RESEARCH ARTICLE SUMMARY

PROTEOMICS

A subcellular map of the human proteome

Peter J. Thul,* Lovisa Åkesson,* Mikaela Wiking, Diana Mahdessian, Aikaterini Geladaki, Hammou Ait Blal, Tove Alm, Anna Asplund, Lars Björk, Lisa M. Breckels, Anna Bäckström, Frida Danielsson, Linn Fagerberg, Jenny Fall, Laurent Gatto, Christian Gnann, Sophia Hober, Martin Hjelmare, Fredric Johansson, Sunjae Lee, Cecilia Lindskog, Jan Mulder, Claire M. Mulvey, Peter Nilsson, Per Oksvold, Johan Rockberg, Rutger Schutten, Jochen M. Schwenk, Åsa Sivertsson, Evelina Sjöstedt, Marie Skogs, Charlotte Stadler, Devin P. Sullivan, Hanna Tegel, Casper Winsnes, Cheng Zhang, Martin Zwahlen, Adil Mardinoglu, Fredrik Pontén, Kalle von Feilitzen, Kathryn S. Lilley, Mathias Uhlén,† Emma Lundberg†

INTRODUCTION: A complete view of human biology can only be achieved by studying the molecular components of its smallest functional unit, the cell. Cells are internally organized into compartments called organelles. The spatial partitioning provided by organelles creates an enclosed environment or surface for chemical reactions tailored to fulfill specific functions. These functions are tightly linked to a specific set of proteins. Therefore, resolving the subcellular location of the human proteome provides information about the function of the organelle and its underlying cellular mechanisms. We present a subcellular map of the human proteome, called the Cell Atlas, to facilitate functional exploration of individual proteins and their role in human biology and disease.

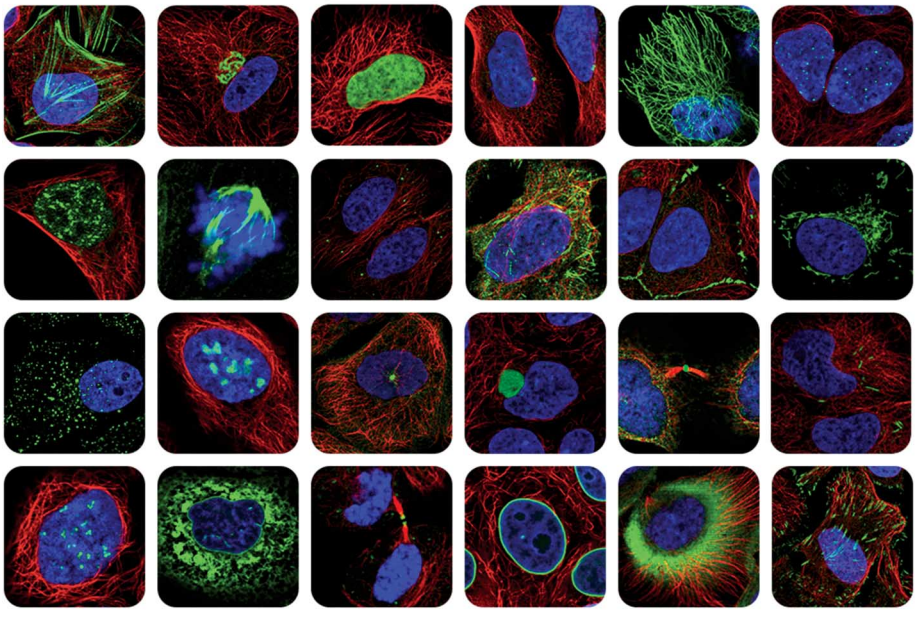
RATIONALE: Immunofluorescence (IF) microscopy was used to systematically resolve the spatial distribution of human proteins in cultivated cell lines and map them to cellular compartments and substructures with single-cell resolution. This approach allowed definition of the precise location of a majority of the human proteins in their cellular context and exploration of single-cell variations in protein expression patterns. The proteome-wide information about protein spatial distribution was validated with an orthogonal proteomics method, and the results were integrated into existing network models of protein-protein interactions for increased accuracy.

RESULTS: We report a high-resolution characterization of the spatial subcellular distribution of the human proteome based on more than 80,000 confocal IF images. A total of 12,003 proteins targeted by 13,993 antibodies were classified into one or several of 30 cellular compartments and substructures, altogether defining the proteomes of 13 major organelles. The organelles with the largest proteomes were the nucleus and its substructures (6245 proteins), such as bodies and speckles, and the cytosol (4279 proteins). However, smaller organelles such as the midbody, rods and rings, and nucleoli also showed a larger diversity than previously recognized. Intriguingly, about half of all proteins were localized to multiple compartments, showing that there is a shared pool of proteins even among functionally unrelated organelles. Single-cell analysis revealed 1855 proteins with variation in their expression pattern, either in terms of expression levels or spatial distribution. Last, the spatial information was used to refine biological networks. Our location-pruned network that restricts protein interaction to the same organelle improved the accuracy of the human interactome model. The analysis also included transcriptomics data for all putative protein-coding genes (19,628) in 56 human cell lines of various origins. On average, cell lines expressed 11,490 genes, with half of them (6295) being expressed across all samples, suggesting a "housekeeping" role.

CONCLUSION: The cellular proteome is compartmentalized and spatiotemporally regulated to a high degree. The high-resolution subcellular map of the human proteome that we provide describes this cellular complexity, with many multilocalizing proteins and single-cell variations. The map is presented as an interactive database called the Cell Atlas, part of the Human Protein Atlas (www.proteinatlas.org). The Cell Atlas constitutes a key resource for a holistic understanding of the human cell and its complex underlying molecular machinery, as well as a major step toward modeling the human cell. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aal3321



Creation of an image-based map of the human subcellular proteome. The subcellular locations of 12,003 proteins were determined by IF microscopy in cell lines of various origins. High-resolution IF images such as those shown above enabled mapping of proteins to distinct subcellular structures. This resulted in the definition of the proteomes of 13 major cellular organelles, revealing multilocalizing proteins, as well as expression variability on a single-cell level.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: mathias.uhlen@scilifelab.se (M.U.); emma.lundberg@scilifelab.se (E.L.)
Cite this article as P. J. Thul *et al.*, *Science* **356**, eaal3321 (2017). DOI: 10.1126/science.aal3321

RESEARCH ARTICLE

PROTEOMICS

A subcellular map of the human proteome

Peter J. Thul,[1]* Lovisa Åkesson,[1]* Mikaela Wiking,[1] Diana Mahdessian,[1] Aikaterini Geladaki,[2,3] Hammou Ait Blal,[1] Tove Alm,[1] Anna Asplund,[4] Lars Björk,[1] Lisa M. Breckels,[2,5] Anna Bäckström,[1] Frida Danielsson,[1] Linn Fagerberg,[1] Jenny Fall,[1] Laurent Gatto,[2,5] Christian Gnann,[1] Sophia Hober,[6] Martin Hjelmare,[1] Fredric Johansson,[1] Sunjae Lee,[1] Cecilia Lindskog,[4] Jan Mulder,[7] Claire M. Mulvey,[2] Peter Nilsson,[1] Per Oksvold,[1] Johan Rockberg,[6] Rutger Schutten,[1] Jochen M. Schwenk,[1] Åsa Sivertsson,[1] Evelina Sjöstedt,[4] Marie Skogs,[1] Charlotte Stadler,[1] Devin P. Sullivan,[1] Hanna Tegel,[6] Casper Winsnes,[1] Cheng Zhang,[1] Martin Zwahlen,[1] Adil Mardinoglu,[1] Fredrik Pontén,[4] Kalle von Feilitzen,[1] Kathryn S. Lilley,[2] Mathias Uhlén,[1]† Emma Lundberg[1]†

Resolving the spatial distribution of the human proteome at a subcellular level can greatly increase our understanding of human biology and disease. Here we present a comprehensive image-based map of subcellular protein distribution, the Cell Atlas, built by integrating transcriptomics and antibody-based immunofluorescence microscopy with validation by mass spectrometry. Mapping the in situ localization of 12,003 human proteins at a single-cell level to 30 subcellular structures enabled the definition of the proteomes of 13 major organelles. Exploration of the proteomes revealed single-cell variations in abundance or spatial distribution and localization of about half of the proteins to multiple compartments. This subcellular map can be used to refine existing protein-protein interaction networks and provides an important resource to deconvolute the highly complex architecture of the human cell.

Spatial partitioning of biological functions is a phenomenon that is fundamental to life. In humans, this spatial partitioning constitutes a hierarchy of specialized systems ranging across scales—from organs to specialized cells to subcellular structures, down to macromolecular complexes. At the cellular level, proteins function at specific times and subcellular locations, such as organelles. These locations provide a specific chemical environment and set of interaction partners that are necessary to fulfill the protein's function. Mislocalization of proteins can be associated with cellular dysfunction and disease (*1*, *2*). Thus, knowledge of the spatial distribution of proteins at a subcellular level is essential for understanding protein function, interactions, and cellular mechanisms.

Several approaches for systematic analysis of protein localizations have been described. Quantitative mass spectrometric readouts allow identification of proteins with similar distribution profiles across fractionation gradients (*3*–*7*) or proteins labeled by proximity-dependent enzymatic reactions in cells (*8*–*11*). In contrast, imaging-based approaches using tagged proteins (*12*–*14*) or affinity reagents (*15*, *16*) enable exploration of the subcellular distribution of proteins in situ in single cells and can also effectively identify cell-to-cell variability and multi-organelle distribution. Complementary to these experimental methods, a number of in silico approaches have been used to predict subcellular localization in eukaryotic cells [e.g., (*17*, *18*)]. The manually curated UniProt database (*19*) is an important resource for protein localization that collects subcellular data from literature and external databases for a large number of species. Despite these efforts, experimental data on subcellular localization are lacking for the majority of human proteins. To address this need, pilot studies have been initiated to probe human proteins by means of immunofluorescence (IF) and high-resolution confocal microscopy (*15*, *20*, *21*) and mass spectrometry (*7*). To date, maps of the subcellular proteome of murine stem cells (*6*), HeLa cells (*7*), and rat liver (*22*) are the best-characterized data sets for mammals.

Here we report the establishment of the Cell Atlas—a comprehensive, proteome-wide knowledge resource for subcellular localization in human cells—within the framework of the Human Protein Atlas (HPA) (*23*, *24*). By integration of transcriptomics data and an antibody-based image-profiling approach, we provide experimental localization data for 12,003 proteins, using a panel of 22 human cell lines and 13,993 antibodies. The spatial distribution of these proteins is resolved to 30 cellular structures and substructures, altogether representing 13 major organelles. Particular emphases were on defining the organelle proteomes and describing multilocalizing proteins and proteins displaying single-cell variability. We expect the availability of localization information for the human proteome to complement other systematic efforts on the DNA (*25*, *26*), RNA (*27*, *28*), and proteome (*19*, *29*) levels and aid in the molecular understanding of the human cell and its interactions.

Cell lines and transcriptomics analysis

The aim in creating the Cell Atlas was to define the proteomes of organelles and subcellular compartments by IF imaging (Fig. 1). To select suitable cell lines for the effort, transcriptomics analysis using RNA sequencing (RNA-seq) was performed on 56 human cell lines from various origins representing different germ layers and tissues (table S1). A hierarchical clustering analysis based on RNA-seq data (Fig. 2A) showed that cell lines of similar origin or phenotype clustered together, indicating a common pattern of gene expression. Prominent clusters included myeloid cell lines, lymphoid cell lines, endothelial cells, and cells immortalized by introduction of telomerase. Twenty-two cell lines were selected for IF imaging—together expressing 84% of all protein-coding genes (16,504 of 19,628) predicted by Ensembl [version 83.38 (*26*)]—based on a transcripts-per-million (TPM) cutoff of ≥1 (table S2). Interestingly, by applying TPM values, the average number of expressed genes in the sequenced cell lines was 11,490 (table S2), and the range spanned from 10,136 in Daudi cells (B lymphoblast) to 12,816 in SCLC-21H cells (small cell lung carcinoma). This is notably less than the previously measured average of ~14,000 transcripts obtained using FPKM values (fragments per kilobase of transcript per million mapped reads) as a normalization method. However, the TPM-based number corresponds more accurately to the number of proteins actually detected in this and other proteomic studies (*30*, *31*).

