then plotted against the log of the fold-change (LFC) for each gene in RNA-seq (red dots, LFC different than 0; blue dots, absolute value of LFC > 1). (D) ChIP-seq normalized signal of different histone marks for Twist and Vimentin.

To investigate correlations between changes in the levels of histone modifications and of gene expression, we assigned an average value of the repressive H3K27me3, as well as activatory H3K4me3 and histone H3 lysine-27 acetylation (H3K27ac) reads across 10 kb surrounding the transcription start site, and for each gene, we plotted the differences in these values between WT and Ki-67 knockout cells against the log2 of the fold change in expression from the RNAseq (Fig. 3C). Indeed, genes with increased H3K27me3 average signal (negative AvgCTL - AvgKO values) have a tendency of being down regulated (negative log2FoldChange values). This trend is inverted for the activatory marks H3K4me3 and H3K27ac. Nevertheless, in all cases, the correlation is stronger for downregulated genes, and for genes with positive values of log2 of the fold change, the difference in the distribution of the marks is less evident. Of nine genes whose down-regulated expression in Ki-67 knockout cells we confirmed by qRT-PCR (Fig. 1D), we only found an obvious increase of H3K27me3 on the EMT-promoting transcription factor Twist1. This correlated with down-regulation of active promoter-associated H3K4me3 and H3K27ac (Fig. 3D, left). There was also a slight increase of H3K27me3 and reduction in H3K27ac at the Vimentin promoter (Fig. 3D, right). Taken together, these results suggest that Ki-67 loss generally increases PRC2-mediated repressive histone marks at down-regulated genes, but most of the expression alterations resulting from Ki-67 knockout may not be directly due to modulation by PRC2. Instead, they are likely knock-on effects of altered expression of other transcriptional regulators such as changes in activatory promoter-associated histone marks.

4.1.3 Ki-67 deletion disrupts the different stages of tumorigenicity

To determine whether these phenotypic alterations affect the tumorigenicity of cancer cells, we first engrafted WT and Ki-67 knockout 4T1 cells orthotopically into mouse mammary fat pads. Since Ki-67 knockout caused alteration of inflammatory response genes, we initially used athymic nude and NOD/SCID mice, allowing us to assess roles of Ki-67 in tumor growth and metastasis while minimizing possible confounding effects of an altered immune response. RNA-Seq of early-stage tumors from WT and Ki-67 mutant 4T1 cells grafted into nude mice showed that Ki-67–dependent transcriptome changes were well preserved *in vivo* (Fig. 4A, B). A total of 7149 genes where differentially expressed upon KI-67 KO (4073 down-regulated; 3076 up-regulated), and direct comparison of the log2 of the fold change of those genes against the results obtained from *in vitro* 4T1 culture showed a fair level of correlation (Fig. 4B). Only 7 genes presented statistically significant inverse LFC values in both conditions, all of which were up-regulated in tumors and

down-regulated in 4T1 cells (bottom-right quadrant Fig. 4B). The similarities in gene expression patterns include down-regulation of mesenchymal genes and up-regulation of epithelial genes and of the Notch pathway, which was validated by increased HES1 staining in tumors (Fig. 4C). Reduced vimentin staining of Ki-67 mutant tumors confirmed the shift in EMT spectrum from mesenchymal towards epithelial phenotypes *in vivo* (Fig. 4D).

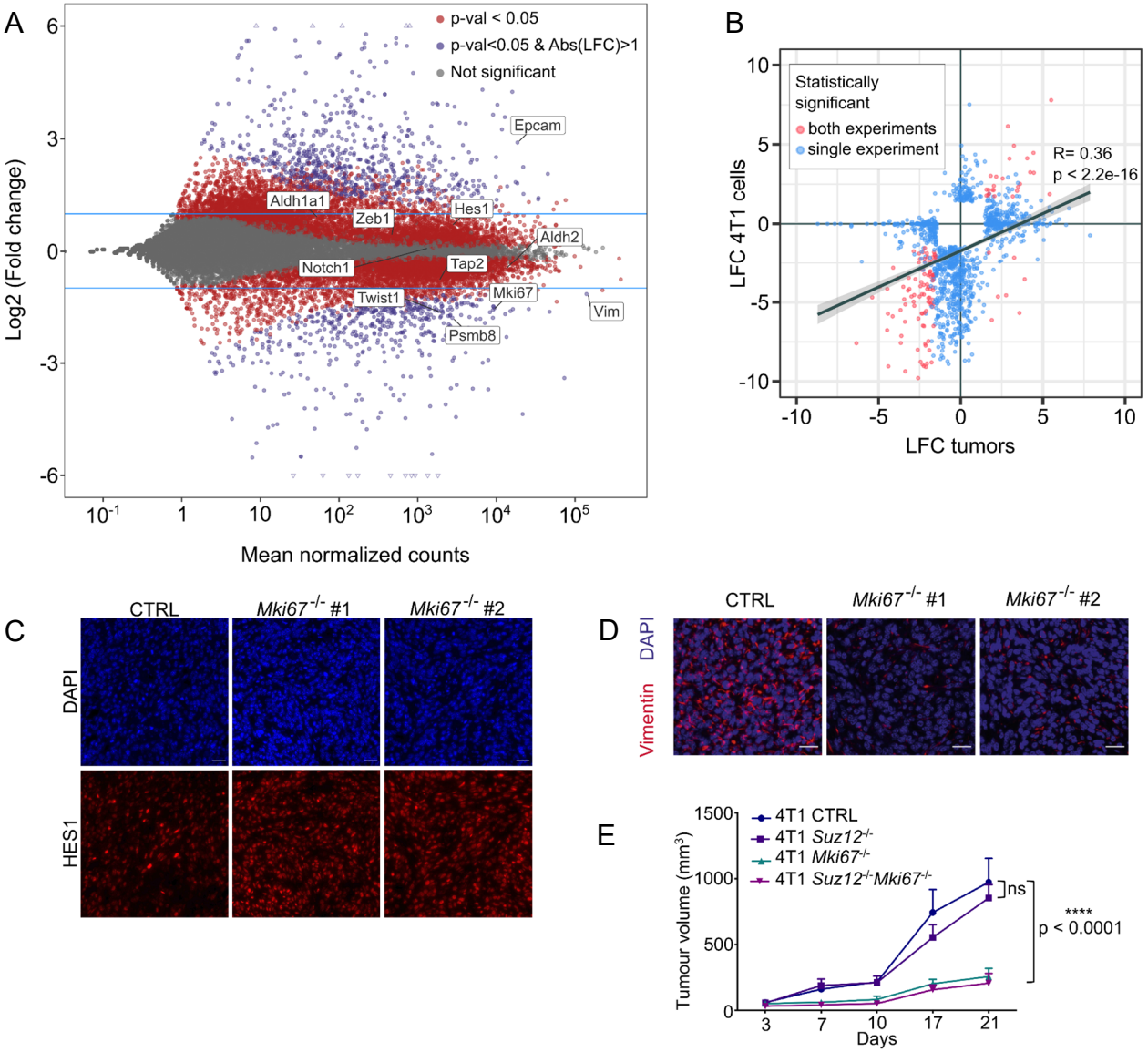


Figure 4.4 | Ki-67 KO transcriptional changes impair tumorigenicity (A) MAplot analysis of differentially expressed genes (DEGs) in MDA-MB-231 MKI67–/– cells. Red dots: DEGs with p-value<0.05; purple dots: log2 fold change (LFC)>1 or <–1 and p-value<0.05; gray dots: not significant (NS). (B) Dot plot analysis of differentially expressed genes (DEG) in 4T1 Mki67–/– cells vs. tumors derived from grafting 4T1 Mki67–/– cells into nude mice. showing a highly significant correlation. Blue dots, DEG in one experiment; Pink dots, DEG in both experiments; LFC, log2 fold change. (C) Fluorescent immunohistochemistry analysis of HES1 in 4T1 CTRL and Mki67–/– tumors in NOD/SCID mice. (D) Fluorescent immunohistochemistry analysis of vimentin in 4T1 CTRL (WT) and Mki67–/– tumors in NOD/SCID mice. (Scale bars: 30 µm.). (E) Tumor growth of 4T1 CTRL, Suz12–/–, Mki67–/–, and Mki67–/– Suz12–/– xenografts over 3 weeks (n=3).

Further analysis showed that cell proliferation in vivo was unaffected by loss of Ki-67 but tumors from Mki67–/– 4T1 cells grew significantly slower than from control cells in both types of immunodeficient mice. These differences could not be totally attributed to changes in apoptosis or necrosis rates and we did not find evidence for increased DNA damage in Ki-67 knockout cells nor tumors. Thus, in vivo differences in growth between WT and Ki-67 knockout tumors cannot be explained either by reduced cell proliferation or by increased DNA damage or cell death. None of the plausible explanations for the reduced tumor growth tested are true in all experimental situations, indicating that effects of Ki-67 loss are probably wide ranging and multifactorial.

We observed also that in orthotopic grafts, control 4T1 cells metastasized to the lungs in 4 wk in nude mice, but metastases were largely absent at this point in mice bearing Mki67–/– tumors. This difference was mostly due to an increased capacity of Mki67–/– cells to seed metastatic lesions.

Last, since many of the genes repressed in the absence of Ki-67 are under the control of PRC2 and concurrent knockout of PRC2 genes seems to partly restore the EMT to Mki67–/– cells, we tested whether the inactivation of PRC2 in cells lacking Ki-67 could restore tumor growth. Concurrent ablation of Ki-67 and the PRC2 component Suz12 did not rescue the tumor growth rate of WT 4T1 cells (Fig. 4E).

4.2 Transcriptional roles of the Mediator kinases CDK8 and CDK19

As mentioned before, not all CDK-cyclin complexes directly control the progress of the cell cycle. A notable case is CDK8, which was discovered as a kinase that binds cyclin C and, like CDK7-cyclin H and CDK9-cyclin T, can regulate transcription by phosphorylating the C-terminal repeat domain (CTD) of RNA polymerase II (Pol II) (Rickert et al, 1999). CDK8 and cyclin C are exceptionally highly conserved in vertebrates, as illustrated by 97% and 98% amino acid identity over the whole sequence between *Xenopus* and human CDK8 and cyclin C, respectively. This unusual level of cross-species conservation implies positive selective pressure over almost all amino acids of these proteins. Indeed, the CDK8 paralog CDK19 also shares these characteristics, and their kinase domains present a high level of identity, too. They mainly diverge in the C-terminal domain that in both cases seems to be disordered, with CDK8 additionally presenting prion-like sequence signatures in this region.

In complex with Med12 and Med13, CDK8-cyclin C form the canonical cyclin dependent kinase module (CKM) of the Mediator transcriptional co-regulator complex, a function that is conserved with the more divergent yeast homologues of CKM subunits (Jeronimo et al, 2016). The latter were revealed as suppressors of a CTD truncation, suggesting a transcriptional-repressive activity of the CKM (Liao et al, 1995). CDK19 also binds cyclin C and interacts with Mediator, in a manner generally thought to be exclusive with CDK8 (Sato et al, 2004; Tsutsui et al, 2008; Knuesel et al, 2009). Mediator is a large multi-subunit complex required for Pol II-dependent transcription in all eukaryotes (Malik & Roeder, 2010). Acute ablation of vertebrate Mediator is lethal for cells and results in a rapid downregulation of the entire transcriptome (El Khattabi et al, 2019). In contrast, the activity of the CKM is apparently non-essential in many cell types, as genes encoding CDK8, CDK19 and cyclin C are not required for survival and proliferation of most cell types in different organisms (Loncle et al, 2007; Kuchin et al, 1995; Li et al, 2014; Postlmayr et al, 2020). However, CDK8 is required for normal development as germline ablation of the *Cdk8* gene is lethal at the pre-implantation stage in mice (Westerling et al, 2007), while conditional deletion using a Sox2 Cre driver is lethal around embryonic day 10.5 (Postlmayr et al, 2020). Cyclin C gene deletion is embryonically lethal at day 10.5 with severe growth defects, and its deletion in adults affects T-cell differentiation (Li et al, 2014), while deletion of Med12 is lethal at late embryonic stages (Rocha et al, 2010). CDK19 deletion has not yet been reported. An essential requirement for CKM subunits in transcriptional regulation in animals cannot, however, be completely ruled out, since differences in the lethality stage of CKM subunit deletions could be due to differential maternal mRNA contributions. Consistent with a repressive role for the CKM in transcription, members of our laboratory and colleagues recently reported that inhibition of CDK8 and CDK19 in human and

mouse pluripotent stem cells is associated with a global overactivation of enhancers and a stabilization of the naive state (Lynch et al, 2020).

CDK8 has also been attributed oncogenic functions in different cancers (Pelish et al, 2015; Firestein et al, 2008; Morris et al, 2008; Kapoor et al, 2010; McDermott et al, 2017; Nakamura et al, 2018; Menzl et al, 2019). Originally proposed to act in intestinal cancers by promoting Wnt transcription (Firestein et al, 2008), CDK8 is also involved in the regulation of other trnascriptional programs such as Notch signaling (Li et al, 2014), NFκB (Chen et al, 2017), HIF1α (Galbraith et al, 2013), the serum-response (Donner et al, 2010), the interferon-γ response (Bancerek et al, 2013; Steinparzer et al, 2019), p53 (Donner et al, 2007), superenhancers (Pelish et al, 2015), histone variant incorporation into chromatin (Kapoor et al, 2010), in pluripotency maintenance (Adler et al, 2012) and the senescence associated tumor-promoting secretory phenotype (Porter et al, 2012). In contrast to CDK8, almost nothing is known about CDK19 roles in cancer, and whether it compensates for loss of CDK8 remains unknown. *In vitro* inhibition or knockdown experiments have suggested that CDK8 and CDK19 control different sets of target genes (Tsutsui et al, 2008; Galbraith et al, 2013; Poss et al, 2016).

Multiple efforts from members of our laboratory have been aimed to investigate the essentiality of the CDK8/19 family in the context of regulation of expression programs during embryonic development and cancer. In this section, I will describe my contribution to those projects that is based mainly on the analysis of RNA-seq data coming from different biological models.

4.2.1 CDK8/19 double knockout impairs proliferation of intestinal cells

To evaluate the possible requirements for CDK8 for cell proliferation and survival in adult vertebrates, Alain Camasses, a former member of our team, designed and generated a conditional knockout allele of *Cdk8* in the mouse by Lox/Cre-mediated recombination targeting exon 2. This removes the critical catalytic lysine-52 and results in a frameshift that truncates over 90% of the protein. In contrast with the work of Postlmayr and colleagues (Postlmayr et al, 2020), this conditional knockout is only induced in the intestine. $Cdk8^{lox/lox}$ mice were crossed with mice expressing a Tamoxifen-inducible Cre under the control of the Villin promoter (el Marjou et al, 2004) (Vill::$Cre^{ERT2/+}$), and efficient deletion of CDK8 in the intestinal epithelium was confirmed. We sought to understand the requirement for CDK8 in the adult mouse intestine since this is one of the most highly proliferative tissues in adults. Mice lacking CDK8 were healthy and did not present any striking phenotypes in the intestine; there was no difference in the number of proliferating cells nor in cell cycle distribution. The number of the different cell types in the intestinal epithelium (stem cells, goblet cells, tuft cells and Paneth) did not change when

compared to wild-type mice, indicating that differentiation programmes were not affected. The deletion was maintained after 2 months, showing that there is no counter-selection for non-recombined intestinal crypts. CDK8 was described as an oncogene in colorectal cancer where it promotes β-catenin-dependent transcription (Firestein et al, 2008; Morris et al, 2008), suggesting that its deletion should inhibit intestinal tumorigenesis. To test this, we used two different systems to induce tumorigenesis: the colitis-associated chemical model of intestinal tumorigenesis induced by azoxymethane-dextran sodium sulfate (AOM-DSS) and the genetic deletion the Adenomatous polyposis coli (Apc) tumor suppressor gene. Despite the difference in morbidity ($APC^{-/-}$ mice presented extremely aggressive tumorigenesis) we observed no effect of Cdk8 deletion on proliferation, number or area of tumors between WT and knockout animals. These data do not support a major role for CDK8 in intestinal tumorigenesis in the mouse. The lack of evident phenotypes of Cdk8 deletion in the adult intestine might suggest a compensatory mechanism of the paralog kinase CDK19. We therefore sought to generate a conditional double knockout. Assuming that a double Cdk8/Cdk19 knockout in adult mice might be lethal, we undertook a conditional deletion using intestinal organoids, which recapitulate many features of intestinal development and morphology while facilitating genetic manipulation in vitro (Clevers, 2016). Dr. Susana Prieto from our team generated intestinal organoids from WT and *Cdk8*$^{lox/lox}$; Vill::$Cre^{ERT2/+}$ mice and disrupted Cdk19 by CRISPR-Cas9-directed gene targeting, using retroviral transduction of Cas9 and a synthetic guide RNA-expressing plasmid. Cdk8 removal was efficient after 7 days of Tamoxifen treatment, were double-knockout-specific loss of Cyclin C and STAT1 Ser 727 phosphorylation , a previously described CDK8 substrate (Bancerek et al, 2013), confirmed redundancy of the two kinases. Growth appeared somewhat slower only in double knockout organoids, with a corresponding larger fraction of non-proliferating cells, as demonstrated by loss of Ki-67 staining. Consistent with slower growth, Cdk8 deletion was counter-selected in Cdk19 knockout organoids, as long-term culture resulted in reappearance of CDK8 in two out of three double knockout organoid populations, presumably due to expansion of a minor unrecombined population.

4.2.2 CDK8/19 regulates gene expression of the CFTR pathway in intestinal epithelium

Since the complex CDK8/19-Cyclin C is a core component of the kinase module of the Mediator, we reckoned that the effects observed are due to changes in the transcriptional regulation. We performed RNA-sequencing analysis of stable populations of single and double Cdk8/Cdk19 knockout organoids. CDK8 loss had a stronger effect (716 genes upregulated, 575

downregulated) than CDK19 loss (158 up, 151 down), while double knockout organoids (1819 up, 1363 down) revealed functional redundancy between CDK8 and CDK19 in regulating gene expression (Fig 5A, B). However, most expression alterations were minor considering that the Mediator is a master regulation of transcription, with only 830 genes deregulated by a factor of two or more. This is consistent with previous studies, none of which have shown sweeping changes in the transcriptome upon downregulation or inhibition of CDK8 or CDK19, rather, alteration of a limited number of specific gene sets, including super-enhancer-associated genes (El Khattabi et al, 2019; Pelish et al, 2015; Galbraith et al, 2013; Steinparzer et al, 2019; Poss et al, 2016; Clarke et al, 2016). In terms of genes controlling cell proliferation, cyclin A (Ccna2), cyclin B (Ccnb2) and cyclin E (Ccne2) were slightly downregulated in knockouts, but this is likely to be a consequence rather than a cause of reduced cell proliferation. CDK8 has previously been found to regulate the p53 and c-Myc pathways (Donner et al, 2007; Adler et al, 2012), but intestinal cells lacking both kinases showed only a slight (though significant) downregulation of c-Myc, while p53 was not affected. In contrast, cyclin G1, a positive mediator of p53 responses and RB functions with a role in cell cycle arrest (Zhao et al, 2003), and p21 (Cdkn1A), a p53 target that inhibits cyclin-dependent kinases to provoke cell cycle arrest, were more strongly upregulated in the double knockout organoids. Gene enrichment analysis in double knockout organoids showed that deregulated genes were enriched in genes characteristic of epithelial tissues, as defined by the mouse gene expression atlas (Fig. 5C, top), that unexpectedly revealed a significant alteration of genes modulated in intestinal knockouts of the cystic fibrosis transmembrane conductance regulator, CFTR (Fig. 5C, bottom). In double knockouts, expression of genes involved in mucus production, Muc2, Muc3, Muc13, Nlrp6, Agr2, Gcnt4, Tff1, were upregulated, while Cftr was reduced. Susana Prieto validated changes of selected genes by qRT-PCR. The loss of Cftr mRNA was also reflected at the protein level, since CFTR protein was lost in double mutant organoids (Fig. 5D).

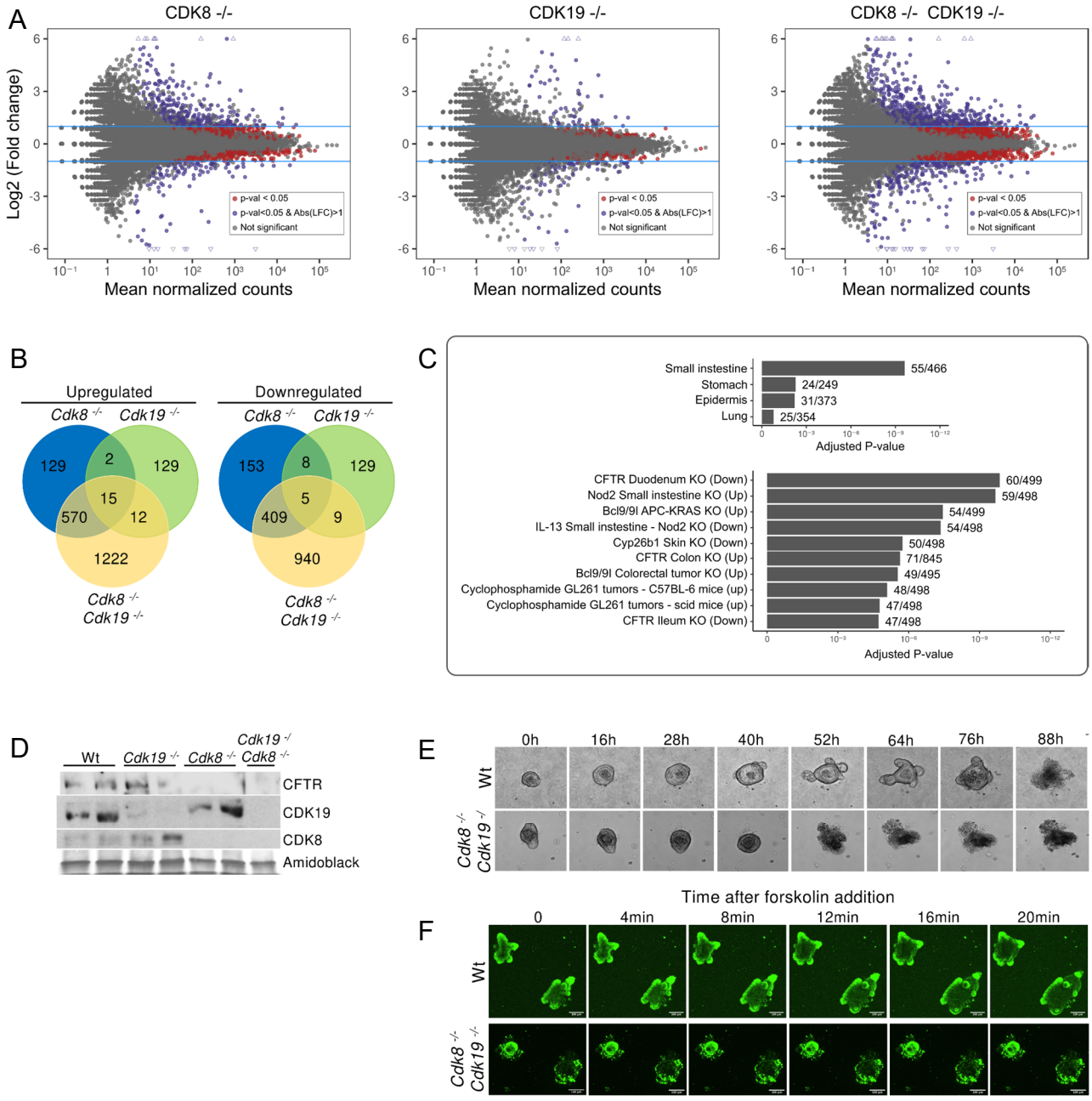


Figure 4.5 | CDK8/19 controls the expression of the CFTR pathways in intestinal organoids (A) MAplots analysis of differentially expressed genes (DEGs) in Cdk8–/– (left), Cdk19–/– (middle) and Cdk8–/–, Cdk19–/– intestinal organoids. Red dots: DEGs with p-value<0.05; purple dots: log2 fold change (LFC)>1 or <–1 and p-value<0.05; gray dots: not significant (NS). (B) Venn diagrams illustrating the overlap of up- and down- regulated genes in the organoids with the three different genotypes (C) Gene set enrichment analysis of the deregulated genes with an absolute value of the Log2 fold change greater than one for the double knockout organoids. Top, sets containing gene expression patterns from different tissues as defined by the mouse expression atlas. Bottom, sets containing genes deregulated after gene knockouts or chemical treatment as reported other studies deposited in the gene expression omnibus (GEO) (D) Western blot of indicated proteins extracted from Wt, Cdk8–/–, Cdk19–/– and Cdk8–/– Cdk19–/– organoids. Amidoblack was used as loading control. (E) Representative phase contrast images of organoids at the indicated time points. mucus release (observed as a dark staining in the center of the organoid. (F) Fluorescence confocal microscopy images of Calcein green-labeled WT and Cdk8–/– Cdk19–/– organoids treated with forskolin. Scale bars, 100 µm.

Cystic fibrosis (CF) is a disease of mucosal epithelia which affects the intestine and other epithelial tissues, and is characterized by excessive mucus accumulation and frequent inflammation (Ehre et al, 2014). We thus wanted to see whether the transcriptome alterations in mutant intestinal organoids translate into a cystic fibrosis-like phenotype. Staining mucin polysaccharides by periodic acid-Schiff (PAS) showed intense mucus accumulation inside *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ organoids, specifically. This was then confirmed by time-lapse video-microscopy (Fig5 E). Functionality of CFTR can be tested in intestinal organoids using forskolin, an adenylate cyclase activator that, if CFTR is functional, induces luminal fluid secretion and organoid swelling (Dekkers et al, 2013). We found that forskolin caused swelling of wild-type but not double mutant organoids (Fig. 5F), indicating that loss of CDK8 and CDK19 functionally recapitulates the *Cftr* mutant phenotype. Additional experiments using the specific CDK8/19 family inhibitor, Senexin B, were performed to confirm the CDK8/19-mediated gene expression regulation of CFTR pathway. By using short and long treatments with the inhibitor and then observing the occurrence of forskolin-mediated swelling, we detected that CDK8/19 inhibition only emulates the CFTR phenotype after treatment for 24 hs., which is compatible with transcriptional regulation times rather than direct CFTR phosphorylation.

Taken together, these data suggest that Mediator kinases are both functionally redundant and largely dispensable for cell survival, proliferation, and differentiation, but may be essential for regulation of specific gene sets in particular cell types; in this case, the CFTR pathway in the intestinal epithelium. Nevertheless, while cell proliferation defects have not previously been reported in HCT-116 cancer cells lacking both kinases (Koehler et al, 2016) our data suggest that in the intestinal epithelium, cells devoid of both CDK8 and CDK19 have an increased tendency to become quiescent, implying that they provide a growth advantage. Our *in vivo* results do not support an oncogenic role for CDK8 in intestinal tumorigenesis, in contrast to early *in vitro* studies (Firestein et al, 2008; Morris et al, 2008).

How direct is the transcriptional regulation of the CFTR pathway and why it is specific to intestinal epithelium are still open questions. We hypothesized that the CDK8/19 kinase family of kinases might regulate specific transcriptional programmes that confer its identity to each cell type. In this way, a milder CF-like phenotype generated by intestinal tissue-specific double knockouts might be benign, allowing normal mouse development, while whole-mouse ablation of CDK8 is embryonic-lethal.

4.2.3 CDK8/19 ablation protects hepatic cells against tumorigenesis

I participated in a collaborative project directed by Dr. Damien Gregoire, from the team formerly directed by Dr. Urszula Hibner (and currently by Dr. Michael Hahne), whose goal was to extend the understanding of the effects of the CDK8/19 family in hepatic carcinogenesis, in the context of the contradicting evidence of the role of CDK8 in different cancer types.

Indeed, quantification by qRT-PCR of CDK8 and CDK19 expression in a large cohort of hepatocellular carcinoma (HCC) patients (n=268) showed that both are significantly overexpressed in HCC tumors compared to non-tumoral counterparts or normal liver. These data are consistent with an oncogenic role of Mediator kinases in hepatocellular carcinoma.

To understand if CDK8 activity is required for normal liver homeostasis, the genetic mouse model described in the section 4.2.1 was used for crosses with transgenic mice expressing the Cre recombinase under the control of the Albumin promoter (Alb-Cre mice), resulting in hepatocyte-specific knockout of CDK8. In accordance with previous reports, the liver-specific deletion of CDK8 was compatible with normal liver development and no change in liver physiology was detected. Moreover, we detected no differences in the expression levels of several β-catenin target genes between the control and the *Cdk8*$^{-/-}$ animals, indicating that CDK8 is not required to regulate this pathway under normal homeostasis conditions.

To study the potential roles of CDK8 in the context of liver cancer a model of hepatic carcinogenesis was used in control and *Cdk8*$^{-/-}$ animals. The hepatotoxic agent diethylnitrosamine (DEN) was injected to young mice giving rise to liver tumors after 6-8 months. After 28 weeks, 9 out of 17 (53%) of the control mice had at least one macroscopic liver tumor while only one out of eleven (9%) *Cdk8*$^{-/-}$ animals developed tumors by this time point. These data support the hypothesis that CDK8 contributes to chemically-induced liver carcinogenesis in mice. To investigate whether the effects of CDK8 loss on carcinogenesis were cell autonomous, primary hepatic progenitor cells (BMEL) from *Cdk8*$^{lLox/Lox}$ embryos were isolated. Upon stable transfection of Cre, we obtained an efficient deletion of CDK8 in these cells. CDK8 removal had no effect on BMEL cell morphology or growth characteristics under standard monolayer culture conditions. As murine liver tumors triggered by DEN injection are often driven by mutated forms of Ras (Buchmann et al., 1991), forced expression of an oncogenic form of Ras, H-RasG12V, was used to investigate effects of CDK8 loss on hepatic cell transformation. Similarly, to results obtained with independent BMEL cell lines (Akkari et al., 2012), H-RasG12V was sufficient to transform primary *Cdk8*$^{lLox/Lox}$ BMEL cells, which then efficiently formed colonies in soft agar. In contrast, CRE-mediated deletion of CDK8 abolished colony formation in this assay, which is consistent with the protective role of CDK8 deletion in DEN-treated livers. Moreover, deletion of CDK8 in

cells previously transformed by RasG12V expression reverted their transformed phenotype. *In vivo* experiments also showed that *Cdk8*$^{-/-}$ mice were specifically protected against formation of orthotopic tumors after CDK8F/F RasG12V or *Cdk8*$^{-/-}$ RasG12V BMEL cells injection into the liver of immunodeficient mice. The colony formation assay results were reproduced in the human hepatoblastoma cell line, HepG2, excluding the possibility that the requirement for CDK8 for transformation is specific for BMEL cells.

Multiple putative pathways involved in carcinogenesis were studied to understand the mechanism of the protective effects of CDK8 ablation against tumor formation. Since CDK8 overexpression in patients correlates with mutant p53 status, we next considered the possibility that CDK8 acts by modulating p53 function. Accordingly, soft agar tests showed that the requirement for CDK8 in Ras induced transformation was abrogated by p53 inactivation via CRISPR/Cas9 editing. Furthermore, depletion of CDK8 did not prevent transformation of the p53-deficient Huh7 cell line. These results were extended to an *in vivo* setting, where simultaneous transfection with N-RasG12D and CRISPR-p53 gave rise to multiple aggressive tumors within 4 weeks. Consistent with results in cellular models, this combination of oncogenic stimuli was also fully efficient in *Cdk8*$^{-/-}$ animals, whose hepatocytes are devoid of CDK8 and failed to properly form tumors upon Ras-mediated transformation only in the presence of p53. Cell lines derived from *Tp53*$^{-/-}$ *Cdk8*$^{lLox/Lox}$ or *Tp53*$^{-/-}$ *Cdk8*$^{-/-}$ tumors were equally capable of giving rise to tumors when injected orthotopically into immunocompetent mice. These results suggest that CDK8 is required for initiation of tumorigenesis by counteracting p53 function.

Interestingly, single ablation of CDK19 recapitulates the protective phenotype of *Cdk8*$^{-/-}$ against transformation in mouse and human cell lines. Thus, both CDK8 and CDK19 are required for Ras-driven transformation of primary hepatic progenitors, neither paralogue being able to compensate for the absence of the other.

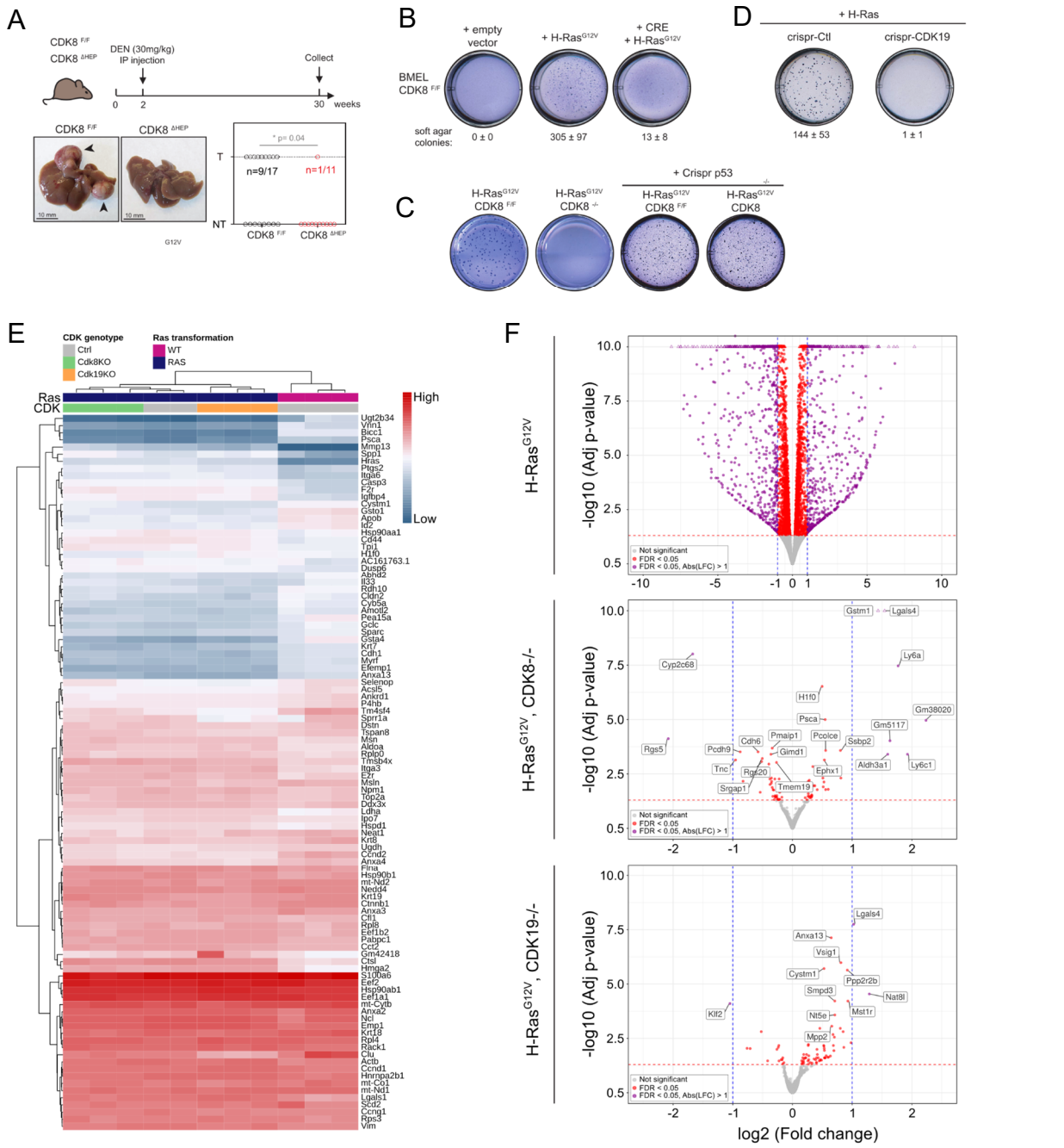


Figure 4.6 | CDK8/19 regulate hepatocyte RAS transformation via p53 independently of their transcriptional activity *(*A) Effect of CDK8 depletion on DEN-induced hepatic carcinogenesis. Pictures of representative livers for each genotype are shown. Arrowheads indicate tumors. Graph represents the repartition of the 28 DEN-injected animals, in each genotype depending if the liver presents a visible tumor (T) or not (NT). P-value of Fisher's exact test is indicated. (B, C, D) Soft agar growth of different BMEL cell lines. Mean number of colonies per well ± SD from (n>=3). B, difference in Ras transformation between WT and CDK8-/- cells; C, replication of the experiment in B with and without p53 activity; D, effects in cells where CDK19 was deleted. (E) Heatmap for the RNASeq experiments. The raw counts were transformed using RPKM and the 100 genes with the higher variance were selected. Hierarchical clustering was used to draw the trees. Colors for RAS transformation: violet, wild type; blue, RASG12V mutation. Colors for Cdk8/19: gray, Control; green, Cdk8 KO; orange, Cdk19 KO. (F) Volcano plots for the differential gene expression. The values of the minus logarithm in base ten of the adjusted p-values (FDR) versus the logarithm in base two of the counts are plotted for each gene and two levels of significance are shown. Red, genes with FDR < 0.05; Purple, FDR < 0.05 and abs (LFC) > 1; Gray, not significant.

As CDK8 and CDK19 constitute the only members of the kinase module regulating transcriptional activity of the Mediator complex, we wanted to explore the hypothesis that transcriptional effects are responsible for the requirement of simultaneous presence of both kinases in HRasG12V-mediated transformation. We analyzed gene expression profiles of the mutant Ras-expressing BMEL cells in the presence or absence of CDK8 or CDK19 by RNA-seq. RasG12V caused deregulation of over 1000 genes, irrespective of the status of Cdk8 and Cdk19 when growing in exponential culture, explaining their indistinguishable morphology. Surprisingly, these acute changes in gene expression were not greatly enhanced by the concomitant disruption of either CDK8 or CDK19. Only a few genes were observed to be deregulated between RasG12V-transformed cells and the same cells without CDK8/19 and in all cases the fold changes were minor. Only around 30 genes showed more than 2-fold differences in expression between cells expressing RasG12V with or without deletion of Cdk8 or Cdk19. At this threshold of fold-change, we did not identify common genes deregulated by the absence of either kinase. This result further shows that, despite the importance of Mediator for gene regulation, neither CDK8 nor CDK19 are required for implementing wide-ranging changes to gene expression, suggesting that their effects in carcinogenesis, which involve p53, are not due to major transcriptional changes.

4.2.4 CDK8/19 regulate transcriptional stress responses in embryonic cells

Similarly to CDK1, the only essential mammalian cell cycle CDK, Cdk8 is lethal in pre-implantation stages of embryogenesis in mice (Westerling et al, 2007). This led scientists to believe that transcriptional regulation by the CDK8/19 kinases was required for normal cell cycle progression. However, we and others (Postlmayr et al, 2020; Prieto et al., 2022 pre-print) have shown that after the initial steps of embryonic development the cell cycle can progress normally in the absence of CDK8 to a point where mice bearing a null mutation in both Cdk8 alleles of specific tissues can be born. This led our team to pursue a project aimed at understanding the roles of the CDK8/19 kinases subfamily during embryogenesis. We utilized the *Cdk8*$^{Lox/Lox}$ mice described in the section 4.2.1 and we crossed them with mice bearing a genetic construct to conditionally express Cre-ERT2 upon tamoxifen treatment. Mice having both genetic constructs can be induced to express the recombinase, which then will generate a CDK8 knockout. Using these animals Dr. Lucie Angevin obtained Mouse Embryonic Fibroblast (MEF) as described by Xu (Xu, 2005), both as primary or immortalized (“3T3 protocol”) cell cultures. These cells where the *Cdk8*-

$^{-/-}$ genotype can be attained by tamoxifen treatment were subjected to CDK19 Crispr-CAS9 gene targeting for the generation of the *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ double knockouts MEF.

The first goal was to characterize the role of the CDK8/19 kinase family on cell proliferation in MEFs. A mild reduction in the growth rate of Cdk8 and double knockouts was observed when compared with wild-type or CDK19 knockout cells. This difference cannot be attributed to an increase in cell death, since cell viability showed similar relative values to those observed in control cells. It has been reported that transcriptional CDK8 activity directly regulates the expression of the cell cycle inhibitor p21 in a p53 independent manner (Donner et al., 2007). We surmised that the difference in growth rate might be due to an increased number of cells entering a non-cycling state or an overall slowdown of the cell cycle. Indeed, flow cytometry analysis of asynchronous cell cultures treated with propidium iodide showed a considerable increase of cells having 2C amounts of DNA in the double knockout compared with the control. Furthermore, levels of p21, as measured by western blot, were also higher in the double knockout. Accordingly, incorporation of 5-ethynyl-2′-deoxyuridine (EdU) after serum starvation was lower in *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ cultures suggesting a decrease in cycling cells. Taken together, these results revealed a diminished capacity of cell cycle progression in CDK8/19 double knockout cells, although our data do not allow for a confident determination whether this is due to a fraction of cells completely arrested or most cells presenting a slower cell cycle.

To understand if the changes observed in *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ MEFs were due to transcriptional regulation we performed differential gene expression analysis from RNAseq data. In accordance with the results shown for intestinal organoids, MEFs presented an extensive change in their transcriptome (4340 deregulated genes) when both kinases are not present (Fig. 7A, B). Individual knockouts of CDK8 and CDK19 result in discrete effects on the expression profiles, with the majority of these deregulated genes not being shared between both genotypes (Fig. 7A, B). Nonetheless, only 373 genes are detected in the double knockout, and barely a few genes in the individual knockouts when a threshold of a 2-fold change in gene expression was established (Fig. 7B, left), which constitute an underwhelming result when considering the crucial role of mediator in transcription. It should be noted that, different from the organoids analysis (section 4.2.2), the log(fold change) threshold in this case is implemented in the hypothesis test, resulting in a more stringent selection of deregulated genes.

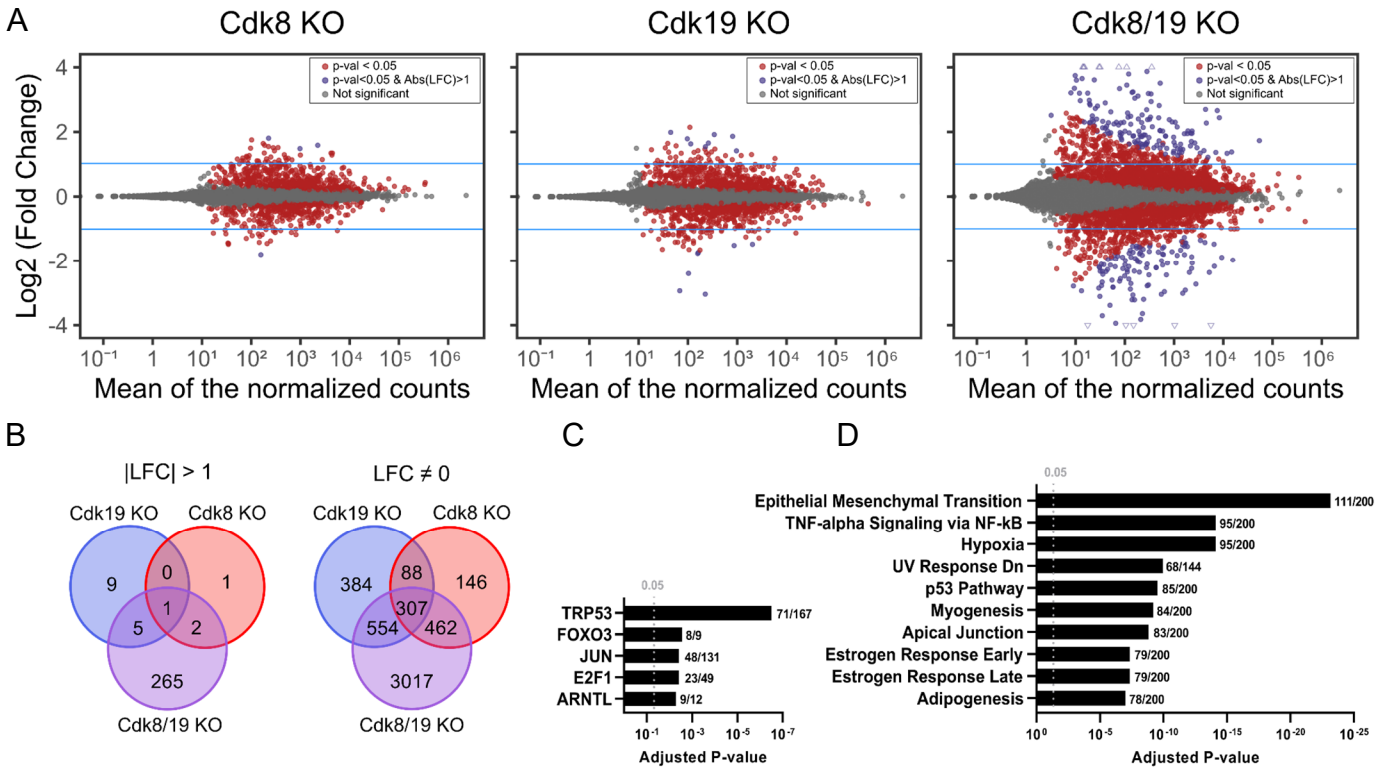


Figure 4.7 | CDK8/19 regulate the transcription of stress response pathways (A) MAplots analysis of differentially expressed genes (DEGs) in Cdk8-/- (left),Cdk19-/- (middle) and Cdk8-/-Cdk19-/- cultured MEF. Red dots: DEGs with p-value<0.05; purple dots: log2 fold change > 1 or < –1 and p-value<0.05; gray dots: not significant (NS). (B) Venn diagrams illustrating the overlap of deregulated genes in MEF with the three different genotypes. Left, genes with log2 fold change > 1 or < −1; right, all deregulated genes. (C, D) Gene set enrichment analysis of deregulated genes. Analysis was performed against the TRRUST transcription factors (C) and MsigDB Hallmarks (D) databases.

Gene set enrichment analysis showed that in the double knockouts, the deregulated genes are enriched in targets of p53-mediated transcriptional regulation (Fig. 7C). This p53-like response could explain, at least partially, the increase of p21 levels and the cell cycle progression hyndrance in *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ MEFs. An enrichment of genes of the EMT, Tumor necrosis factor signaling and Hypoxia pathways was observed in our analysis (Fig. 7D), and all of them were already reported to be associated with CDK8/19 functions (Xu et al., 2015; Serrao et al., 2018; Chen et al., 2017; Yamamoto et al., 2017; Galbraith et al., 2013). Taking these results together, the CDK8/19 family seems to regulate the expression of pathways associated with different types of cellular stress in MEF.

A molecular docking screen of CDK8 was performed by Dr. Franck Borel and Dr. Jean-Luc Ferrer (Institut Biologie Structurale, Genoble), to find small molecules that could bind the active site of CDK8. Out of the $11x10^6$ publicly available candidate compounds, the 100 molecules with the highest affinity for CDK8 were tested by the team of Dr. Sandrine Ruchaud (Station Biologique, Roscoff). In vitro specificity tests of these compounds on CDK8 and 8 other kinases likely to be inhibited by these drugs. Dr. Angevin was in charge of the biological characterization of these CDK8/19 specific inhibitors. By analyzing a model substrate, the phosphorylation of serine 727 of STAT1, she identified the drug Senexin B as the most effective and specific inhibitor for this kinase family in cells. We next sought to compare the trnascriptional deregulation observed in the double knockouts with the effect triggered by treatment with Senexin B. Indeed, treatment with the inhibitor modified the expression of 4340 genes (273 with 2-fold change threshold) (Fig. 8A, left), which is a number comparable to what was observed in the CDK8/19 double knockout. As expected, the treatment of *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ MEFs had only a modest effect on the transcriptome (Fig. 8A, right). Principal components analysis (PCA) of all samples showed a similar tendency, in which Senexin B-treated samples are grouped together regardless of their genotype (Fig. 8B). To better illustrate this behavior, I generated a heatmap with the 100 most variable genes (y axis) across all samples (x axis) for WT and double knockout MEFs, both treated and untreated with Senexin B (Fig. 8C). The expression patterns of the double knockout cells do not change considerably after treatment with the inhibitor, which is the case for WT cells (Fig. 8C). Nevertheless, Senexin B-treated MEFs do not have an approximate behavior to that observed in CDK8/19 KO cells. The comparison between the deregulated genes after deletion of both kinases or Senexin B treatment showed a 40-50 % overlap, illustrating a common mechanism of action. However, there might be genes affected only by the complete ablation of both kinases, while some effects might be due to off-targetting by the inhibitor (Fig. 8D). A similar gene enrichment trend was observed for Senexin B-treated MEFs when compared with the analysis for double KO cells, with a majority of p53 regulated genes (as defined by the TRRUST database) changing their expression levels significatively (Fig. 8E), and a great overlap with the MSigDB pathways (Fig. 8F).

Altogether, these results reveal substantial functional redundancy between CDK8 and CDK19, as well as differences between inhibition and gene knockout. Moreover, this work suggests that these kinases, which are dispensable in basal conditions, are essential for cellular responses to stress.

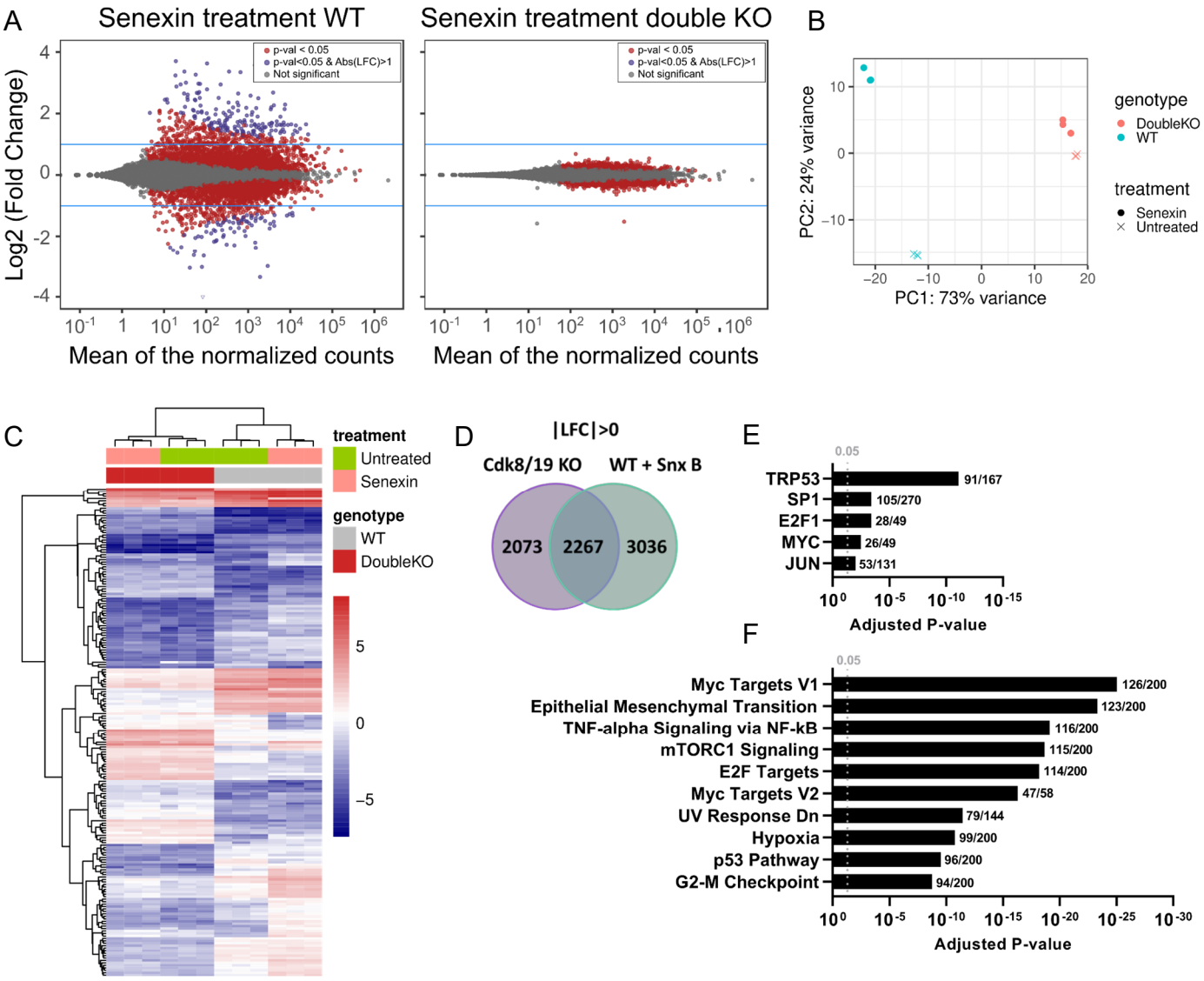


Figure 4.8 | CDK8/19 inhibition partially recapitulate the double knockout phenotype (A) MAplots analysis of differentially expressed genes (DEGs) in Cdk8+/+ Cdk19+/+ (left) and Cdk8-/- Cdk19-/- (right) MEFs after senexin treatment. Red dots: DEGs with p-value<0.05; purple dots: log2 fold change > 1 or < −1 and p-value<0.05; gray dots: not significant (NS). (B) Two-dimensional representation of the principal component analysis (PCA). The two principal components selected explain 97% of the total variance between samples. (C) Heatmap for the RNASeq experiments. The raw counts were transformed using RPKM and the 100 genes with the higher variance were selected. Hierarchical clustering was used to draw the trees. Colors for Senexin B treatment: Green, untreated; Pink, treated. Colors for Cdk8/19: Gray, Control; Red, double KO. (D) Venn diagrams illustrating the overlap of deregulated genes in MEF after CDK8/19 deletion and Senexin B treatment. (E, F) Gene set enrichment analysis of deregulated genes after senexin treatment in WT MEFs. Analysis was performed against the TRRUST transcription factors (E) and MsigDB Hallmarks (F) databases.

Results obtained from the transcriptome analysis in different conditions led Dr. Angevin and Dr Fisher to contemplate the possibility that these kinases act as a stress modulator via the activation of specific transcriptional programmes. Of particular interest was the response to UltraViolet (UV) light exposure, that has not yet been described in the bibliography. Subsequent studies from our

team revealed that the complete loss, but not inhibition, of Cdk8/19 sensitizes MEFs to UV-C-induced apoptosis. In addition, in irradiated Cdk8/19 KO MEFs, p53 stabilization seems to occur more rapidly, and depends on PML protein. While CDK8 was diffused across the nucleus, PML localized in defined foci. Surprisingly, depletion of p53 and PML by siRNA further increased UV-C-induced apoptosis in Cdk8/19 KO MEFs. It therefore seems that the stabilization of p53 via PML limits apoptosis in mutant Cdk8/19 cells, perhaps by blocking the cell cycle.

5. General discussion

From its origins, this thesis aimed to further develop the concept of the “quantitative model” for the cell cycle. Highly inspired by the early work of my director, Dr. Fisher, and, following on studies from members of Paul Nurse’s laboratory (Fisher and Nurse, 1996; Coudreuse and Nurse, 2010) I sought to design a tool to externally command the endogenous CDK1 activity in human cells. Here, I presented the design of a genetic construct to be stealthily inserted as an artificial intron within an exon of the *CDK1* gene locus. This system, based on the work of Andersson-Rolf and colleagues (Andersson-Rolf et al., 2017), can switch from the endogenous CDK1 activity to an analogue-sensitive version of CDK1 fused either with cyclin B1 or cyclin A2, upon Cre-mediated recombination. Its conception and building constituted a veritable *tour de force,* where basic concepts of genetic engineering were applied to generate an intricate tool to control the cell cycle. Previous works have reported that CDK1 is the only essential family member and sufficient for driving the entire cell cycle in mice (Santamaria et al., 2007). Furthermore, CDK2 cannot compensate for its activity even when located in the *CDK1* locus, indicating that it is the protein rather than its transcriptional regulation that makes CDK1 essential (Santamaria et al., 2007). Nonetheless, it has never been demonstrated that the mammalian cell cycle can be driven by a single CDK-cyclin complex as it has been shown in yeast (Coudreuse and Nurse, 2010). This chemical-genetic system would allow us to answer this question, and others, which are greatly relevant for the validation of the “quantitative model” theory that states that is the level of global CDK phosphorylation given by the relative CDK/PPase activity ratio, rather than specific combinations of CDKs and cyclins, that brings about the different events of the cell cycle. Indeed, it has been recently reported that in the absence of CDK1, CDK2 can assume the role of the mitotic kinase when it is overexpressed to generate similar phosphorylation levels to that observed for CDK1 (Lau et al., 2021). Moreover, the initiation of S-phase, historically assigned to the CDK2-cyclin E complex, was recently shown to happen in the absence of CDK2, an requires CDC7 or CDK1, most probably in complex with cyclin B, to phosphorylate the MCM complex proteins (Suski et al., 2022).

Another important feature of this system is that CDK1 activity can be externally regulated in a dose-dependent manner by the addition of a specific inhibitor. Other chemical-genetic systems allowing for the rapid degradation of Cyclins (Hégarat et al., 2020) or CDK (Lau et al., 2021) have been reported, although these “all-or-nothing” approaches lack the potential to observe what happens at intermediate CDK activity levels. Considering previous evidence, CDK-mediated

phosphorylation might operate as a bistable switch and an intermediate degree of CDK activity might be an artificial situation and, thus, one might question the relevance of a dose-dependent modulation system. However, this would allow us to better understand how different CDK1 phosphorylation levels trigger different cell cycle phases and to observe cellular changes during transitions between those levels that otherwise transpire too quickly to be detected. Modulation of CDK1 activity might also be crucial in a context where a single fusion CDK1-cyclin complex drives the cell cycle, if its expression is regulated by the *CDK1* gene locus. Fusion CDK1-cyclin proteins can be degraded by the APC/C, allowing for the metaphase-to-anaphase transition, but it is not clear if cells would be able to cycle without the contribution of cyclin expression regulation to the CDK activity oscillation. In that case, CDK activity can be externally controlled by alternating concentrations of a specific inhibitor.

With this genetic switch I sought to obtain an elegant system to convert the endogenous CDK1 activity to its AS version, that could be extended to multiple different cell models to answer fundamental questions about the control of the cell cycle. Although yet unfruitful considering this main goal, the overall experience resulted in an excellent thought exercise that helped me master important concepts of the genetic organization of the molecular control of the cell cycle.

During the course of my doctoral studies, the team of Piotr Sicinski published a paper in which they accomplished the substitution of *CDK1* locus with a genetic construct expressing AS-CDK1 in mice. Unfortunately, this is lethal for embryos and homozygous animals bearing such construct cannot be obtained. Nevertheless, they obtained homozygous AS-CDK1 ESCs which were used to study the roles of CDK1 in developmental epigenetic regulation. This inspired us to revisit our genetic switch to adapt it to function in mice. In this way, animals having homozygous alleles with the CDK1 switch can be brought to term, and then Cre recombinase could be expressed in any particular tissue to study the dose-dependent effects of specific CDK1 inhibition. Working with animals provides us with the possibility of crossing heterozygous individuals which simplifies the insertion strategy: we no longer have to target exons to ensure that alleles will be null if not mutated. A similar approach, but now based mostly in the FLEx system (Schnütgen et al., 2003), was submitted for a call for proposals to develop genetically modified mouse models by PHENOMIN and it was accepted to be funded. This will become a pivotal tool for the laboratory to develop an updated cell cycle model that takes into account the recent body of evidence.

We hypothesized that a possible mechanism of action for the quantitative model for the cell cycle control might be that coordinated increase in the overall phosphorylation levels can generate physicochemical changes in protein condensates regulating their formation and dissolution.

Several biochemical processes occurring during interphase are reported to happen in phase-separated membraneless compartments such as rRNA biogenesis (Harmon and Jülicher, 2022), cellular stress response (Corpet et al., 2020; Hoischen et al., 2018; Rai et al., 2018), gene expression regulation (Guo et al., 2019; Cramer, 2019), DNA replication and DNA damage repair. Most of these MLOs are dissolved at the entry into mitosis, while the formation of other condensates is observed, including centrosomes, kinetochores and the perichromosomal layer, and whose components are reported to undergo phase separation. Not only these condensates are regulated in a cell cycle manner but, in most cases, here I showed that key proteins driving their formation are phosphorylated mainly by CDKs, and possibly other cell cycle-related kinases. We therefore theorized that global CDK phosphorylation changes regulate the reorganization of the cell to perform different biochemical reactions during the different phases of the cell cycle.

The two main molecular features that determine whether a protein can phase separate are multivalency and intrinsic disorder (Banani et al., 2017; Shin and Brangwynne, 2017). Regions of the proteins with such characteristics are able to form macromolecular condensates with other molecules mediated by homotypic and heterotypic interactions. It has been reported that phosphorylation happens preferably in disordered regions (Iakoucheva et al., 2004) and that, indeed, kinase activity can modulate the process of phase separation (Rai et al., 2018). To inspect if this was the case for CDK-mediated phosphorylation, we first compiled a database of high-confidence human and yeast CDK1-family substrates. Our first observation was that only around 10% of human CDK phosphosites are non-proline-directed, while our analysis of high-confidence CDK substrates in yeast indicates that the corresponding figure may be closer to 20%. This came as a surprise considering that CDK1 can also efficiently phosphorylate non-proline-directed sites (Suzuki et al., 2015), and that recent reports estimated that approximately 30% of CDK1 phosphorylation sites detected in mouse ESCs were non-proline directed (Michowski et al., 2020). There exists a clear discovery bias toward proline-directed, and non-proline-directed sites are almost always filtered out from analysis of CDK substrates.

I then calculated the disordered regions in all proteins detected as being phosphorylated in our yeast and human datasets using multiple disorder predictors. I observed that phosphorylation in general is significantly enriched in disordered regions of the proteins in accordance with previous reports (Iakoucheva et al., 2004), although our results were obtained by normalizing the data in such a way that the composition signature of disorder is taken into account and thus removing the bias that other studies might present due to the differential composition of IDRs. I estimated that there exists an increased risk of CDK-mediated phosphorylation to occur in IDRs, in both yeast and human datasets. To avoid a compositional bias due to the predictor chosen, I replicated

this analysis with three different well-ranked predictors, obtaining the same results. Surprisingly, this trend was analogous to that observed for other cell cycle-related kinases (except NEK kinase family) and MAPK kinase family members that, albeit not being cell cycle kinases, share the tendency to phosphorylate proline-directed sites. Next, I estimated the distribution of disorder percentage for each human and yeast CDK datasets and I compared them against the rest of their phosphorylated proteins. In both cases, using different predictors, I saw an approximately 2-fold increase of disorder proportion in CDK targets when compared with the rest of the phosphoproteome. This was also true for other cell cycle kinases (again, except NEK), and although MAPK targets showed a higher disorder percentage than the rest of the human phosphoproteome, they were significantly less disordered than targets of CDK, PLK, AURK and DYRK. MAP kinases and CDKs share the same proline-directed recognition motif and both have a comparable increased risk of phosphorylating disordered regions, and yet MAPK substrates notably present less proportion of disordered residues than CDK targets. This might mean that evolution is shaping cell cycle-regulated proteins to have a higher proportion of disordered regions, and thus present an increased proportion of their amino acid chain that can be accessible to specialized kinases and consequently phosphorylated. It can be argued that disorder as a sole requirement for cell cycle phosphorylation is an oversimplification, given that cell cycle-related kinases have different phosphorylation motifs. How can each kinase recognize their substrates if they all phosphorylate the same molecular feature? In the first place, there are numerous examples of overlapping kinases that phosphorylate the same substrates. To distinguish when and where to target a given substrate, different kinases might profit from spatio-temporal limits (CDK1 and CDK2 being cytoplasm or nucleus, or at the kinetochore). Moreover, the general idea of intrinsic disorder as multivalent, unstructured and mobile regions of proteins is helpful to generalize certain concepts but it might fail to describe more precise events in the cell. Considering different “flavors” of disorder, as long ago proposed by Vucetic and collaborators (Vucetic et al. 2003), could also contribute to answering the question. Indeed, these different flavors of disorder are detected by general disorder predictors but each one has their own particular compositional signature, structural (or, more correctly, un-structural) properties and they have been associated with distinct biological functions (Vucetic et al. 2003).

We established a collaborative line of work with Maarten Altelaar and Juan Manuel Valverde, that provided us with an additional dataset to include in this analysis. They had obtained mass spectrometry data of phosphorylation on proteins of single *Xenopus* embryos in their first stages of embriogenesis. This naturally occurring highly-synchronous system allowed us to collect cell cycle-resolved phosphorylation data that could be compared with our previous analysis of publicly

available data on human and yeast. They detected a fraction of the phosphorylations changed over time ("dynamic phosphorylations") that could be clustered following four different behaviors. We find evidence that CDKs are responsible for the majority of cell cycle-regulated phosphorylations. We intersected our high-confidence human CDK1-family substrates and found that nearly a quarter of them are represented among the dynamically phosphorylated proteins in *Xenopus* embryos, while more than half of the phosphosites conform to the minimal CDK consensus motif. This is likely an underestimation of the true proportion since our human CDK substrate dataset is probably incomplete and biased towards proline-directed phosphorylations. We then focus on the dynamic phosphorylations that were showing an obvious oscillatory behavior. Using quantitative mass spectrometry in high time-resolution, we demonstrated that the oscillations in phosphorylation levels were, indeed, switch-like. This is consistent with mathematical modeling of the mitotic CDK regulatory network, which predicts switch-like CDK1 activation. Surprisingly, however, this abrupt phosphorylation of many CDK substrates occurred despite progressive downregulation of CDK1 phosphorylation on tyrosine-15, which has been thought to be a key contributor to ultrasensitivity in CDK1 activation (Kim and Ferrell, 2007; Trunnell et al., 2011). Indeed, early work on cell cycles during *Xenopus laevis* development demonstrated that CDK1-Y15 phosphorylation was absent from cell cycles 2-12 (Ferrell et al., 1991), although more recent data suggest low level and minor fluctuations in cell cycles 2-3 (Tsai et al., 2014), consistent with our findings. We also observe similar fluctuations of S38 in *Xenopus* Wee1A, and S120 and S299 of CDC25A, which are likely mediated by CDK1. Our data suggests that the CDC25 positive feedback loop is active and increasing over time, while the WEE1 double-negative loop remains constant, explaining the slight oscillation but progressive general downregulation of CDK1-Y15 over time. If regulated CDK1-Y15 phosphorylation decreases over time, how then could the switch-like dynamics be sustained for a prolonged period after the first cell division? One possibility is that futile cycles of CDK1 and opposing protein phosphatase activity, likely PP2A (Krasinska et al., 2011), are responsible. Similar homeostatic behavior has been recently reported for KNL1, whose phosphorylation status is regulated by the kinase PLK1 and the phosphatase PP2A and controls the recruitment of BUB complex to kinetochores for the assembly of the mitotic checkpoint complex, although no evidence for switch-like behavior is shown (Corno et al., 2022, pre-print). Nevertheless, the Goldbeter-Koshland model of futile cycling predicts switch-like changes in network output upon small variations in the relative activities of opposing enzymes around a critical threshold, even without any feedback between the opposing enzymes (Goldbeter and Koshland, 1981). PP2A is inhibited by Greatwall kinase-mediated phosphorylation of ARRP19/ENSA (Gharbi-Ayachi et al., 2010; Mochida et al., 2010),

which is promoted by CDK1 itself (Yu et al., 2006). Consistent with our data and this model, reconstituting mitotic entry with purified components in Xenopus egg extracts demonstrated that switch-like behavior did not depend on CDK1-Y15 regulation, but, rather, reciprocal regulation between CDK1 and PP2A via Greatwall (Mochida et al., 2016). A second CDK1-Y15 phosphorylation-independent mechanism generates positive feedback in cyclin B1 accumulation, since CDK1-mediated phosphorylation of the ubiquitin ligase subunit NIPA prevents interphase cyclin B1 degradation (Bassermann et al., 2005; Bassermann et al., 2007). Our data suggest that both CDK1-Y15-independent mechanisms may contribute to the switch-like phosphorylation since we find that both Greatwall and NIPA phosphorylation oscillate in mitosis in vivo.

Next, by performing quantitative phosphoproteomic on replicating *Xenopus* eggs extract we showed phosphorylation dynamics of known DNA replication factors and many others during S-phase. MCM4 and RIF1, which are involved in DDK-mediated initiation of DNA replication and replisome stability (Alver et al., 2017), are among those with the highest number of dynamically regulated phosphosites, suggesting a role for hyperphosphorylation of a subset of replication proteins. Our phosphorylation motif analysis revealed that potential DDK phosphosites are upregulated during S-phase, both *in vivo* and *in vitro*. A recent paper has highlighted the overlapping functions between DDK- and CDK-mediated phosphorylation at the S-phase entry, where one of both kinases is required to progress in the cell cycle (Suski et al., 2022). Both kinases share some targets, most notably MCM2, although each kinase seems to have specific sites in those targets (Suski et al., 2022). We observed that phosphosites in aurora kinases phosphorylation motifs were enriched in interphase clusters. Aurora kinases are best known as regulators of the microtubule-kinetochore attachment at prometaphase and of cytokinesis, thus this suggests that phosphorylation of many of their substrates occurs upstream of mitotic CDK phosphorylation, or that Aurora kinases play an under-recognized role during DNA replication. Indeed, it has previously been suggested that Aurora kinases regulate DNA synthesis through their activity in mitosis, by chromatin remodeling and by organizing the replication origin firing program (Koch et al., 2011). Further studies will be required to elucidate interphase roles of Aurora kinases.

Xenopus dynamic phosphoproteins replicated all the results obtained for yeast CDK targets and human cell cycle kinases: a general tendency of phosphosites to locate in IDRs with an increased risk of phosphorylation in these regions for the dynamic phosphosites specifically; moreover, proteins with at least one dynamic phosphosite were significantly more disordered than the rest of the phosphoproteins detected. We hypothesized that cell cycle regulatory mechanisms have been selected to control intrinsically disordered proteins, presumably to modulate the formation

and dissolution of MLOs acting as biochemical hubs in specific phases of the cell cycle. Since CDK sites tend to cluster within IDRs (Holt et al., 2009; Moses et al., 2007), and our data show that many cell cycle-regulated CDK substrates are key components of MLOs, it is likely that CDK-mediated phosphorylation will affect MLO physicochemical properties. Indeed, CDKs can control formation of stress granules (Yahya et al., 2021) and histone bodies (Hur et al., 2020), NPCs (Linder et al., 2017) and possibly many other MLOs. An emerging model is that liquid-liquid phase separation, which depends on weak interactions between IDRs, underlies the self-assembly of many MLOs (Banani et al., 2017; Hyman et al., 2014; Woodruff et al., 2018). Thus, switch-like phosphorylation of IDRs may regulate rapid MLO reorganization in mitosis, when most of these MLOs are dissolved. In support of this model, DYRK3 is a CDK-related kinase whose inhibition disrupts mitotic remodeling of stress granules, splicing speckles and pericentriolar material, all of which are thought to assemble via phase separation (Rai et al., 2018). Here, I presented a collaborative multiscale *in vitro*, *in cellulo* and *in silico* set of experiments on the proliferation marker Ki-67, which clearly shows a CDK1-dependent regulation of condensation. However, no simple conclusion can be driven regarding the net effect of Ki-67 phosphorylation on the process of phase separation, since there seems to exist competing modes of regulation of phase separation by phosphorylation of its repeat domain. I surmise that different levels of phosphorylation and the cellular context will localize Ki-67 to the perinucleolar heterochromatin during interphase and to the perichromosomal layer in mitosis. A recent publication suggests that alternating charged blocks are responsible (and sufficient) to drive Ki-67 phase separation and form “the periphery of mitotic chromosomes”, regardless if these blocks are obtained by phosphorylation of clustered sites or by phosphomimetic mutations (Yamazaki et al., 2022). This supports the idea of a more global and coarse regulatory function of phosphorylation, in which the addition of negative charges alone might control cell reorganization by modulating phase separation processes.

It is not currently feasible to study effects of CDK-inhibition on mitotic MLO phosphorylation and structure since CDK inhibition prevents or reverses entry into mitosis. Nevertheless, the data presented in this thesis will help to lay the theoretical foundations for stating the hypothesis that the function of a considerable proportion of CDK-mediated phosphorylation is to regulate phase separation of key IDRs of MLOs.

Last, I presented results of two different projects to which I contributed, aimed to understand the function of two biological systems intimately related with CDK-mediated regulation of MLOs.

The first is the regulation of chromatin organization by Ki-67 and the consequent changes in gene expression. We found that knockout of Ki-67 brings about massive changes in the transcriptome, although without any obvious cellular phenotype. We noticed extensive alterations in chromatin histone marks, especially the inhibitory H3K27 trimethylation associated with polycomb-mediated repression of gene expression. Nevertheless, those changes do not seem to fully explain the observed global reprogramming of transcription, and modifications in other chromatin regulatory mechanisms might be also responsible for this effect.

We showed that all stages of tumorigenesis were affected upon ablation of Ki-67. RNAseq analysis data revealed alterations of expression of genes involved in the EMT, antigen presentation, drug metabolism, and other cancer-associated hallmarks. Those were later confirmed by the failure to generate intestinal tumors in Ki-67 mutant mice; the reduced ability of some Ki-67 knockout cancer cells to induce angiogenesis; the presence of altered epithelial and mesenchymal characteristics in Ki-67-negative cancer cells; the decrease ability of the latter to colonize other tissues and give rise to metastases; their increased sensitivity to drugs; and their reduced interactions with the immune system.

We detected transcriptional changes of key regulators of the EMT pathway which is transcriptionally regulated by the PRC2 complex. However, inactivation of Suz12 did not restore tumorigenicity to Ki-67 knockout 4T1 cells. Importantly, similar cancer-associated pathways are deregulated in 4T1 and MDA-MB-231 cells, suggesting general conservation of regulatory mechanisms.

Ki-67 might act as an effector of CDK1-mediated regulation by localizing in the perinucleolar heterochromatin and controlling its organization and, consequently, general transcriptional programmes depending on this organization. Nevertheless, complete absence of Ki-67 does not generate any detectable changes in proliferation of any of the cellular models tested, which might indicate that these programmes are important only in specific conditions or that this entire system is heavily backed up.

The second biological system is the Mediator-dependent transcriptional regulation by the CDK8/19 subfamily. This highly conserved subfamily of CDKs, together with cyclin C, is part of the kinase module of the Mediator. A distinctive particularity of these proteins is that they present a disordered C-terminal domain that is predicted to have a prion-like composition. CDK8/19 are able to phosphorylate Pol II CTD, which is also intrinsically disordered. Taking into account that the Mediator and its interaction with enhancers is reported to be mediated by phase separation, we surmised that the kinase module might act as an internal control of this mechanism to regulate gene expression. Indeed, recent work shows evidence for Pol II CTD phosphorylation-mediated

regulation of the switch between condensates having transcriptional functions and those performing mRNA splicing (Guo et al., 2019). However, our results do not support an essential role for Mediator kinases in cell proliferation nor global regulation of transcription, in contrast to Mediator itself. We found that only relatively few genes were highly deregulated for double knockouts for mouse intestinal organoids and MEFs. We noticed a trend where slightly more genes were upregulated than downregulated upon loss of either CDK8 alone or both kinases, while effects of combined deletion were more than additive of effects of single deletions, indicating functional redundancy. In primary hepatic progenitor BMEL cells, the transcriptional effects were also minor considering how drastic the effects of the deletion of other members of the Mediator complex are (El Khattabi et al., 2019). Nevertheless, each one of these biological models showed phenotypical particularities after CDK8/19 knockouts. In intestinal organoids, double knockouts showed only a mild effect in proliferation and growth, although after some weeks of growing, the genotype was reverted, possibly because of negative selection of *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ organoids. Additionally, a cystic fibrosis-like phenotype was observed for the organoids lacking both kinases, with a transcriptional signature similar to that shown upon ablation of the *CFTR* gene in human cells. In BMEL cells, only single knockouts were performed and the deletion of CDK8 protected those cells against malignant transformation and tumorigenesis initiation, indicating that it is required for those processes, possibly by counteracting the effects of p53. Indeed, deregulated genes in CDK8/19 double knockouts in MEFs were clearly enriched in targets of p53 and other stress response pathways. Moreover, these cells also showed a slower growth rate, with accumulation of p21 and more cells in interphase.

In summary, I presented here the results of several collaborative projects aiming to contribute to development of a model for the cell cycle which can integrate new evidence that cannot be satisfactorily explained by the classic model. With the help of experts in other fields such as physics/biophysics, proteomics or transcription, I was able to answer key questions to further advance the quantitative model of the cell cycle. Here, I propose a mechanism that might explain how total CDK phosphorylation levels can regulate the cell cycle in a precise manner: Abrupt changes in self-regulated CDK activity phosphorylate proteins in MLOs, the biochemical centers of the cell, regulating their formation and dissolution in a cell cycle-dependent manner.
Some details about the biochemistry and the physicochemistry aspects have escaped the scope of our research, and a more thorough and comprehensive analysis of how CDK phosphorylation controls different MLOs can be envisaged as a perspective. However, based on personal interests, I would like to finish this work with an unanswered question about the transfer of

biological information. How can the most fundamental processes of the cell rely on the regulation of regions of proteins loosely conserved and presenting a high evolutionary rate? The fact that the passage of evolutionary information on how to regulate cellular organization during the cell cycle does not strongly rely on the primary amino acid sequence might be an explanation for why major mechanistic aspects remain beyond our comprehension. If the sole evolutionary constraint for a CDK target is to have IDRs that assure an accessible cluster of potential phosphorylation sites, it is expected that the specific sequence of protein regions regulated by CDK phosphorylation is not strictly conserved. Moreover, if regulation of the organization of MLOs through the cycle depends on total phosphorylation of the MLOs components, the loss of a phosphosite in one protein might be compensated by the gain of a phosphosite in another. This theory opens the door to a new understanding of protein evolution that does not require passing on strict sequence information to the next generation but rather to maintain a pool of accessible phosphorylatable sites independent of their specific location within the protein and might be independent of the specific protein within an MLO.

6 Materials and methods

Chapter 2

Plasmid cloning

ThermoFisher One Shot™ Stbl3™ Chemically Competent *E. coli* were used in all cloning experiments, unless otherwise specified. Bacteria were transformed using a heat shock protocol as indicated in the manufacturer manual. Plasmid DNA purification was performed with the QIAprep Spin Miniprep Kit for lower yields or with the NucleoBond AX 500 for higher yields, following the provided protocols by the manufacturer.

Genetic manipulation of plasmids

All restriction enzymes used are from New England Biolabs, in their “High Fidelity” (HF) version or compatible with CutSmart® buffer. The standard restriction protocol provided in the New England Biolabs website was used.

T4 ligase (NEB) and Antarctic phosphatase (NEB) reactions were performed as indicated by the vendor.

Genomic DNA extraction

Genomic DNA for the generation of the homology arms was extracted from a SUM 159 cell line, using the PureLink Genomic DNA extraction protocol.

PCR

All PCR reactions were performed using the Phusion™ High-Fidelity DNA Polymerase following the protocol provided by the vendor. In case of low primer specificity, reactions were repeated adding 20% of Betaine 5M and 3% DMSO to the reaction mix.

For the colony PCR, after transforming bacteria and plating them under selection, colonies were identified and partially scraped with a pick and used as template material for the PCR reactions. Each reaction tube had the content of 5 colonies. Colonies from tubes showing the expected band were then tested individually using a similar procedure.

For the Golden Gate Assembly reaction, the protocol provided in the New England Biolabs website was followed and adapted to our construct as indicated below:

ATP ligase buffer	3µl
10mM ATP	3µl
SapI enzyme	1µl
T4 ligase	1µl
Vector backbone (~150ng/µl)	1µl
LHA (~50ng/µl)	3µl
RHA (~50ng/µl)	3µl
H2O	15µl

The reaction was set in the thermocycler for 30 cycles of 5 min:37 °C / 5 min:16 °C; and a final cycle of 15 min:37 °C / 20 min:65 °C (held at 4 °C after the reaction finished).

