Species-wide survey of transcript and protein abundance variation in yeast

Élie Teyssonnière

Species-wide survey of transcript and protein abundance variation in yeast

Élie Teyssonnière
HaploTeam | UMR7156 | CNRS | Université de Strasbourg | 2023

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

UNIVERSITÉ DE STRASBOURG

ECOLE DOCTORAL ED414

UMR7156 Génétique Moléculaire, Génomique et Microbiologie

Thèse présentée par :

Élie TEYSSONNIÈRE

Soutenue le : 20 octobre 2023

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/Spécialité : **Génétique**

Analyse de la variation de l'abondance des transcrits et des protéines à l'échelle de l'espèce chez la levure

Ph.D. supervisé par :

Pr. Joseph SCHACHERER	Professeur, Université de Strasbourg
Dr. Anne FRIEDRICH	Maître de Conférences, Université de Strasbourg

Rapporteurs externes

Pr. Maria MAR ALBÀ	Professeure, Hospital del Mar Medical Research Institute, Barcelone
Dr. Guillaume DISS	Directeur de recherche, Friedrich Miescher Institute for Biomedical Research, Bâle

Examinateurs

Dr. Stéphanie BLANDIN	Chargée de recherche, Université de Strasbourg
Pr. Michael KNOP	Professeur, Zentrum für Molekulare Biologie der Universität Heidelberg

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

UNIVERSITY OF STRASBOURG

DOCTORAL SCHOOL ED414

UMR7156 Molecular Genetics, Genomics and Microbiology

Doctoral dissertation presented by:

Élie TEYSSONNIÈRE

defended on: October 20th, 2023

In partial fulfillment of the requirements of a degree of: **Doctor of Philosophy**

Discipline/Specialty: **Genetics**

Species-wide survey of transcript and protein abundance variation in yeast

Ph.D. advised by:

Pr. Joseph SCHACHERER	Professor, University of Strasbourg
Dr. Anne FRIEDRICH	Associate professor, University of Strasbourg

External reporters

Pr. Maria MAR ALBÀ	Professor, Hospital del Mar Medical Research Institute, Barcelona
Dr. Guillaume DISS	Group leader, Friedrich Miescher Institute for Biomedical Research, Basel

Examiners

Dr. Stéphanie BLANDIN	Group leader, University of Strasbourg
Pr. Michael KNOP	Professor, Zentrum für Molekulare Biologie der Universität Heidelberg

Abstract

An astonishing phenotypic diversity can be observed in natural populations. One of the major goals of modern biology is to unravel the genetic origins of this phenotypic landscape. Gene expression is known to be a main determinant of the relationship between genotypes and phenotypes. In recent decades, several analytical and technical advances have made it possible to study gene expression at every step of the expression process (*e.g.*, transcriptome and proteome) and at very large scales. However, a complete exploration of gene expression across the entire process and at the population scale is still lacking. The goal of this dissertation is to get a more comprehensive view of how each layer of gene expression varies, influences each other, and is related to the natural genetic diversity observed within species. To this end, we analysed the transcriptomes and proteomes of a large natural population of *S. cerevisiae* (bringing together more than 1,000 individuals) and found unsuspected differences between mRNA and protein abundance regulation. Simultaneously, we studied the gene expression process at three different molecular levels (transcriptome, translatome and proteome) and found that important buffering mechanisms underlie the expression variation between individuals.

ACKNOWLEDGEMENTS

First of all, I would like to thank the members of the committee, Pr. Maria Del Mar Alba Soler, Dr. Guillaume Diss, Pr. Michael Knop and Dr. Stéphanie Blandin for accepting to evaluate this work. The following work was carried out at the Department of Molecular Genetics, Genomics and Microbiology (GMGM), UMR7156/CNRS at the University of Strasbourg, under the supervision of Pr. Joseph Schacherer and Dr. Anne Friedrich.

Tout d'abord, merci Joseph de m'avoir accompagné et tant appris depuis 2017 ! D'abord en cours de génétique quantitative où mon aversion pour la génétique que j'avais développée en licence s'est ébranlée, jusqu'à la rédaction de cette thèse, ton exigence scientifique et ta rigueur professionnelle m'ont énormément apporté. Tu m'as permis de repousser certaines limites que je n'aurais pas soupçonnées atteindre (en fait, je pense principalement à l'Ekiden 2021). J'ai réellement découvert ce qu'est la science pendant ces 4 années de thèse et j'y ai pris (en moyenne) beaucoup de plaisir ! Merci de m'avoir débauché en 2019 lorsque je m'orientais vers quelque chose d'horriblement moléculaire, merci de m'aider (pendant mon master et encore aujourd'hui) à réaliser certains projets qui me sont chers (🇯🇵), bref, merci pour tout !

Aux membres de l'Haploteam, vous avez réussi à rendre ces 4 années de vie au laboratoire agréables et légères en dépit des sombres heures qu'un étudiant en thèse devrait rencontrer. Au final, une thèse dans un environnement où les gens sont chouettes, ce n'est vraiment pas si pire. Anne, merci d'avoir été disponible à chaque fois que j'avais la moindre question, ou que pour X ou Y raison, j'avais besoin d'aide. Tu feras partie des personnes que je mettrai dans ma liste des « modèles professionnels » ! Andreas, Victor, Pia, Elodie, Emna, Gauthier, Omar, Carmen, Abhishek, Jan, Sam, Arthur, Chris, merci de m'avoir accompagné à un moment ou un autre pendant ces 4 ans. Entre Karaoké, trampoline, foot, squash, basket, badminton et moult restaurants, je n'ai pas vu le temps passer ! J'espère avoir encore l'occasion de vous mettre une peignée au bad ☺. Jing, merci d'avoir été la source de discussions scientifiques (ou pas) passionnantes et d'avoir été un genre de cheat code quand il m'arrivait d'être bloqué sur mes projets. Au final, ce n'est vraiment pas sorcier la science quand t'es dans le coin ! La liste est longue, mais merci Emilien de m'avoir appris comment manigancer toutes sortes de complots avec finesse et d'avoir été un grand partenaire de foot et de GeoGuessr. À Claudia, Claudine et Isabel, merci de votre aide toujours précieuse et pertinente pendant ces 4 années ! Marion, merci d'avoir sacrifié tant de repas pour céder à mes caprices de restaurants ! Merci d'avoir supporté nos horribles goûts musicaux tous les vendredis et d'avoir été la parfaite partner in crime dans

nos interminables discussions/debriefings de tout ce qui se passait de près ou de loin du labo ! De toutes les Potterheads du GMGM, tu es de loin la meilleure ! continue comme ça ! Pabien, dois-je te mettre avec le labo tant tu squattes ma vie et mon appart… Je te suis extrêmement reconnaissant d'avoir apporté toute ta positivité-anh, autant au travail qu'à côté-anh, tu es un véritable être de lumière et de joie-anh. Merci pour *les mojitos au QG avec Sandrine et Béa*, ton inébranlable patience lors de nos sessions SDA JCE ou HA, THE session karaoké (c'était super, j'imagine ??), les babysittings, ton pitoyable niveau à MK8, les ANH-UNH, les échecs et j'en oublie… Puissent tes grandes et nombreuses ambitions se réaliser V•ᴥ•V.

Et puisque que la vie à l'université ne se limite pas à une équipe, merci à tous ceux qui ont entouré de près ou de loin mon passage à l'Unistra. Noémie, merci d'avoir été une source de distraction continue et chronophage depuis nos premières aventures en IDS dans la troupe Isa. Entre la tite coloc, les innombrables bouffes plus ou moins saines, les aventures aux urgences, au subway, l'écriture en BU et j'en passe, la thèse aurait été beaucoup trop plate sans une téléréalité IRL. Fais plus de sport et arrête de fumer ! Mon petit LF, tu es le meilleur (t)déiste que je connaisse, merci d'avoir été mon second en master et d'avoir survécu à mes innombrables réflexions de vie inutiles ! T'y es presque ! Merci à Aline sans qui ma thèse n'aurait été qu'un combat contre vents et démarches… Merci à Isabelle Caldelari et à Thierry Nadalig de m'avoir permis de découvrir l'enseignement ! J'y pris beaucoup de plaisir, en grande partie parce que j'étais bien accompagné pour ça !

Merci aussi à tous ceux qui de près ou de loin ont rendu cette thèse possible, depuis les équipes du centre Léon Berard à Lyon, à l'équipe pédagogique du master microbio à Strasbourg en passant par celle de la licence sciences de la vie à Villeurbanne. Merci à tous ceux qui m'ont permis de me vider l'esprit pendant ces 4 années, que ce soit au foot, à Festibad ou pendant d'interminable sessions JDR ! Merci à mes amis, la fine équipe, les gros pores, la team Ardéchoise, le caviar… et enfin je souhaite remercier et adresser une pensée à M. Frank Hekking qui m'a donné le coup de pouce dont j'avais besoin pour me lancer dans cette aventure.

Pour finir, merci ne suffit pas pour ma famille (mes grands-parents, mes oncles et tantes, mes cousins plus ou moins germains…) et particulièrement mes (ti-) parents qui m'ont soutenu et ont cru en moi sans discontinuer dans les hauts et bas et qui m'ont toujours guidé vers les choix les plus judicieux et réfléchis ! Vous êtes trop forts et incroyables, je ne serai jamais à votre niveau ! Pareil pour Tinou qui est là depuis que j'ai des souvenirs en étant la pierre angulaire de ma bêtise et qui en dépit de sa modeste taille est tout bonnement the GOAT ! c'est une

véritable fierté d'être ton frère (enfin pas vraiment, faut qu'on t'explique avec Papa et Maman…). Vous êtes les meilleurs et de très loin !!

PiniH, un texte sera insuffisant… Entre les escapades à vélo perdues en Slovénie, les pubs et randos aux fin fond du pays de Galles, juste merci d'être là, de me supporter, de me soutenir et d'être toi. Ma petite Albane, continue de bien grandir, de bien manger et essaye de dormir plus longtemps, j'ai une chance inouïe que tu sois là avec nous !

Analyse de la variation de l’abondance des transcrits et des protéines à l’échelle de l’espèce chez la levure

Introduction

Comprendre l'origine de l'importante diversité phénotypique observée au sein des populations naturelles est au cœur de la biologie moderne. Plus spécifiquement, la détermination des origines génétiques sous-jacentes aux variations de traits observées chez des individus génétiquement distincts est un prérequis indispensable dans de nombreux domaines tels que la médecine, l'agro-alimentaire ou les sciences environnementales. Les liens entre la diversité génétique et phénotypique, aussi appelés relation génotype-phénotype, sont le résultat d'une multitude de facteurs (environnementaux ou internes à l'organisme) influençant un organisme sur plusieurs échelles (moléculaire, cellulaire, tissulaire, par exemple). Le processus d'expression de gènes est quant à lui l'un des principaux moteurs de la relation génotype-phénotype (Aguet et al., 2023; Albert and Kruglyak, 2015; Tak and Farnham, 2015). En effet, de nombreux travaux ont montré l'impact important des modifications de l'expression génique sur de nombreuses pathologies humaines (Corbett, 2018; Lee and Young, 2013). Ainsi, comprendre comment varie l’expression des gènes à travers les individus est essentiel dans l’exploration de la relation génotype-phénotype. En dépit de sa nature linéaire (l’ADN est transcrit en ARN qui est traduit en protéine), le processus d’expression des gènes est incroyablement complexe. Chacune de ses étapes (transcription, traduction, dégradation des ARNm et des protéines…) est finement régulée via plusieurs centaines de facteurs cellulaires et l’aspect hiérarchique du processus cache en fait de nombreuses interactions entre ARN et protéines (Buccitelli and Selbach, 2020; Liu et al., 2016). Aussi, les mécanismes de variations de l’expression des gènes à l’échelle des populations sont à ce jour grandement incompris.

Au cours des deux dernières décennies, de nombreux progrès techniques et analytiques ont débouché sur une large gamme d’outils permettant à la fois des mesures fines et à grande échelle de l’expression des gènes, mais aussi de relier les variations de cette expression au fond

génétique des individus. Sur le plan technique il est possible de citer la mise en place du RNA-sequencing (Stark et al., 2019) pour quantifier le transcriptome des individus, le développement à plus large échelle de la spectrométrie de masse en tandem pour leur protéome (Demichev et al., 2020; Messner et al., 2023, 2022), et enfin la création de techniques plus spécifiques comme le ribosome-profiling permettant de mesurer avec précision le processus de traduction (Ingolia et al., 2019). Au niveau analytique, l'extension à grande échelle des techniques précédemment citées a été accompagnée par le développement des analyses d'associations pangénomiques (Dehghan, 2018) (*Genome-Wide Association Studies* - GWAS). En utilisant de large cohortes d'individus, ces dernières permettent d'associer des variants génétiques, généralement des variants nucléotidiques, aux variations d'un phénotype précis quantifiées dans une population. Les variants génétiques ainsi détectés sont appelés QTL (pour *Quantitative Trait Loci*) et dans le cas de GWAS visant à étudier les variations des niveaux d'ARNm ou de protéines, on parle respectivement d'eQTL (expression QTL) et de pQTL (protein QTL) (Aguet et al., 2023). Cependant et malgré ces différentes avancées, de nombreux aspects concernant les variations d'expression de gènes restent méconnus, notamment à l'échelle des populations. Par exemple, il n'y a toujours pas de consensus sur le degré de similarité entre les variations des transcriptomes et des protéomes à travers de larges groupes d'individus (Buccitelli and Selbach, 2020; Fortelny et al., 2017; Liu et al., 2016). De plus, la similitude entre les eQTL et les pQTL est encore grandement débattue car les résultats des études à ce sujet sont contradictoires.

C'est dans ce contexte que prend place mon projet de thèse. Il s'agit d'explorer simultanément plusieurs étapes de l'expression des gènes en utilisant une population naturelle de la levure *S. cerevisiae* pour laquelle le génome a été entièrement séquencé via la technique Illumina (Peter et al., 2018) et où un large jeu de données lié au transcriptome a déjà été généré (Caudal et al., 2023). Cependant, l'exploration à grande échelle de l'expression génique est encore difficile à cause de certaines limitations techniques, notamment au niveau protéomique où il est difficile de quantifier avec précision un grand nombre de gène lorsqu'on travaille sur de nombreux individus. De ce fait, ma thèse s'est articulée autour de deux principaux projets : l'un portant sur la quantification précise de l'expression des gènes à travers chacunes des étapes de cette dernière et donc quantifiant un grand nombre de gènes à travers un nombre limité d'individus (8 isolats naturels de *S. cerevisiae*), l'autre portant sur l'exploration à travers un très grand nombre d'individus du transcriptome et du protéome de *S. cerevisiae* mais en ayant une couverture génique plus faible.

Resultats

La variation de traduction à travers différents fonds génétiques révèle une signature de l'atténuation post-transcriptionnelle chez la levure

Dans ce projet, nous avons quantifié avec précision l'expression de plus de 4 344 gènes dans 8 isolats naturels de *S. cerevisiae* provenant d'environnements très différents. La quantification de l'expression de gènes s'est faite via du RNA-sequencing, du ribosome-profiling et de la spectrométrie de masse en tandem, permettant d'avoir une vision globale des dynamiques de régulation tout au long du processus d'expression.

Plusieurs points ont été explorés au cours de ce projet. Tout d'abord, en comparant les données de transcription et de traduction (donc, de RNA-sequencing et de ribosome-profiling fait en collaboration avec le Riken Institute - Tokyo), nous avons observé que les variations d'expression de gènes semblaient principalement liées aux préférences trophiques de chacune des souches de levures. Celles-ci ayant des origines écologiques très variées, elles ont probablement adapté la régulation des gènes liés au métabolisme en conséquence. Si ces résultats sont partagés entre la régulation transcriptionnelle et traductionnelle, ces dernières sont pourtant assez différentes en termes d'intensité. En effet, nous avons observé que les variations d'abondance d'ARNm sont plus importantes que les variations observées sur les données de ribosome-profiling. Ceci est dû à un phénomène appelé atténuation post-transcriptionnelle (post-transcriptional buffering) déjà décrit dans des études antérieures (Artieri and Fraser, 2014; Blevins et al., 2019; McManus et al., 2014; Wang et al., 2020). Ce phénomène suggère que les étapes avancées de l'expression de gènes sont plus conservées et donc plus contraintes évolutivement parlant. Nous avons aussi détecté que l'atténuation post-transcriptionnelle affecte préférentiellement certains types de gènes, comme les gènes essentiels, les gènes liés à des complexes protéiques ou même les gènes ayant un faible niveau d'expression. Ceci n'avait pas été démontré jusqu'à présent et permet d'éclaircir les mécanismes sous-jacents à l'atténuation post-transcriptionnelle qui sont encore très méconnus. Enfin nous avons utilisé les données d'abondances transcriptionnelles et traductionnelles pour explorer comment certains gènes présents che*z S. cerevisiae* mais issus d'espèces différentes de levures sont exprimés. Ces gènes sont issus de mécanismes d'échange de matériel génétiques entre espèces (comme les introgressions ou les transferts horizontaux de gènes) et la régulation de leur expression a rarement été explorée (D'Angiolo et al., 2020; Marsit et al.,

2015; Novo et al., 2009; Peter et al., 2018). Nous avons pu observer différents profils d'expression en fonction de l'origine de ces gènes. Les introgressions avait par exemple des niveaux d'expression similaires en comparaison avec leur orthologues alors que les gènes issus de transferts horizontaux sont pour leur part moins traduit que les autres gènes de *S. cerevisiae*.

Nous avons ensuite étendu l'exploration de l'expression de gènes de ces 8 souches au niveau protéique (en collaboration avec le Weizmann Institute of Science - Israël). Plusieurs résultats précédemment obtenus sont retrouvés au niveau protéique. Tout d'abord les variations d'abondance protéique sont principalement liées aux gènes du métabolisme, ce qui confirme les résultats obtenus en utilisant des données de RNA-sequencing et de Ribosome-profiling. De plus, en comparant le niveau de variation au sein de chacune des étapes d'expression de gènes, nous avons aussi pu observer le phénomène d'atténuation post-transcriptionnelle, les variations d'abondance protéique étant les plus faibles, et donc cette étape du processus d'expression est de fait la plus conservée à travers les 8 isolats. Ces résultats confirment la présence du phénomène d'atténuation post-transcriptionnelle. Nous nous sommes par la suite penchés plus précisément sur les différences de contraintes évolutives entre chacune des étapes du processus d'expression. Dans cette optique, nous avons utilisé une base de données décrivant plusieurs centaines de caractéristiques des gènes de *S. cerevisiae*. À l'aide d'une quantification précise de la vitesse évolutive associée à chacune de ces caractéristiques (Wang et al., 2020), nous avons pu montrer que l'évolution de la régulation de l'abondance génique, bien qu'ayant de nombreuses spécificités pour chacune des étapes de l'expression, suit plusieurs principes communs, notamment que les gènes ayant une place centrale dans les interactions entre protéines ou ayant un rôle fondamental dans le fonctionnement cellulaire verront leur expression grandement conservée à travers les individus. À l'inverse et en adéquation avec les résultats précédemment cités, les gènes liés au métabolisme et aux capacités respiratoires des cellules ont une expression qui évolue beaucoup plus rapidement que les reste des gènes.

Les transcriptomes et protéomes quantitatifs à l'échelle de l'espèce révèlent un contrôle génétique distinct de la variation de l'expression génique chez la levure

L’objectif de ce projet était d’explorer à l'échelle d’une population entière (dans ce cas, plus de 900 isolats naturels de *S. cerevisiae*) les variations d’abondance des ARNm et des protéines. Ce projet a été rendu possible grâce à de nouvelles techniques (expérimentales et analytiques) d’exploration des protéomes à très large échelle développées récemment (Demichev et al., 2020; Messner et al., 2023). Dans ce contexte, nous avons combiné des données quantifiant la quasi-totalité des ARNm dans 989 isolats (Caudal et al., 2023) avec des données nouvellement générées en collaboration avec le Charité Institute (Berlin) et le Francis Crick Institute (Londres) dans 942 isolats. La combinaison des données de transcriptomes et de protéomes ont permis d’avoir une estimation complète et précise de 629 gènes à travers 888 souches.

Ce jeu de données à très large échelle nous a permis de répondre à différentes questions concernant les relations et interactions entre le transcriptome et le protéome à l'échelle d’une espèce entière. Tout d’abord, nous nous sommes intéressés à la question de la corrélation entre les abondances d’ARNm et de protéines pour chaque gène à travers tous les isolats. Le niveau de corrélation entre les variations du transcriptome et du protéome à travers de large cohortes d’individus étant à ce jour toujours débattu. Nous avons observé que le degré de corrélation est en moyenne assez faible (corrélation de Spearman moyen = 0.165), ce qui signifie que les variations d’expression de gènes à travers une espèce sont assez différentes si l’on s'intéresse au transcriptome ou au protéome. Nous avons pu observer que les gènes liés au métabolisme étaient généralement associés à de plus hauts niveaux de corrélation ARNm-protéine, ceci étant dû au haut niveau de variation d’abondance d’ARNm et de protéines entre individus de ces gènes. Nous avons aussi pu observer que ces gènes du métabolisme étaient souvent détectés comme signature du processus de domestication de *S. cerevisiae*. De façon similaire, nous avons détecté une sous-expression basale des gènes liés à la respiration chez les souches dites domestiquée, ce qui est probablement dû au fait que ces isolats ont été sélectionnés et utilisés dans des contextes de fermentation. Ce genre de signature est consistant avec des études réalisées précédemment (Lahue et al., 2020).

Une des causes possibles de la dissimilarité globale entre transcriptome et protéome est que les abondances d’ARNm et de protéines sont influencées par des facteurs génétiques propres à chacune des étapes du processus d’expression. Nous nous sommes alors penchés sur l’exploration des origines génétiques des variations du transcriptome et du protéome. Les 888 isolats utilisés pour cette étude ayant été complètement séquencés dans une étude antérieure, nous avons pu utiliser la technique d’association pangénomique. Pour le transcriptome et le

protéome, nous avons détecté près de 1,200 associations entre des polymorphismes nucléotidiques et des variations d'abondance d'ARNm et de protéines (596 eQTL et 598 pQTL). De façon surprenante, seulement 22 QTL sont partagés entre les eQTL et les pQTL. Bien que ceci soit en accord avec le faible niveau de corrélation entre les variations du transcriptome et du protéome, cela reste beaucoup plus bas que ce qui a été observé précédemment (Albert et al., 2014; Battle et al., 2015; Buccitelli and Selbach, 2020; Jiang et al., 2020).

Conclusion

L’ensemble des travaux réalisés pendant ma thèse ont permis d’éclaircir les mécanismes de régulations qui façonnent les variations d'expression de gènes à l’échelle d’une population. Le message principal étant que chacune des étapes du processus d’expression suit des dynamiques de régulations propres, avec une tendance globale à la diminution de la variation d’expression observée entre chaque individu au fur et à mesure de l’expression génique. Dans le cadre de la compréhension de la relation génotype-phénotype, il est donc primordial de considérer l’impact d’un variant tout au cours du processus d’expression génique pour comprendre comment il peut impacter la diversité des traits observés au sein d’une population naturelle.

Résumé graphique du projet

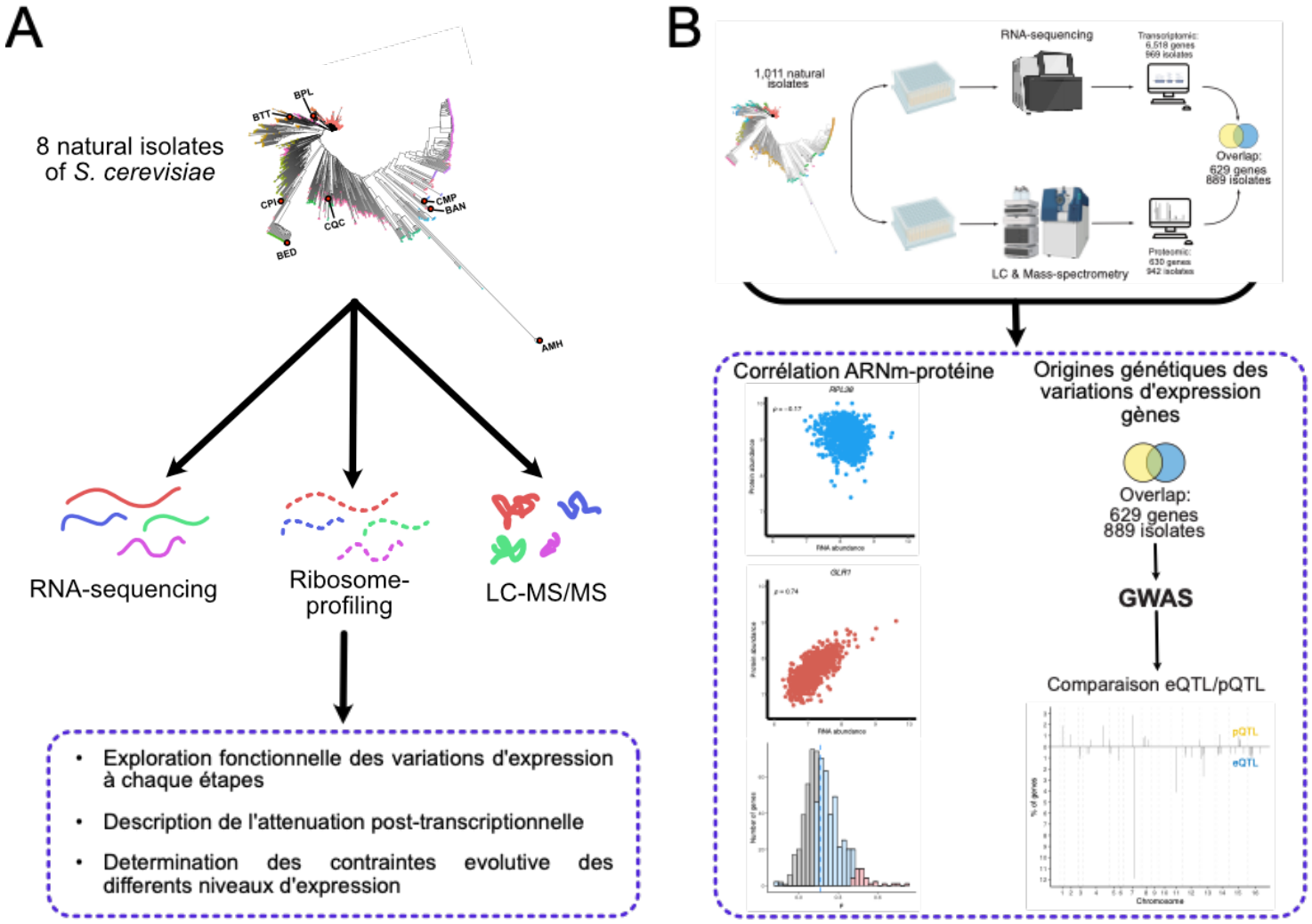


Figure 1 : (A) À l'aide de 8 souches naturelles de *S. cerevisiae*, nous avons quantifié l'abondance d'ARNm et de protéines et la traduction afin d'élucider comment l'expression de gènes varie à chaque étape. Ceci nous a permis de caractériser avec précision le phénomène d'atténuation post-transcriptionnelle et les contraintes évolutives spécifiques à chaque étape du processus d'expression de gènes. (B) Nous avons combiner des jeux de donnés de transcriptomes et de protéomes obtenus à très large échelle pour comparer les variations d'abondance d'ARNm et de protéines. Nous avons observé que ces variations étaient très différentes et que cela peut être expliqué par l'importante différence entre les origines génétique des abondances des ARNm et des protéines.

Bibliographie

Aguet, F., Alasoo, K., Li, Y.I., Battle, A., Im, H.K., Montgomery, S.B., Lappalainen, T., 2023. Molecular quantitative trait loci. Nat. Rev. Methods Primer 3, 1–22. https://doi.org/10.1038/s43586-022-00188-6

Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212. https://doi.org/10.1038/nrg3891

Albert, F.W., Treusch, S., Shockley, A.H., Bloom, J.S., Kruglyak, L., 2014. Genetics of single-cell protein abundance variation in large yeast populations. Nature 506, 494–497. https://doi.org/10.1038/nature12904

Artieri, C.G., Fraser, H.B., 2014. Evolution at two levels of gene expression in yeast. Genome Res. 24, 411–421. https://doi.org/10.1101/gr.165522.113

Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., Gilad, Y., 2015. Impact of regulatory variation from RNA to protein. Science 347, 664–667. https://doi.org/10.1126/science.1260793

Blevins, W.R., Tavella, T., Moro, S.G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., Carey, L.B., Albà, M.M., 2019. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Sci. Rep. 9, 11005. https://doi.org/10.1038/s41598-019-47424-w

Buccitelli, C., Selbach, M., 2020. mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 1–15. https://doi.org/10.1038/s41576-020-0258-4

Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonniere, E., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., 2023. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. https://doi.org/10.1101/2023.05.17.541122

Corbett, A.H., 2018. Post-transcriptional regulation of gene expression and human disease. Curr. Opin. Cell Biol. 52, 96–104. https://doi.org/10.1016/j.ceb.2018.02.011

D'Angiolo, M., De Chiara, M., Yue, J.-X., Irizar, A., Stenberg, S., Persson, K., Llored, A., Barré, B., Schacherer, J., Marangoni, R., Gilson, E., Warringer, J., Liti, G., 2020. A yeast living ancestor reveals the origin of genomic introgressions. Nature 587, 420–425. https://doi.org/10.1038/s41586-020-2889-1

Dehghan, A., 2018. Genome-Wide Association Studies, in: Evangelou, E. (Ed.), Genetic Epidemiology: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 37–49. https://doi.org/10.1007/978-1-4939-7868-7_4

Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., Ralser, M., 2020. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods 17, 41–44. https://doi.org/10.1038/s41592-019-0638-x

Fortelny, N., Overall, C.M., Pavlidis, P., Freue, G.V.C., 2017. Can we predict protein from mRNA levels? Nature 547, E19–E20. https://doi.org/10.1038/nature22293

Ingolia, N.T., Hussmann, J.A., Weissman, J.S., 2019. Ribosome Profiling: Global Views of Translation. Cold Spring Harb. Perspect. Biol. 11. https://doi.org/10.1101/cshperspect.a032698

Jiang, L., Wang, M., Lin, S., Jian, R., Li, Xiao, Chan, J., Dong, G., Fang, H., Robinson, A.E., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., MacArthur, D.G., Meier, S.R., Nedzel, J.L., Nguyen, D.Y., Segrè, A.V., Todres, E., Balliu, B., Barbeira, A.N., Battle, A., Bonazzola, R., Brown, A., Brown, C.D., Castel, S.E., Conrad, D., Cotter, D.J., Cox, N., Das, S., Goede, O.M. de, Dermitzakis, E.T., Engelhardt, B.E., Eskin, E., Eulalio, T.Y., Ferraro, N.M., Flynn, E., Fresard, L., Gamazon, E.R., Garrido-Martín, D., Gay, N.R.,

Guigó, R., Hamel, A.R., He, Y., Hoffman, P.J., Hormozdiari, F., Hou, L., Im, H.K., Jo, B., Kasela, S., Kellis, M., Kim-Hellmuth, S., Kwong, A., Lappalainen, T., Li, Xin, Liang, Y., Mangul, S., Mohammadi, P., Montgomery, S.B., Muñoz-Aguirre, M., Nachun, D.C., Nobel, A.B., Oliva, M., Park, YoSon, Park, Yongjin, Parsana, P., Reverter, F., Rouhana, J.M., Sabatti, C., Saha, A., Skol, A.D., Stephens, M., Stranger, B.E., Strober, B.J., Teran, N.A., Viñuela, A., Wang, G., Wen, X., Wright, F., Wucher, V., Zou, Y., Ferreira, P.G., Li, G., Melé, M., Yeger-Lotem, E., Barcus, M.E., Bradbury, D., Krubit, T., McLean, J.A., Qi, L., Robinson, K., Roche, N.V., Smith, A.M., Sobin, L., Tabor, D.E., Undale, A., Bridge, J., Brigham, L.E., Foster, B.A., Gillard, B.M., Hasz, R., Hunter, M., Johns, C., Johnson, M., Karasik, E., Kopen, G., Leinweber, W.F., McDonald, A., Moser, M.T., Myer, K., Ramsey, K.D., Roe, B., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Jewell, S.D., Rohrer, D.C., Valley, D.R., Davis, D.A., Mash, D.C., Branton, P.A., Barker, L.K., Gardiner, H.M., Mosavel, M., Siminoff, L.A., Flicek, P., Haeussler, M., Juettemann, T., Kent, W.J., Lee, C.M., Powell, C.C., Rosenbloom, K.R., Ruffier, M., Sheppard, D., Taylor, K., Trevanion, S.J., Zerbino, D.R., Abell, N.S., Akey, J., Chen, L., Demanelis, K., Doherty, J.A., Feinberg, A.P., Hansen, K.D., Hickey, P.F., Jasmine, F., Kaul, R., Kibriya, M.G., Li, J.B., Li, Q., Linder, S.E., Pierce, B.L., Rizzardi, L.F., Smith, K.S., Stamatoyannopoulos, J., Tang, H., Carithers, L.J., Guan, P., Koester, S.E., Little, A.R., Moore, H.M., Nierras, C.R., Rao, A.K., Vaught, J.B., Volpi, S., Snyder, M.P., 2020. A Quantitative Proteome Map of the Human Body. Cell 183, 269-283.e19. https://doi.org/10.1016/j.cell.2020.08.036

Lahue, C., Madden, A., Dunn, R., Smukowski Heil, C., 2020. History and Domestication of Saccharomyces cerevisiae in Bread Baking. Front. Genet. 11.

Lee, T.I., Young, R.A., 2013. Transcriptional Regulation and Its Misregulation in Disease. Cell 152, 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Liu, Y., Beyer, A., Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550. https://doi.org/10.1016/j.cell.2016.03.014

Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E., Dequin, S., Galeote, V., 2015. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Mol. Biol. Evol. 32, 1695–1707. https://doi.org/10.1093/molbev/msv057

McManus, C.J., May, G.E., Spealman, P., Shteyman, A., 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422–430. https://doi.org/10.1101/gr.164996.113

Messner, C.B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleder, M., Ralser, M., 2023. Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. PROTEOMICS 23, 2200013. https://doi.org/10.1002/pmic.202200013

Messner, C.B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleder, M., Ralser, M., 2022. Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. PROTEOMICS n/a, 2200013. https://doi.org/10.1002/pmic.202200013

Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L., Wincker, P., Casaregola, S., Dequin, S., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc. Natl. Acad. Sci. 106, 16333–16338. https://doi.org/10.1073/pnas.0904673106

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G.,

Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Stark, R., Grzelak, M., Hadfield, J., 2019. RNA sequencing: the teenage years. Nat. Rev. Genet. 20, 631–656. https://doi.org/10.1038/s41576-019-0150-2

Tak, Y.G., Farnham, P.J., 2015. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin 8, 57. https://doi.org/10.1186/s13072-015-0050-4

Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M.E., Peters, A.H.F.M., Anders, S., Kaessmann, H., 2020. Transcriptome and translatome co-evolution in mammals. Nature 588, 642–647. https://doi.org/10.1038/s41586-020-2899-z

Table of Contents

STATE OF THE ART

The genotype-phenotype relationship: it's complicated…

An exceptional phenotypic diversity can be observed within all species. On both macroscopic and molecular scales, individuals differ in a myriad of observable and quantifiable traits. The origin of this diversity has long been questioned and understanding the mechanisms underlying the phenotypic landscape observed in a population has been a central and long-standing challenge in biology. In fact, as humanity moved from nomadism to sedentism, relying on both agriculture and livestock, it became crucial to improve the traits of interest in domesticated plants and animals. Thus, the control of phenotypic diversity quickly became a central concern.

A complex relationship for complex traits

A long-standing interest

Throughout the history of science, the relationship between heredity and traits has been questioned. Traces of primitive explorations of pedigree have been observed on a 6000 year old Babylonian tablet, which may describe horse breeding, suggesting early hypotheses on the hereditary nature of livestock characteristics (Coonen, 1952). Later, during the Ancient Greek period (400-300 BCE), physicians and philosophers also established early theories of trait heredity and reproduction (Bazopoulou-Kyrkanidou, 1992). However, in the late 19th century, Mendel established the first law of heredity (Mendel, 1866) and, with the rediscovery of his work in the early 20th century by De Vries, Correns and Tschermak, genetics emerged as one of the major disciplines in the life sciences. At the same time, several fundamental concepts in genetics were proposed and established by Bateson and Johannsen, such as "*gene*", "*genotype*" and "*phenotype*" (Bateson et al., 1909; Johannsen, 1911). These terms respectively described the "*unit factor*" of heredity, the "*sum of all the genes*" and the "*direct inspection […] or direct measures of assessment*" (Johannsen, 1911). Johannsen supported the theory of a direct relationship between genotype and phenotype. The highly valuable implications of a precise dissection of the genotype-phenotype relationship in medicine, agriculture, food production, or industry quickly led to intensive research efforts. At the same time, several fundamental discoveries in the 20th century led to a clearer view of the genetic mechanisms underlying heredity. One example is the work of Morgan, Sturtevant, Muller and Bridges on the fruit fly, which combined Mendelian and chromosomal theories (Morgan et al., 1923). This work was a real keystone at the time, on which several other major discoveries were based, such as the

concept of mutation (Muller, 1928). Later, in the mid-20th century, the rise of molecular biology and biochemistry also revolutionized genetics. For example, Beadle and Tatum demonstrated that genes are involved in biochemical reactions (Beadle and Tatum, 1941) through the action of an enzyme. In addition, Rosalind Franklin's exploration of the structure of the deoxyribonucleic acid (DNA) molecule and the results published by Crick and Watson in the 1950s (Watson and Crick, 1953) clearly established the role of DNA as being the vector of heredity. At the end of this period, a major discovery introduced the concept of gene expression by messenger ribonucleic acid (mRNA) molecules and the regulation of gene expression (Jacob and Monod, 1961). From this point on, the relationship between the amazing diversity observed in natural populations and its biological origin became clearer. Genetic variations in each individual, through the process of gene expression (transcription and translation), affect the function or quantity of a given protein, which ultimately affects one or more traits of the individual. This link between genotype and phenotype is commonly referred to as the genotype-phenotype relationship and is still the subject of considerable research in modern biological science.

The decomposition of a phenotype

Despite its seemingly simple nature, the genotype-phenotype relationship is a truly complex and subtle process. First, the number of genes influencing a trait can range from 1 to several thousand. The distribution of the phenotypes within a population according is usually a good approach to determine if the trait is control by one or several genes (Figure 1A, B). If the population exhibits a phenotypic bimodal distribution (Figure 1A), this usually indicates that the phenotype is controlled by a single gene. In humans, about 6,000 thousand diseases (Condò, 2022), such as cystic fibrosis, neurofibromatosis or Duchenne muscular dystrophy, are caused by a single defective gene. Several non-pathological characteristics are also determined by a single gene, such as the ABO blood group (Yamamoto et al., 1990). Conversely, and for most quantifiable traits, the phenotypic distribution within a population follows a normal distribution (Figure 1B) These traits are considered as complex. Autism, Alzheimer's disease, or human height (Akiyama et al., 2019; Nikolac Perkovic and Pivac, 2019; Ramaswami and Geschwind, 2018) are notable examples of complex traits for which hundreds or even thousands of genetic variants have been found to influence the appearance or intensity of the trait. Deciphering and capturing all the genetic factors involved in such traits is a tedious, but crucial, task for predicting and treating complex diseases. It is worth noting that both monogenic and complex

traits are, of course, not solely due to genetic factors. An individual's environmental background can influence a trait as much, if not more, than their genetic background. In the case of autism, for example, several external factors such as zinc deficiency, prenatal and perinatal stress, or parental age are known to be associated with the onset of autism spectrum disorders (Grabrucker, 2013). Another and perhaps more famous example of the impact of environmental factors on a phenotype is the association between lung cancer and smoking (Hecht, 2006).

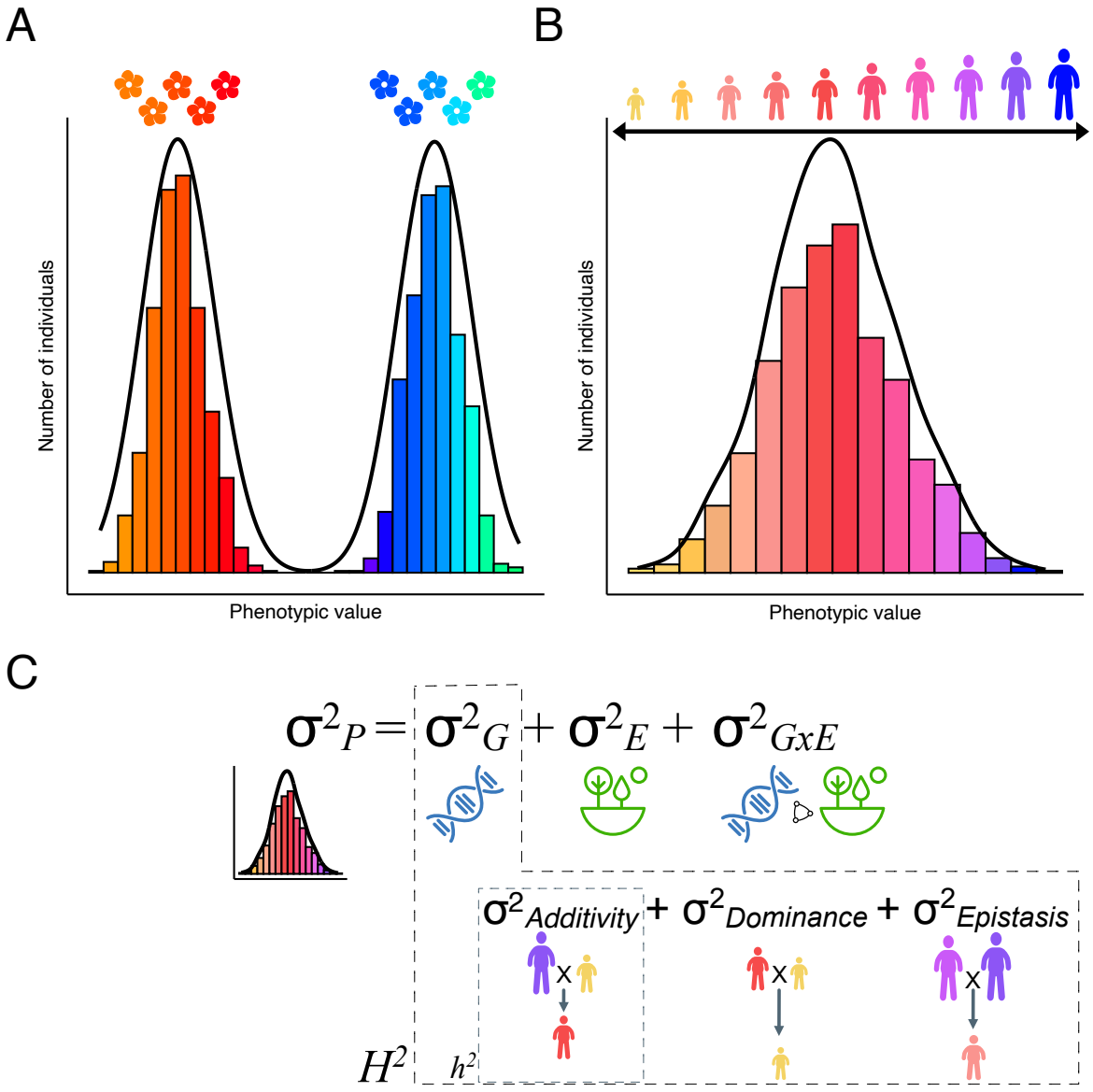


Figure 1: The origin of the phenotypic landscape.

Example of a (A) bimodal or (B) normal distribution of a trait in a population. While traits that follow bimodal distributions are often related to a single gene, the complex or polygenic traits follow normal distributions. (C) Dissection of the phenotypic variation origins of a trait

The phenotypic variance of a trait (σ^2_P) is therefore considered as equal to the sum of the genetic variance (σ^2_G), the environmental variance (σ^2_E) as well as the genetic and environmental interaction variance (σ^2_{GxE}) (Figure 1C). The genetic variance itself is a highly complex composition of additive and non-additive effects (Figure 1C). The fraction of phenotypic variance controlled by the genetic variance is called broad-sense heritability (H^2) while the narrow-sense heritability (h^2) includes only the fraction of the phenotypic variance associated with additive genetic effects (Figure 1C). The non-additive effects are related to phenomena such as epistasis and dominance (Su et al., 2012). Briefly, epistasis describes the interaction between two or more loci that causes a phenotype different from that expected from an additive effect, while dominance describes the situation where one allele at a heterozygous locus masks the phenotype attributed to the alternative allele. Recent estimates in the budding yeast *Saccharomyces cerevisiae* showed that trait variation is mostly due to additive effects, about 55% of phenotypic variation, while non-additive effects account for 29% of phenotypic variation (Bloom et al., 2015, 2013; Fournier et al., 2019).

From molecular changes to macroscopic traits

The relationship between a specific DNA sequence modification and a phenotypic variation depends on multiple mechanisms that affect individuals at several scales. For example, a mutation will influence molecular reactions that affect one or more cellular processes, resulting in tissue, organ, or overall macroscopic trait variation. Studying the molecular mechanisms underlying phenotypic changes is therefore a critical step to fully understand the genotype-phenotype relationship.

Genetic diversity

Identifying and characterizing the genetic variants that cause a change in phenotype is the first step in understanding the molecular processes involved. The simplest and most common type of DNA variant is the modification of a single base (Figure 2A), often called a Single Nucleotide Polymorphism (SNP). Extensive exploration of the genomes of more than 2,500 human individuals has for example resulted in the discovery of more than 88 million SNPs (The 1000 Genomes Project Consortium, 2015). SNPs are often used to describe the genetic differences within species. In fact, the number of SNPs in each individual (compared to a reference

genome), and its integration at a population-scale level is a good, although incomplete, indicator of the genetic diversity within this population. In humans, each individual carries on average 4.5 million SNPs, which corresponds to one SNP every 1,000 bp. In microorganisms, intraspecific genetic diversity can be much higher. For example, in the budding yeast *S. cerevisiae*, an average of 1 SNP per 200 bp has been reported between individuals from a natural population (Peter et al., 2018). The second most common type of genetic variant is insertion or deletion of a few bases, commonly referred to as "indels" (Figure 2A). More than 3.6 million indels have been identified in the 2,500 human genomes (The 1000 Genomes Project Consortium, 2015). Both SNPs and indels can be detected efficiently using short-read sequencing techniques such as Illumina sequencing.

Structural variants (SVs) comprise diverse type of genetic changes that are much larger (at least >50 base pairs) than SNPs and indels. SVs include large chromosomal alterations such as deletions or insertions, translocations, aneuploidies, copy number variants (CNV), inversions and duplications (Figure 2B). The SVs are usually much more difficult to detect than SNPs or indels. Although the detection of these variants is theoretically possible using classical short-read sequencing methods (such as Illumina paired-end sequencing), their detection is more reliable using specific sequencing techniques, such as long-read sequencing (Logsdon et al., 2020; Shi et al., 2016; Wenger et al., 2019). Although they are less common, about 60,000 SVs have been detected in the 2,500 human genomes, they affect a larger number of bases compared to SNPs and indels. On average, an individual will differ from the human reference genome at about 4.5 million genetic positions due to SNPs and indels, while the individual's SVs will affect an average of 20 million bases (The 1000 Genomes Project Consortium, 2015). It is important to note that until recently, the vast majority of large-scale genomic explorations were based on short-read sequencing methods, resulting in poor exploration and characterization of SVs across individuals. Today, the development and democratization of long-read sequencing technologies, such as Oxford Nanopore Sequencing or PacBio technologies, are filling this gap (Audano et al., 2019; He et al., 2023; Zhou et al., 2022). Overall, the nature of a variant will, of course, be a major determinant of its molecular effects, as well as several other characteristics such as its genetic location, its homo- or heterozygous state, or its size.

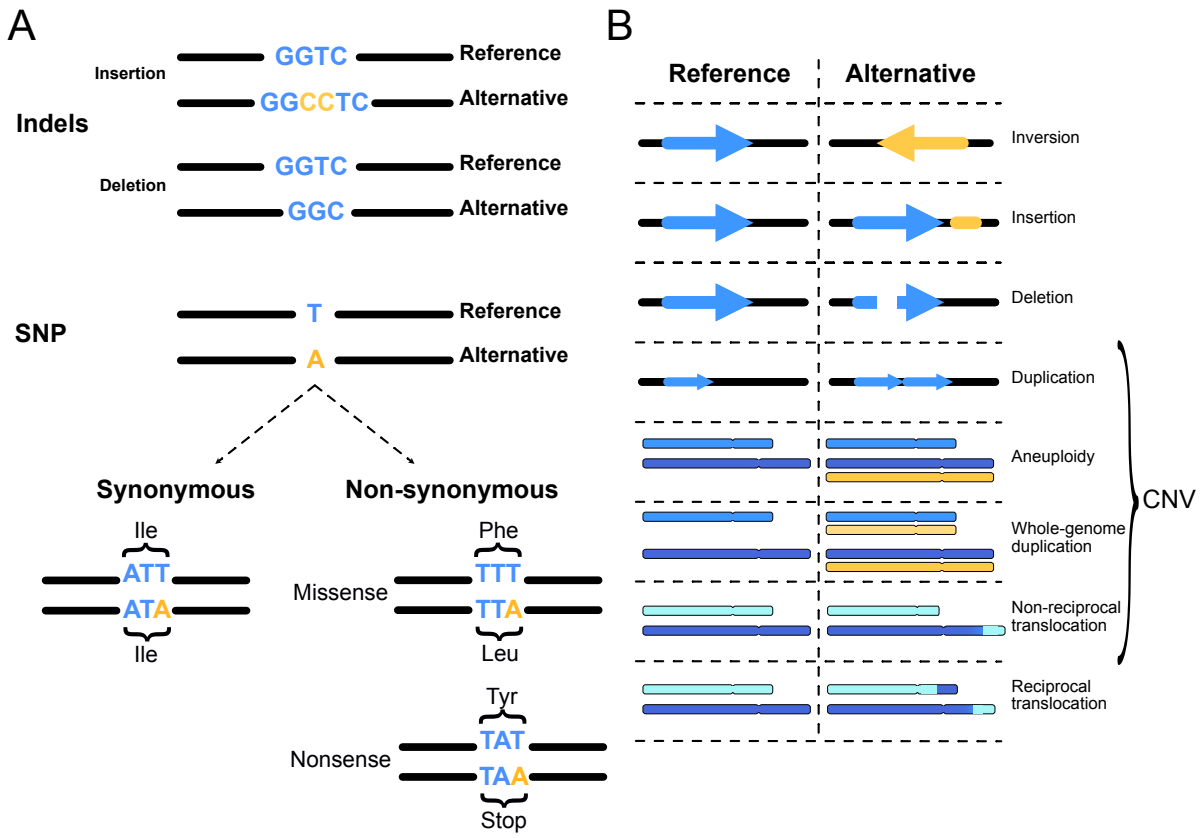


Figure 2: The different types of genetic variants.

(A) Example of indels and SNPs, with the consequences that can be induced by a SNP on coding sequences. (B) Schematic of structural variants (>50 bp), some of which can induce copy number variations (CNVs).

The frequency at which a genetic variant occurs in a population, often measured as minor allele frequency (*i.e.*, the percentage of individuals carrying the 2nd most common allele), is often low (Figure 3). Overall, the vast majority of variants present in an individual are rare (Peter et al., 2018; The 1000 Genomes Project Consortium, 2015). Interestingly, rare variants are thought to play an important role in complex traits, even if detecting their actual contribution to the phenotype can be laborious (Bodmer and Bonilla, 2008; Cirulli and Goldstein, 2010; Gibson, 2012; The UK10K Consortium, 2015). In yeast, for example, rare variants have been shown to play a major role in several growth phenotypes (Bloom et al., 2019; Fournier et al., 2019) and deleterious variants tend to be enriched among the rare variants (Figure 3) (Peter et al., 2018).

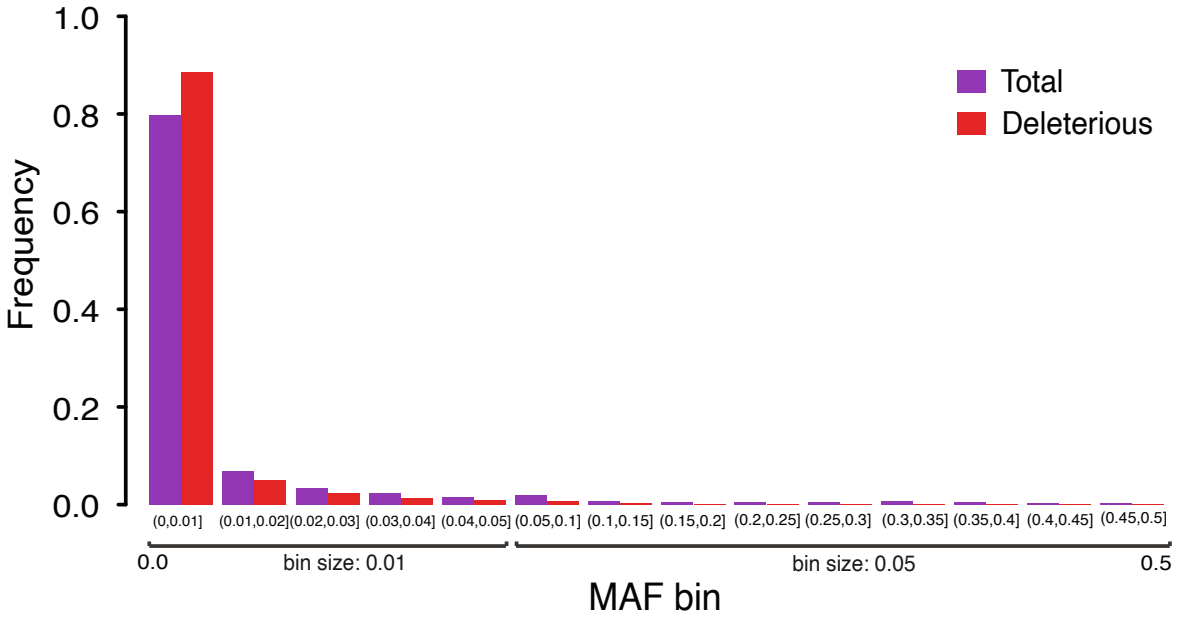


Figure 3: Minor allele frequency distribution of the SNPs in a natural population of 1,011 *S. cerevisiae* isolates.

The rare variants (MAF<0.05) account for 92% of the SNPs in a natural population of *S. cerevisiae*. Among them, deleterious variants tend to be enriched (adapted from Peter et al., 2018).

DNA and molecular changes for macroscopic consequences

The molecular changes induced by genetic variants are diverse and include a wide range of phenotypes. For example, a change in the amino acid composition of a protein or the apparition of a premature stop codon is a possible consequence of non-synonymous substitutions that can be induced by SNPs (Figure 2A). Even though non-synonymous SNPs are less common than synonymous SNPs, they are the most frequently associated with human diseases (Yip et al., 2008). Such a change is obviously a strong modifier of several protein properties. The enzymatic function itself can be affected as it is the case for the human gene *ALDH1L1*, a tumor suppressor gene involved in folate metabolism (Krupenko and Horita, 2019) whose catalytic activity is reduced in individuals with a specific SNP causing an amino acid change (Frosst et al., 1995) leading to an increased risk of cardiovascular disease, birth defects, and cancer. Protein stability is also a known molecular property that can be altered by SNPs or indels (Bromberg and Rost, 2009; Casadio et al., 2011). Severe diseases such as amyotrophic lateral sclerosis (Lindberg et al., 2005; Ling et al., 2010) are associated with an increased or decreased protein stability (Randles et al., 2006; Wang and Moult, 2001). Finally, modification of protein

amino acid sequence can also affect the protein-protein interaction ability of a particular protein (Cheng et al., 2021; Porta-Pardo et al., 2015). Sickle-cell disease, an autosomal recessive pathology, is a famous case of protein interaction modification where a unique SNP that is located in the coding sequence of the β-globin gene (Ingram, 1957; Rees et al., 2010; Sundd et al., 2019) leads to an abnormal protein aggregation into large polymers which ultimately causes an abnormal red cell shape (Bunn, 1997). In addition, it is worth noting that synonymous SNP (*i.e.*, a nucleotide substitution that does not change the final amino acid sequence of a protein, figure 2A), despite their apparently neutral effect, are also known to be associated with human disease (Sauna and Kimchi-Sarfaty, 2013). An elegant example of this is in cystic fibrosis, where a synonymous SNP associated with the disease alters the mRNA structure of the *CFTR* gene, resulting in misfolding of the cognate protein and its dysfunction (Bartoszewski et al., 2010). Although less studied than SNPs, SVs can also induce specific molecular changes. A very recent example of this is the case of rust resistance in wheat, which has been linked to a translocation linked to an introgression event (the transfer of genetic material from one species to another through hybridization), and results in a fused kinase that drives resistance to the pathology (Wang et al., 2023).

The case of missing heritability

In the recent decades, many studies have attempted to unambiguously identify the genetic variants that influence the onset or intensity of human diseases or phenotypes by testing the association of each variant (usually SNPs) with the phenotypic values quantified in a population. This is usually done using analytical tools such as genome-wide association studies or linkage mapping studies, which will be described later. This has become particularly relevant with the development of short-read sequencing techniques (*e.g.*, Illumina sequencing) since 2005, allowing for more and more large-scale variant characterizations. For example, the influence of genetic factors on human height has been extensively studied (Wood et al., 2014; Yang et al., 2015, 2010; Yengo et al., 2018) and the most recent large-scale study using data from 5.4 million individuals identified more than 12,000 SNPs significantly associated with height (Yengo et al., 2022). However, the fraction of phenotypic variation explained by this large number of genetic variants is at most 40% (depending on the ancestry of the individuals), suggesting that the dissection of the origin of height variation is far from complete. Interestingly, for many complex human traits, examination of the genetic fraction of phenotypic variation fails to explain most of the trait variation. For example, in the case of familial

colorectal cancer, less than half of the heritability of this disease is associated with clearly identified genetic variants (Schubert et al., 2020). Another example is the case of Alzheimer's disease, where the various identified genetic variants account for only 7.78% of the phenotypic variance (Ridge et al., 2013), while the estimated heritability of this disease reaches 58% (Gatz et al., 2006). This phenomenon is better known as the “missing heritability” and has been widely discussed and investigated (Génin, 2020; Maher, 2008; Manolio et al., 2009; Owen and Williams, 2021; Young, 2019). Several hypotheses have been formulated to explain the gap between the total estimated heritability and the proportion experimentally associated with genetic variants. Rare genetic variants, which are poorly considered when studying the relationship between a trait and its genetic origin, epigenetic factors as well as genetic interactions could explain part of this missing heritability. Moreover, since SVs have so far been poorly characterized across individuals, their effects on phenotypes have been less studied, which may also explain some of the missing heritability. This large gray area is still far from being resolved, and it highlights how complex and still poorly understood the genotype-phenotype relationship is.

Gene expression as a driver of the genotype-phenotype relationship

Interestingly, despite the strong influence of SNPs in coding regions at the molecular level, the majority of SNPs detected as influencing a specific trait in humans are located in non-coding or intronic regions (Aguet et al., 2023; Tak and Farnham, 2015). This suggests that the effect of these non-coding SNPs is more likely to be on regulatory processes, causing changes in gene expression, which ultimately induce cellular and macroscopic phenotypic changes.

All molecular steps of gene expression can be affected by a genetic variant. The initial accessibility of a gene to the transcriptional machinery is strongly influenced by nearby or distal DNA modifications. Chromatin organization, for example, is a tightly regulated process that is highly sensitive to SNPs or indels (Degner et al., 2012; Delaneau et al., 2019). Similarly, DNA methylation alteration, which has long been associated with cancer (Das and Singal, 2004; Koch et al., 2018), is also a known regulatory step that is tightly regulated and influenced by numerous genetic variants (Hawe et al., 2022; Min et al., 2021). Overall, several diseases are linked to transcriptional modification, as several types of cancer are known to be associated with aberrant transcriptional regulation, such as prostate cancer (Demichelis et al., 2012), melanoma (Huang et al., 2013) or leukemia (Sanda et al., 2012). Dysregulation of mRNA degradation is also a

known factor influencing human disease (Saramago et al., 2019). Furthermore, targeting specific mRNA degradation pathways is a promising avenue for the development of anticancer therapies (Bokhari et al., 2018; Huang et al., 2018; Lindeboom et al., 2019). The molecular effects of genetic variants may also involve post-transcriptional mechanisms. For example, RNA splicing has been shown to be an important link between DNA variation and disease (Y. I. Li et al., 2016). Finally, alterations in translational regulation itself can be a major source of phenotypic variation. A good example in humans is the Fragile X Mental Retardation Protein (FMRP), encoded by the *FRM1* gene (Verkerk et al., 1991). This protein normally regulates and represses translation through various mechanisms (Li et al., 2001; Richter et al., 2015). When FMRP is not expressed, this leads to a global and abnormal translation of many different mRNA (Udagawa et al., 2013), resulting in the Fragile X syndrome, a common inherited form of intellectual disability (Corbett, 2018).

Therefore, because of the central role of gene expression in the genotype-phenotype relationship, this link between the genetic origins of a trait and its establishment is inherently tedious to dissect and understand. The complete expression process of each gene (i.e., transcription, translation, transcript or protein maturation, and degradation) is the combination of tenths of tightly regulated mechanisms, with a plethora of interactions and retro-controls between each step (Figure 4). The multifaceted nature of the genotype-phenotype relationship implies that the consequences of a genetic variant, and thus the biological mechanism underlying a particular phenotype, may affect any of the gene expression steps.

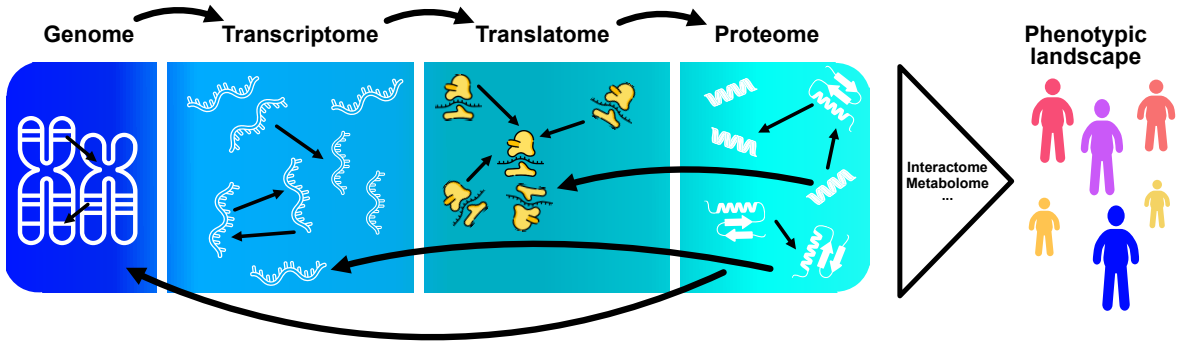


Figure 4: Gene expression is a complex process that underlies the genotype-phenotype relationship.

Gene expression is a tightly regulated process in which each step interacts with the others, resulting in a truly complex regulatory network (adapted from Buccitelli and Selbach, 2020).

Determining the role of gene expression in the genotype-phenotype relationship

As described above, alteration of gene expression is one of the main mechanisms that translates genotypes into phenotypes. The accurate study of gene expression at any level (transcriptome, proteome...) is a constant challenge for modern biology, as is the development of bioinformatic tools to link gene expression and genetic variants. Combined, these two aspects of gene expression studies are crucial to dissect the impact of changes in transcript and protein abundance on the phenotypic landscape of natural populations.

Quantification of gene expression

Together with sequencing techniques, the precise quantification of gene expression at each step of the expression process has been a keystone of modern biology, leading to considerable advances in medicine, agriculture, and biotechnology. Over the last two decades in particular, several tools have allowed a steady increase in the precision and scale of quantification of either mRNA or protein abundance.

Quantification of mRNA

Methods for quantifying mRNA molecules in an individual have been developed since the early days of molecular biology. Northern blotting was developed in the late 1970s and at the time was a robust technique for relative quantification of mRNAs of interest (Reue, 1998; Sambrook and Russell, 2001). The principle was based on the transfer of fractionated and separated (by denaturing gel electrophoresis) mRNAs onto a membrane. The mRNA of interest was then revealed on the membrane by a hybridization step using a cDNA probe, either radiolabeled or linked to an enzyme, and relatively quantified by comparing the label intensity to a control label. Also based on hybridization, the ribonuclease protection assay (RPA) involved a liquid mixture between the total mRNAs and a specific probe (Azrolan and Breslow, 1990). Once hybridization is achieved, single-stranded RNAs are degraded, leaving only the hybrid, which is then electrophoresed on a denaturing gel, allowing relative or even absolute quantification of mRNA using titration reactions (Reue, 1998). However, the precise quantification of mRNA in an absolute manner has been improved and achieved mainly by polymerase chain reaction (PCR) techniques, in particular the combination of reverse transcriptase (RT) and PCR, known

as quantitative RT-PCR (Bustin, 2000; Foley et al., 1993). Theoretically, RT-PCR can quantify a single RNA molecule in a sample. However, the practical limit is closer to ten molecules due to RT inefficiency (Reue, 1998). Absolute quantification using quantitative RT-PCR was developed in the late 1980s by adding an exogenous transcript standard (Becker-André and Hahlbrock, 1989; Gilliland et al., 1990; Wang et al., 1989). Despite the significant advances in mRNA quantification achieved with these techniques, one of their major drawbacks was their limited suitability for probe multiplexing. In other words, the simultaneous quantification of different mRNA was severely limited. This limitation was overcome with the development of DNA microarrays in the early 1990s (Bumgarner, 2013). Briefly, the principle is based on the detection of hybridization of DNA fragments on a surface containing probes corresponding to genic regions in the case of mRNA quantification. The hybridization is detected and quantified by the prior labeling of DNA fragments extracted from the sample or obtained after reverse transcriptase of mRNA. This method allows relative mRNA quantification in a much higher throughput than previous techniques and has therefore been widely used to analyze gene expression (Schena et al., 1995; Tarca et al., 2006).

However, the development and the increased accessibility of RNA-sequencing (or RNA-seq, figure 5) from the late 2000s (Emrich et al., 2007; Lister et al., 2008; Nagalakshmi et al., 2008) has led to a decline of the use of DNA microarrays. RNA-seq, that is originally based on short-read sequencing (mainly Illumina sequencing), allows a theoretical absolute quantification of the cell's transcripts, giving a global view of all mRNAs from a sample, a tissue or an individual (such quantifications are called transcriptome) and is therefore less biased than microarray methods (Stark et al., 2019). In addition, where microarrays required prior knowledge of the genome of the species being studied to construct probes, RNA-seq has no such requirement and can be performed on all species. The classical workflow (Figure 5) is based on an RNA extraction step, followed by a ribosomal RNA depletion or mRNA enrichment, followed by cDNA synthesis and short-read sequencing after a final sequencing adapter ligation (Hrdlickova et al., 2017). The proper quantification of the expression of each gene then relies on bioinformatics pipelines (Corchete et al., 2020). Classically, the sequencing reads are filtered and aligned, with low quality or multi-mapped reads typically discarded after these steps. The expression quantification of each gene is then performed. The resulting data set is typically a raw count of the number of reads aligned to each gene. To formally compare mRNA abundance across datasets, the raw counts are transformed and normalized using, for example, the transcripts per million (TPM) unit (Conesa et al., 2016). The RNA-seq technique has been

widely used to study various gene expression processes, such as RNA splicing (Wang et al., 2008) or RNA-mediated gene regulation (W. Li et al., 2016; Morris and Mattick, 2014). Due to its versatility and its accessibility, RNA-seq has been an important tool to understand or diagnose gene expression alterations in various pathologies (Byron et al., 2016; Doebele et al., 2015; Hong et al., 2020; Wirka et al., 2018) and to explore the transcriptome at the population level (Caudal et al., 2023; The GTEx Consortium, 2015). However, even though RNA-seq has considerably advanced the study of RNA abundance, the fact that this strategy was until recently solely based on short-read sequencing has made it to suffer from several limitations. For example, in the case of mRNA isoforms, short reads sequencing prevents from accurate individual quantification (Djebali et al., 2012; Stark et al., 2019). In organisms where transcripts can be very long and variable, such as humans, where more than half of the transcripts are longer than 2,500 bp (Frankish et al., 2019), this issue is highly relevant and specific methods have emerged as powerful alternatives to short-read-based RNA-seq to account for mRNA isoforms.

Indeed, the development of new techniques to sequence long fragments of DNA (also known as long-read sequencing) is leading to new information on mRNA abundance (Figure 5). These methods, namely Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) (Rhoads and Au, 2015; Wang et al., 2021) allow the capture of the entire mRNA molecule, mainly through cDNA sequencing. They are particularly efficient for mRNA isoform detection and *de novo* transcriptome analysis (Stark et al., 2019). In addition, ONT can also be used to sequence native mRNA molecules, which allows for more precise exploration of mRNA base modifications. However, the throughput of these long-read sequencing methods is still limited.

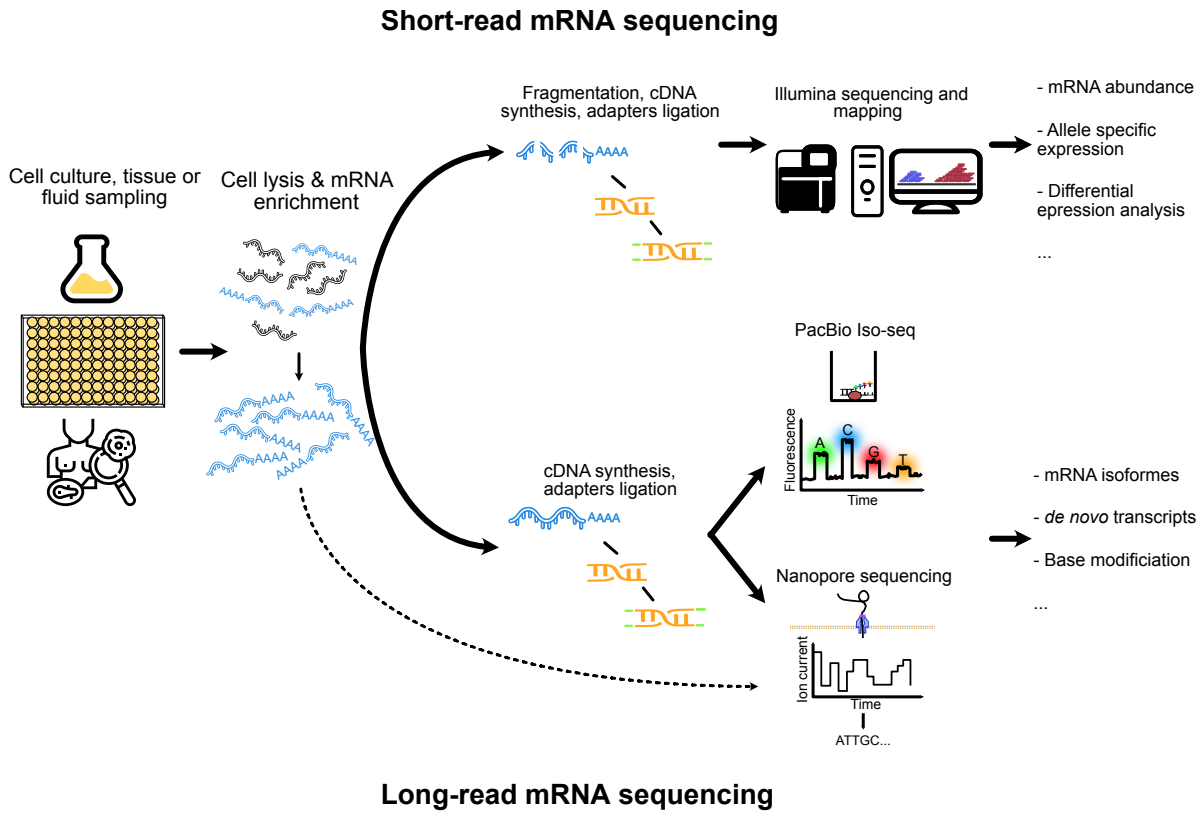


Figure 5: Short- and long-read mRNA sequencing

Short-read RNA sequencing is based on the extraction and isolation/enrichment of mRNA, which is then fragmented, ligated to adapters, and sequenced with Illumina. The reads are typically mapped to the genome, and the number of reads per gene is normalized, *e.g.*, using transcript per million (TPM) normalization. Long-read sequencing is based on either PacBio or Nanopore technologies. Both methods are capable of sequencing full-length mRNA and are therefore very useful for detecting isoform changes in transcripts, *de novo* transcripts or base modification.

Quantification of protein

Similar to mRNA, protein abundance quantification has constantly evolved since its inception in the second half of the 20th century. Although some quantification methods are similar to what can be done for mRNA abundance, such as Western blotting (Burnette, 1981; Towbin et al., 1979), most of the quantification techniques are specific to protein abundance due to the very different chemical nature of peptide chains compared to RNA. Early protein quantification methods were based on spectrophotometric measurements, such as the Bradford or Lowry protein assay (Bradford, 1976; Lowry et al., 1951). However, these methods are not compatible for global quantification of all the different proteins in a sample. One of the first methods to

allow for the quantification of the total set of proteins expressed in a cell, a tissue, or whole organism (such sets of protein are called proteomes) was based on 2-dimensional electrophoresis gels (Magdeldin et al., 2014; O'Farrell, 1975). The principle is simple: all proteins in a sample are extracted and successively separated by two properties on the two dimensions of a gel (usually a polyacrylamide gel). The first dimension of the gel resolves protein molecules according to their isoelectric point (using a pH gradient), while the second dimension resolves them according to their molecular weight. The resulting gel consists of several separated dots that can be excised and quantified using, for example, a coupled mass spectrometry device. This technique has been repeatedly used to quantify the global proteome of many organisms, such as bacteria (Wasinger et al., 1995), yeast (Gygi et al., 1999) or humans (Friedman et al., 2004).

Since the 2000s, however, the evolution of both molecular biology tools and mass spectrometry methods has led to more precise and comprehensive proteome acquisitions. For example, tag-based protein quantifications have been widely used in the last decade. These techniques rely on the construction of libraries in which each protein is individually fused with a tag, such as green fluorescent protein (GFP) or a high-affinity epitope. The tagged proteins are quantified by immuno- or photo-detection. In yeast, several studies have quantified the proteome using either GFP microscopy (Breker et al., 2013; Chong et al., 2015; Dénervaud et al., 2013; Mazumder et al., 2013; Tkach et al., 2012; Yofe et al., 2016), GFP flow cytometry (Davidson et al., 2011; Lee et al., 2007; Newman et al., 2006) or tag immunodetection (Ghaemmaghami et al., 2003). However, these methods suffer from major shortcomings when it comes to large-scale or unbiased quantification. First, library construction can be tedious, and the quantification itself requires each protein to be measured independently. Second, the fusion of a tag to a protein is obviously not biologically neutral, and this can lead to misfunction or dysregulation of abundance. Therefore, other methods related to mass spectrometry techniques have become prominent tool to explore and precisely resolve proteomes in an almost exhaustive manner. For example, liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is particularly suited to multiplexed sample analysis, making it a perfect tool for proteome exploration. The method is generally based on double analysis of peptides (obtained after trypsin digestion and separated by LC) using two coupled mass spectrometers. Quantification and identification of each protein is performed by analyzing the mass-to-charge (m/z) ratio of the peptide fragments on the two spectrometers. Over the past decade, several technological advances have improved the reliability of mass spectrometry proteomics. One

notable example is the rise of data-independent acquisition (DIA) (Chapman et al., 2014; Gillet et al., 2016), a method that allows for a broader range of protein quantification (Li et al., 2021). Several software packages, such as DIA-NN, are specifically designed for large-scale proteomics experiments (Demichev et al., 2020). More importantly, these technological advances have enabled proteome exploration at a much higher throughput and scale in the recent years (Figure 6) (Messner et al., 2023, 2022; Muenzner et al., 2022), although proteomes at population level remain largely uncharacterized, especially compared to transcriptomes at population level.

The limitations of LC-MS/MS have been studied extensively, and one of the major limitations is missing data. Missing peptides is a common and long-standing issue in LC-MS/MS studies, especially for large-scale exploration (Karpievitch et al., 2010; Muenzner et al., 2022). There are several reasons for this. For example, the abundance level of a protein is a major determinant of its detectability: low abundance peptides are often missed by LC-MS/MS, resulting in a biased quantification towards highly abundant proteins. Also, the chemical and physical characteristics of some proteins make them prone to be trapped in the LC column.

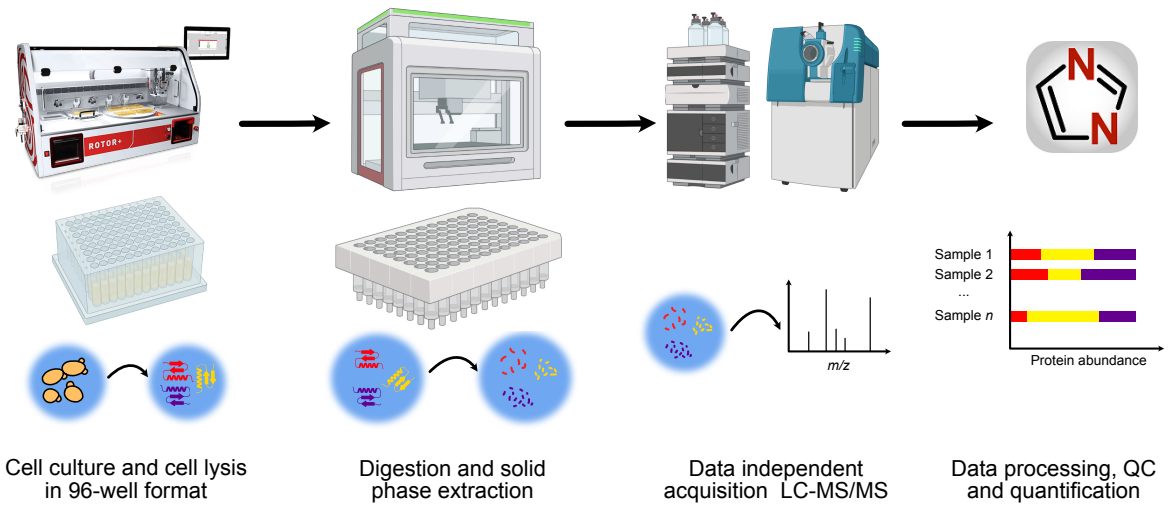


Figure 6: Recent advances in LC-MS/MS have enabled for high-throughput proteomic exploration.

Several advances have increased the throughput of proteomic experiments. Laboratory robots such as plate rotators (Step 1) or automated liquid handlers (Step 2) have greatly facilitated the preparation of proteomic samples. Also, new methods of LC-MS/MS have led to a decrease in time per sample of proteomic experiment (step 3), approximately 19min/sample (Muenzner et al., 2022). Since the resulting dataset of large-scale proteome exploration are usually computationally intensive, new

bioinformatic tools such as DIA-NN (Demichev et al., 2020) have been developed to efficiently and accurately handle and process LC-MS/MS data (step 4). Figure adapted from Muenzner et al., 2022, images from Biorender and Singer websites.

Ribosome profiling

Despite the interconnected nature of mRNA and protein abundance, transcriptomic and proteomic regulation can be very different. Therefore, understanding the quantitative relationship between the two expression levels requires a precise analysis of the translation process. While there are several techniques that allow the exploration of translation (Dermit et al., 2017), such as polysome profiling (Arava et al., 2003), a more precise and robust method was developed more than a decade ago, called ribosome profiling or ribo-seq (Figure 7A) (Ingolia et al., 2009). The principle is to extract intact polysomes from a cell and subject the cell extract to RNase digestion. The mRNA fragments that are actively being translated are protected from the RNase treatment by the translating ribosomes, resulting in a pool of ≈28 nucleotides of mRNA fragments. These fragments are then sequenced to determine which parts of the transcriptome are being translated. Usually, ribo-seq and RNA-seq experiments are performed on the same sample.

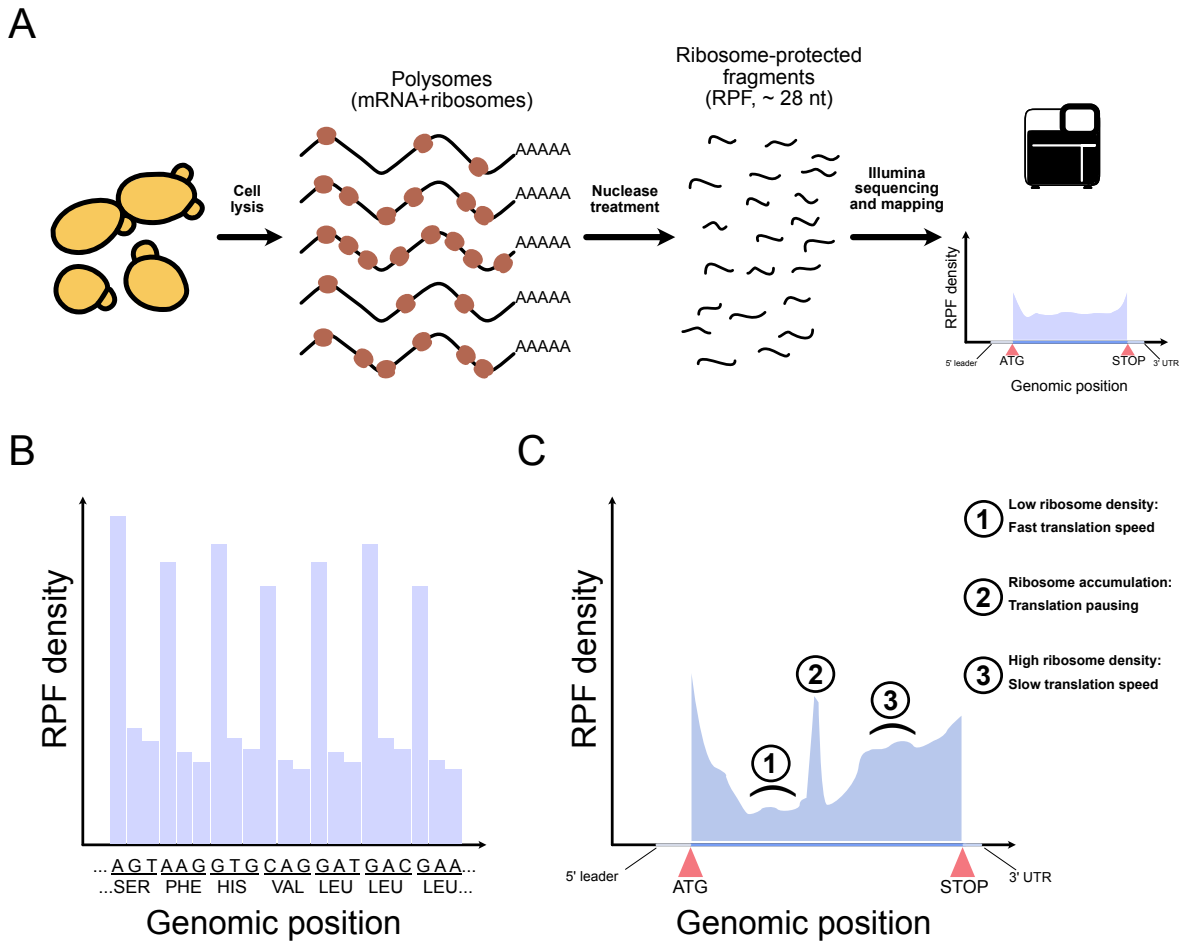


Figure 7: Ribosome profiling is a powerful tool to study translation.

