EVOLUTIONARY BIOLOGY

Proteotype coevolution and quantitative diversity across 11 mammalian species

Qian Ba[1]†, Yuanyuan Hei[1]†, Anasuya Dighe[2,3], Wenxue Li[1], Jamie Maziarz[2,3], Irene Pak[2,3], Shisheng Wang[5], Günter P. Wagner[2,3,4,6]*, Yansheng Liu[1,7]*

Evolutionary profiling has been largely limited to the nucleotide level. Using consistent proteomic methods, we quantified proteomic and phosphoproteomic layers in fibroblasts from 11 common mammalian species, with transcriptomes as reference. Covariation analysis indicates that transcript and protein expression levels and variabilities across mammals remarkably follow functional role, with extracellular matrix–associated expression being the most variable, demonstrating strong transcriptome-proteome coevolution. The biological variability of gene expression is universal at both interindividual and interspecies scales but to a different extent. RNA metabolic processes particularly show higher interspecies versus interindividual variation. Our results further indicate that while the ubiquitin-proteasome system is strongly conserved in mammals, lysosome-mediated protein degradation exhibits remarkable variation between mammalian lineages. In addition, the phosphosite profiles reveal a phosphorylation coevolution network independent of protein abundance.

INTRODUCTION

Despite the scalable nucleotide sequencing performed in evolutionary biology, it is ultimately protein abundances and activities that, to a large part, define the organism phenotype. Recently, a qualitative proteome landscape for 100 taxonomically diverse organisms was established by mass spectrometry (MS)–based analysis (*1*). However, a quantitative evolutionary comparison of proteomes across multiple species, such as mammals, represents, so far, uncharted territory. Ribosome profiling (Ribo-seq) was used as a proxy for quantifying proteins synthesized, which revealed coevolutionary patterns across the transcriptome and translatome in five mammals (*2*). However, both the proteome dynamic range and protein degradation cannot be directly measured by Ribo-seq. On the other hand, the recently developed reproducible proteomic workflow exemplified by data-independent acquisition (DIA) MS has achieved favorable reproducibility and quantitative performance for the global proteome (*3*, *4*), with data quality thoroughly and widely assessed (*5*–*8*). However, a systematic, unbiased multispecies quantitative effort has been lacking to link individual variability to species level variability (*9*) and to understand phosphorylation signaling among multiple species in a comprehensive manner.

RESULTS

Steady-state and diverse proteotype across 11 mammalian species

Proteotype is defined as the proteome complement of a genotype (*10*, *11*). To understand the functional and molecular basis of proteotype evolution in mammals, we profiled the steady-state proteomes and phosphoproteomes of primary skin fibroblasts from 11 common mammalian species (Fig. 1A) by DIA-MS (*12*, *13*). Considering that different cell types present a major variable factor in profiling gene expression (*14*–*16*), we here exclusively analyzed cultured fibroblast cells commonly used for evolutionary studies (fig. S1). The mammalian species we analyzed represent two major phylogenetic clades, *Euarchontoglires* (EAOG: primates, rodents, and their relatives) and *Laurasiatheria* (LAUT: carnivores, hoofed animals, and their relatives), together with an evolutionarily distant species, *Monodelphis domestica* (opossum), a marsupial, as the outgroup. The EAOG taxon samples include *Oryctolagus cuniculus* (rabbit), *Rattus norvegicus* (rat), *Macaca mulatta* (monkey), and *Homo sapiens* (human), whereas the LAUT taxon samples include *Ovis aries* (sheep), *Bos taurus* (cow), *Sus scrofa* (pig), *Canis lupus* (dog), *Felis catus* (cat), and *Equus caballus* (horse). For clarity, the short common names of the species are used hereafter.

Our spectral library-free DIA-MS (*4*, *17*) was able to detect an average of 6490 protein groups [peptide and protein false discovery rates (FDRs) < 1%] (*18*) in the fibroblasts of different species, ranging from 5968 (dog) to 7165 (human) identified proteins. In most species, our phosphoproteomic analysis measured >8000 unique, confidently localized phosphosites (P-sites) despite limited species-specific peptide mixtures acquired. In addition, we profiled the mRNA levels for an average of 12,400 genes in all the samples by mRNA sequencing (RNA-seq) to study possible posttranscriptional regulations (Fig. 1A and tables S1 and S2). To quantify the molecular traits across the 11 species, the mRNA, protein, and P-site levels were further compared after a one-to-one gene ortholog mapping between species, using the Ensembl Biomart tool (see Materials and Methods) (*19*). This gene-centric mapping filtered 4353 transcripts [transcripts per million (TPM) > 1], 1660 proteins, and 546 phosphoproteins (containing 611 P-sites) being overlapping across species (fig. S2A and table S3). All the individual replicates clustered together in the hierarchical clustering analysis based on both mRNA and protein quantities (Pearson $R = 0.9822$ and 0.9855; fig. S2, B to E). The overall coefficient of variations (CVs) of the experimental replicates at both mRNA and protein levels are comparable, most of which (86.59% for transcripts and 88.59% for proteins) are <20% (fig. S2, F and G). Considering the

[1]Yale Cancer Biology Institute, West Haven, CT 06516, USA. [2]Yale Systems Biology Institute, West Haven, CT 06516, USA. [3]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA. [4]Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT 06510, USA. [5]West China-Washington Mitochondria and Metabolism Research Center, West China Hospital, Sichuan University, Chengdu 610041, China. [6]Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI 48202, USA. [7]Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06510, USA.
*Corresponding author. Email: gunter.wagner@yale.edu (G.P.W.); yansheng.liu@yale.edu (Y.L.)
†These authors contributed equally to this work.

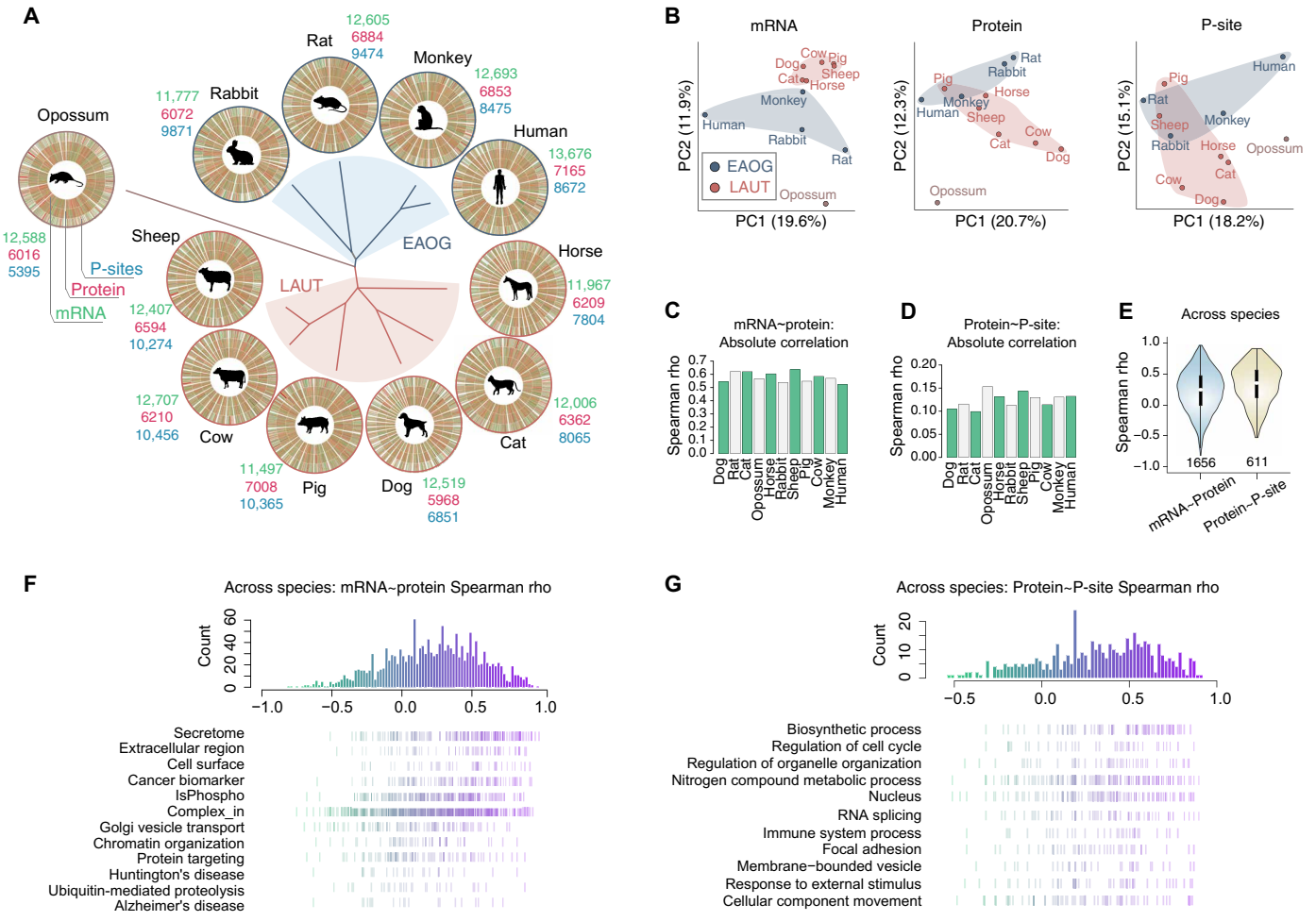

Fig. 1. Identification and quantification overview of transcriptome, proteome, and phosphoproteome across 11 mammalian species. (**A**) Phylogenetic relationships among 11 mammalian species including 10 Boreotherian mammals and the opossum *M. domestica* as an outgroup. In Boreotherian mammals, six LAUT were colored as red, and four EAOG were colored as blue. The transcriptome, proteome, and phosphoproteome in skin fibroblast cells of all these species were shown in Circos plots. The identification numbers of mRNA (TPM > 1), protein, or P-sites in individual species are shown in green, red, and blue colors; created with BioRender.com. (**B**) Principal component analysis (PCA) of mRNA, protein, and P-site profiles in 11 species. (**C** and **D**) Within-species correlation between different layers [mRNA~protein (C) and protein~P-site (D)] of all genes/sites in individual species. Spearman's rho for every species was individually calculated from all detected genes/sites in single species. (**E**) Gene-specific and cross-species correlation between different layers (mRNA~protein and protein~P-site). Spearman's rho for every gene/site was individually calculated from the dataset of 11 species. All rho values were then summarized as violin plots. (**F** and **G**) Representative pathways showing different distribution of mRNA~protein (F) and protein~P-site (G) correlation.

similarly great reproducibility demonstrated for phosphoproteomic DIA workflow (*4*, *20*), together, our datasets enabled a deep and precise quantification of proteotype and covarying signaling at each molecular layer (fig. S3) in mammals.