The point mutagenesis PCR for correcting the plasmid was performed using the Agilent QuickChage II XL site-directed mutagenesis Kit, following the protocol provided by the vendor.

Primers:

puroR cassette

Fw- ACATCGATTGTGAAATTTGTGATGCTATTGC

Rv- ACATCGATGGTTACAAGACAGGTTTAAGG

Left homology arm

Fw-GTTTAAACGCTCTTCTGTGGGCCACACCAGACATTTCTATCATGTGA

Rv- GTTTAAACGCTCTTCTTACCTGATTTTTTTCATGGCTACCACTTGACC

Right homology arm

Fw- GTTTAAACGCTCTTCTTAGACTAGAAAGTGAAGAGGAAGGGGTTCCTAG

Rv- GTTTAAACGCTCTTCTTTATGGACTGTTAGGAAGACAGGTCTCCAAA

Site-directed mutagenesis

Fw-GTATTGGGCGCTATTCCGCTTCCTCGCTCACTG

Rv-CAGTGAGCGAGGAAGCGGAATAGCGCCCAATAC

Colony PCR

Fw-ACAAGTGCCATAGAGCCCAC

Rv-AGCAACAGATGGAAGGCCTC

Gel electrophoresis

All DNA gel electrophoresis were performed in 0.8%-1% agarose gels (depending on the required resolution) using SYBR Safe DNA Gel Stain and an UV transilluminator to detect the bands.

Cell culture

All cellular cultures were maintained at 37 °C and at 21% O2 and 5% CO2.

HAP1 cells were cultured in IMDM (Glibco) with 10% fetal bovine serum (FBS) and 1% Penicillin/Streptomycin. HCT116 were cultured in DMEM supplemented with GlutaMax ® (Glibco), 10% FBS and 1% Penicillin/Streptomycin. SUM159 cells required special culture media based on HAM F12 media, supplemented with 5% 10% FBS, 10 g/ml insulin, 1 g/ml hydrocortisone and 1% Penicillin/Streptomycin

Transfection

Co-transfection of recombinant Cas9 proteins (Synthego), guide RNAs (Synthego) and the vector containing the template DNA for the homologous recombination-mediated knock-in were performed with the Lipofectamine™ CRISPRMAX™ Cas9 Transfection Reagent. The protocol provided by Synthego was adapted as follows:

Three conditions were established for the transfection Cas9 + gRNA + CDK1-CyclinB1 vector, Cas9 + gRNA + CDK1-CyclinA2 vector and Cas9 + gRNA. For each of those, three different guides were tested in different reactions plus a reaction with no gRNA. For each one of those 12 total reactions 0.5 µl of the Cas9 Plus™reagent and 0.5 µl of the Cas9 protein (0.3 µM) were mixed in 12.5µl Opti-MEM™. Then 0.65 µl of the corresponding gRNA (or none, for the controls) were added to the reaction. For the conditions with donor vectors, two final concentrations of plasmid DNA were tested by adding 2 µl and 4 µl, corresponding to 500 ng and 1 µg of material, respectively. Tubes with all the components were incubated for 10 minutes and then 13.25 µl of the solution containing the transfection agent (12.5 µl Opti-MEM™ + 0.75 µl CRISPRMAX™ Reagent) was added to the tubes and mixed by pipetting. After 15 min incubation, the complete transfection solution was mixed with $5x10^5$ cells in 250 µl of media and plated for 24 hours, after which a change of media was required. Selection with Puromycin was started 48 hours after plating the cells.

For lipofection of vectors containing the CDK-cyclin constructs, Lipofectamine™ 2000 Transfection Reagent was used following the protocol provided by the vendor. Two transfection times were tested, 4 hours and 24 hours. After those timepoints the culture media containing the reagents was replaced by fresh media.

Electroporation

The Neon™ transfection system was used to deliver the RNP complexes and the donor plasmid into cells, following the protocol provided by the vendor. The RNP complexes were assembled in the resuspension buffer with the donor DNA as follows:

sgRNA	3 µl (90 pmol)
Cas9	0.5 µl (10 pmol)
Donor DNA	2 µl (2-5 µg/µl)
Resuspension buffer	1.5 µl

In all cases 1-2 x 10^6 cells were transfected with a sgRNA-to-Cas9 ratio of 9:1 and the programs used are indicated below:

Cell line	Pulse voltage (V)	Pulse duration (ms)	Number of pulses
HCT-116	1530	30	1
HAP1	1575	10	3
SUM159	1100	30	2

Software and sequence analysis

Sequence analysis and visualization was performed using the SnapGene Viewer and Unipro UGENE (Okonechnikov et al., 2012) softwares. All primers were designed using the Primer3 Plus web service (Untergasser et al., 2012).

Chapter 3

Data collection for human and yeast CDK1 targets.

Data of CDK1 substrates for *S. cerevisiae* were downloaded from online supplementary information of papers describing two different studies using *in vitro* (Ubersax et al., 2003) and *in vivo* (Holt et al. 2009) approaches, respectively. High confidence yeast CDK1 targets were defined as the intersection of both datasets. Other phosphorylations detected in both studies for which there was no evidence for CDK1 involvement were considered as the non-CDK1-mediated phosphoproteome (universe). For human CDK1 subfamily targets, we extracted information available in the PhosphoSitePlus database (Hornbeck et al., 2015). An additional step of manual curation from several studies (Blethrow et al., 2008; Chi et al., 2008; Hégarat et al., 2020; Wyatt et al. 2013; Orthwein et al., 2014; Linder et al., 2017; Liu et al., 2000; Chi et al., 2020; Klein et al.,

2008; Goto et al., 2006; Curtis et al., 2002; Fourest-Lieuvin et al., 2006; Milner et al., 1993; Lowe et al., 1998; Yun et al., 2003; Kitzmann et al., 1999; Thiel et al., 2002; Li et al., 2006) was performed to obtain a high confidence human CDK1 subfamily targets dataset. The phosphoproteome universe was constructed with all the phosphorylated proteins deposited in the PhosphoSitePlus database after subtraction of the CDK1 subfamily targets. Information for other kinases (MAPK, AURK, PLK, NEK and DYRK) were also extracted from PhosphoSitePlus and manually curated, with no additional data retrieved from bibliography.

Prediction of intrinsically disordered regions.

For the UniProt proteomes of human, yeast and *Xenopus laevis*, disorder information was fetched from MobiDB (Piovesan et al., 2021) with the exception of SPOT disorder predictor, which was calculated for all the proteins of each dataset. For *Xenopus* proteomics studies, where a non-UniProt database was used, the available standalone software of IUPred, VSL2B, and SPOT was used to predict IDRs in all proteins of the database. The same procedure was applied to human and yeast proteomes, to avoid possible inconsistencies in the prediction that might arise from the utilization of different versions of predictors.

Differential disorder composition.

For the three organisms analyzed (*Xenopus*, human and yeast), the amino acid composition for the entire phosphoproteome and for the disordered regions of the phosphoproteome was calculated. For each amino acid, we estimated the differential disorder composition with the equation:

$$Differential\ composition = \frac{composition\ in\ disordered\ regions\ -\ composition\ in\ the\ phosphoproteome}{composition\ in\ the\ phosphoproteome}$$

Positives values show amino acids enriched in disordered regions while negative values represent aminoacids depleted in disordered regions.

Bioinformatic and statistical analysis of disorder and phosphorylation.

All statistical analysis was performed with the R programming language (https://www.r-project.org/) using R studio as an integrated development environment (available at https://rstudio.com/). The packages Tidyverse (Wickham et al., 2019) and Bioconductor (Gentleman et al., 2004) were used for cleaning, manipulation, and graphical representation of the data. Sequence logos were generated using the information content as described (Douglasset al., 2012). Disorder scores were plotted with an ad hoc designed script.

For the contingency table analysis, the disordered regions of human CDK, MAPK, AURK, PLK, NEK and DYRK targets, as well as yeast CDK targets and *Xenopus* dynamic phosphoproteins were calculated with three predictors (IUPred, VSL2B, and SPOT) using their standalone distributable software. For each combination of disorder predictor and phosphorylation dataset, a two-by-two table with the counts of phosphorylatable residues (Ser/Thr) phosphorylated or not, either located in IDRs or in structured regions, was generated. Each table was then analyzed with the Fisher test for estimating the odds ratio and the associated P-value.
The source code for all the analysis conducted in this thesis is available upon request.

Egg collection and in vitro fertilization.

Female *X. laevis* frogs were primed with 50 international units (IU) of human chorionic gonadotropin at least 2 days, and no more than 7-8 days before a secondary injection with 625 IU to induce ovulation. Roughly 16 hours after the second injection, fresh eggs were collected by pelvic massage and kept in 1x Marc's Modified Ringer's (MMR). Next, eggs were placed in a petri dish and checked under the microscope to keep only those that exhibited the healthy pigment pattern (dark animal pole and white vegetal pole).
To perform the *in vitro* fertilization, around 1/3 of a full testis was cut into fine pieces and mixed within 500 µl of 1x MMR. The suspension was pipetted up and down until big clumps were dissolved. Next, buffer was removed from the petri dish and the eggs were collected. Once eggs were well dispersed across the dish, the sperm suspension was added. The dish was then flooded with 0.1x MMR to induce fertilization.

Sample collection.

The first time point was collected immediately before adding the sperm suspension (time 0' corresponds to the unfertilized egg). Eggs were kept at room temperature (18-20°) and under a dissecting microscope after fertilization. At approximately 15 minutes, fertilized eggs underwent shrinkage of the animal hemisphere and rotation within the vitelline membrane, so that the animal hemisphere faced upwards. These changes are known indicators of successful fertilization, so only the eggs that underwent these changes were used for the experiment.
Samples were collected approximately every 15 minutes. Eggs were rapidly placed in individual tubes and snap frozen in liquid nitrogen, to preserve the phosphorylation events occurring at that specific time. Since eggs were monitored under the microscope, we were able to determine if samples were collected before or after a cell division had occurred.

***Xenopus* egg extracts.**

Interphase *Xenopus* egg extracts were prepared, and DNA replication time courses performed, as described previously (Parisis et al., 2017).Mitosis was induced in extracts by adding recombinant GST-Cyclin BΔ90 (40 ng/ml). The CSF extract was a kind gift from Arianne Abrieu, from CRBM-Montpellier.

Embryo lysis.

For cellular lysis of embryos, a similar approach was used to that described by Lindeboom et al. (Lindeboom et al., 2018). Briefly, each sample was thawed and homogenized with 15 µL of ice-cold cell lysis buffer (20mM Tris-HCl pH 8.0, 70mM KCl, 1mM EDTA, 10% glycerol, 5mM DTT, 0.125% Nonidet P-40, 1mM PMSF, 1x complete EDTA-free protease inhibitor, 1xPhoStop). Samples were subsequently centrifuged at max speed on a benchtop Eppendorf centrifuge. 10 µL of soluble material was recovered and snap-frozen in liquid nitrogen. Samples were stored at -80°C until further processing.

Protein digestion and phosphopeptide enrichment.

Cell lysates were digested using the FASP method (Wiśniewski et al., 2009). Briefly: proteins were thawed and immediately reduced and alkylated with 10mM DTT and 0.05M iodoacetamide. Next, proteins were digested with Lys-C (overnight) at 37°C in a wet chamber, followed by addition of trypsin and further incubation under the same conditions for 4 hours. Both enzymes were used at 1:50 enzyme to protein ratio (protein quantification by Bradford assay showed that each individual egg provides ~20 µg of yolk free protein). For the egg extract experiment, each FASP filter was loaded with 200µg of protein. Peptides were cleaned using the Oasis HLB 96 well plates (Waters Corporation) and consequently subjected to phosphopeptide enrichment using Fe(III)-NTA 5 µl cartridges in the automated AssayMAP Bravo Platform (Agilent Technologies), as described by Post et al (Post et al., 2017). Both flow through (peptides) and eluates (phosphopeptides) were dried down and stored at -80°C until further use.

LC-MS/MS analysis.

All samples for label-free shotgun proteomics were analyzed using a UHPLC 1290 system (Agilent Technologies) coupled to an Orbitrap Q Exactive HF mass spectrometer (Thermo Fisher

Scientific). Nano flow rate was achieved using a split flow setup aided by an external valve as described by Meiring and colleagues (Meiring et al., 2002). Peptides were first trapped onto a pre-column (inner diameter [ID] of 100 µm and 2 cm length; packed in-house with 3µm C18 ReproSil particles [Dr. Maisch GmbH]) and eluted for separation into an analytical column (ID of 75 µm and 50cm length; packed in-house with 2.7 µm Poroshell EC-C18 particles (Agilent Technologies). The latter was done using a two-buffer system, consisting of buffer A (0.1% formic acid [FA] in water) and buffer B (0.1% FA in 80% ACN). Peptides were trapped during 5 minutes at 5 µl/min flow-rate with solvent A. For the measurement of the full proteome, we used a 155 min gradient from 10 to 36% of solvent B. For the phosphoproteome, we used a 95 min gradient from 8 to 32% of solvent B. Both methods included a wash with 100% solvent B for 5 minutes followed by a column equilibration with 100% solvent A for the last 10 minutes.

The mass spectrometer was operated in data dependent acquisition (DDA) mode. Full scan MS was acquired from 375-1600 m/z with a 60,000 resolution at 200 m/z. Accumulation target value was set to 3e6 ions with a maximum injection time of 20 ms. Up to 15 (12 for the phosphoproteome) of the most intense precursor ions were isolated (1.4m/z window) for fragmentation using high energy collision induced dissociation (HCD) with a normalized collision energy of 27. For MS2 scans an accumulation target value of 1e5 ions and a maximum injection time of 50 ms were selected. Scans were acquired from 200-2000m/z with a 30,000 resolution at 200m/z. Dynamic exclusion was set at 24s for the proteome and 12 s for the phosphoproteome.
For targeted proteomics, an EASY-nLC 1200 System (Thermo Fisher Scientific) coupled to an Orbitrap Q Exactive HF was used. Peptides were separated using an EASY-Spray analytical column (ID of 75µm and 25cm length; packed with 2µm C18 particles with a 100 Å pore size) (Thermo Fisher Scientific). Gradient lengths were shortened to 60 minutes. Phosphopeptides of interest from the previous experiment were selected and heavy-labeled versions were synthesized (JPT Peptide Technologies). These synthetic standards were used during method development for retention time scheduling and instrument ion fill-time optimisation. Additionally, synthetic heavy peptides were pooled and combined with synthetic retention time peptide standards (iRT, Biognosys) to generate a spectral library, measured in DDA mode using the same LC-MS setup. This spectral library provided fragment intensity and retention time information for quality control assessment of targeted measurements. Samples were reconstituted in 2% FA containing ~200fmol of each synthetic standard. The mass spectrometer was operated in data independent acquisition mode with an inclusion list of targets for parallel reaction monitoring (PRM). The list included the m/z values for the heavy and light versions of the phosphopeptides. Optimal measurement parameters were determined using test samples spiked with the heavy-

labeled standards in order to guarantee optimal sensitivity for detection of endogenous phosphopeptides. We measured the targets of interest in a scheduled fashion, during a four-minute window with a 120,000 resolution, maximum injection time of 246ms and an accumulation target value of 2e5 ions, to ensure maximum specificity and sensitivity.

Data processing.

DDA raw files were processed with MaxQuant56 (v1.6.0.1) using a false discovery rate (FDR) <0.01. The default settings were used, with the following exceptions: variable modifications, specifically methionine oxidation, protein N-term acetylation and serine, threonine and tyrosine phosphorylation were selected. Cysteine carbamidomethylation was selected as a fixed modification. We also enabled the 'match between runs' option with the default values. Fractions were set so that matching was done only among biological replicates and samples of consecutive time points. The database search was conducted against a database generated by Temu and colleagues (Temu et al., 2016). This was particularly helpful since other publicly available databases contained several incomplete and/or poorly annotated sequences, which proved to be impractical for further data analysis.

The data was uploaded to the Perseus platform (Tyanova et al., 2016) for further analysis. Briefly: decoy sequences and potential contaminants were filtered out. Only high confidence localisation (>0.75 localisation probability) phosphosites were conserved for further analysis. Intensities were log2 transformed and then normalized by subtracting the median intensity of each sample. Biological replicates were grouped accordingly by time point; this grouping allowed us to filter the data and keep only those phosphosites that could be detected in at least two out of three biological replicates in any of the time points. Missing values were imputed from a random normal distribution applying a downshift of 1.8 times the standard deviation of the dataset, and a width of 0.3 times the standard deviation. This effectively replaced missing values at the lower end of the intensity distribution. We then performed an ANOVA (Benjamini-Hochberg FDR <0.05) to determine which phosphosites displayed statistically significant changes through the time course. Average phosphosite intensities were grouped using a combination of k-means and hierarchical clustering using the ComplexHeatmap package (Gu et al., 2016) in R. Protein intensities were processed in a similar fashion, removing proteins that were only identified by peptides that carry one or more modified amino acids.

Next, the full list of proteins with significantly changing phosphosites were matched against the human Uniprot database using the Basic Local Alignment Search Tool (BLAST), to render Uniprot identifiers that were compatible with different Gene Ontology (GO) analysis tools. We used the

STRING web tool (Szklarczyk et al, 2019) to gain insight into the relation amongst dynamically phosphorylated proteins. The full list of phosphoproteins was uploaded and analyzed using default settings. Next, the protein network was loaded into Cytoscape for clustering and visualization. Proteins were clustered using GLay community clustering (Su et al., 2010) and enrichment of GO terms per cluster was obtained using BiNGO(Maere et al., 2005).

GO term enrichment was also acquired individually for each group (A-D) obtained after hierarchical clustering of dynamic phosphosites. For this we used STRING and filtered the enriched GO terms to keep only those with $p < 0.01$, fold enrichment >5 and a minimum of 5 proteins per term. The list of terms was further condensed by removal of redundant terms using the Revigo web tool (Supek et al., 2011). Remaining GO terms (including BP, MF and CC) were manually curated to further avoid redundancy. Following the same strategy, we analyzed GO term enrichment for the interphase and mitotic clusters from the in vitro dataset separately.

PRM raw files were analyzed with Skyline software. Signal quality for each target of interest was assessed visually for all samples. Quality control of endogenous signals was done by confirming the perfect co-elution of both peptide forms (heavy and light), assessing their retention time and peak shape. We also used the similarity of the relative intensity of fragment ions (rdotp > 0.9) between light and heavy to exclude signals that showed poor correlation. Quantifications were done with a minimum of three fragments per phosphopeptide. The data was loaded into R for data cleanup and visualization, using the Complex-Heatmap and ggplot (Wickham et al., 2019) packages.

Motif analysis.

The obtained phosphopeptides were aligned by centering them around the phosphosite detected and the conserved motifs for the different kinases were determined using regular expressions by applying the following rules:

PLK: [D/N/E/Y]-X-[S/T]-[Hydrophobic / ^P]

AURA/AURB: [K/R]-X-[S/T]*[^P]

NEK: [L|M|F|W]-X-[S|T]*-[A|V|I||L|F|W|Y|M]-[K|R]

Casein kinase 1: [D/E]-[D/E]-[D/E]-X-X-[S/T]* or [S/T]-X-X-[S/T]*

Casein kinase 2:[S/T]-[S/T]*-X-[E/D/S]

DDK: [S/T]*-[E/D]-X-[E/D] or [S/T]*-[S/T]-P

PKA: R-[R/K]-X-[S/T]*-[Hydrophobic]

Cdk full consensus motif: [S/T]*-P-X-[K/R]

Cdk minimal consensus motif: [S/T]*-P

Where [] groups multiple amino acids for one position, ^ at the left of a certain amino acid informs that it is forbidden for that position, and X represents any amino acid.

Data collection for MLO proteomes.

Data from proteomics studies of the composition of MLOs characterized by liquid-liquid phase separation were obtained from the following sources: stress granules (Jain et al., 2016; Fong et al., 2013), nuclear speckles (Fong et al., 2013; Dopie et al., 2020), PML nuclear bodies ((Fong et al., 2013; Liu et al., 2010), P-bodies (Hubstenbergeret al., 2017), nucleoli (Stenströmet al., 2020; Tafforeau et al., 2013), nuclear pore complexes (Lin and Hoelz, 2019), Cajal bodies (Fong et al., 2013; Machyna et al., 2013), Super-enhancer-Mediator condensates (Quevedo et al., 2019).

Calculation of SCDMs.

Elements (i,j) of SCDM were calculated using equation 1 of Huihui and Ghosh (Huihui and Ghosh, 2021). In this coarse grain model, each amino acid is considered a point with a charge q = -1 for Aspartic and Glutamic acid, charge q = 1 for Arginine and Lysine, charge q = 0 for all other amino acids. Phosphorylation is modeled by replacing neutral charge of Serine, Threonine to q = -2 to mimic the effect of double negative charge of phosphate groups. SCDM maps were made visually continuous between neighboring (i,j) pairs using spline-16 interpolation.

All atom simulations.

All-atom Monte Carlo simulation were performed for MCM4, TICRR, CDT1, Coilin (unphosphorylated and phosphorylated forms) to compute radius of gyration by using CAMPARI (version 2) based on the ABSINTH implicit solvation paradigm (Vitalis and Pappu, 2009a; Vitalis and Pappu, 2009b). Simulations were carried out at 298 K. For CDT1, 16 trajectories were generated with each trajectory having 10 million steps. For MCM4, 27 trajectories were generated with each trajectory having 6.5 million steps. For Coilin, 54 trajectories were generated with each trajectory running for 4 million steps. For TICRR, 90 trajectories were generated with each trajectory having 3 million steps. For each trajectory, irrespective of the sequence, first 1.5 million steps were discarded due to equilibration yielding a cumulative (over all trajectories) of at least 135 million steps for each sequence. Each of the simulations were run using zero salt with only neutralizing Na+ and/or Cl- ions added in a droplet of 400 Angstrom. PDB files were generated every 5000 Monte Carlo steps. Phosphorylated sequences were modeled by replacing Serine or Threonine by two Aspartic acids. ACE cap and NME tail were added to each protein sequence.

Coarse-grained force field.

A one-bead-per-residue coarse-grained (CG) model was adopted that has been shown to capture the structural and phase separation properties of flexible proteins as a function of their sequence (Dignon et al., 2018a) and phosphorylation pattern (Regy et al., 2021). In this framework, bonded interactions between neighboring residues were modeled using a harmonic potential with a constant of 1000 kJ/mol/nm2 and a bond length of 0.38 nm. Electrostatic interactions between charged residues, i.e. Asp, Glu (-1e); Lys, Arg (+1e), His (+0.5e); ph-Ser, ph-Thr (-2e), were computed with a screened Coulomb potential, using a Debye-Huckel length of 1 nm. Short-range interactions were modeled by Lennard-Jones (LJ) potentials defined by

$$V_{ij}(r) = 4\lambda_{ij}\epsilon\left(\left(\sigma_{ij}/r\right)^{12} - \left(\sigma_{ij}/r\right)^{6}\right)$$

where values of λij, ϵ, and σij used reported values (Dignon et al., 2018a; Regy et al., 2021) with the exception of λArgArg that was set equal to 0.01, instead of 0.00, in order not to neglect excluded volume effects while using LJ functional form. In the simulation of full-length Ki-67, the conformation of the N-terminal folded domain (1-128) was restrained by an elastic network based on experimental structure (PDB: 1R21) with a distance cutoff of 2 nm and an elastic constant of 5000 kJ/mol/nm2. All MD simulations were performed with GROMACS 2018.3 (Abraham et al., 2015). Simulations were run in the NVT ensemble, controlling the temperature by means of a Langevin thermostat with a friction constant of 25 ps-1 and a time step of 10 fs.

Single chain MD simulations of full-length Ki-67.

The structure with PDB id 1R21 was used to model the conformation of the first 128 residues of the full-length Ki-67, while TraDES (Feldman and Hogue, 2000; Feldman and Hogue, 2002) was used to generate the initial conformation of the remaining disordered part of the non-phosphorylated molecule. The models were then joined using UCSF Chimera 1.14 (Pettersen et al., 2004). The initial configuration for the phosphorylated full-length Ki-67 was obtained by 'mutating' (i.e. using parameters that have been tuned specifically for ph-Ser/ph-Thr) the amino acids in the non-phosphorylated structure. Initial configurations were inserted in a cubic simulation box with a side of 100 nm with periodic boundary conditions. We employed a Parallel Tempering (PT-MD) scheme, with eleven replicas in the 300-400K range in order to enhance the conformational sampling and obtain a reliable estimate of the radius of gyration of the protein as a function of the temperature. PT-MD simulation ran for 1e8 steps, attempting Monte Carlo exchanges between neighboring replicas every 1e2 steps and saving snapshots every 1e4 steps for further analysis.

Single chain MD simulations of consensus repeats.

The initial extended configuration of the non-phosphorylated consensus repeat of Ki-67 was generated with the tLEAP tool available in AmberTools18, while the phosphorylated structure was generated by 'mutating' (i.e. using parameters that have been tuned specifically for ph-Ser/ph-Thr) the residues in the non-phosphorylated configuration. PT-MD simulation followed the same protocol employed for the simulations of full-length Ki-67, with the only difference being the range of temperatures, between 200K and 500K, with intervals of 10K.

Phase Coexistence MD simulations.

Phase coexistence simulations of monomers and dimers of the non-phosphorylated consensus repeat and of monomers of the phosphorylated consensus repeat were carried out employing the slab method (Dignon et al., 2018a). Initial configurations of the slab simulations for each system were generated by inserting 200 copies of the respective molecules in a 20x20x30nm box in random positions and orientations with the gmx insert-molecules tool available in GROMACS 2018.3 (Abraham et al., 2015), and extending the simulation box in the z direction to 200 nm. Simulations were run for 1.5e8 steps, saving frames every 10000 steps for further analysis, discarding the first 0.5e8 steps of the simulations as equilibration.

Estimation of theta temperatures.

Coil-to-globule transition temperatures (Tθ) were estimated for non-phosphorylated full-length Ki-67 and the monomer of the non-phosphorylated consensus repeat from single chain PT-MD simulations, following a reported approach (Dignon et al., 2018b). Intramolecular distances as a function of the chain separation |i-j| were computed for each temperature replica and fitted with the following expression:

$$R_{ij} = 0.55|i-j|^{\nu}$$

where the scaling exponent ν is the fitting parameter. We then determined Tθ as the temperature corresponding to ν = 0.5

Estimation of binodal curves.

Density profiles from phase coexistence simulations were estimated by means of the gmx density tool available in GROMACS 2018.3, centering the dense phase in the middle of the z-axis. The concentration of the diluted phase was computed by averaging the density in the box at

0nm<z<60nm and 140nm<z<200nm, while the concentration of the dense phase was obtained by averaging the density at 90nm<z<110nm. Following the approach of (Dignon et al., 2018a) the critical temperatures, Tc, for the three systems investigated with phase coexistence simulations, have been evaluated by fitting the following equation:

$$\rho_H - \rho_L = A(T_c - T)^{0.325}$$

Human opto-Ki-67 plasmid construction.

pCDNA5_FRT_Ki67-FL-mCherry-Cry2 was generated by PCR amplification of human Ki67. The fragment was cloned into the AflII/KpnI digested pCDNA5_FRT_TO_TurboID-mCherry-Cry2 (Addgene plasmid # 166504) using the In-Fusion HD Cloning Kit protocol.

Generation of Flp-In T-REx 293 opto-Ki67 cell lines.

Flp-In T-REx 293 (Termo Fisher Scientifc, Darmstadt, Germany) cell line was grown under standard conditions (37˚C, 5% CO2) in Dulbecco's modified Eagle's medium (Merck-Sigma-Aldrich, D5796). The medium was supplemented with 10% fetal bovine serum (FBS), 100 µg/ ml Zeocin and 15 µg/ml Blasticidin. One million HEK-293 T-REx cells were plated in a 6-well plate 24 hours before transfection. The next day, 500 ng of each optoKi67 expression plasmid is combined with 3.5 µg pOG44 encoding the Flp recombinase. Transfection was made with 8 µl Lipofectamine 2000 according to the manufacturer's instructions. 48 hours post transfection, cells were transferred to a 100 mm petri dish. On the next day, selection was performed by adding hygromycin B at a final concentration of 50 µg/mL. Around 14 days after selection, clones were pooled and expanded. The cells were tested for the expression of the construct by immunofluorescence. Flp-InT-Rex 293 opto-Ki67 stable cell lines were maintained with 15 µg/mL Blasticidin and 15 µg/mL Hygromycin.

Opto-Ki-67 activation.

Cells were plated in DMEM on coverslips a day prior to activation. Expression of opto-Ki67 was induced with 2 µg/ml doxycycline for 16 hours. For light activation, plates were transferred into a custom-made illumination box containing an array of 24 LEDs (488nm) delivering 10 mW/cm2 (light intensity measured using a ThorLabs-PM16-121-power meter). Cry2 activation was induced using 4 min of blue light cycles: 4s On followed by 10s Off. Cells were fixed with 4% paraformaldehyde (PFA) for 15 min at RT, counterstained with Hoechst (Invitrogen, Cat H21491)

in PBS-TritonX (0.2%), and mounted on glass slides using Prolong Gold antifade reagent (Invitrogen, Cat P36930). Images were captured using a 63x objective (NA 1.46 oil).
To check the effect of inhibitors on Ki-67 foci formation, the cells were incubated for 1h with 0.5 µM okadaic acid, 5 µM purvalanol A or vehicle (DMSO) for 1 hour prior to light activation and fixation. Foci number was analyzed using FIJI Software and statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise post-hoc comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method. Plots were generated using the ggplot2 library in R.

Ki-67 consensus repeat DNA construct.

The cDNA sequence coding for Ki-67 consensus repeat (Ki-67CR) was ordered from IDT® gene synthesis. It was subsequently cloned into pDB-GFP plasmid between HindIII and XhoI sites to obtain the pDB-GFP-Ki-67-CR vector. In this construct, Ki-67-CR was fused with a (his)6-GFP N-terminal tag. GFP sequence is followed by the HRV 3C (3C) protease recognition site (Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro). Specific cleavage can occur between Gln and Gly, with Gly-Pro remaining at the C terminus Ki-67-CR without any tag.

Ki-67 consensus repeat expression and purification.

The pDB-GFP-Ki-67-CR plasmid was transformed into *E. coli* BL21(DE3); transformed cells were grown overnight at 25°C in N-5052 auto-induced medium (Studier, 2005), supplemented with 50 µg/ml kanamycin and 15N NH4Cl. Cells were harvested by 20 min centrifugation at 6000g at 4°C. The pellet was resuspended in 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 2mM DTT (buffer A) and stored at -80°C. Cells were supplemented with a Complete® EDTA free tablet (Roche), lysed by sonication, insoluble proteins and cell debris were sedimented by centrifugation at 40000g at 4°C for 30 min. Supernatant was supplemented with imidazole to 5 mM final and loaded onto 5ml gravity affinity columns (Ni Sepharose Excel 5ml, Cytiva), equilibrated with buffer A. Columns were washed with 50 ml of buffer A and proteins were eluted with a one-step gradient of buffer B (buffer A containing 500 mM imidazole). The peak fractions were analyzed by SDS-PAGE. Fractions containing tagged Ki-67-CR were pooled and dialysed overnight at 4°C against buffer C (50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT). The dialysed protein was then loaded on a Superdex S200 16/60 (HiLoad 16/600 Superdex 200pg, Cytiva) equilibrated with buffer C. Fractions containing the protein of interest were analyzed by SDS-PAGE and pooled. The purified GFP-Ki-67-CR protein was concentrated to 5 mg/ml with Vivaspin centrifuge concentrator (Sartorius Stedim Biotech).

***In vitro* Ki-67 peptide phosphorylation assay.**

Protein was desalted by using PD10 Mini-Trap column in 50 mM Hepes 7.5, 50 mM NaCl, 2 mM DTT. Phosphorylation reaction was performed in a total volume of 220 µl, and contained 140 µl of GFP-Ki-67-CR (400 µg), 5mM MgCl2, 500 µM ATP, 50mM β-glycerophosphate, recombinant CDK1-cyclin B/CKS1 (5 µg; gift from Jane Endicott and Tony Ly) in 50 mM Hepes pH 7.5, 50 mM NaCl, 2 mM DTT. Reaction was incubated at 30°C for 18 hours and stopped by adding 10 mM EDTA. It was subsequently desalted by using PD10 Mini-Trap column in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT, and used to perform NMR experiments, Phos-tag SDS-PAGE and phosphoproteomics.

NMR experiments and data analysis.

All NMR samples contained final concentrations of 10% D2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Experiments were performed at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient coil, operating at a 1H frequency of 700 and 800 MHz. 15N-HSQC was acquired for each sample in order to determine amide (1HN and 15N) chemical shifts of non-phosphorylated and phosphorylated GFP-tagged Ki-67-CR. 15N-HSQC spectra were acquired for respectively non-phosphorylated (and phosphorylated) proteins at 800 (700) MHz using 32 (128) scans, 128 (256) increments and a spectral width of 22.5 ppm in the indirect dimension. All spectra were processed with TopSpin v3.5 (Bruker Biospin) and analyzed using CCPN-Analysis software (Vranken et al., 2005). Chemical shifts were referenced with respect to the H2O signal relative to DSS using the 1H/X frequency ratio of the zero point according to Markley and colleagues (Markley et al., 1998)

Total internal reflection fluorescence (TIRF) experiments.

In order to measure the ability of the phosphorylated and non-phosphorylated Ki-67 repeat to liquid-liquid phase separate, 50 µM protein was prepared in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT. Dextran was added just before preparing samples on circular glass coverslips (2.5 cm, 165 µm thick, Marienfeld). Coverslips were cleaned with a 15 min cycle of sonication with ultrasounds in 1M KOH, followed by a second cycle of sonication in deionized water. Samples were deposited into wells of Press-to-seal silicone isolator with adhesive (Invitrogen), and covered with a second coverslip to avoid evaporation. Images were acquired with a custom-made TIRF

microscope using a LX 488-50 OBIS laser source (Coherent). Oil immersion objective with a 1.4 numerical aperture (Plan-Apochromat 100x, Zeiss) was used. Fluorescence was collected with an EmCCD iXon Ultra897 (Andor) camera. The setup includes a 1.5x telescope to obtain a final imaging magnification of 150-fold, corresponding to a camera pixel size of 81.3 nm. Fluorescence images at different time points were obtained by averaging 150 individual images, each acquired over 50 ms exposure time.

Phos-tag SDS-PAGE.

For Phos-tag SDS-PAGE (12.5% Phos-tagTM SuperSepTM pre-cast gel 50 µmol/L; Fujifilm Wako Chemicals #193-16571), 500 ng of unphosphorylated and 500 ng of CDK1-cyclin B-CKS1 phosphorylated GFP-Ki-67-CR protein were loaded. SDS-PAGE was performed following standard protocol, with the exception of two additional washes 20 min each in transfer buffer containing 10 mM EDTA, followed by a wash in transfer buffer, preceding wet transfer. For Western blot, anti-GFP antibody (rabbit polyclonal Chromotek PABG1; 1:10 000) was used.

Mapping of Ki-67-CR phosphorylation sites by mass spectrometry.

6.5 µg of GFP-Ki-67-CR protein phosphorylated by CDK1-cyclin B-CKS1 were digested in a FASP filter as described earlier for the other samples. Peptides were subsequently cleaned using C18 cartridges and phosphor-enriched using Fe(III)-NTA cartridges in the AssayMAP Bravo. Phosphopeptides were measured in a technical duplicate, acquiring in DDA mode with an Ultimate 3000 uHPLC system coupled to an Orbitrap Exploris 480 (Thermo Fisher Scientific) during a 60 minutes gradient. Raw files were searched against the Ki-67-CR sequence using MaxQuant with the same parameters as applied to the other phosphoproteomics experiments.

Chapter 4

Cell culture and mice.

4T1 cells were provided by Robert Hipskind (IGMM, Montpellier); MDA-MB-231 cells were obtained from SIRIC, Montpellier. NIH 3T3, 4T1, MDA-MB-231, HeLa S3 and MCF-7 cells were grown in Dulbecco modified Eagle medium (DMEM - high glucose, pyruvate, GlutaMAX – Gibco® Life LifeTechnologies) supplemented with 10% fetal bovine serum (SIGMA or DUTSCHER). Cells were grown under standard conditions at 37°C in humidified incubator containing 5% CO2. 6-8 weeks-old female athymic nude (Hsd:Athymic Foxn1nu/Foxn1+), and NOD.SCID (NOD.CB17-Prkdcscid/NCrHsd) mice were purchased from Envigo.

Cdk8lox/lox mice were generated as follows: An 8076 bp genomic fragment (mouse chromosome 5: 146,254,503 to 146,262,579) enclosing the essential exon 2 (whose deletion results in loss of the essential catalytic lysine residue and causes a frameshift truncating over 90% of the protein) of the CDK8 gene was amplified by PCR from genomic DNA of 129/Sv embryonic stem cells and cloned into pGEM-T-easy. The diphtheria toxin A gene was cloned into the SacII site. 64 bp to the 3' of exon 2, the sequence CTCTAT was mutated to CTCGAG, generating an XhoI site. LoxP sites flanking exon 2 were generated by a combination of conventional cloning and recombineering, using a recombineering approach (Liu et al, 2003). The loxP PGK-Neo cassette was amplified from pL452 plasmid with flanking AvrII/HindIII sites at each end and cloned into the AvrII site upstream of exon 2. Fragment orientation was confirmed by the generation of 3.5 kb HindIII and 2.0 kb NheI sites, and the vector was recombined in E. coli strain SW106 with inducible Cre recombinase expression followed by HindII digestion, generating a single loxP site upstream of exon 2. Into this recombined vector, the FRT-PGK-Neo-FRT-LoxP cassette (amplified from pL451 with flanking XhoI sites) was cloned in the newly generated XhoI site downstream of exon 2, resulting in the "deletion construct". The orientation was confirmed by the generation of 2.2 kb Nhe1 and 3.4 kb BamHI sites. Functionality of the two recombination sites was tested as follows: the FRT site was confirmed by recombination in E. coli strain SW105 with inducible FlpE recombinase expression, deleting the FRT-Neo cassette and generating a 1.4 kb BamHI fragment; the resulting plasmid was transformed in E. coli strain SW106 with inducible Cre recombinase expression, deleting exon 2 and resulting in a 1.1 kb BamHI fragment. The NotI linearised fragment of the deletion construct was transfected by electroporation into 129/Sv embryonic stem cells. 244 Neomycin-resistant colonies were genotyped by PCR and Southern blotting. Two probes were used: one outside the 3' end of the deletion construct, with HindIII digestion site giving a single 9kb fragment for the WT and a 7kb fragment for the correctly-integrated deletion cassette, and one to the 5' end of the deletion cassette, again giving the same 9kb fragment for the Wt but a 3.5 kb fragment for the deletion cassette. 10 colonies showed a correct integration by homologous recombination. These ES cells were injected into blastocysts obtained from pregnant BALB/C mice, and chimeric mice were crossed with C57/Bl6J mice constitutively expressing FlpE recombinase, removing the FRTNeo cassette. Agouti mice were genotyped by PCR, showing correct insertion of the LoxP sites around exon 2. Cdk8 lox/lox mice were crossed with Villin-CreERT2 +/- mice to obtain Cdk8 lox/lox, VillinCreERT2 +/- .

To obtain small intestine organoids Cdk8 lox/lox and Cdk8 lox/lox/Villin-CreERT2 mice were used. Establishment, expansion and maintenance of organoids were performed as described previously (Sato et al, 2009). To induce the Cre-mediated recombination of Cdk8 in vitro, organoids were

cultured during 7 days in medium supplemented with 600 nM 4- Hydroxytamoxifen (Sigma H7904) resuspended in ethanol. Evaluation of knockout efficiency was performed using genotyping, qPCR and Western blotting. CRISPR/Cas9-mediated genome editing was employed to remove CDK19 from the organoids. CRISPR single guide RNA (sgRNA) targeting murine Cdk19 sequence (5'- AAAGTGGGACGCGGCACCTA-3', from Zhang lab database) was cloned as synthetic dsDNA into lentiCRISPRv2 vector as described ((Sanjana et al, 2014); provided by F. Zhang, Addgene plasmid #52961). Lentiviruses encoding the sgRNA targeting sequence were produced in HEK 293T cells transfected with LentiCRISPRv2 (+sgRNA Cdk19), pMD2.G and psPAX2. The viral supernatant (collected in organoids culture media) was passed through a 0.45-µm filter and used the same day for infection. Lentiviral-mediated transduction and antibiotic selection was performed as described previously (Onuma et al, 2013). Briefly, for lentiviral infection, organoids (5 days after seeding) were diluted into 10ml of PBS and dissociated into single cells by passing them 10-15 times through a needle with an insulin syringe. A volume containing 1–5 x 105 intestinal cells was centrifuged at 300 g for 5 minutes and resuspended with 1ml of the viral supernatant produced in HEK 293T cells. This mixture (virus + single stem cells) was layered on top of a Matrigel-covered well (12 well plate). 24 hours later, virus and dead cells containing media were removed and the Matrigel-attached cells were covered with 200 µl of Matrigel + 200 µl of culture medium to create a "sandwich" containing the infected cells inside. After polymerisation of the second Matrigel layer, 1 ml of organoid media per well was used to allow organoid formation inside the Matrigel. 24 hours later, Puromycin was added (5 µg/mL) and selection was conducted for 4 days. Once the organoids appeared (4-5 days after seeding the infected single cells), CDK19 knockout was verified by Western blot. We observed that CDK19 protein was still present, albeit decreased; therefore, we picked individual organoids and allowed them to grow in separated wells until we obtained several populations where CDK19 protein was completely absent, as seen by Western blot. DNA sequencing confirmed the deletion of a fragment of DNA around the sequence corresponding to Cdk19 sgRNA, and qPCR confirmed the absence of Cdk19 mRNA.

BMEL (Bipotential Mouse Embryonic Liver) cell line was isolated from CDK8F/F mouse. Cells were grown on collagen-coated plates in RPMI medium (Gibco/Life Technologies) supplemented with 10% fetal calf serum (Pan-Biotech), insulin 10 µg/mL (Sigma), IGFII 30 ng/mL (Peprotech), EGF 50 ng/mL (Peprotech), 100 units/ml penicillin, and 100 mg/ml streptomycin. Cells were grown under standard conditions at 37°C in a humidified incubator containing 5% CO2. All cells were routinely tested to confirm absence of mycoplasma contamination. Generation of cell lines pMSCV retroviral vectors (Clontech) encoding CRE recombinase, tamoxifen-inducible CREERT2

(addgene plasmid #22776) or human H-RasG12V were used to generate stable BMEL cell lines. For induction of CRE-ERT2 activity, 4-OH tamoxifen was added to the culture medium at a concentration of 2 µM for two weeks.

DEN induced carcinogenesis: Diethylnitrosamine (DEN) (30 mg/kg) was injected intraperitoneally in 14 days old male mice. Mice were sacrificed and livers collected after a period of 8 months.

Cell culture:

MEFs were grown under hypoxic conditions at 1% O2 and 5% CO2, 3T3 MEFs and NIH-3T3s at 21% O2 and 5% CO2. All cultures were performed in a humid incubator at 37°C. Cells were frozen in a solution of 90% fetal calf serum + 10% DMSO (dimethyl sulfoxide).

MEFs are derived from Cdk8Lox/Lox Polr2atm1(cre/ERT2)Bbd and Cdk8+/+ Polr2atm1(cre/ERT2)Bbd embryos produced by the crossing of Cdk8Lox/+ mice and Polr2atm1(cre/ERT2)Bbd mice. The preparation of MEFs was done according to a standard protocol (Xu, 2005). At E12.5, the mothers were sacrificed, the embryos were collected in PBS and dissected to remove internal organs, head, and limbs. The remaining carcasses were dissociated mechanically with a scalpel blade and chemically in a trypsin solution, and then cultured separately (P0), each embryo in a 10cm diameter dish. The primary cells that had adhered and proliferated after several days of culture were amplified and frozen. MEFs were immortalized by serial passages following the protocol 3T3 protocol, which consists of transferring 3,105 cells every three days to a new dish (Todaro and Green, 1963; Xu, 2005).

RNA extraction.

Total RNA was extracted using Trizol (Life Technologies) following manufacturer's instructions from wild type and two clones of Ki-67 knockout NIH/3T3 or 4T1 cell lines, as well as tumors. For MDA-MB-231 cells, MEFs and mouse intestinal organoids, total RNA was extracted using Trizol until the aqueous phase extraction step, that was followed by gDNA eliminator column step from the RNAeasy kit from Qiagen following the manufacturer’s instructions.

Tumors were removed from Trizol, ground in liquid nitrogen using a mortar and pestle then put back in Trizol to complete cell lysis. Chloroform was added, samples were centrifuged, the upper aqueous layer was isolated and ethanol was added.

For BMEL cells, RNA was extracted from exponentially growing subconfluent cultures.

In all cases RNA was then bound to a RNeasy column. Contaminating genomic DNA was degraded DNaseI treatment. RNA integrity was analyzed on Agilent 2100 bioanalyzer.

RNA sequencing.

For library preparation from RNA of NIH/3T3 cells, 4T1 cells and tumors, cDNA synthesis was performed on rRNA-depleted samples using the TruSeq Stranded Total RNA Library Preparation (RS-122-2301). All sequencing libraries were prepared in two or three biological replicates. Indexed cDNA libraries were sequenced by MGX (Montpellier) on Illumina HiSeq2000 with a single 50 bp read and a 10 bp index read. For library preparation from MDA-MB-231 cell RNA, cDNA synthesis was performed with oligo-dT enrichment and random N6 primers using the MGISP-100RS pipeline (BGI-Shenzhen, China). A-Tailing Mix and RNA Index Adapters were added by end repair. cDNA fragments obtained from the previous step were amplified by PCR, and products were purified by Ampure XP Beads and validated on the Agilent Technologies 2100 bioanalyzer for quality control. The double stranded PCR products were heat denatured and circularized by the splint oligo sequence to get the final library. The single strand circle DNA (ssCir DNA) was formatted as the final library. The final library was amplified with phi29 to make DNA nanoball (DNB) which had more than 300 copies of each molecule, DNBs were loaded into the patterned nanoarray and paired-end 100 base reads were generated on DNBSEQ-G400 platform (BGI-Shenzhen, China).

For intestinal organoids and MEFs all conditions were prepared as biological triplicates and sent to BGI Tech Solutions (Hong Kong)- Co for library preparation and RNA sequencing. Purification of mRNA from total RNA was achieved using oligo(dT)-attached magnetic beads and then fragmented for random hexamer-primed reverse transcription, followed by a second-strand cDNA synthesis. Sequencing was performed with the BGISEQ, SE50 platform to obtain an average of 50 million single-end, 50bp reads per sample.

For BMEL cells, the preparation of the library was done with the TruSeq Stranded mRNA Sample Preparation kit (Illumina). The sequencing was performed in an Illumina Hiseq 2500 sequencer by the Sequencing Platform of Montpellier, with 50 base pairs (bp) single end reads to an estimated depth of 25 million reads per sample.

In all cases the raw reads obtained in fastq format were subject to quality control using the FastQC software. The reads passing the quality control were aligned to the reference genome (mouse: GRCm38.p6; human: GRCh38.p13) and the counts per gene were quantified using the tool STAR 2.6.0a (Dobin et al., 2012). The mouse (Ensembl v93) and human (Genecode v35) genome annotations were used for establishing the coordinates of each gene and their transcripts. Differential expression analysis was performed in R using the DESeq2 (2) library embedded in an *ad hoc* script. After normalization of the read counts per gene, a negative binomial generalized linear model was fitted considering single factor design for assessing the

differential gene expression. Wald tests were performed for assessing statistical significance on the differential expression of each gene, then tests were independently filtered and corrected by multiple hypothesis testing (Benjamini–Hochberg).

Gene set enrichment analysis.

Over-representation of gene set analysis was performed using the enrichR (Kuleshov et al., 2016) and clusterProfiler (Yu et al., 2012) R libraries, against gene sets present in the Enrichr collection databases (Kuleshov et al., 2016). The p value was computed with the Fisher Exact test and then adjusted for multiple hypothesis testing in order to obtain the FDR (false discovery rate) adjusted p-value. GSEA gene set enrichment analysis (Subramanian et al., 2005) was performed using the javaGSEA desktop application with a log-fold-change pre-ranked gene list.

Chromatin Immunoprecipitation and sequencing (ChIP-Seq).

For chromatin immunoprecipitation (ChIP) of histones, 4T1 WT cells and Ki-67 KO 4T1 cells were cultured as described above. Before fixation, culturing media was removed and plates were washed with cold PBS. Cells were cross-linked in 1% formaldehyde (28908, Thermo Fisher Scientific) for 10 min at RT with intermittent agitation every 2 min. After fixation, cells were washed 3 times with ice-cold PBS and scraped in PBS with 1× protease inhibitor cocktail mix (05056489001, Sigma-Aldrich). Cells were collected by centrifugation (8 min, 1500 rpm, 4°C) and resuspended (108 cells) in 50 ml of lysis buffer (50 mM HEPES [pH 7.5], 140 mM NaCl, 0.1% Na-deoxycholate, 1 mM EDTA, 0.1% SDS, 1% NP-40 [IPEGAL], 1X protease inhibitor cocktail mix) for 30 min on ice. Nuclei were collected by brief centrifugation (2,500 rpm, 10 min, 4°C) and resuspended (nuclei/10M cells) in 300 µl of sonication buffer (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA, 1% SDS, and 1X protease inhibitor cocktail mix) and sonicated using a Diagenode bioruptor to generate 200- to 1,000-bp fragments, as determined by agarose gel electrophoresis. 30 µl magnetic beads (Protein A, 10002D, Life technologies) slurry was pre-washed three times with 1ml of PBS (1% BSA) before antibody incubation. Beads were collected and resuspended in 300 µl of PBS (1% BSA). 5µl of H3K27me3 antibody (17-622; Merck-Millipore), 2 µl of H3K4me3 antibody (ab8580; Abcam) and 5 µl of H3K27ac antibody (ab4729; Abcam) were added for each ChIP and incubated for 4 hours at 4°C. Antibody coupled magnetic beads were washed three times with PBS/1% BSA, and 300 µl of sonicated cell lysates were diluted 6-fold in binding buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl [pH 8.0]). The diluted chromatin was incubated overnight with antibody coupled magnetic beads at

4°C. The immunoprecipitates were washed 5 times with RIPA buffer (50 mM HEPES [pH 7.6], 1 mM EDTA, 0.7% Na-deoxycholate, 1% NP-40, 0.5M LiCl) and twice with TE buffer (10 mM Tris-HCl [pH 8.0], 0.1 mM EDTA). The DNA was recovered by reversing the cross-links in 1% SDS, 0.1 M NaHCO3 for 6 h at 65°C, purified using ChIP DNA clean and concentrator kit (Ozyme/Cell Signaling Technology) and eluted in 30 µl of elution buffer provided with the kit. Eluted DNA was quantified using Qubit ds DNA high sensitivity kit (Life Technologies) and treated with RNase A at 37°C for 1hour. Two columns were combined to get the desired amount of DNA for library preparation (30 ng) for respective histone marks and experiments were done in duplicates for each histone mark. For the spike-in, 2 µl of spike-in-antibody (61686, Active motif) was added with the respective anti-histone antibodies during antibody incubation with the beads and 7 µl of spike-in chromatin (53083, Active motif) was added with the sonicated chromatin (300 µl) for each CHIP experiment. The libraries were prepared by MGX Sequencing platform (Montpellier, France) using TruSeq CHIP library preparation kit (Illumina) following guidelines and sequenced with HiSeq-2500 apparatus.

For the raw reads obtained, the base calling was performed with Illumina RTA software, and FastQC was used to perform quality control of the sequencing. The reads that passed the quality control were aligned with Bowtie2 (v2.3.5.1) (Langmead and Salzberg, 2012) to the GRCm38.p6 version of the mouse genome. The total count of reads mapped to the fly genome (BDGP6.28) was used for the spike-in normalization. Read pairs were only considered if both were mapped correctly to the genome (SAM flag=2). Mapped reads were subject to spike-normalization, input (IgG) subtraction and binning using the deepTools API (Ramírez et al., 2016). The same tool was used for generating the heatmaps.

qRT-PCR.

For reverse transcription (cDNA synthesis), 500 ng of purified RNA in total volume of 10µl, extracted by RNeasy Mini Kit (Qiagen), were mixed with 1µl of 10mM dNTPs mix (LifeTechnologies) and 1µl of 50µM random hexaprimers (NEB). Samples were incubated at 65°C for 5 minutes, transferred to ice, 5µl 5x First Strand Buffer, 2µl100mM DTT and 1µl RNasin RNase Inhibitor (Promega) were added and samples were incubated at 25°C for 10 minutes, 42°C for 2 minutes. 1µl of SuperScript® III Reverse Transcriptase (LifeTechnologies) was added to each sample and incubated at 50°C for 60 minutes, 70°C for 15 minutes. qPCR was performed using LightCycler® 480 SYBR Green I Master (Roche) and LightCycler® 480 qPCR machine, with the PCR primers shown in Table S2. The reaction contained 5ng of cDNA, 2µl of 1µM qPCR primer pair, 5µl 2x Master Mix, and final volume made up to 10µl with DNase free water. Primers used

for both mouse (Mki67) and human (MKI67) Ki-67 in addition to the housekeeping genes are similar to those used by Sobecki et al., 2016. qPCR was conducted at 95°C for 10 min, and then 40 cycles of 95°C 20s, 58°C 20s and 72°C 20s. The specificity of the reaction was verified by melting curve analysis.

Actb	Fw- GGCCCAGAGCAAGAGAGGTATCC
	Rv- ACGCACGATTTCCCTCTCAGC
Gapdh	Fw- CGTCCCGTAGACAAAATGGT
	Rv- TGACTGTGCCGTTGAATTTG
B2m	Fw- ACGTAACACAGTTCCACCCG
	Rv- CAGTCTCAGTGGGGGTGAAT
Gusb	Fw- AACAACACACTGACCCCTCA
	Rv- ACCACAGATCGATGCAGTCC
Zeb2	Fw- GGCAAGGCCTTCAAGTACAA
	Rv- AAGCGTTTCTTGCAGTTTGG
Twist1	Fw- AGCGGGTCATGGCTAACG
	Rv- GGACCTGGTACAGGAAGTCGA
Tap2	Fw- CTGGCGGACATGGCTTTACTT
	Rv- CTCCCACTTTTAGCAGTCCCC
Psmb8	Fw- ATGGCGTTACTGGATCTGTGC
	Rv- CGCGGAGAAACTGTAGTGTCC
Zeb1	Fw- CGCCATGAGAAGAACGAGGAC
	Rv- CTGTGAATCCGTAAGTGCTCTTT
Vim	Fw- CTGCTTCAAGACTCGGTGGAC
	Rv- ATCTCCTCCTCGTACAGGTCG
Epcam	Fw- GGAGTCCCTGTTCCATTCTTCT
	Rv- GCGATGACTGCTAATGACACCA
Hes1	Fw- GATAGCTCCCGGCATTCCAAG
	Rv- GCGCGGTATTTCCCCAACA
Lef1	Fw- GCCACCGATGAGATGATCCC
	Rv- TTGATGTCGGCTAAGTCGCC
Aldh1l2	Fw- TTTCCTGAGGGGATCAAGGC
	Rv- GACCTCGAATCCAGTTATGCAA
Aldh3a1	Fw-AATATCAGTAGCATCGTGAACCG

Rv- GGAGAGCCCCTTAATCGTGAAA

Antibodies.

Western blot and immunofluorescence: Ki-67 (clone SP6; Abcam), cyclin A2 (6E6; Novocastra), PCNA (ab18197; Abcam), β-catenin (BD610154; BD-Bioscience), actin (A2066; Sigma), vimentin (D21H3, #5741; CST), CDK8 (sc-1521 C19 Polyclonal; Santa Cruz), CDK19 (HPA007053; polyclonal Sigma), CFTR (ab2784; Abcam)

Dilutions were 1:1000 for western blot and 1:500 for immunofluorescence.

ChIP-seq: H3K27me3 (17-622; Merck-Millipore), H3K4me3 (ab8580; Abcam), H3K27ac (ab4729; Abcam).

DNA replication assay, EdU labeling, drug treatments.

Cells were treated with 10 µM 5-ethynyl-2'-deoxyuridine (EdU; LifeTechnologies) for the indicated time, harvested, washed once with cold PBS, resuspended in 300µL cold PBS and fixed with 700µL ice-cold 100% ethanol. Click reaction was performed according to the manufacturer instructions (Click-iT™ Plus EdU Alexa Fluor™ 647 Flow Cytometry Assay Kit; Invitrogen) and cells were analyzed by flow-cytometry (BD FACSCanto II). FlowJo® software was used for analysis.

Cell extracts and Western-blotting.

Frozen cell pellets (harvested by trypsinization, washed with cold PBS) were lysed directly in Laemmli buffer at 95°C. Protein concentrations were determined by BCA protein assay (Pierce Biotechnology). Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (7.5%, 12.5% and 15% gels) and transferred to Immobilon membranes (Milipore) at 1.15 mA/cm2 for 120 min with a semi-dry blotting apparatus. Membranes were blocked in TBS-T pH 7.6 (20 mM Tris, 140mM NaCl, 0.1% Tween-20) containing non-fat dry milk (5%), incubated with the primary antibody for 2 hours at RT or overnight at 4°C, washed several times with TBS-T for a total of 45 minutes, incubated with secondary antibody at 1/5000 dilution for 1 hour at RT and washed several times in TBS-T. Signals were detected using Western Lightning Plus-ECL (PerkinElmer) and Amersham HyperfilmTM (GE Healthcare).

For intestinal organoids, Matrigel was disrupted by pipetting up and down several times the media in each well over the dome of Matrigel. This mix was spun down at 200g for 5 minutes at 4°C and the pellet was washed twice with 1ml of PBS. Pellets were snap frozen in liquid nitrogen and kept at -80°C until use. For organoids extracts, frozen pellets were lysed by incubation at 4°C for 20

minutes in lysis buffer with protease and phosphatase inhibitors (50mM Tris, pH 7.4, 100mM NaCl, 50mM NaF, 40mM beta-glycerophosphate, 2.5mM Na-Vanadate, 5mM EDTA, 1mM EGTA, 1mM DTT, 1% Triton X-100 and Protease inhibitor cocktail (Sigma P8340) diluted 1/400. The solubilised proteins were recovered from the supernatant after centrifugation at 16000g for 20 minutes at 4°C, snap frozen in liquid nitrogen and stored at -80°C until use. Protein concentrations were determined by BCA protein assay (Pierce Biotechnology). 30µg of of total proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (7% 10% and 12.5% gels) and transferred to Immobilon membranes (Milipore) at 90 volts for 120 min with a wet blotting apparatus. Membranes were blocked in TBS-T pH 7.6 (20mM Tris, 140mM NaCl, 0.1% Tween-20) containing non-fat dry milk (5%), incubated with the primary antibody in TBS-T + 3% BSA for 2 hours at RT or overnight at 4°C, washed 3 times with TBS-T for a total of 15 minutes, incubated with secondary antibody at 1/10000 dilution in TBS-T + 5% nonfat dried milk for 30 minutes at RT, and washed 3 times in TBS-T for a total of 15 minutes. Signals were detected using Western Lightning Plus-ECL (PerkinElmer) and Amersham HyperfilmTM ECL (GE Healthcare).

Xenografts.

Animals were housed in the animal facility of IGMM and were maintained in a pathogen-free environment and fed ad libitum. To generate primary tumors, 106 cells (4T1) or 3x106 cells (MDA-MB-231) of log-phase viable ‘mouse pathogen-free’ (Test: IMPACT1, Iddex) were implanted into the fourth inguinal mammary gland (in 50 µl PBS (4T1) or 200 µl PBS (MDA-MB-231)). Primary tumor volume was measured every week by electronic caliper using the formula “p/6*S2 (Smaller radius)*L (Larger radius)”. For tail-vein injections, 106 4T1 cells were injected into the tail vein of BALB/c or nude mice. At the end of the experiments, following sacrifice, primary tumors were excised and fixed overnight in neutral buffered formalin (10%) before paraffin embedding (see above).

Forskolin-induced swelling.

To remove exon 2 of Cdk8 from Cdk8lox/lox/Villin-Cre-ERT2+/- /Cdk19-/- , organoids were treated with 600nM 4-hydrxoytamoxifen for 7 days. Once the Cdk8/Cdk19 double KO was obtained, forskolin-induced swelling was measured as indicated (Dekkers et al, 2013). Organoids were transferred to CELLview culture dishes PS 35/10 mm, glass bottom, 4 compartments (Greiner Bio-One, 627870), two days before imaging. Confocal spinning disk (Dragonfly, Andor, Oxford Instruments) microscope equipped with heated chamber allowing constant temperature (37°C) and CO2 flow (5% CO2), EMCCD iXon888 camera (Lifer Andor, pixel = 13 µm), objective 10x/0.45

DRY, correction Plan Apo Lambda, 4mm working distance, was used for imaging, with Fusion acquisition software. Images of a single organoid, previously selected, were taken every 2 minutes during 20 minutes after forskolin addition (5µM, or DMSO vehicle control) to the media. For data analysis, a macro was created using Fiji software. It consisted of recognising and filling the structures imaged through the alexa-488 track, to calculate the increase of total organoid area in single organoids over the different time points.

Soft agar.

10^5 cells for each cell line were mixed with medium supplemented with 0.5% agarose and placed on top of the 1% agarose layer. 1 mL medium was added to the solidified layer and changed every 2-3 days. After 4 to 6 weeks, soft agar was stained with crystal violet 0,005% in 4% PFA for 1h. Colonies visible to the naked eye were counted manually.

7. Bibliography

1. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001
2. Adler, A. S., McCleland, M. L., Truong, T., Lau, S., Modrusan, Z., Soukup, T. M., Roose-Girma, M., Blackwood, E. M., & Firestein, R. (2012). CDK8 maintains tumor dedifferentiation and embryonic stem cell pluripotency. *Cancer Research*, *72*(8), 2129–2139. https://doi.org/10.1158/0008-5472.CAN-11-3886
3. Akkari, L., Grégoire, D., Floc'H, N., Moreau, M., Hernandez, C., Simonin, Y., Rosenberg, A. R., Lassus, P., & Hibner, U. (2012). Hepatitis C viral protein NS5A induces EMT and participates in oncogenic transformation of primary hepatocyte precursors. *Journal of Hepatology*, *57*(5), 1021–1028. https://doi.org/10.1016/j.jhep.2012.06.027
4. Aladjem, M. I., & Redon, C. E. (2017). Order from clutter: Selective interactions at mammalian replication origins. *Nature Reviews Genetics*, *18*(2), 101–116. https://doi.org/10.1038/nrg.2016.141
5. Alarcón, C., Zaromytidou, A. I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A. N., Manova-Todorova, K., Macias, M. J., Sapkota, G., Pan, D., & Massagué, J. (2009). Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. *Cell*, *139*(4), 757–769. https://doi.org/10.1016/J.CELL.2009.09.035
6. Aleem, E., Kiyokawa, H., & Kaldis, P. (2005). Cdc2-cyclin E complexes regulate the G1/S phase transition. *Nature Cell Biology*, *7*(8), 831–836. https://doi.org/10.1038/ncb1284
7. Al-Husini, N., Tomares, D. T., Bitar, O., Childers, W. S., & Schrader, J. M. (2018). α-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies. *Molecular Cell*, *71*(6), 1027-1039.e14. https://doi.org/10.1016/J.MOLCEL.2018.08.003/ATTACHMENT/B7D8F3C4-3820-4743-8393-F1AACDC4B5BC/MMC1.PDF
8. Al-Husini, N., Tomares, D. T., Pfaffenberger, Z. J., Muthunayake, N. S., Samad, M. A., Zuo, T., Bitar, O., Aretakis, J. R., Bharmal, M. H. M., Gega, A., Biteen, J. S., Childers, W. S., & Schrader, J. M. (2020). BR-Bodies Provide Selectively Permeable Condensates that Stimulate mRNA Decay and Prevent Release of Decay Intermediates. *Molecular Cell*, *78*(4), 670-682.e8. https://doi.org/10.1016/J.MOLCEL.2020.04.001
9. Alver, R. C., Chadha, G. S., Gillespie, P. J., & Blow, J. J. (2017). Reversal of DDK-Mediated MCM Phosphorylation by Rif1-PP1 Regulates Replication Initiation and Replisome Stability Independently of ATR/Chk1. *Cell Reports*, *18*(10), 2508–2520. https://doi.org/10.1016/j.celrep.2017.02.042
10. Ameri, M., Nezafat, N., & Eskandari, S. (2022). The potential of intrinsically disordered regions in vaccine development. In *Expert Review of Vaccines* (Vol. 21, Issue 1, pp. 1–3). Taylor & Francis. https://doi.org/10.1080/14760584.2022.1997600
11. Amin, P., Awal, S., Vigneron, S., Roque, S., Mechali, F., Labbé, J. C., Lorca, T., & Castro, A. (2021). PP2A-B55: substrates and regulators in the control of cellular functions. *Oncogene 2021 41:1*, *41*(1), 1–14. https://doi.org/10.1038/s41388-021-02068-x
12. Amon, A., Tyers, M., Futcher, B., & Nasmyth, K. (1993). Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell*, *74*(6), 993–1007. https://doi.org/10.1016/0092-8674(93)90722-3
13. Anderson, L., Henderson, C., & Adachi, Y. (2001). Phosphorylation and Rapid Relocalization of 53BP1 to Nuclear Foci upon DNA Damage. *Molecular and Cellular Biology*, *21*(5), 1719–1729.

https://doi.org/10.1128/MCB.21.5.1719-1729.2001/ASSET/F10BF90B-4BC1-48A3-BAF9-1C06113EFCAC/ASSETS/GRAPHIC/MB0511463008.JPEG

14. Andersson-Rolf, A., Mustata, R. C., Merenda, A., Kim, J., Perera, S., Grego, T., Andrews, K., Tremble, K., Silva, J. C. R., Fink, J., Skarnes, W. C., & Koo, B. K. (2017). One-step generation of conditional and reversible gene knockouts. *Nature Methods*, *14*(3), 287–289. https://doi.org/10.1038/nmeth.4156
15. Andrade, L. E. C., Tan, E. M., & Chan, E. K. L. (1993). Immunocytochemical analysis of the coiled body in the cell cycle and during cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(5), 1947–1951. https://doi.org/10.1073/pnas.90.5.1947
16. Anfinsen, C. B., & Haber, E. (1961). Studies on the reduction and re-formation of protein disulfide bonds. *The Journal of Biological Chemistry*, *236*(5), 1361–1363. https://doi.org/10.1016/s0021-9258(18)64177-8
17. Aparicio, T., Ibarra, A., & Méndez, J. (2006). Cdc45-MCM-GINS, a new power player for DNA replication. *Cell Division*, *1*(1), 1–5. https://doi.org/10.1186/1747-1028-1-18/FIGURES/1
18. Aravamudhan, P., Goldfarb, A. A., & Joglekar, A. P. (2015). The kinetochore encodes a mechanical switch to disrupt spindle assembly checkpoint signalling. *Nature Cell Biology 2014 17:7*, *17*(7), 868–879. https://doi.org/10.1038/ncb3179
19. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. In *Nature Genetics* (Vol. 25, Issue 1, pp. 25–29). Nature Publishing Group. https://doi.org/10.1038/75556
20. Asquith, T. N., Uhlig, J., Mehansho, H., Putman, L., Carlson, D. M., & Butler, L. (1987). Binding of Condensed Tannins to Salivary Proline-Rich Glycoproteins: The Role of Carbohydrate. *Journal of Agricultural and Food Chemistry*, *35*(3), 331–334. https://doi.org/10.1021/jf00075a012
21. Axton, J. M., Dombrádi, V., Cohen, P. T. W., & Glover, D. M. (1990). One of the protein phosphatase 1 isoenzymes in Drosophila is essential for mitosis. *Cell*, *63*(1), 33–46. https://doi.org/10.1016/0092-8674(90)90286-N
22. Azaldegui, C. A., Vecchiarelli, A. G., & Biteen, J. S. (2021). The emergence of phase separation as an organizing principle in bacteria. In *Biophysical Journal* (Vol. 120, Issue 7, pp. 1123–1138). https://doi.org/10.1016/j.bpj.2020.09.023
23. Bah, A., & Forman-Kay, J. D. (2016). Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry*, *291*(13), 6696–6705. https://doi.org/10.1074/jbc.R115.695056
24. Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, *18*(5), 285–298. https://doi.org/10.1038/nrm.2017.7
25. Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R., & Rosen, M. K. (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell*, *166*(3), 651–663. https://doi.org/10.1016/j.cell.2016.06.010
26. Bancerek, J., Poss, Z. C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dölken, L., Strobl, B., Müller, M., Taatjes, D. J., & Kovarik, P. (2013). CDK8 Kinase Phosphorylates Transcription Factor STAT1 to Selectively Regulate the Interferon Response. *Immunity*, *38*(2), 250–262. https://doi.org/10.1016/j.immuni.2012.10.017
27. Barr, A. R., Heldt, F. S., Zhang, T., Bakal, C., & Novák, B. (2016). A Dynamical Framework for the All-or-None G1/S Transition. *Cell Systems*, *2*(1), 27–37. https://doi.org/10.1016/J.CELS.2016.01.001

28. Bassermann, F., Von Klitzing, C., Illert, A. L., Münch, S., Morris, S. W., Pagano, M., Peschel, C., & Duyster, J. (2007). Multisite phosphorylation of nuclear interaction partner of ALK (NIPA) at G2/M involves cyclin B1/Cdk1. *Journal of Biological Chemistry*, *282*(22), 15965–15972. https://doi.org/10.1074/JBC.M610819200/ATTACHMENT/19E339F5-A737-4CB8-9C23-F4F9013BF422/MMC1.PDF

29. Bassermann, F., Von Klitzing, C., Münch, S., Bai, R. Y., Kawaguchi, H., Morris, S. W., Peschel, C., & Duyster, J. (2005). NIPA defines an SCF-type mammalian E3 ligase that regulates mitotic entry. *Cell*, *122*(1), 45–57. https://doi.org/10.1016/j.cell.2005.04.034

30. Basu, S., Greenwood, J., Jones, A. W., & Nurse, P. (2022). Core control principles of the eukaryotic cell cycle. *Nature 2022*, 1–6. https://doi.org/10.1038/s41586-022-04798-8

31. Bazile, F., St-Pierre, J., & D'Amours, D. (2010). Three-step model for condensin activation during mitotic chromosome condensation. *Cell Cycle*, *9*(16), 3263–3275. https://doi.org/10.4161/cc.9.16.12620

32. Bell, S. P., & Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, *357*(6374), 128–134. https://doi.org/10.1038/357128a0

33. Berchtold, D., Battich, N., & Pelkmans, L. (2018). A Systems-Level Study Reveals Regulators of Membrane-less Organelles in Human Cells. *Molecular Cell*, *72*(6), 1035-1049.e5. https://doi.org/10.1016/J.MOLCEL.2018.10.036/ATTACHMENT/55E95006-EC9C-4506-89AD-F6E82716F8A0/MMC3.XLSX

34. Berdougo, E., Nachury, M. V., Jackson, P. K., & Jallepalli, P. V. (2008). The nucleolar phosphatase Cdc14B is dispensable for chromosome segregation and mitotic exit in human cells. *Http://Dx.Doi.Org.Insb.Bib.Cnrs.Fr/10.4161/Cc.7.9.5792*, *7*(9), 1184–1190. https://doi.org/10.4161/CC.7.9.5792

35. Bialojan, C., & Takai, A. (1988). Inhibitory effect of a marine-sponge toxin, okadaic acid, on protein phosphatases. Specificity and kinetics. *Biochemical Journal*, *256*(1), 283. https://doi.org/10.1042/BJ2560283

36. Bishop, A. C., Ubersax, J. A., Pøtsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., Shimizu, E., Tsien, J. Z., Schultz, P. G., Rose, M. D., Wood, J. L., Morgan, D. O., & Shokat, K. M. (2000). A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*, *407*(6802), 395–401. https://doi.org/10.1038/35030148

37. Blanco, M. A., Sánchez-Díaz, A., De Prada, J. M., & Moreno, S. (2000). APC(ste9/srw1) promotes degradation of mitotic cyclins in G1 and is inhibited by cdc2 phosphorylation. *EMBO Journal*, *19*(15), 3945–3955. https://doi.org/10.1093/emboj/19.15.3945

38. Blethrow, J. D., Glavy, J. S., Morgan, D. O., & Shokat, K. M. (2008). Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1442–1447. https://doi.org/10.1073/pnas.0708966105

39. Bloom, J., Cristea, I. M., Procko, A. L., Lubkov, V., Chait, B. T., Snyder, M., & Cross, F. R. (2011). Global analysis of Cdc14 phosphatase reveals diverse roles in mitotic processes. *Journal of Biological Chemistry*, *286*(7), 5434–5445. https://doi.org/10.1074/JBC.M110.205054/ATTACHMENT/2C08C7E9-32E4-474F-9F65-867973953346/MMC1.ZIP

40. Bloomfield, M., Chen, J., & Cimini, D. (2021). Spindle Architectural Features Must Be Considered Along With Cell Size to Explain the Timing of Mitotic Checkpoint Silencing. *Frontiers in Physiology*, *11*, 1842. https://doi.org/10.3389/FPHYS.2020.596263/BIBTEX

41. Bøe, S. O., Haave, M., Jul-Larsen, Å., Grudic, A., Bjerkvig, R., & Lønning, P. E. (2006). Promyelocytic leukemia nuclear bodies are predetermined processing sites for damaged DNA. *Journal of Cell Science*, *119*(16), 3284–3295. https://doi.org/10.1242/jcs.03068

42. Booher, R., & Beach, D. (1988). Involvement of cdc13+ in mitotic control in Schizosaccharomyces pombe: possible interaction of the gene product with microtubules. *The EMBO Journal*, *7*(8), 2321–2327. https://doi.org/10.1002/j.1460-2075.1988.tb03075.x

43. Boos, D., Sanchez-Pulido, L., Rappas, M., Pearl, L. H., Oliver, A. W., Ponting, C. P., & Diffley, J. F. X. (2011). Regulation of DNA Replication through Sld3-Dpb11 Interaction Is Conserved from Yeast to Humans. *Current Biology*, *21*(13), 1152–1157. https://doi.org/10.1016/J.CUB.2011.05.057

44. Booth, D. G., & Earnshaw, W. C. (2017). Ki-67 and the Chromosome Periphery Compartment in Mitosis. *Trends in Cell Biology*, *27*(12), 906–916. https://doi.org/10.1016/j.tcb.2017.08.001

45. Booth, D. G., Takagi, M., Sanchez-Pulido, L., Petfalski, E., Vargiu, G., Samejima, K., Imamoto, N., Ponting, C. P., Tollervey, D., Earnshaw, W. C., & Vagnarelli, P. (2014). Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *ELife*, *2014*(3). https://doi.org/10.7554/eLife.01641.001

46. Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., Sottini, A., Soranno, A., Buholzer, K. J., Nettels, D., Kragelund, B. B., Best, R. B., & Schuler, B. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature*, *555*(7694), 61–66. https://doi.org/10.1038/nature25762

47. Bothmer, A., Robbiani, D. F., Di Virgilio, M., Bunting, S. F., Klein, I. A., Feldhahn, N., Barlow, J., Chen, H. T., Bosque, D., Callen, E., Nussenzweig, A., & Nussenzweig, M. C. (2011). Regulation of DNA End Joining, Resection, and Immunoglobulin Class Switch Recombination by 53BP1. *Molecular Cell*, *42*(3), 319–329. https://doi.org/10.1016/j.molcel.2011.03.019

48. Bothmer, A., Robbiani, D. F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., & Nussenzweig, M. C. (2010). 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *Journal of Experimental Medicine*, *207*(4), 855–865. https://doi.org/10.1084/jem.20100244

49. Boze, H., Marlin, T., Durand, D., Pérez, J., Vemhet, A., Canon, F., Sami-Manchado, P., Cheynier, V., & Cabane, B. (2010). Proline-rich salivary proteins have extended conformations. *Biophysical Journal*, *99*(2), 656–665. https://doi.org/10.1016/j.bpj.2010.04.050

50. Bradbury, E. M., Inglis, R. J., & Matthews, H. R. (1974). Control of cell division by very lysine rich histone (F1) phosphorylation. *Nature*, *247*(5439), 257–261. https://doi.org/10.1038/247257a0

51. Bradbury, E. M., Inglis, R. J., Matthews, H. R., & Langan, T. A. (1974). Molecular basis of control of mitotic cell division in eukaryotes. *Nature*, *249*(5457), 553–556. https://doi.org/10.1038/249553a0

52. Brandeis, M., Rosewell, I., Carrington, M., Crompton, T., Jacobs, M. A., Kirk, J., Gannon, J., & Hunt, T. (1998). Cyclin B2-null mice develop normally and are fertile whereas cyclin B1-null mice die in utero. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(8), 4344–4349. https://doi.org/10.1073/PNAS.95.8.4344

53. Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., & Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, *324*(5935), 1729–1732. https://doi.org/10.1126/science.1172046

54. Brangwynne, C. P., Mitchison, T. J., & Hyman, A. A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4334–4339. https://doi.org/10.1073/pnas.1017150108

55. Brown, C. J., Johnson, A. K., & Daughdrill, G. W. (2010). Comparing Models of Evolution for Ordered and Disordered Proteins. *Molecular Biology and Evolution*, *27*(3), 609–621. https://doi.org/10.1093/molbev/msp277

56. Brown, C. J., Johnson, A. K., Dunker, A. K., & Daughdrill, G. W. (2011). Evolution and disorder. *Current Opinion in Structural Biology*, *21*(3), 441–446. https://doi.org/10.1016/j.sbi.2011.02.005

57. Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J., & Keith Dunker, A. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of Molecular Evolution*, *55*(1), 104–110. https://doi.org/10.1007/s00239-001-2309-6

58. Brown, N. R., Korolchuk, S., Martin, M. P., Stanley, W. A., Moukhametzianov, R., Noble, M. E. M., & Endicott, J. A. (2015). CDK1 structures reveal conserved and unique features of the essential cell cycle CDK. *Nature Communications*, *6*(1), 1–12. https://doi.org/10.1038/ncomms7769

59. Brown, N. R., Noble, M. E. M., Endicott, J. A., & Johnson, L. N. (1999). The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nature Cell Biology*, *1*(7), 438–443. https://doi.org/10.1038/15674

60. Buchmann, A., Bauer-Hofmann, R., Mahr, J., Drinkwater, N. R., Luz, A., & Schwarz, M. (1991). Mutational activation of the c-Ha-ras gene in liver tumors of different rodent strains: Correlation with susceptibility to hepatocarcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(3), 911–915. https://doi.org/10.1073/pnas.88.3.911

61. Bueno, A., & Russell, P. (1993). Two fission yeast B-type cyclins, cig2 and Cdc13, have different functions in mitosis. *Molecular and Cellular Biology*, *13*(4), 2286–2297. https://doi.org/10.1128/mcb.13.4.2286-2297.1993

62. Bueno, A., Richardson, H., Reed, S. I., & Russell, P. (1991). A fission yeast B-type cyclin functioning early in the cell cycle. *Cell*, *66*(1), 149–159. https://doi.org/10.1016/0092-8674(91)90147-Q

63. Burley, S. K. (2021). Impact of structural biologists and the Protein Data Bank on small-molecule drug discovery and development. In *Journal of Biological Chemistry* (Vol. 296, p. 100559). Elsevier. https://doi.org/10.1016/j.jbc.2021.100559

64. Burset, M., Seledtsov, I. A., & Solovyev, V. V. (2000). Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research*, *28*(21), 4364–4375. https://doi.org/10.1093/nar/28.21.4364

65. Burton, J. L., Xiong, Y., & Solomon, M. J. (2011). Mechanisms of pseudosubstrate inhibition of the anaphase promoting complex by Acm1. *EMBO Journal*, *30*(9), 1818–1829. https://doi.org/10.1038/emboj.2011.90

66. Cao, L., Chen, F., Yang, X., Xu, W., Xie, J., & Yu, L. (2014). Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evolutionary Biology*, *14*(1), 1–16. https://doi.org/10.1186/1471-2148-14-10