(A) Ribosome-profiling consists of the sequencing of mRNA fragments that have been protected from a nuclease treatment by the translating ribosomes. The reads have a classical length of 28nt and are mostly mapped to the ORF sequence (the 5' leader and the 3' UTR and usually not covered). (B) The ribosome-profiling reads usually show a 3nt periodicity, reflecting the codon-wise movement of the ribosome on the mRNA. (C) The density of the ribosome along the mRNA is a powerful revelator of translations dynamics, such as the elongation speed or translation pauses.

Thanks to the mechanisms underlying the translation process, the ribosome profiling results have some specific features. First, the average length of the reads corresponds to the size of the mRNA covered by the ribosome, *i.e.*, 28 nucleotides (Ingolia, 2010). Second, due to the codon-wise movement of the ribosome along the mRNA, ribosome profiling libraries exhibit a characteristic 3-nucleotide periodicity (Figure 7B), supporting that single codon resolution is achievable with this technique (Ingolia et al., 2009). In addition, the distribution of reads along the gene sequence can be indicative of several translational features, such as changes in elongation rate and ribosomal frameshifting or stalling (Figure 7C) (Ingolia et al., 2019; Michel et al., 2012; Napthine et al., 2017).

Because of this versatility, ribosome profiling has been a powerful tool for the precise study of translation (Brar and Weissman, 2015). In fact, studies have been conducted on quantitative, mechanistic, and spatial aspects of the translation process (Guydosh and Green, 2014; G.-W. Li et al., 2014; Williams et al., 2014). Finally, ribosome profiling has also been a powerful tool for exploring functional genome evolution, and notably how *de novo* genes are expressed and fixed (Blevins et al., 2021; Wacholder et al., 2023).

Currently, ribosome-profiling still faces several limitations (Brar and Weissman, 2015). At the experimental level, rapidly stopping translating ribosomes to obtain an accurate snapshot of translation is a critical step. Translation elongation inhibitors, such as cycloheximide, have been widely used, but they are known to induce ribosome distribution biases, particularly around the translation start site (Guydosh and Green, 2014; Hussmann et al., 2015; Ingolia et al., 2009). In this regard, flash freezing with liquid nitrogen is a robust alternative to effectively stop the ribosome movement (Ingolia et al., 2012). Other experimental biases are known, such as contamination of mRNA fragments that are not ribosome-protected fragments, but rather structured RNA. *In silico* data processing is usually required to address such issues (Ingolia et al., 2014). Finally, one of the main limitations of ribosome profiling is that due to the higher level of sample processed (compared to mRNA), large amounts of cellular material are required to accurately quantify translation (Ingolia et al., 2012). As a result, ribosome profiling remains challenging to scale for single-cell approaches or high-throughput studies. However, recent advances in the sensitivity of the method and in the analytical workflow, especially with the incorporation of machine learning steps, have enabled the development of single-cell ribosome profiling (VanInsberghe et al., 2021). Regarding the adaptation of ribosome profiling for large-scale studies, the linear amplification of ribosome-protected fragments after their extraction is a promising solution and may allow the exploration of translation at the population scale (Mito et al., 2023).

Analytical link between genotype and phenotype variations

Accurately determining the genetic loci involved in phenotypic variation has been a constant and arduous challenge over the past 50 years. The development of molecular genetics and biology, together with advances in DNA sequencing through the Sanger technique (Sanger et

al., 1977), led to classical approaches to link genotypes and phenotypes, such as random mutagenesis. In this technique, chemical or physical mutagenesis agents were applied to individuals such as mice or yeast to induce point mutations. These latter were then associated to wide ranges of phenotypic changes, such as cell division in the fission yeast *Schizosaccharomyces pombe* (Nurse et al., 1976; Nurse and Thuriaux, 1980) or motility in the nematode *Caenorhabditis elegans* (Brenner, 1974). In parallel, the advent of the PCR amplification techniques allowed more targeted genetic disruption, especially in model organisms with efficient recombination capacities, such as *S. cerevisiae* (Shortle et al., 1982; Wach et al., 1994). In this case, systematic deletion (or knock-out, KO) of each gene in the genome allows precise characterization of the effect of each gene (Giaever and Nislow, 2014). More recently, targeted mutagenesis was deeply developed with the introduction of CRISPR-CAS9 technologies (Cong et al., 2013; Doudna and Charpentier, 2014; Jinek et al., 2012). CRISPR-CAS9 allows for highly efficient and nucleotide resolution modifications that are difficult to achieve in human cells, for example.

However, while these methods are highly efficient for characterizing the cellular or molecular effects of particular genetic variants, they are not well suited for exploring the influence of natural genetic variation on phenotype, especially for complex traits. Such investigations require either a large cohort of individuals or a model organism for which large-scale segregant generation is possible. Genetic association studies are the main strategy for such investigations. Two main types of association studies are commonly used: linkage analysis (or linkage mapping) and genome-wide association studies (GWAS). The goal of these strategies is to detect and associate genetic regions with quantitative phenotypic variation (*e.g.*, human height, mRNA or protein abundance, growth on a particular medium). The results of these studies are called quantitative trait loci (or QTL). Depending on the method used to link genetic regions to quantitative phenotypes, the resolution of QTLs can vary from large chromosomal regions to single nucleotide variants. Gene expression, specifically mRNA and protein abundance, can be used as a phenotype to investigate the genetic origins of gene expression variation between individuals. In the case of mRNA abundance, the detected loci are usually referred to as eQTL (for expression QTL) (Nica and Dermitzakis, 2013). In the case of protein abundance, they are referred as pQTL (for protein QTL) (Ferkingstad et al., 2021).

Linkage mapping studies

Linkage mapping studies are based on the generation of a large number of segregants from an original cross between two genetically and phenotypically distinct individuals (Figure 8A). Due to meiotic recombination, the parental genetic variants are shuffled among the offspring, resulting in unique genotype in each segregant. By combining the phenotypic measurement of both parents and segregants with their genotypes (Figure 8B), it is possible to recover the causal regions of the phenotype under study. In fact, individuals with a similar phenotype will most likely share the genomic regions that carry the causal variants, while the vast majority of the other parental variants will be randomly distributed along the genome (Figure 8C). While early linkage mapping studies relied on genetic markers such as restriction fragment length polymorphisms (Botstein et al., 1980), most linkage mapping analyses are based on SNPs (Albert et al., 2018; Brem et al., 2002). More recently, SVs were also integrated in such analyses (Weller et al., 2023). The budding yeast *S. cerevisiae* is an important tool for linkage mapping studies because it combines several characteristics that are crucial for an efficient QTL detection: a small genome, a short sexual generation time and, more importantly, a high meiotic recombination rate (Fay, 2013; Liti and Louis, 2012).

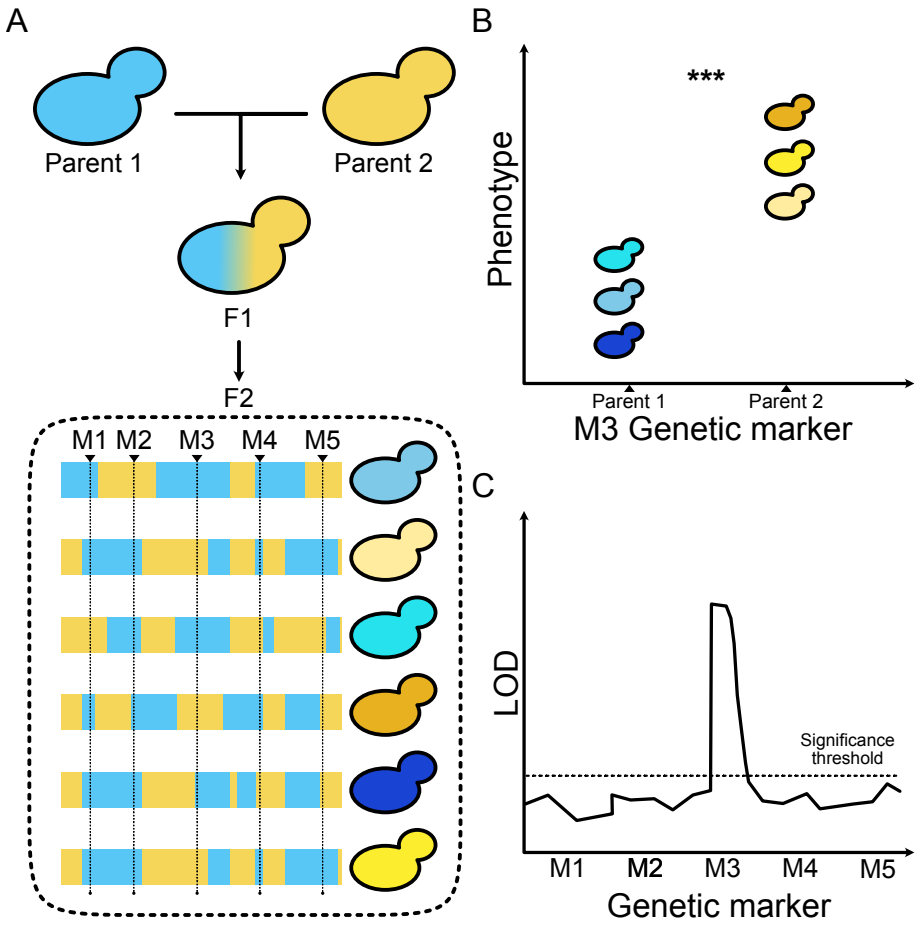


Figure 8: Overview of the linkage mapping studies.

(A) In linkage mapping studies, 2 individuals with different genotypes and phenotypes are crossed and the phenotype of their offspring is associated with genetic markers (M1, M2... M5) distributed throughout the genome (*e.g.*, SNPs). (B) A genetic marker associated with the phenotype difference between the parents will also show this association in the segregants. (C) All markers ultimately have an association score (e.*g.*, LOD, logarithm of odds) that must exceed a certain threshold to be considered significant.

Disentangling the genetic origins of inter-individual variation in gene expression has often been investigated using linkage mapping. Studies of mRNA abundance have been performed in several organisms, ranging from plants (Jansen and Nap, 2001) to animals (Schadt et al., 2003) and of course, yeast (Albert et al., 2018; Brem et al., 2002), resulting in the discovery of thousands of eQTL. Protein abundance has also been explored and genetically mapped using linkage mapping studies. For example, in yeast, linkage mapping and proteomics have been combined twice (Albert et al., 2014; Foss et al., 2007), but due to technical limitations, the number of proteins included in these studies was limited. In the case of the earlier study (Foss et al., 2007), the LC-MS/MS used at the time had a high signal-to-noise ratio, which made it difficult to repeatedly cover a large number of proteins in the samples. In the end, 221 proteins were used for linkage mapping. In the latter study (Albert et al., 2014), the proteins were quantified using a single-cell measurement, where the proteins are fused to GFP and the signal is measured using fluorescence-activated cell sorting. To ensure good statistical power and the possibility to control the results, only 160 proteins were ultimately used.

While linkage mapping approach is a powerful method to determine the genomic region associated with a phenotype, it has some inherent limitations. First, the genetic diversity captured by this method is limited to that found in the two parental individuals which does not recapitulate the complete genetic diversity of the species. Also, depending on the recombination rate of the organism under study, the resolution of linkage mapping may be limited, especially if the study focuses on the first generation of offspring (Flint et al., 2005). As a result, large regions are associated with the phenotype under study, and precise identification of the causal variant can be tedious. Several tools have been developed specifically to address these issues. For example, in mice, an entire lineage has been generated from 8 inbred individuals (Collaborative Cross Consortium, 2012) leading to an increase in both genetic diversity and resolution (Gatti et al., 2014). In yeast, the generation of large population by crossing 16

genetically distinct individuals has also been achieved to tackle the genetic diversity limitation (Bloom et al., 2019). Given the limited resolution of linkage mapping studies, one possible solution is to map the genetic origin of a phenotype using the F6 generation of the parental cross (Jakobson and Jarosz, 2019).

Genome-wide association studies

The development of large-scale sequencing projects over the last two decades has led to a clearer view of intraspecific genetic diversity. In the mid-2000s, using large cohorts of individuals whose allelic status is clearly defined across their genome, a new method for linking genotypes to phenotypes was developed: genome-wide association studies (GWAS) (Dewan et al., 2006; Klein et al., 2005; Wellcome Trust Case Control Consortium, 2007). This method is based on testing the association of hundreds of thousands or even millions of variants with phenotypes of interest. Until recently, the tested genetic variants were mostly SNPs (Uffelmann et al., 2021) sometimes completed with CNVs data (Wellcome Trust Case Control Consortium, 2010).

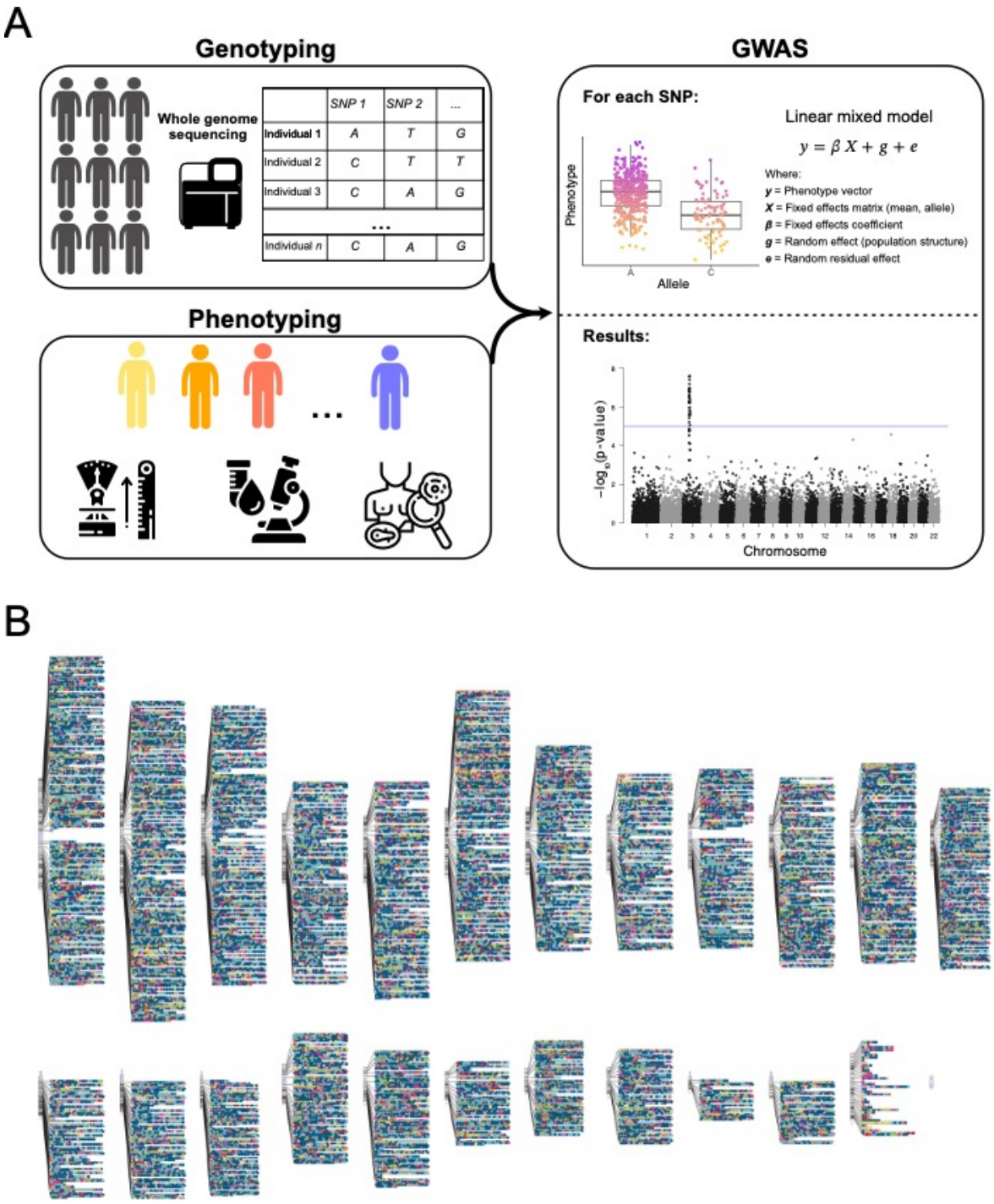


Figure 9: Overview of GWAS.

(A) GWAS are performed by combining genotyping data, where the allelic state of each individual is known, and phenotypic data for each individual in the population. The association between the phenotype and each variant is performed using an LMM that considers confounding factors, such as population structure. The final result of a GWAS is typically presented using Manhattan plots where the association of each SNP along the genome has an association score ($-log_{10}(p\text{-}value)$) that must exceed a threshold (represented by the horizontal line) to be considered as significant. (B) Overview of the human GWAS where each colored dot corresponds to an association between a locus and a

human phenotype. The colors represent the type of phenotype, see www.ebi.ac.uk/gwas/ for more details. Plot and data were generated and obtained from the NHGRI-EBI catalog (MacArthur et al., 2017).

Overall, the experimental workflow of a GWAS follows several steps (Figure 9A). First, a large population is genotyped to resolve the allelic status of each individual. Historically, microarray data were used to resolve the genotype of each individual, but nowadays whole-genome sequencing is preferred and has the advantage to capture nearly all genotypic variation across the genome. Each of the variants (generally the bi-allelic SNPs) is then tested for association with the phenotype of interest. There are several ways to perform the association, among which the used of either linear mixed regression models (Figure 9A) or logistic mixed regression models, depending on the nature of the phenotype (either continuous or discrete) (Uffelmann et al., 2021). As population structure or familial relatedness is a major confounding factor in GWAS (Balding, 2006; Kang et al., 2010; Zhang et al., 2010; Zhao et al., 2007), linear mixed models (LMM) are powerful statistical methods to correct for such confounding. As LMM can be computationally intensive, several tools have been developed to increase the accessibility of these methods, such as TASSEL or FaST-LMM (Lippert et al., 2011; Zhang et al., 2010). Because associations are made across thousands to millions of SNPs, false discovery rate correction is a critical part of GWAS. Depending on the population and the species, different types of significance correction are used. In humans, a common method is to use a Bonferroni corrected p-value to detect significant associations, resulting in a p-value threshold of 5×10^{-8} (Uffelmann et al., 2021). For more complex GWAS focusing on multiple phenotypes simultaneously, a trait-specific p-value can be defined by performing permutation tests (Caudal et al., 2023; Peter et al., 2018). The results of a GWAS can be easily visualized using a Manhattan plot (Figure 9A), where for each SNP across the chromosomes (x-axis), an association score (usually the $-\log_{10}$ transformation of the association p-value) is plotted (y-axis). Since its development, GWAS have been used to study a wide variety of phenotypes. As of April 2023, more than 6,000 human GWAS have been published, for which more than 500,000 associations has been highlighted (Figure 9B) (MacArthur et al., 2017). Human diseases have naturally caught lot of attention and examples include cancer (Sud et al., 2017), type 2 diabetes (Zhao et al., 2017), and psychological or mental disorders (Duncan et al., 2017; Hyde et al., 2016; Jansen et al., 2019; Li et al., 2017). Non-pathological complex phenotypes can also be investigated with GWAS such as body mass index (Yengo et al., 2018), educational attainment (Lee et al., 2018), or even musical beat synchronization (Niarchou et al., 2022).

Of course, gene expression itself has been studied with GWAS. In this case, a hundred to a few thousand phenotypes are analyzed simultaneously, *i.e.*, mRNA or protein abundance. Technically, GWAS focusing on gene expression can be tedious to perform, as gene expression exploration (e.g., RNA sequencing or LC-MS/MS) has to be performed in each individual. In addition, as mentioned above, specific significance thresholds need to be set in such studies, as the risk of false-positive discovery is high (Liu et al., 2019). A common way to visualize GWAS performed on mRNA or protein abundance is to plot the genomic location of the eQTL or pQTL against the location of its associated trait (*i.e.*, the affected gene) (Figure 10A). Several studies have focused on the genetic origins of mRNA abundance, the most famous in human being the Genotype-Tissue Expression (GTEx) (The GTEx Consortium, 2020, 2017, 2015), in which transcript abundance was monitored in 49 human tissues from 838 postmortem donors. Recently, a catalog of human eQTL has been established (Kerimov et al., 2021). Plants (Lan et al., 2021) and of course, yeast (Caudal et al., 2023) have also been used for eQTL exploration by GWAS. Fewer large-scale studies have been performed for pQTL exploration with GWAS (Suhre et al., 2021), mainly because large-scale quantification of protein abundance has been a major limiting factor. Notable studies focused on the plasma and serum proteome (Ferkingstad et al., 2021; Gudjonsson et al., 2022). Until now, finding a good trade-off between the number of samples in a proteomic GWAS and the number of proteins included has been tedious.

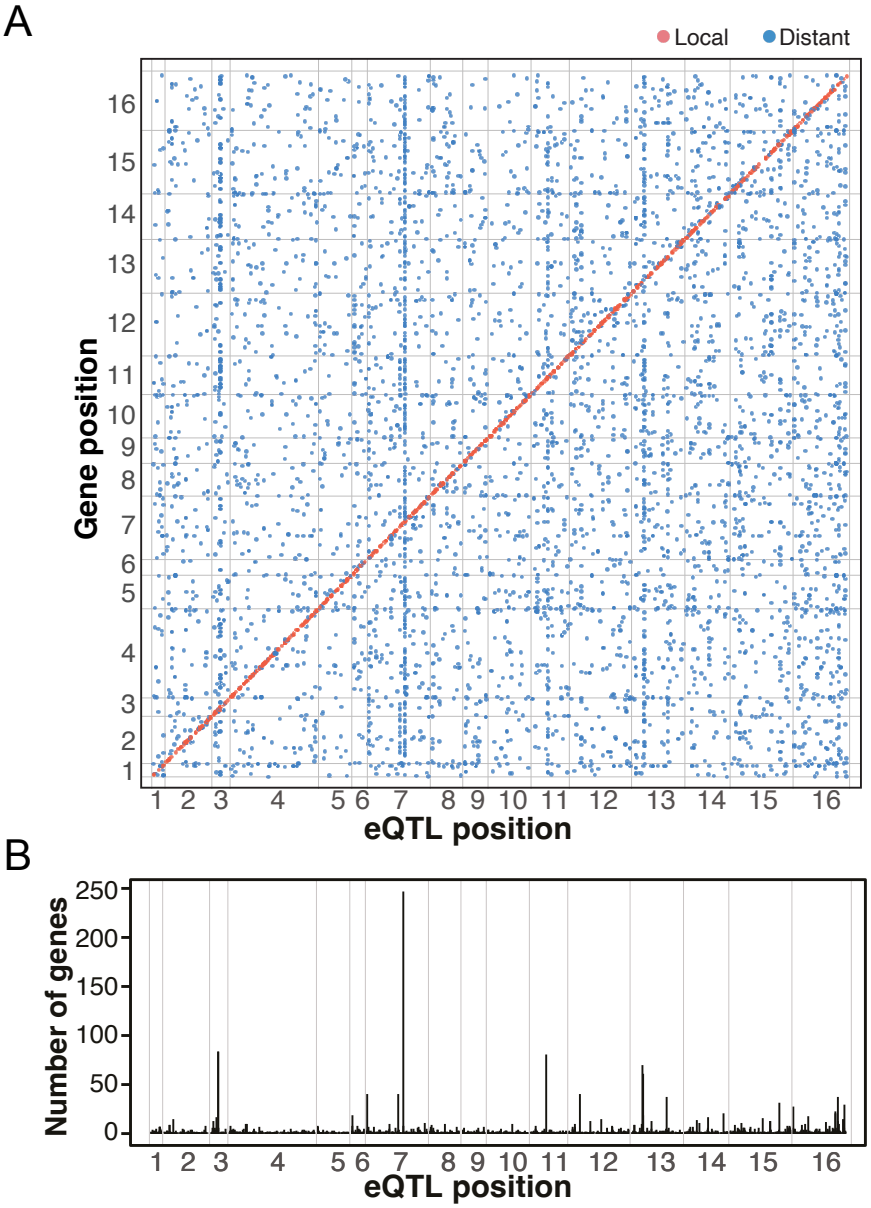


Figure 10: Expression-based GWAS are powerful tools for detecting the mechanisms of gene expression regulation.

(A) Expression-based GWAS are typically represented by plotting the eQTL or pQTL position against the target gene position. The diagonal line (red dots) represents the eQTL or pQTL affecting the abundance of nearby genes. The accumulation of points on a precise vertical line reveals the presence of QTL hotspots that affect numerous genes across the genome. (B) The QTL hotspots are easily observed by looking at the number of genes affected by the QTL in a genetic window (here 25kb). Both plots are adapted from Caudal et al., 2023.

Despite the significant progress that has been made with GWAS, the technique still suffers from limitations (Tam et al., 2019). First, due to the stringency applied to avoid false-positive associations, genetic variants are likely to be missed. Therefore, only a fraction of the

heritability of a complex trait is captured by GWAS (Dudbridge and Gusnanto, 2008; Manolio et al., 2009) and this fraction will be biased towards high effect variants. Conversely, small effect variants are more difficult to detect. In addition, a large fraction of the heritability of traits is also missed because several sources of genetic variants are poorly considered (Manolio et al., 2009; Zuk et al., 2014). For example, SVs are rarely accounted for in GWAS, mainly because of limitations for their accurate detection that were only recently overcome, but also because of the difficulty to integrate these data in GWAS. The democratization of long-read sequencing and the development of pangenome graph-based GWAS are promising solutions to account for the SV effect (He et al., 2023; Li et al., 2022; Logsdon et al., 2020; Zhou et al., 2022). Rare variants are also poorly considered in GWAS because their low representation in the cohorts studied makes it difficult to detect their association with sufficient statistical power. In this regard, the construction of diallel crosses allows to artificially increase the allele frequency of rare variants (Fournier et al., 2019). Finally, due to linkage disequilibrium (the non-random association between two or more alleles on different loci), finding the correct causal variant can be difficult (Altshuler et al., 2008), especially when the QTL is in a non-coding region, and sometimes requires additional studies to confirm the effect of a genetic variant. For this reason, the predictive power of GWAS in the clinical context remains limited (Janssens and van Duijn, 2008; Loos and Janssens, 2017). Increasing the size of cohorts is a good solution to overcome such difficulties (Tam et al., 2019). In this context, the development of high-throughput molecular phenotyping (Caudal et al., 2023; Messner et al., 2022) is a promising approach to overcome some of the current limitations of GWAS.

Gene expression and SNP

As described above, both linkage mapping approaches and GWAS have been used to dissect the genetic origins of variations in transcript and protein abundance across individuals. At the transcriptomic level, extensive efforts have been made to understand which and how genetic variants, more specifically SNPs, affect mRNA abundance (Caudal et al., 2023; Foss et al., 2007; Gan et al., 2011; Ghazalpour et al., 2011; The GTEx Consortium, 2015). These studies have led to the discovery of thousands of eQTL that affect gene expression through a variety of mechanisms. Some eQTL affect gene expression of nearby genes, while others affect their target genes in a distant manner (Figure 11). The former, also known as local- or *cis*-eQTL (or local- and *cis*-pQTL in the case of protein abundance), usually affects the ability of the transcription machinery to bind to the promoter of its target gene (Figure 11A). Any element

involved in the direct regulation of a gene's transcription can be affected, such as the core promoter (Lubliner et al., 2015; Tirosh et al., 2009), enhancer regions (Garieri et al., 2017; Kikuchi et al., 2019), nearby chromatin accessibility (Keele et al., 2020), or terminal regions (Hill et al., 2021). In addition, the local-eQTL tend to have a greater effect on their target genes compared to the distant-eQTL (Albert et al., 2018; Caudal et al., 2023). In humans, late lactose tolerance is a famous case of local DNA variants inducing a change of gene expression, resulting in a persistent lactase expression. Briefly, while populations with the ancestral allele have a decreased expression of the lactase (*LCT*) gene after childhood, single mutations (either alone or in combination) in an enhancer region of *LCT* create a new binding site for a transcription factor leading to a non-downregulated pathway for *LCT* expression (Enattah et al., 2002; Fang et al., 2012; Lewinsky et al., 2005; Olds and Sibley, 2003). Graphically, *cis* regulation is easily observed as a diagonal line when plotting the eQTL position vs the target gene position (reflecting a similar location of the QTL and the trait on the genome, Figure 10A).

Conversely, eQTL or pQTL described as distant (or *trans*-) can be located anywhere in the genome, either on the same or different chromosomes (Figure 11B). The effects of distant-QTL are usually achieved through the proteins or RNA involved in transcriptional or translational regulation such as mRNA binding protein, transcription factors or non-coding RNA (He et al., 2020; Lutz et al., 2019). Interestingly, distant-QTL tend to be located in hotspots (Albert et al., 2018; Qu et al., 2018; Yao et al., 2017) and have more pleiotropic effects than local regulatory variants (Lemos et al., 2008; Prud'homme et al., 2007). On a plot of QTL position versus affected gene position, distant QTL hotspots are graphically indicated by the vertical accumulation of points (Figure 10B). Due to their multiple effects, distant regulatory variants tend to be more deleterious and less beneficial than local regulatory variants (Coolon et al., 2015; Emerson et al., 2010; Schaefke et al., 2013). Therefore, they are an important driver of the relationship between diseases and their genetic origins (Westra et al., 2013), as it is the case for several autoimmune pathologies such as type 1 diabetes or systemic lupus erythematosus (Han et al., 2009; Heinig et al., 2010).

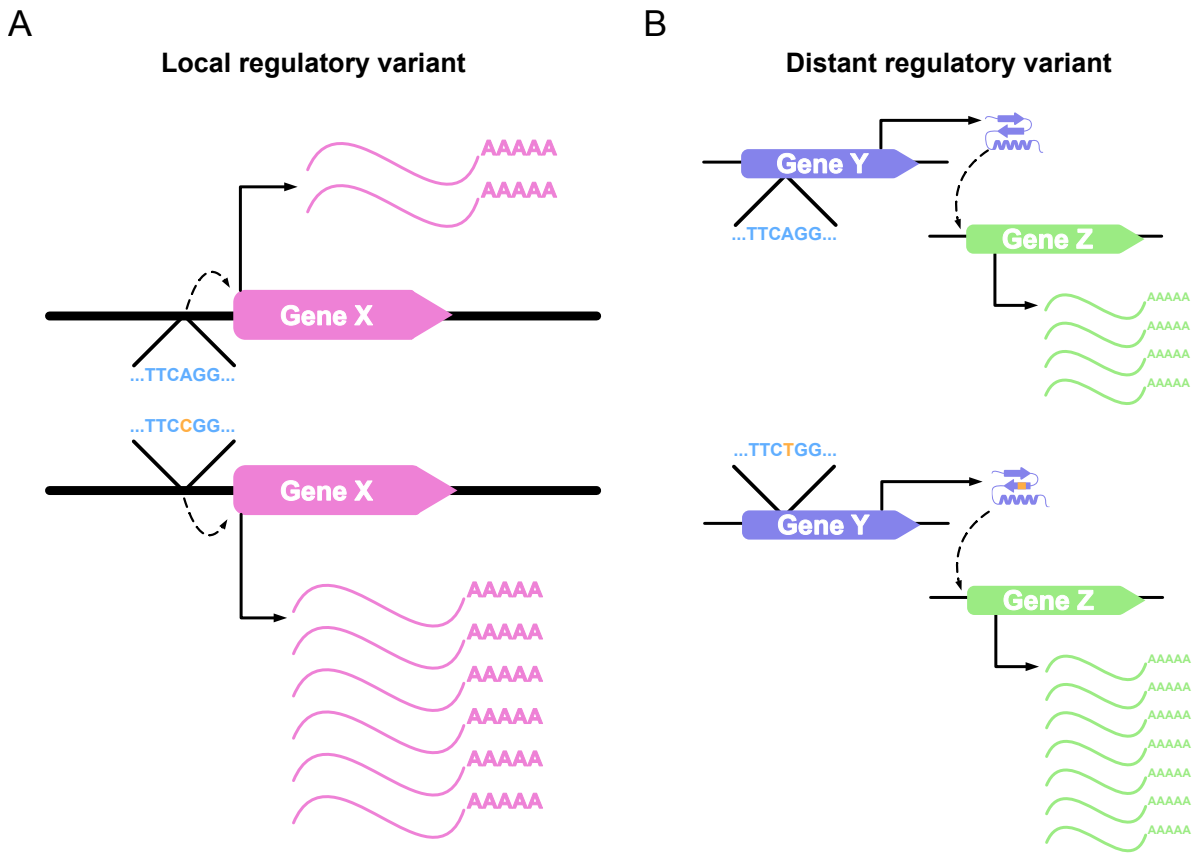


Figure 11: The genetic variants associated with changes in gene expression can affect the expression either locally or distantly.

(A) Local regulatory variants are located close to their target gene, usually in the promoter region and thus alter the ability of various proteins, such as transcription factors, to bind to DNA. (B) Distant regulatory variants affect the expression of genes located on another chromosome or far away on the same chromosome. This is usually achieved through the action of an intermediary protein.

Gene expression and SV

Structural variants are also major contributors to variability in mRNA or protein abundance. Down syndrome (or trisomy 21) is a notable example, where individuals carry a full or partial extra copy of the 21st chromosome. This results in a huge macroscopic phenotypic impact and increased risk for a wide range of diseases (Patterson, 2009). At the molecular level, the consequences of this aneuploidy are extensive, affecting gene expression of genes located on all chromosomes (Letourneau et al., 2014; Prandini et al., 2007). In plants, a large-scale study of 100 tomato lines revealed that SVs (which in this organism are mainly associated with transposons) are a major source of variation in gene expression across the species (Alonge et al., 2020). Another elegant example of this is the case of the sulfite tolerance in *Saccharomyces cerevisiae*. Several *S. cerevisiae* strains isolated from winemaking environment

have shown an enhanced tolerance to sulfite (Pérez-Ortín et al., 2002), a common chemical used to suppress the growth of various non-*Saccharomyces* yeasts or lactic bacteria (Ribéreau-Gayon et al., 2006). This enhanced tolerance was associated with a reciprocal translocation between chromosomes VIII and XVI, resulting in an overexpression of the *SSU1* gene, a sulfite pump whose promoter is altered by the translocation (Pérez-Ortín et al., 2002). Later studies showed that three different chromosomal rearrangements (two translocations and one inversion) can induce *SSU1* overexpression in different yeast isolates (García-Ríos et al., 2019; Yuasa et al., 2004; Zimmer et al., 2014). However, due to the tedious nature of detecting and characterizing SVs in large populations, their impact on gene expression, especially at the population level, remains largely unexplored. Recently however, CNV has been taken into consideration to explore variation in transcript abundance in yeast (Caudal et al., 2023).

The transcriptome and proteome relationship

In the last two decades, both technological (*e.g.*, RNA sequencing, large-scale LC-MS/MS) and analytical (*e.g.*, GWAS) developments have led to a better understanding of the molecular mechanisms underlying the genotype-phenotype relationship. However, when both transcript and protein abundance studies are numerous, their conclusions can be very contradictory, especially when it comes to the relationship between the transcriptome and the proteome. As mentioned previously, the apparent linear and hierarchical nature of gene expression hides a complex and tightly regulated phenomenon (Buccitelli and Selbach, 2020; Liu et al., 2016; Vogel and Marcotte, 2012). Thus, the final protein abundance at any-given time in a cell is the result of a subtle balance between transcription rate, mRNA half-life, translation rate, protein half-life, and the cell cycle progression (Baum et al., 2019; Buccitelli and Selbach, 2020). Therefore, many aspects of the relationship between the transcriptome and the proteome are poorly understood. For example, how well protein abundance can be predicted from transcript abundance is still an ongoing debate. Similarly, it is still unclear if and how proteome variation reflects transcriptome variation. Large-scale exploration of mRNA and protein abundance is a promising tool to address these grey areas, as it allows precise quantification of expression variation and deep exploration of the genetic origin of these variations. However, at the population level, studies focusing on the relationship between the transcriptome and the proteome are sparse.

Transcript-protein correlation

A good way to explore the interaction between the transcriptome and the proteome is to focus on the correlation between mRNA and protein abundance. An important distinction must be made, as two types of mRNA-protein correlations can be calculated (Buccitelli and Selbach, 2020; Liu et al., 2016): the first focuses on the correlation of all genes in a sample (*e.g.*, a tissue, a cell, a strain...), while the other highlights the variation in mRNA and protein abundance across different conditions. The first type is referred to as the "across-gene" correlation while the second is referred to as the "within-gene" correlation. The confusion between these two types of correlations is common, even in the scientific literature (Fortelny et al., 2017). Correlations are often calculated using either Spearman or Pearson coefficients, and the resulting value is highly dependent on the type of correlation (*i.e.*, across- or within-gene).

Across-gene correlation

As explained above, the across-gene correlation focuses of the comparison between the mRNA and protein levels within a single sample. Graphically, this is often analyzed by plotting the transcript and peptide abundance together (Figure 12A), where each point corresponds to a gene. The across-gene correlation has been extensively studied in several species such as humans (Battle et al., 2015; Edfors et al., 2016; Gautier et al., 2016; Salovska et al., 2020; Wang et al., 2019; Wilhelm et al., 2014; Zhang et al., 2014), mice and rats (Aydin et al., 2023; J. J. Li et al., 2014; Moritz et al., 2019; Schwanhäusser et al., 2011), fruit flies (Becker et al., 2018), maize (Ponnala et al., 2014) and, of course, yeast (Gygi et al., 1999; Ingolia et al., 2009; Marguerat et al., 2012). In the vast majority of these studies, across-gene correlation typically shows medium-high to high correlation indexes (0.4-0.8) (Figure 12B). This observation is consistent through all the species in which this has been examined. Overall, this means that highly abundant transcripts encode for highly abundant proteins. It is important to note that across-gene correlation may be sensitive to the time point at which the mRNA and proteome abundances were surveyed. Indeed, the hierarchical nature of gene expression will introduce a delay between transcription and translation (Fournier et al., 2010; Gedeon and Bokes, 2012). Therefore, when biological samples are exposed to changing conditions, a change in gene expression will follow, and for most genes, this will first affect transcription and then translation. Ultimately, two measures of mRNA and protein abundance during the steady state and transition phase will most likely have different correlation levels, with the steady state

correlation index being higher than the transition state correlation index. Also, biological features such as the group of genes being monitored or the context of the study (*e.g.*, tissue, cell type...) are known to influence the mRNA-protein correlation level (Buccitelli and Selbach, 2020). For example, in a study of several cell lines and tissues targeting specific proteins, the Pearson coefficient between mRNA and protein abundance varied from 0.39 to 0.79 (Edfors et al., 2016).

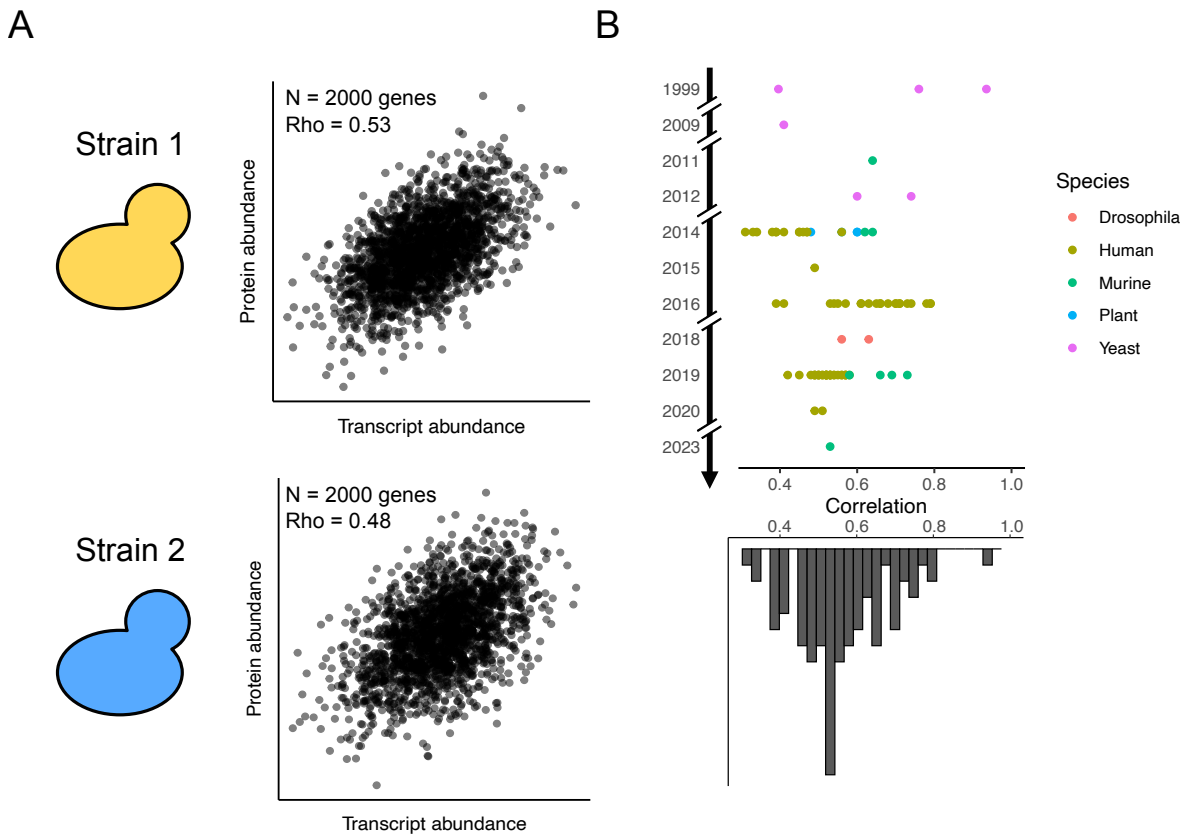


Figure 12: The across-gene correlation shows a good match between transcript and protein levels within a sample.

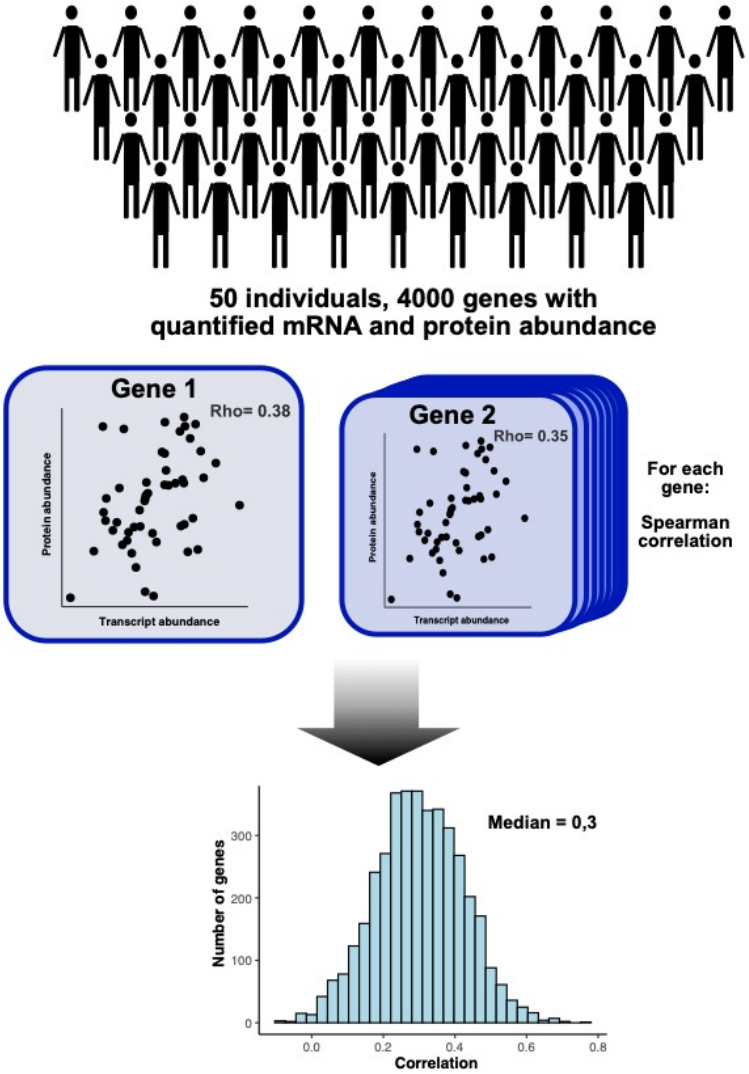
(A) The across-gene correlation quantifies the correlation between transcript and protein abundance within a sample, such as a cell line, a tissue, or a strain (as shown here). This correlation is calculated using the overlap between protein and transcript measurements from the sample. (B) Over the past two decades, several measurements or the across-gene correlation have been made across multiple species and the calculated coefficients typically fall between 0.4 and 0.8. Data gathered from Buccitelli and Selbach, 2020, and from Aydin et al., 2023.

Overall, this type of correlation is a good indicator to empirically quantify the impact of mRNA abundance on the final protein abundance (Liu et al., 2016). Investigations on mouse data showed that between 56% and 84% of the variance in protein levels is explained by mRNA

levels (J. J. Li et al., 2014). This highlights that at steady state and within a sample, protein abundance is primarily explained by mRNA abundance.

Within-gene correlation

While across-gene correlation provides valuable information about how mRNA and protein levels are coupled in a specific sample, it does not provide information about how transcript and protein abundance variations are reflected at larger scales. Within-gene correlation is a powerful tool to interrogate and investigate gene expression changes across multiple samples (tissues, cell types, strains, growth conditions...). In this approach, the mRNA-protein correlation is calculated for each gene using the sample transcript and protein levels (Figure 13). Interestingly, there is no clear consensus across studies for the within gene correlation (Archer et al., 2018; Aydin et al., 2023; Battle et al., 2015; Chick et al., 2016; Ghazalpour et al., 2011; Huang et al., 2017; Jiang et al., 2020; Mertins et al., 2016; Mirauta et al., 2020; Mun et al., 2019; Vasaikar et al., 2019; Wang et al., 2019; Zhang et al., 2014, 2016). Within gene correlation indexes range from 0.14 to 0.59 (Battle et al., 2015; Upadhya and Ryan, 2022), so there is no clear consensus on whether mRNA and protein changes are correlated or not.



Figures 13: The within-gene correlation highlights how the protein abundance variations match the mRNA abundance variations.

In the within-gene correlation, the transcript and protein abundances of all the individuals are correlated for each gene. Graphically, a plot can be drawn for each gene where each dot corresponds to a sample. The correlation indexes are calculated for each gene and used to get an overall correlation median.

There are several reasons for this uncertainty. Technical biases can be an influencing factor, as proteins that are preferentially captured by previous proteomic methods will show up at higher correlation levels (Alam et al., 2016; Upadhya and Ryan, 2022). The type of data used in the correlation calculation (*i.e.*, absolute, or relative mRNA and protein quantification) is also a determinant of the overall correlation levels. Indeed, absolute transcript and protein levels span several orders of magnitude, while the relative expression change of protein across samples remains in a much narrower range (Marguerat et al., 2012; Messner et al., 2023). Similarly, the

magnitude of variation across the samples also strongly influences the mRNA-protein correlation: genes with important expression variation are more likely to have a high within-gene correlation because the changes are more likely to affect both expression levels (Buccitelli and Selbach, 2020; Wang et al., 2019). In part because of this, the cellular function of the gene is also an important determinant of within-gene correlation. For example, metabolism-related genes tend to be associated with high levels of correlation (Buccitelli and Selbach, 2020; Wang et al., 2019). It is worth noting that these genes are known to have highly variable expression across individuals (Caudal et al., 2023). Conversely, ribosomal genes tend to show no correlation or slight anticorrelation (Buccitelli and Selbach, 2020; Wang et al., 2019). Expression noise is also a confounding factor for mRNA-protein correlation, especially for low expressed genes where biological signal variations are of the same magnitude as the noise. More generally, the precision of proteome and transcriptome exploration will strongly influence the mRNA-protein correlation (Buccitelli and Selbach, 2020). Finally, the aforementioned investigation included at most a few hundred samples (192 mouse samples in Chick et al., 2016). To truly explore both proteome and transcriptome variation at the population level, larger cohorts of individuals are required. Until recent advances (Messner et al., 2022), this was technically difficult to achieve.

Post-transcriptional buffering

Although tightly regulated, gene expression is a noisy process, especially at the transcriptional level. External factors can induce gene expression noise through surface receptors, as cells live in a highly fluctuating environment where important changes in condition must be distinguished from rapid and noisy signals (Liu et al., 2016). Similarly, internal factors such as random transcription initiation can also lead to expression noise (Chalancon et al., 2012; Gandhi et al., 2011). Erroneous or inappropriate gene expression can ultimately lead to proteome imbalance and is obviously detrimental to proper cellular homeostasis. If the external factors can be compensated by annealing incorrect cellular signals before they affect gene expression (Chalancon et al., 2012; Hornung and Barkai, 2008), the internal factors are more likely to activate transcription. Cells must therefore cope with expression noise. More generally, the overall effect of gene expression variations can alter several key cellular functions. Interestingly, some of these central functions tend to show robustness to expression variation

(Félix and Barkoulas, 2015). Post-transcriptional buffering is a good example of the cellular mechanisms that deal with deleterious transcriptional noise or variation.

Different contexts, one phenomenon

The phenomenon of post-transcriptional buffering describes the fact that transcriptional variation tends to be buffered as the gene expression process progresses (Figure 14A). Over the past decades, this phenomenon has been repeatedly observed in several contexts and has thus emerged as a crucial determinant of the relationship between transcripts and proteins. Several early investigations of post-transcriptional buffering were made by comparing proteome and transcriptome changes associated with CNV in cancer cells (Geiger et al., 2010; Stingele et al., 2012).These studies, along with later investigations on larger sample in both human cancer cells and yeast (Dephoure et al., 2014; Gonçalves et al., 2017; Liu et al., 2017; Zhang et al., 2014) highlighted that the proteome composition is not as sensitive to gene dosage variation as the transcriptome, which is known to typically reflects CNVs (Fehrmann et al., 2015; Schlattl et al., 2011). Accordingly, early interspecies comparisons between proteome and transcriptome also highlighted that proteome variation were more constrained than transcriptome variation (Khan et al., 2013; Laurent et al., 2010; Schrimpf et al., 2009).

Taken together, these results indicate that protein abundance is more constrained and conserved than transcript abundance. Conceptually, this fits with the scheme that proteome variation will more directly affect the final phenotypic landscape compared to the transcriptome. Thus, changes in protein abundance are more likely to be deleterious and will be under stronger selective pressure. In addition to mRNA and protein comparisons, several ribosome profiling experiments showed that transcript abundance fluctuations between individuals, species or even different conditions are also buffered at the translational level (Artieri and Fraser, 2014; Blevins et al., 2019; McManus et al., 2014; Wang et al., 2015, 2020). This highlights that post-transcriptional buffering is a multilayered phenomenon, suggesting that multiple mechanisms underlie the phenomenon.

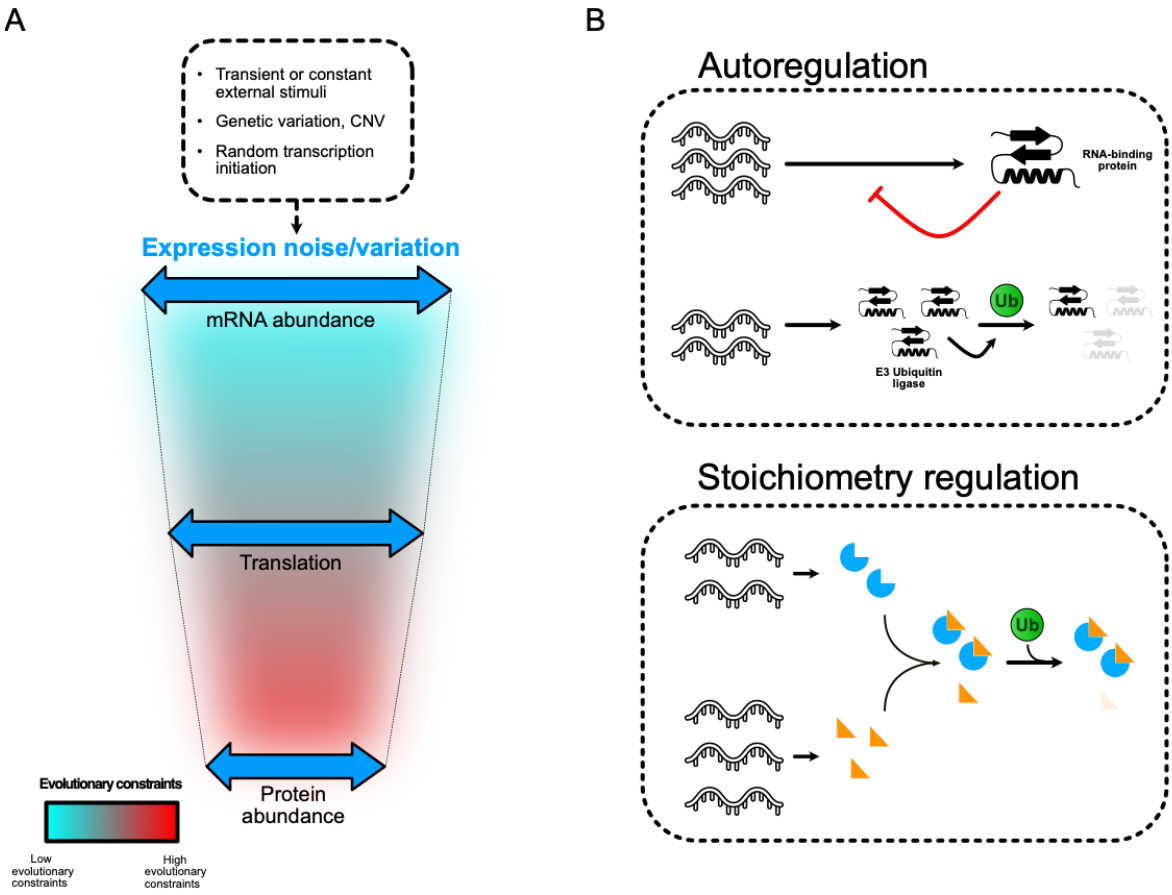


Figure 14: Post-transcriptional buffering is a central phenomenon in dealing with transcriptional variation.

(A) Post-transcriptional buffering allows the cell to counteract the effect of transcriptional variation and results in a reduced variation at the translatome and proteome levels. This suggests that these two levels are under stronger evolutionary constraints than transcriptome. (B) The mechanisms underlying post-transcriptional buffering are still under debate. Autoregulation and stoichiometry regulation are thought to be part of these mechanisms. Adapted from Buccitelli and Selbach, 2020.

Mechanisms underlying the post-transcriptional buffering

Although post-transcriptional buffering is frequently observed when comparing transcription, translation, and protein abundance together, the cellular mechanisms underlying this phenomenon remain elusive and poorly understood. Since post-transcriptional buffering is observed in both ribosome profiling and proteomic data, it is likely that multiple and distinct processes are involved.

Conceptually, post-transcriptional buffering requires some kind of feedback in which the cell detects and copes with expression fluctuations. Autoregulation may be a simple mechanism that allows this type of feedback (Figure 14B), as many transcription factors and RNA-binding

proteins can inhibit their own expression (Grönlund et al., 2013; Müller-McNicoll et al., 2019). E3 ubiquitin ligases are also under strong autoregulation as they are able to recognize and target their excess proteins, thereby inducing proteasome-mediated degradation (de Bie and Ciechanover, 2011). However, autoregulation is limited to proteins that can influence their own expression or abundance (transcription factors, mRNA-binding proteins, ubiquitin ligases…) and is obviously not sufficient to explain the extensive and frequently observed buffering. Moreover, it is mechanistically difficult to imagine a global mechanism capable of detecting fluctuations in protein abundance and driving translation in response to these variations. Accordingly, a previous study in aneuploid yeast has not found feedback mechanisms on protein synthesis in the case of expression variation for molecular complex-related proteins (Taggart and Li, 2018). Yet, the proteins with the most robust abundance to variation seem to be precisely molecular complex proteins (Dephoure et al., 2014; Gonçalves et al., 2017; Liu et al., 2017; Stingele et al., 2012). Indeed, protein degradation of the unassociated complex components may play an important role in post-transcriptional buffering at the proteome level (Gonçalves et al., 2017; Juszkiewicz and Hegde, 2018; Taggart et al., 2020). The strong robustness of the complex-related protein is consistent with the numerous investigations highlighting that gene dosage imbalance can be highly deleterious (Deutschbauer et al., 2005; Morrill and Amon, 2019; Ohnuki and Ohya, 2018; Veitia and Potier, 2015). Again, these findings represent only a subset of genes. What other processes enable the cell to cope with expression variation at the proteome level is still unknown.

At the level of translation, knowledge about post-transcriptional buffering is even more scarce. Although the phenomenon has been detected in several studies, no clear description of post-transcriptional buffering at this level has been made. The few insights into the mechanisms underlying translational buffering suggest that variations in mRNA abundance at the translation level tend to be attenuated by modulation of translation efficiency (*i.e.*, the ratio between the quantification of translation for a gene and its mRNA abundance) (McManus et al., 2014). Overall, although post-transcriptional buffering plays a major role in expression variation across multiple scales (intra- and interspecific, across conditions...) and levels (translation and protein abundance), it is still poorly characterized.

Overlap in the genetic origins of transcript and protein abundance

The final abundance of both mRNA and protein is the consequence of tight genetic regulation. Both GWAS and linkage mapping have been used to explore the genetic origin of gene expression resulting in the discovery of thousands of eQTL and pQTL. The question of whether pQTL mirror eQTL has of course been explored on several occasions, but to date, no clear consensus has been reached. Due to the different correlation levels of the across- and within-gene correlation (Buccitelli and Selbach, 2020), it is puzzling to answer to this question intuitively. On the one hand, the high degree of similarity between mRNA abundances within the sample (*i.e.*, the across-gene correlation) suggests that protein abundance is largely the result of mRNA abundance and therefore the genetic regulation of the proteome could only result from that of the transcriptome. Conversely, the lower and still undetermined mRNA-protein correlation across samples (*i.e.*, the within-gene correlation) could emphasize that the abundance variations between individuals are highly expression layer specific, and thus the genetic origins of the transcriptome and proteome should be as well.

A debated similarity

The simultaneous exploration of mRNA and protein abundance has several prerequisites in addition to those necessary for GWAS or linkage mapping studies. First, the conditions between transcriptomic and proteomic exploration need to be as similar as possible to ensure that the observed gene expression variation is not related to condition biases and is mostly due to genetic variation between individuals. Moreover, GWAS or linkage mapping exploration must be performed on the same cohort of individuals to obtain comparable genetic origins.

A significant number of studies have questioned the overlap between the genetic origin of a protein and its transcript abundance. In yeast, for example, two major studies focused on this issue (Albert et al., 2014; Foss et al., 2007). In both cases, these studies were performed on segregants from a cross between two isolates (a laboratory strain (BY) and a wine strain (RM11)). The early study found that the comparison between eQTL and pQTL was very modest: about 10% of the eQTL also affected the abundance of the related protein. This was consistent with later findings focusing on protein networks, which showed that protein co-regulation across natural yeast isolates was mostly different from mRNA co-regulation (Foss et al., 2011). However, the more recent eQTL/pQTL exploration in yeast (Albert et al., 2014) showed very contrasting results: in this work, more than 60% of the eQTL had a corresponding pQTL. It is important to note that the two studies were technically different, as the latter was

performed and based on a single-cell approach with quantification based on a GFP tag, whereas the previous study used LC-MS/MS quantification.

The similarity between eQTL and pQTL has also been investigated in the mouse model on several occasions (Chick et al., 2016; Ghazalpour et al., 2011). Here, the two studies are technically similar for the protein quantification (based on tandem mass-spectrometry) but slightly different for mRNA quantification (the latest used RNA-seq while the first one used RNA-microarray). It is also worth noting that in one case the eQTL/pQTL was performed on an inbred population (Ghazalpour et al., 2011) while in the other case, an outbred population was used (Chick et al., 2016). Again, there is a large discrepancy between the two studies: in the case of the inbred mouse set, approximately 5 to 6% of the eQTL have a corresponding pQTL, while in the case of the outbred mouse set, this value reaches 33%. Despite the drastically different conclusions of these studies, they agree on the fact that the local eQTL and pQTL tend to be more shared than the distant ones. For example, the overlap between the two types of QTL in Chick et al. (2016) is almost exclusively related to local QTL (1,392 local eQTL out of a total of 1,401 overlapping eQTL). Another contrasting result from these two studies needs to be emphasized: while the two studies had a similar within-gene correlation (between 0.25 and 0.30), suggesting that the mRNA-protein correlation across a population may not be an accurate predictor of the similarities between the genetic origins of mRNA and peptide abundance, one of them highlights that the genes with overlapping eQTL and pQTL tend to have higher within-gene correlations (Chick et al., 2016). Interestingly, a survey on human lymphoblastoid cell lines revealed a similarly high overlap: 33% of the eQTL replicated in the set of pQTL (Battle et al., 2015). Consistent with the eQTL/pQTL overlap, the correspondence between the regulatory hotspots affecting mRNA or protein abundance is often inconsistent across the studies (Albert et al., 2014; Foss et al., 2007), especially since pQTL or eQTL hotspots are not always detected (Ghazalpour et al., 2011).

Limitation of the explorations

The important difference in the overlap between eQTL and pQTL may be due to several reasons. First, the precision of quantification: in the early studies (Foss et al., 2007; Ghazalpour et al., 2011), the quantification of mRNA abundance was based on microarray technology, which is known to be less sensitive to subtle changes in mRNA abundance compared to RNA sequencing (Mantione et al., 2014). However, the study with the higher overlap (≈60%) was

also microarray-based (Albert et al., 2014; Smith and Kruglyak, 2008). Thus, quantification precision is unlikely to be the main reason for the contrasting results.

Another reason that could affect the reliability of the previous studies and therefore cause the observed discrepancies is the size of the populations in which mRNA and protein abundance was monitored. Overall, the more recent the study, the larger the cohort: the first studies on yeast and mice included 94 and 97 individuals, respectively (Foss et al., 2007; Ghazalpour et al., 2011) whereas the later studies included 114 and 192 individuals, respectively (Albert et al., 2014; Chick et al., 2016). The study on human lymphoblastoid cell lines focused on a total of 62 lines (Battle et al., 2015). Despite the tendency to study larger cohorts, these numbers are relatively small compared to the dimensionality of the problem, as only a very small fraction of the genetic diversity of the species is studied. Moreover, the genetic backgrounds of both the yeast and mouse strains in the aforementioned studies suffer from major limitations in terms of natural diversity: the yeast strains were for the most part generated from a simple cross between only two isolates (Albert et al., 2014; Foss et al., 2007), and the 97 and 192 mice are either inbred lines (Ghazalpour et al., 2011) or outbred lines derived from 8 inbred individuals (Chick et al., 2016). Again, this is a major limitation in terms of natural genetic diversity. It is therefore difficult to extend these results to larger scales. Ideally, a reliable approach to explore the similarities between the genetic origins of mRNA and protein abundance at the species level would be based on a larger scale exploration of gene expression, which is now possible as both RNA-seq and LC-MS/MS have been developed to reach population scales (Caudal et al., 2023; Messner et al., 2023, 2022; The GTEx Consortium, 2015). Equally important, to ensure truly reliable genetic diversity, the population studied should consist of natural rather than constructed strains or isolates.

Saccharomyces cerevisiae, a powerful model to explore gene expression variation

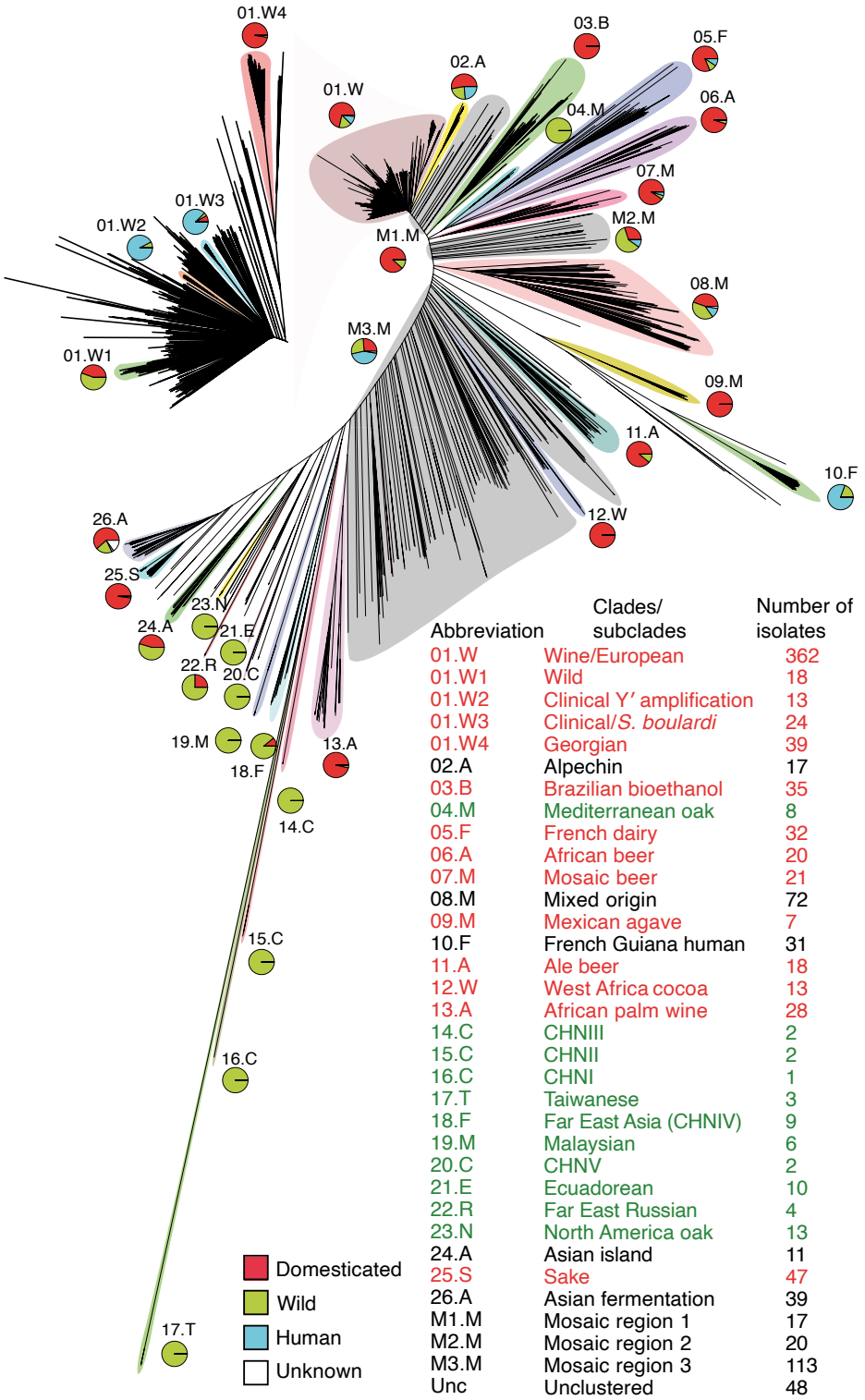
As mentioned in the previous chapter, large-scale exploration of gene expression is a fundamental step in dissecting the genotype-phenotype relationship. Although technological advances have made population-scale studies more accessible, this type of exploration remains laborious. In this context, *S. cerevisiae* is a convenient and safe organism on which most molecular techniques have either been developed or adapted. This yeast is an ascomycetous fungus with a 12 Mb nuclear genome, distributed across 16 chromosomes resulting from an

ancestral whole genome duplication likely caused by an hybridization event and forming an allopolyploid (Marcet-Houben and Gabaldón, 2015; Wolfe and Shields, 1997). The genome is very compact compared to other eukaryotes as 70% of the genome corresponds to coding sequences and only 2% of the protein coding genes have an intron (Hooks et al., 2014). This yeast can be found is a very large diversity of natural and anthropized natural environment. The natural history of this species has obviously been strongly influenced by its extensive use in the context of anthropized fermented substrates (De Guidi et al., 2023).

S. cerevisiae as a model to explore genome, transcriptome and proteome variations

A deeply characterized genome

S. cerevisiae, which was the first eukaryote to be fully sequenced in 1996 (Goffeau et al., 1996), has been a central model organism for biological science. There are more than 6,000 genes in the *S. cerevisiae* genome, although this number varies slightly between isolates (Peter et al., 2018). As of today (June 2023), complete genome sequencing of the *S. cerevisiae* genome with Illumina sequencing has been conducted on more than 3,000 isolates from a vast diversity of geographical and ecological origins (Basile et al., 2021; Duan et al., 2018; Gallone et al., 2016; Lee et al., 2022; Peter et al., 2018; Strope et al., 2015). Among these studies, and to date, the more complete genomic exploration of *S. cerevisiae* has been performed by fully sequencing the genomes of 1,011 natural isolates with Illumina technology (Peter et al., 2018). The studied population includes both wild and domesticated strains, with diverse ecological and geographical origins. Indeed, the strains were sampled from all 5 continents and come from different isolation sources such as clinical, wild (*e.g.*, flower, soil, tree, water), wine, bread, bioethanol production. In this population, more than 1.6 million SNPs and 125,000 indels were detected, highlighting a high nucleotide diversity within this species, reaching up to 1.8% between the most distantly related isolates. The vast majority of SNPs are present at low frequency within the population, as 92% of these polymorphic positions have a MAF less than 5%. The genetic diversity observed within this population results from specific evolutionary events and determinants that have shaped the genomes of the strains during the evolution of the species. Though the construction of a neighbor-joining tree based on the complete dataset of bi-allelic SNPs, 26 subpopulations were identified, recapitulating for most of them the ecological origin of the isolates (Figure 15) and allowing a clear distinction between wild and domesticated subpopulations.



Abbreviation	Clades/ subclades	Number of isolates
01.W	Wine/European	362
01.W1	Wild	18
01.W2	Clinical Y′ amplification	13
01.W3	Clinical/*S. boulardi*	24
01.W4	Georgian	39
02.A	Alpechin	17
03.B	Brazilian bioethanol	35
04.M	Mediterranean oak	8
05.F	French dairy	32
06.A	African beer	20
07.M	Mosaic beer	21
08.M	Mixed origin	72
09.M	Mexican agave	7
10.F	French Guiana human	31
11.A	Ale beer	18
12.W	West Africa cocoa	13
13.A	African palm wine	28
14.C	CHNIII	2
15.C	CHNII	2
16.C	CHNI	1
17.T	Taiwanese	3
18.F	Far East Asia (CHNIV)	9
19.M	Malaysian	6
20.C	CHNV	2
21.E	Ecuadorean	10
22.R	Far East Russian	4
23.N	North America oak	13
24.A	Asian island	11
25.S	Sake	47
26.A	Asian fermentation	39
M1.M	Mosaic region 1	17
M2.M	Mosaic region 2	20
M3.M	Mosaic region 3	113
Unc	Unclustered	48

Figure 15: Neighbor-joining tree of the collection of 1011 *S. cerevisiae* isolates.

This neighbor-joining tree constructed from the biallelic SNPs, highlights the subpopulations and their ecological origins (domesticated, wild, human, and unknown). Figure obtained from Peter et al., 2018

Interestingly, this population also displays an important diversity in terms of CNVs. Indeed, each gene has a duplicated or deleted copy at least once across the 1,011 isolates, and CNVs were highlighted in each isolate. This study allowed to define the pangenome of the species, *i.e.*, the complete set of genes found in the species based on this population, that is composed of 7,796 genes, as well as the core genome, *i.e.*, the set of genes shared by all the isolates, that gathered 4,940 genes. In-between, the 2,856 remaining genes are considered as accessory, as only a fraction, that can be large or small, of the population carries them (Li et al., 2019; Peter et al., 2018). Accessory genes are particularly interesting because they reveal much about the evolutionary history of *S. cerevisiae*. For example, many accessory genes originate from introgressions with the closely related species *Saccharomyces paradoxus* (D'Angiolo et al., 2020; Peter et al., 2018). The uneven distribution of introgressed accessory genes across subpopulations allowed to trace different hybridization events between the two species. The Mexican agave, Alpechin, and French Guyana subpopulations have, on average, more introgressed genes than the other subpopulations. Horizontal gene transfer is also a mechanism that has led to the accumulation of several accessory genes, especially in wine isolates, where several genes coming from *Zygosaccharomyces bailii* and *Torulaspora microellipsoides* confer evolutionary advantage to the recipient isolates, especially in the winemaking environment (Marsit et al., 2015; Novo et al., 2009).

Another fundamental aspect of the *S. cerevisiae* genome is that the isolates show very different levels of ploidy. While the majority of natural isolates are in a diploid state, haploid or higher ploidy levels (3n, 4n and 5n) are common place (Peter et al., 2018). More specifically, some anthropized subpopulations, such as the beer ones, are enriched in polyploids. Among the 1,011 studied isolates, more than 200 isolates have at least one chromosome in an aneuploid state. The aneuploidies are both related to chromosome gain and loss and are unevenly distributed across the genome: the smallest chromosomes, *i.e.*, 1st and 9th, are preferentially affected. Interestingly, despite the known preference for asexual reproduction in *S. cerevisiae*, about 63% of isolates (mainly domesticated isolates) show heterozygosity punctuated by loss of heterozygosity (LOH) regions of varying size, depending on the subpopulation. For example, the Sake subpopulation, that is mostly diploid, has an average of 80% of its genome affected by LOH.

Even if a large diversity of genetic variants was detected in this population, it is important to emphasize that the genotyping relied on a short-read sequencing approach (Peter et al., 2018),

which missed a large fraction of SVs (translocations, inversions, insertions, deletions, etc.). Since this type of variant is known to have a strong influence on yeast phenotype and gene expression (Gorkovskiy and Verstrepen, 2021; Hou et al., 2014; Zimmer et al., 2014), further genomic characterizations using long-read sequencing techniques at large scale will be an essential step to have a comprehensive view of genetic diversity among the *S. cerevisiae* species.

Population-scale gene expression exploration in *S. cerevisiae*

Because of its safe and easy manipulation, combined with rapid growth capabilities and a wide range of available technics for yeast, *S. cerevisiae* is indeed a very good model to study gene expression. As already mentioned, several studies of gene expression have been carried out using *S. cerevisiae*, for each step of gene expression (Albert et al., 2014; Artieri and Fraser, 2014; Brem and Kruglyak, 2005; Foss et al., 2011; Gygi et al., 1999; Khan et al., 2009; McManus et al., 2014; Smith and Kruglyak, 2008). However, the large-scale exploration of gene expression is a very recent advance for this species (Caudal et al., 2023; Messner et al., 2023).

The 1,011 population described above (Peter et al., 2018) was recently used for an extensive transcriptomic survey (Caudal et al., 2023). In this study, high-quality transcriptomes of 969 isolates were generated in a synthetic complete medium. The detected transcripts comprised 6,445 open reading frames (ORF), of which 4,977 belonged to the core genome and 1,468 to the accessory genome. This is one of the largest population transcriptome explorations to date. Due to its very large scale, this study allowed for an in-depth characterization of the transcriptome variation between individuals. Surprisingly, the results showed that, overall, accessory genes have a very specific transcriptional behavior, being less expressed than the other genes, but more variable in expression across the individuals. However, this behavior was also related to the type of accessory genes. For example, there was a large variability in the mRNA abundance of accessory ORF originating from HGT evens across the population, while this was not the case for the ORF that were acquired through introgression events. In this particular case, it was even possible to compare the expression of the genes introgressed from *S. paradoxus* with the expression of their orthologs in the context of allelic heterozygosity (*i.e.*, for an ORF, an isolate has one copy of the *S. cerevisiae* allele and one copy of the *S. paradoxus* allele). Using allele-specific expression, no difference was found between the expression of the

S. paradoxus allele and the *S. cerevisiae* allele. By focusing on the subpopulation level, it was possible to detect specific differentially expressed genes (DEG) for most subpopulations (Caudal et al., 2023). The DEG were strongly associated with the environment from which the strains were isolated. For example, the *GAL* pathway was overexpressed in dairy fermentation-related isolates, even in the presence of glucose. This type of metabolic switch represents a key adaptation to lactose-rich media (Boocock et al., 2021; Duan et al., 2019), and highlights the central role of gene expression modulation in adaptation to anthropized processes. Finally, this investigation allowed for a precise association between mRNA abundance and genetic variants (SNPs and CNVs) present in this population. Using GWAS, it was possible to detect 7,273 SNP-eQTL and 2,197 CNV-eQTL affecting a total of 3,471 genes. Several fundamental aspects underlying the variation in mRNA abundance were observed with this GWAS. First, both *cis* and *trans*-eQTL were detected, with *cis* having a larger effect on their target genes. Regarding the CNV-eQTL, many of them were in fact related to the aneuploidies of the 1st, 3rd, 8th, 9th, and 11th chromosomes. Interestingly, a difference was found between the effect of CNV-eQTL and SNP-eQTL, with the latter having a more important effect on mRNA abundance.

On the proteomic side, a recent study was carried out on the same set of isolates (Muenzner et al., 2022): the proteomes of 613 isolates grown on a synthetic minimal medium were accurately monitored using a high-throughput LC-MS/MS, resulting in the quantification of 1,563 proteins. The main focus of this study was to investigate the effect of aneuploidies on the proteome and to compare this with the transcriptome. Consistently with the post-transcriptional buffering phenomenon, a general dosage compensation was observed at the proteome level. In fact, the mRNA-abundance reflected more accurately the chromosomal imbalance than the protein abundance. In addition, the ubiquitin-proteasome system was likely a major process in variation buffering. Yet, due to the differences in the culture conditions, it is difficult to extend the comparison between the population's proteome and the transcriptome to other aspects of the gene expression population, as the medium difference could be a major confounding factor, especially when comparing GWAS results. Therefore, a large-scale and exact comparison between the population transcriptome and proteome is still lacking.

The S. cerevisiae domestication and its consequences on its evolutionnary history

The aforementioned investigations of both the genomes and transcriptomes of the *S. cerevisiae* population have emphasized the drastic impact of domestication on the *S. cerevisiae* evolutionnary history. As this species is likely to be the most widely used in food making because of its efficient fermentation capabilities, the domestication of *S. cerevisiae* has been extensively studied. However, due to the large number of strains that are used in different industrial contexts, the impact of domestication, particularly at the molecular level, have yet to be fully characterized.

Domestication and industrial use of *S. cerevisiae*

Evidence of deliberate fermentation has been found in early human history, dating back to prehistorical time (Gallone et al., 2016; McGovern et al., 2004; Michel et al., 1992; Samuel, 1996). The clear detection and association between yeast and fermentation came as modern science was in its early stage, when Louis Pasteur described the role of *S. cerevisiae* in alcoholic fermentation (Pasteur, 1858). The case of the domestication of microorganisms is somewhat peculiar, since human selection, until the beginning of the last century, was mainly based on the indirect assessment of metabolic capacities, whereas the domestication and selection of animals or plants is based on visual and more quantifiable phenotypes (*e.g.*, biomass production, size) (De Guidi et al., 2023). Because of this, it is expected that the domesticated strains will have very specific metabolic capacities depending on their isolation origin.

In the 1,011-population described above, the ecological backgrounds of the domesticated isolates are diverse (Figure 15, 16) (Peter et al., 2018). Alcoholic beverages are the main source of domesticated isolates. As observed through the structure of the population: the largest subpopulation is related to European wine isolates and 3 subpopulations mainly include beer isolates (African beer, mosaic beer, ale beer). In addition, 3 subpopulations include strains from other alcoholic beverages (sake, African palm wine and Mexican agave-related beverage). Isolates related to food fermentation also grouped in specific subpopulations such as the ones related to cheese making (French dairy) while bakery related strain are mostly dispersed within the population. Finally, 35 domesticated isolates come from bioethanol production sites and are grouped in a specific subpopulation (Brazilian bioethanol). It is worth noting that, although being selected for fermentative purposes, the ecological niches of the domesticated isolates are very different. The microbial interactions, the temperature, the consistency of the medium and many other fundamental factors that are known to affect cell biology are specific to each

ecological origin. Therefore, characterization of the molecular impact of domestication may require subpopulation-specific investigations.

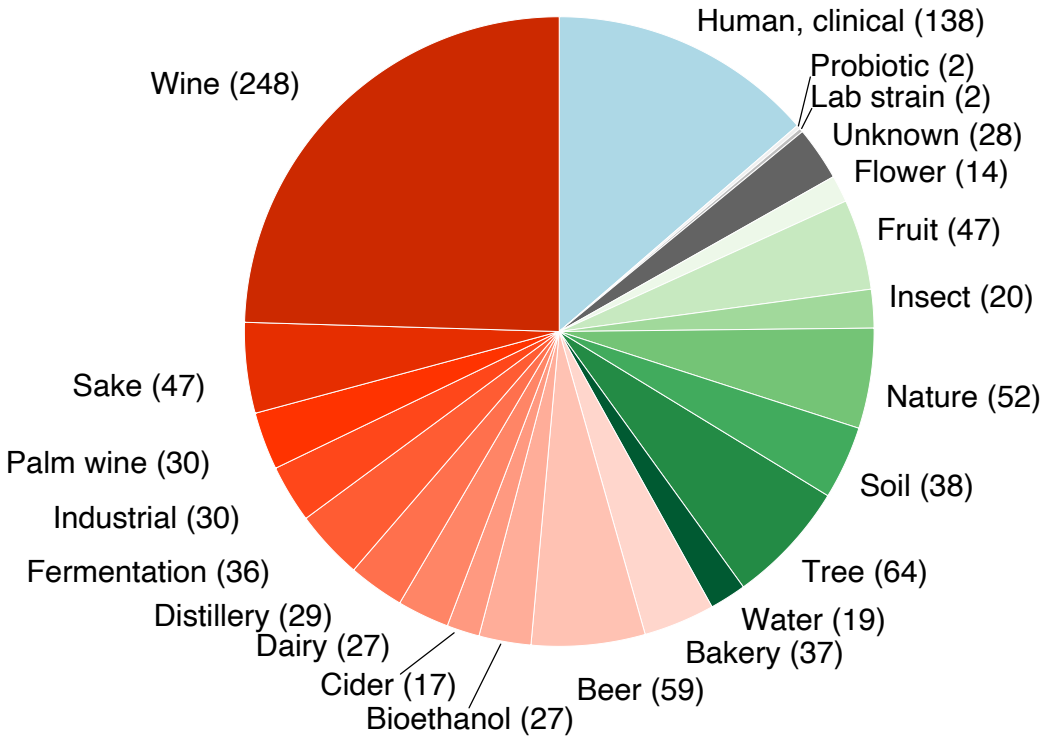


Figure 16: Environmental origin of the 1011-population.

The colors correspond to domesticated (shades of red), wild (shade of green) or clinical (blue) ecological environments. The domesticated environments comprise very different trophic and high condition diversity. Adapted from Peter et al., 2018.

It is still unclear whether a single or multiple events led to the domestication of *S. cerevisiae*. The general consensus states that the species originated from eastern China (Duan et al., 2018; Peter et al., 2018). However, while the 1,011-isolates collection supports the fact that multiple independent events shaped the domestication of *S. cerevisiae* (Peter et al., 2018), a study of Asian strains showing that domesticated isolates form two main groups (depending on the liquid or solid state on which they were isolated) suggests that domestication resulted from a single bottleneck event (Duan et al., 2018). Regarding the Chinese domesticated isolates, their population structure indeed seems to support the bottleneck hypothesis (Duan et al., 2018). However, the larger study of the 1,011 isolates includes a greater genetic diversity and may therefore be more reliable in accurately describing the domestication history of the species as a whole.