We subsequently explored the extent of regulation at different molecular layers. First, according to principal components analysis (PCA) (Fig. 1B), we found that, compared to the matched mRNA data, the proteome data showed a smaller power in separating EAOG and LAUT (90 Ma ago), indicating substantial proteotype variability. The phosphoproteome data also showed a limited separating power, which might be partially due to the low coverage of phosphoproteome across species. The transcriptome and proteome of opossum (160 Ma) are both quantitatively distant from EAOG and LAUT species, as expected. Second, we analyzed mRNA~protein and protein~P-site relationships on the absolute scale. Within each species, the mRNA-protein correlation is as high as 0.52 to 0.64 (Spearman rho; Fig. 1C and fig. S4), similar to previous single-species reports (*21*–*23*). Thus, mRNA quantities seem to determine the protein abundances in all mammalian cells tested to a globally similar extent. In contrast, the absolute protein abundances only poorly predict the P-site intensities (rho = 0.10 to 0.15; Fig. 1D) (*4*). Third, we determined the gene-specific and P-site–specific quantitative correlations in the relative scale. We found that cross-species mRNA~protein correlation is centered at Spearman's rho of merely 0.224, based on all quantified mRNA~protein pairs across species (n = 1656; Fig. 1E), arguing for pervasive protein level remodeling between species. Because of the removal of detectability bias among P-sites, the cross-species protein~P-site correlation is higher than the within-species comparison but still weak for many P-sites (mean of rho for all P-sites, 0.326). Furthermore, the rho distributions for mRNA~protein and protein~P-site tend to be gene function class dependent (Fig. 1, F and G). We discovered that, for example, the mRNA~protein correlation across species is much higher

for "secretome" (mean of rho = 0.464) than other processes such as "ubiquitin-mediated proteolysis" (rho = 0.027). The protein~P-site correlation rho was merely 0.245 for "cellular component movement." Therefore, both posttranscriptional and phosphorylation-mediated regulations among mammals are extensive and gene function specific.

Transcriptome-proteome coevolution across gene classes

Phylogenetic relationships can cause nominal correlations among variables even when the variables are in fact evolving independently. To associate mRNA and protein levels with evolutionary process, we performed phylogenetically independent contrast (PIC) analysis to remove the effect of phylogenetic history (see also Materials and Methods and fig. S5) (*24*). The thus determined mRNA and protein PICs across phylogenetical nodes are still positively correlated (averaged $R = 0.306$; fig. S5, A and B), confirming strong transcriptome-proteome coevolution along the phylogeny. The top mRNA~protein coevolving genes are WDR13, XRCC5, GCN1, and others (fig. S5C). To compare coevolution between strongly correlated ($R > 0.8$) and noncorrelated ($|R| < 0.2$) genes, we determined the signed geometric mean of mRNA and protein PIC values as a proxy, which integrates the direction and size of evolution. We found that highly correlated mRNA-protein pairs tend to coevolve at all episode nodes throughout the phylogenetical tree (fig. S6 for C12 to C21). For example, at node C13 that separates EAOG and LAUT clades, geometric means of mRNA and protein PIC are mostly positive and much higher if mRNA~protein correlation is high ($P = 2.4 \times 10^{-13}$) (Fig. 2A). By mapping the number of protein-protein interactions in the STRING (search tool for retrieval of interacting genes/proteins) database (*25*)

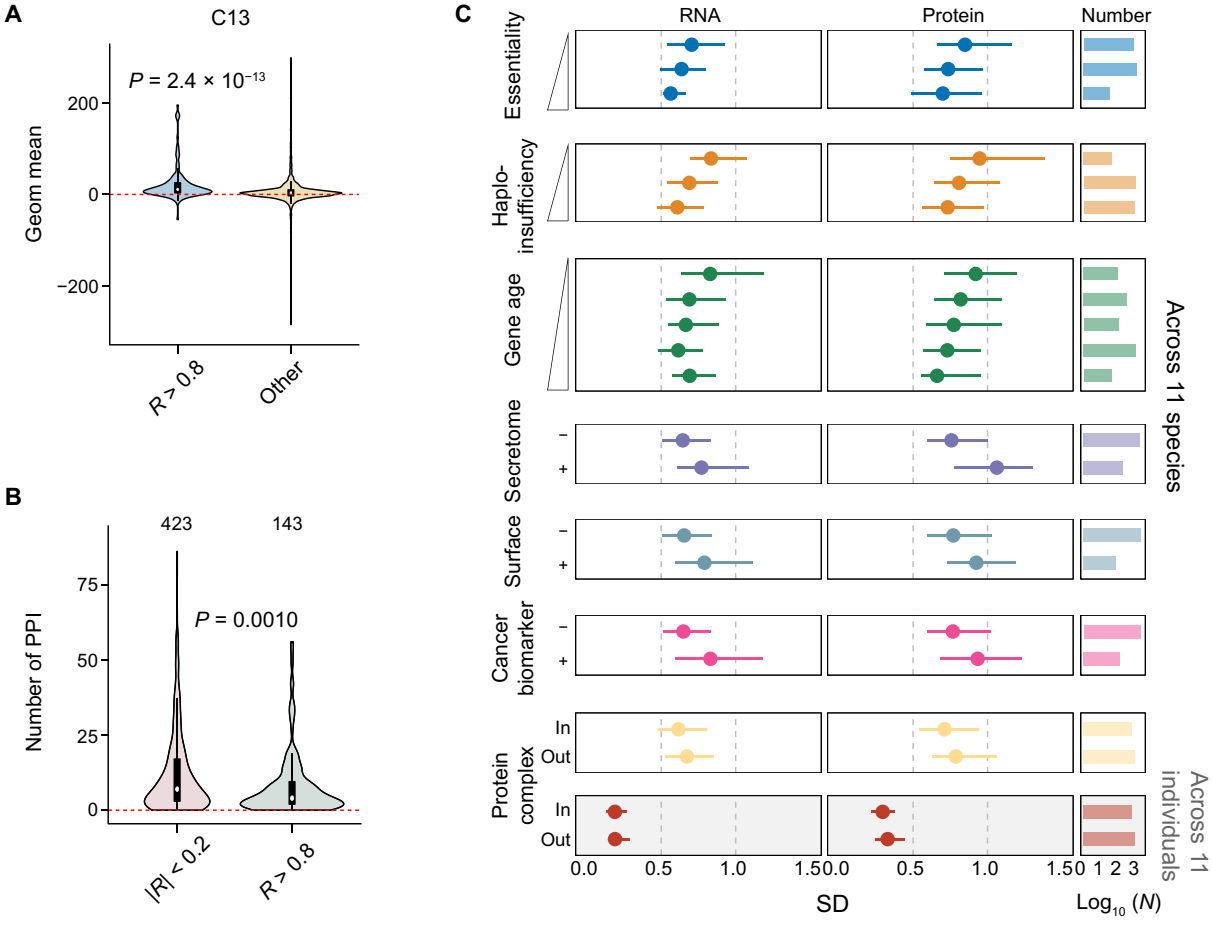

Fig. 2. Biological features associated with mRNA-protein coevolution and their quantitative variability across 11 mammalian species. (**A**) Highly correlated mRNA-protein pairs tend to coevolve. The distribution of signed geometric (geom) mean of mRNA and protein PIC values at node C13 for each gene with high mRNA-protein PIC correlations (Pearson's $R > 0.8$) and other genes. Node C13 separates LAUT from Eurchontoglires. *P* values, Wilcoxon rank sum test. (**B**) Number of protein-protein interactions (PPI) of each protein with high (Pearson's $R > 0.8$) or low (|Pearson's *R*| <0.2) mRNA-protein PIC correlations. *P* values, Wilcoxon rank sum test. (**C**) Profiling of SD for mRNA and protein levels and their relationship to gene essentiality (Wilcoxon rank sum test $P = 9.948 \times 10^{-9}$ and 0.005369 for nonessential versus conditional essential at the mRNA level and 2.644×10^{-15} and 8.203×10^{-4} at the protein level), haploinsufficiency ($P = 2.776 \times 10^{-4}$, 3.374×10^{-7}, and 7.44×10^{-9} for haplo-insensitive versus medium, insensitive versus sensitive, and medium versus sensitive at the mRNA level and 0.005101, 3.243×10^{-5}, and 1.675×10^{-7} at the protein level), secretome ($P = 1.137 \times 10^{-8}$ and 2.303×10^{-15} at the mRNA and protein levels), surfaceome ($P = 6.176 \times 10^{-4}$ and 4.607×10^{-4} at the mRNA and protein levels), and cancer biomarker annotations ($P = 1.514 \times 10^{-7}$ and 1.036×10^{-4} at the mRNA and protein levels) and whether the corresponding proteins are in stable protein complex in the CORUM database (Complex_in) or not (Complex_out). Note that $P = 4.556 \times 10^{-4}$ and 2.393×10^{-6} were obtained at the mRNA and protein levels, respectively, for comparing Complex_In and Out groups in cross-species comparison and $P = 0.2582$ and 3.816×10^{-6} in the cross-individual comparison. The point range showed the median, 25th, and 75th quantiles, and the histogram showed gene number in each class. SD, SD of the relative expression across 11 species or 11 individuals.

among proteins, we found that correlated evolution of RNA and protein levels (e.g., $R > 0.8$ versus $|R| < 0.2$; Fig. 2B) is prevalent among proteins with a low number of protein-protein interactions.

To discern the functional implications of gene expression variability, we calculated the SD for mRNA and protein levels among species (see Materials and Methods). By mapping SDs to a gene essentiality database (*26*), we found that the expression levels of essential or conditionally essential gene products tend to be much less variable during evolution than the expressions of nonessential genes (Fig. 2C). Likewise, haploinsufficient genes (*27*) are expressed with much more stable transcript and protein levels than other genes. Also, the newly evolved mammal-specific genes during evolution (i.e., "younger" genes) (*28*) show higher expression divergence among mammals, especially at the protein level, than those of "older" eukaryote-specific genes (Fig. 2C and fig. S7A). Furthermore, secretory (*29*) and cell surface proteins (*30*) display much less stability between species compared to nonsecretory or proteins inside the cells, consistent to their adaptable roles. A curated list of cancer biomarkers (*31*) also showed larger variability, indicating that these proteins might be prone to be dysregulated in disease states. Together, the interspecies SD of mRNA and protein abundances provides an evolutionary angle to understand gene functional diversity.

To broadly map mRNA and protein species variability to gene functions in an unbiased manner, we used a two-dimensional (2D) enrichment plot (Fig. 3A) (*32*). Such a 2D plot essentially summarizes gene classes expressed with either significantly more variable or stable transcript or protein levels, as compared to other classes. First, we found that most functional classes exhibit highly correlated mRNA and protein variabilities—when mRNA levels are variable, the protein concentrations also tend to be diverse across species and vice versa. This overall trend endorses the notion of strong coevolution of transcriptome and proteome across gene classes (table S4 and fig. S7B) (*2*). Second, among the gene classes, extracellular matrix–related genes have most variable expressions at both, the transcript and the protein levels (highlighted in blue, Fig. 3A). This agrees well with the role of extracellular matrix in interacting with the environment. On the other hand, the intracellular protein and phosphorylation removal systems such as the proteasome complex and the serine/threonine phosphatase complexes showed the least variable concentrations among all functional classes. This may imply that those cell adaptive responses, correcting excessive protein copies (*33*) and abnormal phosphorylation (*34*), are evolutionarily conserved. Third, the gene categories with the most deviation from transcriptome-proteome coevolution are translation-related genes, especially genes involved in translational elongation and initiation, for which the protein quantitative variabilities are strongly buffered compared to that of mRNA. Last, we confirmed that the lower abundant transcripts and proteins are not associated with functional categories that are more variable between mammalian species, strongly endorsing that the above observations are independent of any intensity-associated effects (fig. S7, C and D). In summary, both PIC analysis and functional annotations demonstrate global coevolutionary dynamics between transcriptome and proteome levels across gene classes and among species, with deviating classes mainly due to the protein level buffering and protein-protein interaction constraints.