A classification of the RNA expression levels according to the principle previously described in (*24*) was performed to define genes expressed in all cell lines and those expressed in a cell line–restricted manner (fig. S1). About one-third (6295) of the protein-coding genes were expressed in all cell lines, suggesting a "housekeeping" role, whereas 45% showed a more variable expression. Eleven percent (2090) were not detected in any of the analyzed cell lines. Of these genes, 1225 were detected in tissues, suggesting that they code for proteins restricted to a smaller number of specialized cell types or representing specific developmental stages (table S3). Functional annotations from Gene Ontology (GO) support

[1]Science for Life Laboratory, School of Biotechnology, KTH Royal Institute of Technology, SE-171 21 Stockholm, Sweden. [2]Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. [3]Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. [4]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden. [5]Computational Proteomics Unit, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. [6]Department of Proteomics, School of Biotechnology, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden. [7]Science for Life Laboratory, Department of Neuroscience, Karolinska Institute, SE-171 77 Stockholm, Sweden.
*These authors contributed equally to this work.
†Corresponding author. Email: mathias.uhlen@scilifelab.se (M.U.); emma.lundberg@scilifelab.se (E.L.)

this hypothesis, showing enrichment for tissue-restricted proteins, such as receptors in the sensory cells or reproduction-related proteins (table S4).

Creation of a subcellular map

As an integrated part of the HPA effort (*23*), antibodies have been generated, affinity-purified using the antigen, and validated by protein microarray analysis to ensure specific and selective binding to the intended target antigen (*32*). These antibodies cover the majority of all predicted human protein-coding genes. A systematic workflow for subcellular localization of proteins was established that uses IF and high-resolution confocal microscopy, as described in fig. S2 (*15*, *16*). Altogether, 13,993 antibodies (13,073 antibodies generated by the HPA project, complemented with 920 commercially available antibodies) were selected to be included in the Cell Atlas after reliability analysis. Every antibody was used for immunostaining of the bone osteosarcoma–derived U-2 OS cell line and two additional cell lines from the panel showing a high expression of the target gene. In addition to the antibody of interest, reference markers outlining the nucleus, microtubules, and endoplasmic reticulum (ER) were included in each sample (fig. S3). For all proteins, the spatial expression patterns observed in the confocal images were assigned to one or more of 30 cellular organelles and substructures (Fig. 1 and table S5) and classified by a location-specific reliability score, as outlined below. The images and primary data are presented in the Cell Atlas in a gene-centric manner, including the classification of all images and a description of the validation and reliability of the antibodies and identified locations. Furthermore, the images were annotated by a citizen science approach through the Project Discovery platform within EVE Online, a massive multiplayer online game; more than 180,000 players across the world have contributed more than 7 million minutes of active participation to date (*33*). In total, the Cell Atlas (in version 16.1 of the HPA) contains 82,152 high-resolution annotated images covering 61% of all human protein-coding genes and 73% of the genes expressed in the IF cell line panel. The complete localization data set containing the results for all proteins in the Cell Atlas, as well as all successful stainings obtained in the different cell lines, are given in tables S6 and S7, respectively.

Validation of data in the Cell Atlas

Recently, there have been many articles questioning the quality and use of antibodies in research [e.g., (*34*, *35*)]. Because off-target antibody binding can cause false-positive results, efforts have gone into manually annotating all antibodies regarding their reliability and quality of the staining. In the Cell Atlas, we provide a reliability score for every annotated location and protein on a four-tiered scale: "validated," "supported," "approved," and "uncertain." Locations obtained the score "validated" if the antibody was validated according to one of the validation "pillars" proposed by an international working group (*36*) as suitable for IF: (i) genetic methods using

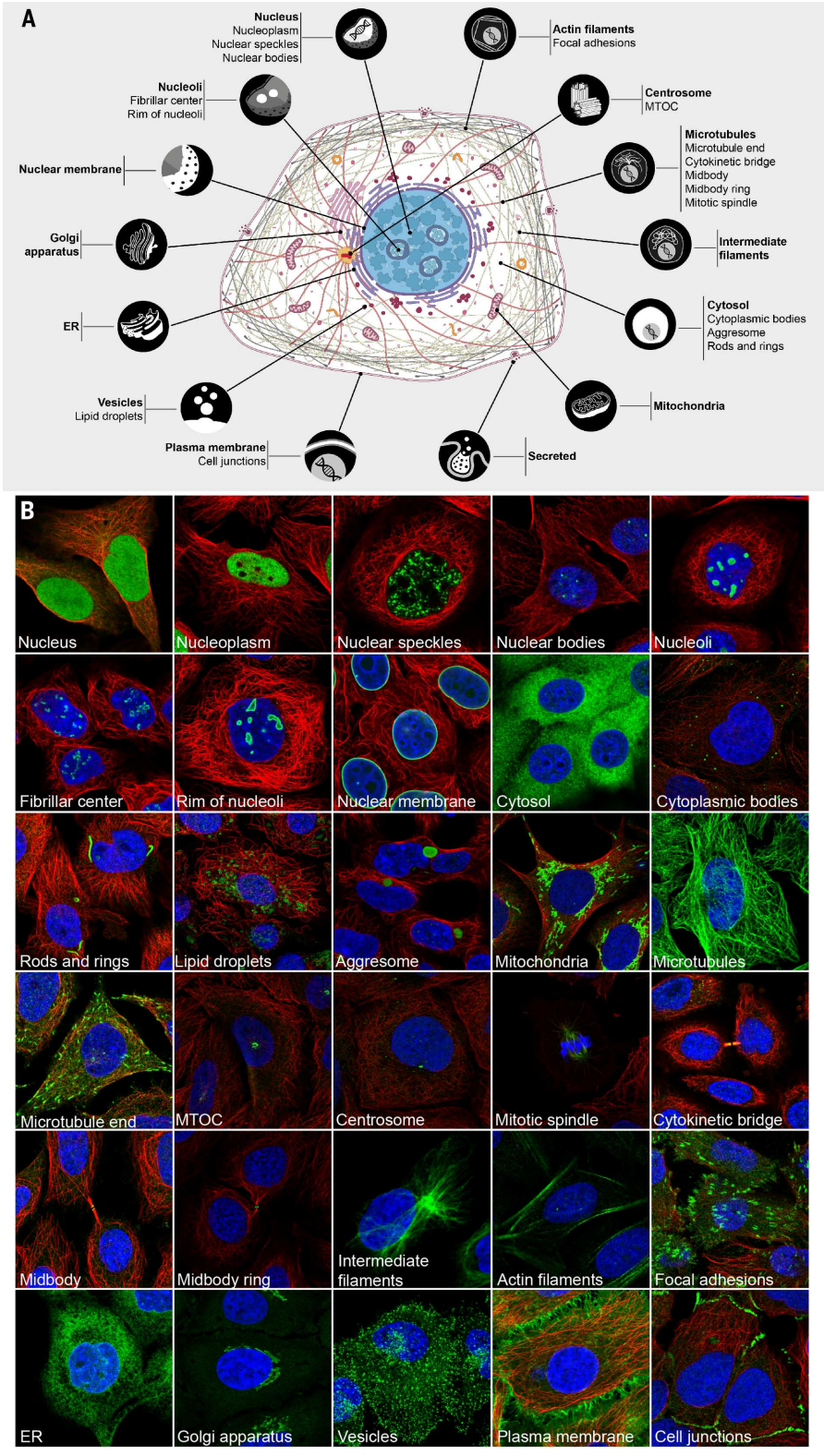


Fig. 1. Subcellular locations in the Cell Atlas. (**A**) Schematic overview of the cell. Thirteen subcellular proteomes, as well as a proteome of secreted proteins, were defined in the Cell Atlas by determining the localization of proteins to 30 subcellular structures. (**B**) Subcellular structures annotated in the Cell Atlas by immunofluorescence (IF) microscopy. Examples of proteins (green) localizing to each annotated structure in the representative set of human cell lines used in the Cell Atlas. Microtubules are marked with an antibody against tubulin (red); the nucleus is counterstained with DAPI (blue). The side of an image is 64 μm. Information about cell lines, antibodies, and proteins is given in table S6.

short interfering RNA (siRNA) silencing (37) or CRISPR-Cas9 knockout, (ii) expression of a fluorescent protein–tagged protein at endogenous levels (38), or (iii) independent antibodies targeting different epitopes (see fig. S4 for examples). The second tier, "supported" locations, is defined by agreement with external experimental data from the UniProt database. An "approved" location score indicates a lack of external experimental information about the protein location. Last, an "uncertain" location is contradictory to complementary information, such as literature or transcriptomics data. "Uncertain" locations are only shown when it cannot be ruled out that the data are correct. In fig. S5, the distributions of scores for all proteins are shown. Forty-three percent of the protein locations are in the top two tiers, representing a high degree of certainty in the results, and half of the proteins are in the "approved" category. Although these proteins have no external evidence to support their location, the antibodies passed our quality tests and showed a consistent IF staining. Nevertheless, the likelihood of false-positive results may be higher and should be taken into consideration when looking at individual proteins, whereas the effect on global proteomic analyses is negligible (fig. S6).

The human organelle proteomes

The spatial information provided by the IF images enabled the development of a subcellular map. The distribution of 12,003 proteins into 30 cellular compartments and substructures is shown in Fig. 2B and detailed in table S8. We were able to describe the proteomes for 13 major organelles. In addition, we defined a secretome containing proteins secreted through the classical pathway by combining three bioinformatic methods for signal peptide recognition with seven prediction methods for transmembrane regions (24), which indicated that 2918 proteins are secreted (table S9). Most proteins in the Cell Atlas were found in the nucleoplasm and its substructures (6245). The number of nuclear proteins considerably exceeds previously reported numbers. Although false nuclear localizations can be observed because of cross-reactivity of antibodies (21), the fraction of nuclear locations are similar in the higher- and lower-reliability tiers. The second largest number of proteins was identified in the cytosol (4279), followed by vesicles (1806), including transport vesicles and small membrane-bound organelles such as endosomes or peroxisomes. The nucleoli, including their fibrillar center, contained 1270 different proteins, which is a more diverse proteome than that of the mitochondria or Golgi apparatus, although nucleoli are more restricted in their known function. In total, we acquired subcellular experimental evidence for 5662 proteins (47%) lacking an experimentally determined GO term for a cellular compartment. Furthermore, we refined or confirmed subcellular locations for 6341 (53%) proteins already classified by experimentally determined GO terms (fig. S7).

We further investigated the enrichment of RNA classification categories for the defined organelle proteomes. Figure 2C shows that proteins located in the mitochondria, nucleus, nucleoli, and ER are more often expressed in all cell lines, which emphasizes their housekeeping role and important function for cellular survival. In contrast, proteins with RNA expression patterns categorized as "enriched" (expression in a cell line at least five times as high as in all other cell lines) and "enhanced" (expression in one or more lines five times as high as the mean expression across all cell lines) are more commonly secreted or located in the plasma membrane, vesicles, and cytoskeleton, which indicates that these compartments play important roles in intercellular communication and adaptation to the surrounding microenvironment. An analogous pattern was seen in the RNA class distribution across 59 human tissues (fig. S8), indicating general similarities in organelle organization between cell lines and tissues.