Phenotypic and molecular impact of the domestication

Domestication usually results in a profound modification of the phenotypic landscape of a species. In the case of yeast, a wide range of trait variation can be observed when comparing wild and domesticated isolates. On the metabolic side, the fermentative capacity of the domesticated isolates are increased (Bell et al., 2001) and, more globally, there is a shift towards fermentation rather than respiration (Lahue et al., 2020). Interestingly, each domesticated subpopulation has undergone specific metabolic evolution due to the diversity of the ecological niches. As mentioned above, the diary subpopulation has a particular shift between glucose and galactose metabolism, which conferred the cell an increased ability to ferment lactose (Boocock et al., 2021; Caudal et al., 2023; Duan et al., 2019). Beer isolates have on their side a better fitness when grown with maltose as a carbon source (Gallone et al., 2016).

However, the adaptation of carbon or energy metabolism is only a small part of the drastic changes that result from domestication. Several other cellular or molecular phenotypes have been associated with the domesticated isolates of *S. cerevisiae*. The general signatures (other than metabolic adaptation) of the domesticated isolate include improved osmotic stress tolerance and reduced sporulation (despite the higher proportion of heterozygotes) (De Guidi et al., 2023). The Sake isolates for instance went through morphological changes (Ohnuki et al., 2017) and are now highly resistant to high ethanol concentration (Shiroma et al., 2014; Watanabe et al., 2011). Stress resistance is also an important domestication trait in the wine subpopulation, as these isolates tend to be more resistant to the presence of copper or sulfite (Brandolini et al., 2002; Yuasa et al., 2004). Interestingly, several SVs underlie this increase in stress resistance in wine isolates. For copper tolerance, this is associated to an increased copy number of the *CUP1* gene (Fogel and Welch, 1982; Peter et al., 2018; Steenwyk and Rokas, 2018). As mentioned above, the sulfite resistance is associated to large chromosomal rearrangements such as translocations and inversions (García-Ríos et al., 2019; Pérez-Ortín et al., 2002; Yuasa et al., 2004; Zimmer et al., 2014). Chromosomal alterations is also frequently observed in beer isolates, as many of them are polyploids (3n, 4n and 5n) (Gallone et al., 2016; Peter et al., 2018; Saada et al., 2022), which is interesting as high ploidy levels are also a hallmark of plant domestication (Purugganan and Fuller, 2009).

References

Aguet, F., Alasoo, K., Li, Y.I., Battle, A., Im, H.K., Montgomery, S.B., Lappalainen, T., 2023. Molecular quantitative trait loci. Nat. Rev. Methods Primer 3, 1–22. https://doi.org/10.1038/s43586-022-00188-6

Akiyama, M., Ishigaki, K., Sakaue, S., Momozawa, Y., Horikoshi, M., Hirata, M., Matsuda, K., Ikegawa, S., Takahashi, A., Kanai, M., Suzuki, S., Matsui, D., Naito, M., Yamaji, T., Iwasaki, M., Sawada, N., Tanno, K., Sasaki, M., Hozawa, A., Minegishi, N., Wakai, K., Tsugane, S., Shimizu, A., Yamamoto, M., Okada, Y., Murakami, Y., Kubo, M., Kamatani, Y., 2019. Characterizing rare and low-frequency height-associated variants in the Japanese population. Nat. Commun. 10, 4393. https://doi.org/10.1038/s41467-019-12276-5

Alam, M.T., Zelezniak, A., Mülleder, M., Shliaha, P., Schwarz, R., Capuano, F., Vowinckel, J., Radmaneshfar, E., Krüger, A., Calvani, E., Michel, S., Börno, S., Christen, S., Patil, K.R., Timmermann, B., Lilley, K.S., Ralser, M., 2016. The metabolic background is a global player in Saccharomyces gene expression epistasis. Nat. Microbiol. 1, 1–10. https://doi.org/10.1038/nmicrobiol.2015.30

Albert, F.W., Bloom, J.S., Siegel, J., Day, L., Kruglyak, L., 2018. Genetics of trans-regulatory variation in gene expression. eLife 7, e35471. https://doi.org/10.7554/eLife.35471

Albert, F.W., Treusch, S., Shockley, A.H., Bloom, J.S., Kruglyak, L., 2014. Genetics of single-cell protein abundance variation in large yeast populations. Nature 506, 494–497. https://doi.org/10.1038/nature12904

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T.H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A.L., Tieman, D.M., Klee, H., Kirsche, M., Aganezov, S., Ranallo-Benavidez, T.R., Lemmon, Z.H., Kim, J., Robitaille, G., Kramer, M., Goodwin, S., McCombie, W.R., Hutton, S., Van Eck, J., Gillis, J., Eshed, Y., Sedlazeck, F.J., van der Knaap, E., Schatz, M.C., Lippman, Z.B., 2020. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell 182, 145-161.e23. https://doi.org/10.1016/j.cell.2020.05.021

Altshuler, D., Daly, M.J., Lander, E.S., 2008. Genetic mapping in human disease. Science 322, 881–888. https://doi.org/10.1126/science.1156409

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., Herschlag, D., 2003. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. 100, 3889–3894. https://doi.org/10.1073/pnas.0635171100

Archer, T.C., Ehrenberger, T., Mundt, F., Gold, M.P., Krug, K., Mah, C.K., Mahoney, E.L., Daniel, C.J., LeNail, A., Ramamoorthy, D., Mertins, P., Mani, D.R., Zhang, H., Gillette, M.A., Clauser, K., Noble, M., Tang, L.C., Pierre-François, J., Silterra, J., Jensen, J., Tamayo, P., Korshunov, A., Pfister, S.M., Kool, M., Northcott, P.A., Sears, R.C., Lipton, J.O., Carr, S.A., Mesirov, J.P., Pomeroy, S.L., Fraenkel, E., 2018. Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups. Cancer Cell 34, 396-410.e8. https://doi.org/10.1016/j.ccell.2018.08.004

Artieri, C.G., Fraser, H.B., 2014. Evolution at two levels of gene expression in yeast. Genome Res. 24, 411–421. https://doi.org/10.1101/gr.165522.113

Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., Warren, W.C., Magrini, V., McGrath, S.D., Li, Y.I., Wilson, R.K., Eichler, E.E., 2019. Characterizing the Major Structural Variant Alleles of the Human Genome. Cell 176, 663-675.e19. https://doi.org/10.1016/j.cell.2018.12.019

Aydin, S., Pham, D.T., Zhang, T., Keele, G.R., Skelly, D.A., Paulo, J.A., Pankratz, M., Choi, T., Gygi, S.P., Reinholdt, L.G., Baker, C.L., Churchill, G.A., Munger, S.C., 2023. Genetic dissection of the pluripotent proteome through multi-omics data integration. Cell Genomics 0. https://doi.org/10.1016/j.xgen.2023.100283

Azrolan, N., Breslow, J.L., 1990. A solution hybridization/RNase protection assay with riboprobes to determine absolute levels of apoB, A-I, and E mRNA in human hepatoma cell lines. J. Lipid Res. 31, 1141–1146. https://doi.org/10.1016/S0022-2275(20)42754-3

Balding, D.J., 2006. A tutorial on statistical methods for population association studies. Nat. Rev. Genet. 7, 781–791. https://doi.org/10.1038/nrg1916

Bartoszewski, R.A., Jablonsky, M., Bartoszewska, S., Stevenson, L., Dai, Q., Kappes, J., Collawn, J.F., Bebok, Z., 2010. A Synonymous Single Nucleotide Polymorphism in ΔF508 CFTR Alters the Secondary Structure of the mRNA and the Expression of the Mutant Protein. J. Biol. Chem. 285, 28741–28748. https://doi.org/10.1074/jbc.M110.154575

Basile, A., De Pascale, F., Bianca, F., Rossi, A., Frizzarin, M., De Bernardini, N., Bosaro, M., Baldisseri, A., Antoniali, P., Lopreiato, R., Treu, L., Campanaro, S., 2021. Large-scale sequencing and comparative analysis of oenological Saccharomyces cerevisiae strains supported by nanopore refinement of key genomes. Food Microbiol. 97, 103753. https://doi.org/10.1016/j.fm.2021.103753

Bateson, W., Bateson, W., Mendel, G., Leighton, A.G., 1909. Mendel's principles of heredity, by W. Bateson. University Press, Cambridge [Eng.]. https://doi.org/10.5962/bhl.title.1057

Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., Gilad, Y., 2015. Impact of regulatory variation from RNA to protein. Science 347, 664–667. https://doi.org/10.1126/science.1260793

Baum, K., Schuchhardt, J., Wolf, J., Busse, D., 2019. Of Gene Expression and Cell Division Time: A Mathematical Framework for Advanced Differential Gene Expression and Data Analysis. Cell Syst. 9, 569-579.e7. https://doi.org/10.1016/j.cels.2019.07.009

Bazopoulou-Kyrkanidou, E., 1992. Genetic concepts in Greek literature from the eighth to the fourth century B.C. Hum. Genet. 88, 500–507. https://doi.org/10.1007/BF00219335

Beadle, G.W., Tatum, E.L., 1941. Genetic Control of Biochemical Reactions in Neurospora. Proc. Natl. Acad. Sci. 27, 499–506. https://doi.org/10.1073/pnas.27.11.499

Becker, K., Bluhm, A., Casas-Vila, N., Dinges, N., Dejung, M., Sayols, S., Kreutz, C., Roignant, J.-Y., Butter, F., Legewie, S., 2018. Quantifying post-transcriptional regulation in the development of Drosophila melanogaster. Nat. Commun. 9, 4970. https://doi.org/10.1038/s41467-018-07455-9

Becker-André, M., Hahlbrock, K., 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcipt titration assay (PATTY). Nucleic Acids Res. 17, 9437–9446. https://doi.org/10.1093/nar/17.22.9437

Bell, P.J., Higgins, V.J., Attfield, P.V., 2001. Comparison of fermentative capacities of industrial baking and wild-type yeasts of the species Saccharomyces cerevisiae in different sugar media. Lett. Appl. Microbiol. 32, 224–229. https://doi.org/10.1046/j.1472-765x.2001.00894.x

Blevins, W.R., Ruiz-Orera, J., Messeguer, X., Blasco-Moreno, B., Villanueva-Cañas, J.L., Espinar, L., Díez, J., Carey, L.B., Albà, M.M., 2021. Uncovering de novo gene birth in yeast using deep transcriptomics. Nat. Commun. 12, 604. https://doi.org/10.1038/s41467-021-20911-3

Blevins, W.R., Tavella, T., Moro, S.G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., Carey, L.B., Albà, M.M., 2019. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Sci. Rep. 9, 11005. https://doi.org/10.1038/s41598-019-47424-w

Bloom, J.S., Boocock, J., Treusch, S., Sadhu, M.J., Day, L., Oates-Barker, H., Kruglyak, L., 2019. Rare variants contribute disproportionately to quantitative trait variation in yeast. eLife 8, e49212. https://doi.org/10.7554/eLife.49212

Bodmer, W., Bonilla, C., 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nat. Genet. 40, 695–701. https://doi.org/10.1038/ng.f.136

Bokhari, A., Jonchere, V., Lagrange, A., Bertrand, R., Svrcek, M., Marisa, L., Buhard, O., Greene, M., Demidova, A., Jia, J., Adriaenssens, E., Chassat, T., Biard, D.S., Flejou, J.-F., Lejeune, F., Duval, A., Collura, A., 2018. Targeting nonsense-mediated mRNA decay in colorectal cancers with microsatellite instability. Oncogenesis 7, 70. https://doi.org/10.1038/s41389-018-0079-x

Boocock, J., Sadhu, M.J., Durvasula, A., Bloom, J.S., Kruglyak, L., 2021. Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. Science 371, 415–419. https://doi.org/10.1126/science.aba0542

Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314–331.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Brandolini, V., Tedeschi, P., Capece, A., Maietti, A., Mazzotta, D., Salzano, G., Paparella, A., Romano, P., 2002. Saccharomyces cerevisiae wine strains differing in copper resistance exhibit different capability to reduce copper content in wine. World J. Microbiol. Biotechnol. 18, 499–503. https://doi.org/10.1023/A:1016306813502

Brar, G.A., Weissman, J.S., 2015. Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat. Rev. Mol. Cell Biol. 16, 651–664. https://doi.org/10.1038/nrm4069
Breker, M., Gymrek, M., Schuldiner, M., 2013. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. J. Cell Biol. 200, 839–850. https://doi.org/10.1083/jcb.201301120
Brem, R.B., Kruglyak, L., 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc. Natl. Acad. Sci. 102, 1572–1577. https://doi.org/10.1073/pnas.0408709102
Brem, R.B., Yvert, G., Clinton, R., Kruglyak, L., 2002. Genetic dissection of transcriptional regulation in budding yeast. Science 296, 752–755. https://doi.org/10.1126/science.1069516
Brenner, S., 1974. The genetics of Caenorhabditis elegans. Genetics 77, 71–94. https://doi.org/10.1093/genetics/77.1.71
Bromberg, Y., Rost, B., 2009. Correlating protein function and stability through the analysis of single amino acid substitutions. BMC Bioinformatics 10, S8. https://doi.org/10.1186/1471-2105-10-S8-S8
Buccitelli, C., Selbach, M., 2020. mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 21, 630–644. https://doi.org/10.1038/s41576-020-0258-4
Bumgarner, R., 2013. DNA microarrays: Types, Applications and their future. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al 0 22, Unit-22.1. https://doi.org/10.1002/0471142727.mb2201s101
Bunn, H.F., 1997. Pathogenesis and Treatment of Sickle Cell Disease. N. Engl. J. Med. 337, 762–769. https://doi.org/10.1056/NEJM199709113371107
Burnette, W.N., 1981. "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. 112, 195–203. https://doi.org/10.1016/0003-2697(81)90281-5
Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endocrinol. 25, 169–193. https://doi.org/10.1677/jme.0.0250169
Byron, S.A., Van Keuren-Jensen, K.R., Engelthaler, D.M., Carpten, J.D., Craig, D.W., 2016. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat. Rev. Genet. 17, 257–271. https://doi.org/10.1038/nrg.2016.10
Casadio, R., Vassura, M., Tiwari, S., Fariselli, P., Luigi Martelli, P., 2011. Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. Hum. Mutat. 32, 1161–1170. https://doi.org/10.1002/humu.21555
Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonniere, E., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., 2023. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. https://doi.org/10.1101/2023.05.17.541122
Chalancon, G., Ravarani, C.N.J., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., Babu, M.M., 2012. Interplay between gene expression noise and regulatory network architecture. Trends Genet. 28, 221–232. https://doi.org/10.1016/j.tig.2012.01.006
Chapman, J.D., Goodlett, D.R., Masselon, C.D., 2014. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. Mass Spectrom. Rev. 33, 452–470. https://doi.org/10.1002/mas.21400
Cheng, F., Zhao, J., Wang, Y., Lu, W., Liu, Z., Zhou, Y., Martin, W.R., Wang, R., Huang, J., Hao, T., Yue, H., Ma, J., Hou, Y., Castrillon, J.A., Fang, J., Lathia, J.D., Keri, R.A., Lightstone, F.C., Antman, E.M., Rabadan, R., Hill, D.E., Eng, C., Vidal, M., Loscalzo, J., 2021. Comprehensive characterization of protein–protein interactions perturbed by disease mutations. Nat. Genet. 53, 342–353. https://doi.org/10.1038/s41588-020-00774-y
Chick, J.M., Munger, S.C., Simecek, P., Huttlin, E.L., Choi, K., Gatti, D.M., Raghupathy, N., Svenson, K.L., Churchill, G.A., Gygi, S.P., 2016. Defining the consequences of genetic variation on a proteome-wide scale. Nature 534, 500–505. https://doi.org/10.1038/nature18270
Chong, Y.T., Koh, J.L.Y., Friesen, H., Duffy, S.K., Cox, M.J., Moses, A., Moffat, J., Boone, C., Andrews, B.J., 2015. Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. Cell 161, 1413–1424. https://doi.org/10.1016/j.cell.2015.04.051
Cirulli, E.T., Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat. Rev. Genet. 11, 415–425. https://doi.org/10.1038/nrg2779
Collaborative Cross Consortium, 2012. The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics 190, 389–401. https://doi.org/10.1534/genetics.111.132639
Condò, I., 2022. Rare Monogenic Diseases: Molecular Pathophysiology and Novel Therapies. Int. J. Mol. Sci. 23, 6525. https://doi.org/10.3390/ijms23126525

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13. https://doi.org/10.1186/s13059-016-0881-8

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823. https://doi.org/10.1126/science.1231143

Coolon, J.D., Stevenson, K.R., McManus, C.J., Yang, B., Graveley, B.R., Wittkopp, P.J., 2015. Molecular Mechanisms and Evolutionary Processes Contributing to Accelerated Divergence of Gene Expression on the Drosophila X Chromosome. Mol. Biol. Evol. 32, 2605–2615. https://doi.org/10.1093/molbev/msv135

Coonen, L.P., 1952. Books, Battles, and Biology. Sci. Mon. 74, 211–217.

Corbett, A.H., 2018. Post-transcriptional regulation of gene expression and human disease. Curr. Opin. Cell Biol. 52, 96–104. https://doi.org/10.1016/j.ceb.2018.02.011

Corchete, L.A., Rojas, E.A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N.C., Burguillo, F.J., 2020. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. Sci. Rep. 10, 19737. https://doi.org/10.1038/s41598-020-76881-x

D'Angiolo, M., De Chiara, M., Yue, J.-X., Irizar, A., Stenberg, S., Persson, K., Llored, A., Barré, B., Schacherer, J., Marangoni, R., Gilson, E., Warringer, J., Liti, G., 2020. A yeast living ancestor reveals the origin of genomic introgressions. Nature 587, 420–425. https://doi.org/10.1038/s41586-020-2889-1

Das, P.M., Singal, R., 2004. DNA Methylation and Cancer. J. Clin. Oncol. 22, 4632–4642. https://doi.org/10.1200/JCO.2004.07.151

Davidson, G.S., Joe, R.M., Roy, S., Meirelles, O., Allen, C.P., Wilson, M.R., Tapia, P.H., Manzanilla, E.E., Dodson, A.E., Chakraborty, S., Carter, M., Young, S., Edwards, B., Sklar, L., Werner-Washburne, M., 2011. The proteomics of quiescent and nonquiescent cell differentiation in yeast stationary-phase cultures. Mol. Biol. Cell 22, 988–998. https://doi.org/10.1091/mbc.E10-06-0499

de Bie, P., Ciechanover, A., 2011. Ubiquitination of E3 ligases: self-regulation of the ubiquitin system via proteolytic and non-proteolytic mechanisms. Cell Death Differ. 18, 1393–1402. https://doi.org/10.1038/cdd.2011.16

De Guidi, I., Legras, J.-L., Galeote, V., Sicard, D., 2023. Yeast domestication in fermented food and beverages: past research and new avenues. Curr. Opin. Food Sci. 51, 101032. https://doi.org/10.1016/j.cofs.2023.101032

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., Stephens, M., Gilad, Y., Pritchard, J.K., 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 482, 390–394. https://doi.org/10.1038/nature10808

Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., Kumar, S., Ongen, H., Popadin, K., Marbach, D., Ambrosini, G., Bielser, D., Hacker, D., Romano, L., Ribaux, P., Wiederkehr, M., Falconnet, E., Bucher, P., Bergmann, S., Antonarakis, S.E., Reymond, A., Dermitzakis, E.T., 2019. Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science 364, eaat8266. https://doi.org/10.1126/science.aat8266

Demichelis, F., Setlur, S.R., Banerjee, S., Chakravarty, D., Chen, J.Y.H., Chen, C.X., Huang, J., Beltran, H., Oldridge, D.A., Kitabayashi, N., Stenzel, B., Schaefer, G., Horninger, W., Bektic, J., Chinnaiyan, A.M., Goldenberg, S., Siddiqui, J., Regan, M.M., Kearney, M., Soong, T.D., Rickman, D.S., Elemento, O., Wei, J.T., Scherr, D.S., Sanda, M.A., Bartsch, G., Lee, C., Klocker, H., Rubin, M.A., 2012. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. Proc. Natl. Acad. Sci. U. S. A. 109, 6686–6691. https://doi.org/10.1073/pnas.1117405109

Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., Ralser, M., 2020. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods 17, 41–44. https://doi.org/10.1038/s41592-019-0638-x

Dénervaud, N., Becker, J., Delgado-Gonzalo, R., Damay, P., Rajkumar, A.S., Unser, M., Shore, D., Naef, F., Maerkl, S.J., 2013. A chemostat array enables the spatio-temporal analysis of the yeast proteome. Proc. Natl. Acad. Sci. U. S. A. 110, 15842–15847. https://doi.org/10.1073/pnas.1308265110

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., Torres, E.M., 2014. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife 3, e03023. https://doi.org/10.7554/eLife.03023

Dermit, M., Dodel, M., Mardakheh, F.K., 2017. Methods for monitoring and measurement of protein translation in time and space. Mol. Biosyst. 13, 2477–2488. https://doi.org/10.1039/c7mb00476a

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., Giaever, G., 2005. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. Genetics 169, 1915–1925. https://doi.org/10.1534/genetics.104.036871

Dewan, A., Liu, M., Hartman, S., Zhang, S.S.-M., Liu, D.T.L., Zhao, C., Tam, P.O.S., Chan, W.M., Lam, D.S.C., Snyder, M., Barnstable, C., Pang, C.P., Hoh, J., 2006. HTRA1 promoter polymorphism in wet age-related macular degeneration. Science 314, 989–992. https://doi.org/10.1126/science.1133807

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., Gingeras, T.R., 2012. Landscape of transcription in human cells. Nature 489, 101–108. https://doi.org/10.1038/nature11233

Doebele, R.C., Davis, L.E., Vaishnavi, A., Le, A.T., Estrada-Bernal, A., Keysar, S., Jimeno, A., Varella-Garcia, M., Aisner, D.L., Li, Y., Stephens, P.J., Morosini, D., Tuch, B.B., Fernandes, M., Nanda, N., Low, J.A., 2015. An Oncogenic NTRK Fusion in a Patient with Soft-Tissue Sarcoma with Response to the Tropomyosin-Related Kinase Inhibitor LOXO-101. Cancer Discov. 5, 1049–1057. https://doi.org/10.1158/2159-8290.CD-15-0443

Doudna, J.A., Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096. https://doi.org/10.1126/science.1258096

Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., Bai, F.-Y., 2018. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nat. Commun. 9, 2690. https://doi.org/10.1038/s41467-018-05106-7

Duan, S.-F., Shi, J.-Y., Yin, Q., Zhang, R.-P., Han, P.-J., Wang, Q.-M., Bai, F.-Y., 2019. Reverse Evolution of a Classic Gene Network in Yeast Offers a Competitive Advantage. Curr. Biol. 29, 1126-1136.e5. https://doi.org/10.1016/j.cub.2019.02.038

Dudbridge, F., Gusnanto, A., 2008. Estimation of significance thresholds for genomewide association scans. Genet. Epidemiol. 32, 227–234. https://doi.org/10.1002/gepi.20297

Duncan, L., Yilmaz, Z., Gaspar, H., Walters, R., Goldstein, J., Anttila, V., Bulik-Sullivan, B., Ripke, S., Eating Disorders Working Group of the Psychiatric Genomics Consortium, Thornton, L., Hinney, A., Daly, M., Sullivan, P.F., Zeggini, E., Breen, G., Bulik, C.M., 2017. Significant Locus and Metabolic Genetic Correlations Revealed in Genome-Wide Association Study of Anorexia Nervosa. Am. J. Psychiatry 174, 850–858. https://doi.org/10.1176/appi.ajp.2017.16121402

Edfors, F., Danielsson, F., Hallström, B.M., Käll, L., Lundberg, E., Pontén, F., Forsström, B., Uhlén, M., 2016. Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol. Syst. Biol. 12, 883. https://doi.org/10.15252/msb.20167144

Emerson, J.J., Hsieh, L.-C., Sung, H.-M., Wang, T.-Y., Huang, C.-J., Lu, H.H.-S., Lu, M.-Y.J., Wu, S.-H., Li, W.-H., 2010. Natural selection on cis and trans regulation in yeasts. Genome Res. 20, 826–836. https://doi.org/10.1101/gr.101576.109

Emrich, S.J., Barbazuk, W.B., Li, L., Schnable, P.S., 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res. 17, 69–73. https://doi.org/10.1101/gr.5145806

Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., Järvelä, I., 2002. Identification of a variant associated with adult-type hypolactasia. Nat. Genet. 30, 233–237. https://doi.org/10.1038/ng826

Fang, L., Ahn, J.K., Wodziak, D., Sibley, E., 2012. The human lactase persistence-associated SNP -13910*T enables in vivo functional persistence of lactase promoter-reporter transgene expression. Hum. Genet. 131, 1153–1159. https://doi.org/10.1007/s00439-012-1140-z

Fay, J.C., 2013. The molecular basis of phenotypic variation in yeast. Curr. Opin. Genet. Dev. 23, 672–677. https://doi.org/10.1016/j.gde.2013.10.005

Fehrmann, R.S.N., Karjalainen, J.M., Krajewska, M., Westra, H.-J., Maloney, D., Simeonov, A., Pers, T.H., Hirschhorn, J.N., Jansen, R.C., Schultes, E.A., van Haagen, H.H.H.B.M., de Vries, E.G.E., te Meerman, G.J., Wijmenga, C., van Vugt, M.A.T.M., Franke, L., 2015. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat. Genet. 47, 115–125. https://doi.org/10.1038/ng.3173

Félix, M.-A., Barkoulas, M., 2015. Pervasive robustness in biological systems. Nat. Rev. Genet. 16, 483–496. https://doi.org/10.1038/nrg3949

Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., Jensson, B.O., Zink, F., Halldorsson, G.H., Masson, G., Arnadottir, G.A., Katrinardottir, H., Juliusson, K., Magnusson, M.K., Magnusson, O.T., Fridriksdottir, R., Saevarsdottir, S., Gudjonsson, S.A., Stacey, S.N., Rognvaldsson, S., Eiriksdottir, T., Olafsdottir, T.A., Steinthorsdottir, V., Tragante, V., Ulfarsson, M.O., Stefansson, H., Jonsdottir, I., Holm, H., Rafnar, T., Melsted, P., Saemundsdottir, J., Norddahl, G.L., Lund, S.H., Gudbjartsson, D.F., Thorsteinsdottir, U., Stefansson, K., 2021. Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. 53, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w

Flint, J., Valdar, W., Shifman, S., Mott, R., 2005. Strategies for mapping and cloning quantitative trait genes in rodents. Nat. Rev. Genet. 6, 271–286. https://doi.org/10.1038/nrg1576

Fogel, S., Welch, J.W., 1982. Tandem gene amplification mediates copper resistance in yeast. Proc. Natl. Acad. Sci. U. S. A. 79, 5342–5346. https://doi.org/10.1073/pnas.79.17.5342

Foley, K.P., Leonard, M.W., Engel, J.D., 1993. Quantitation of RNA using the polymerase chain reaction. Trends Genet. 9, 380–385. https://doi.org/10.1016/0168-9525(93)90137-7

Fortelny, N., Overall, C.M., Pavlidis, P., Freue, G.V.C., 2017. Can we predict protein from mRNA levels? Nature 547, E19–E20. https://doi.org/10.1038/nature22293

Foss, E.J., Radulovic, D., Shaffer, S.A., Goodlett, D.R., Kruglyak, L., Bedalov, A., 2011. Genetic Variation Shapes Protein Networks Mainly through Non-transcriptional Mechanisms. PLOS Biol. 9, e1001144. https://doi.org/10.1371/journal.pbio.1001144

Foss, E.J., Radulovic, D., Shaffer, S.A., Ruderfer, D.M., Bedalov, A., Goodlett, D.R., Kruglyak, L., 2007. Genetic basis of proteome variation in yeast. Nat. Genet. 39, 1369–1375. https://doi.org/10.1038/ng.2007.22

Fournier, M.L., Paulson, A., Pavelka, N., Mosley, A.L., Gaudenz, K., Bradford, W.D., Glynn, E., Li, H., Sardiu, M.E., Fleharty, B., Seidel, C., Florens, L., Washburn, M.P., 2010. Delayed Correlation of mRNA and Protein Expression in Rapamycin-treated Cells and a Role for Ggc1 in Cellular Sensitivity to Rapamycin*. Mol. Cell. Proteomics 9, 271–284. https://doi.org/10.1074/mcp.M900415-MCP200

Fournier, T., Abou Saada, O., Hou, J., Peter, J., Caudal, E., Schacherer, J., 2019. Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. eLife 8, e49258. https://doi.org/10.7554/eLife.49258

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I.T., García Girón, C., Gonzalez, J.M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O.G., Lagarde, J., Martin, F.J., Martínez, L., Mohanan, S., Muir, P., Navarro, F.C.P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B.M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J.S., Gerstein, M., Guigó, R., Hubbard, T.J.P., Kellis, M., Paten, B., Reymond, A., Tress, M.L., Flicek, P., 2019. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 47, D766–D773. https://doi.org/10.1093/nar/gky955

Friedman, D.B., Hill, S., Keller, J.W., Merchant, N.B., Levy, S.E., Coffey, R.J., Caprioli, R.M., 2004. Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics 4, 793–811. https://doi.org/10.1002/pmic.200300635

Frosst, P., Blom, H.J., Milos, R., Goyette, P., Sheppard, C.A., Matthews, R.G., Boers, G.J.H., den Heijer, M., Kluijtmans, L. a. J., van den Heuve, L.P., Rozen, R., 1995. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat. Genet. 10, 111–113. https://doi.org/10.1038/ng0595-111

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., Verstrepen, K.J., 2016. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. Cell 166, 1397-1410.e16. https://doi.org/10.1016/j.cell.2016.08.020

Gan, X., Stegle, O., Behr, J., Steffen, J.G., Drewe, P., Hildebrand, K.L., Lyngsoe, R., Schultheiss, S.J., Osborne, E.J., Sreedharan, V.T., Kahles, A., Bohnert, R., Jean, G., Derwent, P., Kersey, P., Belfield, E.J., Harberd, N.P., Kemen, E., Toomajian, C., Kover, P.X., Clark, R.M., Rätsch, G., Mott, R., 2011. Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature 477, 419–423. https://doi.org/10.1038/nature10414

Gandhi, S.J., Zenklusen, D., Lionnet, T., Singer, R.H., 2011. Transcription of functionally related constitutive genes is not coordinated. Nat. Struct. Mol. Biol. 18, 27–34. https://doi.org/10.1038/nsmb.1934

García-Ríos, E., Nuévalos, M., Barrio, E., Puig, S., Guillamón, J.M., 2019. A new chromosomal rearrangement improves the adaptation of wine yeasts to sulfite. Environ. Microbiol. 21, 1771–1781. https://doi.org/10.1111/1462-2920.14586

Garieri, M., Delaneau, O., Santoni, F., Fish, R.J., Mull, D., Carninci, P., Dermitzakis, E.T., Antonarakis, S.E., Fort, A., 2017. The effect of genetic variation on promoter usage and enhancer activity. Nat. Commun. 8, 1358. https://doi.org/10.1038/s41467-017-01467-7

Gatti, D.M., Svenson, K.L., Shabalin, A., Wu, L.-Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., Palmer, A., Chesler, E.J., Broman, K.W., Churchill, G.A., 2014. Quantitative Trait Locus Mapping Methods for Diversity Outbred Mice. G3 GenesGenomesGenetics 4, 1623–1633. https://doi.org/10.1534/g3.114.013748

Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A., Pedersen, N.L., 2006. Role of Genes and Environments for Explaining Alzheimer Disease. Arch. Gen. Psychiatry 63, 168–174. https://doi.org/10.1001/archpsyc.63.2.168

Gautier, E.-F., Ducamp, S., Leduc, M., Salnot, V., Guillonneau, F., Dussiot, M., Hale, J., Giarratana, M.-C., Raimbault, A., Douay, L., Lacombe, C., Mohandas, N., Verdier, F., Zermati, Y., Mayeux, P., 2016. Comprehensive Proteomic Analysis of Human Erythropoiesis. Cell Rep. 16, 1470–1484. https://doi.org/10.1016/j.celrep.2016.06.085

Gedeon, T., Bokes, P., 2012. Delayed Protein Synthesis Reduces the Correlation between mRNA and Protein Fluctuations. Biophys. J. 103, 377–385. https://doi.org/10.1016/j.bpj.2012.06.025

Geiger, T., Cox, J., Mann, M., 2010. Proteomic Changes Resulting from Gene Copy Number Variations in Cancer Cells. PLOS Genet. 6, e1001090. https://doi.org/10.1371/journal.pgen.1001090

Génin, E., 2020. Missing heritability of complex diseases: case solved? Hum. Genet. 139, 103–113. https://doi.org/10.1007/s00439-019-02034-4

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741. https://doi.org/10.1038/nature02046

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P.-Z., Brewer, H., Weitz, K., Ii, D.G.C., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D.J., Smith, R.D., Lusis, A.J., 2011. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. PLOS Genet. 7, e1001393. https://doi.org/10.1371/journal.pgen.1001393

Giaever, G., Nislow, C., 2014. The Yeast Deletion Collection: A Decade of Functional Genomics. Genetics 197, 451–465. https://doi.org/10.1534/genetics.114.161620

Gibson, G., 2012. Rare and common variants: twenty arguments. Nat. Rev. Genet. 13, 135–145. https://doi.org/10.1038/nrg3118

Gillet, L.C., Leitner, A., Aebersold, R., 2016. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. Annu. Rev. Anal. Chem. 9, 449–472. https://doi.org/10.1146/annurev-anchem-071015-041535

Gilliland, G., Perrin, S., Blanchard, K., Bunn, H.F., 1990. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. Proc. Natl. Acad. Sci. 87, 2725–2729. https://doi.org/10.1073/pnas.87.7.2725

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver,

S.G., 1996. Life with 6000 Genes. Science 274, 546–567. https://doi.org/10.1126/science.274.5287.546
Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., Beltrao, P., 2017. Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. Cell Syst. 5, 386-398.e4. https://doi.org/10.1016/j.cels.2017.08.013
Gorkovskiy, A., Verstrepen, K.J., 2021. The Role of Structural Variation in Adaptation and Evolution of Yeast and Other Fungi. Genes 12, 699. https://doi.org/10.3390/genes12050699
Grabrucker, A., 2013. Environmental Factors in Autism. Front. Psychiatry 3.
Grönlund, A., Lötstedt, P., Elf, J., 2013. Transcription factor binding kinetics constrain noise suppression via negative feedback. Nat. Commun. 4, 1864. https://doi.org/10.1038/ncomms2867
Gudjonsson, A., Gudmundsdottir, V., Axelsson, G.T., Gudmundsson, E.F., Jonsson, B.G., Launer, L.J., Lamb, J.R., Jennings, L.L., Aspelund, T., Emilsson, V., Gudnason, V., 2022. A genome-wide association study of serum proteins reveals shared loci with common diseases. Nat. Commun. 13, 480. https://doi.org/10.1038/s41467-021-27850-z
Guydosh, N.R., Green, R., 2014. Dom34 rescues ribosomes in 3' untranslated regions. Cell 156, 950–962. https://doi.org/10.1016/j.cell.2014.02.006
Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R., 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19, 1720–1730. https://doi.org/10.1128/MCB.19.3.1720
Han, J.-W., Zheng, H.-F., Cui, Y., Sun, L.-D., Ye, D.-Q., Hu, Z., Xu, Jin-Hua, Cai, Z.-M., Huang, W., Zhao, G.-P., Xie, H.-F., Fang, H., Lu, Q.-J., Xu, Jian-Hua, Li, X.-P., Pan, Y.-F., Deng, D.-Q., Zeng, F.-Q., Ye, Z.-Z., Zhang, X.-Y., Wang, Q.-W., Hao, F., Ma, L., Zuo, X.-B., Zhou, F.-S., Du, W.-H., Cheng, Y.-L., Yang, J.-Q., Shen, S.-K., Li, J., Sheng, Y.-J., Zuo, X.-X., Zhu, W.-F., Gao, F., Zhang, P.-L., Guo, Q., Li, B., Gao, M., Xiao, F.-L., Quan, C., Zhang, C., Zhang, Z., Zhu, K.-J., Li, Yang, Hu, D.-Y., Lu, W.-S., Huang, J.-L., Liu, S.-X., Li, H., Ren, Y.-Q., Wang, Z.-X., Yang, C.-J., Wang, P.-G., Zhou, W.-M., Lv, Y.-M., Zhang, A.-P., Zhang, S.-Q., Lin, D., Li, Yi, Low, H.Q., Shen, M., Zhai, Z.-F., Wang, Y., Zhang, F.-Y., Yang, S., Liu, J.-J., Zhang, X.-J., 2009. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat. Genet. 41, 1234–1237. https://doi.org/10.1038/ng.472
Hawe, J.S., Wilson, R., Schmid, K.T., Zhou, L., Lakshmanan, L.N., Lehne, B.C., Kühnel, B., Scott, W.R., Wielscher, M., Yew, Y.W., Baumbach, C., Lee, D.P., Marouli, E., Bernard, M., Pfeiffer, L., Matías-García, P.R., Autio, M.I., Bourgeois, S., Herder, C., Karhunen, V., Meitinger, T., Prokisch, H., Rathmann, W., Roden, M., Sebert, S., Shin, J., Strauch, K., Zhang, W., Tan, W.L.W., Hauck, S.M., Merl-Pham, J., Grallert, H., Barbosa, E.G.V., Illig, T., Peters, A., Paus, T., Pausova, Z., Deloukas, P., Foo, R.S.Y., Jarvelin, M.-R., Kooner, J.S., Loh, M., Heinig, M., Gieger, C., Waldenberger, M., Chambers, J.C., 2022. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. Nat. Genet. 54, 18–29. https://doi.org/10.1038/s41588-021-00969-x
He, B., Shi, J., Wang, X., Jiang, H., Zhu, H.-J., 2020. Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. BMC Biol. 18, 97. https://doi.org/10.1186/s12915-020-00830-3
He, Q., Tang, S., Zhi, H., Chen, J., Zhang, Jun, Liang, H., Alam, O., Li, H., Zhang, H., Xing, Lihe, Li, X., Zhang, W., Wang, Hailong, Shi, J., Du, H., Wu, H., Wang, L., Yang, P., Xing, Lu, Yan, H., Song, Z., Liu, J., Wang, Haigang, Tian, X., Qiao, Z., Feng, G., Guo, R., Zhu, W., Ren, Y., Hao, H., Li, M., Zhang, A., Guo, E., Yan, F., Li, Q., Liu, Y., Tian, B., Zhao, X., Jia, R., Feng, B., Zhang, Jiewei, Wei, J., Lai, J., Jia, G., Purugganan, M., Diao, X., 2023. A graph-based genome and pan-genome variation of the model plant Setaria. Nat. Genet. 1–11. https://doi.org/10.1038/s41588-023-01423-w
Hecht, S.S., 2006. Cigarette smoking: cancer risks, carcinogens, and mechanisms. Langenbecks Arch. Surg. 391, 603–613. https://doi.org/10.1007/s00423-006-0111-z
Heinig, M., Petretto, E., Wallace, C., Bottolo, L., Rotival, M., Lu, H., Li, Y., Sarwar, R., Langley, S.R., Bauerfeind, A., Hummel, O., Lee, Y.-A., Paskas, S., Rintisch, C., Saar, K., Cooper, J., Buchan, R., Gray, E.E., Cyster, J.G., Cardiogenics Consortium, Erdmann, J., Hengstenberg, C., Maouche, S., Ouwehand, W.H., Rice, C.M., Samani, N.J., Schunkert, H., Goodall, A.H., Schulz, H., Roider, H.G., Vingron, M., Blankenberg, S., Münzel, T., Zeller, T., Szymczak, S., Ziegler, A., Tiret, L., Smyth, D.J., Pravenec, M., Aitman, T.J., Cambien, F., Clayton, D., Todd, J.A., Hubner, N., Cook, S.A., 2010. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature 467, 460–464. https://doi.org/10.1038/nature09386

Hill, M.S., Vande Zande, P., Wittkopp, P.J., 2021. Molecular and evolutionary processes generating variation in gene expression. Nat. Rev. Genet. 22, 203–215. https://doi.org/10.1038/s41576-020-00304-w
Hong, M., Tao, S., Zhang, L., Diao, L.-T., Huang, X., Huang, S., Xie, S.-J., Xiao, Z.-D., Zhang, H., 2020. RNA sequencing: new technologies and applications in cancer research. J. Hematol. Oncol.J Hematol Oncol 13, 166. https://doi.org/10.1186/s13045-020-01005-x
Hooks, K.B., Delneri, D., Griffiths-Jones, S., 2014. Intron Evolution in Saccharomycetaceae. Genome Biol. Evol. 6, 2543–2556. https://doi.org/10.1093/gbe/evu196
Hornung, G., Barkai, N., 2008. Noise Propagation and Signaling Sensitivity in Biological Networks: A Role for Positive Feedback. PLOS Comput. Biol. 4, e8. https://doi.org/10.1371/journal.pcbi.0040008
Hou, J., Friedrich, A., de Montigny, J., Schacherer, J., 2014. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in Saccharomyces cerevisiae. Curr. Biol. CB 24, 1153–1159. https://doi.org/10.1016/j.cub.2014.03.063
Hrdlickova, R., Toloue, M., Tian, B., 2017. RNA-Seq methods for transcriptome analysis. Wiley Interdiscip. Rev. RNA 8, 10.1002/wrna.1364. https://doi.org/10.1002/wrna.1364
Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., Garraway, L.A., 2013. Highly recurrent TERT promoter mutations in human melanoma. Science 339, 957–959. https://doi.org/10.1126/science.1229259
Huang, K.-L., Li, S., Mertins, P., Cao, S., Gunawardena, H.P., Ruggles, K.V., Mani, D.R., Clauser, K.R., Tanioka, M., Usary, J., Kavuri, S.M., Xie, L., Yoon, C., Qiao, J.W., Wrobel, J., Wyczalkowski, M.A., Erdmann-Gilmore, P., Snider, J.E., Hoog, J., Singh, P., Niu, B., Guo, Z., Sun, S.Q., Sanati, S., Kawaler, E., Wang, X., Scott, A., Ye, K., McLellan, M.D., Wendl, M.C., Malovannaya, A., Held, J.M., Gillette, M.A., Fenyö, D., Kinsinger, C.R., Mesri, M., Rodriguez, H., Davies, S.R., Perou, C.M., Ma, C., Reid Townsend, R., Chen, X., Carr, S.A., Ellis, M.J., Ding, L., 2017. Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. Nat. Commun. 8, 14864. https://doi.org/10.1038/ncomms14864
Huang, L., Low, A., Damle, S.S., Keenan, M.M., Kuntz, S., Murray, S.F., Monia, B.P., Guo, S., 2018. Antisense suppression of the nonsense mediated decay factor Upf3b as a potential treatment for diseases caused by nonsense mutations. Genome Biol. 19, 4. https://doi.org/10.1186/s13059-017-1386-9
Hussmann, J.A., Patchett, S., Johnson, A., Sawyer, S., Press, W.H., 2015. Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. PLOS Genet. 11, e1005732. https://doi.org/10.1371/journal.pgen.1005732
Hyde, C.L., Nagle, M.W., Tian, C., Chen, X., Paciga, S.A., Wendland, J.R., Tung, J.Y., Hinds, D.A., Perlis, R.H., Winslow, A.R., 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat. Genet. 48, 1031–1036. https://doi.org/10.1038/ng.3623
Ingolia, N.T., 2010. Chapter 6 - Genome-Wide Translational Profiling by Ribosome Footprinting, in: Methods in Enzymology, Guide to Yeast Genetics: Functional Genomics, Proteomics, and Other Systems Analysis. Academic Press, pp. 119–142. https://doi.org/10.1016/S0076-6879(10)70006-9
Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., Weissman, J.S., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat. Protoc. 7, 1534–1550. https://doi.org/10.1038/nprot.2012.086
Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., Weissman, J.S., 2014. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. Cell Rep. 8, 1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045
Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 324, 218–223. https://doi.org/10.1126/science.1168978
Ingolia, N.T., Hussmann, J.A., Weissman, J.S., 2019. Ribosome Profiling: Global Views of Translation. Cold Spring Harb. Perspect. Biol. 11. https://doi.org/10.1101/cshperspect.a032698
Ingram, V.M., 1957. Gene Mutations in Human Hæmoglobin: the Chemical Difference Between Normal and Sickle Cell Hæmoglobin. Nature 180, 326–328. https://doi.org/10.1038/180326a0
Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356. https://doi.org/10.1016/S0022-2836(61)80072-7
Jakobson, C.M., Jarosz, D.F., 2019. Molecular Origins of Complex Heritability in Natural Genotype-to-Phenotype Relationships. Cell Syst. 8, 363-379.e3. https://doi.org/10.1016/j.cels.2019.04.002
Jansen, P.R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A.R., de Leeuw, C.A., Benjamins, J.S., Muñoz-Manchado, A.B., Nagel, M., Savage, J.E., Tiemeier, H., White, T.,

23andMe Research Team, Tung, J.Y., Hinds, D.A., Vacic, V., Wang, X., Sullivan, P.F., van der Sluis, S., Polderman, T.J.C., Smit, A.B., Hjerling-Leffler, J., Van Someren, E.J.W., Posthuma, D., 2019. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat. Genet. 51, 394–403. https://doi.org/10.1038/s41588-018-0333-3

Jansen, R.C., Nap, J.P., 2001. Genetical genomics: the added value from segregation. Trends Genet. TIG 17, 388–391. https://doi.org/10.1016/s0168-9525(01)02310-1

Janssens, A.C.J.W., van Duijn, C.M., 2008. Genome-based prediction of common diseases: advances and prospects. Hum. Mol. Genet. 17, R166-173. https://doi.org/10.1093/hmg/ddn250

Jiang, L., Wang, M., Lin, S., Jian, R., Li, Xiao, Chan, J., Dong, G., Fang, H., Robinson, A.E., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., MacArthur, D.G., Meier, S.R., Nedzel, J.L., Nguyen, D.Y., Segrè, A.V., Todres, E., Balliu, B., Barbeira, A.N., Battle, A., Bonazzola, R., Brown, A., Brown, C.D., Castel, S.E., Conrad, D., Cotter, D.J., Cox, N., Das, S., Goede, O.M. de, Dermitzakis, E.T., Engelhardt, B.E., Eskin, E., Eulalio, T.Y., Ferraro, N.M., Flynn, E., Fresard, L., Gamazon, E.R., Garrido-Martín, D., Gay, N.R., Guigó, R., Hamel, A.R., He, Y., Hoffman, P.J., Hormozdiari, F., Hou, L., Im, H.K., Jo, B., Kasela, S., Kellis, M., Kim-Hellmuth, S., Kwong, A., Lappalainen, T., Li, Xin, Liang, Y., Mangul, S., Mohammadi, P., Montgomery, S.B., Muñoz-Aguirre, M., Nachun, D.C., Nobel, A.B., Oliva, M., Park, YoSon, Park, Yongjin, Parsana, P., Reverter, F., Rouhana, J.M., Sabatti, C., Saha, A., Skol, A.D., Stephens, M., Stranger, B.E., Strober, B.J., Teran, N.A., Viñuela, A., Wang, G., Wen, X., Wright, F., Wucher, V., Zou, Y., Ferreira, P.G., Li, G., Melé, M., Yeger-Lotem, E., Barcus, M.E., Bradbury, D., Krubit, T., McLean, J.A., Qi, L., Robinson, K., Roche, N.V., Smith, A.M., Sobin, L., Tabor, D.E., Undale, A., Bridge, J., Brigham, L.E., Foster, B.A., Gillard, B.M., Hasz, R., Hunter, M., Johns, C., Johnson, M., Karasik, E., Kopen, G., Leinweber, W.F., McDonald, A., Moser, M.T., Myer, K., Ramsey, K.D., Roe, B., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Jewell, S.D., Rohrer, D.C., Valley, D.R., Davis, D.A., Mash, D.C., Branton, P.A., Barker, L.K., Gardiner, H.M., Mosavel, M., Siminoff, L.A., Flicek, P., Haeussler, M., Juettemann, T., Kent, W.J., Lee, C.M., Powell, C.C., Rosenbloom, K.R., Ruffier, M., Sheppard, D., Taylor, K., Trevanion, S.J., Zerbino, D.R., Abell, N.S., Akey, J., Chen, L., Demanelis, K., Doherty, J.A., Feinberg, A.P., Hansen, K.D., Hickey, P.F., Jasmine, F., Kaul, R., Kibriya, M.G., Li, J.B., Li, Q., Linder, S.E., Pierce, B.L., Rizzardi, L.F., Smith, K.S., Stamatoyannopoulos, J., Tang, H., Carithers, L.J., Guan, P., Koester, S.E., Little, A.R., Moore, H.M., Nierras, C.R., Rao, A.K., Vaught, J.B., Volpi, S., Snyder, M.P., 2020. A Quantitative Proteome Map of the Human Body. Cell 183, 269-283.e19. https://doi.org/10.1016/j.cell.2020.08.036

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821. https://doi.org/10.1126/science.1225829

Johannsen, W., 1911. The Genotype Conception of Heredity. Am. Nat. 45, 129–159.

Juszkiewicz, S., Hegde, R.S., 2018. Quality Control of Orphaned Proteins. Mol. Cell 71, 443–457. https://doi.org/10.1016/j.molcel.2018.07.001

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S., Freimer, N.B., Sabatti, C., Eskin, E., 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42, 348–354. https://doi.org/10.1038/ng.548

Karpievitch, Y.V., Polpitiya, A.D., Anderson, G.A., Smith, R.D., Dabney, A.R., 2010. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann. Appl. Stat. 4, 1797–1823. https://doi.org/10.1214/10-AOAS341

Keele, G.R., Quach, B.C., Israel, J.W., Chappell, G.A., Lewis, L., Safi, A., Simon, J.M., Cotney, P., Crawford, G.E., Valdar, W., Rusyn, I., Furey, T.S., 2020. Integrative QTL analysis of gene expression and chromatin accessibility identifies multi-tissue patterns of genetic regulation. PLOS Genet. 16, e1008537. https://doi.org/10.1371/journal.pgen.1008537

Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J., Burdett, T., Jupp, S., Parkinson, H., Papatheodorou, I., Yates, A.D., Zerbino, D.R., Alasoo, K., 2021. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat. Genet. 53, 1290–1299. https://doi.org/10.1038/s41588-021-00924-w

Khan, Z., Bloom, J.S., Garcia, B.A., Singh, M., Kruglyak, L., 2009. Protein quantification across hundreds of experimental conditions. Proc. Natl. Acad. Sci. 106, 15544–15548. https://doi.org/10.1073/pnas.0904100106

Khan, Z., Ford, M.J., Cusanovich, D.A., Mitrano, A., Pritchard, J.K., Gilad, Y., 2013. Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures. Science 342, 1100–1104. https://doi.org/10.1126/science.1242379

Kikuchi, M., Hara, N., Hasegawa, M., Miyashita, A., Kuwano, R., Ikeuchi, T., Nakaya, A., 2019. Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. BMC Med. Genomics 12, 128. https://doi.org/10.1186/s12920-019-0574-8

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.-Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., Hoh, J., 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308, 385–389. https://doi.org/10.1126/science.1109557

Koch, A., Joosten, S.C., Feng, Z., de Ruijter, T.C., Draht, M.X., Melotte, V., Smits, K.M., Veeck, J., Herman, J.G., Van Neste, L., Van Criekinge, W., De Meyer, T., van Engeland, M., 2018. Analysis of DNA methylation in cancer: location revisited. Nat. Rev. Clin. Oncol. 15, 459–466. https://doi.org/10.1038/s41571-018-0004-4

Krupenko, S.A., Horita, D.A., 2019. The Role of Single-Nucleotide Polymorphisms in the Function of Candidate Tumor Suppressor ALDH1L1. Front. Genet. 10.

Lahue, C., Madden, A., Dunn, R., Smukowski Heil, C., 2020. History and Domestication of Saccharomyces cerevisiae in Bread Baking. Front. Genet. 11.

Lan, Y., Sun, R., Ouyang, J., Ding, W., Kim, M.-J., Wu, J., Li, Y., Shi, T., 2021. AtMAD: Arabidopsis thaliana multi-omics association database. Nucleic Acids Res. 49, D1445–D1451. https://doi.org/10.1093/nar/gkaa1042

Laurent, J.M., Vogel, C., Kwon, T., Craig, S.A., Boutz, D.R., Huse, H.K., Nozue, K., Walia, H., Whiteley, M., Ronald, P.C., Marcotte, E.M., 2010. Protein abundances are more conserved than mRNA abundances across diverse taxa. PROTEOMICS 10, 4209–4212. https://doi.org/10.1002/pmic.201000327

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M.A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P.N., Walters, R.K., Willoughby, E.A., Yengo, L., 23andMe Research Team, COGENT (Cognitive Genomics Consortium), Social Science Genetic Association Consortium, Alver, M., Bao, Y., Clark, D.W., Day, F.R., Furlotte, N.A., Joshi, P.K., Kemper, K.E., Kleinman, A., Langenberg, C., Mägi, R., Trampush, J.W., Verma, S.S., Wu, Y., Lam, M., Zhao, J.H., Zheng, Z., Boardman, J.D., Campbell, H., Freese, J., Harris, K.M., Hayward, C., Herd, P., Kumari, M., Lencz, T., Luan, J., Malhotra, A.K., Metspalu, A., Milani, L., Ong, K.K., Perry, J.R.B., Porteous, D.J., Ritchie, M.D., Smart, M.C., Smith, B.H., Tung, J.Y., Wareham, N.J., Wilson, J.F., Beauchamp, J.P., Conley, D.C., Esko, T., Lehrer, S.F., Magnusson, P.K.E., Oskarsson, S., Pers, T.H., Robinson, M.R., Thom, K., Watson, C., Chabris, C.F., Meyer, M.N., Laibson, D.I., Yang, J., Johannesson, M., Koellinger, P.D., Turley, P., Visscher, P.M., Benjamin, D.J., Cesarini, D., 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50, 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lee, M.-W., Kim, B.-J., Choi, H.-K., Ryu, M.-J., Kim, S.-B., Kang, K.-M., Cho, E.-J., Youn, H.-D., Huh, W.-K., Kim, S.-T., 2007. Global protein expression profiling of budding yeast in response to DNA damage. Yeast Chichester Engl. 24, 145–154. https://doi.org/10.1002/yea.1446

Lee, T.J., Liu, Y.-C., Liu, W.-A., Lin, Y.-F., Lee, H.-H., Ke, H.-M., Huang, J.-P., Lu, M.-Y.J., Hsieh, C.-L., Chung, K.-F., Liti, G., Tsai, I.J., 2022. Extensive sampling of Saccharomyces cerevisiae in Taiwan reveals ecology and evolution of predomesticated lineages. Genome Res. 32, 864–877. https://doi.org/10.1101/gr.276286.121

Lemos, B., Araripe, L.O., Fontanillas, P., Hartl, D.L., 2008. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proc. Natl. Acad. Sci. U. S. A. 105, 14471–14476. https://doi.org/10.1073/pnas.0805160105

Letourneau, A., Santoni, F.A., Bonilla, X., Sailani, M.R., Gonzalez, D., Kind, J., Chevalier, C., Thurman, R., Sandstrom, R.S., Hibaoui, Y., Garieri, M., Popadin, K., Falconnet, E., Gagnebin, M., Gehrig, C., Vannier, A., Guipponi, M., Farinelli, L., Robyr, D., Migliavacca, E., Borel, C., Deutsch, S., Feki, A., Stamatoyannopoulos, J.A., Herault, Y., van Steensel, B., Guigo, R., Antonarakis, S.E., 2014. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature 508, 345–350. https://doi.org/10.1038/nature13200

Lewinsky, R.H., Jensen, T.G.K., Møller, J., Stensballe, A., Olsen, J., Troelsen, J.T., 2005. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum. Mol. Genet. 14, 3945–3953. https://doi.org/10.1093/hmg/ddi418
Li, G., Ji, B., Nielsen, J., 2019. The pan-genome of Saccharomyces cerevisiae. FEMS Yeast Res. 19, foz064. https://doi.org/10.1093/femsyr/foz064
Li, G.-W., Burkhardt, D., Gross, C., Weissman, J.S., 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157, 624–635. https://doi.org/10.1016/j.cell.2014.02.033
Li, H., Wang, S., Chai, S., Yang, Z., Zhang, Q., Xin, H., Xu, Y., Lin, S., Chen, X., Yao, Z., Yang, Q., Fei, Z., Huang, S., Zhang, Z., 2022. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. Nat. Commun. 13, 682. https://doi.org/10.1038/s41467-022-28362-0
Li, J., Smith, L.S., Zhu, H.-J., 2021. Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. Drug Discov. Today Technol. 39, 49–56. https://doi.org/10.1016/j.ddtec.2021.06.006
Li, J.J., Bickel, P.J., Biggin, M.D., 2014. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. PeerJ 2, e270. https://doi.org/10.7717/peerj.270
Li, W., Notani, D., Rosenfeld, M.G., 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet. 17, 207–223. https://doi.org/10.1038/nrg.2016.4
Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., Pritchard, J.K., 2016. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604. https://doi.org/10.1126/science.aad9417
Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., Song, Z., Ji, W., Wang, M., Zhou, J., Chen, B., Liu, Y., Wang, J., Wang, P., Yang, P., Wang, Q., Feng, G., Liu, B., Sun, W., Li, B., He, G., Li, Weidong, Wan, C., Xu, Q., Li, Wenjin, Wen, Z., Liu, K., Huang, F., Ji, J., Ripke, S., Yue, W., Sullivan, P.F., O'Donovan, M.C., Shi, Y., 2017. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat. Genet. 49, 1576–1583. https://doi.org/10.1038/ng.3973
Li, Z., Zhang, Y., Ku, L., Wilkinson, K.D., Warren, S.T., Feng, Y., 2001. The fragile X mental retardation protein inhibits translation via interacting with mRNA. Nucleic Acids Res. 29, 2276–2283. https://doi.org/10.1093/nar/29.11.2276
Lindberg, M.J., Byström, R., Boknäs, N., Andersen, P.M., Oliveberg, M., 2005. Systematically perturbed folding patterns of amyotrophic lateral sclerosis (ALS)-associated SOD1 mutants. Proc. Natl. Acad. Sci. 102, 9754–9759. https://doi.org/10.1073/pnas.0501957102
Lindeboom, R.G.H., Vermeulen, M., Lehner, B., Supek, F., 2019. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nat. Genet. 51, 1645–1651. https://doi.org/10.1038/s41588-019-0517-5
Ling, S.-C., Albuquerque, C.P., Han, J.S., Lagier-Tourenne, C., Tokunaga, S., Zhou, H., Cleveland, D.W., 2010. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc. Natl. Acad. Sci. U. S. A. 107, 13318–13323. https://doi.org/10.1073/pnas.1008227107
Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., Heckerman, D., 2011. FaST linear mixed models for genome-wide association studies. Nat. Methods 8, 833–835. https://doi.org/10.1038/nmeth.1681
Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., Ecker, J.R., 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523–536. https://doi.org/10.1016/j.cell.2008.03.029
Liti, G., Louis, E.J., 2012. Advances in quantitative trait analysis in yeast. PLoS Genet. 8, e1002912. https://doi.org/10.1371/journal.pgen.1002912
Liu, B., Gloudemans, M.J., Rao, A.S., Ingelsson, E., Montgomery, S.B., 2019. Abundant associations with gene expression complicate GWAS follow-up. Nat. Genet. 51, 768–769. https://doi.org/10.1038/s41588-019-0404-0
Liu, Y., Beyer, A., Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550. https://doi.org/10.1016/j.cell.2016.03.014
Liu, Y., Borel, C., Li, L., Müller, T., Williams, E.G., Germain, P.-L., Buljan, M., Sajic, T., Boersema, P.J., Shao, W., Faini, M., Testa, G., Beyer, A., Antonarakis, S.E., Aebersold, R., 2017. Systematic

proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. Nat. Commun. 8, 1212. https://doi.org/10.1038/s41467-017-01422-6
Logsdon, G.A., Vollger, M.R., Eichler, E.E., 2020. Long-read human genome sequencing and its applications. Nat. Rev. Genet. 21, 597–614. https://doi.org/10.1038/s41576-020-0236-x
Loos, R.J.F., Janssens, A.C.J.W., 2017. Predicting Polygenic Obesity Using Genetic Information. Cell Metab. 25, 535–543. https://doi.org/10.1016/j.cmet.2017.02.013
Lowry, OliverH., Rosebrough, NiraJ., Farr, A.L., Randall, RoseJ., 1951. PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT. J. Biol. Chem. 193, 265–275. https://doi.org/10.1016/S0021-9258(19)52451-6
Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., Segal, E., 2015. Core promoter sequence in yeast is a major determinant of expression level. Genome Res. 25, 1008–1017. https://doi.org/10.1101/gr.188193.114
Lutz, S., Brion, C., Kliebhan, M., Albert, F.W., 2019. DNA variants affecting the expression of numerous genes in trans have diverse mechanisms of action and evolutionary histories. PLOS Genet. 15, e1008375. https://doi.org/10.1371/journal.pgen.1008375
MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z.M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., Parkinson, H., 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901. https://doi.org/10.1093/nar/gkw1133
Magdeldin, S., Enany, S., Yoshida, Y., Xu, B., Zhang, Y., Zureena, Z., Lokamani, I., Yaoita, E., Yamamoto, T., 2014. Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. Clin. Proteomics 11, 16. https://doi.org/10.1186/1559-0275-11-16
Maher, B., 2008. Personal genomes: The case of the missing heritability. Nature 456, 18–21. https://doi.org/10.1038/456018a
Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. https://doi.org/10.1038/nature08494
Mantione, K.J., Kream, R.M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J.M., Stefano, G.B., 2014. Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq. Med. Sci. Monit. Basic Res. 20, 138–141. https://doi.org/10.12659/MSMBR.892101
Marcet-Houben, M., Gabaldón, T., 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. PLoS Biol. 13, e1002220. https://doi.org/10.1371/journal.pbio.1002220
Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., Bähler, J., 2012. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell 151, 671–683. https://doi.org/10.1016/j.cell.2012.09.019
Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E., Dequin, S., Galeote, V., 2015. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Mol. Biol. Evol. 32, 1695–1707. https://doi.org/10.1093/molbev/msv057
Mazumder, A., Pesudo, L.Q., McRee, S., Bathe, M., Samson, L.D., 2013. Genome-wide single-cell-level screen for protein abundance and localization changes in response to DNA damage in S. cerevisiae. Nucleic Acids Res. 41, 9310–9324. https://doi.org/10.1093/nar/gkt715
McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nuñez, A., Butrym, E.D., Richards, M.P., Wang, Chen-shan, Cheng, G., Zhao, Z., Wang, Changsui, 2004. Fermented beverages of pre- and proto-historic China. Proc. Natl. Acad. Sci. 101, 17593–17598. https://doi.org/10.1073/pnas.0407921102
McManus, C.J., May, G.E., Spealman, P., Shteyman, A., 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422–430. https://doi.org/10.1101/gr.164996.113
Mendel, G., 1866. Versuche über Pflanzen-Hybriden. Verhandlungen Naturforschenden Vereines Brünn Bd.4 (1865-1866), 3–47.
Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J.T., Gatza, M.L., Wilkerson, M., Perou, C.M., Yellapantula, V., Huang, K., Lin, C., McLellan, M.D., Yan, P., Davies, S.R.,

Townsend, R.R., Skates, S.J., Wang, J., Zhang, B., Kinsinger, C.R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A.G., Fenyö, D., Ellis, M.J., Carr, S.A., NCI CPTAC, 2016. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature 534, 55–62. https://doi.org/10.1038/nature18003

Messner, C.B., Demichev, V., Muenzner, J., Aulakh, S.K., Barthel, N., Röhl, A., Herrera-Domínguez, L., Egger, A.-S., Kamrad, S., Hou, J., Tan, G., Lemke, O., Calvani, E., Szyrwiel, L., Mülleder, M., Lilley, K.S., Boone, C., Kustatscher, G., Ralser, M., 2023. The proteomic landscape of genome-wide genetic perturbations. Cell 186, 2018-2034.e21. https://doi.org/10.1016/j.cell.2023.03.026

Messner, C.B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleder, M., Ralser, M., 2022. Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. PROTEOMICS n/a, 2200013. https://doi.org/10.1002/pmic.202200013

Michel, A.M., Choudhury, K.R., Firth, A.E., Ingolia, N.T., Atkins, J.F., Baranov, P.V., 2012. Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res. 22, 2219–2229. https://doi.org/10.1101/gr.133249.111

Michel, R.H., McGovern, P.E., Badler, V.R., 1992. Chemical evidence for ancient beer. Nature 360, 24–24. https://doi.org/10.1038/360024b0

Min, J.L., Hemani, G., Hannon, E., Dekkers, K.F., Castillo-Fernandez, J., Luijk, R., Carnero-Montoro, E., Lawson, D.J., Burrows, K., Suderman, M., Bretherick, A.D., Richardson, T.G., Klughammer, J., Iotchkova, V., Sharp, G., Al Khleifat, A., Shatunov, A., Iacoangeli, A., McArdle, W.L., Ho, K.M., Kumar, A., Söderhäll, C., Soriano-Tárraga, C., Giralt-Steinhauer, E., Kazmi, N., Mason, D., McRae, A.F., Corcoran, D.L., Sugden, K., Kasela, S., Cardona, A., Day, F.R., Cugliari, G., Viberti, C., Guarrera, S., Lerro, M., Gupta, R., Bollepalli, S., Mandaviya, P., Zeng, Y., Clarke, T.-K., Walker, R.M., Schmoll, V., Czamara, D., Ruiz-Arenas, C., Rezwan, F.I., Marioni, R.E., Lin, T., Awaloff, Y., Germain, M., Aïssi, D., Zwamborn, R., van Eijk, K., Dekker, A., van Dongen, J., Hottenga, J.-J., Willemsen, G., Xu, C.-J., Barturen, G., Català-Moll, F., Kerick, M., Wang, C., Melton, P., Elliott, H.R., Shin, J., Bernard, M., Yet, I., Smart, M., Gorrie-Stone, T., Shaw, C., Al Chalabi, A., Ring, S.M., Pershagen, G., Melén, E., Jiménez-Conde, J., Roquer, J., Lawlor, D.A., Wright, J., Martin, N.G., Montgomery, G.W., Moffitt, T.E., Poulton, R., Esko, T., Milani, L., Metspalu, A., Perry, J.R.B., Ong, K.K., Wareham, N.J., Matullo, G., Sacerdote, C., Panico, S., Caspi, A., Arseneault, L., Gagnon, F., Ollikainen, M., Kaprio, J., Felix, J.F., Rivadeneira, F., Tiemeier, H., van IJzendoorn, M.H., Uitterlinden, A.G., Jaddoe, V.W.V., Haley, C., McIntosh, A.M., Evans, K.L., Murray, A., Räikkönen, K., Lahti, J., Nohr, E.A., Sørensen, T.I.A., Hansen, T., Morgen, C.S., Binder, E.B., Lucae, S., Gonzalez, J.R., Bustamante, M., Sunyer, J., Holloway, J.W., Karmaus, W., Zhang, H., Deary, I.J., Wray, N.R., Starr, J.M., Beekman, M., van Heemst, D., Slagboom, P.E., Morange, P.-E., Trégouët, D.-A., Veldink, J.H., Davies, G.E., de Geus, E.J.C., Boomsma, D.I., Vonk, J.M., Brunekreef, B., Koppelman, G.H., Alarcón-Riquelme, M.E., Huang, R.-C., Pennell, C.E., van Meurs, J., Ikram, M.A., Hughes, A.D., Tillin, T., Chaturvedi, N., Pausova, Z., Paus, T., Spector, T.D., Kumari, M., Schalkwyk, L.C., Visscher, P.M., Davey Smith, G., Bock, C., Gaunt, T.R., Bell, J.T., Heijmans, B.T., Mill, J., Relton, C.L., 2021. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat. Genet. 53, 1311–1321. https://doi.org/10.1038/s41588-021-00923-x

Mirauta, B.A., Seaton, D.D., Bensaddek, D., Brenes, A., Bonder, M.J., Kilpinen, H., HipSci Consortium, Stegle, O., Lamond, A.I., 2020. Population-scale proteome variation in human induced pluripotent stem cells. eLife 9, e57390. https://doi.org/10.7554/eLife.57390

Mito, M., Shichino, Y., Iwasaki, S., 2023. Thor-Ribo-Seq: ribosome profiling tailored for low input with RNA-dependent RNA amplification. https://doi.org/10.1101/2023.01.15.524129

Morgan, T.H., Sturtevant, A.H., Muller, H.J., Bridges, C.B., 1923. The mechanism of Mendelian heredity. H. Holt.

Moritz, C.P., Mühlhaus, T., Tenzer, S., Schulenborg, T., Friauf, E., 2019. Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause. J. Neurochem. 149, 582–604. https://doi.org/10.1111/jnc.14664

Morrill, S.A., Amon, A., 2019. Why haploinsufficiency persists. Proc. Natl. Acad. Sci. 116, 11866–11871. https://doi.org/10.1073/pnas.1900437116

Morris, K.V., Mattick, J.S., 2014. The rise of regulatory RNA. Nat. Rev. Genet. 15, 423–437. https://doi.org/10.1038/nrg3722

Muenzner, J., Trébulle, P., Agostini, F., Messner, C.B., Steger, M., Lehmann, A., Caudal, E., Egger, A.-S., Amari, F., Barthel, N., Chiara, M.D., Mülleder, M., Demichev, V., Liti, G., Schacherer, J.,

Gossmann, T., Berman, J., Ralser, M., 2022. The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids. https://doi.org/10.1101/2022.04.06.487392
Muller, H.J., 1928. The Production of Mutations by X-Rays. Proc. Natl. Acad. Sci. U. S. A. 14, 714–726.
Müller-McNicoll, M., Rossbach, O., Hui, J., Medenbach, J., 2019. Auto-regulatory feedback by RNA-binding proteins. J. Mol. Cell Biol. 11, 930–939. https://doi.org/10.1093/jmcb/mjz043
Mun, D.-G., Bhin, J., Kim, S., Kim, Hyunwoo, Jung, J.H., Jung, Yeonjoo, Jang, Y.E., Park, J.M., Kim, Hokeun, Jung, Yeonhwa, Lee, Hangyeore, Bae, J., Back, S., Kim, S.-J., Kim, Jieun, Park, H., Li, H., Hwang, K.-B., Park, Y.S., Yook, J.H., Kim, B.S., Kwon, S.Y., Ryu, S.W., Park, D.Y., Jeon, T.Y., Kim, D.H., Lee, J.-H., Han, S.-U., Song, K.S., Park, D., Park, J.W., Rodriguez, H., Kim, Jaesang, Lee, Hookeun, Kim, K.P., Yang, E.G., Kim, H.K., Paek, E., Lee, S., Lee, S.-W., Hwang, D., 2019. Proteogenomic Characterization of Human Early-Onset Gastric Cancer. Cancer Cell 35, 111-124.e10. https://doi.org/10.1016/j.ccell.2018.12.003
Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320, 1344–1349. https://doi.org/10.1126/science.1158441
Napthine, S., Ling, R., Finch, L.K., Jones, J.D., Bell, S., Brierley, I., Firth, A.E., 2017. Protein-directed ribosomal frameshifting temporally regulates gene expression. Nat. Commun. 8, 15582. https://doi.org/10.1038/ncomms15582
Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., Weissman, J.S., 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441, 840–846. https://doi.org/10.1038/nature04785
Niarchou, M., Gustavson, D.E., Sathirapongsasuti, J.F., Anglada-Tort, M., Eising, E., Bell, E., McArthur, E., Straub, P., McAuley, J.D., Capra, J.A., Ullén, F., Creanza, N., Mosing, M.A., Hinds, D.A., Davis, L.K., Jacoby, N., Gordon, R.L., 2022. Genome-wide association study of musical beat synchronization demonstrates high polygenicity. Nat. Hum. Behav. 6, 1292–1309. https://doi.org/10.1038/s41562-022-01359-x
Nica, A.C., Dermitzakis, E.T., 2013. Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. B Biol. Sci. 368, 20120362. https://doi.org/10.1098/rstb.2012.0362
Nikolac Perkovic, M., Pivac, N., 2019. Genetic Markers of Alzheimer's Disease. Adv. Exp. Med. Biol. 1192, 27–52. https://doi.org/10.1007/978-981-32-9721-0_3
Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L., Wincker, P., Casaregola, S., Dequin, S., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc. Natl. Acad. Sci. 106, 16333–16338. https://doi.org/10.1073/pnas.0904673106
Nurse, P., Thuriaux, P., 1980. Regulatory genes controlling mitosis in the fission yeast Schizosaccharomyces pombe. Genetics 96, 627–637. https://doi.org/10.1093/genetics/96.3.627
Nurse, P., Thuriaux, P., Nasmyth, K., 1976. Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. Mol. Gen. Genet. MGG 146, 167–178. https://doi.org/10.1007/BF00268085
O'Farrell, P.H., 1975. High Resolution Two-Dimensional Electrophoresis of Proteins. J. Biol. Chem. 250, 4007–4021.
Ohnuki, S., Ohya, Y., 2018. High-dimensional single-cell phenotyping reveals extensive haploinsufficiency. PLOS Biol. 16, e2005130. https://doi.org/10.1371/journal.pbio.2005130
Ohnuki, S., Okada, H., Friedrich, A., Kanno, Y., Goshima, T., Hasuda, H., Inahashi, M., Okazaki, N., Tamura, H., Nakamura, R., Hirata, D., Fukuda, H., Shimoi, H., Kitamoto, K., Watanabe, D., Schacherer, J., Akao, T., Ohya, Y., 2017. Phenotypic Diagnosis of Lineage and Differentiation During Sake Yeast Breeding. G3 GenesGenomesGenetics 7, 2807–2820. https://doi.org/10.1534/g3.117.044099
Olds, L.C., Sibley, E., 2003. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. Hum. Mol. Genet. 12, 2333–2340. https://doi.org/10.1093/hmg/ddg244
Owen, M.J., Williams, N.M., 2021. Explaining the missing heritability of psychiatric disorders. World Psychiatry 20, 294–295. https://doi.org/10.1002/wps.20870
Pasteur, L., 1858. Nouveaux faits concernant l'histoire de la fermentation alcoolique. Comptes Rendus Chim 47, 1011–1013.
Patterson, D., 2009. Molecular genetic analysis of Down syndrome. Hum. Genet. 126, 195–214. https://doi.org/10.1007/s00439-009-0696-8

Pérez-Ortín, J.E., Querol, A., Puig, S., Barrio, E., 2002. Molecular Characterization of a Chromosomal Rearrangement Involved in the Adaptive Evolution of Yeast Strains. Genome Res. 12, 1533–1539. https://doi.org/10.1101/gr.436602

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Ponnala, L., Wang, Y., Sun, Q., van Wijk, K.J., 2014. Correlation of mRNA and protein abundance in the developing maize leaf. Plant J. Cell Mol. Biol. 78, 424–440. https://doi.org/10.1111/tpj.12482

Porta-Pardo, E., Garcia-Alonso, L., Hrabe, T., Dopazo, J., Godzik, A., 2015. A Pan-Cancer Catalogue of Cancer Driver Protein Interaction Interfaces. PLOS Comput. Biol. 11, e1004518. https://doi.org/10.1371/journal.pcbi.1004518

Prandini, P., Deutsch, S., Lyle, R., Gagnebin, M., Vivier, C.D., Delorenzi, M., Gehrig, C., Descombes, P., Sherman, S., Bricarelli, F.D., Baldo, C., Novelli, A., Dallapiccola, B., Antonarakis, S.E., 2007. Natural Gene-Expression Variation in Down Syndrome Modulates the Outcome of Gene-Dosage Imbalance. Am. J. Hum. Genet. 81, 252–263. https://doi.org/10.1086/519248

Prud'homme, B., Gompel, N., Carroll, S.B., 2007. Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. U. S. A. 104 Suppl 1, 8605–8612. https://doi.org/10.1073/pnas.0700488104

Purugganan, M.D., Fuller, D.Q., 2009. The nature of selection during plant domestication. Nature 457, 843–848. https://doi.org/10.1038/nature07895

Qu, W., Gurdziel, K., Pique-Regi, R., Ruden, D.M., 2018. Lead Modulates trans- and cis-Expression Quantitative Trait Loci (eQTLs) in Drosophila melanogaster Heads. Front. Genet. 9.

Ramaswami, G., Geschwind, D.H., 2018. Genetics of autism spectrum disorder. Handb. Clin. Neurol. 147, 321–329. https://doi.org/10.1016/B978-0-444-63233-3.00021-X

Randles, L.G., Lappalainen, I., Fowler, S.B., Moore, B., Hamill, S.J., Clarke, J., 2006. Using Model Proteins to Quantify the Effects of Pathogenic Mutations in Ig-like Proteins. J. Biol. Chem. 281, 24216–24226. https://doi.org/10.1074/jbc.M603593200

Rees, D.C., Williams, T.N., Gladwin, M.T., 2010. Sickle-cell disease. The Lancet 376, 2018–2031. https://doi.org/10.1016/S0140-6736(10)61029-X

Reue, K., 1998. mRNA Quantitation Techniques: Considerations for Experimental Design and Application. J. Nutr. 128, 2038–2044. https://doi.org/10.1093/jn/128.11.2038

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics, SI: Metagenomics of Marine Environments 13, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Ribéreau-Gayon, P., Dubourdieu, D., Donèche, B., Lonvaud, A., 2006. Handbook of Enology, Volume 1: The Microbiology of Wine and Vinifications. John Wiley & Sons.