Biological variability analysis: Interindividual versus interspecies

How does evolutionary gene expression variability compare to variation at other scales? We here refer to a published dataset, part of which reported the proteome variability among 11 unrelated healthy human individuals by using the same cell type and DIA-MS technique (*35*). This comparison thus intriguingly presents an interspecies ($n = 11$) versus interindividual ($n = 11$) comparison of biodiversity. Combined PCA (see Materials and Methods) indicates that the global variability between mammalian species is naturally much larger than the variability between human individuals (fig. S7E). Besides the extent of variation, proteins participating in any heteromeric protein complex [i.e., Complex_In, according to the annotation in the CORUM database (*36*)] exhibited significantly lower interindividual and interspecies SDs than other "Complex_Out" proteins (Fig. 2C, bottom two). This demonstrates the proteostasis control through protein complex stoichiometry, which was previously found posttranscriptionally in many studies (*35*, *37*–*39*). However, our data indicate that the mRNA variability in the Complex_In group is also notably lower than that of Complex_Out. The gene-gene correlation analysis demonstrates a consistent trend (fig. S3D). Hence, the evolutionary diversity might have already started to intensively regulate the transcript abundance toward protein-level usage of, e.g., protein complexes.

Next, we used a 2D enrichment plot (*32*) to illustrate the functional convergence between interindividual diversity and interspecies diversity at the proteome level (Fig. 3B). We found that, at both individual and species scales, the members of the proteasome complex, again, manifest the lowest protein abundance variability, whereas the extracellular matrix proteins show the highest. Compared to the same items enriched in Fig. 3A, it is thus appealing to deduce that abundance control of the subcellular proteomes can extend over both scales of biodiversity (table S5 and fig. S7F). RNA processing and cell division pathways show a particular higher protein abundance variability at the interspecies scale than at the interindividual scale, indicating that these pathways play a particularly important role in mammalian evolution. In addition, at the mRNA level, a similar analysis suggests that while the most variable classes are again extracellular matrix related at both biodiversity scales, the most stable classes are RNA processing–associated pathways (fig. S8). Thus, during mammalian evolution, RNA processing and splicing, although stable between individuals, are tightly regulated at the protein level. To summarize, the interindividual and interspecies proteome variabilities are in fact broadly correlated and function dependent.

Divergent evolutionary conservation for proteasome- and lysosome-mediated protein degradation

Because of the prominent conservation of proteasome expression between individuals and between species, we next sought to interrogate the interspecies stability of the general protein degradation machineries in the cell. We found that the transcript and protein abundance profiles of lysosomal hydrolases and ubiquitin (Ub) enzymes including both deubiquitylating enzymes (DUBs) and E3 Ub ligases demonstrate distinctly higher variabilities than the proteasome components across both, individuals and species (Fig. 4A and fig. S9). Whereas the distribution of Ub enzyme levels is wide, the lysosomal proteases display an even higher protein expression diversity than the average level proteome wide ($P = 5.38 \times 10^{-05}$; fig. S9). This result suggests that the lysosome-mediated degradation pathway is fast evolving among mammals, despite proteasomal degradation is evolutionarily conserved. Only two proteasome proteins showed exceptional instability—proteasome 20S subunit alpha type-4 (PSMA4) and proteasome activator complex subunit 2 (PSME2), the latter was implicated in immunoproteasome assembly (*40*). Many

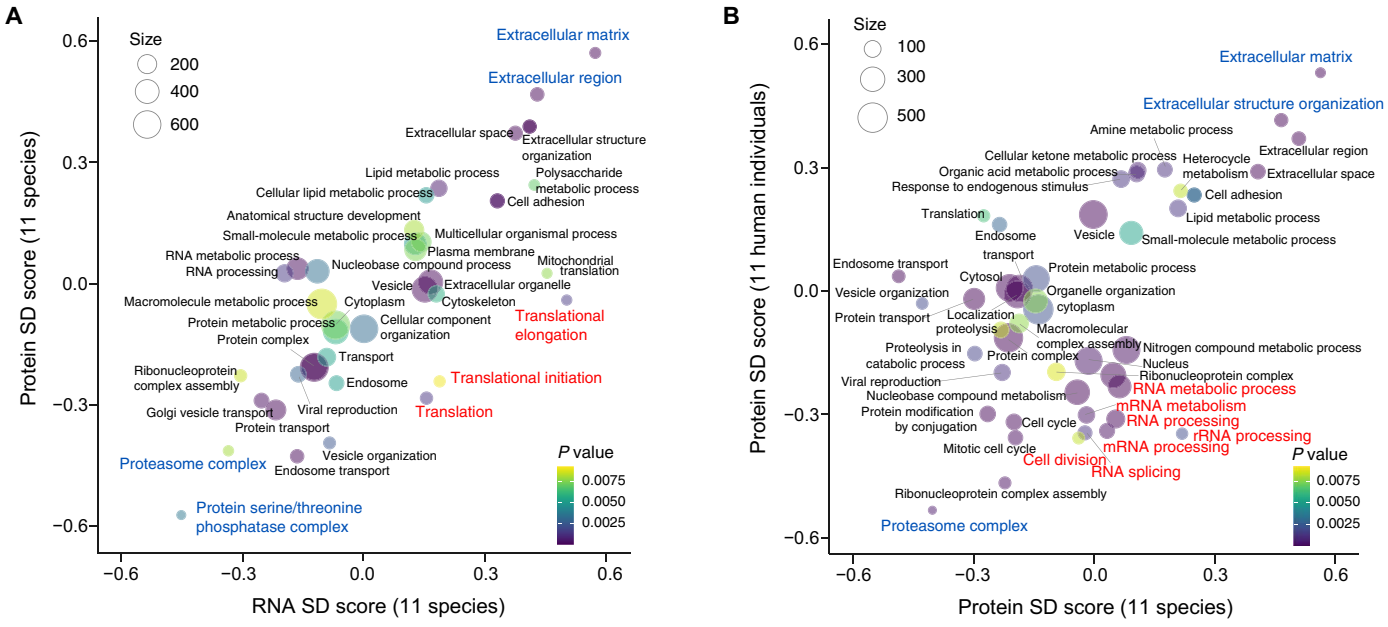

Fig. 3. Functional enrichment analysis revealing protein expression variability control at the posttranscriptional level and between the individual and species scales. (**A**) 2D enrichment plot of SD selected GOBPs (gene ontology biological processes) or GOCCs (gene ontology cellular components). The axes denote enrichment scores for SD from the average at mRNA (*x* axis) and protein (*y* axis) levels across 11 species. (**B**) 2D enrichment plot of SD selected GOBPs or GOCCs denoting gene expression variability analysis across species and human individuals. The axes denote enrichment scores for the protein SD from average across 11 species (*x* axis) and 11 individuals (*y* axis). rRNA, ribosomal RNA.

lysosome hydrolases seem to be expressed lower in opossum than EAOG and LAUT species in which they become more variable. Certain DUBs such as ubiquitin carboxyl-terminal hydrolase 5 (USP5), USP7, USP8, USP14, and USP47 are as conserved as proteasome core subunits, whereas USP19 and USP48 are quite dynamic among mammals. A summary network analysis not only reinforces similar rates of mRNA and protein divergence in these cellular protein removal processes but also reveals that a few Ub enzymes such as the deubiquitinating protein VCPIP1 (valosin containing protein interacting protein 1), the ubiquitin thioesterase (OTU deubiquitinase with linear linkage specificity), and the E3 ubiquitin-protein ligases XIAP (X-linked inhibitor of apoptosis) , and TRIP12 (thyroid hormone receptor interactor 12) underwent extensive posttranscriptional regulation during evolution (Fig. 4B). We additionally checked that the CVs of experimental replicates for all the above proteins, confirming the biological variability between the functional classes, are not associated with any measurement bias ($P = 0.335$; Fig. 4C).

To corroborate these results, we performed an independent Ub modification search (*41*), to focus on Ub chains with specific Ub chain linkage to different lysine (K) residues. We were able to detect the signature peptides representing modified lysine positions in Ub, including K11, K27, K48, and K63, successfully among all species with the modified representative peptides (see Materials and Methods). For the most abundant K48 and K63 DIA-MS signals, we used an unmodified peptide between K48 and K63 peptide sequences as a normalization reference. The relative comparison suggests that the K48-linked Ub chains harbor a smaller cross-species variability than K63 chains (CV 27.2% versus 77.4%; Fig. 4D). Because of the canonical role of K48 polyUb chains in proteasome-mediated degradation and the diverse signaling roles of K63, the Ub modification searching effort agrees well to the above variability analysis.

Common and covarying phosphoproteomic signatures across mammalian species

To interrogate how phosphoproteomics could shed light on molecular evolution, we first evaluated cross-species SDs among transcripts, proteins, and phosphoproteins. We found that P-sites are molecularly much more dynamic than mRNAs and proteins ($P < 2.2 \times 10^{-16}$; Fig. 5A and fig. S10). In addition, the protein abundance "regressed out" P-site values (or P-site_reg) (*42*) are still credibly higher ($P < 2.2 \times 10^{-16}$), indicating that phosphoproteomes captured activity dynamics that is not reflected by transcript and protein abundances. Following, we focused on the 611 P-sites that were aligned across 11 species (hereafter, common P-sites). Compared to all the other P-sites, we detected that, in human, the common P-sites are predicted with an overall lower "sift_ala_score," a computational score predicting the system tolerance if the P-site residue is mutated to alanine ($P = 6.6 \times 10^{-12}$; Fig. 5B) (*43*, *44*). Besides, the common P-sites exhibit a noteworthy higher functional fitness score (*43*) than other P-sites ($P = 3.6 \times 10^{-10}$). Moreover, by mapping to a dataset reporting P-site–specific melting temperature (T_m) (*45*), we found a small but significant difference, indicating that the common P-sites may bring more structural thermal stability to proteins than other P-sites ($P = 3.5 \times 10^{-4}$). These results supported the conservative and essential role of common P-sites in evolution and organismal fitness. As for sequence features, the frequency of amino acids surrounding all the common P-sites revealed a diverse amino acid distribution and enriched for a few signature motifs (Fig. 5, C and D). These motifs, such as (SP), (SP.R), and (R..S), can potentially match to the substrate motifs of, e.g., cyclin-dependent kinases (CDKs) and calmodulin-dependent protein kinase or protein kinase A. Last, a relative amino acid frequency comparison to human background revealed notable depletion of glutamic acid (E) and enrichment of arginine (R) and serine