67. Card, G. L. (2000). Crystal structure of a gamma-herpesvirus cyclin-cdk complex. *The EMBO Journal*, *19*(12), 2877–2888. https://doi.org/10.1093/emboj/19.12.2877

68. Carette, J. E., Raaben, M., Wong, A. C., Herbert, A. S., Obernosterer, G., Mulherkar, N., Kuehne, A. I., Kranzusch, P. J., Griffin, A. M., Ruthel, G., Cin, P. D., Dye, J. M., Whelan, S. P., Chandran, K., & Brummelkamp, T. R. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature*, *477*(7364), 340–343. https://doi.org/10.1038/nature10348

69. Carmo-Fonseca, M., Ferreira, J., & Lamond, A. I. (1993). Assembly of snRNP-containing coiled bodies is regulated in interphase and mitosis - Evidence that the coiled body is a kinetic nuclear structure. *Journal of Cell Biology*, *120*(4), 841–852. https://doi.org/10.1083/jcb.120.4.841

70. Celetti, G., Paci, G., Caria, J., VanDelinder, V., Bachand, G., & Lemke, E. A. (2020). The liquid state of FG-nucleoporins mimics permeability barrier properties of nuclear pore complexes. *Journal of Cell Biology*, *219*(1). https://doi.org/10.1083/jcb.201907157

71. Cerqueira, A., Santamaría, D., Martínez-Pastor, B., Cuadrado, M., Fernández-Capetillo, O., & Barbacid, M. (2009). Overall Cdk activity modulates the DNA damage response in mammalian cells. *Journal of Cell Biology*, *187*(6), 773–780. https://doi.org/10.1083/jcb.200903033

72. Cescutti, R., Negrini, S., Kohzaki, M., & Halazonetis, T. D. (2010). TopBP1 functions with 53BP1 in the G1 DNA damage checkpoint. *EMBO Journal*, *29*(21), 3723–3732. https://doi.org/10.1038/EMBOJ.2010.238

73. Chapman, J. R., Sossick, A. J., Boulton, S. J., & Jackson, S. P. (2012). BRCA1-associated exclusion of 53BP1 from DNA: Damage sites underlies temporal control of DNA repair. *Journal of Cell Science*, *125*(15), 3529–3534. https://doi.org/10.1242/jcs.105353

74. Chase, A., & Cross, N. C. P. (2011). Aberrations of EZH2 in cancer. *Clinical Cancer Research*, *17*(9), 2613–2618. https://doi.org/10.1158/1078-0432.CCR-10-2156

75. Chebaro, Y., Ballard, A. J., Chakraborty, D., & Wales, D. J. (2015). Intrinsically disordered energy landscapes. *Scientific Reports*, *5*(1), 1–12. https://doi.org/10.1038/srep10386

76. Chen, F., Archambault, V., Kar, A., Lio', P., D'Avino, P. P., Sinka, R., Lilley, K., Laue, E. D., Deak, P., Capalbo, L., & Glover, D. M. (2007). Multiple Protein Phosphatases Are Required for Mitosis in Drosophila. *Current Biology*, *17*(4), 293–303. https://doi.org/10.1016/J.CUB.2007.01.068

77. Chen, J. W., Romero, P., Uversky, V. N., & Dunker, A. K. (2006). Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. *Journal of Proteome Research*, *5*(4), 888–898. https://doi.org/10.1021/pr060049p

78. Chen, J. W., Romero, P., Uversky, V. N., & Dunker, A. K. (2006). Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *Journal of Proteome Research*, *5*(4), 879–887. https://doi.org/10.1021/pr060048x

79. Chen, J., & Liu, J. (2014). Spatial-temporal model for silencing of the mitotic spindle assembly checkpoint. *Nature Communications 2014 5:1*, *5*(1), 1–13. https://doi.org/10.1038/ncomms5795

80. Chen, M., Liang, J., Ji, H., Yang, Z., Altilia, S., Hu, B., Schronce, A., McDermott, M. S. J., Schools, G. P., Lim, C. U., Oliver, D., Shtutman, M. S., Lu, T., Stark, G. R., Porter, D. C., Broude, E. V., & Roninson, I. B. (2017). CDK8/19 Mediator kinases potentiate induction of transcription by NFκB. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), 10208–10213. https://doi.org/10.1073/pnas.1710467114

81. Chen, S. C. C., Chen, F. C., & Li, W. H. (2010). Phosphorylated and nonphosphorylated serine and threonine residues evolve at different rates in mammals. *Molecular Biology and Evolution*, *27*(11), 2548–2554. https://doi.org/10.1093/molbev/msq142

82. Cheng, J., Sweredoski, M. J., & Baldi, P. (2005). Accurate prediction of protein disordered regions by mining protein structure data. *Data Mining and Knowledge Discovery*, *11*(3), 213–222. https://doi.org/10.1007/s10618-005-0001-y

83. Cheng, S., Cetinkaya, M., & Gräter, F. (2010). How sequence determines elasticity of disordered proteins. *Biophysical Journal*, *99*(12), 3863–3869. https://doi.org/10.1016/j.bpj.2010.10.011

84. Chi, Y., Carter, J. H., Swanger, J., Mazin, A. V., Moritz, R. L., & Clurman, B. E. (2020). A novel landscape of nuclear human CDK2 substrates revealed by in situ phosphorylation. *Science Advances*, *6*(16), 9899–9916. https://doi.org/10.1126/SCIADV.AAZ9899/SUPPL_FILE/AAZ9899_TABLE_S5.XLSX

85. Chi, Y., Welcker, M., Hizli, A. A., Posakony, J. J., Aebersold, R., & Clurman, B. E. (2008). Identification of CDK2 substrates in human cell lysates. *Genome Biology*, *9*(10), 1–12. https://doi.org/10.1186/gb-2008-9-10-r149

86. Chica, N., Rozalén, A. E., Pérez-Hidalgo, L., Rubio, A., Novak, B., & Moreno, S. (2016). Nutritional control of cell size by the greatwall-endosulfine-PP2A·B55 pathway. *Current Biology*, *26*(3), 319–330. https://doi.org/10.1016/j.cub.2015.12.035

87. Choi, J. M., Holehouse, A. S., & Pappu, R. V. (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. In *Annual Review of Biophysics* (Vol. 49, pp. 107–133). Annual Reviews. https://doi.org/10.1146/annurev-biophys-121219-081629

88. Chung, I., Osterwald, S., Deeg, K. I., & Rippe, K. (2012). PML body meets telomere. *Nucleus*, *3*(3), 263–275. https://doi.org/10.4161/nucl.20326

89. Cidado, J., Wong, H. Y., Marc Rosen, D., Cimino-Mathews, A., Garay, J. P., Fessler, A. G., Rasheed, Z. A., Hicks, J., Cochran, R. L., Croessmann, S., Zabransky, D. J., Mohseni, M., Beaver, J. A., Chu, D., Cravero, K., Christenson, E. S., Medford, A., Mattox, A., De Marzo, A. M., … Park, B. H. (2016). Ki-67 is required for maintenance of cancer stem cells but not cell proliferation. *Oncotarget*, *7*(5), 6281–6293. https://doi.org/10.18632/oncotarget.7057

90. Ciosk, R., Zachariae, W., Michaelis, C., Shevchenko, A., Mann, M., & Nasmyth, K. (1998). An ESP1/PDS1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast. *Cell*, *93*(6), 1067–1076. https://doi.org/10.1016/S0092-8674(00)81211-8

91. Clarke, P. G. H., & Clarke, S. (1996). Nineteenth century research on naturally occurring cell death and related phenomena. *Anatomy and Embryology*, *193*(2), 81–99. https://doi.org/10.1007/BF00214700

92. Clevers, H. (2016). Modeling Development and Disease with Organoids. *Cell*, *165*(7), 1586–1597. https://doi.org/10.1016/j.cell.2016.05.082

93. Clift, D., & Schuh, M. (2013). Restarting life: fertilization and the transition from meiosis to mitosis. *Nature Reviews Molecular Cell Biology 2013 14:9*, *14*(9), 549–562. https://doi.org/10.1038/nrm3643

94. Colnot, S., Niwa-Kawakita, M., Hamard, G., Godard, C., Le Plenier, S., Houbron, C., Romagnolo, B., Berrebi, D., Giovannini, M., & Perret, C. (2004). Colorectal cancers in a new mouse model of familial adenomatous polyposis: Influence of genetic and environmental modifiers. *Laboratory Investigation*, *84*(12), 1619–1630. https://doi.org/10.1038/labinvest.3700180

95. Connolly, T., & Beach, D. (1994). Interaction between the Cig1 and Cig2 B-type cyclins in the fission yeast cell cycle. *Molecular and Cellular Biology*, *14*(1), 768–776. https://doi.org/10.1128/mcb.14.1.768-776.1994

96. Cook, J. G., Park, C. H., Burke, T. W., Leone, G., DeGregori, J., Engel, A., & Nevins, J. R. (2002). Analysis of Cdc6 function in the assembly of mammalian prereplication complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1347–1352. https://doi.org/10.1073/pnas.032677499

97. Cooper, S. (2019). The synchronization manifesto: a critique of whole-culture synchronization. *FEBS Journal*, *286*(23), 4650–4656. https://doi.org/10.1111/FEBS.15050

98. Cooper, S. (2021). The Anti-G0 Manifesto: Should a problematic construct (G0) with no biological reality be removed from the cell cycle? Yes! *BioEssays*, *43*(3). https://doi.org/10.1002/bies.202000270

99. Cooper, S. (2003). Reappraisal of serum starvation, the restriction point, G0, and G1 phase arrest points. *The FASEB Journal*, *17*(3), 333–340. https://doi.org/10.1096/fj.02-0352rev

100. Corbett, K. D. (2017). Molecular Mechanisms of Spindle Assembly Checkpoint Activation and Silencing. *Progress in Molecular and Subcellular Biology*, *56*, 429–455. https://doi.org/10.1007/978-3-319-58592-5_18

101. Corno, A., Cordeiro, M. H., Allan, L. A., Wei, Q., Harrington, E., Smith, R. J., & Saurin, A. T. (n.d.). *A bifunctional kinase-phosphatase module integrates mitotic checkpoint and error-correction signalling to ensure mitotic fidelity*. https://doi.org/10.1101/2022.05.22.492960

102. Corpet, A., Kleijwegt, C., Roubille, S., Juillard, F., Jacquet, K., Texier, P., & Lomonte, P. (2020). Survey and summary PML nuclear bodies and chromatin dynamics: Catch me if you can! *Nucleic Acids Research*, *48*(21), 11890–11912. https://doi.org/10.1093/nar/gkaa828

103. Coudreuse, D., & Nurse, P. (2010). Driving the cell cycle with a minimal CDK control network. *Nature*, *468*(7327), 1074–1080. https://doi.org/10.1038/nature09543

104. Courvalin, J. C., Segil, N., Blobel, G., & Worman, H. J. (1992). The lamin B receptor of the inner nuclear membrane undergoes mitosis- specific phosphorylation and is a substrate for p34(cdc2)-type protein kinase. *Journal of Biological Chemistry*, *267*(27), 19035–19038. https://doi.org/10.1016/s0021-9258(18)41734-6

105. Coverley, D., Laman, H., & Laskey, R. A. (2002). Distinct roles for cyclins E and A during DNA replication complex assembly and activation. *Nature Cell Biology 2002 4:7*, *4*(7), 523–528. https://doi.org/10.1038/ncb813

106. Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4

107. Crick, F. H. C., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, *192*(4809), 1227–1232. https://doi.org/10.1038/1921227a0

108. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., & Matunis, M. L. J. (2002). Proteomic analysis of the mammalian nuclear pore complex. *Journal of Cell Biology*, *158*(5), 915–927. https://doi.org/10.1083/jcb.200206106

109. Cross, F. R. (1988). DAF1, a mutant gene affecting size control, pheromone arrest, and cell cycle kinetics of Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *8*(11), 4675–4684. https://doi.org/10.1128/mcb.8.11.4675-4684.1988

110. Cross, F. R., Buchler, N. E., & Skotheim, J. M. (2011). Evolution of networks and sequences in eukaryotic cell cycle control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1584), 3532–3544. https://doi.org/10.1098/rstb.2011.0078

111. Cueille, N., Salimova, E., Esteban, V., Blanco, M., Moreno, S., Bueno, A., & Simanis, V. (2001). Flp1, a fission yeast orthologue of the S. cerevisiae CDC14 gene, is not required for cyclin degradation or rum1p stabilisation at the end of mitosis. *Journal of Cell Science*, *114*(14), 2649–2664. https://doi.org/10.1242/JCS.114.14.2649

112. Curtis, M., Nikolopoulos, S. N., & Turner, C. E. (2002). Actopaxin is phosphorylated during mitosis and is a substrate for cyclin B1/cdc2 kinase. *Biochemical Journal*, *363*(2), 233–242. https://doi.org/10.1042/bj3630233

113. Cuylen, S., Blaukopf, C., Politi, A. Z., Muller-Reichert, T., Neumann, B., Poser, I., Ellenberg, J., Hyman, A. A., & Gerlich, D. W. (2016). Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. *Nature*, *535*(7611), 308–312. https://doi.org/10.1038/nature18610

114. Cyert, M. S., & Kirschner, M. W. (1988). Regulation of MPF activity in vitro. *Cell*, *53*(2), 185–195. https://doi.org/10.1016/0092-8674(88)90380-7

115. Cyert, M. S., & Thorner, J. (1989). Putting It On and Taking It Off: Minireview Phosphoprotein Phosphatase Involvement in Cell Cycle Regulation. *Cell*, *57*, 891–893.

116. Daily, K. M., Radivojac, P., & Dunker, A. K. (2005). Intrinsic disorder and prote in modifications: Building an SVM predictor for methylation. *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB '05*, *2005*. https://doi.org/10.1109/cibcb.2005.1594957

117. Daniels, A. J., Williams, R. J. P., & Wright, P. E. (1978). The character of the stored molecules in chromaffin granules of the adrenal medulla: A nuclear magnetic resonance study. *Neuroscience*, *3*(6), 573–585. https://doi.org/10.1016/0306-4522(78)90022-2

118. Daniels, D. L. (2013). Mutual Exclusivity of MED12/MED12L, MED13/13L, and CDK8/19 Paralogs Revealed within the CDK-Mediator Kinase Module. *Journal of Proteomics & Bioinformatics*, *01*(S2). https://doi.org/10.4172/jpb.s2-004

119. Dao, T. P., Kolaitis, R. M., Kim, H. J., O'Donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J. P., & Castañeda, C. A. (2018). Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Molecular Cell*, *69*(6), 965-978.e6. https://doi.org/10.1016/j.molcel.2018.02.004

120. Darling, A. L., Liu, Y., Oldfield, C. J., & Uversky, V. N. (2018). Intrinsically Disordered Proteome of Human Membrane-Less Organelles. *Proteomics*, *18*(5–6), 1700193. https://doi.org/10.1002/pmic.201700193

121. Darling, A. L., & Uversky, V. N. (2018). Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. *Frontiers in Genetics*, *9*(MAY), 158. https://doi.org/10.3389/fgene.2018.00158

122. Darling, A. L., & Uversky, V. N. (2017). Intrinsic disorder in proteins with pathogenic repeat expansions. In *Molecules* (Vol. 22, Issue 12, p. 2027). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/molecules22122027

123. Daughdrill, G. W., Narayanaswami, P., Gilmore, S. H., Belczyk, A., & Brown, C. J. (2007). Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *Journal of Molecular Evolution*, *65*(3), 277–288. https://doi.org/10.1007/s00239-007-9011-2

124. de Bruin, R. A. M., Kalashnikova, T. I., Chahwan, C., McDonald, W. H., Wohlschlegel, J., Yates, J., Russell, P., & Wittenberg, C. (2006). Constraining G1-Specific Transcription to Late G1 Phase: The MBF-Associated Corepressor Nrm1 Acts via Negative Feedback. *Molecular Cell*, *23*(4), 483–496. https://doi.org/10.1016/J.MOLCEL.2006.06.025/ATTACHMENT/D0FA0778-DBB8-4120-99FF-8A0E6A886A57/MMC1.PDF

125. De Bruin, R. A. M., McDonald, W. H., Kalashnikova, T. I., Yates, J., & Wittenberg, C. (2004). Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell*, *117*(7), 887–898. https://doi.org/10.1016/J.CELL.2004.05.025/ATTACHMENT/3DAE3333-72FF-44BE-8B67-E076E5FFE023/MMC1.PDF

126. Dechat, T., Gotzmann, J., Stockinger, A., Harris, C. A., Talle, M. A., Siekierka, J. J., & Foisner, R. (1998). Detergent-salt resistance of LAP2α in interphase nuclei and phosphorylation-dependent association with chromosomes early in nuclear assembly implies functions in nuclear structure dynamics. *EMBO Journal*, *17*(16), 4887–4902. https://doi.org/10.1093/emboj/17.16.4887

127. Dekkers, J. F., Wiegerinck, C. L., De Jonge, H. R., Bronsveld, I., Janssens, H. M., De Winter-De Groot, K. M., Brandsma, A. M., De Jong, N. W. M., Bijvelds, M. J. C., Scholte, B. J., Nieuwenhuis, E. E. S., Van Den Brink, S., Clevers, H., Van Der Ent, C. K., Middendorp, S., & Beekman, J. M. (2013). A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nature Medicine*, *19*(7), 939–945. https://doi.org/10.1038/nm.3201

128. Delbarre, E., Ivanauskiene, K., Küntziger, T., & Collas, P. (2013). DAXX-dependent supply of soluble (H3.3-H4) dimers to PML bodies pending deposition into chromatin. *Genome Research*, *23*(3), 440–451. https://doi.org/10.1101/gr.142703.112

129. Delbarre, E., Ivanauskiene, K., Spirkoski, J., Shah, A., Vekterud, K., Moskaug, J., Bøe, S. O., Wong, L. H., Küntziger, T., & Collas, P. (2017). PML protein organizes heterochromatin domains where it regulates histone H3.3 deposition by ATRX/DAXX. *Genome Research*, *27*(6), 913–921. https://doi.org/10.1101/gr.215830.116

130. Dellaire, G., Ching, R. W., Dehghani, H., Ren, Y., & Bazett-Jones, D. P. (2006). The number of PML nuclear bodies increases in early S phase by a fission mechanism. *Journal of Cell Science*, *119*(6), 1026–1033. https://doi.org/10.1242/jcs.02816

131. Delmolino, L. M., Saha, P., & Dutta, A. (2001). Multiple Mechanisms Regulate Subcellular Localization of Human CDC6. *Journal of Biological Chemistry*, *276*(29), 26947–26954. https://doi.org/10.1074/jbc.M101870200

132. DePamphilis, M. L. (2003). The "ORC cycle": A novel pathway for regulating eukaryotic DNA replication. *Gene*, *310*(1–2), 1–15. https://doi.org/10.1016/S0378-1119(03)00546-8

133. Diffley, J. F. X., Cocker, J. H., Dowell, S. J., & Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell*, *78*(2), 303–316. https://doi.org/10.1016/0092-8674(94)90299-2

134. DiFiore, B., Davey, N. E., Hagting, A., Izawa, D., Mansfeld, J., Gibson, T. J., & Pines, J. (2015). The ABBA Motif binds APC/C activators and is shared by APC/C substrates and regulators. *Developmental Cell*, *32*(3), 358–372. https://doi.org/10.1016/J.DEVCEL.2015.01.003/ATTACHMENT/A1BFDD3D-D2A5-40C2-8127-5905D0DD7EF1/MMC2.XLSX

135. Dignon, G. L., Zheng, W., Best, R. B., Kim, Y. C., & Mittal, J. (2018). Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(40), 9929–9934. https://doi.org/10.1073/pnas.1804177115

136. Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B., & Mittal, J. (2018). Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Computational Biology*, *14*(1), e1005941. https://doi.org/10.1371/journal.pcbi.1005941

137. Dirick, L., Böhm, T., & Nasmyth, K. (1995). Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of Saccharomyces cerevisiae. *EMBO Journal*, *14*(19), 4803–4813. https://doi.org/10.1002/j.1460-2075.1995.tb00162.x

138. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

139. Dobro, M. J., Oikonomou, C. M., Piper, A., Cohen, J., Guo, K., Jensen, T., Tadayon, J., Donermeyer, J., Park, Y., Solis, B. A., Kjær, A., Jewett, A. I., McDowall, A. W., Chen, S., Chang, Y. W., Shi, J., Subramanian, P., Iancu, C. V., Li, Z., … Jensen, G. J. (2017). Uncharacterized bacterial structures revealed by electron cryotomography. *Journal of Bacteriology*, *199*(17). https://doi.org/10.1128/JB.00100-17

140. Dominski, Z., & Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Molecular and Cellular Biology*, *11*(12), 6075–6083. https://doi.org/10.1128/mcb.11.12.6075-6083.1991

141. Dong, Z., Zhu, C., Zhan, Q., Jiang, W., & Yao, X. (2018). Cdk phosphorylation licenses Kif4A chromosome localization required for early mitotic progression. *Journal of Molecular Cell Biology*, *10*(4), 358–370. https://doi.org/10.1093/jmcb/mjy033

142. Donner, A. J., Ebmeier, C. C., Taatjes, D. J., & Espinosa, J. M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature Structural & Molecular Biology 2010 17:2*, *17*(2), 194–201. https://doi.org/10.1038/NSMB.1752

143. Donner, A. J., Szostek, S., Hoover, J. M., & Espinosa, J. M. (2007). CDK8 Is a Stimulus-Specific Positive Coregulator of p53 Target Genes. *Molecular Cell*, *27*(1), 121–133. https://doi.org/10.1016/j.molcel.2007.05.026

144. Donovan, S., Harwood, J., Drury, L. S., & Diffley, J. F. X. (1997). Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(11), 5611–5616. https://doi.org/10.1073/pnas.94.11.5611

145. Dopie, J., Sweredoski, M. J., Moradian, A., & Belmont, A. S. (2020). Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *Journal of Cell Biology*, *219*(9). https://doi.org/10.1083/jcb.201910207

146. Dosztányi, Z., Csizmók, V., Tompa, P., & Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of Molecular Biology*, *347*(4), 827–839. https://doi.org/10.1016/j.jmb.2005.01.071

147. Douglass, J., Gunaratne, R., Bradford, D., Saeed, F., Hoffert, J. D., Steinbach, P. J., Knepper, M. A., & Pisitkun, T. (2012). Identifying protein kinase target preferences using mass spectrometry. *American Journal of Physiology - Cell Physiology*, *303*(7), 715–727. https://doi.org/10.1152/ajpcell.00166.2012

148. Dranovsky, A., Vincent, I., Gregori, L., Schwarzman, A., Colflesh, D., Enghild, J., Strittmatter, W., Davies, P., & Goldgaber, D. (2001). Cdc2 phosphorylation of nucleolin demarcates mitotic stages and Alzheimer's disease pathology. *Neurobiology of Aging*, *22*(4), 517–528. https://doi.org/10.1016/S0197-4580(00)00248-7

149. Dultz, E., Zanin, E., Wurzenberger, C., Braun, M., Rabut, G., Sironi, L., & Ellenberg, J. (2008). Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells. *Journal of Cell Biology*, *180*(5), 857–865. https://doi.org/10.1083/jcb.200707026

150. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, *41*(21), 6573–6582. https://doi.org/10.1021/bi012159+

151. Dunphy, W. G., Brizuela, L., Beach, D., & Newport, J. (1988). The Xenopus cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis. *Cell*, *54*(3), 423–431. https://doi.org/10.1016/0092-8674(88)90205-X

152. Dunphy, W. G., & Kumagai, A. (1991). The cdc25 protein contains an intrinsic phosphatase activity. *Cell*, *67*(1), 189–196. https://doi.org/10.1016/0092-8674(91)90582-J

153. Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology 2005 6:3*, *6*(3), 197–208. https://doi.org/10.1038/nrm1589

154. Ehre, C., Ridley, C., & Thornton, D. J. (2014). Cystic fibrosis: An inherited disease affecting mucin-producing organs. *International Journal of Biochemistry and Cell Biology*, *52*, 136–145. https://doi.org/10.1016/j.biocel.2014.03.011

155. El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K. R., Park, S., Wang, X., Krebs, J., Tripathi, S., Sakabe, N., Sobreira, D. R., Huang, S. C., Rao, S. S. P., Pruett, N., Chauss, D., Sadler, E., … Casellas, R. (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell*, *178*(5), 1145-1158.e20. https://doi.org/10.1016/j.cell.2019.07.011

156. El Marjou, F., Janssen, K. P., Chang, B. H. J., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., & Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis*, *39*(3), 186–193. https://doi.org/10.1002/gene.20042

157. Elzen, N. Den, & Pines, J. (2001). Cyclin A is destroyed in prometaphase and can delay chromosome alignment and anaphase. *Journal of Cell Biology*, *153*(1), 121–135. https://doi.org/10.1083/jcb.153.1.121

158. Emanuele, M. J., Lan, W., Jwa, M., Miller, S. A., Chan, C. S. M., & Stukenberg, P. T. (2008). Aurora B kinase and protein phosphatase 1 have opposing roles in modulating kinetochore assembly. *Journal of Cell Biology*, *181*(2), 241–254. https://doi.org/10.1083/JCB.200710019

159. Endl, E., & Gerdes, J. (2000). The Ki-67 protein: Fascinating forms and an unknown function. *Experimental Cell Research*, *257*(2), 231–237. https://doi.org/10.1006/excr.2000.4888

160. Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, *3*(11), e3647. https://doi.org/10.1371/journal.pone.0003647

161. Enserink, J. M. (2015). Sumo and the cellular stress response. *Cell Division*, *10*(1), 1–13. https://doi.org/10.1186/s13008-015-0010-1

162. Erdel, F., Rademacher, A., Vlijm, R., Tünnermann, J., Frank, L., Weinmann, R., Schweigert, E., Yserentant, K., Hummert, J., Bauer, C., Schumacher, S., Al Alwash, A., Normand, C., Herten, D. P., Engelhardt, J., & Rippe, K. (2020). Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. *Molecular Cell*, *78*(2), 236-249.e7. https://doi.org/10.1016/j.molcel.2020.02.005

163. Errico, A., Deshmukh, K., Tanaka, Y., Pozniakovsky, A., & Hunt, T. (n.d.). *Identification of substrates for cyclin dependent kinases*. https://doi.org/10.1016/j.advenzreg.2009.12.001

164. Espinoza, F. H., Farrell, A., Erdjument-Bromage, H., Tempst, P., & Morgan, D. O. (1996). A cyclin-dependent kinase-activating kinase (CAK) in budding yeast unrelated to vertebrate CAK. *Science (New York, N.Y.)*, *273*(5282), 1714–1717. https://doi.org/10.1126/SCIENCE.273.5282.1714

165. Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D., & Hunt, T. (1983). Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell*, *33*(2), 389–396. https://doi.org/10.1016/0092-8674(83)90420-8

166. Everett, R. D., Lomonte, P., Sternsdorf, T., Van Driel, R., & Orr, A. (1999). Cell cycle regulation of PML modification and ND10 composition. *Journal of Cell Science*, *112*(24), 4581–4588. https://doi.org/10.1242/jcs.112.24.4581

167. Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., & Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(48), 20240–20245. https://doi.org/10.1073/pnas.0911500106

168. Feldman, H. J., & Hogue, C. W. V. (2000). A fast method to sample real protein conformational space. *Proteins: Structure, Function and Genetics*, *39*(2), 112–131. https://doi.org/10.1002/(SICI)1097-0134(20000501)39:2<112::AID-PROT2>3.0.CO;2-B

169. Feldman, H. J., & Hogue, C. W. V. (2002). Probabilistic sampling of protein conformations: New hope for brute force? *Proteins: Structure, Function and Genetics*, *46*(1), 8–23. https://doi.org/10.1002/prot.1163

170. Felix, M. A., Cohen, P., & Karsenti, E. (1990). Cdc2 H1 kinase is negatively regulated by a type 2A phosphatase in the Xenopus early embryonic cell cycle: evidence from the effects of okadaic acid. *The EMBO Journal*, *9*(3), 675–683. https://doi.org/10.1002/J.1460-2075.1990.TB08159.X

171. Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V., & Brangwynne, C. P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, *165*(7), 1686–1697. https://doi.org/10.1016/j.cell.2016.04.047

172. Fernandez-Capetillo, O., Chen, H. T., Celeste, A., Ward, I., Romanienko, P. J., Morales, J. C., Naka, K., Xia, Z., Camerini-Otero, R. D., Motoyama, N., Carpenter, P. B., Bonner, W. M., Chen, J., & Nussenzweig, A. (2002). DNA damage-induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biology*, *4*(12), 993–997. https://doi.org/10.1038/ncb884

173. Ferrell, J. E., Wu, M., Gerhart, J. C., & Martin, G. S. (1991). Cell cycle tyrosine phosphorylation of p34cdc2 and a microtubule-associated protein kinase homolog in Xenopus oocytes and eggs. *Molecular and Cellular Biology*, *11*(4), 1965–1971. https://doi.org/10.1128/mcb.11.4.1965-1971.1991

174. Ferrell, J. E. (1996). Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends in Biochemical Sciences*, *21*(12), 460–466. https://doi.org/10.1016/S0968-0004(96)20026-X

175. Ferrezuelo, F., Colomina, N., Futcher, B., & Aldea, M. (2010). The transcriptional network activated by Cln3 cyclin at the G1-to-S transition of the yeast cell cycle. *Genome Biology*, *11*(6), 1–18. https://doi.org/10.1186/gb-2010-11-6-r67

176. Firestein, R., Bass, A. J., Kim, S. Y., Dunn, I. F., Silver, S. J., Guney, I., Freed, E., Ligon, A. H., Vena, N., Ogino, S., Chheda, M. G., Tamayo, P., Finn, S., Shrestha, Y., Boehm, J. S., Jain, S., Bojarski, E., Mermel, C., Barretina, J., … Hahn, W. C. (2008). CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature*, *455*(7212), 547–551. https://doi.org/10.1038/nature07179

177. Fisher, C. K., & Stultz, C. M. (2011). Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology*, *21*(3), 426–431. https://doi.org/10.1016/j.sbi.2011.04.001

178. Fisher, D. L., & Nurse, P. (1996). A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S-phase and mitosis in the absence of G1 cyclins. *EMBO Journal*, *15*(4), 850–860. https://doi.org/10.1002/j.1460-2075.1996.tb00420.x

179. Fisher, D., & Krasinska, L. (2022). Explaining Redundancy in CDK-Mediated Control of the Cell Cycle: Unifying the Continuum and Quantitative Models. *Cells 2022, Vol. 11, Page 2019*, *11*(13), 2019. https://doi.org/10.3390/CELLS11132019

180. Fisher, D., Krasinska, L., Coudreuse, D., & Novák, B. (2012). Phosphorylation network dynamics in the control of cell cycle transitions. *Journal of Cell Science*, *125*(20), 4703–4711. https://doi.org/10.1242/JCS.106351

181. Fisher, R. P., & Morgan, D. O. (1994). A novel cyclin associates with M015/CDK7 to form the CDK-activating kinase. *Cell*, *78*(4), 713–724. https://doi.org/10.1016/0092-8674(94)90535-5

182. Flick, K., & Kaiser, P. (2012). Protein degradation and the stress response. *Seminars in Cell and Developmental Biology*, *23*(5), 515–522. https://doi.org/10.1016/j.semcdb.2012.01.019

183. Fong, K. wing, Li, Y., Wang, W., Ma, W., Li, K., Qi, R. Z., Liu, D., Songyang, Z., & Chen, J. (2013). Whole-genome screening identifies proteins localized to distinct nuclear bodies. *Journal of Cell Biology*, *203*(1), 149–164. https://doi.org/10.1083/jcb.201303145

184. Forsburg, S. L., & Nurse, P. (1991). Identification of a Gl-type cyclin puc1 + in the fission yeast Schizosaccharomyces pombe. *Nature*, *351*(6323), 245–248. https://doi.org/10.1038/351245a0

185. Fourest-Lieuvin, A., Peris, L., Gache, V., Garcia-Saez, I., Juillan-Binard, C., Lantez, V., & Job, D. (2006). Microtubule regulation in mitosis: Tubulin phosphorylation by the cyclin-dependent kinase Cdk1. *Molecular Biology of the Cell*, *17*(3), 1041–1050. https://doi.org/10.1091/mbc.E05-07-0621

186. Fryer, C. J., White, J. B., & Jones, K. A. (2004). Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Molecular Cell*, *16*(4), 509–520. https://doi.org/10.1016/j.molcel.2004.10.014

187. Fukuchi, S., Sakamoto, S., Nobe, Y., Murakami, S. D., Amemiya, T., Hosoda, K., Koike, R., Hiroaki, H., & Ota, M. (2012). IDEAL: Intrinsically disordered proteins with extensive annotations and literature. *Nucleic Acids Research*, *40*(D1). https://doi.org/10.1093/nar/gkr884

188. Furuno, N., Elzen, N. Den, & Pines, J. (1999). Human cyclin A is required for mitosis until mid prophase. *Journal of Cell Biology*, *147*(2), 295–306. https://doi.org/10.1083/jcb.147.2.295

189. Fuxreiter, M., Tóth-Petróczy, Á., Kraut, D. A., Matouschek, A. T., Lim, R. Y. H., Xue, B., Kurgan, L., & Uversky, V. N. (2014). Disordered proteinaceous machines. In *Chemical Reviews* (Vol. 114, Issue 13, pp. 6806–6843). American Chemical Society. https://doi.org/10.1021/cr4007329

190. Gaggioli, V., Zeiser, E., Rivers, D., Bradshaw, C. R., Ahringer, J., & Zegerman, P. (2014). CDK phosphorylation of SLD-2 is required for replication initiation and germline development in C. elegans. *Journal of Cell Biology*, *204*(4), 507–522. https://doi.org/10.1083/jcb.201310083

191. Galbraith, M. D., Allen, M. A., Bensard, C. L., Wang, X., Schwinn, M. K., Qin, B., Long, H. W., Daniels, D. L., Hahn, W. C., Dowell, R. D., & Espinosa, J. M. (2013). XHIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell*, *153*(6), 1327. https://doi.org/10.1016/j.cell.2013.04.048

192. Galea, C. A., Wang, Y., Sivakolundu, S. G., & Kriwacki, R. W. (2008). Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. In *Biochemistry* (Vol. 47, Issue 29, pp. 7598–7609). https://doi.org/10.1021/bi8006803

193. Gall, J. G. (2000). Cajal bodies: The first 100 years. *Annual Review of Cell and Developmental Biology*, *16*, 273–300. https://doi.org/10.1146/annurev.cellbio.16.1.273

194. Gambus, A., Khoudoli, G. A., Jones, R. C., & Blow, J. J. (2011). MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. *Journal of Biological Chemistry*, *286*(13), 11855–11864. https://doi.org/10.1074/jbc.M110.199521

195. Ganuza, M., Sáiz-Ladera, C., Cañamero, M., Gómez, G., Schneider, R., Blasco, M. A., Pisano, D., Paramio, J. M., Santamaría, D., & Barbacid, M. (2012). Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion. *The EMBO Journal*, *31*(11), 2498–2510. https://doi.org/10.1038/EMBOJ.2012.94

196. Gao, J., & Xu, D. (2012). Correlation between posttranslational modification and intrinsic disorder in Protein. *Pacific Symposium on Biocomputing*, 94–103. https://doi.org/10.1142/9789814366496_0010

197. Garcia, D. A., Johnson, T. A., Presman, D. M., Fettweis, G., Wagh, K., Rinaldi, L., Stavreva, D. A., Paakinaho, V., Jensen, R. A. M., Mandrup, S., Upadhyaya, A., & Hager, G. L. (2021). An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. *Molecular Cell*, *81*(7), 1484-1498.e6. https://doi.org/10.1016/j.molcel.2021.01.013

198. Gareau, J. R., & Lima, C. D. (2010). The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition. *Nature Reviews Molecular Cell Biology*, *11*(12), 861–871. https://doi.org/10.1038/nrm3011

199. Garg, P., & Burgers, P. M. J. (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. *Critical Reviews in Biochemistry and Molecular Biology*, *40*(2), 115–128. https://doi.org/10.1080/10409230590935433

200. Gautier, J., Norbury, C., Lohka, M., Nurse, P., & Maller, J. (1988). Purified maturation-promoting factor contains the product of a Xenopus homolog of the fission yeast cell cycle control gene cdc2+. *Cell*, *54*(3), 433–439. https://doi.org/10.1016/0092-8674(88)90206-1

201. Gautier, J., Solomon, M. J., Booher, R. N., Bazan, J. F., & Kirschner, M. W. (1991). Cdc25 Is a Specific Tyrosine Phosphatase That Directly Activates P34Cdc2. *Cell*, *67*(1), 197–211. https://doi.org/10.1016/0092-8674(91)90583-K

202. Gavet, O., & Pines, J. (2010). Activation of cyclin B1-Cdk1 synchronizes events in the nucleus and the cytoplasm at mitosis. *Journal of Cell Biology*, *189*(2), 247–259. https://doi.org/10.1083/jcb.200909144

203. Gavet, O., & Pines, J. (2010). Progressive Activation of CyclinB1-Cdk1 Coordinates Entry to Mitosis. *Developmental Cell*, *18*(4), 533–543. https://doi.org/10.1016/j.devcel.2010.02.013

204. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., … Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology 2004 5:10*, *5*(10), 1–16. https://doi.org/10.1186/GB-2004-5-10-R80

205. Gerhart, J., Wu, M., & Kirschner, M. (1984). Cell cycle dynamics of an M-phase-specific cytoplasmic factor in Xenopus laevis oocytes and eggs. *Journal of Cell Biology*, *98*(4), 1247–1255. https://doi.org/10.1083/jcb.98.4.1247

206. Gharbi-Ayachi, A., Labbé, J. C., Burgess, A., Vigneron, S., Strub, J. M., Brioudes, E., Van-Dorsselaer, A., Castro, A., & Lorca, T. (2010). The substrate of Greatwall kinase, Arpp19, controls mitosis by inhibiting protein phosphatase 2A. *Science*, *330*(6011), 1673–1677. https://doi.org/10.1126/science.1197048

207. Ghodke, I., Remisova, M., Furst, A., Kilic, S., Reina-San-Martin, B., Poetsch, A. R., Altmeyer, M., & Soutoglou, E. (2021). AHNAK controls 53BP1-mediated p53 response by restraining 53BP1 oligomerization and phase separation. *Molecular Cell*, *81*(12), 2596-2610.e7. https://doi.org/10.1016/j.molcel.2021.04.010

208. Gill, G. (2004). SUMO and ubiquitin in the nucleus: Different functions, similar mechanisms? *Genes and Development*, *18*(17), 2046–2059. https://doi.org/10.1101/gad.1214604

209. Glotzer, M., Murray, A. W., & Kirschner, M. W. (1991). Cyclin is degraded by the ubiquitin pathway. *Nature 1991 349:6305*, *349*(6305), 132–138. https://doi.org/10.1038/349132a0

210. Golan, A., Yudkovsky, Y., & Hershko, A. (2002). The cyclin-ubiquitin ligase activity of cyclosome/APC is jointly activated by protein kinases Cdk1-cyclin B and Plk. *Journal of Biological Chemistry*, *277*(18), 15552–15557. https://doi.org/10.1074/jbc.M111476200

211. Goldbeter, A., & Koshland, D. E. (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(11 II), 6840–6844. https://doi.org/10.1073/pnas.78.11.6840

212. Gomes, E., & Shorter, J. (2019). The molecular language of membraneless organelles. *Journal of Biological Chemistry*, *294*(18), 7115–7127. https://doi.org/10.1074/jbc.TM118.001192

213. Gong, D., Pomerening, J. R., Myers, J. W., Gustavsson, C., Jones, J. T., Hahn, A. T., Meyer, T., & Ferrell, J. E. (2007). Cyclin A2 Regulates Nuclear-Envelope Breakdown and the Nuclear Accumulation of Cyclin B1. *Current Biology*, *17*(1), 85–91. https://doi.org/10.1016/j.cub.2006.11.066

214. Gorr, I. H., Boos, D., & Stemmann, O. (2005). Mutual inhibition of separase and Cdk1 by two-step complex formation. *Molecular Cell*, *19*(1), 135–141. https://doi.org/10.1016/j.molcel.2005.05.022

215. Goto, H., Kiyono, T., Tomono, Y., Kawajiri, A., Urano, T., Furukawa, K., Nigg, E. A., & Inagaki, M. (2006). Complex formation of Plk1 and INCENP required for metaphase-anaphase transition. *Nature Cell Biology*, *8*(2), 180–187. https://doi.org/10.1038/ncb1350

216. Gotoh, E., & Durante, M. (2006). Chromosome condensation outside of mitosis: Mechanisms and new tools. *Journal of Cellular Physiology*, *209*(2), 297–304. https://doi.org/10.1002/jcp.20720

217. Gould, K. L., & Nurse, P. (1989). Tyrosine phosphorylation of the fission yeast cdc2+ protein kinase regulates entry into mitosis. *Nature*, *342*(6245), 39–45. https://doi.org/10.1038/342039a0

218. Greening, C., & Lithgow, T. (2020). Formation and function of bacterial organelles. *Nature Reviews Microbiology*, *18*(12), 677–689. https://doi.org/10.1038/s41579-020-0413-0

219. Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

220. Guerrier, P., Dorée, M., & Freyssinet, G. (1975). Stimulation précoce des activités protéines kinases au cours du processus hormonal de réinitiation de la méiose dans les ovocytes d'Etoiles de mer. *Comptes rendus hebdomadaires des seances de l"Academie des sciences. Serie D: Sciences naturelles*, *281*(20), 1475–1478.

221. Guharoy, M., Szabo, B., Martos, S. C., Kosol, S., & Tompa, P. (2013). Intrinsic structural disorder in cytoskeletal proteins. *Cytoskeleton*, *70*(10), 550–571. https://doi.org/10.1002/cm.21118

222. Gunbin, K. V., Suslov, V. V., Turnaev, I. I., Afonnikov, D. A., & Kolchanov, N. A. (2011). Molecular evolution of cyclin proteins in animals and fungi. *BMC Evolutionary Biology 2011 11:1*, *11*(1), 1–20. https://doi.org/10.1186/1471-2148-11-224

223. Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., Spille, J. H., Afeyan, L. K., Zamudio, A. V., Shrinivas, K., Abraham, B. J., Boija, A., Decker, T. M., Rimel, J. K., Fant, C. B., Lee, T. I., Cisse, I. I., Sharp, P. A., Taatjes, D. J., & Young, R. A. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, *572*(7770), 543–548. https://doi.org/10.1038/s41586-019-1464-0

224. Hadwiger, J. A., Wittenberg, C., Richardson, H. E., De Barros Lopes, M., & Reed, S. I. (1989). A family of cyclin homologs that control the G1 phase in yeast). *Proceedings of the National Academy of Sciences of the United States of America*, *86*(16), 6255–6259. https://doi.org/10.1073/pnas.86.16.6255

225. Hagan, I., Hayles, J., & Nurse, P. (1988). Cloning and sequencing of the cyclin-related cdc13+ gene and a cytological study of its role in fission yeast mitosis. *Journal of Cell Science*, *91 (Pt 4)*(4), 587–595. https://doi.org/10.1242/jcs.91.4.587

226. Hagting, A., Den Elzen, N., Vodermaier, H. C., Waizenegger, I. C., Peters, J. M., & Pines, J. (2002). Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. *Journal of Cell Biology*, *157*(7), 1125–1137. https://doi.org/10.1083/jcb.200111001

227. Hancock, R. (2004). A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *Journal of Structural Biology*, *146*(3), 281–290. https://doi.org/10.1016/j.jsb.2003.12.008

228. Handwerger, K. E., Murphy, C., & Gall, J. G. (2003). Steady-state dynamics of Cajal body components in the Xenopus germinal vesicle. *Journal of Cell Biology*, *160*(4), 495–504. https://doi.org/10.1083/jcb.200212024

229. Hanson, J., Yang, Y., Paliwal, K., & Zhou, Y. (2017). Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics*, *33*(5), 685–692. https://doi.org/10.1093/bioinformatics/btw678

230. Harkey, T., Govind Kumar, V., Hettige, J., Tabari, S. H., Immadisetty, K., & Moradi, M. (2019). The Role of a Crystallographically Unresolved Cytoplasmic Loop in Stabilizing the Bacterial Membrane Insertase YidC2. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-51052-9

231. Harmon, T. S., & Jülicher, F. (2022). Molecular Assembly Lines in Active Droplets. *Physical Review Letters*, *128*(10). https://doi.org/10.1103/PhysRevLett.128.108102

232. Harper, J. W., & Elledge, S. J. (1998). The role of Cdk7 in CAK function, a retro-retrospective. *Genes & Development*, *12*(3), 285–289. https://doi.org/10.1101/GAD.12.3.285

233. Hartwell, L. H., Culotti, J., & Reid, B. (1970). Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *66*(2), 352–359. https://doi.org/10.1073/pnas.66.2.352

234. Hartwell, L. H., Culotti, J., Pringle, J. R., & Reid, B. J. (1974). Genetic control of the cell division cycle in yeast. In *Science* (Vol. 183, Issue 4120, pp. 46–51). https://doi.org/10.1126/science.183.4120.46

235. Hase, M. E., & Cordes, V. C. (2003). Direct interaction with Nup153 mediates binding of Tpr to the periphery of the nuclear pore complex. *Molecular Biology of the Cell*, *14*(5), 1923–1940. https://doi.org/10.1091/mbc.E02-09-0620

236. Hayashi, Y., Kato, K., & Kimura, K. (2017). The hierarchical structure of the perichromosomal layer comprises Ki67, ribosomal RNAs, and nucleolar proteins. *Biochemical and Biophysical Research Communications*, *493*(2), 1043–1049. https://doi.org/10.1016/J.BBRC.2017.09.092

237. Heald, R., & McKeon, F. (1990). Mutations of phosphorylation sites in lamin A that prevent nuclear lamina disassembly in mitosis. *Cell*, *61*(4), 579–589. https://doi.org/10.1016/0092-8674(90)90470-Y

238. Hearst, S. M., Gilder, A. S., Negi, S. S., Davis, M. D., George, E. M., Whittom, A. A., Toyota, C. G., Husedzinovic, A., Gruss, O. J., & Hebert, M. D. (2009). Cajal-body formation correlates with differential coilin phosphorylation in primary and transformed cell lines. *Journal of Cell Science*, *122*(11), 1872–1881. https://doi.org/10.1242/jcs.044040

239. Hégarat, N., Crncec, A., Suarez Peredo Rodriguez, M. F., Echegaray Iturra, F., Gu, Y., Busby, O., Lang, P. F., Barr, A. R., Bakal, C., Kanemaki, M. T., Lamond, A. I., Novak, B., Ly, T., & Hochegger, H. (2020). Cyclin A triggers Mitosis either via the Greatwall kinase pathway or Cyclin B. *The EMBO Journal*, *39*(11), e104419. https://doi.org/10.15252/embj.2020104419

240. Hégarat, N., Rata, S., & Hochegger, H. (2016). Bistability of mitotic entry and exit switches during open mitosis in mammalian cells. *BioEssays*, *38*(7), 627–643. https://doi.org/10.1002/bies.201600057

241. Hégarat, N., Vesely, C., Vinod, P. K., Ocasio, C., Peter, N., Gannon, J., Oliver, A. W., Novák, B., & Hochegger, H. (2014). PP2A/B55 and Fcp1 Regulate Greatwall and Ensa Dephosphorylation during Mitotic Exit. *PLoS Genetics*, *10*(1), e1004004. https://doi.org/10.1371/journal.pgen.1004004

242. Hellmuth, S., Gómez-H, L., Pendás, A. M., & Stemmann, O. (2020). Securin-independent regulation of separase by checkpoint-induced shugoshin–MAD2. *Nature*, *580*(7804), 536–541. https://doi.org/10.1038/s41586-020-2182-3

243. Heppner, G. H., Miller, F. R., & Shekhar, P. V. M. (2000). Nontransgenic models of breast cancer. *Breast Cancer Research*, *2*(5), 331–334. https://doi.org/10.1186/bcr77

244. Hernandez-Verdun, D. (2011). Assembly and disassembly of the nucleolus during the cell cycle. *Nucleus*, *2*(3), 189–194. https://doi.org/10.4161/nucl.2.3.16246

245. Hindley, J., & Phear, G. A. (1984). Sequence of the cell division gene CDC2 from Schizosaccharomyces pombe; patterns of splicing and homology to protein kinases. *Gene*, *31*(1–3), 129–134. https://doi.org/10.1016/0378-1119(84)90203-8

246. Hiraga, S. I., Alvino, G. M., Chang, F. J., Lian, H. Y., Sridhar, A., Kubota, T., Brewer, B. J., Weinreich, M., Raghuraman, M. K., & Donaldson, A. D. (2014). Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes & Development*, *28*(4), 372–383. https://doi.org/10.1101/GAD.231258.113

247. Hirano, T. (2005). Condensins: Organizing and segregating the genome. *Current Biology*, *15*(7), R265–R275. https://doi.org/10.1016/j.cub.2005.03.037

248. Hiruma, Y., Sacristan, C., Pachis, S. T., Adamopoulos, A., Kuijt, T., Ubbink, M., Von Castelmur, E., Perrakis, A., & Kops, G. J. P. L. (2015). Competition between MPS1 and microtubules at kinetochores regulates spindle checkpoint signaling. *Science*, *348*(6240), 1264–1267. https://doi.org/10.1126/science.aaa4055

249. Hisamoto, N., Sugimoto, K., & Matsumoto, K. (1994). The Glc7 type 1 protein phosphatase of Saccharomyces cerevisiae is required for cell cycle progression in G2/M. *Molecular and Cellular Biology*, *14*(5), 3158–3165. https://doi.org/10.1128/MCB.14.5.3158-3165.1994

250. Hochegger, H., Dejsuphong, D., Sonoda, E., Saberi, A., Rajendra, E., Kirk, J., Hunt, T., & Takeda, S. (2007). An essential role for Cdk1 in S phase control is revealed via chemical genetics in vertebrate cells. *Journal of Cell Biology*, *178*(2), 257–268. https://doi.org/10.1083/jcb.200702034

251. Hoffmann, I., Clarke, P. R., Marcote, M. J., Karsenti, E., & Draetta, G. (1993). Phosphorylation and activation of human cdc25-C by cdc2-cyclin B and its involvement in the self-amplification of MPF at mitosis. *EMBO Journal*, *12*(1), 53–63. https://doi.org/10.1002/j.1460-2075.1993.tb05631.x

252. Hofweber, M., & Dormann, D. (2019). Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics. *Journal of Biological Chemistry*, *294*(18), 7137–7150. https://doi.org/10.1074/jbc.TM118.001189

253. Hoischen, C., Monajembashi, S., Weisshart, K., & Hemmerich, P. (2018). Multimodal light microscopy approaches to reveal structural and functional properties of promyelocytic leukemia nuclear bodies. *Frontiers in Oncology*, *8*(MAY), 125. https://doi.org/10.3389/fonc.2018.00125

254. Holland, A. J., & Taylor, S. S. (2008). Many faces of separase regulation. *SEB Experimental Biology Series*, *59*, 99–112. https://www.researchgate.net/publication/5483457

255. Holmes, J. K., & Solomon, M. J. (1996). A Predictive Scale for Evaluating Cyclin-dependent Kinase Substrates: A COMPARISON OF p34cdc2 AND p33cdk2*. *Journal of Biological Chemistry*, *271*(41), 25240–25246. https://doi.org/10.1074/JBC.271.41.25240

256. Holmstrom, E. D., Liu, Z., Nettels, D., Best, R. B., & Schuler, B. (2019). Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-10356-0

257. Holt, C. (2013). Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. In *Current Opinion in Structural Biology* (Vol. 23, Issue 3, pp. 420–425). https://doi.org/10.1016/j.sbi.2013.02.010

258. Holt, L. J., Tuch, B. B., Villen, J., Johnson, A. D., Gygi, S. P., & Morgan, D. O. (2009). Global analysis of cdk1 substrate phosphorylation sites provides insiqhts into evolution. *Science*, *325*(5948), 1682–1686. https://doi.org/10.1126/science.1172867

259. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*, *43*(D1), D512–D520. https://doi.org/10.1093/NAR/GKU1267

260. Huang, C. Y. F., & Ferrell, J. E. (1996). Dependence of Mos-induced Cdc2 activation on MAP kinase function in a cell-free system. *The EMBO Journal*, *15*(9), 2169–2173. https://doi.org/10.1002/J.1460-2075.1996.TB00570.X

261. Hubbard, S. J., Eisenmenger, F., & Thornton, J. M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Science*, *3*(5), 757–768. https://doi.org/10.1002/pro.5560030505

262. Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J. B., Munier, A., Fradet, M., Daunesse, M., Bertrand, E., Pierron, G., Mozziconacci, J., Kress, M., & Weil, D. (2017). P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell*, *68*(1), 144-157.e5. https://doi.org/10.1016/j.molcel.2017.09.003

263. Hughes, M. P., Sawaya, M. R., Boyer, D. R., Goldschmidt, L., Rodriguez, J. A., Cascio, D., Chong, L., Gonen, T., & Eisenberg, D. S. (2018). Atomic structures of low-complexity protein segments reveal kinked b sheets that assemble networks. *Science*, *359*(6376), 698–701. https://doi.org/10.1126/science.aan6398

264. Huihui, J., & Ghosh, K. (2020). An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *The Journal of Chemical Physics*, *152*(16), 161102. https://doi.org/10.1063/5.0004619

265. Huihui, J., & Ghosh, K. (2021). Intrachain interaction topology can identify functionally similar intrinsically disordered proteins. *Biophysical Journal*, *120*(10), 1860–1868. https://doi.org/10.1016/j.bpj.2020.11.2282

266. Hülsmann, B. B., Labokha, A. A., & Görlich, D. (2012). The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell*, *150*(4), 738–751. https://doi.org/10.1016/j.cell.2012.07.019

267. Hume, S., Dianov, G. L., & Ramadan, K. (2020). A unified model for the G1/S cell cycle transition. *Nucleic Acids Research*, *48*(22), 12483–12501. https://doi.org/10.1093/nar/gkaa1002

268. Hur, W., Kemp, J. P., Tarzia, M., Deneke, V. E., Marzluff, W. F., Duronio, R. J., & Di Talia, S. (2020). CDK-Regulated Phase Separation Seeded by Histone Genes Ensures Precise Growth and Function of Histone Locus Bodies. *Journal of Cleaner Production*, *54*(3), 379-394.e6. https://doi.org/10.1016/j.devcel.2020.06.003

269. Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annual Review of Cell and Developmental Biology*, *30*, 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325

270. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, *32*(3), 1037–1049. https://doi.org/10.1093/nar/gkh253

271. Iqbal, S., & Hoque, M. T. (2015). DisPredict: A predictor of disordered protein using optimized RBF kernel. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0141551

272. Ira, G., Pellicioll, A., Balijja, A., Wang, X., Florani, S., Carotenuto, W., Liberi, G., Bressan, D., Wan, L., Hollingsworth, N. M., Haber, J. E., & Folani, M. (2004). DNA end resection, homologous

recombination and DNA damage checkpoint activation require CDK1. *Nature*, *431*(7011), 1011–1017. https://doi.org/10.1038/nature02964

273. Ishida, S., Huang, E., Zuzan, H., Spang, R., Leone, G., West, M., & Nevins, J. R. (2001). Role for E2F in Control of Both DNA Replication and Mitotic Functions as Revealed from DNA Microarray Analysis. *Molecular and Cellular Biology*, *21*(14), 4684–4699. https://doi.org/10.1128/MCB.21.14.4684-4699.2001/ASSET/8C86675A-77F0-441B-A0DF-B892DAA4E954/ASSETS/GRAPHIC/MB1410025006.JPEG

274. Ivanyi-Nagy, R., Davidovic, L., Khandjian, E. W., & Darlix, J. L. (2005). Disordered RNA chaperone proteins: From functions to disease. In *Cellular and Molecular Life Sciences* (Vol. 62, Issue 13, pp. 1409–1417). https://doi.org/10.1007/s00018-005-5100-9

275. Jackson, L. P., Reed, S. I., & Haase, S. B. (2006). Distinct Mechanisms Control the Stability of the Related S-Phase Cyclins Clb5 and Clb6. *Molecular and Cellular Biology*, *26*(6), 2456–2466. https://doi.org/10.1128/mcb.26.6.2456-2466.2006

276. Jain, S., Wheeler, J. R., Walters, R. W., Agrawal, A., Barsic, A., & Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell*, *164*(3), 487–498. https://doi.org/10.1016/j.cell.2015.12.038

277. Janin, J., & Sternberg, M. J. E. (2013). Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biology Reports*, *5*(1). https://doi.org/10.3410/B5-2

278. Janke, C., & Magiera, M. M. (2020). The tubulin code and its role in controlling microtubule properties and functions. In *Nature Reviews Molecular Cell Biology* (Vol. 21, Issue 6, pp. 307–326). Nature Publishing Group. https://doi.org/10.1038/s41580-020-0214-3

279. Jaspersen, S. L., Charles, J. F., & Morgan, D. O. (1999). Inhibitory phosphorylation of the APC regulator Hct1 is controlled by the kinase Cdc28 and the phosphatase Cdc14. *Current Biology*, *9*(5), 227–236. https://doi.org/10.1016/S0960-9822(99)80111-0

280. Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J., & Pavletich, N. P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*, *376*(6538), 313–320. https://doi.org/10.1038/376313a0

281. Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science*, *293*(5532), 1074–1080. https://doi.org/10.1126/science.1063127

282. Jeronimo, C., Langelier, M. F., Bataille, A. R., Pascal, J. M., Pugh, B. F., & Robert, F. (2016). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Molecular Cell*, *64*(3), 455–466. https://doi.org/10.1016/j.molcel.2016.09.002

283. Ji, Z., Gao, H., & Yu, H. (2015). Kinetochore attachment sensed by competitive Mps1 and microtubule binding to Ndc80C. *Science*, *348*(6240), 1260–1264. https://doi.org/10.1126/science.aaa4029

284. Jiang, W., Wells, N. J., & Hunter, T. (1999). Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 6193–6198. https://doi.org/10.1073/pnas.96.11.6193

285. Johnson, D. G., Ohtani, K., & Nevins, J. R. (1994). Autoregulatory control of E2F1 expression in response to positive and negative regulators of cell cycle progression. *Genes and Development*, *8*(13), 1514–1525. https://doi.org/10.1101/gad.8.13.1514

286. Kalaszczynska, I., Geng, Y., Iino, T., Mizuno, S. ichi, Choi, Y., Kondratiuk, I., Silver, D. P., Wolgemuth, D. J., Akashi, K., & Sicinski, P. (2009). Cyclin A Is Redundant in Fibroblasts but Essential in Hematopoietic and Embryonic Stem Cells. *Cell*, *138*(2), 352–365. https://doi.org/10.1016/j.cell.2009.04.062

287. Kao, L., Wang, Y. T., Chen, Y. C., Tseng, S. F., Jhang, J. C., Chen, Y. J., & Teng, S. C. (2014). Global analysis of cdc14 dephosphorylation sites reveals essential regulatory role in mitosis and cytokinesis. *Molecular and Cellular Proteomics*, *13*(2), 594–605.

https://doi.org/10.1074/MCP.M113.032680/ATTACHMENT/017D91D7-EF09-41B7-864F-DF1F4F2BAB00/MMC1.ZIP

288. Kapoor, A., Goldberg, M. S., Cumberland, L. K., Ratnakumar, K., Segura, M. F., Emanuel, P. O., Menendez, S., Vardabasso, C., LeRoy, G., Vidal, C. I., Polsky, D., Osman, I., Garcia, B. A., Hernando, E., & Bernstein, E. (2010). The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature*, *468*(7327), 1105–1111. https://doi.org/10.1038/nature09590

289. Kara, N., Hossain, M., Prasanth, S. G., & Stillman, B. (2015). Orc1 binding to mitotic chromosomes precedes spatial patterning during G1 phase and assembly of the origin recognition complex in human cells. *Journal of Biological Chemistry*, *290*(19), 12355–12369. https://doi.org/10.1074/jbc.M114.625012

290. Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., & McKnight, S. L. (2012). Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell*, *149*(4), 753–767. https://doi.org/10.1016/j.cell.2012.04.017

291. Katuwawala, A., Oldfield, C. J., & Kurgan, L. (2020). Accuracy of protein-level disorder predictions. In *Briefings in Bioinformatics* (Vol. 21, Issue 5, pp. 1509–1522). https://doi.org/10.1093/bib/bbz100

292. Kausch, I., Lingnau, A., Endl, E., Sellmann, K., Deinert, I., Ratliff, T. L., Jocham, D., Sczakiel, G., Gerdes, J., & Böhle, A. (2003). Antisense treatment against Ki-67 mRNA inhibits proliferation and tumor growth in vitro and in vivo. *International Journal of Cancer*, *105*(5), 710–716. https://doi.org/10.1002/ijc.11111

293. Keiten-Schmitz, J., Röder, L., Hornstein, E., Müller-McNicoll, M., & Müller, S. (2021). SUMO: Glue or Solvent for Phase-Separated Ribonucleoprotein Complexes and Molecular Condensates? *Frontiers in Molecular Biosciences*, *8*, 307. https://doi.org/10.3389/fmolb.2021.673038

294. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, *181*(4610), 662–666. https://doi.org/10.1038/181662a0

295. Kendrew, J. C., & Parrish, R. . (1957). The crystal structure of myoglobin III. Sperm-whale myoglobin. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *238*(1214), 305–324. https://doi.org/10.1098/rspa.1957.0002

296. Keul, N. D., Oruganty, K., Schaper Bergman, E. T., Beattie, N. R., McDonald, W. E., Kadirvelraj, R., Gross, M. L., Phillips, R. S., Harvey, S. C., & Wood, Z. A. (2018). The entropic force generated by intrinsically disordered segments tunes protein function. *Nature 2018 563:7732*, *563*(7732), 584–588. https://doi.org/10.1038/s41586-018-0699-5

297. Kilic, S., Lezaja, A., Gatti, M., Bianco, E., Michelena, J., Imhof, R., & Altmeyer, M. (2019). Phase separation of 53 BP 1 determines liquid-like behavior of DNA repair compartments . *The EMBO Journal*, *38*(16), e101379. https://doi.org/10.15252/embj.2018101379

298. Kim, J.-E., McAvoy, S. A., Smith, D. I., & Chen, J. (2005). Human TopBP1 Ensures Genome Integrity during Normal S Phase. *Molecular and Cellular Biology*, *25*(24), 10907–10915. https://doi.org/10.1128/mcb.25.24.10907-10915.2005

299. Kim, S. Y., & Ferrell, J. E. (2007). Substrate Competition as a Source of Ultrasensitivity in the Inactivation of Wee1. *Cell*, *128*(6), 1133–1145. https://doi.org/10.1016/j.cell.2007.01.039

300. Kinoshita, N., Ohkura, H., & Yanagida, M. (1990). Distinct, Essential Roles of Type 1 and 2A Protein Phosphatases in the Control of the Fission Yeast Cell Division Cycle. *Cell*, *63*, 405–415.

301. Kinoshita, N., Yamano, H., Niwa, H., Yoshida, T., & Yanagida, M. (1993). Negative regulation of mitosis by the fission yeast protein phosphatase ppa2. *Genes and Development*, *7*(6), 1059–1071. https://doi.org/10.1101/gad.7.6.1059

302. Kitzmann, M., Vandromme, M., Schaeffer, V., Carnac, G., Labbé, J.-C., Lamb, N., & Fernandez, A. (1999). cdk1- and cdk2-Mediated Phosphorylation of MyoD Ser200 in Growing C2 Myoblasts: Role in Modulating MyoD Half-Life and Myogenic Activity. *Molecular and Cellular Biology*, *19*(4), 3167–3176. https://doi.org/10.1128/mcb.19.4.3167

303. Klein, U. R., Haindl, M., Nigg, E. A., & Muller, S. (2009). RanBP2 and SENP3 function in a mitotic SUMO2/3 conjugation-deconjugation cycle on borealin. *Molecular Biology of the Cell*, *20*(1), 410–418. https://doi.org/10.1091/mbc.E08-05-0511

304. Knuesel, M. T., Meyer, K. D., Bernecky, C., & Taatjes, D. J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes and Development*, *23*(4), 439–451. https://doi.org/10.1101/gad.1767009

305. Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J. P., Gannon, J., & Hunt, T. (1992). Identification of the domains in cyclin A required for binding to, and activation of, p34cdc2 and p32cdk2 protein kinase subunits. *Molecular Biology of the Cell*, *3*(11), 1279–1294. https://doi.org/10.1091/mbc.3.11.1279

306. Koch, A., Krug, K., Pengelley, S., Macek, B., & Hauf, S. (2011). Mitotic substrates of the kinase aurora with roles in chromatin regulation identified through quantitative phosphoproteomics of fission yeast. *Science Signaling*, *4*(179). https://doi.org/10.1126/scisignal.2001588

307. Koehler, M. F. T., Bergeron, P., Blackwood, E. M., Bowman, K., Clark, K. R., Firestein, R., Kiefer, J. R., Maskos, K., McCleland, M. L., Orren, L., Salphati, L., Schmidt, S., Schneider, E. V., Wu, J., & Beresini, M. H. (2016). Development of a Potent, Specific CDK8 Kinase Inhibitor Which Phenocopies CDK8/19 Knockout Cells. *ACS Medicinal Chemistry Letters*, *7*(3), 223–228. https://doi.org/10.1021/acsmedchemlett.5b00278

308. Koschubs, T., Seizl, M., Larivière, L., Kurth, F., Baumli, S., Martin, D. E., & Cramer, P. (2009). Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *The EMBO Journal*, *28*(1), 69–80. https://doi.org/10.1038/EMBOJ.2008.254

309. Kozar, K., Ciemerych, M. A., Rebel, V. I., Shigematsu, H., Zagozdzon, A., Sicinska, E., Geng, Y., Yu, Q., Bhattacharya, S., Bronson, R. T., Akashi, K., & Sicinski, P. (2004). Mouse development and cell proliferation in the absence of D-cyclins. *Cell*, *118*(4), 477–491. https://doi.org/10.1016/j.cell.2004.07.025

310. Kraft, C., Herzog, F., Gieffers, C., Mechtler, K., Hagting, A., Pines, J., & Peters, J. M. (2003). Mitotic regulation of the human anaphase-promoting complex by phosphorylation. *EMBO Journal*, *22*(24), 6598–6609. https://doi.org/10.1093/emboj/cdg627

311. Kramer, E. R., Scheuringer, N., Podtelejnikov, A. V., Mann, M., & Peters, J. M. (2000). Mitotic regulation of the APC activator proteins CDC20 and CDH1. *Molecular Biology of the Cell*, *11*(5), 1555–1569. https://doi.org/10.1091/mbc.11.5.1555

312. Krasinska, L., Domingo-Sananes, M. R., Kapuy, O., Parisis, N., Harker, B., Moorhead, G., Rossignol, M., Novák, B., & Fisher, D. (2011). Protein Phosphatase 2A Controls the Order and Dynamics of Cell-Cycle Transitions. *Molecular Cell*, *44*(3), 437–450. https://doi.org/10.1016/J.MOLCEL.2011.10.007

313. Kreitz, S., Ritzi, M., Baack, M., & Knippers, R. (2001). The Human Origin Recognition Complex Protein 1 Dissociates from Chromatin during S Phase in HeLa Cells. *Journal of Biological Chemistry*, *276*(9), 6337–6342. https://doi.org/10.1074/jbc.M009473200

314. Krüger, T., Zentgraf, H., & Scheer, U. (2007). Intranucleolar sites of ribosome biogenesis defined by the localization of early binding ribosomal proteins. *Journal of Cell Biology*, *177*(4), 573–578. https://doi.org/10.1083/JCB.200612048

315. Kuchin, S., Yeghiayan, P., & Carlson, M. (1995). Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(9), 4006–4010. https://doi.org/10.1073/pnas.92.9.4006

316. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Maayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(1), W90–W97. https://doi.org/10.1093/nar/gkw377

317. Kumada, K., Nakamura, T., Nagao, K., Funabiki, H., Nakagawa, T., & Yanagida, M. (1998). Cut1 is loaded onto the spindle by binding to Cut2 and promotes anaphase spindle movement upon Cut2 proteolysis. *Current Biology*, *8*(11), 633–641. https://doi.org/10.1016/S0960-9822(98)70250-7

318. Kumagai, A., & Dunphy, W. G. (1991). The cdc25 protein controls tyrosine dephosphorylation of the cdc2 protein in a cell-free system. *Cell*, *64*(5), 903–914. https://doi.org/10.1016/0092-8674(91)90315-P

319. Kumagai, A., Shevchenko, A., Shevchenko, A., & Dunphy, W. G. (2011). Direct regulation of Treslin by cyclin-dependent kinase is essential for the onset of DNA replication. *Journal of Cell Biology*, *193*(6), 995–1007. https://doi.org/10.1083/JCB.201102003

320. Kumagai, A., Shevchenko, A., Shevchenko, A., & Dunphy, W. G. (2010). Treslin Collaborates with TopBP1 in Triggering the Initiation of DNA Replication. *Cell*, *140*(3), 349–359. https://doi.org/10.1016/J.CELL.2009.12.049/ATTACHMENT/3600B096-5DA0-4053-AF07-EB1D784336B1/MMC1.PDF

321. Labbe, J. C., Capony, J. P., Caput, D., Cavadore, J. C., Derancourt, J., Kaghad, M., Lelias, J. M., Picard, A., & Doree, M. (1989). MPF from starfish oocytes at first meiotic metaphase is a heterodimer containing one molecule of cdc2 and one molecule of cyclin B. *EMBO Journal*, *8*(10), 3053–3058. https://doi.org/10.1002/j.1460-2075.1989.tb08456.x

322. Labbe, J. C., Lee, M. G., Nurse, P., Picard, A., & Doree, M. (1988). Activation at M-phase of a protein kinase encoded by a starfish homologue of the cell cycle control gene cdc2+. *Nature*, *335*(6187), 251–254. https://doi.org/10.1038/335251a0

323. Labib, K. (2010). Howdo Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes and Development*, *24*(12), 1208–1219. https://doi.org/10.1101/gad.1933010

324. Labokha, A. A., Gradmann, S., Frey, S., Hülsmann, B. B., Urlaub, H., Baldus, M., & Görlich, D. (2013). Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. *EMBO Journal*, *32*(2), 204–218. https://doi.org/10.1038/emboj.2012.302

325. Ladouceur, A. M., Parmar, B. S., Biedzinski, S., Wall, J., Tope, S. G., Cohn, D., Kim, A., Soubry, N., Reyes-Lamothe, R., & Weber, S. C. (2020). Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid-liquid phase separation. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(31), 18540–18549. https://doi.org/10.1073/pnas.2005019117

326. Lallemand-Breitenbach, V., & de Thé, H. (2018). PML nuclear bodies: from architecture to function. *Current Opinion in Cell Biology*, *52*, 154–161. https://doi.org/10.1016/j.ceb.2018.03.011

327. Lallemand-Breitenbach, V., & de Thé, H. (2010). PML nuclear bodies. *Cold Spring Harbor Perspectives in Biology*, *2*(5). https://doi.org/10.1101/cshperspect.a000661

328. Landry, C. R., Levy, E. D., & Michnick, S. W. (2009). Weak functional constraints on phosphoproteomes. In *Trends in Genetics* (Vol. 25, Issue 5, pp. 193–197). https://doi.org/10.1016/j.tig.2009.03.003

329. Lane, N., & Martin, W. (2010). The energetics of genome complexity. *Nature*, *467*(7318), 929–934. https://doi.org/10.1038/nature09486

330. Lång, A., Lång, E., & Bøe, S. O. (2019). PML Bodies in Mitosis. *Cells*, *8*(8), 893. https://doi.org/10.3390/cells8080893

331. Lang, M., Jegou, T., Chung, I., Richter, K., Münch, S., Udvarhelyi, A., Cremer, C., Hemmerich, P., Engelhardt, J., Hell, S. W., & Rippe, K. (2010). Three-dimensional organization of promyelocytic

leukemia nuclear bodies. *Journal of Cell Science*, *123*(3), 392–400. https://doi.org/10.1242/jcs.053496

332. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

333. Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, *547*(7662), 236–240. https://doi.org/10.1038/nature22822

334. Lau, H. W., Ma, H. T., Yeung, T. K., Tam, M. Y., Zheng, D., Chu, S. K., & Poon, R. Y. C. (2021). Quantitative differences between cyclin-dependent kinases underlie the unique functions of CDK1 in human cells. *Cell Reports*, *37*(2). https://doi.org/10.1016/j.celrep.2021.109808

335. Laurell, E., Beck, K., Krupina, K., Theerthagiri, G., Bodenmiller, B., Horvath, P., Aebersold, R., Antonin, W., & Kutay, U. (2011). Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. *Cell*, *144*(4), 539–550. https://doi.org/10.1016/j.cell.2011.01.012

336. Lawrence, R. T., Searle, B. C., Llovet, A., & Villén, J. (2016). Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nature Methods 2016 13:5*, *13*(5), 431–434. https://doi.org/10.1038/NMETH.3811

337. Lee, M. G., & Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature*, *327*(6117), 31–35. https://doi.org/10.1038/327031a0

338. Lee, T. H., Solomon, M. J., Mumby, M. C., & Kirschner, M. W. (1991). INH, a Negative Regulator of MPF, Is a Form of Protein Phosphatase 2A. *Cell*, *64*, 415–423.