The goal of proteomic studies lies in the large-scale localization of previously uncharacterized proteins to achieve a complete picture of organelle function. IF images are particularly advantageous in the identification of protein constituents of compartments that are challenging to purify or have distinct substructures. For example, specialized domains within a compartment, such as cell junctions in the plasma membrane, are easily visible in IF—for example, in the case of the uncharacterized protein C4orf19 (Fig. 2D). Other compartments, such as the cytokinetic bridge, correspond to a rare cellular event and are thus challenging for proteomic studies. However, with our high-resolution images, we were able not only to identify 88 proteins located in the cytokinetic bridge (Fig. 2E), but also to analyze the underlying components midbody (36 proteins; Fig. 2F) and midbody ring (12 proteins; Fig. 2G). The detection of well-known constituents such as CHMP1B in the midbody, as well as less well-characterized proteins such as APC2 in the midbody ring or CCSAP in the cytokinetic bridge, provides an enhanced understanding of the final step of cell division. In nucleoli, we identified proteins such as MKI67 that are localized in the rim around the nucleolus and reorganize to line the condensed chromosomes during mitosis (Fig. 2H). As described below, additional tailored assays to complement the Cell Atlas further increase the available information about subcellular locations. The largely uncharacterized dynamic structure termed rods and rings (RR) previously had only three known members, including IMPDH1 and IMPDH2 (Fig. 2I) (39). We discovered and confirmed 21 RR candidates by actively inducing RR formation with the compound ribavirin (39). The assignment of additional proteins to the RR sheds new light on this structure and provides opportunities for better understanding its origin, composition, and function. In the nucleus, the PML body (marked by SP100; Fig. 2J) was a prominent substructure. This location can be further explored for selected proteins, because the Cell Atlas contains additional images generated by superresolution microscopy, allowing a distinction between proteins localizing to the surface (SP100; Fig. 2K, lower image) versus to the core (ZBTB8A; Fig. 2K, upper image) of the PML body.

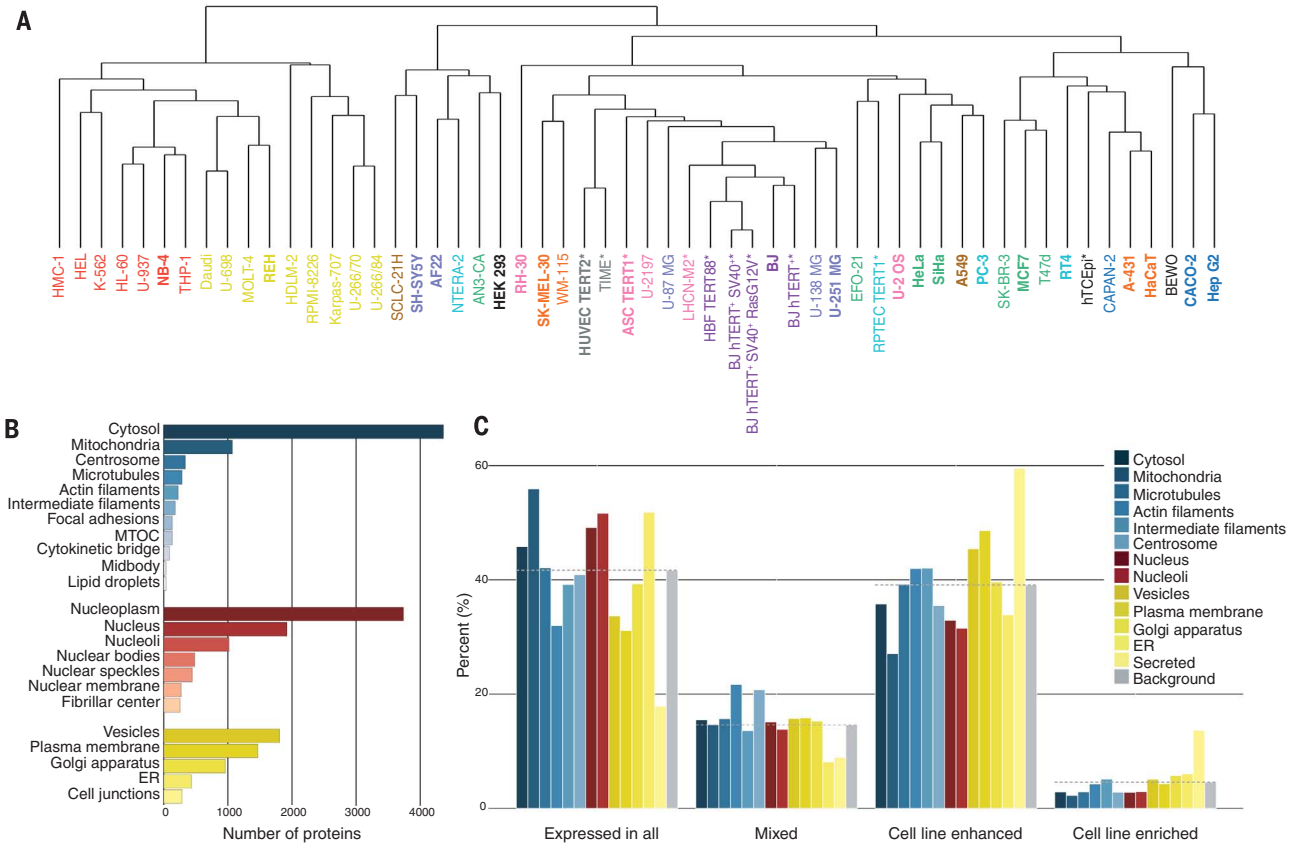
Validation with other proteome-wide data sets

To evaluate the overall validity of our data, we assessed its agreement with functional protein information from independent proteome-wide databases. First, we performed a GO "biological process" term analysis of the proteome of each organelle. The significantly enriched terms were all related to known key processes of the respective organelle (table S10). Second, we analyzed the location enrichment of a set of proteins by a hypergeometric statistical test. In this manner, we could demonstrate that the nuclear receptors according to nucleaRDB (40) and their co-regulators as defined by the Nuclear Receptor Signaling Atlas (41) were enriched in the nucleus (Fig. 3A and fig. S9) and that the group of predicted secreted proteins were enriched in the organelles of the secretory pathway (Golgi apparatus, vesicles, and ER) (Fig. 3A). Third, enrichment tests with the mammalian complex database CORUM (42) showed similar results (Fig. 3A and fig. S9). Known complexes were significantly enriched in the respective organelle, with the exception of the cytoskeleton.

Validation by mass spectrometry

Proteome databases contain information about the subcellular localizations of already characterized proteins; however, our data set contains a large portion of proteins with a previously uncharacterized location. Therefore, we used an independent approach to reliably validate our annotations. The Cell Atlas data were compared with a high-resolution spatial protein map generated by a mass spectrometry–based method called hyperLOPIT (hyperplexed localization of organelle proteins by isotope tagging). HyperLOPIT aims to resolve all subcellular compartments in a single experiment by combining biochemical cell fractionation with quantitative mass spectrometry and robust multivariate statistical analysis (3, 6). This enables global identification and quantification of proteins and assignment to their respective subcellular compartments (43). The technique does not rely on absolute organelle purification but is based on the measurement of the distribution of cellular proteins across multiple density gradient fractions. Protein localization is assigned by comparing the distributions of proteins of unknown subcellular location with those of unambiguous organelle markers.

The hyperLOPIT approach was applied to create a subcellular map of the U-2 OS cell line. Spatial distribution profiles of 5020 proteins were determined, and a support vector machine was used to classify 1971 proteins to 12 discrete subcellular compartments, which were customized to match with the annotations in the Cell Atlas (Fig. 3B). Localization information for a total of 3626 proteins was available in both the Cell Atlas (U-2 OS only; table S11) and hyperLOPIT results (table S12). Of these, 1426 proteins were unambiguously



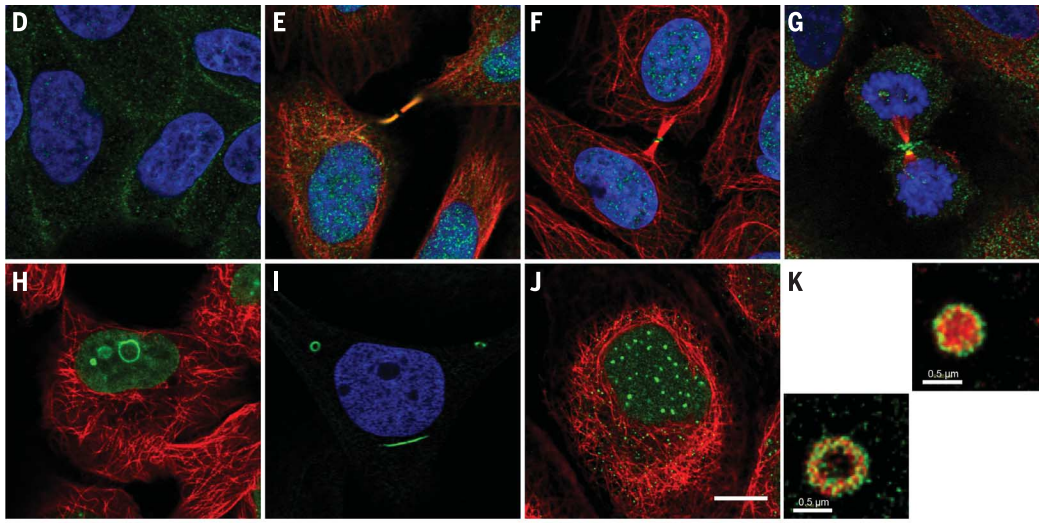


Fig. 2. Transcriptomics and proteomics. (**A**) mRNA deep sequencing was performed for 56 cell lines. The cell lines were clustered on the basis of gene expression patterns. The color of the cell line name represents its origin: red, myeloid; yellow, lymphoid; brown, lung; periwinkle, brain; turquoise, renal, urinary, and male reproductive system; green, breast and female reproductive system; pink, sarcoma; purple, fibroblast; blue, abdominal; orange, skin; black, miscellaneous. Cells immortalized by the introduction of telomerase are indicated by an asterisk. Cell lines in bold are included in the Cell Atlas cell line panel. (**B**) Number of proteins per subcellular location. A total of 12,003 proteins were localized to one or more subcellular compartments in this study. Locations are sorted and color-coded according to the number of proteins and the meta-compartments in which they occur [cytoplasm (cytosol and embedded organelles; shades of blue), nucleus (nuclear and nucleolar structures; shades of red), and secretory pathway (ER, Golgi apparatus, vesicles, and plasma membrane; shades of yellow)]. Some locations are merged: aggresomes and RR to cytosol, microtubule ends and mitotic spindle to microtubules, and midbody ring to midbody. (**C**) RNA classification categories per major organelle (nucleus and nuclear membrane are merged) compared with the background of genes in the Cell Atlas. Genes with a TPM value of ≥1 were considered as expressed and classified either as expressed in all cell lines, enriched (expression in one cell line at least fivefold as high as in all other cell lines), enhanced (average TPM level fivefold as high in one or more cell lines as the mean TPM of all cell lines), or mixed (expressed, but not in one of the other categories). (**D**) C4orf19 (detected by antibody HPA043458 in RT4 cells) localized to cell junctions, a subdomain of the plasma membrane. (**E** to **G**) Protein localization at the final stage of cytokinesis: (E) CCSAP to the cytokinetic bridge (detected by HPA028402 in U-2 OS cells) (E), CHMP1B to the midbody (detected by HPA061997 in SiHa cells) (F), and APC2 to the midbody ring (detected by HPA078002 in U-2 OS cells) (G). (**H**) MKI67 (detected by CAB000058 in U-251 MG cells) localized to the rim of nucleoli. (**I**) Previously uncharacterized protein C21orf59 (detected by CAB034170 in U-2 OS cells) localized to RR, whose formation was induced by ribavirin. (**J** and **K**) Conventional IF images in the Cell Atlas (J) and superresolution images acquired by stimulated emission depletion microscopy (K) of a PML body (a type of nuclear body) show the surface of the body marked by PML (red) and the shell protein SP100 (HPA016707, green) or the core protein ZBTB8A (HPA031768, green). Scale bars, 10 µm in (J) [applicable to (D) to (I)] and 0.05 µm in (K).

classified to a single location by hyperLOPIT. Within this group, 799 were also assigned a single location in the Cell Atlas, whereas the remaining 627 proteins had Cell Atlas annotations for more than one location.