Richter, J.D., Bassell, G.J., Klann, E., 2015. Dysregulation and restoration of translational homeostasis in fragile X syndrome. Nat. Rev. Neurosci. 16, 595–605. https://doi.org/10.1038/nrn4001

Ridge, P.G., Mukherjee, S., Crane, P.K., Kauwe, J.S.K., Consortium, A.D.G., 2013. Alzheimer's Disease: Analyzing the Missing Heritability. PLOS ONE 8, e79771. https://doi.org/10.1371/journal.pone.0079771

Saada, O.A., Tsouris, A., Large, C., Friedrich, A., Dunham, M.J., Schacherer, J., 2022. Phased polyploid genomes provide deeper insight into the multiple origins of domesticated Saccharomyces cerevisiae beer yeasts. Curr. Biol. 32, 1350-1361.e3. https://doi.org/10.1016/j.cub.2022.01.068

Salovska, B., Zhu, H., Gandhi, T., Frank, M., Li, W., Rosenberger, G., Wu, C., Germain, P.-L., Zhou, H., Hodny, Z., Reiter, L., Liu, Y., 2020. Isoform-resolved correlation analysis between mRNA abundance regulation and protein level degradation. Mol. Syst. Biol. 16, e9170. https://doi.org/10.15252/msb.20199170

Sambrook, J., Russell, D.W. (David W., 2001. Molecular cloning: a laboratory manual. Vol. 1, 3rd ed. ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Samuel, D., 1996. Investigation of Ancient Egyptian Baking and Brewing Methods by Correlative Microscopy. Science 273, 488–490. https://doi.org/10.1126/science.273.5274.488

Sanda, T., Lawton, L.N., Barrasa, M.I., Fan, Z.P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, M.A., Jamieson, C.H.M., Staudt, L.M., Young, R.A., Look, A.T., 2012. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell 22, 209–221. https://doi.org/10.1016/j.ccr.2012.06.007

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.
Saramago, M., da Costa, P.J., Viegas, S.C., Arraiano, C.M., 2019. The Implication of mRNA Degradation Disorders on Human DISease: Focus on DIS3 and DIS3-Like Enzymes, in: Romão, L. (Ed.), The MRNA Metabolism in Human Disease, Advances in Experimental Medicine and Biology. Springer International Publishing, Cham, pp. 85–98. https://doi.org/10.1007/978-3-030-19966-1_4
Sauna, Z.E., Kimchi-Sarfaty, C., 2013. Synonymous Mutations as a Cause of Human Genetic Disease, in: ELS. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0025173
Schadt, E.E., Monks, S.A., Drake, T.A., Lusis, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G., Linsley, P.S., Mao, M., Stoughton, R.B., Friend, S.H., 2003. Genetics of gene expression surveyed in maize, mouse and man. Nature 422, 297–302. https://doi.org/10.1038/nature01434
Schaefke, B., Emerson, J.J., Wang, T.-Y., Lu, M.-Y.J., Hsieh, L.-C., Li, W.-H., 2013. Inheritance of gene expression level and selective constraints on trans- and cis-regulatory changes in yeast. Mol. Biol. Evol. 30, 2121–2133. https://doi.org/10.1093/molbev/mst114
Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. Science 270, 467–470. https://doi.org/10.1126/science.270.5235.467
Schlattl, A., Anders, S., Waszak, S.M., Huber, W., Korbel, J.O., 2011. Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. Genome Res. 21, 2004–2013. https://doi.org/10.1101/gr.122614.111
Schrimpf, S.P., Weiss, M., Reiter, L., Ahrens, C.H., Jovanovic, M., Malmström, J., Brunner, E., Mohanty, S., Lercher, M.J., Hunziker, P.E., Aebersold, R., von Mering, C., Hengartner, M.O., 2009. Comparative Functional Analysis of the Caenorhabditis elegans and Drosophila melanogaster Proteomes. PLoS Biol. 7. https://doi.org/10.1371/journal.pbio.1000048
Schubert, S.A., Morreau, H., de Miranda, N.F.C.C., van Wezel, T., 2020. The missing heritability of familial colorectal cancer. Mutagenesis 35, 221–231. https://doi.org/10.1093/mutage/gez027
Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M., 2011. Global quantification of mammalian gene expression control. Nature 473, 337–342. https://doi.org/10.1038/nature10098
Shi, L., Guo, Y., Dong, C., Huddleston, J., Yang, H., Han, X., Fu, A., Li, Q., Li, N., Gong, S., Lintner, K.E., Ding, Q., Wang, Z., Hu, J., Wang, D., Wang, F., Wang, L., Lyon, G.J., Guan, Y., Shen, Y., Evgrafov, O.V., Knowles, J.A., Thibaud-Nissen, F., Schneider, V., Yu, C.-Y., Zhou, L., Eichler, E.E., So, K.-F., Wang, K., 2016. Long-read sequencing and de novo assembly of a Chinese genome. Nat. Commun. 7, 12065. https://doi.org/10.1038/ncomms12065
Shiroma, S., Jayakody, L.N., Horie, K., Okamoto, K., Kitagaki, H., 2014. Enhancement of ethanol fermentation in Saccharomyces cerevisiae sake yeast by disrupting mitophagy function. Appl. Environ. Microbiol. 80, 1002–1012. https://doi.org/10.1128/AEM.03130-13
Shortle, D., Haber, J.E., Botstein, D., 1982. Lethal Disruption of the Yeast Actin Gene by Integrative DNA Transformation. Science 217, 371–373. https://doi.org/10.1126/science.7046050
Smith, E.N., Kruglyak, L., 2008. Gene–Environment Interaction in Yeast Gene Expression. PLOS Biol. 6, e83. https://doi.org/10.1371/journal.pbio.0060083
Stark, R., Grzelak, M., Hadfield, J., 2019. RNA sequencing: the teenage years. Nat. Rev. Genet. 20, 631–656. https://doi.org/10.1038/s41576-019-0150-2
Steenwyk, J.L., Rokas, A., 2018. Copy Number Variation in Fungi and Its Implications for Wine Yeast Genetic Diversity and Adaptation. Front. Microbiol. 9, 288. https://doi.org/10.3389/fmicb.2018.00288
Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., Storchova, Z., 2012. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol. Syst. Biol. 8, 608. https://doi.org/10.1038/msb.2012.40
Strope, P.K., Skelly, D.A., Kozmin, S.G., Mahadevan, G., Stone, E.A., Magwene, P.M., Dietrich, F.S., McCusker, J.H., 2015. The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. Genome Res. 25, 762–774. https://doi.org/10.1101/gr.185538.114
Su, G., Christensen, O.F., Ostersen, T., Henryon, M., Lund, M.S., 2012. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single

Nucleotide Polymorphism Markers. PLoS ONE 7, e45293. https://doi.org/10.1371/journal.pone.0045293
Sud, A., Kinnersley, B., Houlston, R.S., 2017. Genome-wide association studies of cancer: current insights and future perspectives. Nat. Rev. Cancer 17, 692–704. https://doi.org/10.1038/nrc.2017.82
Suhre, K., McCarthy, M.I., Schwenk, J.M., 2021. Genetics meets proteomics: perspectives for large population-based studies. Nat. Rev. Genet. 22, 19–37. https://doi.org/10.1038/s41576-020-0268-2
Sundd, P., Gladwin, M.T., Novelli, E.M., 2019. Pathophysiology of Sickle Cell Disease. Annu. Rev. Pathol. Mech. Dis. 14, 263–292. https://doi.org/10.1146/annurev-pathmechdis-012418-012838
Taggart, J.C., Li, G.-W., 2018. Production of Protein-Complex Components is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Syst. 7, 580-589.e4. https://doi.org/10.1016/j.cels.2018.11.003
Taggart, J.C., Zauber, H., Selbach, M., Li, G.-W., McShane, E., 2020. Keeping the Proportions of Protein Complex Components in Check. Cell Syst. 10, 125–132. https://doi.org/10.1016/j.cels.2020.01.004
Tak, Y.G., Farnham, P.J., 2015. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin 8, 57. https://doi.org/10.1186/s13072-015-0050-4
Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. 20, 467–484. https://doi.org/10.1038/s41576-019-0127-1
Tarca, A.L., Romero, R., Draghici, S., 2006. Analysis of microarray experiments of gene expression profiling. Am. J. Obstet. Gynecol. 195, 373–388. https://doi.org/10.1016/j.ajog.2006.07.001
The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393
The GTEx Consortium, 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330. https://doi.org/10.1126/science.aaz1776
The GTEx Consortium, 2017. Genetic effects on gene expression across human tissues. Nature 550, 204–213. https://doi.org/10.1038/nature24277
The GTEx Consortium, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110
The UK10K Consortium, 2015. The UK10K project identifies rare variants in health and disease. Nature 526, 82–90. https://doi.org/10.1038/nature14962
Tirosh, I., Barkai, N., Verstrepen, K.J., 2009. Promoter architecture and the evolvability of gene expression. J. Biol. 8, 95. https://doi.org/10.1186/jbiol204
Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C., Davis, T.N., Nislow, C., Brown, G.W., 2012. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. Nat. Cell Biol. 14, 966–976. https://doi.org/10.1038/ncb2549
Towbin, H., Staehelin, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. 76, 4350–4354. https://doi.org/10.1073/pnas.76.9.4350
Udagawa, T., Farny, N.G., Jakovcevski, M., Kaphzan, H., Alarcon, J.M., Anilkumar, S., Ivshina, M., Hurt, J.A., Nagaoka, K., Nalavadi, V.C., Lorenz, L.J., Bassell, G.J., Akbarian, S., Chattarji, S., Klann, E., Richter, J.D., 2013. Genetic and acute CPEB1 depletion ameliorate fragile X pathophysiology. Nat. Med. 19, 1473–1477. https://doi.org/10.1038/nm.3353
Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., Posthuma, D., 2021. Genome-wide association studies. Nat. Rev. Methods Primer 1, 1–21. https://doi.org/10.1038/s43586-021-00056-9
Upadhya, S.R., Ryan, C.J., 2022. Experimental reproducibility limits the correlation between mRNA and protein abundances in tumor proteomic profiles. Cell Rep. Methods 2, 100288. https://doi.org/10.1016/j.crmeth.2022.100288
VanInsberghe, M., van den Berg, J., Andersson-Rolf, A., Clevers, H., van Oudenaarden, A., 2021. Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. Nature 597, 561–565. https://doi.org/10.1038/s41586-021-03887-4
Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., Gritsenko, M.A., Zimmerman, L.J., McDermott, J.E., Clauss, T.R., Moore, R.J., Zhao, R., Monroe, M.E., Wang, Y.-T., Chambers, M.C., Slebos, R.J.C., Lau, K.S., Mo, Q., Ding, L., Ellis, M., Thiagarajan, M., Kinsinger, C.R., Rodriguez, H., Smith, R.D., Rodland, K.D., Liebler, D.C., Liu, T.,

Zhang, B., Pandey, A., Paulovich, A., Hoofnagle, A., Mani, D.R., Chan, D.W., Ransohoff, D.F., Fenyo, D., Tabb, D.L., Levine, D.A., Boja, E.S., Kuhn, E., White, F.M., Whiteley, G.A., Zhu, H., Zhang, H., Shih, I.-M., Bavarva, J., Whiteaker, J., Ketchum, K.A., Clauser, K.R., Ruggles, K., Elburn, K., Hannick, L., Watson, M., Oberti, M., Mesri, M., Sanders, M.E., Borucki, M., Gillette, M.A., Snyder, M., Edwards, N.J., Vatanian, N., Rudnick, P.A., McGarvey, P.B., Mertins, P., Townsend, R.R., Thangudu, R.R., Rivers, R.C., Payne, S.H., Davies, S.R., Cai, S., Stein, S.E., Carr, S.A., Skates, S.J., Madhavan, S., Hiltke, T., Chen, X., Zhao, Y., Wang, Y., Zhang, Z., 2019. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell 177, 1035-1049.e19. https://doi.org/10.1016/j.cell.2019.03.030

Veitia, R.A., Potier, M.C., 2015. Gene dosage imbalances: action, reaction, and models. Trends Biochem. Sci. 40, 309–317. https://doi.org/10.1016/j.tibs.2015.03.011

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65, 905–914. https://doi.org/10.1016/0092-8674(91)90397-h

Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. 13, 227–232. https://doi.org/10.1038/nrg3185

Wach, A., Brachat, A., Pöhlmann, R., Philippsen, P., 1994. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast Chichester Engl. 10, 1793–1808. https://doi.org/10.1002/yea.320101310

Wacholder, A., Parikh, S.B., Coelho, N.C., Acar, O., Houghton, C., Chou, L., Carvunis, A.-R., 2023. A vast evolutionarily transient translatome contributes to phenotype and fitness. Cell Syst. 14, 363-381.e8. https://doi.org/10.1016/j.cels.2023.04.002

Wang, A.M., Doyle, M.V., Mark, D.F., 1989. Quantitation of mRNA by the polymerase chain reaction. Proc. Natl. Acad. Sci. 86, 9717–9721. https://doi.org/10.1073/pnas.86.24.9717

Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L., Meng, C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H., Kuster, B., 2019. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol. Syst. Biol. 15, e8503. https://doi.org/10.15252/msb.20188503

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. https://doi.org/10.1038/nature07509

Wang, Yunhao, Zhao, Y., Bollas, A., Wang, Yuru, Au, K.F., 2021. Nanopore sequencing technology, bioinformatics and applications. Nat. Biotechnol. 39, 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wang, Z., Moult, J., 2001. SNPs, protein structure, and disease. Hum. Mutat. 17, 263–270. https://doi.org/10.1002/humu.22

Wang, Z., Sun, X., Zhao, Y., Guo, X., Jiang, H., Li, H., Gu, Z., 2015. Evolution of Gene Regulation during Transcription and Translation. Genome Biol. Evol. 7, 1155–1167. https://doi.org/10.1093/gbe/evv059

Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M.E., Peters, A.H.F.M., Anders, S., Kaessmann, H., 2020. Transcriptome and translatome co-evolution in mammals. Nature 1–6. https://doi.org/10.1038/s41586-020-2899-z

Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., Humphery-Smith, I., 1995. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis 16, 1090–1094. https://doi.org/10.1002/elps.11501601185

Watanabe, D., Nogami, S., Ohya, Y., Kanno, Y., Zhou, Y., Akao, T., Shimoi, H., 2011. Ethanol fermentation driven by elevated expression of the G1 cyclin gene CLN3 in sake yeast. J. Biosci. Bioeng. 112, 577–582. https://doi.org/10.1016/j.jbiosc.2011.08.010

Watson, J.D., Crick, F.H.C., 1953. The Structure of Dna. Cold Spring Harb. Symp. Quant. Biol. 18, 123–131. https://doi.org/10.1101/SQB.1953.018.01.020

Wellcome Trust Case Control Consortium, 2010. Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464, 713–720. https://doi.org/10.1038/nature08979

Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678. https://doi.org/10.1038/nature05911

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. 37, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., Zhernakova, A., Zhernakova, D.V., Veldink, J.H., Van den Berg, L.H., Karjalainen, J., Withoff, S., Uitterlinden, A.G., Hofman, A., Rivadeneira, F., 't Hoen, P.A.C., Reinmaa, E., Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Nalls, M.A., Homuth, G., Nauck, M., Radke, D., Völker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S.A., Enquobahrie, D.A., Lumley, T., Montgomery, G.W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R.C., Visscher, P.M., Knight, J.C., Psaty, B.M., Ripatti, S., Teumer, A., Frayling, T.M., Metspalu, A., van Meurs, J.B.J., Franke, L., 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243. https://doi.org/10.1038/ng.2756

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., Kuster, B., 2014. Mass-spectrometry-based draft of the human proteome. Nature 509, 582–587. https://doi.org/10.1038/nature13319

Williams, C.C., Jan, C.H., Weissman, J.S., 2014. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science 346, 748–751. https://doi.org/10.1126/science.1257522

Wirka, R.C., Pjanic, M., Quertermous, T., 2018. Advances in Transcriptomics. Circ. Res. 122, 1200–1220. https://doi.org/10.1161/CIRCRESAHA.117.310910

Wolfe, K.H., Shields, D.C., 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387, 708–713. https://doi.org/10.1038/42711

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Buchkovich, M.L., Croteau-Chonka, D.C., Day, F.R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A.U., Karjalainen, J., Lo, K.S., Locke, A.E., Mägi, R., Mihailov, E., Porcu, E., Randall, J.C., Scherag, A., Vinkhuyzen, A.A.E., Westra, H.-J., Winkler, T.W., Workalemahu, T., Zhao, J.H., Absher, D., Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Fraser, R.M., Goel, A., Gong, J., Justice, A.E., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Lui, J.C., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Nalls, M.A., Nyholt, D.R., Palmer, C.D., Pasko, D., Pechlivanis, S., Prokopenko, I., Ried, J.S., Ripke, S., Shungin, D., Stancáková, A., Strawbridge, R.J., Sung, Y.J., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S.W., van Setten, J., Van Vliet-Ostaptchouk, J.V., Wang, Z., Yengo, L., Zhang, W., Afzal, U., Arnlöv, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C., Blüher, M., Bolton, J.L., Böttcher, Y., Boyd, H.A., Bruinenberg, M., Buckley, B.M., Buyske, S., Caspersen, I.H., Chines, P.S., Clarke, R., Claudi-Boehm, S., Cooper, M., Daw, E.W., De Jong, P.A., Deelen, J., Delgado, G., Denny, J.C., Dhonukshe-Rutten, R., Dimitriou, M., Doney, A.S.F., Dörr, M., Eklund, N., Eury, E., Folkersen, L., Garcia, M.E., Geller, F., Giedraitis, V., Go, A.S., Grallert, H., Grammer, T.B., Gräßler, J., Grönberg, H., de Groot, L.C.P.G.M., Groves, C.J., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hannemann, A., Hartman, C.A., Hassinen, M., Hayward, C., Heard-Costa, N.L., Helmer, Q., Hemani, G., Henders, A.K., Hillege, H.L., Hlatky, M.A., Hoffmann, W., Hoffmann, P., Holmen, O., Houwing-Duistermaat, J.J., Illig, T., Isaacs, A., James, A.L., Jeff, J., Johansen, B., Johansson, Å., Jolley, J., Juliusdottir, T., Junttila, J., Kho, A.N., Kinnunen, L., Klopp, N., Kocher, T., Kratzer, W., Lichtner, P., Lind, L., Lindström, J., Lobbens, S., Lorentzon, M., Lu, Y., Lyssenko, V., Magnusson, P.K.E., Mahajan, A., Maillard, M., McArdle, W.L., McKenzie, C.A., McLachlan, S., McLaren, P.J., Menni, C., Merger, S., Milani, L., Moayyeri, A., Monda, K.L., Morken, M.A., Müller, G., Müller-Nurasyid, M., Musk, A.W., Narisu, N., Nauck, M., Nolte, I.M., Nöthen, M.M., Oozageer, L., Pilz, S., Rayner,

N.W., Renstrom, F., Robertson, N.R., Rose, L.M., Roussel, R., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F.R., Schunkert, H., Scott, R.A., Sehmi, J., Seufferlein, T., Shi, J., Silventoinen, K., Smit, J.H., Smith, A.V., Smolonska, J., Stanton, A.V., Stirrups, K., Stott, D.J., Stringham, H.M., Sundström, J., Swertz, M.A., Syvänen, A.-C., Tayo, B.O., Thorleifsson, G., Tyrer, J.P., van Dijk, S., van Schoor, N.M., van der Velde, N., van Heemst, D., van Oort, F.V.A., Vermeulen, S.H., Verweij, N., Vonk, J.M., Waite, L.L., Waldenberger, M., Wennauer, R., Wilkens, L.R., Willenborg, C., Wilsgaard, T., Wojczynski, M.K., Wong, A., Wright, A.F., Zhang, Q., Arveiler, D., Bakker, S.J.L., Beilby, J., Bergman, R.N., Bergmann, S., Biffar, R., Blangero, J., Boomsma, D.I., Bornstein, S.R., Bovet, P., Brambilla, P., Brown, M.J., Campbell, H., Caulfield, M.J., Chakravarti, A., Collins, R., Collins, F.S., Crawford, D.C., Cupples, L.A., Danesh, J., de Faire, U., den Ruijter, H.M., Erbel, R., Erdmann, J., Eriksson, J.G., Farrall, M., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N.G., Forrester, T., Gansevoort, R.T., Gejman, P.V., Gieger, C., Golay, A., Gottesman, O., Gudnason, V., Gyllensten, U., Haas, D.W., Hall, A.S., Harris, T.B., Hattersley, A.T., Heath, A.C., Hengstenberg, C., Hicks, A.A., Hindorff, L.A., Hingorani, A.D., Hofman, A., Hovingh, G.K., Humphries, S.E., Hunt, S.C., Hypponen, E., Jacobs, K.B., Jarvelin, M.-R., Jousilahti, P., Jula, A.M., Kaprio, J., Kastelein, J.J.P., Kayser, M., Kee, F., Keinanen-Kiukaanniemi, S.M., Kiemeney, L.A., Kooner, J.S., Kooperberg, C., Koskinen, S., Kovacs, P., Kraja, A.T., Kumari, M., Kuusisto, J., Lakka, T.A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lupoli, S., Madden, P.A.F., Männistö, S., Manunta, P., Marette, A., Matise, T.C., McKnight, B., Meitinger, T., Moll, F.L., Montgomery, G.W., Morris, A.D., Morris, A.P., Murray, J.C., Nelis, M., Ohlsson, C., Oldehinkel, A.J., Ong, K.K., Ouwehand, W.H., Pasterkamp, G., Peters, A., Pramstaller, P.P., Price, J.F., Qi, L., Raitakari, O.T., Rankinen, T., Rao, D.C., Rice, T.K., Ritchie, M., Rudan, I., Salomaa, V., Samani, N.J., Saramies, J., Sarzynski, M.A., Schwarz, P.E.H., Sebert, S., Sever, P., Shuldiner, A.R., Sinisalo, J., Steinthorsdottir, V., Stolk, R.P., Tardif, J.-C., Tönjes, A., Tremblay, A., Tremoli, E., Virtamo, J., Vohl, M.-C., Electronic Medical Records and Genomics (eMEMERGEGE) Consortium, MIGen Consortium, PAGEGE Consortium, LifeLines Cohort Study, Amouyel, P., Asselbergs, F.W., Assimes, T.L., Bochud, M., Boehm, B.O., Boerwinkle, E., Bottinger, E.P., Bouchard, C., Cauchi, S., Chambers, J.C., Chanock, S.J., Cooper, R.S., de Bakker, P.I.W., Dedoussis, G., Ferrucci, L., Franks, P.W., Froguel, P., Groop, L.C., Haiman, C.A., Hamsten, A., Hayes, M.G., Hui, J., Hunter, D.J., Hveem, K., Jukema, J.W., Kaplan, R.C., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N.G., März, W., Melbye, M., Moebus, S., Munroe, P.B., Njølstad, I., Oostra, B.A., Palmer, C.N.A., Pedersen, N.L., Perola, M., Pérusse, L., Peters, U., Powell, J.E., Power, C., Quertermous, T., Rauramaa, R., Reinmaa, E., Ridker, P.M., Rivadeneira, F., Rotter, J.I., Saaristo, T.E., Saleheen, D., Schlessinger, D., Slagboom, P.E., Snieder, H., Spector, T.D., Strauch, K., Stumvoll, M., Tuomilehto, J., Uusitupa, M., van der Harst, P., Völzke, H., Walker, M., Wareham, N.J., Watkins, H., Wichmann, H.-E., Wilson, J.F., Zanen, P., Deloukas, P., Heid, I.M., Lindgren, C.M., Mohlke, K.L., Speliotes, E.K., Thorsteinsdottir, U., Barroso, I., Fox, C.S., North, K.E., Strachan, D.P., Beckmann, J.S., Berndt, S.I., Boehnke, M., Borecki, I.B., McCarthy, M.I., Metspalu, A., Stefansson, K., Uitterlinden, A.G., van Duijn, C.M., Franke, L., Willer, C.J., Price, A.L., Lettre, G., Loos, R.J.F., Weedon, M.N., Ingelsson, E., O'Connell, J.R., Abecasis, G.R., Chasman, D.I., Goddard, M.E., Visscher, P.M., Hirschhorn, J.N., Frayling, T.M., 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46, 1173–1186. https://doi.org/10.1038/ng.3097

Yamamoto, F., Clausen, H., White, T., Marken, J., Hakomori, S., 1990. Molecular genetic basis of the histo-blood group ABO system. Nature 345, 229–233. https://doi.org/10.1038/345229a0

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., LifeLines Cohort Study, Esko, T., Milani, L., Mägi, R., Metspalu, A., Hamsten, A., Magnusson, P.K.E., Pedersen, N.L., Ingelsson, E., Soranzo, N., Keller, M.C., Wray, N.R., Goddard, M.E., Visscher, P.M., 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47, 1114–1120. https://doi.org/10.1038/ng.3390

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569. https://doi.org/10.1038/ng.608

Yao, C., Joehanes, R., Johnson, A.D., Huan, T., Liu, C., Freedman, J.E., Munson, P.J., Hill, D.E., Vidal, M., Levy, D., 2017. Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. Am. J. Hum. Genet. 100, 571–580. https://doi.org/10.1016/j.ajhg.2017.02.003

Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., the GIANT Consortium, 2018. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum. Mol. Genet. 27, 3641–3649. https://doi.org/10.1093/hmg/ddy271

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S., Miao, J., Arias, J.D., Graham, S.E., Mukamel, R.E., Spracklen, C.N., Yin, X., Chen, S.-H., Ferreira, T., Highland, H.H., Ji, Y., Karaderi, T., Lin, K., Lüll, K., Malden, D.E., Medina-Gomez, C., Machado, M., Moore, A., Rüeger, S., Sim, X., Vrieze, S., Ahluwalia, T.S., Akiyama, M., Allison, M.A., Alvarez, M., Andersen, M.K., Ani, A., Appadurai, V., Arbeeva, L., Bhaskar, S., Bielak, L.F., Bollepalli, S., Bonnycastle, L.L., Bork-Jensen, J., Bradfield, J.P., Bradford, Y., Braund, P.S., Brody, J.A., Burgdorf, K.S., Cade, B.E., Cai, H., Cai, Q., Campbell, A., Cañadas-Garre, M., Catamo, E., Chai, J.-F., Chai, X., Chang, L.-C., Chang, Y.-C., Chen, C.-H., Chesi, A., Choi, S.H., Chung, R.-H., Cocca, M., Concas, M.P., Couture, C., Cuellar-Partida, G., Danning, R., Daw, E.W., Degenhard, F., Delgado, G.E., Delitala, A., Demirkan, A., Deng, X., Devineni, P., Dietl, A., Dimitriou, M., Dimitrov, L., Dorajoo, R., Ekici, A.B., Engmann, J.E., Fairhurst-Hunter, Z., Farmaki, A.-E., Faul, J.D., Fernandez-Lopez, J.-C., Forer, L., Francescatto, M., Freitag-Wolf, S., Fuchsberger, C., Galesloot, T.E., Gao, Y., Gao, Z., Geller, F., Giannakopoulou, O., Giulianini, F., Gjesing, A.P., Goel, A., Gordon, S.D., Gorski, M., Grove, J., Guo, X., Gustafsson, S., Haessler, J., Hansen, T.F., Havulinna, A.S., Haworth, S.J., He, J., Heard-Costa, N., Hebbar, P., Hindy, G., Ho, Y.-L.A., Hofer, E., Holliday, E., Horn, K., Hornsby, W.E., Hottenga, J.-J., Huang, H., Huang, J., Huerta-Chagoya, A., Huffman, J.E., Hung, Y.-J., Huo, S., Hwang, M.Y., Iha, H., Ikeda, D.D., Isono, M., Jackson, A.U., Jäger, S., Jansen, I.E., Johansson, I., Jonas, J.B., Jonsson, A., Jørgensen, T., Kalafati, I.-P., Kanai, M., Kanoni, S., Kårhus, L.L., Kasturiratne, A., Katsuya, T., Kawaguchi, T., Kember, R.L., Kentistou, K.A., Kim, H.-N., Kim, Y.J., Kleber, M.E., Knol, M.J., Kurbasic, A., Lauzon, M., Le, P., Lea, R., Lee, J.-Y., Leonard, H.L., Li, S.A., Li, Xiaohui, Li, Xiaoyin, Liang, J., Lin, H., Lin, S.-Y., Liu, Jun, Liu, X., Lo, K.S., Long, J., Lores-Motta, L., Luan, J., Lyssenko, V., Lyytikäinen, L.-P., Mahajan, A., Mamakou, V., Mangino, M., Manichaikul, A., Marten, J., Mattheisen, M., Mavarani, L., McDaid, A.F., Meidtner, K., Melendez, T.L., Mercader, J.M., Milaneschi, Y., Miller, J.E., Millwood, I.Y., Mishra, P.P., Mitchell, R.E., Møllehave, L.T., Morgan, A., Mucha, S., Munz, M., Nakatochi, M., Nelson, C.P., Nethander, M., Nho, C.W., Nielsen, A.A., Nolte, I.M., Nongmaithem, S.S., Noordam, R., Ntalla, I., Nutile, T., Pandit, A., Christofidou, P., Pärna, K., Pauper, M., Petersen, E.R.B., Petersen, L.V., Pitkänen, N., Polašek, O., Poveda, A., Preuss, M.H., Pyarajan, S., Raffield, L.M., Rakugi, H., Ramirez, J., Rasheed, A., Raven, D., Rayner, N.W., Riveros, C., Rohde, R., Ruggiero, D., Ruotsalainen, S.E., Ryan, K.A., Sabater-Lleal, M., Saxena, R., Scholz, M., Sendamarai, A., Shen, B., Shi, J., Shin, J.H., Sidore, C., Sitlani, C.M., Slieker, R.C., Smit, R.A.J., Smith, A.V., Smith, J.A., Smyth, L.J., Southam, L., Steinthorsdottir, V., Sun, L., Takeuchi, F., Tallapragada, D.S.P., Taylor, K.D., Tayo, B.O., Tcheandjieu, C., Terzikhan, N., Tesolin, P., Teumer, A., Theusch, E., Thompson, D.J., Thorleifsson, G., Timmers, P.R.H.J., Trompet, S., Turman, C., Vaccargiu, S., van der Laan, S.W., van der Most, P.J., van Klinken, J.B., van Setten, J., Verma, S.S., Verweij, N., Veturi, Y., Wang, C.A., Wang, C., Wang, L., Wang, Z., Warren, H.R., Bin Wei, W., Wickremasinghe, A.R., Wielscher, M., Wiggins, K.L., Winsvold, B.S., Wong, A., Wu, Y., Wuttke, M., Xia, R., Xie, T., Yamamoto, K., Yang, Jingyun, Yao, J., Young, H., Yousri, N.A., Yu, L., Zeng, L., Zhang, W., Zhang, X., Zhao, J.-H., Zhao, W., Zhou, W., Zimmermann, M.E., Zoledziewska, M., Adair, L.S., Adams, H.H.H., Aguilar-Salinas, C.A., Al-Mulla, F., Arnett, D.K., Asselbergs, F.W., Åsvold, B.O., Attia, J., Banas, B., Bandinelli, S., Bennett, D.A., Bergler, T., Bharadwaj, D., Biino, G., Bisgaard, H., Boerwinkle, E., Böger, C.A., Bønnelykke, K., Boomsma, D.I., Børglum, A.D., Borja, J.B., Bouchard, C., Bowden, D.W., Brandslund, I., Brumpton, B., Buring, J.E., Caulfield, M.J., Chambers, J.C., Chandak, G.R., Chanock, S.J., Chaturvedi, N., Chen, Y.-D.I., Chen, Z., Cheng, C.-Y., Christophersen, I.E., Ciullo, M., Cole, J.W., Collins, F.S., Cooper, R.S., Cruz, M., Cucca, F., Cupples, L.A., Cutler, M.J., Damrauer, S.M., Dantoft, T.M., de Borst, G.J., de Groot, L.C.P.G.M., De Jager, P.L., de Kleijn, D.P.V., Janaka de Silva, H., Dedoussis, G.V., den Hollander, A.I., Du, S., Easton, D.F., Elders, P.J.M., Eliassen, A.H., Ellinor, P.T., Elmståhl, S., Erdmann, J., Evans, M.K., Fatkin, D., Feenstra, B., Feitosa, M.F., Ferrucci, L., Ford, I., Fornage, M., Franke, A., Franks, P.W., Freedman, B.I., Gasparini, P., Gieger,

C., Girotto, G., Goddard, M.E., Golightly, Y.M., Gonzalez-Villalpando, C., Gordon-Larsen, P., Grallert, H., Grant, S.F.A., Grarup, N., Griffiths, L., Gudnason, V., Haiman, C., Hakonarson, H., Hansen, T., Hartman, C.A., Hattersley, A.T., Hayward, C., Heckbert, S.R., Heng, C.-K., Hengstenberg, C., Hewitt, A.W., Hishigaki, H., Hoyng, C.B., Huang, P.L., Huang, W., Hunt, S.C., Hveem, K., Hyppönen, E., Iacono, W.G., Ichihara, S., Ikram, M.A., Isasi, C.R., Jackson, R.D., Jarvelin, M.-R., Jin, Z.-B., Jöckel, K.-H., Joshi, P.K., Jousilahti, P., Jukema, J.W., Kähönen, M., Kamatani, Y., Kang, K.D., Kaprio, J., Kardia, S.L.R., Karpe, F., Kato, N., Kee, F., Kessler, T., Khera, A.V., Khor, C.C., Kiemeney, L.A.L.M., Kim, B.-J., Kim, E.K., Kim, H.-L., Kirchhof, P., Kivimaki, M., Koh, W.-P., Koistinen, H.A., Kolovou, G.D., Kooner, J.S., Kooperberg, C., Köttgen, A., Kovacs, P., Kraaijeveld, A., Kraft, P., Krauss, R.M., Kumari, M., Kutalik, Z., Laakso, M., Lange, L.A., Langenberg, C., Launer, L.J., Le Marchand, L., Lee, H., Lee, N.R., Lehtimäki, T., Li, H., Li, L., Lieb, W., Lin, X., Lind, L., Linneberg, A., Liu, C.-T., Liu, Jianjun, Loeffler, M., London, B., Lubitz, S.A., Lye, S.J., Mackey, D.A., Mägi, R., Magnusson, P.K.E., Marcus, G.M., Vidal, P.M., Martin, N.G., März, W., Matsuda, F., McGarrah, R.W., McGue, M., McKnight, A.J., Medland, S.E., Mellström, D., Metspalu, A., Mitchell, B.D., Mitchell, P., Mook-Kanamori, D.O., Morris, A.D., Mucci, L.A., Munroe, P.B., Nalls, M.A., Nazarian, S., Nelson, A.E., Neville, M.J., Newton-Cheh, C., Nielsen, C.S., Nöthen, M.M., Ohlsson, C., Oldehinkel, A.J., Orozco, L., Pahkala, K., Pajukanta, P., Palmer, C.N.A., Parra, E.J., Pattaro, C., Pedersen, O., Pennell, C.E., Penninx, B.W.J.H., Perusse, L., Peters, A., Peyser, P.A., Porteous, D.J., Posthuma, D., Power, C., Pramstaller, P.P., Province, M.A., Qi, Q., Qu, J., Rader, D.J., Raitakari, O.T., Ralhan, S., Rallidis, L.S., Rao, D.C., Redline, S., Reilly, D.F., Reiner, A.P., Rhee, S.Y., Ridker, P.M., Rienstra, M., Ripatti, S., Ritchie, M.D., Roden, D.M., Rosendaal, F.R., Rotter, J.I., Rudan, I., Rutters, F., Sabanayagam, C., Saleheen, D., Salomaa, V., Samani, N.J., Sanghera, D.K., Sattar, N., Schmidt, B., Schmidt, H., Schmidt, R., Schulze, M.B., Schunkert, H., Scott, L.J., Scott, R.J., Sever, P., Shiroma, E.J., Shoemaker, M.B., Shu, X.-O., Simonsick, E.M., Sims, M., Singh, J.R., Singleton, A.B., Sinner, M.F., Smith, J.G., Snieder, H., Spector, T.D., Stampfer, M.J., Stark, K.J., Strachan, D.P., ‘t Hart, L.M., Tabara, Y., Tang, H., Tardif, J.-C., Thanaraj, T.A., Timpson, N.J., Tönjes, A., Tremblay, A., Tuomi, T., Tuomilehto, J., Tusié-Luna, M.-T., Uitterlinden, A.G., van Dam, R.M., van der Harst, P., Van der Velde, N., van Duijn, C.M., van Schoor, N.M., Vitart, V., Völker, U., Vollenweider, P., Völzke, H., Wacher-Rodarte, N.H., Walker, M., Wang, Y.X., Wareham, N.J., Watanabe, R.M., Watkins, H., Weir, D.R., Werge, T.M., Widen, E., Wilkens, L.R., Willemsen, G., Willett, W.C., Wilson, J.F., Wong, T.-Y., Woo, J.-T., Wright, A.F., Wu, J.-Y., Xu, H., Yajnik, C.S., Yokota, M., Yuan, J.-M., Zeggini, E., Zemel, B.S., Zheng, W., Zhu, X., Zmuda, J.M., Zonderman, A.B., Zwart, J.-A., Chasman, D.I., Cho, Y.S., Heid, I.M., McCarthy, M.I., Ng, M.C.Y., O’Donnell, C.J., Rivadeneira, F., Thorsteinsdottir, U., Sun, Y.V., Tai, E.S., Boehnke, M., Deloukas, P., Justice, A.E., Lindgren, C.M., Loos, R.J.F., Mohlke, K.L., North, K.E., Stefansson, K., Walters, R.G., Winkler, T.W., Young, K.L., Loh, P.-R., Yang, Jian, Esko, T., Assimes, T.L., Auton, A., Abecasis, G.R., Willer, C.J., Locke, A.E., Berndt, S.I., Lettre, G., Frayling, T.M., Okada, Y., Wood, A.R., Visscher, P.M., Hirschhorn, J.N., 2022. A saturated map of common genetic variants associated with human height. Nature 610, 704–712. https://doi.org/10.1038/s41586-022-05275-y

Yip, Y.L., Famiglietti, M., Gos, A., Duek, P.D., David, F.P.A., Gateau, A., Bairoch, A., 2008. Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase. Hum. Mutat. 29, 361–366. https://doi.org/10.1002/humu.20671

Yofe, I., Weill, U., Meurer, M., Chuartzman, S., Zalckvar, E., Goldman, O., Ben-Dor, S., Schütze, C., Wiedemann, N., Knop, M., Khmelinskii, A., Schuldiner, M., 2016. One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. Nat. Methods 13, 371–378. https://doi.org/10.1038/nmeth.3795

Young, A.I., 2019. Solving the missing heritability problem. PLOS Genet. 15, e1008222. https://doi.org/10.1371/journal.pgen.1008222

Yuasa, N., Nakagawa, Y., Hayakawa, M., Iimura, Y., 2004. Distribution of the sulfite resistance gene SSU1-R and the variation in its promoter region in wine yeasts. J. Biosci. Bioeng. 98, 394–397. https://doi.org/10.1016/S1389-1723(04)00303-2

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., Davies, S.R., Wang, S., Wang, P., Kinsinger, C.R., Rivers, R.C., Rodriguez, H., Townsend, R.R., Ellis, M.J.C., Carr, S.A., Tabb, D.L., Coffey, R.J., Slebos, R.J.C., Liebler, D.C., 2014. Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387. https://doi.org/10.1038/nature13438

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, Bai, McDermott, J.E., Zhou, J.-Y., Petyuk, V.A., Chen, Li, Ray, D., Sun, S., Yang, F., Chen, Lijun, Wang, J., Shah, P., Cha, S.W., Aiyetan, P., Woo, S., Tian, Y., Gritsenko, M.A., Clauss, T.R., Choi, C., Monroe, M.E., Thomas, S., Nie, S., Wu, C., Moore, R.J., Yu, K.-H., Tabb, D.L., Fenyö, D., Bafna, V., Wang, Y., Rodriguez, H., Boja, E.S., Hiltke, T., Rivers, R.C., Sokoll, L., Zhu, H., Shih, I.-M., Cope, L., Pandey, A., Zhang, Bing, Snyder, M.P., Levine, D.A., Smith, R.D., Chan, D.W., Rodland, K.D., Carr, S.A., Gillette, M.A., Klauser, K.R., Kuhn, E., Mani, D.R., Mertins, P., Ketchum, K.A., Thangudu, R., Cai, S., Oberti, M., Paulovich, A.G., Whiteaker, J.R., Edwards, N.J., McGarvey, P.B., Madhavan, S., Wang, P., Chan, D.W., Pandey, A., Shih, I.-M., Zhang, H., Zhang, Z., Zhu, H., Cope, L., Whiteley, G.A., Skates, S.J., White, F.M., Levine, D.A., Boja, E.S., Kinsinger, C.R., Hiltke, T., Mesri, M., Rivers, R.C., Rodriguez, H., Shaw, K.M., Stein, S.E., Fenyo, D., Liu, T., McDermott, J.E., Payne, S.H., Rodland, K.D., Smith, R.D., Rudnick, P., Snyder, M., Zhao, Y., Chen, X., Ransohoff, D.F., Hoofnagle, A.N., Liebler, D.C., Sanders, M.E., Shi, Z., Slebos, R.J.C., Tabb, D.L., Zhang, Bing, Zimmerman, L.J., Wang, Y., Davies, S.R., Ding, L., Ellis, M.J.C., Townsend, R.R., 2016. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell 166, 755–765. https://doi.org/10.1016/j.cell.2016.05.069

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., Buckler, E.S., 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42, 355–360. https://doi.org/10.1038/ng.546

Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., Nordborg, M., 2007. An Arabidopsis Example of Association Mapping in Structured Samples. PLOS Genet. 3, e4. https://doi.org/10.1371/journal.pgen.0030004

Zhao, W., Rasheed, A., Tikkanen, E., Lee, J.-J., Butterworth, A.S., Howson, J.M.M., Assimes, T.L., Chowdhury, R., Orho-Melander, M., Damrauer, S., Small, A., Asma, S., Imamura, M., Yamauch, T., Chambers, J.C., Chen, P., Sapkota, B.R., Shah, N., Jabeen, S., Surendran, P., Lu, Y., Zhang, W., Imran, A., Abbas, S., Majeed, F., Trindade, K., Qamar, N., Mallick, N.H., Yaqoob, Z., Saghir, T., Rizvi, S.N.H., Memon, A., Rasheed, S.Z., Memon, F.-U.-R., Mehmood, K., Ahmed, N., Qureshi, I.H., Tanveer-Us-Salam, null, Iqbal, W., Malik, U., Mehra, N., Kuo, J.Z., Sheu, W.H.-H., Guo, X., Hsiung, C.A., Juang, J.-M.J., Taylor, K.D., Hung, Y.-J., Lee, W.-J., Quertermous, T., Lee, I.-T., Hsu, C.-C., Bottinger, E.P., Ralhan, S., Teo, Y.Y., Wang, T.-D., Alam, D.S., Di Angelantonio, E., Epstein, S., Nielsen, S.F., Nordestgaard, B.G., Tybjaerg-Hansen, A., Young, R., CHD Exome+ Consortium, Benn, M., Frikke-Schmidt, R., Kamstrup, P.R., EPIC-CVD Consortium, EPIC-Interact Consortium, Michigan Biobank, Jukema, J.W., Sattar, N., Smit, R., Chung, R.-H., Liang, K.-W., Anand, S., Sanghera, D.K., Ripatti, S., Loos, R.J.F., Kooner, J.S., Tai, E.S., Rotter, J.I., Chen, Y.-D.I., Frossard, P., Maeda, S., Kadowaki, T., Reilly, M., Pare, G., Melander, O., Salomaa, V., Rader, D.J., Danesh, J., Voight, B.F., Saleheen, D., 2017. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. Nat. Genet. 49, 1450–1457. https://doi.org/10.1038/ng.3943

Zhou, Y., Zhang, Zhiyang, Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., Zhang, J., Lyu, H., Lin, T., Gao, Q., Saha, S., Mueller, L., Fei, Z., Städler, T., Xu, S., Zhang, Zhiwu, Speed, D., Huang, S., 2022. Graph pangenome captures missing heritability and empowers tomato breeding. Nature 606, 527–534. https://doi.org/10.1038/s41586-022-04808-9

Zimmer, A., Durand, C., Loira, N., Durrens, P., Sherman, D.J., Marullo, P., 2014. QTL Dissection of Lag Phase in Wine Fermentation Reveals a New Translocation Responsible for Saccharomyces cerevisiae Adaptation to Sulfite. PLOS ONE 9, e86298. https://doi.org/10.1371/journal.pone.0086298

Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., Lander, E.S., 2014. Searching for missing heritability: Designing rare variant association studies. Proc. Natl. Acad. Sci. 111, E455–E464. https://doi.org/10.1073/pnas.1322563111

Overview of the project

Unraveling the genetic origins of the large phenotypic diversity observed in nature is a central goal of modern biology. In the case of complex traits such as cancer, height or autism, the association between genetic variants and the phenotype is often tedious to achieve, and overall, the mechanisms linking the variant to its cognate phenotype are still elusive. Gene expression is one key determinant of the genotype-phenotype relationship. More specifically, the regulation of gene expression plays a major role in translating genotypes into phenotypes, in particular when the genetic variants associated to complex traits are localized in non-coding or regulatory genomic regions. However, this regulation is truly puzzling because it involves regulatory processes that affect each level of gene expression. Over the past two decades, several technological and analytical advances have greatly facilitated the study of gene expression. Quantification of gene expression using RNA-seq, ribosome profiling as well as LC-MS/MS allowed for accurate quantification of each step of the process, while GWAS or linkage mapping helped to precisely map the origin of transcript or protein abundance variation on the genome. Nevertheless, large-scale studies of gene expression are sparse, and many aspects of gene expression remain to be explored and elucidated. For example, the similarities between the genetic origins of mRNA or protein abundance, and more globally, how protein and mRNA variation fit together, are still ongoing debates.

In this context, my project aims to take advantage of these technological advances and of the powerful *S. cerevisiae* model to study the variation of gene expression among individuals at the population level. A large collection of natural isolates of *S. cerevisiae* is available in our laboratory, which gathers more than a thousand strains for which genomes were completely sequenced using Illumina technology. The collection show very diverse ecological origins and includes both domesticated and wild strains, resulting in an accurate representation of the species diversity.

During these 4 years, I explored each level of gene expression in order to investigate their variation in a natural population. In a first chapter, I will describe the survey of gene expression at the transcriptional and translational levels, with the aim of characterizing a known determinant of variation between individuals, namely post-transcriptional buffering (Figure 1A). This phenomenon describes the fact that transcriptional variation tends to be buffered for as long as the expression process, suggesting increased evolutionary constraints on the later steps of gene expression. Although frequently observed, this phenomenon remains poorly understood, particularly at the level of translation. In a collaboration with the Riken Institute

(Japan), we performed ribosome profiling and RNA sequencing on 8 natural isolates of *S. cerevisiae*. Our results covered a large number of genes, 3,755 in total. We found that the transcriptional and translational variations were associated with metabolism-related genes. We detected post-transcriptional buffering in our dataset and found that modulation of translational efficiency is an important mechanism underlying this phenomenon. Interestingly, essential genes, protein complex-related genes as well as less transcribed genes were preferentially affected by post-transcriptional buffering. In addition, we investigated the translation of a subset of the *S. cerevisiae* pangenome, the accessory genes, focusing on the introgressed and Horizontal Gene Transfers (HGT) ORFs. We found that the introgressed genes were translated similarly to their orthologous ORF in the other isolates, whereas the HGT ORFs showed a lower translation efficiency. Overall, our results provide insights into the mechanisms underlying post-transcriptional buffering at the translational level and its specificity. For example, the cellular systems that cope with complex imbalance toxicity could be one of the drivers of the phenomenon already at the translational level, as it is at the proteomic level.

We sought to extend these findings at the proteome level by measuring protein abundances in these 8 isolates, which is described in the second chapter (Figure 1A). We analyzed the aforementioned RNA sequencing and ribosome profiling data together with proteomic data obtained by an LC-MS/MS approach. This dataset was generated in collaboration with the Weizmann Institute of Science (Israel) and includes 3,635 proteins. The 3-layer quantification (transcriptome, translatome and proteome) was possible for a total of 2,840 genes. We found that protein abundance variations were also mainly associated with metabolism- or respiration-related genes. Again, we found that post-transcriptional buffering was a major determinant of protein abundance variation across our isolate. We even observed that the more advanced the gene expression process, the stronger this phenomenon was. Therefore, protein abundance variations are more constrained, and different from what can be found at the transcriptome or translatome level. The difference in variation can also be observed by looking at the correlation between each of the gene expression steps. Despite being thought as a proxy for protein abundance, the ribosome profiling data was only slightly more correlated to protein data in comparison with RNA-seq data for the across-gene correlation, meaning that the proteome is slightly better reflected by the translatome than the transcriptome. When looking at the within-gene correlation, both RNA-seq and ribosome profiling data had a mediocre correlation with the protein abundance, meaning that inter-individual proteome variation are barely captured by transcriptome and proteome. Taken together, this suggests that each gene expression layer may

be subject to different evolutionary constraints. We sought to verify this by quantifying gene expression evolution at each step for more than 700 features characterizing each gene. We found that although each gene expression layer has some specificities, and that there are some general rules underlying gene expression evolution. For example, genes that are central to cellular networks (*i.e.*, that interact a lot with other proteins or play a fundamental process for the cell) tend to have more constrained gene expression regulation, while genes related to metabolism tend to have faster gene expression evolution. This is strongly consistent with the previous findings on post-transcriptional buffering.

Finally, we explored the relationship between the transcriptome and proteome, this time at the population level, in order to get an accurate view of how the transcriptome and proteome vary at the species level (Figure 1B). In the third and final chapter, I describe the comparison between two large-scale surveys related to the transcriptomes and the proteomes of the 1,011 *S. cerevisiae* strains from our collection. While the transcriptome dataset was generated in our laboratory for a previous project, the proteome dataset was generated in collaboration with the Charité University of Medicine in Berlin and the Francis Crick Institute in London. The combined proteome and transcriptome dataset covers 629 genes from 889 isolates. As our data was one of the largest mRNA and protein abundance comparisons to date, we explored several gene expression phenomena that have never been studied at this scale in yeast, such as gene co-expression networks, post-transcriptional buffering, or domestication-related proteomic signatures. More importantly, we detailed the within-gene correlation in our population and showed that it was rather weak, around 0.16. Although the level of this correlation has been debated, this is significantly lower than what has been found in previous studies. Interestingly, the correlation level tended to be gene-dependent, with metabolism- and respiration-related genes showing high correlation levels, while ribosome-related genes tended to be uncorrelated or anticorrelated. The overall weak correlation between mRNA and protein abundance variation suggested that the genetic origins of the transcriptome and proteome are different. We performed GWAS to unravel the association between genome variation (taking both SNPs and CNVs into account) and proteome or transcriptome variation. We found that the overlap between eQTL and pQTL was very modest, especially when looking at the SNP-based GWAS. Only 3.6% of SNP-QTLs were shared between transcriptome and proteome. The CNV-based GWAS, on the other hand, showed a higher similarity between the genetic origins of mRNA and protein abundance, but this was mainly related to the presence of large aneuploid segments on specific chromosomes, which have a strong impact on gene expression. No overlap between

CNV-eQTL or pQTL nonrelated to aneuploidy was observed. Taken together, our results show that at the population level, the proteome and the transcriptome are two very distinct layers of gene expression, with very specific mechanisms underlying inter-individual variation.

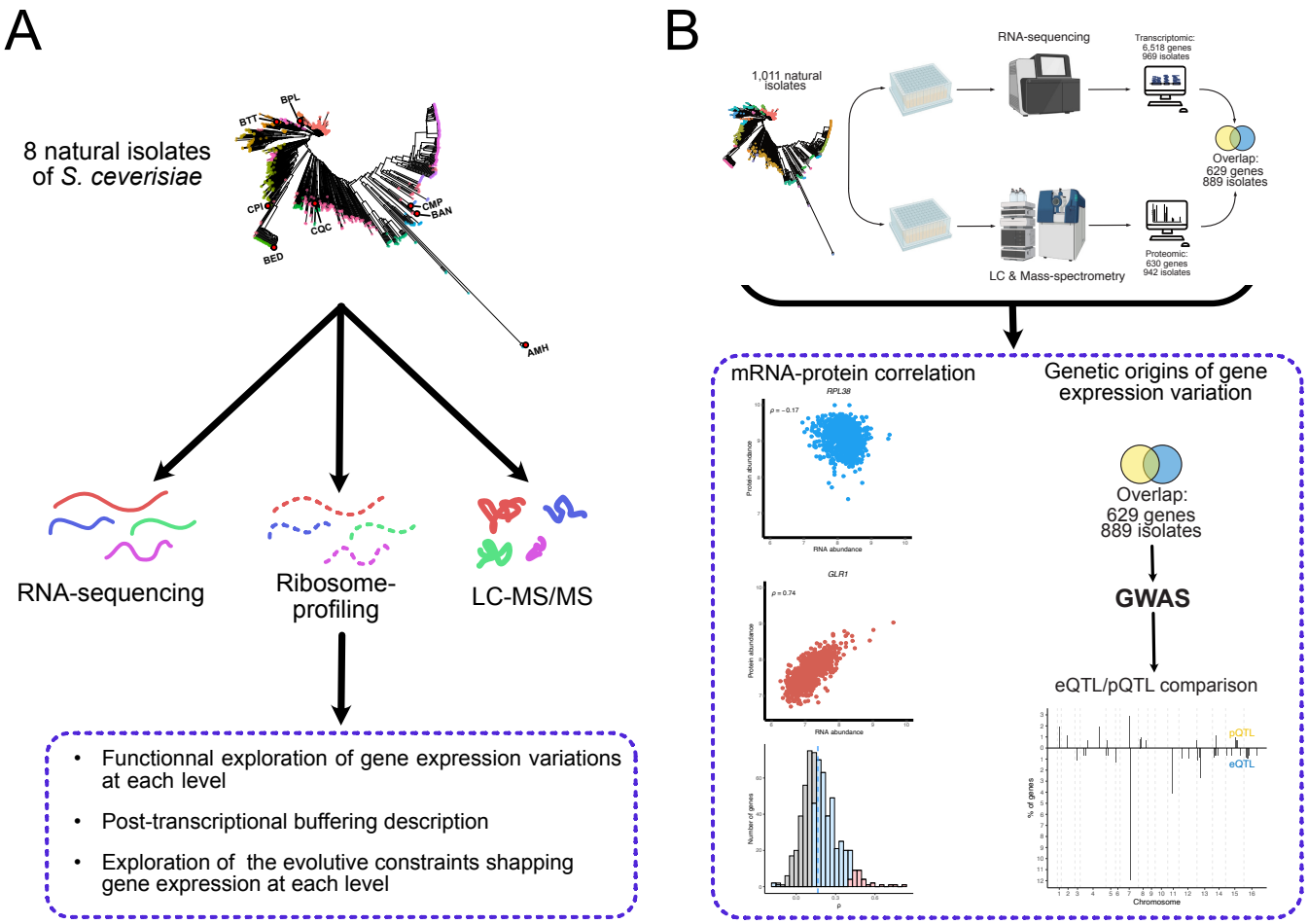


Figure 1: Summary of the PhD project.

(A) A first project focuses on gene expression variation across 8 isolates for which RNA-seq, ribo-seq and LC-MS/MS data are available. In a first chapter I will focus on the comparison between transcriptional and translational variation and on post-transcriptional buffering at the translational level. In a second chapter, I will combine these datasets with new proteomic data from one of the 8 isolates and explore the evolutionary constraints on gene expression evolution across the gene expression process. (B) In a third and final chapter, I will present the population-level exploration of both transcriptome and proteome across 889 isolates. The main focus and findings will be related to within-gene mRNA-protein correlation and the genetic origins of transcript and protein abundances.

CHAPTER I

Translation variation across genetic backgrounds reveals a post-transcriptional buffering signature in yeast

Collaborative work from:

Elie M. Teyssonniere[1], Yuichi Shichino[2], Anne Friedrich[1], Shintaro Iwasaki[2,3,4] and Joseph Schacherer[1,5]

1. Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France

2. RNA Systems Biochemistry Laboratory, RIKEN Cluster for Pioneering Research, Wako, Saitama, 351-0198 Japan

3. Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan

4. AMED-CREST, Japan Agency for Medical Research and Development, Wako, Saitama 351-0198, Japan

5. Institut Universitaire de France (IUF), Paris, France

Abstract

Gene expression is known to vary among individuals, and this variability can impact the phenotypic diversity observed in natural populations. While the transcriptome and proteome have been extensively studied, little is known about the translation process itself. Here, we therefore performed ribosome and transcriptomic profiling on a genetically and ecologically diverse subset of natural isolates of the *Saccharomyces cerevisiae* yeast. Interestingly, we found that the Euclidean distances between each profile and the expression fold changes in each pairwise isolate comparison were approximately 10% higher at the transcriptomic level. This observation clearly indicates that the transcriptional variation observed in the different isolates is buffered through a phenomenon known as post-transcriptional buffering at the translation level. Furthermore, this phenomenon seemed to have a specific signature by preferentially affecting essential genes as well as genes involved in complex-forming proteins, and low transcribed genes. We also explored the translation of the *S. cerevisiae* pangenome and found that the accessory genes related to introgression events displayed similar transcription and translation levels as the core genome. By contrast, genes acquired through horizontal gene transfer events tended to be less efficiently translated. Together, our results highlight both the extent and signature of the post-transcriptional buffering.

Introduction

Transcript and protein abundance variations are well-known sources of phenotypic diversity across individuals. Protein abundance is influenced by both transcriptional and post-transcriptional regulations, which ultimately affect the final phenotypes. There are several cellular mechanisms involved in the modulation of final protein abundance, including mRNA stability, translation initiation and protein degradation (Buccitelli and Selbach, 2020). In the last decades, various technologies have greatly facilitated the detailed exploration of all these steps. Some of these technologies include DNA high-throughput sequencing methods such as RNA sequencing (Hrdlickova et al., 2017), and mass spectrometry (Lu et al., 2007), which enable a global description of transcriptomic and proteomic dynamics. By associating these data with genomic data, we can greatly improve our understanding of the mRNA and protein abundance regulation at the population level (Jiang et al., 2020; Kita et al., 2017; Messner et al., 2022; Suhre et al., 2020; The GTEx Consortium, 2015). Furthermore, the core of the translational process can be precisely dissected with the development of ribosome profiling (or Ribo-Seq) (Ingolia et al., 2019, 2009). This strategy relies on the sequencing of mRNA fragments covered by the ribosomes during the translation process, revealing which parts of the transcriptome are actively being translated. This method can quantify translation in mRNA-wise (number of fragments of the corresponding mRNA) and also the behavior of the ribosomes at the given codon (density of the ribosomes along the mRNA) (Ingolia, 2014).

The budding yeast *Saccharomyces cerevisiae* has been a powerful model for ribosome profiling experiments, as this technique was developed on this organism (Ingolia et al., 2009). Translational variation in yeast has been explored with ribosome profiling on several occasions (Albert et al., 2014; Artieri and Fraser, 2014; Blevins et al., 2021, 2019; McManus et al., 2014; Wang et al., 2015). Interestingly, several of these studies highlighted that the transcriptional variations tended to be buffered when looking at the translational variations (Artieri and Fraser, 2014; Blevins et al., 2019; McManus et al., 2014; Wang et al., 2015). This phenomenon is known as post-transcriptional buffering and has also been observed when comparing transcriptomic and proteomic datasets (Gonçalves et al., 2017; Kustatscher et al., 2017). However, despite recurring observations, the mechanisms underlying this post-transcriptional buffering are still poorly understood. Moreover, while transcription and protein abundance have been extensively monitored, translation itself has been considerably less studied, and no clear description of the phenomenon has been made at this level. More globally, translational

variation remains largely unexplored, and several known sources of expression variation, such as accessory ORF (open reading frames), have yet to be investigated at the translational layer. Here, we conducted ribosome profiling and RNA sequencing in the same conditions on eight *S. cerevisiae* natural isolates coming from very diverse ecological environments and being genetically different (Peter et al., 2018). We first compared the transcriptional and translational variations, and found that they had similar functional patterns. Metabolism-related genes tended to be more variable across the eight isolates while essential genes and genes involved in molecular complexes had more conserved transcription and translation regulation. Interestingly, we found that the transcriptional profiles were less correlated to each other compared to the translational profiles. Accordingly, Euclidean distances and expression variations (quantified using the absolute log2 transformed foldchanges for each gene in each isolate pairwise comparisons) were approximately 10% higher in the transcriptomic data, indicating that post-transcriptional buffering is a strong determinant of the translational variations. More importantly, we found that this phenomenon has a specific signature in terms of affected genes. We observed that essential genes and protein complex-related genes as well as lowly transcribed genes tended to be preferentially buffered. Furthermore, we investigated the transcription and translation of accessory open reading frames (ORFs) present in the eight isolates, particularly those acquired through introgression or horizontal gene transfer (HGT) events. We observed that introgression-related ORFs were similarly transcribed and translated compared to their orthologs, while HGT-related ORFs displayed a significantly lower translation efficiency than the rest of genes. Together, our results provide an overview of translational variations as well as an accurate description of post-transcriptional buffering.

Results

Ribosome profiling and RNA sequencing across eight natural isolates

We performed both RNA sequencing (RNA-seq) and ribosome profiling (Ribo-seq) on eight genetically diverse *S. cerevisiae* isolates (Table S1), which were cultivated and harvested in the exact same condition. These isolates were selected to represent the genetic diversity of the species (Figure S1) and were grown on a synthetic complete medium. All the genomes of the isolates were all previously sequenced (Peter et al., 2018), and in addition to their very different genetic backgrounds, they also came from very diverse environments. After TPM (transcript per million) normalization of RNA-seq and Ribo-seq raw counts (see Methods), we computed a translation efficiency (TE) value of each gene in each isolate by dividing its Ribo-seq TPM value by its RNA-seq TPM value. In total, we analyzed 3,755 genes and our results showed a strong correlation between RNA-seq and Ribo-seq data (Spearman correlation test between 0.769 and 0.867), highlighting the relationship between transcription and translation. To gain a global view of intraspecific variation, we performed pairwise Spearman correlation tests on the two datasets for each strain (Figure 1A). The RNA-seq and Ribo-seq correlation matrices showed similar patterns, indicating that transcriptional variations were largely reflected at the translational level. However, the correlation coefficients were generally higher in the Ribo-seq matrix than the average value in the RNA-seq matrix (Figure 1B), suggesting that translational profiles were more similar than transcriptional profiles.

Next, we sought to identify genes that did not follow these correlation trends in RNA-seq and Ribo-seq data to detect genes with variable regulation of transcription and translation. To achieve this, we used a combination of two different methods to make a pairwise comparison of all isolates. The first method was a Mahalanobis distance calculation to detect outliers in the pairwise comparison (See Methods, Figure S2A) (Ho et al., 2018). The second method relied on the selection of genes displaying residual (obtained from a linear regression model computed on the isolate pairwise comparison), which are in 2.5% highest or 2.5% lowest residual quantiles. We selected the genes that overlapped between the two methods (Figure S2B) and identified a total of 357 genes in Ribo-seq data and 352 genes for RNA-seq, with 179 overlaps between the two expression layers (Figure S3A and B). These genes are later mentioned as "variable genes". Most of these genes were only detected once in the 28 pairwise comparisons (128 out of 357 Ribo-seq variables genes, 118 out of 352 RNA-seq variables genes, Figure 1C-D). The number of variable genes detected in the pairwise comparisons against the 7 other

strains ranged from 125 to 204 for Ribo-seq (median=172) and from 128 and 204 to RNA-seq (median=177.5).

To investigate functional enrichment among these identified genes, we conducted a Gene Ontology analysis (GO) (Ashburner et al., 2000; Gene Ontology Consortium, 2021) on both Ribo-seq and RNA-seq variable genes, as well as on the 179 overlapping gene set. The RNA-seq variable genes yielded a relatively low number of terms compared to the other dataset (20 terms, Table S2). However, the majority of its terms were detected in the Ribo-seq variable genes, suggesting that despites the difference between the two genes groups, the functions of the genes are mostly shared. We also found that 63 terms were shared between the Ribo-seq variable genes (out of 79 terms, Table S3) and the overlapping variable genes (out of 86 terms, Table S4). Many of the shared results were related to metabolism terms (such as “glycolytic process”, “pentose-phosphate shunt” and “glucose metabolic process”). This observation may be explained by the fact that the eight isolates used in this study were obtained from distinct environments (Table S1) and might have adapted their regulation of several metabolic functions to different trophic conditions.

Interestingly, our analysis revealed that the GO term displaying the lowest p-value was the depletion of the “protein-containing complex” term (Ribo-seq variable genes: 1.08e-17, RNA-seq variable genes: 7.55e-09, overlapping variable genes: 6.35e-18), suggesting that genes encoding protein involved in protein complexes are underrepresented in the variable gene sets. To confirm this observation, we checked if genes previously annotated as related to protein complexes (Pu et al., 2009) were indeed significantly depleted from the variable genes set. Similarly, we explored if genes characterized as essential (Dowell et al., 2010; Giaever et al., 2002) were enriched or depleted in the variable genes since these characteristics (essentiality and being involved in protein complexes) have been cited in previous studies exploring evolutionary constraints (Morrill and Amon, 2019; Pál et al., 2006; Rancati et al., 2018). Using Fisher's exact test (FET), we found that essential genes were strongly depleted in the variable gene set (RNA-seq: odds ratio = 0.40, p-value = 1.55e-10; Ribo-seq: odds ratio = 0.22, p-value = 8.89e-16). The results were very similar when examining protein complex-related genes (RNA-seq: odds ratio = 0.24, p-value = 1.81e-28; Ribo-seq: odds ratio = 0.17, p-value = 1.79e-38). Interestingly, the depletion was lower with the Ribo-seq variable genes for both essential genes and protein complex-related genes, suggesting that these genes were less likely to exhibit variable regulation at the translational level compared to the transcriptional level.

Together, our results highlight that expression variation is unequal among the genes. While metabolism related genes display important transcriptional and translational variations among

the eight isolates, essential genes and protein complex related genes are related to a lower expression variation.

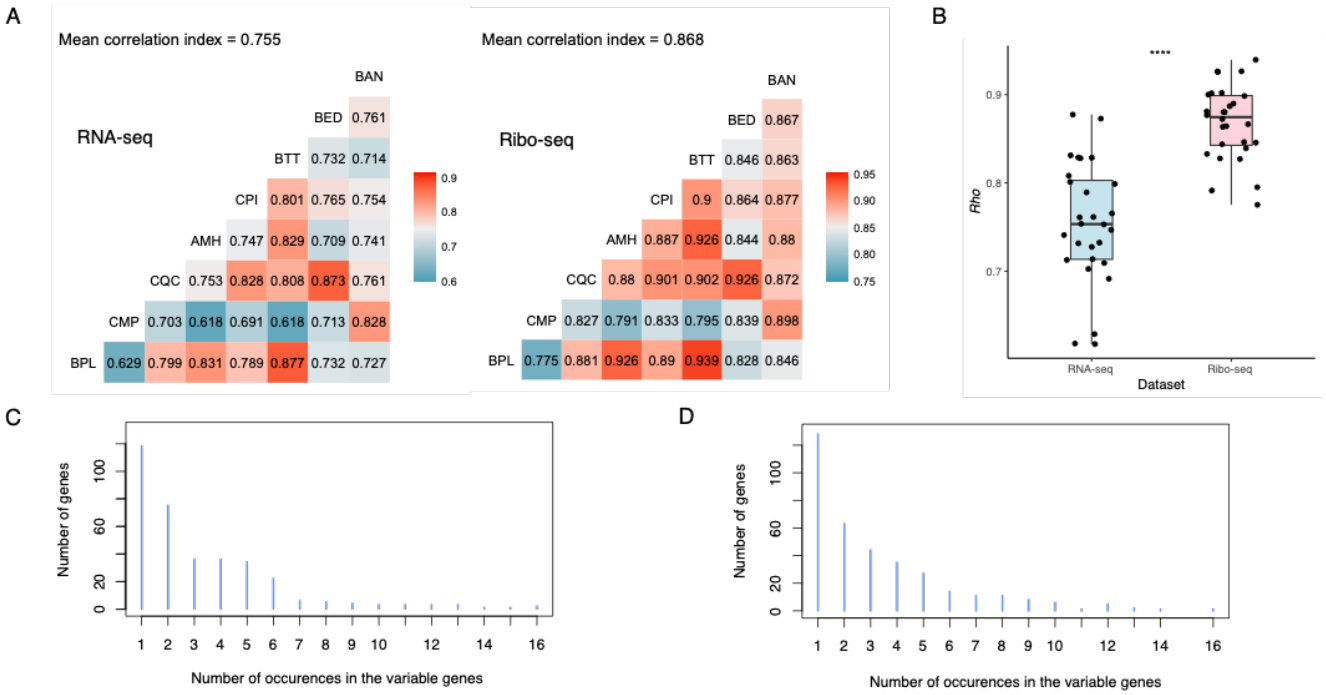


Figure 1. Exploration of the transcriptional and translational variations. (A) Correlation matrixes (Spearman correlation test) of each RNA-seq and Ribo-seq isolate pairwise comparison (all the coefficients displayed significant P-values). (B) Difference between the RNA-seq and Ribo-seq correlation levels (Wilcoxon test p-value = 1.4×10^{-9}). (C, D) Number of occurrences in the variable genes using (C) RNA-seq and (D) Ribo-seq data.

Post-transcriptional buffering at the translation level across isolates

Several results suggest lower variability at the translational level compared to the transcriptomic level, such as the higher correlation in the Ribo-seq dataset (Figure 1B). To confirm these observations, we first computed the Euclidean distances and checked if the distances between each profile were higher in RNA-seq or Ribo-seq $\log_{10}$ data. We found that the distance between the strains were approximately 10% higher in RNA-seq data (t-test, p-value = 0.036) (Figure 2A), suggesting that the transcriptional variations tended to be higher than the translational ones. Consistently, the expression variance of each gene across the 8 isolates was significantly higher at the transcription level (Figure S4).

We also quantified gene expression variation in the two expression layers by computing the absolute value of the $\log_2$ transformed foldchange ($|log2(FC|$, see Methods) for each gene in

the pairwise comparisons of the isolates (n = 105,140). The higher this value will be, the stronger the transcriptional or translational variation will be. We found that |*log2(FC*| was approximately 10% higher in RNA-seq data (mean = 0.975) compared to Ribo-seq data (mean = 0.872, Figure 2B), suggesting that overall, the transcriptional variations tended to be buffered. These results are consistent with the phenomenon of post transcriptional buffering that has been observed previously (Artieri and Fraser, 2014; Blevins et al., 2019; McManus et al., 2014; Wang et al., 2015). In their study, they showed that the buffering of transcriptional variation may be linked to modification of translational efficiency. We sought to confirm this by comparing the RNA-seq, Ribo-seq and TE $\log_2$(FC) values in each pairwise comparison. We observed that RNA-seq $\log_2$(FC) values were strongly anti-correlated with TE $\log_2$(FC) values (Figure 2C, S5A, Table S5). Conversely, we found that the comparisons between RNA-seq and Ribo-seq $\log_2$(FC) always showed a positive correlation (Figure 2D, S5B, Table S5). Finally, we found no correlation between Ribo-seq and TE $\log_2$(FC) (Figure S5C, Table S5). Together, our results suggest that even though transcriptional and translation variations are similar in direction, their strength are strongly attenuated at the translation level by an opposite change in translation efficiency.

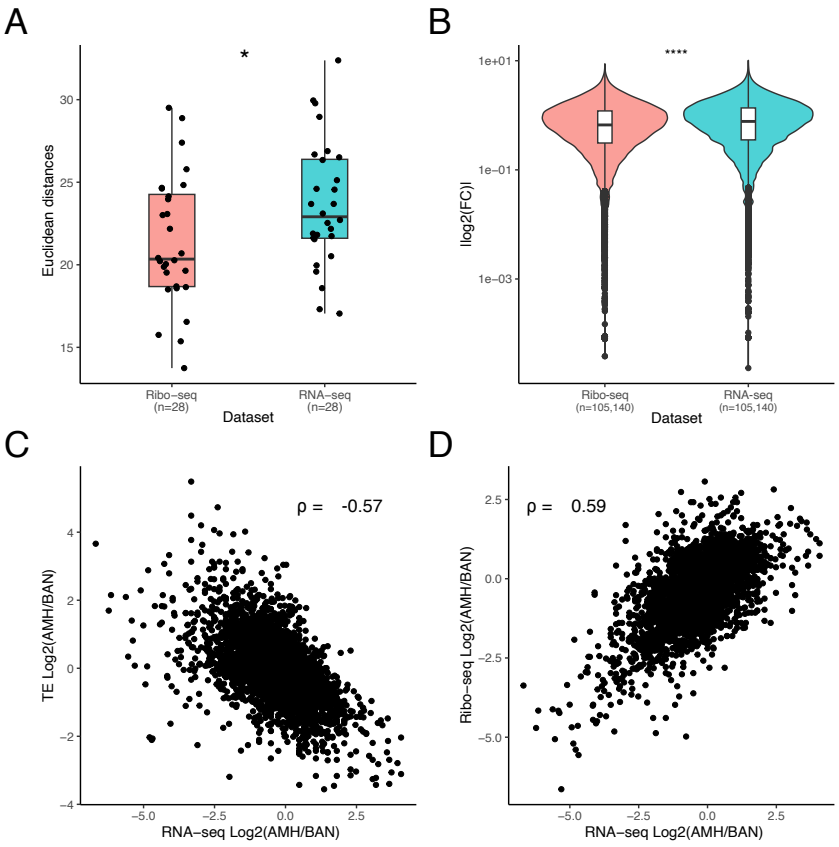


Figure 2. The transcription variations are buffered because of TE modulation. (A) Euclidean distance comparison between RNA-seq or Ribo-seq profiles. The distances were significantly higher in RNA-seq data (Wilcoxon test p-value = 0.0478). (B) Differences between the RNA-seq and Ribo-seq |*log2(FC)*| values obtained by comparing each gene's TPM in each pairwise comparison (Wilcoxon test p-value = $3.3x10^{-193}$). (C) Comparison between the RNA-seq log2(FC) and the TE log2(FC) in the AMH vs BAN isolate pairwise comparison (Spearman correlation p-value = $2.8x10^{-317}$). (D) Comparison between the RNA-seq log2(FC) and the Ribo-seq log2(FC) in the AMH vs BAN isolate pairwise comparison (Spearman correlation p-value < $4.94x10^{-324}$)

Signature of the post-transcriptional buffering at the translation level

Despite several observations of post-transcriptional buffering, this phenomenon remains largely unknown, especially at the functional level (Artieri and Fraser, 2014; Blevins et al., 2019; McManus et al., 2014; Wang et al., 2015). We sought to further characterize the general rules underlying this phenomenon by looking for genes that would be preferentially affected by the post-transcriptional buffering at the translational level.

With that in mind, we split each of the RNA-seq vs TE $\log_2$(FC) pairwise comparisons in orthogonal spaces using a TE and RNA-seq fold-change threshold of 1.5 (Figure 3A). This method made it possible to distinguish two categories of genes. First, we detected genes with transcriptional variation using the RNA-seq $\log_2$(FC) threshold. Second, using the TE $\log_2$(FC) threshold, we were able to identify genes with buffered variation from those not affected by the buffering (Figure 3A). The number of genes affected by post-transcriptional buffering ranged from 854 to 1,319 across the pairwise comparisons, with a mean value of 1,051 per comparison (Figure S6). No genes were buffered in all pairwise comparisons, and the proportion of buffered genes among the genes with transcriptional variation averaged 46.8%, ranging from 37.2% to 58.8% (Figure S7).

We then selected genes whose transcriptional variation was recurrently buffered or unbuffered (see Methods). We detected 361 and 507 genes whose transcriptional variation was recurrently buffered and unbuffered, respectively. We searched for a functional signature among the buffered genes but found no significant enrichment using GO analysis. We then focused on the content of essential (Dowell et al., 2010; Giaever et al., 2002) and protein complex-related genes (Pu et al., 2009) in recurrently buffered genes in comparison to recurrently unbuffered genes. The proportion of essential genes was significantly higher in the buffered gene set compared to the unbuffered genes (Figure 3B). Similarly, protein complex-related genes were also in higher proportion among the buffered gene (Figure 3C). Together, these results support

that the genes are unequally affected by the post-transcriptional buffering at the translational level, with essential genes and protein complex-related genes preferentially buffered.

Interestingly, essentiality and protein-protein interactions are features debated for their influence on protein sequence evolution (Fraser et al., 2002; Pál et al., 2006; Rancati et al., 2018; Zhang and Yang, 2015). Another major determinant of sequence evolution is gene expression level, as highly expressed proteins are known to be more conserved (Drummond et al., 2006, 2005; Rocha, 2006; Zhang and Yang, 2015). We therefore sought to see if the transcription level was also involved in preferential buffering. Surprisingly, we found that recurrently buffered genes tended to be less transcribed (median TPM = 49.8) than recurrently unbuffered genes (median TPM = 230.1, Figure 3D). The same results were observed when comparing the recurrently buffered genes to the rest of the genes (Figure S8A). The results were also similar when looking at the average expression levels of buffered and not buffered genes in the pairwise comparisons (Figure S9). Together, these results highlight that transcription level is also a determinant of the phenomenon of post-transcriptional buffering, since buffered genes tend to be less transcribed.

We sought to confirm these results by exploring the codon usage bias of the buffered genes since this feature is known to be related to expression level (Coghlan and Wolfe, 2000; Plotkin and Kudla, 2011) and essentiality (Dilucca et al., 2015). We computed a codon usage bias index for each gene using tRNA Adaptation Index (tAI) (dos Reis et al., 2004, 2003), and we confirmed that this index correlated with RNA-seq or Ribo-seq data (Figure S10). We then compared the tAI values of the buffered group to those of the unbuffered group. We observed a significantly lower tAI in the buffered group (median tAI = 0.34) compared to the unbuffered group (median tAI = 0.38) (Figure 3E). Similar results were observed when comparing the recurrently buffered genes with the rest of the genes (Figure S8B). This observation supported the previous results of expression level difference between the two groups.

Overall, these results highlight the fact that the phenomenon of post-transcriptional buffering preferentially affects essential, protein complex-related genes, or genes with lower transcription levels and therefore has a specific signature. This behavior toward some specific categories of genes has never been shown before and highlights evolutionary constraints affecting the translational regulation of these genes.

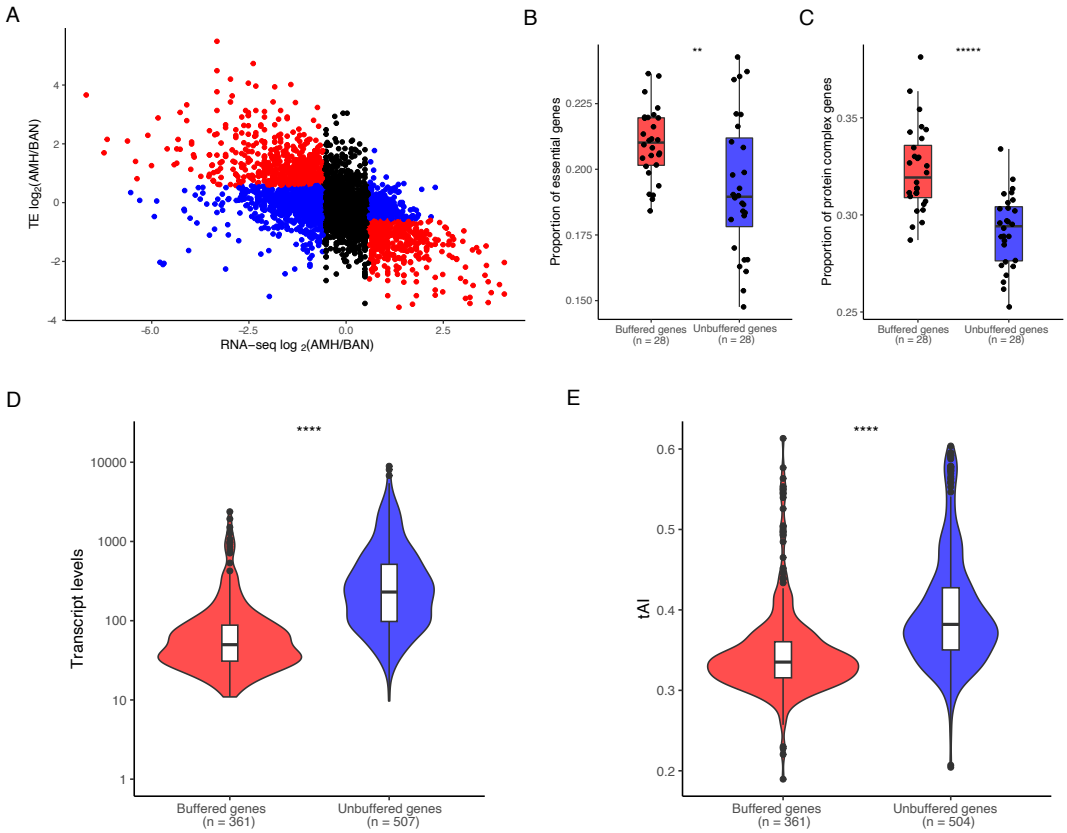


Figure 3. Post-transcriptional buffering has a specific signature. (A) Detection of the buffered and unbuffered genes according to a 1.5-fold change threshold (corresponding to the lines on the plot) in the AMH vs BAN isolate pairwise comparison, blue points = unbuffered variation; red points = buffered variation. (B, C) Proportion of, respectively, essential gene and protein complex related genes among the two gene groups in the 28 pairwise comparison. The proportion are in both cases higher in the buffered group (Wilcoxon test). (D) RNA-seq and (E) tAI levels of the recurrently buffered or unbuffered genes (respective Wilcoxon test p-value: $2.6x10^{-67}$ and $1.03x10^{-40}$).

Transcription and translation variation of accessory genes

Recent advances in *S. cerevisiae* population genomics have highlighted the presence of more than 1,700 variable ORFs (accessory genes) in this species (Peter et al., 2018). Our translation exploration across multiple individuals from very different genetic and environmental origins is a unique opportunity to explore the translation of such ORFs, which was omitted in previous yeast ribosome profiling investigations. In our eight isolates, the number of these ORFs varies between 63 to 215, corresponding to a total of 446 unique accessory ORFs (median = 94 accessory ORFs per strain) (Figure S11), but depending on the isolate, between 36% and 72% of them were expressed (Figure S11). We observed that the unexpressed ORFs tended to be smaller (on average 148.81 bp) than the expressed ones (on average 365.2 bp, Figure S12).

Overall, our eight isolates displayed variable profiles in terms of accessory ORFs origins (Figure S11). Two strains differed notably to the others in their compositions: the CPI isolate due to a very high number of ORFs acquired by introgression and the BPL isolate due to genes acquired through horizontal gene transfer (HGT).

Regarding the CPI isolate, this strain was originally isolated in Mexico and has been described as part of the "Mexican agave clade" (Peter et al., 2018). This subpopulation has a high number of introgressed ORFs coming from the yeast *Saccharomyces paradoxus* (median = 161 ORFs per strain vs 25.75 in the overall population). The CPI isolate had 87 expressed ORFs coming from introgression events, and 45 ORFs had known orthologs in S288C. In order to strictly explore the impact on transcription and translation, we focused on 18 out of the 45 ORFs that were homozygous for the *S. paradoxus* allele in the CPI isolate and found no expression difference between these ORFs and their orthologs in the 7 other strains (again at the transcriptional and translational level) (Figure 4A, B). These results imply that the transcriptional and translational regulation of ORFs acquired by introgressions from *S. paradoxus* is similar to their regulation of their orthologs.

We then focused on the expression of the 16 accessory ORFs coming from HGT in the case of the BPL isolate (Table S7). This strain has already been described as part of a wine subpopulation (Peter et al., 2018) and the occurrence of HGT events in this type of strain has already been observed (Marsit et al., 2015; Novo et al., 2009). Briefly, the coexistence of *S. cerevisiae* with other yeast species in the wine environment led to gene transfer that can confer evolutionary advantage in the winemaking environment. We compare the expression of these ORFs to other genes in the BPL isolate (Figure S13). Surprisingly, HGT ORFs almost all showed lower Ribo-seq values than RNA-seq, suggesting a lower TE than the rest of the genes. We then compared HGT ORFs with the rest of the genes and found a significant difference where HGT ORFs were less efficiently translated than other genes in the BPL isolate (mean HGT TE = 0.545, mean HGT other genes = 0.865) (Figure 4C). This observation clearly shows that HGT-related ORFs exhibited significantly lower translation efficiency compared to the rest of the genome.

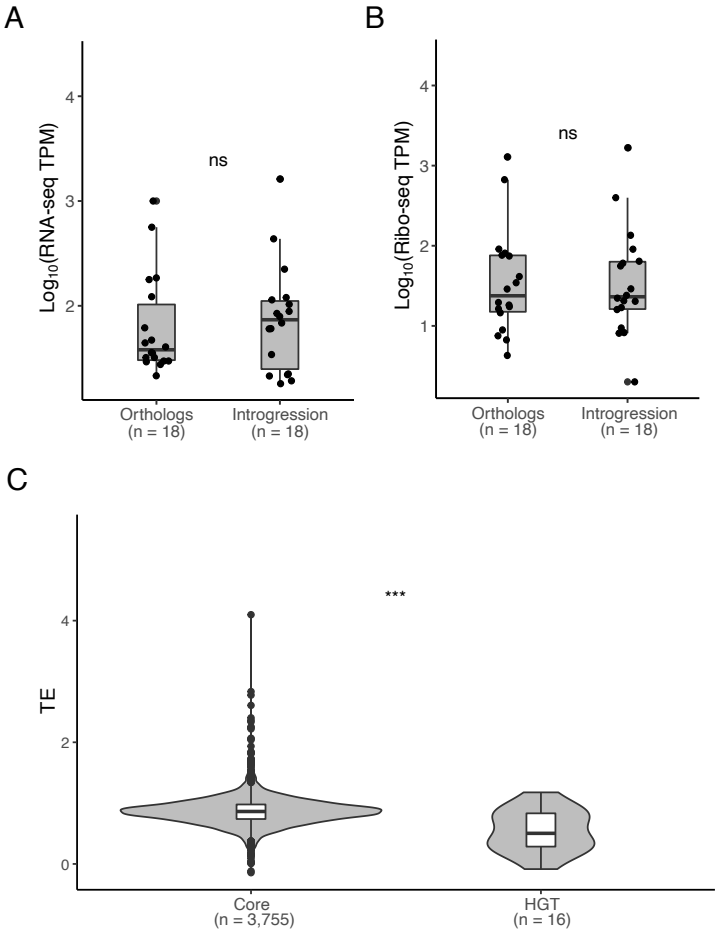


Figure 4. Translation levels of the *S. cerevisiae* pangenome. (A, B) RNA-seq and Ribo-seq level of the ORFs acquired through introgression event in the CPI isolate and being homozygous for the *S. paradoxus* allele (n = 18) and their orthologs in the other isolates. No difference in term of transcription of translation were observed between the introgression related ORFs and their orthologs. (C) TE difference between the ORFs acquired through HGT in the BPL isolate and the other ORFs (Wilcoxon test p-value = 0.0002).

Discussion

Translational variation is a major determinant of the transcriptome-proteome relationship, and therefore plays a central role in the phenotypic diversity observed in natural populations. However, the translation process itself remains largely unexplored, and the central mechanisms driving translation variations are still poorly understood. In this study, we have precisely monitored translation variations across natural isolates of *S. cerevisiae* using ribosome profiling.

Gene expression is known to differ across individual. We observed that the translational and transcriptional variations were functionally similar, with metabolism-related genes displaying the greatest variation while the translation regulation of essential genes and protein complex-related genes were more conserved. These results are consistent with recent large-scale exploration of mRNA abundance (Caudal et al., Submitted) and highlight that gene expression plasticity might be driven by the metabolism preferences between isolates coming from very different environments (Hodgins-Davis et al., 2012). Conversly, expression variation of genes with central and essential functions is likely to be deleterious and tend to be therefore more conserved (Fraser et al., 2004).

Our dataset also allowed to get better insight into the phenomenon of post-transcriptional buffering at the translational level (Artieri and Fraser, 2014; Blevins et al., 2019; McManus et al., 2014; Wang et al., 2015). Using $|log2(FC)|$, we quantitatively measured gene expression variation across both layers of expression (transcription and translation), and found that the median transcriptional value $|log2(FC)|$ was 10% higher than the median translational $|log2(FC)|$. Together with the differences in gene expression variance and correlations coefficients between isolates, this clearly indicates that post-transcriptional buffering is detected in our dataset. As previously suggested (McManus et al., 2014), we observed that TE modulation plays a central role in compensating for variations in mRNA abundance.