339. Lei, M., Kawasaki, Y., Young, M. R., Kihara, M., Sugino, A., & Tye, B. K. (1997). Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis. *Genes and Development*, *11*(24), 3365–3374. https://doi.org/10.1101/gad.11.24.3365

340. Lénárt, P., Rabut, G., Daigle, N., Hand, A. R., Terasaki, M., & Ellenberg, J. (2003). Nuclear envelope breakdown in starfish oocytes proceeds by partial NPC disassembly followed by a rapidly spreading fenestration of nuclear membranes. *Journal of Cell Biology*, *160*(7), 1055–1068. https://doi.org/10.1083/jcb.200211076

341. Lewis, C. D., & Laemmli, U. K. (1982). Higher order metaphase chromosome structure: Evidence for metalloprotein interactions. *Cell*, *29*(1), 171–181. https://doi.org/10.1016/0092-8674(82)90101-5

342. Li, C. H., Coffey, E. L., Dall'Agnese, A., Hannett, N. M., Tang, X., Henninger, J. E., Platt, J. M., Oksuz, O., Zamudio, A. V., Afeyan, L. K., Schuijers, J., Liu, X. S., Markoulaki, S., Lungjangwa, T., LeRoy, G., Svoboda, D. S., Wogram, E., Lee, T. I., Jaenisch, R., & Young, R. A. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature*, *586*(7829), 440–444. https://doi.org/10.1038/s41586-020-2574-4

343. Li, N., Fassl, A., Chick, J., Inuzuka, H., Li, X., Mansour, M. R., Liu, L., Wang, H., King, B., Shaik, S., Gutierrez, A., Ordureau, A., Otto, T., Kreslavsky, T., Baitsch, L., Bury, L., Meyer, C. A., Ke, N., Mulry, K. A., … Sicinski, P. (2014). Cyclin C is a haploinsufficient tumour suppressor. *Nature Cell Biology*, *16*(11), 1080–1091. https://doi.org/10.1038/ncb3046

344. Liang, C., & Stillman, B. (1997). Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes and Development*, *11*(24), 3375–3386. https://doi.org/10.1101/gad.11.24.3375

345. Liao, S. M., Zhang, J., Jeffery, D. A., Koleske, A. J., Thompson, C. M., Chao, D. M., Viljoen, M., van Vuuren, H. J. J., & Young, R. A. (1995). A kinase–cyclin pair in the RNA polymerase II holoenzyme. *Nature*, *374*(6518), 193–196. https://doi.org/10.1038/374193a0

346. Lin, D. H., & Hoelz, A. (2019). The structure of the nuclear pore complex (An Update). *Annual Review of Biochemistry*, *88*, 725–783. https://doi.org/10.1146/annurev-biochem-062917-011901

347. Lindeboom, R. G. H., Smits, A. H., Perino, M., Veenstra, G. J. C., & Vermeulen, M. (2019). Mass spectrometry-based absolute quantification of single Xenopus embryo proteomes. *Cold Spring Harbor Protocols*, *2019*(6), 479–485. https://doi.org/10.1101/pdb.prot098376

348. Linder, M. I., Köhler, M., Boersema, P., Weberruss, M., Wandke, C., Marino, J., Ashiono, C., Picotti, P., Antonin, W., & Kutay, U. (2017). Mitotic Disassembly of Nuclear Pore Complexes Involves CDK1- and PLK1-Mediated Phosphorylation of Key Interconnecting Nucleoporins. *Developmental Cell*, *43*(2), 141-156.e7. https://doi.org/10.1016/j.devcel.2017.08.020

349. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction: Implications for structural proteomics. *Structure*, *11*(11), 1453–1459. https://doi.org/10.1016/j.str.2003.10.002

350. Liu, D., Vader, G., Vromans, M. J. M., Lampson, M. A., & Lens, S. M. A. (2009). Sensing chromosome bi-orientation by spatial separation of Aurora B kinase from kinetochore substrates. *Science*, *323*(5919), 1350–1353. https://doi.org/10.1126/science.1167000

351. Liu, H., Rankin, S., & Yu, H. (2013). Phosphorylation-enabled binding of SGO1-PP2A to cohesin protects sororin and centromeric cohesion during mitosis. *Nature Cell Biology*, *15*(1), 40–49. https://doi.org/10.1038/ncb2637

352. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry*, *45*(22), 6873–6888. https://doi.org/10.1021/bi0602718

353. Liu, J., Song, Y., Tian, B., Qian, J., Dong, Y., Liu, J., Liu, B., & Sun, Z. (2010). Functional proteomic analysis of promyelocytic leukaemia nuclear bodies in irradiation-induced MCF-7 cells. *Journal of Biochemistry*, *148*(6), 659–667. https://doi.org/10.1093/jb/mvq105

354. Liu, J. L., Hebert, M. D., Ye, Y., Templeton, D. J., Kung, H. J., & Matera, A. G. (2000). Cell cycle-dependent localization of the CDK2-cyclin E complex in Cajal (coiled) bodies. *Journal of Cell Science*, *113*(9), 1543–1552. https://doi.org/10.1242/jcs.113.9.1543

355. Liu, P., Jenkins, N. A., & Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Research*, *13*(3), 476–484. https://doi.org/10.1101/gr.749203

356. Liu, X., Liu, X., Wang, H., Dou, Z., Ruan, K., Hill, D. L., Li, L., Shi, Y., & Yao, X. (2020). Phase separation drives decision making in cell division. *Journal of Biological Chemistry*, *295*(39), 13419–13431. https://doi.org/10.1074/jbc.REV120.011746

357. Liu, Y., Wang, X., & Liu, B. (2019). A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. In *Briefings in Bioinformatics* (Vol. 20, Issue 1). https://doi.org/10.1093/bib/bbx126

358. Lohka, M. J., Hayes, M. K., & Maller, J. L. (1988). Purification of maturation-promoting factor, an intracellular regulator of early mitotic events. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(9), 3009–3013. https://doi.org/10.1073/PNAS.85.9.3009

359. Lombard-Banek, C., Moody, S. A., Manzini, M. C., & Nemes, P. (2019). Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-Cell Proteomics in Complex Tissues: Developing Cell Clones in Live Xenopus laevis and Zebrafish Embryos. *Analytical Chemistry*, *91*(7), 4797–4805. https://doi.org/10.1021/ACS.ANALCHEM.9B00345/SUPPL_FILE/AC9B00345_SI_002.XLSX

360. Loncle, N., Boube, M., Joulia, L., Boschiero, C., Werner, M., Cribbs, D. L., & Bourbon, H. M. (2007). Distinct roles for Mediator Cdk8 module subunits in Drosophila development. *EMBO Journal*, *26*(4), 1045–1054. https://doi.org/10.1038/sj.emboj.7601566

361. Loog, M., & Morgan, D. O. (2005). Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, *434*(7029), 104–108. https://doi.org/10.1038/nature03329

362. Lörincz, A. T., & Reed, S. I. (1984). Primary structure homology between the product of yeast cell division control gene CDC28 and vertebrate oncogenes. *Nature*, *307*(5947), 183–185. https://doi.org/10.1038/307183a0

363. Lowe, M., Rabouille, C., Nakamura, N., Watson, R., Jackman, M., Jämsä, E., Rahman, D., Pappin, D. J. C., & Warrenz, G. (1998). Cdc2 kinase directly phosphorylates the cis-Golgi matrix protein GM130 and is required for Golgi fragmentation in mitosis. *Cell*, *94*(6), 783–793. https://doi.org/10.1016/S0092-8674(00)81737-7

364. Lu, D., Hsiao, J. Y., Davey, N. E., van Voorhis, V. A., Foster, S. A., Tang, C., & Morgan, D. O. (2014). Multiple mechanisms determine the order of APC/C substrate degradation in mitosis. *Journal of Cell Biology*, *207*(1), 23–39. https://doi.org/10.1083/jcb.201402041

365. 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., & Bohr, V. A. (2017). Cell cycle-dependent phosphorylation regulates RECQL4 pathway choice and ubiquitination in DNA double-strand break repair. *Nature Communications*, *8*(1), 1–14. https://doi.org/10.1038/s41467-017-02146-3

366. Lucena, R., Alcaide-Gavilán, M., Anastasia, S. D., & Kellogg, D. R. (2017). Wee1 and Cdc25 are controlled by conserved PP2A-dependent mechanisms in fission yeast. *Cell Cycle*, *16*(5), 428–435. https://doi.org/10.1080/15384101.2017.1281476

367. Ludlow, J. W., Glendening, C. L., Livingston, D. M., Decaprio, J. A., Ludlow, J. W., Shon, J., Pipas, J. M., Livingston, D. M., & Decaprio, J. A. (1993). Specific enzymatic dephosphorylation of the retinoblastoma protein. *Molecular and Cellular Biology*, *13*(1), 367–372. https://doi.org/10.1128/MCB.13.1.367-372.1993

368. Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R. S., Grøhfte, M., Chan, K. L., Hickson, I. D., Bartek, J., & Lukas, J. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nature Cell Biology*, *13*(3), 243–253. https://doi.org/10.1038/ncb2201

369. Lundgren, K., Walworth, N., Booher, R., Dembski, M., Kirschner, M., & Beach, D. (1991). Mik1 and Wee1 Cooperate in the Inhibitory Tyrosine Phosphorylation of Cdc2. *Cell*, *64*(6), 1111–1122. https://doi.org/10.1016/0092-8674(91)90266-2

370. Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., Stratton, M. R., & Lamond, A. I. (2014). A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *ELife*, *2014*(3). https://doi.org/10.7554/ELIFE.01630

371. Ly, T., Endo, A., & Lamond, A. I. (2015). Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *ELife*, *2015*(4). https://doi.org/10.7554/ELIFE.04534

372. Ly, T., Whigham, A., Clarke, R., Brenes-Murillo, A. J., Estes, B., Madhessian, D., Lundberg, E., Wadsworth, P., & Lamond, A. I. (2017). Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *ELife*, *6*. https://doi.org/10.7554/ELIFE.27574

373. Lynch, C. J., Bernad, R., Martínez-Val, A., Shahbazi, M. N., Nóbrega-Pereira, S., Calvo, I., Blanco-Aparicio, C., Tarantino, C., Garreta, E., Richart-Ginés, L., Alcazar, N., Graña-Castro, O., Gómez-Lopez, G., Aksoy, I., Muñoz-Martín, M., Martinez, S., Ortega, S., Prieto, S., Simboeck, E., … Serrano, M. (2020). Global hyperactivation of enhancers stabilizes human and mouse naive pluripotency through inhibition of CDK8/19 Mediator kinases. *Nature Cell Biology 2020 22:10*, *22*(10), 1223–1238. https://doi.org/10.1038/s41556-020-0573-1

374. Macaulay, C., Meier, E., & Forbes, D. J. (1995). Differential Mitotic Phosphorylation of Proteins of the Nuclear Pore Complex (∗). *Journal of Biological Chemistry*, *270*(1), 254–262. https://doi.org/10.1074/JBC.270.1.254

375. Machyna, M., Heyn, P., & Neugebauer, K. M. (2013). Cajal bodies: Where form meets function. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 17–34. https://doi.org/10.1002/wrna.1139

376. Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics*, *21*(16), 3448–3449. https://doi.org/10.1093/bioinformatics/bti551

377. Maeshima, K., & Laemmli, U. K. (2003). A Two-step scaffolding model for mitotic chromosome assembly. *Developmental Cell*, *4*(4), 467–480. https://doi.org/10.1016/S1534-5807(03)00092-3

378. Malik, S., & Roeder, R. G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics*, *11*(11), 761–772. https://doi.org/10.1038/nrg2901

379. Malumbres, M. (2014). Cyclin-dependent kinases. *Genome Biology*, *15*(6), 1–10. https://doi.org/10.1186/GB4184/FIGURES/4

380. Malumbres, M., Sotillo, R., Santamaría, D., Galán, J., Cerezo, A., Ortega, S., Dubus, P., & Barbacid, M. (2004). Mammalian Cells Cycle without the D-Type Cyclin-Dependent Kinases Cdk4 and Cdk6. *Cell*, *118*(4), 493–504. https://doi.org/10.1016/J.CELL.2004.08.002

381. Manke, I. A., Lowery, D. M., Nguyen, A., & Yaffe, M. B. (2003). BRCT Repeats As Phosphopeptide-Binding Modules Involved in Protein Targeting. *Science*, *302*(5645), 636–639. https://doi.org/10.1126/science.1088877

382. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science*, *298*(5600), 1912–1934. https://doi.org/10.1126/science.1075762

383. Margolis, S. S., Perry, J. A., Weitzel, D. H., Freel, C. D., Yoshida, M., Haystead, T. A., & Kornbluth, S. (2006). A role for PP1 in the Cdc2/cyclin B-mediated positive feedback activation of Cdc25. *Molecular Biology of the Cell*, *17*(4), 1779–1789. https://doi.org/10.1091/MBC.E05-08-0751/ASSET/IMAGES/LARGE/ZMK0040676040009.JPEG

384. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., & Wüthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *European Journal of Biochemistry*, *256*(1), 1–15. https://doi.org/10.1046/j.1432-1327.1998.2560001.x

385. Martín-Castellanos, C., Blanco, M. A., De Prada, J. M., & Moreno, S. (2000). The puc1 cyclin regulates the G1 phase of the fission yeast cell cycle in response to cell size. *Molecular Biology of the Cell*, *11*(2), 543–554. https://doi.org/10.1091/mbc.11.2.543

386. Martinsson, H. S., Starborg, M., Erlandsson, F., & Zetterberg, A. (2005). Single cell analysis of G1 check points - The relationship between the restriction point and phosphorylation of pRb. *Experimental Cell Research*, *305*(2), 383–391. https://doi.org/10.1016/j.yexcr.2005.01.023

387. Masui, Y., & Markert, C. L. (1971). Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. *Journal of Experimental Zoology*, *177*(2), 129–145. https://doi.org/10.1002/jez.1401770202

388. Maxwell, B. A., Gwon, Y., Mishra, A., Peng, J., Nakamura, H., Zhang, K., Kim, H. J., & Taylor, J. P. (2021). Ubiquitination is essential for recovery of cellular activities after heat shock. *Science*, *372*(6549), 2021.04.22.440934. https://doi.org/10.1126/science.abc3593

389. Mazumdar, M., Sundareshan, S., & Misteli, T. (2004). Human chromokinesin KIF4A functions in chromosome condensation and segregation. *Journal of Cell Biology*, *166*(5), 613–620. https://doi.org/10.1083/jcb.200401142

390. McDermott, M. S. J., Chumanevich, A. A., Lim, C. U., Liang, J., Chen, M., Altilia, S., Oliver, D., Rae, J. M., Shtutman, M., Kiaris, H., Gyorffy, B., Roninson, I. B., & Broude, E. V. (2017). Inhibition of CDK8 mediator kinase suppresses estrogen dependent transcription and the growth of estrogen receptor positive breast cancer. *Oncotarget*, *8*(8), 12558–12575. https://doi.org/10.18632/oncotarget.14894

391. McElhinny, S. A. N., Gordenin, D. A., Stith, C. M., Burgers, P. M. J., & Kunkel, T. A. (2008). Division of Labor at the Eukaryotic Replication Fork. *Molecular Cell*, *30*(2), 137–144. https://doi.org/10.1016/j.molcel.2008.02.022

392. McGowan, C. H., & Russell, P. (1993). Human Wee1 kinase inhibits cell division by phosphorylating p34(cdc2) exclusively on Tyr15. *EMBO Journal*, *12*(1), 75–85. https://doi.org/10.1002/j.1460-2075.1993.tb05633.x

393. McSwiggen, D. T., Mir, M., Darzacq, X., & Tjian, R. (2019). Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. In *Genes & development* (Vol. 33, Issues 23–24, pp. 1619–1634). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.331520.119

394. Meiring, H. D., Van Der Heeft, E., Ten Hove, G. J., & De Jong, A. P. J. M. (2002). Nanoscale LC-MS(n): Technical design and applications to peptide and protein analysis. *Journal of Separation Science*, *25*(9), 557–568. https://doi.org/10.1002/1615-9314(20020601)25:9<557::AID-JSSC557>3.0.CO;2-F

395. Méndez, J., Zou-Yang, X. H., Kim, S. Y., Hidaka, M., Tansey, W. P., & Stillman, B. (2002). Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Molecular Cell*, *9*(3), 481–491. https://doi.org/10.1016/S1097-2765(02)00467-7

396. Meng, F., Uversky, V. N., & Kurgan, L. (2017). Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. In *Cellular and Molecular Life Sciences* (Vol. 74, Issue 17, pp. 3069–3090). https://doi.org/10.1007/s00018-017-2555-4

397. Menzl, I., Witalisz-Siepracka, A., & Sexl, V. (2019). CDK8-novel therapeutic opportunities. *Pharmaceuticals*, *12*(2), 92. https://doi.org/10.3390/ph12020092

398. Merrick, K. A., Larochelle, S., Zhang, C., Allen, J. J., Shokat, K. M., & Fisher, R. P. (2008). Distinct Activation Pathways Confer Cyclin-Binding Specificity on Cdk1 and Cdk2 in Human Cells. *Molecular Cell*, *32*(5), 662–672. https://doi.org/10.1016/j.molcel.2008.10.022

399. Merrick, K. A., Wohlbold, L., Zhang, C., Allen, J. J., Horiuchi, D., Huskey, N. E., Goga, A., Shokat, K. M., & Fisher, R. P. (2011). Switching Cdk2 On or Off with Small Molecules to Reveal Requirements in Human Cell Proliferation. *Molecular Cell*, *42*(5), 624–636. https://doi.org/10.1016/J.MOLCEL.2011.03.031

400. Meyerson, M., Enders, G. H., Wu, C. L., Su, L. K., Gorka, C., Nelson, C., Harlow, E., & Tsai, L. H. (1992). A family of human cdc2-related protein kinases. *EMBO Journal*, *11*(8), 2909–2917. https://doi.org/10.1002/j.1460-2075.1992.tb05360.x

401. Michowski, W., Chick, J. M., Chu, C., Kolodziejczyk, A., Wang, Y., Suski, J. M., Abraham, B., Anders, L., Day, D., Dunkl, L. M., Li Cheong Man, M., Zhang, T., Laphanuwat, P., Bacon, N. A., Liu, L., Fassl, A., Sharma, S., Otto, T., Jecrois, E., … Sicinski, P. (2020). Cdk1 Controls Global Epigenetic Landscape in Embryonic Stem Cells. *Molecular Cell*, *78*(3), 459-476.e13. https://doi.org/10.1016/j.molcel.2020.03.010

402. Miller, S. L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science*, *117*(3046), 528–529. https://doi.org/10.1126/SCIENCE.117.3046.528

403. Milner, R. E., Busaan, J. L., Holmes, C. F. B., Wang, J. H., & Michalak, M. (1993). Phosphorylation of dystrophin. The carboxyl-terminal region of dystrophin is a substrate for in vitro phosphorylation by p34(cdc2) protein kinase. *Journal of Biological Chemistry*, *268*(29), 21901–21905. https://doi.org/10.1016/s0021-9258(20)80626-7

404. Min, J., Wright, W. E., & Shay, J. W. (2019). Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52. *Genes and Development*, *33*(13–14), 814–827. https://doi.org/10.1101/gad.324905.119

405. Minshull, J., Golsteyn, R., Hill, C. S., & Hunt, T. (1990). The A- and B-type cyclin associated cdc2 kinases in Xenopus turn on and off at different times in the cell cycle. *The EMBO Journal*, *9*(9), 2865–2875. https://doi.org/10.1002/J.1460-2075.1990.TB07476.X

406. Minton, A. P. (2000). Implications of macromolecular crowding for protein assembly. In *Current Opinion in Structural Biology* (Vol. 10, Issue 1, pp. 34–39). https://doi.org/10.1016/S0959-440X(99)00045-7

407. Misteli, T. (2001). The concept of self-organization in cellular architecture. *Journal of Cell Biology*, *155*(2), 181–185. https://doi.org/10.1083/jcb.200108110

408. Mochida, S., Maslen, S. L., Skehel, M., & Hunt, T. (2010). Greatwall phosphorylates an inhibitor of protein phosphatase 2A that is essential for mitosis. *Science*, *330*(6011), 1670–1673. https://doi.org/10.1126/science.1195689

409. Mochida, S., Rata, S., Hino, H., Nagai, T., & Novák, B. (2016). Two Bistable Switches Govern M Phase Entry. *Current Biology*, *26*(24), 3361–3367. https://doi.org/10.1016/j.cub.2016.10.022

410. Monterroso, B., Zorrilla, S., Sobrinos-Sanguino, M., Robles-Ramos, M. A., López-Álvarez, M., Margolin, W., Keating, C. D., & Rivas, G. (2019). Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA. *EMBO Reports*, *20*(1), e45946. https://doi.org/10.15252/EMBR.201845946

411. Moreno, S., Hayles, J., & Nurse, P. (1989). Regulation of p34cdc2 protein kinase during mitosis. *Cell*, *58*(2), 361–372. https://doi.org/10.1016/0092-8674(89)90850-7

412. Morgan, D. O. (1997). CYCLIN-DEPENDENT KINASES: Engines, Clocks, and Microprocessors. *Annu. Rev. Cell Dev. Biol*, *13*, 261–291. www.annualreviews.org

413. Morgan, D. O. (2007). *The Cell Cycle, Principles of Control.* (1st ed.). New Science Press Ltd; ; Oxford University Press. https://morganlab.ucsf.edu/sites/g/files/tkssra2561/f/wysiwyg/Morgan Cell Cycle Book.pdf

414. Morgan, D. O. (2016). Cell division: Mitotic regulation comes into focus. *Nature*, *536*(7617), 407–408. https://doi.org/10.1038/nature19423

415. Morris, E. J., Ji, J. Y., Yang, F., Di Stefano, L., Herr, A., Moon, N. S., Kwon, E. J., Haigis, K. M., Näär, A. M., & Dyson, N. J. (2008). E2F1 represses β-catenin transcription and is antagonized by both pRB and CDK8. *Nature*, *455*(7212), 552–556. https://doi.org/10.1038/nature07310

416. Moses, A. M., Hériché, J. K., & Durbin, R. (2007). Clustering of phosphorylation site recognition motifs can be exploited to predict the targets of cyclin-dependent kinase. *Genome Biology*, *8*(2), 1–14. https://doi.org/10.1186/gb-2007-8-2-r23

417. Mrouj, K., Andrés-Sánchez, N., Dubra, G., Singh, P., Sobecki, M., Chahar, D., Al Ghoul, E., Aznar, A. B., Prieto, S., Pirot, N., Bernex, F., Bordignon, B., Hassen-Khodja, C., Villalba, M., Krasinska, L., & Fisher, D. (2021). Ki-67 regulates global gene expression and promotes sequential stages of carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10). https://doi.org/10.1073/pnas.2026507118

418. Mueller, P. R., Coleman, T. R., Kumagai, A., & Dunphy, W. G. (1995). Myt1: A membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15. *Science*, *270*(5233), 86–90. https://doi.org/10.1126/science.270.5233.86

419. Mühlhäusser, P., & Kutay, U. (2007). An in vitro nuclear disassembly system reveals a role for the RanGTPase system and microtubule-dependent steps in nuclear envelope breakdown. *Journal of Cell Biology*, *178*(4), 595–610. https://doi.org/10.1083/jcb.200703002

420. Müller, S., Matunis, M. J., & Dejean, A. (1998). Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO Journal*, *17*(1), 61–70. https://doi.org/10.1093/emboj/17.1.61

421. Murray, D. T., Kato, M., Lin, Y., Thurber, K. R., Hung, I., McKnight, S. L., & Tycko, R. (2017). Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*, *171*(3), 615-627.e16. https://doi.org/10.1016/j.cell.2017.08.048

422. Musacchio, A. (2015). The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Current Biology*, *25*(20), R1002–R1018. https://doi.org/10.1016/j.cub.2015.08.051

423. Musacchio, A. (2022). On the role of phase separation in the biogenesis of membraneless compartments. *The EMBO Journal*, *41*(5), e109952. https://doi.org/10.15252/embj.2021109952

424. Muthunayake, N. S., Tomares, D. T., Childers, W. S., & Schrader, J. M. (2020). Phase-separated bacterial ribonucleoprotein bodies organize mRNA decay. In *Wiley Interdisciplinary Reviews: RNA* (Vol. 11, Issue 6, p. e1599). John Wiley & Sons, Ltd. https://doi.org/10.1002/wrna.1599

425. Nakamura, A., Nakata, D., Kakoi, Y., Kunitomo, M., Murai, S., Ebara, S., Hata, A., & Hara, T. (2018). CDK8/19 inhibition induces premature G1/S transition and ATRdependent cell death in prostate cancer cells. *Oncotarget*, *9*(17), 13474–13487. https://doi.org/10.18632/oncotarget.24414

426. Narasimha, A. M., Kaulich, M., Shapiro, G. S., Choi, Y. J., Sicinski, P., & Dowdy, S. F. (2014). Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. *ELife*, *3*. https://doi.org/10.7554/elife.02872

427. Narasumani, M., & Harrison, P. M. (2018). Discerning evolutionary trends in post-translational modification and the effect of intrinsic disorder: Analysis of methylation, acetylation and ubiquitination sites in human proteins. *PLoS Computational Biology*, *14*(8), e1006349. https://doi.org/10.1371/journal.pcbi.1006349

428. Nash, R., Tokiwa, G., Anand, S., Erickson, K., & Futcher, A. B. (1988). The WHI1+ gene of Saccharomyces cerevisiae tethers cell division to cell size and is a cyclin homolog. *The EMBO Journal*, *7*(13), 4335–4346. https://doi.org/10.1002/j.1460-2075.1988.tb03332.x

429. Nasmyth, K. (1995). *Evolution of the cell cycle*. https://royalsocietypublishing.org/

430. Necci, M., Piovesan, D., Dosztanyi, Z., & Tosatto, S. C. E. (2017). MobiDB-lite: Fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*, *33*(9), 1402–1404. https://doi.org/10.1093/bioinformatics/btx015

431. Necci, M., Piovesan, D., Hoque, M. T., Walsh, I., Iqbal, S., Vendruscolo, M., Sormanni, P., Wang, C., Raimondi, D., Sharma, R., Zhou, Y., Litfin, T., Galzitskaya, O. V., Lobanov, M. Y., Vranken, W., Wallner, B., Mirabello, C., Malhis, N., Dosztányi, Z., … Tosatto, S. C. E. (2021). Critical assessment of protein intrinsic disorder prediction. *Nature Methods*, *18*(5), 472–481. https://doi.org/10.1038/s41592-021-01117-3

432. Nemet, J., Jelicic, B., Rubelj, I., & Sopta, M. (2014). The two faces of Cdk8, a positive/negative regulator of transcription. *Biochimie*, *97*(1), 22–27. https://doi.org/10.1016/j.biochi.2013.10.004

433. Newport, J. W., & Kirschner, M. W. (1984). Regulation of the cell cycle during early Xenopus development. *Cell*, *37*(3), 731–742. https://doi.org/10.1016/0092-8674(84)90409-4

434. Newport, J., & Kirschner, M. (1982). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*, *30*(3), 675–686. https://doi.org/10.1016/0092-8674(82)90272-0

435. Nielsen, J. T., & Mulder, F. A. A. (2019). Quality and bias of protein disorder predictors. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41644-w

436. Nishiyama, T., Sykora, M. M., Huis, P. J., Mechtler, K., & Peters, J. M. (2013). Aurora B and Cdk1 mediate Wapl activation and release of acetylated cohesin from chromosomes by phosphorylating Sororin. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), 13404–13409. https://doi.org/10.1073/pnas.1305020110

437. Niwa-Kawakita, M., Ferhi, O., Soilihi, H., Le Bras, M., Lallemand-Breitenbach, V., & de Thé, H. (2017). PML is a ROS sensor activating p53 upon oxidative stress. *Journal of Experimental Medicine*, *214*(11), 3197–3206. https://doi.org/10.1084/jem.20160301

438. Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., & Baldwin, A. J. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*, *57*(5), 936–947. https://doi.org/10.1016/j.molcel.2015.01.013

439. Novak, B., & Tyson, J. J. (1993). Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *Journal of Cell Science*, *106*(4), 1153–1168. https://doi.org/10.1242/jcs.106.4.1153

440. Novak, B., Kapuy, O., Domingo-Sananes, M. R., & Tyson, J. J. (2010). Regulated protein kinases and phosphatases in cell cycle decisions. *Current Opinion in Cell Biology*, *22*(6), 801–808. https://doi.org/10.1016/J.CEB.2010.07.001

441. Novak, B., & Tyson, J. J. (2022). Mitotic kinase oscillation governs the latching of cell cycle switches. *Current Biology*, *32*(12), 2780-2785.e2. https://doi.org/10.1016/J.CUB.2022.04.016/ATTACHMENT/9602B395-3757-4AFF-BBA3-4DF507BD22AF/MMC2.PDF

442. Novák, B., & Tyson, J. J. (2021). Mechanisms of signalling-memory governing progression through the eukaryotic cell cycle. *Current Opinion in Cell Biology*, *69*, 7–16. https://doi.org/10.1016/J.CEB.2020.12.003

443. Nurse, P. (1975). Genetic control of cell size at cell division in yeast. *Nature*, *256*(5518), 547–551. https://doi.org/10.1038/256547a0

444. Nurse, P., & Bissett, Y. (1981). Gene required in G1 for commitment to cell cycle and in G 2 for control of mitosis in fission yeast. *Nature*, *292*(5823), 558–560. https://doi.org/10.1038/292558a0

445. Nurse, P., Thuriaux, P., & Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *MGG Molecular & General Genetics*, *146*(2), 167–178. https://doi.org/10.1007/BF00268085

446. Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztányi, Z., Uversky, V. N., Obradovic, Z., Kurgan, L., Dunker, A. K., & Gough, J. (2013). D2P2: Database of disordered protein predictions. *Nucleic Acids Research*, *41*(D1), D508–D516. https://doi.org/10.1093/nar/gks1226

447. Obara-Ishihara, T., & Okayama, H. (1994). A B-type cyclin negatively regulates conjugation via interacting with cell cycle “start” genes in fission yeast. *EMBO Journal*, *13*(8), 1863–1872. https://doi.org/10.1002/j.1460-2075.1994.tb06455.x

448. Ohta, S., Bukowski-Wills, J. C., Sanchez-Pulido, L., Alves, F. de L., Wood, L., Chen, Z. A., Platani, M., Fischer, L., Hudson, D. F., Ponting, C. P., Fukagawa, T., Earnshaw, W. C., & Rappsilber, J. (2010). The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. *Cell*, *142*(5), 810–821. https://doi.org/10.1016/j.cell.2010.07.047

449. Okamoto, K., & Sagata, N. (2007). Mechanism for inactivation of the mitotic inhibitory kinase Wee1 at M phase. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(10), 3753–3758. https://doi.org/10.1073/PNAS.0607357104

450. Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O. G., Kandrov, D., Rasputin, K., Syabro, M., & Tleukenov, T. (2012). Unipro UGENE: A unified bioinformatics toolkit. In *Bioinformatics* (Vol. 28, Issue 8, pp. 1166–1167). Oxford Academic. https://doi.org/10.1093/bioinformatics/bts091

451. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., & Mann, M. (2010). Quantitative phosphoproteomics revealswidespread full phosphorylation site occupancy during mitosis. *Science Signaling*, *3*(104). https://doi.org/10.1126/SCISIGNAL.2000475/SUPPL_FILE/SUPPLEMENTARYTABLE_S9.XLS

452. Olugbile, S., Kulangara, C., Bang, G., Bertholet, S., Suzarte, E., Villard, V., Frank, G., Audran, R., Razaname, A., Nebie, I., Awobusuyi, O., Spertini, F., Kajava, A. V., Felger, I., Druilhe, P., &

Corradin, G. (2009). Vaccine potentials of an intrinsically unstructured fragment derived from the blood stage-associated Plasmodium falciparum protein PFF0165c. *Infection and Immunity*, *77*(12), 5701–5709. https://doi.org/10.1128/IAI.00652-09

453. Onischenko, E. A., Gubanova, N. V., Kiseleva, E. V., & Hallberg, E. (2005). Cdk1 and okadaic acid-sensitive phosphatases control assembly of nuclear pore complexes in Drosophila embryos. *Molecular Biology of the Cell*, *16*(11), 5152–5162. https://doi.org/10.1091/mbc.E05-07-0642

454. Onuma, K., Ochiai, M., Orihashi, K., Takahashi, M., Imai, T., Nakagama, H., & Hippo, Y. (2013). Genetic reconstitution of tumorigenesis in primary intestinal cells. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(27), 11127–11132. https://doi.org/10.1073/PNAS.1221926110

455. Orthwein, A., Fradet-Turcotte, A., Noordermeer, S. M., Canny, M. D., Brun, C. M., Strecker, J., Escribano-Diaz, C., & Durocher, D. (2014). Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. *Science*, *344*(6180), 189–193. https://doi.org/10.1126/SCIENCE.1248024/SUPPL_FILE/ORTHWEIN.SM.PDF

456. Owen, I., & Shewmaker, F. (2019). The role of post-translational modifications in the phase transitions of intrinsically disordered proteins. *International Journal of Molecular Sciences*, *20*(21), 5501. https://doi.org/10.3390/ijms20215501

457. Pagano, M., Pepperkok, R., Verde, F., Ansorge, W., & Draetta, G. (1992). Cyclin A is required at two points in the human cell cycle. *The EMBO Journal*, *11*(3), 961–971. https://doi.org/10.1002/J.1460-2075.1992.TB05135.X

458. Pagano, M., & Draetta, G. (1991). Cyclin A, cell cycle control and oncogenesis. *Progress in Growth Factor Research*, *3*(4), 267–277. https://doi.org/10.1016/0955-2235(91)90004-N

459. Pagliuca, F. W., Collins, M. O., Lichawska, A., Zegerman, P., Choudhary, J. S., & Pines, J. (2011). Quantitative Proteomics Reveals the Basis for the Biochemical Specificity of the Cell-Cycle Machinery. *Molecular Cell*, *43*(3), 406–417. https://doi.org/10.1016/j.molcel.2011.05.031

460. Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V., & Rosen, M. K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Molecular Cell*, *63*(1), 72–85. https://doi.org/10.1016/j.molcel.2016.05.042

461. Pardee, A. B. (1974). A restriction point for control of normal animal cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(4), 1286–1290. https://doi.org/10.1073/pnas.71.4.1286

462. Parisis, N., Krasinska, L., Harker, B., Urbach, S., Rossignol, M., Camasses, A., Dewar, J., Morin, N., & Fisher, D. (2017). Initiation of DNA replication requires actin dynamics and formin activity . *The EMBO Journal*, *36*(21), 3212–3231. https://doi.org/10.15252/embj.201796585

463. Parker, L. L., & Piwnica-Worms, H. (1992). Inactivation of the p34cdc2-cyclin B complex by the human WEE1 tyrosine kinase. *Science*, *257*(5078), 1955–1957. https://doi.org/10.1126/science.1384126

464. Parker, L. L., Atherton-Fessler, S., & Piwnica-Worms, H. (1992). P107Weel Is a Dual-Specificity Kinase That Phosphorylates P34Cdc2 on Tyrosine 15 (Cell Cycle/Baculovirus Expression). *Cell Biology*, *89*, 2917–2921.

465. Parker, M. W., Botchan, M. R., & Berger, J. M. (2017). Mechanisms and regulation of DNA replication initiation in eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology*, *52*(2), 107–144. https://doi.org/10.1080/10409238.2016.1274717

466. Partscht, P., Uddin, B., & Schiebel, E. (2021). Human cells lacking CDC14A and CDC14B show differences in ciliogenesis but not in mitotic progression. Journal of Cell Science, 134(2). https://doi.org/10.1242/JCS.255950/266820/AM/HUMAN-CELLS-LACKING-CDC14A-AND-CDC14B-SHOW

467. Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., & Alberti, S. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, *162*(5), 1066–1077. https://doi.org/10.1016/j.cell.2015.07.047

468. Pederson, T. (2011). The Nucleolus. *Cold Spring Harbor Perspectives in Biology*, *3*(3), a000638. https://doi.org/10.1101/CSHPERSPECT.A000638

469. Pejaver, V., Hsu, W. L., Xin, F., Dunker, A. K., Uversky, V. N., & Radivojac, P. (2014). The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Science*, *23*(8), 1077–1093. https://doi.org/10.1002/pro.2494

470. Pelish, H. E., Liau, B. B., Nitulescu, I. I., Tangpeerachaikul, A., Poss, Z. C., Da Silva, D. H., Caruso, B. T., Arefolov, A., Fadeyi, O., Christie, A. L., Du, K., Banka, D., Schneider, E. V., Jestel, A., Zou, G., Si, C., Ebmeier, C. C., Bronson, R. T., Krivtsov, A. V., … Shair, M. D. (2015). Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature*, *526*(7572), 273–276. https://doi.org/10.1038/nature14904

471. Peng, Q. Y., & Zhang, Q. F. (2006). Precise positions of Phoebe determined with CCD image-overlapping calibration. *Monthly Notices of the Royal Astronomical Society*, *366*(1), 208–212. https://doi.org/10.1186/1471-2105-7-208

472. Peng, Z., Mizianty, M. J., Xue, B., Kurgan, L., & Uversky, V. N. (2012). More than just tails: Intrinsic disorder in histone proteins. *Molecular BioSystems*, *8*(7), 1886–1901. https://doi.org/10.1039/c2mb25102g

473. Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., Hu, G., Uversky, V. N., & Kurgan, L. (2014). Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences*, *72*(1), 137–151. https://doi.org/10.1007/S00018-014-1661-9/FIGURES/7

474. Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., Hu, G., Uversky, V. N., & Kurgan, L. (2014). Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences*, *72*(1), 137–151. https://doi.org/10.1007/s00018-014-1661-9

475. Perkins, G., Drury, L. S., & Diffley, J. F. X. (2001). Separate SCFCDC4 recognition elements target Cdc6 for proteolysis in S phase and mitosis. *EMBO Journal*, *20*(17), 4836–4845. https://doi.org/10.1093/emboj/20.17.4836

476. Perutz, M. F., Steinrauf, L. K., Stockell, A., & Bangham, A. D. (1959). Chemical and crystallographic study of the two fractions of adult horse haemoglobin. *Journal of Molecular Biology*, *1*(4–5), 402–404. https://doi.org/10.1016/S0022-2836(59)80024-3

477. Peter, M., Nakagawa, J., Dorée, M., Labbé, J. C., & Nigg, E. A. (1990). In vitro disassembly of the nuclear lamina and M phase-specific phosphorylation of lamins by cdc2 kinase. *Cell*, *61*(4), 591–602. https://doi.org/10.1016/0092-8674(90)90471-P

478. Petersen, B. O., Wagener, C., Marinoni, F., Kramer, E. R., Melixetian, M., Denchi, E. L., Gieffers, C., Matteucci, C., Peters, J. M., & Helin, K. (2000). Cell cycle– and cell growth–regulated proteolysis of mammalian CDC6 is dependent on APC–CDH1. *Genes & Development*, *14*(18), 2330–2343. https://doi.org/10.1101/GAD.832500

479. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular and Cellular Proteomics*, *11*(11), 1475–1488. https://doi.org/10.1074/MCP.O112.020131/ATTACHMENT/C97541F6-42EB-4896-AFF1-62DAB7E6D879/MMC1.ZIP

480. Petri, E. T., Errico, A., Escobedo, L., Hunt, T., & Basavappa, R. (2007). The crystal structure of human cyclin B. *Cell Cycle*, *6*(11), 1342–1349. https://doi.org/10.4161/cc.6.11.4297

481. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

482. Peuchen, E. H., Cox, O. F., Sun, L., Hebert, A. S., Coon, J. J., Champion, M. M., Dovichi, N. J., & Huber, P. W. (2017). Phosphorylation Dynamics Dominate the Regulated Proteome during Early Xenopus Development. *Scientific Reports 2017 7:1*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-15936-y

483. Pfleger, C. M., & Kirschner, M. W. (2000). The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes & Development*, *14*(6), 655–665. https://doi.org/10.1101/GAD.14.6.655

484. Piatti, S., Böhm, T., Cocker, J. H., Diffley, J. F. X., & Nasmyth, K. (1996). Activation of S-phase-promoting CDKs in late G1 defines a “point of no return” after which Cdc6 synthesis cannot promote DNA replication in yeast. *Genes and Development*, *10*(12), 1516–1531. https://doi.org/10.1101/gad.10.12.1516

485. Piggott, J. R., Rai, R., & Carter, B. L. A. (1982). A bifunctional gene product involved in two phases of the yeast cell cycle. *Nature*, *298*(5872), 391–393. https://doi.org/10.1038/298391a0

486. Pines, J. (2006). Mitosis: A matter of getting rid of the right protein at the right time. *Trends in Cell Biology*, *16*(1), 55–63. https://doi.org/10.1016/j.tcb.2005.11.006

487. Piovesan, D., Necci, M., Escobedo, N., Monzon, A. M., Hatos, A., Mičetić, I., Quaglia, F., Paladin, L., Ramasamy, P., Dosztányi, Z., Vranken, W. F., Davey, N. E., Parisi, G., Fuxreiter, M., & Tosatto, S. C. E. (2021). MobiDB: Intrinsically disordered proteins in 2021. *Nucleic Acids Research*, *49*(D1), D361–D367. https://doi.org/10.1093/nar/gkaa1058

488. Planas-Silva, M. D., & Weinberg, R. A. (1997). The restriction point and control of cell proliferation. *Current Opinion in Cell Biology*, *9*(6), 768–772. https://doi.org/10.1016/S0955-0674(97)80076-2

489. Platani, M., Goldberg, I., Swedlow, J. R., & Lamond, A. I. (2000). In vivo analysis of Cajal body movement, separation, and joining in live human cells. *Journal of Cell Biology*, *151*(7), 1561–1574. https://doi.org/10.1083/jcb.151.7.1561

490. Pomerening, J. R., Sontag, E. D., & Ferrell, J. E. (2003). Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biology 2003 5:4*, *5*(4), 346–351. https://doi.org/10.1038/ncb954

491. Pomerening, J. R., Sun, Y. K., & Ferrell, J. E. (2005). Systems-level dissection of the cell-cycle oscillator: Bypassing positive feedback produces damped oscillations. *Cell*, *122*(4), 565–578. https://doi.org/10.1016/j.cell.2005.06.016

492. Pomerening, J. R., Ubersax, J. A., & Ferrell, J. E. (2008). Rapid cycling and precocious termination of G1 phase in cells expressing CDK1AF. *Molecular Biology of the Cell*, *19*(8), 3426–3441. https://doi.org/10.1091/mbc.E08-02-0172

493. Porter, D. C., Farmaki, E., Altilia, S., Schools, G. P., West, D. K., Chen, M., Chang, B. D., Puzyrev, A. T., Lim, C. U., Rokow-Kittell, R., Friedhoff, L. T., Papavassiliou, A. G., Kalurupalle, S., Hurteau, G., Shi, J., Baran, P. S., Gyorffy, B., Wentland, M. P., Broude, E. V., … Roninson, I. B. (2012). Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(34), 13799–13804. https://doi.org/10.1073/pnas.1206906109

494. Poss, Z. C., Ebmeier, C. C., Odell, A. T., Tangpeerachaikul, A., Lee, T., Pelish, H. E., Shair, M. D., Dowell, R. D., Old, W. M., & Taatjes, D. J. (2016). Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. *Cell Reports*, *15*(2), 436–450. https://doi.org/10.1016/j.celrep.2016.03.030

495. Post, H., Penning, R., Fitzpatrick, M. A., Garrigues, L. B., Wu, W., Macgillavry, H. D., Hoogenraad, C. C., Heck, A. J. R., & Altelaar, A. F. M. (2017). Robust, Sensitive, and Automated

Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *Journal of Proteome Research*, *16*(2), 728–737. https://doi.org/10.1021/ACS.JPROTEOME.6B00753/SUPPL_FILE/PR6B00753_SI_001.PDF

496. Postlmayr, A., Dumeau, C. E., & Wutz, A. (2020). Cdk8 is required for establishment of H3K27me3 and gene repression by Xist and mouse development. *Development (Cambridge, England)*, *147*(11). https://doi.org/10.1242/dev.175141

497. Presler, M., Van Itallie, E., Klein, A. M., Kunz, R., Coughlin, M. L., Peshkin, L., Gygi, S. P., Wühr, M., & Kirschner, M. W. (2017). Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the Xenopus egg. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(50), E10838–E10847. https://doi.org/10.1073/PNAS.1709207114

498. Prieto, S., Dubra, G., Camasses, A., Simboeck, E., Aznar, A. B., Begon-Pescia, C., Pirot, N., Gerbe, F., Angevin, L., Jay, P., Krasinska, L., & Fisher, D. (n.d.). CDK8 and CDK19 act redundantly to control the CFTR pathway in the intestinal epithelium. *BioRxiv*. https://doi.org/10.1101/2022.01.28.478171

499. Pursell, Z. F., Isoz, I., Lundström, E. B., Johansson, E., & Kunkel, T. A. (2007). Yeast DNA polymerase ε participates in leading-strand DNA replication. *Science*, *317*(5834), 127–130. https://doi.org/10.1126/SCIENCE.1144067/SUPPL_FILE/PURSELL_SOM.PDF

500. Purvis, J. E., & Lahav, G. (2013). Encoding and Decoding Cellular Information through Signaling Dynamics. *Cell*, *152*(5), 945–956. https://doi.org/10.1016/J.CELL.2013.02.005

501. Qiao, R., Weissmann, F., Yamaguchi, M., Brown, N. G., VanderLinden, R., Imre, R., Jarvis, M. A., Brunner, M. R., Davidson, I. F., Litos, G., Haselbach, D., Mechtler, K., Stark, H., Schulman, B. A., & Peters, J. M. (2016). Mechanism of APC/CCDC20 activation by mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(19), E2570–E2578. https://doi.org/10.1073/pnas.1604929113

502. Qie, S., & Diehl, J. A. (2020). Cyclin D degradation by E3 ligases in cancer progression and treatment. *Seminars in Cancer Biology*, *67*, 159–170. https://doi.org/10.1016/J.SEMCANCER.2020.01.012

503. Quaglia, F., Mészáros, B., Salladini, E., Hatos, A., Pancsa, R., Chemes, L. B., Pajkos, M., Lazar, T., Peña-Díaz, S., Santos, J., Ács, V., Farahi, N., Fichó, E., Aspromonte, M. C., Bassot, C., Chasapi, A., Davey, N. E., Davidović, R., Dobson, L., … Piovesan, D. (2022). DisProt in 2022: Improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Research*, *50*(D1), D480–D487. https://doi.org/10.1093/nar/gkab1082

504. Quevedo, M., Meert, L., Dekker, M. R., Dekkers, D. H. W., Brandsma, J. H., van den Berg, D. L. C., Ozgür, Z., IJcken, W. F. J. va., Demmers, J., Fornerod, M., & Poot, R. A. (2019). Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-10502-8

505. Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., Goebl, M. G., & Iakoucheva, L. M. (2010). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Structure, Function and Bioinformatics*, *78*(2), 365–380. https://doi.org/10.1002/PROT.22555

506. Rahmanzadeh, R., Hüttmann, G., Gerdes, J., & Scholzen, T. (2007). Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis. *Cell Proliferation*, *40*(3), 422–430. https://doi.org/10.1111/j.1365-2184.2007.00433.x

507. Rai, A. K., Chen, J. X., Selbach, M., & Pelkmans, L. (2018). Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature*, *559*(7713), 211–216. https://doi.org/10.1038/s41586-018-0279-8

508. Ramamurthy, M., Sankar, S., Abraham, A. M., Nandagopal, B., & Sridharan, G. (2019). B cell epitopes in the intrinsically disordered regions of neuraminidase and hemagglutinin proteins of

H5N1 and H9N2 avian influenza viruses for peptide-based vaccine development. *Journal of Cellular Biochemistry*, *120*(10), 17534–17544. https://doi.org/10.1002/jcb.29017

509. Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/NAR/GKW257

510. Rata, S., Suarez Peredo Rodriguez, M. F., Joseph, S., Peter, N., Echegaray Iturra, F., Yang, F., Madzvamuse, A., Ruppert, J. G., Samejima, K., Platani, M., Alvarez-Fernandez, M., Malumbres, M., Earnshaw, W. C., Novak, B., & Hochegger, H. (2018). Two Interlinked Bistable Switches Govern Mitotic Control in Mammalian Cells. *Current Biology*, *28*(23), 3824-3832.e6. https://doi.org/10.1016/J.CUB.2018.09.059/ATTACHMENT/840D6348-0935-42E5-82DE-B0E48897FBFE/MMC1.PDF

511. Regy, R. M., Thompson, J., Kim, Y. C., & Mittal, J. (2021). Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science*, *30*(7), 1371–1379. https://doi.org/10.1002/pro.4094

512. Remmerie, M., & Janssens, V. (2019). PP2A: A promising biomarker and therapeutic target in endometrial cancer. *Frontiers in Oncology*, *9*(JUN), 462. https://doi.org/10.3389/FONC.2019.00462/BIBTEX

513. Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., & Diffley, J. F. X. (2009). Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell*, *139*(4), 719–730. https://doi.org/10.1016/j.cell.2009.10.015

514. Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R. A., & Dynlacht, B. D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G2/M checkpoints. *Genes & Development*, *16*(2), 245–256. https://doi.org/10.1101/GAD.949802

515. Ren, C., Zheng, Y., Liu, C., Mencius, J., Wu, Z., & Quan, S. (2022). Molecular Characterization of an Intrinsically Disordered Chaperone Reveals Net-Charge Regulation in Chaperone Action: Net-charge regulated chaperone activity of an IDP. *Journal of Molecular Biology*, *434*(5). https://doi.org/10.1016/j.jmb.2021.167405

516. Renaud, J. P., Chari, A., Ciferri, C., Liu, W. T., Rémigy, H. W., Stark, H., & Wiesmann, C. (2018). Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nature Reviews Drug Discovery*, *17*(7), 471–492. https://doi.org/10.1038/nrd.2018.77

517. Riback, J. A., Eeftens, J. M., Lee, D. S., Quinodoz, S. A., Beckers, L., Becker, L. A., & Brangwynne, C. P. (2022). Viscoelastic RNA entanglement and advective flow underlies nucleolar form and function. *Biophysical Journal*, *121*(3), 473a. https://doi.org/10.1016/j.bpj.2021.11.419

518. Riback, J. A., Zhu, L., Ferrolino, M. C., Tolbert, M., Mitrea, D. M., Sanders, D. W., Wei, M. T., Kriwacki, R. W., & Brangwynne, C. P. (2020). Composition-dependent thermodynamics of intracellular phase separation. *Nature*, *581*(7807), 209–214. https://doi.org/10.1038/s41586-020-2256-2

519. Richardson, H. E., Wittenberg, C., Cross, F., & Reed, S. I. (1989). An essential G1 function for cyclin-like proteins in yeast. *Cell*, *59*(6), 1127–1133. https://doi.org/10.1016/0092-8674(89)90768-X

520. Rickert, P., Corden, J. L., & Lees, E. (1999). Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases. *Oncogene*, *18*(4), 1093–1102. https://doi.org/10.1038/sj.onc.1202399

521. Rocha, P. P., Scholze, M., Bleiß, W., & Schrewe, H. (2010). Med12 is essential for early mouse development and for canonical Wnt and Wnt/PCP signaling. *Development*, *137*(16), 2723–2731. https://doi.org/10.1242/dev.053660

522. Roll-Mecak, A. (2015). Intrinsically disordered tubulin tails: Complex tuners of microtubule functions? *Seminars in Cell and Developmental Biology*, *37*, 11–19. https://doi.org/10.1016/j.semcdb.2014.09.026

523. Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. E., & Dunker, A. K. (1997). Identifying disordered regions in proteins from amino acid sequence. *IEEE International Conference on Neural Networks - Conference Proceedings*, *1*, 90–95. https://doi.org/10.1109/ICNN.1997.611643

524. Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., & Chait, B. T. (2000). The yeast nuclear pore complex: Composition, architecture, transport mechanism. *Journal of Cell Biology*, *148*(4), 635–651. https://doi.org/10.1083/jcb.148.4.635

525. Rowles, A., Tada, S., & Blow, J. J. (1999). Changes in association of the Xenopus origin recognition complex with chromatin on licensing of replication origins. *Journal of Cell Science*, *112*(12), 2011–2018. https://doi.org/10.1242/jcs.112.12.2011

526. Rubinstein, M., & Colby, R. H. (2003). *Polymer physics*. Oxford University Press.

527. Rudner, A. D., & Murray, A. W. (2000). Phosphorylation by Cdc28 activates the Cdc20-dependent activity of the anaphase-promoting complex. *Journal of Cell Biology*, *149*(7), 1377–1390. https://doi.org/10.1083/jcb.149.7.1377

528. Russell, P., & Nurse, P. (1986). cdc25+ functions as an inducer in the mitotic control of fission yeast. *Cell*, *45*(1), 145–153. https://doi.org/10.1016/0092-8674(86)90546-5

529. Russell, P., & Nurse, P. (1987). Negative regulation of mitosis by wee1+, a gene encoding a protein kinase homolog. *Cell*, *49*(4), 559–567. https://doi.org/10.1016/0092-8674(87)90458-2

530. Saibil, H. (2000). Molecular chaperones: containers and surfaces for folding, stabilising or unfolding proteins. *Current Opinion in Structural Biology*, *10*(2), 251–258. https://doi.org/10.1016/S0959-440X(00)00074-9

531. Sanjana, N. E., Shalem, O., & Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods 2014 11:8*, *11*(8), 783–784. https://doi.org/10.1038/NMETH.3047

532. Santamaría, D., Barrière, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., Cáceres, J. F., Dubus, P., Malumbres, M., & Barbacid, M. (2007). Cdk1 is sufficient to drive the mammalian cell cycle. *Nature 2007 448:7155*, *448*(7155), 811–815. https://doi.org/10.1038/nature06046

533. Sanulli, S., Trnka, M. J., Dharmarajan, V., Tibble, R. W., Pascal, B. D., Burlingame, A. L., Griffin, P. R., Gross, J. D., & Narlikar, G. J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature*, *575*(7782), 390–394. https://doi.org/10.1038/s41586-019-1669-2

534. Sato, S., Tomomori-Sato, C., Parmely, T. J., Florens, L., Zybailov, B., Swanson, S. K., Banks, C. A. S., Jin, J., Cai, Y., Washburn, M. P., Conaway, J. W., & Conaway, R. C. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Molecular Cell*, *14*(5), 685–691. https://doi.org/10.1016/j.molcel.2004.05.006

535. Sato, T., Vries, R. G., Snippert, H. J., Van De Wetering, M., Barker, N., Stange, D. E., Van Es, J. H., Abo, A., Kujala, P., Peters, P. J., & Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*, *459*(7244), 262–265. https://doi.org/10.1038/nature07935

536. Sawyer, I. A., Hager, G. L., & Dundr, M. (2016). Specific genomic cues regulate Cajal body assembly. *Https://Doi.Org/10.1080/15476286.2016.1243648*, *14*(6), 791–803. https://doi.org/10.1080/15476286.2016.1243648

537. Schad, E., Tompa, P., & Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biology*, *12*(12), 1–13. https://doi.org/10.1186/gb-2011-12-12-r120

538. Schaefer, C., Schlessinger, A., & Rost, B. (2010). Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. *Bioinformatics*, *26*(5), 625–631. https://doi.org/10.1093/bioinformatics/btq012

539. Schluter, C., Duchrow, M., Wohlenberg, C., Becker, M. H. G., Key, G., Flad -, H. D., & Gerdes, J. (1993). The cell proliferation-associated antigen of antibody Ki-67: A very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins. *Journal of Cell Biology*, *123*(3), 513–522. https://doi.org/10.1083/jcb.123.3.513

540. Schmidlin, T., Debets, D. O., van Gelder, C. A. G. H., Stecker, K. E., Rontogianni, S., van den Eshof, B. L., Kemper, K., Lips, E. H., van den Biggelaar, M., Peeper, D. S., Heck, A. J. R., & Altelaar, M. (2019). High-Throughput Assessment of Kinome-wide Activation States. *Cell Systems*, *9*(4), 366-374.e5. https://doi.org/10.1016/J.CELS.2019.08.005

541. Schmidt, H. B., & Görlich, D. (2016). Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. In *Trends in Biochemical Sciences* (Vol. 41, Issue 1, pp. 46–61). Elsevier. https://doi.org/10.1016/j.tibs.2015.11.001

542. Schnütgen, F., Doerflinger, N., Calléja, C., Wendling, O., Chambon, P., & Ghyselinck, N. B. (2003). A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nature Biotechnology*, *21*(5), 562–565. https://doi.org/10.1038/nbt811

543. Schulman, B. A., Lindstrom, D. L., & Harlow, E. (1998). Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10453–10458. https://doi.org/10.1073/pnas.95.18.10453

544. Seker, H., Rubbi, C., Linke, S. P., Bowman, E. D., Garfield, S., Hansen, L., Borden, K. L. B., Milner, J., & Harris, C. C. (2003). UV-C-induced DNA damage leads to p53-dependent nuclear trafficking of PML. *Oncogene*, *22*(11), 1620–1628. https://doi.org/10.1038/sj.onc.1206140

545. Seki, T., & Diffley, J. F. X. (2000). Stepwise assembly of initiation proteins at budding yeast replication origins in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14115–14120. https://doi.org/10.1073/pnas.97.26.14115

546. Serrao, A., Jenkins, L. M., Chumanevich, A. A., Horst, B., Liang, J., Gatza, M. L., Lee, N. Y., Roninson, I. B., Broude, E. V., & Mythreye, K. (2018). Mediator kinase CDK8/CDK19 drives YAP1-dependent BMP4-induced EMT in cancer. *Oncogene*, *37*(35), 4792–4808. https://doi.org/10.1038/s41388-018-0316-y

547. Seshacharyulu, P., Pandey, P., Datta, K., & Batra, S. K. (2013). Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer Letters*, *335*(1), 9–18. https://doi.org/10.1016/J.CANLET.2013.02.036

548. Sha, Z., Blyszcz, T., González-Prieto, R., Vertegaal, A. C. O., & Goldberg, A. L. (2019). Inhibiting ubiquitination causes an accumulation of SUMOylated newly synthesized nuclear proteins at PML bodies. *Journal of Biological Chemistry*, *294*(42), 15218–15234. https://doi.org/10.1074/jbc.RA119.009147

549. Shapiro, D. M., Ney, M., Eghtesadi, S. A., & Chilkoti, A. (2021). Protein Phase Separation Arising from Intrinsic Disorder: First-Principles to Bespoke Applications. *Journal of Physical Chemistry B*, *125*(25), 6740–6759. https://doi.org/10.1021/acs.jpcb.1c01146

550. Sharma, K., D'Souza, R. C. J., Tyanova, S., Schaab, C., Wiśniewski, J. R., Cox, J., & Mann, M. (2014). Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Reports*, *8*(5), 1583–1594. https://doi.org/10.1016/j.celrep.2014.07.036

551. Shen, T. H., Lin, H. K., Scaglioni, P. P., Yung, T. M., & Pandolfi, P. P. (2006). The Mechanisms of PML-Nuclear Body Formation. *Molecular Cell*, *24*(3), 331–339. https://doi.org/10.1016/j.molcel.2006.09.013

552. Sherr, C. J. (2000). *The pezcoller lecture: Cancer cell cycles revisited*. Cancer Research. https://aacrjournals.org/cancerres/article/60/14/3689/506403/The-Pezcoller-Lecture-Cancer-Cell-Cycles-Revisited

553. Sheu, Y. J., & Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature*, *463*(7277), 113–117. https://doi.org/10.1038/nature08647

554. Shi, L., & Tu, B. P. (2013). Acetyl-CoA induces transcription of the key G1 cyclin CLN3 to promote entry into the cell division cycle in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(18), 7318–7323. https://doi.org/10.1073/pnas.1302490110

555. Shin, Y., Berry, J., Pannucci, N., Haataja, M. P., Toettcher, J. E., & Brangwynne, C. P. (2017). Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell*, *168*(1–2), 159-171.e14. https://doi.org/10.1016/j.cell.2016.11.054

556. Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. In *Science* (Vol. 357, Issue 6357). American Association for the Advancement of Science. https://doi.org/10.1126/science.aaf4382

557. Shively, J. M., Ball, F., Brown, D. H., & Saunders, R. E. (1973). Functional organelles in prokaryotes: Polyhedral inclusions (carboxysomes) of thiobacillus neapolitanus. *Science*, *182*(4112), 584–586. https://doi.org/10.1126/science.182.4112.584

558. Shupp, A., Casimiro, M. C., Pestell, R. G., Shupp, A., Casimiro, M. C., & Pestell, R. G. (2017). Biological functions of CDK5 and potential CDK5 targeted clinical treatments. *Oncotarget*, *8*(10), 17373–17382. https://doi.org/10.18632/ONCOTARGET.14538

559. Simanis, V., & Nurse, P. (1986). The cell cycle control gene cdc2+ of fission yeast encodes a protein kinase potentially regulated by phosphorylation. *Cell*, *45*(2), 261–268. https://doi.org/10.1016/0092-8674(86)90390-9

560. Sirri, V., Hernandez-Verdun, D., & Roussel, P. (2002). Cyclin-dependent kinases govern formation and maintenance of the nucleolus. *Journal of Cell Biology*, *156*(6), 969–981. https://doi.org/10.1083/jcb.200201024

561. Siu, K. T., Rosner, M. R., & Minella, A. C. (2012). An integrated view of cyclin E function and regulation. *Cell Cycle*, *11*(1), 57. https://doi.org/10.4161/CC.11.1.18775

562. Skotheim, J. M., Di Talia, S., Siggia, E. D., & Cross, F. R. (2008). Positive feedback of G1 cyclins ensures coherent cell cycle entry. *Nature*, *454*(7202), 291–296. https://doi.org/10.1038/nature07118

563. Sobecki, M., Mrouj, K., Camasses, A., Parisis, N., Nicolas, E., Llères, D., Gerbe, F., Prieto, S., Krasinska, L., David, A., Eguren, M., Birling, M. C., Urbach, S., Hem, S., Déjardin, J., Malumbres, M., Jay, P., Dulic, V., Lafontaine, D. L. J., … Fisher, D. (2016). The cell proliferation antigen Ki-67 organises heterochromatin. *ELife*, *5*(MARCH2016). https://doi.org/10.7554/eLife.13722

564. Söding, J., Zwicker, D., Sohrabi-Jahromi, S., Boehning, M., & Kirschbaum, J. (2020). Mechanisms for Active Regulation of Biomolecular Condensates. *Trends in Cell Biology*, *30*(1), 4–14. https://doi.org/10.1016/j.tcb.2019.10.006

565. Sola, M. M., Langan, T., & Cohen, P. (1991). p34cdc2 phosphorylation sites in histone H1 are dephosphorylated by protein phosphatase 2A1. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1094*(2), 211–216. https://doi.org/10.1016/0167-4889(91)90011-L

566. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., & Cantley, L. C. (1994). Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Current Biology*, *4*(11), 973–982. https://doi.org/10.1016/S0960-9822(00)00221-9

567. Spencer, S. L., Cappell, S. D., Tsai, F. C., Overton, K. W., Wang, C. L., & Meyer, T. (2013). XThe proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell*, *155*(2), 369. https://doi.org/10.1016/j.cell.2013.08.062

568. Spies, J., Lukas, C., Somyajit, K., Rask, M. B., Lukas, J., & Neelsen, K. J. (2019). 53BP1 nuclear bodies enforce replication timing at under-replicated DNA to limit heritable DNA damage. *Nature Cell Biology*, *21*(4), 487–497. https://doi.org/10.1038/s41556-019-0293-6

569. Stadler, M., Chelbi-Alix, M. K., Koken, M. H. M., Venturini, L., Lee, C., Saïb, A., Quignon, F., Pelicano, L., Guillemin, M. C., Schindler, C., & De Thé, H. (1995). Transcriptional induction of the PML growth suppressor gene by interferons is mediated through an ISRE and a GAS element. *Oncogene*, *11*(12), 2564–2573.

570. Starborg, M., Gell, K., Brundell, E., & Höög, C. (1996). The murine Ki-67 cell proliferation antigen accumulates in the nucleolar and heterochromatic regions of interphase cells and at the periphery of the mitotic chromosomes in a process essential for cell cycle progression. *Journal of Cell Science*, *109*(1), 143–153. https://doi.org/10.1242/jcs.109.1.143

571. Stegmeier, F., & Amon, A. (2002). *CLOSING MITOSIS: The Functions of the Cdc14 Phosphatase and Its Regulation*. https://doi.org/10.1146/annurev.genet.38.072902.093051

572. Steinparzer, I., Sedlyarov, V., Rubin, J. D., Eislmayr, K., Galbraith, M. D., Levandowski, C. B., Vcelkova, T., Sneezum, L., Wascher, F., Amman, F., Kleinova, R., Bender, H., Andrysik, Z., Espinosa, J. M., Superti-Furga, G., Dowell, R. D., Taatjes, D. J., & Kovarik, P. (2019). Transcriptional Responses to IFN-γ Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions. *Molecular Cell*, *76*(3), 485-499.e8. https://doi.org/10.1016/J.MOLCEL.2019.07.034

573. Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P., & Kirschner, M. W. (2001). Dual inhibition of sister chromatid separation at metaphase. *Cell*, *107*(6), 715–726. https://doi.org/10.1016/S0092-8674(01)00603-1

574. Stenström, L., Mahdessian, D., Gnann, C., Cesnik, A. J., Ouyang, W., Leonetti, M. D., Uhlén, M., Cuylen-Haering, S., Thul, P. J., & Lundberg, E. (2020). Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. *Molecular Systems Biology*, *16*(8), e9469. https://doi.org/10.15252/msb.20209469

575. Stern, B., & Nurse, P. (1996). A quantitative model for the cdc2 control of S phase and mitosis in fission yeast. *Trends in Genetics*, *12*(9), 345–350. https://doi.org/10.1016/S0168-9525(96)80016-3

576. Strausfeld, U., Labbé, J. C., Fesquet, D., Cavadore, J. C., Picard, A., Sadhu, K., Russell, P., & Dorée, M. (1991). Dephosphorylation and activation of a p34cdc2/cyclin B complex in vitro by human CDC25 protein. *Nature*, *351*(6323), 242–245. https://doi.org/10.1038/351242a0

577. Strickfaden, H., Tolsma, T. O., Sharma, A., Underhill, D. A., Hansen, J. C., & Hendzel, M. J. (2020). Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell*, *183*(7), 1772-1784.e13. https://doi.org/10.1016/j.cell.2020.11.027

578. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016

579. Su, G., Kuchinsky, A., Morris, J. H., States, D. J., & Meng, F. (2010). GLay: Community structure analysis of biological networks. *Bioinformatics*, *26*(24), 3135–3137. https://doi.org/10.1093/bioinformatics/btq596

580. 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., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

581. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

582. Suski, J. M., Ratnayeke, N., Braun, M., Zhang, T., Strmiska, V., Michowski, W., Can, G., Simoneau, A., Snioch, K., Cup, M., Sullivan, C. M., Wu, X., Nowacka, J., Branigan, T. B., Pack, L. R., DeCaprio, J. A., Geng, Y., Zou, L., Gygi, S. P., … Sicinski, P. (2022). CDC7-independent G1/S transition revealed by targeted protein degradation. *Nature*, *605*(7909), 357–365. https://doi.org/10.1038/s41586-022-04698-x

583. Sutter, M., Greber, B., Aussignargues, C., & Kerfeld, C. A. (2017). Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science*, *356*(6344), 1293–1297. https://doi.org/10.1126/science.aan3289

584. Swanton, C., Mann, D. J., Fleckenstein, B., Neipel, F., Peters, G., & Jones, N. (1997). Herpes viral cyclin/Cdk6 complexes evade inhibition by CDK inhibitor proteins. *Nature*, *390*(6656), 184–187. https://doi.org/10.1038/36606

585. Swiss Institute of Bioinformatics. (n.d.). *SWISS-MODEL H. sapiens*. https://swissmodel.expasy.org/repository/species/9606

586. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

587. Szmyd, R., Niska-Blakie, J., Diril, M. K., Renck Nunes, P., Tzelepis, K., Lacroix, A., van Hul, N., Deng, L. W., Matos, J., Dreesen, O., Bisteau, X., & Kaldis, P. (2019). Premature activation of Cdk1 leads to mitotic events in S phase and embryonic lethality. *Oncogene*, *38*(7), 998–1018. https://doi.org/10.1038/s41388-018-0464-0

588. Tafforeau, L., Zorbas, C., Langhendries, J. L., Mullineux, S. T., Stamatopoulou, V., Mullier, R., Wacheul, L., & Lafontaine, D. L. J. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Molecular Cell*, *51*(4), 539–551. https://doi.org/10.1016/j.molcel.2013.08.011

589. Takagi, M., Natsume, T., Kanemaki, M. T., & Imamoto, N. (2016). Perichromosomal protein Ki67 supports mitotic chromosome architecture. *Genes to Cells*, *21*(10), 1113–1124. https://doi.org/10.1111/gtc.12420

590. Tanaka, T. U., Rachidi, N., Janke, C., Pereira, G., Galova, M., Schiebel, E., Stark, M. J. R., & Nasmyth, K. (2002). Evidence that the Ipl1-Sli15 (Aurora Kinase-INCENP) complex promotes chromosome bi-orientation by altering kinetochore-spindle pole connections. *Cell*, *108*(3), 317–329. https://doi.org/10.1016/S0092-8674(02)00633-5

591. Tanaka, T., Knapp, D., & Kim, N. (1997). Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell*, *90*(4), 649–660. https://doi.org/10.1016/S0092-8674(00)80526-7

592. Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I. F., Tang, W., Cisneros, D. A., Bhaskara, V., Nishiyama, T., Vaziri, A., Wutz, A., Ellenberg, J., & Peters, J. M. (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature*, *501*(7468), 564–568. https://doi.org/10.1038/nature12471

593. Temu, T., Mann, M., Räschle, M., & Cox, J. (2016). Homology-driven assembly of NOn-redundant protEin sequence sets (NOmESS) for mass spectrometry. *Bioinformatics*, *32*(9), 1417–1419. https://doi.org/10.1093/bioinformatics/btv756

594. Thiel, D. A., Reeder, M. K., Pfaff, A., Coleman, T. R., Sells, M. A., & Chernoff, J. (2002). Cell Cycle-Regulated Phosphorylation of p21-Activated Kinase 1. *Current Biology*, *12*(14), 1227–1232. https://doi.org/10.1016/S0960-9822(02)00931-4

595. Thron, C. D. (1996). A model for a bistable biochemical trigger of mitosis. *Biophysical Chemistry*, *57*(2–3), 239–251. https://doi.org/10.1016/0301-4622(95)00075-5

596. Tiwari, N., Tiwari, V. K., Waldmeier, L., Balwierz, P. J., Arnold, P., Pachkov, M., Meyer-Schaller, N., Schübeler, D., vanNimwegen, E., & Christofori, G. (2013). Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. *Cancer Cell*, *23*(6), 768–783. https://doi.org/10.1016/j.ccr.2013.04.020

597. Todaro, G. J., & Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *The Journal of Cell Biology*, *17*(2), 299–313. https://doi.org/10.1083/jcb.17.2.299

598. Tompa, P. (2003). Intrinsically unstructured proteins evolve by repeat expansion. In *BioEssays* (Vol. 25, Issue 9, pp. 847–855). https://doi.org/10.1002/bies.10324

599. Tompa, P., & Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal*, *18*(11), 1169–1175. https://doi.org/10.1096/fj.04-1584rev

600. Tompa, P., & Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, *33*(1), 2–8. https://doi.org/10.1016/j.tibs.2007.10.003

601. Tompa, P., Szász, C., & Buday, L. (2005). Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences*, *30*(9), 484–489. https://doi.org/10.1016/j.tibs.2005.07.008

602. Tóth-Petróczy, Á., Oldfield, C. J., Simon, I., Takagi, Y., Dunker, A. K., Uversky, V. N., & Fuxreiter, M. (2008). Malleable machines in transcription regulation: The Mediator complex. *PLoS Computational Biology*, *4*(12), e1000243. https://doi.org/10.1371/journal.pcbi.1000243

603. Trautmann, S., Wolfe, B. A., Jorgensen, P., Tyers, M., Gould, K. L., & McCollum, D. (2001). Fission yeast Clp1p phosphatase regulates G2/M transition and coordination of cytokinesis with cell cycle progression. *Current Biology*, *11*(12), 931–940. https://doi.org/10.1016/S0960-9822(01)00268-8

604. Travesa, A., Kuo, D., De Bruin, R. A. M., Kalashnikova, T. I., Guaderrama, M., Thai, K., Aslanian, A., Smolka, M. B., Yates, J. R., Ideker, T., & Wittenberg, C. (2012). DNA replication stress differentially regulates G1/S genes via Rad53-dependent inactivation of Nrm1. *The EMBO Journal*, *31*(7), 1811–1822. https://doi.org/10.1038/EMBOJ.2012.28

605. Trunnell, N. B., Poon, A. C., Kim, S. Y., & Ferrell, J. E. (2011). Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Molecular Cell*, *41*(3), 263–274. https://doi.org/10.1016/j.molcel.2011.01.012

606. Tsai, T. Y. C., Theriot, J. A., & Ferrell, J. E. (2014). Changes in Oscillatory Dynamics in the Cell Cycle of Early Xenopus laevis Embryos. *PLoS Biology*, *12*(2), e1001788. https://doi.org/10.1371/journal.pbio.1001788

607. Tsang, B., Pritišanac, I., Scherer, S. W., Moses, A. M., & Forman-Kay, J. D. (2020). Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell*, *183*(7), 1742–1756. https://doi.org/10.1016/j.cell.2020.11.050

608. Tseng, L. C., & Chen, R. H. (2011). Temporal control of nuclear envelope assembly by phosphorylation of lamin B receptor. *Molecular Biology of the Cell*, *22*(18), 3306–3317. https://doi.org/10.1091/mbc.E11-03-0199

609. Tsuji, T., Ficarro, S. B., & Jiang, W. (2006). Essential role of phosphorylation of MCM2 by Cdc7/Dbf4 in the initiation of DNA replication in mammalian cells. *Molecular Biology of the Cell*, *17*(10), 4459–4472. https://doi.org/10.1091/mbc.E06-03-0241

610. Tsutsui, T., Umemura, H., Tanaka, A., Mizuki, F., Hirose, Y., & Ohkuma, Y. (2008). Human mediator kinase subunit CDK11 plays a negative role in viral activator VP16-dependent transcriptional regulation. *Genes to Cells*, *13*(8), 817–826. https://doi.org/10.1111/J.1365-2443.2008.01208.X

611. Turoverov, K. K., Kuznetsova, I. M., & Uversky, V. N. (2010). The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Progress in Biophysics and Molecular Biology*, *102*(2–3), 73–84. https://doi.org/10.1016/j.pbiomolbio.2010.01.003

612. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, *13*(9), 731–740. https://doi.org/10.1038/nmeth.3901

613. Tyson, J. J., & Novak, B. (2001). Regulation of the Eukaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions. *Journal of Theoretical Biology*, *210*(2), 249–263. https://doi.org/10.1006/JTBI.2001.2293

614. Tyson, J. J., & Novák, B. (2022). Time-keeping and decision-making in the cell cycle. *Interface Focus*, *12*(4). https://doi.org/10.1098/RSFS.2021.0075

615. Tyson, J. J., Novak, B., Odell, G. M., Chen, K., & Thron, C. D. (1996). Chemical kinetic theory: understanding cell-cycle regulation. *Trends in Biochemical Sciences*, *21*(3), 89–96. https://doi.org/10.1016/S0968-0004(96)10011-6

616. Ubersax, J. A., Woodbury, E. L., Quang, P. N., Paraz, M., Blethrow, J. D., Shah, K., Shokat, K. M., & Morgan, D. O. (2003). Targets of the cyclin-dependent kinase Cdk1. *Nature 2003 425:6960*, *425*(6960), 859–864. https://doi.org/10.1038/nature02062

617. Uebe, R., & Schüler, D. (2016). Magnetosome biogenesis in magnetotactic bacteria. *Nature Reviews Microbiology*, *14*(10), 621–637. https://doi.org/10.1038/nrmicro.2016.99

618. Uhlmann, F., Wernic, D., Poupart, M. A., Koonin, E. V., & Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell*, *103*(3), 375–386. https://doi.org/10.1016/S0092-8674(00)00130-6

619. Ungricht, R., & Kutay, U. (2017). Mechanisms and functions of nuclear envelope remodelling. *Nature Reviews Molecular Cell Biology*, *18*(4), 229–245. https://doi.org/10.1038/nrm.2016.153

620. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115–e115. https://doi.org/10.1093/nar/gks596

621. Uversky, V. N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology*, *44*, 18–30. https://doi.org/10.1016/j.sbi.2016.10.015

622. Uversky, V. N. (2019). Intrinsically disordered proteins and their “Mysterious” (meta)physics. *Frontiers in Physics*, *7*(FEB), 10. https://doi.org/10.3389/fphy.2019.00010

623. Uversky, V. N., & Dunker, A. K. (2013). The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. *F1000 Biology Reports*, *5*(1). https://doi.org/10.3410/B5-1

624. Vagnarelli, P. (2012). Mitotic chromosome condensation in vertebrates. *Experimental Cell Research*, *318*(12), 1435–1441. https://doi.org/10.1016/j.yexcr.2012.03.017

625. Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., & Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. In *Chemical Reviews* (Vol. 114, Issue 13, pp. 6589–6631). https://doi.org/10.1021/cr400525m

626. van Mierlo, G., Jansen, J. R. G., Wang, J., Poser, I., van Heeringen, S. J., & Vermeulen, M. (2021). Predicting protein condensate formation using machine learning. *Cell Reports*, *34*(5), 108705. https://doi.org/10.1016/J.CELREP.2021.108705

627. Vigneron, S., Brioudes, E., Burgess, A., Labbé, J. C., Lorca, T., & Castro, A. (2009). Greatwall maintains mitosis through regulation of PP2A. *EMBO Journal*, *28*(18), 2786–2793. https://doi.org/10.1038/emboj.2009.228

628. Visintin, R., Craig, K., Hwang, E. S., Prinz, S., Tyers, M., & Amon, A. (1998). The Phosphatase Cdc14 Triggers Mitotic Exit by Reversal of Cdk-Dependent Phosphorylation. *Molecular Cell*, *2*(6), 709–718. https://doi.org/10.1016/S1097-2765(00)80286-5

629. Voth, W., & Jakob, U. (2017). Stress-Activated Chaperones: A First Line of Defense. *Trends in Biochemical Sciences*, *42*(11), 899–913. https://doi.org/10.1016/j.tibs.2017.08.006

630. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, *59*(4), 687–696. https://doi.org/10.1002/prot.20449

631. Vucetic, S., Brown, C. J., Dunker, A. K., & Obradovic, Z. (2003). Flavors of protein disorder. *Proteins: Structure, Function and Genetics*, *52*(4), 573–584. https://doi.org/10.1002/prot.10437

632. Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L. M., Cortese, M. S., Lawson, J. D., Brown, C. J., Sikes, J. G., Newton, C. D., & Dunker, A. K. (2005). DisProt: A database of protein disorder. *Bioinformatics*, *21*(1), 137–140. https://doi.org/10.1093/bioinformatics/bth476

633. Walsh, I., Martin, A. J. M., Di domenico, T., & Tosatto, S. C. E. (2012). Espritz: Accurate and fast prediction of protein disorder. *Bioinformatics*, *28*(4), 503–509. https://doi.org/10.1093/bioinformatics/btr682

634. Wang, B., Matsuoka, S., Carpenter, P. B., & Elledge, S. J. (2002). 53BP1, a mediator of the DNA damage checkpoint. *Science*, *298*(5597), 1435–1438. https://doi.org/10.1126/science.1076182

635. Wang, Q., Sawyer, I. A., Sung, M. H., Sturgill, D., Shevtsov, S. P., Pegoraro, G., Hakim, O., Baek, S., Hager, G. L., & Dundr, M. (2016). Cajal bodies are linked to genome conformation. *Nature Communications*, *7*(1), 1–17. https://doi.org/10.1038/ncomms10966

636. Ward, I. M., & Klein, P. G. (2007). Polymer Physics. In *eMagRes* (Vol. 2007). Oxford University Press. https://doi.org/10.1002/9780470034590.emrstm0404

637. Wasserman, W. J., & Smith, L. D. (1978). The cyclic behavior of a cytoplasmic factor controlling nuclear membrane breakdown. *Journal of Cell Biology*, *78*(1). https://doi.org/10.1083/jcb.78.1.r15

638. Watson, M. L. (1954). Pores in the mammalian nuclear membrane. *BBA - Biochimica et Biophysica Acta*, *15*(4), 475–479. https://doi.org/10.1016/0006-3002(54)90004-9

639. Wei-Shan, H., Amit, V. C., & Clarke, D. J. (2019). Cell cycle regulation of condensin Smc4. *Oncotarget*, *10*(3), 263–276. https://doi.org/10.18632/oncotarget.26467

640. Welburn, J. P. I., Tucker, J. A., Johnson, T., Lindert, L., Morgan, M., Willis, A., Noble, M. E. M., & Endicott, J. A. (2007). How tyrosine 15 phosphorylation inhibits the activity of cyclin-dependent kinase 2-cyclin A. *Journal of Biological Chemistry*, *282*(5), 3173–3181. https://doi.org/10.1074/JBC.M609151200/ATTACHMENT/C5C35172-8699-4E41-80D2-8CB367CB3E70/MMC1.PDF

641. Welburn, J. P. I., Vleugel, M., Liu, D., Yates, J. R., Lampson, M. A., Fukagawa, T., & Cheeseman, I. M. (2010). Aurora B Phosphorylates Spatially Distinct Targets to Differentially Regulate the Kinetochore-Microtubule Interface. *Molecular Cell*, *38*(3), 383–392. https://doi.org/10.1016/j.molcel.2010.02.034

642. Westerling, T., Kuuluvainen, E., & Mäkelä, T. P. (2007). Cdk8 Is Essential for Preimplantation Mouse Development. *Molecular and Cellular Biology*, *27*(17), 6177–6182. https://doi.org/10.1128/mcb.01302-06

643. White, F. H. (1961). Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease. *The Journal of Biological Chemistry*, *236*(5), 1353–1360. https://doi.org/10.1016/s0021-9258(18)64176-6

644. Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

645. Wickham, H., Averick, M., Bryan, J., Chang, W., D', L., Mcgowan, A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S. M., Müller, K.,

Ooms, J., Robinson, D., Seidel, D. P., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/JOSS.01686

646. Winkler, C., Munter, S. De, Dessel, N. Van, Lesage, B., Heroes, E., Boens, S., Beullens, M., Eynde, A. Van, & Bollen, M. (2015). The selective inhibition of protein phosphatase-1 results in mitotic catastrophe and impaired tumor growth. *Journal of Cell Science*, *128*(24), 4526–4537. https://doi.org/10.1242/JCS.175588/VIDEO-9

647. Wippich, F., Bodenmiller, B., Trajkovska, M. G., Wanka, S., Aebersold, R., & Pelkmans, L. (2013). Dual specificity kinase DYRK3 couples stress granule condensation/ dissolution to mTORC1 signaling. *Cell*, *152*(4), 791–805. https://doi.org/10.1016/j.cell.2013.01.033

648. Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–362. https://doi.org/10.1038/nmeth.1322

649. Wood, D. J., & Endicott, J. A. (2018). Structural insights into the functional diversity of the CDK–cyclin family. In *Open Biology* (Vol. 8, Issue 9). The Royal Society. https://doi.org/10.1098/rsob.180112

650. Woodruff, J. B. (2018). Assembly of Mitotic Structures through Phase Separation. *Journal of Molecular Biology*, *430*(23), 4762–4772. https://doi.org/10.1016/j.jmb.2018.04.041

651. Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, *293*(2), 321–331. https://doi.org/10.1006/jmbi.1999.3110

652. Wu, J. Q., Guo, J. Y., Tang, W., Yang, C. S., Freel, C. D., Chen, C., Nairn, A. C., & Kornbluth, S. (2009). PP1-mediated dephosphorylation of phosphoproteins at mitotic exit is controlled by inhibitor-1 and PP1 phosphorylation. *Nature Cell Biology 2009 11:5*, *11*(5), 644–651. https://doi.org/10.1038/ncb1871

653. Wyatt, H. D. M., Sarbajna, S., Matos, J., & West, S. C. (2013). Coordinated actions of SLX1-SLX4 and MUS81-EME1 for holliday junction resolution in human cells. *Molecular Cell*, *52*(2), 234–247. https://doi.org/10.1016/j.molcel.2013.08.035

654. Xu, J. (2005). Preparation, Culture, and Immortalization of Mouse Embryonic Fibroblasts. *Current Protocols in Molecular Biology*, *70*(1), 28.1.1-28.1.8. https://doi.org/10.1002/0471142727.mb2801s70

655. Xu, N., Libertini, S., Black, E. J., Lao, Y., Hegarat, N., Walker, M., & Gillespie, D. A. (2012). Cdk-mediated phosphorylation of Chk1 is required for efficient activation and full checkpoint proficiency in response to DNA damage. *Oncogene*, *31*(9), 1086–1094. https://doi.org/10.1038/onc.2011.310

656. Xu, W., Wang, Z., Zhang, W., Qian, K., Li, H., Kong, D., Li, Y., & Tang, Y. (2015). Mutated K-ras activates CDK8 to stimulate the epithelial-to-mesenchymal transition in pancreatic cancer in part via the Wnt/β-catenin signaling pathway. *Cancer Letters*, *356*(2), 613–627. https://doi.org/10.1016/j.canlet.2014.10.008

657. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., & Uversky, V. N. (2010). PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1804*(4), 996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011

658. Xue, B., Dunker, A. K., & Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics*, *30*(2), 137–149. https://doi.org/10.1080/07391102.2012.675145

659. Yahya, G., Pérez, A. P., Mendoza, M. B., Parisi, E., Moreno, D. F., Artés, M. H., Gallego, C., & Aldea, M. (2021). Stress granules display bistable dynamics modulated by Cdk. *Journal of Cell Biology*, *220*(3). https://doi.org/10.1083/JCB.202005102

660. Yamaguchi, S., Okayama, H., & Nurse, P. (2000). Fission yeast Fizzy-related protein srw1p is a G1-specific promoter of mitotic cyclin B degradation. *EMBO Journal*, *19*(15), 3968–3977. https://doi.org/10.1093/emboj/19.15.3968

661. Yamamoto, S., Hagihara, T., Horiuchi, Y., Okui, A., Wani, S., Yoshida, T., Inoue, T., Tanaka, A., Ito, T., Hirose, Y., & Ohkuma, Y. (2017). Mediator cyclin-dependent kinases upregulate transcription of inflammatory genes in cooperation with NF-κB and C/EBPβ on stimulation of Toll-like receptor 9. *Genes to Cells*, *22*(3), 265–276. https://doi.org/10.1111/gtc.12475

662. Yamano, H., Gannon, J., & Hunt, T. (1996). The role of proteolysis in cell cycle progression in Schizosaccharomyces pombe. *EMBO Journal*, *15*(19), 5268–5279. https://doi.org/10.1002/j.1460-2075.1996.tb00912.x

663. Yamano, H., Kitamura, K., Kominami, K. ichiro, Lehmann, A., Katayama, S., Hunt, T., & Toda, T. (2000). The spike of S phase cyclin Cig2 expression at the G1-S border in fission yeast requires both APC and SCF ubiquitin ligases. *Molecular Cell*, *6*(6), 1377–1387. https://doi.org/10.1016/S1097-2765(00)00135-0

664. Yamano, H., Kominami, K. I., Harrison, C., Kitamura, K., Katayama, S., Dhut, S., Hunt, T., & Toda, T. (2004). Requirement of the SCFPPop1/Pop2 Ubiquitin Ligase for Degradation of the Fission Yeast S Phase Cyclin Cig2. *Journal of Biological Chemistry*, *279*(18), 18974–18980. https://doi.org/10.1074/jbc.M311060200

665. Yamashita, K., Yasuda, H., Pines, J., Yasumoto, K., Nishitani, H., Ohtsubo, M., Hunter, T., Sugimura, T., & Nishimoto, T. (1990). Okadaic acid, a potent inhibitor of type 1 and type 2A protein phosphatases, activates cdc2/H1 kinase and transiently induces a premature mitosis-like state in BHK21 cells. *The EMBO Journal*, *9*(13), 4331–4338. https://doi.org/10.1002/J.1460-2075.1990.TB07882.X

666. Yamazaki, H., Takagi, M., Kosako, H., Hirano, T., & Yoshimura, S. H. (2022). Cell cycle-specific phase separation regulated by protein charge blockiness. *Nature Cell Biology*, *24*(5), 625–632. https://doi.org/10.1038/s41556-022-00903-1

667. Yang, Q., & Ferrell, J. E. (2013). The Cdk1–APC/C cell cycle oscillator circuit functions as a time-delayed, ultrasensitive switch. *Nature Cell Biology 2013 15:5*, *15*(5), 519–525. https://doi.org/10.1038/ncb2737

668. Yao, G., Lee, T. J., Mori, S., Nevins, J. R., & You, L. (2008). A bistable Rb-E2F switch underlies the restriction point. *Nature Cell Biology*, *10*(4), 476–482. https://doi.org/10.1038/ncb1711

669. Yao, R. W., Xu, G., Wang, Y., Shan, L., Luan, P. F., Wang, Y., Wu, M., Yang, L. Z., Xing, Y. H., Yang, L., & Chen, L. L. (2019). Nascent Pre-rRNA Sorting via Phase Separation Drives the Assembly of Dense Fibrillar Components in the Human Nucleolus. *Molecular Cell*, *76*(5), 767-783.e11. https://doi.org/10.1016/j.molcel.2019.08.014

670. Ye, X., Zhu, C., & Harper, J. W. (2001). A premature-termination mutation in the Mus musculus cyclin-dependent kinase 3 gene. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(4), 1682–1686. https://doi.org/10.1073/pnas.98.4.1682

671. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A., & Diffley, J. F. X. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*, *519*(7544), 431–435. https://doi.org/10.1038/nature14285

672. Yeong, F. M., Lim, H. H., Wang, Y., & Surana, U. (2001). Early Expressed Clb Proteins Allow Accumulation of Mitotic Cyclin by Inactivating Proteolytic Machinery during S Phase. *Molecular and Cellular Biology*, *21*(15), 5071–5081. https://doi.org/10.1128/mcb.21.15.5071-5081.2001

673. Yoo, S. H., & Albanesi, J. P. (1990). Ca2+-induced conformational change and aggregation of chromogranin A. Journal of Biological Chemistry, 265(24), 14414–14421. https://doi.org/10.1016/s0021-9258(18)77318-3

674. Yoon, M. K., Mitrea, D. M., Ou, L., & Kriwacki, R. W. (2012). Cell cycle regulation by the intrinsically disordered proteins p21 and p27. In *Biochemical Society Transactions* (Vol. 40, Issue 5, pp. 981–988). https://doi.org/10.1042/BST20120092

675. Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/OMI.2011.0118/ASSET/IMAGES/LARGE/FIGURE1.JPEG

676. Yu, J., Zhao, Y., Li, Z. X., Galas, S., & Goldberg, M. L. (2006). Greatwall Kinase Participates in the Cdc2 Autoregulatory Loop in Xenopus Egg Extracts. *Molecular Cell*, *22*(1), 83–91. https://doi.org/10.1016/j.molcel.2006.02.022

677. Yu, J., Raia, P., Ghent, C. M., Raisch, T., Sadian, Y., Cavadini, S., Sabale, P. M., Barford, D., Raunser, S., Morgan, D. O., & Boland, A. (2021). Structural basis of human separase regulation by securin and CDK1–cyclin B1. *Nature*, *596*(7870), 138–142. https://doi.org/10.1038/s41586-021-03764-0

678. Yu, X., Chini, C. C. S., He, M., Mer, G., & Chen, J. (2003). The BRCT Domain Is a Phospho-Protein Binding Domain. *Science*, *302*(5645), 639–642. https://doi.org/10.1126/science.1088753

679. Yun, J., Chae, H. D., Choi, T. S., Kim, E. H., Bang, Y. J., Chung, J., Choi, K. S., Mantovani, R., & Shin, D. Y. (2003). Cdk2-dependent phosphorylation of the NF-Y transcription factor and its involvement in the p53-p21 signaling pathway. *Journal of Biological Chemistry*, *278*(38), 36966–36972. https://doi.org/10.1074/jbc.M305178200

680. Zachariae, W., Schwab, M., Nasmyth, K., & Seufert, W. (1998). Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the anaphase promoting complex. *Science*, *282*(5394), 1721–1724. https://doi.org/10.1126/science.282.5394.1721

681. Zamudio, A. V., Dall'Agnese, A., Henninger, J. E., Manteiga, J. C., Afeyan, L. K., Hannett, N. M., Coffey, E. L., Li, C. H., Oksuz, O., Sabari, B. R., Boija, A., Klein, I. A., Hawken, S. W., Spille, J. H., Decker, T. M., Cisse, I. I., Abraham, B. J., Lee, T. I., Taatjes, D. J., … Young, R. A. (2019). Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Molecular Cell*, *76*(5), 753-766.e6. https://doi.org/10.1016/j.molcel.2019.08.016

682. Zandany, N., Lewin, L., Nirenberg, V., Orr, I., & Yifrach, O. (2015). Entropic clocks in the service of electrical signaling: "Ball and chain" mechanisms for ion channel inactivation and clustering. *FEBS Letters*, *589*(19), 2441–2447. https://doi.org/10.1016/j.febslet.2015.06.010

683. Zaslavsky, B. Y., & Uversky, V. N. (2018). In Aqua Veritas: The Indispensable yet Mostly Ignored Role of Water in Phase Separation and Membrane-less Organelles. *Biochemistry*, *57*(17), 2437–2451. https://doi.org/10.1021/acs.biochem.7b01215

684. Zegerman, P. (2015). Evolutionary conservation of the CDK targets in eukaryotic DNA replication initiation. *Chromosoma*, *124*(3), 309–321. https://doi.org/10.1007/s00412-014-0500-y

685. Zhang, C., Kenski, D. M., Paulson, J. L., Bonshtien, A., Sessa, G., Cross, J. V., Templeton, D. J., & Shokat, K. M. (2005). A second-site suppressor strategy for chemical genetic analysis of diverse protein kinases. *Nature Methods*, *2*(6), 435–441. https://doi.org/10.1038/nmeth764

686. Zhang, H., Zhao, R., Tones, J., Liu, M., Dilley, R. L., Chenoweth, D. M., Greenberg, R. A., & Lampson, M. A. (2020). Nuclear body phase separation drives telomere clustering in ALT cancer cells. *Molecular Biology of the Cell*, *31*(18), 2048–2056. https://doi.org/10.1091/mbc.E19-10-0589

687. Zhang, L., Geng, X., Wang, F., Tang, J., Ichida, Y., Sharma, A., Jin, S., Chen, M., Tang, M., Pozo, F. M., Wang, W., Wang, J., Wozniak, M., Guo, X., Miyagi, M., Jin, F., Xu, Y., Yao, X., & Zhang, Y. (2022). 53BP1 regulates heterochromatin through liquid phase separation. *Nature Communications*, *13*(1), 1–16. https://doi.org/10.1038/s41467-022-28019-y

688. Zhang, T., Faraggi, E., Xue, B., Dunker, A. K., Uversky, V. N., & Zhou, Y. (2012). Spine-d: Accurate prediction of short and long disordered regions by a single neural-network based method. *Journal of Biomolecular Structure and Dynamics*, *29*(4), 799–813. https://doi.org/10.1080/073911012010525022

689. Zhao, L., Samuels, T., Winckler, S., Korgaonkar, C., Tompkins, V., Horne, M. C., & Quelle, D. E. (2003). Cyclin G1 has growth inhibitory activity linked to the ARF-Mdm2-p53 and pRb tumor suppressor pathways. *Molecular Cancer Research : MCR*, *1*(3), 195–206. https://pubmed-ncbi-nlm-nih-gov.insb.bib.cnrs.fr/12556559/

690. Zhao, X., Feng, D., Wang, Q., Abdulla, A., Xie, X. J., Zhou, J., Sun, Y., Yang, E. S., Liu, L. P., Vaitheesvaran, B., Bridges, L., Kurland, I. J., Strich, R., Ni, J. Q., Wang, C., Ericsson, J., Pessin, J. E., Ji, J. Y., & Yang, F. (2012). Regulation of lipogenesis by cyclin-dependent kinase 8 - Mediated control of SREBP-1. *Journal of Clinical Investigation*, *122*(7), 2417–2427. https://doi.org/10.1172/JCI61462

691. Zheng, J. N., Pei, D. S., Mao, L. J., Liu, X. Y., Mei, D. D., Zhang, B. F., Shi, Z., Wen, R. M., & Sun, X. Q. (2009). Inhibition of renal cancer cell growth in vitro and in vivo with oncolytic adenovirus armed short hairpin RNA targeting Ki-67 encoding mRNA. *Cancer Gene Therapy*, *16*(1), 20–32. https://doi.org/10.1038/cgt.2008.61

692. Zheng, J. N., Ma, T. X., Cao, J. Y., Sun, X. Q., Chen, J. C., Li, W., Wen, R. M., Sun, Y. F., & Pei, D. S. (2006). Knockdown of Ki-67 by small interfering RNA leads to inhibition of proliferation and induction of apoptosis in human renal carcinoma cells. *Life Sciences*, *78*(7), 724–729. https://doi.org/10.1016/j.lfs.2005.05.064

Appendix 1

A CDK-mediated phosphorylation switch of disordered protein condensation

A CDK-mediated phosphorylation switch of disordered protein condensation

Maarten Altelaar (m.altelaar@uu.nl)
Utrecht University https://orcid.org/0000-0001-5093-5945

Juan Valverde
Utrecht University

Geronimo Dubra
CNRS

Henk W.P. Van den Toorn
Utrecht University https://orcid.org/0000-0002-0270-5763

Guido van Mierlo
EPFL https://orcid.org/0000-0001-5883-0339

Michiel Vermeulen
Radboud University Nijmegen https://orcid.org/0000-0003-0836-6894

Albert Heck
Utrecht University https://orcid.org/0000-0002-2405-4404

Carlos Elena-Real
CBS, University of Montpellier

Aurélie Fournet
CBS, University of Montpellier

Emile Al Ghoul
IGH, University of Montpellier

Dhanvantri Chahar
IGMM, University of Montpellier

Austin Haider
University of Denver

Matteo Paloni
CBS, University of Montpellier https://orcid.org/0000-0003-4841-9321

Angelos Constantinou
Institute of Human Genetics, UMR9002 CNRS-UM, 141 rue de la Cardonille, 34396 Montpellier, France.