Two comparisons between the data sets were performed: First, a comparison of proteins shown to be present in only one location in the Cell Atlas data ("unique match," table S13), and second, a comparison of all available proteins—including those shown to reside in more than one subcellular class in the Cell Atlas—with one unambiguous assignment in the hyperLOPIT data set ("partial match," table S13). Of the 799 proteins assigned by the Cell Atlas to a single location we found 76% agreement (unique match) with hyperLOPIT subcellular assignments. For the 1426 proteins common between the two data sets, 82% agreement (partial match) was observed between subcellular assignments. However, the overall agreement differed between the four reliability tiers of the Cell Atlas and was only 46% for the "approved" tier, which makes up 51% of the Cell Atlas data set (table S13). At the organelle level (table S13), the agreement ranged from 91 and 92% for the ER and mitochondria, respectively, to 60% for vesicles. This lower overlap is expected, because vesicles, as defined in the Cell Atlas, group together several organelles and structures that could be analyzed separately using hyperLOPIT. It is clear from the principal components analysis (PCA) shown in Fig. 3C that many Cell Atlas "vesicular" proteins reside in the unclassified intermediate area of the hyperLOPIT data set. Vesicles are highly dynamic structures that are generated in, and traffic between, different parts of the cell, and hence the steady-state location of their protein constituents is likely to involve multiple locations, which in the hyperLOPIT data would result in no single, unique classification. The hyperLOPIT workflow involves fractionation of chromatin-associated proteins and nucleoplasm and nucleolus, and this additional fractionation manifests itself as discrete protein correlation patterns. Interrogation of the hyperLOPIT data with Cell Atlas nuclear assignments revealed a nucleolar-like subcluster in the hyperLOPIT data; this demonstrates the power of combining data obtained using orthogonal methods (Fig. 3D).

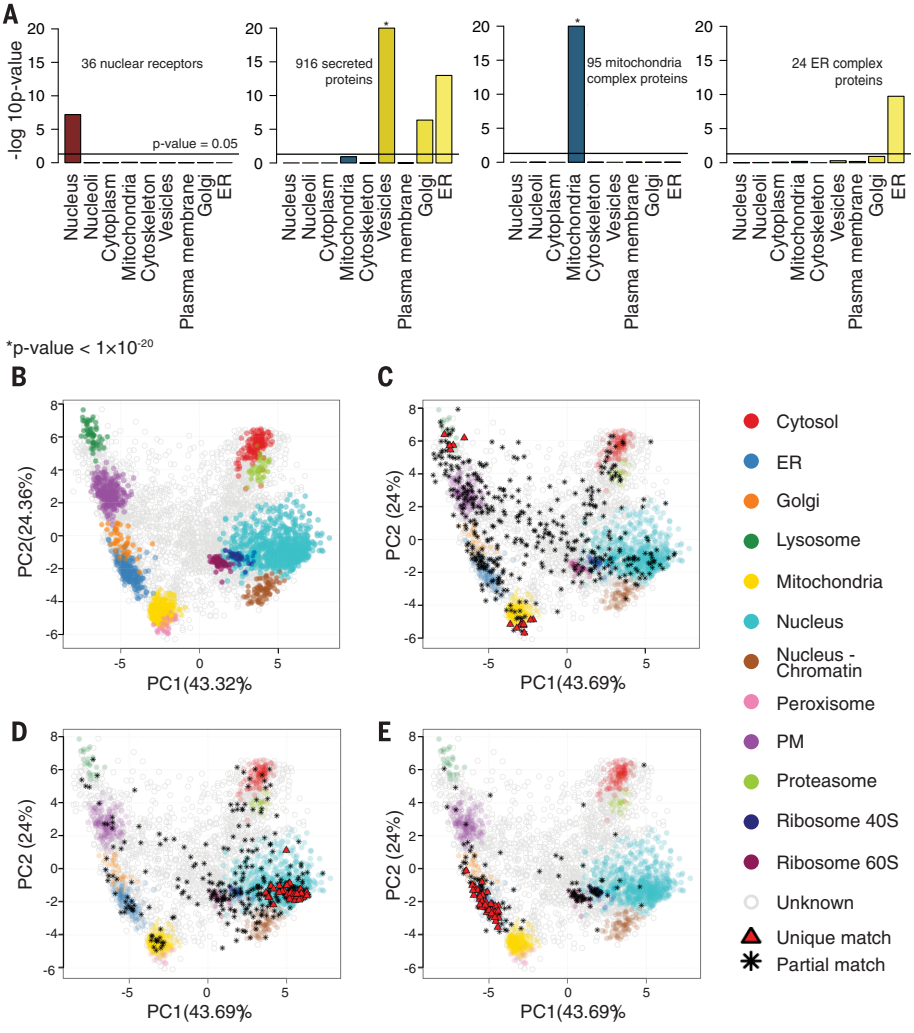


Fig. 3. Validation by proteome-wide databases and hyperLOPIT. (**A**) Location enrichment analyses of different protein sets. Hypergeometric tests were performed to evaluate subcellular locations ($P < 0.05$). Nuclear receptors were enriched in the nucleus meta-compartment. Predicted secreted proteins were enriched in organelles of the secretory pathway: ER, Golgi apparatus, and vesicles. Members of known complexes according to the CORUM database were enriched in the respective organelles—for instance, mitochondria and ER. Color-coding is as in Fig. 2. (**B**) A PCA representation of the human U-2 OS cell hyperLOPIT data (5020 proteins common across two hyperLOPIT replicates). One point represents one protein, and proteins cluster according to their density gradient distribution. Colored circles correspond to subcellular compartments that have been classified by a support vector machine. For the statistical comparison to the Cell Atlas, hyperLOPIT subcellular annotations were matched with their equivalent Cell Atlas definition. (**C** to **E**) PCA plots of the U-2 OS human data set for (C) vesicles, (D) nucleoli, and (E) the ER. Proteins occurring in both the Cell Atlas and hyperLOPIT data sets are visualized (3626 proteins). Black stars represent partial matches (a single assignment in the hyperLOPIT data, more than one in the HPA data set), and red triangles represent unique matches (a single assignment in both the HPA and hyperLOPIT data sets). PM, plasma membrane.

In the hyperLOPIT data set, 60% of the proteins identified fell into the "unclassified" category. This unclassified category may represent several dynamic scenarios, such as proteins localized to unannotated subcellular structures or multilocalizing proteins. A separate analysis was conducted for the 1755 proteins that were labeled by hyperLOPIT as "unclassified" but that contained subcellular information in the Cell Atlas (fig. S10). Interestingly, the majority of the hyperLOPIT-unclassified proteins were found in the HPA classes "nucleoplasm," "vesicles," "nucleoplasm and cytosol," and "plasma membrane and cytosol," reflecting the highly dynamic localization of the majority of cellular proteins.

To show the complementary nature of the Cell Atlas and hyperLOPIT for predicting subcellular location, we applied a transfer learning method (*44*) to integrate the two data sources. Transfer learning allows one to meaningfully integrate heterogeneous data. By combining labeled marker proteins common to both data sets, a significant increase in classifier accuracy was obtained (fig. S11) relative to that obtained using the Cell Atlas alone ($P < 2 \times 10^{-16}$). This highlights the strength of integrating the two approaches for the optimal classification of proteins to organelles.

Proteins localized to multiple compartments

In a pilot for this study (*15*), we concluded that many of the studied proteins are not restricted to a single organelle but rather localized to one or more additional locations. This observation is supported by the hyperLOPIT data described above and by data for yeast, in which 54.3% of the proteins were assigned to multiple localizations (*14*).

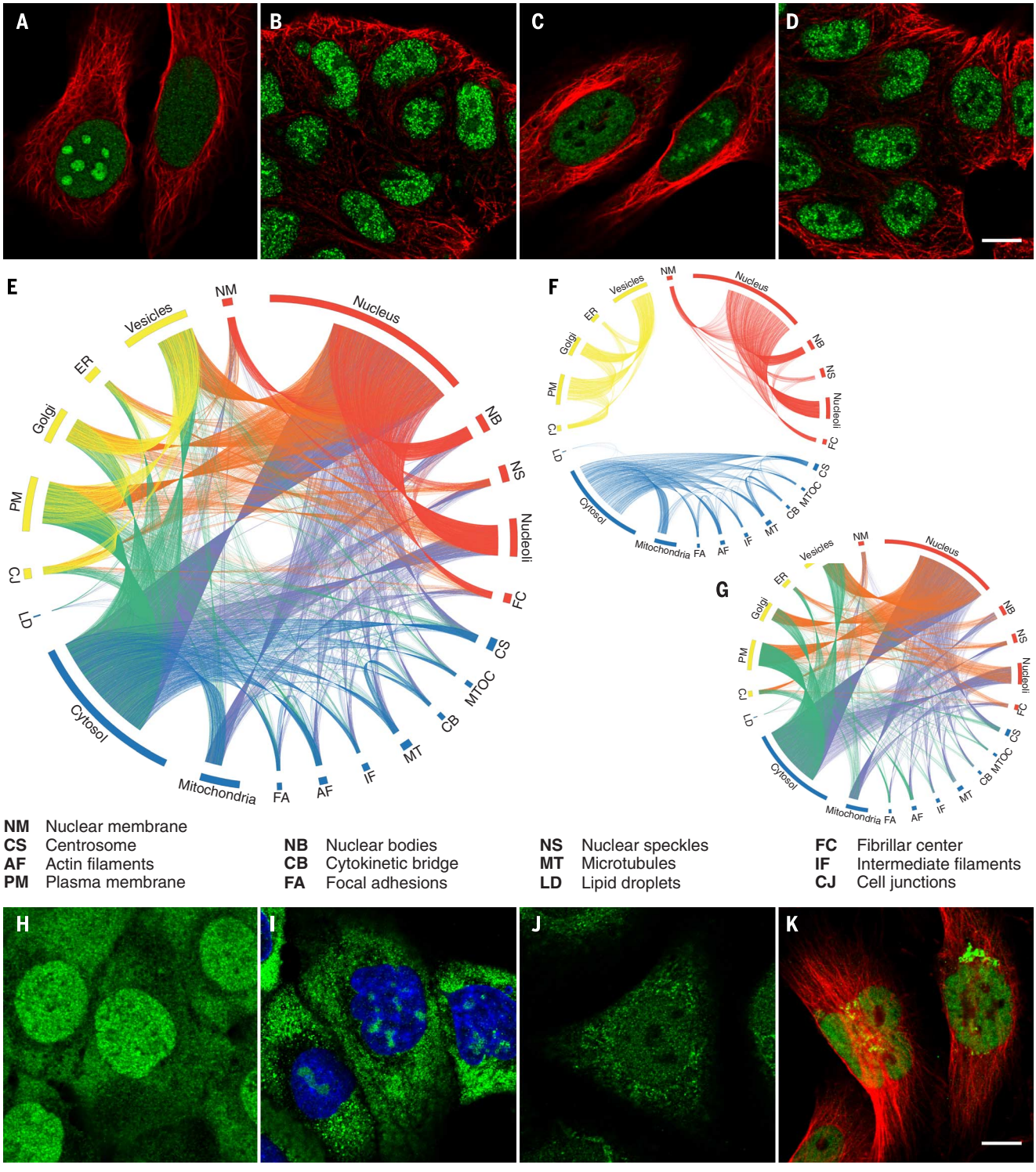


Fig. 4. Multilocalizing proteins in the human proteome. (**A** to **D**) ZNF554 is an example of a protein with a cell line–dependent subcellular localization. Two antibodies, HPA060247 [left, (A) and (B)] and HPA063358 [right, (C) and (D)], binding different epitopes detected ZNF554 in both the nucleoplasm and nucleoli in U-2 OS cells, whereas it was only detected in the nucleoplasm in RT4 and SH-SY5Y (not shown). The nucleolar expression was detected in just a fraction of the U-2 OS cells and thus additionally showed a single-cell variation. Scale bar, 10 μm. (**E** to **G**) Circular plots with the identified proteins of each compartment presented and sorted by meta-compartments. Multilocalizing proteins appearing more than once in the plots are connected by a line. Color-coding is as in Fig. 2, with secondary colors representing multilocalization across meta-compartments. The plots show (E) connections among all meta-compartments and proteins, (F) connections only within a meta-compartment, and (G) connections only across meta-compartments. (**H** to **K**) Examples of dual localizations: (H) UBE2L3 in nucleus and cytosol (detected by HPA062415 in A-431 cells), (I) 60S ribosomal protein L19 in nucleoli and cytosol (detected by HPA043014 in U-2 OS cells), (J) MTIF in nucleus and mitochondria (detected by HPA039791 in U-2 OS cells), and (K) CCAR1 in Golgi apparatus and nucleoplasm (detected by HPA007856 in U-251 MG cells). Scale bar, 10 μm.