Interestingly, we found that the post-transcriptional buffering has a specific signature. It preferentially affects genes that are essential or related to protein complexes, as well as genes with low transcript levels. Several reasons could underlie this preferential buffering. For complex-forming protein, it is well established that the imbalance of complex components can be deleterious for (Deutschbauer et al., 2005; Ohnuki and Ohya, 2018; Veitia and Potier, 2015), partly for stoichiometric reasons (Morrill and Amon, 2019). More generally, protein complex-related genes are known to have stronger regulatory control at the protein level rather than mRNA level (Jüschke et al., 2013) and programmed translation of complex components

precisely proportional to stoichiometry was not only found in yeast (Taggart and Li, 2018), but also in bacteria and plants (Chotewutmontri and Barkan, 2016; Li et al., 2014; Lukoszek et al., 2016; Trösch et al., 2018). Essential genes are known to carry a central and highly conserved function in the cell (Costanzo et al., 2016), which can lead to higher constraints on gene expression evolution (Wang et al., 2020). However, the conditional nature of essentiality (Larrimore and Rancati, 2019; Papp et al., 2004) and the ongoing debate on the importance of essentiality on evolutionary constraint (Pál et al., 2006; Rancati et al., 2018) suggest that the link between expression conservation and essentiality remains unclear. Regarding the fact that buffered genes tend to be less transcribed than other genes, this is surprising since very abundant proteins are usually well conserved (Drummond et al., 2006, 2005; Rocha, 2006; Zhang and Yang, 2015). Thus, we would have expected that the regulation of highly expressed genes would also be highly conserved as well, and therefore preferentially affected by the phenomenon of post-transcriptional buffering.

Finally, working with genetically distinct natural isolates allowed to explore the translation of a part of the *S. cerevisiae* accessory genome, which is something that has been barely investigated so far. Accessory genes had very different translation dynamics depending on their origins of acquisition. While introgression-related ORFs displayed similar levels of translation compared to their orthologs, HGT-related ORF were less translated, resulting in low translation efficiency. These results on the pangenome nevertheless remain limited due to the low representation of the entire pangenome of *S. cerevisiae* (Peter et al., 2018). More generally, a broader view of *S. cerevisiae* population translation would improve our understanding of the post-transcriptional buffering phenomenon.

Overall, our results highlight the importance of the post-transcriptional buffering at the translation level, as well as its specific signature. Moreover, they give one of the first insight into the translation dynamics of a specific part of the genome such as accessory genes.

Materials and methods

Strain, culture, and flash freezing

The complete list of isolates used in this study is available and described in Table S1. The strains were grown in liquid SC medium (Yeast Nitrogen Base with ammonium sulfate 6.7 $g.l^{-1}$, MPbio, OH, USA; amino acid mixture 2 $g.l^{-1}$, MPbio; glucose 20 $g.l^{-1}$, Euromedex, France). The culture was maintained until the strains reached their growth mid log phase using an optical plate reader (Tecan infinite F200 pro). The cells were then filtered using 0.45 µm MCE membrane (Merk Millipore, France). The filters were then plunged into a 50 mL tube containing liquid nitrogen and stocked in a -80°C freezer before being used for ribosome profiling and RNA sequencing experiment.

Ribosome profiling and RNA sequencing

The library preparation for ribosome profiling was performed as previously described with modifications (McGlincy and Ingolia, 2017; Mito et al., 2020). Cells on the filters were mixed with frozen droplets of 600 µL lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 100 µg/mL cycloheximide, and 1% Triton X-100) and crushed using Multi-beads Shocker (Yasui Kikai, Japan). Lysate containing 20 µg of total RNA was digested with 10 U of RNase I (Lucigen, WI, USA) for 45 min at 25ºC. Ribosomes were precipitated by sucrose cushion and ultracentrifugation, suspended into EDTA lysis buffer, in which 5 mM $MgCl_2$ in lysis buffer was substituted with 5 mM ethylenediaminetetraacetic acid (EDTA), and transferred to Amicon Ultra-0.5 Ultracel-100 (Merck Millipore, MA, USA) to separate footprints from ribosome subunits (the details will be described elsewhere [M.M. and S.I., unpublished data]). RNAs ranging from 17 to 34 nt were excised from a polyacrylamide TBE-Urea gel. The rRNA was depleted using RiboMinus Transcriptome Isolation Kit (yeast) (Thermo Fisher Scientific, MA, USA).

For RNA-seq, total RNA was purified using TRIzol LS reagent (Thermo Fisher Scientific) and Direct-zol RNA Miniprep Kit (Zymo research, CA, USA). Following the removal of rRNA by RiboMinus Transcriptome Isolation Kit (yeast), the sequencing library was prepared with TruSeq Stranded mRNA Library Prep Kit (Illumina, CA, USA). The ribosome profiling and

RNA-Seq libraries were sequenced on a HiSeq 4000 platform (Illumina) with a single-end 50 bp.

Sequence data alignment, quantification, and normalization

Alignment and quantification of ribosome profiling and RNA-Seq data were performed as previously described with modifications (McGlincy and Ingolia, 2017). After the removal of the linker sequence and the splitting based on sample barcode, we removed reads that mapped to non-coding RNA (ncRNA) sequences using STAR 2.7.0a (Dobin et al., 2013). Despite the fact that we used an rRNA depletion method for both RNA-seq and Ribo-seq, our libraries were highly contaminated with ncRNA (Table S7), resulting in a relatively low reads number input for the alignment: between 92,676 and 350,221 reads for RNA-seq and between 248,228 and 1,049,718 reads for Ribo-seq. Remaining reads were aligned to the S288C *S. cerevisiae* genome using STAR 2.7.0a (Dobin et al., 2013). For the analysis of the accessory ORFs (open reading frames), the reads were also aligned to all the ORF detected in the pangenome of *S. cerevisiae* (Peter et al., 2018). The A-site offsets of ribosome footprints were determined according to the location of the 5′ end of reads mapped to start codons. For RNA-seq, offsets were set to 15 for all mRNA fragments. Reads corresponding to the first and last five codons of each coding sequence (CDS) were excluded from the analysis. For calculation of transcript per million (TPM) values for each CDS, we normalized read counts by CDS length minus 10 and adjusted sum of all normalized values to one million. The custom scripts will be available upon requests. We finally calculated a translation efficiency (TE) value by dividing the Ribo-seq TPM value by the RNA-seq TPM value for each gene in each isolate.

Expression variation analysis

The TPM normalized datasets for Ribo-seq and RNA-seq were $\log_{10}$-transformed. A first general overview of the strain variation was obtained using Spearman correlation test. Next, on each strains vs. strain pairwise comparisons, we applied Mahalanobis distance (Ho et al., 2018) using the check_outlier() function (from the R package “performance”), where the genes with a distance higher than 10.59 (χ^2 distribution, with a 0.005 alpha level and 2 degrees of freedom) were selected. These genes were then filtered using a linear model on the strains vs. strain pairwise comparison: the residuals coming from the linear model were used to select the genes

displaying 2.5% highest and lowest residuals. We kept the genes that were overlapping between the Mahalanobis detection method and the linear regression residual method. The detected genes are later mentioned as “variable genes”.

Variable genes characteristics

Using genes descriptive data (Dowell et al., 2010; Giaever et al., 2002; Pu et al., 2009), we focused on describing the characteristics of the variable genes. Firstly, we questioned if the variable genes detected earlier displayed any enrichment or depletion of essential genes or genes part of protein complexes with Fisher's exact tests (FET). Gene ontology (GO) analysis (Ashburner et al., 2000) was performed on the geneontology.org website (Gene Ontology Consortium, 2021; Gene Ontology Consortium, 2019), using the subset of genes (N = 3755) encompassed by our Ribo-seq and RNA-seq experiments as the reference list. The p-values were corrected using Bonferroni correction. This was performed on RNA-seq or Ribo-seq variable genes, and on the overlapping variable genes (between the two datasets).

Detection of a post-transcriptional buffering phenomenon

In order to see if expression variation across the 8 isolates was stronger in RNA-seq or Ribo-seq data, we firstly computed the Euclidean distances between each strain using the $\log_{10}$-transformed data from both datasets. We also obtained for each gene the expression variance using the $\log_{10}$-transformed data from both datasets. In addition, we generated $\log_2$ foldchange ($\log_2$(FC)) values in RNA-seq, Ribo-seq and TE datasets, where for each gene in each pairwise comparison:

$$\text{Log}_2(\text{FC}) = log_2(\frac{TPM\ or\ TE\ count\ in\ the\ strain\ 1}{TPM or\ TE\ count\ in\ the\ strain\ 2})$$

Then, in each pairwise comparison, we used Spearman correlation test to compare the behavior of the 3 kinds of $\log_2$(FC) dataset (RNA-seq, Ribo-seq, and TE) against each other. We finally quantified the expression variation (in both RNA-seq and Ribo-seq data) based on the $\log_2$(FC):

$$\text{variation}_{\text{ gene X}} = |log_2(\frac{TPM\ gene\ X\ strain\ A}{TPM\ gene\ X\ strain\ B})|$$

The more this value increases, the more the difference between the TPM values is important.

Buffered and conserved regulation genes characteristics

In each isolate pairwise comparison and using the RNA-seq log_2(FC) vs. TE log_2(FC) comparison, we defined 2 groups of genes by applying a 1.5-foldchange threshold (log_2(FC)≈0.58 or log_2(FC)≈-0.58) for both RNA-seq log_2(FC) vs. TE log_2(FC) (see Figure 3A). This enables to capture genes with a buffered transcriptional variation (in red in Figure 3A) among the genes displaying at least a 1.5 foldchange transcriptional variation. The gene displaying transcriptional variation that were not capture among the buffered genes were considered as unbuffered genes (in blue in Figure 3A). We checked the percentage of genes concerned by post-transcriptional buffering among the genes that displayed transcriptional variation (using a minimal 1.5-foldchange threshold in the pairwise comparison). Genes were considered as recurrently buffered or unbuffered if they were detected in the corresponding group at least in more than half of the isolate pairwise comparison (i.e., detected 15 or more as buffered or unbuffered). We then performed Fisher's exact test with the two gene sets (buffered or unbuffered) to detect enrichment of essential or protein complex-related genes. We also check the proportion of essential genes and protein complex-related genes in the two groups in each pairwise comparison.

We then compare the transcription or translation level of the buffered and unbuffered groups by using for each gene the mean RNA-seq TPM value across the 8 isolates.

Codon usage bias influence

We used the tAI (tRNA adaptation index) index (dos Reis et al., 2004, 2003) to estimate the codon bias usage of each gene. Briefly, tAI is an index showing how much a gene is adapted to the tRNA genome structure in terms of codon usage. In this perspective, we first calculated the tRNA copy number of our 8 strains with tRNAscan-SE with default parameters (Chan and Lowe, 2019; Lowe and Chan, 2016) using the assembled genome sequences from (Peter et al., 2018). Then, for each isolate, we used the tRNA copy number to compute the tAI using a Perl program available on https://github.com/mariodosreis/tai (dos Reis et al., 2004, 2003) using default parameters. The resulting dataset was a tAI value for more than 98.8% of the 3,755 genes (some genes were discarded during the calculation) in each isolate. Ultimately, we could calculate an overall tAI (mean of the 8 or less tAI values) for 3,746 genes. For each isolate, we correlated the expression levels and TE of each gene to its tAI index.

Accessory ORF analysis

Using the mapping done on the *S. cerevisiae* pangenome, we selected the accessory ORFs that were previously detected in each of our 8 strains and selected the ones that had TPM values higher than 0 in both RNA-seq and Ribo-seq data. Then, we calculated the TE of each accessory ORF.

We first focused on the CPI isolate ORF acquired through introgression events (with the *Saccharomyces paradoxus* species). We selected the ORFs known to have an ortholog in *S. cerevisiae* genes. Then using homo/heterozygosity data (adapted from (Peter et al., 2018) gene presence/absence data), we selected the ORF that were homozygous for the *S. paradoxus* allele (n = 18) and we compared their expression with their orthologs in *S. cerevisiae*. We then compared the introgression TPM values vs. their orthologs mean TPM value (obtained from the 7 other strains) to see if we could observe over- or under-expression of these ORFs in comparison with their orthologs.

We explored then the expression levels of the BPL isolate accessory ORFs, especially the ones acquired through horizontal gene transfer (HGT) by comparing their expressions (RNA-seq and Ribo-seq) and TE with the other BPL gene values.

Data availability

All sequencing reads are available in the Gene Expression Omnibus (GEO) under the accession number GSE173654.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173654

Supplementary Material

Supplementary tables available at:
https://www.dropbox.com/scl/fi/782fpyfyarxkbwqas8494/Sup_Tables.xlsx?rlkey=2oztn6rye71qy3rqgotq91nwv&dl=0

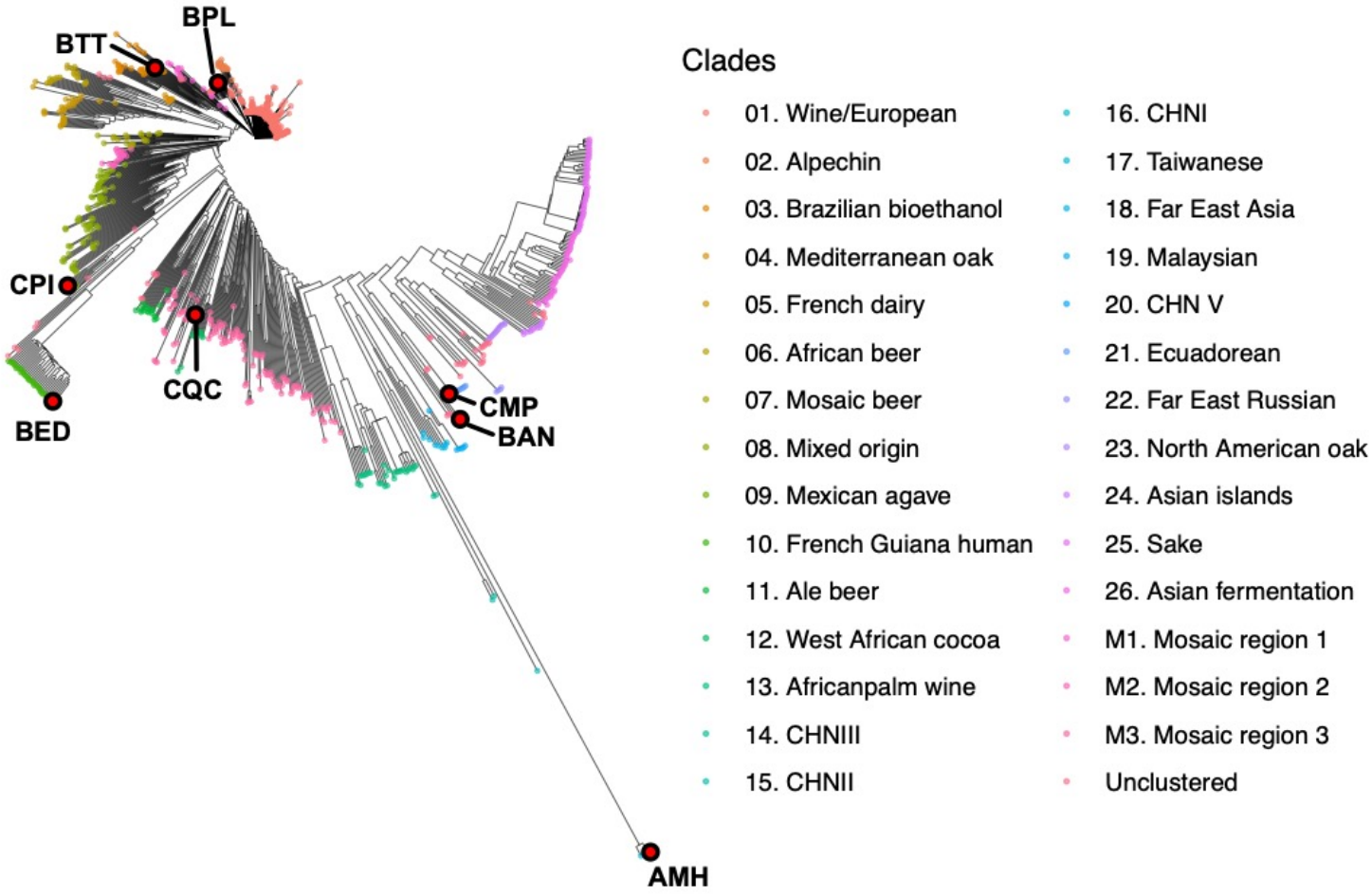


Figure S1. Eight isolates were selected to ensure a maximum genetic diversity Neighbor-joining tree obtained from the biallelic SNPs of 1011 isolates (1). The colors of the points correspond to their clade, and the 8 isolates are highlighted in red.

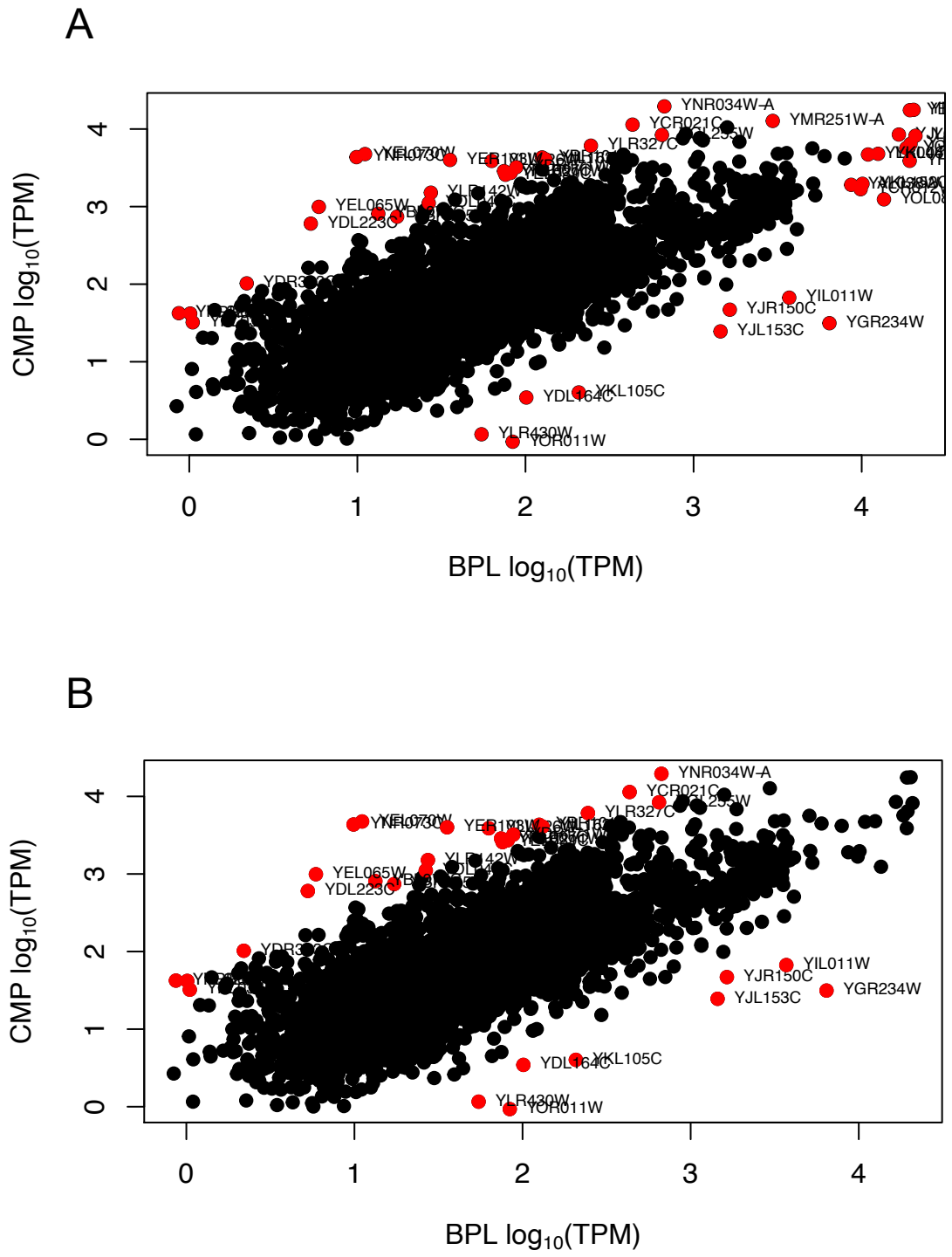


Figure S2. Detection of the genes displaying variable transcription and translation regulation. BPL (*x* axis) vs CMP (*y* axis) pairwise comparison of the $\log_{10}$ transformed Ribo-seq data. (A) The points highlighted in red are the variable genes detected using Mahalanobis distance only. (B) The points highlighted in red are the genes detected using the combination of Mahalanobis distance and linear regression residuals.

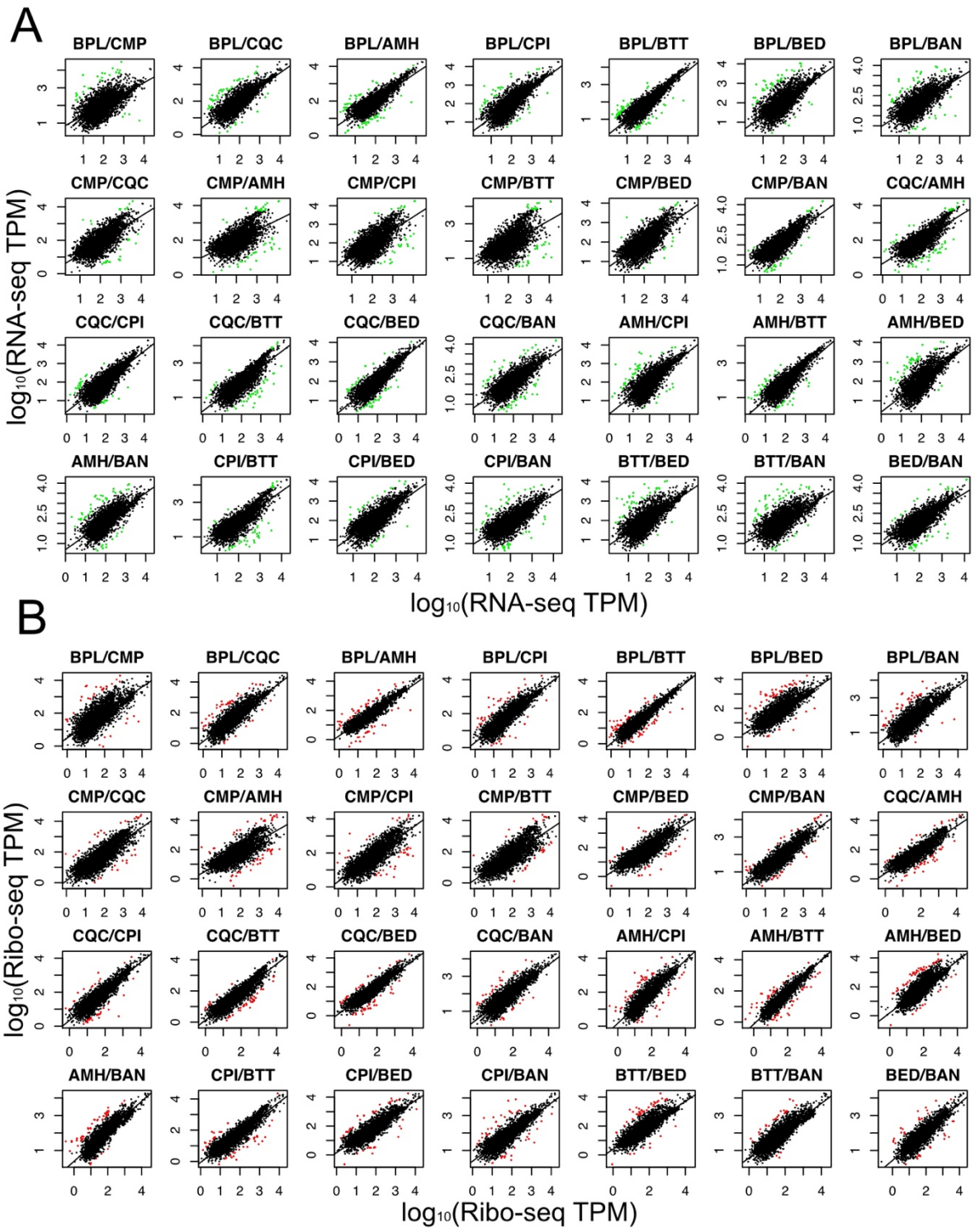


Figure S3. Total variable genes detected using both RNA-seq and Ribo-seq data Variable genes detection in the 28 (A) RNA-seq and (B) Ribo-seq pairwise comparisons. The variable genes for RNA-seq are highlighted in green and the ones for Ribo-seq in red. The black lines correspond to the linear regression obtained in each pairwise comparison.

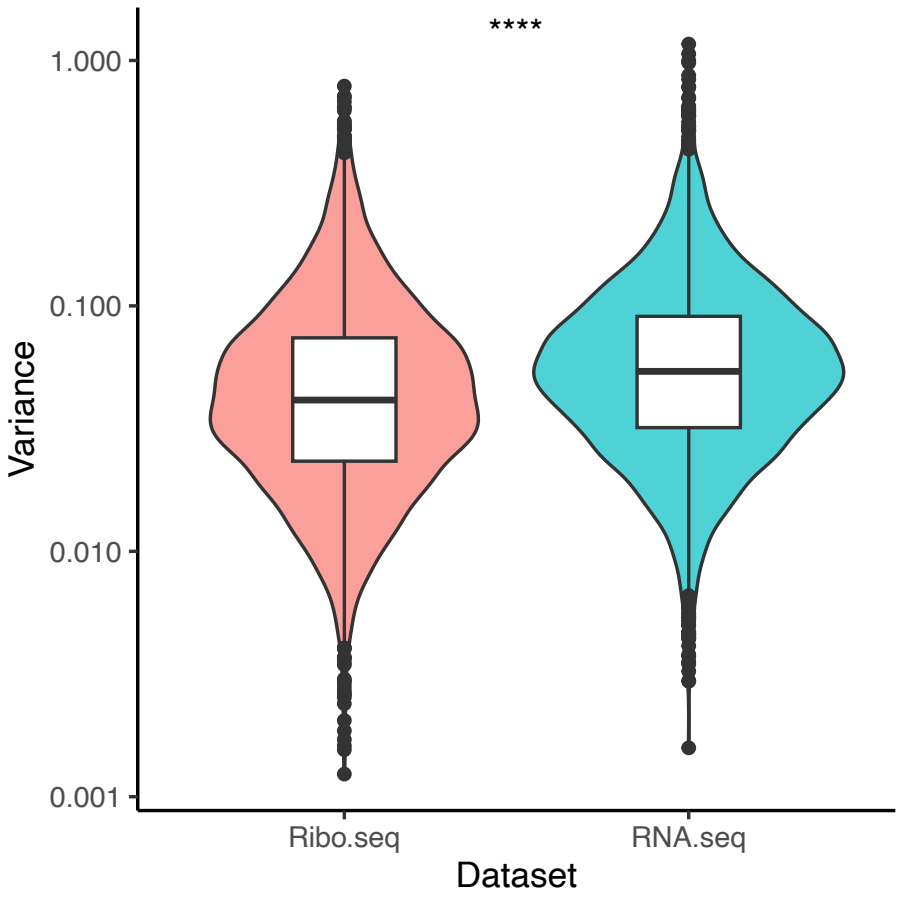


Figure S4. Gene expression variance is higher at the transcriptional level. Gene-wise variance using the $\log_{10}$ transformed data of the two datasets (Wilcoxon test p-value = 5.09×10^{-38}).

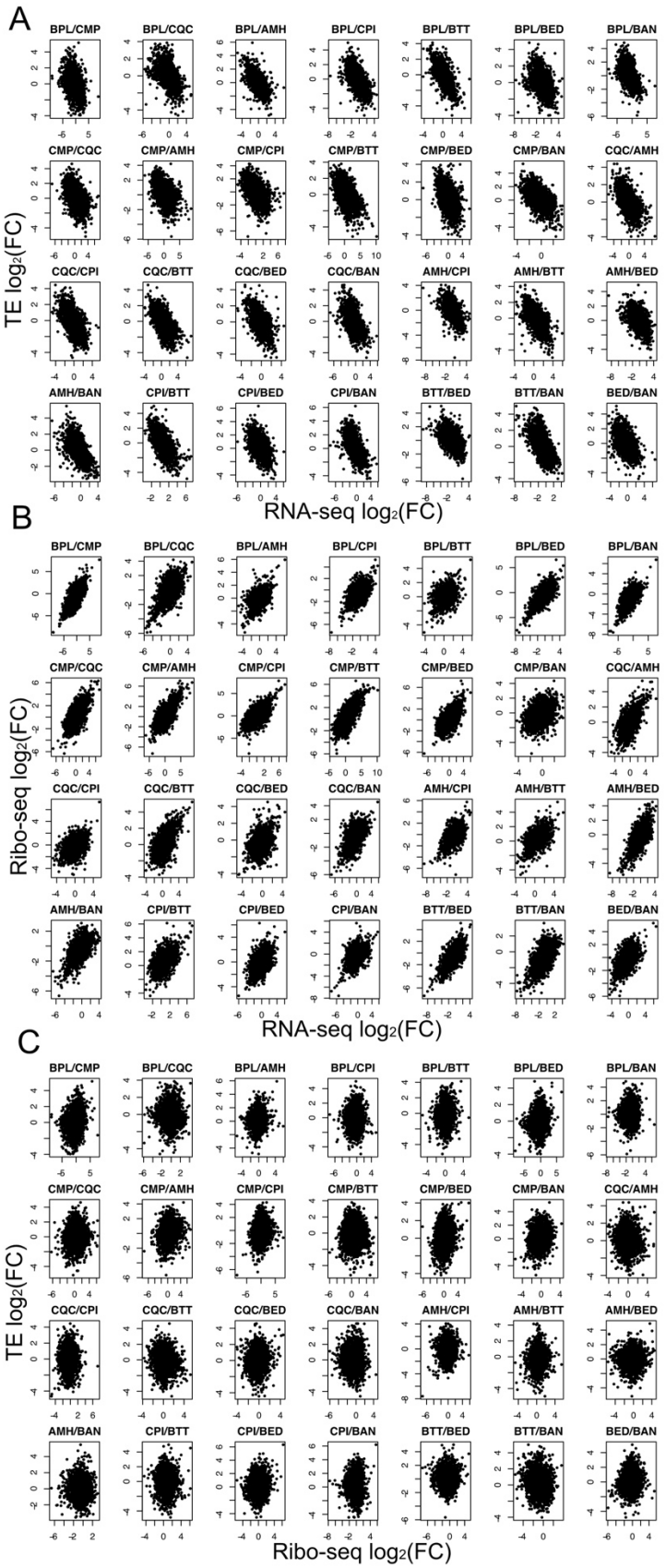


Figure S5. TE modulation is determinant for post-transcriptional buffering. For each pairwise comparison, correlation between: (A) the RNA-seq and TE log_2(FC), (B) the RNA-seq and Ribo-seq log_2(FC) and (C) the Ribo-seq and TE log_2(FC).

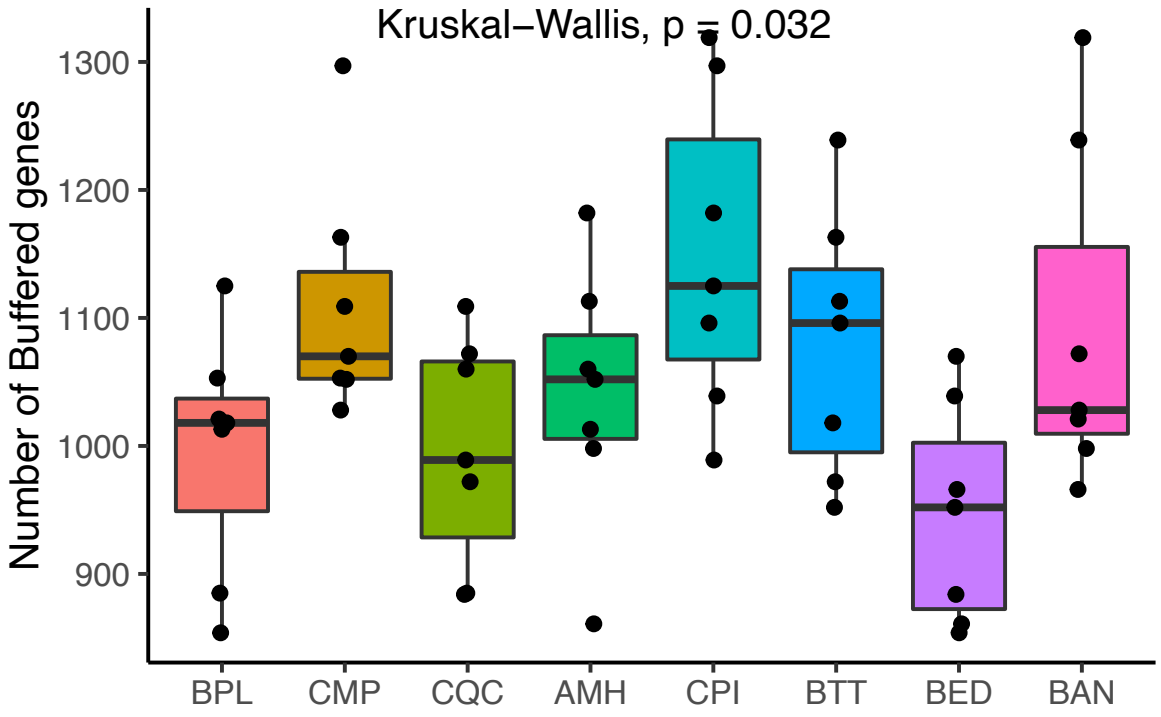


Figure S6. Number of buffered genes in each isolate pairwise comparison. Number of genes in the buffered group for each isolate (each isolate has 7 values corresponding to the 7 pairwise comparisons against the other isolates). The number of genes is slightly different across the isolates (Kruskal-Wallis test).

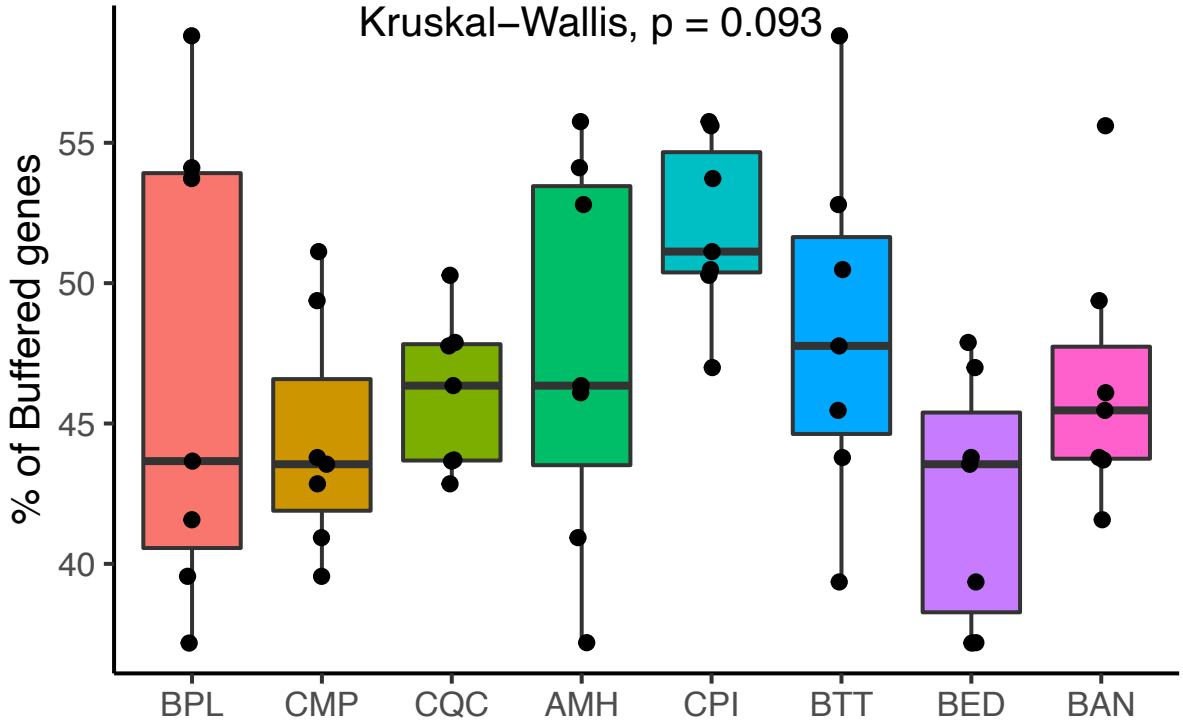


Figure S7. Proportion of buffered genes in each isolate pairwise comparison. Percentage of gene having their variations buffered among the ones that displayed divergent transcription for each isolate (each isolate has 7 percentage values corresponding to the 7 pairwise comparisons against the other isolates). There is no difference in term of percentage among the isolates (Kruskal-Wallis test).

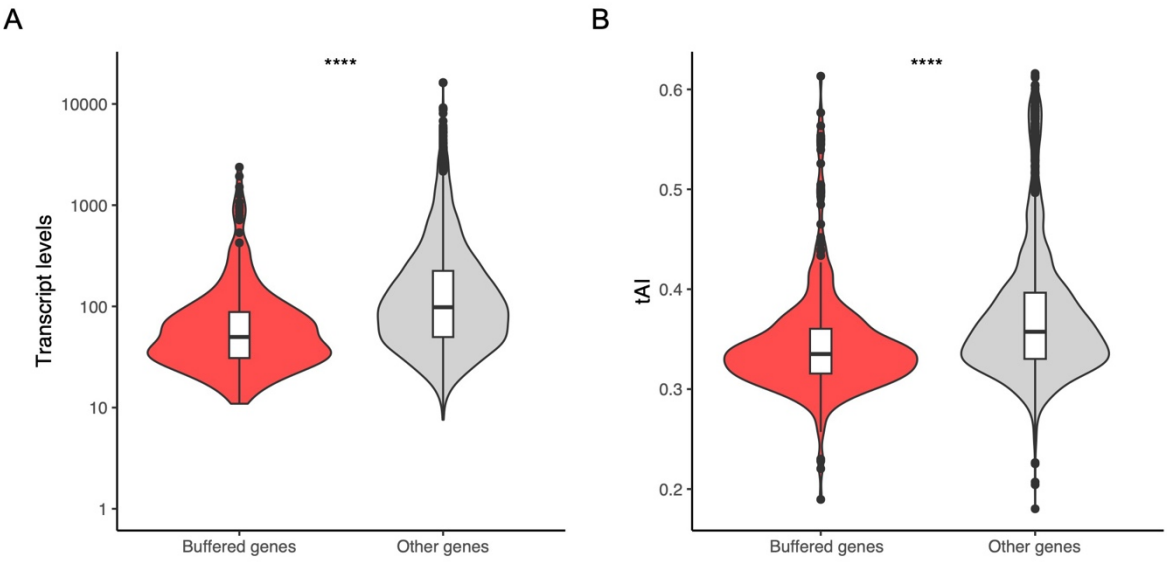


Figure S8. mRNA abundance and tAI difference between the buffered gene the rest of the genes. Difference in the (A) mRNA levels and (B) tAI levels of the buffered genes against all the other genes. Respective Wilcoxon test p-values: 6.15×10^{-33} and 2.42×10^{-22}.

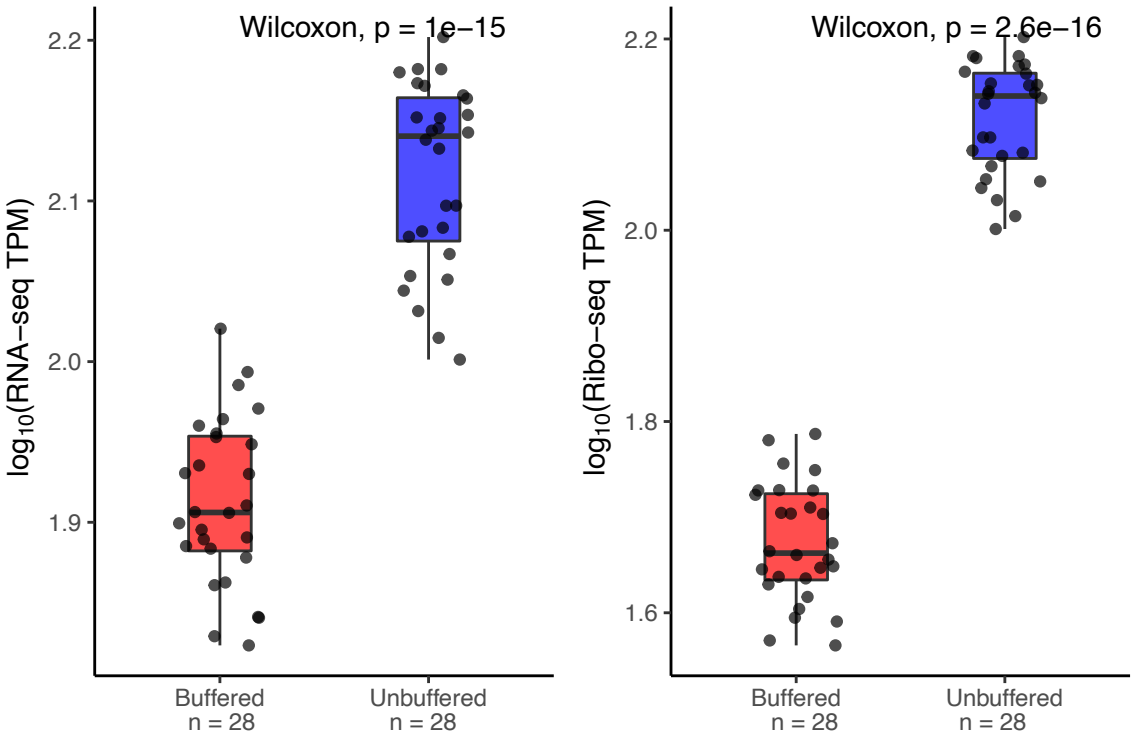


Figure S9. RNA-seq and Ribo-seq levels of the buffered and unbuffered genes in each pairwise comparison. Difference between the average expression level of each gene group (buffered and unbuffered) in each pairwise comparison using both RNA-seq and Ribo-seq data.

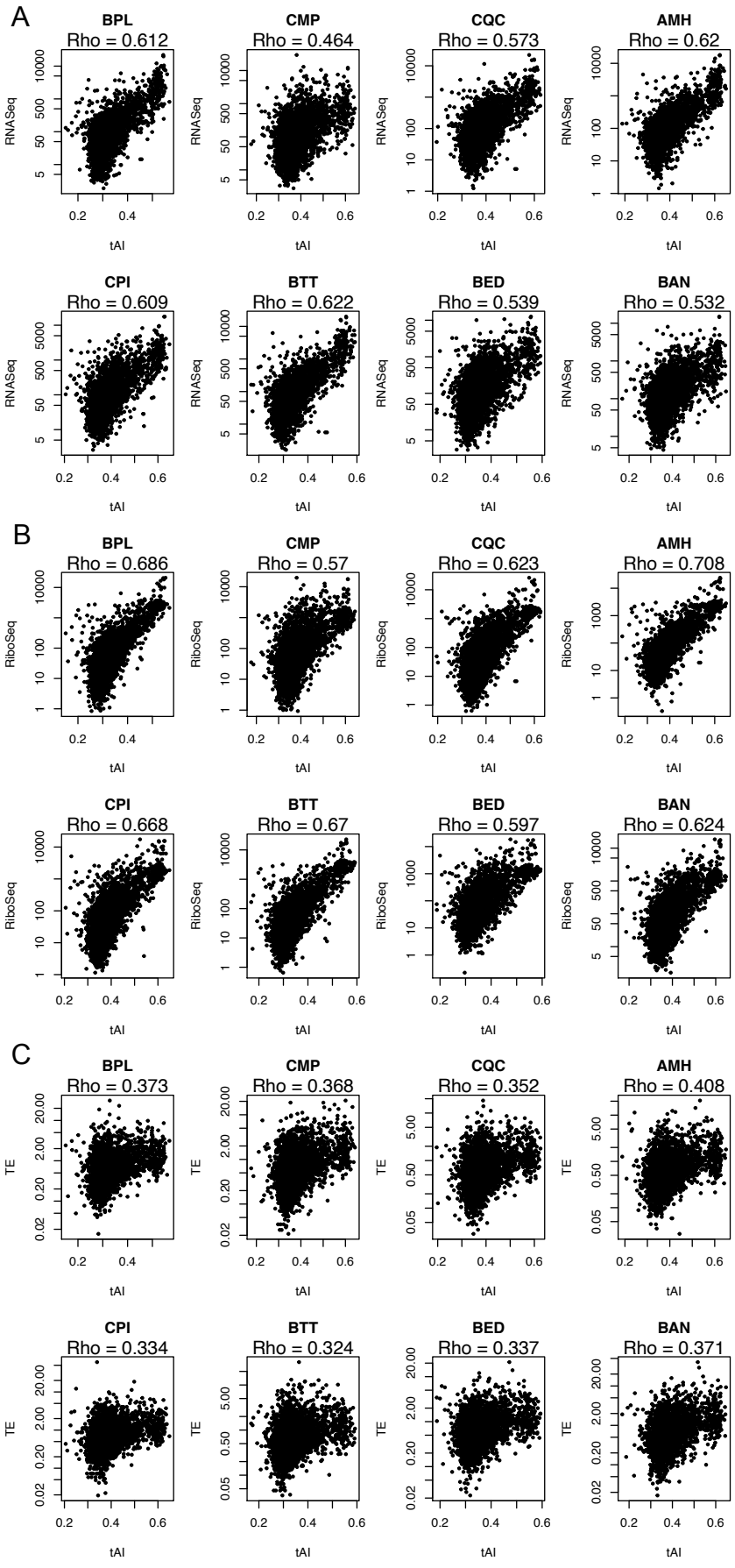

A
BPL
Rho = 0.612
CMP
Rho = 0.464
CQC
Rho = 0.573
AMH
Rho = 0.62
CPI
Rho = 0.609
BTT
Rho = 0.622
BED
Rho = 0.539
BAN
Rho = 0.532
RNASeq
tAI
B
BPL
Rho = 0.686
CMP
Rho = 0.57
CQC
Rho = 0.623
AMH
Rho = 0.708
CPI
Rho = 0.668
BTT
Rho = 0.67
BED
Rho = 0.597
BAN
Rho = 0.624
RiboSeq
tAI
C
BPL
Rho = 0.373
CMP
Rho = 0.368
CQC
Rho = 0.352
AMH
Rho = 0.408
CPI
Rho = 0.334
BTT
Rho = 0.324
BED
Rho = 0.337
BAN
Rho = 0.371
TE
tAI

Figure S10. Comparison between the tAI and the expression levels. Correlation between tAI and $\log_{10}$ value of: (A) RNA-seq, (B) Ribo-seq and (C) TE data. The spearman correlation coefficient is above each plot, all of the coefficient where significant (P-value<0.05)

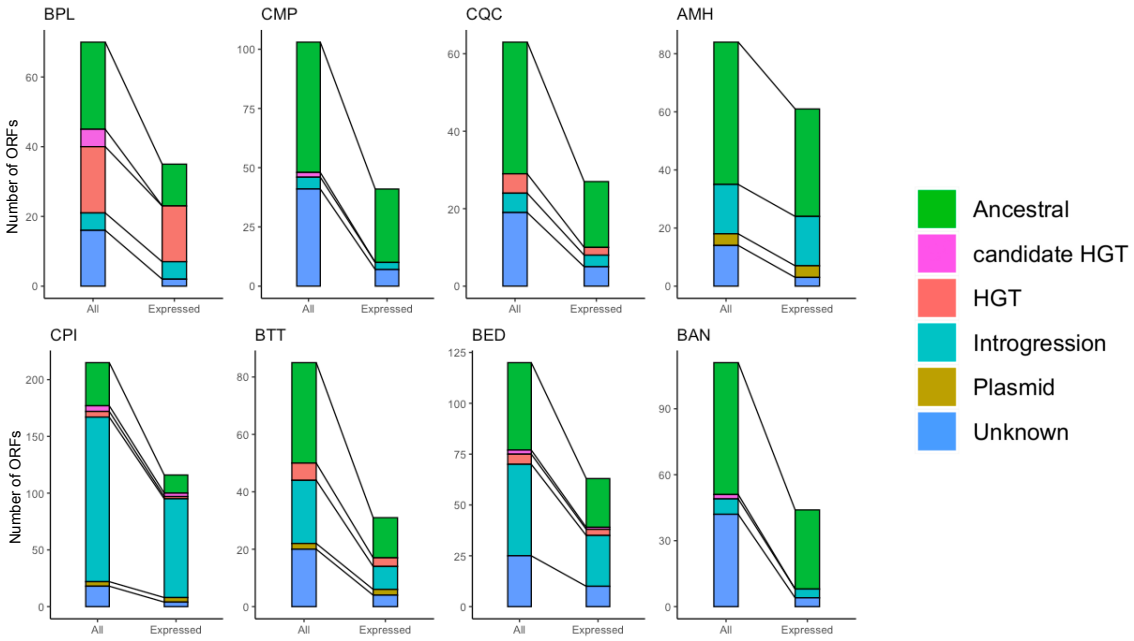


Figure S11. Expression status of the accessory ORF for each isolate. Number of each kind of accessory ORFs in each isolate. For each isolate, the two bar plots correspond to the expected (using the data of Peter *et al*., 2018) accessory ORFs (All) and the ones that were actually transcribed and translated (Expressed).

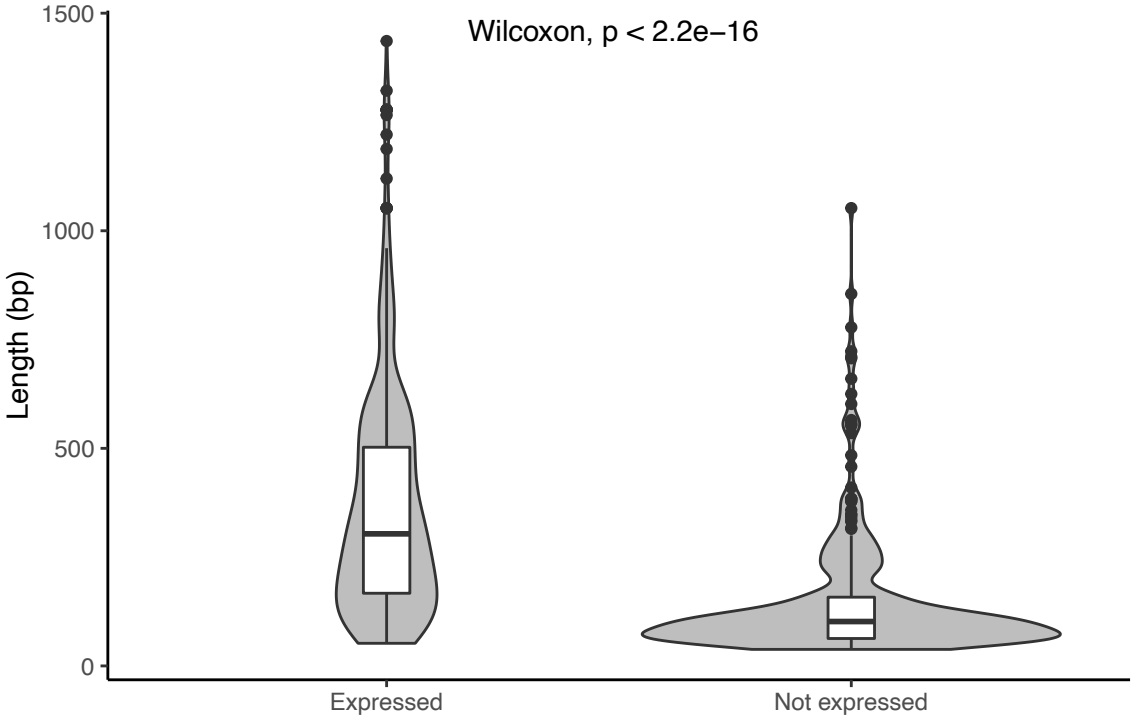


Figure S12. Small accessory ORF tend to be less expressed. Length (bp) difference between the expressed and not expressed accessory ORFs among our 8 isolates. The not-expressed ORFs tended to be shorter (Wilcoxon test)

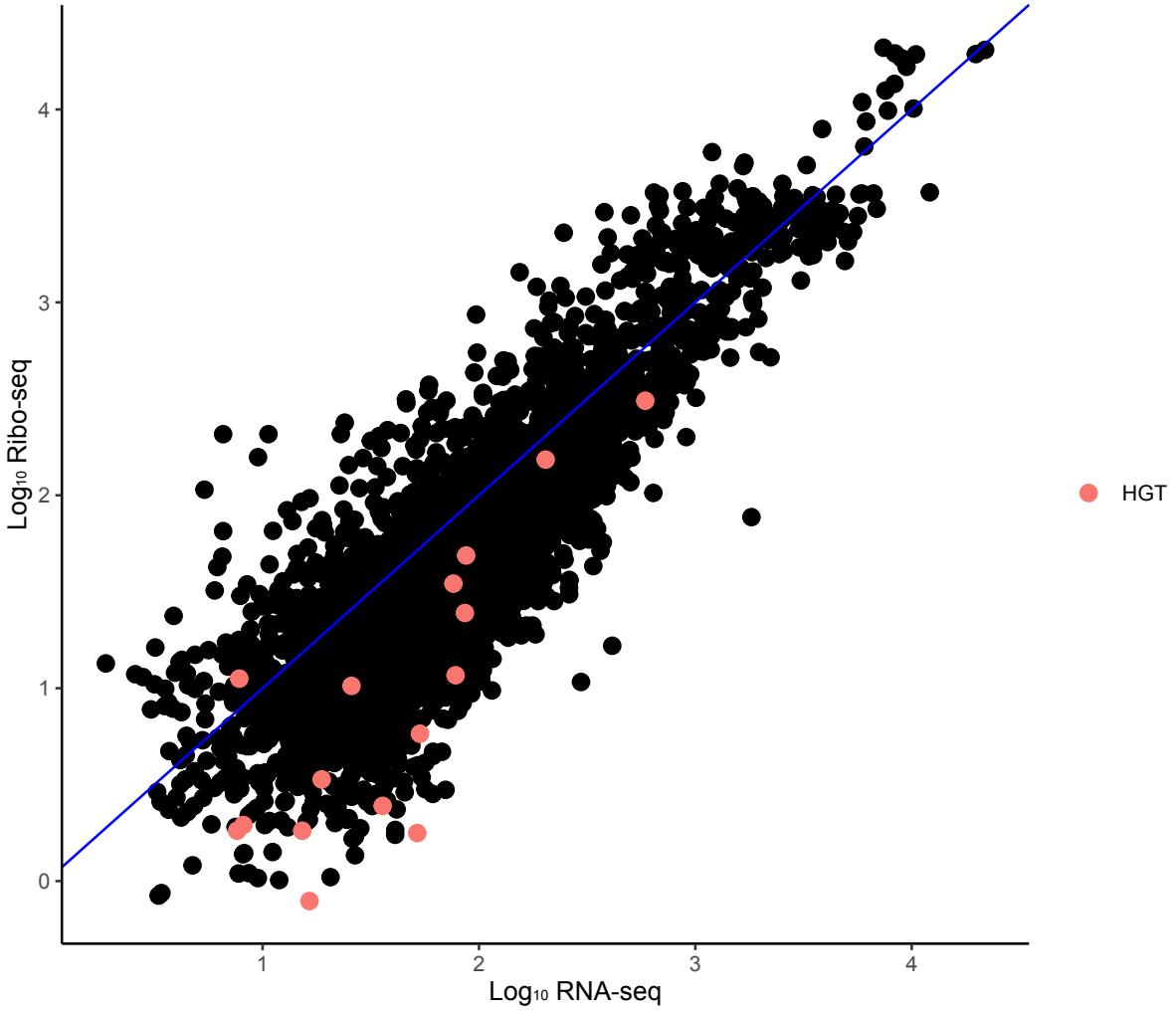


Figure S13. HGT-related ORF are less translated. Log_{10} TPM value of RNA-seq (x axis) and Ribo-seq (y axis) for the strain BPL. The pink points correspond to the HGT accessory ORF. The blue line corresponds to the x=y line.

Bibliography

Albert, F.W., Muzzey, D., Weissman, J.S., Kruglyak, L., 2014. Genetic Influences on Translation in Yeast. PLOS Genet. 10, e1004692. https://doi.org/10.1371/journal.pgen.1004692

Artieri, C.G., Fraser, H.B., 2014. Evolution at two levels of gene expression in yeast. Genome Res. 24, 411–421. https://doi.org/10.1101/gr.165522.113

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. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556

Blevins, W.R., Ruiz-Orera, J., Messeguer, X., Blasco-Moreno, B., Villanueva-Cañas, J.L., Espinar, L., Díez, J., Carey, L.B., Albà, M.M., 2021. Uncovering de novo gene birth in yeast using deep transcriptomics. Nat. Commun. 12, 604. https://doi.org/10.1038/s41467-021-20911-3

Blevins, W.R., Tavella, T., Moro, S.G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., Carey, L.B., Albà, M.M., 2019. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Sci. Rep. 9, 11005. https://doi.org/10.1038/s41598-019-47424-w

Buccitelli, C., Selbach, M., 2020. mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 21, 630–644. https://doi.org/10.1038/s41576-020-0258-4

Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonnière, E., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., Submitted. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast.

Chan, P.P., Lowe, T.M., 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol. Biol. Clifton NJ 1962, 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Chotewutmontri, P., Barkan, A., 2016. Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. PLoS Genet. 12, e1006106. https://doi.org/10.1371/journal.pgen.1006106

Coghlan, A., Wolfe, K.H., 2000. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. Yeast 16, 1131–1145. https://doi.org/10.1002/1097-0061(20000915)16:12<1131::AID-YEA609>3.0.CO;2-F

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., Pelechano, V., Styles, E.B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J.S., Srikumar, T., Bahr, S., Chen, Y., Deshpande, R., Kurat, C.F., Li, S.C., Li, Z., Usaj, M.M., Okada, H., Pascoe, N., San Luis, B.-J., Sharifpoor, S., Shuteriqi, E., Simpkins, S.W., Snider, J., Suresh, H.G., Tan, Y., Zhu, H., Malod-Dognin, N., Janjic, V., Przulj, N., Troyanskaya, O.G., Stagljar, I., Xia, T., Ohya, Y., Gingras, A.-C., Raught, B., Boutros, M., Steinmetz, L.M., Moore, C.L., Rosebrock, A.P., Caudy, A.A., Myers, C.L., Andrews, B., Boone, C., 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science 353, aaf1420. https://doi.org/10.1126/science.aaf1420

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., Giaever, G., 2005. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. Genetics 169, 1915–1925. https://doi.org/10.1534/genetics.104.036871

Dilucca, M., Cimini, G., Semmoloni, A., Deiana, A., Giansanti, A., 2015. Codon Bias Patterns of E. coli's Interacting Proteins. PLOS ONE 10, e0142127. https://doi.org/10.1371/journal.pone.0142127

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. Bioinforma. Oxf. Engl. 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

dos Reis, M., Savva, R., Wernisch, L., 2004. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. 32, 5036–5044. https://doi.org/10.1093/nar/gkh834

dos Reis, M., Wernisch, L., Savva, R., 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Res. 31, 6976–6985. https://doi.org/10.1093/nar/gkg897

Dowell, R.D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D.A., Rolfe, P.A., Heisler, L.E., Chin, B., Nislow, C., Giaever, G., Phillips, P.C., Fink, G.R., Gifford, D.K., Boone, C., 2010. Genotype to phenotype: a complex problem. Science 328, 469. https://doi.org/10.1126/science.1189015

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H., 2005. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. 102, 14338–14343. https://doi.org/10.1073/pnas.0504070102

Drummond, D.A., Raval, A., Wilke, C.O., 2006. A Single Determinant Dominates the Rate of Yeast Protein Evolution. Mol. Biol. Evol. 23, 327–337. https://doi.org/10.1093/molbev/msj038

Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J., Eisen, M.B., 2004. Noise Minimization in Eukaryotic Gene Expression. PLOS Biol. 2, e137. https://doi.org/10.1371/journal.pbio.0020137

Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W., 2002. Evolutionary rate in the protein interaction network. Science 296, 750–752. https://doi.org/10.1126/science.1068696

Gene Ontology Consortium, 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334. https://doi.org/10.1093/nar/gkaa1113

Gene Ontology Consortium, T., 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338. https://doi.org/10.1093/nar/gky1055

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A.P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kötter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., Johnston, M., 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391. https://doi.org/10.1038/nature00935

Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., Beltrao, P., 2017. Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. Cell Syst. 5, 386-398.e4. https://doi.org/10.1016/j.cels.2017.08.013

Ho, B., Baryshnikova, A., Brown, G.W., 2018. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Syst. 6, 192-205.e3. https://doi.org/10.1016/j.cels.2017.12.004

Hodgins-Davis, A., Adomas, A.B., Warringer, J., Townsend, J.P., 2012. Abundant Gene-by-Environment Interactions in Gene Expression Reaction Norms to Copper within Saccharomyces cerevisiae. Genome Biol. Evol. 4, 1061–1079. https://doi.org/10.1093/gbe/evs084

Hrdlickova, R., Toloue, M., Tian, B., 2017. RNA-Seq methods for transcriptome analysis. Wiley Interdiscip. Rev. RNA 8. https://doi.org/10.1002/wrna.1364

Ingolia, N.T., 2014. Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. 15, 205–213. https://doi.org/10.1038/nrg3645

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218–223. https://doi.org/10.1126/science.1168978

Ingolia, N.T., Hussmann, J.A., Weissman, J.S., 2019. Ribosome Profiling: Global Views of Translation. Cold Spring Harb. Perspect. Biol. 11. https://doi.org/10.1101/cshperspect.a032698

Jiang, L., Wang, M., Lin, S., Jian, R., Li, Xiao, Chan, J., Dong, G., Fang, H., Robinson, A.E., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., MacArthur, D.G., Meier, S.R., Nedzel, J.L., Nguyen, D.Y., Segrè, A.V., Todres, E., Balliu, B., Barbeira, A.N., Battle, A., Bonazzola, R., Brown, A., Brown, C.D., Castel, S.E., Conrad, D., Cotter, D.J., Cox, N., Das, S., Goede, O.M. de, Dermitzakis, E.T., Engelhardt, B.E., Eskin, E., Eulalio, T.Y., Ferraro, N.M., Flynn, E., Fresard, L., Gamazon, E.R., Garrido-Martín, D., Gay, N.R., Guigó, R., Hamel, A.R., He, Y., Hoffman, P.J., Hormozdiari, F., Hou, L., Im, H.K., Jo, B., Kasela, S., Kellis, M., Kim-Hellmuth, S., Kwong, A., Lappalainen, T., Li, Xin, Liang, Y., Mangul, S., Mohammadi, P., Montgomery, S.B., Muñoz-Aguirre, M., Nachun, D.C., Nobel, A.B., Oliva, M., Park, YoSon, Park, Yongjin, Parsana, P., Reverter, F., Rouhana, J.M., Sabatti, C., Saha, A., Skol, A.D., Stephens, M., Stranger, B.E., Strober, B.J., Teran, N.A., Viñuela, A., Wang, G., Wen, X., Wright, F., Wucher, V., Zou, Y., Ferreira, P.G., Li, G., Melé, M., Yeger-Lotem, E., Barcus, M.E., Bradbury, D., Krubit, T., McLean, J.A., Qi, L., Robinson, K., Roche, N.V., Smith, A.M., Sobin, L., Tabor, D.E., Undale, A., Bridge, J., Brigham, L.E., Foster, B.A., Gillard, B.M., Hasz, R., Hunter, M., Johns, C., Johnson, M., Karasik, E., Kopen, G., Leinweber, W.F., McDonald, A., Moser, M.T., Myer, K., Ramsey, K.D., Roe, B., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Jewell, S.D., Rohrer, D.C., Valley, D.R., Davis, D.A., Mash, D.C., Branton, P.A., Barker, L.K., Gardiner, H.M., Mosavel, M., Siminoff, L.A., Flicek, P., Haeussler, M., Juettemann, T., Kent, W.J.,

Lee, C.M., Powell, C.C., Rosenbloom, K.R., Ruffier, M., Sheppard, D., Taylor, K., Trevanion, S.J., Zerbino, D.R., Abell, N.S., Akey, J., Chen, L., Demanelis, K., Doherty, J.A., Feinberg, A.P., Hansen, K.D., Hickey, P.F., Jasmine, F., Kaul, R., Kibriya, M.G., Li, J.B., Li, Q., Linder, S.E., Pierce, B.L., Rizzardi, L.F., Smith, K.S., Stamatoyannopoulos, J., Tang, H., Carithers, L.J., Guan, P., Koester, S.E., Little, A.R., Moore, H.M., Nierras, C.R., Rao, A.K., Vaught, J.B., Volpi, S., Snyder, M.P., 2020. A Quantitative Proteome Map of the Human Body. Cell 183, 269-283.e19. https://doi.org/10.1016/j.cell.2020.08.036

Jüschke, C., Dohnal, I., Pichler, P., Harzer, H., Swart, R., Ammerer, G., Mechtler, K., Knoblich, J.A., 2013. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. Genome Biol. 14, r133. https://doi.org/10.1186/gb-2013-14-11-r133

Kita, R., Venkataram, S., Zhou, Y., Fraser, H.B., 2017. High-resolution mapping of cis-regulatory variation in budding yeast. Proc. Natl. Acad. Sci. U. S. A. 114, E10736–E10744. https://doi.org/10.1073/pnas.1717421114

Kustatscher, G., Grabowski, P., Rappsilber, J., 2017. Pervasive coexpression of spatially proximal genes is buffered at the protein level. Mol. Syst. Biol. 13, 937. https://doi.org/10.15252/msb.20177548

Larrimore, K.E., Rancati, G., 2019. The conditional nature of gene essentiality. Curr. Opin. Genet. Dev., Evolutionary genetics 58–59, 55–61. https://doi.org/10.1016/j.gde.2019.07.015

Li, G.-W., Burkhardt, D., Gross, C., Weissman, J.S., 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157, 624–635. https://doi.org/10.1016/j.cell.2014.02.033

Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57. https://doi.org/10.1093/nar/gkw413

Lu, P., Vogel, C., Wang, R., Yao, X., Marcotte, E.M., 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat. Biotechnol. 25, 117–124. https://doi.org/10.1038/nbt1270

Lukoszek, R., Feist, P., Ignatova, Z., 2016. Insights into the adaptive response of Arabidopsis thaliana to prolonged thermal stress by ribosomal profiling and RNA-Seq. BMC Plant Biol. 16, 221. https://doi.org/10.1186/s12870-016-0915-0

Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E., Dequin, S., Galeote, V., 2015. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Mol. Biol. Evol. 32, 1695–1707. https://doi.org/10.1093/molbev/msv057

McGlincy, N.J., Ingolia, N.T., 2017. Transcriptome-wide measurement of translation by ribosome profiling. Methods San Diego Calif 126, 112–129. https://doi.org/10.1016/j.ymeth.2017.05.028

McManus, C.J., May, G.E., Spealman, P., Shteyman, A., 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422–430. https://doi.org/10.1101/gr.164996.113

Messner, C.B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleder, M., Ralser, M., 2022. Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. PROTEOMICS n/a, 2200013. https://doi.org/10.1002/pmic.202200013

Mito, M., Mishima, Y., Iwasaki, S., 2020. Protocol for Disome Profiling to Survey Ribosome Collision in Humans and Zebrafish. STAR Protoc. 1, 100168. https://doi.org/10.1016/j.xpro.2020.100168

Morrill, S.A., Amon, A., 2019. Why haploinsufficiency persists. Proc. Natl. Acad. Sci. 116, 11866–11871. https://doi.org/10.1073/pnas.1900437116

Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L., Wincker, P., Casaregola, S., Dequin, S., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc. Natl. Acad. Sci. 106, 16333–16338. https://doi.org/10.1073/pnas.0904673106

Ohnuki, S., Ohya, Y., 2018. High-dimensional single-cell phenotyping reveals extensive haploinsufficiency. PLOS Biol. 16, e2005130. https://doi.org/10.1371/journal.pbio.2005130

Pál, C., Papp, B., Lercher, M.J., 2006. An integrated view of protein evolution. Nat. Rev. Genet. 7, 337–348. https://doi.org/10.1038/nrg1838

Papp, B., Pál, C., Hurst, L.D., 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429, 661–664. https://doi.org/10.1038/nature02636

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011

Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5
Plotkin, J.B., Kudla, G., 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12, 32–42. https://doi.org/10.1038/nrg2899
Pu, S., Wong, J., Turner, B., Cho, E., Wodak, S.J., 2009. Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. 37, 825–831. https://doi.org/10.1093/nar/gkn1005
Rancati, G., Moffat, J., Typas, A., Pavelka, N., 2018. Emerging and evolving concepts in gene essentiality. Nat. Rev. Genet. 19, 34–49. https://doi.org/10.1038/nrg.2017.74
Rocha, E.P.C., 2006. The quest for the universals of protein evolution. Trends Genet. 22, 412–416. https://doi.org/10.1016/j.tig.2006.06.004
Suhre, K., McCarthy, M.I., Schwenk, J.M., 2020. Genetics meets proteomics: perspectives for large population-based studies. Nat. Rev. Genet. 1–19. https://doi.org/10.1038/s41576-020-0268-2
Taggart, J.C., Li, G.-W., 2018. Production of Protein-Complex Components is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Syst. 7, 580-589.e4. https://doi.org/10.1016/j.cels.2018.11.003
The GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110
Trösch, R., Barahimipour, R., Gao, Y., Badillo-Corona, J.A., Gotsmann, V.L., Zimmer, D., Mühlhaus, T., Zoschke, R., Willmund, F., 2018. Commonalities and differences of chloroplast translation in a green alga and land plants. Nat. Plants 4, 564–575. https://doi.org/10.1038/s41477-018-0211-0
Veitia, R.A., Potier, M.C., 2015. Gene dosage imbalances: action, reaction, and models. Trends Biochem. Sci. 40, 309–317. https://doi.org/10.1016/j.tibs.2015.03.011
Wang, Z., Sun, X., Zhao, Y., Guo, X., Jiang, H., Li, H., Gu, Z., 2015. Evolution of gene regulation during transcription and translation. Genome Biol. Evol. 7, 1155–1167. https://doi.org/10.1093/gbe/evv059
Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M.E., Peters, A.H.F.M., Anders, S., Kaessmann, H., 2020. Transcriptome and translatome co-evolution in mammals. Nature 588, 642–647. https://doi.org/10.1038/s41586-020-2899-z
Zhang, J., Yang, J.-R., 2015. Determinants of the rate of protein sequence evolution. Nat. Rev. Genet. 16, 409–420. https://doi.org/10.1038/nrg3950

CHAPTER II

Metabolism adaptation is a main driver of protein abundance evolution in the yeast Saccharomyces cerevisiae.

Collaborative work from:
Elie M. Teyssonniere[1], Benjamin Dubreuil[2], Emmanuel D. Levy[2] and Joseph Schacherer[1,3]

1. Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France
2. Department of Chemical and Structural Biology, Weizmann Institute of Science. Rehovot, Israel
3. Institut Universitaire de France (IUF), Paris, France

Abstract

Variation in gene expression among individuals is one of the major causes of the phenotypic diversity observed in natural populations. Expression variation can occur at any step of the gene expression process, and numerous human diseases have been linked to transcriptional or post-transcriptional variation. However, the mechanisms that influence the evolution of each level of gene expression remain poorly understood. Here, we monitored the proteome of 8 natural isolates of *Saccharomyces cerevisiae* and compared it with the transcriptome and translatome previously obtained from the same set of strains. We found that the proteome variations were mainly related to metabolism or respiration related genes. Interestingly, we found that the proteome variations differed from the transcriptome and proteome variations in part due to post-transcriptional buffering. In addition, we observed that translational variation (measured by ribosome profiling) was only slightly better at reflecting variation in protein abundance than transcriptional variation. Finally, we examined the factors that influence gene expression evolution at each gene expression level. We found that despite expression level specificities, similar evolutionary constraints affect all steps of gene expression. For example, genes encoding highly interacting proteins showed more conserved gene expression regulation, while metabolism-related genes showed faster gene expression evolution. Our results highlight that adaptation to different trophic conditions is a major driver of gene expression evolution across individuals.

Introduction

Gene expression is the main driver of the relationship between the phenotypic landscape observed in natural populations and the genetic mechanisms underlying this large phenotypic diversity. Modification of gene expression, and more specifically protein abundance, is a known source of phenotypic variation (Albert and Kruglyak, 2015; Maurano et al., 2012). In recent decades, the increased accessibility of various gene expression quantification methods, such as RNA sequencing or LC-MS/MS, has allowed several studies to be conducted to examine mRNA or protein levels across healthy or natural individuals (Battle et al., 2015; Ferkingstad et al., 2021; The GTEx Consortium, 2015). In addition, the development of techniques such as ribosome profiling has led to the precise study of the translation process (Ingolia et al., 2009). This technique is based on the sequencing of mRNA fragments protected by ribosomes during translation and is considered a better proxy for protein abundance than transcript abundance (Brar and Weissman, 2015; Ingolia, 2010). While several pathologies are known to be associated with changes in gene expression (Corbett, 2018; Lee and Young, 2013), physiological and benign changes in either transcript or protein abundance are repeatedly observed across individuals (Battle et al., 2015; Jiang et al., 2020; Niu et al., 2023; The GTEx Consortium, 2015).

However, it is well known that the natural variation in gene expression between individuals differs depending on the level at which gene expression is considered. In fact, several studies have highlighted that expression variation tends to decrease as the gene expression process progresses. This phenomenon, called post-transcriptional buffering, has been observed when both proteomic and ribosome profiling data are compared with transcriptomic data (Artieri and Fraser, 2014; Blevins et al., 2019; Dephoure et al., 2014; Gonçalves et al., 2017; McManus et al., 2014; Wang et al., 2020). The mechanisms underlying this phenomenon remain unclear. Several cellular processes such as autoregulation or stoichiometry control are thought to be involved in such a phenomenon, but they are not sufficient to fully explain the extent of post-transcriptional buffering (Buccitelli and Selbach, 2020). Overall, this phenomenon suggests that protein abundance is more conserved than transcript abundance and is therefore subject to different evolutionary constraints. Yet, the comparison between the determinants of gene expression evolution between each layer of expression is still lacking, mainly because few studies have focused on all expression levels together.

To explore what are the determinants of gene expression evolution across individuals, we precisely quantified the protein abundance of 8 natural isolates of *Saccharomyces cerevisiae* previously used for RNA sequencing and ribosome profiling experiments (Teyssonniere et al., Submitted). We combined these 3 datasets and observed that the variation in protein abundance was mainly related to metabolic genes, which is consistent with the variation observed at the transcriptome and translatome levels (Teyssonniere et al., Submitted). Consistent with the previous observation of post-transcriptional buffering, we observed that protein abundance was less variable than both transcript abundance and ribosome-protected mRNA fragments (RPF) abundance. Interestingly, we observed that the ribosome profiling data was only slightly more correlated with protein abundance than the transcriptomic data when looking at the mRNA-protein correlation of each individual. In addition, when looking at the gene-wise mRNA-protein correlation (*i.e.* the correlation between the abundance variation across the isolate), RPF was not more correlated with protein abundance than mRNA abundance. Finally, we explored the determinants of gene expression evolution at each level (mRNA, RPF and protein abundance) and found that if the proteome tends to have specific constraints, several general rules shaped the evolution of each gene expression level. Consistent with the results above, metabolism-related genes tended to have the faster evolutionary rate across all expression layers. Conversely, highly interacting genes or genes involved in central cellular processes were associated with higher evolutionary constraints in each layer. Taken together, our results provide a more accurate picture of how gene expression evolves at each step of the expression process.

Results

Proteome quantification of 8 S. cerevisiae isolates

We quantitatively profiled the proteomes of eight isolated strains of *S. cerevisiae* (Table S1) enabling the detection of 28,800 (±3000) peptides per strain, including 93.8% of unique peptides (Figure S1). Our protein identification against the *S. cerevisiae* S288C reference proteome clustered these peptides in 3635 protein groups. Before analysis, we removed 20 hits corresponding to common lab contaminants, 52 hits corresponding to reverse sequences and 77 hits matching multiple-proteins (mostly duplicated proteins). For quantitative analysis, we only consider proteins identified by at least two unique peptides yielding 3,429 single-protein hits (Figure 1A, Table S2). The protein levels were quantified based on average peptide intensities determined by label-free quantification method (Cox et al., 2014). Therefore, the profiled intracellular proteome covers about half of the reference S288C *S. cerevisiae* proteome, and accounts for ~70% of cytoplasmic proteins in budding yeast. A significant proportion of the detected proteins (66%, n=2,280) were ubiquitously expressed across all eight strains while the remaining hits (34%, n=1,149) had partially missing protein levels within samples. Among those hits, a majority (73%, n=837) had been quantified in at least one replicate of every strain, and more than half (57%, n=660) in at least two replicates of every strain. To avoid discarding potentially valuable data, we decided to impute the missing protein intensities based on the distribution of quantified hits. Thus, 8% of protein intensities were imputed among all samples (ranging from 5% for AMH to 12% for CQC) for 10-30% of proteins per strain where at least one sample had a missing value (Figure S2A). For every strain, the variation of protein expression due to imputation should be negligible (below 2-fold between the 25th-75th percentiles) and at most comprised between 2 to 4-fold (5th-95th percentiles of $\Delta_{intensity}$), except for CQC where this difference rises to about 10-fold. Nevertheless, the difference of intensity ($\Delta_{intensity}$) between observed and imputed expression was centered on 0 for most strains, and slightly shifted for CQC (-0.12) and CMP (-0.04) towards lower expression after imputation (Figure S2B).

Overall, the 3,429 proteins encompassed genes with high transcripts level (Figure 1B) and a functional enrichment using GO annotation (Ashburner et al., 2000; Gene Ontology Consortium, 2021) revealed that several groups of genes tended to be more captured by our proteome exploration: genes related to protein transport, translation, ribosome, transcription and finally metabolism paths were overrepresented among the 3,429 proteins (Figure S3, table S3). Additionally, we observed an enrichment of essential genes and genes related to protein

complexes among the captured proteins. Overall, the proteome profiles were very similar among the 8 isolates as indicated by the strong correlation of their average protein expression (Rho>0.9, Figure 1C), with BPL displaying a slightly different profile from the others. Furthermore, protein levels are considerably stable within strains compared to their variation across different strains. Indeed, we note a very weak variation between biological and technical replicates among each strain (R>0.97), except for CMP and CQC (Figure S4A). The CMP biological replicates are slightly less correlated than their technical replicates (R_{bio}>0.95 vs. R_{tech}>0.97). For CQC, the correlation between biological replicates is also not as strong (0.86>R_{bio}>0.90) with one pair of technical replicates being quite reproducible ($R_{tech.1}$=0.97) unlike the other pair ($R_{tech.2}$=0.93). The principal component analysis captures 53% of the variability in expression among samples (Figure S4B) with noticeable higher biological variation for CMP and CQC strains as seen by the greater distance between their samples compared to the other strains. In the case of CQC, we can also observe a higher distance among the technical replicates, indicating their lack of reproducibility.

In addition, we also measured the variation of expression at the gene-level. We calculated the CV (measured as a percentage) for each gene and found that it ranged from 3 to 282%, with a median of 27% (Figure S5). Using gene set enrichment analysis (GSEA) with the R package *fgsea* (Korotkevich et al., 2021; Subramanian et al., 2005), we found that metabolism-related genes were enriched among the most variable genes (Table S4).

Recently, RNA sequencing and ribosome profiling were performed on the above 8 isolates (Teyssonnière et al., submitted). We sought to compare the transcriptomes, translatomes, and proteomes of these 8 isolates, which overlapped on 2,840 genes. Using the mean protein abundance for each gene across the 4 replicates and by combining it with the RNA-seq and ribo-seq data, we performed LOESS normalization of the 3 datasets together to accurately analyze the difference in abundance variation across expression levels (Figure S6, Table S5).

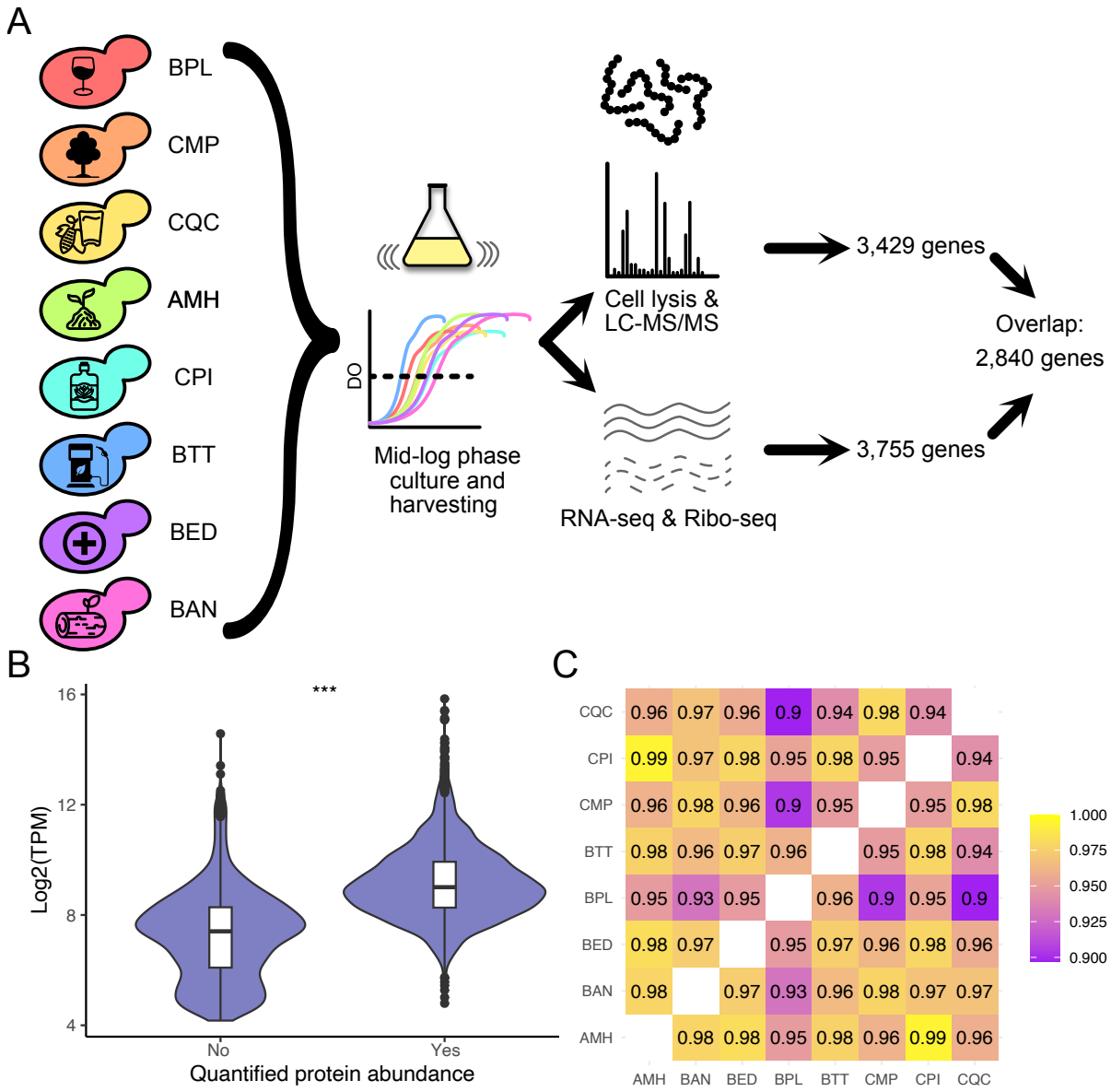


Figure 1: Generation of the proteomes of 8 isolates of *S. cerevisiae*.