Alessandro Barducci
Centre de Biochimie Structurale

Kingshuk Ghosh
University of Denver

Nathalie Sibille

CBS, University of Montpellier

Pau Bernadó

CBS

Puck Knipscheer

https://orcid.org/0000-0003-4198-0132

Liliana Krasinska

CNRS https://orcid.org/0000-0002-6858-0852

Daniel Fisher

French National Centre for Scientific Research https://orcid.org/0000-0002-0822-3482

Biological Sciences - Article

Keywords:

Posted Date: February 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1370895/v1

A CDK-mediated phosphorylation switch of disordered protein condensation

Authors: Juan Manuel Valverde[1,2†], Geronimo Dubra[3,4†], Henk van den Toorn[1,2], Guido van Mierlo[5], Michiel Vermeulen[5], Albert J.R. Heck[1,2], Carlos Elena-Real[6], Aurélie Fournet[6], Emile Al Ghoul[7], Dhanvantri Chahar[3,4], Austin Haider[8], Matteo Paloni[6], Angelos Constantinou[7], Alessandro Barducci[6], Kingshuk Ghosh[8], Nathalie Sibille[6], Pau Bernado[6], Puck Knipscheer[9], Liliana Krasinska[3,4‡], Daniel Fisher[3,4‡*], Maarten Altelaar[1,2‡*]

Affiliations:

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, 3584 CH Utrecht, Netherlands.

[2]Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, Netherlands.

[3]IGMM, University of Montpellier, CNRS, Inserm, Montpellier, France.

[4]Equipe Labellisée LIGUE 2018, Ligue Nationale Contre le Cancer, Paris, France.

[5]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands.

[6]CBS, University of Montpellier, INSERM, CNRS, Montpellier, France.

[7]IGH, University of Montpellier, CNRS, Montpellier, France.

[8]Department of Physics and Astronomy, and Department of Molecular and Cellular Biophysics, University of Denver, Denver, Colorado 80208, USA.

[9]Oncode Institute, Hubrecht Institute–KNAW and University Medical Center, Utrecht, 3584 CT, Netherlands.

*Correspondence to: m.altelaar@uu.nl and daniel.fisher@igmm.cnrs.fr

†‡ Equal contributions

Cell cycle transitions arise from collective changes in protein phosphorylation states triggered by cyclin-dependent kinases (CDKs), but conceptual and mechanistic explanations for the abrupt cellular reorganisation that occurs upon mitotic entry are lacking. Specific interactions between distinct CDK-cyclin complexes and sequence motifs encoded in substrates might result in highly ordered phosphorylation[1], while bistability in the mitotic CDK1 control network can trigger switch-like phosphorylation[2]. Yet the dynamics of mitotic phosphorylation has not been demonstrated *in vivo*, and the roles of most cell cycle-regulated phosphorylations are unclear. Here, we show evidence that switch-like phosphorylation of intrinsically disordered proteins (IDPs) by CDKs contributes to mitotic cellular reorganisation by controlling protein-protein interactions and phase separation. We studied protein phosphorylation in single *Xenopus* embryos throughout synchronous cell cycles, performed parallel assignment of cell cycle phases using egg extracts, and analysed dynamics of mitotic phosphorylation using quantitative targeted phosphoproteomics. This provided a high-resolution map of dynamic phosphosites from the egg to the 16-cell embryo and showed that mitotic phosphorylation occurs on entire protein complexes involved in diverse subcellular processes and is switch-like *in vivo*. Most cell cycle-regulated phosphosites occurred in CDK consensus motifs and located to intrinsically disordered regions. We found that substrates of CDKs and other cell cycle kinases are significantly more disordered than phosphoproteins in general, a principle conserved from yeast to humans, while around half are components of membraneless organelles (MLOs), whose assembly is thought to involve phase separation. Analytical modelling predicts modulation of homotypic IDP interactions by CDK-mediated phosphorylation, which was confirmed by biophysical and biochemical analysis of a model IDP, Ki-67. These results highlight the dynamic control of intrinsic disorder as a conserved hallmark of the cell cycle and suggest a mechanism for CDK-mediated mitotic cellular reorganisation.

Main

To explain behaviour of complex systems, such as the cell cycle, two general strategies have been used[3]: top-down identification of system components, such as screens which have identified hundreds of CDK substrates[4–9] and cell cycle-regulated proteins[10], and bottom-up molecular analysis of the structural effects of individual phosphorylations on single proteins[11]. Yet it has proven challenging to use studies performed at such different scales to reconcile different models of CDK-mediated phosphorylation. We reasoned that understanding how

thousands of mitotic phosphorylations[12] bring about an ordered cell cycle transition would require a multidisciplinary quantitative approach involving cell biology, biochemistry, bioinformatics and biophysics. A *sine qua non* is a time-resolved map of *in vivo* cell cycle phosphorylation in a system devoid of artifacts arising from cell synchronisation[13,14], and with temporal resolution that alternative approaches, like centrifugal elutriation[15] or FACS[16] cannot provide. Dynamic phosphorylation states cannot be determined from cell populations[17], while single-cell proteomics studies[18,19] currently have insufficient sensitivity and reproducibility for low stoichiometry and dynamic targets.

A high-resolution map of *in vivo* cell cycle phosphorylation

We took advantage of the naturally synchronous early cell cycles of *Xenopus laevis* embryos[20,21] to perform quantitative phosphoproteomics *in vivo*, using a sensitive phosphopeptide enrichment strategy[22]. We collected single embryos at 15-minute intervals while recording visual cues of cell divisions. Phosphopeptides from each embryo were purified, separated by nano-LC and analysed by mass spectrometry (Fig. 1a). We identified 4583 high-confidence phosphosites mapping to 1843 proteins (Extended data Fig. 1a; Data S1), most being phosphoserines (Extended data Fig. 1b). Individual embryo phosphorylation states strongly correlated (Extended data Fig. 1c). We thus generated a dynamic map of protein phosphorylation from an unfertilised egg to a 16-cell embryo.

We focused on 1032 sites whose variation in phosphorylation over time was statistically significant (hereafter denoted "dynamic phosphosites") which occurred on 646 proteins. Gene ontology (GO) and network analysis revealed high functional association and interconnectivity between groups of proteins involved in RNA binding and the nuclear pore complex (NPC), DNA replication and chromatin remodeling, and microtubule regulation (Fig. 1b). Hierarchical clustering uncovered four distinct groups that reflect cell cycle-regulated behaviour (Fig. 1c; Data S1). The levels of clusters A and B phosphosites were highest in eggs and post-fertilisation, and decreased during the first round of DNA replication, suggesting that dephosphorylation of these sites may prepare the zygote for upcoming cell divisions[23]. GO analysis for group A highlighted proteins involved in RNA regulation and nuclear organisation, including the NPC and nuclear transport, chromosomal structure and segregation (Extended data Fig. 1d), as also observed in a recent study on meiosis exit[24]. Cluster B phosphosites were enriched in regulators of RNA biosynthesis and stability, translation, actin, DNA replication and repair (Extended data Fig. 1d). Cluster C phosphosites progressively increased after meiotic exit, while cluster D phosphosites had a clear oscillating signature with upregulation

preceding each cell division. GO analysis of cluster C shows dominance of interphase cell cycle processes including DNA replication, RNA-related processes and chromosome organisation (Extended data Fig. 1d), and included phosphosites displaying a reciprocal oscillating trend and a lower amplitude compared to cluster D sites. Several such sites, *e.g.* S31 of the replication licensing protein MCM4, were from monophosphorylated peptides, while the multiphosphorylated forms were found in cluster D (Extended data Fig. 1e). Thus, cluster C contains the earliest phosphorylations of proteins that are highly phosphorylated at mitosis. Cluster D shows coordinated phosphorylation of multiple members of protein complexes involved in diverse processes, suggesting a common mechanism of regulation (Extended data Fig. 1f). Importantly, phosphoproteome changes were not simply a reflection of changes in abundance of the corresponding proteins (Extended data Fig. 2), which are generally negligible during *Xenopus* early development[25].

We assigned *in vivo* embryo phosphosites to different cell cycle stages by comparing with phosphorylation patterns of replicating or mitotic egg extracts (Fig. 1d). Replication was initiated by adding purified sperm chromatin to interphase egg extracts and quantified over time (Fig. 1e, top), while mitosis was triggered by adding recombinant cyclin B and verified microscopically. We also used egg extracts arrested at meiotic metaphase II (Cytostatic Factor, CSF-arrested). Overall, we identified 6937 phosphosites, which included 71% of the sites identified *in vivo* (Fig. 1f, Data S1). 1728 sites varied between S and M-phase, including 693 sites upregulated in S-phase and 1035 in mitosis (Fig. 1e, Data S1). GO analysis of interphase and mitotic sites revealed processes enriched in *in vivo* cluster C and cluster D, respectively (Extended data Fig. 3a). Several DNA-replication factors, including MCM4 and RIF1, showed multi-site phosphorylation specifically in S-phase (Extended data Fig. 3b). This phosphoproteomics dataset greatly increases the known repertoire of phosphorylation sites upregulated during S-phase[12].

We next analysed the cell cycle behaviour of dynamic phosphosites that we found *in vivo* (Extended data Fig. 3c). Most embryo cluster A sites were upregulated in both CSF-arrested meiotic extracts and mitotic extracts, highlighting the global similarities of regulation of meiotic and mitotic M-phase, despite the additional activity of the Mos/MEK/MAP kinase pathway in meiosis. Around half of embryo cluster B sites were present only in interphase, while the rest showed a minimum phosphorylation in late S-phase, confirming their dephosphorylation during the first round of DNA replication. As expected, most sites from embryo clusters C and D were part of the *in vitro* S-phase and mitotic groups, respectively.

Therefore, single embryo data can successfully identify cell cycle-dependent phosphorylation. In mitosis, as expected, monophosphorylated species are reduced because multisite phosphorylation emerges (Extended data Fig. 3d; Extended data Fig. 1e).

Predominance of CDK targets

Analysis of kinase consensus motifs showed that proline-directed (S/T-P) sites, which conform to the minimal consensus for CDKs, comprise 51% of all detected phosphosites *in vivo* and 60% of dynamic sites (Extended data Fig. 4a). Around 10% of all phosphosites matched the full CDK1-family consensus site: S/TPxK/R. Replicating and mitotic extracts displayed a similar trend (Extended data Fig. 4a). Putative CDK targets dominated all clusters, with 80% of sites in cluster D *in vivo* and mitotic clusters *in vitro* conforming to the minimal CDK motif (Fig.1g, Extended data Fig. 4b, c). Consensus sites of other kinases such as Aurora, Polo-like kinase (PLK), DBF4-dependent kinase (DDK) and Casein kinase I and II were present to a lesser extent (Extended data Fig. 4b, d). In meiotic M-phase, MAP kinases, which have the same consensus motif as CDKs, are likely responsible for sites specific to embryo cluster A or CSF extracts, but these kinases are inactivated during early embryonic cell cycles[26], suggesting that most of the other dynamic proline-directed phosphorylations are due to CDKs.

Although few direct CDK substrates have been characterised in *Xenopus*, they are likely conserved between vertebrates. We therefore manually curated a set of 654 human CDK1-subfamily targets (Data S2; see Supplementary Methods for sources). 303 of these have *Xenopus* homologues among the 1843 phosphoproteins we detected, and 149 were present among the 646 proteins with dynamic phosphosites in *Xenopus* embryos (Fig. 1h). Thus, the predominance of CDK motifs among dynamic phosphosites reflects a high proportion of *bona fide* CDK substrates. This is a conservative estimate, since we only considered proline-directed sites as CDK motifs, although we found that 10-20% of human and yeast CDK substrates (Data S2; see Supplementary Methods for sources) were non-proline-directed (Extended data Fig. 4e), confirming a recent finding[33]. These data reinforce the dominant role of CDKs in cell cycle-regulated phosphorylation.

Mitotic phosphorylation is switch-like *in vivo*

We next determined whether mitotic phosphorylation of individual phosphosites is progressive or switch-like *in vivo*. We analysed dynamics of 64 cluster D sites from diverse protein complexes in single embryos every 180-seconds using quantitative targeted phosphoproteomics[27–29] by parallel reaction monitoring[30], thereby obtaining a quantitative

158 description of mitotic phosphorylation *in vivo* at extremely high-time resolution (Fig. 2a). This
159 revealed parallel and abrupt upregulation of all phosphosites preceding each cell division (Fig.
160 2b, c), indicating switch-like phosphorylation of diverse protein complexes at mitotic onset.
161 This was not due to oscillation of CDK1-Y15 inhibitory phosphorylation, which was
162 downregulated over time (Fig. 2d), as previously reported[31], consistent with lack of
163 corresponding phosphorylation of the CDK1-Y15-regulatory enzymes, CDC25 and WEE1. In
164 contrast, oscillating phosphorylations on NIPA and the APC/C, which regulate mitotic cyclin
165 accumulation, as well as Greatwall kinase, which activates the PP2A inhibitors Arpp19/ENSA,
166 were apparent (Extended data Fig. 5a). These data suggest that control of mitotic cyclin levels
167 and PP2A activity, and therefore the overall CDK/phosphatase activity ratio[2], suffices for
168 switch-like mitotic phosphorylation whereas regulated CDK1-Y15 phosphorylation is not
169 essential (Extended data Fig. 5b). This is consistent with the self-sufficiency of futile cycles of
170 opposing enzymes in generating switch-like network output in the absence of allosteric
171 regulation[32].

172 **The cell cycle phosphoproteome is intrinsically disordered**

173 We wondered whether the diverse dynamic phosphoproteins share common structural features
174 facilitating switch-like CDK-mediated phosphorylation. Phosphosites in general are often
175 located in intrinsically disordered regions (IDRs) of proteins[34], which is also true for yeast and
176 mouse CDK sites[35,8,9]. Yet previous analyses did not exclude the possibility that this is an
177 artefact due to the enrichment of serine, threonine and proline in disordered regions, which is
178 consistently predicted across the entire proteome of *Xenopus*, human and yeast (Extended data
179 Fig. 6a). We corrected for this compositional bias, and found that phosphorylatable residues in
180 IDR are indeed more highly phosphorylated than those in ordered regions (Fig. 3a-c). This
181 enrichment was increased for proteins with at least one site displaying dynamic
182 phosphorylation; the same was true for human CDK substrates (Fig. 3b, c). To estimate the
183 differential phosphorylation of disordered sites globally, we calculated the ratio of dynamically
184 phosphorylated (*Xenopus*) or CDK-phosphorylated (yeast, human) to non-phosphorylated
185 serine and threonine in both disordered and structured regions (Extended data Fig. 6b; see
186 Methods). This confirmed that cell cycle-regulated phosphorylation is largely skewed towards
187 disordered regions and that CDKs preferentially phosphorylate disordered sites (Fig. 3d,
188 Extended data Fig. 6c). We then asked whether this is also true for substrates of other protein
189 kinases. We analysed the mitotic PLK and Aurora kinases, DYRK kinases, which promote
190 mitotic phosphorylation of several IDPs[36], NEK kinases, which have roles in centrosome

duplication and various stages of mitosis, and MAP kinases, which share the proline-directed S/T consensus site. For each kinase, documented phosphosites were strongly enriched in IDRs (Extended data Fig. 6c, d), supporting the idea that phosphorylation of residues in IDRs is kinetically favoured[34].

To explain the dominance of CDK-mediated phosphorylation in the cell cycle, we surmised that their substrates might be more disordered than phosphoproteins in general. We therefore determined the percentage of disordered residues of proteins in our datasets, compared to the rest of their respective phosphoproteomes (Data S3). This revealed that, on average, both *Xenopus* dynamic phosphoproteins and human and yeast CDK substrates contain approximately twice the proportion of disordered amino acids as other phosphoproteins (Fig. 3e, Extended data Fig. 6e), putting them among the top quartile of proteins with the most disorder in the proteome. If this reflects the importance of disordered proteins for the cell cycle generally, then substrates of other cell cycle kinases might also be more disordered than other phosphoproteins. Indeed, targets of most cell cycle kinases are significantly more disordered than targets of MAP kinase (Fig. 3f), whose phosphosites are also proline-directed and preferentially located in IDRs (Extended data Fig. 6d).

Enrichment of MLO components among CDK substrates

We thus reasoned that phosphorylation may have been selected to regulate the functions of IDPs during the cell cycle. IDPs are key components of membrane-less organelles (MLO), many of which (*e.g.* Cajal bodies, nucleoli, nuclear pore complexes, splicing speckles) are thought to arise by phase separation (PS)[37], are disassembled in mitosis, and can be regulated by phosphorylation[36,38,39]. To corroborate our hypothesis, we analysed available data on cellular localisation for each of our curated human CDK substrates. We found that 257 (39.2%) are present in MLOs, including key IDPs such as coilin (Cajal bodies), nucleophosmin, nucleolin and Ki-67 (nucleoli), 53BP1 (53BP1 bodies), nucleoporins (NPC) and PML (PML bodies) (Fig. 3g). We then manually curated an MLO proteome from human proteomics studies (Data S4; See Supplementary Methods for sources). Homologues of 204 dynamic *Xenopus* phosphoproteins (31.6%) localise to MLOs, as do 73 of the 149 proteins (50%) that show dynamic phosphorylation in *Xenopus* and are CDK substrates in human (Fig. 3g). The vast majority of proline-directed phosphosites and confirmed CDK sites in these proteins were located in predicted IDRs (Extended data Fig. 7).

CDKs regulate IDR phase separation

Both stochastic and specific interactions between IDPs contribute to PS and MLO assembly[37,40,41]. We hypothesised that cell cycle kinase-mediated phosphorylation might modulate such interactions. We first applied a machine learning classifier[42] to predict whether cell cycle-regulated phosphoproteins have an increase in average propensity for PS (PSAP score). Indeed, we observed a sharp increase in the PSAP score, from the proteome to the phosphoproteome, and a further increase for dynamic phosphoproteins, with the highest score for mitotic cluster D (Extended data Fig. 8a). Similarly, the propensity for PS is far higher amongst targets of most cell cycle kinases (CDK, Aurora, PLK, but not NEK) and DYRK kinases than the overall phosphoproteome, but less so for MAP kinase substrates.

Next, to better understand the biochemical effects of their cell cycle-regulated phosphorylation, we analysed a selection of IDRs from CDK substrates. We applied a general heteropolymer theory that uses sequence charge decoration matrices (SCDM), based on electrostatic interactions only, to identify intra-chain interaction topology[43,44]. Since this should correlate with inter-chain interactions that promote PS, SCDMs provide indirect insights to propensity to phase separate. Of the 12 IDPs tested, 7 (nucleolin, nucleophosmin, NUP53, ELYS, MCM4, 53BP1 and the splicing factor SF3B1) had SCDM maps showing visibly decreased self-association propensity (increased red regions in Extended data Fig. 8b), implying reduced propensity to phase separate, upon CDK-site phosphorylation. Conversely, for SRRM2, CDK-mediated phosphorylation is predicted to increase intra-chain attraction (Extended data Fig. 8b) and hence PS tendency. For 4 proteins (MDC1, TICRR, COILIN, and CDT1), SCDM maps were inconclusive. To further analyse these trends, we calculated radius of gyration of several IDRs using all-atom simulation. Effects of phosphorylation on CDT1 (28.4Å to 30.3Å), TICRR (56.2Å to 57.3Å) and coilin (39 Å to 37.9 Å) were minor, while MCM4 IDR expands upon phosphorylation (21.9Å to 26.3Å), consistent with SCDM analysis. Overall, these data suggest that phosphorylation is a key regulator of homotypic interactions, an important element of PS propensity, of most IDRs.

To test this hypothesis, we focused on a model CDK substrate, Ki-67, an IDP that organises heterochromatin structure[45] and perichromosomal layer formation from nucleolar components in mitosis[46,47]. Ki-67 contains a multivalent Ki-67 repeat domain that is highly phosphorylated in mitosis by CDKs (Fig. 4a), which regulates its perichromosomal localisation[48]. SCDM analysis predicted that phosphorylation of full-length Ki-67 should promote self-interaction and thus PS, but this cannot be attributed to interactions of its repeat motif alone, since phosphorylation of the latter is predicted to reduce homotypic interactions (Fig. 4b). In

agreement, coarse-grained (CG) molecular dynamics (MD) simulations (Extended data Movies 1 and 2) showed that the radius of gyration of full-length Ki-67 decreased upon phosphorylation (Fig. 4c, left) while that of a single consensus repeat motif increased (Fig. 4c, right). MD simulations also showed that PS is enhanced by increasing repeat valency and counteracted by phosphorylation (Fig. 4d), consistent with SCDM analysis. To test these predictions experimentally, we first used the optogenetic Cry2 "optodroplet" system[49] with full length Ki-67 or a series of deletion mutants. Full-length Ki-67 localised to the nucleolus, as expected, but exposure to blue light caused rapid appearance of small round foci in the nucleoplasm, which was dependent on the level of induced Ki-67 expression, consistent with PS (Extended data Fig. 9a). Importantly, promoting CDK-mediated phosphorylation by inhibiting PP2A with okadaic acid[2] led to foci formation in the absence of blue light, while pan-CDK inhibition with purvalanol A prevented induction of foci upon light (Fig. 4e, f). These results indicate that, as predicted by SCDM and MD, phosphorylation of full-length Ki-67 promotes PS. Results were similar for constructs lacking the C-terminal LR domain, that binds chromatin, or the N-terminal domain, which is required for the nucleolar localisation of Ki-67 (Extended data Fig. 9b). Finally, we purified a consensus repeat polypeptide (Extended data Fig. 10a) and phosphorylated it *in vitro* with recombinant CDK complexes. Nuclear Magnetic Resonance spectroscopy showed a reduced amide proton spectral dispersion typical for an IDP, and confirmed appearance of 7 phosphorylated residues upon incubation with purified CDKs and ATP (Fig. 4g). We mapped phosphorylation sites and intensity by phosphoproteomics and Phos-Tag-SDS-PAGE, indicating stoichiometric phosphorylation (Extended data Fig. 10b, c). Purified GFP-tagged Ki-67 repeat motif could phase-separate *in vitro*, and, as predicted, this was abolished upon full phosphorylation by CDK (Fig. 4h). Taken together, these results confirm that CDK-mediated phosphorylation is able to both promote or inhibit homotypic interactions that contribute to PS, and suggest that Ki-67 may have several competing modes of PS that are differentially regulated by phosphorylation. Our data suggest a mechanism for Ki-67-mediated mitotic targeting of nucleolar components to the perichromosomal layer[45,46] via CDK-mediated phosphorylation, which reduces PS of several major nucleolar IDPs, thus triggering nucleolar disassembly, while simultaneously promoting PS of Ki-67 bound to chromatin to recruit nucleolar components.

In conclusion, this work reveals *in vivo* that CDK-dependent mitotic phosphorylation occurs in a switch-like manner on diverse proteins whose common denominators are a high level of disorder and localisation to MLOs. Furthermore, our data show that CDK-mediated

phosphorylation regulates homotypic interactions between IDPs, which may coordinate diverse cellular processes during the cell cycle. While this is not incompatible with models in which high-affinity interactions contribute to MLO formation by PS[50,51], it suggests that cell cycle control may be less specific than previously thought.

Acknowledgments: We thank Merlijn Witte for technical assistance with the *Xenopus laevis* egg fertilization experiments, Ariane Abrieu for a gift of CSF egg extracts, and Markus Raschle from the Technical University of Kaiserslautern for providing the *Xenopus laevis* protein database. **Funding:** AJRH and MA acknowledge support from the Horizon 2020 program INFRAIA project Epic-XS (Project 823839) and the NWO funded Netherlands Proteomics Centre through the National Road Map for Large-scale Infrastructures program X-Omics (Project 184.034.019) of the Netherlands Proteomics Centre. JMV is supported by scholarships from the Ministry of Science and Technology of Costa Rica (MICITT) and the University of Costa Rica (UCR). PK and MV are funded by the Oncode Institute which is financed by the Dutch Cancer Society and by the gravitation program CancerGenomiCs.nl from the Netherlands Organisation for Scientific Research (NWO). DF and LK are Inserm employees. GD is funded by the Institut National de Cancer, France (INCa) PRT-K programme (PRT-K17 n° 2018-023). The Fisher lab is funded by the Ligue Nationale Contre le Cancer, France (EL2018.LNCC/DF) and INCa (PLBIO18-094). The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05).

Author contributions: MA and DF conceived and supervised the project. JMV, PK and LK designed and interpreted experiments. JMV, HT, AH, GvM, LK and GD performed experiments and interpreted the data. MV supervised GvM. JMV, LK, GD, DF and MA wrote the paper.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data is available in the main text or the supplementary materials. All data, code, and materials are available on request.

Supplementary Materials:

Materials and Methods

Extended data figures 1-10

Data S1-S4

Movies S1, S2

References

1. Örd, M. *et al.* Multisite phosphorylation code of CDK. *Nat Struct Mol Biol* **26**, 649–658 (2019).
2. Krasinska, L. *et al.* Protein Phosphatase 2A Controls the Order and Dynamics of Cell-Cycle Transitions. *Mol Cell* **44**, 437–50 (2011).
3. Hyman, A. A. Whither systems biology. *Philos Trans R Soc Lond B Biol Sci* **366**, 3635–3637 (2011).
4. Errico, A., Deshmukh, K., Tanaka, Y., Pozniakovsky, A. & Hunt, T. Identification of substrates for cyclin dependent kinases. *Adv. Enzyme Regul.* **50**, 375–399 (2010).
5. Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859–64 (2003).
6. Chi, Y. *et al.* Identification of CDK2 substrates in human cell lysates. *Genome Biol* **9**, R149 (2008).
7. Blethrow, J. D., Glavy, J. S., Morgan, D. O. & Shokat, K. M. Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci U S A* **105**, 1442–7 (2008).
8. Holt, L. J. *et al.* Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682–6 (2009).
9. Michowski, W. *et al.* Cdk1 Controls Global Epigenetic Landscape in Embryonic Stem Cells. *Mol Cell* **78**, 459-476.e13 (2020).
10. Mahdessian, D. *et al.* Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* **590**, 649–654 (2021).
11. Orlicky, S., Tang, X., Willems, A., Tyers, M. & Sicheri, F. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* **112**, 243–256 (2003).
12. Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3 (2010).
13. Cooper, S. The synchronization manifesto: a critique of whole-culture synchronization. *FEBS J* **286**, 4650–4656 (2019).
14. Ly, T., Endo, A. & Lamond, A. I. Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* **4**, e04534 (2015).
15. Ly, T. *et al.* A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* **3**, e01630 (2014).
16. Ly, T. *et al.* Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *eLife* **6**, e27574 (2017).

17. Purvis, J. E. & Lahav, G. Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956 (2013).
18. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
19. Lombard-Banek, C., Moody, S. A., Manzini, M. C. & Nemes, P. Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-Cell Proteomics in Complex Tissues: Developing Cell Clones in Live Xenopus laevis and Zebrafish Embryos. *Anal Chem* **91**, 4797–4805 (2019).
20. Newport, J. & Kirschner, M. A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686 (1982).
21. Newport, J. W. & Kirschner, M. W. Regulation of the cell cycle during early Xenopus development. *Cell* **37**, 731–42 (1984).
22. Post, H. *et al.* Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J. Proteome Res.* **16**, 728–737 (2017).
23. Clift, D. & Schuh, M. Restarting life: fertilization and the transition from meiosis to mitosis. *Nat Rev Mol Cell Biol* **14**, 549–562 (2013).
24. Presler, M. *et al.* Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the Xenopus egg. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10838–E10847 (2017).
25. Peuchen, E. H. *et al.* Phosphorylation Dynamics Dominate the Regulated Proteome during Early Xenopus Development. *Sci Rep* **7**, 15647 (2017).
26. Ferrell, J. E., Jr., Wu, M., Gerhart, J. C. & Martin, G. S. Cell cycle tyrosine phosphorylation of p34cdc2 and a microtubule-associated protein kinase homolog in Xenopus oocytes and eggs. *Mol Cell Biol* **11**, 1965–71 (1991).
27. Lawrence, R. T., Searle, B. C., Llovet, A. & Villén, J. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat Methods* **13**, 431–434 (2016).
28. Schmidlin, T. *et al.* Assessment of SRM, MRM3, and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *PROTEOMICS* **16**, 2193–2205 (2016).
29. Schmidlin, T. *et al.* High-Throughput Assessment of Kinome-wide Activation States. *Cell Systems* **9**, 366-374.e5 (2019).

30. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics *. *Molecular & Cellular Proteomics* **11**, 1475–1488 (2012).
31. Tsai, T. Y.-C., Theriot, J. A. & Jr, J. E. F. Changes in Oscillatory Dynamics in the Cell Cycle of Early Xenopus laevis Embryos. *PLOS Biology* **12**, e1001788 (2014).
32. Goldbeter, A. & Koshland, D. E., Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A* **78**, 6840–4 (1981).
33. Suzuki, K. *et al.* Identification of non-Ser/Thr-Pro consensus motifs for Cdk1 and their roles in mitotic regulation of C2H2 zinc finger proteins and Ect2. *Scientific Reports* **5**, 7929 (2015).
34. Iakoucheva, L. M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).
35. Moses, A. M., Hériché, J.-K. & Durbin, R. Clustering of phosphorylation site recognition motifs can be exploited to predict the targets of cyclin-dependent kinase. *Genome Biol.* **8**, R23 (2007).
36. Rai, A. K., Chen, J.-X., Selbach, M. & Pelkmans, L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216 (2018).
37. Shimobayashi, S. F., Ronceray, P., Sanders, D. W., Haataja, M. P. & Brangwynne, C. P. Nucleation landscape of biomolecular condensates. *Nature* **599**, 503–506 (2021).
38. Berchtold, D., Battich, N. & Pelkmans, L. A Systems-Level Study Reveals Regulators of Membrane-less Organelles in Human Cells. *Mol. Cell* **72**, 1035-1049.e5 (2018).
39. Hur, W. *et al.* CDK-Regulated Phase Separation Seeded by Histone Genes Ensures Precise Growth and Function of Histone Locus Bodies. *Dev Cell* **54**, 379-394.e6 (2020).
40. Lin, Y.-H., Wu, H., Jia, B., Zhang, M. & Chan, H. S. Assembly of model postsynaptic densities involves interactions auxiliary to stoichiometric binding. *Biophys J* **121**, 157–171 (2022).
41. Ghosh, K. Stoichiometric versus stochastic interaction in models of liquid-liquid phase separation. *Biophys J* **121**, 4–6 (2022).
42. van Mierlo, G. *et al.* Predicting protein condensate formation using machine learning. *Cell Rep* **34**, 108705 (2021).
43. Huihui, J. & Ghosh, K. An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *J Chem Phys* **152**, 161102 (2020).

44. Huihui, J. & Ghosh, K. Intrachain interaction topology can identify functionally similar intrinsically disordered proteins. *Biophys J* **120**, 1860–1868 (2021).
45. Sobecki, M. *et al.* The cell proliferation antigen Ki-67 organises heterochromatin. *Elife* **5**, e13722 (2016).
46. Booth, D. G. *et al.* Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *Elife* **3**, e01641 (2014).
47. Hayashi, Y., Kato, K. & Kimura, K. The hierarchical structure of the perichromosomal layer comprises Ki67, ribosomal RNAs, and nucleolar proteins. *Biochemical and Biophysical Research Communications* **493**, 1043–1049 (2017).
48. Hégarat, N. *et al.* Cyclin A triggers Mitosis either via the Greatwall kinase pathway or Cyclin B. *The EMBO Journal* **39**, e104419 (2020).
49. Shin, Y. *et al.* Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171.e14 (2017).
50. Feng, Z., Jia, B. & Zhang, M. Liquid–Liquid Phase Separation in Biology: Specific Stoichiometric Molecular Interactions vs Promiscuous Interactions Mediated by Disordered Sequences. *Biochemistry* **60**, 2397–2406 (2021).
51. Musacchio, A. On the role of phase separation in the biogenesis of membraneless compartments. *EMBO J* e109952 (2022) doi:10.15252/embj.2021109952.

Figures

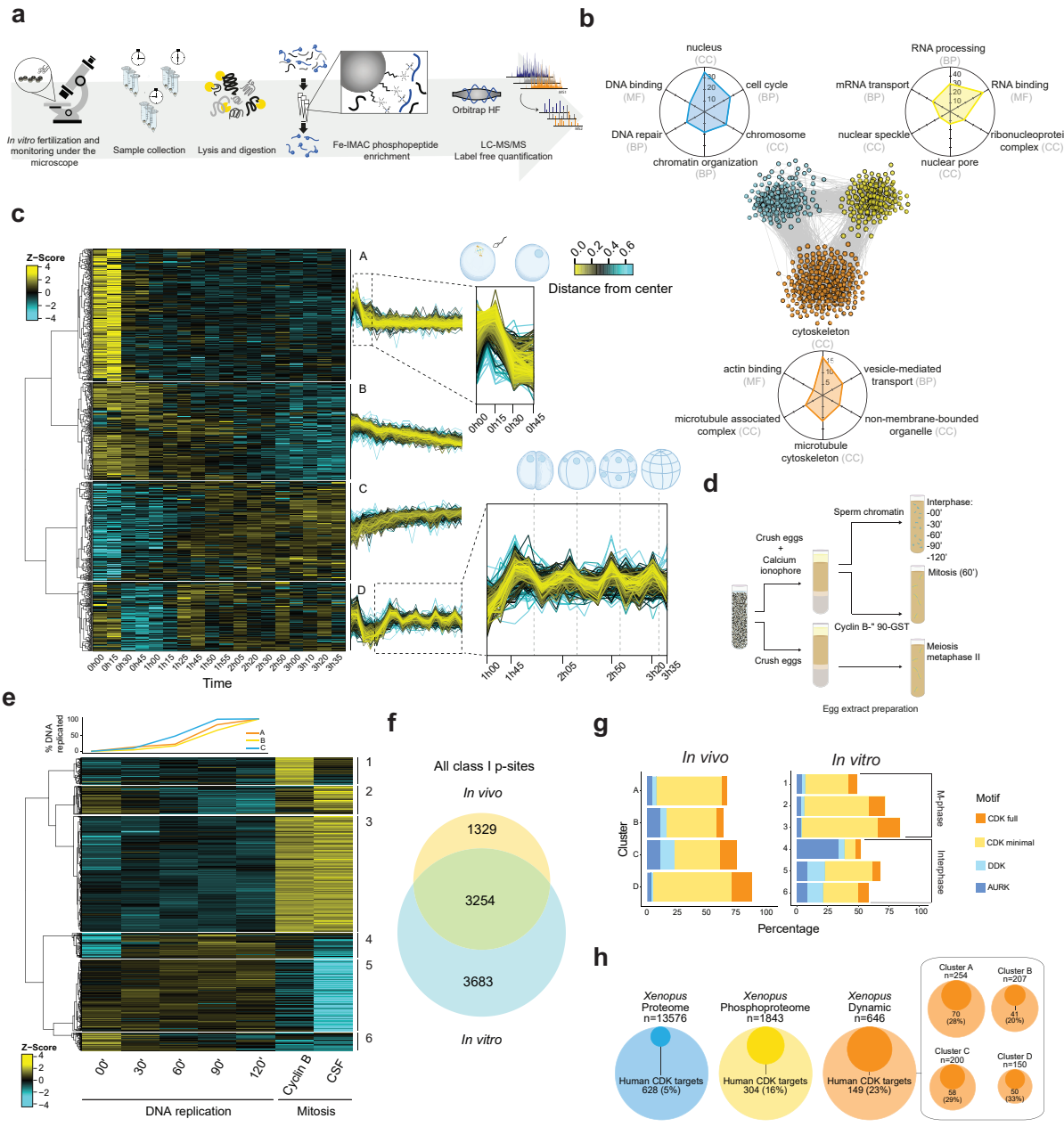


Figure 1. The time-resolved phosphoproteome from a single-cell to a 16-cell embryo and its cell cycle assignment. (a) Schematic representation of the workflow. Single *Xenopus* eggs and embryos were collected followed by cell lysis, protein digestion, phosphopeptide enrichment and high-resolution proteomics analysis. (b) STRING network of functionally associated proteins undergoing dynamic phosphorylation (each node represents a protein). Vicinity clustering reveals three main groups (yellow, blue and orange) with a high degree of association. Radar plots show the corresponding GO terms (adjusted p value <0.05) for each group (axes show -Log_{10}(adj p value) for each GO term). (c) Hierarchical clustering of

significantly changing phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05), reveals 4 clusters with distinct regulation (A-D). Dashed boxes in clusters A and D are zoomed-in to highlight dynamic phosphorylation patterns (dashed lines depict the time points of cell division). (d) Scheme of the experiment in the *Xenopus* egg extract. (e) Top: quantification of DNA replication in each biological replicate. Below: Hierarchical clustering of dynamic phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05) reveals differential regulation of phosphosites during S-phase and mitosis. (f) Overlap between *in vivo* (embryo) and *in vitro* (egg extract) phosphoproteomics. (g) Proportion of phosphosites according to their potential upstream kinase for each cluster in the *in vivo* (top) and *in vitro* (bottom) experiments. (h) Circle plots presenting enrichment of homologues of human CDK substrates among *Xenopus* phosphoproteins detected *in vivo* and those with dynamic phosphosites.

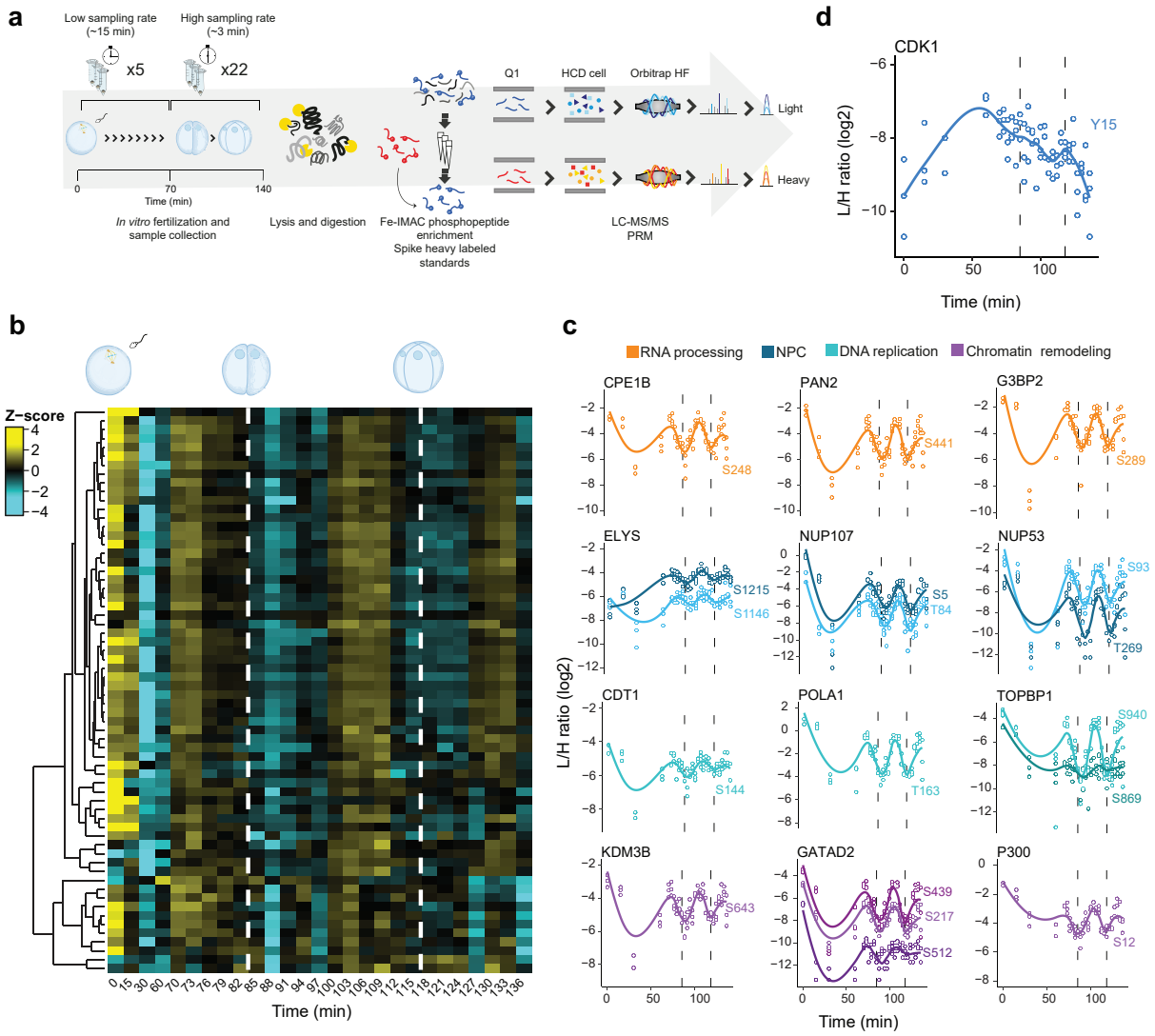


Figure 2. Switch-like mitotic phosphorylation *in vivo*. (a) Schematic representation of the workflow. Samples were collected over two cell divisions and enriched phosphopeptides were subjected to targeted proteomics analysis. (b) Heat map shows a highly synchronous wave of phosphorylation preceding each of the two cell divisions. Dashed lines depict times when cell divisions were recorded. (c) Single phosphosite plots from selected proteins. Each dot represents a biological replicate (n=3). Dashed lines depict times when cell divisions were recorded. (d) Single phosphosite plot of CDK1 inhibitory phosphorylation (Y15).

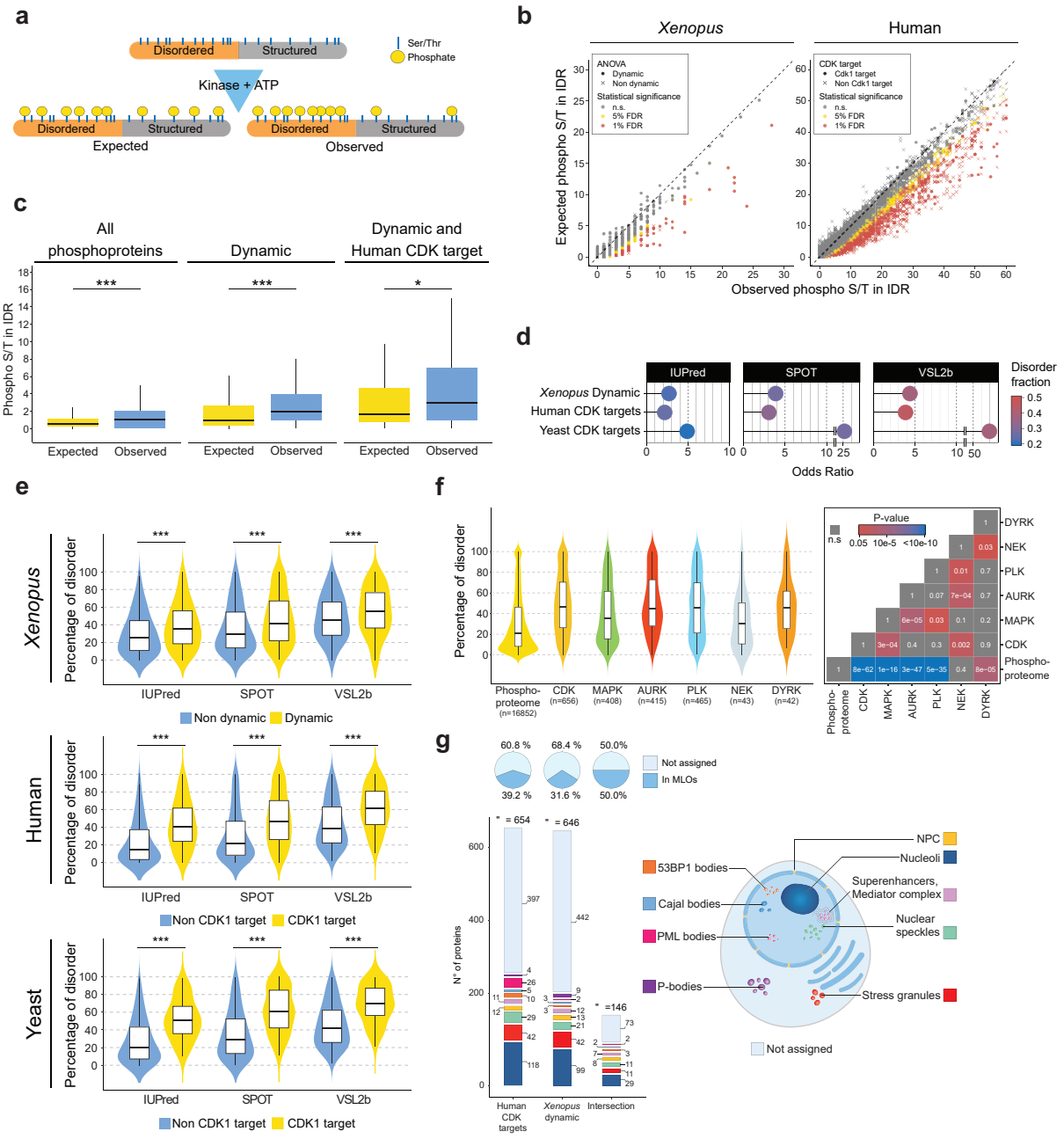


Fig. 3. The cell cycle phosphoproteome is characterised by intrinsic disorder and MLO components. (a) Scheme illustrating hypothetical enrichment of phosphorylation in disordered regions when taking into account amino acid compositional bias. (b) Scatter plot of expected vs observed phosphorylated Ser/Thr for each protein of human and *Xenopus* phosphoprotein datasets. FDR thresholds of 5% and 1% are marked in yellow and red respectively. Circles: proteins with at least one dynamic phosphorylation in *Xenopus*, or human CDK1 subfamily substrates, respectively. (c) Boxplots showing expected vs observed phosphorylated Ser/Thr among all phosphoproteins detected (left), phosphoproteins with at least one dynamic phosphosite (middle), and dynamic phosphoproteins also detected as CDK1 subfamily targets in humans (right). Distributions were compared with the Wilcoxon signed-rank test. *p<0.05,

**p<0.01, ***p<0.001. (d) Plots showing the common Odds Ratio of Ser/Thr phosphorylation in structured and ordered regions calculated with the Fisher's test (see Extended data Fig. 6b, c). For all organisms, the disordered regions were calculated with three different disorder predictors. The disordered fraction is presented in a colour scale. (e) Violin plots of the distribution of disordered residues per protein for CDK targets vs the rest of the phosphoproteome for human and yeast, and dynamic phosphoproteins vs the rest of the phosphoproteome for *Xenopus*. Intrinsic disorder was calculated with three different predictors (IUPred, SPOT, and VSL2b). Statistical significance was evaluated with the Wilcoxon–Mann–Whitney test; ***p<0.001. (f) Violin plot (left) showing the distribution of disordered residues per protein for CDK, MAPK, Aurora, PLK, NEK and DYRK kinase targets vs the rest of the phosphoproteome for human targets. Statistical significance was assessed by Kruskal-Wallis ANOVA, and pairwise comparisons were performed with Dunn's post-hoc tests. The adjusted p-values (Benjamini-Hochberg) are shown in a tile plot (right). (g) Human CDK1 subfamily targets, *Xenopus* dynamic phosphoproteins, and the intersection of both sets, that are present in our manually curated proteome of membraneless organelles.

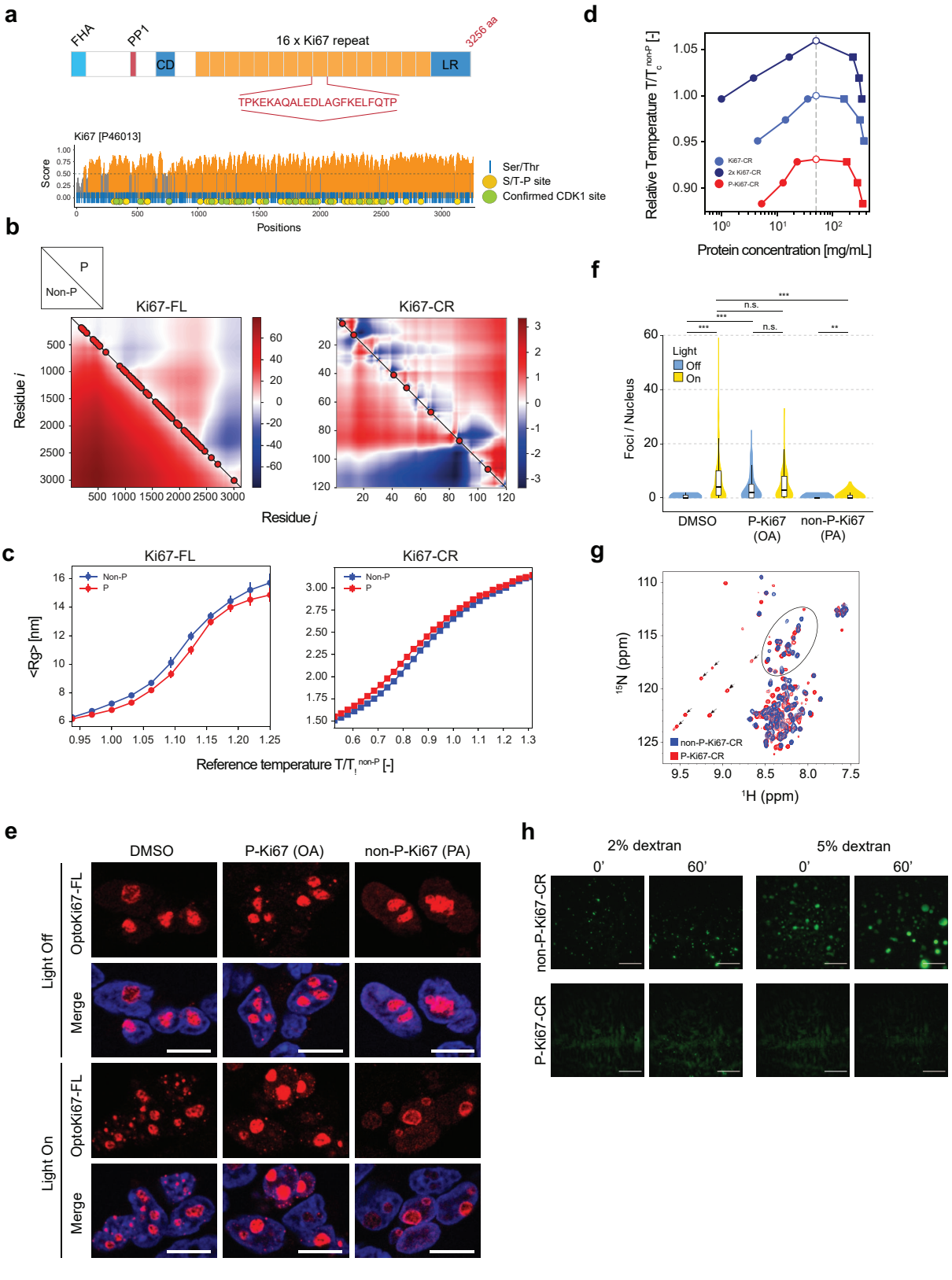


Fig. 4. CDK-mediated phosphorylation regulates phase separation of a model IDP. (a) Top, scheme of the human Ki-67 protein (FHA, forkhead-associated domain; PP1, PP1 phosphatase-binding domain; CD, conserved domain; LR, leucine arginine-rich domain).

Highlighted, Ki-67 repeat consensus motif. Bottom, diagram of IUPred score over the length of human Ki-67. Regions with scores >0.5 (orange) are considered to be disordered, and <0.5 (grey) structured. Blue vertical lines indicate Ser and Thr residues; yellow circles, known Ser/Thr-Pro phosphosites; green circles, confirmed CDK1 subfamily phosphorylations. (b) Sequence Charge Decoration Matrix (SCDM) maps for full length Ki-67 (FL, left) and Ki-67 consensus repeat (CR, right), depicting the contribution of electrostatic interaction dictating the distance between two amino acid residues i and j (shown in x and y axes). The values of SCDM for different residue pairs (i,j) are shown using colour schemes with red and blue denoting positive (repulsive) and negative (attractive) values, respectively. The lower and upper triangles indicate SCDM map for the unphosphorylated (non-P) and phosphorylated (P) sequences, respectively. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are indicated with red circles. (c) Dependency of the radius of gyration (Rg) on the simulation temperature in single-chain MD simulations for full chain Ki-67 (left) and consensus repeat (right). The reference temperature is the θ temperature of the non-phosphorylated molecule for full chain and consensus repeat, respectively. Reported error bars are obtained by block analysis over 10 blocks. (d) Binodal curves from phase coexistence simulations of the Ki-67 consensus repeat sequence. For each temperature, filled circles indicate the dilute phase density and squares indicate the coexisting dense phase density. Empty circles indicate the fitted critical temperature (T_c) of each system. The T_c of the non-phosphorylated monomer (light blue empty circle) was the reference for the normalisation of the temperature values. The light gray dashed line indicates the total concentration used in the simulations. The reference temperature is the θ temperature of the non-phosphorylated molecule for full chain and consensus repeat, respectively. Reported error bars are obtained by block analysis over 10 blocks. (e) Representative fluorescent images of HEK-293 cells expressing opto-Ki-67 (FL) construct before (Light Off) and after (Light On) exposure to blue light. Cells were pretreated for 1h with either vehicle (DMSO), 0.5 μM okadaic acid (OA), to inhibit protein phosphatase 2, or 5 μM purvalanol A (PA), to inhibit CDKs. DNA was stained with Hoechst 33258; scale bars, 10μm. (f) Violin plot presenting quantification of results from (e); the number of foci per nucleus was counted. Statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise *post-hoc* comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method. (g) Overlaid NMR ^{1}H-^{15}N HSQC of unphosphorylated (blue) and CDK-phosphorylated (red) GFP-tagged Ki-67 consensus repeat. Each cross-peak corresponds to one residue. The seven new deshielded cross peaks (highlighted by a black flag)

appearing above 8.5 ppm in ^{1}H correspond to phosphorylated serines or threonines (^{1}H downfield chemical shift perturbation on phosphorylated Ser/Thr residues due to phosphate electronegativity). Non phosphorylated Ser/Thr residues are surrounded by a black oval. (h) Representative fluorescence images of *in vitro* phase separation assay with purified GFP-tagged Ki-67 consensus repeat (CR), non-phosphorylated (non-P) or *in vitro* phosphorylated with recombinant CDK1-cyclin B-CKS1 (P), at indicated dextran concentrations and time points; scale bars, 10μm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- DATAS1Phosphoproteomicsdata.xlsx
- DATAS2CDKphosphorylationdata.xlsx
- DATAS3Proteindisorderprediction.xlsx
- DATAS4HumanproteinsinMLOs.xlsx
- SupplementarymovieS1.mpg
- SupplementarymovieS2A.mpg
- SupplementarymovieS2B.mpg

Appendix 2

Ki-67 regulates global gene expression and promotes sequential stages of carcinogenesis

Ki-67 regulates global gene expression and promotes sequential stages of carcinogenesis

Karim Mrouj[a,b,1], Nuria Andrés-Sánchez[a,b], Geronimo Dubra[a,b], Priyanka Singh[a,b,2], Michal Sobecki[a,b,3], Dhanvantri Chahar[a,b], Emile Al Ghoul[a,b], Ana Bella Aznar[a,b], Susana Prieto[a,b], Nelly Pirot[c,d], Florence Bernex[c,d], Benoit Bordignon[e], Cedric Hassen-Khodja[e], Martin Villalba[f], Liliana Krasinska[a,b], and Daniel Fisher[a,b,4]

[a]Institut de Génétique Moléculaire de Montpellier, CNRS, INSERM, University of Montpellier, 34293 Montpellier, France; [b]Equipe Labellisée Ligue 2018, Ligue Nationale Contre le Cancer, 75013 Paris, France; [c]Institut de Recherche en Cancérologie de Montpellier, Institut du Cancer de Montpellier, INSERM, University of Montpellier, 34090 Montpellier, France; [d]BioCampus, Le Réseau d'Histologie Expérimentale de Montpellier, CNRS, INSERM, University of Montpellier, 34090 Montpellier, France; [e]BioCampus, Montpellier Ressources Imagerie, CNRS, INSERM, University of Montpellier, 34090 Montpellier, France; and [f]Institute of Regenerative Medicine and Biotherapies, INSERM, University of Montpellier, 34295 Montpellier, France

Edited by Anton Berns, Netherlands Cancer Institute, Amsterdam, The Netherlands, and approved January 28, 2021 (received for review December 29, 2020)

Ki-67 is a nuclear protein that is expressed in all proliferating vertebrate cells. Here, we demonstrate that, although Ki-67 is not required for cell proliferation, its genetic ablation inhibits each step of tumor initiation, growth, and metastasis. Mice lacking Ki-67 are resistant to chemical or genetic induction of intestinal tumorigenesis. In established cancer cells, Ki-67 knockout causes global transcriptome remodeling that alters the epithelial–mesenchymal balance and suppresses stem cell characteristics. When grafted into mice, tumor growth is slowed, and metastasis is abrogated, despite normal cell proliferation rates. Yet, Ki-67 loss also down-regulates major histocompatibility complex class I antigen presentation and, in the 4T1 syngeneic model of mammary carcinoma, leads to an immune-suppressive environment that prevents the early phase of tumor regression. Finally, genes involved in xenobiotic metabolism are down-regulated, and cells are sensitized to various drug classes. Our results suggest that Ki-67 enables transcriptional programs required for cellular adaptation to the environment. This facilitates multiple steps of carcinogenesis and drug resistance, yet may render cancer cells more susceptible to antitumor immune responses.

Ki-67 | cancer | genetically modified mice | transcription

Ki-67 is a nuclear protein expressed only in proliferating vertebrate cells, a property underlying its widespread use in oncology as a biomarker (1). Its expression is routinely assessed in histopathology to grade tumors; there are also indications for its use as a prognostic marker (2), although uncertainty over the relationship between Ki-67 index and prognosis remains. The cellular functions of Ki-67 are not well understood, and whether it is involved in tumorigenesis is unclear.

For a long time, Ki-67 was thought to be required for cell proliferation (3–8), and early work suggested that it promotes ribosomal RNA transcription (4, 9). However, recent genetic studies have shown that despite promoting formation of the perichromosomal layer of mitotic chromosomes (9–12), it is not required for cell proliferation (10–13). It is also dispensable for ribosomal RNA synthesis and processing, and mice lacking Ki-67 develop and age normally (11). In addition, Ki-67 is not overexpressed in cancers; rather, Ki-67 expression is controlled by cell cycle regulators, including cyclin-dependent kinases (CDKs), the activating subunit of the ubiquitin ligase APC/C-CDH1, which is required for destruction of mitotic cyclins, and the cell cycle transcription factor B-Myb (11, 14, 15). These recent studies raise the question of whether Ki-67 plays any role in tumorigenesis, which has not been addressed genetically.

Despite not being essential for cell proliferation, Ki-67 might be important in carcinogenesis for other reasons. We previously identified over 50 Ki-67–interacting proteins that are involved in transcription and chromatin regulation. We also found that Ki-67 organizes heterochromatin: in NIH/3T3 mouse embyonic fibroblast cells with transcription activator-like effector nucleases (TALEN)-mediated biallelic disruption of the Ki-67 gene or human cancer cells with stable knockdown of Ki-67, repressive histone marks histone H3 lysine-9 tri-methylation (H3K9me3) and histone H4 lysine-20 tri-methylation (H4K20me3) are dispersed, the heterochromatin is less compact, and association between centromeres and nucleoli is disrupted. Conversely, overexpression of Ki-67 caused ectopic heterochromatin formation (11). Involvement of Ki-67 in chromatin organization was corroborated by a study showing that Ki-67 is required to maintain heterochromatin marks at inactive X chromosomes in nontransformed cells (16). Heterochromatin is a phenotypic marker of multiple cancers (17), suggesting that it might be involved in carcinogenesis. We thus tested whether and how Ki-67 is required for different steps of carcinogenesis.

Results

Ki-67 Is Dispensable in Cancer Cell Lines yet Is Rarely Mutated in Human Cancers. Ki-67 expression is widely used as a marker for cell proliferation in cancer, but whether it is important for

Significance

Ki-67 is a nuclear protein present in all proliferating vertebrate cells and is widely used as a marker in clinical cancer histopathology. However, its cellular functions have remained largely mysterious, and whether it plays any roles in cancer was unknown. In this work, we show genetically that Ki-67 is not required for cell proliferation in tumors, but it is required for all stages of carcinogenesis. The effects on cell transformation, tumor growth, metastasis, and drug sensitivity correlate with genome-scale changes in gene expression that modify cellular programs implicated in cancer. Thus, Ki-67 expression is advantageous for cancer cells, but it comes with an Achilles' heel: it makes them more visible to the immune system.

Author contributions: K.M., L.K., and D.F. designed research; K.M., N.A.-S., P.S., M.S., D.C., E.A.G., A.B.A., S.P., N.P., F.B., B.B., C.H.-K., and M.V. performed research; K.M., N.A.-S., G.D., P.S., M.S., D.C., E.A.G., N.P., F.B., B.B., C.H.-K., M.V., L.K., and D.F. analyzed data; and L.K. and D.F. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Present address: Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305.

[2]Present address: Section on Nutrient Control of Gene Expression, National Institute of Child Health and Human Development, National Institute of Health, Bethesda, MD 20892.

[3]Present address: Institute of Anatomy, University of Zurich, CH - 8057 Zurich, Switzerland.

[4]To whom correspondence may be addressed. Email: daniel.fisher@igmm.cnrs.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2026507118/-/DCSupplemental.

Published March 3, 2021.

carcinogenesis is unclear. We previously found no adverse effects on cell proliferation of either shRNA-mediated knockdown of Ki-67 in U2OS or HeLa cells or TALEN-mediated disruption of the gene encoding Ki-67 in NIH/3T3 cells (11). To see whether this is true across different cancer types, we interrogated the Cancer Dependency Map project Dependency Mapper (DepMap) (https://depmap.org/portal/) (18) using the latest CRISPR (Avana 20Q1) and RNA interference (RNAi; Broad Institute) datasets. As controls, we compared Ki-67 with two proto-oncogenes that have roles in cell proliferation, *MYC* and *CCND1* encoding c-Myc and cyclin D1, respectively, and with *PCNA*, encoding Proliferating Cell Nuclear Antigen, whose expression profile is similar to that of Ki-67.

Ki-67, PCNA, and c-Myc were universally expressed, while cyclin D1 was expressed in most cell lines, although not all (Fig. 1*A*), possibly due to compensation by a different D-type cyclin. In CRISPR-Cas9 screens, PCNA was essential for proliferation of 775 of 777 cell lines tested, while 719 of 726 lines required c-Myc, and 579 of 726 required cyclin D1 (Fig. 1*B*). In contrast, Ki-67 knockout did not affect cell proliferation in 725 of 739 cell lines (Fig. 1*B*), confirming that it is generally dispensable for cell proliferation in human cancer cells. Nevertheless, like *PCNA*, *MKI67* showed almost no copy number variations among the different cell lines, in contrast to the two cancer driver oncogenes *MYC* and *CCND1*, which were frequently amplified (Fig. 1*C*). The same is true in data from clinical samples of cancer patients in the cBioPortal database (https://www.cbioportal.org/), which further showed that only 5% of cancers presented mutations in *MKI67* (Fig. 1*D*). Almost all of these were missense mutations, which might be passenger mutations.

Thus, even though Ki-67 is dispensable for cell proliferation in virtually all cancer cells, it is ubiquitously expressed at similar levels in all cancers, corroborating our previous findings that variability in Ki-67 expression is accounted for by its regulation through the cell cycle (14). Together, these observations suggest that Ki-67 may provide benefit to cancer cells. In contrast, its overexpression may be counterselected in cancers, which would fit with our finding that increased levels of Ki-67 arrest cell proliferation (11).

The lack of overexpression or deletion of Ki-67 in cancers implies that correlating Ki-67 expression with patient survival is

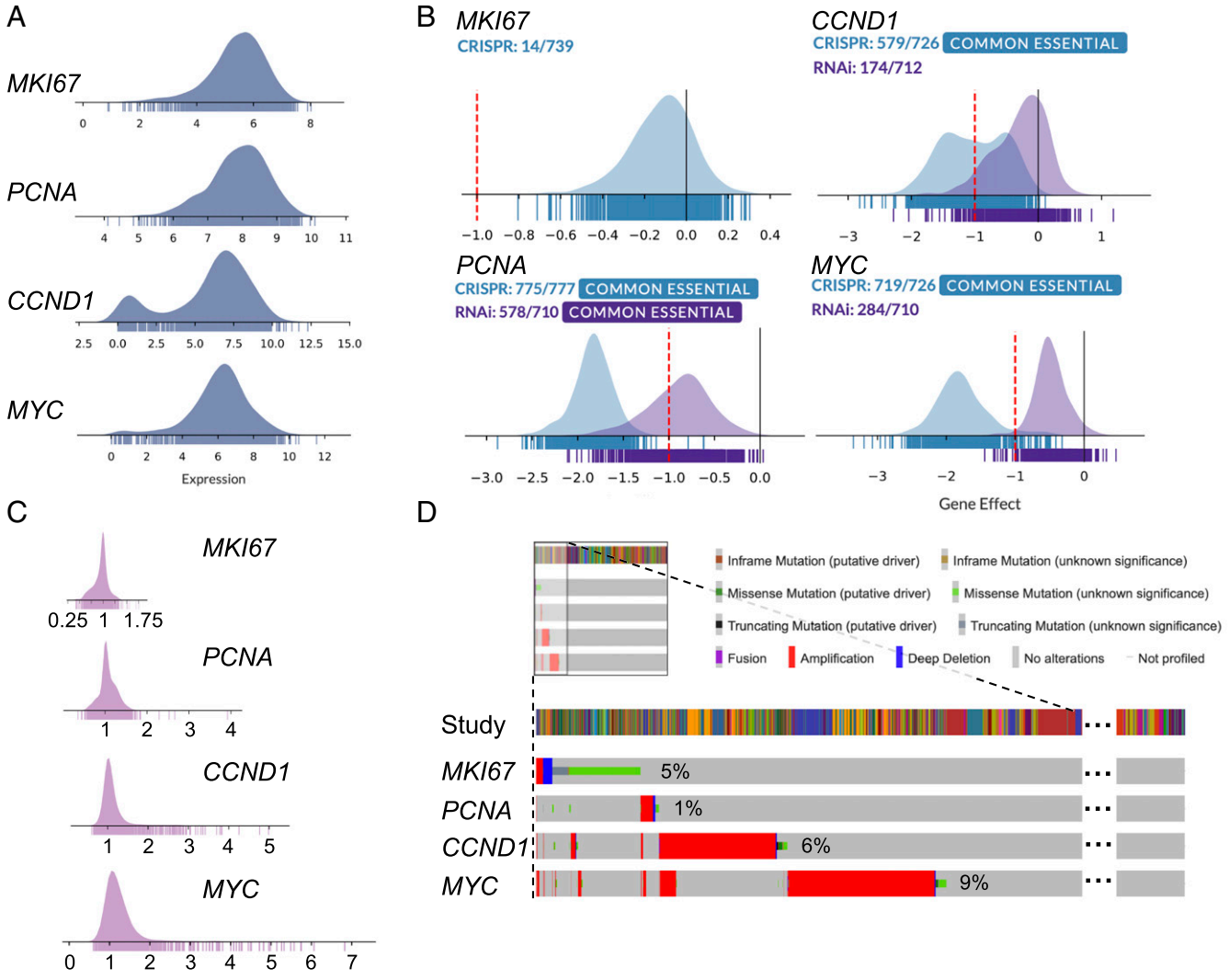


Fig. 1. Ki-67 is dispensable for human cancer cells but is rarely mutated. (*A*) Expression levels of Ki-67 (MKI67), PCNA, cyclin D1 (CCND1), and c-Myc (MYC) in cell lines from the Cancer Cell Line Encyclopedia. Values are $\log_2$(transcripts per kilobase million [TPM] + 1) of TPM-normalized RNA-Seq data. (*B*) Relative copy number values for each gene from Sanger and Broad Institute whole-exome sequencing data of cancer cell lines. (*C*) DepMap analysis of dependency of cancer cell lines on each gene, from genome-wide CRISPR (blue) and RNAi (violet) screens. Lower values of gene effect indicate higher likelihood of dependency in a given cell line. Zero corresponds to a nonessential gene; −1 is the median of all common essential genes. (*D*) cBioPortal mutation analysis of each gene in pancancer TCGA data.

likely to simply reflect the impact of the fraction of proliferating cells. If so, then PCNA expression should give similar results. To address this question, we queried survival correlations in different cancer types with *MKI67* and *PCNA* messenger RNA levels in The Cancer Genome Atlas expression data using OncoLnc (19). For both genes, there was either no correlation (breast and colorectal cancer), a modest positive correlation (lung cancer), or a strong negative correlation (liver cancer and renal cancer) of expression levels with survival (*SI Appendix*, Fig. S1). Thus, in several different cancers, despite the contrasting requirements for PCNA and Ki-67 for cell proliferation, the correlation of expression of each gene with survival is very similar.