One of the strengths of imaging-based spatial protein analysis is the ability to localize a protein in situ and simultaneously visualize protein distribution among multiple cellular structures, thus identifying multilocalizing proteins (MLPs). Here we have classified the main and additional locations for each protein on the basis of a clear difference either in the signal strength or in the occurrence across the tested cell lines. More than 50% (6163) of the proteins were detected at more than one location, of which 27% (1649) were detected at three or more locations (table S8). ER and mitochondria mainly contained specifically located proteins, whereas the proteomes of the plasma membrane and the nuclear substructures contained mainly MLPs, consistent with the hyperLOPIT data (Fig. 3E and fig. S12). This finding is consistent with the known biological function of the organelles. Whereas the proteome of the mitochondria is more self-contained, the nucleus, plasma membrane, and cytosol contain many proteins that operate across organelles to regulate metabolic reactions or gene expression or to transmit information from the surrounding environment. Also observed were MLPs that varied in their cell-to-cell spatial distribution, as well as MLPs such as ZNF554 that showed a cell line–dependent location, with different localization in the three cell lines tested (Fig. 4, A to D). In total, 3546 MLPs showed a cell line–dependent localization (table S14).

To investigate whether MLPs are organized in superordinate structures, we grouped the individual organelles and substructures into three meta-compartments—nucleus (nuclear and nucleolar structures), cytoplasm (cytosol, mitochondria, and the different types of cytoskeleton), and the secretory pathway (ER, Golgi apparatus, vesicles, and plasma membrane)—and searched for distinct patterns within and across these meta-compartments by aligning the proteins on a circular plot (Fig. 4, E to G). Within the cytoplasm meta-compartment, most MLPs appeared between the cytosol and the cytoskeletal structures and other organelles embedded in it (Fig. 4F). Similarly, most MLPs in the nucleus could be identified as a combination of nucleoplasm and the fine structures within, such as nucleoli or nuclear bodies, and likely reflect dynamic translocations of proteins between these proximal compartments (Fig. 4F). The MLPs in the secretory pathway exhibit a sequential pattern, likely reflecting the directional protein trafficking (Fig. 4F). This analysis was repeated with stratification according to the reliability of locations to control for the effect of data quality on our results (fig. S6). The patterns of multilocalization were highly similar regardless of the data set used.

Frequent patterns of multilocalization across meta-compartments included cytosol and nucleus, cytosol and nucleoli, and mitochondria and nucleoli (Fig. 4G). Enrichment analysis of GO "biological process" terms of these proteins (table S15) revealed that MLPs of the nucleus and the cytosol are related to transcription and cell cycle regulation, such as UBE2L3 (Fig. 4H); MLPs of the cytosol and nucleoli are enriched for ribosomal proteins, such as 60*S* ribosomal protein L19, which can be also found on the ER (Fig. 4I); and proteins found in both the mitochondria and nucleus are related to protein translation and cellular respiration, such as MTIF3 (Fig. 4J) and NDUFA9, respectively. Intriguingly, the meta-compartments secretory pathway and nucleus shared a very high number of MLPs, despite not being in direct physical contact with each other. These MLPs are characterized by their involvement in the regulation of transcription or cell cycle–dependent processes—for example, CCAR1 (Fig. 4K). This indicated that the proteomes of

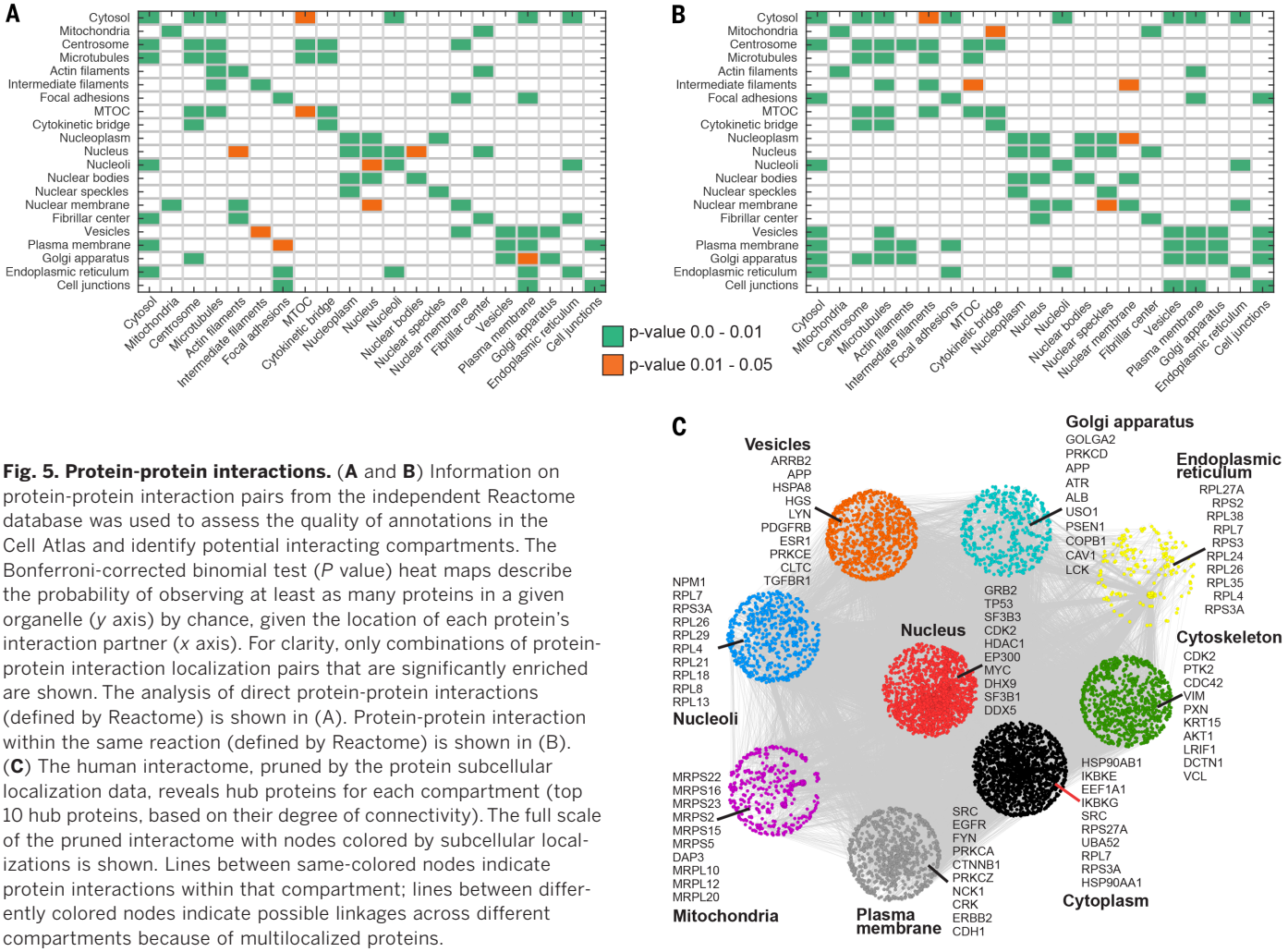


Fig. 5. Protein-protein interactions. (**A** and **B**) Information on protein-protein interaction pairs from the independent Reactome database was used to assess the quality of annotations in the Cell Atlas and identify potential interacting compartments. The Bonferroni-corrected binomial test (*P* value) heat maps describe the probability of observing at least as many proteins in a given organelle (*y* axis) by chance, given the location of each protein's interaction partner (*x* axis). For clarity, only combinations of protein-protein interaction localization pairs that are significantly enriched are shown. The analysis of direct protein-protein interactions (defined by Reactome) is shown in (A). Protein-protein interaction within the same reaction (defined by Reactome) is shown in (B). (**C**) The human interactome, pruned by the protein subcellular localization data, reveals hub proteins for each compartment (top 10 hub proteins, based on their degree of connectivity). The full scale of the pruned interactome with nodes colored by subcellular localizations is shown. Lines between same-colored nodes indicate protein interactions within that compartment; lines between differently colored nodes indicate possible linkages across different compartments because of multilocalized proteins.

the ER, Golgi apparatus, and vesicles are more functionally versatile and should not be reduced to their role in protein secretion. In fact, the MLPs create a range of interactions between functionally distant organelles and include them in a network of regulatory processes, which are primarily associated with the nucleus. This may be an indication of the complex network of events surrounding how the cell conveys signals from the exterior to the nucleus.

Spatial information refines biological networks

The biological function of an organelle is not only defined by the presence or absence of proteins, but also by its underlying chain of reactions, which in turn are often conducted by protein-protein interactions. We used the spatial information of the Cell Atlas to examine the relationship between protein interaction partners. For every annotated structure in the Cell Atlas, we investigated the subcellular locations for the direct protein interaction partners, according to the Reactome database (*45*). Figure 5A shows a heat map of the probability that proteins in one cellular compartment interact directly with proteins in the same or other compartments. Within this stringent constraint, the majority of the significant enrichments ($P < 0.05$) for an interaction pair were found within the same organelle. This compartmental enrichment was even observed for small structures such as nuclear bodies and nucleoli fibrillar centers. The exception was the microtubule-organizing center (MTOC), which showed significant enrichment for interactors found in the centrosome and microtubules. For some structures, proximal structures were also found to be significantly enriched. Proteins in the plasma membrane, for example, showed increased probability of directly interacting with proteins in the plasma membrane, cell junctions, Golgi apparatus, vesicles, focal adhesions, and cytosol. These results support the quality of the locations annotated in the Cell Atlas, given that direct protein-protein interactions occur in the same or connected compartments. To explore how cellular signaling expands across cellular compartments through reaction pathways, the same analysis was performed for the organelle proteomes, looking at protein interactions within reaction pathways defined by Reactome (Fig. 5B). In this analysis, the meta-compartments became more prominent, especially in terms of interactions between the organelles of the secretory pathway and signaling between compartments. Unexpected cross-talk between compartments included apparent interactions between the cytokinetic bridge and nuclear bodies.

We examined whether existing protein-protein interaction networks would benefit from a more comprehensive annotation of a protein's subcellular location, given that it constrains the possible number of interaction partners. The localization data was integrated, as spatial boundaries, into the human interactome that was recently used to systemically uncover the molecular background of human diseases (*46*). The interactome included annotations for 79,020 interactions of 7827 proteins. By taking the subcellular main location into account, the number decreased to 51,885 (65.7%) interactions of 6985 proteins that were found to be likely to occur in vivo (fig. S13). However, a substantial number of protein interactions were found when additional locations were included, raising the total to 62,352 (78.9%) interactions of 7494 proteins (fig. S13). This further supports the important functional role of MLPs. With this new location-pruned interaction data set, we generated a map to identify the most connective proteins, also called hub proteins, of each compartment (Fig. 5C). The hub proteins of each compartment were mostly different from hubs of the original, nonannotated interactome (table S16); hence, our data set led to the identification of previously unrecognized driver genes within the network. The localization-annotated interactome is available in table S17.

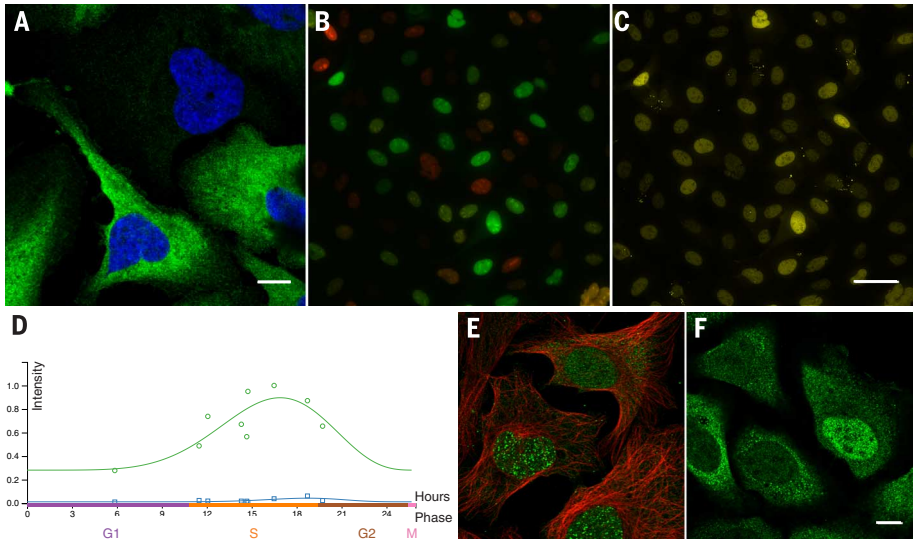


Fig. 6. Single-cell variation in protein expression. (**A**) CRYAB (detected by CAB002053 in U-2 OS cells) showed a single-cell variation in the cytosolic signal strength. (**B** and **C**) U-2 OS FUCCI cells expressed the cell cycle regulators CDT1 (red) during the G_1 phase and geminin (green) during the S and G_2 phases. An antibody targeting ANLN (yellow) stained only cells in the S and G_2 phases, marked by the green fluorescence. (**D**) Pattern of expression of ANLN across the cell cycle in U-2 OS cells by pseudo-temporal analysis using a time-regressive computational model. (**E**) The protein abundance of PCNA (detected by HPA030522 in U-2 OS cells) at nuclear bodies varied during the cell cycle. (**F**) PSMC6 (detected by HPA042823 in U-2 OS cells) changed its spatial distribution from nucleoplasm to cytosol during the cell cycle, based on data from U-2 OS FUCCI cells. Scale bars, 10 μm in (A) and (F) [applies to (E)] and 50 μm in (C) [applies to (B)].