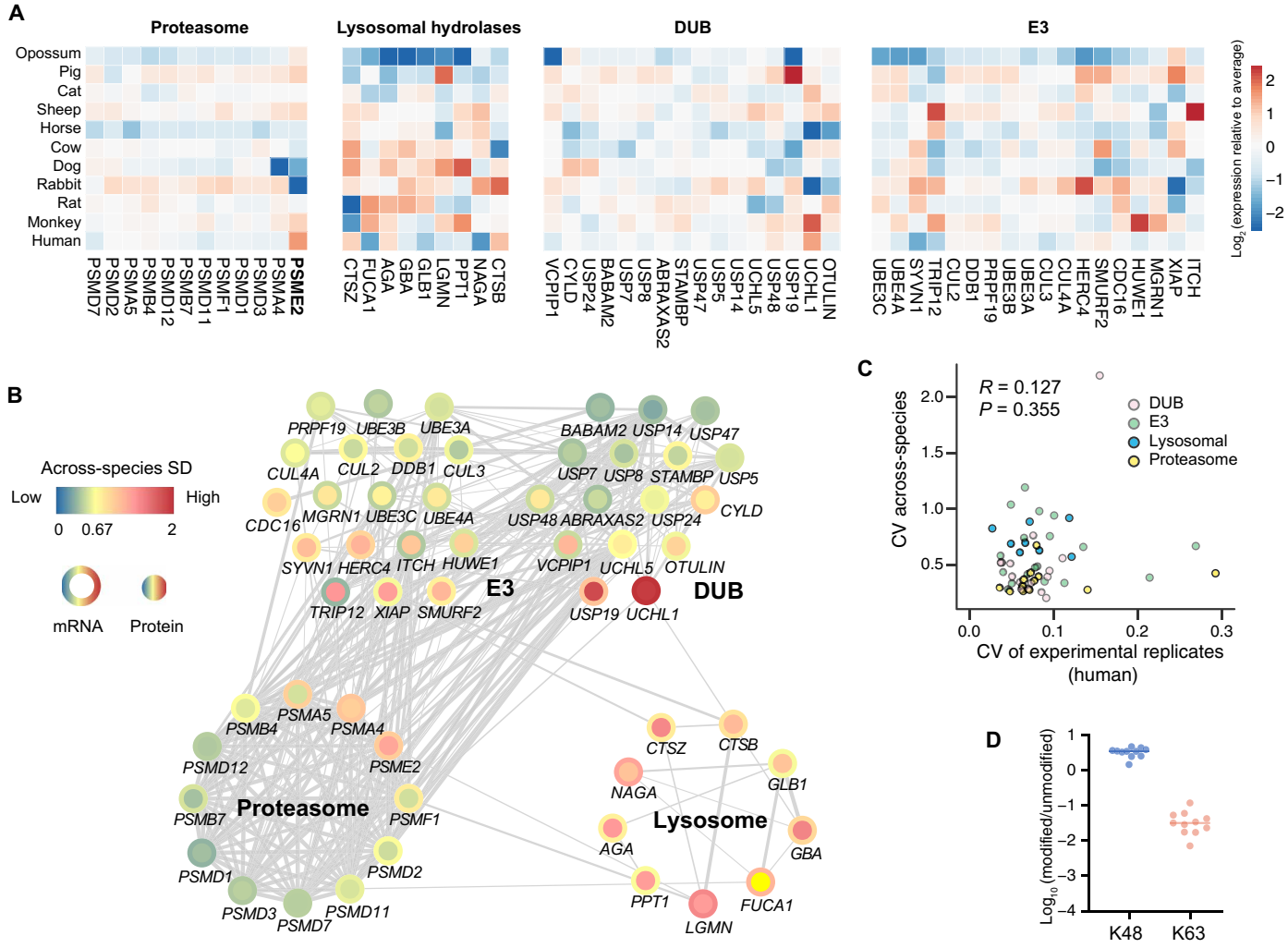

Fig. 4. The distinctive interspecies quantitative diversity for proteasomal and lysosomal protein degradation. (**A**) Heatmaps visualizing the protein-specific variability in 11 species for proteins detected and annotated as proteasome, lysosomal hydrolases, Ub enzymes including both DUBs and E3 Ub ligases. The color bar represents expressive values relative to the average, as calculated by the $\log_2$ (intensity) in the particular species subtracted by mean of the $\log_2$ (intensity) across species. (**B**) Network analysis visualizing STRING interactions between proteins. The outer ring and inner circle are colored for mRNA and protein SDs, respectively. SD, SD of the relative expression across 11 species. (**C**) The distribution of the quantitative CVs among experimental replicates for human along with the CVs across 11 species. CV, coefficient of variation of the raw scale (before log transformation) expression. (**D**) The log-scale ratios of modified Ub peptide versus unmodified peptide reference for K48 and K63 that were quantified across species.

(S) around the common P-sites (Fig. 5E). These motifs and rules might be useful for predicting P-site evolutionary conservation in mammals.

Previously, little has been known about the structure and evolution of phosphorylation signaling networks. We here built a phosphorylation coevolution network (Fig. 5F and fig. S11 for all P-site identities). Each node represents a particular P-site. In addition, each edge denotes a strong site-to-site Pearson correlation ($P < 0.001$) based on the P-site_reg levels across the 11 mammalian species. We found that the majority (i.e., 96.23%) site-to-site correlations are strongly positive and coevolving in the network. Also, the biological annotation suggests that most of the P-sites (i.e., 74.85%) could be functionally annotated to five processes including transport, RNA processing, protein modification process, organelle organization, and developmental process. The top nine kinases, including three CDKs (1, 2, and 5), could govern 18.69% of the P-sites in this network (Fig. 5F), indicating their prevalent roles in mammalian evolution.

In summary, our phosphoproteomics analysis has revealed conservative P-site motifs and a pilot phosphorylation coevolutionary network containing variance independent of protein abundance. We lastly generated a website to facilitate the navigation of the multi-omic data basis (https://yslproteomics.shinyapps.io/Evolutome/; fig. S12).

DISCUSSION

Our proteomic data delineate both protein expression and activity regulation underlying evolutionary diversity on a large scale. In agreement with Wang *et al.* (*2*) that measured translational rates, we found that the old, essential, and housekeeping genes tend to be expressed with much smaller between-species divergence than other

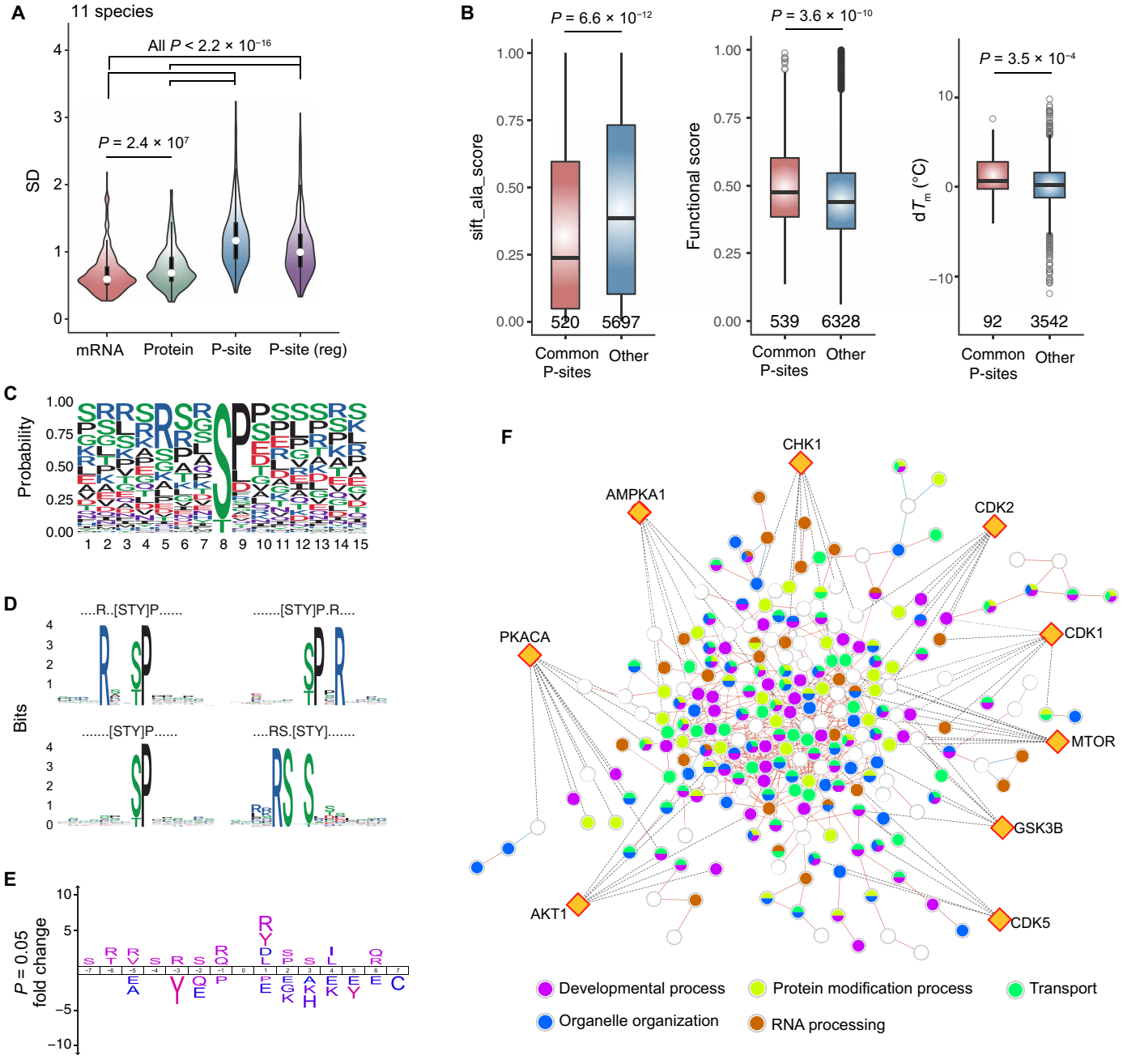

Fig. 5. P-site characteristics among 11 mammalian species. (**A**) Profiling of SD across 11 species from the average at mRNA, protein, P-site, and protein-corrected P-site levels from the intersecting list between layers. SD, SD of the relative quantification across 11 species. (**B**) The site-specific parameters and features of the common P-sites (detected in all 11 species) and other human P-sites. *P* values, Wilcoxon rank sum test. dT_m, difference in melting temperature. (**C**) The sequence logo of surrounding amino acids (±7 amino acids) of common P-sites (detected in all 11 species). (**D**) Motif analysis of the flanking amino acids (±7 amino acids) around the common P-sites (detected in all 11 species). The representative enriched motifs (four motif examples with >7-fold of enrichment) were shown. (**E**) Sequence analysis of the flanking amino acids (±7 amino acids) around the P-sites (common P-sites versus other human P-sites). The fold changes (FCs) of significant residues ($P < 0.05$) were shown. (**F**) By inferring the binary, quantitative correlation analysis between any P-site pairs, the phosphorylation coevolution network was built on the significantly associated common P-sites (Pearson's correlation, $P < 0.001$) after correction by total protein changes. Red lines indicate positive associations while blue lines indicate negative associations. The representative GOBPs of phosphoproteins were highlighted in different colors. The top nine kinases curated from the P-sites as substrates were shown as diamonds, and the kinase-substrates pairs were shown as dashed lines.

genes. We also found that coevolution of expression layers is strong and prevalent across gene classes. In our results, the global proteome variability is comparable to or even slightly higher than the transcriptome level across species (see Fig. 5A and fig. S10), emphasizing the evolution of lineage adaptions, rather than the genome-wide compensatory evolution (*2*), at the proteome level. Processes such as negative regulation of G protein–coupled receptor signaling and mRNA methylation, as examples that can be extracted from table S4, seem to show a higher variability at the protein level than at the RNA level. In addition, our data further suggest that the involvement in protein complexes or protein-protein interactions presents additional constraints potentially reducing the transcriptome-proteome coevolution.