Mice Lacking Ki-67 Are Resistant to Intestinal Tumorigenesis. The fact that Ki-67 is ubiquitously expressed in cancers but is not required for cell proliferation raises the question of whether it has functional roles in carcinogenesis. To see whether Ki-67 knockout affects initiation of tumorigenesis in vivo, we used a germline TALEN disruption of Ki-67 ($Mki67^{2ntΔ/2ntΔ}$) that we generated (11). We first employed chemical induction of colon carcinogenesis by azoxymethane/dextran sodium sulfate (AOM-DSS) treatment (20) in wild-type (WT) and $Mki67^{2ntΔ/2ntΔ}$ mice. We observed that dextran sodium sulfate (DSS) alone induced a similar decrease in body weight in control and mutated mice compared with the controls (*SI Appendix*, Fig. S2*A*). Histopathological examination of colonic sections revealed a typical DSS-induced colitis in both genotypes (*SI Appendix*, Fig. S2*B*), with increased numbers of immune cells in the lamina propria, moderate crypt cells damage, and hyperplasia. As expected, AOM-DSS efficiently induced colon tumors within 16 wk in both WT and $Mki67^{2ntΔ/+}$ mice (indeed, heterozygous mice had bigger lesions; future studies will address effects of Ki-67 gene dosage on tumor growth). However, no macroscopic lesions were observed in $Mki67^{2ntΔ/2ntΔ}$ mice (Fig. 2 *A* and *B*). This suggests that Ki-67 is specifically required for initiation of tumorigenesis. To confirm these findings, we used a genetic model of intestinal tumorigenesis. We crossed $Mki67^{2ntΔ/2ntΔ}$ mice with $Apc^{Δ14/+}$ mice, which rapidly develop tumors in the intestine due to loss of the second allele of the *Apc* tumor suppressor gene (21). While as expected, $Apc^{Δ14/+}$ $Mki67^{2ntΔ/+}$ mice formed multiple colon tumors, tumor burden was strongly reduced in $Apc^{Δ14/+}$ $Mki67^{2ntΔ/2ntΔ}$ mice (Fig. 2 *C* and *D*).

Thus, Ki-67 is required for efficient initiation of tumorigenesis induced by chemical mutagenesis or loss of a tumor suppressor in vivo, suggesting that it might be required for cell transformation. To test this, we transduced our previously generated $Mki67^{+/+}$ or TALEN-mutated $Mki67^{-/-}$ NIH/3T3 fibroblasts (11) with oncogenic mutant H-Ras (G12V), and evaluated colony formation as an indicator of transformation. While $H\text{-}Ras^{G12V}$-transduced control 3T3 cells efficiently formed colonies, $Mki67^{-/-}$ 3T3 cells did not, despite having similar rates of cell proliferation (*SI Appendix*, Fig. S3). This indicates that Ki-67 expression facilitates oncogene-induced transformation in these cells.

Loss of Ki-67 Causes Global Transcriptome Changes and Deregulates Pathways Involved in Cancer. Since we previously found that knockdown of Ki-67 in cancer cells altered their chromatin organization and affected gene expression (11), we hypothesized that the resistance of Ki-67 knockout cells to transformation might also result from gene expression changes. We first performed RNA-sequencing (RNA-seq) analysis of Ki-67 WT and knockout NIH/3T3 cells. This revealed surprisingly wide-ranging effects of Ki-67 loss on the transcriptome, with 2,558 genes significantly deregulated in independent clones of $Mki67^{-/-}$ cells ($q < 0.05$) (Fig. 3*A* and Dataset S1). This level of transcriptome alteration suggested a global effect on chromatin rather than a direct involvement of Ki-67 in controlling specific pathways or transcription factors, which is consistent with our previous finding that Ki-67 interacts with many general chromatin regulators in the U2OS cancer cell line (11). We therefore expected that Ki-67 knockout would also extensively affect the transcriptome of established cancer cells, with possible consequences for tumorigenicity. To investigate this, we used the syngeneic 4T1 mouse mammary carcinoma model, which is derived from BALB/c mice. This cell line mimics human triple-negative breast cancer, is highly invasive, and spontaneously metastasizes to distant organs (22, 23). As expected, 4T1 cell proliferation rates were unaffected by CRISPR-Cas9–mediated *Mki67* gene knockout (*SI Appendix*, Fig. S4). In 4T1 cells, *Mki67* knockout caused even more extensive gene expression alterations than in NIH/3T3 cells: 4,979 genes were deregulated, of which 1,239 and 585 genes were more than twofold down-regulated and up-regulated, respectively (Fig. 3 *B* and *C* and Dataset S2). There was little overlap in the deregulated genes between $Mki67^{-/-}$ 4T1 (epithelial) and NIH/3T3 (mesenchymal) cells (Fig. 3*C* and *SI Appendix*, Tables S1 and S2) in accordance with our hypothesis that, by organizing chromatin, Ki-67 enables global gene regulation in different cell types rather than directly controlling specific genes.

To see whether the global effect of Ki-67 knockout on gene expression is conserved across cancer cell types and species, we next disrupted the *MKI67* gene by CRISPR-Cas9 in human MDA-MB-231 triple-negative breast cancer cells (*SI Appendix*, Fig. S5 *A* and *B*), which is a highly mesenchymal-like cell line due to an extensive epithelial–mesenchymal transition (EMT). As expected, $MKI67^{-/-}$ MDA-MB-231 cells proliferated normally in vitro (*SI Appendix*, Fig. S5 *C* and *D*). Transcriptome analysis by RNA-Seq showed that Ki-67 knockout in this cell line also caused genome-scale alterations in gene expression (Fig. 3*D*), with 9,127 genes deregulated, 914 of which were up- or down-regulated by a factor of more than two (Dataset S3).

We investigated whether the extensive transcriptome changes seen in cancer cells upon Ki-67 knockout affect pathways involved in tumorigenesis. In 4T1 cells, bioinformatic analysis of the most up- and down-regulated genes revealed deregulation of various components of inflammation, apoptosis, p53, the EMT, estrogen response, K-Ras signaling, and hypoxia (Fig. 3*E*). We also manually analyzed transcriptome data and noticed up-regulation of the Notch pathway, down-regulation of the EMT, the Wnt pathway, antigen presentation, and aldehyde metabolism, which we validated by qRT-PCR (Fig. 3*F*). Down-regulated genes were enriched in targets of nuclear factor erythroid 2-related factor 2, one of the major orchestrators of responses to oxidative stress; polycomb-repression complex 2 (PRC2), which mediates Histone H3 lysine-27 trimethylation (H3K27me3) and is a well-characterized regulator of the EMT (24, 25); the pluripotency factors Nanog and Sox2; and interferon regulatory factor 8 (Fig. 3*G*). All of these pathways have previously been implicated in tumorigenesis. In MDA-MB-231 cells, like 4T1 cells, pathway analysis also revealed genes involved in the EMT, inflammatory response, early estrogen response, K-RAS signaling, and hypoxia, while a significant portion of the deregulated genes was under the control of PRC2 and estrogen receptor 1 (Fig. 3 *H* and *I* and Dataset S3). In summary, similar pathways involved in cancer are affected upon Ki-67 knockout in different cancer cell lines.

The prevalence of down-regulation of gene expression in Ki-67 knockout cells, the enrichment in PRC2 targets among these genes, and our previous observations that Ki-67 associates with the essential PRC2 component SUZ12 (11) prompted us to ask whether loss of Ki-67 affects genome-wide distribution of H3K27me3. To answer this question, we performed chromatin immunoprecipitation with high-throughput sequencing (ChIP-Seq) on WT and Ki-67 knockout 4T1 cells. While there were no genome-wide changes in H3K27me3 distribution, a substantial subfraction of genes showed an increase in this mark

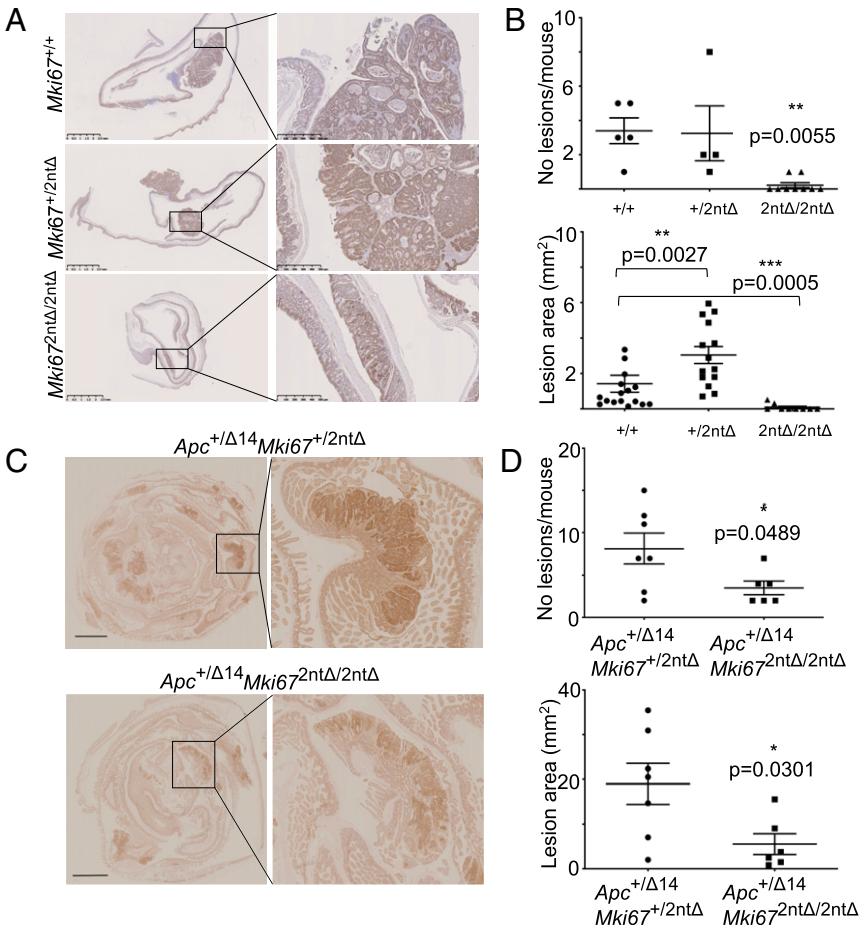


Fig. 2. Germline disruption of Ki-67 protects mice against intestinal tumorigenesis. (*A* and *C*) immunohistochemistry staining of β-catenin in whole intestines from 6- to 7-mo-old mice. (*A*) WT, $Mki67^{+/2ntΔ}$ and $Mki67^{2ntΔ/2ntΔ}$ mice were treated with AOM-DSS and analyzed 16 wk later. Zoomed *Insets* show accumulation of β-catenin in nuclei. (Scale bars: *Left*, 2.5 mm; *Inset*, 500 μm.) (*C*) $Apc^{+/Δ14}Mki67^{+/2ntΔ}$ and $Apc^{+/Δ14}Mki67^{2ntΔ/2ntΔ}$ mice were analyzed after 6 mo (scale bars: 3 mm). (*B* and *D*) Quantification of the number (*Upper*) and total area (*Lower*) of neoplastic lesions. Error bars, SEM ($n = 7$ $Apc^{+/Δ14}Mki67^{+/2ntΔ}$ mice; $n = 6$ $Apc^{+/Δ14}Mki67^{2ntΔ/2ntΔ}$ mice). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

(*SI Appendix*, Fig. S6*A*). This was particularly evident for genes strongly repressed in Ki-67 knockout cells (Fig. 3 *J*, *Left*). To investigate correlations between changes in the levels of histone modifications and of gene expression, we assigned an average value of the repressive H3K9me3, as well as activatory H3K4me3 and histone H3 lysine-27 acetylation (H3K27ac) reads across 10 kb surrounding the transcription start site, and plotted the differences in these values between WT and Ki-67 knockout cells with gene expression changes. This showed a strong correlation between the change in gene expression, H3K27me3, and H3K4me3 (Fig. 3*K*), while H3K27ac levels only correlated well with the most highly down-regulated genes (*SI Appendix*, Fig. S6*B*). Of nine genes whose down-regulated expression in Ki-67 knockout cells we confirmed by qRT-PCR, we only found an obvious increase of H3K27me3 on the EMT-promoting transcription factor Twist1. This correlated with down-regulation of active promoter-associated H3K4me3 and H3K27ac (Fig. 3*L*). There was also a slight increase of H3K27me3 and reduction in H3K27ac at the Vimentin promoter (*SI Appendix*, Fig. S6*C*). Taken together, these results suggest that Ki-67 loss generally increases PRC2-mediated repressive histone marks at down-regulated genes, but most of the expression alterations resulting from Ki-67 knockout may not be directly due to modulation by PRC2. Instead, they are likely knock-on effects of altered expression of other transcriptional regulators that result in changes in active promoter-associated histone marks.

Ki-67 Promotes Stem Cell Characteristics and Controls the EMT in Mammary Carcinoma. We next investigated the biological consequences resulting from these global transcriptome alterations in cancer cells lacking Ki-67. Importantly, although the Notch pathway is oncogenic in T cell acute lymphoid leukemia, it can act as a tumor suppressor in specific cellular contexts (26), can block Wnt signaling (27, 28) (a driver of tumorigenesis, cell stemness, and the EMT), and induce drug resistance (29). We confirmed Notch pathway up-regulation at the protein level (Fig. 4*A*). While the EMT is closely associated with a stem-like state (30–32), the most stem-like states appear to show a hybrid expression of both epithelial and mesenchymal characteristics, with the most epithelial and mesenchymal cells losing stemness (33–35). We found that 4T1 cells express both E-cadherin and vimentin, suggesting a highly stem-like state (Fig. 4*B*), but Ki-67 knockout 4T1 cells had reduced expression of the mesenchymal marker vimentin and up-regulated E-cadherin (Fig. 4 *B* and *C*). To see whether this translates to a loss of stem-like character, we analyzed aldehyde dehydrogenase (ALDH) activity, which is a bona fide marker of stem and progenitor cells (36, 37). ALDH activity was strongly reduced in 4T1 Ki-67 knockout cells (Fig. 4*D*). Furthermore, the ability to form spheroids in the absence of adhesion to a surface, another characteristic of stem cells (30, 38), was also largely decreased (Fig. 4*E*).

To test whether repression of the EMT in Ki-67 knockouts depends on PRC2, we additionally disrupted PRC2 components

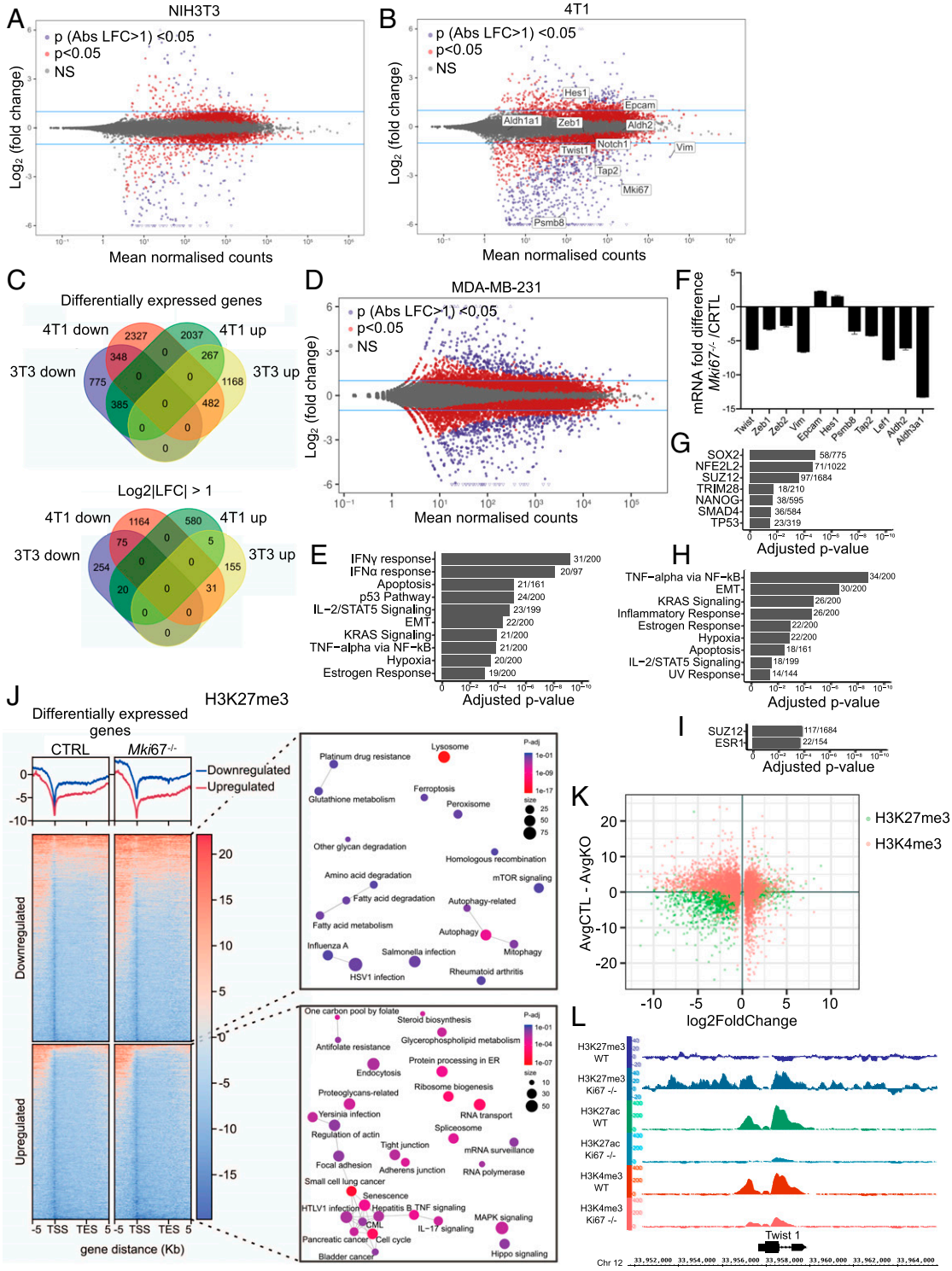


Fig. 3. Ki-67 ablation deregulates global gene expression programs in 4T1 cells. Dot plot analysis of differentially expressed genes (DEGs) in NIH/3T3 *Mki67*$^{-/-}$ cells (*A*), 4T1 *Mki67*$^{-/-}$ cells (*B*), and MDA-MB-231 *MKI67*$^{-/-}$ cells (*D*). Red dots: DEGs with *P* value < 0.05; purple dots: $\log_2$ fold change (LFC) >1 or <−1, *P* value < 0.05; gray dots: not significant (NS). (*C*) Venn diagrams of DEGs in NIH/3T3 and 4T1 *Mki67*$^{-/-}$ cells under condition of *P* value < 0.05 (*Upper*) and *P* value (LFC > 1 or < −1) < 0.05 (*Lower*). Gene set enrichment analysis of highly deregulated genes in 4T1 *Mki67*$^{-/-}$ cells (*E*) and MDA-MB-231 *MKI67*$^{-/-}$ cells (*H*). (*F*) qRT-PCR analysis of DEGs in 4T1 *Mki67*$^{-/-}$ cells; fold change in expression ± SD is shown. Gene set enrichment analysis of ENCODE and ChEA Consensus transcription factors from ChIP-X gene sets in *Mki67*$^{-/-}$ 4T1 cells (*G*) and MDA-MB-231 *MKI67*$^{-/-}$ cells (*I*). False discovery rate-adjusted *P* values. (*J*, *Left*) Heat maps of ChIP-Seq analysis of H3K27me3 in most down- and up-regulated genes in WT (CTRL) and *Mki67*$^{-/-}$ cells (TSS, transcription start site; TES, transcription end site). (*J*, *Right*) Gene set enrichment analysis associated with these genes (ER, estrogen receptor; HSV1, Herpex simplex virus 1; mTOR, mechanistic target of rapamycin; MAPK, mitogen-activated protein kinase; CML, chronic myeloid leukemia; HTLV1, human T cell lymphotropic virus type 1). (*K*) The average values of the H3K27me3 (in green) and H3K4me3 (in apricot) ChIP-Seq reads over the 10-kb region surrounding the gene in 4T1 *Mki67*$^{-/-}$ cells (AvgKO) were subtracted from the values of WT cells (AvgCTL) and then plotted against the log of the fold change for each gene in RNA-Seq. (*L*) ChIP-Seq profiles of histone marks at the Twist1 locus.

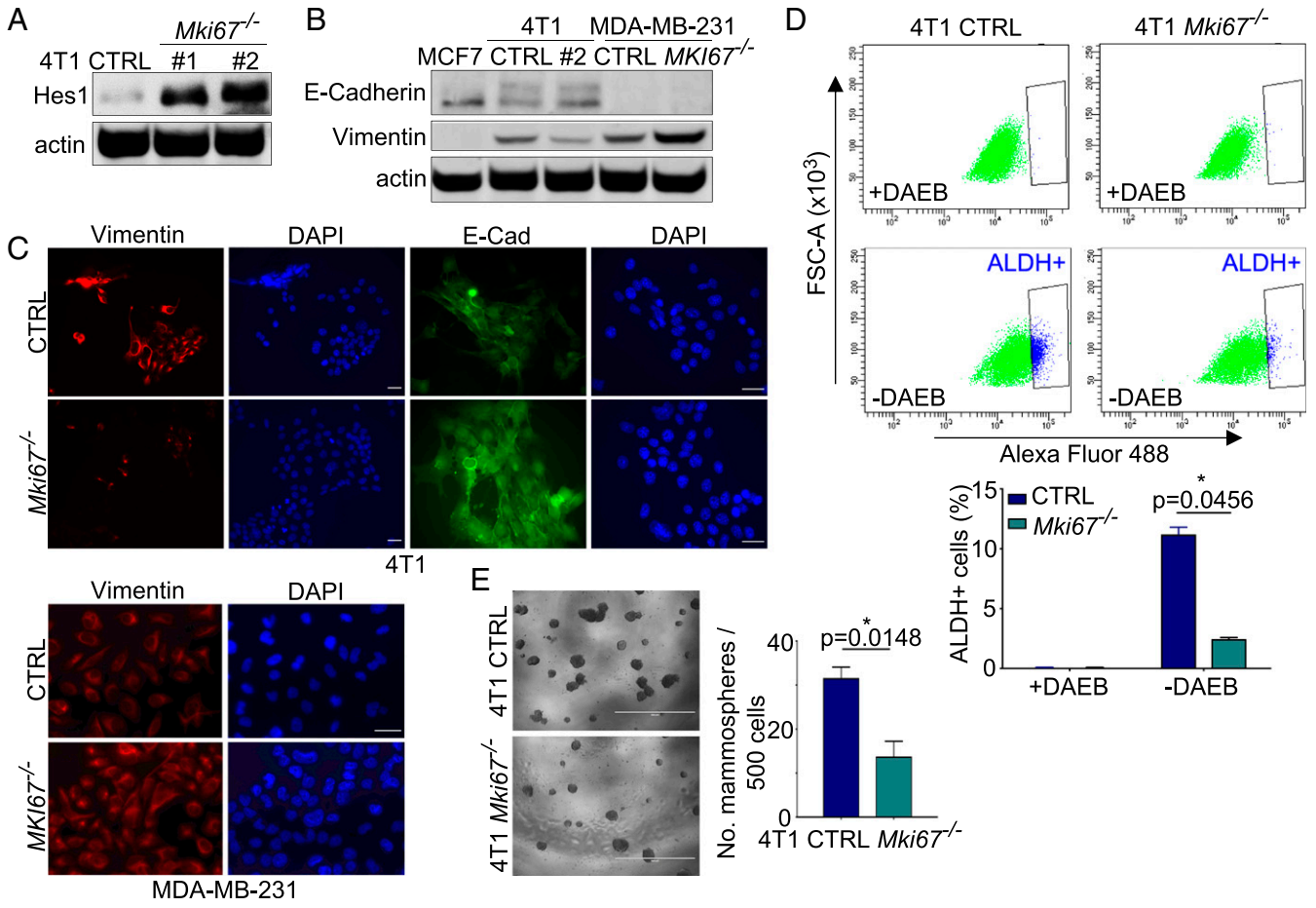


Fig. 4. Ki-67 promotes stem cell characteristics and controls the EMT in mammary carcinoma. (*A*) Immunoblot analysis of Hes1 expression in CTRL and two different clones (#1 and #2) of *Mki67*$^{-/-}$ 4T1 cells. Immunoblot (*B*) and immunofluorescence (*C*) analysis of vimentin and E-cadherin (E-Cad) expression in CTRL, *Mki67*$^{-/-}$ 4T1, and MDA-MB-231 cells. DNA was stained with DAPI. (Scale bars: 30 μm.) (*D*) ALDH activity measured using a flow cytometry assay in 4T1 CTRL and *Mki67*$^{-/-}$ cells. DAEB (*N*,*N*-diethylaminobenzaldehyde), an inhibitor of ALDH, was used as a negative control. (*Upper*) Flow cytometry profiles (FSC-A, forward scatter area). (*Lower*) Quantification of ALDH+ cells. Error bars, SEM ($n = 2$ independent analyses). (*E*) Mammosphere formation assay of 4T1 CTRL or *Mki67*$^{-/-}$ cells. Representative images (*Left*) and quantification (*Right*; error bars, SEM, $n = 10$) after 7 d. (Scale bars: 400 μm.)*$P < 0.05$.

Suz12 or *Ezh2* in WT and *Mki67*$^{-/-}$ 4T1 cells using CRISPR-Cas9 (*SI Appendix*, Fig. S7*A*). This partially rescued vimentin expression, indicative of an increased ability to undergo an EMT (*SI Appendix*, Fig. S7*B*), but did not restore it to levels observed in WT cells. This corroborates the results of the H3K27me3 ChIP-Seq analysis described above, indicating that the influence of PRC2 activity on gene expression changes observed in Ki-67 knockout cells is limited.

Analysis of EMT markers in MDA-MB-231 cells showed that Ki-67 knockout cells retained a mesenchymal character, with vimentin but no E-cadherin staining (Fig. 4 *B* and *C*), suggesting that they are far from a hybrid-like EMT state. Importantly, Ki-67 knockout cells had even higher vimentin expression, implying a further distance from the putative stem-like hybrid state. In agreement with this hypothesis, neither WT nor Ki-67 knockout MDA-MB-231 cells showed significant ALDH activity (*SI Appendix*, Fig. S8*A*). Importantly, Ki-67 knockout cells were unable to form mammospheres, indicating a conserved requirement for Ki-67 in maintaining ability to seed formation of new cell colonies (*SI Appendix*, Fig. S8*B*). As such, loss of Ki-67 reverses cells to a more epithelial or a more mesenchymal state, depending on the cell type of origin. This perturbation of the EMT correlates with a reduction of stem-like character in different cell lines.

Ki-67 Expression Promotes Tumor Growth. To determine whether these phenotypic alterations affect the tumorigenicity of cancer cells, we first engrafted WT and Ki-67 knockout 4T1 cells orthotopically into mouse mammary fat pads. Since Ki-67 knockout caused alteration of inflammatory response genes (Fig. 3*E*), we initially used athymic nude and NOD/SCID mice, allowing us to assess roles of Ki-67 in tumor growth and metastasis while minimizing possible confounding effects of an altered immune response. RNA-Seq of early-stage tumors from WT and Ki-67 mutant 4T1 cells grafted into nude mice showed that Ki-67–dependent transcriptome changes were well preserved in vivo (Fig. 5*A* and Dataset S4), including down-regulation of mesenchymal genes and up-regulation of epithelial genes and of the Notch pathway, which was validated by increased HES1 staining in tumors (*SI Appendix*, Fig. S9*A*). Reduced vimentin staining of Ki-67 mutant tumors confirmed the loss of EMT in vivo (Fig. 5*B*). As assessed by PCNA and phosphohistone H3S10 staining, cell proliferation in vivo was unaffected by loss of Ki-67 (Fig. 5 *C* and *D*). However, tumors from *Mki67*$^{-/-}$ 4T1 cells grew significantly slower than from control cells in both types of immunodeficient mice (Fig. 5 *E* and *F*). There were no apparent differences in apoptosis upon Ki-67 knockout in 4T1 cells in vitro (*SI Appendix*, Fig. S9*B*), and while there was significantly increased apoptosis in one of the two 4T1 Ki-67 knockout clones in vivo (*SI Appendix*, Fig. S9 *C* and *D*), both clones had similarly slowed tumor progression. Analysis of necrosis revealed variability between tumors and clones (*SI*

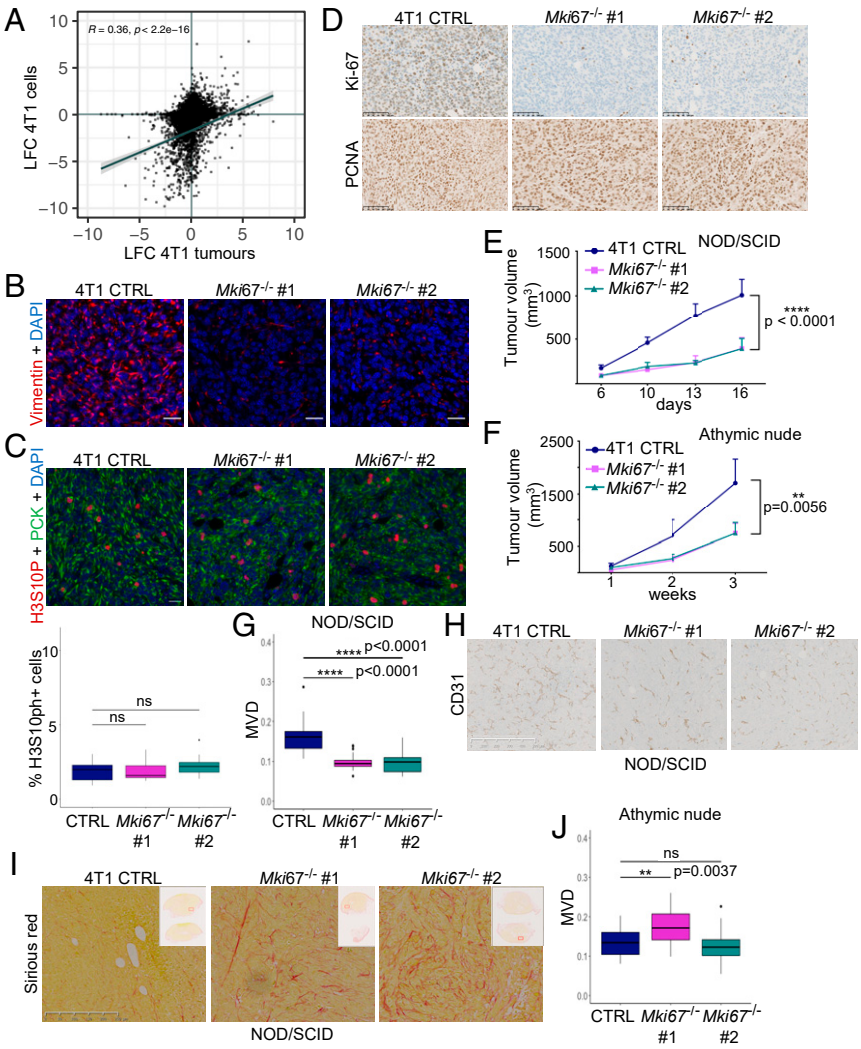


Fig. 5. Ki-67 promotes tumor growth. (*A*) Dot plot analysis of differentially expressed genes in 4T1 *Mki67*$^{-/-}$ cells vs. tumors derived from grafting 4T1 *Mki67*$^{-/-}$ cells into nude mice, showing a highly significant correlation. LFC, $\log_2$ fold change. (*B*) Fluorescent immunohistochemistry analysis of vimentin in 4T1 CTRL (WT) and *Mki67*$^{-/-}$ tumors in NOD/SCID mice. (Scale bars: 30 μm.) (*C*) Fluorescent IHC analysis of H3S10ph in 4T1 CTRL and *Mki67*$^{-/-}$ tumors in NOD/SCID mice (red; pan cytokeratin [PCK] in green, and DNA stained with DAPI in blue; *Upper*) and quantification (*Lower*). (Scale bars: 30 μm.) (*D*) IHC staining for Ki-67 and PCNA in 4T1 CTRL or *Mki67*$^{-/-}$ tumors. (Scale bars: 100 μm.) The 4T1 CTRL or *Mki67*$^{-/-}$ orthotopic xenografts in NOD/SCID (*E*) and nude (*F*) mice. Tumor growth was monitored for 3 wk. Error bars, SEM ($n = 6$ mice). CD-31 staining of endothelial cells in CTRL and *Mki67*$^{-/-}$ tumors in NOD-SCID mice: quantification of mean vessel density (MVD; *G*) and representative IHC images (*H*). (Scale bar: 500 μm.) (*I*) Sirius red staining of collagen in CTRL and *Mki67*$^{-/-}$ tumors in NOD-SCID mice. *Insets* indicate tumor areas presented at higher magnification. (Scale bar: 250 μm.) (*J*) Quantification of MVD in CTRL and *Mki67*$^{-/-}$ tumors in nude mice. $^{**}P < 0.001$; $^{****}P < 0.0001$; ns, nonsignificant.

Appendix, Fig. S9 *E* and *F*). As assessed by γ-H2A.X (histone H2A.X phosphorylated on serine-139) staining, we did not find evidence for increased DNA damage in Ki-67 knockout cells nor tumors (*SI Appendix*, Fig. S9 *G* and *H*). Thus, in vivo differences in tumor growth between WT and Ki-67 knockout tumors cannot be explained either by reduced cell proliferation or by increased DNA damage or cell death. It is theoretically possible that the slower tumor growth of Ki-67 knockouts might be explained by a marginally lower intrinsic cell proliferation rate that we did not detect when culturing WT and knockout cell lines individually but might become visible over longer timescales and when cocultured with WT cells. To test this, we performed competition experiments in vitro between WT and knockout 4T1 cells over 12 d. We found that the initial 50:50 ratio of WT and Ki-67 knockout 4T1 cells was maintained after 2 wk of coculture (*SI Appendix*, Fig. S10), ruling out this possible explanation.

Next, since Ki-67 4T1 knockout cells had down-regulated expression of several angiogenic factors, including angiopoietin-1, both in cell culture and in tumors (Datasets S2 and S4), we analyzed angiogenesis in WT and Ki-67 knockout tumors by CD31 staining of endothelial cells. Indeed, we found that the mean blood vessel density was significantly reduced in tumors from both Ki-67 knockout clones in NOD/SCID mice (Fig. 5 *G* and *H*). Tumors lacking Ki-67 also appeared more fibrotic than controls, as indicated by Sirius red staining of collagen (Fig. 5*I*). However, blood vessel density was comparable between WT and Ki-67 knockout tumors in nude mice (Fig. 5*J*). Moreover, the distribution of vessels of different sizes was similar between all genotypes and mouse backgrounds (*SI Appendix*, Fig. S9*I*). Therefore, while we cannot rule out a contribution of either reduced or altered angiogenesis to the slow growth of knockout tumors, this cannot explain the differences seen in all situations.

We also analyzed tumors from xenografts of WT and Ki-67 mutant MDA-MB-231 cells in nude mice. We similarly observed reduced tumor growth rate despite normal cell proliferation (*SI*

Appendix, Fig. S11 *A–C*). As with 4T1 cells, there was no detectable increase in apoptosis in MDA-MB-231 cells lacking Ki-67 in vitro (*SI Appendix*, Fig. S9*B*). However, apoptosis, but not DNA damage (*SI Appendix*, Fig. S11*D*), was significantly increased in Ki-67 knockout MDA-MB-231 tumors (*SI Appendix*, Fig. S11*E*). Necrosis was reduced (*SI Appendix*, Fig. S11*F*), and fibrosis was more apparent (*SI Appendix*, Fig. S11*G*), while mean blood vessel density and distribution of vessels of different sizes were comparable with control tumors (*SI Appendix*, Fig. S11 *H* and *I*).

To assess the generality of these observations in a different tumor type, we stably knocked down Ki-67 by short hairpin RNA (shRNA) in a commonly used aggressive human cervical cancer cell line, HeLa S3, and grafted the resulting cells and an shRNA control line into opposing flanks of athymic nude mice. Ki-67 knockdown was maintained in vivo (*SI Appendix*, Fig. S12 *A* and *B*). Again, tumor growth was severely impaired (*SI Appendix*, Fig. S12*C*), despite unaffected cell proliferation (as shown by unchanged PCNA and mitotic indices) (*SI Appendix*, Fig. S12 *D* and *E*), while there was increased necrosis and apoptosis (*SI Appendix*, Fig. S12 *F* and *G*). It is interesting to note that shRNA of Ki-67 was sufficient to induce strong phenotypes.

In conclusion, Ki-67 knockout or knockdown in several different established cancer cell lines consistently results in slower tumor growth upon grafts in mice despite unchanged cell proliferation and often leads to noncell-autonomous increases in apoptosis or fibrosis, with necrosis and angiogenesis more variably affected. However, none of these plausible explanations for the reduced tumor growth are true in all experimental situations. These results show that effects of Ki-67 loss are wide ranging and multifactorial.

Lastly, since many of the genes repressed in the absence of Ki-67 are under the control of PRC2 and concurrent knockout of PRC2 genes partly restores the EMT to Ki-67 KO cells (*SI Appendix*, Fig. S7), we tested whether the inactivation of PRC2 in cells lacking Ki-67 could restore tumor growth. We injected *Suz12*$^{-/-}$, *Mki67*$^{-/-}$, and *Suz12*$^{-/-}$ *Mki67*$^{-/-}$ 4T1 cells orthotopically into nude mice and found that *Suz12* knockout did not restore *Mki67*$^{-/-}$ 4T1 tumor growth rates, nor affect tumorigenicity of control cells (*SI Appendix*, Fig. S13). As such, PRC2 contributes to suppressing the EMT in the absence of Ki-67, but it does not in itself have either pro- or antitumor activity in this context, in contrast to other systems (25, 39). This result also suggests that Ki-67 roles in tumorigenesis depend on several chromatin regulatory pathways, consistent with it being a hub for interactions with multiple chromatin regulators.

Ki-67 Promotes Metastasis but Enables Efficient Antitumor Immune Responses. We observed that in orthotopic grafts, control 4T1 cells metastasized to the lungs in 4 wk in nude mice, but metastases were largely absent at this point in mice bearing *Mki67*$^{-/-}$ tumors (Fig. 6*A*). We tested whether this was due to differences in detachment from the primary tumor or ability to seed metastases. To do this, we quantified rates of metastasis formation of control and Ki-67 knockout 4T1 cells by injecting cells directly into the tail vein, then dissociating lung tissue after 3 wk, and growing cells in the presence of 6-thioguanine, to which 4T1 cells are resistant. The number of metastatic cells that formed colonies was reduced nearly 100-fold in Ki-67 knockouts (Fig. 6*B*). This points to an essential requirement for Ki-67 in seeding metastasis, in accordance with its apparent role in conferring stem-like characteristics. We could not test whether this reduced metastatic capacity was conserved in MDA-MB-231 cells since, in our experiments, neither Ki-67 knockout nor control MDA-MB-231 xenografts generated any visible lung metastases, consistent with their lack of stem-like character.

We next investigated how Ki-67 expression affects tumorigenesis and metastasis in the context of an intact immune system by engrafting WT or *Mki67*$^{-/-}$ 4T1 cells into immune-proficient BALB/c mice. As expected (40), control 4T1 tumors established quickly and initially regressed before regrowing (Fig. 6*C*). This initial regression has been attributed to a strong antitumor immune response (40). Surprisingly, no such initial regression occurred when *Mki67*$^{-/-}$ 4T1 cells were engrafted (Fig. 6*C*), suggesting that Ki-67 knockout cells fail to trigger an efficient immune response. Nevertheless, despite the consequently higher tumor burden, there were either similar or less metastases than from control 4T1 cells (Fig. 6*D*), although differences failed to reach statistical significance ($P = 0.13$). We surmised that injection of 4T1 cells directly into the circulation via the tail vein should circumvent the immune targeting of tumor cells at the primary site, allowing direct assessment of the capacity of cells to establish a metastatic niche. Importantly, WT 4T1 cells again more efficiently colonized lungs in this setting than *Mki67*$^{-/-}$ 4T1 cells (Fig. 6*E*), despite the higher overall tumor burden in the latter. Together, the results of these experiments underline the requirement for Ki-67 in seeding metastasis. They also suggest a defective immune response to Ki-67 knockout tumors. In agreement, immunohistological analysis revealed increased infiltration of myeloid-derived suppressor cells (MDSCs) (41) in knockout tumors (using the MDSC marker, GR1) (Fig. 6*F*), possibly indicating an immunosuppressive environment that protects the tumors against immune-mediated cytotoxicity. To investigate why Ki-67 knockout cells fail to induce an efficient antitumor immune response, we analyzed the expression of the mouse major histocompatibility complex (MHC) class I, responsible for antigen presentation. Transcriptome analysis showed down-regulated expression of factors implicated in antigen processing and loading on MHC I: Tapasin, Tap1, Tap2, Psmb8, and Psmb9 in *Mki67*$^{-/-}$ 4T1 cells; tumors showed a further down-regulation of Erap1 and B2M (Fig. 3 *B* and *F* and Datasets S2 and S4). Flow cytometry revealed lower expression of MHC class I molecules H2D and H2K (Fig. 6*G*). MHC class I expression was also down-regulated in *MKI67*$^{-/-}$ MDA-MB-231 cells (*SI Appendix*, Fig. S14), suggesting that roles of Ki-67 in maintaining MHC expression are conserved across species.

Cells Lacking Ki-67 Are Sensitized to Drugs. The above results show that Ki-67 enables cell transformation, tumor growth, and metastasis, yet also confers efficient targeting of tumors by the immune system. Finally, we asked whether Ki-67 would also influence drug responses, as, in addition to stem-like characteristics, the EMT has also been associated with resistance to chemotherapeutic drugs (42). We noticed that 26 genes involved in drug metabolism were down-regulated in Ki-67 knockout 4T1 cells, while only 1 was up-regulated, suggesting that Ki-67 expression might affect sensitivity to chemotherapeutic drugs (Fig. 7*A* and *SI Appendix*, Fig. S15*A*). To test this, we performed an automated gene–drug screen using the Prestwick chemical library, composed of 1,283 Food and Drug Administration–approved small molecules. We also included salinomycin, a positive control found to target cancer stem cells (CSCs) (43), and 6-thioguanine, which was originally used to isolate 4T1 cells (22). Control 4T1 cells were sensitive to 102 drugs at 10 μM concentration, while the two *Mki67*$^{-/-}$ clones were sensitive to 99 and 98, with 82 hits common to the three cell lines (*SI Appendix*, Fig. S15*B* and Dataset S5). This suggests that Ki-67 loss does not qualitatively alter the drug sensitivity profiles. We next determined the concentration of drug needed for 50% growth inhibition (IC_{50}) of 10 hits commonly used in cancer therapy. Importantly, *Mki67*$^{-/-}$ cells were markedly more sensitive to all the molecules tested (Fig. 7*B*). As such, by supporting expression of xenobiotic metabolism genes, Ki-67 provides cancer cells with a degree of protection against therapeutic drugs.

Discussion

The above results show that Ki-67 is not required for cancer cell proliferation in vitro or in vivo in any cell type tested, but its expression critically influences all steps of tumorigenesis, including

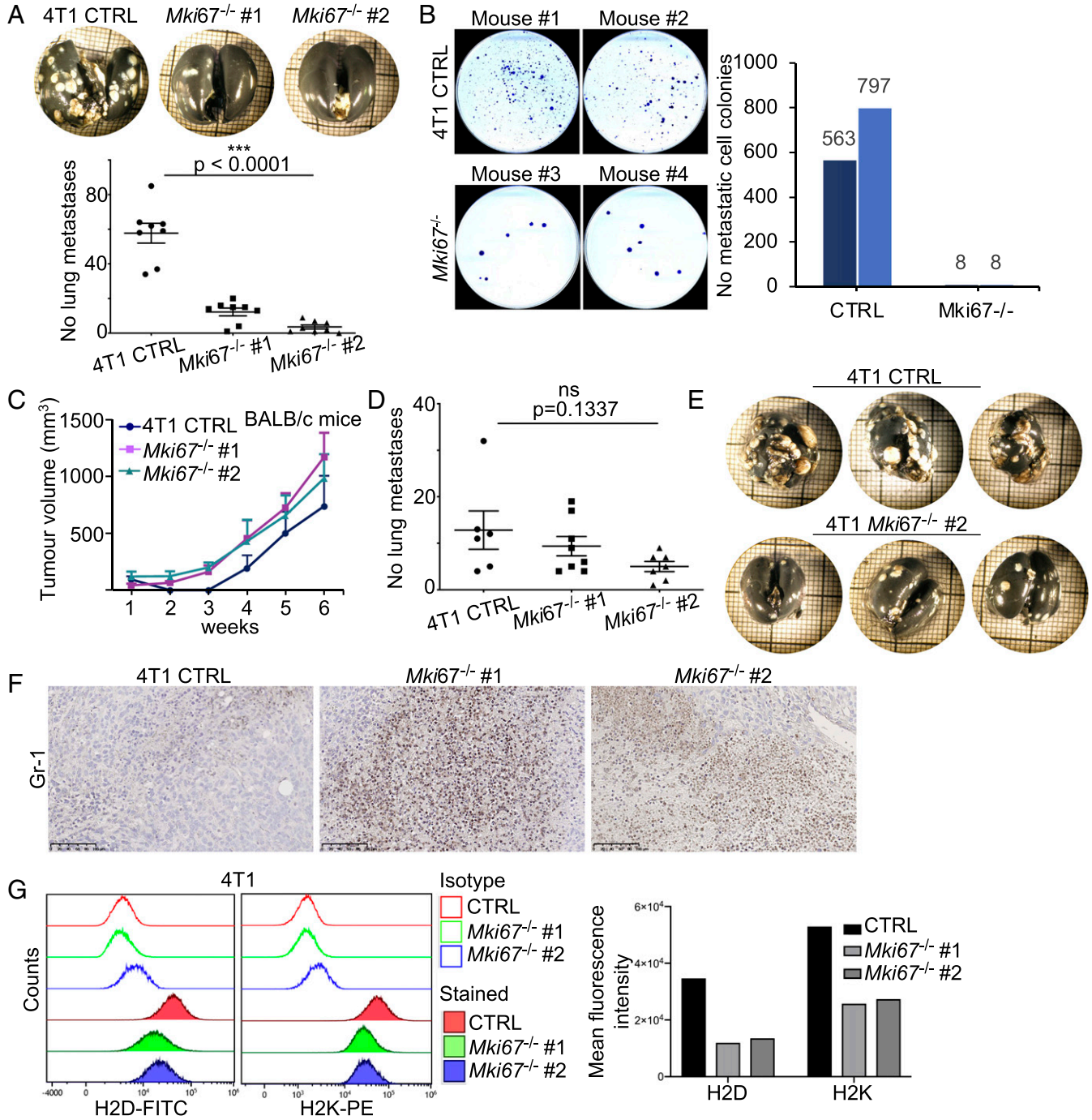
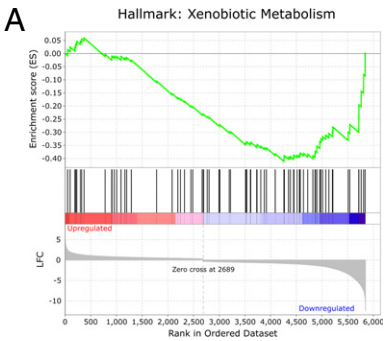


Fig. 6. Ki-67 promotes metastasis and antitumor immune responses. (*A*) Quantification of lung metastases in nude mice injected orthotopically with 4T1 CTRL (WT) or *Mki67*$^{-/-}$ cells. Error bars, SEM. (*A*, *Upper*) Representative images of lungs stained to visualize metastases (white nodules). (background scale is mm; ***$P < 0.0001$). (*B*) Lung tissue from nude mice injected via tail vein with 4T1 CTRL or *Mki67*$^{-/-}$ cells (two per condition) was dissociated after 2 wk, and resulting cells were maintained in the presence of 6-thioguanine to select for 4T1 cells. (*B*, *Left*) Crystal violet staining of resulting colonies. (*B*, *Right*) Quantification. (*C*) Tumor growth over 6 wk of 4T1 CTRL or *Mki67*$^{-/-}$ orthotopic xenografts in immunocompetent BALB/c mice. Error bars, SEM ($n = 8$ mice), and quantification of lung metastases in each group (*D*). Error bars, SEM; ns, non significant. (*E*) The 4T1 CTRL or *Mki67*$^{-/-}$ #2 cells were injected via tail vein into immune-competent BALB/c mice. Representative images of stained lungs (day 21 postinjection); metastases are white. (background scale is mm) (*F*) Immunochistochemistry analysis of 4T1 CTRL or *Mki67*$^{-/-}$ tumors (week 4 posttransplantation) stained for Gr-1, an MDSC marker. (Scale bars: 100 μm.) (*G*) The 4T1 CTRL and *Mki67*$^{-/-}$ cells were stained with anti–H2D^{d} fluorescein isothiocyanate (FITC)-conjugated antibody and anti–H2K^{d} Phycoerythrin (PE)-conjugated antibody, or control isotypes. (*G*, *Left*) Flow cytometry profiles. (*G*, *Right*) Quantification ($n = 1$ of each clone).

initiation, progression, and metastasis, as well as immune responses and drug sensitivity. In both human and mouse mammary carcinoma cells, we present evidence that this is because Ki-67 sustains transcriptional programs needed for tumor cells to adapt to their environment. This is indicated by the failure to generate intestinal tumors in Ki-67 mutant mice; the reduced ability of some Ki-67 knockout cancer cells to induce angiogenesis; the presence of both epithelial and mesenchymal characteristics in Ki-67–positive



B

Compound	IC_{50} (μM)		
	4T1 CTRL	Mki67$^{-/-}$#1	Mki67$^{-/-}$#2
Camptothecin	4.1	1.5	1.0
Doxorubicin	1.2	0.5	0.6
Epirubicin	0.7	0.3	0.34
Daunorubicin	0.64	0.3	0.3
Mitoxantrone	1.3	0.4	0.49
Etoposide	21.4	7.8	9.1
Topotecan	15.8	2.6	1.9
Cytarabine	3.0	1.3	1.5
Floxiuridine	0.1	0.1	0.0
Vorinostat	5.7	2.3	2.8
6-Thioguanine	N.A	N.A	N.A
Salinomycine	4.4	2.6	2.7

Fig. 7. Ki-67 promotes cancer cell drug resistance. (*A*) Xenobiotic metabolism is a hallmark of genes down-regulated in *Mki67*$^{-/-}$ cells. LFC, log2 fold change. (*B*) IC_{50} (concentration of drug needed for 50% growth inhibition) of 4T1 WT (CTRL) or *Mki67*$^{-/-}$ cells to indicated compounds, derived from dose–response experiments.

epithelial cancer cells, which are altered in cells lacking Ki-67; the inability of the latter to colonize other tissues and give rise to metastases; their increased sensitivity to drugs; and their reduced interactions with the immune system. These phenotypes correlate well with alterations of expression of genes involved in the EMT, antigen presentation, drug metabolism, and other cancer-associated hallmarks.

Our data suggest that Ki-67 does not work by directly controlling expression of specific genes, since in Ki-67 knockout mouse fibroblast or epithelial cancer cells, there are relatively few common deregulated genes. Instead, we find evidence that Ki-67 regulates general chromatin states that affect expression of genes in fundamental biological processes. This is supported by the fact that gene expression alterations are global (i.e., well within an order of magnitude of the number of genes in the genome), with around 2,500 genes altered in nontransformed fibroblasts, nearly 5,000 genes deregulated in Ki-67 knockout mouse mammary carcinoma cells, and over 9,000 (i.e., the majority of expressed genes) in human mammary cancer cells. We compared changes in expression of genes involved in biological processes that are significantly enriched in Ki-67 knockout 3T3, 4T1, and 4T1 tumor grafts and human MDA-MB-231 cells (*SI Appendix*, Fig. S16). This clearly shows the similarity of affected pathways, and even genes, among these different knockouts, and highlights the fact that within common pathways different individual genes may be affected among different cell types. The similarity between the changes in different mammary carcinoma cells is apparent both in vitro and in tumors.

Our interpretation is also consistent with the extensive alterations in chromatin histone marks. With such wide-ranging transcriptome changes, it is almost impossible to attribute the loss of tumorigenicity to changes in expression of candidate genes, or even single regulatory mechanisms. Indeed, inactivation of the PRC2 complex did not restore tumorigenicity to Ki-67 knockout 4T1 cells. However, similar cancer-associated pathways are deregulated in 4T1 and MDA-MB-231 cells, suggesting general conservation of regulatory mechanisms.

A role in cancer-promoting transcriptional programs provides a plausible explanation for the ubiquitous expression of Ki-67 across essentially all cancer types. Indeed, we show that Ki-67 is required for efficient tumor growth, independently of cell proliferation, in different human and mouse cancer cell types. Although to some observers it might appear paradoxical that tumor progression is reduced in Ki-67 knockouts without changes in cell proliferation rates, our data suggest several possible mechanisms. We observed noncell-autonomous effects in vivo, including increases in apoptosis or fibrosis, to which altered angiogenesis could be a contributor. Other forms of cell death cannot be ruled out, but we did not find a consistent increase in necrosis. Additionally, it is possible that tumors might arise from a smaller number of initiating cells, due to a loss of stem-like character. In support of this, spheroid formation is decreased in both highly epithelial (4T1) and highly mesenchymal (MDA-MB-231) cancer cells. The fact that Ki-67 knockout reduces ability to seed metastases is consistent with this idea.

Whether or not Ki-67 expression maintains CSCs as such is debatable. A recent study in the human colorectal cancer cell line DLD-1 showed that Ki-67 knockout reduces the number of cells expressing the antigen CD133, commonly assumed to be a marker of CSCs (13). However, in intestinal crypts, CD133 expression is not specific to stem cells, and CD133-negative cancer cells were equally capable of sustaining tumorigenesis in a long-term serial transplantation model (44). Furthermore, the concept that CSCs are rare preexisting populations of cells with hard-wired CSC properties, a model that emerged from xenograft experiments using sorted populations of hematopoietic stem cells, no longer appears valid, at least for solid tumors (45). Lineage-tracing experiments in the intestine showed that tumors arise from stem cells, which divide rapidly, make up around 10% of the cell in intestinal crypts, and can be regenerated from nonstem cells (45). A more recent concept is that epithelial cells are phenotypically plastic, in a manner dependent on EMT-inducing transcription factors (46), with no such fixed entity as CSC or non-CSC. Emerging evidence suggests that cells can reside in various phenotypic stages along the EMT spectrum with concurrent expression of epithelial and mesenchymal traits and that such hybrid states are important for carcinogenesis (33–35). Our data showing that ALDH activity, which is a bona fide marker for intestinal stem cells and breast and colon CSCs (36, 37), is expressed by a significant fraction (~10%) of WT epithelial cancer cells but far fewer (<3%) Ki-67 knockout cells, are consistent with this model.

The EMT and stemness have previously been correlated with drug metabolism (42, 47). We find that, in cancer cells with a reversion of the EMT, Ki-67 loss leads to a reduced expression of genes encoding xenobiotic metabolism factors, which translates into increased sensitivity to all drugs tested in 4T1 cells. This is likely a conserved phenotype of Ki-67 loss since Ki-67 knockout HeLa, DLD1, and MCF10A cells were also found to be more sensitive to all drugs tested (10, 13).

Previous studies of links between stem cell characteristics and carcinogenesis have focused on transcriptional states, converging on Myc-regulated targets as the strongest link between stem cell and cancer cell transcriptional signatures (48). However, while in Ki-67 knockout NIH/3T3 cells, Myc was slightly down-regulated (Dataset S1), it was significantly up-regulated in 4T1 cells (Dataset S2), ruling out the possibility that the reduced oncogenicity of these cells is simply due

to loss of Myc, and again favoring a global effect of Ki-67 loss on the chromatin state.

Our hypothesis is consistent with our previous identification of many general chromatin regulators interacting with Ki-67, including components of PRC1 and PRC2, REST, NuRD, NIF1, NuA4, MLL, SET1, NoRC, and NCOA6 complexes (11). Disrupting one such interactor, the PRC2 complex, partially, but not fully, rescued mesenchymal traits in Ki-67 mutant epithelial cancer cells, consistent with the idea that Ki-67 acts through multiple chromatin regulatory complexes. In accordance, loss of PRC2 did not restore tumorigenicity to *Mki67*$^{-/-}$ 4T1 cells. Identifying the exact biochemical mechanisms by which Ki-67 affects the chromatin state and gene expression will require further studies. Ki-67 has no enzymatic activity and appears to be a largely intrinsically disordered protein, potentially providing hub-like properties for protein–protein interactions. How this confers the ability to adapt to the environment is not clear, but one can speculate that its binding of a large number of general transcription regulators is involved in maintenance of metastable states between different transcriptional programs—in other words, transcriptional plasticity.

In conclusion, Ki-67, which is universally expressed in proliferating cells, enables multiple steps in carcinogenesis in different cancer types, which require drastic changes in transcriptional programs. Therapeutic targeting of Ki-67 itself will likely be challenging since it is an intrinsically disordered protein with no inherent enzyme activity. However, it could be of therapeutic benefit to inhibit its effectors in control of cellular adaptation to the environment. Alternatively, our results also suggest that Ki-67 confers an Achilles' heel to cancer cells, namely their recognition by the immune-system. It will be interesting to see if nonproliferating cells, which do not express Ki-67, are consequently more resistant to immune-mediated killing. If this is the case, then, paradoxically, it might be advantageous to promote, rather than hinder, cell proliferation in order for immunotherapy to be optimally effective.

Materials and Methods

Details of all materials and methods can be found in *SI Appendix*. Materials include all animals, cell lines, antibodies, plasmids, PCR primers, and reagents. Methods include CRISPR-Cas9–mediated genome editing, AOM-DSS–mediated colon carcinogenesis, DNA replication assay, mammosphere assay, ALDH activity assay, xenografts, visualization of lung metastases, cell extracts and western blotting, histology and immunostaining, qRT-PCR, colony formation assay, RNA-sequencing library preparation and sequencing, ChIP-Seq, sequence data processing, gene set enrichment analysis, automated drug library screen, and statistical analysis.

Data Availability. All RNA-Seq and ChIP-Seq raw data have been deposited in the Gene Expression Omnibus as a SuperSeries on December 13, 2020 (accession no. GSE163114).

ACKNOWLEDGMENTS. We thank D. Grimanelli for initial guidance on analyzing transcriptomes; A. Covinhes and Y. Buscail for immunohistochemistry development; Montpellier Resources in Imaging for imaging and flow cytometry services; MGX for sequencing services; and P. Jay, D. Santamaria, M. E. Hochberg, and M. Serrano for comments on the manuscript. K.M and M.S. were funded by the Ligue Nationale Contre le Cancer (LNCC). K.M. and P.S. received funding from World-Wide Cancer Research (WWCR). N.A.-S. was funded by the University of Montpellier. G.D. and D.C. were funded by the French National Cancer Institute (INCa). This work was initiated and finalized with support from LNCC Grants EL2013.LNCC/DF and EL2018.LNCC/DF and continued with support from WWCR Grant 16-0006 and INCa Grant PLBIO18-094. The chemical library screen relied on support provided by the Programme Opérationnel FEDER-FSE 2014-2020 Languedoc Roussillon. The Réseau d'Histologie Expérimentale de Montpellier histology facility was supported by SIRIC Montpellier Cancer Grant INCa_Inserm_DGOS_12553, the European Regional Development Foundation, and the Occitanie region (FEDER-FSE 2014-2020 Languedoc Roussillon).

1. E. Endl, J. Gerdes, The Ki-67 protein: Fascinating forms and an unknown function. *Exp. Cell Res.* **257**, 231–237 (2000).
2. J. Cuzick *et al.*, Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J. Clin. Oncol.* **29**, 4273–4278 (2011).
3. I. Kausch *et al.*, Antisense treatment against Ki-67 mRNA inhibits proliferation and tumor growth in vitro and in vivo. *Int. J. Cancer* **105**, 710–716 (2003).
4. R. Rahmanzadeh, G. Hüttmann, J. Gerdes, T. Scholzen, Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis. *Cell Prolif.* **40**, 422–430 (2007).
5. C. Schlüter *et al.*, The cell proliferation-associated antigen of antibody Ki-67: A very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins. *J. Cell Biol.* **123**, 513–522 (1993).
6. M. Starborg, K. Gell, E. Brundell, C. Höög, The murine Ki-67 cell proliferation antigen accumulates in the nucleolar and heterochromatic regions of interphase cells and at the periphery of the mitotic chromosomes in a process essential for cell cycle progression. *J. Cell Sci.* **109**, 143–153 (1996).
7. J. N. Zheng *et al.*, Inhibition of renal cancer cell growth in vitro and in vivo with oncolytic adenovirus armed short hairpin RNA targeting Ki-67 encoding mRNA. *Cancer Gene Ther.* **16**, 20–32 (2009).
8. J. N. Zheng *et al.*, Knockdown of Ki-67 by small interfering RNA leads to inhibition of proliferation and induction of apoptosis in human renal carcinoma cells. *Life Sci.* **78**, 724–729 (2006).
9. D. G. Booth *et al.*, Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *eLife* **3**, e01641 (2014).
10. S. Cuylen *et al.*, Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. *Nature* **535**, 308–312 (2016).
11. M. Sobecki *et al.*, The cell proliferation antigen Ki-67 organises heterochromatin. *eLife* **5**, e13722 (2016).
12. M. Takagi, T. Natsume, M. T. Kanemaki, N. Imamoto, Perichromosomal protein Ki67 supports mitotic chromosome architecture. *Genes Cells* **21**, 1113–1124 (2016).
13. J. Cidado *et al.*, Ki-67 is required for maintenance of cancer stem cells but not cell proliferation. *Oncotarget* **7**, 6281–6293 (2016).
14. M. Sobecki *et al.*, Cell-cycle regulation accounts for variability in Ki-67 expression levels. *Cancer Res.* **77**, 2722–2734 (2017).
15. I. Miller *et al.*, Ki67 is a graded rather than a binary marker of proliferation versus quiescence. *Cell Rep.* **24**, 1105–1112.e5 (2018).
16. X. Sun *et al.*, Ki-67 contributes to normal cell cycle progression and inactive X heterochromatin in p21 checkpoint-proficient human cells. *Mol. Cell. Biol.* **37**, e00569-16 (2017).
17. J. T. Norton *et al.*, Perinucleolar compartment prevalence is a phenotypic pancancer marker of malignancy. *Cancer* **113**, 861–869 (2008).
18. A. Tsherniak *et al.*, Defining a cancer dependency Map. *Cell* **170**, 564–576.e16 (2017).
19. J. Anaya, OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Comput. Sci.* **2**, e67 (2016).
20. M. De Robertis *et al.*, The AOM/DSS murine model for the study of colon carcinogenesis: From pathways to diagnosis and therapy studies. *J. Carcinog.* **10**, 9 (2011).
21. S. Colnot *et al.*, Colorectal cancers in a new mouse model of familial adenomatous polyposis: Influence of genetic and environmental modifiers. *Lab. Invest.* **84**, 1619–1630 (2004).
22. D. L. Dexter *et al.*, Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res.* **38**, 3174–3181 (1978).
23. G. H. Heppner, F. R. Miller, P. M. Shekhar, Nontransgenic models of breast cancer. *Breast Cancer Res.* **2**, 331–334 (2000).
24. N. Tiwari *et al.*, Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. *Cancer Cell* **23**, 768–783 (2013).
25. A. Chase, N. C. P. Cross, Aberrations of EZH2 in cancer. *Clin. Cancer Res.* **17**, 2613–2618 (2011).
26. C. S. Nowell, F. Radtke, Notch as a tumour suppressor. *Nat. Rev. Cancer* **17**, 145–159 (2017).
27. M. Nicolas *et al.*, Notch1 functions as a tumor suppressor in mouse skin. *Nat. Genet.* **33**, 416–421 (2003).
28. V. Devgan, C. Mammucari, S. E. Millar, C. Brisken, G. P. Dotto, p21WAF1/Cip1 is a negative transcriptional regulator of Wnt4 expression downstream of Notch1 activation. *Genes Dev.* **19**, 1485–1495 (2005).
29. R. D. Meng *et al.*, γ-Secretase inhibitors abrogate oxaliplatin-induced activation of the Notch-1 signaling pathway in colon cancer cells resulting in enhanced chemosensitivity. *Cancer Res.* **69**, 573–582 (2009).
30. S. A. Mani *et al.*, The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
31. A.-P. Morel *et al.*, Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* **3**, e2888 (2008).
32. U. Wellner *et al.*, The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* **11**, 1487–1495 (2009).
33. C. Kröger *et al.*, Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7353–7362 (2019).
34. I. Pastushenko *et al.*, Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).
35. X. Ye *et al.*, Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).

36. C. Ginestier *et al*., ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555–567 (2007).
37. E. H. Huang *et al*., Aldehyde dehydrogenase 1 is a marker for normal and malignant human colonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. *Cancer Res.* **69**, 3382–3389 (2009).
38. T. Ishiguro *et al*., Tumor-derived spheroids: Relevance to cancer stem cells and clinical applications. *Cancer Sci.* **108**, 283–289 (2017).
39. A. P. Bracken, K. Helin, Polycomb group proteins: Navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* **9**, 773–784 (2009).
40. K. Tao, M. Fang, J. Alroy, G. G. Sahagian, Imagable 4T1 model for the study of late stage breast cancer. *BMC Cancer* **8**, 228 (2008).
41. Y. Yang, C. Li, T. Liu, X. Dai, A. V. Bazhin, Myeloid-derived suppressor cells in tumors: From mechanisms to antigen specificity and microenvironmental regulation. *Front. Immunol.* **11**, 1371 (2020).
42. X. Zheng *et al*., Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).
43. P. B. Gupta *et al*., Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138**, 645–659 (2009).
44. S. V. Shmelkov *et al*., CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. *J. Clin. Invest.* **118**, 2111–2120 (2008).
45. E. Batlle, H. Clevers, Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).
46. C. L. Chaffer *et al*., Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* **154**, 61–74 (2013).
47. T. Shibue, R. A. Weinberg, EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).
48. J. Kim *et al*., A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–324 (2010).

Appendix 3

CDK8 and CDK19 act redundantly to control the CFTR pathway in the intestinal epithelium

CDK8 and CDK19 control the CFTR pathway in the intestinal epithelium

Susana Prieto[1,2], Elsie Hodimont[1,2], Lucie Angevin[1,2], Geronimo Dubra[1,2], Ana Bella Aznar[1,2], Alain Camasses[1,2], Christina Begon-Pescia[1,7], Nelly Pirot[4,5], François Gerbe[2,3], Philippe Jay[2,3], Liliana Krasinska[1,2] and Daniel Fisher[1,2]*

[1]IGMM, University of Montpellier, CNRS, Inserm, Montpellier, France.
[2]Equipe Labellisée LIGUE 2018, Ligue Nationale Contre le Cancer, Paris, France.
[3]IGF, University of Montpellier, CNRS, Inserm, Montpellier, France.
[4]IRCM, University of Montpellier, ICM, INSERM, Montpellier, France
[5]BioCampus, RHEM, University of Montpellier, CNRS, INSERM, Montpellier, France
[6]Current address: Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[7]Current address: LPHI, University of Montpellier, Montpellier, France

Contact information: * correspondence to: daniel.fisher@igmm.cnrs.fr

Keywords CDK8; CDK19; Mediator; intestinal epithelium; CFTR.

Word count:

Section	*Words*	*Chrs (inc spaces)*
Abstract	150	1118
Main text	3468	22969
Materials and Methods	3872	25890
Acknowledgements, author contributions, COI statement	153	1069
References	2273	13689
Figure legends and EV figure legends	1829	11523

Abstract
CDK8 and CDK19 form a conserved cyclin-dependent kinase subfamily that regulates the Mediator complex. However, cells lacking either kinase have only minor transcriptional alterations, suggesting functional redundancy. Here, we demonstrate that CDK8 and CDK19 act together to control the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) pathway in mice. We find that CDK8 is dispensable both for normal intestinal homeostasis and efficient tumourigenesis, and individual knockout of genes encoding CDK8 or CDK19 in intestinal organoids has only limited effects on gene expression. In contrast, their combined deletion, although not cell-lethal, causes progressive loss of proliferative capacity, accompanied by more extensive gene expression changes. In particular, the CFTR pathway is downregulated, leading to mucus accumulation and increased secretion by goblet cells. Pharmacological inhibition indicates that the CFTR pathway is dependent on CDK8/19 kinase activity. We conclude that Mediator kinases are not essential in all cells, but cooperate to regulate tissue-specific transcriptional programmes.

Introduction
CDK8 was discovered as a kinase that binds cyclin C and, like CDK7-cyclin H and CDK9-cyclin T, can promote transcription by phosphorylating the C-terminal repeat domain (CTD) of RNA polymerase II (Pol II) (Rickert *et al*, 1999). CDK8 and cyclin C are exceptionally highly conserved in vertebrates, as illustrated by 97% amino acid identity over the whole sequence between *Xenopus* and human CDK8, and 98% identity between cyclin C proteins from these species. This unusual level of cross-species conservation implies critical functions in fundamental cellular processes. In complex with Med12 and Med13, CDK8-cyclin C forms the canonical cyclin-dependent kinase module (CKM) of the Mediator transcriptional co-regulator complex, a function that is conserved with the more divergent yeast homologues of CKM subunits (Jeronimo *et al*, 2016). The latter were revealed as suppressors of a CTD truncation, suggesting a transcription-repressive activity of the CKM (Liao *et al*, 1995). In vertebrates, a second member of the CDK8 subfamily, CDK19, almost identical in the kinase domain with CDK8, also binds cyclin C and interacts with Mediator, in a manner generally thought to be exclusive with CDK8 (Sato *et al*, 2004; Tsutsui *et al*, 2008; Knuesel *et al*, 2009).

Mediator is a large multi-subunit complex required for Pol II-dependent transcription in all eukaryotes (Malik & Roeder, 2010). Acute ablation of vertebrate Mediator is lethal for cells and results in a rapid downregulation of the entire transcriptome (El Khattabi *et al*, 2019). The tail subunits of Mediator integrate enhancer-associated transcription factor activity into conformational changes of the head- and middle complex. This regulates Pol II interactions with the basal transcriptional machinery at promoters, as well as phosphorylation of the CTD (Malik & Roeder, 2010). Biochemical analysis of yeast CKM provided evidence that it negatively regulates Mediator. Data from *in vitro* experiments suggest that it hinders basal transcription by sterically blocking CTD-dependent recruitment of PolII to Mediator middle subunits (Elmlund *et al*, 2006; Tsai *et al*, 2013); while *in vivo*, the CKM binds to the same promoters as core Mediator but with low stoichiometry (Andrau *et al*, 2006). This appears to be due to negative regulation of Mediator binding to upstream enhancer sequences and release of the CKM module upon Mediator-pol II interactions (Jeronimo *et al*, 2016).

In contrast to Mediator, the activity of the CKM is apparently non-essential in many cell types, as genes encoding CDK8, CDK19 and cyclin C are not required for survival and proliferation of most cell types in different organisms (Loncle *et al*, 2007; Kuchin *et al*, 1995; Li *et al*, 2014; Postlmayr *et al*, 2020). However, they are required for normal development. Germline ablation of *Cdk8* is lethal at the pre-implantation stage in mice (Westerling *et al*, 2007), while conditional deletion using a *Sox2* Cre driver is lethal around embryonic day 10.5 (Postlmayr *et al*, 2020). The difference in lethality stage between the two genotypes suggests that CDK8 might be essential in zygotes, prior to *Sox2* expression.

Deletions of other CKM subunits in mice have variable phenotypes. Cyclin C gene deletion is embryonic lethal at day 10.5 with severe growth defects, and its deletion in adults affects T-cell

differentiation (Li *et al*, 2014), while deletion of *Med12* is lethal at late embryonic stages, preventing neural-tube closure, axis elongation and organ morphogenesis (Rocha *et al*, 2010). CDK19 deletion has not yet been reported. An essential requirement for CKM subunits in transcriptional regulation in animals cannot, however, be completely ruled out, since differences in the lethality stage of CKM subunit deletions could be due to differential maternal mRNA contributions.

Consistent with a repressive role for the CKM in transcription, we recently reported that inhibition of CDK8 and CDK19 in human and mouse pluripotent stem cells is associated with a global overactivation of enhancers and a stabilisation of the naiive state (Lynch *et al*, 2020). Similarly, in acute myeloid leukaemia, CDK8/19 bind superenhancers and their chemical inhibition further activates enhancer activity (Pelish *et al*, 2015).

CDK8 has been attributed oncogenic functions in different cancers, including Wnt-dependent colorectal cancer, melanoma, breast and prostate cancer, acute myeloid leukaemia and B-cell leukaemia (Pelish *et al*, 2015; Firestein *et al*, 2008; Morris *et al*, 2008; Kapoor *et al*, 2010; McDermott *et al*, 2017; Nakamura *et al*, 2018; Menzl *et al*, 2019). Originally proposed to act in intestinal cancers by promoting Wnt transcription (Firestein *et al*, 2008), CDK8 is also involved in transcription dependent on Notch signalling (Li *et al*, 2014), HIF1α-gene expression (Galbraith *et al*, 2013), the serum-response (Donner *et al*, 2010), the interferon-γ response (Steinparzer *et al*, 2019), p53 (Donner *et al*, 2007), superenhancers (Pelish *et al*, 2015), histone variant incorporation into chromatin (Kapoor *et al*, 2010), in pluripotency maintenance (Adler *et al*, 2012) and the senescence-associated tumour-promoting secretory phenotype (Porter *et al*, 2012). It also restrains NK-mediated cell toxicity and tumour surveillance (Hofmann *et al*, 2020). Yet genetic confirmation of requirements for CDK8 in cancers *in vivo* has generally been lacking. Conditional knockout of *Cdk8* in the intestinal epithelium did not hinder intestinal tumour development in *Apc* mutant mice; rather, it appeared to enhance tumourigenesis, and knockouts were reported to have lost the Polycomb group 2-mediated repressive histone mark, H3 lysine-27 trimethylation, thus upregulating oncogenic transcription (McCleland *et al*, 2015).

In contrast to CDK8, almost nothing is known about CDK19 roles in cancer, and whether it compensates for loss of CDK8 remains unknown. *In vitro* inhibition or knockdown experiments have suggested that CDK8 and CDK19 control different sets of target genes (Tsutsui *et al*, 2008, 11; Galbraith *et al*, 2013, 11; Poss *et al*, 2016). By genetic ablation in mice liver cells, we recently found that CDK8 and CDK19 are both required for hepatic carcinogenesis, and highlighted genetic interaction with p53 as critical for their roles in tumourigenesis (Bacevic *et al*, 2019).

A number of potent CDK8 pharmacological inhibitors have been developed (Pelish *et al*, 2015; Porter *et al*, 2012; Hofmann *et al*, 2020; Bergeron *et al*, 2016; Koehler *et al*, 2016; Dale *et al*, 2015; Schiemann *et al*, 2016), which are expected to also target CDK19. Anti-cancer activity of CDK8/19 inhibitors has been somewhat limited, and there may be only a small therapeutic window, due to systemic toxicity. However, debate about whether CDK8/19 inhibitor toxicity is on-target, *i.e.* due to inhibition of CDK8 and CDK19 (Clarke *et al*, 2016), or off-target, due to inhibition of other kinases (Chen *et al*, 2019), continues. Furthermore, at least some CDK8/19-mediated phenotypes appear to be kinase-independent (Steinparzer *et al*, 2019; Audetat *et al*, 2017; Menzl *et al*, 2019).

Thus, despite their established roles as regulators of Mediator, and considerable interest in their therapeutic targeting in cancer, we do not yet fully understand the redundant and specific roles of CDK8 and CDK19. We therefore used gene-targeting in mice to address these questions, and determine whether their combined deletion is lethal. We confirm that knockout of the *Cdk8* gene in the intestinal epithelium has little or no effect on cell proliferation or differentiation. Furthermore, double deletion of *Cdk8* and *Cdk19* is compatible with cell proliferation in intestinal organoids. However, the double knockout reveals redundant functions in long-term control of cell proliferation and gene expression programmes. We uncover an unexpected requirement for these kinases in control of the CFTR pathway, a key player in cystic fibrosis.

Results and discussion

To evaluate the possible requirements for CDK8 for cell proliferation and survival in adult vertebrates, we designed and generated a conditional knockout allele of *Cdk8* in the mouse by Lox/Cre targeting exon 2 (Fig EV1A). This removes the critical catalytic lysine-52 and results in a frameshift that truncates over 90% of the protein. A similar conditional *Cdk8* allele was independently generated (McCleland *et al*, 2015). We studied the requirement for CDK8 in the adult mouse intestine since this is one of the most highly proliferative tissues in adults. We crossed $Cdk8^{lox/lox}$ mice to mice expressing a Tamoxifen-inducible Cre under the control of the Villin promoter (el Marjou *et al*, 2004), and verified efficient deletion of *Cdk8* in the intestinal epithelium by genotyping and Western blotting (Fig EV1C, D). Mice lacking CDK8 were healthy and did not present any phenotypes in the intestine; there was no difference in the number of proliferating cells nor in cell cycle distribution as assessed by the number of BrdU positive cells after a two-hour pulse (Fig 1A, B, Fig EV2A). The number of stem cells, goblet cells, tuft cells and Paneth cells was similar to wild-type mice, indicating that differentiation programmes were not affected (Fig 1B, Fig EV2A). The deletion was maintained after 2 months, showing that there is no counter-selection for non-recombined intestinal crypts (Fig EV2B). We performed RNA-sequencing from the intestinal epithelium of wild-type and knockout mice, but could identify only 4 genes (*Wdr72*, *Trim12a*, *Npr3*, *Trim30d*) with statistically significant expression alterations (all of which were upregulated), suggesting that CDK8 loss has only minor effects on gene expression in the intestine that are obscured by biological variability between animals.

CDK8, like CDK7, can phosphorylate PolII CTD (Rickert *et al*, 1999). Since cyclin C deletion in mice abolishes CDK8 activity yet does not affect PolII CTD phosphorylation (Li *et al*, 2014), while CTD S5 phosphorylation is normal in a non-proliferative tissue (liver) of *Cdk7* knockout mice (Ganuza *et al*, 2012), we asked whether CDK7 and CDK8 can compensate each other in CTD phosphorylation. We crossed single floxed *Cdk7* and *Cdk8* mice to generate $Cdk7^{lox/lox}$; $Cdk8^{lox/lox}$ mice, with Cre-expressed under control of a ubiquitous promoter (RPB1). While complete CDK8 loss occurred in both intestine and liver, CDK7 loss was incomplete in both tissues (Fig EV3A, B). In the intestine this incomplete deletion was expected since CDK7 is required for cell proliferation; thus, rare non-recombined crypts repopulate the epithelium (Ganuza *et al*, 2012). The strong reduction of CDK7 combined with ablation of CDK8 only led to a slight reduction in phosphorylation of PolII CTD S5, while S2 and S7 phosphorylation were normal (Fig EV3A, B). These results do not indicate a critical role for CDK8 in this phosphorylation *in vivo*, and show that a low level of CDK7 is sufficient for PolII CTD-phosphorylation.