Single-cell variations in protein expression

Protein profiling by IF microscopy allows analysis of expression patterns on a single-cell level to reveal variations in a protein's expression across the analyzed cells. In the Cell Atlas, we labeled proteins with an observed single-cell variation (SCV), such as the nucleolar localization of ZNF554 (Fig. 4, A and C). SCV can be observed either in protein expression levels (IF signal intensity) or in the spatial distribution pattern. Of the 12,003 detected proteins, 1855 (15%) showed a SCV (table S18). Further studies are needed to reveal whether SCV is due to dynamic protein regulation or stochastic events. The majority of these proteins showed a variation in protein expression levels (1671)—for example, CRYAB (Fig. 6A)—whereas 222 proteins showed a variation in spatial distribution (38 proteins fall into both categories). The organelles with the most SCV proteins were the cytosol (394), nucleoplasm (381), nucleoli (230), and mitochondria (206) (table S8)—organelles that also contain most known cell cycle–dependent proteins.

In addition to being related to the subcellular structures that only appear during cell division (mitotic spindle, cytokinetic bridge, midbody, and midbody ring), it is plausible to expect a majority of these SCVs to also be related to the cell cycle, because the cells in the images were growing under asynchronous conditions. To confirm this, we used two approaches for a subset of the proteins. First, we stained selected proteins with an observed SCV in the U-2 OS FUCCI [fluorescence ubiquitination cell cycle indicator (*47*)] cell line, which allows monitoring of the cell cycle; by this method, we verified a cell cycle–dependent expression of 64 proteins, including, for example, ANLN (Fig. 6, B and C; see the list of proteins in table S19). The second approach used a computational model to infer the cell cycle position on the basis of features of the microtubule and nucleus reference markers. In this manner, the cell cycle position of the cells in the images could be determined in a continuous model, and a pseudo-temporal reconstruction allowed the pattern of cell cycle dependency to be modeled. Figure 6D

shows such a plot for ANLN, which is expressed in cells in the S and G_2 phases, according to both FUCCI colocalization and the pseudo-temporal computational modeling. Like for SCV, cell cycle–dependent variation could be observed either in a change of the intensity—for example, in the case of PCNA (Fig. 6E)—or in a change of the localization, illustrated by the translocation of PSMC6 from nucleoplasm to cytosol (Fig. 6F).

Discussion

Here we present the most comprehensive map of the subcellular distribution of the human proteome, generated by high-resolution IF images on a single-cell level. The results are presented in an interactive resource, the Cell Atlas, as part of the Human Protein Atlas (www.proteinatlas.org). This allows exploration of the organelle proteomes, their substructures, single-location and multilocalizing proteins, and proteins exhibiting single-cell variations in expression or cell cycle–dependent expression. These defined categories can furthermore be explored in terms of gene expression patterns across a multitude of cell lines on the basis of transcriptome data. To facilitate integration with other biological resources, all data are available for download from the Human Protein Atlas and through collaborations with efforts such as UniProt (*19*), NextProt (*29*), GO (*48*), and the pan-European ELIXIR project (*49*).

Spatial partitioning of biological reactions by compartmentalization is an important cellular mechanism for allowing multiple cellular reactions to occur in parallel while avoiding cross-talk. Intriguingly, we identified more than 50% of the analyzed proteins as localizing to more than one compartment at the same time. The fact that proteins are localized at multiple sites increases the complexity of the cell from a systems perspective. On one level, it can function as a spatial confinement to control the timing of the molecular function in the designated compartment. On another level, multilocalizing proteins are more prone to have diverse protein-protein interactions because of an increased number of potential interaction partners. This is of particular relevance for network analyses and the identification of key hub proteins that play a crucial role in linking complexes to smaller subnetworks, leading to a cellular-wide network. Moreover, proteins that localize to more than one compartment may have context-specific functions, increasing the functionality of the proteome. The fact that proteins "moonlight" in different parts of the cell is now well accepted (*50*, *51*). The high percentage of proteins in multiple locations, as indicated by the complementary IF and hyperLOPIT data sets, may be an indicator of the scale on which moonlighting occurs. The more complex a system is, the greater the number of parts that must be sustained in their proper place, and the lesser the tolerance for errors; therefore, a high degree of regulation and control is required. To understand cellular function, and particularly in the context of health and disease, detailed knowledge about the cellular system is needed. We demonstrated that current network models benefit from integration of the Cell Atlas localization data as spatial boundaries to remove false-positive interactions.

The proteome of a single cell is compartmentalized and spatiotemporally regulated to a high degree. Protein expression and localization change over time and enable the cell to react to intrinsic or extrinsic factors. Although only presenting a snapshot of the current state of a few cells, our single-cell analysis gives insight into this dynamic process. The high-resolution map of the subcellular localization of 12,003 human proteins provided by the Cell Atlas is a key resource for a comprehensive understanding of the human cell and its complex underlying molecular machinery, as well as a major step toward modeling the human cell.

Material and methods

Tissue culture cell line cultivation

All cell lines were cultivated at 37°C in a 5% CO_2 humidified environment in the following growth media: Roswell Park Memorial Institute medium (A-431, REH, RH-30, SiHa, SK-MEL-30; Sigma-Aldrich); Dulbecco's Modified Eagle Medium (A549, BJ, HaCaT, HeLa, NTERA, SH-S5Y5; Sigma-Aldrich); Eagle's Minimal Essential Medium (CACO-2, HEK293, HepG2, MCF-7, U-251 MG; Sigma-Aldrich); McCoy's 5A modified (RT-4, U-2 OS; Sigma-Aldrich). Media were always supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich); additional cell line-specific supplements were: 1% non-essential amino acids (CACO-2, HeLa, HRK293, HepG2, MCF-7), 1% L-glutamine (CACO-2, HaCaT, HepG2, MCF-7, NTERA, RT-4, U-2 OS), 5% horse serum (NTERA). No antibiotics were used.

AF22 cells were kindly provided by A. Falk. They were cultivated in DMEM/F12 supplemented with N-2 (Cat#17502048, Thermo Fisher) and Pen/Strep (Sigma-Aldrich), with freshly added B-27 (1:1000, Cat#12587010, Thermo Fisher), EGF (10 ng/ml, AF-100-15, PeproTech) and FGF2 (10 ng/ml, 100-18B, PeproTech), flask and plates were coated in two steps with poly-N-ornithine (Sigma-Aldrich) and laminin (Sigma-Aldrich). Telomerase-immortalized cell line HUVEC/TERT2 (Cat# MHT-006-2) and ASC/TERT1 (Cat# MHS-001) were a kind gift by Evercyte GmbH, Vienna, Austria, and were cultured in EndoUp2 and AdipoUp, respectively. U2-OS FUCCI cells were developed and kindly provided by A. Miyawaki (*47*). The cells were cultivated in McCoy's 5A modified medium supplemented with 1% L-glutamine and 10% FBS. HeLa-Kyoto cell lines stably expressing an enhanced green fluorescent protein (EGFP)–tagged protein encoded on Bacterial Artificial Chromosome (BAC) were a kind gift from A. Hyman, Max Planck Institute Dresden, Germany, and were cultivated as described in Skogs *et al.* (*38*, *46*). CRISPR-Cas9 knockout and GFP-expressing cells were a kind gift by Horizon Discovery, Cambridge, UK. Their designed HAP1 cell lines were cultivated in IMDM (Iscove's Modified Dulbecco's Medium, Sigma-Aldrich) media supplemented with 10% FBS and 1% Pen/Strep. All cells were harvested at 60 to 70% confluency by trypsinization (Trypsin-EDTA solution from Sigma-Aldrich) for splitting or preparing in glass bottom plates.

Antibodies

All antibodies generated and validated within the HPA project were rabbit polyclonal antibodies. They were designed to bind specifically to as many isoforms of the target protein as possible. The antigens consisted of recombinant protein epitope signature tags (PrEST) with a typical length between 50 and 100 amino acids (*52*). The resulting antibodies were affinity purified using the antigen as affinity ligand (*32*). All antibodies used were first approved for sensitivity and lack of cross-reactivity to other proteins, on arrays consisting of glass slides with spotted PrEST fragments. Commercial antibodies were provided by the suppliers and used according to the supplier's recommendations.

Sample preparation for indirect immunofluorescence

A standardized protocol optimized for proteome-wide immunofluorescence localization studies was used, which has previously been described in detail by Stadler *et al.* (*16*). Briefly, cells were seeded in 96-well glass bottom plates (Whatman, Cat# 7716-2370, GE Healthcare, UK, and Greiner Sensoplate Plus, Cat# 655892, Greiner Bio-One, Germany) coated with fibronectin (VWR, Sigma-Aldrich) and grown to a confluency of 60 to 70% (log-phase growth). PBS-washed cells were fixed in 4% paraformaldehyde (PFA) in growth media supplemented with 10% FBS for 15 min, followed by permeabilization with 0.1% Triton X-100 in PBS for 3×5 min. After a washing step with PBS, cells were incubated with the primary antibody overnight at 4°C. Rabbit polyclonal HPA antibodies were diluted to 2 to 4 μg/ml in blocking buffer (PBS with 4% FBS) containing 1 μg/ml mouse anti-tubulin (Abcam, ab7291, RRID:AB_2241126, Cambridge, UK), and 1 μg/mL chicken anti-calreticulin (Abcam, ab14234, RRID:AB_2228460) or rat anti-KDEL antibody (MAC 256) (Abcam, ab50601, RRID:AB_880636), respectively. On the next day after 4×10 min washes with PBS, the cells were incubated for 90 min at room temperature with the following secondary antibodies (all from ThermoFisher Scientific) diluted to 1 μg/ml in blocking buffer: goat anti-rabbit AlexaFluor 488 (A11034, RRID:AB_2576217), goat anti-mouse AlexaFluor 555 (A21424, RRID:AB_2535845), and goat anti-chicken AlexaFlour 647 (A-21449, RRID:AB_2535866), or goat anti-rat AlexaFluor 647 (A21247, RRID:AB_1056356), respectively. Cells were subsequently counterstained with 4′,6-diamidino-2-phenylindole (DAPI) for 10 min. After washing with PBS, the wells were completely filled with 78% glycerol in PBS and sealed.

Fluorescence image acquisition

Fluorescent images were acquired with a Leica SP5 confocal microscope (DM6000CS) equipped with a 63× HCX PL APO 1.40 oil CS objective (Leica Microsystems, Mannheim, Germany). The settings for each image were as follows: Pinhole

1 Airy unit, 16-bit acquisition, and a pixel size of 0.08 μm. The detector gain measuring the signal of each antibody was adjusted to a maximum of 800 V to avoid strong background noise. The majority of the images were acquired manually from at least two representative field-of-views (FOVs). For proteins displaying single cell variations in their expression pattern, at least six different FOVs were acquired. A small part of the plates were imaged automatically using the MatrixScreener M3 in LAS AF software (Leica Microsystem, Mannheim, Germany). Here, *z*-stacks at six FOVs were acquired and afterward two images were manually selected for display in the Cell Atlas. All images on the Cell Atlas are unprocessed with a small compression due to conversion from TIFF to JPEG file format.