The biological variabilities between individuals and between species jointly shape the biological diversity on Earth. Our data addressed two principles underlying biodiversity. The first principle is that the proteome largely reflects variability at the transcriptional level. As shown in Fig. 3A, gene expression variabilities at mRNA and corresponding protein levels are overall tightly controlled within classes of genes of the same molecular functions. Extracellular matrix proteins overall show a very high evolvability. The dynamics of gene products participating in the extracellular matrix might render each mammalian species a rapid environment adaptive response. In contrast, the expression of proteasome components seems to be evolutionarily conserved as reported (*33*). According to our data, the tight expression regulation is already detectable at the transcriptional level, probably due to the high cost of protein synthesis (*46*). Accordingly, the translation-relevant gene expressions are particularly stabilized at the protein level across species (Fig. 3A). This reinforces the previous finding that translation efficiency profile is highly conserved in evolution (*47*).

In a mammalian cell, the proteasome and lysosome represent the two major machineries for protein degradation (*48*). Previously, Ub-proteasome pathway was found to be evolutionarily conserved (*33*), but little is known about the evolutionary conservation of the lysosomal proteolysis pathway and the entire Ub system. We discovered a much higher variability of the lysosomal degradation pathway. Considering that the endosome transportation–associated protein expressions are quite stable across species, our discovery might be associated with lysosomal exocytosis for remodeling extracellular proteins (*49*). Consistent to this finding, previous studies about the Ub code (*50*) suggested that K48-linked Ub, the most stable Ub codes we found here, is mainly involved in canonical protein degradation via proteasome. In contrast, other lysine-linked Ubs carrying diverse functions, such as endocytosis and DNA damage responses (K63), were quantitatively more variable between species in our results. Even for proteasome itself, the immunoproteasome components such as PSME2 (*11*) manifested high interspecies variation than other proteasome components. Together, our data delineate a complex and manifold relationship between protein degradation and biodiversity.

The second principle underlying biodiversity at the molecular level is that proteome variability tends to be parallel at both the individual and species scales (Fig. 3B). Although it is expected that interspecies protein variability globally exceeds the interindividual variability, proteins in extracellular matrix and proteasome again showed the highest and the lowest variable abundances, respectively, also among humans. Functional classes such as mRNA splicing, mRNA metabolism, and cell division that showed exceptionally higher interspecies variability are intriguing and warrant future investigations. For example, Keren *et al.* (*51*) have summarized various changing models of mRNA alternative splicing (AS) in different eukaryotic lineages. AS provides a strategy for relaxing negative selection pressure against evolutionary change (*52*). In our data, the evolutionary variability of the AS pathway is strong at the protein level, which might enhance evolvability and proteotype diversity. The cell division–associated protein variability among mammals was further supported by phosphoproteomic data. Our cross-scale analysis represents a critical step toward comprehensive understanding of inherent protein expression variability. For many proteins such as those in and related to the extracellular matrix–mediated pathways, their expressions exhibit marked individual differences through intrinsic regulatory machineries, the evolutionary selective forces in speciation and phylogeny would prefer to just expand the magnitude of their intraspecific protein diversity. For the other type of proteins, they are regulated particularly along the evolution axis, irrespective of their biological variability within species, such as RNA processing. It is thus essential to classify two types of protein variabilities for future evolutionary research.

Compared to early studies (*53*), our phosphoproteome data are fairly large in coverage and quantitative. Previously, little was known about the evolution of phosphorylation signaling networks. Our covariance analysis between P-sites provides a first coevolutionary phosphorylation network independent of protein levels. The enriched sequence motifs around the common P-sites pointed out similar kinases such as CDKs heavily evolving in mammalian evolution. We found the patterns of depleted glutamic acid (E) and enriching arginine (R) residues around the common P-site. This result agrees well to our previous study that showed that such patterns tend to accelerate protein turnover when the P-sites get phosphorylated (*20*). Thus, the timely phosphoprotein turnover may render active selection on P-sites in mammalian evolution.

The present study only analyzed fibroblast cells, whereas enormous single-cell and multitissue studies have demonstrated the complexity and diversity of gene expression among different cell types (*14*–*16*). Our datasets could provide additional broad impacts, because skin-derived fibroblast cells play an essential role during cutaneous wound healing and were used to discover different molecular traits associated with the longevity between different mammals and other species previously (*54*, *55*), and skin fibroblasts are critical determinants of skin cancer malignancy across species (*56*, *57*). Moreover, cryobanking of fibroblast cells presents a national-wide tool that facilitates the biological diversity characterization and contributes to ex situ conservation of genetic resources (*58*). Although most of our conclusions, such as transcriptome-proteome coevolution and biological variability control, are unlikely to be limited to fibroblast cells, our results promisingly anchor future proteotype characterization studies on other cell types and multiple tissues across species. In addition, the species differences of molecular events might be more apparent in studies of the disturbed or dynamic systems. Moreover, as already shown in studies analyzing protein turnover (*9*, *59*), thermal stability (*60*), protein-protein interaction (*61*), and specific phosphorylation signaling process (*62*, *63*), the MS-based quantitative proteomic analysis across multiple species will provide additional insights into answering fundamental evolutionary questions and beyond. We expect the establishment of quantitative landscape of proteins and posttranslational modifications across species to further contribute to our understanding of biological variability and biodiversity on Earth.

MATERIALS AND METHODS

Skin fibroblast cell culture

Human skin fibroblast (SF) cells were purchased from American Type Culture Collection (CRL-4001). *B. taurus* (cow), *C. lupus* (dog), *E. caballus* (horse), *F. catus* (cat), *M. mulatta* (monkey), *M. domestica* (opossum), *O. cuniculus* (rabbit), *O. aries* (sheep), *R. norvegicus* (rat), and *S. scrofa* (pig) SFs were obtained from fresh skin tissue following established protocols (*56*, *64*). Briefly, a small piece of skin was collected with hair removed, washed in phosphate-buffered saline (PBS) buffer, and cut into strips approximately 1.0 cm^2. Dermis was separated from epidermis by enzymatic digestion (30 min in 0.25% trypsin buffer at 37°C, followed by dissociation buffer [collagenase (1 mg/ml), dispase (1 mg/ml), and deoxyribonuclease I (400 μg/ml)] for 45 min at 37°C. Epidermis was then removed, and 2-mm pieces were cut from the dermis and transferred to a 12-well plate and covered with media. Fibroblasts emerged from the explants and grew to confluency in growth media with extra tissue removed. Fibroblast cell cultures were then established in 10-cm dishes with high glucose Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum at 5% CO_2 (fig. S1). The approximate doubling times for all the species are about every 3 to 4 days for *H. sapiens*, 3 to 4 days for *B. taurus*, 3 to 7 days for *C. lupus*, 4 days for *E. caballus*, 7 days for *F. catus*, 7 days for *M. mulatta*, 4 to 5 days for *M. domestica*, 2 to 3 days for *O. cuniculus*, 5 to 7 days for *O. aries*, 6 to 7 days for *R. norvegicus*, and every 3 to 5 days for *S. scrofa*. The sample preparation procedures were approved by Institutional Animal Care and Use Committee under the no. 2021-11483. Three replicates (each of 60 to 80% confluency) were processed for RNA-seq and proteomic analyses, respectively, with the exception that five replicates for human and two for monkey proteomic profiling. After estimating the experimental reproducibility (fig. S2), all the replicate measurements were accepted and averaged per each of the 11 mammalian species for transcriptomic and proteomic profiling.

RNA isolation, sequencing, and data procession

RNA was isolated using an RNeasy micro kit (QIAGEN) and resuspended in 15 μl of water. The RNA samples were measured at The Yale Center for Genome Analysis on the Agilent Bioanalyzer 2100 to determine RNA quality, prepared mRNA libraries, and sequenced on Illumina HiSeq 2500 to generate 30 to 40 million reads per sample [single-end 75–base pair (bp) reads].

RNA-seq data obtained were quantified using the transcript-based quantification approach provided in the "kallisto" program (*65*). Reads are aligned to a reference transcriptome using a fast hashing of k-mers together with a directed de Bruijn graph of the transcriptome. This rapid quantification technique produces transcript-wise abundances that are then normalized and mapped to individual genes and ultimately reported in terms of TPM (*66*). The Ensembl release 100 (May 2020 version) (*67*) gene annotation model was used, and raw sequence reads (single-end 75 bp) for SFs from the 11 species were aligned to GRCh38.p13, ARS-UCD1.2, CanFam3.1, Felis_catus_9.0, EquCab3.0, Mmul_10, MonDom5 (Release 97), OryCun2.0, Oar_v3.1, Rnor_6.0, and Sscrofa11.1 reference transcriptome assemblies. The Ensembl Biomart tool was used to obtain a dataset of one-to-one orthologs. Specifically, the "Multi-species Comparisons" and "Homologues" filters were used in addition to "Homology type" filter to obtain one-to-one orthologs from a pair of species. This was done in an iterative fashion for each pair of species. An intersection operator was applied to these pairwise gene lists to obtain a final set of 8138 orthologs between 11 species. To facilitate gene expression across species, this one-to-one ortholog dataset was formulated across the 11 species such that the sum of TPMs across these genes for each species totals to 1×10^6. The TPMs across replicates of the same species were averaged and $\log_2$-transformed for the following bioinformatic analysis.

Protein extraction, alkylation, and digestion

The fibroblast cell proteomes were harvested as previously described (*35*). Cells were washed with PBS twice and scaped off from the dish using the lysis buffer containing 8 M urea containing complete protease inhibitor cocktail (Roche) and Halt Phosphatase Inhibitor (Thermo Fisher Scientific). The cell pellets were then ultrasonically lysed at 4°C for 2 min using a VialTweeter device (Hielscher-Ultrasound Technology) and centrifuged at 18,000*g* for 1 hour to remove the insoluble material. Protein concentrations were then determined with a Bradford assay (Bio-Rad, Hercules, CA, USA). The supernatant protein samples were reduced with 10 mM dithiothreitol for 1 hour at 57°C and alkylated by 20 mM iodoacetamide in the dark for 1 hour at room temperature. All samples were further diluted five times using 100 mM NH_4HCO_3 and were digested in-solution with sequencing-grade porcine trypsin (Promega) overnight at 37°C as previously described (*37*). The resulted peptide mixture was desalted with a C18 column (MarocoSpin Columns, NEST Group INC). The final peptide amounts were determined by NanoDrop (Thermo Fisher Scientific).

Phosphopeptide enrichment

Besides ~4 μg of peptides digested per sample that were used for proteomic analysis, all the peptides from different replicates of equal amount were pooled for phosphopeptide enrichment and phosphoproteomics, due to the limited peptide amounts yielded in individual replicate. The phosphopeptide enrichment was performed using the High-Select Fe-NTA kit (Thermo Fisher Scientific, A32992) according to the manufacturer's instructions (*68*). Briefly, the resins of spin-column in the Fe-NTA kit were aliquoted and incubated with 80 to 300 μg of total peptides for 30 min at room temperature. The resins were then transferred into a filter tip (TF-20-L-R-S, Axygen), so that the supernatant was removed by centrifugation. Then, the resins were washed sequentially with 200 μl of washing buffer (80% acetonitrile (ACN) and 0.1% trifluoroacetic acid) three times and 200 μl of liquid chromatography (LC)–MS grade H_2O two times to remove the nonspecifically adsorbed peptides. The enriched phosphopeptides were then eluted off the resins by 100 μl of elution buffer (50% ACN and 5% $NH_3{\cdot}H_2O$) two times. All the centrifugation steps were kept at 500*g*, 30 s. The eluates per species were combined and dried by speed-vac and stored in −80°C.