(A) We generated the accurate proteome of 8 *S. cerevisiae* isolates from different ecological origins from mid-log phase culture in synthetic complete medium. We combined these data with RNA-seq and ribo-seq data generated on the same set of strains (Teyssonnière et al., submitted). (B) Enclosed proteins tend to be more transcribed than the rest of the proteins. (C) Correlation matrix between the isolates. Overall, the proteomic profiles tend to be very similar.

Strain specific proteome variations

Protein abundance is known to vary among individuals and is a major driver of phenotypic variation and adaptation to environmental fluctuations. Since the 8 strains used in our study were isolates from very different environments, we sought to detect possible adaptation in protein abundance. To this end, we performed a differential expression analysis for each strain.

Based on the replicated protein level data, we compared the abundance of each protein between isolates using the linear regression-based method LIMMA (Ritchie et al., 2015; Smyth, 2005). We used a one-vs-all strategy, which allowed us to accurately determine the strain-specific under- or over-abundance of each protein (Figure 2A, Figure S7). We considered a gene to be differentially expressed if the fold change and the FDR-adjusted p-value reported by LIMMA were respectively higher than 1.2 and lower than 10^{-5}. Overall, across the 8 isolates, we detected 317 cases of differentially expressed proteins (DEP), corresponding to 244 unique proteins (Table S6, Figure S8). In addition, protein expression has been quantified in all samples for 68% of the DEP (*i.e.,* no imputed expression), and for 92% of the DEP in half of the samples at least and at most completely missing from one or two strains. Hence, our imputation of missing values should have a low impact on the result of the differential expression analysis, since we also average the expression between all strains. Surprisingly, we found that the DEP number greatly varied across the isolate. It ranged between 9 for the AMH isolate and 193 for the BPL isolate. The high number of DEP in BPL is in line with its different proteome profile (Figure 1C). We looked for functional enrichment among the DEP using the gene ontology (GO) annotation and found that the enriched features were mainly related to either amino-acid metabolism or respiration function (Table S9, Figure S9). This suggests that intraspecies variability in terms of protein abundance seems to be the result of metabolic adaptations to different trophic conditions, as it is observed using the RNA-seq data (Caudal et al., 2023).

We then focused on isolate-specific adaptation of protein abundance. It was possible to relate some of the DEP to the environmental origin of the isolates. For example, BPL showed a very important overexpression of *PDC5*, a minor isoform of pyruvate decarboxylase that is essential for alcoholic fermentation. Since BPL is a wine strain, such an adjustment in protein abundance may reflect how selection for ethanol synthesis performance can affect protein abundance. However, to get a broader view of protein abundance specificities for each strain, we performed GSEA on the Log2(FC) value obtained for each gene in each one-vs-all comparison (Table S8). We found several signatures that recapitulated the environmental condition of the isolate. For example, BTT, a bioethanol strain, tended to overexpress the *lipid metabolic pathway* genes (Figure 2B), which is consistent with the central role of lipid adaptation and turn-over for ethanol tolerance in *S. cerevisiae* (Eardley and Timson, 2020; Ma and Liu, 2010; Vanegas et al., 2012). However, a large proportion of the enriched features in all the isolates were related to ATP metabolism and respiration (Table S8). The switch between respiration and fermentation being one of the signatures of *S. cerevisiae* domestication (Lahue et al., 2020), we checked if this signature was observable among the domesticated (namely, BPL, BTT, CQC

and CPI). Interestingly, we found that domesticated isolates had different DEP pattern for the respiration related genes (Figure 2C). If BPL and BTT showed a clear underexpression of respiration-related genes, CPI showed no particular enrichment and CQC even showed an overexpression of many respiration-related genes. This could be consistent with the fact that, unlike wine or bioethanol production (for BPL and BTT), where yeasts are inoculated either by backsloping or addition of starter culture, cocoa fermentation in West Africa (the ecological origin of CQC) is a spontaneous fermentative process (De Vuyst et al., 2023; De Vuyst and Weckx, 2016; De Vuyst and Leroy, 2020; Díaz-Muñoz et al., 2022; Fernández Maura et al., 2016; Leroy and De Vuyst, 2004). This likely explains why CQC has not shifted its metabolism contrarily to BTT or BPL and still primarily bases its energy production on respiration rather than fermentation. Taken together, our results show that protein abundance signature measured as DEP is a marker of environmental adaptation, but also of evolutionary history.

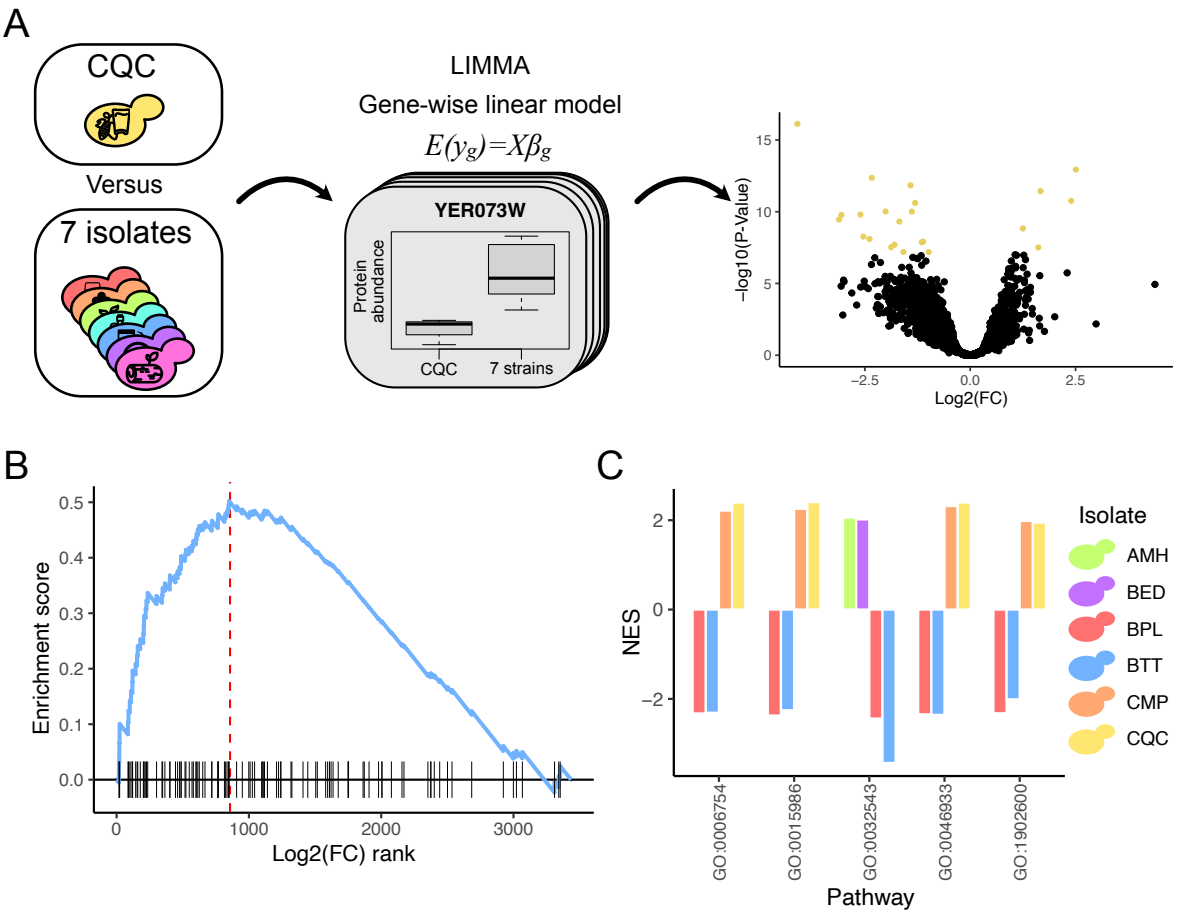


Figure 2: DEP survey reveals specific signature related to environmental adaptation.

(A) For each isolate (here CQC is used as an example), we detected DEP by applying a gene-wise linear model using the R package LIMMA. This allowed a one-against-all strain comparison to accurately detect isolate-specific over- or under-expressed genes. The p-value and Log2(FC) are

calculated and generated by LIMMA. (B) The genes annotated as *lipid metabolic pathway* in the GO annotation tend to be enriched in the over-expressed genes in the BTT isolate. The blue line corresponds to the enrichment of *lipid metabolic pathway* genes along the Log2(FC) ranks. The red line represents the maximum enrichment score (as reported by the R package *fgsea*). The vertical segments represent the position of each *lipid metabolic pathway* gene along the Log2(FC). (C) Normalized enrichment scores (NES, high values correspond to enrichment among over-expressed proteins, while low values correspond to enrichment among under-expressed proteins) of significant respiration-related GO annotations (from left to right: *ATP biosynthetic process*; *proton motive force-driven ATP synthesis*; *mitochondrial translation*; *proton-transporting ATP synthase activity, rotational mechanism*; *proton transmembrane transport*. See Methods) for each strain.

Transcriptional variation across the isolate is buffered across the gene expression

Interestingly, the proteome variations (mostly related to metabolism and respiration genes) seem at first highly similar to what was observed when looking at transcriptome and translatome variations (Teyssonniere et al., Submitted). We sought to compare more precisely the variations across the expression layers. We explored the inter-strain expression variations at each layer using two main approaches. First, we looked at the 28 pairwise correlations (corresponding to the correlations of each isolate against another) and observed a significant increase similarity between the expression profiles as long as the expression process goes on (median $\rho_{\text{transcriptome}}$ < median $\rho_{\text{Ribo-seq}}$ < median ρ_{proteome}, figure 3A). This was also observed using the non-normalized abundance (Figure S10A). Consistently, we quantified the variations level using an absolute log2 transformed fold change for each gene in each isolate pairwise comparison (|log2(FC)|, see methods). Briefly, the more this value increases, the more variable is the expression of a gene between two isolates. We found that the |log2(FC)| median value across the 28 pairwise comparisons were significantly decreasing as long as the expression process progresses (median $|\log2(FC)|_{\text{transcriptome}}$ > median $|\log2(FC)|_{\text{Ribo-seq}}$ > median $|\log2(FC)|_{\text{proteome}}$, figure 3B). Again, this was also observed using the non-normalized data (Figure S10B). The tendency of variation diminution at each step was also observed using Euclidean distances and gene-wise variance (Figure S10C, D). Taken together, these findings imply that gene expression is more constrained and therefore more conserved at the later steps of the process. This is in line with a phenomenon called post-transcriptional buffering that has been observed using both proteomics and ribo-seq data (Artieri and Fraser, 2014; Blevins et al., 2019; Dephoure et al., 2014; Gonçalves et al., 2017; McManus et al., 2014; Wang et al., 2020). Our data confirms that

this phenomenon takes place in each post-transcriptional step and therefore, different mechanisms are certainly involved in its establishment.

Interestingly, our results suggest that post-transcriptional buffering affects the proteome more than the translatome. Overall, we observed that proteome variation was consistently lower than translatome variation. As ribosome profiling has been considered as a proxy for protein abundance (Brar and Weissman, 2015), we sought to question the similarities between the translatome and the proteome and the ability of ribosome profiling to reflect protein abundance. In this context, we explore the correlations between each expression layer. To explore the relationship between the proteome and the other gene expression layers, two types of correlations are typically calculated: the across-gene correlation and the within-gene correlation (Buccitelli and Selbach, 2020; Liu et al., 2016). While the across-gene correlation explores the relationship between expression levels in a sample (here, an isolate), the within-gene correlation compares how each expression level varies in each gene and provides a better view of the similarities in variation between the transcriptome, translatome, and proteome. Computed with the non-normalized data, the across-gene correlation revealed that the mRNA-protein correlations were on average reaching 0.53 (Spearman correlation tests, Figure 3C) which is in line with previous explorations (Buccitelli and Selbach, 2020). Surprisingly, the translatome-proteome correlation was only a little higher than this: 0.59 (Figure 3C). The difference was nonetheless significant (paired Wilcoxon test p-value = 0.0078). Yet, this is greatly lower than the transcriptome-translatome correlation that reached 0.83. This suggests that although ribosome profiling is a better proxy for protein abundance than RNA sequencing, it is very limited as a predictor of protein abundance within an individual. Using the normalized data, this was even more apparent, as no difference was observed between the mRNA-protein correlation and the translatome proteome correlation (Figure S11A). For the within-gene correlation, even though the comparisons include 8 samples for each gene, the large number of genes compensates for this and allows us to compare variation differences between expression levels. Here, we found no difference between the predictability of protein abundance across the 8 isolates using either transcriptome or translatome data (Figure 3D) The transcriptome-proteome correlation and the translatome-proteome correlation had an average within-gene correlation of 0.11 and 0.11, respectively, while the transcriptome-translatome correlation reached 0.52. The results were identical when using the normalized data (Figure S11B). This highlights that ribosome profiling data may not be a reliable proxy for protein abundance across

individuals. Furthermore, it suggests that the constraints shaping proteome variation might be different from those shaping transcriptome and proteome variation.

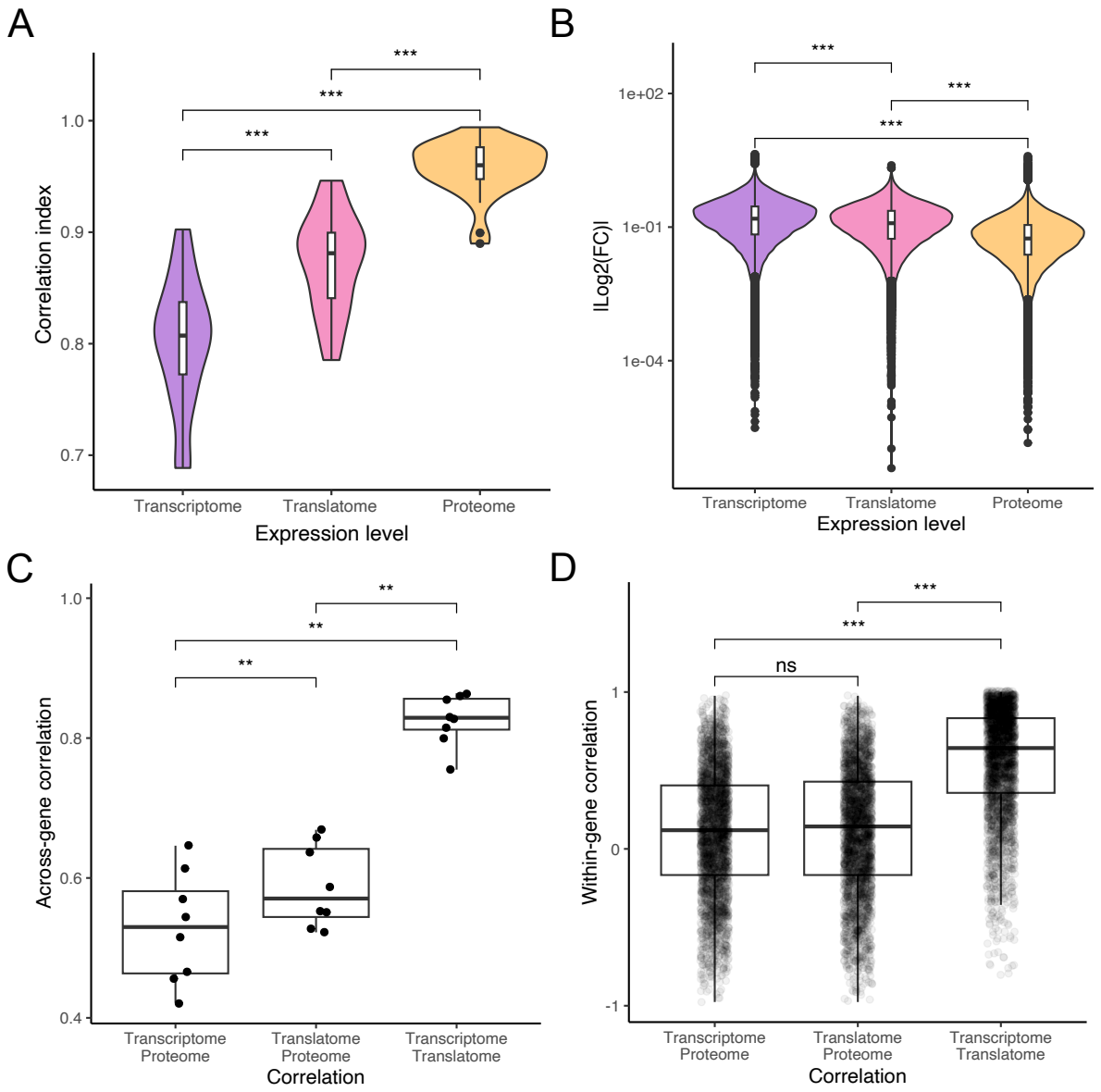


Figure 3: Gene expression variations across the gene expression levels.

(A) Pairwise correlations (n=28) of the 8 isolates in each dataset show that the proteomic profiles are more similar than those obtained on the transcriptome and translatome. All p-values were obtained from Wilcoxon tests and were lower than $1x10^{-5}$. (B) Comparison of |log2(FC)| (n= 79,520) obtained in each dataset revealed that transcriptional variations are buffered during the gene expression process. All p-values were obtained from Wilcoxon tests and were lower than $1x10^{-20}$. (C) The across-gene correlation of translatome vs. proteome is only slightly higher than the cross-gene correlation of transcriptome vs. proteome. All p-values were obtained from paired Wilcoxon tests and were all equal to 0.0078. (D) The within-gene correlation of the translatome vs. proteome is not higher than the

across-gene correlation of the transcriptome vs. proteome. All p-values were obtained by paired Wilcoxon tests. The *** correspond to p-values lower than $1x10^{-20}$.

Gene expression evolution is gene specific

We sought to explore the determinants of gene expression evolution at each stage of the expression process. We used all categorical features from the *Yeastomics* dataset, which collects 3,685 gene characteristics from 27 studies (see full list and citations on https://github.com/benjamin-elusers/yeastomics). These characteristics include many gene and protein features related to *e.g.,* chromosomal location, cellular function, interaction capabilities. The overlap between this dataset and our expression dataset reached 2,308 genes. We selected the characteristics that affected at least 10 genes, resulting in a list of 793 features (Table S6). To test whether each of the characteristics had an impact on the evolution of gene expression, we adapted a method previously published on mammalian expression data (Wang et al., 2020). Briefly, this method relies on the construction of Euclidean distance trees based on gene expression across the 8 isolates (Figure 4A). The total length of the resulting tree serves as a measure of gene expression evolution. Using all the normalized expression of all genes, we observed that the resulting trees had different size: the proteomic based tree was the shortest while the transcriptomic based tree was the longest (Figure 4A). This is in line with our aforementioned observations on post-transcriptional buffering.

For each category, we then compared the length of the tree resulting from the included genes to a randomly generated length (see Methods), as the length of the tree is strongly correlated with the number of genes used for its computation (Figure 4B, Figure S12). For example, the category *cat_genomics.sgd.chr_A* (genes located on chromosome 1) encompassed 30 genes. Using normalized abundance, the resulting tree has a total branch length of 21.72 (arbitrary unit), which is not significantly different from the lengths obtained with randomly generated trees (using 30 genes) (Figure 4C). The ratio between the median random tree lengths and the computed tree length are used as a measure of gene expression evolution. We tested the significance of the difference between the length and using a corrected (FDR) p-value threshold of 0.001, we detected 59 features influencing gene expression variation in at least one step of gene expression (mRNA abundance, translation, protein abundance) (Table S7), revealing that gene expression evolution is unequal across the genes. Several categories were based on previous expression variation exploration (Lahtvee et al., 2017) and were associated

respectively with high and low expression variation in our dataset, supporting the reliability of the dataset and the tree-based exploration of expression evolution constraints.

We found that the overlap between the expression layer was overall small. Only 3 features influenced both mRNA and protein expression evolution. Similarly, 4 features influenced both translation and protein expression evolution while 22 features influenced both transcription and translation regulation evolution. At first glance, this suggests that the gene expression evolutionary constraints are layer specific, even if mRNA and RPF abundance seem to face similar constraints. However, we observed several trends that were conserved across each expression step. For instance, we found that the categories related to metabolism (Figure 4D) were associated with faster gene expression evolution compared to the rest of the genes. Consistently, in each dataset, the category associated with the fastest gene expression evolution across our 8 isolates was always related to metabolism (Table S7). This is in line with the DEP detected previously and other exploration in yeast highlighting that metabolism genes are usually among the most variable genes across individuals (Caudal et al., 2023). This important plasticity in gene expression is most likely a mechanism allowing for an optimal adaptation to different trophic specificities across our 8 isolates and reflects the environmental differences between the ecosystem in which each strain naturally occurs. Inversely, we found that several categories were associated with strong evolutionary constraints. For example, the feature associated with protein interactions resulted in the construction of short expression trees in each layer (Figure 4E). Accordingly, the most conserved categories were all related to interaction features. More generally, we found that central and essential cell functions were associated with evolutionary constraints. For instance, genes annotated as the core essential group were associated with constrained protein abundance evolution, while cytoplasmic translation related genes displayed a more conserved translation regulation.

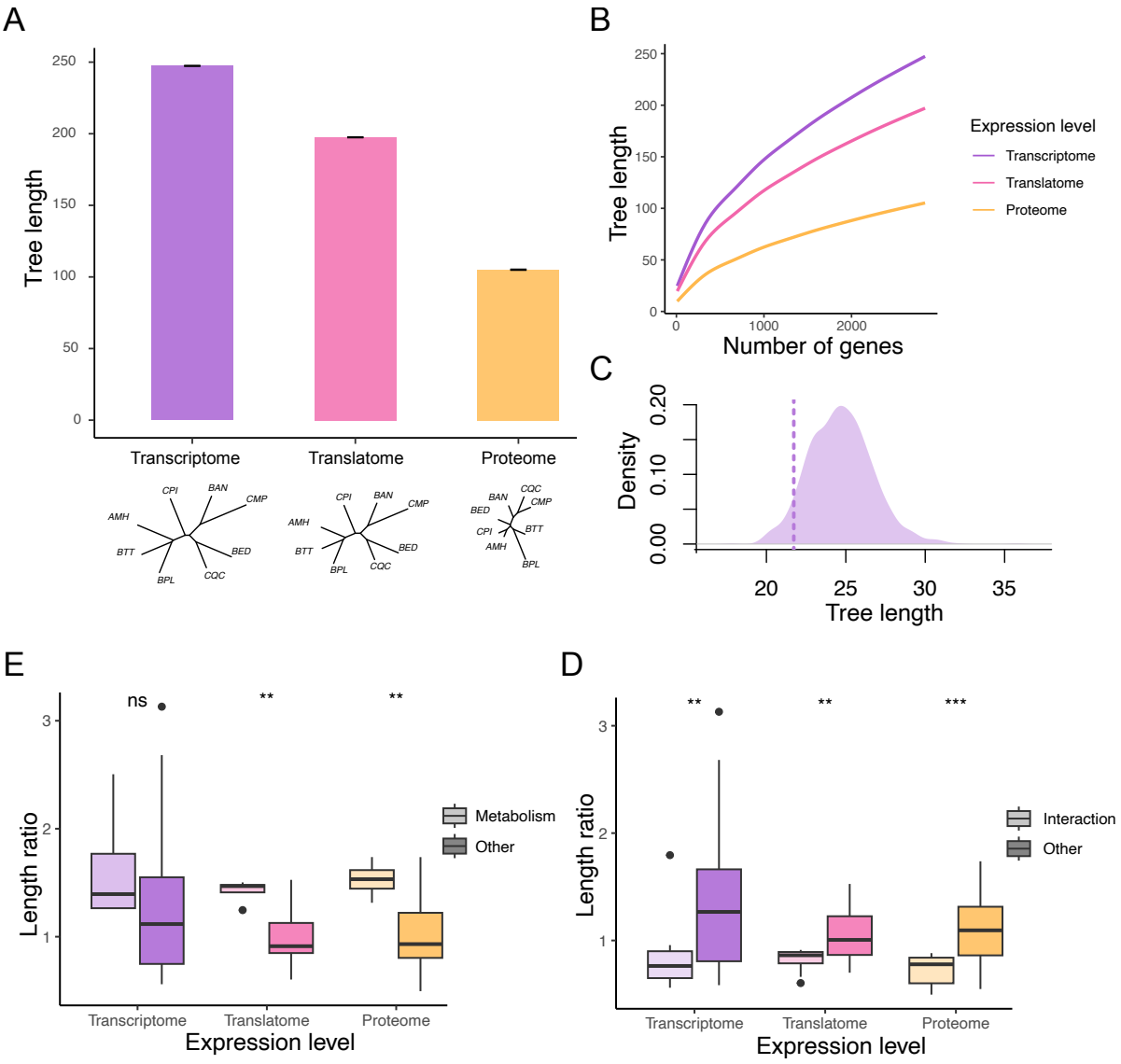


Figure 4: Exploration of evolutionary constraints on gene expression in *S. cerevisiae*.

(A) Trees constructed from the gene expression-based Euclidean distances and their respective lengths. Consistent with post-transcriptional buffering, the proteomic tree is smaller than the others. The error bars correspond to the bootstrapping (n=1000) performed on each tree. (B) The number of genes is a strong determinant of the length of the trees. The lines are constructed by smoothing the 1000 lengths obtained for each number of genes in each expression layer (see full plot in figure S7). (C) For each gene feature (here *cat_genomics.sgd.chr_A*, i.e., the genes located on chromosome I) in each expression level (here RNA-seq), the length of the tree resulting from the selected gene is calculated (the dashed purple line). This length is then compared to the 1000 tree lengths generated using exactly the same number of genes (here n=30) if it is significantly shorter or longer. (D) The categories related to metabolism are associated with faster gene expression evolution in the translatome and proteome. The p-values are all computed with Wilcoxon test and those annotated

with ** are below 0.01. (E) The categories related to interactions are often associated with stronger constraints on the evolution of gene expression. The p-values are all calculated using the Wilcoxon test and are all below 0.01.

Discussion

Gene expression is one of the major determinants of phenotypic variation observed between individuals. It is a complex phenomenon in which each step is tightly regulated. Here, we deeply quantified the proteome of 8 natural isolates of *S. cerevisiae* and combined these data with previously generated transcriptomes and translatomes on the same set of isolates (Teyssonniere et al., Submitted), allowing for a precise exploration of the expression variation throughout the gene expression process.

Our results highlight that protein abundance variation tended to be mainly related to metabolic pathways. Furthermore, we observed that protein abundance signatures of each isolate (detected as DEPs) often included genes related to respiration, which is a known marker of metabolic adaptation in the domestication of *S. cerevisiae*. These signatures adequately matched the wild or domesticated origin of the strain, and even highlighted differences within domesticated isolates related to their different uses. These observations are consistent with previous transcriptome and translatome surveys (Caudal et al., 2023; Teyssonniere et al., Submitted).

When comparing the variation between different expression levels, we observed that the expression variation between isolates tends to be buffered as the gene expression process progresses. This phenomenon, known as post-transcriptional buffering (Artieri and Fraser, 2014; Blevins et al., 2019; Dephoure et al., 2014; Gonçalves et al., 2017; McManus et al., 2014; Wang et al., 2020), highlights that the constraints on gene expression evolution are greater in the final stages of the expression process, which is consistent with the fact that proteins are the drivers of cellular machinery and functions. Abnormal changes in protein abundance can therefore be highly deleterious, and post-transcriptional buffering is often considered as a coping mechanism for faulty expression regulation (Buccitelli and Selbach, 2020; Liu et al., 2016). Our results show that this phenomenon is multilayered and its effect increases with the course of gene expression. Interestingly, we observed that although ribosome profiling is known to be a proxy for protein abundance (Brar and Weissman, 2015), the translatome was only slightly better at reflecting protein abundance within isolates than the transcriptome. When looking at the similarity between transcriptome, translatome and proteome variation across the expression layer in a gene-wise manner, we found that both translatome and transcriptome poorly reflected the proteome variation observed across isolates.

Finally, we focused on exploring the constraints that shape gene expression evolution. By adapting an analysis method previously developed (Wang et al., 2020), we could detect several gene characteristics that seemed to be involved in either fast or slow gene expression evolution.

Consistently with our previous findings on gene expression variation, the genes involved in different metabolism pathways were associated with fast gene expression evolution. As the 8 isolates came from different environments with very different conditions in terms of nutrient and resource availability, this highlights that trophic constraints play an important role in shaping gene expression. Conversely, genes that are involved in multiple protein interactions or that tend to play an essential role in cellular functions are likely to be associated with strong constraints on expression evolution. This is consistent with previous findings on the deleterious effects of unbalanced complex component (Deutschbauer et al., 2005; Morrill and Amon, 2019; Ohnuki and Ohya, 2018; Veitia and Potier, 2015) and with the highly stochiometric expression of gene involved in protein complexes (Chotewutmontri and Barkan, 2016; Jüschke et al., 2013; Li et al., 2014; Lukoszek et al., 2016; Taggart and Li, 2018; Trösch et al., 2018).

Overall, our study highlights that, in yeast, differences in gene expression between individuals, and thus the evolution of gene expression, are tightly linked to both environmental constraints and constraints on gene function. Furthermore, these constraints tend to differ across levels of gene expression, with protein abundance tending to be the most constrained level.

Materials and methods

Sample preparation for proteomics profiling

In this study, we conducted a comprehensive proteomics profiling experiment using a subset of eight *S. cerevisiae* strains representing the diverse range of ecological, geographical, and genetic characteristics from a population of 1,011 natural isolates. The selected strains were cultured on synthetic defined media (SD), and their growth was closely monitored by measuring the optical density (OD). We specifically harvested cells when they reached the mid-log phase (OD ~0.5). Prior to sample processing, we performed two rounds of washing using Phosphate-Buffered Saline solution (PBS) to remove extraneous contaminants. The cell pellets were then flash-frozen in liquid nitrogen. In total, we prepared 32 samples, with OD values ranging from 0.4 to 0.8 units. Four replicates were prepared for each strain, consisting of two biological replicates derived from distinct colonies and two technical replicates that underwent identical sample preparation, only separating them before freezing them. Subsequently, all samples were sent to the proteomics facility for further analysis. The frozen cell pellets were then lysed and submitted to in-solution tryptic digestion using the S-Trap method (by Protifi). A solid phase extraction cleaning step using Oasis HLB was employed to purify the resulting peptides. The purified peptides were then subjected to nanoflow liquid chromatography (nanoAcquity) coupled with high-resolution, high-mass accuracy mass spectrometry (Thermo Exploris 480).

Proteomics identification and database searching

For data analysis, each sample was analyzed separately on the mass spectrometer in a randomized order during the discovery mode. The raw data acquired from the instrument were processed using MaxQuant v1.6.6.0. The Andromeda search engine was utilized to search the data against a database comprising protein sequences of Saccharomyces cerevisiae obtained from Uniprot.org. This database was supplemented with common lab protein contaminants. During the search, we considered fixed modifications such as cysteine carbamidomethylation and variable modifications of methionine oxidation and/or protein N-terminal acetylation. Quantitative comparisons were performed using Perseus v1.6.0.7. Decoy hits were filtered out, and only proteins detected in at least two replicates of at least one experimental group were retained for further analysis. This rigorous methodology ensured high-quality data for subsequent interpretation and downstream analysis. Ultimately, our proteomic dataset comprised 3,429 genes. We performed a GO analysis on this set of genes using the R package

gprofiler2 (Kolberg et al., 2020; Raudvere et al., 2019) to functionally characterize the genes encompassed in our dataset. We used semantic similarity to reduce the number of detected GO annotations (biological process) with the *rrvgo* R package (Sayols, 2022).

Quantitative analysis of proteomes and differentially expressed proteins

The quantitative analysis of detected proteins was performed in R. First, we applied a log-transformation on expression intensities with a normalization of the data by the median of each sample to obtain relative protein expression within and across samples. To minimize information loss, we imputed missing protein expression values using the Bayesian Principal Component Analysis algorithm (Oba et al., 2003) as long as the protein was detected in 2 replicates of the same strain and among at least 4 strains. We computed the CV for each protein across the 8 strain and performed a GSEA using the R package FGSEA (REF) and with the GO annotation to explore which genes tended to display expression variability.

The LIMMA package (Ritchie et al., 2015; Smyth, 2005) was used for identifying differentially regulated proteins, using the average across all four replicates, since the variance within strains was found to be much lower than between samples (cf Heatmap). We corrected the statistical significance for multiple testing of differential expression using the False-Discovery Rate procedure (Benjamini and Hochberg, 1995). Then, a protein was considered differentially expressed if the fold change was higher than 1.2 and the adjusted p-value was lower than 10-5, as reported by LIMMA. For each isolate, we used the log2 transformed fold change of each gene to compute GSEA to detect strain-specific over or under expressed cellular pathway.

Expression variation exploration and comparison

We sought to explore gene expression variation all along the gene expression process. We combined our proteomic data with data previously generated on the same set of strains and under the same culture conditions (Teyssonniere et al., Submitted). This data, generated using RNA-seq and ribo-seq, comprises the precise measurement of 3,755 genes at the transcriptome and translatome level. The overlap between the two datasets covers 2,840 genes. We applied a LOESS normalization on the RNA-seq, ribo-seq and proteomic data together in order to strictly compare biological variations.

Using this dataset, we computed and compared the pairwise correlation levels of each isolate versus another using Spearman correlation test in each dataset. And we also computed the Euclidean distances between each isolate and computed the gene wise variance as well and

checked if the resulting values were different depending on the expression level. Additionally, we computed for each gene, and in each isolate pairwise comparison, the absolute value of the log2 transformed fold change (|log2(FC)|) between two isolates as follow:

$$|log2(FC)|_{gene\ X\ in\ comparison\ strain\ 1\ vs\ strain\ 2} = |log2(\frac{\text{normalized expression}_{\text{gene X strain 1}}}{\text{normalized expression}_{\text{gene X strain 2}}})|$$

For both pairwise correlations and absolute log2 fold change, the analyses were also performed on the non-normalized data.

Correlation between the expression levels

We sought to explore the proximity between Ribo-seq data and proteomic data. To do so, we computed two types of gene expression correlations: the across-gene correlation and the within-gene correlation (Buccitelli and Selbach, 2020; Liu et al., 2016). The across-gene correlation is computed for each isolate (so in our case, 8 correlations per expression level) and it is based on the comparison between the gene expression data of all genes (in our case, the correlation between 2,840 values) in each expression level correlation (i.e., transcriptome vs translatome, transcriptome vs proteome and translatome vs proteome). We computed the correlation using a Spearman correlation test. The within-gene correlation is computed for each gene (in our case, 2,840 correlation per expression level) and is based on the gene expression values across the 8 isolates. Again, this was computed using a Spearman correlation for each expression level correlation: transcriptome vs translatome, transcriptome vs proteome and translatome vs proteome.

Gene expression evolution constraints

We looked for the determinants of gene expression evolution. We did so by using the Yeastomics database available online (https://github.com/benjamin-elusers/yeastomics for the full list, its construction, and citations). This database encompasses 3,685 gene characteristics that are either numeric (ex: specific codon composition, gene length, variance level in previous studies…) or boolean (ex: is on chromosome A, part of specific GO annotation, essentiality…). We used the boolean characteristics for which at least 10 genes fulfilled the characteristics, which resulted in a list of 793 characteristics. The overlap between the gene characterized by this data and our 2,840 genes reached 2,308 genes. We used these characteristics to explore the constraints on gene expression evolution by adapting a tree-based approach previously published (Wang et al., 2020). This method relies on the construction of trees using the

Euclidean distances generated from the 3 gene expressions. The total length of the resulting tree branches is used as a measure of gene expression evolution. Using all genes, we observed a difference in tree length depending on the expression level (statistically confirmed by a 1,000-step bootstrapping) which was in line with the post-transcriptional buffering phenomenon and was a first control to support the reliability of this method. The tree length being highly dependent on the number of genes used for the tree construction, we computed for each number the lengths expected by chance for each characteristic by generating 1,000 trees constructed with randomly selected genes. For each gene characteristic, we selected the corresponding genes and constructed a tree for each gene expression level. We compared the resulting tree length with the 1,000 random lengths using a simple normal density probability test as the random lengths were normally distributed. The p-values obtained were FDR-corrected using and a threshold of 0,001 was considered to detect characteristics significantly associated with gene expression evolution. The ratio between the computed tree-length and the median random tree length was used to detect characteristics associated with expression evolutionary constraint (ratio <1) or with fast expression evolution (ratio>1).

Supplementary Material

Supplementary tables available at:

https://www.dropbox.com/scl/fi/dljfmouqew8z7gkeb3dag/S_table.xlsx?rlkey=2v1slof41uecvgl21t5sguxkx&dl=0

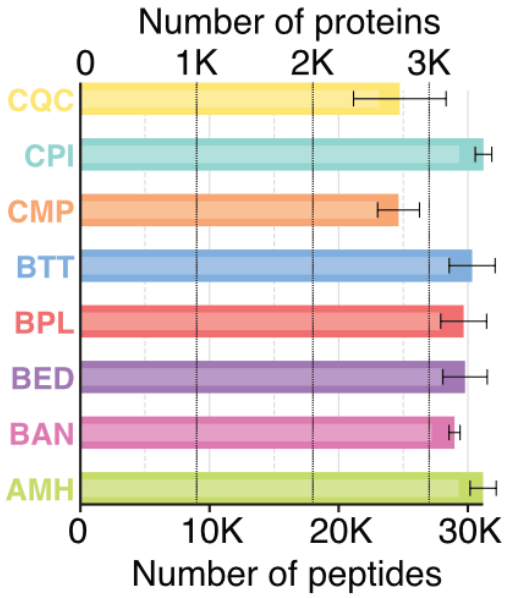


Figure S1: Number of quantified peptides and proteins for each isolate.
The length of the colored bars corresponds to the average number of detected peptides in each strain, with the standard deviation indicated as the T-shaped error bar in black. The fraction of unique peptides detected is shown as a narrow bar with dimmed colors within each horizontal bar. The top x-axis is scaled to indicate how many unique proteins are matched by any detected peptide.

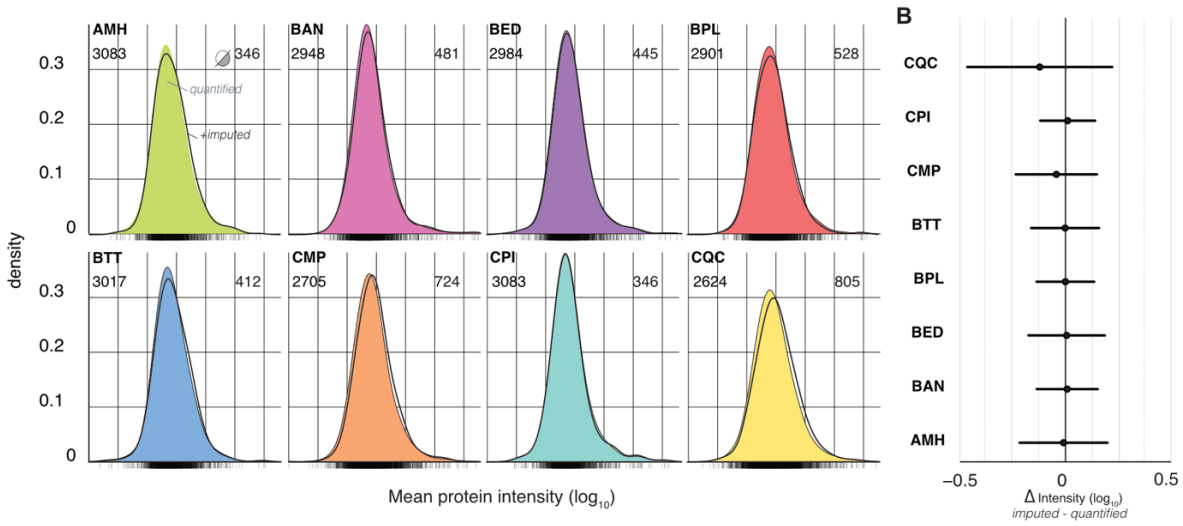


Figure S2: Quantification and imputation of protein expression across strains.
(A) The density area represents the distribution of protein expression determined from the average logarithmic intensity (log_{10}) of detected peptides for each strain. A compact representation is also shown as a one-dimensional marginal distribution horizontally below each plot. The number of proteins quantified in each strain is written on the top left corner. A black density line indicates the distribution of protein expression after the imputation of missing values for proteins that were partially undetected among samples. The number of proteins partially detected is indicated in the top right corner. (B) The position of the circles relative to the x-axis is given by the average difference between imputed and quantified protein expression ($\Delta_{intensity}$) in each strain. For each strain, two arms extend in opposite directions to highlight the range of $\Delta_{intensity}$ within one standard deviation from the mean represented by the circles.

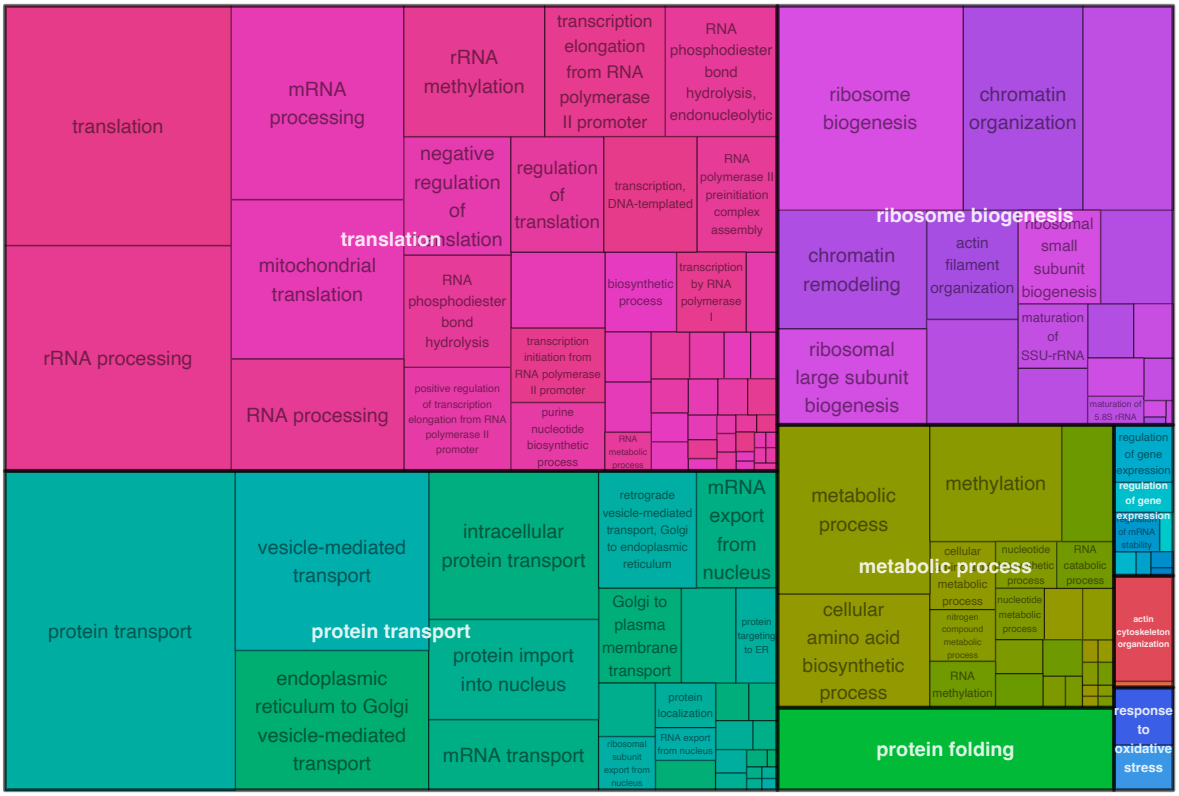


Figure S3: GO enrichment analysis of the encompassed proteins.
Graphical representation of the GO enrichment analysis on the included genes from the proteomic data. The GO categories (in white) were obtained using semantic similarities on the terms detected using the *gprofiler2* package in R. The semantic similarity was performed using the *rrvgo* package in R.

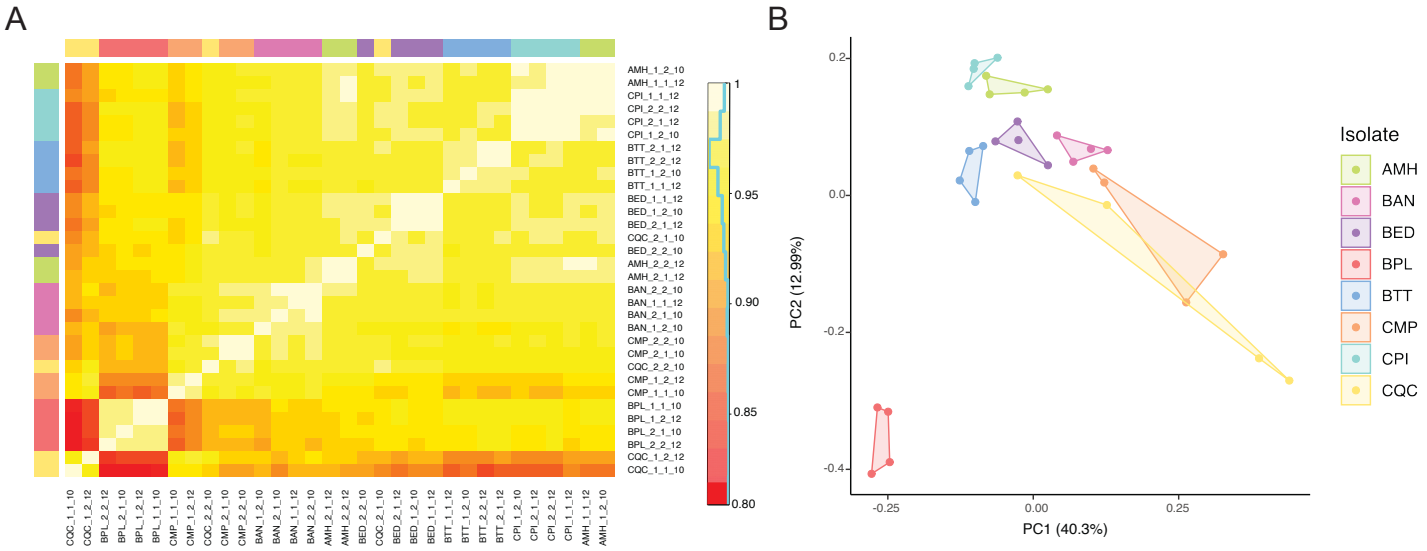


Figure S4: Inter and intra protein abundance variation.

(A) Heatmap of the Pearson correlation coefficients for pairwise comparisons of protein expression between proteomics samples. A hierarchical clustering based on the complete Euclidean distances between samples' expression profiles was applied to both rows and columns. The colors represent each isolate as shown in figure 1. (B) Principal component analysis of samples expression profile. The expression profiles of all samples are plotted as points scatterplot against the first (x-axis) and second (y-axis) principal components, which capture 52% of the variability between them. Points are colored according to their matching strains. The shaded regions delimit the range of variability among samples of the same strain.

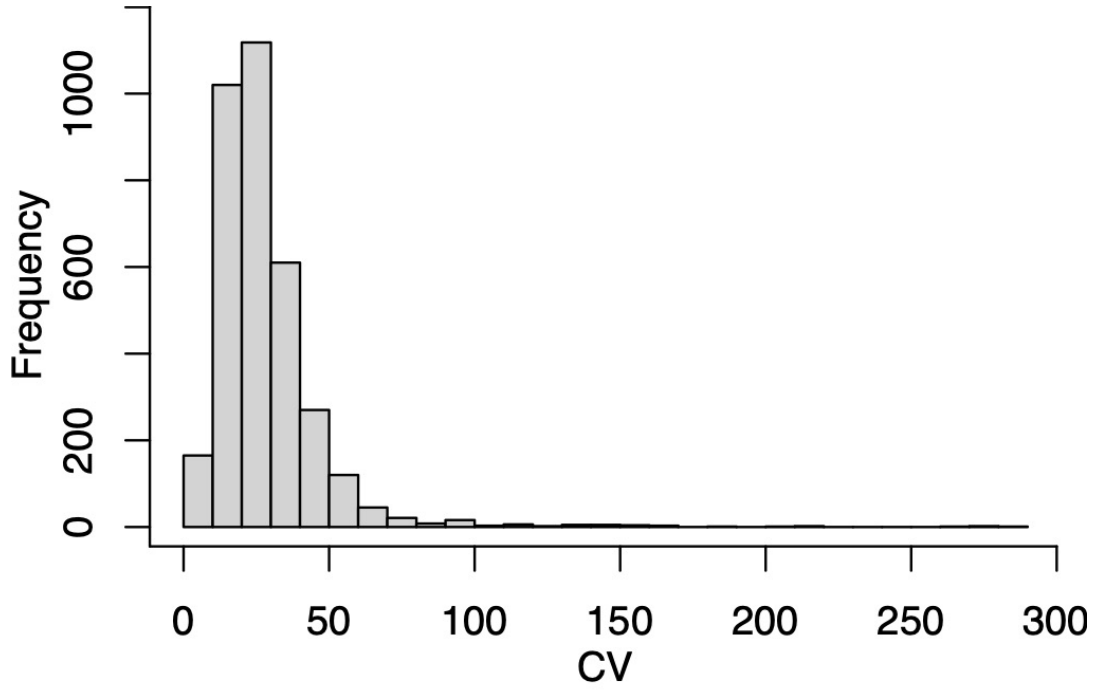


Figure S5: Distribution of the CVs across from each quantified protein.
Histogram of the CV calculated on the median protein abundance of each of the 3,429 genes quantified in the proteomic data.

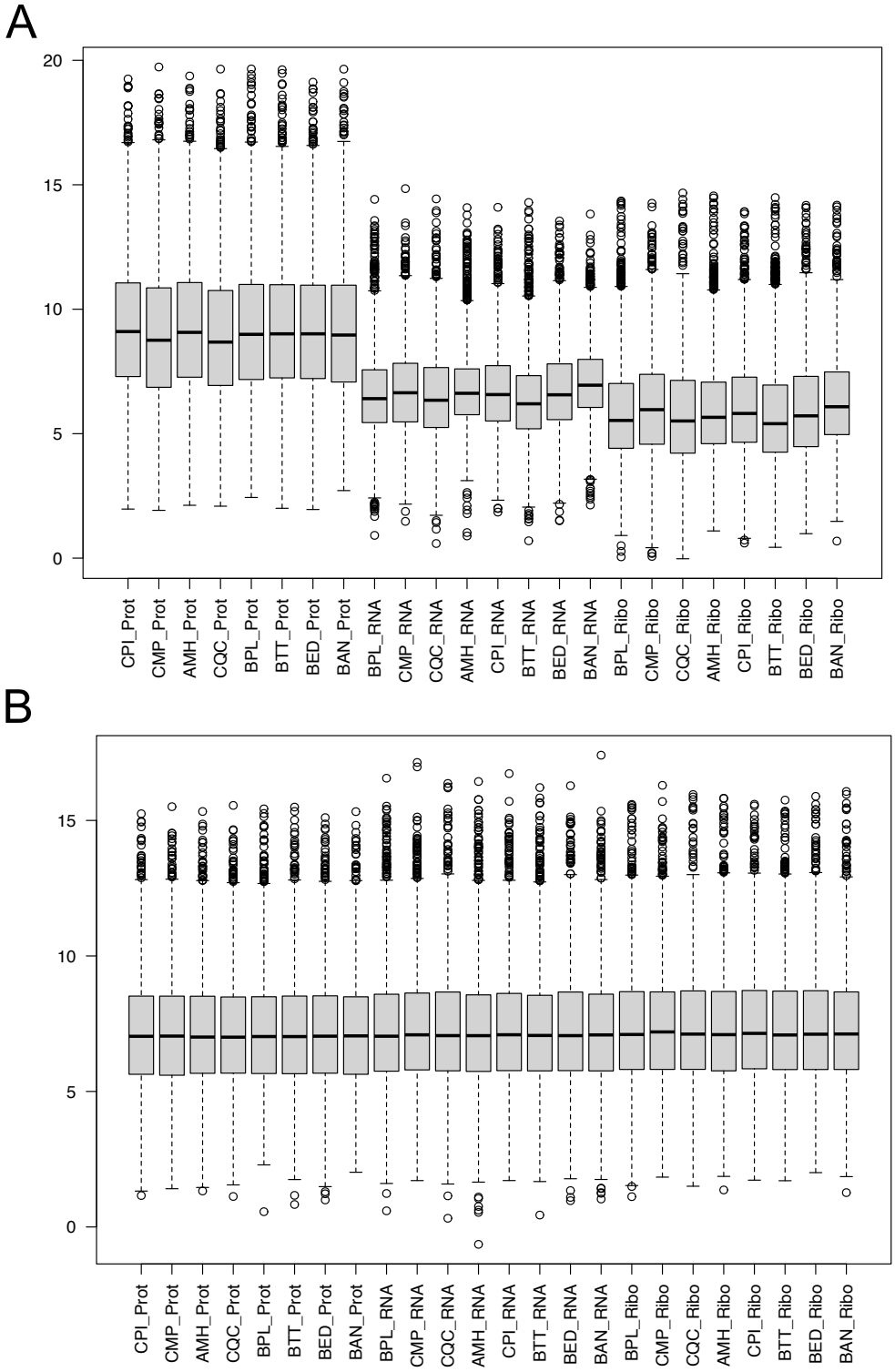


Figure S6: LOESS normalization of the 3 datasets.

(A) Overview of the abundance values across the 8 isolates in the 3 data before normalization. (B) Overview of the abundance values across the 8 isolates in the 3 data after normalization.

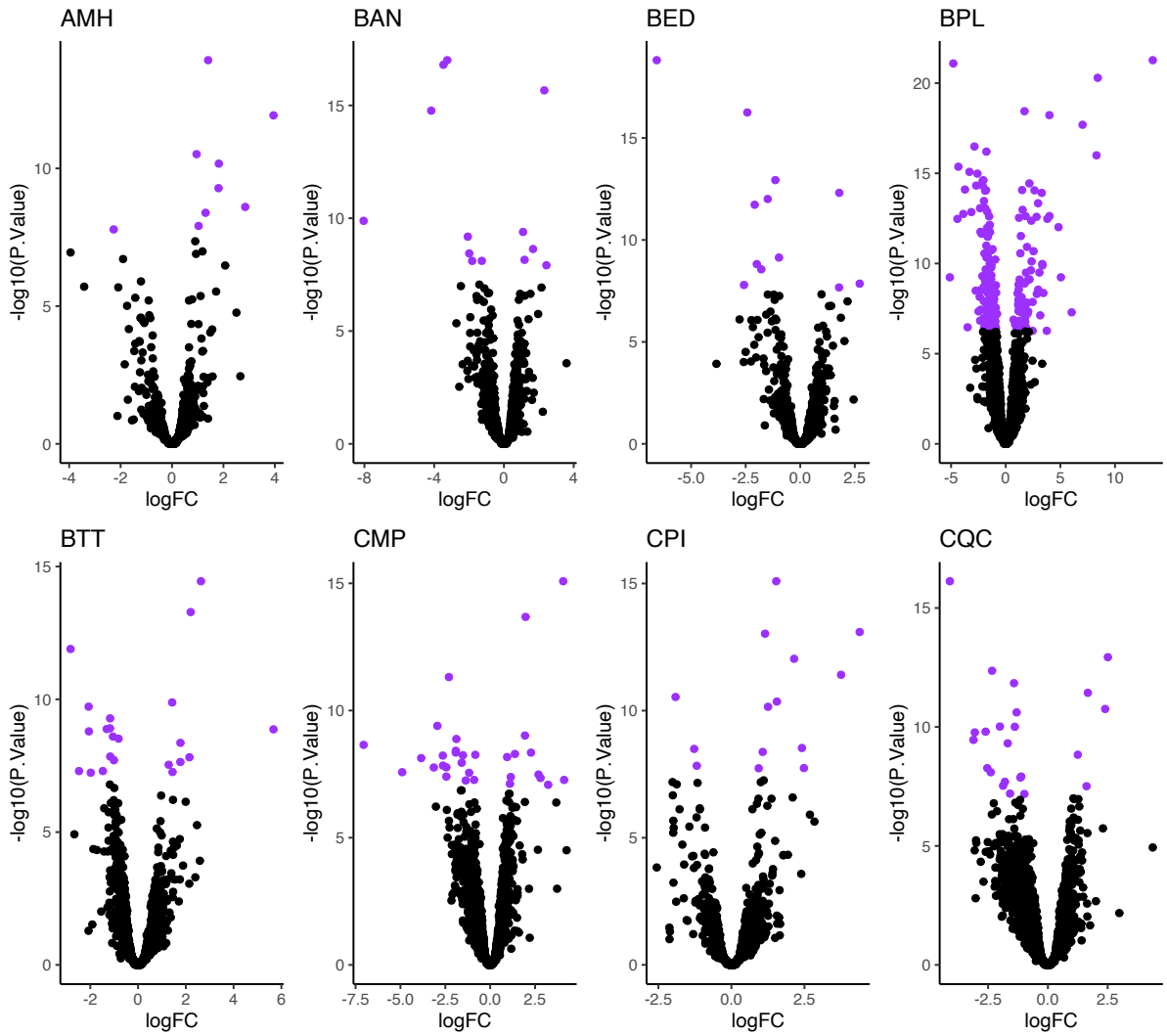


Figure S7: DEP detection in each isolate.

Volcano plot for each isolate. The points highlighted in purple are the DEP. The detection was performed using the LIMMA R package.

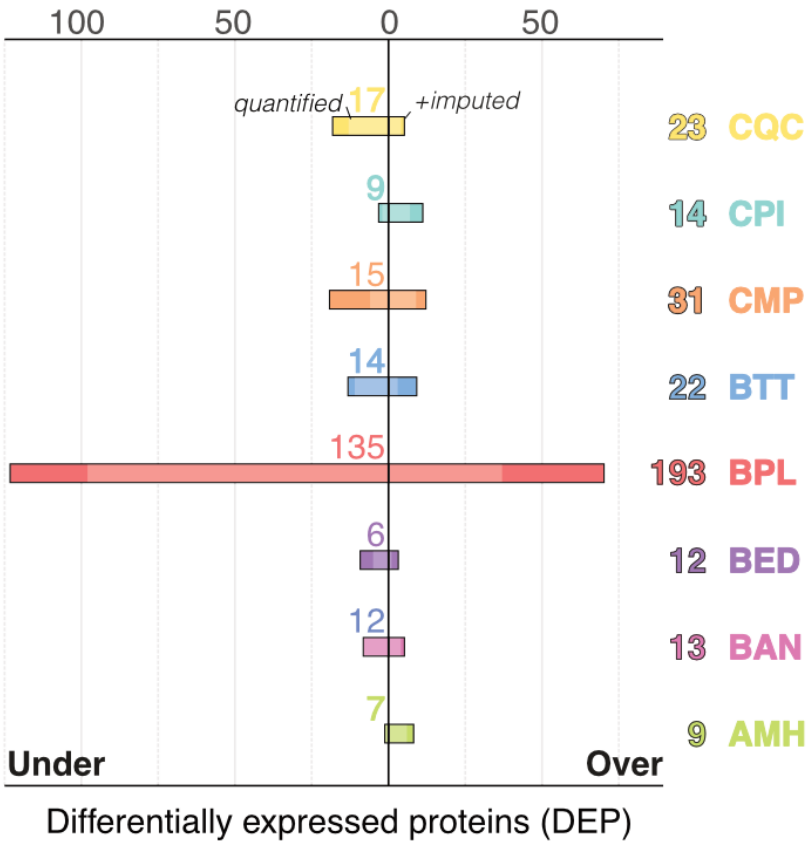


Figure S8: Number of differentially expressed proteins for each isolate.

Differential protein expression was calculated using LIMMA (see Methods) comparing the average values across samples for one strain *versus* the average across samples for the remaining strains (one-vs-all) for every protein. We consider the decrease/increase of protein expression significant when the absolute fold-change reached at least 1.2 and if the adjusted p-value was greater than 10^{-5}, at a false discovery rate of 5%. The number of differentially expressed proteins (DEP) is reported along the x-axis, with underexpressed and overexpressed proteins respectively shown as bars in the left and right directions. The total number of differentially expressed proteins for each strain is written on the right. The number of differentially expressed proteins using only quantified protein expression (*i.e.* non-imputed) is also shown in lighter colors and represented by dimmed colored bars.

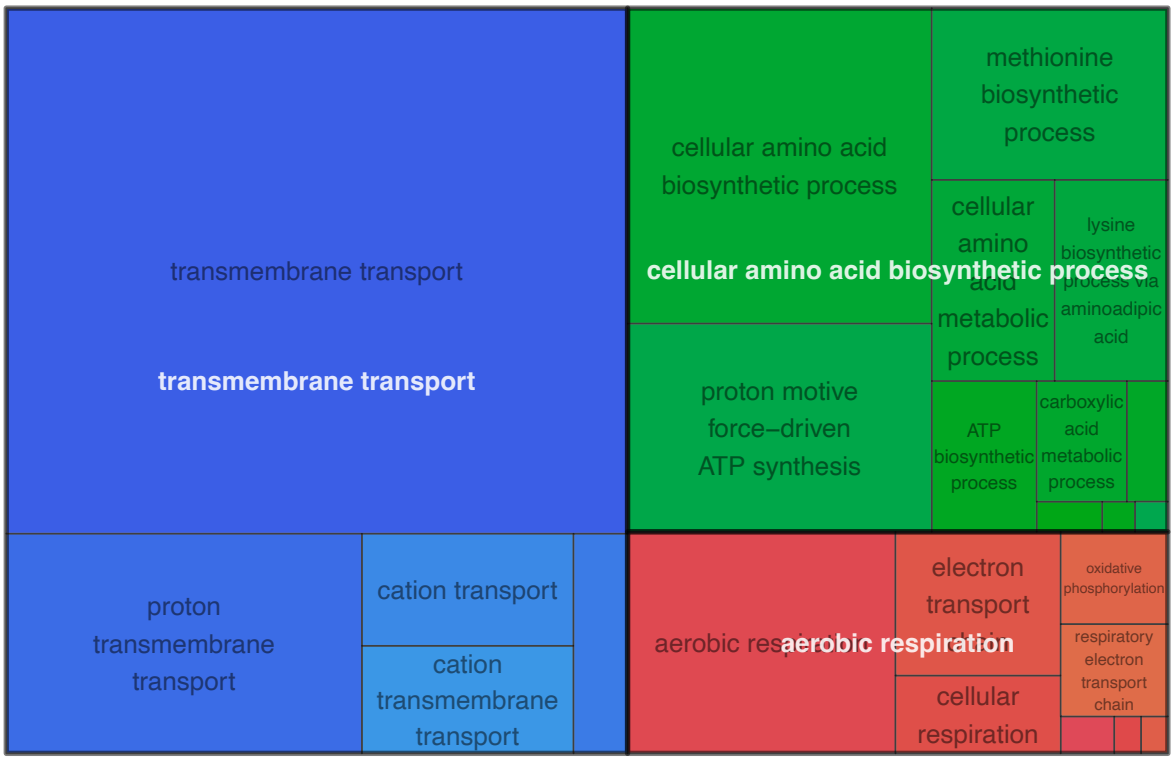


Figure S9: Overview of the BP GO features enriched in the DEP

Graphical representation of the GO enrichment analysis on the DEP. The main GO categories (in white) were obtained using semantic similarities on the terms detected using the *gprofiler2* package in R. The semantic similarity was performed using the *rrvgo* package in R.

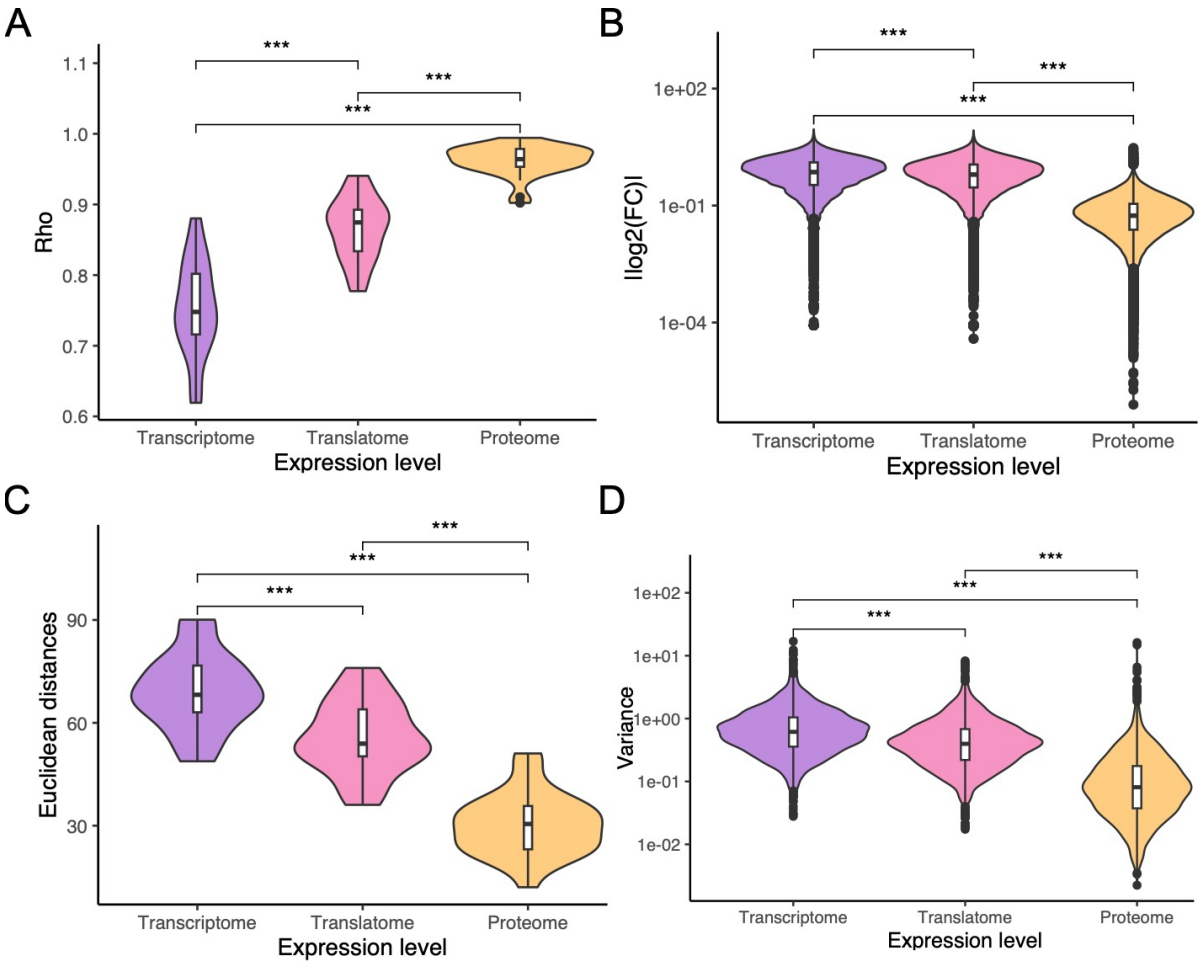


Figure S10: Transcriptional variation is buffered as long as the expression process progresses. (A) Pairwise correlations (n=28) of the 8 isolates in each dataset (using the non-normalized abundance) show that the proteomic profiles are more similar than those obtained on the transcriptome and translatome. All p-values were obtained from Wilcoxon tests and were less than $1x10^{-5}$. (B) Comparison of |log2(FC)| (n= 79,520) obtained in each dataset revealed that transcriptional variations are buffered during the gene expression process (using the non-normalized abundance). All p-values were obtained from Wilcoxon tests and were less than $1x10^{-20}$. (C) The comparison of Euclidean distances (n=28) between each isolate obtained in each dataset are in line with the post-transcriptional buffering phenomenon. All p-values were obtained from Wilcoxon tests and were less than $5x10^{-5}$. (D) Gene-wise variance across all the expressions level also support the presence of post-transcriptional buffering. All p-values were obtained from Wilcoxon tests and were less than $1x10^{-20}$.

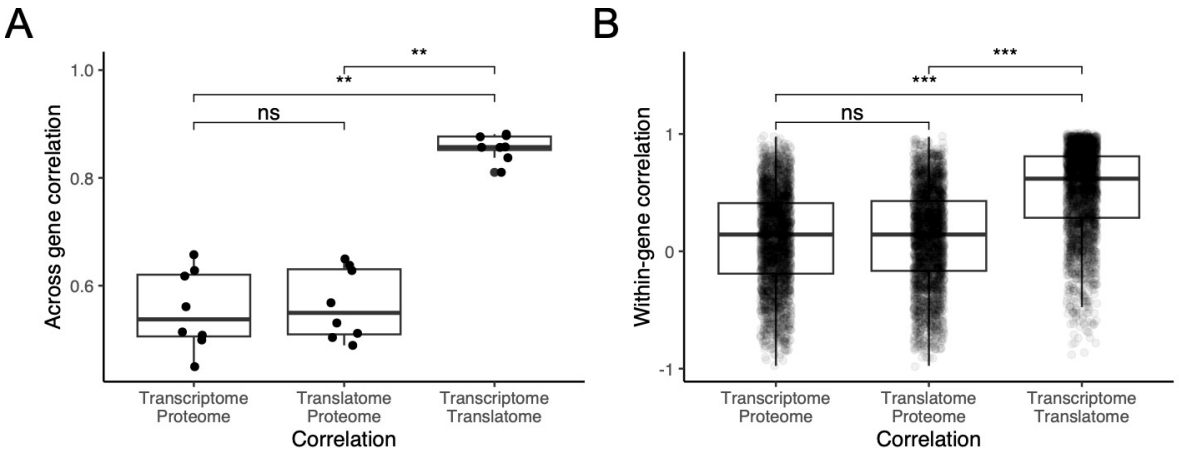


Figure S11: Across and within gene correlation using the normalized data.

(A) The across-gene correlation of translatome vs. proteome is only slightly higher than the cross-gene correlation of transcriptome vs. proteome. The p-values were obtained from paired Wilcoxon tests and the ** correspond to p-values equal to 0.0078. (B) The within-gene correlation of the translatome vs. proteome is not higher than the across-gene correlation of the transcriptome vs. proteome. All p-values were obtained by paired Wilcoxon test. The *** correspond to p-values less than $1x10^{-20}$.

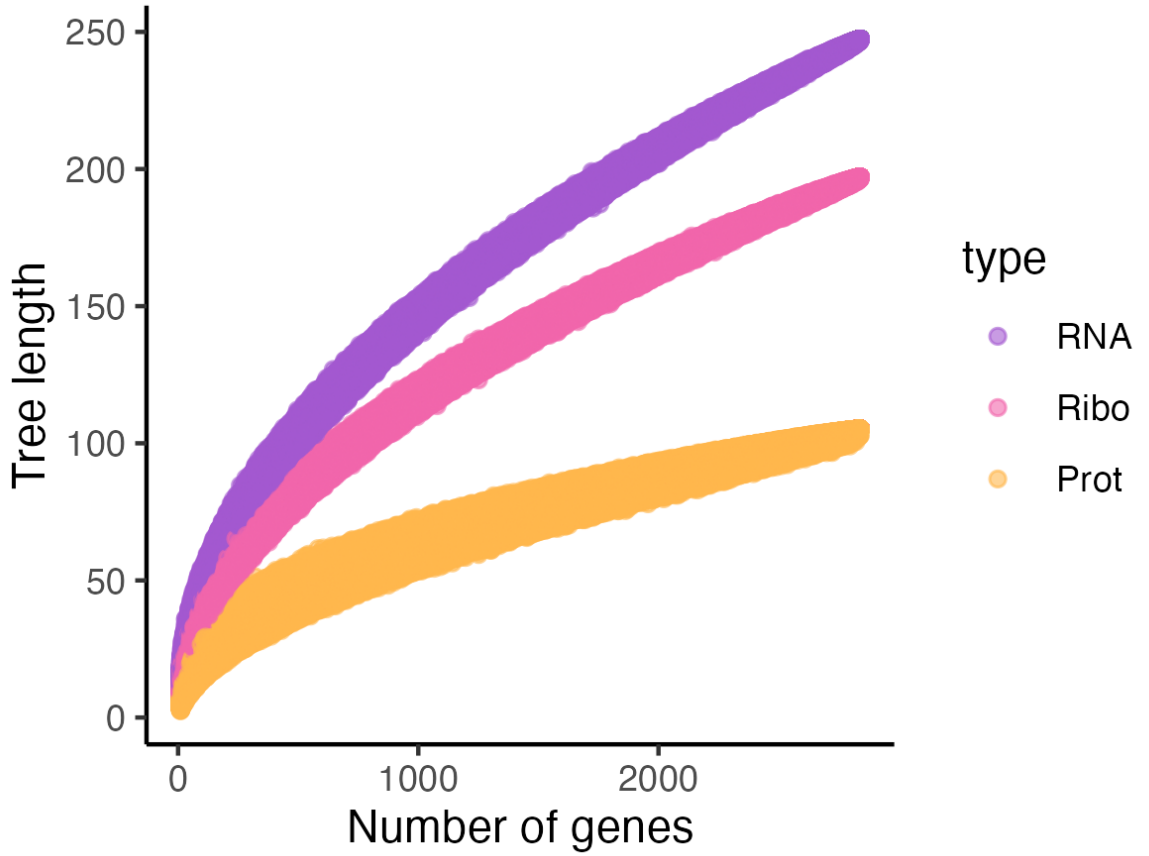


Figure S12: Effect of the gene number on the tree length in each dataset.
For each gene number, 1,000 trees were constructed by randomly selecting the corresponding number of genes and the total length was computed.

Bibliography

Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212. https://doi.org/10.1038/nrg3891

Artieri, C.G., Fraser, H.B., 2014. Evolution at two levels of gene expression in yeast. Genome Res. 24, 411–421. https://doi.org/10.1101/gr.165522.113

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. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556

Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., Gilad, Y., 2015. Impact of regulatory variation from RNA to protein. Science 347, 664–667. https://doi.org/10.1126/science.1260793

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Blevins, W.R., Tavella, T., Moro, S.G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., Carey, L.B., Albà, M.M., 2019. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Sci. Rep. 9, 11005. https://doi.org/10.1038/s41598-019-47424-w

Brar, G.A., Weissman, J.S., 2015. Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat. Rev. Mol. Cell Biol. 16, 651–664. https://doi.org/10.1038/nrm4069

Buccitelli, C., Selbach, M., 2020. mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 21, 630–644. https://doi.org/10.1038/s41576-020-0258-4

Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonniere, E., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., 2023. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. https://doi.org/10.1101/2023.05.17.541122

Chotewutmontri, P., Barkan, A., 2016. Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. PLoS Genet. 12, e1006106. https://doi.org/10.1371/journal.pgen.1006106

Corbett, A.H., 2018. Post-transcriptional regulation of gene expression and human disease. Curr. Opin. Cell Biol. 52, 96–104. https://doi.org/10.1016/j.ceb.2018.02.011

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., Mann, M., 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics MCP 13, 2513–2526. https://doi.org/10.1074/mcp.M113.031591

De Vuyst, L., Comasio, A., Kerrebroeck, S.V., 2023. Sourdough production: fermentation strategies, microbial ecology, and use of non-flour ingredients. Crit. Rev. Food Sci. Nutr. 63, 2447–2479. https://doi.org/10.1080/10408398.2021.1976100

De Vuyst, L., Weckx, S., 2016. The cocoa bean fermentation process: from ecosystem analysis to starter culture development. J. Appl. Microbiol. 121, 5–17. https://doi.org/10.1111/jam.13045

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., Torres, E.M., 2014. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife 3, e03023. https://doi.org/10.7554/eLife.03023

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., Giaever, G., 2005. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. Genetics 169, 1915–1925. https://doi.org/10.1534/genetics.104.036871

De Vuyst, L., Leroy, F., 2020. Functional role of yeasts, lactic acid bacteria and acetic acid bacteria in cocoa fermentation processes. FEMS Microbiol. Rev. 44, 432–453. https://doi.org/10.1093/femsre/fuaa014

Díaz-Muñoz, C., Verce, M., De Vuyst, L., Weckx, S., 2022. Phylogenomics of a Saccharomyces cerevisiae cocoa strain reveals adaptation to a West African fermented food population. iScience 25, 105309. https://doi.org/10.1016/j.isci.2022.105309

Eardley, J., Timson, D.J., 2020. Yeast Cellular Stress: Impacts on Bioethanol Production. Fermentation 6, 109. https://doi.org/10.3390/fermentation6040109

Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., Jensson, B.O., Zink, F., Halldorsson, G.H., Masson, G., Arnadottir, G.A., Katrinardottir, H., Juliusson, K., Magnusson, M.K., Magnusson, O.T., Fridriksdottir, R., Saevarsdottir, S., Gudjonsson, S.A., Stacey, S.N., Rognvaldsson, S., Eiriksdottir, T., Olafsdottir, T.A., Steinthorsdottir, V., Tragante, V., Ulfarsson, M.O., Stefansson, H., Jonsdottir, I., Holm, H., Rafnar, T., Melsted, P., Saemundsdottir, J., Norddahl,

G.L., Lund, S.H., Gudbjartsson, D.F., Thorsteinsdottir, U., Stefansson, K., 2021. Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. 53, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w

Fernández Maura, Y., Balzarini, T., Clapé Borges, P., Evrard, P., De Vuyst, L., Daniel, H.-M., 2016. The environmental and intrinsic yeast diversity of Cuban cocoa bean heap fermentations. Int. J. Food Microbiol. 233, 34–43. https://doi.org/10.1016/j.ijfoodmicro.2016.06.012

Gene Ontology Consortium, 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334. https://doi.org/10.1093/nar/gkaa1113

Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., Beltrao, P., 2017. Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. Cell Syst. 5, 386-398.e4. https://doi.org/10.1016/j.cels.2017.08.013

Ingolia, N.T., 2010. Chapter 6 - Genome-Wide Translational Profiling by Ribosome Footprinting, in: Methods in Enzymology, Guide to Yeast Genetics: Functional Genomics, Proteomics, and Other Systems Analysis. Academic Press, pp. 119–142. https://doi.org/10.1016/S0076-6879(10)70006-9

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 324, 218–223. https://doi.org/10.1126/science.1168978

Jiang, L., Wang, M., Lin, S., Jian, R., Li, Xiao, Chan, J., Dong, G., Fang, H., Robinson, A.E., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., MacArthur, D.G., Meier, S.R., Nedzel, J.L., Nguyen, D.Y., Segrè, A.V., Todres, E., Balliu, B., Barbeira, A.N., Battle, A., Bonazzola, R., Brown, A., Brown, C.D., Castel, S.E., Conrad, D., Cotter, D.J., Cox, N., Das, S., Goede, O.M. de, Dermitzakis, E.T., Engelhardt, B.E., Eskin, E., Eulalio, T.Y., Ferraro, N.M., Flynn, E., Fresard, L., Gamazon, E.R., Garrido-Martín, D., Gay, N.R., Guigó, R., Hamel, A.R., He, Y., Hoffman, P.J., Hormozdiari, F., Hou, L., Im, H.K., Jo, B., Kasela, S., Kellis, M., Kim-Hellmuth, S., Kwong, A., Lappalainen, T., Li, Xin, Liang, Y., Mangul, S., Mohammadi, P., Montgomery, S.B., Muñoz-Aguirre, M., Nachun, D.C., Nobel, A.B., Oliva, M., Park, YoSon, Park, Yongjin, Parsana, P., Reverter, F., Rouhana, J.M., Sabatti, C., Saha, A., Skol, A.D., Stephens, M., Stranger, B.E., Strober, B.J., Teran, N.A., Viñuela, A., Wang, G., Wen, X., Wright, F., Wucher, V., Zou, Y., Ferreira, P.G., Li, G., Melé, M., Yeger-Lotem, E., Barcus, M.E., Bradbury, D., Krubit, T., McLean, J.A., Qi, L., Robinson, K., Roche, N.V., Smith, A.M., Sobin, L., Tabor, D.E., Undale, A., Bridge, J., Brigham, L.E., Foster, B.A., Gillard, B.M., Hasz, R., Hunter, M., Johns, C., Johnson, M., Karasik, E., Kopen, G., Leinweber, W.F., McDonald, A., Moser, M.T., Myer, K., Ramsey, K.D., Roe, B., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Jewell, S.D., Rohrer, D.C., Valley, D.R., Davis, D.A., Mash, D.C., Branton, P.A., Barker, L.K., Gardiner, H.M., Mosavel, M., Siminoff, L.A., Flicek, P., Haeussler, M., Juettemann, T., Kent, W.J., Lee, C.M., Powell, C.C., Rosenbloom, K.R., Ruffier, M., Sheppard, D., Taylor, K., Trevanion, S.J., Zerbino, D.R., Abell, N.S., Akey, J., Chen, L., Demanelis, K., Doherty, J.A., Feinberg, A.P., Hansen, K.D., Hickey, P.F., Jasmine, F., Kaul, R., Kibriya, M.G., Li, J.B., Li, Q., Linder, S.E., Pierce, B.L., Rizzardi, L.F., Smith, K.S., Stamatoyannopoulos, J., Tang, H., Carithers, L.J., Guan, P., Koester, S.E., Little, A.R., Moore, H.M., Nierras, C.R., Rao, A.K., Vaught, J.B., Volpi, S., Snyder, M.P., 2020. A Quantitative Proteome Map of the Human Body. Cell 183, 269-283.e19. https://doi.org/10.1016/j.cell.2020.08.036

Jüschke, C., Dohnal, I., Pichler, P., Harzer, H., Swart, R., Ammerer, G., Mechtler, K., Knoblich, J.A., 2013. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. Genome Biol. 14, r133. https://doi.org/10.1186/gb-2013-14-11-r133

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., Peterson, H., 2020. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. https://doi.org/10.12688/f1000research.24956.2

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., Sergushichev, A., 2021. Fast gene set enrichment analysis. https://doi.org/10.1101/060012

Lahtvee, P.-J., Sánchez, B.J., Smialowska, A., Kasvandik, S., Elsemman, I.E., Gatto, F., Nielsen, J., 2017. Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. Cell Syst. 4, 495-504.e5. https://doi.org/10.1016/j.cels.2017.03.003

Lahue, C., Madden, A., Dunn, R., Smukowski Heil, C., 2020. History and Domestication of Saccharomyces cerevisiae in Bread Baking. Front. Genet. 11.