IF image annotation

The subcellular location of each protein was manually determined based on the signal pattern and relation to the markers for nucleus (DAPI), microtubules, and endoplasmic reticulum. The annotated locations were as follows: actin filaments, aggresome, cell junctions, centrosome, cytokinetic bridge, cytoplasmic bodies, cytosol, endoplasmic reticulum, focal adhesions, Golgi apparatus, intermediate filaments, lipid droplets, microtubule organizing center (MTOC), microtubules, microtubule ends, midbody, midbody ring, mitochondria, mitotic spindle, nuclear bodies, nuclear membrane, nuclear speckles, nucleolar fibrillar center, nucleolar rim, nucleoli, nucleoplasm, nucleus, plasma membrane, rods and rings, and vesicles. If more than one location was detected, they were defined as main or additional location depending on the relative signal strength between the location and the most common location when including all cell lines. Variation between single cells were annotated either as a variation in the intensity or spatial distribution based on a visual inspection. The staining was not annotated if considered negative or unspecific.

Prediction of the human secretome

For the prediction of the human secretome, the analysis was performed as previously described (*24*). Briefly, a majority decision approach was used based on results from three methods for the prediction of signal peptides (SP): SignalP4.0 (*53*), Phobius (*54*), and SPOCTOPUS (*55*). SignalP4.0 is solely focused on the prediction of SPs whereas the two latter combine the prediction of transmembrane (TM) segments and SPs. In addition, results from the prediction of the human membrane proteome (*56*) were included to classify proteins with a predicted SP as well as one or more TM regions as membrane-spanning. The resulting list of potentially secreted proteins consists of all proteins with a predicted signal peptide by two out of three methods and not including a predicted TM region.

Classification of location reliability

Detected locations were classified based on the reliability of the antibodies and their respective stainings. A score was used for the classification, which incorporated several factors: reproducibility of the antibody staining in different cell lines (also taken in account when the signal strength correlates with RNA expression); reproducibility of the staining using antibodies binding to different epitopes on the target protein; validation data for the specificity of the antibody (knockdown by siRNA or CRISPR-Cas9 knockout mutants, matching signal with fluorescent-tagged protein); experimental evidence for location described in literature. There were also soft factors such as antibody validation by non–IF-related methods such as Western blot or immunohistochemistry. The final score led either to the failing of the antibody (~50% of all tested antibodies failed) or to the assignment into one of the following four classes: (i) "validated," if at least one antibody is validated—for example, two independent antibodies show the same localization, that was also observed in experiments outside the HPA or it was supported by, e.g., siRNA silencing; (ii) "supported," if there is external experimental data for the location; (iii) "approved," if the localization of the protein has not been previously described and was detected by only one antibody without additional validation; and (iv) "uncertain," if the antibody staining is contradictory to experimental data or no expression is detected on the RNA level.

RNA sequencing

Cell lines were selected for IF imaging based on RNA expression of genes (*57*). RNA was extracted from the cells using the RNeasy kit (Qiagen), generating high-quality total RNA (i.e., RIN > 8) that was used as input material for library construction with Illumina TruSeq Stranded mRNA reagents. Duplicate samples were sequenced on the Illumina HiSeq2500 platform. Raw sequences were mapped to the human reference genome GrCh38 and further quantified using the Kallisto software (*58*) to generate normalized transcript per million (TPM) values. TPM values for genes were generated by summing up TPM values for the corresponding transcripts generated by Kallisto. Genes with a TPM value ≥1 were considered expressed.

Location enrichment of protein sets by hypergeometric test

Enrichment of a group of proteins in subcellular locations was examined by hypergeometric tests. In each subcellular location enrichment test, only proteins with subcellular location annotated were considered. Predicted secreted proteins were collected from the HPA (*24*), nuclear receptors from nucleaRDB (*40*), nuclear receptor co-regulators from nuclear receptor signaling atlas (*41*), and subcellular location-specific protein complexes from CORUM (*42*). In CORUM database, nuclear complex proteins were taken from a term "nucleus" in the database; nucleoli complex proteins from "nucleolus"; cytoskeleton complex proteins from "actin cytoskeleton," "microtubule cytoskeleton," and "centrosome" complexes; mitochondria complex proteins from "mitochondrion"; vesicle complex proteins from "intracellular transport vesicle," "peroxisome," and "vacuole or lysosome"; ER complex proteins from "endoplasmic reticulum"; plasma membrane complex proteins from "plasma membrane/membrane attached" and "cell junction"; and cytoplasm complex proteins from "cytoplasm."

HyperLOPIT comparison with Cell Atlas annotations

To compare the subcellular assignments by both methods it was necessary to match the 12 subcellular organelle definitions used by hyperLOPIT to the 30 image categories defined in the Cell Atlas. The comparison was broken down into the following subclasses: all Cell Atlas subnuclear categories ("nucleus," "nucleoplasm," "nuclear speckles," "nuclear bodies," "nucleoli," "nucleoli fibrillar center," and "nuclear membrane") were individually compared with a single hyperLOPIT nuclear class encompassing both hyperLOPIT terms "nucleus" and "nuclear chromatin"; the Cell Atlas term for "vesicles" was compared with the combined hyperLOPIT terms for "lysosome" and "peroxisome" (consistent with the Cell Atlas definition for vesicles); and the Cell Atlas class "cell junctions" was compared with the hyperLOPIT term "plasma membrane." For the Cell Atlas terms called "plasma membrane," "mitochondria," "endoplasmic reticulum," "Golgi apparatus" and "cytosol/cytoplasm," the same terms are also available for hyperLOPIT and thus a direct comparison was performed. Proteins that were assigned by hyperLOPIT to the large protein complexes such as ribosomal subunits and proteasome were excluded from the comparison. A detailed description of the hyperLOPIT approach is provided in the supplementary materials.

Heat maps for protein-protein interaction

Protein-protein interaction pairs were obtained from the independent Reactome database (downloaded 20 September 2016) (*45*). A binomial test was used to compare the observed frequency of a target protein (Protein B) localizing to a given compartment with the expected frequency based on all annotations in the Cell Atlas. Here, the likelihood of localizations of the first protein in the pair (Protein A) can be ignored, as under the null hypothesis it has no impact on the localization of Protein B. The test therefore becomes the probability that we observe at least as many instances of Protein B in a specific compartment given the number of "tries" (instances of Protein A) and the background distribution of proteins over the locations in the Cell Atlas. The background distribution of locations was constructed by taking the frequency of each annotated location for proteins in in the Cell Atlas over the total number of proteins annotated in the Cell Atlas.

The results of the test were visualized using a heat map of *P* values (Fig. 5, A and B) where rows represent the location of Protein A and columns represent the location of Protein B. Values are therefore the probability of seeing Protein B in the given compartment at least as frequently as it was actually observed assuming the background distribution. The Bonferroni multiple-hypothesis

correction applied per-row to correct for the number of locations being tested for in each pairing. By then considering the correlation of the protein-protein interaction pair locations, key insights into the nature and quality of the data in the Cell Atlas can be gained.

The Reactome database contains several types of protein-protein interactions that can be used to assess different properties of the Cell Atlas annotations. To assess the quality of annotation, we first analyzed direct interactions reasoning that interacting proteins must occupy the same physical space at some point in the cell cycle and therefore should be localized either to the same compartment or adjacent compartments (Fig. 5A).

The same analysis was further performed for protein pairs listed as belonging to the same reaction pathway as defined by the Reactome database to assess what compartments potentially interact through signal cascades (Fig. 5B). This analysis was created using MATLAB2016a.

Figure generation

Plots were generated using R studio (v. 3.3.1) and the additional ggplot2 package. The cell line hierarchical clustering was based on the Spearman correlation of the RNA sequencing data for each cell line. The average distance was used to determine the hierarchical clusters and visualized then by the R package ggdendro. The circular plots showing distribution of multilocalizing proteins were created using the Circos software (v. 0.69) (*59*). The image montages were created using FIJI ImageJ (v. 2.0.0-rc-49/1.51f).

Gene Ontology terms and functional enrichment

To check the overlap with GO annotations for proteins in the Cell Atlas, the web-based tool QuickGO (*60*) was used to acquire GO annotations for all genes using filters for cellular component and information source (downloaded 15 February 2017). The GO annotations based on data from the Cell Atlas were removed, and the Ensembl IDs for all Cell Atlas genes were then used for checking the overlap of genes with experimental evidence for any GO annotation. The functional annotation clustering for the genes not expressed in the Cell Atlas cell line panel was performed using the web based tool DAVID (Database for Annotation, Visualization, and Integrated Discovery v. 6.8) (*61*). All human genes were used as a background and the GO domain "biological process" terms with Bonferroni value of less than 0.01 were regarded as significantly enriched.

Location-pruned protein-protein interactions

Proteins interactions were obtained from published protein interactome data (*46*); among those protein interactions, only interactions with "signaling," "kinase," "complex," "literature," and "binary" types were taken; this indicates direct protein interactions. Those protein interactions were pruned to proteins localized in the same subcellular locations, in either cytoplasm or plasma membrane, or in either cytoplasm or cytoskeleton. Location-pruned protein interactions were visualized (Fig. 5C) through the edge-weighted spring embedded layout of Cytoscape (*62*) and their nodes were colored by the least frequent one of subcellular locations they have. In each subcellular location, hub proteins from protein interactions of given subcellular locations were examined based on their degree connectivity.