LC separation coupled with MS

All peptide-level samples (and their pooled mixtures per species) were resolved in 2% ACN and 0.1% formic acid for LC-MS measurements, with 2 μg of peptides or 0.2 to 0.5 μg of the enriched phosphopeptides injected per measurement. The single-shot DIA-MS analysis of 2.5 hours was performed as previously described (*4*, *69*). The LC used was an EASY-nLC 1200 system (Thermo Fisher Scientific, San Jose, CA) harboring a 75 μm by 50 cm C18 column packed with 100A C18 material. A 150-min LC separation was configured on the basis of the mix of buffer A (0.1% formic acid in H_2O) and buffer B (80% acetonitrile containing 0.1% formic acid):

Buffer B was made to increase from 4 to 34% in 139 min, then to surge to 100% in 3 min, and then kept at 100% for 8 min. The LC-MS flow rate was kept at 300 nl/min with the temperature-controlled at 60°C by a PRSO-V1 column oven (Sonation GmbH, Biberach, Germany). The additional column re-equilibration was performed in about 10 to 15 min using the high-flow rate up to ~800 nl/min.

DIA-MS measurements

The Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) instrument coupled to a nanoelectrospray ion source (NanoFlex, Thermo Fisher Scientific) was used as the DIA-MS platform for both proteomic and phosphoproteomic analyses (*4*). Spray voltage was set to 2000 V and heating capillary temperature was set at 275°C. All the DIA-MS methods consisted of 1 MS1 scan and 33 MS2 scans of variable windows by quadrupole isolation (*70*). This schema was composed of 350 to 373.775, 373.25 to 393.75, 393.25 to 410.75, 410.25 to 427.75, 427.25 to 443.75, 443.25 to 459.75, 459.25 to 474.75, 474.25 to 489.75, 489.25 to 503.75, 503.25 to 518.75, 518.25 to 533.75, 533.25 to 547.75, 547.25 to 562.75, 562.25 to 577.75, 577.25 to 592.75, 592.25 to 608.75, 608.25 to 623.75, 623.25 to 639.75, 639.25 to 656.75, 656.25 to 674.75, 674.25 to 692.75, 692.25 to 711.75, 711.25 to 732.75, 732.25 to 754.75, 754.25 to 778.75, 778.25 to 803.75, 803.25 to 833.75, 833.25 to 866.75, 866.25 to 905.75, 905.25 to 955.75, 955.25 to 1023.75, 1023.25 to 1134.75, and 1134.225 to 1650 with 0.5 mass/charge ratio (m/z) overlapping between windows. The MS1 scan range was 350 to 1650 m/z and the MS1 resolution was 120,000 at 200 m/z. The MS1 full scan automatic gain control (AGC) target value was set to be 2.0×10^6, and the maximum injection time was 50 ms. The MS2 scan range was set to be 200 to 1800 m/z, and the MS2 resolution was 30,000 at 200 m/z. The normalized HCD collision energy was set at 28%. The MS2 AGC was set to be 1.5×10^6, and the maximum injection time was 52 ms. The default peptide charge state was set to 2.

Database search for proteomics and phosphoproteomics

DIA-MS data procession was performed using Spectronaut v14 (*17*, *71*) with the "DirectDIA," an optimal spectral library-free pipeline (*72*). For both proteomic and phosphoproteomic results (*4*), the DIA-MS raw datasets were searched directly against the Ensembl species-specific protein fasta files (zipped files with name ending as "pep.all.fa.gz" at useast.ensembl.org/index.html). These files include Bos_taurus.ARS-UCD1.2.pep.all.fa (for "cow"), Canis_lupus_familiaris.CanFam3.1.pep.all.fa (for "dog"), Cavia_porcellus.Cavpor3.0.pep.all.fa (for "opossum"), Equus_caballus.EquCab3.0.pep.all.fa (for "horse"), Felis_catus.Felis_catus_9.0.pep.all.fa (for "cat"), Homo_sapiens.GRCh38.pep.all.fa (for "human"), Macaca_mulatta.Mmul_10.pep.all.fa (for "monkey"), Oryctolagus_cuniculus.OryCun2.0.pep.all.fa (for "rabbit"), Ovis_aries.Oar_v3.1.pep.all.fa (for "sheep"), Rattus_norvegicus.Rnor_6.0.pep.all.fa (for "rat"), and Sus_scrofa.Sscrofa11.1.pep.all.fa, (for "pig").

In particular, for the total proteomic identification in each species, the possibilities of oxidation at methionine and acetylation at the protein N terminus were set as variable modifications, whereas carbamidomethylation at cysteine was set as a fixed modification. For the phosphoproteomic identification, the additional possibility of phosphorylation at serine/threonine/tyrosine (S/T/Y) was enabled as the variable modification. For both proteomic and phosphoproteomic datasets per species, both peptide and protein FDR (based on Q value) were both controlled at 1%. In particular, the PTM localization option in Spectronaut v14 was enabled to locate P-sites (*6*, *73*) in each species with the probability score cutoff of >0.75 (*73*), which ensures class I peptides (*74*), in which each P-site is confidently localized onto one S, T, or Y in the peptide sequence, to be identified, quantified, and reported in the results. For each localized P-site, the corresponding phosphopeptide precursors (if more than one) were averaged for quantification.

Database search for Ub linkage types

In addition, to search and determine the different types of Ub chains via its lysine residue (i.e., the "Ub code"), a "Gly-Gly" or diGly modification was set up as a variable modification in a separated search in all mammalian species, by searching the total proteomic data against the same fasta files. After the identical FDR control (i.e., 1%) at both peptide and protein levels, the diGly localization scoring in peptide sequence, the search results on the peptide level were manually inspected. The detection of K11-, K27-, K48-, and K63-linked chains were inferred on the basis of the identification of peptide precursors for TLTGK$_{GG}$TITLEVEPSDTIENVK (K11), TITLEVEPSDTIENVK$_{GG}$AK (K27), LIFAGK$_{GG}$QLEDGR (K48), and TLSDYNIQK$_{GG}$ESTLHLVLR (K63), respectively, all mapped to Ub sequence. To determine the relative quantitative variability between K48- and K63-linked chains, the tryptic peptide TITLEVEPSDTIENVK that is in the middle between K48 and K63 peptides was used for normalization purpose.

Data analysis, representation, and statistics

All the other Spectronaut settings for identification and quantification were kept as default (*4*). This means that, for example, the "missed cleavages" allowed was set at 2, the "inference correction" was enabled, the "global normalization" (on "median") was used, the quantification was performed at the MS2 level using peak areas, the "protein inference algorithm" was implemented using "IDPicker," and the top 3 peptide precursors ("min: 1 and max: 3") were summed on the basis of MS2-level peak areas for representing protein quantities in all DIA analyses. The quantitative data reported by Spectronaut analysis for proteins and P-sites were then $\log_2$-transformed for downstream statistical analysis if applicable. As for multispecies analysis, due to the database searching against Ensembl species-specific protein fasta files, the proteomic quantitative results could be directly added to the "one-to-one ortholog" data table consisting of 8138 genes (see above for RNA-seq data procession) using the ensembl gene identities, allowing the transcriptomic and proteomic quantifications to be summarized with a "gene-centric" perspective. For the absolute scale analysis, the $\log_2$-transformed mRNA, protein, and P-site quantification data are compared between molecular layers or between species. For the relative scale analysis, the mRNA, protein, and P-site quantification data of each species were compared to the averaged values across all 11 species, summarized as fold changes (FCs), and the $\log_2$-transformed FCs (i.e., individual species/averaged values of 11 species) were used for relative correlation analysis and variability analysis (i.e., by determining the SD). Alternatively, the CVs were calculated using data before log transformation for evaluating data deviation. Pearson and Spearman's correlation coefficients were calculated using R [functions cor() or cor.test() to infer statistical significance]. Wilcoxon rank sum test was used to calculate P values of SDs between functional categories.

Bioinformatic analysis

Circos-0.69-9 (http://circos.ca) (*75*) was used for the circle visualization (Fig. 1A). Functional annotation was carried out in David

Functional Annotation Tool v6.8 (https://david.ncifcrf.gov/summary.jsp) (*76*) with all detected proteins in this study as background (fig. S3). PCA was performed by NIA (National Institute of Aging) Array Analysis (Fig. 1B) (*77*). The colored scatterplots were visualized by the "heatscatter" function in R package "LSD" based on a 2D kernel density estimation (figs. S4 and S7). The annotation of surfaceome, secretome, cancer biomarkers, gene essentiality, haploinsufficiency, and age were annotated, respectively, by Cell Surface Protein Atlas (*30*), human secretome map (*29*), Human Protein Atlas (www.proteinatlas.org) (*31*), OGEE v2 (*26*), HIPred scores (*27*), and modeAge (*28*) (Fig. 2). Protein complex information was extracted from the CORUM database (*36*) (Fig. 2). The gene ontology and Kyoto Encyclopedia of Genes and Genomes annotation and 2D enrichment analysis were performed in Perseus v1.6.8.0 (*78*) (Fig. 3 and figs. S7 and S8). The retrieve of flanking amino acid sequences (±7 amino acids) of P-sites and the motif enrichment were performed by motifeR (*79*) (Fig. 5). Sequence analysis (Fig. 4) was conducted and visualized by IceLogo (https://iomics.ugent.be/icelogoserver) (*80*). The functional score can reflect the importance of P-site for organismal fitness (Fig. 5) (*43*). The sift score predicts the functional impact of missense variants based on sequence homology and the physicochemical properties of the amino acids (Fig. 5) (*81*). The melting temperature (T_m, °C) values (Fig. 5) for each P-site were taken from the reported datasets (*45*). The net phosphorylation changes were detected by total protein change correction through linear regression as reported previously (*42*). After regression, the phospho-site pairs with significant correlation (Pearson, $P < 0.001$) were used to construct interaction network by Cytoscape (*82*) (Fig. 5 and fig. S11). The kinase-substrate relations were curated from PhosphoSitePlus (*83*). To identify the homologous P-sites across 11 species from our phosphorylation datasets, the sequence windows (±7 amino acid flanking sequence) of detected P-sites in one-to-one orthologous proteins were aligned by the "pairwiseAlignment" function in R package "Biostrings." The sequence windows with a score ≥ 10 and in orthologous proteins were considered as homologous. Data from other species were first mapped to human data, and then the common P-sites were obtained as the intersection of all species. The number and relationship of protein-protein interactions were curated using STRING v11.0 (https://string-db.org) (*25*).