Lee, T.I., Young, R.A., 2013. Transcriptional Regulation and Its Misregulation in Disease. Cell 152, 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Leroy, F., De Vuyst, L., 2004. Lactic acid bacteria as functional starter cultures for the food fermentation industry. Trends Food Sci. Technol. 15, 67–78. https://doi.org/10.1016/j.tifs.2003.09.004

Li, G.-W., Burkhardt, D., Gross, C., Weissman, J.S., 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157, 624–635. https://doi.org/10.1016/j.cell.2014.02.033

Liu, Y., Beyer, A., Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550. https://doi.org/10.1016/j.cell.2016.03.014

Lukoszek, R., Feist, P., Ignatova, Z., 2016. Insights into the adaptive response of Arabidopsis thaliana to prolonged thermal stress by ribosomal profiling and RNA-Seq. BMC Plant Biol. 16, 221. https://doi.org/10.1186/s12870-016-0915-0

Ma, M., Liu, Z.L., 2010. Mechanisms of ethanol tolerance in Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. 87, 829–845. https://doi.org/10.1007/s00253-010-2594-3

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., Stamatoyannopoulos, J.A., 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337, 1190–1195. https://doi.org/10.1126/science.1222794

McManus, C.J., May, G.E., Spealman, P., Shteyman, A., 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422–430. https://doi.org/10.1101/gr.164996.113

Morrill, S.A., Amon, A., 2019. Why haploinsufficiency persists. Proc. Natl. Acad. Sci. 116, 11866–11871. https://doi.org/10.1073/pnas.1900437116

Niu, L., Stinson, S.E., Holm, L.A., Lund, M.A.V., Fonvig, C.E., Cobuccio, L., Meisner, J., Juel, H.B., Thiele, M., Krag, A., Holm, J.-C., Rasmussen, S., Hansen, T., Mann, M., 2023. Plasma Proteome Variation and its Genetic Determinants in Children and Adolescents. https://doi.org/10.1101/2023.03.31.23287853

Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., Ishii, S., 2003. A Bayesian missing value estimation method for gene expression profile data. Bioinforma. Oxf. Engl. 19, 2088–2096. https://doi.org/10.1093/bioinformatics/btg287

Ohnuki, S., Ohya, Y., 2018. High-dimensional single-cell phenotyping reveals extensive haploinsufficiency. PLOS Biol. 16, e2005130. https://doi.org/10.1371/journal.pbio.2005130

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198. https://doi.org/10.1093/nar/gkz369

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007

Sayols, S., 2022. rrvgo: Reduce + Visualize GO. https://doi.org/10.18129/B9.bioc.rrvgo

Smyth, G.K., 2005. limma: Linear Models for Microarray Data, in: Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A., Dudoit, S. (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Statistics for Biology and Health. Springer, New York, NY, pp. 397–420. https://doi.org/10.1007/0-387-29362-0_23

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. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

Taggart, J.C., Li, G.-W., 2018. Production of Protein-Complex Components is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Syst. 7, 580-589.e4. https://doi.org/10.1016/j.cels.2018.11.003

Teyssonniere, E., Shichino, Y., Friedrich, A., Iwasaki, S., Schacherer, J., Submitted. Translation variation across genetic backgrounds reveals a post-transcriptional buffering signature in yeast.

The GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110

Trösch, R., Barahimipour, R., Gao, Y., Badillo-Corona, J.A., Gotsmann, V.L., Zimmer, D., Mühlhaus, T., Zoschke, R., Willmund, F., 2018. Commonalities and differences of chloroplast translation in a green alga and land plants. Nat. Plants 4, 564–575. https://doi.org/10.1038/s41477-018-0211-0

Vanegas, J.M., Contreras, M.F., Faller, R., Longo, M.L., 2012. Role of Unsaturated Lipid and Ergosterol in Ethanol Tolerance of Model Yeast Biomembranes. Biophys. J. 102, 507–516. https://doi.org/10.1016/j.bpj.2011.12.038

Veitia, R.A., Potier, M.C., 2015. Gene dosage imbalances: action, reaction, and models. Trends Biochem. Sci. 40, 309–317. https://doi.org/10.1016/j.tibs.2015.03.011

Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M.E., Peters, A.H.F.M., Anders, S., Kaessmann, H., 2020. Transcriptome and translatome co-evolution in mammals. Nature 1–6. https://doi.org/10.1038/s41586-020-2899-z

CHAPTER III

Species-wide quantitative transcriptomes and proteomes reveal distinct genetic control of gene expression variation in yeast

Collaborative work from:

E. Teyssonnière[1], P. Trébulle[3], J. Muenzner[2], V. Loegler[1], D. Ludwig[2, 4], F. Amari[2, 4], M. Mülleder[4], A. Friedrich[1], J. Hou[1], M. Ralser[2,3,5], and J. Schacherer[1,6]

1. Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France
2. Charité Universitätsmedizin Berlin, Berlin, Germany
3. The Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, UK
4. Core Facility High Throughput Mass Spectrometry, Charité Universitätsmedizin, Berlin, Germany
5. Max Planck Institute for Molecular Genetics, Berlin, Germany
6. Institut Universitaire de France (IUF), Paris, France

Summary

Gene expression varies between individuals and corresponds to a key step linking genotypes to phenotypes. However, our knowledge regarding the species-wide genetic control of protein abundance, including its dependency on transcript levels, is very limited. Here, we have determined quantitative proteomes of a large population of 942 diverse natural Saccharomyces cerevisiae yeast isolates. We found that mRNA and protein abundances are weakly correlated at the population gene level. While the protein co-expression network recapitulates major biological functions, differential expression patterns reveal proteomic signatures related to specific populations. Comprehensive genetic association analyses highlight that genetic variants associated with variation in protein (pQTL) and transcript (eQTL) levels poorly overlap (3.6%). Our results demonstrate that transcriptome and proteome are governed by distinct genetic bases, likely explained by protein turnover. It also highlights the importance of integrating these different levels of gene expression to better understand the genotype-phenotype relationship.

Introduction

Understanding the genetic basis of phenotypic variation in natural populations is one of the main goals of modern biology. Gene expression differs among individuals and is known to be a main determinant of phenotypic variation (Albert and Kruglyak, 2015; Maurano et al., 2012). In humans, the onset and development of numerous diseases have been linked to abnormal regulation of gene expression (Cookson et al., 2009). It is therefore essential to understand how genomic information is expressed through the different layers of gene regulation (*i.e.,* transcriptomes and proteomes). Over the past decades, the development of methods for high-throughput quantification of mRNA and protein abundance has made it possible to explore both the proteome and the transcriptome on a larger scale (Messner et al., 2022b; Moyerbrailean et al., 2015). These approaches facilitated the detection of numerous genetic loci (quantitative trait loci, QTL) affecting either transcript (eQTL) or protein (pQTL) levels (Chick et al., 2016; Ferkingstad et al., 2021; Folkersen et al., 2020; The GTEx Consortium, 2020, 2017, 2015). However, the relationship between transcript and protein levels remains debated and poorly understood at the population level (Buccitelli and Selbach, 2020).

The transcript-protein correlation provides a first global view of the dependency of the two gene expression layers. Two types of mRNA-protein correlation can be determined, across- and within-gene, reflecting very different dynamics (Buccitelli and Selbach, 2020; Fortelny et al., 2017; Liu et al., 2016). The across-gene correlation analysis focuses on the overall correlation of a large set of genes coming from the same sample under a given condition to find out how well the absolute abundances of mRNAs and proteins are correlated. This correlation has been widely investigated in several species, such as human (Battle et al., 2015; Edfors et al., 2016; Gautier et al., 2016; Salovska et al., 2020; Wang et al., 2019; Wilhelm et al., 2014; Zhang et al., 2014), rats and mice (Aydin et al., 2023; Li et al., 2014; Moritz et al., 2019; Schwanhäusser et al., 2011), flies (Becker et al., 2018), plants (Ponnala et al., 2014) or yeast (Gygi et al., 1999; Ingolia et al., 2009; Marguerat et al., 2012). Across-gene correlations are consistently high and range from 0.4 to 0.8, suggesting that the absolute number of transcripts and proteins are globally correlated. Therefore, very abundant transcripts generally lead to very abundant proteins and vice versa.

However, the relationship between the transcript and protein abundance at the population level is explored via their variation across samples (*e.g.*, individuals, tissues, or cell lines). Within-gene correlation analysis gives a view on how the protein level of each gene tracks its mRNA level in a population. Different studies have investigated this within-gene correlation in

different contexts and organisms, but they often show divergent results. Several surveys of tumors, normal human tissues, as well as pluripotent stem cells have highlighted this discrepancy in estimates with median within-gene correlation coefficients ranging from 0.14 to 0.59 (Archer et al., 2018; Aydin et al., 2023; Battle et al., 2015; Huang et al., 2017; Jiang et al., 2020; Mertins et al., 2016; Mirauta et al., 2020; Mun et al., 2019; Upadhya and Ryan, 2022; Vasaikar et al., 2019; Wang et al., 2019; Zhang et al., 2014, 2016). Similarly, the overlap of the detected loci influencing mRNA (eQTL) and protein (pQTL) abundance greatly differed across the datasets. It ranges from a very weak overlap of 5.5% in a study on 97 inbred and recombinant mice to nearly 35% in human (n = 62) and mice (n = 192) (Battle et al., 2015; Chick et al., 2016; Ghazalpour et al., 2011).

Part of the diverging results might have been driven by technical limitations. For instance, it has been shown that by selecting the most representative peptides in prior proteomic methods, the overall correlation of global transcript and mRNA abundance improves significantly (Alam et al., 2016; Upadhya and Ryan, 2022). A key difference is also whether the goal of the survey is to correlate absolute number of transcripts and proteins, or relative changes in protein or mRNA levels, which differ between samples. While the absolute number of transcripts and proteins spans several orders of magnitude, the relative expression differences of any individual protein across samples varies within a much narrower range (Marguerat et al., 2012; Messner et al., 2022a). Finally, a main limitation of these studies is that the sample size is much lower than the dimensionality of the problem.

To determine to which extent differences in relative changes in mRNA and protein levels are correlated and the genetic origins of their abundance variation are shared, a large-scale population survey exploring these two facets in a quantitative way was therefore necessary. Here, we took advantage of the 1,011 yeast *Saccharomyces cerevisiae* population we genome-sequenced and for which we have a species-level understanding of the natural genetic and phenotypic diversity (Peter et al., 2018). In order to be fully able to compare and analyze at unprecedented detail the relationship between these two layers of gene regulation, we therefore generated 942 quantitative proteomes in which cells were also cultured in synthetic complete medium supplemented with amino acids using high-throughput mass-spectrometry. We found that protein levels are molecular traits that exhibit considerable variation between individuals and specific signatures related to certain subpopulations. This large available population also makes it possible to generate a detailed map of loci involved in the variation of protein abundance (pQTL) at the species level, via genome-wide association studies (GWAS). Interestingly, local pQTL are less frequent than distant ones (8% of the total set of pQTL) but

they have a higher impact on their respective traits. Integration of proteomic and transcriptomic datasets acquired in parallel under similar conditions allowed comparison of accurate quantification of the mRNA and protein abundance of 629 genes across 889 natural isolates (Caudal et al., 2023). Based on these unique datasets, we clearly demonstrated that the degree of within-gene correlation between protein and mRNA abundance is very low (*Rho* = 0.165). Consistently, we found that the genetic variants influencing protein and mRNA abundance are very dissimilar. Our study highlights that population-scale proteomes are essential and add a new dimension to the characterization of the genotype-phenotype relationship when integrated with genomic and transcriptomic information.

Results

Quantitative proteomes of a large collection of natural isolates

We generated a quantitative proteomic dataset for strains of the 1,011 strains collection (Peter et al., 2018) from cells cultivated in synthetic complete medium with amino acids in order to match the growth medium used for RNA sequencing (Caudal et al., 2023) (Figure 1A).We had previously acquired a proteome dataset of the 1,011 strains collection, measured with microflow chromatography and SWATH MS (Muenzner et al., 2022). For the acquisition of this new dataset we used a proteomic method that allows for an even higher throughput, using analytical flowrate chromatography and Scanning-SWATH MS with a 3 min gradient (Messner et al., 2021). After cultivation of the yeast isolates in 96 well plates, proteins were extracted, and subjected to reduction, alkylation, and trypsination in a semi-automated workflow using liquid handling robotics (Messner et al., 2020). Peptide preparations were separated using a 3-minute high-flow rate (800 µl/min) chromatographic gradient using an Infinity II chromatographic system (Agilent Technologies), coupled to a 6600 Triple TOF instrument (Sciex). Data was recorded using Scanning SWATH acquisition (Messner et al., 2021) and the raw data was processed using the DIA-NN software (version 1.8), which was specifically developed for large scale proteomic exploration (Demichev et al., 2020). We applied several quality filters where poor-quality samples were removed from the analysis, and we excluded peptides that were not detected in more than 80% of the samples (see Methods). The generated dataset hence encompasses protein abundance quantification for 630 proteins among 942 isolates (Table S1 and Table S2). This dataset therefore covers the overall genetic diversity of the species and captures the subpopulations that were defined as part of the 1,011 yeast genomes project, including both domesticated and wild clades (Peter et al., 2018) (Figure S1A). We combined the proteomic dataset with transcriptomic data obtained from the 1,011 strains collection (Caudal et al., 2023), which gave access to the quantified expression of both levels for 629 genes across 889 isolates (Figure 1B, Table S1). To be able to properly compare these two datasets, we normalized them with quantile normalization after imputing the missing values using the KNN method (Table S3, Figure S1B-C).

To characterize the quantified proteins in our study, we first compared the level of transcription of both the identified and unidentified proteins. Low abundance transcripts are less likely to be quantified by proteomics as compared to high abundant transcripts (Figure S1D). Indeed, 489

out of 629 consistently quantified proteins fall into the 20% highest transcribed genes (n = 1,304). In total, 537 out of 629 quantified proteins were found in the two highest abundance deciles as defined in a recent yeast protein abundance meta-analysis (Ho et al., 2018) (Figure S1E). Overall, proteins related to essential genes and involved in molecular complexes were both significantly enriched in the set of proteins quantified by Scanning SWATH (odd-ratio = 3.5 and 2.2 respectively, Fisher's exact test, p-values < $2.2x10^{-16}$) (Dowell et al., 2010; Giaever et al., 2002; Pu et al., 2009). Function-wise, we found that metabolism-related genes were overrepresented among the 629 genes included in our study (Table S4).

We then investigated the level of variation in protein abundance by calculating the coefficient of variation (CV) for each protein using the non-normalized dataset. We found an average CV of 31%, varying between 12% and 98% and one high outlier reaching 300% (PDC5, a pyruvate decarboxylase). The precursor-level CVs across quality control samples (15.15%) were much lower than the precursor-level CVs across the natural isolate samples (34.21%), confirming that a biological signal was observed across the isolates (see Methods). Gene set enrichment analyses (GSEA) were performed using the CVs and significant enrichment of genes related to amino acid metabolism, respiration or pyruvate metabolism was found for proteins with a high CV, indicating that they vary the most (Table S5). By contrast, proteins with a low CV were significantly related to genes involved in tRNA aminoacylation or protein degradation.

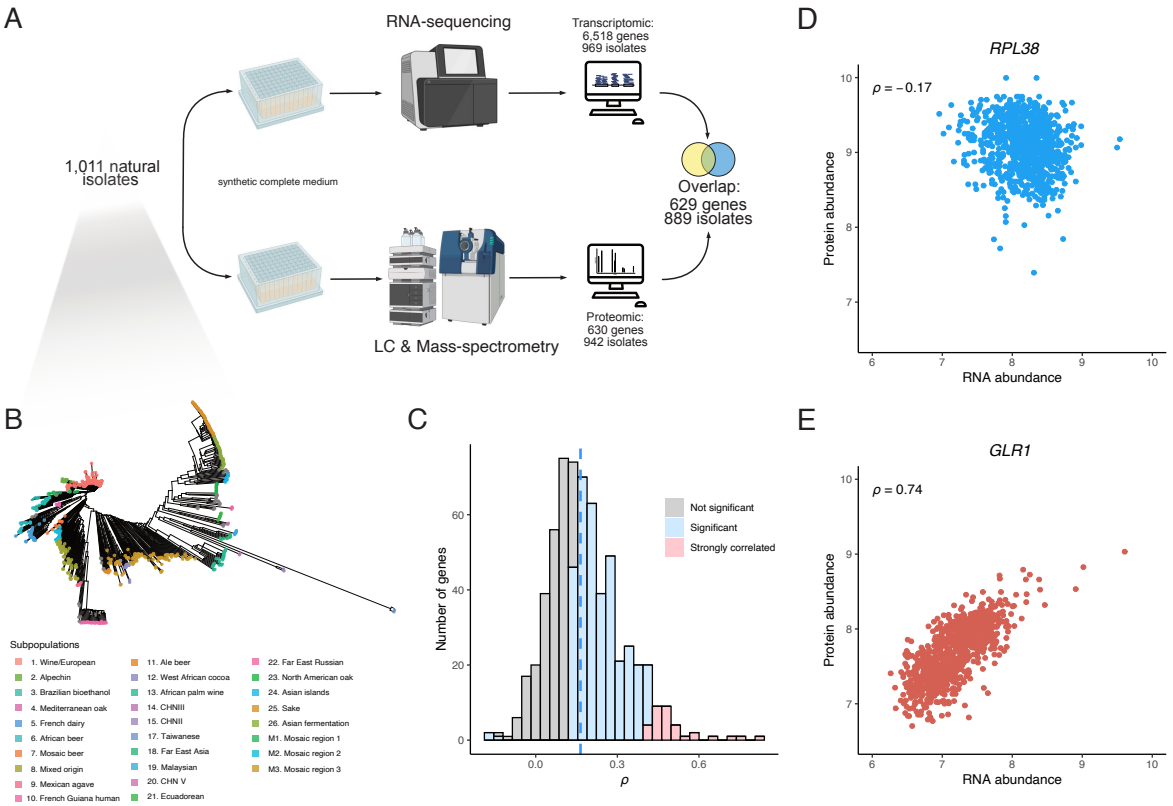


Figure 1. Quantitative proteomes and transcriptomes of a large *S. cerevisiae* population.

A. The proteomic dataset was generated on isolates grown in synthetic complete (SC) medium with amino acids using a semi-automated sample preparation workflow, and Scanning-SWATH MS (see Methods). The overlap between this dataset and the recently generated transcriptomic dataset on the same population in the same condition (Caudal et al., 2023) resulted in 629 protein/transcript abundances across 889 isolates. **B.** Phylogenetic trees of the isolates used in this study. Colors correspond to previously defined subpopulations (Peter et al., 2018). **C.** Gene-wise correlation coefficients (Spearman correlation test) between the proteome and the transcriptome. **D.** and **E.** mRNA-protein within-gene correlation across isolates for the *RPL38* and *GLR1* genes (ρ corresponds to the Spearman correlation coefficient with p-values of 4.8×10^{-7} and 2.3×10^{-153}, respectively).

Transcript and protein abundances are weakly correlated at the gene level across isolates

As proteomes and transcriptomes were obtained using the same growth media, our dataset allowed us to characterize the different types of correlation between mRNA and protein abundance across a natural population. We first determined the across-gene correlation, *i.e.* the concordance between protein and transcript abundance for each isolate, and found a very high correlation (median *rho* = 0.53, interquartile range of 0.06, Figure S2), which is consistent with what was previously described (Battle et al., 2015; Becker et al., 2018; Edfors et al., 2016; Gautier et al., 2016; Moritz et al., 2019; Ponnala et al., 2014; Salovska et al., 2020; Wang et al., 2019; Wilhelm et al., 2014; Zhang et al., 2014). We next computed the correlation between the protein and mRNA normalized abundance for each gene across the 889 natural isolates (Figure 1C-D-E, Table S6). While the across-gene correlation levels were in line with previous explorations, we found an overall very low within-gene correlation level (median ρ = 0.165, interquartile range of 0.17). This value is much lower than the one determined with smaller samples in mice (approximately 0.25) (Chick et al., 2016; Ghazalpour et al., 2011) and in human healthy tissues (0.35 and 0.46) (Jiang et al., 2020; Wang et al., 2019), but it is in line with what was found in human lymphoblastoid cell lines (0.14) (Battle et al., 2015). For a total of 385 out of the 629 quantified proteins, the level is significantly correlated with RNA level (Bonferroni corrected p-value < 0.05). Out of these 385 proteins, only 3 show a negative correlation: Rps13, Asc1 and Rpl38 (Figure 1D), all ribosomal related proteins. This observation is consistent with previous surveys pointing out that some ribosome-related proteins are negatively correlated with their cognate transcripts (Buccitelli and Selbach, 2020; Wang et al., 2019). But overall, this correlated set of 385 proteins/transcripts is significantly enriched of genes related to several

metabolism pathways (Table S7). Moreover, the most strongly correlated set of proteins/transcripts (n = 33) show functional enrichment of genes related to mitochondrial respiration (Table S8) (see Methods). Interestingly, it points out that this specific pathway has similar gene regulation at both levels. Finally, we observed that four genes with very high mRNA-protein correlation were located outside the main correlation index distribution (Figure 1C). These genes all have correlation coefficients greater than 0.6: *SFA1* (alcohol dehydrogenase), *HBN1* (unknown function), *GLR1* (glutathione oxidoreductase, Figure 1E) and *YLR179C* (unknown function). Such a high correlation clearly points to common regulatory mechanisms and genetic bases underlying the two levels of variation, as we have seen below.

Gene expression is more constrained at the proteome level

By combining these proteomic and transcriptomic datasets, we are in a position to simultaneously explore and compare the variation of these two gene expression layers at the population level. We therefore computed the absolute *Log2(fold change)* value (*i.e.,* $|Log_2(FC)|$) for each gene in each pair of isolates and found that this value is 32% lower on average for the proteome (Figure 2A), suggesting that protein abundance is less variable and more constrained than mRNA abundance. Furthermore, a higher correlation was observed between proteomes (*rho* = 0.92) compared to transcriptomes (*rho* = 0.83) (Figure 2B). Finally, the variance observed for each gene was lower for the proteomic data (Figure S3A) and the Euclidean distances between each isolate were smaller when computed with the protein abundance dataset (Figure S3B). Overall, these observations reflect and highlight the presence of a global post-transcriptional buffering of the transcriptome variations.

Despite recurrent observations (Blevins et al., 2019; Kustatscher et al., 2017; McManus et al., 2014; Muenzner et al., 2022; Wang et al., 2020), the post-transcriptional buffering phenomenon remains largely functionally uncharacterized and poorly understood. We sought to better understand this phenomenon at the genetic level by examining the cellular functions that tended to be most affected by post-transcriptional buffering. Briefly, we constructed neighbor-joining trees using the proteome or transcriptome Euclidean distances between each isolate (Figure 2C and see Methods) (Wang et al., 2020). Total branch length was used as a measure of expression variation and evolution at the species level. We then calculated the ratio between the lengths of the proteome and transcriptome tree branches to quantify the strength of the post-transcriptional buffering phenomenon. The lengths of branches from the proteome-based tree were shorter than those from the transcriptome-based tree, resulting in a length ratio of 0.93 (Figure 2C, Figure

S3C). This observation is consistent with the differences in Euclidean distances observed previously (Figure 2B). We then applied the same procedure to 101 sets of genes, representing central biological processes obtained from a reduced list of gene ontology (GO) annotations (Table S9). We found that a total of 16 sets display a ratio lower than 0.93 and a significant difference between the proteome and transcriptome branch lengths, meaning that these sets are strongly affected by the phenomenon of post-transcriptional buffering (Figure 2D, Table S10). Interestingly, 6 out of the 16 sets include genes with functions related to protein production and maturation (Figure 2D), highlighting that the evolution of the cellular machinery involved in protein production and maturation is highly constrained. The other set of genes are related to several metabolism processes and detected as strongly buffered, despite being highly variable in the proteomic data (Table S5). This observation could be due to the fact that metabolism-related genes are among the genes with the greatest variation in mRNA abundance at the species level (Caudal et al., 2023). This variation is largely attenuated at the proteome level but remains important, reflecting differences in metabolic preferences within the population. Moreover, we also found 3 sets with a ratio higher than 0.93 and a significant difference between the proteome and transcriptome trees, which means that the expression variation of these genes is greater at the proteome level (Figure 2D, Table S10). Interestingly, all of them are related to protein catabolism, highlighting a difference in post-transcriptional mechanism for this specific functional category. Taken together, these results provide new insights into post-transcriptional buffering as well as its functional impact.

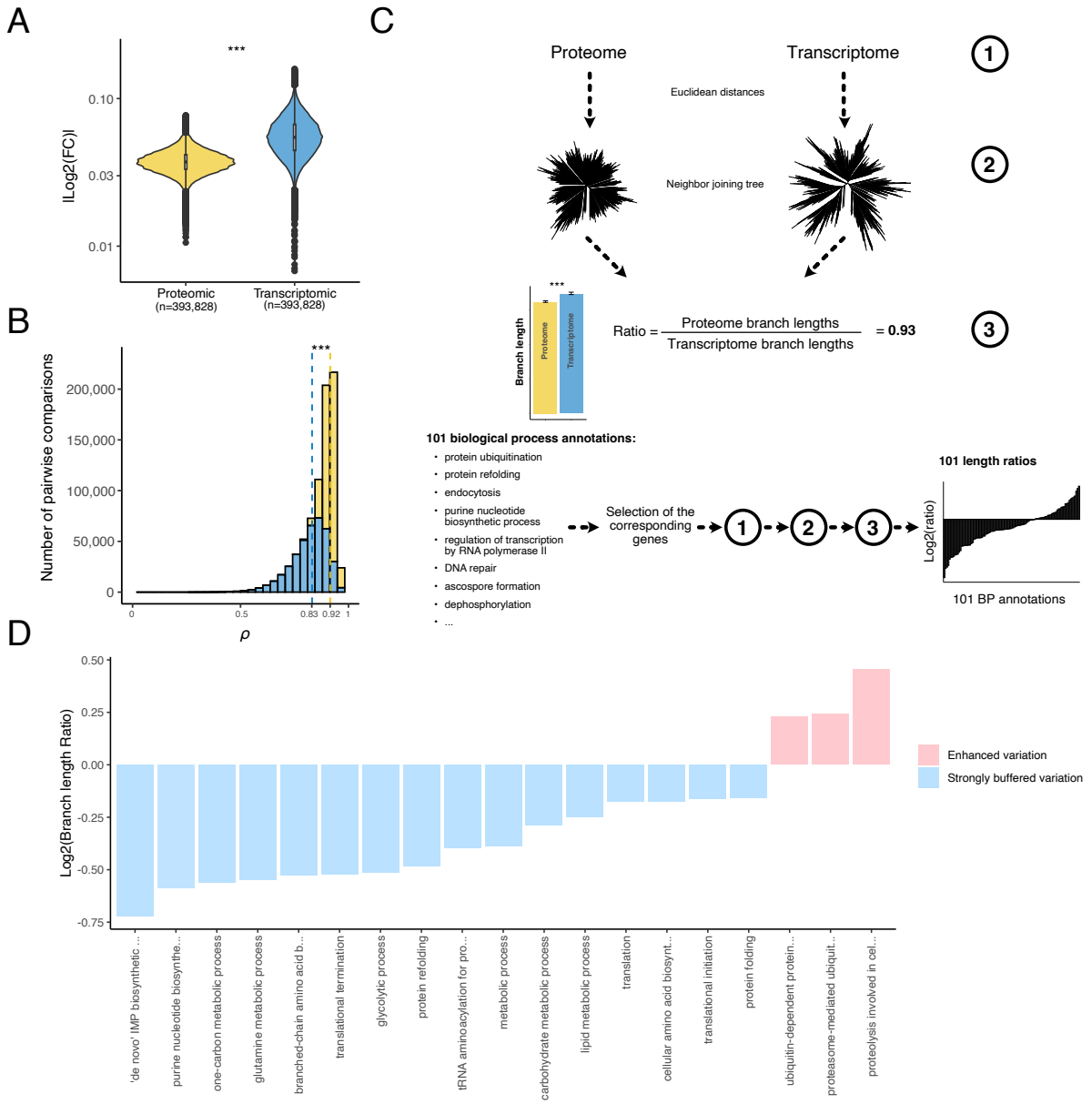


Figure 2. Detection and functional description of the post-transcriptional buffering.

A. Median | *log2(fold changes)* | computed in each isolate pairwise comparison using both proteomic and transcriptomic data (*** = Wilcoxon test, p-value < $2.2x10^{-16}$) (see Methods). **B.** Correlation coefficients from the isolate pairwise comparisons using both protein and transcript abundance (*** = Wilcoxon test, p-value < $2.2x10^{-16}$). The dotted lines correspond to the median correlation index for the proteomic (yellow) and transcriptomic (blue) data. **C.** Cellular functions that are preferentially affected by post-transcriptional buffering. Briefly, using either the proteome and the transcriptome abundances -1- we constructed expression-based neighbor joining trees -2- and compared the total sum of the branch lengths. We computed a ratio -3- defined by the proteome total branch lengths divided by the transcriptome total branch lengths. Using all the genes, this ratio was equal to 0.93 (overall, the expression evolution is more constrained at the proteome level). We performed the same procedure using subsets of genes corresponding to 101 biological process annotations. The biological processes

displaying a ratio lower than 0.93 and a significant difference in terms of branch lengths (see Methods) were considered as strongly buffered. The biological processes displaying a ratio higher than 1 and a significant difference in term of branch lengths had an enhanced abundance variation at the proteome level. **D.** Biological processes detected as strongly buffered or with an enhanced variation using the procedure detailed in C.

Architecture of the proteome landscape

Using these datasets, we then sought to understand the main determinants shaping the proteome architecture at the population level. The *S. cerevisiae* yeast species exhibit a clear population structure, which potentially can impact the proteome landscape (Peter et al., 2018) (Table S1). We performed a principal component analysis (PCA) with the protein abundance data and found that no clear grouping emerged from the subpopulations when plotting together the 6 first principal components (Figure S4A-B-C). The same results were observed for transcriptomes (Figure S4D-E-F). To confirm this, we also computed the Euclidean distance across transcript and protein levels between every pair of isolates and used these to construct a neighbor-joining tree (Figure 3B-C). We observed that none of the subpopulations present in the genetic-based tree merged in either the proteome- or transcriptome-based tree (Figure 3A-B-C). Together, these results highlight that population structure does not impact transcriptomes and proteomes in the *S. cerevisiae* species.

One potential determinant of the proteome organization could be related to co-expression networks that strongly influence the coordination of gene expression or various cellular processes. Using Weighted Gene Co-Expression Network Analysis (WGCNA) (Zhang and Horvath, 2005) on the normalized protein abundance data, we detected seven co-expression modules (Figure 3D, Table S11). Each of these modules corresponds to a specific biological function (Table S12, Figure S5) and encompasses between 38 (*Cellular amino acid biosynthetic process*) and 114 (*Ribosome biogenesis*) genes. Interestingly, very similar modules were found applying the same procedure on the mRNA normalized data. Five co-expression modules were detected (Figure 3E, Figure S6, Table S13, Table S14), and all of them were detected in the seven proteomic modules, suggesting that co-expression patterns recapitulate central cell functions are conserved across the two expression layers (Figure 3D-E).

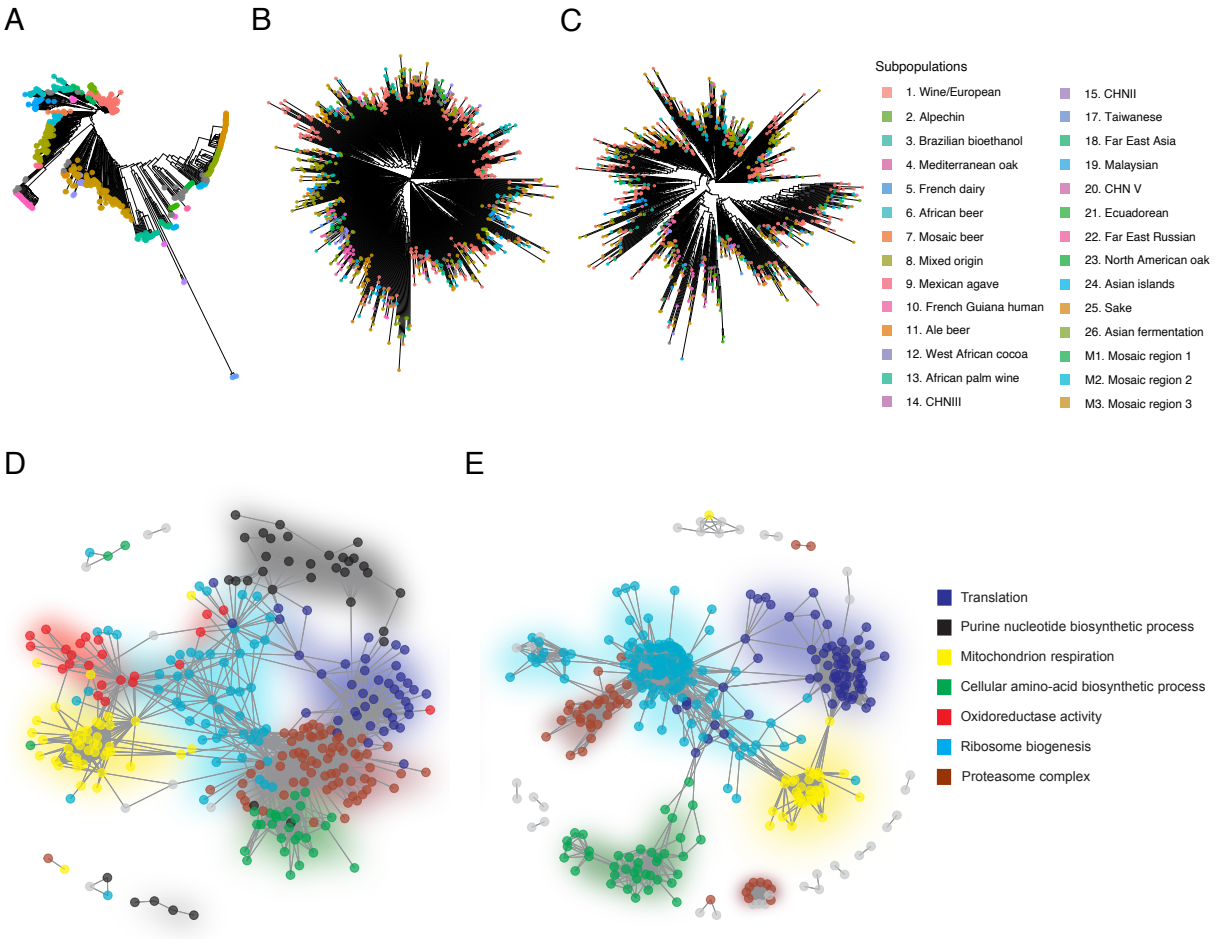


Figure 3. Co-expression network is a major determinant of the proteome organization while the population structure is not.

A-B-C. Comparison between the phylogenetic tree (A) obtained using the bi-allelic SNP (as in Peter et al. 2018) and the trees obtained from the Euclidean distances based on protein (B) or transcript (C) abundance. Colors correspond to the subpopulations. **D-E**. Cellular co-expression network computed with WGCNA using proteomic (D) or transcriptomic (E) data. Colors represent the cellular pathway detected for each co-expression module.

Insight into subpopulation-specific protein expression

We further wanted to explore and determine the presence of subpopulation-specific signatures. We therefore sought to identify differential protein expression patterns by comparing each clade to the rest of the population and we detected a total number of 1,129 differentially expressed proteins (DEPs) (corresponding to 465 unique proteins, Figure S7, Table S15). An average of 59 DEPs was found per clade, ranging from 218 for the Wine clade to 0 for wild Asian clades represented by a small sample (*e.g.* CHN, Taiwanese and Far East Russian) (Figure S8A). Several DEPs were adequately related to the ecological origin of the different subpopulations.

For example, several subpopulations related to alcoholic fermentation show overexpression of alcohol dehydrogenases, such as ADH4 in Wine and Brazilian bioethanol clades as well as ADH3 in the Sake subpopulation. In the French Dairy subpopulation, we also observed an underexpression of SEC23, a GTPase-activating protein involved in the COPII related vesicle formation, which could reflect an adaptation to this secretory pathway to the cheese-making environment (Celińska and Nicaud, 2019). Overall, these observations suggest that domestication and more generally, ecological constraints are drivers of the proteomic landscape evolution in a natural population. We then performed GSEA based on differential expressed proteins in each subpopulation and found significant enrichments for various biological processes (Figure 4A, Table S16). Many enriched functional categories were associated with respiration related genes (*e.g.* "*respiratory electron chain transport*"). Interestingly, we observed that while most wild clades (8 out of 13) tend to have overexpression of respiration-related proteins, these are underexpressed in domesticated subpopulations (5 out of 7).

We therefore further explored the impact of domestication on the proteome at the population level. Using the same DEP detection method, we assessed the proteome differences between the domesticated and wild isolates (Peter et al., 2018) and found a total of 133 DEPs (Table S17). Among these proteins, other alcohol dehydrogenases such as SFA1 and ADH3 were highly abundant in domesticated isolates. A GSEA performed on this set of DEPs clearly shows an enrichment of underexpressed respiration-related proteins in domesticated clades (Figure 4B, Table S18). Unlike wild isolates, domesticated isolates were selected for fermentation purposes, likely leading to this specific signature. This observation is in line with the previous finding pointing out that the switch from a preference between respiration and fermentation is one of the hallmarks of domestication in yeast (Lahue et al., 2020). In addition, significant enrichment of the functional category "*chaperon mediated protein folding*" points to overexpression of this set of proteins in the domesticated isolates (Figure 4B), which may be an adaptative response to long-term exposure to ethanol, known to induce protein denaturation (Auesukaree, 2017). By performing the same analysis on transcriptomic data (Figure S8B, Table S17), similar results, showing overexpression of respiration-related genes in domesticated clades, were obtained (Table S19).

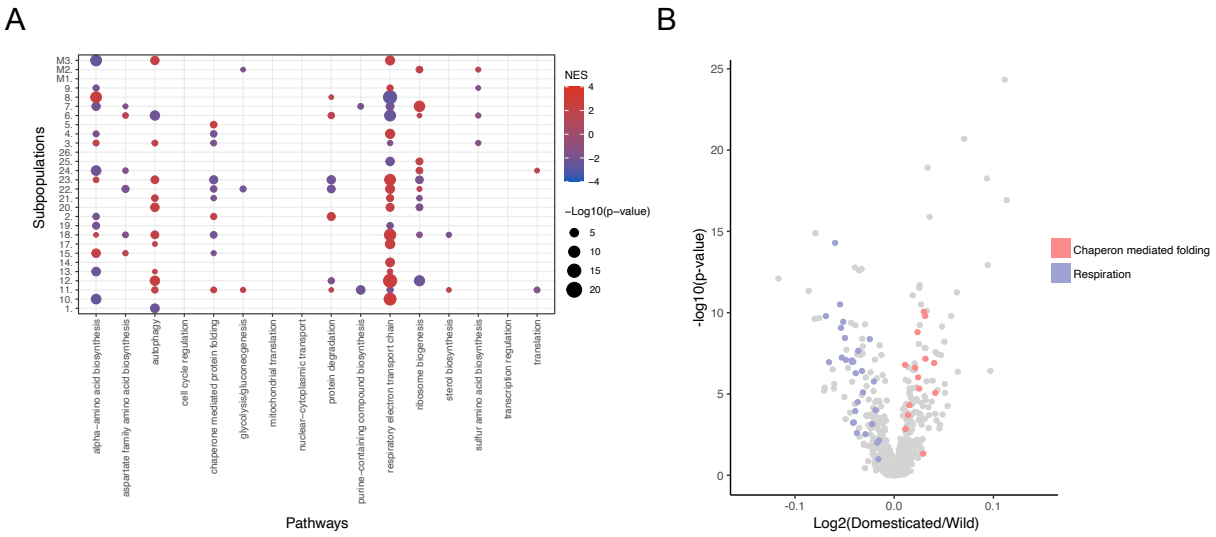


Figure 4. DEPs reveal domestication- and subpopulation-specific metabolic adaptation.

A. GSEA results on the DEPs (using 16 broad functional annotations from *44*) of each subpopulation. Colors represent the normalized enrichment score (NES): Red – overexpression, blue – underexpression in subpopulation. **B.** Volcano plot of the comparison between wild and domesticated isolates. Colors highlight the genes belonging to two functional annotations related to chaperon mediated folding and respiration.

The genetic bases of protein abundance at the population scale

To uncover the genetic origins of the proteome variation at the population-scale, we performed genome-wide association studies (GWAS) and considered both SNPs and CNVs that were characterized previously(Peter et al., 2018). We focused on isolates for which both proteomic and transcriptomic data were available, resulting in a set of 889 isolates. In this population, a total of 84,633 SNPs and 1,019 CNVs were considered, with a minor allele frequency higher than 5%. We performed GWAS using the raw protein abundances of the genes for which we have both levels of expression (*i.e.*, 629 genes). Overall, we detected a total of 598 SNP-pQTL after colliding SNP affected by linkage disequilibrium (R2 > 0.6), and 4,528 CNV-pQTL corresponding to 501 and 520 loci and affecting 300 and 93 genes, respectively (Figure 5A-B, Table S20, Table S21, data file 1).

Among the SNP-pQTL, 8% (n = 50) were local-pQTL, showing that regulation of protein abundance is primarily achieved through *trans* regulation. This fraction is consistent with previous exploration in yeast (Foss et al., 2007) and lower than what is usually found at the transcriptome level (Albert et al., 2018; Caudal et al., 2023). Nonetheless, we observed that the local SNP-pQTL have a higher effect size compared to *trans* SNP-pQTL (Figure 5C) and tend to be located near the transcription starting site of the gene (Figure S9). We found no strong

SNP-pQTL hotspots, suggesting that most of the distant pQTL are evenly distributed throughout the genome (Figure 5D).

In contrast, CNVs impacting protein abundance had a biased location on the chromosomes 1, 3, 8, 9 and 11 (Figure 5B). Out of 4,528 CNV-pQTL, a total of 4,303 were located on these chromosomes and affected a gene on their respective chromosome. This observed bias is due to the presence of aneuploidies on these chromosomes in our population (Peter et al., 2018). These CNV-pQTL have also a higher impact on the protein abundance variation compared to the other CNV-pQTL, suggesting that aneuploidies represent a major source of proteome variation at the population level (Figure S10). Only 24 local CNV-pQTL out of 4,528 were detected, and no significant effect size between local and distant CNV-pQTL was found (Figure 5C).

We then looked at the extent to which the genetic bases of protein abundance are common with those underlying the abundance of transcripts. We performed GWAS using the transcriptomic dataset and detected 596 SNP-eQTL and 4,877 CNV-eQTL (Figure S11, data file 2), which is of the same order of magnitude as the GWAS proteome results. Surprisingly, the overlap between the SNP-pQTL and the SNP-eQTL is very low, with only 3.6% of shared SNP-QTL (n = 22). Interestingly, 18 out of 22 were related to local regulation, meaning that 36% of the local SNP-pQTL (18 out of 50) also impact the cognate transcripts of their target protein. This observation is consistent with previous findings showing that the common regulation between mRNA and protein abundances is mainly related to local regulation (Chick et al., 2016; Ghazalpour et al., 2011). Overall, we observed that genes with a strong correlation between transcript and protein abundance, such as the top four most correlated genes previously mentioned (*SFA1*, *HBN1*, *GLR1* and *YLR179C*), tend to have a shared pQTL and eQTL (Figure S12). Additionally, we found that the SNP-pQTL distribution across the genome did not match the SNP-eQTL distribution, where a QTL hotspot could be detected around the *CTT1* gene (Caudal et al., 2023; Stuecker et al., 2018). The reasons for the weak overlap are likely multifactorial, but protein-specific regulation, such as protein degradation, may play a central role. We sought to confirm this by looking at the average protein turnover (Muenzner et al., 2022) of the proteins with and without overlapping pQTL and eQTL (Figure S13A, see Methods). We found that proteins, for which an overlap between pQTL and eQTL was detected, show a lower turnover rate compared to the other proteins. Consistently, the half-life of proteins with an overlapping SNP-QTL was higher than the rest of the proteome (Figure S13B). This observation suggests that protein degradation is probably involved in the large differences observed between the genetic origins of mRNA and protein abundance.

In contrast, the overlap between the two sets of CNV-QTL is much higher, as 3,097 QTLs were shared between the transcriptome and proteome, *i.e.*, approximately 68% of the CNV-pQTL. However, these shared CNV-QTLs are all aneuploidy-related CNVs, suggesting that the effect of aneuploidies is persistent through the expression layers (Muenzner et al., 2022). None of the non-aneuploidy CNV-QTL (22 CNV-eQTL and 216 CNV-pQTL) were shared. Together, our results highlight that the genetic bases underlying population-level protein abundance are very distinct from those underlying mRNA abundance.

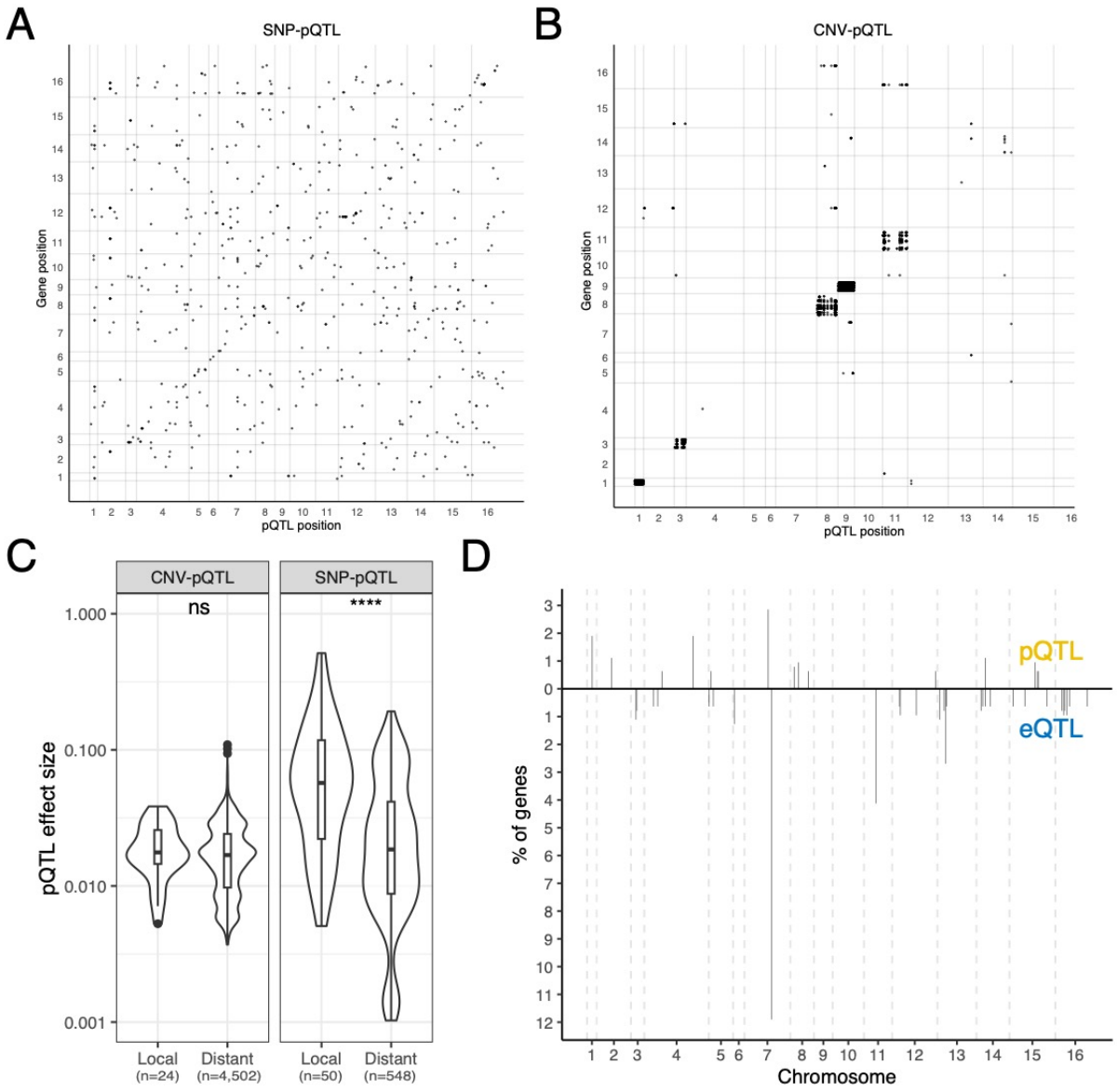


Figure 5. SNP- and CNV-pQTL detection highlights strong differences in the genetic origin of transcript and protein abundance.

A-B. Map of the SNP- (A) and CNV- (B) pQTL. The x-axis is the QTL position on the genome and the y-axis the position of the affected gene on the genome. The x and y-axis numbers represent the 16

chromosomes of *S. cerevisiae*. **C.** Effect size difference between the local and distant pQTL for the SNP (p-value= $6.2x10^{-8}$) and CNV pQTL (p-value=0.38). **D.** Distribution of the SNP-pQTL and SNP-eQTL hotspots along the genome. The y-axis represents the percentage of the 629 genes that by each hotspot (defined as a 20 kb window containing 4 or more distinct SNP).

Discussion

Quantifying transcripts and proteins expressed in a large natural population is fundamental for having a better understanding of the genotype-phenotype relationship. In this study, we have quantitatively analyzed the proteome of 942 natural isolates of *S. cerevisiae*, allowing in-depth exploration of protein abundance and precise characterization of the genetic origins of its variation at the species level.

The *S. cerevisiae* species is characterized by a complex population structure, with domesticated and wild subpopulations (Peter et al., 2018). Structured populations are also observed in a large number of other species, such as humans, and their impact on the proteome remains unexplored. In our dataset, the population structure had no significant impact on the proteomic landscape. This observation is consistent with previous results obtained with the transcriptomes of *S. cerevisiae* isolates (Caudal et al., 2023; Kita et al., 2017). In fact, most subpopulations are characterized by specific signatures related to a small set of genes but not to a general pattern.

This dataset allowed us to have better insight into the architecture of the species-wide proteome variation. First, we found that the co-expression network captures main biological functions and is globally conserved across the species. Second, we detected differential protein expression signatures specific to subpopulations, reflecting an adaptation to specific ecological conditions, such as domesticated environments. Similar expression signatures can be also observed using transcriptomic data (Caudal et al., 2023; Hodgins-Davis et al., 2012), highlighting that gene expression plasticity at both levels is a key mechanism of environmental adaptation.

The species-wide proteomes and transcriptomes obtained in the same condition represent a unique opportunity to compare the gene regulation at both levels. The overall agreement between protein and transcript within each isolate appears to be high and this in the whole population, showing again that very abundant transcripts generally lead to very abundant proteins and vice versa (Battle et al., 2015; Becker et al., 2018; Edfors et al., 2016; Gautier et al., 2016; Moritz et al., 2019; Ponnala et al., 2014; Salovska et al., 2020; Wang et al., 2019; Wilhelm et al., 2014; Zhang et al., 2014). However, our data allow for the first time to have an accurate estimation of the correlation per gene at the population level and we found that this gene-wise correlation is very weak with a median of 0.165, which is lower than most previous estimates based on much smaller human and mice populations (Chick et al., 2016; Ghazalpour et al., 2011; Jiang et al., 2020; Upadhya and Ryan, 2022; Wang et al., 2019). Consistent with this result, genome-wide association studies also highlighted that SNPs related to variation in protein (pQTL) and transcript (eQTL) levels poorly overlap (3.6%), with mostly common local

QTL. This result is consistent with one of the first eQTL/pQTL comparisons (Foss et al., 2007) but unlike other studies, showing a higher overlap (Albert et al., 2014). However, we should emphasize that we were not able to map the genetic basis of the entire *S. cerevisiae* proteome and therefore the eQTL/pQTL overlap might be biased and underestimated.

Mechanistically, our results suggest that the regulation of protein degradation has an impact on the variation of the proteome, and therefore on its genetic basis. Proteins with a high turnover rate will be more affected by proteome-specific regulation and will therefore show a weaker correspondence with the transcriptome. Conversely, proteins with a low turnover rate are more likely to be impacted by variation in transcript abundance. They will therefore likely reflect variation in mRNA abundance.

Although mass spectrometers are highly sensitive, it should be noted the limitation that proteomic methods are biased towards quantification of highly abundant proteins. Indeed, the fraction of the proteome quantified constitutes the vast majority of the total proteomic mass of a cell and is enriched for essential genes as well as in genes most connected in functional networks. Our dataset captures many of the fundamental processes. Yet, results related to low abundant proteins are missed by this approach.

Overall, our study clearly highlights that the dependency between transcript and protein levels is complex, pointing to the importance of post-transcriptional regulation of protein abundance. Proteome and transcriptome are indeed two distinct layers of gene regulation, which need to be further explored to understand the genotype-phenotype relationship. As gene function is ultimately executed by the proteome, while mRNA is the messenger, more proteomic approaches will be needed to create a better understanding of the phenotypic diversity. Our study provides a first species-wide insight into the genetics that underlies both proteome and transcriptome diversity in a natural population.

Materials and methods

Cultivation of library for proteomics

The yeast isolate collection was grown on agar containing synthetic complete medium (SC; 6.7 g/L yeast nitrogen base (MP Biomedicals, Cat#114027512-CF), 20 g/L glucose, 2 g/L synthetic complete amino acid mixture (MP Biomedicals, Cat#114400022)). After 48 h, colonies were inoculated in 200 µL SC liquid medium using a Singer Rotor and incubated at 30 °C overnight without shaking. These pre-cultures were then mixed by pipetting up and down, and diluted 20x by transferring 80 µL per culture to deep-well plates pre-filled with 1.55 mL SC liquid medium and one borosilicate glass bead per well. Plates were sealed with a permeable membrane and grown for 8 h at 1000 rpm, 30°C to exponential phase. The optical density at harvest was measured using an Infinite M Nano (Tecan). Per culture, 1.4 mL of cell suspension were harvested by transferring into a new deep-well plate and subsequent centrifugation (3,220 x g, 5 min, 4°C). The supernatant was removed by inverting the plates. Cell pellets were immediately cooled on dry ice and stored at −80°C.

Sample preparation

Samples for proteomics were prepared as previously described (Messner et al., 2022a, 2020; Muenzner et al., 2022). In brief, samples were processed in 96-well format, with lysis being achieved by beat beating using a Spex Geno/Grinder and 200 µL of lysis buffer (100 mM ammonium bicarbonate, 7 M urea). Samples were reduced and alkylated using DTT (20 µL, 55 mM) and iodoacetamide (20 µL, 120 mM), respectively, diluted with 1 mL 100 mM ammonium bicarbonate, and 500 µL per sample were digested using 2 µg Trypsin/LysC (Promega, Cat#V5072). After 17 h of incubation at 37°C, 25 µL 20% formic acid were added to the samples, and peptides were purified using solid-phase extraction as described previously (Messner et al., 2020). Eluted samples were vacuum-dried and subsequently dissolved in 70 µL 0.1% formic acid. An equivoluminal pool of all samples was generated to be used as technical controls (QCs) during MS measurements. The peptide concentration of this pool was determined using a fluorimetric peptide assay kit (Thermo Scientific, Cat#23290). Peptide concentrations per sample were estimated by multiplying the optical density recorded at harvest with the ratio between pool peptide concentration and the median at-harvest optical density.

LC–MS/MS measurements

In brief, peptides were separated on a 3-min high-flow chromatographic gradient and recorded by mass spectrometry using Scanning SWATH (Messner et al., 2021) using an online coupled 1290 Infinity II LC system (Agilent) - 6600+ TripleTOF platform (Sciex). 5 µg of sample were injected onto a reverse phase HPLC column (Luna®Omega 1.6µm C18 100A, 30 × 2.1 mm, Phenomenex) and resolved by gradient elution at a flow rate of 800 µL/min and column temperature of 30 °C. All solvents were of LC-MS grade. The gradient program used 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B) and was as follows: 1% to 40% B in 3 min, increase to 80% B at 1.2 mL over 0.5 min, which was maintained for 0.2 min and followed by equilibration with starting conditions for 1 min. For mass spectrometry analysis, the scanning swath precursor isolation window was 10 m/z; the bin size was set to 1/5th of the window size, the cycle time was 0.7 s, the precursor range 400 m/z to 900 m/z, the fragment range 100 m/z to 1500 m/z as previously described(Messner et al., 2021). An IonDrive TurboV source was used with ion source gas 1 (nebulizer gas), ion source gas 2 (heater gas), and curtain gas set to 50 psi, 40 psi and 25 psi respectively. The source temperature was set to 450 °C and the ion spray voltage to 5500 V.

Data processing

The mass spectrometry files were processed following the approach previously described (Muenzner et al., 2022). Briefly, an experimental spectral library obtained using the S288c was filtered to reduce the search space to peptides well shared across the strains. This library was then used with the software DIA-NN (Demichev et al., 2020) (Version 1.8) and the following parameters: missed cleavages: 0, Mass accuracy: 20, Mass accuracy MS1: 12, scan windows: 6. The option 'MBR' was used to process the data. As the peptides selected were not necessarily present ubiquitously in all the strains, an additional step was required to remove false positives (entries where a peptide is detected in a strain where it should be absent). This represents only ~1% of the total entries of the report.

Samples and entries with insufficient MS2 signal quality (< 1/3 of median MS2 signal) and with entries with Q.Value (> 0.01), PG.Q.Value (> 0.01), Global.Protein.Q.Value (> 0.01), Global.PG.Q.Value (> 0.01) were removed. A similar threshold was applied to Lib.PG.Q.Value and Lib.Q.Value to account for the MBR option used. Non-proteotypic precursors were also excluded. Outlier samples were detected based on the total ion chromatograms (TIC) and number of identified precursors per sample (Z-Score > 2.5) and were excluded from further

analysis. Precursors were filtered according to their detection rate in the samples, with a threshold set at 80% of detection rate across all the strains, while precursors with a coefficient of variation (CV) above 0.3 in the QC samples were excluded. The CVs of QCs and wild isolates samples were calculated and had a median CV of 15.15% and 34.21%, respectively (Figure S14; table S22). Batch correction was carried out at the precursor level using median batch correction, which consists in bringing the median value of the precursors in the different batches to the same level. Proteins were then quantified from the peptide abundance using the maxLFQ (Cox et al., 2014) function implemented in the DIA-NN R package. The resulting dataset consists of 630 proteins for 942 strains. We imputed the missing value for further exploration using the KNN imputation method from the *impute* R package (Hastie et al., 2022).

Combination of transcriptomic and proteomic data

Unless specified, all the analysis performed below were conducted using R version 4.1.2. The transcriptomic data was generated previously (Caudal et al., 2023). We used the log2 transcript per million (TPM) data, where the overlap with proteomic data was encompassing 629 genes across 889 isolates, for the genome wide association studies (see later for the method). For the exploration of gene expression variation, subpopulation related DEG and gene expression network, we used the variance stabilized data obtained directly from the log2 TPM data. In this case one gene was removed from the analysis and the reference strain data was not considered, which resulted in an overlap of 628 genes across 888 isolates. To only focus on real expression variation difference between the expression layers, we normalized the proteomic and transcript abundance using quantile normalization. Unless specified, all the analyses described below use the quantile normalized transcriptomic and proteomic data. We recomputed the raw protein abundance coefficient of variation (CV) of each gene by dividing the standard deviation by the mean (using the non-normalized abundance) and transformed it to a percentage. Based on the CV, we performed a functional exploration by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) using the *fgsea* R package (Korotkevich et al., 2021) for the gene ontology annotation (Ashburner et al., 2000; Gene Ontology Consortium, 2021) to detect cellular pathways with a conserved regulation across the population. The within- and across-gene mRNA-protein correlation was performed for each gene or each isolate using a Spearman correlation test. We selected the genes with a mRNA-protein correlation index higher than 0.42 (> 95% percentile) and performed gene ontology (GO) enrichment analysis using the biological process (BP) database using the *topGO* R package (Alexa and Rahnenfuhrer, 2022). For the

GO analysis looking at the functional enrichment present in the 630, the gene list reference was the genes encompassed in the transcriptomic data (Caudal et al., 2023). The others GO analyses used the 628 genes as the reference list. All the others GO analyses were performed using the same procedure, unless specified.

Expression variation exploration

We measured the strength of protein and transcript abundance variation using several methods. We computed an absolute transformed *Log2(fold change)* value ($|Log_2(FC)|$) where in each isolate pairwise comparison (ex: strain A vs strain B) and for each gene, we performed:

$$|log_2(\frac{normalized\ abundance\ of\ gene\ X\ in\ strain\ A}{normalized\ abundance\ of\ gene\ X\ in\ strain\ B})|$$

Briefly, the more this value increases, the more different is protein abundance between two isolates for a specific gene. We also computed a pairwise spearman correlation between the isolates using the normalized proteomic and transcriptomic data. We also gathered the Euclidean distances between the expression profiles of each isolate, as well as the gene expression variance per gene.

We explored the post-transcriptional buffering phenomenon using an approach based on the computation of expression trees (Wang et al., 2020). First, on both protein and transcript normalized abundances, we constructed a neighbor-joining tree based on the Euclidean distance between each isolate. We computed the total branch length of these two trees and created a ratio of the proteome tree length on the transcriptome tree length. The ratio was equal to 0.93 which is line with the difference in Euclidean distance between the transcriptome and proteome. We performed 100 bootstrapping tests and used the resulting branch lengths to test the difference between the proteome and the transcriptome tree. We sought to check if some cellular pathways tended to be more affected by the post transcriptional buffering phenomenon. To do so, we gathered a reduced biological process GO annotation by computing the similarity between each GO term using the *rrvgo* R package and the 'Resnik" method (Sayols, 2022). We discarded terms that are at least 50% overlapping with another term and the terms encompassing no more than 5 genes, which resulted in a list of 101 terms. For each of these terms, we performed the same tree exploration, but this time with the genes encompassed by each term. We obtained therefore 101 tree length ratios. We selected the terms displaying a ratio lower than 0.93 or higher than 1, and for which the total branch length between the proteome and the transcriptome

was significantly different after 10 bootstrapping steps (Bonferroni corrected Wilcoxon test p-value < 0.001).

Transcriptome and proteome landscape exploration

We sought to check if the genetic structure of the population had an impact on the transcriptome and proteome structure. We obtained the genetic distances from (Peter et al., 2018) between pairs of isolates and compared them to the pairwise isolate correlation (Spearman correlation test) obtained with the normalized transcript or protein abundances. We also used both normalized protein and mRNA abundance data to perform principal component analysis (PCA) using the *prcomp* function from the *stats* R package. For the 2 PCA (transcriptomic and proteomic), we plotted the 6 first principal components (PC) together (PC1-PC2, PC3-PC4 and PC5-PC6) and looked for eventual grouping according to the subpopulation as defined previously(Peter et al., 2018). We then computed a Weighted Gene Co-Expression Network Analysis (WGCNA) using the *WGCNA* R package (Langfelder and Horvath, 2008) to detect co-expression module in both mRNA and peptide normalized abundance. To do so, we generated a Topological Overlap Matrix (TOM) using the *blockwiseModules* function. The TOM were calculated based on a signed adjacency matrix with the power of 9 for the mRNA abundance data and 5 for the peptide abundance data. The *blockwiseModules* automatically detected the co expression modules by generating a clustering from a dissimilarity matrix (1-TOM) using the following option: *detectCutHeight* = 0.995; *minModuleSize* = 30. This resulted in the detection of 5 and 7 transcriptome and proteome modules respectively. We computed an overrepresentation analysis for each co-expression module with the GO terms as annotation and using the *mod_ora* function from the *CEMiTool* R package (Russo et al., 2018) and used the most representative GO terms as the final annotation for each detected module. The two co-expression networks were generated for plotting by computing an adjacency matrix from the TOM matrix (generated previously) and ultimately plotted using the *ggnet2* function from the *GGally* R package.

Transcriptome and proteome differentially expressed gene detection

We used the normalized protein abundance to detect subpopulation- specific (Peter et al., 2018) differentially expressed proteins (DEPs). The goal was to detect either over- or underexpressed genes by comparing the normalized expression of all the isolates from a subpopulation against the rest of the population using a Wilcoxon test for each gene. The p-value of the test was

corrected using a Bonferroni correction with the *p.adjust* function in R. A gene was considered as differentially expressed if the corrected p-value of the Wilcoxon test was below 0.05. We computed as well a log2 transformed fold change (log2(FC)) value for each gene in each subpopulation using the mean expression of the subpopulation divided by the mean expression of the rest of the population. To further characterize the detected DEPs, we performed a functional exploration using GSEA (with the *fgsea* function from the *fgsea* R package) using the log2(FC) value from the DEP exploration as score rankings. In order to have a global view of the pathways that were significantly differentially expressed in each subpopulation, we used the 16 co-expression modules detected and defined previously using the population transcriptome data (Caudal et al., 2023) as biological function annotations for the GSEA. We performed the same procedure but this time comparing the domesticated against the wild isolate using the clade wise annotation from (Peter et al., 2018). This time, the test was performed on both normalized protein and transcript abundances.

Proteome and transcriptome genome-wide association studies

We computed GWAS with a linear mixed model-based method as described previously (Caudal et al., 2023; Peter et al., 2018) using FaST-LMM (Lippert et al., 2011). We performed the GWAS using either the transcriptome log2 transformed TPM data or the protein abundance. For each dataset, we performed two separated GWAS, one based on SNP as genotype, and one based on the CNV as genotype. The SNP GWAS was run with total of 84,633 SNP displaying a minor allele frequency (MAF) > 5% and that were not located in the telomeric regions (< 20kb away from the chromosome ends). The CNV GWAS was run on a total of 1,019 CNV (MAF > 5%). We used the SNP matrix for both SNP and CNV GWAS, thus evaluating the kinship between the isolate to account for the population structure. We set a phenotype-specific p-value threshold using 100 permutation tests where the phenotypes were randomly permutated between the isolates. We use the 5% lowest p-value quantile from these permutation tests to define the significance threshold. We finally scaled the significance thresholds of the CNV GWAS to account for the size difference between the SNP and CNV matrices.

Regarding the SNP GWAS, the detected QTL were filtered to avoid false positives detection due to linkage disequilibrium among the SNP as described previously (Caudal et al., 2023). This resulted in the filtration of 81 eQTL and 131 pQTL (out of respectively 677 and 729 QTL). The QTL were considered as “local” QTL when they were located 25 kb around their affected

phenotype. We also sought to detect QTL hotspots in both transcriptome and proteome GWAS. We defined a hotspot as a concentration of at least 4 QTL in a 20 kb window.

We compared the protein turnover rate (Muenzner et al., 2022) of obtained on 619 proteins encompassed in our dataset to see whether turnover rate had an impact on the overlap between SNP-eQTL and SNP-pQTL. This data comprises protein degradation rates for 1,836 gene across 55 natural isolates. We computed an average turnover rate per gene and used this value to compare the level of protein degradation of the protein with or without an overlapping QTL.

Supplementary Material

Supplementary tables available at:

https://www.dropbox.com/s/ukah0o3q2b8e4pw/Sup_tables.xlsx?dl=0

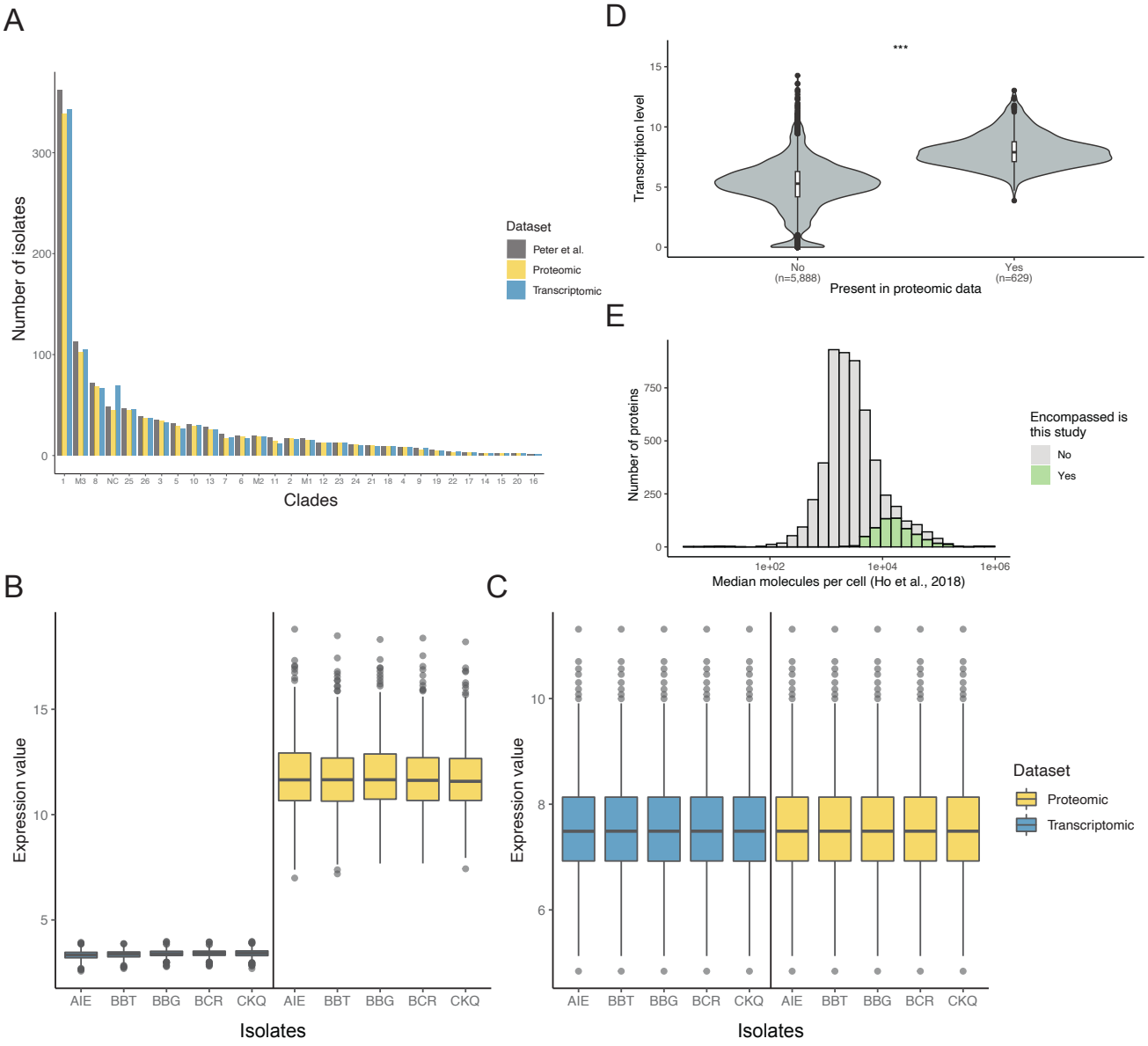


Figure S1. **Description of the characteristics and the normalization of the population proteome.** (A) number of isolates encompassed in the proteomic datasets, in the transcriptomic dataset (Caudal et al., submitted) and in the overall population (Peter et al., 2018). The x-axis corresponds to the clades (or subpopulations) as defined previously (Peter et al., 2018). (B, C) Expression values of 5 randomly selected isolates for protein and transcript abundance before (B) and after (C) quantile normalization. (D) mRNA levels of the gene encompassed or not in the proteomic data (*** = p-value < 2.2×10^{-16}, Wilcoxon test). (E) Protein levels (as defined in Ho et al., 2018) of the genes encompassed by our proteomic data.

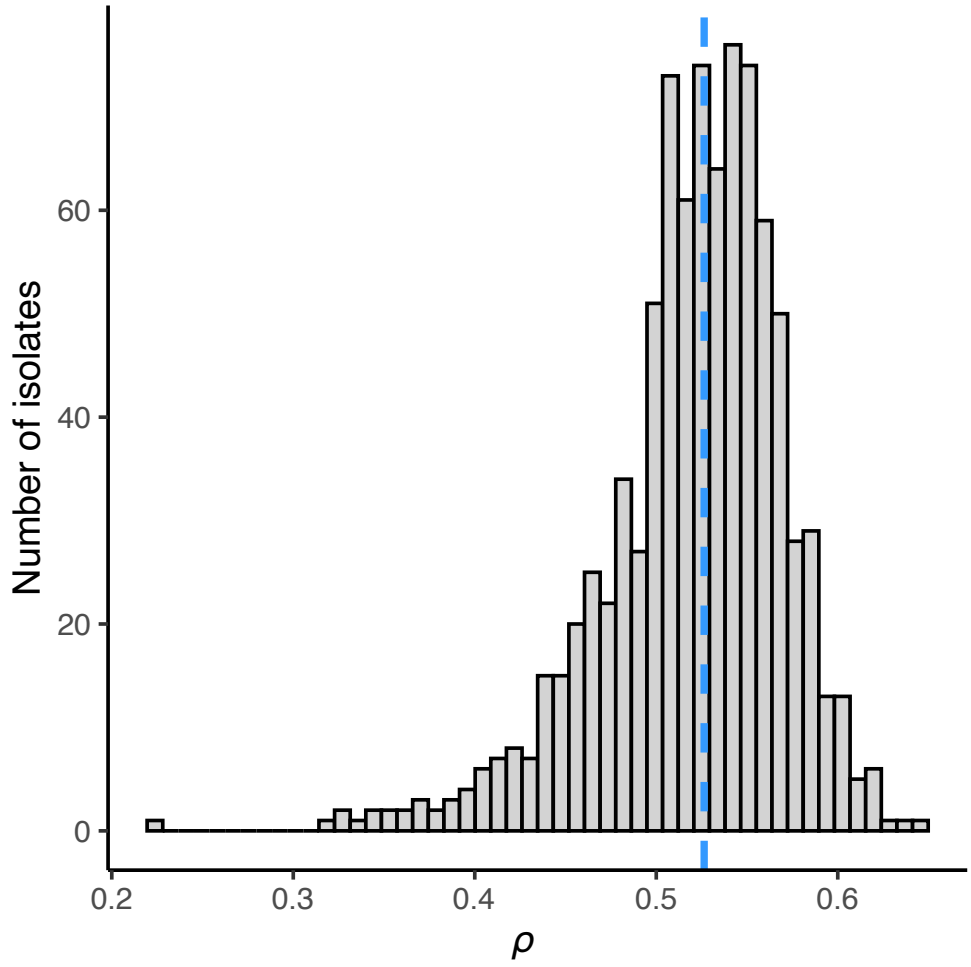


Figure S2. Across-gene correlation.

mRNA-protein correlation in each isolate (across-gene correlation). The blue line represents the median (0.53).

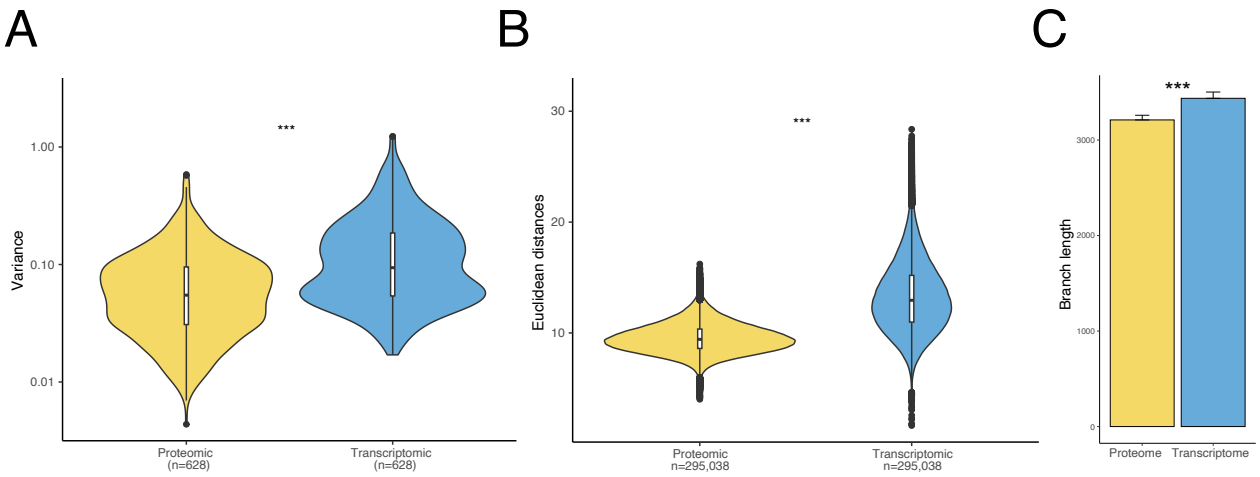


Figure S3. Detection of the post-transcriptional buffering.

(A) Comparison between the gene-wise protein and transcript normalized abundance variance (*** = p-value < $2.2x10^{-16}$, Wilcoxon test). (B) Euclidean distances between each isolate using the protein or transcript normalized abundance (*** = p-value < $2.2x10^{-16}$, Wilcoxon test). (C) Branch length difference between the proteome and the transcriptome-based tree. The error bars correspond to 100 bootstrapping steps. We used the bootstrap values to test if the difference in branch length is significant between the two trees (*** = p-value < $2.2x10^{-16}$, Wilcoxon test).

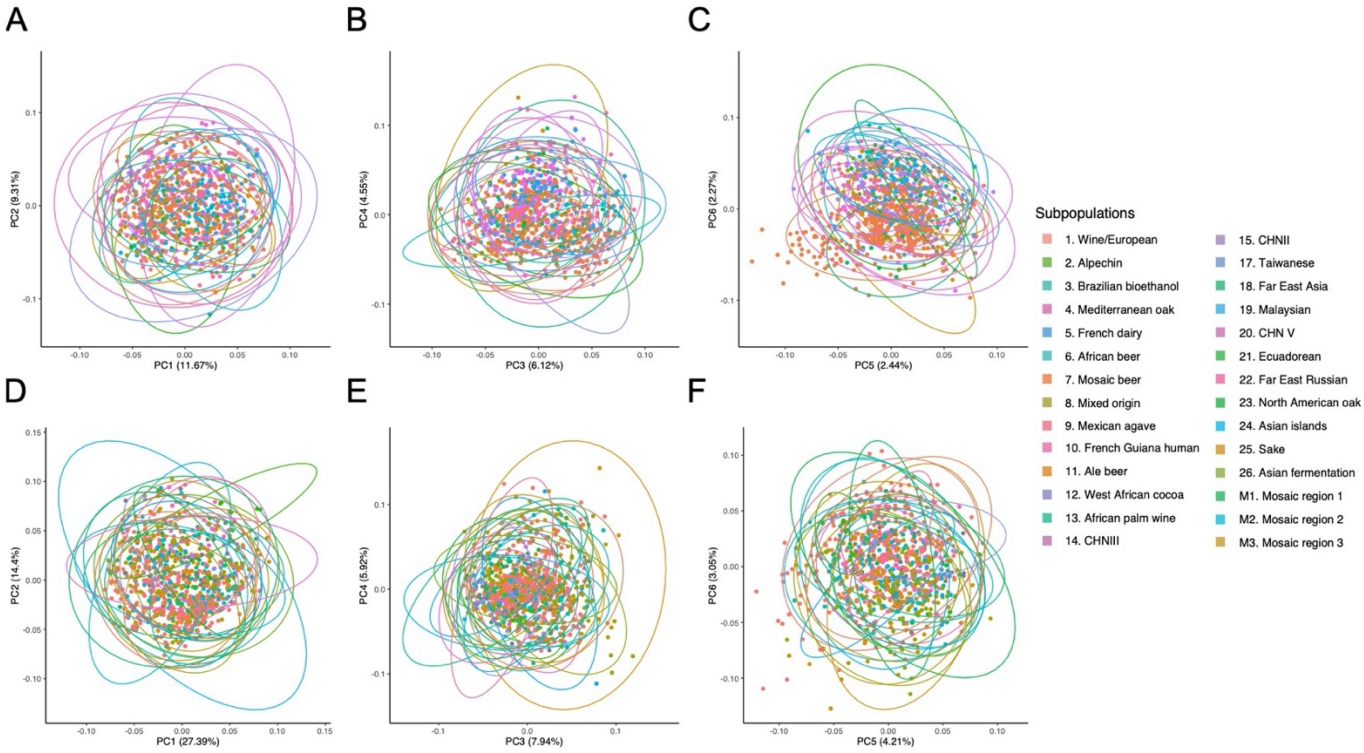


Figure S4. Population structure is not reflected on the proteome based PCA.

PCA using protein (A, B, C) or transcript (D, E, F) abundance. The 6 first PC are plotted together, and the colors correspond to the subpopulations (clades).

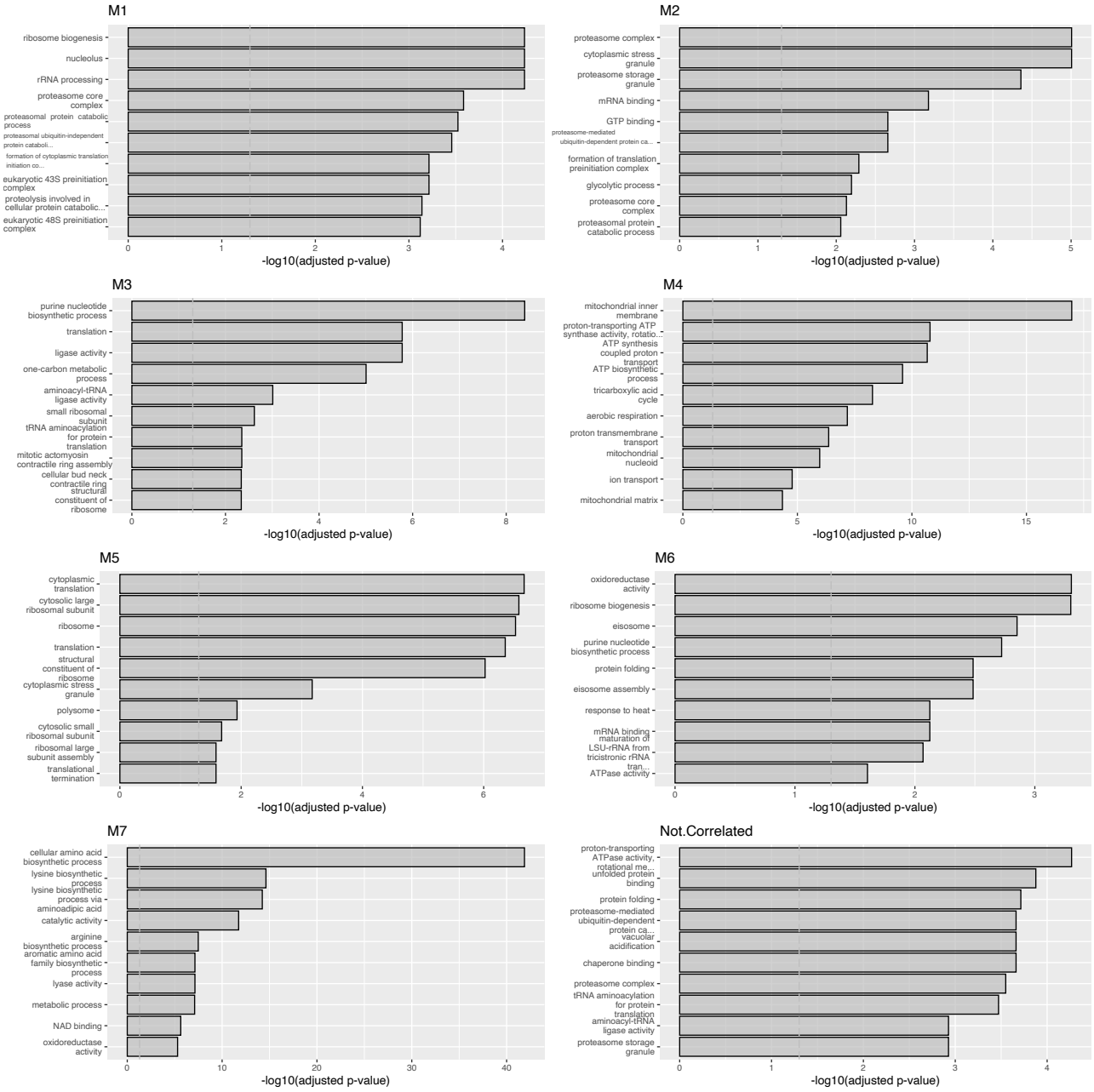


Figure S5. Functional exploration of the proteome WGCNA modules.

Functional enrichment of each co-expression module detected using WGCNA on protein abundance data. The enrichment was performed using the CEMiTool package. The dotted lines on each graph represent the significance threshold.