REFERENCES AND NOTES

1. K. Laurila, M. Vihinen, Prediction of disease-related mutations affecting protein localization. *BMC Genomics* **10**, 122 (2009). doi: 10.1186/1471-2164-10-122; pmid: 19309509
2. S. Park *et al*., Protein localization as a principal feature of the etiology and comorbidity of genetic diseases. *Mol. Syst. Biol.* **7**, 494 (2011). doi: 10.1038/msb.2011.29; pmid: 21613983
3. T. P. Dunkley, R. Watson, J. L. Griffin, P. Dupree, K. S. Lilley, Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell. Proteomics* **3**, 1128–1134 (2004). doi: 10.1074/mcp.T400009-MCP200; pmid: 15295017
4. L. J. Foster *et al*., A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006). doi: 10.1016/j.cell.2006.03.022; pmid: 16615899
5. L. Jakobsen *et al*., Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J.* **30**, 1520–1535 (2011). doi: 10.1038/emboj.2011.63; pmid: 21399614
6. A. Christoforou *et al*., A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7**, 9992 (2016). doi: 10.1038/ncomms9992; pmid: 26754106
7. D. N. Itzhak, S. Tyanova, J. Cox, G. H. Borner, Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016). doi: 10.7554/eLife.16950; pmid: 27278775
8. K. J. Roux, D. I. Kim, B. Burke, BioID: A screen for protein-protein interactions. *Curr. Protoc. Protein Sci.* **74**, 19.23.1–19.23.14 (2013). pmid: 24510646
9. H.-W. Rhee *et al*., Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013). doi: 10.1126/science.1230593; pmid: 23371551
10. V. Hung *et al*., Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol. Cell* **55**, 332–341 (2014). doi: 10.1016/j.molcel.2014.06.003; pmid: 25002142
11. S.-Y. Lee *et al*., APEX fingerprinting reveals the subcellular localization of proteins of interest. *Cell Rep.* **15**, 1837–1847 (2016). doi: 10.1016/j.celrep.2016.04.064; pmid: 27184847
12. J. C. Simpson, R. Wellenreuther, A. Poustka, R. Pepperkok, S. Wiemann, Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1**, 287–292 (2000). doi: 10.1093/embo-reports/kvd058; pmid: 11256614
13. W.-K. Huh *et al*., Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003). doi: 10.1038/nature02026; pmid: 14562095
14. Y. T. Chong *et al*., Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413–1424 (2015). doi: 10.1016/j.cell.2015.04.051; pmid: 26046442
15. L. Barbe *et al*., Toward a confocal subcellular atlas of the human proteome. *Mol. Cell. Proteomics* **7**, 499–508 (2008). doi: 10.1074/mcp.M700325-MCP200; pmid: 18029348
16. C. Stadler, M. Skogs, H. Brismar, M. Uhlén, E. Lundberg, A single fixation protocol for proteome-wide immunofluorescence localization studies. *J. Proteomics* **73**, 1067–1078 (2010). doi: 10.1016/j.jprot.2009.10.012; pmid: 19896565
17. P. Horton *et al*., WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **35**, W585–W587 (2007). doi: 10.1093/nar/gkm259; pmid: 17517783
18. K. C. Chou, Z. C. Wu, X. Xiao, iLoc-Hum: Using the accumulation-label scale to predict subcellular locations of human proteins with both single and multiple sites. *Mol. Biosyst.* **8**, 629–641 (2012). doi: 10.1039/C1MB05420A; pmid: 22134333
19. UniProt Consortium, UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015). doi: 10.1093/nar/gku989; pmid: 25348405
20. L. Fagerberg *et al*., Mapping the subcellular protein distribution in three human cell lines. *J. Proteome Res.* **10**, 3766–3777 (2011). doi: 10.1021/pr200379a; pmid: 21675716
21. C. Stadler *et al*., Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat. Methods* **10**, 315–323 (2013). doi: 10.1038/nmeth.2377; pmid: 23435261
22. M. Jadot *et al*., Accounting for protein subcellular localization: A compartmental map of the rat liver proteome. *Mol. Cell. Proteomics* **16**, 194–212 (2017). pmid: 27923875
23. M. Uhlen *et al*., Towards a knowledge-based human protein atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010). doi: 10.1038/nbt1210-1248; pmid: 21139605
24. M. Uhlén *et al*., Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi: 10.1126/science.1260419; pmid: 25613900
25. K. D. Pruitt, T. Tatusova, G. R. Brown, D. R. Maglott, NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135 (2012). doi: 10.1093/nar/gkr1079; pmid: 22121212
26. A. Yates *et al*., Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016). doi: 10.1093/nar/gkv1157; pmid: 26687719
27. H. Kawaji *et al*., Update of the FANTOM web resource: From mammalian transcriptional landscape to its dynamic regulation. *Nucleic Acids Res.* **39**, D856–D860 (2011). pmid: 21075797
28. K. G. Ardlie *et al*., The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi: 10.1126/science.1262110; pmid: 25954001
29. P. Gaudet *et al*., The neXtProt knowledgebase on human proteins: Current status. *Nucleic Acids Res.* **43**, D764–D770 (2015). doi: 10.1093/nar/gku1178; pmid: 25593349
30. M. Beck *et al*., The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549–549 (2011). doi: 10.1038/msb.2011.82; pmid: 22068332
31. T. Geiger, A. Wehner, C. Schaab, J. Cox, M. Mann, Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11**, 014050 (2012). doi: 10.1074/mcp.M111.014050; pmid: 22278370
32. P. Nilsson *et al*., Towards a human proteome atlas: High-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5**, 4327–4337 (2005). doi: 10.1002/pmic.200500072; pmid: 16237735
33. M. Peplow, Citizen science lures gamers into Sweden's Human Protein Atlas. *Nat. Biotechnol.* **34**, 452–453 (2016). doi: 10.1038/nbt0516-452c
34. J. Bordeaux *et al*., Antibody validation. *Biotechniques* **48**, 197–209 (2010). doi: 10.2144/000113382; pmid: 20359301
35. M. Baker, Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274–276 (2015). doi: 10.1038/521274a; pmid: 25993940
36. M. Uhlen *et al*., A proposal for validation of antibodies. *Nat. Methods* **13**, 823–827 (2016). pmid: 27595404
37. C. Stadler *et al*., Systematic validation of antibody binding and protein subcellular localization using siRNA and confocal microscopy. *J. Proteomics* **75**, 2236–2251 (2012). doi: 10.1016/j.jprot.2012.01.030; pmid: 22361696
38. M. Skogs *et al*., Antibody validation in bioimaging applications based on endogenous expression of tagged proteins. *J. Proteome Res.* **16**, 147–155 (2017). pmid: 27723985
39. G. Covini *et al*., Cytoplasmic rods and rings autoantibodies developed during pegylated interferon and ribavirin therapy in patients with chronic hepatitis C. *Antivir. Ther.* **17**, 805–811 (2012). doi: 10.3851/IMP1993; pmid: 22293655
40. B. Vroling *et al*., NucleaRDB: Information system for nuclear receptors. *Nucleic Acids Res.* **40**, D377–D380 (2012). doi: 10.1093/nar/gkq1009; pmid: 22064856
41. S. A. Ochsner, C. M. Watkins, B. S. LaGrone, D. L. Steffen, N. J. McKenna, Research resource: Tissue-specific transcriptomics and cistromics of nuclear receptor signaling: a web research resource. *Mol. Endocrinol.* **24**, 2065–2069 (2010). doi: 10.1210/me.2010-0216; pmid: 20685849
42. A. Ruepp *et al*., CORUM: The comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* **38**, D497–D501 (2010). doi: 10.1093/nar/gkp914; pmid: 19884131
43. L. Gatto, L. M. Breckels, S. Wieczorek, T. Burger, K. S. Lilley, Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. *Bioinformatics* **30**, 1322–1324 (2014). doi: 10.1093/bioinformatics/btu013; pmid: 24413670
44. L. M. Breckels *et al*., Learning from heterogeneous data sources: An application in spatial proteomics. *PLOS Comput. Biol.* **12**, e1004920 (2016). doi: 10.1371/journal.pcbi.1004920; pmid: 27175778

45. A. Fabregat *et al*., The Reactome pathway knowledgebase. *Nucleic Acids Res.* **44**, D481–D487 (2016). doi: 10.1093/nar/gkv1351; pmid: 26656494
46. J. Menche *et al*., Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015). doi: 10.1126/science.1257601; pmid: 25700523
47. A. Sakaue-Sawano *et al*., Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008). doi: 10.1016/j.cell.2007.12.033; pmid: 18267078
48. Gene Ontology Consortium, Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015). doi: 10.1093/nar/gku1179; pmid: 25428369
49. L. C. Crosswell, J. M. Thornton, ELIXIR: A distributed infrastructure for European biological data. *Trends Biotechnol.* **30**, 241–242 (2012). doi: 10.1016/j.tibtech.2012.02.002; pmid: 22417641
50. C. E. Chapple *et al*., Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.* **6**, 7412 (2015). doi: 10.1038/ncomms8412; pmid: 26054620
51. K.-W. Min, S.-H. Lee, S. J. Baek, Moonlighting proteins in cancer. *Cancer Lett.* **370**, 108–116 (2016). doi: 10.1016/j.canlet.2015.09.022; pmid: 26499805
52. M. Lindskog, J. Rockberg, M. Uhlén, F. Sterky, Selection of protein epitopes for antibody production. *Biotechniques* **38**, 723–727 (2005). doi: 10.2144/05385ST02; pmid: 15945371
53. T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011). doi: 10.1038/nmeth.1701; pmid: 21959131
54. L. Käll, A. Krogh, E. L. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004). doi: 10.1016/j.jmb.2004.03.016; pmid: 15111065
55. H. Viklund, A. Bernsel, M. Skwark, A. Elofsson, SPOCTOPUS: A combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **24**, 2928–2929 (2008). doi: 10.1093/bioinformatics/btn550; pmid: 18945683
56. L. Fagerberg, K. Jonasson, G. von Heijne, M. Uhlén, L. Berglund, Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010). doi: 10.1002/pmic.200900258; pmid: 20175080
57. F. Danielsson *et al*., RNA deep sequencing as a tool for selection of cell lines for systematic subcellular localization of all human proteins. *J. Proteome Res.* **12**, 299–307 (2013). doi: 10.1021/pr3009308; pmid: 23227862
58. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016). doi: 10.1038/nbt.3519; pmid: 27043002
59. M. Krzywinski *et al*., Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009). doi: 10.1101/gr.092759.109; pmid: 19541911
60. D. Binns *et al*., QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009). doi: 10.1093/bioinformatics/btp536; pmid: 19744993
61. 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 (2008). doi: 10.1038/nprot.2008.211; pmid: 19131956
62. P. Shannon *et al*., Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). doi: 10.1101/gr.1239303; pmid: 14597658

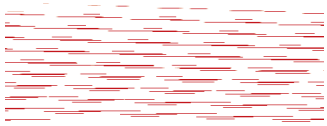
ACKNOWLEDGMENTS

We acknowledge the staff of the Human Protein Atlas program and Science for Life Laboratory for valuable contributions. We also acknowledge support from these facilities of the Science for Life Laboratory: the National Genomics Infrastructure and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure, the National Laboratories for Chemical Biology at Karolinska Institutet for access to imaging infrastructure, and the Advanced Light Microscopy facility for superresolution microscopy. We acknowledge H. Masai (Tokyo Metropolitan Institute of Medical Science) for providing the stable U-2 OS FUCCI cell line; A. Hyman (Max Planck Institute) for HeLa cells expressing GFP-tagged proteins; A. Falk (Karolinska Institute) for the AF22 cells; Evercyte for the ASC TERT1, HUVEC TERT2, HBF TERT88, HTCEpi, LHCN-M2, and RPTEC TERT cells; Horizon Discoveries for the Hap1 CRISPR knockout cells; and R. Pepperkok and B. Neumann for prepared solid-phase siRNA plates. We acknowledge CCP Games and MMOS Srl for the creation of Project Discovery and the players of EVE Online for classification of protein patterns in images through Project Discovery. Funding was provided by the Knut and Alice Wallenberg Foundation and the Erling Persson Foundation to M.U. and the Science for Life Laboratory to E.L. C.M.M. and L.M.B. were funded by Wellcome Trust grant 108441/Z/15/Z. A.G. was funded through the Alexander S. Onassis Public Benefit Foundation, the Foundation for Education and European Culture, and the Embiricos Trust Scholarship of Jesus College Cambridge. M.U., F.P., P.N., and S.H. are cofounders and M.U. and F.P. are on the board of Atlas Antibodies. All antibodies are available from various commercial providers, including Atlas Antibodies. The data are available in the supplementary materials. Images and Cell Atlas transcriptome and proteome data are available in the Human Protein Atlas (www.proteinatlas.org/humancell).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/356/6340/eaal3321/suppl/DC1
Additional Materials and Methods
Figs. S1 to S13
Tables S1 to S19
References (*63*–*66*)

2 November 2016; accepted 31 March 2017
Published online 11 May 2017
10.1126/science.aal3321



A subcellular map of the human proteome

Peter J. Thul, Lovisa Åkesson, Mikaela Wiking, Diana Mahdessian, Aikaterini Geladaki, Hammou Ait Blal, Tove Alm, Anna Asplund, Lars Björk, Lisa M. Breckels, Anna Bäckström, Frida Danielsson, Linn Fagerberg, Jenny Fall, Laurent Gatto, Christian Gnann, Sophia Hober, Martin Hjelmare, Fredric Johansson, Sunjae Lee, Cecilia Lindskog, Jan Mulder, Claire M. Mulvey, Peter Nilsson, Per Oksvold, Johan Rockberg, Rutger Schutten, Jochen M. Schwenk, Åsa Sivertsson, Evelina Sjöstedt, Marie Skogs, Charlotte Stadler, Devin P. Sullivan, Hanna Tegel, Casper Winsnes, Cheng Zhang, Martin Zwahlen, Adil Mardinoglu, Fredrik Pontén, Kalle von Feilitzen, Kathryn S. Lilley, Mathias Uhlén and Emma Lundberg

(May 11, 2017)
Science **356** (6340), . [doi: 10.1126/science.aal3321] originally published online May 11, 2017

Editor's Summary

Mapping the proteome

Proteins function in the context of their environment, so an understanding of cellular processes requires a knowledge of protein localization. Thul *et al.* used immunofluorescence microscopy to map 12,003 human proteins at a single-cell level into 30 cellular compartments and substructures (see the Perspective by Horwitz and Johnson). They validated their results by mass spectroscopy and used them to model and refine protein-protein interaction networks. The cellular proteome is highly spatiotemporally regulated. Many proteins localize to multiple compartments, and many show cell-to-cell variation in their expression patterns. Presented as an interactive database called the Cell Atlas, this work provides an important resource for ongoing efforts to understand human biology.

Science, this issue p. eaal3321; see also p. 806