Species are the product of a process of lineage splitting (speciation) and divergence and are thus not statistically independent random events. The most broadly used method to account for phylogenetic structure is the method of PICs where observations on N species are transformed into N-1 contrasts (differences) (*84*, *85*). The method estimates the most likely history of evolutionary change in variables such that the contrasts reflect independent evolutionary changes. PIC values were computed by R package APE (analysis of phylogenetics and evolution) using the function "pic" (*84*, *85*) where the gene and protein expression values were iteratively fed into the pic function to obtain an array of resultant PICs (*86*).

Website inventory for proteome-centric multispecies navigation

Because of the well-matched multi-omic layers especially the application of consistent DIA-MS, we consider our dataset a high-quality resource for future mammalian evolution and gene expression studies. We thus generated a website https://yslproteomics.shinyapps.io/Evolutome/ to facilitate the navigation of the data basis (fig. S12). This website interactively provides queries about the abundances for any transcript or protein in every species. It additionally offers heatmaps and scatterplots between molecular layers for individual gene or gene sets of interest.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn0756

REFERENCES AND NOTES

1. J. B. Müller, P. E. Geyer, A. R. Colaço, P. V. Treit, M. T. Strauss, M. Oroshi, S. Doll, S. V. Winter, J. M. Bader, N. Köhler, F. Theis, A. Santos, M. Mann, The proteome landscape of the kingdoms of life. *Nature* **582**, 592–596 (2020).
2. Z.-Y. Wang, E. Leushkin, A. Liechti, S. Ovchinnikova, K. Mößinger, T. Brüning, C. Rummel, F. Grützner, M. Cardoso-Moreira, P. Janich, D. Gatfield, B. Diagouraga, B. de Massy, M. E. Gill, A. H. F. M. Peters, S. Anders, H. Kaessmann, Transcriptome and translatome co-evolution in mammals. *Nature* **588**, 642–647 (2020).
3. R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
4. E. Gao, W. Li, C. Wu, W. Shao, Y. Dia, Y. Liu, Data-independent acquisition-based proteome and phosphoproteome profiling across six melanoma cell lines reveals determinants of proteotypes. *Mol. Omics* **17**, 413–425 (2021).
5. P. Navarro, J. Kuharev, L. C. Gillet, O. M. Bernhardt, B. M. Lean, H. L. Röst, S. A. Tate, C.-C. Tsou, L. Reiter, U. Distler, G. Rosenberger, Y. Perez-Riverol, A. I. Nesvizhskii, R. Aebersold, S. Tenzer, A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* **34**, 1130–1136 (2016).
6. G. Rosenberger, Y. Liu, H. L. Röst, C. Ludwig, A. Buil, A. Bensimon, M. Soste, T. D. Spector, E. T. Dermitzakis, B. C. Collins, L. Malmström, R. Aebersold, Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS. *Nat. Biotechnol.* **35**, 781–788 (2017).
7. B. C. Collins, C. L. Hunter, Y. Liu, B. Schilling, G. Rosenberger, S. L. Bader, D. W. Chan, B. W. Gibson, A.-C. Gingras, J. M. Held, M. Hirayama-Kurogi, G. Hou, C. Krisp, B. Larsen, L. Lin, S. Liu, M. P. Molloy, R. L. Moritz, S. Ohtsuki, R. Schlapbach, N. Selevsek, S. N. Thomas, S.-C. Tzeng, H. Zhang, R. Aebersold, Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **8**, 291 (2017).
8. Y. Xuan, N. W. Bateman, T. P. Conrads, Standardization and harmonization of distributed multi-center proteotype analysis supporting precision medicine studies. *Nat. Commun.* **11**, 5248 (2020).
9. K. Swovick, K. A. Welle, J. R. Hryhorenko, A. Seluanov, V. Gorbunova, S. Ghaemmaghami, Cross-species comparison of proteome turnover kinetics. *Mol. Cell. Proteomics* **17**, 580–591 (2018).
10. P. Picotti, M. Clément-Ziza, H. Lam, D. S. Campbell, A. Schmidt, E. W. Deutsch, H. Röst, Z. Sun, O. Rinner, L. Reiter, Q. Shen, J. J. Michaelson, A. Frei, S. Alberti, U. Kusebauch, B. Wollscheid, R. L. Moritz, A. Beyer, R. Aebersold, A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266–270 (2013).
11. N. Romanov, M. Kuhn, R. Aebersold, A. Ori, M. Beck, P. Bork, Disentangling genetic and environmental effects on the proteotypes of individuals. *Cell* **177**, 1308–1318.e10 (2019).
12. J. D. Venable, M. Q. Dong, J. Wohlschlegel, A. Dillin, J. R. Yates, Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).
13. L. C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).
14. C. Ding, Y. Li, F. Guo, Y. Jiang, W. Ying, D. Li, D. Yang, X. Xia, W. Liu, Y. Zhao, Y. He, X. Li, W. Sun, Q. Liu, L. Song, B. Zhen, P. Zhang, X. Qian, J. Qin, F. He, A cell-type-resolved liver proteome. *Mol. Cell. Proteomics* **15**, 3190–3202 (2016).
15. D. P. Nusinow, J. Szpyt, M. Ghandi, C. M. Rose, E. R. McDonald III, M. Kalocsay, J. Jané-Valbuena, E. Gelfand, D. K. Schweppe, M. Jedrychowski, J. Golji, D. A. Porter, T. Rejtar, Y. K. Wang, G. V. Kryukov, F. Stegmeier, B. K. Erickson, L. A. Garraway, W. R. Sellers, S. P. Gygi, Quantitative proteomics of the cancer cell line encyclopedia. *Cell* **180**, 387–402.e16 (2020).
16. S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, S. Kasela, V. Wucher, S. E. Castel, A. R. Hamel, A. Viñuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N. Barbeira, D. Garrido-Martín, B. B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria; GTEx Consortium, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigó, A. V. Segrè, B. E. Stranger, K. G. Ardlie, T. Lappalainen, Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).

17. R. Bruderer, O. M. Bernhardt, T. Gandhi, Y. Xuan, J. Sondermann, M. Schmidt, D. Gomez-Varela, L. Reiter, Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* **16**, 2296–2309 (2017).
18. G. Rosenberger, I. Bludau, U. Schmitt, M. Heusel, C. L. Hunter, Y. Liu, M. J. MacCoss, B. X. MacLean, A. I. Nesvizhskii, P. G. A. Pedrioli, L. Reiter, H. L. Röst, S. Tate, Y. S. Ting, B. C. Collins, R. Aebersold, Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat. Methods* **14**, 921–927 (2017).
19. K. L. Howe, P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M. Ridwan Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, B. E. Houdaigui, R. Fatima, A. Gall, C. G. Giron, T. Grego, C. Guijarro-Clarke, L. Haggerty, A. Hemrom, T. Hourlier, O. G. Izuogu, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, J. C. Marugán, T. Maurel, A. C. McMahon, S. Mohanan, B. Moore, M. Muffato, D. N. Oheh, D. Paraschas, A. Parker, A. Parton, I. Prosovetskaia, M. P. Sakthivel, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, E. Steed, M. Szpak, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, B. Walts, A. Winterbottom, M. Chakiachvili, A. Chaubal, N. De Silva, B. Flint, A. Frankish, S. E. Hunt, G. R. Ilsley, N. Langridge, J. E. Loveland, F. J. Martin, J. M. Mudge, J. Morales, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, F. Cunningham, A. D. Yates, D. R. Zerbino, P. Flicek, Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
20. C. Wu, Q. Ba, D. Lu, W. Li, B. Salovska, P. Hou, T. Mueller, G. Rosenberger, E. Gao, Y. Di, H. Zhou, E. F. Fornasiero, Y. Liu, Global and site-specific effect of phosphorylation on protein turnover. *Dev. Cell* **56**, 111–124.e6 (2021).
21. Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
22. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
23. C. Vogel, E. M. Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
24. T. Garland Jr., A. R. Ives, Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* **155**, 346–364 (2000).
25. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. von Mering, STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
26. W. H. Chen, G. Lu, X. Chen, X. M. Zhao, P. Bork, OGEE v2: An update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res.* **45**, D940–D944 (2017).
27. H. A. Shihab, M. F. Rogers, C. Campbell, T. R. Gaunt, HIPred: An integrative approach to predicting haploinsufficient genes. *Bioinformatics* **33**, 1751–1757 (2017).
28. B. J. Liebeskind, C. D. McWhite, E. M. Marcotte, Towards consensus gene ages. *Genome Biol. Evol.* **8**, 1812–1823 (2016).
29. M. Uhlén, M. J. Karlsson, A. Hober, A.-S. Svensson, J. Scheffel, D. Kotol, W. Zhong, A. Tebani, L. Strandberg, F. Edfors, E. Sjöstedt, J. Mulder, A. Mardinoglu, A. Berling, S. Ekblad, M. Dannemeyer, S. Kanje, J. Rockberg, M. Lundqvist, M. Malm, A.-L. Volk, P. Nilsson, A. Månberg, T. Dodig-Crnkovic, E. Pin, M. Zwahlen, P. Oksvold, K. von Feilitzen, R. S. Häussler, M.-G. Hong, C. Lindskog, F. Ponten, B. Katona, J. Vuu, E. Lindström, J. Nielsen, J. Robinson, B. Ayoglu, D. Mahdessian, D. Sullivan, P. Thul, F. Danielsson, C. Stadler, E. Lundberg, G. Bergström, A. Gummesson, B. G. Voldborg, H. Tegel, S. Hober, B. Forsström, J. M. Schwenk, L. Fagerberg, Å. Sivertsson, The human secretome. *Science Signal.* **12**, eaaz0274 (2019).
30. D. Bausch-Fluck, U. Goldmann, S. Müller, M. van Oostrum, M. Müller, O. T. Schubert, B. Wollscheid, The in silico human surfaceome. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10988–E10997 (2018).
31. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. M. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
32. J. Cox, M. Mann, 1D and 2D annotation enrichment: A statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13** (Suppl. 16), S12 (2012).
33. A. Rousseau, A. Bertolotti, An evolutionarily conserved pathway controls proteasome homeostasis. *Nature* **536**, 184–189 (2016).
34. A.-C. Gingras, M. Caballero, M. Zarske, A. Sanchez, T. R. Hazbun, S. Fields, N. Sonenberg, E. Hafen, B. Raught, R. Aebersold, A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell. Proteomics* **4**, 1725–1740 (2005).
35. Y. Liu, C. Borel, L. Li, T. Müller, E. G. Williams, P.-L. Germain, M. Buljan, T. Sajic, P. J. Boersema, W. Shao, M. Faini, G. Testa, A. Beyer, S. E. Antonarakis, R. Aebersold, Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. *Nat. Commun.* **8**, 1212 (2017).
36. M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: The comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).
37. Y. Liu, Y. Mi, T. Mueller, S. Kreibich, E. G. Williams, A. Van Drogen, C. Borel, M. Frank, P.-L. Germain, I. Bludau, M. Mehnert, M. Seifert, M. Emmenlauer, I. Sorg, F. Bezrukov, F. S. Bena, H. Zhou, C. Dehio, G. Testa, J. Saez-Rodriguez, S. E. Antonarakis, W.-D. Hardt, R. Aebersold, Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat. Biotechnol.* **37**, 314–322 (2019).
38. N. Dephoure, S. Hwang, C. O'Sullivan, S. E. Dodgson, S. P. Gygi, A. Amon, E. M. Torres, Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *eLife* **3**, e03023 (2014).
39. C. Jüschke, I. Dohnal, P. Pichler, H. Harzer, R. Swart, G. Ammerer, K. Mechtler, J. A. Knoblich, Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. *Genome Biol.* **14**, r133 (2013).
40. S. Murata, H. Udono, N. Tanahashi, N. Hamada, K. Watanabe, K. Adachi, T. Yamano, K. Yui, N. Kobayashi, M. Kasahara, K. Tanaka, T. Chiba, Immunoproteasome assembly and antigen presentation in mice lacking both PA28alpha and PA28beta. *EMBO J.* **20**, 5898–5907 (2001).
41. W. Kim, E. J. Bennett, E. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. Sowa, R. Rad, J. Rush, M. J. Comb, J. W. Harper, S. P. Gygi, Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).
42. T. I. Roumeliotis, S. P. Williams, E. Gonçalves, C. Alsinet, M. Del Castillo Velasco-Herrera, N. Aben, F. Z. Ghavidel, M. Michaut, M. Schubert, S. Price, J. C. Wright, L. Yu, M. Yang, R. Dienstmann, J. Guinney, P. Beltrao, A. Brazma, M. Pardo, O. Stegle, D. J. Adams, L. Wessels, J. Saez-Rodriguez, U. M. Dermott, J. S. Choudhary, Genomic determinants of protein abundance variation in colorectal cancer cells. *Cell Rep.* **20**, 2201–2214 (2017).
43. D. Ochoa, A. F. Jarnuczak, C. Viéitez, M. Gehre, M. Soucheray, A. Mateus, A. A. Kleefeldt, A. Hill, L. Garcia-Alonso, F. Stein, N. J. Krogan, M. M. Savitski, D. L. Swaney, J. A. Vizcaíno, K.-M. Noh, P. Beltrao, The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **38**, 365–373 (2020).
44. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
45. J. X. Huang, G. Lee, K. E. Cavanaugh, J. W. Chang, M. L. Gardel, R. E. Moellering, High throughput discovery of functional protein modifications by hotspot thermal profiling. *Nat. Methods* **16**, 894–901 (2019).
46. J. R. Warner, The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
47. T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
48. I. Dikic, Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).
49. B. Tancini, S. Buratta, F. Delo, K. Sagini, E. Chiaradia, R. M. Pellegrino, C. Emiliani, L. Urbanelli, Lysosomal exocytosis: The extracellular role of an intracellular organelle. *Membranes* **10**, 406 (2020).
50. M. Tracz, W. Bialek, Beyond K48 and K63: Non-canonical protein ubiquitination. *Cell. Mol. Biol. Lett.* **26**, 1 (2021).
51. H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: Diversification, exon definition and function. *Nat. Rev. Genet.* **11**, 345–355 (2010).
52. Y. Xing, C. Lee, Alternative splicing and RNA selection pressure—Evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* **7**, 499–509 (2006).
53. J. Boekhorst, B. van Breukelen, A. Heck Jr., B. Snel, Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol.* **9**, R144 (2008).
54. A. M. Pickering, M. Lehr, C. M. Gendron, S. D. Pletcher, R. A. Miller, Mitochondrial thioredoxin reductase 2 is elevated in long-lived primate as well as rodent species and extends fly mean lifespan. *Aging Cell* **16**, 683–692 (2017).
55. S. Ma, A. Upneja, A. Galecki, Y.-M. Tsai, C. F. Burant, S. Raskind, Q. Zhang, Z. D. Zhang, A. Seluanov, V. Gorbunova, C. B. Clish, R. A. Miller, V. N. Gladyshev, Cell culture-based profiling across mammals reveals DNA repair and metabolism as determinants of species longevity. *eLife* **5**, e19130 (2016).
56. Kshitiz, J. Afzal, J. D. Maziarz, A. Hamidzadeh, C. Liang, E. M. Erkenbrack, H. N. Kim, J.-D. Haeger, C. Pfarrer, T. Hoang, T. Ott, T. Spencer, M. Pavličev, D. F. Antczak, A. Levchenko, G. P. Wagner, Evolution of placental invasion and cancer metastasis are causally linked. *Nat. Ecol. Evol.* **3**, 1743–1753 (2019).