CDK8 was described as an oncogene in colorectal cancer where it promotes beta-catenin-dependent transcription (Firestein *et al*, 2008; Morris *et al*, 2008), suggesting that its deletion should inhibit intestinal tumorigenesis. To test this, we used a colitis-associated chemical model of intestinal tumourigenesis. We treated adult control mice or mice lacking *Cdk8* with azoxymethane-dextran sodium sulphate (AOM-DSS) to chemically induce intestinal tumours, and sacrificed mice 2 months after the first DSS treatment (Fig EV4A). We observed no effect of *Cdk8* deletion on colitis-induced weight loss (Fig EV4B) and no counter-selection for *Cdk8* deletion in the tumours (Fig EV4C, D). There was no difference in number or area of tumours between WT and knockout animals (Fig 1C-E). We next tested whether CDK8 loss affects acute activation of the Wnt pathway, by concomitant homozygous deletion of the *Adenomatous polyposis coli* (*Apc*) tumour suppressor gene. Both $Apc^{-/-}$; $Cdk8^{+/+}$ and $Apc^{-/-}$; $Cdk8^{-/-}$ mice showed rapid morbidity necessitating sacrifice 5 days after Tamoxifen treatment, and hyperplastic intestinal epithelium. Intestinal epithelium lacking both *Cdk8* and *Apc* had no difference in number of Ki-67-positive cells when compared to mutant *Apc* alone, indicating similar cell proliferation. (Fig EV5A-D). Taken together, our data do not support a major role for CDK8 in intestinal tumourigenesis in the mouse.

The lack of striking phenotypes of *Cdk8* deletion in the adult intestine, along with the almost complete conservation of the kinase domain between CDK8 and CDK19 and across vertebrate species (Fig EV6A) suggested that if CDK8 has essential functions, CDK19 might be able to compensate for its loss. We therefore sought to generate a conditional double knockout. Since

double *Cdk8/Cdk19* knockout in adult mice might be lethal, we undertook a conditional deletion using intestinal organoids, which recapitulate many features of intestinal development and morphology while facilitating genetic manipulation *in vitro* (Clevers, 2016). We thus generated intestinal organoids from WT and *Cdk8*$^{lox/lox}$*; Vill::Cre*$^{ERT2/+}$ mice and disrupted *Cdk19* by CRISPR-Cas9-directed gene targeting (Fig EV6B), using retroviral transduction of Cas9 and a synthetic guide RNA-expressing plasmid. *Cdk8* removal was efficient after 7 days of Tamoxifen treatment (Fig 2A, B). Cyclin C was lost specifically in double knockout organoids (Fig 2B) and correlated with a loss of STAT1 phosphorylated on S727, a previously described CDK8 substrate (Bancerek *et al*, 2013, 727), confirming redundancy of the two kinases. Growth appeared somewhat slower only in double knockout organoids (Fig 2C, D), with a corresponding larger fraction of non-proliferating cells, as demonstrated by loss of Ki-67 staining (Fig 2E, F). Consistent with slower growth, *Cdk8* deletion was counter-selected in *Cdk19* knockout organoids, as long-term culture resulted in reappearance of CDK8 in two out of three double knockout organoid populations, presumably due to expansion of a minor unrecombined population (Fig EV7).

To determine effects of loss of the Mediator kinases on gene expression, we performed RNA-sequencing analysis of stable populations of single and double *Cdk8/Cdk19* knockout organoids. CDK8 loss had a stronger effect (716 genes upregulated, 575 downregulated) than CDK19 loss (158 up, 151 down), while double knockout organoids (1819 up, 1363 down; Fig 3A, B) revealed functional redundancy between CDK8 and CDK19 in regulating gene expression. However, most expression alterations were minor, with only 830 genes deregulated by a factor of two or more. This is consistent with previous studies, none of which have shown sweeping changes in the transcriptome upon downregulation or inhibition of CDK8 or CDK19, but rather, alteration of a limited number of specific gene sets, including super-enhancer associated genes (El Khattabi *et al*, 2019; Pelish *et al*, 2015; Galbraith *et al*, 2013; Steinparzer *et al*, 2019; Poss *et al*, 2016; Clarke *et al*, 2016). In terms of genes controlling cell proliferation, cyclin A (*Ccna2)* and cyclin B (*Ccnb1)* were slightly downregulated in knockouts, but this is likely to be a consequence rather than a cause of reduced cell proliferation. CDK8 has previously been found to regulate the p53 and c-Myc pathways (Donner *et al*, 2007; Adler *et al*, 2012), but intestinal cells lacking both kinases showed only a slight (though significant) downregulation of c-Myc, while p53 was not affected. In contrast, cyclin G1, a positive mediator of p53 responses and RB functions with a role in cell cycle arrest (Zhao *et al*, 2003), and p21 (*Cdkn1A)*, a p53 target that inhibits cyclin-dependent kinases to provoke cell cycle arrest, were more strongly upregulated.

Pathway analysis in double knockout organoids unexpectedly revealed a significant alteration of genes also modulated in intestinal knockouts of the Cystic Fibrosis Transmembrane conductance Regulator, CFTR (Fig 3C). In double knockouts, expression of genes involved in mucus production, *Muc2, Muc3, Muc13, Nlrp6, Agr2, Gcnt4, Tff1*, were upregulated, while *Cftr* was reduced. We validated changes of selected genes by qRT-PCR (Fig 3C, D). The loss of *Cftr* mRNA was also reflected at the protein level, since CFTR protein was lost in double mutant organoids (Fig 3E).

Cystic fibrosis is a disease of mucosal epithelia which also affects the intestine, and is characterised by excessive mucus accumulation and frequent inflammation (Ehre *et al*, 2014). We thus wanted to see whether the transcriptome alterations in mutant intestinal organoids translate into a cystic fibrosis-like phenotype. Staining mucin polysaccharides by periodic acid-Schiff (PAS) showed intense mucus accumulation in goblet cells specifically in *Cdk8*$^{-/-}$ *Cdk19*$^{-/-}$ organoids (Fig 4A, B). Time-lapse video-microscopy confirmed accelerated mucus release from double mutant organoids (Fig 4C, D). Functionality of CFTR can be tested in intestinal organoids using forskolin, an adenylate cyclase activator that, if CFTR is functional, induces luminal fluid secretion and organoid swelling (Dekkers *et al*, 2013). We found that forskolin caused swelling of wild-type but not double mutant organoids (Fig 4E, F; Movie EV1), indicating that CFTR downregulation upon loss of CDK8 and CDK19 functionally recapitulates the *Cftr* mutant phenotype.

These results implicate CDK8 and CDK19 as functionally redundant regulators of the CFTR pathway in the small intestine. Since transcriptional regulation by the Mediator CKM module might

be in part independent of CDK8/19 kinase activity, we investigated whether inhibiting CDK8/19 would recapitulate their genetic disruption. We treated wild-type organoids with the CDK8/CDK19 inhibitor, Senexin B, for 1h or 24h before adding forskolin. We reasoned that if effects of CDK8/19 inhibition depend on transcriptional changes, they might take 24h to become detectable, whereas if they depend only on post-transcriptional regulation of CFTR, they might be seen after 1h. Fig 4F shows that there is a dose-dependent reduction of swelling after 24h, but not 1 hour, of Senexin B treatment, implicating that loss of CDK8/19 kinase activity recapitulates a *Cftr*-mutant phenotype. To see whether this correlates with downregulation of CFTR expression, we performed qRT-PCR analysis on organoids treated with Senexin B over a time course. We found that Senexin B treatment for 24h leads to the downregulation of *Cftr* and upregulation of *Muc3* expression (Fig 4G) seen upon genetic ablation of both *Cdk8* and *Cdk19*, indicating that kinase activity of CDK8/19 controls CFTR pathway gene expression.

This study shows that Mediator kinases are both functionally redundant and largely dispensable for cell survival, proliferation, and differentiation, but may be essential for regulation of specific gene sets in particular cell types; in this case, the CFTR pathway in the intestinal epithelium. Nevertheless, while cell proliferation defects have not previously been reported in HCT-116 cancer cells lacking both kinases (Koehler *et al*, 2016) our data suggest that in the intestinal epithelium, cells devoid of both CDK8 and CDK19 have an increased tendency to become quiescent, implying that they provide a growth advantage.

Our *in vivo* results do not support an oncogenic role for CDK8 in intestinal tumourigenesis, in contrast to early *in vitro* studies (Firestein *et al*, 2008; Morris *et al*, 2008). Another recent study using the heterozygous germline Apc^{min} mutant model of intestinal tumourigenesis also concluded that *Cdk8* deletion does not hinder tumourigenesis; on the contrary, in this model, while there was no difference in micro-adenoma formation, detectable increases in tumour number, size and fraction of proliferating cells were observed upon deletion of *Cdk8* (McCleland *et al*, 2015). The reasons for the slight difference in effects of *Cdk8* deletion between chemical carcinogenesis and Apc^{min} mutation on tumours are currently unclear, but, taken together, these studies suffice to conclude that *Cdk8* has neither oncogenic nor strong tumour suppressor activity in the mouse intestine.

Our results also do not support an essential role for Mediator kinases in general gene expression, in contrast to Mediator itself, since relatively few genes were highly deregulated in double knockouts. Slightly more genes were upregulated than downregulated upon loss of either CDK8 alone or both kinases, while effects of combined deletion were more than additive of effects of single deletions, indicating functional redundancy. Organoid growth and differentiation were not prevented by knockout of both kinases, indicating that, generally, CDK8 and CDK19 are not essential for implementation of new transcriptional programmes. However, we found that there was a strong overlap between transcriptome changes of double knockout organoids and intestinal knockout of the gene encoding CFTR, a chloride and bicarbonate ion-channel that regulates fluid homeostasis in epithelia, and whose mutation causes cystic fibrosis (CF), a disease associated with mucus retention and inflammation of epithelia. Double knockout organoids showed increased mucin expression and strong accumulation of mucins in goblet cells, coupled with a precocious secretion of mucus, as well as a lack of forskolin-induced swelling, which depends on CFTR (Dekkers *et al*, 2013), indicating that CDK8/19 regulate fluid and/or mucus homeostasis. This appears to depend on their kinase activity, as specific inhibition of both kinases for 24 hours using Senexin B impaired forskolin-induced swelling in a dose-dependent manner. Since acute CDK8/19 inhibition in organoids for one hour prior to the forskolin assay had no effect, this appears to be due to transcriptional downregulation of *Cftr*. Whether or not CDK8 or CDK19 are implicated in the pathogenesis of cystic fibrosis, however, remains an open question. Almost all CF patients harbour genetic mutations in the *CFTR* gene, yet identical mutations do not have identical disease severity, and the variability between patients is associated with different genetic loci (Wright *et al*, 2011; Corvol *et al*, 2015). There are also significant but variable gene expression alterations in *CFTR*

mutant cells and upon therapeutic interventions, some of which may influence disease phenotypes (Hodos *et al*, 2020). The genes encoding CDK8, CDK19 and cyclin C have not so far been associated with CF. *CDK19* is downregulated upon several model therapeutic interventions, including overexpression of the micro-RNA miR-138, which promotes CFTR expression (Hodos *et al*, 2020; Ramachandran *et al*, 2012). However, in our study, CDK19 knockout alone was insufficient to cause a CF phenotype in the intestine, suggesting that variation in CDK19 expression does not affect CFTR. Identifying the mechanisms by which CDK8 and CDK19 affect expression of genes in the CFTR pathway will be important to better understand the pathophysiology of cystic fibrosis, but will require further studies.

Materials and Methods

Animal studies.

All animal experiments were performed in accordance with international ethics standards and were subjected to approval by the Animal Experimentation Ethics Committee of Languedoc Roussillon.

Cdk8 *conditional knockout mice.*

Cdk8$^{lox/lox}$ mice were generated as follows: An 8076 bp genomic fragment (mouse chromosome 5: 146,254,503 to 146,262,579) enclosing the essential exon 2 (whose deletion results in loss of the essential catalytic lysine residue and causes a frameshift truncating over 90% of the protein) of the CDK8 gene was amplified by PCR from genomic DNA of 129/Sv embryonic stem cells and cloned into pGEM-T-easy. The diphtheria toxin A gene was cloned into the SacII site. 64 bp to the 3′ of exon 2, the sequence CTCTAT was mutated to CTCGAG, generating an XhoI site. LoxP sites flanking exon 2 were generated by a combination of conventional cloning and recombineering, using a recombineering approach (Liu *et al*, 2003). The loxP PGK-Neo cassette was amplified from pL452 plasmid with flanking AvrII/HindIII sites at each end and cloned into the AvrII site upstream of exon 2. Fragment orientation was confirmed by the generation of 3.5 kb HindIII and 2.0 kb NheI sites, and the vector was recombined in *E. coli* strain SW106 with inducible Cre recombinase expression followed by HindII digestion, generating a single loxP site upstream of exon 2. Into this recombined vector, the FRT-PGK-Neo-FRT-LoxP cassette (amplified from pL451 with flanking XhoI sites) was cloned in the newly generated XhoI site downstream of exon 2, resulting in the "deletion construct". The orientation was confirmed by the generation of 2.2 kb Nhe1 and 3.4 kb BamHI sites. Functionality of the two recombination sites was tested as follows: the FRT site was confirmed by recombination in *E. coli* strain SW105 with inducible FlpE recombinase expression, deleting the FRT-Neo cassette and generating a 1.4 kb BamHI fragment; the resulting plasmid was transformed in *E. coli* strain SW106 with inducible Cre recombinase expression, deleting exon 2 and resulting in a 1.1 kb BamHI fragment. The NotI linearised fragment of the deletion construct was transfected by electroporation into 129/Sv embryonic stem cells. 244 Neomycin-resistant colonies were genotyped by PCR and Southern blotting. Two probes were used: one outside the 3′ end of the deletion construct, with HindIII digestion site giving a single 9kb fragment for the WT and a 7kb fragment for the correctly-integrated deletion cassette, and one to the 5′ end of the deletion cassette, again giving the same 9kb fragment for the Wt but a 3.5 kb fragment for the deletion cassette. 10 colonies showed a correct integration by homologous recombination. These ES cells were injected into blastocysts obtained from pregnant BALB/C mice, and chimeric mice were crossed with C57/Bl6J mice constitutively expressing

FlpE recombinase, removing the FRT-Neo cassette. Agouti mice were genotyped by PCR, showing correct insertion of the LoxP sites around exon 2.

Cdk8 $^{lox/lox}$ mice were crossed with *Villin-Cre-*$^{ERT2\ +/-}$ mice to obtain *Cdk8* $^{lox/lox}$, *Villin-Cre-*$^{ERT2\ +/-}$.

Tamoxifen treatment of mice to induce Lox recombination.

Mice were first injected intraperitoneally (IP) with 100µl of 20mg/ml tamoxifen solution (in corn oil). After the injection, they were fed during 5 days with cookies containing 400mg tamoxifen citrate per kg diet (Envigo, Ref TD.130859).

Apc/Cdk8 conditional knockout mice.

C57BL/6 *Apc* $^{lox/lox}$ mice (Colnot *et al*, 2004) were provided by Philippe Jay (IGF, Montpellier). These mice were crossed with *Cdk8* $^{lox/lox}$*/Villin-Cre-*$^{ERT2\ +/-}$ to obtain *Apc* $^{lox/lox}$*/Cdk8* $^{lox/lox}$*/Villin-Cre-*$^{ERT2\ +/-}$ mice. IP injection with tamoxifen during 5 days induced *Cdk8* exon 2 deletion and *Apc* exon 14 deletion in the intestines of mice containing the *Villin-Cre-*^{ERT2} gene. Small intestine and colon samples from these mice were genotyped and analysed by IHC and Western blotting.

Cdk7/Cdk8 conditional knockout mice.

Cdk8$^{lox/lox}$ mice were crossed with RERT mutant mice expressing the inducible Cre-ERT2 from the endogenous *Polr2a* locus (Guerra et al, 2003). The Cdk8$^{lox/lox}$ RERT mice were then crossed with Cdk7$^{lox/lox}$ mice (Ganuza *et al*, 2012) to obtain Cdk8$^{lox/lox}$/Cdk7$^{lox/lox}$, RERT mice in which CDK8 and CDK7 proteins should be removed from the whole body after tamoxifen treatment. Animals were sacrificed after tamoxifen treatment and intestinal epithelium was collected as indicated in the Sample preparation section below. Proteins were extracted and analysed by Western blotting.

AOM/DSS-induced colon carcinogenesis.

11 *Cdk8* $^{lox/lox}$ and 11 *Cdk8* $^{lox/lox}$*/Villin-Cre-*$^{ERT2\ +/-}$ mice were treated with tamoxifen as described above. 4 days later, mice (*Cdk8* $^{lox/lox}$ and *Cdk8* $^{-/-}$) were given a single intraperitoneal injection of AOM (10mg/kg in 0.9% saline; A5486, Sigma-Aldrich); 5 days later, 2.5% Dextran Sodium Sulfate (DSS; MP Biomedicals) was administered in the drinking water during 5 consecutive days. DSS treatment was repeated two more times with 16 days intervals without DSS for recovery (see scheme, Fig. EV3A). Mice were sacrificed 16 days after the third DSS treatment. Colons were flushed with PBS and either used for intestinal epithelium extraction (see Sample preparation for details) or used for IHC studies. Colons used for IHC were fixed overnight in neutral buffered formalin (10%) before paraffin embedding. Briefly, 4µm thick sections were dewaxed in xylene and rehydrated in graded alcohol baths. Slides were incubated in 3% H_2O_2 for 20 min and washed in PBS to quench endogenous peroxidase activity. Antigen retrieval was performed by boiling slides for 20 min in 10mM sodium citrate buffer, pH 6.0. Nonspecific binding sites were blocked in blocking buffer (TBS, pH 7.4, 5% dried milk, 0.5% Triton X-100) for 60 min at RT. Sections were incubated with anti-β-catenin antibody diluted in blocking buffer overnight at 4°C. Envision+(Dako) was used as a secondary reagent. Signals were developed with Fast DAB (Sigma-Aldrich). After dehydration, sections were mounted in Pertex (Histolab), imaged using the Nanozoomer-XR Digital slide Scanner C12000-01 (Hamamatsu) and analysed using NDP.view 2 program (Hamamatsu).

Small intestine organoids.

Cdk8 $^{lox/lox}$ and *Cdk8* $^{lox/lox}$*/Villin-Cre-*^{ERT2} mice were used to obtain small intestine organoids. Establishment, expansion and maintenance of organoids were performed as described previously (Sato *et al*, 2009). To induce the Cre-mediated recombination of *Cdk8 in vitro*, organoids were cultured during 7 days in medium supplemented with 600nM 4-Hydroxytamoxifen (Sigma H7904) resuspended in ethanol. Evaluation of knockout efficiency was performed using genotyping, qPCR and Western blotting.

CRISPR/Cas9-mediated genome editing was employed to remove CDK19 from the organoids. CRISPR single guide RNA (sgRNA) targeting murine *Cdk19* sequence (5'-AAAGTGGGACGCGGCACCTA-3', from Zhang lab database) was cloned as synthetic dsDNA into lentiCRISPRv2 vector as described ((Sanjana *et al*, 2014); provided by F. Zhang, Addgene plasmid #52961). Lentiviruses encoding the sgRNA targeting sequence were produced in HEK 293T cells transfected with LentiCRISPRv2 (+sgRNA *Cdk19*), pMD2.G and psPAX2. The viral supernatant (collected in organoids culture media) was passed through a 0.45-µm filter and used the same day for infection. Lentiviral-mediated transduction and antibiotic selection was performed as described previously (Onuma *et al*, 2013). Briefly, for lentiviral infection, organoids (5 days after seeding) were diluted into 10ml of PBS and dissociated into single cells by passing them 10-15 times through a needle with an insulin syringe. A volume containing 1–5 x 10^5 intestinal cells was centrifuged at 300g for 5 minutes and resuspended with 1ml of the viral supernatant produced in HEK 293T cells. This mixture (virus + single stem cells) was layered on top of a Matrigel-covered well (12 well plate). 24 hours later, virus and dead cells containing media were removed and the Matrigel-attached cells were covered with 200µl of Matrigel + 200µl of culture medium to create a "sandwich" containing the infected cells inside. After polymerisation of the second Matrigel layer, 1 ml of organoid media per well was used to allow organoid formation inside the Matrigel. 24 hours later, Puromycin was added (5µg/mL) and selection was conducted for 4 days. Once the organoids appeared (4-5 days after seeding the infected single cells), CDK19 knockout was verified by Western blot. We observed that CDK19 protein was still present, albeit decreased; therefore, we picked individual organoids and allowed them to growth in separated wells until we obtained several populations where CDK19 protein was completely absent, as seen by Western blot. DNA sequencing confirmed the deletion of a fragment of DNA around the sequence corresponding to *Cdk19* sgRNA, and qPCR confirmed the absence of *Cdk19* mRNA.

Mouse intestine epithelium and organoid sample preparation.

For intestine epithelium samples, a fragment of intestine was cut and flushed with PBS. It was incubated for 5-10 minutes in EDTA-containing buffer (500ml RPMI, 20mM Hepes, 1% Penicillin/Streptomycin (Sigma P4333), 12.5µg/ml DTT, 2mM EDTA pH 7.4 and 10% FBS) to allow easy detachment of the intestine epithelium. The intestinal tube was opened longitudinally and put over a horizontal plate to allow scrapping of the epithelial cells by trawling two needles in opposite directions over the tissue. Cells were recovered from the plate by wetting them with a small volume (200µl) of PBS. They were put into an Eppendorf tube where they were spun down to remove most of the PBS. Pellets were snap frozen and conserved at -80°C.

Frozen intestinal epithelium samples were resuspended in 250µl of lysis buffer with protease and phosphatase inhibitors (5mM Tris, pH 7.4, 100mM NaCl, 50mM NaF, 40mM beta-glycerophosphate, 2.5mM Na-Vanadate, 5mM EDTA, 1mM EGTA, 1mM DTT, 1%

Triton X-100 and Protease inhibitor cocktail diluted 1/400 (Sigma P8340)) and, after addition of 100µl of stainless-steel beads (0.2 mm diameter, 1lb, Next Advance, SSB02), they were disrupted in a bullet blender storm 24 (Next Advance) by shaking during 4 minutes at 4°C with an intensity level of 8. The lysate was incubated during 20 more minutes at 4°C (without shaking) and the solubilised proteins were recovered from the supernatant by centrifugation at 16000g for 20 minutes at 4°C, frozen in liquid nitrogen and stored at -80°C until use. Protein concentrations were determined by BCA protein assay (Pierce Biotechnology).

For organoids samples, Matrigel was disrupted by pipetting up and down several times the media in each well over the dome of Matrigel. This mix was spun down at 200g for 5 minutes at 4°C and the pellet was washed twice with 1ml of PBS. Pellets were snap frozen in liquid nitrogen and kept at -80°C until use.

For organoids extracts, frozen pellets were lysed by incubation at 4°C for 20 minutes in lysis buffer with protease and phosphatase inhibitors (50mM Tris, pH 7.4, 100mM NaCl, 50mM NaF, 40mM beta-glycerophosphate, 2.5mM Na-Vanadate, 5mM EDTA, 1mM EGTA, 1mM DTT, 1% Triton X-100 and Protease inhibitor cocktail (Sigma P8340) diluted 1/400. The solubilised proteins were recovered from the supernatant after centrifugation at 16000g for 20 minutes at 4°C, snap frozen in liquid nitrogen and stored at -80°C until use. Protein concentrations were determined by BCA protein assay (Pierce Biotechnology).

Western-blotting.

For intestinal epithelium and organoids, 30µg of of total proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (7% 10% and 12.5% gels) and transferred to Immobilon membranes (Milipore) at 90 volts for 120 min with a wet-blotting apparatus. Membranes were blocked in TBS-T pH 7.6 (20mM Tris, 140mM NaCl, 0.1% Tween-20) containing non-fat dry milk (5%), incubated with the primary antibody in TBS-T + 3% BSA for 2 hours at RT or over-night at 4°C, washed 3 times with TBS-T for a total of 15 minutes, incubated with secondary antibody at 1/10000 dilution in TBS-T + 5% nonfat dried milk for 30 minutes at RT, and washed 3 times in TBS-T for a total of 15 minutes. Signals were detected using Western Lightning Plus-ECL (PerkinElmer) and Amersham Hyperfilm^{TM} ECL (GE Healthcare).

RNA isolation and qRT-PCR.

RNA was extracted from organoids and purified using RNeasy mini kit (Qiagen) according to manufacturer's protocol. For reverse transcription (cDNA synthesis), 1µg of purified RNA in total volume of 13µl, extracted by RNeasy Mini Kit (Qiagen), was mixed with 1µl of 10mM dNTPs mix (LifeTechnologies) and 1µl of 10µM oligo(dT) 20-primer. Samples were incubated at 65°C for 5 minutes, transferred to ice. 4µl of 5x First Strand Buffer, 1µL of 100mM DTT, 1µl of RNase OUT RNase Inhibitor (Invitrogen) and 1µl of SuperScript® III Reverse Transcriptase (LifeTechnologies) were added to each sample and incubated at 50°C for 1 hour. The reaction was inactivated at 70°C for 15 minutes.

qPCR was performed using LightCycler® 480 II (Roche). The reaction contained 5µl of a 1/10 dilution of the cDNA obtained after RT (25ng of cDNA), 1µl of 10µM qPCR primer pair, 10µl 2x Master Mix in a final volume made up to 20µl with DNase free water. qPCR was conducted at 95°C for 5 min, followed by 45 cycles of 95°C 10s, 59°C 20s and 72°C 15s. The specificity of the reaction was verified by melting curve analysis. β2-mioglobulin (B2M) was used as housekeeping gene.

Genomic DNA extraction and genotyping.

For colon tumors, genomic DNA was extracted using the KAPA Mouse genotyping kit from Clinisciences (KK7352) directly over 3-4 tumors that had been removed from the colon before the extraction of intestinal epithelium.

Genomic DNA was extracted from intestinal epithelium pellets or organoids using KAPA Mouse genotyping kit from Clinisciences (KK7352).

The forward and revers primers used for amplification of the *Cdk8* wild-type allele (650 bp), the *Cdk8*$^{Lox/Lox}$ (850 bp) and the fragment obtained after tamoxifen-induced Cre recombination of *Cdk8*$^{Lox/Lox}$ (*Cdk8* Lox + Cre (-)) (340 bp) are indicated in the table below as CDK8 Fw (genotyping) and CDK8 Rev (genotyping). These fragments were amplified from genomic DNA using the following PCR protocol: 95°C for 5 min; 35 cycles at 95°C for 15 s; 57°C for 15 s; 72°C for 1min 30 s. Amplified DNA fragments were migrated in a 1,5% agarose gels and stained with Ethidium bromide for detection.

Primers used for qPCR and genotyping.

Gene name	Sequence	Use
Mouse B2M Fw Mouse B2M Rev	5'-GGTCTTTCTGGTGCTTGTCT-3' 5'-GCAGTTCAGTATGTTCGGCTT-3'	(qRT-PCR)
CDK8 (1,2)-3 Fw CDK8 (1,2)-3 Rev	5'-GTGGGAGA@AGGAAGGACGAT 5'-GCCATACTTTCCGATCAGCA-3'	(qRT-PCR)
CDK19 (5-6) Fw CDK19 (5-6) Rev	5'-TTCTCCCCTAAAGCCACTCG-3' 5'-ATGGGTTCTGAAGTCAAGAGTT-3'	(qRT-PCR)
Cyc C (10-11) Fw Cyc C (10-11) Rev	5'-CCGAAACCAAAACCACCTCC-3' 5'-TCCCAATATGCTTGACAGAAACA-3'	(qRT-PCR)
CFTR (1-2) Fw CFTR (1-2) Rev	5'-TAAAAGGGACGAGCCAAAAG-3' 5'-CCCTTTCCTCAAAATTGGTG-3'	(qRT-PCR)
Muc2 (15-16) Fw Muc2 (15-16) Rev	5'-AACAACGAGGACTGCATGTG-3' 5'-ACAGGTGCAAATCCCTTGAG-3'	(qRT-PCR)
Muc3 (2-3) Fw Muc3 (2-3) Rev	5'-GTCCGTGGAAGTGAGTGTGA-3' 5'-ATAACCCCTTCATACTCCGGTA-3'	(qRT-PCR)
CDK8 Fw	5'-ACATGCCTTACAGCCTAGTCTTAC-3'	(genotyping)
CDK8 Rev	5'-CCAAATAAATGTATACTCTGCAAG-3'	(genotyping)
CDK19 CRISP-R22 Fw	5'-GAGGAGTCCCTTGCTGAAG-3'	(sequencing)
CDK19 CRISP-R22 Rev	5'-CAGTGCCTCCGAGTTAGC-3'	(sequencing)
APC Genotyping Fw	5'-CTGTTCTGCAGTATGTTATCA-3'	(genotyping)
APC Genotyping Rev	5'-CTATGAGTCAACACAGGATTA-3'	(genotyping)
Villin-Cre-^{ERT2} Fw	5'-CAAGCCTGGCTCGACGGCC-3'	(genotyping)
Villin-Cre-^{ERT2} Rev	5'-CGCGAACATCTTCAGGTTCT-3'	(genotyping)

Antibodies.

Name	Clone	Source	Species	Cat#
CDK8 (C19)	Polyclonal	Santa Cruz	Goat	sc-1521
CDK19	Polyclonal	Sigma	Rabbit	HPA007053
Cyclin C	Polyclonal	gift	Rabbit	
PCNA	Monoclonal	Lab Vision Corporation	Mouse	#MS-106-PO
Lysozyme	Polyclonal	Dako	Rabbit	A0099
Anti-Dclk1	Polyclonal	Abcam	Rabbit	Ab31704
Olfm4	Monoclonal	Cell signaling	Rabbit	39141
β-Catenin	Monoclonal	BD BioSciences	Mouse	610154
BrdU	Monoclonal	DSHB	Mouse	G3G4
RNA pol II Ser2P	Monoclonal	JC Andrau's lab	Rat	(3E10)
RNA pol II Ser5P	Monoclonal	JC Andrau's lab	Rat	(3E8)
RNA pol II Ser7P	Monoclonal	JC Andrau's lab	Rat	(4E12)
Ki-67	SP6	Spring Bioscience	Rabbit	M3064
p53	CM5	Novocastra	Rabbit	NCL-p53-CM5p
Cleaved Caspase 3	ASP176	Cell signaling	Rabbit	9661S
Stat1	Monoclonal	BD transduction laboratories	Mouse	#610115
Phospho-Stat1 (Ser727)	Polyclonal	Cell signaling	Rabbit	#9177
CFTR	Monoclonal	Abcam	Mouse	ab2784
β-Actin	Monoclonal	Sigma	Mouse	a5441
GAPDH	Polyclonal	Sigma	Rabbit	G9545

CycC purified antibody. Rabbit anti-Cyclin C serum was a kind gift from Jacques Piette (IGMM Montpellier, France (Barette *et al*, 2001)). Cyclin C specific antibodies were purified from serum by incubation with a membrane containing Cyclin C protein (ProQinase GmbH). The antibodies were eluted with 200µl of 0,2M glycine pH 2.5 and neutralised rapidly with 21µl of 1M Tris. RNA Pol II Ser2P (3E10), RNA Pol II Ser5P (3E8) and RNA Pol II Ser7P (4E12) antibodies were a kind gift from Jean-Christophe Andrau's lab (IGMM, CNRS Montpellier, France;(Chapman *et al*, 2007)).

Immunohistochemistry (IHC) and tissue staining.

Whole intestines were flushed with PBS and turned inside-out on a wooden stick. They were collected and fixed 24h in neutral buffered formalin 10%, dehydrated, and embedded in paraffin.

Organoids were collected and fixed for 1h in 4% paraformaldehyde at room temperature. They were washed with PBS (x2) and resuspended into 100µl of Histogel

(Fisher Scientific, Ref 12006679), previously thawed in a hot water bath at 60°C. Each drop containing the organoids and Histogel was dried on top of a flat surface and embebed in paraffin. Paraffin-embedded intestines or organoids were cut into 3-µm-thick sections, mounted on slides, then dried at 37°C overnight. Tissue sections were stained with hematoxylin-eosin (HE) with HMS 740 autostainer (MM France) for preliminary analysis. For mucosubstances, tissue sections were stained with Periodic Acid Schiff's (PAS) staining (J. Bancroft & A. Stevens, 1982).

For Ki67, p53 and Caspase 3, IHC was performed as described previously (Rahmanzadeh *et al*, 2007), on a VENTANA Discovery Ultra automated staining instrument (Ventana Medical Systems), using VENTANA reagents, according to the manufacturer's instructions. Briefly, slides were de-paraffinised, then epitope retrieval was performed with CC1 solution (cat# 950-124) at high temperature (95-100°C) or for CC2 solution (cat# 950-123) at high temperature (91°C) for a period time that is suitable for each specific antibody. Endogenous peroxidases were blocked with Discovery Inhibitor (cat# 760-4840).

For CDK8 immunostaining, signal enhancement was performed using the rabbit antibody anti-goat (Vector Laboratories, cat#BA-5000, 1:2000,) for 16min at 37°C then using DISCOVERY OmniMap anti-rabbit HRP detection Kit (cat# 05269679001) according to the manufacturer's instructions. The same kit was used for all other rabbit primary antibodies to amplify the signal. Slides were incubated with DAB (cat# 05266645001), then counterstained with hematoxylin II (cat# 790-2208) for 8 min, followed by Bluing reagent (cat# 760-2037) for 4 min. Slides were then dehydrated with Leica autostainer and coverslipped with Pertex mounting medium with CM6 coverslipper (Microm). Brightfield stained slides were digitalised with a Hamamatsu NanoZoomer 2.0-HT scanner and images were visualised with the NDP.view 1.2.47 software.

For Paneth cells, tuft cells, β-catenin and Brdu detection, the IHCs were performed manually. After deparaffination and rehydration, demasking of antigenic sites was performed by boiling the slides for 20' in 10mM Na-Citrate pH 6.4. After cooling down, slides were treated with 3% H_2O_2, 5' at RT for peroxidase inhibition. Samples were blocked in blocking solution (TBS, 0.5% Triton, 5% dry milk) for 20'. First antibody was diluted 1/400 in blocking solution and the slides were incubated O/N at 4°C in a humid chamber. Slides were then washed with TBS + 0.1% Tween 20 (3 times) and once with TBS without Tween. Secondary antibodies (ImmPRESS™ reagent kit peroxidase anti-rabbit (MP-7451) or anti-Mouse (MP-7402), from Vector Laboratories) were incubated for 30' at RT. Slides were washed twice with TBS + 0.1% Tween 20 and the final wash was done in H_2O. Peroxidase staining was performed using Sigma Fast DAB tablet set (D4293-50SET). After Hematoxylin staining of the nucleus with Gill's Hematoxylin solution N°2 (CAS 517-28-2) from Santa Cruz (SC-24973), slides were rehydrated and mounted in Pertex® (Histolab 00811).

Sequence alignment.

CDK8 and CDK19 protein sequences from different species were aligned using the BoxShade server (https://embnet.vital-it.ch/software/BOX_form.html).

RNA sequencing.

After 7 days of treatment with 600nM 4-Hydroxytamoxifen, organoids were lysed and RNA was extracted following the Trizol RNA isolation protocol (W.M. Keck Foundation Biotechnology Microarryay Resource Laboratory at Yale University) until the end of the

phase-separation step. Total RNA cleanup with DNase digestion was performed by addition of 1.5 volumes of absolute ethanol on top of the aqueous phase obtained after the phase-separation and following the Qiagen RNeasy protocol (W.M. Keck Foundation Biotechnology Microarryay Resource Laboratory at Yale University). RNA integrity was analysed on Agilent 2100 bioanalyzer. All conditions were prepared as biological triplicates and sent to BGI Tech Solutions (HongKong)- Co for library preparation and RNA sequencing. Purification of mRNA from total RNA was achieved using oligo(dT)-attached magnetic beads and then fragmented for random hexamer-primed reverse transcription, followed by a second-strand cDNA synthesis. Sequencing was performed with the BGISEQ, SE50 platform to obtain an average of 50 million single-end, 50bp reads per sample.

Bioinformatic analysis.
The raw reads obtained in fastq format were subject to quality control using the FastQC software. The reads passing the quality control were aligned to the mouse reference genome (GRCm38.p6) and the counts per gene were quantified using the tool STAR 2.6.0a(2). The Ensembl mouse genome annotations (release 93) were used to map the reads to each gene and their corresponding transcripts. Differential gene expression analysis was performed in R using the DESeq2 library. MA-plots showing the deferentially expressed genes were generated with an in-house script and the gene set enrichment analysis were performed using the "enrichR" library, in both cases using the R programming language.

Data availability.
The RNA-sequencing data have been deposited in the Gene Expression Omnibus (GEO, NCBI) repository, and are accessible through GEO Series accession number GSE138808.

CRISPR Cas9-targeting of Cdk19 *gene.*
CRISPR single guide RNA (sgRNA) targeting exon 1 in mouse *Cdk19* (CDK19_Mouse: 5'-AAAGTGGGACGCGGCACCTA-3' from Zhang lab database) was cloned as synthetic dsDNA into lentiCRISPRv2 vector (provided by F. Zhang, Addgene plasmid #52961) as described (Sanjana et al., 2014). Generation of lentiviral particles and infection of organoids were carried out following classical procedures, as described previously (Onuma *et al.,* 2013). Successfully infected cells were selected with puromycin (5µg/ml) during 4 days. Single organoids were picked for clonal expansion. Effects of targeted deletion were verified by sequencing. Genomic DNA was extracted from organoids using KAPA Mouse genotyping kit from Clinisciences (KK7352). The 2 primers used for amplification of the *Cdk19* allele are indicated in the table above as CDK19 CRISP-R22 Fw and CDK19 CRISP-R22 Rev (sequencing). They were amplified as follows: 95°C for 5 min; 30 cycles at 95°C for 45''; 57°C for 30''; 72°C for 1min 30 s. Absence of CDK19 protein in the organoids was also verified by Western blotting.

Time-lapse microscopy.
Organoids (Wt, $Cdk8^{Lox/Lox}$, $Cdk19^{-/-}$ *and* $Cdk8^{Lox/Lox}/Cdk19^{-/-}$) were treated with 600nM 4-hydroxytamoxifen during 7 days to induce recombination of the LoxP sites flanking *Cdk8* exon 2. Images were taken every 4 hours during 7 days on an inverted microscope (Axio Observer, Zeiss) equipped with a heated chamber allowing constant temperature (37°C) and CO_2 flow (5% CO_2). CCD camera (Princeton Instruments (Micromax), pixel = 6,7 µm), with 10x/0.3 DRY PH1objective, correction ECPLAN Neofluar, 5.2 mm working distance.

Acquisition software was MetaMorph 7.8 (Molecular Devices, LLC). Images were analysed using Image J software to calculate the time for release of mucus in the lumen of the *organoid* (observed as a dark staining in the center of the organoid).

Forskolin-induced swelling.
To remove exon 2 of *Cdk8* from *Cdk8*$^{lox/lox}$/Villin-Cre-ERT2$^{+/-}$/*Cdk19*$^{-/-}$, organoids were treated with 600nM 4-hydrxoytamoxifen for 7 days. Once the *Cdk8/Cdk19* double KO was obtained, forskolin-induced swelling was measured as indicated (Dekkers *et al*, 2013). Organoids were transferred to CELLview culture dishes PS 35/10 mm, glass bottom, 4 compartments (Greiner Bio-One, 627870), two days before imaging. Confocal spinning disk (Dragonfly, Andor, Oxford Instruments) microscope equipped with heated chamber allowing constant temperature (37°C) and CO_2 flow (5% CO_2), EMCCD iXon888 camera (Lifer Andor, pixel = 13 µm), objective 10x/0.45 DRY, correction Plan Apo Lambda, 4mm working distance, was used for imaging, with Fusion acquisition software. Images of a single organoid, previously selected, were taken every 2 minutes during 20 minutes after forskolin addition (5µM, or DMSO vehicle control) to the media. For data analysis, a macro was created using Fiji software. It consisted of recognising and filling the structures imaged through the alexa-488 track, to calculate the increase of total organoid area in single organoids over the different time points.

Statistics.
Graphs and statistical analyses were performed using Microsoft Excel 16.50 and GraphPad Prism6 using analyses described in legends.

Acknowledgements
L.A. was funded by the Montpellier FHU Cancer programme and by the Fondation pour la Recherche Medicale; E.H. was funded by the Ligue Nationale Contre le Cancer (LNCC); G.D. was funded by the National Cancer Institute (INCa); S.P., C.B-P. and F.G. are CNRS employees; D.F., P.J. and L.K. are Inserm employees. This work was undertaken with support from INCa (PLBIO10-068 and PLBIO15-005), and the LNCC (EL2013.LNCC/DF and EL2018.LNCC/DF). RHEM histology facility was supported by SIRIC Montpellier Cancer Grant INCa_Inserm_DGOS_12553, the European regional development foundation and the Occitanie region (FEDER-FSE 2014-2020 Languedoc Roussillon).

Author contributions
D.F. conceived the study with assistance from P.J. S.P., L.K. and D.F. designed experiments. S.P., E.H., L.A., A.B.A, A.C., C.B-P., N.P. and F.G. performed experiments. G.D. performed bioinformatics analysis. S.P., G.D., P.J., L.K. and D.F. interpreted data. S.P., L.K. and D.F. wrote the paper.

Conflict of interest statement
The authors declare that they have no conflict of interest.

References
Adler AS, McCleland ML, Truong T, Lau S, Modrusan Z, Soukup TM, Roose-Girma M, Blackwood EM & Firestein R (2012) CDK8 maintains tumor dedifferentiation and embryonic stem cell pluripotency. *Cancer Res* 72: 2129–39

Andrau J-C, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp MG, van de Peppel J, Werner M & Holstege FCP (2006) Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* 22: 179–192

Audetat KA, Galbraith MD, Odell AT, Lee T, Pandey A, Espinosa JM, Dowell RD & Taatjes DJ (2017) A Kinase-Independent Role for Cyclin-Dependent Kinase 19 in p53 Response. *Mol Cell Biol* 37

Bacevic K, Prieto S, Caruso S, Camasses A, Dubra G, Ursic-Bedoya J, Lozano A, Butterworth J, Zucman-Rossi J, Hibner U, *et al* (2019) CDK8 and CDK19 kinases have non-redundant oncogenic functions in hepatocellular carcinoma. *bioRxiv*: 789586

Bancerek J, Poss ZC, Steinparzer I, Sedlyarov V, Pfaffenwimmer T, Mikulic I, Dölken L, Strobl B, Müller M, Taatjes DJ, *et al* (2013) CDK8 Kinase Phosphorylates Transcription Factor STAT1 to Selectively Regulate the Interferon Response. *Immunity* 38: 250–262

Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, *et al* (2017) QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7: 16878

Barette C, Jariel-Encontre I, Piechaczyk M & Piette J (2001) Human cyclin C protein is stabilized by its associated kinase cdk8, independently of its catalytic activity. *Oncogene* 20: 551–562

Bergeron P, Koehler MFT, Blackwood EM, Bowman K, Clark K, Firestein R, Kiefer JR, Maskos K, McCleland ML, Orren L, *et al* (2016) Design and Development of a Series of Potent and Selective Type II Inhibitors of CDK8. *ACS Med Chem Lett* 7: 595–600

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E & Eick D (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* 318: 1780–1782

Chen M, Li J, Liang J, Thompson ZS, Kathrein K, Broude EV & Roninson IB (2019) Systemic Toxicity Reported for CDK8/19 Inhibitors CCT251921 and MSC2530818 Is Not Due to Target Inhibition. *Cells* 8: 1413

Clarke PA, Ortiz-Ruiz M-J, TePoele R, Adeniji-Popoola O, Box G, Court W, Czasch S, Bawab SE, Esdar C, Ewan K, *et al* (2016) Assessing the mechanism and therapeutic potential of modulators of the human Mediator complex-associated protein kinases. *eLife* 5: e20722

Clevers H (2016) Modeling Development and Disease with Organoids. *Cell* 165: 1586–1597

Colnot S, Niwa-Kawakita M, Hamard G, Godard C, Plenier SL, Houbron C, Romagnolo B, Berrebi D, Giovannini M & Perret C (2004) Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers. *Lab Invest* 84: 1619–1630

Corvol H, Blackman SM, Boëlle P-Y, Gallins PJ, Pace RG, Stonebraker JR, Accurso FJ, Clement A, Collaco JM, Dang H, *et al* (2015) Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun* 6: 8382

Dale T, Clarke PA, Esdar C, Waalboer D, Adeniji-Popoola O, Ortiz-Ruiz M-J, Mallinger A, Samant RS, Czodrowski P, Musil D, *et al* (2015) A selective chemical probe for exploring the role of CDK8 and CDK19 in human disease. *Nat Chem Biol* 11: 973–980

Dekkers JF, Wiegerinck CL, de Jonge HR, Bronsveld I, Janssens HM, de Winter-de Groot KM, Brandsma AM, de Jong NWM, Bijvelds MJC, Scholte BJ, *et al* (2013) A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nat Med* 19: 939–945

Donner AJ, Ebmeier CC, Taatjes DJ & Espinosa JM (2010) CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nat Struct Mol Biol* 17: 194–201

Donner AJ, Szostek S, Hoover JM & Espinosa JM (2007) CDK8 is a stimulus-specific positive coregulator of p53 target genes. *Mol Cell* 27: 121–33

Ehre C, Ridley C & Thornton DJ (2014) Cystic fibrosis: an inherited disease affecting mucin-producing organs. *Int J Biochem Cell Biol* 52: 136–145

El Khattabi L, Zhao H, Kalchschmidt J, Young N, Jung S, Van Blerkom P, Kieffer-Kwon P, Kieffer-Kwon K-R, Park S, Wang X, *et al* (2019) A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* 178: 1145-1158.e20

Elmlund H, Baraznenok V, Lindahl M, Samuelsen CO, Koeck PJ, Holmberg S, Hebert H & Gustafsson CM (2006) The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proc Natl Acad Sci U A* 103: 15788–93

Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S, *et al* (2008) CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 455: 547–51

Galbraith MD, Allen MA, Bensard CL, Wang X, Schwinn MK, Qin B, Long HW, Daniels DL, Hahn WC, Dowell RD, *et al* (2013) HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* 153: 1327–39

Ganuza M, Saiz-Ladera C, Canamero M, Gomez G, Schneider R, Blasco MA, Pisano D, Paramio JM, Santamaria D & Barbacid M (2012) Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion. *EMBO J* 31: 2498–510

Guerra C, Mijimolle N, Dhawahir A, Dubus P, Barradas M, Serrano M, Campuzano V & Barbacid M (2003) Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer Cell* 4: 111–120

Hodos RA, Strub MD, Ramachandran S, Li L, McCray PB & Dudley JT (2020) Integrative genomic meta-analysis reveals novel molecular insights into cystic fibrosis and ΔF508-CFTR rescue. *Sci Rep* 10: 20553

Hofmann MH, Mani R, Engelhardt H, Impagnatiello MA, Carotta S, Kerenyi M, Lorenzo-Herrero S, Böttcher J, Scharn D, Arnhof H, *et al* (2020) Selective and Potent CDK8/19 Inhibitors Enhance NK-Cell Activity and Promote Tumor Surveillance. *Mol Cancer Ther* 19: 1018–1030

J. Bancroft & A. Stevens (1982) Theory and Practice of Histological Techniques. In pp 188–190. New York: Churchill Livingstone

Jeronimo C, Langelier M-F, Bataille AR, Pascal JM, Pugh BF & Robert F (2016) Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Mol Cell* 64: 455–466

Kapoor A, Goldberg MS, Cumberland LK, Ratnakumar K, Segura MF, Emanuel PO, Menendez S, Vardabasso C, Leroy G, Vidal CI, *et al* (2010) The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature* 468: 1105–9

Knuesel MT, Meyer KD, Donner AJ, Espinosa JM & Taatjes DJ (2009) The human CDK8 subcomplex is a histone kinase that requires Med12 for activity and can function independently of mediator. *Mol Cell Biol* 29: 650–61

Koehler MF, Bergeron P, Blackwood EM, Bowman K, Clark KR, Firestein R, Kiefer JR, Maskos K, McCleland ML, Orren L, *et al* (2016) Development of a Potent, Specific CDK8 Kinase Inhibitor Which Phenocopies CDK8/19 Knockout Cells. *ACS Med Chem Lett* 7: 223–228

Kuchin S, Yeghiayan P & Carlson M (1995) Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast. *Proc Natl Acad Sci U A* 92: 4006–10

Li N, Fassl A, Chick J, Inuzuka H, Li X, Mansour MR, Liu L, Wang H, King B, Shaik S, *et al* (2014) Cyclin C is a haploinsufficient tumour suppressor. *Nat Cell Biol* 16: 1080–1091

Liao SM, Zhang J, Jeffery DA, Koleske AJ, Thompson CM, Chao DM, Viljoen M, van Vuuren HJ & Young RA (1995) A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* 374: 193–6

Liu P, Jenkins NA & Copeland NG (2003) A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* 13: 476–484

Loncle N, Boube M, Joulia L, Boschiero C, Werner M, Cribbs DL & Bourbon HM (2007) Distinct roles for Mediator Cdk8 module subunits in Drosophila development. *EMBO J* 26: 1045–54

Lynch CJ, Bernad R, Martínez-Val A, Shahbazi MN, Nóbrega-Pereira S, Calvo I, Blanco-Aparicio C, Tarantino C, Garreta E, Richart-Ginés L, *et al* (2020) Global hyperactivation of enhancers stabilizes human and mouse naive pluripotency through inhibition of CDK8/19 Mediator kinases. *Nat Cell Biol* 22: 1223–1238

Malik S & Roeder RG (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet* 11: 761–72

el Marjou F, Janssen KP, Chang BH, Li M, Hindie V, Chan L, Louvard D, Chambon P, Metzger D & Robine S (2004) Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* 39: 186–93

McCleland ML, Soukup TM, Liu SD, Esensten JH, de Sousa e Melo F, Yaylaoglu M, Warming S, Roose-Girma M & Firestein R (2015) Cdk8 deletion in the Apc(Min) murine tumour model represses EZH2 activity and accelerates tumourigenesis. *J Pathol* 237: 508–519

McDermott MSJ, Chumanevich AA, Lim C, Liang J, Chen M, Altilia S, Oliver D, Rae JM, Shtutman M, Kiaris H, *et al* (2017) Inhibition of CDK8 mediator kinase suppresses estrogen dependent transcription and the growth of estrogen receptor positive breast cancer. *Oncotarget* 8

Menzl I, Zhang T, Berger-Becvar A, Grausenburger R, Heller G, Prchal-Murphy M, Edlinger L, Knab VM, Uras IZ, Grundschober E, *et al* (2019) A kinase-independent role for CDK8 in BCR-ABL1+ leukemia. *Nat Commun* 10: 4741

Morris EJ, Ji JY, Yang F, Di Stefano L, Herr A, Moon NS, Kwon EJ, Haigis KM, Naar AM & Dyson NJ (2008) E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. *Nature* 455: 552–6

Nakamura A, Nakata D, Kakoi Y, Kunitomo M, Murai S, Ebara S, Hata A & Hara T (2018) CDK8/19 inhibition induces premature G1/S transition and ATR-dependent cell death in prostate cancer cells. *Oncotarget* 9: 13474–13487

Onuma K, Ochiai M, Orihashi K, Takahashi M, Imai T, Nakagama H & Hippo Y (2013) Genetic reconstitution of tumorigenesis in primary intestinal cells. *Proc Natl Acad Sci* 110: 11127–11132

Pelish HE, Liau BB, Nitulescu II, Tangpeerachaikul A, Poss ZC, Da Silva DH, Caruso BT, Arefolov A, Fadeyi O, Christie AL, *et al* (2015) Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* 526: 273–276

Porter DC, Farmaki E, Altilia S, Schools GP, West DK, Chen M, Chang BD, Puzyrev AT, Lim CU, Rokow-Kittell R, *et al* (2012) Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. *Proc Natl Acad Sci U A* 109: 13799–804

Poss ZC, Ebmeier CC, Odell AT, Tangpeerachaikul A, Lee T, Pelish HE, Shair MD, Dowell RD, Old WM & Taatjes DJ (2016) Identification of Mediator Kinase Substrates in

Human Cells using Cortistatin A and Quantitative Phosphoproteomics. *Cell Rep* 15: 436–450

Postlmayr A, Dumeau CE & Wutz A (2020) Cdk8 is required for establishment of H3K27me3 and gene repression by Xist and mouse development. *Dev Camb Engl* 147

Rahmanzadeh R, Huttmann G, Gerdes J & Scholzen T (2007) Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis. *Cell Prolif* 40: 422–30

Ramachandran S, Karp PH, Jiang P, Ostedgaard LS, Walz AE, Fisher JT, Keshavjee S, Lennox KA, Jacobi AM, Rose SD, *et al* (2012) A microRNA network regulates expression and biosynthesis of wild-type and ΔF508 mutant cystic fibrosis transmembrane conductance regulator. *Proc Natl Acad Sci U S A* 109: 13362–13367

Rickert P, Corden JL & Lees E (1999) Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases. *Oncogene* 18: 1093–102

Rocha PP, Scholze M, Bleiss W & Schrewe H (2010) Med12 is essential for early mouse development and for canonical Wnt and Wnt/PCP signaling. *Dev Camb Engl* 137: 2723–2731

Sanjana NE, Shalem O & Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11: 783–784

Sato S, Tomomori-Sato C, Parmely TJ, Florens L, Zybailov B, Swanson SK, Banks CA, Jin J, Cai Y, Washburn MP, *et al* (2004) A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell* 14: 685–91

Sato T, Vries RG, Snippert HJ, Wetering M van de, Barker N, Stange DE, Es JH van, Abo A, Kujala P, Peters PJ, *et al* (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* 459: 262–265

Schiemann K, Mallinger A, Wienke D, Esdar C, Poeschke O, Busch M, Rohdich F, Eccles SA, Schneider R, Raynaud FI, *et al* (2016) Discovery of potent and selective CDK8 inhibitors from an HSP90 pharmacophore. *Bioorg Med Chem Lett* 26: 1443–1451

Steinparzer I, Sedlyarov V, Rubin JD, Eislmayr K, Galbraith MD, Levandowski CB, Vcelkova T, Sneezum L, Wascher F, Amman F, *et al* (2019) Transcriptional Responses to IFN-γ Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions. *Mol Cell* 76: 485-499.e8

Tsai KL, Sato S, Tomomori-Sato C, Conaway RC, Conaway JW & Asturias FJ (2013) A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nat Struct Mol Biol* 20: 611–9

Tsutsui T, Umemura H, Tanaka A, Mizuki F, Hirose Y & Ohkuma Y (2008) Human mediator kinase subunit CDK11 plays a negative role in viral activator VP16-dependent transcriptional regulation. *Genes Cells Devoted Mol Cell Mech* 13: 817–826

Westerling T, Kuuluvainen E & Makela TP (2007) Cdk8 is essential for preimplantation mouse development. *Mol Cell Biol* 27: 6177–82

Wright FA, Strug LJ, Doshi VK, Commander CW, Blackman SM, Sun L, Berthiaume Y, Cutler D, Cojocaru A, Collaco JM, *et al* (2011) Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet* 43: 539–546

Zhao L, Samuels T, Winckler S, Korgaonkar C, Tompkins V, Horne MC & Quelle DE (2003) Cyclin G1 has growth inhibitory activity linked to the ARF-Mdm2-p53 and pRb tumor suppressor pathways. *Mol Cancer Res MCR* 1: 195–206

Figure legends

Fig. 1. CDK8 knockout does not affect adult mouse intestine homeostasis nor chemically-induced carcinogenesis. (A) Immunohistochemical staining of CKD8 in mouse small intestines collected two months after tamoxifen treatment. $CDK8^{+/+}$ depict mice with floxed Cdk8 alleles. (B) Analysis of cell differentiation (left) and proliferation (right) in the intestine after CDK8 deletion (as in A). Olfm4, Lysozyme, PAS and Dclk1 staining was used to reveal, respectively, stem, Paneth, goblet and tuft cells. β-Catenin staining allows detection of cancer cells (cytoplasmic vs nuclear localisation). Cell proliferation was assessed by PCNA, Ki-67 and BrdU (after 1h pulse) staining. Scatter plots represent the percentage of the area stained by each antibody (relative to the area occupied by hematoxylin). For Paneth cells, BrdU, PCNA and Ki67, only crypts were analysed. For goblet cells, crypts and villi were analysed. For Tuft cells quantification, Dclk1 positive cells were counted in 50 villi. Colour code depicts small intestine (green), proximal colon (blue), and distant colon (red). Mean ± SEM is shown. P-value of unpaired two-tailed t-test is indicated (ns, not significant; $p > 0.05$). Scale bars, 25μm (Olfm4, Lysozyme, Dclk1 and β-Catenin) and 50μm (PAS, BrdU, Ki-67 and PCNA). (C-E) Analysis of mouse colon after AOM/DSS treatment. (C) Quantification of the number of neoplastic lesions (n = 10 for $Cdk8^{+/+}$, and n = 7 for $Cdk8^{-/-}$ mice). P-value of umpaired t-test is indicated: ns, not significant ($p > 0.05$). Mean ± SD is shown. (D) Quantification of the percentage of the colon surface occupied by tumours. Intestine samples were stained for β-Catenin, and tumour regions with nuclear β-Catenin localisation were quantified using NDP.view software. Two-tailed p-value of unpaired t-test is indicated; ns, not significant ($p > 0.05$). Mean ± SD is shown. **(E)** Example of IHC with β-Catenin staining of tumour-free regions with membrane β-catenin localization *(left)* and tumour regions with nuclear β-catenin localisation *(right)*. Scale bars, 50μm.

Fig. 2. Double CDK8/CDK19 knockout intestinal organoids show decreased cell proliferation. (A) Genotyping confirms the loss of *Cdk8* exon 2 in $Cdk8^{-/-}$ and $Cdk8^{-/-}/Cdk19^{-/-}$ organoids after 7 days of OH-tamoxifen treatment. Control plasmids (a, b and c) are described in Fig. EV1C. **(B)** WB of organoid samples after 7 days of OH-tamoxifen treatment; β-actin was used as loading control. **(C)** Phase contrast images of organoids before and after 6 days of OH-tamoxifen treatment. Scale bars, 150μm. **(D)** Quantification of organoid size at day 0 and 6, as in B (mean + SD are shown). P-value, ordinary one-way ANOVA: (*) $p \leq 0.05$, (***) $p \leq 0.001$. **(E)** IHC staining of organoids (day 7 of OH-tamoxifen treatment) with Ki-67 antibody. Scale bars, 100μm. **(F)** Quantification of Ki-67 positive

area (% of the total area of the organoids; mean + SD) in the four different genotypes presented in (E). Areas with positive Ki-67 signal were detected and quantified using QuPath and ImageJ programs. Adjusted p-values of ordinary one-way ANOVA followed by Tukey's multiple comparison test are indicated: (***) p-value ≤ 0.001; (**) p-values ≤ 0.01; (*) p-values ≤ 0.05; ns, not significant (p > 0.05).

Fig. 3. Functional redundancy between CDK8 and CDK19 in regulation of gene expression. **(A)** Dot plot analysis of differentially expressed genes (DEGs). Red dots: DEGs with p-value ≤ 0.05; purple dots: log_2 fold change (LFC) >1 or <-1, p-value ≤ 0.05; grey dots: not significant, NS. Numbers inside plots indicate the number of genes deregulated more than 2-fold. **(B)** Venn diagrams indicating intersection of genes with altered expression in the indicated genotypes. **(C)** Gene set enrichment analysis (using Enrichr database) of highly deregulated genes in *Cdk8*$^{-/-}$/*Cdk19*$^{-/-}$ organoids. Manually curated signatures extracted from RNA-seq studies in GEO where gene expression was measured before and after drug treatment, gene perturbation or disease. **(D)** qRT-PCR analysis of indicated mRNA levels in Wt, *Cdk8*$^{-/-}$, *Cdk19*$^{-/-}$ and *Cdk8*$^{-/-}$/*Cdk19* $^{-/-}$ organoids. **(E)** WB of indicated proteins extracted from Wt, *Cdk8* $^{-/-}$, *Cdk19* $^{-/-}$ and *Cdk8* $^{-/-}$/*Cdk19* $^{-/-}$ organoids.

Fig. 4. CDK8 and CDK19 regulate the CFTR pathway in the small intestine. (A) Histological PAS staining of organoids treated for 7 days with OH-tamoxifen. Scale bar, 50μm. **(B)** Quantification of PAS signal (% of total organoid area; mean ± SD are shown) in the four different genotypes presented in (A). Areas containing positive PAS staining were detected and quantified using QuPath and ImageJ programs. Adjusted p-values of ordinary one-way Anova followed by Tukey's multiple comparison test are indicated: (***) p-value ≤ 0.001; ns: not significant (p > 0.05). **(C)** Representative phase contrast images of organoids at the indicated time points after 7 days of OH-tamoxifen treatment are shown. Scale bar, 100 μm. **(D)** Quantification of the time needed for mucus release (observed as a dark staining in the center of the organoid; mean ± SD are shown). Adjusted p-values of ordinary one-way Anova followed by Tukey's multiple comparison test are indicated: (***) p-value ≤ 0.001; ns: not significant (p > 0.05); (n= 17 for Wt, n=11 for *Cdk8* $^{-/-}$, n=18 for *Cdk19* $^{-/-}$, n= 12 for *Cdk8* $^{-/-}$/ *Cdk19* $^{-/-}$). **(E)** Fluorescence confocal microscopy images of Calcein green–labeled WT and Cdk8$^{-/-}$/Cdk19$^{-/-}$ organoids treated with forskolin. Scale bars, 100 μm. **(F)** Quantification of forskolin-induced swelling in WT organoids treated for 1hour or 24 hours with 0.1μM, 1μM or 10μM Senexin B (SenB), as indicated, or double KO organoids; DMSO vehicle was used as control. The surface area of individual organoids at different time points relative to the area at t = 0 (100%) was measured (mean ± SD, n=8). Linear regression lines are shown. **(G)** qRT-PCR analysis of Cftr and Muc3 mRNA levels in WT organoids either not treated (t=0), or treated with 10μM Senexin B for 2 or 24 hrs.

Fig. EV1. Mouse *Cdk8* conditional knockout by Lox/Cre targeting of exon 2. (A) Schematic representation of the strategy used for the generation of *Cdk8*^{Lox} alleles from a genomic fragment of mouse enclosing exon 2 of the *Cdk8* gene. See Materials and Methods section for details. (**B)** Southern blot analysis of genomic DNA obtained from mice carrying the *Cd8*^{LoxFrt} and *Cdk8*^{Lox} alleles. DNA was digested with HindIII and probed with 2 different probes (5' and 3'), whose positions are shown in the scheme in (A). The length of the fragments obtained after HindIII digestion of the wild type and the recombinant alleles are indicated (see also the scheme in (A)). **(C)** Left, scheme representing the control plasmids (a, b, and c) for WT, floxed and recombined *Cdk8* exon 2. The position of the oligos (Fw and Rev) used for PCR amplification of genomic DNA and control plasmids is indicated. Right, genotyping of *Cdk8* exon 2 in the mouse intestinal epithelium. All mice were treated with OH-tamoxifen to induce recombination of the LoxP sites. The recombined fragment appears as a 340 bp band in the *Cdk8* $^{-/-}$ and *Cdk8* $^{+/-}$ mice (*VillinCre*^{ERT2} recombinase-positive), and is absent in the *Cdk8* $^{Lox/+}$ mouse that does not contain the *VillinCre*^{ERT2} gene.

Fig. EV2. CDK8 is not required for cell proliferation nor differentiation in mouse intestine. (A) Representative immunohistochemistry images of Fig 1B. Small intestine, proximal and distal colon samples were stained for BrdU, Ki67 or PCNA antibodies. Goblet cells were detected with PAS staining. Scale bars: 100µm for small intestine and proximal colon, 50µm for distal colon. **(B)** WB analysis of mouse intestine epithelium showing the absence of CDK8 protein 2 months after OH-tamoxifen feeding. Mice 4, 5 and 6 did not have the *VillinCre*^{ERT2} gene; mice 1, 2 and 3 had the *VillinCre*^{ERT2}. GAPDH protein was used as loading control.

Fig. EV3. Effects of CDK8 and CDK7 knockout on RNA pol II CTD phosphorylation. WB analysis of the indicated proteins in mouse intestinal epithelium **(A)** or liver **(B)** samples from WT and *Cdk7* $^{lox/lox}$, *Cdk8* $^{lox/lox}$, *Rpb-Cre-Ert2* $^{KI/KI}$ mice after tamoxifen treatment. Amido-black staining was used as loading control.

Fig. EV4. CDK8 loss does not affect chemically-induced intestinal carcinogenesis. (A) Scheme showing the steps of the AOM/DSS carcinogenesis experiment. See Materials and Methods for more detailed information. **(B)** Graphs showing female (left; n=6 in both groups*)* and male (right; n=5 in both groups) weight evolution over 21 days following the last DSS treatment. **(C)** Genotyping after AOM/DSS treatment confirms the recombination and loss of *Cdk8* exon 2 in colon tumors from *Cdk8* $^{-/-}$ mice. Control plasmids (a and b), are described in Fig. EV1C. PCR amplification with *Villin-Cre*^{ERT2}-specific primers confirms the presence of the *Cre*^{ERT2} recombinase gene. **(D)** WB with the same colon tumour samples presented in (C) confirm the disappearance of CDK8 protein in the *Cdk8* $^{-/-}$ mice. β-actin was used as the loading control.

Fig. EV5. CDK8 deletion does not prevent Apc-loss-dependent tumourigenesis in mouse intestine. **(A)** Genotyping confirms the loss of *Cdk8* exon 2 in intestine epithelium from *Apc* $^{-/-}$*/Cdk8* $^{-/-}$ mice. Control plasmids (a, b and c) are described in Fig. EV1-C. **(B)** WB with intestine epithelium samples from mice presented in (A) confirm the absence of CDK8 protein in *Cdk8* $^{-/-}$ mice. CDK19 levels vary independently of the presence or absence of CDK8. Amido-black staining was used as loading control. **(C)** IHC staining of CDK8 and Ki-67 in small intestine and colon samples from *Apc* $^{-/-}$*/Cdk8* $^{+/+}$ and *Apc* $^{-/-}$*/Cdk8* $^{-/-}$ mice. Scale bars, 100µm. **(D)** Quantification of the Ki-67 positive area (% of the total area of the intestine presenting positive staining, quantified using QuPath (Bankhead *et al*, 2017) and Image J software; mean ± SD) in the IHC shown in (C). Ki-67 (n=16). Two-tailed p-value of unpaired t-test is indicated: ns, not significant ($p > 0.05$).

Fig. EV6. Amino acid sequence conservation of CDK8 and its paralogue CDK19 between different vertebrates. **(A**) Sequence alignment of Cdk8 (blue) and Cdk19 (green) proteins from: *Homo sapiens, Mus musculus, Xenopus tropicalis, Xenopus laevis, and Danio rerio*. Homologous sequences are black. The C-terminal region of both kinases, containing a disordered region, is highlighted in red. The consensus (> 80%) is presented below the alignment. **(B)** Scheme indicating the fragment of *Cdk19* exon 1 removed by CRISPR-Cas9 (highlighted in red) in intestinal organoids. The arrow indicates the sequence of the sgRNA used. The sequence trace obtained after gene editing is presented below. The colour-code in the sequence in the box corresponds to the sequence trace.

Fig. EV7. CDK8/CDK19 KO organoids are counter-selected. WB indicating the levels of CDK8, CDK19 and phospho-Stat1-S727 in organoids after 14 days of tamoxifen treatment. Two out of the three *Cdk8* $^{-/-}$*/Cdk19* $^{-/-}$ clones show a reappearance of the CDK8 protein: compare with Fig. 2B where proteins were extracted from the same samples, but one week earlier. (★) indicates the two clones where CDK8 protein is detected; this was observed only in organoids where double KO had been induced.

Movie EV1. Live-cell microscopy shows a rapid expansion of both the lumen and total organoid surface area in WT organoids after the addition of forskolin. $Cdk8^{-/-}$/$Cdk19^{-/-}$ organoids do not swell after forskolin addition. Three different sizes of organoids are presented in each condition: big, (top); medium (middle) and small (bottom). Scale bar, 200 µm.

Figure 1.

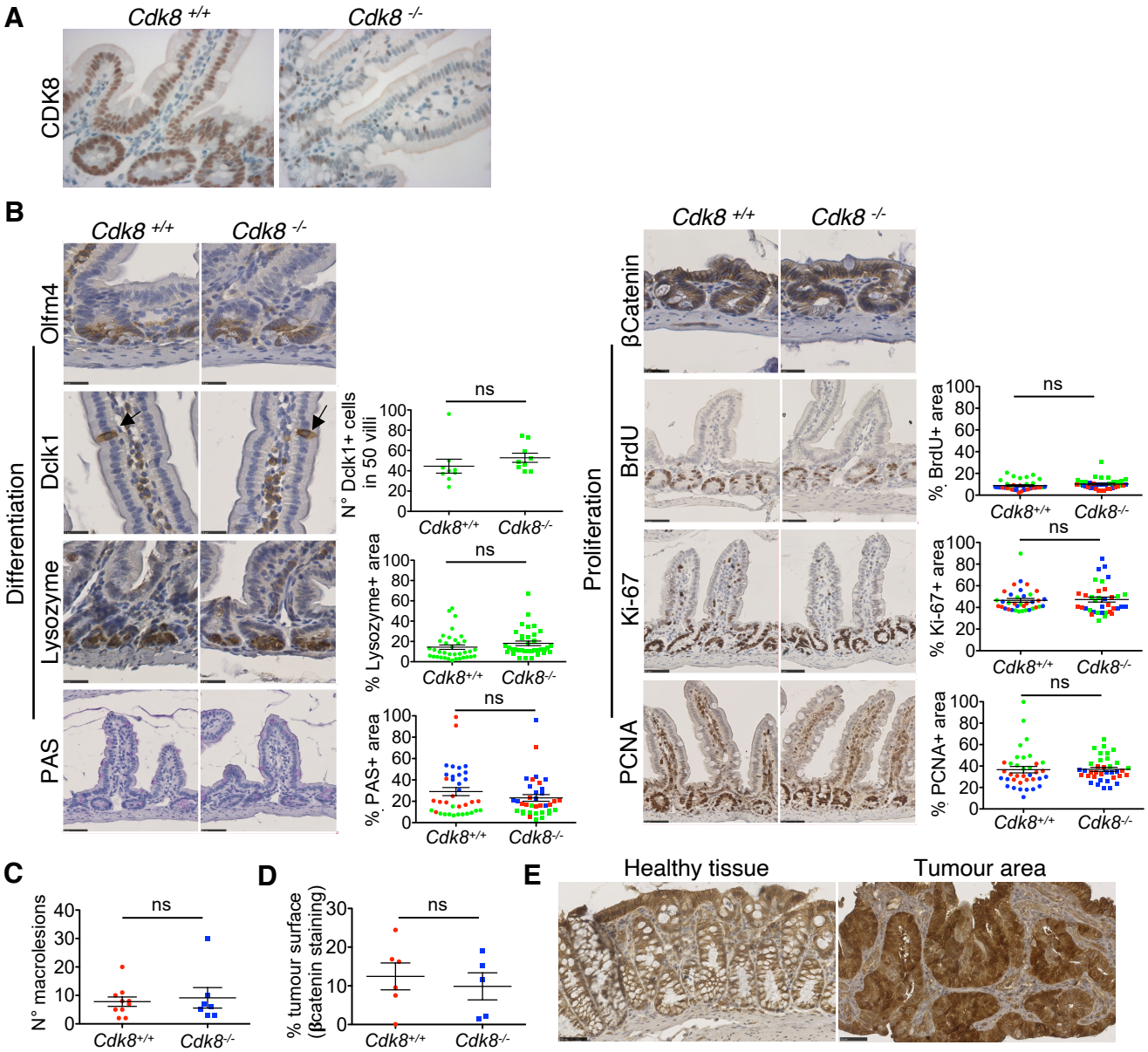

A
Cdk8 +/+
Cdk8 -/-
CDK8
B
Cdk8 +/+
Cdk8 -/-
Differentiation
Olfm4
Dclk1
Lysozyme
PAS
N° Dclk1+ cells in 50 villi
% Lysozyme+ area
% PAS+ area
ns
Cdk8+/+
Cdk8-/-
Proliferation
βCatenin
BrdU
Ki-67
PCNA
% BrdU+ area
% Ki-67+ area
% PCNA+ area
C
N° macrolesions
D
% tumour surface (βcatenin staining)
E
Healthy tissue
Tumour area

Figure 2.

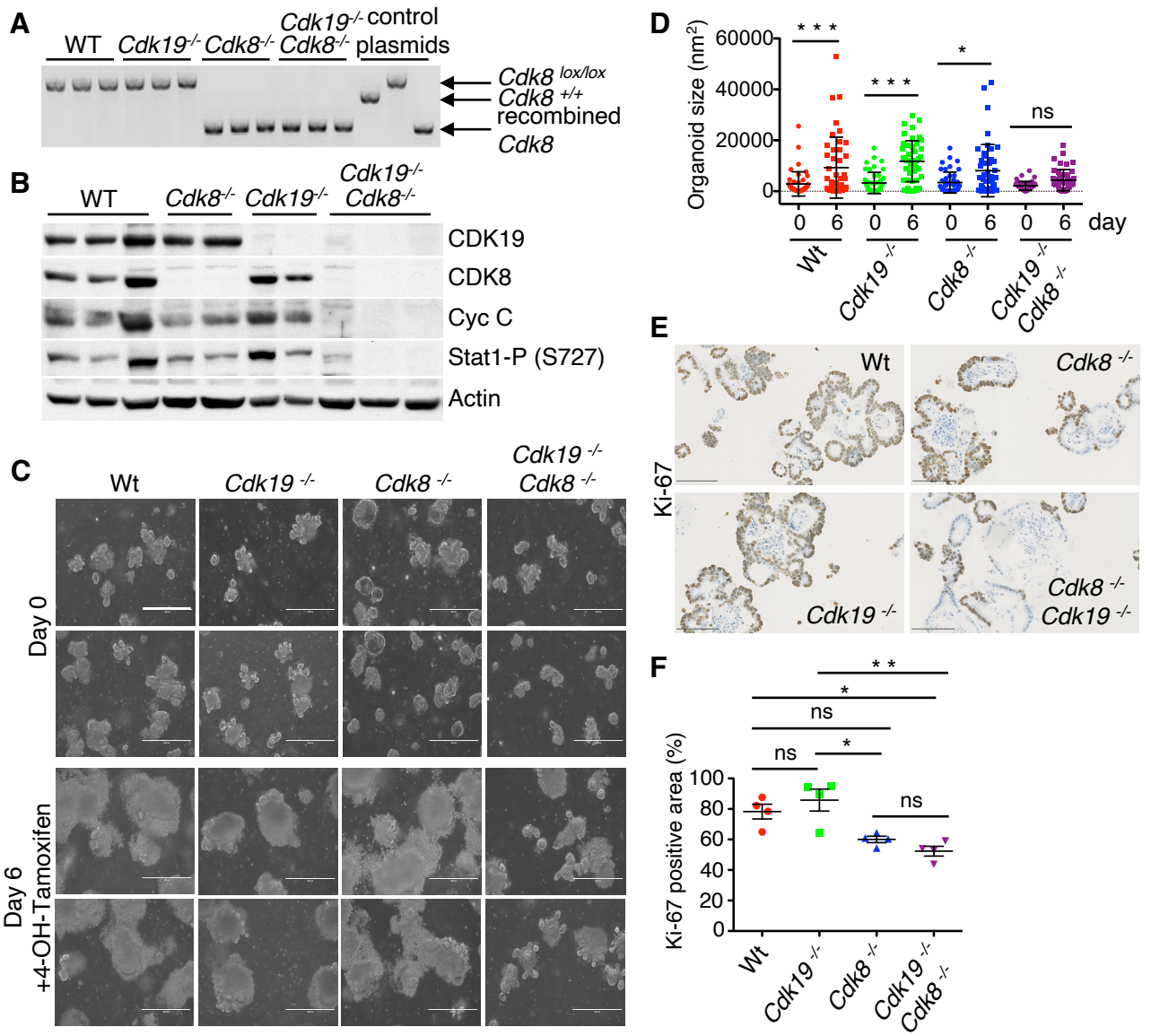

A
WT
Cdk19-/-
Cdk8-/-
Cdk19-/- Cdk8-/-
control plasmids
Cdk8 lox/lox
Cdk8 +/+
recombined Cdk8
B
WT
Cdk8-/-
Cdk19-/-
Cdk19-/- Cdk8-/-
CDK19
CDK8
Cyc C
Stat1-P (S727)
Actin
C
Wt
Cdk19 -/-
Cdk8 -/-
Cdk19 -/- Cdk8 -/-
Day 0
Day 6 +4-OH-Tamoxifen
D
Organoid size (nm2)
60000
40000
20000
0
* * *
* * *
*
ns
0 6 0 6 0 6 0 6 day
Wt
Cdk19 -/-
Cdk8 -/-
Cdk19 -/- Cdk8 -/-
E
Ki-67
Wt
Cdk8 -/-
Cdk19 -/-
Cdk8 -/- Cdk19 -/-
F
Ki-67 positive area (%)
100
80
60
40
20
0
* *
*
ns
ns
*
ns
Wt
Cdk19 -/-
Cdk8 -/-
Cdk19 -/- Cdk8 -/-

Figure 3.