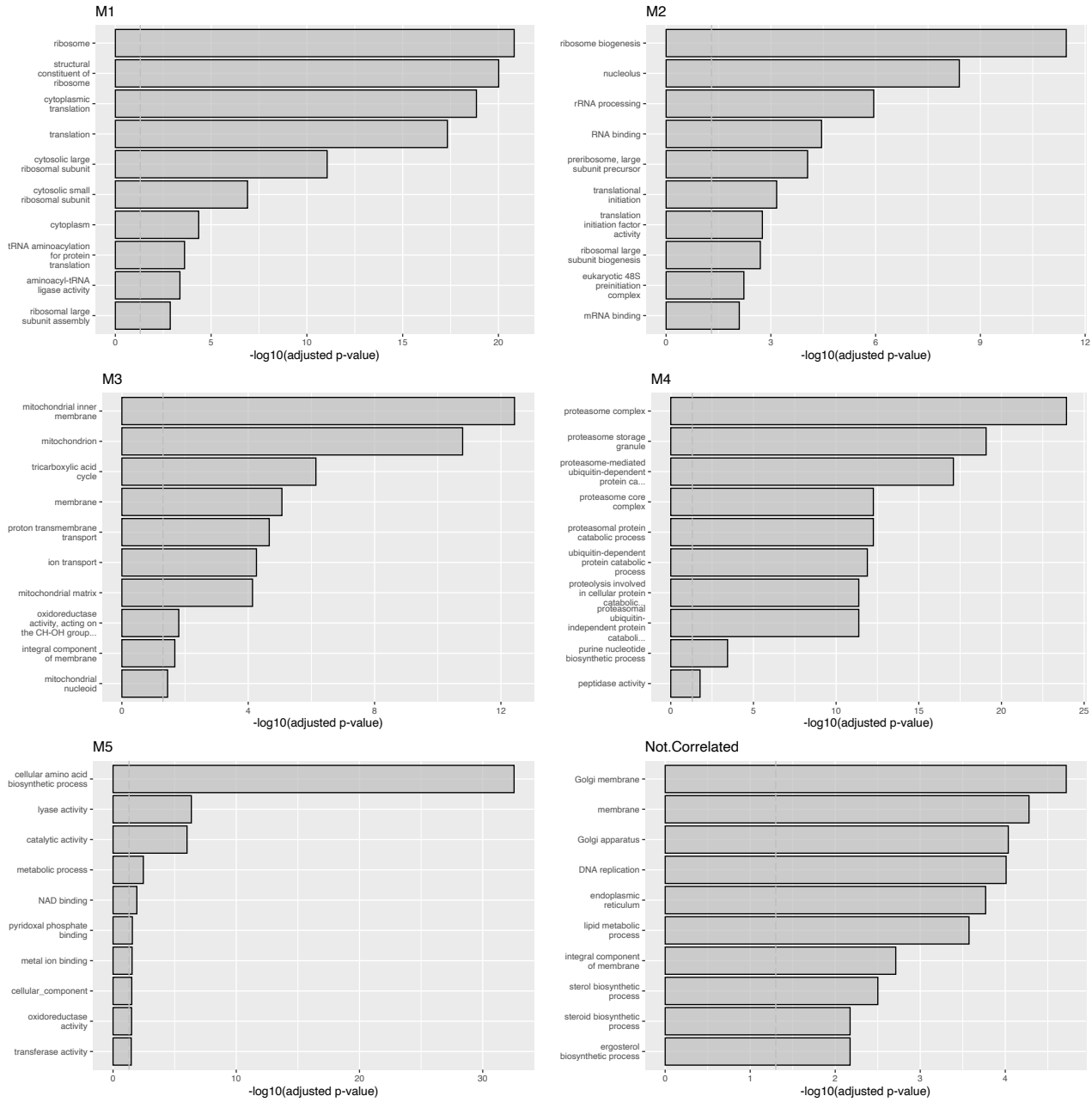


Figure S6. **Functional exploration of the transcriptome WGCNA modules.**

Functional enrichment of each co-expression module detected using WGCNA on transcript abundance data. The enrichment was performed using the CEMiTool package. The dotted lines on each graph represent the significance threshold.

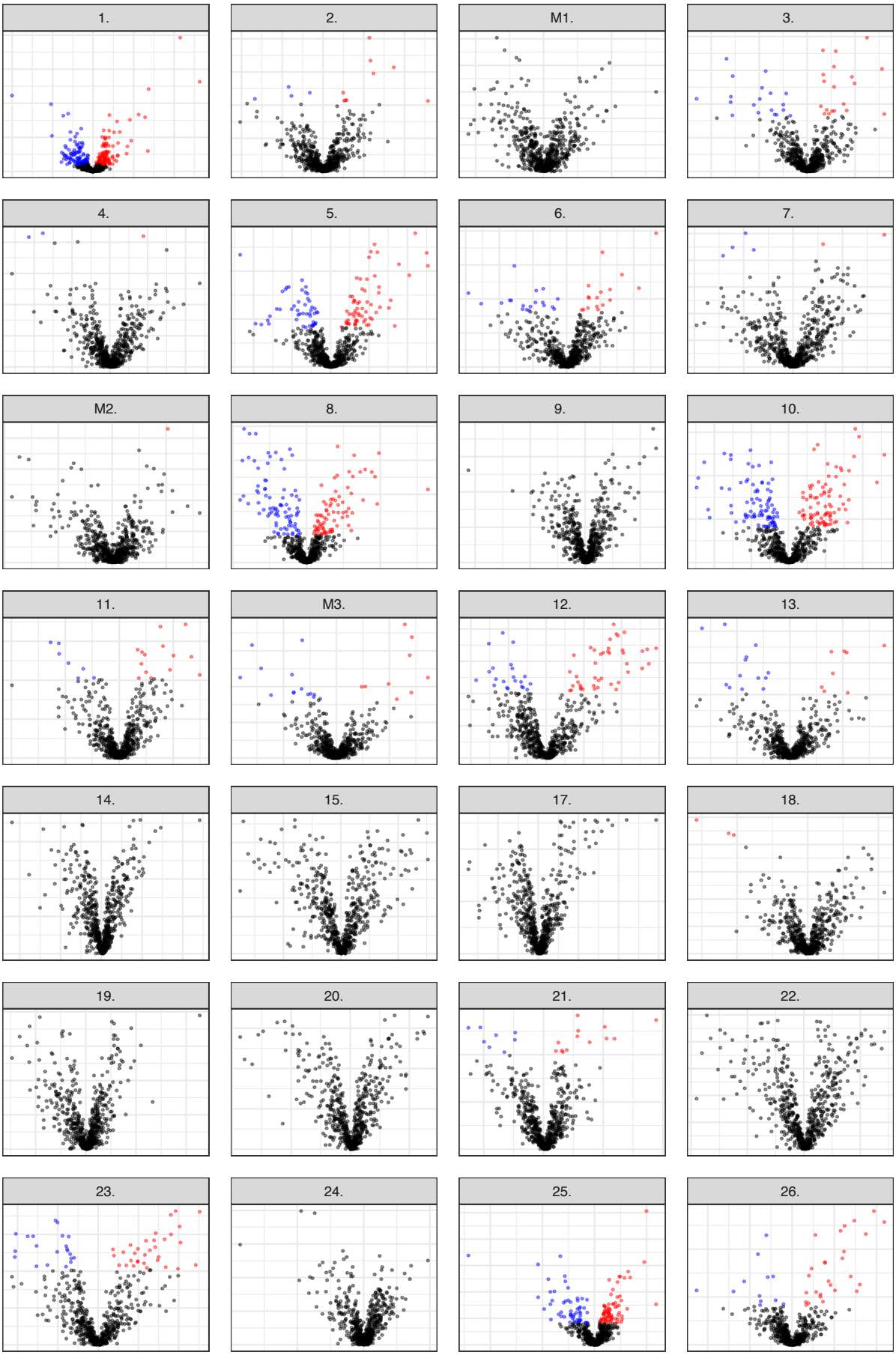

1.
2.
M1.
3.
4.
5.
6.
7.
M2.
8.
9.
10.
11.
M3.
12.
13.
14.
15.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.

Figure S7. DEPs detected in each subpopulation.

Volcano plots for each subpopulation highlighting the DEPs. The blue points correspond to under-expressed gene in a subpopulation while the red points correspond to over-expressed genes.

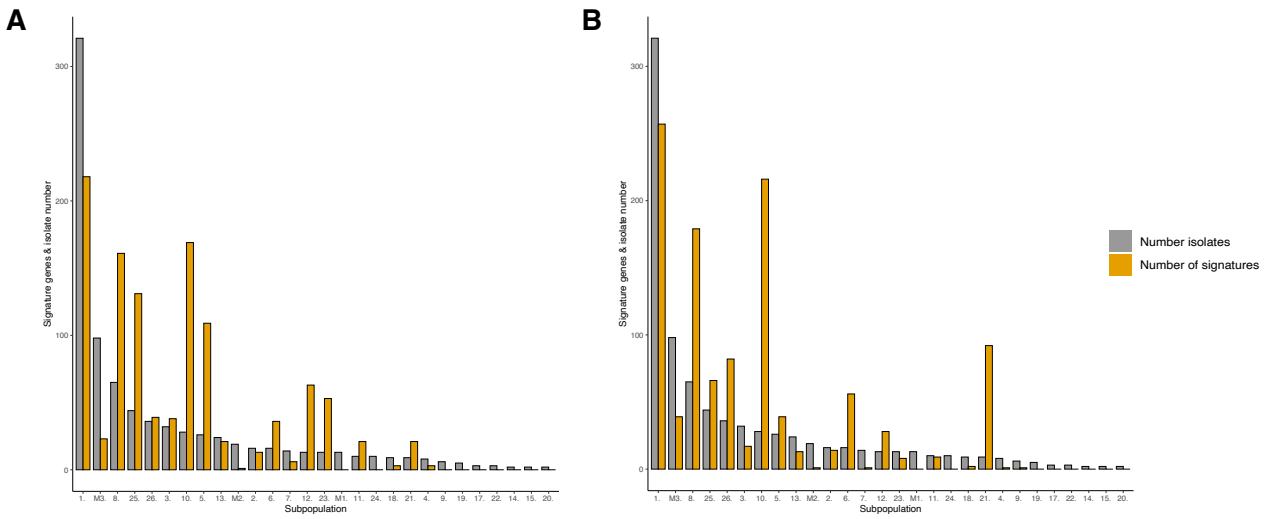


Figure S8. Number of DEP and differentially expressed transcripts.

(A, B) Number of proteome (A) and transcriptome (B) DEPs (or differentially expressed transcripts for the transcriptome) in each subpopulation together with the number of isolates in each subpopulation.

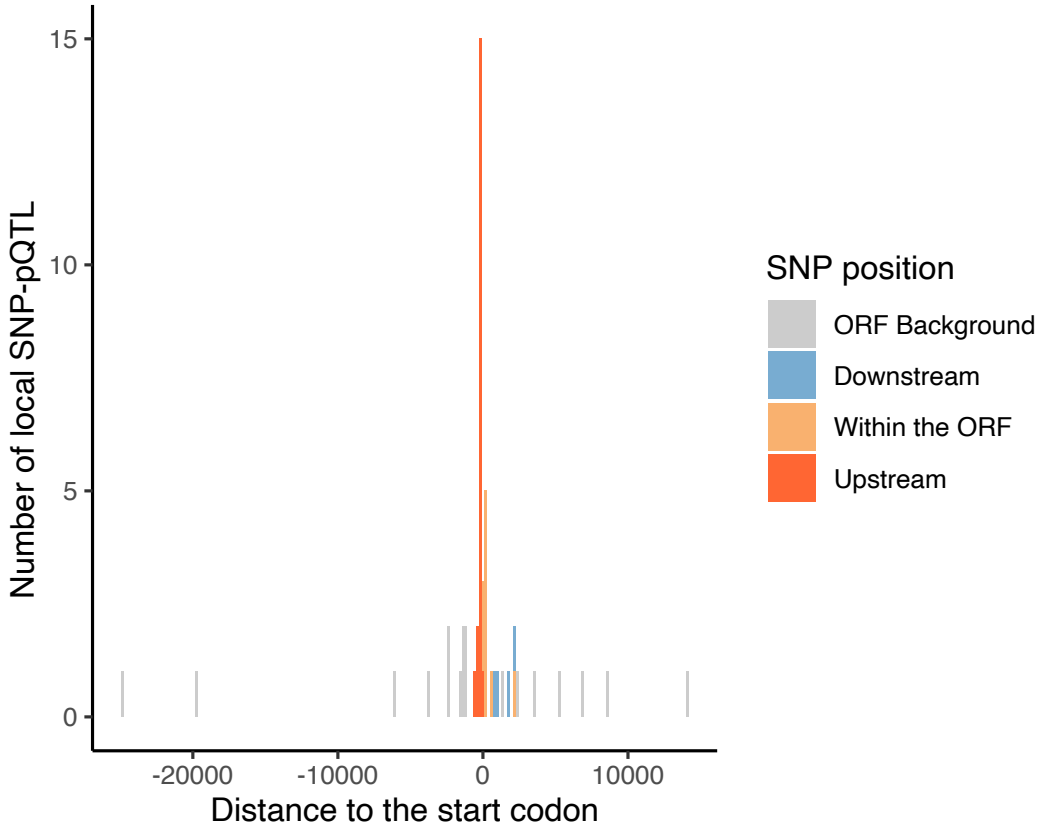


Figure S9. Location of the local SNP-pQTL.

Distribution of the local SNP-pQTL around the start codon of their target gene. Downstream pQTL correspond to QTL located between the stop codon and 200 bp after the stop codon, upstream correspond to pQTL located between the start codon and 1,000 bp before the start codon.

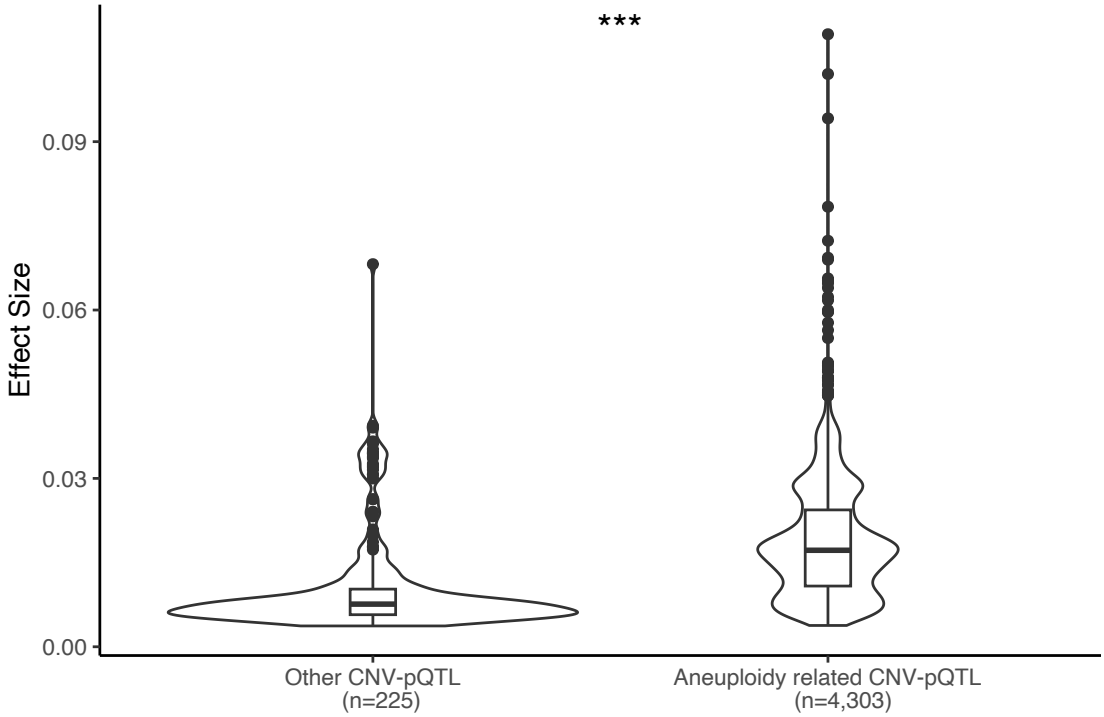


Figure S10. Aneuploidy related CNV-pQTL have a higher effect-size than the other CNV-pQTL. Difference in effect size between the aneuploidy related CNV-pQTL and the other CNV-pQTL (*** = p-value < $2.2x10^{-16}$, Wilcoxon test).

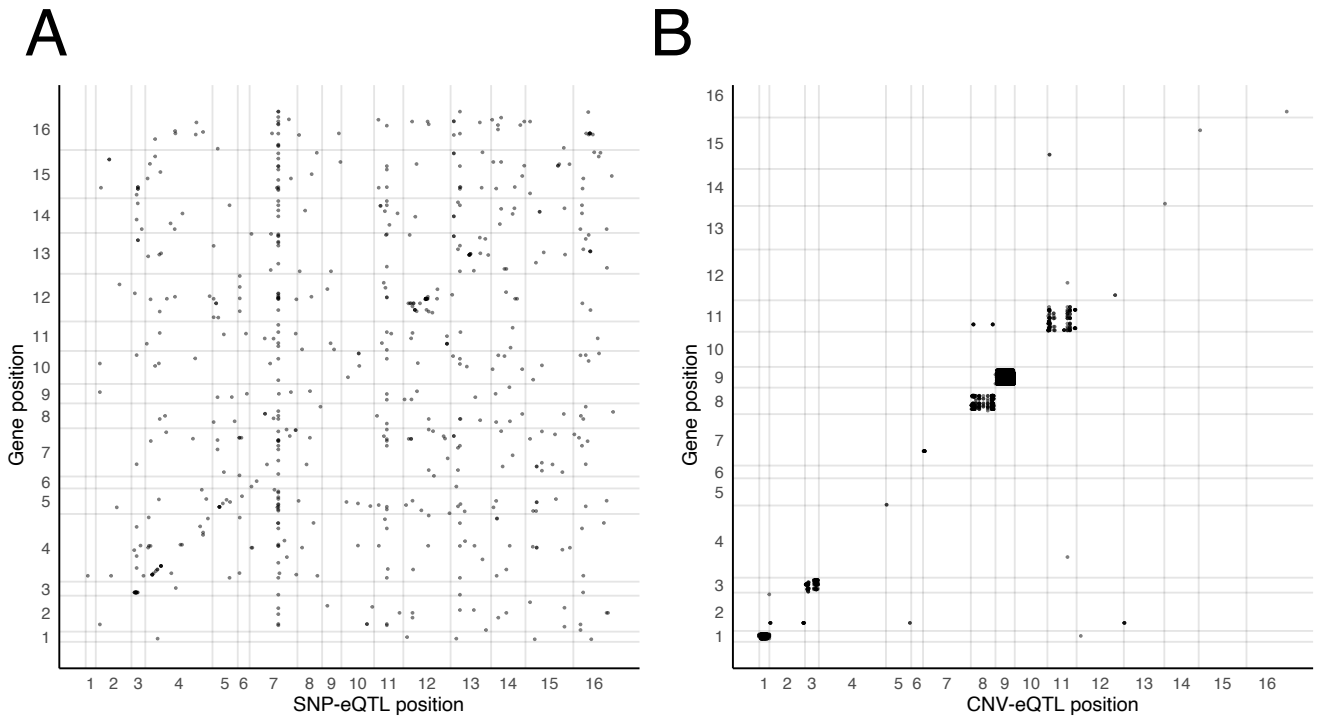


Figure S11. Genomic location of the SNP- and CNV-eQTL.

(A, B) Map of the SNP (A) and CNV (B) eQTL. The x-axis is the QTL positions on the genome and the y-axis the position of the affected genes on the genome. The x and y-axis numbers represent the 16 chromosomes of *S. cerevisiae*.

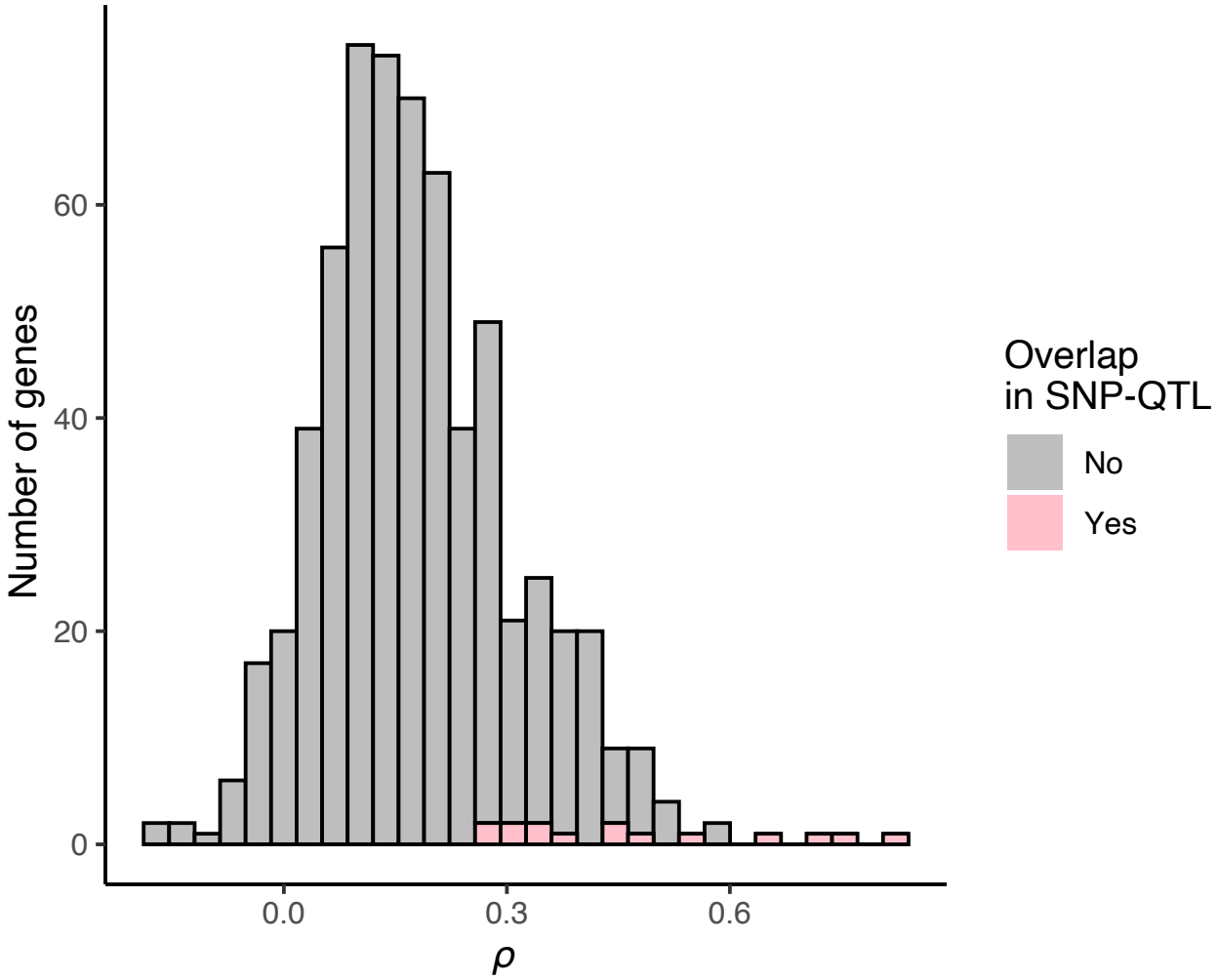


Figure S12. The genes with an overlapping SNP-QTL tend to have a high within-gene mRNA-protein correlation.

Within-gene correlation coefficients (Spearman correlation test) between the proteome and the transcriptome. The genes with an overlapping SNP-pQTL and SNP-eQTL are highlighted in pink.

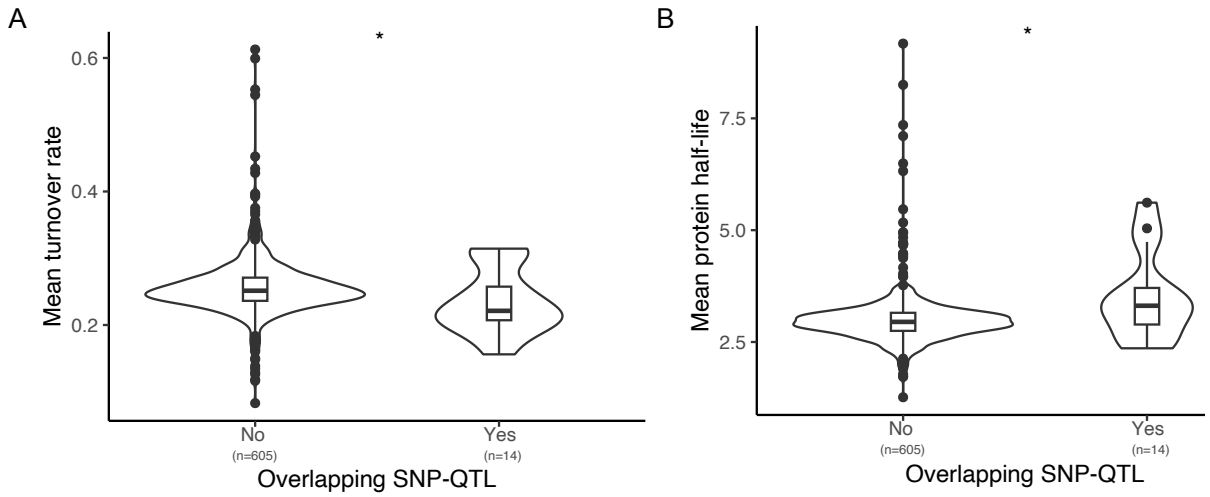


Figure S13. Turnover rate and half-life of the proteins with or without an overlapping SNP-QTL.

(A, B) The turnover rates (A) and protein half-life values (B) were obtained from Muenzner et al., 2022. The difference was tested using a Wilcoxon test (respective p-values = 0.028 and 0.026).

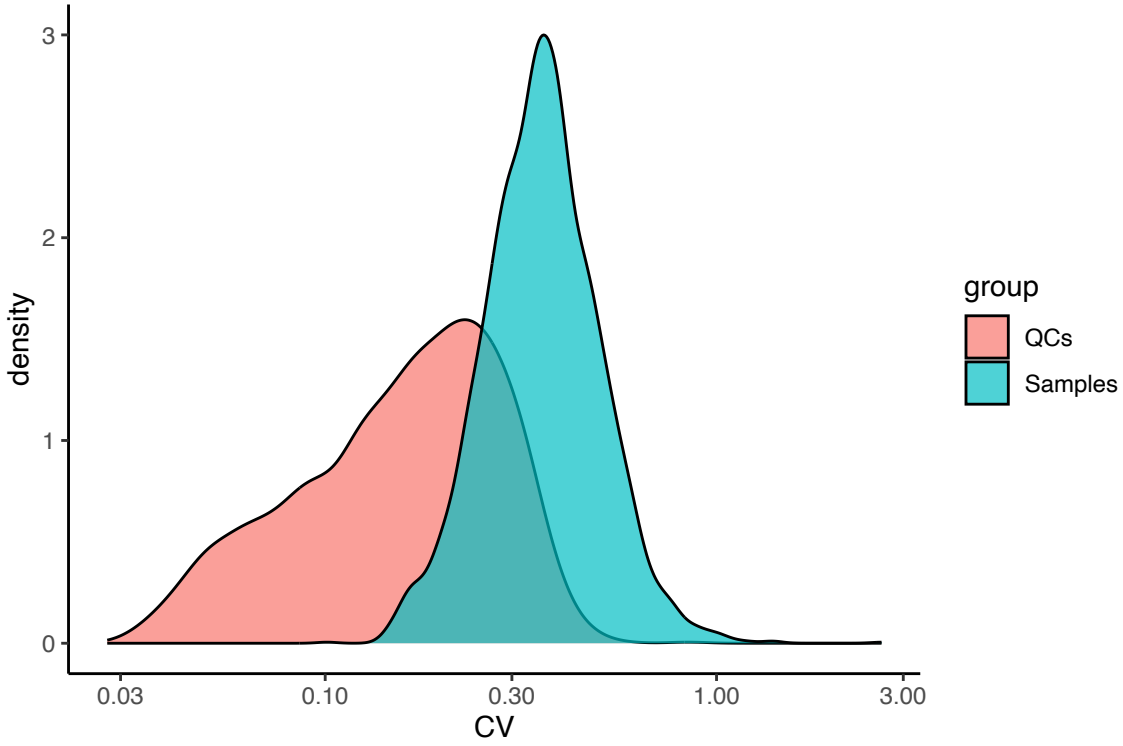


Figure S14. CV from the QCs and samples precursors.

The CV was computing either the QCs set or the sample set.

References

Alam, M.T., Zelezniak, A., Mülleder, M., Shliaha, P., Schwarz, R., Capuano, F., Vowinckel, J., Radmaneshfar, E., Krüger, A., Calvani, E., Michel, S., Börno, S., Christen, S., Patil, K.R., Timmermann, B., Lilley, K.S., Ralser, M., 2016. The metabolic background is a global player in Saccharomyces gene expression epistasis. Nat. Microbiol. 1, 1–10. https://doi.org/10.1038/nmicrobiol.2015.30

Albert, F.W., Bloom, J.S., Siegel, J., Day, L., Kruglyak, L., 2018. Genetics of trans-regulatory variation in gene expression. eLife 7, e35471. https://doi.org/10.7554/eLife.35471

Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212. https://doi.org/10.1038/nrg3891

Albert, F.W., Treusch, S., Shockley, A.H., Bloom, J.S., Kruglyak, L., 2014. Genetics of single-cell protein abundance variation in large yeast populations. Nature 506, 494–497. https://doi.org/10.1038/nature12904

Alexa, A., Rahnenfuhrer, J., 2022. topGO: Enrichment Analysis for Gene Ontology. https://doi.org/10.18129/B9.bioc.topGO

Archer, T.C., Ehrenberger, T., Mundt, F., Gold, M.P., Krug, K., Mah, C.K., Mahoney, E.L., Daniel, C.J., LeNail, A., Ramamoorthy, D., Mertins, P., Mani, D.R., Zhang, H., Gillette, M.A., Clauser, K., Noble, M., Tang, L.C., Pierre-François, J., Silterra, J., Jensen, J., Tamayo, P., Korshunov, A., Pfister, S.M., Kool, M., Northcott, P.A., Sears, R.C., Lipton, J.O., Carr, S.A., Mesirov, J.P., Pomeroy, S.L., Fraenkel, E., 2018. Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups. Cancer Cell 34, 396-410.e8. https://doi.org/10.1016/j.ccell.2018.08.004

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. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556

Auesukaree, C., 2017. Molecular mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation. J. Biosci. Bioeng. 124, 133–142. https://doi.org/10.1016/j.jbiosc.2017.03.009

Aydin, S., Pham, D.T., Zhang, T., Keele, G.R., Skelly, D.A., Paulo, J.A., Pankratz, M., Choi, T., Gygi, S.P., Reinholdt, L.G., Baker, C.L., Churchill, G.A., Munger, S.C., 2023. Genetic dissection of the pluripotent proteome through multi-omics data integration. Cell Genomics 0. https://doi.org/10.1016/j.xgen.2023.100283

Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., Gilad, Y., 2015. Impact of regulatory variation from RNA to protein. Science 347, 664–667. https://doi.org/10.1126/science.1260793

Becker, K., Bluhm, A., Casas-Vila, N., Dinges, N., Dejung, M., Sayols, S., Kreutz, C., Roignant, J.-Y., Butter, F., Legewie, S., 2018. Quantifying post-transcriptional regulation in the development of Drosophila melanogaster. Nat. Commun. 9, 4970. https://doi.org/10.1038/s41467-018-07455-9

Blevins, W.R., Tavella, T., Moro, S.G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., Carey, L.B., Albà, M.M., 2019. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Sci. Rep. 9, 11005. https://doi.org/10.1038/s41598-019-47424-w

Buccitelli, C., Selbach, M., 2020. mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 1–15. https://doi.org/10.1038/s41576-020-0258-4

Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonniere, E., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., 2023. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. https://doi.org/10.1101/2023.05.17.541122

Celińska, E., Nicaud, J.-M., 2019. Filamentous fungi-like secretory pathway strayed in a yeast system: peculiarities of Yarrowia lipolytica secretory pathway underlying its extraordinary performance. Appl. Microbiol. Biotechnol. 103, 39–52. https://doi.org/10.1007/s00253-018-9450-2

Chick, J.M., Munger, S.C., Simecek, P., Huttlin, E.L., Choi, K., Gatti, D.M., Raghupathy, N., Svenson, K.L., Churchill, G.A., Gygi, S.P., 2016. Defining the consequences of genetic variation on a proteome-wide scale. Nature 534, 500–505. https://doi.org/10.1038/nature18270

Cookson, W., Liang, L., Abecasis, G., Moffatt, M., Lathrop, M., 2009. Mapping complex disease traits with global gene expression. Nat. Rev. Genet. 10, 184–194. https://doi.org/10.1038/nrg2537

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., Mann, M., 2014. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ*. Mol. Cell. Proteomics 13, 2513–2526. https://doi.org/10.1074/mcp.M113.031591

Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., Ralser, M., 2020. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods 17, 41–44. https://doi.org/10.1038/s41592-019-0638-x

Dowell, R.D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D.A., Rolfe, P.A., Heisler, L.E., Chin, B., Nislow, C., Giaever, G., Phillips, P.C., Fink, G.R., Gifford, D.K., Boone, C., 2010. Genotype to Phenotype: A Complex Problem. Science 328, 469–469. https://doi.org/10.1126/science.1189015

Edfors, F., Danielsson, F., Hallström, B.M., Käll, L., Lundberg, E., Pontén, F., Forsström, B., Uhlén, M., 2016. Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol. Syst. Biol. 12, 883. https://doi.org/10.15252/msb.20167144

Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., Jensson, B.O., Zink, F., Halldorsson, G.H., Masson, G., Arnadottir, G.A., Katrinardottir, H., Juliusson, K., Magnusson, M.K., Magnusson, O.T., Fridriksdottir, R., Saevarsdottir, S., Gudjonsson, S.A., Stacey, S.N., Rognvaldsson, S., Eiriksdottir, T., Olafsdottir, T.A., Steinthorsdottir, V., Tragante, V., Ulfarsson, M.O., Stefansson, H., Jonsdottir, I., Holm, H., Rafnar, T., Melsted, P., Saemundsdottir, J., Norddahl, G.L., Lund, S.H., Gudbjartsson, D.F., Thorsteinsdottir, U., Stefansson, K., 2021. Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. 53, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w

Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D.H., Hedman, Å.K., Schork, A., Page, K., Zhernakova, D.V., Wu, Y., Peters, J., Eriksson, N., Bergen, S.E., Boutin, T.S., Bretherick, A.D., Enroth, S., Kalnapenkis, A., Gådin, J.R., Suur, B.E., Chen, Y., Matic, L., Gale, J.D., Lee, J., Zhang, W., Quazi, A., Ala-Korpela, M., Choi, S.H., Claringbould, A., Danesh, J., Davey Smith, G., de Masi, F., Elmståhl, S., Engström, G., Fauman, E., Fernandez, C., Franke, L., Franks, P.W., Giedraitis, V., Haley, C., Hamsten, A., Ingason, A., Johansson, Å., Joshi, P.K., Lind, L., Lindgren, C.M., Lubitz, S., Palmer, T., Macdonald-Dunlop, E., Magnusson, M., Melander, O., Michaelsson, K., Morris, A.P., Mägi, R., Nagle, M.W., Nilsson, P.M., Nilsson, J., Orho-Melander, M., Polasek, O., Prins, B., Pålsson, E., Qi, T., Sjögren, M., Sundström, J., Surendran, P., Võsa, U., Werge, T., Wernersson, R., Westra, H.-J., Yang, J., Zhernakova, A., Ärnlöv, J., Fu, J., Smith, J.G., Esko, T., Hayward, C., Gyllensten, U., Landen, M., Siegbahn, A., Wilson, J.F., Wallentin, L., Butterworth, A.S., Holmes, M.V., Ingelsson, E., Mälarstig, A., 2020. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat. Metab. 2, 1135–1148. https://doi.org/10.1038/s42255-020-00287-2

Fortelny, N., Overall, C.M., Pavlidis, P., Freue, G.V.C., 2017. Can we predict protein from mRNA levels? Nature 547, E19–E20. https://doi.org/10.1038/nature22293

Foss, E.J., Radulovic, D., Shaffer, S.A., Ruderfer, D.M., Bedalov, A., Goodlett, D.R., Kruglyak, L., 2007. Genetic basis of proteome variation in yeast. Nat. Genet. 39, 1369–1375. https://doi.org/10.1038/ng.2007.22

Gautier, E.-F., Ducamp, S., Leduc, M., Salnot, V., Guillonneau, F., Dussiot, M., Hale, J., Giarratana, M.-C., Raimbault, A., Douay, L., Lacombe, C., Mohandas, N., Verdier, F., Zermati, Y., Mayeux, P., 2016. Comprehensive Proteomic Analysis of Human Erythropoiesis. Cell Rep. 16, 1470–1484. https://doi.org/10.1016/j.celrep.2016.06.085

Gene Ontology Consortium, 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334. https://doi.org/10.1093/nar/gkaa1113

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P.-Z., Brewer, H., Weitz, K., Ii, D.G.C., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D.J., Smith, R.D., Lusis, A.J., 2011. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. PLOS Genet. 7, e1001393. https://doi.org/10.1371/journal.pgen.1001393

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A.P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J.H., Hempel, S., Herman, Z.,

Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kötter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., Johnston, M., 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391. https://doi.org/10.1038/nature00935

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R., 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19, 1720–1730. https://doi.org/10.1128/MCB.19.3.1720

Hastie, T., Tibshirani, R., Narasimhan, B., Chu, G., 2022. impute: impute: Imputation for microarray data. https://doi.org/10.18129/B9.bioc.impute

Ho, B., Baryshnikova, A., Brown, G.W., 2018. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Syst. 6, 192-205.e3. https://doi.org/10.1016/j.cels.2017.12.004

Hodgins-Davis, A., Adomas, A.B., Warringer, J., Townsend, J.P., 2012. Abundant Gene-by-Environment Interactions in Gene Expression Reaction Norms to Copper within Saccharomyces cerevisiae. Genome Biol. Evol. 4, 1061–1079. https://doi.org/10.1093/gbe/evs084

Huang, K.-L., Li, S., Mertins, P., Cao, S., Gunawardena, H.P., Ruggles, K.V., Mani, D.R., Clauser, K.R., Tanioka, M., Usary, J., Kavuri, S.M., Xie, L., Yoon, C., Qiao, J.W., Wrobel, J., Wyczalkowski, M.A., Erdmann-Gilmore, P., Snider, J.E., Hoog, J., Singh, P., Niu, B., Guo, Z., Sun, S.Q., Sanati, S., Kawaler, E., Wang, X., Scott, A., Ye, K., McLellan, M.D., Wendl, M.C., Malovannaya, A., Held, J.M., Gillette, M.A., Fenyö, D., Kinsinger, C.R., Mesri, M., Rodriguez, H., Davies, S.R., Perou, C.M., Ma, C., Reid Townsend, R., Chen, X., Carr, S.A., Ellis, M.J., Ding, L., 2017. Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. Nat. Commun. 8, 14864. https://doi.org/10.1038/ncomms14864

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 324, 218–223. https://doi.org/10.1126/science.1168978

Jiang, L., Wang, M., Lin, S., Jian, R., Li, Xiao, Chan, J., Dong, G., Fang, H., Robinson, A.E., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., MacArthur, D.G., Meier, S.R., Nedzel, J.L., Nguyen, D.Y., Segrè, A.V., Todres, E., Balliu, B., Barbeira, A.N., Battle, A., Bonazzola, R., Brown, A., Brown, C.D., Castel, S.E., Conrad, D., Cotter, D.J., Cox, N., Das, S., Goede, O.M. de, Dermitzakis, E.T., Engelhardt, B.E., Eskin, E., Eulalio, T.Y., Ferraro, N.M., Flynn, E., Fresard, L., Gamazon, E.R., Garrido-Martín, D., Gay, N.R., Guigó, R., Hamel, A.R., He, Y., Hoffman, P.J., Hormozdiari, F., Hou, L., Im, H.K., Jo, B., Kasela, S., Kellis, M., Kim-Hellmuth, S., Kwong, A., Lappalainen, T., Li, Xin, Liang, Y., Mangul, S., Mohammadi, P., Montgomery, S.B., Muñoz-Aguirre, M., Nachun, D.C., Nobel, A.B., Oliva, M., Park, YoSon, Park, Yongjin, Parsana, P., Reverter, F., Rouhana, J.M., Sabatti, C., Saha, A., Skol, A.D., Stephens, M., Stranger, B.E., Strober, B.J., Teran, N.A., Viñuela, A., Wang, G., Wen, X., Wright, F., Wucher, V., Zou, Y., Ferreira, P.G., Li, G., Melé, M., Yeger-Lotem, E., Barcus, M.E., Bradbury, D., Krubit, T., McLean, J.A., Qi, L., Robinson, K., Roche, N.V., Smith, A.M., Sobin, L., Tabor, D.E., Undale, A., Bridge, J., Brigham, L.E., Foster, B.A., Gillard, B.M., Hasz, R., Hunter, M., Johns, C., Johnson, M., Karasik, E., Kopen, G., Leinweber, W.F., McDonald, A., Moser, M.T., Myer, K., Ramsey, K.D., Roe, B., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Jewell, S.D., Rohrer, D.C., Valley, D.R., Davis, D.A., Mash, D.C., Branton, P.A., Barker, L.K., Gardiner, H.M., Mosavel, M., Siminoff, L.A., Flicek, P., Haeussler, M., Juettemann, T., Kent, W.J., Lee, C.M., Powell, C.C., Rosenbloom, K.R., Ruffier, M., Sheppard, D., Taylor, K., Trevanion, S.J., Zerbino, D.R., Abell, N.S., Akey, J., Chen, L., Demanelis, K., Doherty, J.A., Feinberg, A.P., Hansen, K.D., Hickey, P.F., Jasmine, F., Kaul, R., Kibriya, M.G., Li, J.B., Li, Q., Linder, S.E., Pierce, B.L., Rizzardi, L.F., Smith, K.S., Stamatoyannopoulos, J., Tang, H., Carithers, L.J., Guan, P., Koester, S.E., Little, A.R., Moore, H.M., Nierras, C.R., Rao, A.K., Vaught, J.B., Volpi, S., Snyder, M.P., 2020. A Quantitative Proteome Map of the Human Body. Cell 183, 269-283.e19. https://doi.org/10.1016/j.cell.2020.08.036

Kita, R., Venkataram, S., Zhou, Y., Fraser, H.B., 2017. High-resolution mapping of cis-regulatory variation in budding yeast. Proc. Natl. Acad. Sci. U. S. A. 114, E10736–E10744. https://doi.org/10.1073/pnas.1717421114

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., Sergushichev, A., 2021. Fast gene set enrichment analysis. https://doi.org/10.1101/060012
Kustatscher, G., Grabowski, P., Rappsilber, J., 2017. Pervasive coexpression of spatially proximal genes is buffered at the protein level. Mol. Syst. Biol. 13, 937. https://doi.org/10.15252/msb.20177548
Lahue, C., Madden, A., Dunn, R., Smukowski Heil, C., 2020. History and Domestication of Saccharomyces cerevisiae in Bread Baking. Front. Genet. 11.
Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. https://doi.org/10.1186/1471-2105-9-559
Li, J.J., Bickel, P.J., Biggin, M.D., 2014. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. PeerJ 2, e270. https://doi.org/10.7717/peerj.270
Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., Heckerman, D., 2011. FaST linear mixed models for genome-wide association studies. Nat. Methods 8, 833–835. https://doi.org/10.1038/nmeth.1681
Liu, Y., Beyer, A., Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550. https://doi.org/10.1016/j.cell.2016.03.014
Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., Bähler, J., 2012. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell 151, 671–683. https://doi.org/10.1016/j.cell.2012.09.019
Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., Stamatoyannopoulos, J.A., 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337, 1190–1195. https://doi.org/10.1126/science.1222794
McManus, C.J., May, G.E., Spealman, P., Shteyman, A., 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422–430. https://doi.org/10.1101/gr.164996.113
Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J.T., Gatza, M.L., Wilkerson, M., Perou, C.M., Yellapantula, V., Huang, K., Lin, C., McLellan, M.D., Yan, P., Davies, S.R., Townsend, R.R., Skates, S.J., Wang, J., Zhang, B., Kinsinger, C.R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A.G., Fenyö, D., Ellis, M.J., Carr, S.A., NCI CPTAC, 2016. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature 534, 55–62. https://doi.org/10.1038/nature18003
Messner, C.B., Demichev, V., Bloomfield, N., Yu, J.S.L., White, M., Kreidl, M., Egger, A.-S., Freiwald, A., Ivosev, G., Wasim, F., Zelezniak, A., Jürgens, L., Suttorp, N., Sander, L.E., Kurth, F., Lilley, K.S., Mülleder, M., Tate, S., Ralser, M., 2021. Ultra-fast proteomics with Scanning SWATH. Nat. Biotechnol. 39, 846–854. https://doi.org/10.1038/s41587-021-00860-4
Messner, C.B., Demichev, V., Muenzner, J., Aulakh, S., Röhl, A., Herrera-Domínguez, L., Egger, A.-S., Kamrad, S., Lemke, O., Calvani, E., Mülleder, M., Lilley, K.S., Kustatscher, G., Ralser, M., 2022a. The Proteomic Landscape of Genome-Wide Genetic Perturbations. https://doi.org/10.1101/2022.05.17.492318
Messner, C.B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Mülleder, M., Ralser, M., 2022b. Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. PROTEOMICS n/a, 2200013. https://doi.org/10.1002/pmic.202200013
Messner, C.B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S.I., Egger, A.-S., Kreidl, M., Ludwig, D., Kilian, C., Agostini, F., Zelezniak, A., Thibeault, C., Pfeiffer, M., Hippenstiel, S., Hocke, A., von Kalle, C., Campbell, A., Hayward, C., Porteous, D.J., Marioni, R.E., Langenberg, C., Lilley, K.S., Kuebler, W.M., Mülleder, M., Drosten, C., Suttorp, N., Witzenrath, M., Kurth, F., Sander, L.E., Ralser, M., 2020. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. Cell Syst. 11, 11-24.e4. https://doi.org/10.1016/j.cels.2020.05.012
Mirauta, B.A., Seaton, D.D., Bensaddek, D., Brenes, A., Bonder, M.J., Kilpinen, H., HipSci Consortium, Stegle, O., Lamond, A.I., 2020. Population-scale proteome variation in human induced pluripotent stem cells. eLife 9, e57390. https://doi.org/10.7554/eLife.57390

Moritz, C.P., Mühlhaus, T., Tenzer, S., Schulenborg, T., Friauf, E., 2019. Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause. J. Neurochem. 149, 582–604. https://doi.org/10.1111/jnc.14664

Moyerbrailean, G.A., Davis, G.O., Harvey, C.T., Watza, D., Wen, X., Pique-Regi, R., Luca, F., 2015. A high-throughput RNA-seq approach to profile transcriptional responses. Sci. Rep. 5, 14976. https://doi.org/10.1038/srep14976

Muenzner, J., Trébulle, P., Agostini, F., Messner, C.B., Steger, M., Lehmann, A., Caudal, E., Egger, A.-S., Amari, F., Barthel, N., Chiara, M.D., Mülleder, M., Demichev, V., Liti, G., Schacherer, J., Gossmann, T., Berman, J., Ralser, M., 2022. The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids. https://doi.org/10.1101/2022.04.06.487392

Mun, D.-G., Bhin, J., Kim, S., Kim, Hyunwoo, Jung, J.H., Jung, Yeonjoo, Jang, Y.E., Park, J.M., Kim, Hokeun, Jung, Yeonhwa, Lee, Hangyeore, Bae, J., Back, S., Kim, S.-J., Kim, Jieun, Park, H., Li, H., Hwang, K.-B., Park, Y.S., Yook, J.H., Kim, B.S., Kwon, S.Y., Ryu, S.W., Park, D.Y., Jeon, T.Y., Kim, D.H., Lee, J.-H., Han, S.-U., Song, K.S., Park, D., Park, J.W., Rodriguez, H., Kim, Jaesang, Lee, Hookeun, Kim, K.P., Yang, E.G., Kim, H.K., Paek, E., Lee, S., Lee, S.-W., Hwang, D., 2019. Proteogenomic Characterization of Human Early-Onset Gastric Cancer. Cancer Cell 35, 111-124.e10. https://doi.org/10.1016/j.ccell.2018.12.003

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Ponnala, L., Wang, Y., Sun, Q., van Wijk, K.J., 2014. Correlation of mRNA and protein abundance in the developing maize leaf. Plant J. Cell Mol. Biol. 78, 424–440. https://doi.org/10.1111/tpj.12482

Pu, S., Wong, J., Turner, B., Cho, E., Wodak, S.J., 2009. Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. 37, 825–831. https://doi.org/10.1093/nar/gkn1005

Russo, P.S.T., Ferreira, G.R., Cardozo, L.E., Bürger, M.C., Arias-Carrasco, R., Maruyama, S.R., Hirata, T.D.C., Lima, D.S., Passos, F.M., Fukutani, K.F., Lever, M., Silva, J.S., Maracaja-Coutinho, V., Nakaya, H.I., 2018. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. BMC Bioinformatics 19, 56. https://doi.org/10.1186/s12859-018-2053-1

Salovska, B., Zhu, H., Gandhi, T., Frank, M., Li, W., Rosenberger, G., Wu, C., Germain, P.-L., Zhou, H., Hodny, Z., Reiter, L., Liu, Y., 2020. Isoform-resolved correlation analysis between mRNA abundance regulation and protein level degradation. Mol. Syst. Biol. 16, e9170. https://doi.org/10.15252/msb.20199170

Sayols, S., 2022. rrvgo: Reduce + Visualize GO. https://doi.org/10.18129/B9.bioc.rrvgo

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M., 2011. Global quantification of mammalian gene expression control. Nature 473, 337–342. https://doi.org/10.1038/nature10098

Stuecker, T.N., Scholes, A.N., Lewis, J.A., 2018. Linkage mapping of yeast cross protection connects gene expression variation to a higher-order organismal trait. PLOS Genet. 14, e1007335. https://doi.org/10.1371/journal.pgen.1007335

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. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

The GTEx Consortium, 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330. https://doi.org/10.1126/science.aaz1776

The GTEx Consortium, 2017. Genetic effects on gene expression across human tissues. Nature 550, 204–213. https://doi.org/10.1038/nature24277

The GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110

Upadhya, S.R., Ryan, C.J., 2022. Experimental reproducibility limits the correlation between mRNA and protein abundances in tumor proteomic profiles. Cell Rep. Methods 2, 100288. https://doi.org/10.1016/j.crmeth.2022.100288

Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., Gritsenko, M.A., Zimmerman, L.J., McDermott, J.E., Clauss, T.R., Moore, R.J., Zhao, R.,

Monroe, M.E., Wang, Y.-T., Chambers, M.C., Slebos, R.J.C., Lau, K.S., Mo, Q., Ding, L., Ellis, M., Thiagarajan, M., Kinsinger, C.R., Rodriguez, H., Smith, R.D., Rodland, K.D., Liebler, D.C., Liu, T., Zhang, B., Pandey, A., Paulovich, A., Hoofnagle, A., Mani, D.R., Chan, D.W., Ransohoff, D.F., Fenyo, D., Tabb, D.L., Levine, D.A., Boja, E.S., Kuhn, E., White, F.M., Whiteley, G.A., Zhu, H., Zhang, H., Shih, I.-M., Bavarva, J., Whiteaker, J., Ketchum, K.A., Clauser, K.R., Ruggles, K., Elburn, K., Hannick, L., Watson, M., Oberti, M., Mesri, M., Sanders, M.E., Borucki, M., Gillette, M.A., Snyder, M., Edwards, N.J., Vatanian, N., Rudnick, P.A., McGarvey, P.B., Mertins, P., Townsend, R.R., Thangudu, R.R., Rivers, R.C., Payne, S.H., Davies, S.R., Cai, S., Stein, S.E., Carr, S.A., Skates, S.J., Madhavan, S., Hiltke, T., Chen, X., Zhao, Y., Wang, Y., Zhang, Z., 2019. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell 177, 1035-1049.e19. https://doi.org/10.1016/j.cell.2019.03.030

Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L., Meng, C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H., Kuster, B., 2019. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol. Syst. Biol. 15, e8503. https://doi.org/10.15252/msb.20188503

Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M.E., Peters, A.H.F.M., Anders, S., Kaessmann, H., 2020. Transcriptome and translatome co-evolution in mammals. Nature 1–6. https://doi.org/10.1038/s41586-020-2899-z

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., Kuster, B., 2014. Mass-spectrometry-based draft of the human proteome. Nature 509, 582–587. https://doi.org/10.1038/nature13319

Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, Article17. https://doi.org/10.2202/1544-6115.1128

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., Davies, S.R., Wang, S., Wang, P., Kinsinger, C.R., Rivers, R.C., Rodriguez, H., Townsend, R.R., Ellis, M.J.C., Carr, S.A., Tabb, D.L., Coffey, R.J., Slebos, R.J.C., Liebler, D.C., 2014. Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387. https://doi.org/10.1038/nature13438

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, Bai, McDermott, J.E., Zhou, J.-Y., Petyuk, V.A., Chen, Li, Ray, D., Sun, S., Yang, F., Chen, Lijun, Wang, J., Shah, P., Cha, S.W., Aiyetan, P., Woo, S., Tian, Y., Gritsenko, M.A., Clauss, T.R., Choi, C., Monroe, M.E., Thomas, S., Nie, S., Wu, C., Moore, R.J., Yu, K.-H., Tabb, D.L., Fenyö, D., Bafna, V., Wang, Y., Rodriguez, H., Boja, E.S., Hiltke, T., Rivers, R.C., Sokoll, L., Zhu, H., Shih, I.-M., Cope, L., Pandey, A., Zhang, Bing, Snyder, M.P., Levine, D.A., Smith, R.D., Chan, D.W., Rodland, K.D., Carr, S.A., Gillette, M.A., Klauser, K.R., Kuhn, E., Mani, D.R., Mertins, P., Ketchum, K.A., Thangudu, R., Cai, S., Oberti, M., Paulovich, A.G., Whiteaker, J.R., Edwards, N.J., McGarvey, P.B., Madhavan, S., Wang, P., Chan, D.W., Pandey, A., Shih, I.-M., Zhang, H., Zhang, Z., Zhu, H., Cope, L., Whiteley, G.A., Skates, S.J., White, F.M., Levine, D.A., Boja, E.S., Kinsinger, C.R., Hiltke, T., Mesri, M., Rivers, R.C., Rodriguez, H., Shaw, K.M., Stein, S.E., Fenyo, D., Liu, T., McDermott, J.E., Payne, S.H., Rodland, K.D., Smith, R.D., Rudnick, P., Snyder, M., Zhao, Y., Chen, X., Ransohoff, D.F., Hoofnagle, A.N., Liebler, D.C., Sanders, M.E., Shi, Z., Slebos, R.J.C., Tabb, D.L., Zhang, Bing, Zimmerman, L.J., Wang, Y., Davies, S.R., Ding, L., Ellis, M.J.C., Townsend, R.R., 2016. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell 166, 755–765. https://doi.org/10.1016/j.cell.2016.05.069

CONCLUSION & PERSPECTIVES

In the context of unraveling the links between genetic variation and observable traits within a population or species, it is critical to better characterize intermediate traits (*i.e.* molecular traits). In this perspective, we aimed at exploring gene expression through mRNA abundance, protein abundance, as well as translation itself in the budding yeast *Saccharomyces cerevisiae*. The study of their variation among individuals as well as their correlation and interaction were conducted to better understand how genetic diversity shapes such expression variation at the population level. We have used two complementary approaches to study gene expression: a deep gene coverage approach and a population-level approach. The combination of these two types of approaches is still essential today, as each approach emphasizes different aspects of gene expression variation while suffering from some specific limitations.

Complementary approaches for a better understanding of gene expression variation

Our investigation of gene expression variation using a large gene coverage on a limited number of individuals showed that despite functional similarities between the expression levels (transcriptome, translatome and proteome), there is an important and general trend of variation buffering: the transcriptional variations are buffered at the translatome and proteome levels. We observed that this phenomenon increases as long as the expression process progresses, suggesting higher constraints on the translatome and even more on the proteome. We showed that this buffering, usually referred to as post-transcriptional buffering (Blevins et al., 2019; Gonçalves et al., 2017; McManus et al., 2014; Wang et al., 2020), affects genes unevenly. Indeed, genes such as essential genes or the ones involved in protein complexes are preferentially buffered, this trend being observed at both the translatome and the proteome level. This suggests that several mechanisms underlie this phenomenon. Some already described mechanisms are obviously good candidates, such as those preventing unassembled protein complex components (Chotewutmontri and Barkan, 2016; Jüschke et al., 2013; Lukoszek et al., 2016; Trösch et al., 2018) or, more generally, protein degradation (Gonçalves et al., 2017). However, this phenomenon remains largely elusive and should be further analyzed and explored to clarify it.

Overall, despite allowing for higher gene coverage and more precision, exploring gene expression in a limited number of strains is unlikely to reveal large-scale effects within a large population and poorly suited for the systematic exploration of the genetic origins of gene expression variation.

Population-level exploration of gene expression variation is a more appropriate strategy in this regard as some specific explorations, such as co-expression network or mRNA-protein variation correlation, can only be considered with large-scale gene expression surveys. We therefore examined mRNA and protein abundance at the population level and found that transcriptome and proteome variation were poorly correlated across individuals, in contrast to previous observations in smaller datasets (Albert et al., 2014; Aydin et al., 2023; Buccitelli and Selbach, 2020; Wang et al., 2019). We observed that genetic regulation of protein abundance is highly distinct from genetic regulation of mRNA abundance at the population level and that this is partly due to variation in protein degradation, which may play a central role in proteome-specific regulation. While the relationship and dependency between transcript and protein levels has been debated (Buccitelli and Selbach, 2020; Liu et al., 2016; Upadhya and Ryan, 2022), it remains poorly understood to date. Our dataset represents the first population-level, multi-omics exploration and demonstrates that transcriptomes and proteomes are clearly two distinct layers of regulation, governed by different genetic bases in natural populations, highlighting the importance of integrating these different levels of gene expression to better understand the genotype-phenotype relationship. However, and despite these promising results, systematic studies at the population-scale level are still tedious and costly to perform. Indeed, the implementation of these strategies usually requires the reduction of experimental time and cost per individual. This represents a bottleneck at the proteomic side, for which gene coverage remains relatively low when applied at high-throughput. Consequently, large-scale proteome quantifications still require a sharp trade-off between the number of considered isolates and the proteome coverage.

Towards a larger view of gene expression

Despite representing one of the largest gene expression explorations to date, several crucial points were beyond the scope of our work, either because of technical limitations or the fact that gene expression encompasses tenths of different mechanisms.

A more exhaustive transcriptome and proteome exploration

A larger coverage of the proteome

As mentioned above, large-scale proteome research still suffers from gene coverage limitation. This is related to the methodology used for high-throughput proteomics where several technical

aspects such as the chromatograph time and liquid flow has been drastically reduced. This was an adaptation of a previously published study that already allowed a theoretical throughput of 180 samples per day (Messner et al., 2020). Covering a set of 630 proteins in more than 900 isolates at this rate is a novelty in itself. However, several improvements are still needed, especially in peptide signal acquisition or proteomic data handling, as this is still a limited number of genes covering only 10% of the theoretical proteome, that is moreover biased toward highly expressed genes. Increasing the proteome coverage would be crucial, in particular to capture some more population-specific trends. For example, the *LAC* and *MAL* genes that were related to strong subpopulation-specific transcription signals in the French dairy and beers populations, respectively (Caudal et al., 2023) were not quantified in our population-scale proteome exploration.

Exploration of other culture conditions

Throughout this work, yeast cultures have been performed on complete synthetic medium (at 30°C) to provide a nutrient-rich and controlled environment. However, such culture conditions obviously do not reflect the natural environment for the vast majority of the 1,011 isolates we worked with, especially since *S. cerevisiae* is ubiquitously distributed on earth across both human and wild niches, and thus face a vast diversity of trophic conditions (Bai et al., 2022; Peter et al., 2018; Wang et al., 2012). Therefore, an interesting continuation of this work could be the study of gene expression in various growth culture conditions, closer to the environmental constraints faced by certain subpopulations, for which subpopulation-specific gene expression could be much stronger. Indeed, the strains adapted to specific conditions would most probably have an improved gene expression network to cope with stresses they commonly face in their natural environment, such as high copper or sulfite concentrations for wine isolates, or lactose-rich environment for dairy isolates. This could also allow to better characterize the impact of the *S. cerevisiae* pangenome on gene expression, as several accessory genes are known to be advantageous in specific contexts such as vinification for HGT-related ORF in wine isolates (Marsit et al., 2015). Finally, performing GWAS on gene expression data coming from different culture conditions would be a promising way to explore in depth the overall genotype-phenotype relationship, especially when GWAS based on growth data was already performed using the 1,011 collection (Peter et al., 2018).

Transcript and peptide degradation

Gene expression is a complex mechanism where the final abundance of transcripts and peptides in a cell results from the combination of several factors, including the rate of synthesis, but also the rate of degradation (Buccitelli and Selbach, 2020). Therefore, both transcript and protein degradation can be considered as a determinant mechanism underlying gene expression, and their study could represent an interesting follow-up to this work. Protein degradation has been shown to be highly important for buffering mechanisms and partly underlies the phenomenon of post-transcriptional buffering (de Bie and Ciechanover, 2011; Gonçalves et al., 2017; Juszkiewicz and Hegde, 2018; Taggart et al., 2020). We have also shown that the rate of degradation has a significant influence on the overlap between the genetic origin of mRNA and protein abundance: proteins with high turnover tend to be more associated with proteome-specific regulation.

However, to date there is no high-throughput method for both transcript and protein degradation measurement, and current methods are still too laborious to be applied to more than a thousand samples.

Population-wide exploration of translation regulation

Thor-Ribo-Seq

The study of translation regulation is of particular interest as it represents the central link between transcriptome and proteome. Defective translation regulation can have phenotypic consequences, and has for example been shown to be implicated in the pathogenesis of many diseases such as cancer (Robichaud et al., 2019). However, compared to the transcriptome (Caudal et al., 2023; The GTEx Consortium, 2015) or proteome (Ferkingstad et al., 2021; Messner et al., 2023), the translatome has been poorly explored, especially at the population level where no large-scale studies have been conducted so far, most probably due to technical limitations.

Recently, a team developed a high-throughput approach for ribosome profiling called Thor-Ribo-Seq (Mito et al., 2023). This method relies on the use of a small amount of substrate by linear amplification of mRNA fragments covered by the ribosomes (Figure 1). This is a major advance in the field as it allows for ribosome profiling scalable to large number of samples. This newly developed Thor-Ribo-Seq technique allows to consider generating the translatome for the whole collection of 1,000 natural isolates.

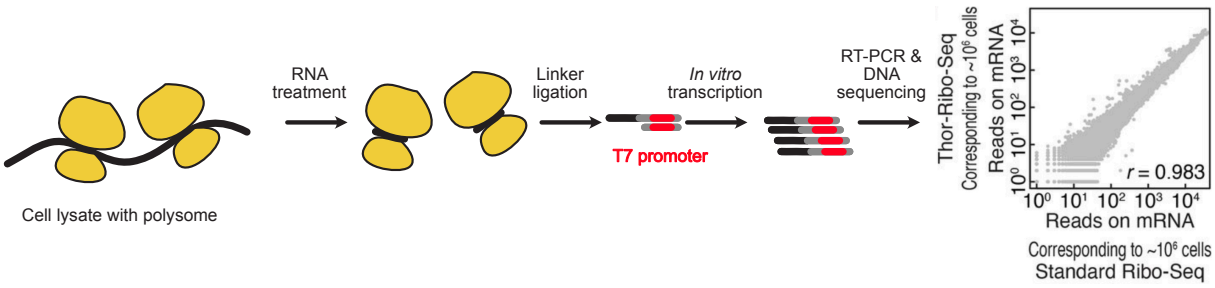


Figure 1: Thor-Ribo-Seq is a powerful approach to perform ribosome profiling on low input sample.

With the Thor-Ribo-Seq method, inputs with low mRNA quantities are amplified via *in vitro* transcription after a fusion of the mRNA fragments with a T7 promoter. The resulting ribosome profiling data is highly similar to data obtained with standard ribosome profiling. Figure adapted from (Mito et al., 2023)

Expected insights

A large-scale ribosome profiling approach would help revealing several fundamental aspects of translation. First, the genetic origins of translation variation across individuals could be explored in depth and would help gain a more precise view of the mechanisms underlying translation regulation (*e.g.*, local, or distant regulation, presence of regulatory hotspots). Equally important will be the comparison of these results with those obtained on both transcriptomes and proteomes (Caudal et al., 2023; Muenzner et al., 2022). The comparative analysis of these datasets would represent an incredible opportunity to decipher the interactions between each expression layer and would provide an exhaustive view of the gene expression process. Finally, translation-specific regulation, such as ribosome velocity diversity and population-scale frameshift catalogs, will be achievable for the first time with this type of data.

Accounting for missing heritability

We performed genome-wide associations to identify the genetic origins of variation in both mRNA and protein abundance. Our study focused on SNPs and CNVs, yet other types of variants can impact phenotypes in general, and in gene expression in particular. Among them, structural variants (SVs) are central in modifying gene expression (Alonge et al., 2020). Their frequency and effects have already been studied in yeast (Dephoure et al., 2014; Muenzner et al., 2022; O'Donnell et al., 2023), and the largest catalog of SVs available for this species so far was constructed from a set of 142 natural isolates mostly out of the 1,011 population (O'Donnell et al., 2023). In this work, SVs were observed as impacting gene expression, especially near their breakpoints. This is in line with previous findings showing that SV-like

inversions or translocations can directly affect the promoter of a gene and thus lead to a modification of gene expression, as has been shown in the case of sulfite resistance in some wine-related *S. cerevisiae* isolates (García-Ríos et al., 2019; Marullo et al., 2020; Pérez-Ortín et al., 2002; Yuasa et al., 2004; Zimmer et al., 2014). Yet, 142 isolates represent only a fraction of the total 1,011 population, suggesting that a large part of the SVs is certainly missed. A project aiming to sequence the entire 1,011-population with long-read sequencing method (Nanopore sequencing technology) is currently ongoing is our laboratory. As the experimental part is now completed, an exhaustive catalog of the SVs across the 1,011 isolates should be released within the next few months. The best methodology to perform GWAS on SVs is still debated, however pangenome graph-based association studies are a promising approach (He et al., 2023; Li et al., 2022; Logsdon et al., 2020; Zhou et al., 2022)

Finally, rare variants (with a frequency in the population below 5%) have also been poorly considered in association studies so far (Génin, 2020; Manolio et al., 2009). The exploration of low-frequency variants in a natural population can be performed by artificially increasing the frequency of the variants using a diallel cross strategy (Fournier et al., 2019). A recent project in our lab investigated mRNA abundance within such a diallel cross and emphasized the significant impact of rare variants on gene expression (Tsouris et al., 2023). The phenotypic characterization of this population could for example be expanded to protein abundance which could help clarifying the effect of rare variants on the proteome.

References

Albert, F.W., Treusch, S., Shockley, A.H., Bloom, J.S., Kruglyak, L., 2014. Genetics of single-cell protein abundance variation in large yeast populations. Nature 506, 494–497. https://doi.org/10.1038/nature12904

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T.H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A.L., Tieman, D.M., Klee, H., Kirsche, M., Aganezov, S., Ranallo-Benavidez, T.R., Lemmon, Z.H., Kim, J., Robitaille, G., Kramer, M., Goodwin, S., McCombie, W.R., Hutton, S., Van Eck, J., Gillis, J., Eshed, Y., Sedlazeck, F.J., van der Knaap, E., Schatz, M.C., Lippman, Z.B., 2020. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell 182, 145-161.e23. https://doi.org/10.1016/j.cell.2020.05.021

Aydin, S., Pham, D.T., Zhang, T., Keele, G.R., Skelly, D.A., Paulo, J.A., Pankratz, M., Choi, T., Gygi, S.P., Reinholdt, L.G., Baker, C.L., Churchill, G.A., Munger, S.C., 2023. Genetic dissection of the pluripotent proteome through multi-omics data integration. Cell Genomics 3, 100283. https://doi.org/10.1016/j.xgen.2023.100283

Bai, F.-Y., Han, D.-Y., Duan, S.-F., Wang, Q.-M., 2022. The Ecology and Evolution of the Baker's Yeast Saccharomyces cerevisiae. Genes 13, 230. https://doi.org/10.3390/genes13020230

Blevins, W.R., Tavella, T., Moro, S.G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., Carey, L.B., Albà, M.M., 2019. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Sci. Rep. 9, 11005. https://doi.org/10.1038/s41598-019-47424-w

Buccitelli, C., Selbach, M., 2020. mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 21, 630–644. https://doi.org/10.1038/s41576-020-0258-4

Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonniere, E., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., 2023. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. https://doi.org/10.1101/2023.05.17.541122

Chotewutmontri, P., Barkan, A., 2016. Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. PLoS Genet. 12, e1006106. https://doi.org/10.1371/journal.pgen.1006106

de Bie, P., Ciechanover, A., 2011. Ubiquitination of E3 ligases: self-regulation of the ubiquitin system via proteolytic and non-proteolytic mechanisms. Cell Death Differ. 18, 1393–1402. https://doi.org/10.1038/cdd.2011.16

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., Torres, E.M., 2014. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife 3, e03023. https://doi.org/10.7554/eLife.03023

Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., Jensson, B.O., Zink, F., Halldorsson, G.H., Masson, G., Arnadottir, G.A., Katrinardottir, H., Juliusson, K., Magnusson, M.K., Magnusson, O.T., Fridriksdottir, R., Saevarsdottir, S., Gudjonsson, S.A., Stacey, S.N., Rognvaldsson, S., Eiriksdottir, T., Olafsdottir, T.A., Steinthorsdottir, V., Tragante, V., Ulfarsson, M.O., Stefansson, H., Jonsdottir, I., Holm, H., Rafnar, T., Melsted, P., Saemundsdottir, J., Norddahl, G.L., Lund, S.H., Gudbjartsson, D.F., Thorsteinsdottir, U., Stefansson, K., 2021. Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. 53, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w

Fournier, T., Abou Saada, O., Hou, J., Peter, J., Caudal, E., Schacherer, J., 2019. Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. eLife 8, e49258. https://doi.org/10.7554/eLife.49258

García-Ríos, E., Nuévalos, M., Barrio, E., Puig, S., Guillamón, J.M., 2019. A new chromosomal rearrangement improves the adaptation of wine yeasts to sulfite. Environ. Microbiol. 21, 1771–1781. https://doi.org/10.1111/1462-2920.14586

Génin, E., 2020. Missing heritability of complex diseases: case solved? Hum. Genet. 139, 103–113. https://doi.org/10.1007/s00439-019-02034-4

Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., Beltrao, P., 2017. Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. Cell Syst. 5, 386-398.e4. https://doi.org/10.1016/j.cels.2017.08.013

He, Q., Tang, S., Zhi, H., Chen, J., Zhang, Jun, Liang, H., Alam, O., Li, H., Zhang, H., Xing, Lihe, Li, X., Zhang, W., Wang, Hailong, Shi, J., Du, H., Wu, H., Wang, L., Yang, P., Xing, Lu, Yan, H., Song,

Z., Liu, J., Wang, Haigang, Tian, X., Qiao, Z., Feng, G., Guo, R., Zhu, W., Ren, Y., Hao, H., Li, M., Zhang, A., Guo, E., Yan, F., Li, Q., Liu, Y., Tian, B., Zhao, X., Jia, R., Feng, B., Zhang, Jiewei, Wei, J., Lai, J., Jia, G., Purugganan, M., Diao, X., 2023. A graph-based genome and pan-genome variation of the model plant Setaria. Nat. Genet. 1–11. https://doi.org/10.1038/s41588-023-01423-w

Jüschke, C., Dohnal, I., Pichler, P., Harzer, H., Swart, R., Ammerer, G., Mechtler, K., Knoblich, J.A., 2013. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. Genome Biol. 14, r133. https://doi.org/10.1186/gb-2013-14-11-r133

Juszkiewicz, S., Hegde, R.S., 2018. Quality Control of Orphaned Proteins. Mol. Cell 71, 443–457. https://doi.org/10.1016/j.molcel.2018.07.001

Li, H., Wang, S., Chai, S., Yang, Z., Zhang, Q., Xin, H., Xu, Y., Lin, S., Chen, X., Yao, Z., Yang, Q., Fei, Z., Huang, S., Zhang, Z., 2022. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. Nat. Commun. 13, 682. https://doi.org/10.1038/s41467-022-28362-0

Liu, Y., Beyer, A., Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550. https://doi.org/10.1016/j.cell.2016.03.014

Logsdon, G.A., Vollger, M.R., Eichler, E.E., 2020. Long-read human genome sequencing and its applications. Nat. Rev. Genet. 21, 597–614. https://doi.org/10.1038/s41576-020-0236-x

Lukoszek, R., Feist, P., Ignatova, Z., 2016. Insights into the adaptive response of Arabidopsis thaliana to prolonged thermal stress by ribosomal profiling and RNA-Seq. BMC Plant Biol. 16, 221. https://doi.org/10.1186/s12870-016-0915-0

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. https://doi.org/10.1038/nature08494

Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E., Dequin, S., Galeote, V., 2015. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Mol. Biol. Evol. 32, 1695–1707. https://doi.org/10.1093/molbev/msv057

Marullo, P., Claisse, O., Raymond Eder, M.L., Börlin, M., Feghali, N., Bernard, M., Legras, J.-L., Albertin, W., Rosa, A.L., Masneuf-Pomarede, I., 2020. SSU1 Checkup, a Rapid Tool for Detecting Chromosomal Rearrangements Related to the SSU1 Promoter in Saccharomyces cerevisiae: An Ecological and Technological Study on Wine Yeast. Front. Microbiol. 11.

McManus, C.J., May, G.E., Spealman, P., Shteyman, A., 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422–430. https://doi.org/10.1101/gr.164996.113

Messner, C.B., Demichev, V., Muenzner, J., Aulakh, S.K., Barthel, N., Röhl, A., Herrera-Domínguez, L., Egger, A.-S., Kamrad, S., Hou, J., Tan, G., Lemke, O., Calvani, E., Szyrwiel, L., Mülleder, M., Lilley, K.S., Boone, C., Kustatscher, G., Ralser, M., 2023. The proteomic landscape of genome-wide genetic perturbations. Cell 186, 2018-2034.e21. https://doi.org/10.1016/j.cell.2023.03.026

Messner, C.B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S.I., Egger, A.-S., Kreidl, M., Ludwig, D., Kilian, C., Agostini, F., Zelezniak, A., Thibeault, C., Pfeiffer, M., Hippenstiel, S., Hocke, A., von Kalle, C., Campbell, A., Hayward, C., Porteous, D.J., Marioni, R.E., Langenberg, C., Lilley, K.S., Kuebler, W.M., Mülleder, M., Drosten, C., Suttorp, N., Witzenrath, M., Kurth, F., Sander, L.E., Ralser, M., 2020. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. Cell Syst. 11, 11-24.e4. https://doi.org/10.1016/j.cels.2020.05.012

Mito, M., Shichino, Y., Iwasaki, S., 2023. Thor-Ribo-Seq: ribosome profiling tailored for low input with RNA-dependent RNA amplification. https://doi.org/10.1101/2023.01.15.524129

Muenzner, J., Trébulle, P., Agostini, F., Messner, C.B., Steger, M., Lehmann, A., Caudal, E., Egger, A.-S., Amari, F., Barthel, N., Chiara, M.D., Mülleder, M., Demichev, V., Liti, G., Schacherer, J., Gossmann, T., Berman, J., Ralser, M., 2022. The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids. https://doi.org/10.1101/2022.04.06.487392

O'Donnell, S., Yue, J.-X., Saada, O.A., Agier, N., Caradec, C., Cokelaer, T., De Chiara, M., Delmas, S., Dutreux, F., Fournier, T., Friedrich, A., Kornobis, E., Li, J., Miao, Z., Tattini, L., Schacherer, J., Liti, G., Fischer, G., 2023. Telomere-to-telomere assemblies of 142 strains characterize the genome

structural landscape in Saccharomyces cerevisiae. Nat. Genet. 55, 1390–1399. https://doi.org/10.1038/s41588-023-01459-y
Pérez-Ortín, J.E., Querol, A., Puig, S., Barrio, E., 2002. Molecular Characterization of a Chromosomal Rearrangement Involved in the Adaptive Evolution of Yeast Strains. Genome Res. 12, 1533–1539. https://doi.org/10.1101/gr.436602
Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5
Robichaud, N., Sonenberg, N., Ruggero, D., Schneider, R.J., 2019. Translational Control in Cancer. Cold Spring Harb. Perspect. Biol. 11, a032896. https://doi.org/10.1101/cshperspect.a032896
Taggart, J.C., Zauber, H., Selbach, M., Li, G.-W., McShane, E., 2020. Keeping the Proportions of Protein Complex Components in Check. Cell Syst. 10, 125–132. https://doi.org/10.1016/j.cels.2020.01.004
The GTEx Consortium, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110
Trösch, R., Barahimipour, R., Gao, Y., Badillo-Corona, J.A., Gotsmann, V.L., Zimmer, D., Mühlhaus, T., Zoschke, R., Willmund, F., 2018. Commonalities and differences of chloroplast translation in a green alga and land plants. Nat. Plants 4, 564–575. https://doi.org/10.1038/s41477-018-0211-0
Tsouris, A., Brach, G., Friedrich, A., Hou, J., Schacherer, J., 2023. Diallel panel reveals a significant impact of low-frequency genetic variants on gene expression variation in yeast. https://doi.org/10.1101/2023.07.21.550015
Upadhya, S.R., Ryan, C.J., 2022. Experimental reproducibility limits the correlation between mRNA and protein abundances in tumor proteomic profiles. Cell Rep. Methods 2, 100288. https://doi.org/10.1016/j.crmeth.2022.100288
Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L., Meng, C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H., Kuster, B., 2019. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol. Syst. Biol. 15, e8503. https://doi.org/10.15252/msb.20188503
Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., Bai, F.-Y., 2012. Surprisingly diverged populations of Saccharomyces cerevisiae in natural environments remote from human activity. Mol. Ecol. 21, 5404–5417. https://doi.org/10.1111/j.1365-294X.2012.05732.x
Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M.E., Peters, A.H.F.M., Anders, S., Kaessmann, H., 2020. Transcriptome and translatome co-evolution in mammals. Nature 588, 642–647. https://doi.org/10.1038/s41586-020-2899-z
Yuasa, N., Nakagawa, Y., Hayakawa, M., Iimura, Y., 2004. Distribution of the sulfite resistance gene SSU1-R and the variation in its promoter region in wine yeasts. J. Biosci. Bioeng. 98, 394–397. https://doi.org/10.1016/S1389-1723(04)00303-2
Zhou, Y., Zhang, Zhiyang, Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., Zhang, J., Lyu, H., Lin, T., Gao, Q., Saha, S., Mueller, L., Fei, Z., Städler, T., Xu, S., Zhang, Zhiwu, Speed, D., Huang, S., 2022. Graph pangenome captures missing heritability and empowers tomato breeding. Nature 606, 527–534. https://doi.org/10.1038/s41586-022-04808-9
Zimmer, A., Durand, C., Loira, N., Durrens, P., Sherman, D.J., Marullo, P., 2014. QTL Dissection of Lag Phase in Wine Fermentation Reveals a New Translocation Responsible for Saccharomyces cerevisiae Adaptation to Sulfite. PLOS ONE 9, e86298. https://doi.org/10.1371/journal.pone.0086298

APPENDIX

List of publications

Submitted

Teyssonnière, E., Schichino, Y., Friedrich, A., Iwasaki, S., Schacherer, J., 2023. Translation variation across genetic backgrounds reveals a post-transcriptional buffering signature in yeast. (In revision - Nucleic Acid Research)

Teyssonnière, E., Trébulle, P., Muenzner, J., Loegler, V., Ludwig, D., Amari, F., Mülleder, M., Friedrich, A., Hou, J., Ralser, M., Schacherer, J. Species-wide quantitative transcriptomes and proteomes reveal distinct genetic control of gene expression variation in yeast. (Submitted to Cell, September 2023)

Caudal, E., Loegler, V., Dutreux, F., Vakirlis, N., **Teyssonnière, E**., Caradec, C., Friedrich, A., Hou, J., Schacherer, J., 2023. Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. https://doi.org/10.1101/2023.05.17.541122 (In revision - Nature Genetics)

In preparation

Teyssonnière, E., Dubreuil, B., Levy, E.D., Schacherer, J., Metabolism adaptation is a main driver of protein abundance evolution in the yeast Saccharomyces cerevisiae.

Oral Communications

Oral

International colloquium: iGénolevure "Yeast Biodiversity": **Teyssonniere, E.**, Schichino, Y., Friedrich, A., Iwasaki, S., Schacherer, J. Translation variation across genetic backgrounds reveals a post-transcriptional buffering signature in yeast. 15-16 November 2021.

ED days, flash talk: **Teyssonniere, E.**, Schichino, Y., Friedrich, A., Iwasaki, S., Schacherer, J. Translation variation across genetic backgrounds reveals a post-transcriptional buffering signature in yeast. April 2021.

CNRS-Weizmann Institute of science PhD Joint Program Workshop: **Teyssonnière, E.,** Dubreuil, B., Levy, E., Schacherer, J. Evolution of gene expression and gene expression in evolution. 13 April 2022.

Poster

EMBO Workshop: From functional genomics to systems biology: **Teyssonnière, E.,** Trébulle, P., Muenzner, J., Loegler, V., Ludwig, D., Amari, F., Mülleder, M., Friedrich, A., Hou, J., Ralser, M., Schacherer, J. Species-wide transcriptome and proteome survey reveals differential regulation of the two gene expression layers in yeast. 15 - 18 November 2022.

Teaching

2020-2022 (2 years) University of Strasbourg. Teaching assistant in microbiology technical course for undergraduate students.

Abstract

An astonishing phenotypic diversity can be observed in natural populations. One of the major goals of modern biology is to unravel the genetic origins of this phenotypic landscape. Gene expression is known to be a main determinant of the relationship between genotypes and phenotypes. In recent decades, several analytical and technical advances have made it possible to study gene expression at every step of the expression process (*e.g.*, transcriptome and proteome) and at very large scales. However, a complete exploration of gene expression across the entire process and at the population scale is still lacking. The goal of this dissertation is to get a more comprehensive view of how each layer of gene expression varies, influences each other, and is related to the natural genetic diversity observed within species. To this end, we analysed the transcriptomes and proteomes of a large natural population of *S. cerevisiae* (bringing together more than 1,000 individuals) and found unsuspected differences between mRNA and protein abundance regulation. Simultaneously, we studied the gene expression process at three different molecular levels (transcriptome, translatome and proteome) and found that important buffering mechanisms underlie the expression variation between individuals.

Elie Teyssonnière

Infraspecific variation and genome evolution
Department of molecular genetics, genomics and microbiology
UMR7156 / CNRS, University of Strasbourg
Ph.D advised by
Pr. Joseph Schacherer
Dr. Anne Friedrich

Haplo Team
Proud to be yeast geneticists

gm
Génétique Moléculaire ▪ Génomique ▪ Microbiologie

cnrs

Université de Strasbourg

Fondation pour la Recherche Médicale