57. G. P. Wagner, Kshitiz, A. Dighe, A. Levchenko, The coevolution of placentation and cancer. *Annu. Rev. Anim. Biosci.* **10**, 259–279 (2022).
58. K. Benirschke, The frozen zoo concept. *Zoo Biol.* **3**, 325–328 (1984).
59. K. Swovick, D. Firsanov, K. A. Welle, J. R. Hryhorenko, J. P. Wise Sr., C. George, T. L. Sformo, A. Seluanov, V. Gorbunova, S. Ghaemmaghami, Interspecies differences in proteome turnover kinetics are correlated with life spans and energetic demands. *Mol. Cell. Proteomics* **20**, 100041 (2021).
60. A. Jarzab, N. Kurzawa, T. Hopf, M. Moerch, J. Zecha, N. Leijten, Y. Bian, E. Musiol, M. Maschberger, G. Stoehr, I. Becher, C. Daly, P. Samaras, J. Mergner, B. Spanier, A. Angelov, T. Werner, M. Bantscheff, M. Wilhelm, M. Klingenspor, S. Lemeer, W. Liebl, H. Hahne, M. M. Savitski, B. Kuster, Meltome atlas-thermal proteome stability across the tree of life. *Nat. Methods* **17**, 495–503 (2020).
61. C. Wan, B. Borgeson, S. Phanse, F. Tu, K. Drew, G. Clark, X. Xiong, O. Kagan, J. Kwan, A. Bezginov, K. Chessman, S. Pal, G. Cromar, O. Papoulas, Z. Ni, D. R. Boutz, S. Stoilova, P. C. Havugimana, X. Guo, R. H. Malty, M. Sarov, J. Greenblatt, M. Babu, W. Brent Derry, E. R. Tillier, J. B. Wallingford, J. Parkinson, E. M. Marcotte, A. Emili, Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).
62. M. E. Nelson, B. L. Parker, J. G. Burchfield, N. J. Hoffman, E. J. Needham, K. C. Cooke, T. Naim, L. Sylow, N. X. Ling, D. Francis, D. M. Norris, R. Chaudhuri, J. S. Oakhill, E. A. Richter, G. S. Lynch, J. Stöckli, D. E. James, Phosphoproteomics reveals conserved exercise-stimulated signaling and AMPK regulation of store-operated calcium entry. *EMBO J.* **38**, e102578 (2019).
63. A. Sebé-Pedrós, M. I. Peña, S. Capella-Gutiérrez, M. Antó, T. Gabaldón, I. Ruiz-Trillo, E. Sabidó, High-throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. *Dev. Cell* **39**, 186–197 (2016).
64. M. A. Kisiel, A. S. Klar, Isolation and culture of human dermal fibroblasts. *Methods Mol. Biol.* **1993**, 71–78 (2019).
65. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
66. G. P. Wagner, K. Kin, V. J. Lynch, Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
67. A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. Ridwan Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. M. Dowall, A. M. Mahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. IIsley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, P. Flicek, Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
68. Q. Gao, H. Zhu, L. Dong, W. Shi, R. Chen, Z. Song, C. Huang, J. Li, X. Dong, Y. Zhou, Q. Liu, L. Ma, X. Wang, J. Zhou, Y. Liu, E. Boja, A. I. Robles, W. Ma, P. Wang, Y. Li, L. Ding, B. Wen, B. Zhang, H. Rodriguez, D. Gao, H. Zhou, J. Fan, Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* **179**, 561–577.e22 (2019).
69. M. Mehnert, W. Li, C. Wu, B. Salovska, Y. Liu, Combining rapid data independent acquisition and CRISPR gene deletion for studying potential protein functions: A case of HMGN1. *Proteomics* , e1800438 (2019).
70. B. Salovska, W. Li, Y. Di, Y. Liu, BoxCarmax: A high-selectivity data-independent acquisition mass spectrometry method for the analysis of protein turnover and complex samples. *Anal. Chem.* **93**, 3103–3111 (2021).
71. R. Bruderer, O. M. Bernhardt, T. Gandhi, S. M. Miladinović, L.-Y. Cheng, S. Messner, T. Ehrenberger, V. Zanotelli, Y. Butscheid, C. Escher, O. Vitek, O. Rinner, L. Reiter, Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410 (2015).
72. C.-C. Tsou, D. Avtonomov, B. Larsen, M. Tucholska, H. Choi, A.-C. Gingras, A. I. Nesvizhskii, DIA-Umpire: Comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12**, 258–264 (2015).
73. D. B. Bekker-Jensen, O. M. Bernhardt, A. Hogrebe, A. Martinez-Val, L. Verbeke, T. Gandhi, C. D. Kelstrup, L. Reiter, J. V. Olsen, Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, 787 (2020).
74. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
75. M. Krzywinski, J. Schein, İ. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, M. A. Marra, Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
76. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
77. A. A. Sharov, D. B. Dudekula, M. S. Ko, A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* **21**, 2548–2549 (2005).
78. S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
79. S. Wang, Y. Cai, J. Cheng, W. Li, Y. Liu, H. Yang, motifeR: An integrated web software for identification and visualization of protein posttranslational modification motifs. *Proteomics* **19**, e1900245 (2019).
80. N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **6**, 786–787 (2009).
81. R. Vaser, S. Adusumalli, S. N. Leng, M. Sikic, P. C. Ng, SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
82. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
83. P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham, E. Skrzypek, PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
84. J. Felsenstein, Phylogenies and the comparative method. *Am. Naturalist* **125**, 1–15 (1985).
85. P. H. Harvey, M. D. Pagel, *The Comparative Method in Evolutionary Biology* (Oxford series in ecology and evolution, Oxford Univ. Press, 1991), pp. viii, 239 pp.
86. E. Paradis, *Analysis of Phylogenetics and Evolution with R. Use R!* (Springer, 2006), pp. xii, 211 pp.
87. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yılmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Acknowledgments

Funding: Y.L. thanks the support from the National Institute of General Medical Sciences (NIGMS) and National Institutes of Health (NIH) through grant R01GM137031. This research was supported in part by a Pilot Grant from Yale Cancer Center and in part by a Career Enhancement Program Grant from the Yale SPORE in Lung Cancer (1P50CA196530), as well as a grant from the NCI with the number U54:CA209992. We thank L. Reiter and T. Gandhi (Biognosys AG, Switzerland) for technical support in Spectronaut and M. A. Lemmon and A. Levchenko (Yale) for helpful discussions. **Author contributions:** Conceptualization: Q.B., G.P.W., and Y.L. Methodology: Q.B., A.D., S.W., G.P.W., and Y.L. Investigation: Q.B., Y.H., A.D., J.M., and I.P. Visualization: Q.B., Y.L., and G.P.W. Supervision: Y.L. and G.P.W. Writing—original draft: Q.B. and Y.L. Writing—review and editing: A.D., W.L., J.M., I.P., S.W., G.P.W., and Y.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The RNA-seq data were uploaded to GEO repository and are available at Sequence Read Archive (SRA) of NCBI, under BioProject: PRJNA674672. The MS data and reports have been deposited to the ProteomeXchange Consortium via the PRIDE (*87*) partner repository with the dataset identifier PXD028979. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 1 November 2021
Accepted 25 July 2022
Published 9 September 2022
10.1126/sciadv.abn0756