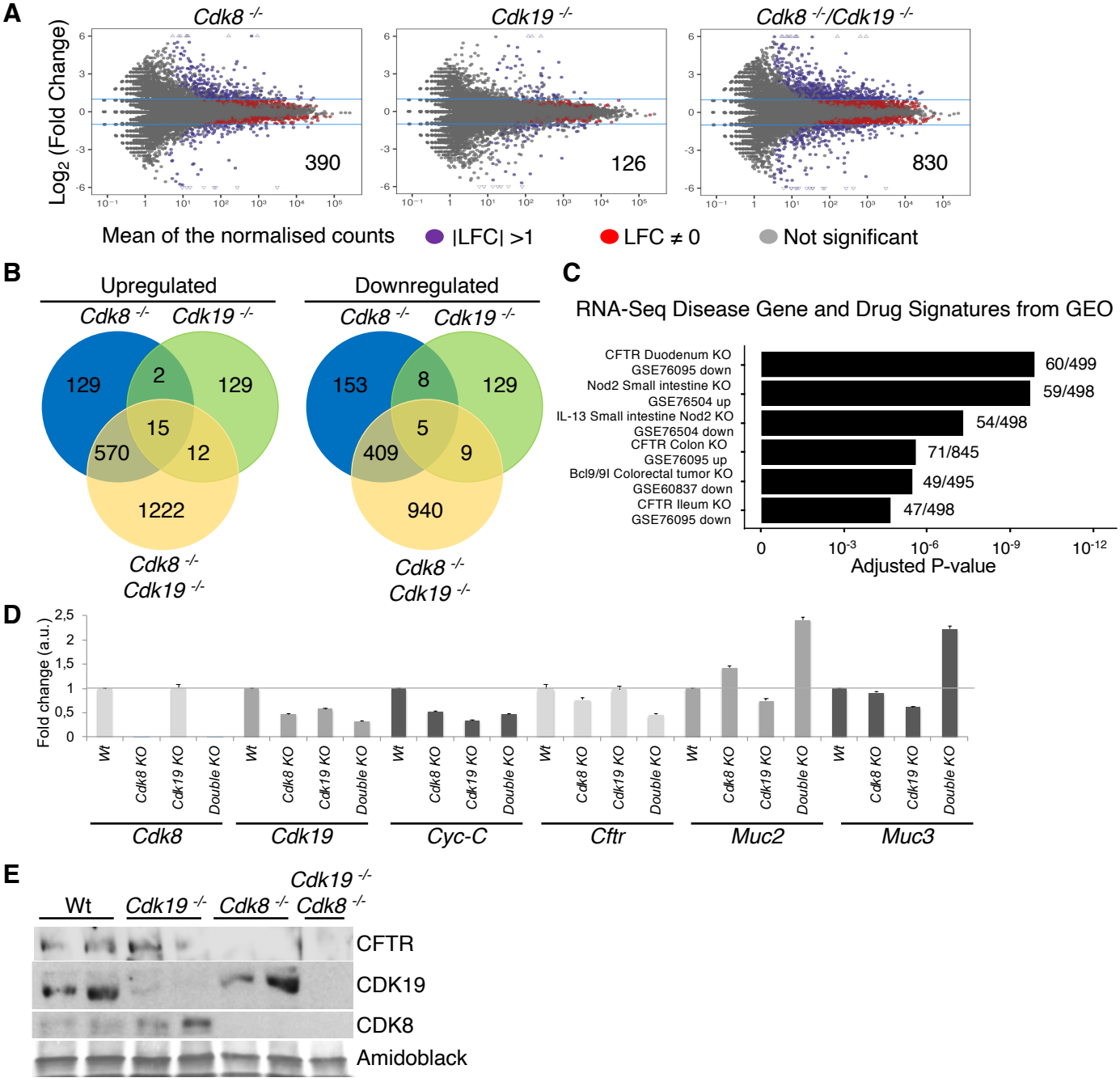
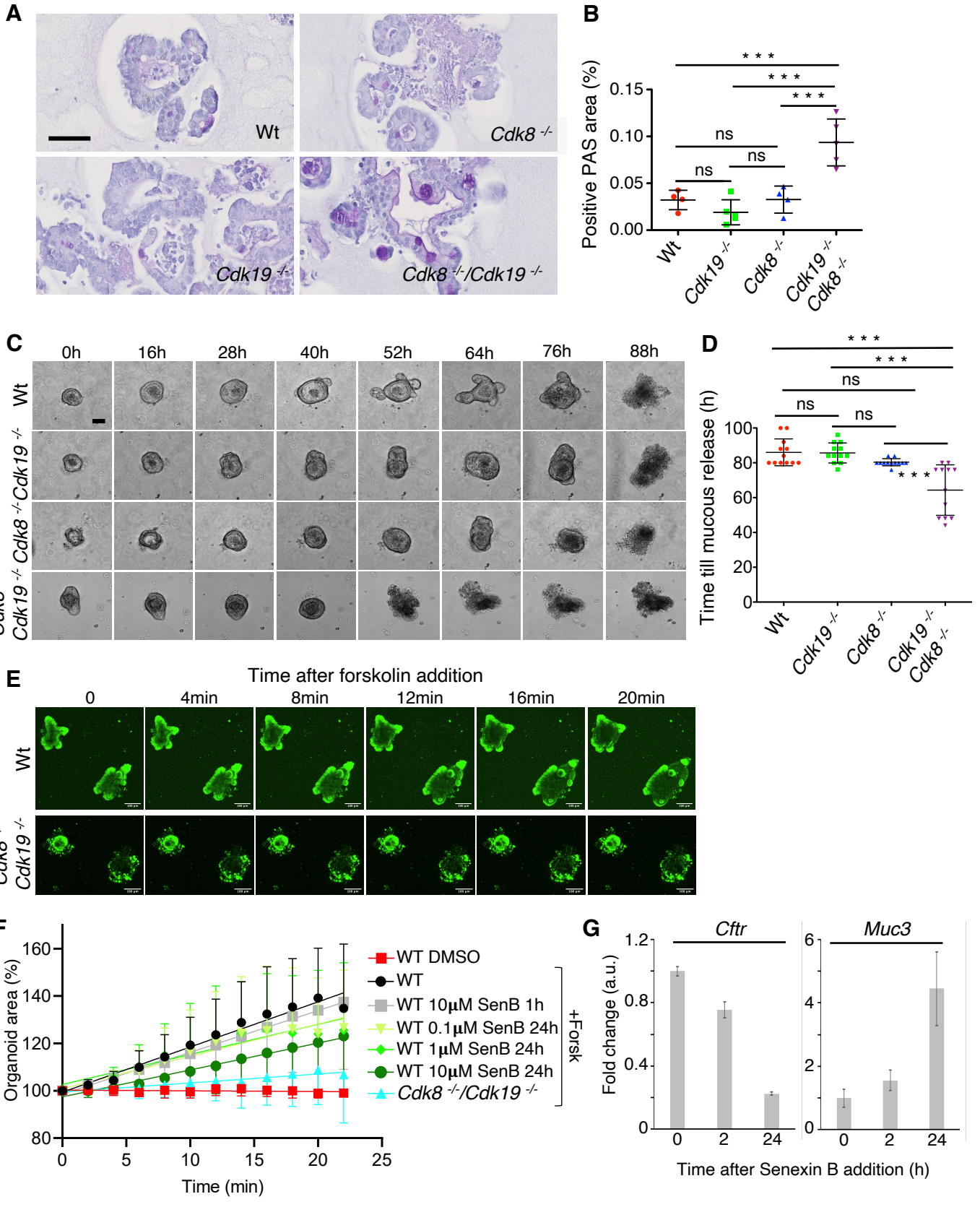

A
Cdk8 -/-
Cdk19 -/-
Cdk8 -/-/Cdk19 -/-
Log2 (Fold Change)
390
126
830
Mean of the normalised counts
|LFC| >1
LFC ≠ 0
Not significant
B
Upregulated
Downregulated
Cdk8 -/-
Cdk19 -/-
129
2
129
15
570
12
1222
153
8
129
5
409
9
940
C
RNA-Seq Disease Gene and Drug Signatures from GEO
CFTR Duodenum KO GSE76095 down
Nod2 Small intestine KO GSE76504 up
IL-13 Small intestine Nod2 KO GSE76504 down
CFTR Colon KO GSE76095 up
Bcl9/9l Colorectal tumor KO GSE60837 down
CFTR Ileum KO GSE76095 down
60/499
59/498
54/498
71/845
49/495
47/498
Adjusted P-value
D
Fold change (a.u.)
Wt
Cdk8 KO
Cdk19 KO
Double KO
Cdk8
Cdk19
Cyc-C
Cftr
Muc2
Muc3
E
Wt
Cdk19 -/-
Cdk8 -/-
Cdk8 -/- Cdk19 -/-
CFTR
CDK19
CDK8
Amidoblack

Figure 4.


A
Wt
Cdk8 -/-
Cdk19 -/-
Cdk8 -/-/Cdk19 -/-
B
Positive PAS area (%)
0.15
0.10
0.05
0.00
* * *
* * *
* * *
ns
ns
ns
Wt
Cdk19 -/-
Cdk8 -/-
Cdk19 -/- Cdk8 -/-
C
0h
16h
28h
40h
52h
64h
76h
88h
Wt
Cdk19 -/-
Cdk8 -/-
Cdk8 -/- Cdk19 -/-
D
Time till mucous release (h)
100
80
60
40
20
0
* * *
* * *
ns
ns
ns
* * *
Wt
Cdk19 -/-
Cdk8 -/-
Cdk19 -/- Cdk8 -/-
E
Time after forskolin addition
0
4min
8min
12min
16min
20min
Wt
Cdk8 -/- Cdk19 -/-
F
Organoid area (%)
160
140
120
100
80
0
5
10
15
20
25
Time (min)
WT DMSO
WT
WT 10µM SenB 1h
WT 0.1µM SenB 24h
WT 1µM SenB 24h
WT 10µM SenB 24h
Cdk8 -/-/Cdk19 -/-
+Forsk
G
Cftr
Muc3
Fold change (a.u.)
1.2
0.8
0.4
0
6
4
2
0
0
2
24
0
2
24
Time after Senexin B addition (h)

Appendix 4

CDK8 and CDK19 kinases have non-redundant oncogenic functions in hepatocellular carcinoma

CDK8 and CDK19 kinases have non-redundant oncogenic functions in hepatocellular carcinoma

Katarina Bacevic [1], Susana Prieto [2], Stefano Caruso [3,4], Alain Camasses [2], Geronimo Dubra [2], José Ursic-Bedoya [1,5], Anthony Lozano [1], Jacqueline Butterworth [1], Jessica Zucman-Rossi [3,4,6], Urszula Hibner [1], Daniel Fisher [2,*,✉] and Damien Gregoire [1,*,✉]

1. Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France
2. Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France ; Equipe labellisée Ligue Contre le Cancer
3. Centre de Recherche des Cordeliers, Sorbonne Universités, Inserm, UMRS-1138, F-75006 Paris, France
4. Functional Genomics of Solid Tumors, USPC, Université Paris Descartes, Université Paris Diderot, Université Paris 13, Labex Immuno-Oncology ; Equipe labellisée Ligue Contre le Cancer, F-75000 Paris, France
5. Department of hepatogastroenterology, Hepatology and Liver Transplantation Unit, Saint Eloi Hospital, University of Montpellier, France
6. European Hospital Georges Pompidou, AP-HP, F-75015, Paris, France

* Equal contribution

✉ Corresponding authors

Abstract

Hepatocellular carcinoma (HCC) is a common cancer with high mortality. The limited therapeutic options for advanced disease include treatment with Sorafenib, a multi-kinase inhibitor whose targets include the Mediator kinase CDK8. Since CDK8 has reported oncogenic activity in Wnt-dependent colorectal cancer, we investigated whether it is also involved in HCC. We find that CDK8 and its paralogue CDK19 are significantly overexpressed in HCC patients, where high levels correlate with poor prognosis. Liver-specific genetic deletion of CDK8 in mice is well supported and protects against chemical carcinogenesis. Deletion of either CDK8 or CDK19 in hepatic precursors had little effect on gene expression in exponential cell growth but prevented oncogene-induced transformation. This phenotype was reversed by concomitant deletion of TP53. These data support important and non-redundant roles for mediator kinases in liver carcinogenesis, where they genetically interact with the TP53 tumor suppressor.

Introduction

The cyclin-dependent kinase CDK8 is the principal catalytic subunit of the Mediator complex kinase module (1). In mice, genetic deletion of CDK8 or its activating subunit, cyclin C, is embryonic lethal (2, 3). Vertebrate genomes encode a second paralogue of CDK8, CDK19, which also binds cyclin C and can replace CDK8 in the kinase module of Mediator (4, 5). CDK8 may act as an oncogene in several tumor types, including melanoma and colorectal cancer (CRC) (6, 7). Amplification of the CDK8 gene was found in 14% of CRC patients, and reversing CDK8 over-expression in CRC cells with amplified CDK8 reduced cell proliferation by interfering with the β-catenin pathway (6, 8). CDK8 was also uncovered in a mouse transposon-mediated mutagenesis screen for genes whose alteration contributes to intestinal cancer (9). However, intestine-specific *Cdk8* deletion in mice failed to confirm an oncogenic role in intestinal tumorigenesis triggered by mutation of the tumor suppressor *Apc* ((10), and our unpublished results).

CDK8 acts as a co-factor of many transcriptional activators and participates in regulating the expression of a large number of genes, including immediate early genes (11–14), targets of β-catenin (6), p53 (15), c-Myc (16), Hypoxia-inducible factor 1α (HIF1α) (17, 18), Nuclear-factor kappa B (NFκB) (19) and Notch1 (3, 20). CDK8 expression appears to maintain tumors in an undifferentiated state by regulating c-Myc programmes (16). CDK8 was also identified as a crucial regulator of tumor-promoting activity of senescent cells (21). It is presently unclear how these multiple pathways participate in tumor promotion by CDK8, and whether its paralogue CDK19 has similar roles in tumorigenesis.

Since the oncogenic roles of CDK8 in Wnt-dependent CRC have been well documented in the context of constitutive Wnt signalling, we wondered whether it also has roles in hepatocellular carcinoma (HCC), another tumor type frequently associated with activation of Wnt/β-catenin signalling (22). HCC is the main primary liver tumor and one of the most deadly cancers worldwide (23). Mutations in the *CTNNB1* gene, coding for β-catenin, are found in around 30% of patients and are mainly associated with the well differentiated, less aggressive, class of HCC (groups G5 - G6) (22, 24). CDK8 is expressed in the liver, where it may regulate lipogenesis (25). Importantly, lipid accumulation as well as *de novo* lipid biosynthesis and the resulting lipotoxicity lead to hepatic inflammation, constituting major risk factors for liver tumorigenesis (22).

In this study, we show that although CDK8 has a limited role in liver homeostasis both CDK8 and CDK19 paralogues are required for hepatic carcinogenesis in the context of wild-type p53.

Results

CDK8 and *CDK19* are highly expressed in p53-mutated hepatocellular carcinoma

To test whether expression of CDK8 or that of its paralogue CDK19 is altered in hepatic carcinogenesis, we first quantified by qRT-PCR their expression in a large cohort of HCC patients (n=268, Supplemental Table 1). We found that both CDK8 and CDK19 are significantly overexpressed in HCC tumors compared to non-tumoral counterparts or normal liver (Figure 1A). Moreover, we detected a correlation between the expression of the two kinases in HCC (Figure 1B), as previously observed in breast cancer (26). Analysis revealed no correlation with a specific aetiology (alcohol, viral infection, metabolic syndrome). However, there was a highly significant difference in CDK8/19 expression among the HCC subgroups defined by the classification based on clinical and molecular features (24, 27): CDK8 or CDK19 high expressors were enriched in the aggressive G1-G3 subsets as compared to G4-G6 (Figure 1C). Coherently, high CDK8 or CDK19 expression was correlated with mutant p53 status, molecular prognostic 5-gene score (28) and macroscopic vascular invasion (Figure 1D, E and F). Finally, high level expression of CDK8 or CDK19 correlated with poor prognosis (Figure 1G). Thus our data are consistent with an oncogenic role for Mediator kinases in hepatocellular carcinoma.

CDK8 is dispensable for liver homeostasis

To investigate possible roles of the Mediator kinases in liver function and hepatic carcinogenesis, we focused on CDK8. We generated a genetically modified mouse with loxP sites flanking exon 2 of the *Cdk8* gene (Figure 2A). Crossing these animals with transgenic mice expressing the *Cre* recombinase under the control of the *Albumin* promoter (*Alb-Cre* mice(29)) gives rise to hepatocyte-specific deletion of the essential exon 2, and induces a frameshift that results in a stop codon at position 52, eliminating CDK8 protein in the liver (CDK8$^{\Delta hep}$ animals) (Figure 2B).

In accordance with the absence of overt phenotype of ubiquitous Cdk8 deletion in adult animals (10), the liver-specific deletion of CDK8 was compatible with normal liver development and we detected no change in liver physiology at the age of 3 months (Figure 2C). Importantly, the β-catenin-driven metabolic liver zonation was not affected by the CDK8 ablation, as judged by the purely centrilobular expression of glutamine synthetase (Figure 2C). Moreover, we detected no differences in the expression levels of several β-catenin target genes between the control and the CDK8$^{\Delta hep}$ animals (Figure 2C). These results indicate that CDK8 is not required to regulate the β-catenin pathway under normal homeostasis conditions. However, in aging animals (> 6 months) CDK8 deficiency led to increased body weight and higher liver steatosis score (Figure 2D), confirming an involvement of CDK8 in liver lipogenesis (25). Older CDK8$^{\Delta hep}$ animals (12-15 months) did not show any sign of liver tumors (n= 12).

CDK8 is required for chemically induced liver carcinogenesis and hepatic cell transformation

We next used a model of hepatic carcinogenesis in CDK8$^{\Delta hep}$ animals to investigate potential roles of CDK8 in liver cancer. In this model, a single injection of hepatotoxic agent diethylnitrosamine (DEN) to young mice gives rise to liver tumors after 6-8 months (30). We sacrificed DEN-treated animals at 28 weeks, a relatively early time point in the kinetics of tumor formation, to allow detection of both positive and negative changes in tumor burden. As expected, 9 out 17 (53%) of the control *Cdk8*$^{F/F}$ mice had at least one macroscopic liver tumor at sacrifice (Figure 3A). In contrast, only one out of eleven (9%) CDK8$^{\Delta hep}$ animals had developed tumors by this time point. Cell death and regenerative response of the livers shortly after the DEN treatment were both indistinguishable between controls and CDK8$^{\Delta hep}$ mice (Supplemental Figure 1), indicating that CDK8 ablation acted by inhibiting tumorigenesis rather than by interfering with the initial hepatoxicity of the treatment. Taken together, these data support the hypothesis that CDK8 contributes to chemically-induced liver carcinogenesis in mice.

To investigate whether the effects of CDK8 loss on carcinogenesis were cell autonomous, we next isolated primary hepatic progenitor cells (BMEL) (31) from *Cdk8*$^{F/F}$ embryos. Upon stable transfection of Cre, we obtained an efficient deletion of CDK8 from these cells (Figure 3B). CDK8 loss had no effect on BMEL cell morphology

or growth characteristics under standard monolayer culture conditions (Figure 3B). As murine liver tumors triggered by DEN injection are often driven by mutated forms of Ras (30), we used forced expression of an oncogenic form of Ras, H-Ras^{G12V}, to investigate effects of CDK8 loss on hepatic cell transformation. Similarly to previous results obtained with independent BMEL cell lines (32), H-Ras^{G12V} was sufficient to transform primary *Cdk8$^{F/F}$* BMEL cells, which then efficiently formed colonies in soft agar (Figure 3C). In contrast, CRE-mediated deletion of CDK8 abolished colony formation in this assay (Figure 3C), which is consistent with the protective role of CDK8 deletion in DEN-treated livers. Next, we used a model of tamoxifen-inducible Cre^{ERT2} activation. This confirmed the CDK8 requirement for Ras-induced transformation. Strikingly, deletion of CDK8 in cells previously transformed by Ras^{G12V} expression reverted their transformed phenotype (Figure 3C). To validate *in vivo* that CDK8 deletion protects hepatic progenitors from Ras-induced transformation, we injected CDK8$^{F/F}$ Ras^{G12V} or CDK8$^{-/-}$ Ras^{G12V} BMEL cells into the liver of immunodeficient mice. While the *Cdk8$^{F/F}$* BMEL expressing Ras^{G12V} gave rise to orthotopic tumors, their counterparts devoid of CDK8 did not (Figure 3D).

We next disrupted *Cdk8* by by CRISPR/Cas9-mediated gene disruption, wich confirmed both the absence of apparent phenotype in CDK8-depleted cells and the requirement for CDK8 for cell transformation (Figure 3E). Finally, to exclude that the requirement for CDK8 for transformation is specific for BMEL cells, we disrupted it by CRISPR/Cas9 in a human hepatoblastoma cell line, HepG2. As expected, the cells grew well in the absence of CDK8, but again did not form colonies in soft agar (Figure 3E). Altogether, our results indicate that CDK8 is required for H-Ras^{G12V}-driven oncogenic transformation of hepatocytes and of hepatic progenitor cells.

CDK8 deletion impairs Ras^{G12V}-driven transformation in a p53-dependent manner

We then investigated possible molecular mechanisms by which CDK8 removal could impair Ras^{G12V}-driven transformation. Deletion of CDK8 did not abrogate Ras^{G12V}-induced cell shape remodelling or ERK phosphorylation (Supplemental Figure 2A, B), indicating that the kinase is dispensable for Ras pathway activation. Second, constitutive activation of the β-catenin pathway in the BMEL cells did not rescue CDK8 deficiency (Supplemental Figure 2C), indicating that tumor-promoting activity of CDK8 in the liver does not rely on the activation of the β-catenin pathway, which is consistent

with the lack of β-catenin-related phenotype in CDK8$^{\Delta hep}$ livers (Figure 2). Third, CDK8 has been proposed as a regulator of glycolysis (33), and we therefore tested whether CDK8 removal affects the metabolism of BMEL cells. However, analysis of glycolysis (Seahorse Glycolysis stress test) and mitochondrial respiration (Seahorse Mito stress test) did not reveal any effect of CDK8 ablation (Supplemental Figure 3), indicating that this regulation is not present in hepatic progenitor cells and that it does not account for the failure of CDK8-mutant cells to grow in soft agar.

Since *CDK8* overexpression in patients correlates with mutant p53 status, we next considered the possibility that CDK8 acts by modulating p53 function. Although CDK8 can act as a coactivator of the p53 transcriptional program (34), we observed that removal of CDK8 rather increases the level of p53 protein (Figure 4A), suggesting that CDK8 might restrain the tumor-suppressive functions of p53 in hepatic cells. In agreement with this idea, soft agar tests showed that the requirement for CDK8 in Ras-induced transformation was abrogated by p53 inactivation via CRISPR/Cas9 editing (Figure 4B). Furthermore, depletion of CDK8 did not prevent transformation of the p53-deficient Huh7 cell line (Supplemental Figure 4). To extend these results to an *in vivo* setting, we triggered tumorigenesis via hydrodynamic gene delivery (HGD) (35) of the activated form of Ras together with CRISPR/Cas9-mediated inactivation of endogenous p53. HGD with Ras alone did not generate tumors, as previously described (36), whereas simultaneous transfection with N-Ras^{G12D} and CRISPR-p53 gave rise to multiple aggressive tumors within 4 weeks (Figure 4C). Consistent with the results of the cellular models, this combination of oncogenic stimuli was also fully efficient in CDK8$^{\Delta hep}$ animals, whose hepatocytes are devoid of CDK8 (Figure 4C). Cell lines derived from CDK8$^{F/F}$ or CDK8$^{-/-}$ HGD-tumors were equally capable of giving rise to tumors when injected orthotopically into immunocompetent mice (Figure 4D). Our results suggest that CDK8 is required for initiation of tumorigenesis by counteracting p53 function.

Both CDK8 and CDK19 are required for Ras^{G12V}-driven transformation

An interesting inference from the above results is that CDK19, whose expression is preserved in CDK8-deleted cells (Figure 5A), is insufficient to compensate for loss of CDK8, despite its high homology and redundant roles in Mediator. Furthermore, both CDK8 and CDK19 are overexpressed in a significant number of HCC patients. We therefore wondered whether CDK19 is also required for

Ras-induced tumorigenesis. To test this, we deleted CDK19 in BMEL cells by CRISPR-Cas9 editing. The resulting mutant cells had unaltered morphology and proliferation kinetics (Figure 5A). Interestingly, CDK19 removal caused a significant upregulation of CDK8 protein levels, but not of mRNA levels, indicating a posttranscriptional feedback regulation of CDK8 in the absence of CDK19 (Figure 5A).

Similarly to CDK8 deletion, CDK19 ablation prevented the acquisition of the transformed phenotype upon ectopic expression of the oncogenic form of Ras, as judged by the lack of anchorage-independent growth (Figure 5B). Again, we obtained similar results upon CDK19 deletion in human HepG2 cells (Supplemental Figure 4). Thus, both CDK8 and CDK19 are required for Ras-driven transformation of primary hepatic progenitors, neither paralogue being able to compensate for the absence of the other.

As CDK8 and CDK19 constitute the only members of the CDK module regulating transcriptional activity of the mediator complex, one hypothesis to account for the requirement of simultaneous presence of both kinases is that they control different subsets of target genes, both of which are needed for transformation by H-Ras^{G12V}. To test this hypothesis, we analyzed gene expression profiles of the Ras-expressing BMEL cells in the presence or absence of CDK8 or CDK19 by RNA-seq (Figure 5C). Surprisingly, disruption of either CDK8 or CDK19 had only minor effects on gene expression, with only around 30 genes reproducibly showing more than 2-fold differences in expression between cells expressing Ras^{G12V} with or without deletion of *Cdk8* or *Cdk19*. At this threshold of fold-change, we did not identify common genes deregulated by the absence of either kinase (Figure 5D). However, even though neither CDK8$^{-/-}$ nor CDK19$^{-/-}$ cells can be transformed by the oncogenic form of Ras, Ras^{G12V} caused deregulation of over 1000 genes, irrespective of the status of *Cdk8* and *Cdk19* when growing in exponential culture, explaining their indistinguishable morphology. This result further shows that, despite the importance of Mediator for gene regulation, neither CDK8 nor CDK19 are required for implementing wide-ranging changes to gene expression, suggesting that their effects in carcinogenesis, which involve p53, are not due to major transcriptional changes.

Discussion

Combining patient data and mouse models, we provide evidence that both Mediator kinases, CDK8 and CDK19, are involved in hepatocellular carcinoma. Our

data indicate that these CDKs are required for initiation of tumorigenesis induced by mutation of a strong oncogene, *Ras*. To our knowledge, this is the first indication that the CDK8 paralogue CDK19 is required for cell transformation, indicating non-redundant functions for both kinases in facilitating tumorigenesis.

We find that removal of CDK8 in hepatocytes has no major consequences for liver functions, as previously indicated by the ubiquitous inactivation of the kinase in adult animals (10). However, we observed that CDK8 depletion induces steatosis in aging animals, in accordance with a role for the kinase as repressor of lipogenesis (25). In contrast to its mild effects on liver physiology, CDK8 disruption has a strong impact on hepatic carcinogenesis. We observed a major reduction of chemically induced tumorigenesis, and a complete protection from Ras-induced cell transformation. This requirement is further highlighted by the fact that CDK8 ablation occurring after Ras^{G12V}-induced transformation is able to restore the non-transformed phenotype.

CDK8 thus apparently has diverse context-specific involvement in various cancers. CDK8 was initially described to have oncogenic properties in Wnt-dependent colorectal cancers, where it controls the beta-catenin pathway (6, 8) and maintains Myc functions (37). However, this was not confirmed genetically in mouse models; indeed, if anything, CDK8 knockout marginally increased progression from early lesions to tumors (10). A recent study also found no effect of CDK8 knockdown on growth of colorectal tumors in syngeneic mice, but it prevented liver metastases (38). Furthermore, CDK8 promotes proliferation of melanoma cell lines (7) and is apparently required for efficient Estrogen Receptor (ER)-dependent transcription in ER-positive breast cancer cells (39). However, most reports of anti-tumor activity caused by interfering with CDK8 function have used kinase inhibitors (39–45) that target both CDK8 and CDK19. The most likely explanation for this is that CDK inhibitors are rarely very specific for one kinase subfamily (46) and some “CDK8/19 inhibitors” may inhibit other kinases required for efficient tumor development. This highlights the need for a genetic approach *in vivo*.

Other studies have suggested that as well as possessing oncogenic activity in some contexts, CDK8 may rather have tumor-suppressive activity in others, including endometrial cancers (47), intestinal cancer (10) and T-cell acute lymphoblastic leukemia (3).

One way of reconciling these apparently different roles for Mediator kinases is to invoke possible tissue-specificity of action. We found that CDK8 deletion has no effect on hepatic cell growth, contrary to certain melanoma or gastric cancer cell lines, in which decrease of proliferation has been described (7, 48). Tissue specificity is further demonstrated by the fact that while knockout of either kinase is well tolerated by hepatic precursor cells, germline-deletion of the *Cdk8* gene in mice is lethal at pre-implantation embryogenic stages (2). Our results suggest that oncogenic activities of CDK8 and CDK19 in HCC are independent of the β-catenin pathway, and therefore likely operate through different mechanisms in colorectal and hepatic carcinogenesis. One key finding of our study is that both CDK8 and CDK19 are required for cell transformation by H-Ras^{G12V}. This is not due to downregulation of one kinase when the other one is removed, as, on the contrary, we found that CDK19 depletion stabilizes CDK8 protein, identifying a post transcriptional regulation of CDK8 levels. Of note, this cross-regulation is not bidirectional, as removal of CDK8 does not seem to stabilize CDK19 protein. A possible explanation for these results might be that CDK8 and CDK19 regulate transcription of specific gene sets that are both required for hepatic cell transformation. The results of our RNA-seq analysis renders this hypothesis unlikely, as very few genes are differentially transcribed in either knockout. However, we cannot exclude this entirely, as 12 genes were differentially regulated in knockouts of both *Cdk8* and *Cdk19* when the threshold for fold-change is removed: Adora1, Nt5e, Fgfbp1, Spp1, Gm5781, Lgals4, Rpl10-ps3, Akap12, Gm8349, Anxa10, Scd1, Ly6a. Yet the magnitude of the gene expression change is likely too low to be biologically meaningful; the regulation of several, including Anxa10 (encoding Annexin 10A), Akap12 (encoding A Kinase Anchoring Protein-12) and Nt5e (encoding 5' nucleotidase) is in opposite directions in each knockout; and no clear roles in cancer for any of these genes have been reported. A possibility that we cannot exclude at this stage is that both CDK8 and CDK19 are required for transcription or repression of genes whose expression is only activated upon growth in foci. However, the minimal effects of deletion of either gene in shaping the transcriptome induced by expression of oncogenic Ras suggests that neither kinase is required to implement large-scale changes to gene expression, and that non-transcriptome effects should be considered.

We suggest an alternative model, which is compatible with both apparent oncogenic activity of Mediator kinases in some circumstances and lack of effects in others. In this model, CDK8/19 are not in themselves oncogenes and therefore their

overexpression does not transform cells, but rather provides a favorable terrain enabling the initiation of tumorigenesis by additional oncogenic factors, including *bona fide* oncogenes. For example, it is well established that, rather than transforming cells directly, expression of strong oncogenes such as Ras^{G12V} promotes premature cell senescence with accumulation of p53 and p16 (49). BMEL cells are immortal, but have a functional wild-type p53 (50). In this context CDK8/19 expression may attenuate the function of p53, thus unleashing the oncogenic potential of mutant Ras. We speculate that loss of p53 by mutation "fixes" the initial advantage conferred by overexpression of CDK8/19. This agrees both with our finding that in the liver, CDK8/19 are only required for cell transformation in the presence, but not the absence, of p53, and with the genetic interaction between *CDK8/19* and *TP53* in patients. If such a scenario was indeed true, it would suggest a therapeutic niche for CDK8/19 kinase inhibitors in cancer: they would be expected to trigger differentiation or death of tumor cells that have not yet acquired p53 mutations. Our results warrant further investigation of CDK8/19 inhibitors as therapeutic agents for p53-positive hepatocellular carcinoma.

Material and Methods

Patients

A total of 268 fresh-frozen tissue samples of HCC, associated with various etiologies, were included in this study. Patients and tumor features were already described in previously published studies and summarized in Supplemental Table 1. Written informed consent was obtained from all subjects in accordance with French legislation.

CDK18 and *CDK19* mRNA expression levels were assessed by quantitative RT-PCR using Fluidigm 96.96 Dynamic Arrays and specific TaqMan predesigned assays (CDK8= Hs00176209_m1; CDK19= Hs00292369_m1; Life Technologies, Carlsbad, CA). Data were calibrated with the RNA ribosomal 18S and changes in mRNA expression levels were determined using a comparative CT method using 5 normal tissue samples as control.

Mice experiments

All reported animal procedures were carried out in accordance with the rules of the French Institutional Animal Care and Use Committee and European Community Council (2010/63/EU). Animal studies were approved by institutional ethical committee (Comité d'éthique en expérimentation animale Languedoc-Roussillon (#36)) and by the Ministère de

l'Enseignement Supérieur, de la Recherche et de l'Innovation (D. Gregoire: APAFIS#11196-2018090515538313v2).

Cdk8$^{F/F}$ mice were generated as following: An 8076 bp genomic fragment (mouse chromosome 5: 146,254,503 to 146,262,579) enclosing the essential exon 2 (whose deletion results in loss of the essential catalytic lysine residue and causes a frameshift truncating over 90% of the protein) of the CDK8 gene was amplified by PCR from genomic DNA of 129/Sv embryonic stem cells and cloned into pGEM-T-easy. The diphtheria toxin A gene was cloned into the SacII site. 64 bp to the 3' of exon 2, the sequence CTCTAT was mutated to CTCGAG, generating an XhoI site. LoxP sites flanking exon 2 were generated by a combination of conventional cloning and recombineering, using an approach published in Liu et al., Genome Res. 2003 Mar;13(3):476-84. The loxP PGK-Neo cassette was amplified from pL452 plasmid with flanking AvrII/HindIII sites at each end and cloned into the AvrII site upstream of exon 2. Fragment orientation was confirmed by the generation of 3.5 kb HindII and 2.0 kb NheI sites, and the vector was recombined in *E.coli* strain SW106 with inducible Cre recombinase expression followed by HindII digestion, generating a single loxP site upstream of exon 2. Into this recombined vector, the FRT-PGK-Neo-FRT-LoxP cassette (amplified from pL451 with flanking XhoI sites) was cloned in the newly generated XhoI site downstream of exon 2, resulting in the "deletion construct". The orientation was confirmed by the generation of 2.2 kb Nhe1 and 3.4 kb BamHI sites. Functionality of the two recombination sites was tested as follows: the FRT site was confirmed by recombination in *E.coli* strain SW105 with inducible FlpE recombinase expression, deleting the FRT-Neo cassette and generating a 1.4 kb BamHI fragment; and the resulting plasmid was transformed in *E.coli* strain SW106 with inducible Cre recombinase expression, deleting exon 2 and resulting in a 1.1 kb BamHI fragment. The NotI linearized fragment of the deletion construct was transfected by electroporation into 129/Sv embryonic stem cells. 244 Neomycin-resistant colonies were genotyped by PCR and Southern blotting. Two probes were used: one outside the 3' end of the deletion construct, with HindIII digestion giving a single 9kb fragment for the WT and a 7kb fragment for the correctly-integrated deletion cassette, and one to the 5' end of the deletion cassette, again giving the same 9kb fragment for the WT but a 3.5 kb fragment for the deletion cassette. 10 colonies showed a correct integration by homologous recombination. These ES cells were injected into blastocysts obtained from pregnant Balb/c mice, and chimeric mice were crossed with C57/Bl6J mice constitutively expressing FlpE recombinase, removing the FRT-Neo cassette. Agouti mice were genotyped by PCR, showing correct insertion of the loxP sites around exon 2. Alb-Cre mice were described previously (29).

Allografts: Athymic Nude mice (*Hsd:Athymic Nude-Foxn1^{nu},* Envigo) or CDK8$^{F/F}$ mice were anesthetized with intra-peritoneal injection of Xylazine-Ketamine mixture. After incision of abdominal wall and peritoneum, the left lateral lobe of the liver was pulled out of the mouse

body. 50000 cells, resuspended in 5 µL of 25% Matrigel (BD) - PBS, were injected using a 10µL Hamilton syringe in the left lobe of the liver. After injection, liver was put back in normal position and the abdomen was sutured. Tumors were allowed to grow out for 4 weeks, then collected and fixed following classical procedures.

DEN induced carcinogenesis**:** Diethylnitrosamine (DEN) (30 mg/kg) was injected intra-peritoneally in 14 days old male mice. Mice were sacrificed and livers collected after a period of 8 months.

Hydrodynamic Gene Delivery**:** Hydrodynamic injections were performed in 6-8 week-old female mice as described previously (35). Briefly, 0.1 mL/g of a solution of sterile saline (0.9% NaCl) containing plasmids of interest were injected into lateral tail vein in 8-10 s. LentiCRISPRv2-sgTp53 (12.5 µg) and pT3-EF1a-N-RAS^{G12D}-GFP (12.5 µg) were injected together with sleeping beauty transposase SB100X (2.5 µg, ratio of 5:1). pCMV(CAT)T7-SB100 was a gift from Zsuzsanna Izsvak (Addgene plasmid # 34879).

Cell lines

BMEL (Bipotential Mouse Embryonic Liver) cell line was isolated from CDK8$^{F/F}$ mouse. Cells were grown on collagen-coated plates in RPMI medium (Gibco/Life Technologies) supplemented with 10% fetal calf serum (Pan-Biotech), insulin 10 µg/mL (Sigma), IGFII 30 ng/mL (Peprotech), EGF 50 ng/mL (Peprotech), 100 units/ml penicillin, and 100 mg/ml streptomycin. Other cell lines used, HepG2, Huh7, and HEK293T, were grown in Dulbecco modified Eagle medium (DMEM–high glucose, pyruvate, GlutaMAX–Gibco® LifeTechnologies) supplemented with 10% fetal bovine serum (Pan-Biotech). Cells were grown under standard conditions at 37°C in a humidified incubator containing 5% CO_2. All cells were routinely tested to confirm absence of mycoplasma contamination.

Generation of cell lines

pMSCV retroviral vectors (Clontech) encoding CRE recombinase, tamoxifen-inducible CRE-ERT2 (addgene plasmid #22776) or human H-Ras^{G12V} were used to generate stable BMEL cell lines. For induction of CRE-ERT2 activity, 4-OH tamoxifen was added to culture medium at a concentration of 2 µM for two weeks. CRISPR subgenomic RNA (sgRNA) targeting murine or human Cdk8, Cdk19 or Trp53 (CDK8 Mouse: 5'-ATCCCTGTGCAACACCCAGT-3'. CDK8_Human 5'-CGAGGACCTGTTTGAATACG-3' CDK19_Mouse: 5'-AAAGTGGGACGCGGCACCTA-3', CDK19_Human, 5'-ATTATGCAGAGCATGACTTG-3' Trp53 5'-ATAAGCCTGAAAATGTCTCC-3' ; all sequences from Zhang lab database) were cloned as synthetic dsDNA into lentiCRISPRv2 vector as described (51) (provided by F. Zhang, Addgene plasmid #52961). Generation of lentiviral particles and infection of BMEL,

Huh7 or HepG2 cells were carried out following classical procedures, and described previously (52). Successfully infected cells were selected with puromycin (2 µg/mL) or hygromycin (150 µg/mL) during 48h. Cell lines were further propagated and transgene expression or effects of targeted deletions verified by Western blot. Polyclonal cell lines were used for all subsequent experiments.

Soft agar

10^5 cells for each cell line were mixed with medium supplemented with 0.5% agarose and placed on top of the 1% agarose layer. 1 mL medium was added to the solidified layer and changed every 2-3 days. After 4 to 6 weeks, soft agar was stained with crystal violet 0,005% in 4% PFA for 1h. Colonies visible to the naked eye were counted manually.

Immunohistochemistry

Livers were fixed for 24 h in 10% neutral buffered formalin, embedded in paraffin, sectioned at 4 µm and stained with haematoxylin-eosin safran (HES) or subjected to immunohistochemical staining. Immunohistochemistry staining of GS (BD Transduction Laboratories, 610517, 1:200) and CDK8 (Santa Cruz, sc-1521, 1:400) were performed using classical procedures, with antigen retrieval in citrate buffer, and biotinylated secondary antibody coupled to streptavidin–peroxidase complex (ABC Vectastain kit; Vector Laboratories). Slides were digitally processed using the Nanozoomer scanner (Hamamatsu).

Protein isolation and Western blotting

Protein lysates of cells or tissues were prepared with lysis buffer (150mM NaCl, 50mM Tris pH 7.5, 0.2% Triton, 1mM EDTA; freshly added 1mM DTT and protease inhibitors cocktail (Roche)) and incubated on ice for 30 min. Samples were centrifuged for 10 min at 13000rpm and supernatant collected. Protein concentrations were determined by BCA protein assay (Pierce Biotechnology). All the samples were mixed 1:1 with Laemmli buffer and heated at 95°C for 5 min. Equal amounts of proteins were separated by SDS–PAGE (Biorad; usually 12% gels). Proteins were transferred onto PVDF membranes (Milipore). Red ponceau or amidoblack staining was done to check transfer efficiency and equal amount of protein loading. Primary antibodies used were: Cdk8 (Santa Cruz #sc1521, dilution 1:1000); Cdk19 (Abcam, dilution 1:1000); p53 (Cell signalling #2524 1:1000), Actin (Sigma A1978, 1:20000), Tubulin (DSHB, 1:400). Antibodies were diluted in 5% BSA in TBS-Tween and incubated overnight at 4°C. Secondary antibodies were either anti-goat IgG-HRP or anti-mouse antibodies IgG-HRP (Jackson immunoresearch). Band intensities were calculated using imageJ Lane Analysis.

RNA isolation, qPCR and RNA-Seq analysis

The RNA was extracted from either cells or liver tissue and purified using RNeasy mini kit (Qiagen) according to manufacturer's protocol. Reverse transcription of total RNA (1µg) was done with QuantiTect Reverse Transcription kit (Qiagen), and cDNA quantified using LC Fast stard DNA Master SYBR Green I Mix (Roche) with primers detailed below on LightCycler480 apparatus (Roche). Gene expression levels were normalized with hypoxanthine phospho-ribosyltransferase (HPRT). Primer pairs used for qPCR: Cdk8 5'-GAATTTCTATGTCGGCATGCAG-3' and 5'-ATAGTCAAAGAGAAGCCATACTTTCC-3', Glul 5'-TAGCTGTCACAAAGCGGGTGTA-3' and 5'-AGTGGAAATGTCAATCTCAGCC-3', Axin2 5'-ACCGGTCACAGGATGTC-3' and 5'-GACTCCAATGGGTAGCTCTTTC-3', c-myc 5'-CCGAGTGCATTGACCCCTCA-3' and 5'- GAGAAGGCCGTGGAATCGGA-3'. Hprt 5'-GCAGTACAGCCCCAAAATGG-3' and 5'-GGTCCTTTTCACCAGCAAGCT-3'.

For RNA-Seq analysis, RNA was extracted from exponentially growing subconfluent BMEL cells in three independent experiments for each cell line, using RNeasy mini kit (Qiagen) with DNAse treatment. RNA integrity was validated using RNA BioAnalyzer (Agilent), all RIN $\geq$ 9.5. The preparation of the library was done with the TruSeq Stranded mRNA Sample Preparation kit (Illumina). The sequencing was performed in an Illumina Hiseq 2500 sequencer by the Sequencing Platform of Montpellier (GenomiX, MGX, France; www.mgx.cnrs.fr), with 50 base pairs (bp) single end reads to an estimated depth of 25 million reads per sample. In order to perform a quality control of the sequencing, FastQC over the fastq files containing the raw reads. All the reads that passed the quality control were aligned to the mouse reference genome (GRCm38.p6) and the counts per gene were quantified using the tool STAR 2.6.0a(2). The Ensembl mouse genome annotations (release 93) were used for establishing the coordinates of each gene and their corresponding transcripts. Differential gene expression analysis was performed in R using the DESeq2 library. After normalization of the read counts per gene, a negative binomial generalized linear model was fitted considering single factor design for assessing the differential expression between CDK8/CDK19 knock-out BMEL Ras^{G12V} transformed cells and BMEL Ras^{G12V} transformed cells (as the control group). Wald test are performed for assessing statistical significance on the differential expression of each gene, then test are independently filtered and corrected by multiple hypothesis testing (Benjamini–Hochberg).

Statistical Analysis

Data sets were tested with 2-tailed unpaired Student *t* tests or Mann-Whitney U tests, correlations were analyzed with Pearson's χ^2 test using Prism Software version 8 (GraphPad). Significant *P* values are shown as: **P* <0.05, ***P* <0.01, ****P* <0.001, and *****P* <0.0001.

Human samples: Data visualization and statistical analysis were performed using R software version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org) and Bioconductor packages. Comparisons of the mRNA expression levels between groups were assessed using Mann-Whitney U test. Spearman's rank-order correlation was used to test the association between continuous variables. Univariate survival analysis was performed using Kaplan-Meier curve with log-rank test. The median *CDK8* and *CDK19* expression levels on the total number of analyzed samples was used to determine the low- and high- expression groups. P-value < 0.05 was considered as significant.

Data availability

The RNA-sequencing data have been deposited in the Gene Expression Omnibus (GEO, NCBI) repository.

Author contributions

KB, SP, DG performed and analyzed experiments. SC and JZR acquired and analyzed human patients data. AC generated CDK8$^{F/F}$ transgenic mouse line. GD analyzed RNA-Seq data. JUB, AL, JB contributed to *in vivo* experiments. DG made the figures. UH, DF and DG designed experiments, analyzed data, supervised the study and wrote the manuscript.

Acknowledgments

We acknowledge Montpellier Biocampus facilities: the imaging facility (MRI), the “Réseau d’Histologie Expérimentale de Montpellier” (RHEM) and Montpellier Genomix (MGX). We are grateful to zootechnicians of IGMM animal housing facility for their work. We thank Christina Begon-Pescia (IGMM transgenesis facility) for help in generation of CDK8 transgenic mouse line. We thank Scott Lowe for the gift of pT3-EF1a-N-RAS^{G12D}-GFP plasmid and Leila Akkari for help setting up hydrodynamic injections. We thank members of our labs for helpful discussions and comments.

This work was funded by Institut National du Cancer (INCa; PLBIO 2015-132, DF and UH teams) and EVA-Plan cancer INSERM THE (UH, JZR), and supported by SIRIC Montpellier Cancer Grant INCa_Inserm_DGOS_12553. DF is funded by the Ligue Nationale contre le Cancer (EL2017-LNCC/DF). DG benefited from support by

Association Française pour l'Etude du Foie (AFEF). JZR team is supported by the Ligue Nationale Contre le Cancer (Equipe Labellisée), Labex OncoImmunology (Investissement d'Avenir), the Fondation Bettencourt-Schueller (coup d'élan Award), the Ligue Contre le Cancer Comité de Paris (Duquesne award) and the Fondation pour la Recherche Médicale (Raymond Rosen award). SC is supported by a funding from "Labex OncoImunology", KB by INCa;PLBIO and JB by ANRS. The funders had no role in study design, data collection and analysis or publication process.

ID ORCID

Susana Prieto https://orcid.org/000-002-9746-2396

José Ursic-Bedoya: https://orcid.org/0000-0003-0076-2059

Stefano Caruso: https://orcid.org/0000-0002-6319-3642

Anthony Lozano : https://orcid.org/0000-0002-9749-0573

Jessica Zucman-Rossi: https://orcid.org/0000-0002-5687-0334

Urszula Hibner: https://orcid.org/0000-0002-5520-7311

Daniel Fisher: https://orcid.org/0000-0002-0822-3482

Damien Gregoire: https://orcid.org/0000-0002-1105-8115

References

1. Soutourina J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* 2018;19(4):262–274.

2. Westerling T, Kuuluvainen E, Makela TP. Cdk8 Is Essential for Preimplantation Mouse Development. *Mol. Cell. Biol.* 2007;27(17):6177–6182.

3. Li N et al. Cyclin C is a haploinsufficient tumour suppressor. *Nat. Cell Biol.* 2014;16(11):1080–1091.

4. Galbraith MD, Donner AJ, Espinosa JM. CDK8: A positive regulator of transcription. *Transcription* 2010;1(1):4–12.

5. Malumbres M et al. Cyclin-dependent kinases: a family portrait. *Nat. Cell Biol.* 2009;11(11):1275–1276.

6. Firestein R et al. CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature* 2008;455(7212):547–551.

7. Kapoor A et al. The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature* 2010;468(7327):1105–1109.

8. Morris EJ et al. E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. *Nature* 2008;455(7212):552–556.

9. Starr TK et al. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer.. *Science* 2009;323(5922):1747–50.

10. McCleland ML et al. *Cdk8* deletion in the *Apc* ^{Min} murine tumour model represses EZH2 activity and accelerates tumourigenesis: *Cdk8* knockout promotes tumorigenesis in Apc ^{Min} mice. *J. Pathol.* 2015;237(4):508–519.

11. Donner AJ, Ebmeier CC, Taatjes DJ, Espinosa JM. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nat. Struct. Mol. Biol.* 2010;17(2):194–201.

12. Galbraith MD, Espinosa JM. Lessons on transcriptional control from the serum response network. *Curr. Opin. Genet. Dev.* 2011;21(2):160–166.

13. Tsutsui T et al. Mediator complex recruits epigenetic regulators via its two cyclin-dependent kinase subunits to repress transcription of immune response genes.. *J. Biol. Chem.* 2013;288(29):20955–65.

14. Bahrami S, Drabløs F. Gene regulation in the immediate-early response process. *Adv. Biol. Regul.* 2016;62:37–49.

15. Donner AJ, Szostek S, Hoover JM, Espinosa JM. CDK8 Is a Stimulus-Specific Positive Coregulator of p53 Target Genes. *Mol. Cell* 2007;27(1):121–133.

16. Adler AS et al. CDK8 Maintains Tumor Dedifferentiation and Embryonic Stem Cell Pluripotency. *Cancer Res.* 2012;72(8):2129–2139.

17. Galbraith MD et al. HIF1A Employs CDK8-Mediator to Stimulate RNAPII Elongation in Response to Hypoxia. *Cell* 2013;153(6):1327–1339.

18. Perez-Perri JI et al. The TIP60 Complex Is a Conserved Coactivator of HIF1A.. *Cell Rep.* 2016;16(1):37–47.

19. Chen M et al. CDK8/19 Mediator kinases potentiate induction of transcription by NFκB. *Proc. Natl. Acad. Sci.* 2017;114(38):10208–10213.

20. Chiang MY et al. Identification of a conserved negative regulatory sequence that influences the leukemogenic activity of NOTCH1.. *Mol. Cell. Biol.* 2006;26(16):6261–71.

21. Porter DC et al. Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities.. *Proc. Natl. Acad. Sci. U. S. A.* 2012;109(34):13799–804.

22. Dhanasekaran R, Nault J-C, Roberts LR, Zucman-Rossi J. Genomic Medicine and Implications for Hepatocellular Carcinoma Prevention and Therapy. *Gastroenterology* 2019;156(2):492–509.

23. Ferlay J et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 2015;136(5):E359-386.

24. Boyault S et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 2007;45(1):42–52.

25. Zhao X et al. Regulation of lipogenesis by cyclin-dependent kinase 8–mediated control of SREBP-1. *J. Clin. Invest.* 2012;122(7):2417–2427.

26. Broude EV et al. Expression of CDK8 and CDK8-interacting Genes as Potential Biomarkers in Breast Cancer. *Curr. Cancer Drug Targets* 2015;15(8):739–749.

27. Calderaro J et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J. Hepatol.* 2017;67(4):727–738.

28. Nault J et al. A Hepatocellular Carcinoma 5-Gene Score Associated With Survival of Patients After Liver Resection. *Gastroenterology* 2013;145(1):176–187.

29. Postic C et al. Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J. Biol. Chem.* 1999;274(1):305–315.

30. Buchmann A et al. Mutational activation of the c-Ha-ras gene in liver tumors of different rodent strains: correlation with susceptibility to hepatocarcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 1991;88(3):911–915.

31. Strick-Marchand H, Weiss MC. Inducible differentiation and morphogenesis of bipotential liver cell lines from wild-type mouse embryos. *Hepatol. Baltim. Md* 2002;36(4 Pt 1):794–804.

32. Akkari L et al. Hepatitis C viral protein NS5A induces EMT and participates in oncogenic transformation of primary hepatocyte precursors. *J. Hepatol.* 2012;57(5):1021–1028.

33. Galbraith MD et al. CDK8 Kinase Activity Promotes Glycolysis. *Cell Rep.* 2017;21(6):1495–1506.

34. Donner AJ, Szostek S, Hoover JM, Espinosa JM. CDK8 is a stimulus-specific positive coregulator of p53 target genes. *Mol. Cell* 2007;27(1):121–133.

35. Chen X, Calvisi DF. Hydrodynamic transfection for generation of novel mouse models for liver cancer research. *Am. J. Pathol.* 2014;184(4):912–923.

36. Kang T-W et al. Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* 2011;479(7374):547–551.

37. Adler AS et al. CDK8 Maintains Tumor Dedifferentiation and Embryonic Stem Cell Pluripotency. *Cancer Res.* 2012;72(8):2129–2139.

38. Liang J et al. CDK8 Selectively Promotes the Growth of Colon Cancer Metastases in the Liver by Regulating Gene Expression of TIMP3 and Matrix Metalloproteinases. *Cancer Res.* 2018;78(23):6594–6606.

39. McDermott MSJ et al. Inhibition of CDK8 mediator kinase suppresses estrogen dependent transcription and the growth of estrogen receptor positive breast cancer [Internet]. *Oncotarget* 2017;8(8). doi:10.18632/oncotarget.14894

40. Clarke PA et al. Assessing the mechanism and therapeutic potential of modulators of the human Mediator complex-associated protein kinases. *eLife* 2016;5:e20722.

41. Czodrowski P et al. Structure-Based Optimization of Potent, Selective, and Orally Bioavailable CDK8 Inhibitors Discovered by High-Throughput Screening. *J. Med. Chem.* 2016;59(20):9337–9349.

42. Dale T et al. A selective chemical probe for exploring the role of CDK8 and CDK19 in human disease. *Nat. Chem. Biol.* 2015;11(12):973–980.

43. Nakamura A et al. CDK8/19 inhibition induces premature G1/S transition and ATR-dependent cell death in prostate cancer cells [Internet]. *Oncotarget* 2018;9(17). doi:10.18632/oncotarget.24414

44. Pelish HE et al. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* 2015;526(7572):273–276.

45. Porter DC et al. Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. *Proc. Natl. Acad. Sci.* 2012;109(34):13799–13804.

46. Jorda R et al. How Selective Are Pharmacological Inhibitors of Cell-Cycle-Regulating Cyclin-Dependent Kinases?. *J. Med. Chem.* 2018;61(20):9105–9120.

47. Gu W et al. Tumor-suppressive effects of CDK8 in endometrial cancer cells. *Cell Cycle* 2013;12(6):987–999.

48. Kim M-Y, Han SI, Lim S-C. Roles of cyclin-dependent kinase 8 and β-catenin in the oncogenesis and progression of gastric adenocarcinoma. *Int. J. Oncol.* 2011;38(5):1375–1383.

49. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras Provokes Premature Cell Senescence Associated with Accumulation of p53 and p16INK4a. *Cell* 1997;88(5):593–602.

50. Akkari L et al. Cell shape and TGF-beta signaling define the choice of lineage during in vitro differentiation of mouse primary hepatic precursors. *J. Cell. Physiol.* 2010;225(1):186–195.

51. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* 2014;11(8):783–784.

52. Vegna S et al. NOD1 Participates in the Innate Immune Response Triggered by Hepatitis C Virus Polymerase. *J. Virol.* 2016;90(13):6022–6035.

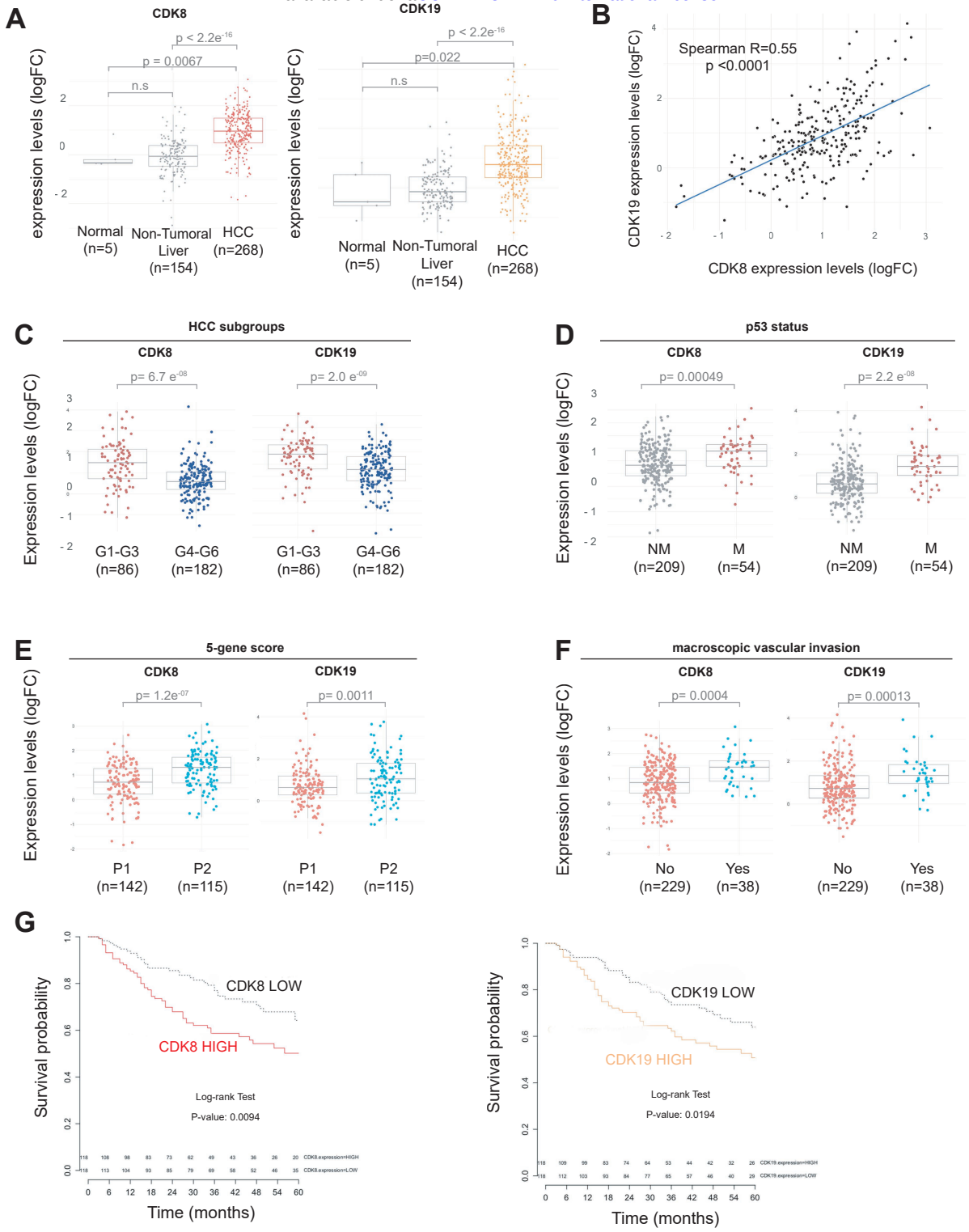


Figure 1: CDK8 and CDK19 are highly expressed in human HCC and associated with mutated p53 status. (**A**) Levels of CDK8 and CDK19 mRNA quantified by qPCR in control liver, non-tumoral part or malignant (HCC) tumor of patients. The fold change (LogFC) in gene expression is presented relative to mean expression level of the corresponding gene in five normal liver samples. Data are represented as Tukey's boxplots where box indicates the first and third quartiles, bar indicates median, whiskers indicate 1.5 interquartile range (IQR). P-values obtained with the Wilcoxon rank-sum test are indicated. (**B**) Correlation between the expression of CDK8 and CDK19 mRNA levels in human HCC samples (n=268) using Spearman's rank-order correlation (**C-F**) Levels of CDK8 and CDK19 mRNA in HCC classified according to molecular classification of HCC (Boyault et al., 2007), mutated status of p53, 5 gene score (Nault et al., 2013), or presence of macroscopic vascular invasion (histological identification), respectively. The number of patients in each class is indicated. The fold change (LogFC) in gene expression is presented relative to mean expression level of the corresponding gene in five normal liver samples. Data are represented as Tukey's boxplots where box indicates the first and third quartiles, bar indicates median, whiskers indicate 1.5 interquartile range (IQR). P-values obtained with the Wilcoxon rank-sum test are indicated. (**G**) Kaplan-Meier plots of overall survival in HCC patients (from(Nault et al., 2013)) with high or low expression of CDK8 or CDK19 mRNA. High/low expression groups were defined by the median expression levels of the total number of analyzed samples. Statistical differences were assessed by Log-rank Test. P-value obtained are indicated.

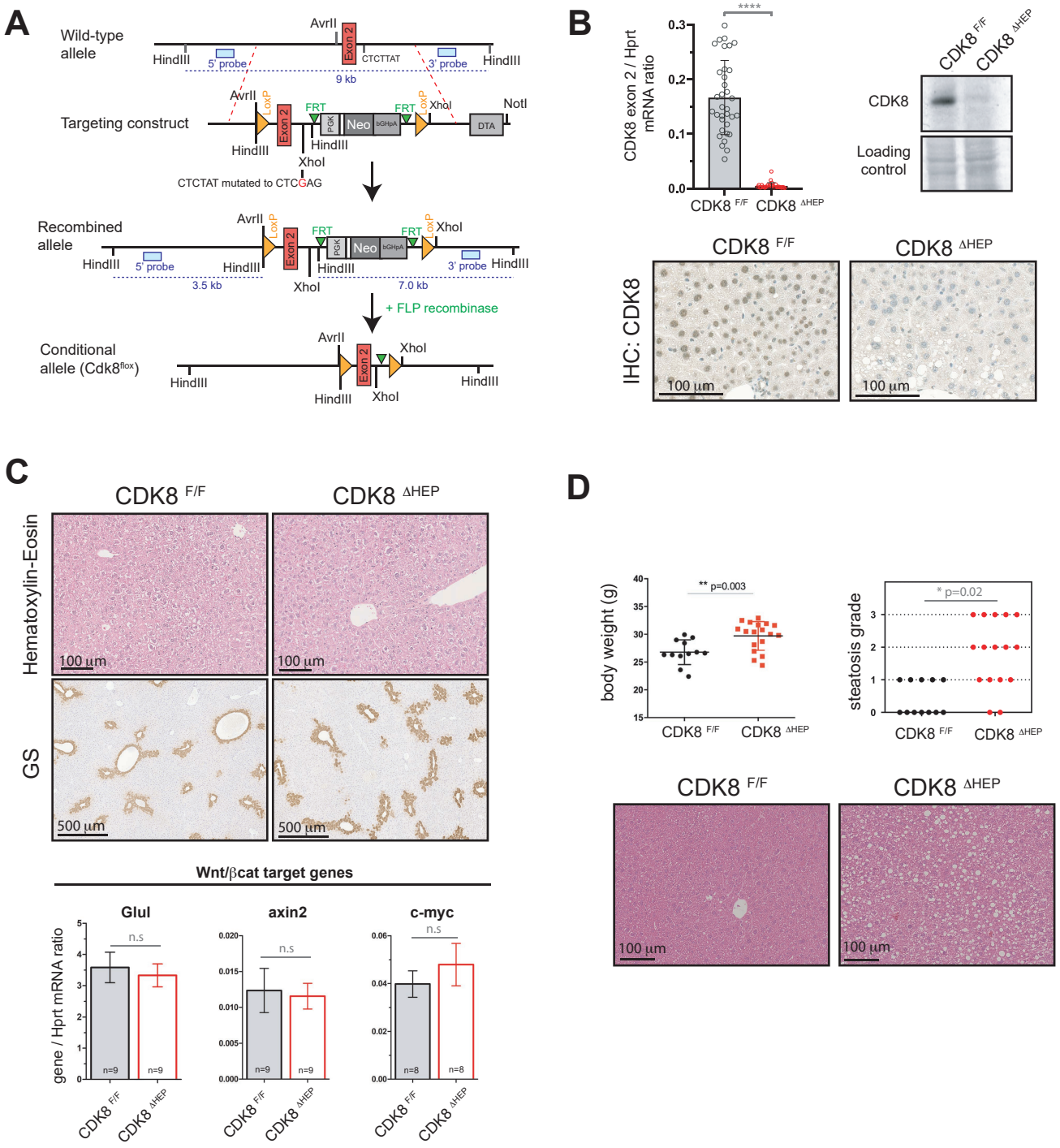
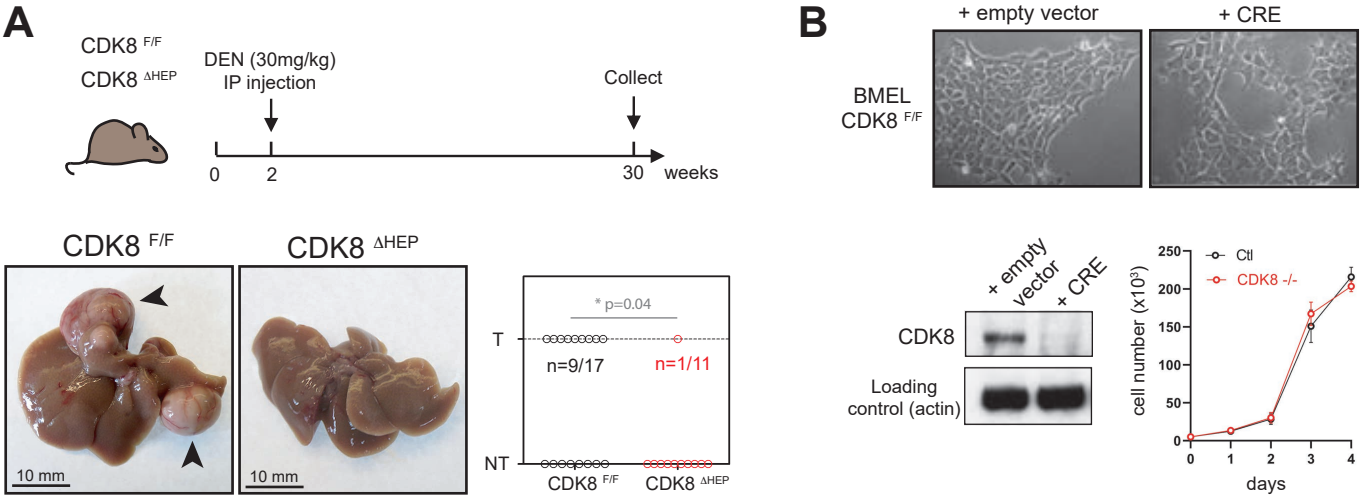
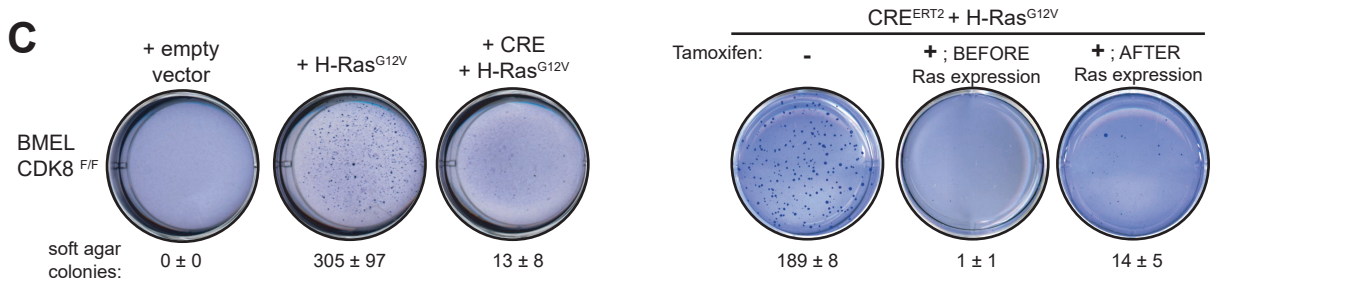


Figure 2: Hepato-specific ablation of CDK8 does not affect liver homeostasis, but induces steatosis. (**A**) Map of CDK8 genomic locus and strategy used to generate CDK8$^{F/F}$ transgenic mouse. See material & methods section for details. (**B**) qPCR, Western-blot and immunohistochemical analysis of CDK8 removal in hepatocytes of CDK8$^{\Delta HEP}$ mice. For qPCR quantification, males and females mice of different ages were used. CDK8 $^{F/F}$ (n=34) or CDK8$^{\Delta HEP}$ (n=30). WB and IHC show results for 3 month old female mice. No differences were observed depending on sex or age of the animal. (**C**) HE staining of, GS staining, and qPCR analysis of β-catenin target genes (*Glul, axin2, c-myc*) in wild-type and CDK8 depleted 3 months old mice. n.s: not significant (t-test). (**D**) Body weight and steatosis grade score of 6 months old CDK8$^{F/F}$ and CDK8$^{\Delta HEP}$ female mice. P value of Fisher's exact test is indicated. Representative liver sections (HES staining) of mice from both genotypes are shown.





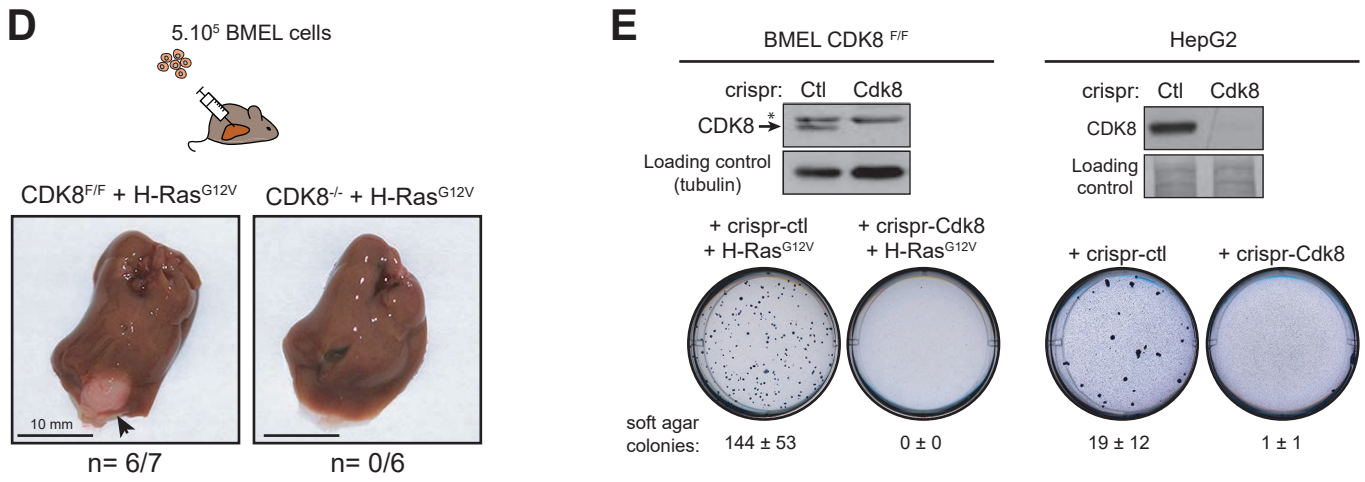


Figure 3: CDK8 is required for hepatic carcinogenesis. (**A**) Effect of CDK8 depletion on DEN-induced hepatic carcinogenesis. Pictures of representative livers for each genotype are shown. Arrowheads indicate tumors. Graph represents the repartition of the 28 DEN-injected animals, in each genotype depending if the liver presents a visible tumor (T) or not (NT). P-value of Fisher's exact test is indicated. (**B**) Pictures and growth curve of CDK8$^{F/F}$ hepatic progenitors (BMEL) WT or depleted of CDK8 by CRE recombination. (**C**) Soft agar growth of different BMEL cell lines. A representative well for each cell line is shown. Mean number of colonies per well ± SD from at least three independent experiments are indicated. Tamoxifen treatment (+Tam) was realized before or after transformation induced by H-Ras^{G12V} expression. (**D**) Result of orthotopic allografts of Ras expressing BMEL depleted or not of CDK8. A representative liver is shown, and the number of livers carrying a tumor out of total liver injected is indicated. Arrowhead points out to the tumor. (**E**) Western-blot characterization of CDK8 removal by crispr/cas9 gene editing and consequences on growth in soft agar for BMEL (left panel) and HepG2 (right panel) cells. Asterisk indicates an aspecific band sometimes detected in BMEL cells. Mean ± SD number of colonies per well from at least three independent experiments are indicated.

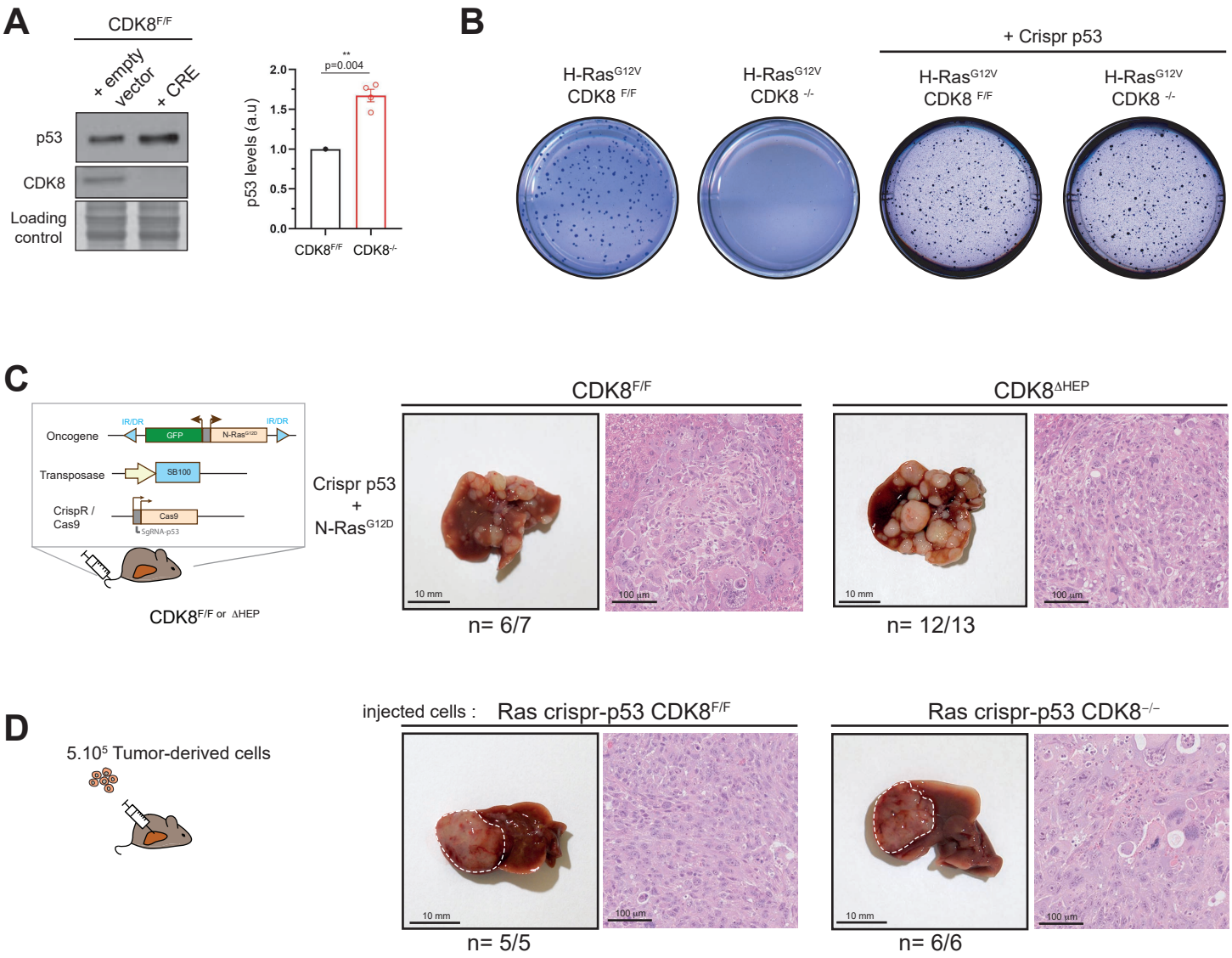


Figure 4: p53 deficiency abrogates CDK8 requirement for hepatic transformation. (**A**) Western Blot and quantification of p53 protein levels in BMEL cell lines, showing stabilization of p53 in absence of CDK8. A representative western blot and quantification from 4 independent experiments (mean ± SD) are shown. For each experiment, the band intensity value of the $CDK8^{F/F}$ sample was normalized to 1. P-value from one-sample t test is indicated. (**B**) Soft agar test for indicated BMEL cell lines. (**C**) Hydrodynamic gene delivery of $N\text{-}Ras^{G12D}$ and crispr-p53 vectors into livers of $CDK8^{F/F}$ or $CDK8^{\Delta HEP}$ female mice. Representative HES staining of tumors are shown. For each genotype, number or mice presenting liver tumors out of number of mice injected is indicated. (**D**) orthotopic allografts of tumor cell lines derived from $CDK8^{F/F}$ or $CDK8^{\Delta HEP}$ Ras crispr-p53. White dashed lines indicated tumor periphery. Representative HES staining of tumors are shown. For each genotype, number or mice presenting liver tumors out of number of mice injected is indicated.

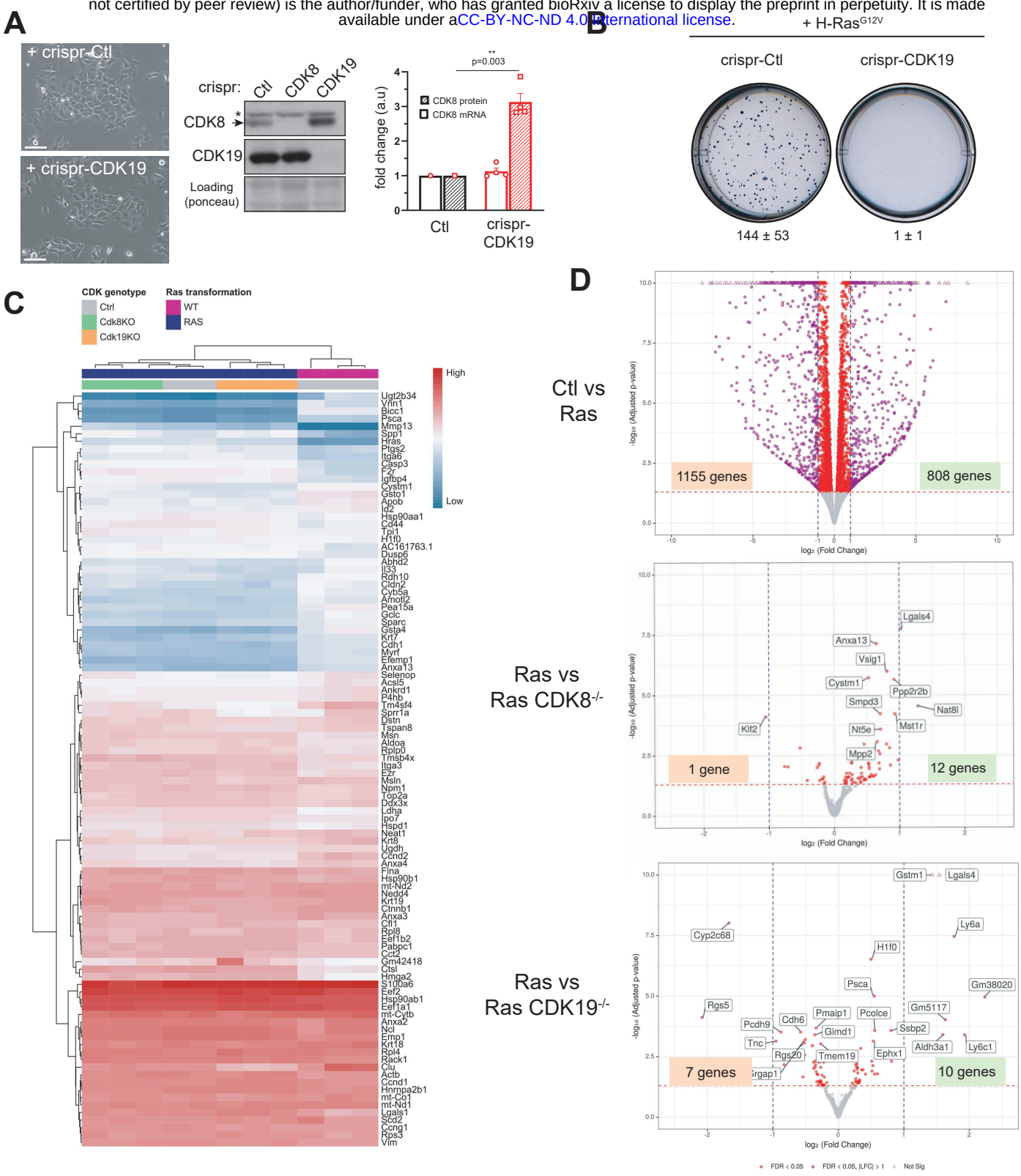


Figure 5: CDK19 and CDK8 are both required for transformation by H-Ras^{G12V} . (**A**) Pictures, western blot and quantification of BMEL cell lines depleted of CDK19 by crispr/Cas9. A representative western blot and quantification from 4 independent experiments (mean ± SD) are shown. * indicates an aspecific band sometimes detected with CDK8 antibody. For each experiment, the band intensity value or cDNA content of the crispr-Ctl sample was normalized to 1. P-value from one-sample t test is indicated. (**B**) Soft agar test for indicated BMEL cell lines. A representative well for each cell line is shown. Mean number of colonies per well ± SD from at least three independent experiments are indicated. (**C**) Heatmap for the RNASeq experiments. The raw counts were transform using the regularized log and the variance was calculated between all the samples for all the genes. We selected the 100 genes with the higher variance. Distance between genes and samples were computed using the euclidean measure and the hierarchical cluster analysis was perform in order to set the distance trees. Two level of colors were set to describe the sample genotypes for RAS transformation (Violet: wild type; Blue: RAS^{G12V} mutation) and for Cdk8 or Cdk19 knock-outs (Gray: Control; Green: Cdk8 KO; Orange: Cdk19 KO). (**D**) Volcano plots for the differential gene expression. The values of the minus logarithm in base ten versus the logarithm in base two of the counts are plotted for each gene and two levels of significance are showed in red (genes with adjusted p-value less than 0.05) and in purple (genes with adjusted p-value less than 0.05 and with absolute values of the log2 of the fold change greater than one).