

## RESEARCH ARTICLE SUMMARY

## PROTEOMICS

# Tissue-based map of the human proteome

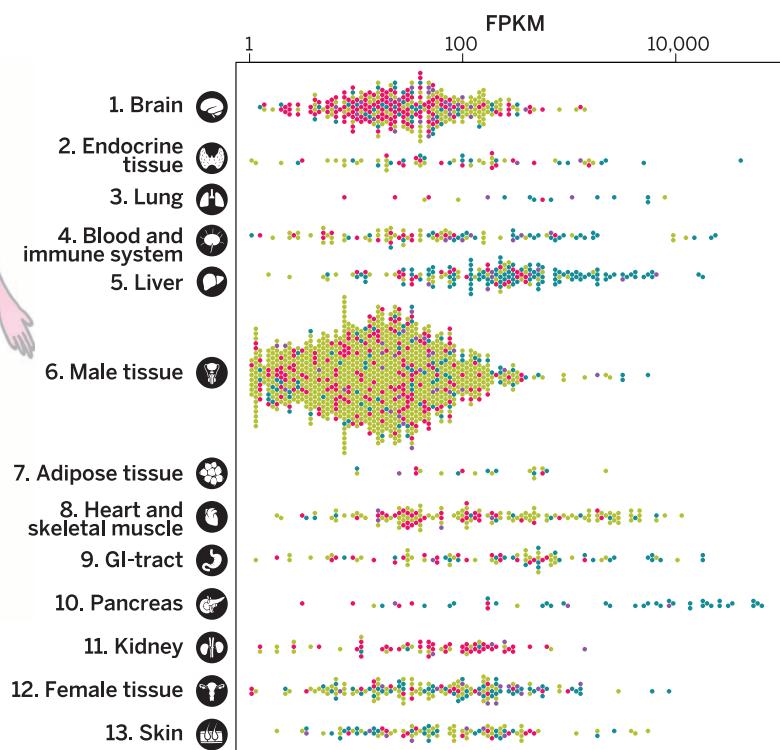
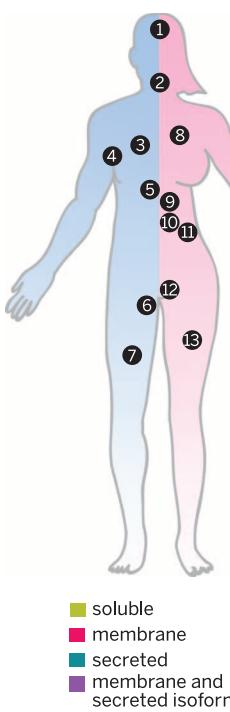
Mathias Uhlen,\* Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M. Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, Fredrik Pontén

**INTRODUCTION:** Resolving the molecular details of proteome variation in the different tissues and organs of the human body would greatly increase our knowledge of human biology and disease. Here, we present a map of the human tissue proteome based on quantitative transcriptomics on a tissue and organ level combined with protein profiling using microarray-based immunohistochemistry to achieve spatial localization of proteins down to the single-cell level. We provide a global analysis of the secreted and membrane proteins, as well as an analysis of the expression profiles for all proteins targeted by pharmaceutical drugs and proteins implicated in cancer.

**RATIONALE:** We have used an integrative omics approach to study the spatial human proteome. Samples representing all major tissues and organs ( $n = 44$ ) in the human body have been analyzed based on 24,028 antibodies corresponding to 16,975 protein-encoding genes, complemented with RNA-seq data for 32 of the tissues. The antibodies have been used to produce more than 13 million tissue-based immunohistochemistry images, each annotated by pathologists for all sampled tissues. To facilitate integration with other biological resources, all data are available for download and cross-referencing.

**RESULTS:** We report a genome-wide analysis of the tissue specificity of RNA and protein expression covering more than 90% of the putative protein-

coding genes, complemented with analyses of various subproteomes, such as predicted secreted proteins ( $n = 3171$ ) and membrane-bound proteins ( $n = 5570$ ). The analysis shows that almost half of the genes are expressed in all analyzed tissues, which suggests that the gene products are needed in all cells to maintain “housekeeping” functions such as cell growth, energy generation, and basic metabolism. Furthermore, there is enrichment in metabolism among these genes, as 60% of all



**The human tissue-enriched proteins.** All tissue-enriched proteins are shown for 13 representative tissues or groups of tissues, stratified according to their predicted subcellular localization. Enriched proteins are mainly intracellular in testis, mainly membrane bound in brain and kidney, and mainly secreted in pancreas and liver.

metabolic enzymes are expressed in all analyzed tissues. The largest number of tissue-enriched genes is found in the testis, followed by the brain and the liver.

## ON OUR WEB SITE

Read the full article at <http://dx.doi.org/10.1126/science.1260419>

Analysis of the 618 proteins targeted by clinically approved drugs unexpectedly showed that 30% are expressed in all analyzed tissues. An analysis of metabolic activity based on genome-scale metabolic models (GEMS) revealed liver as the most metabolically active tissue, followed by adipose tissue and skeletal muscle.

**CONCLUSIONS:** A freely available interactive resource is presented as part of the Human Protein Atlas portal ([www.proteinatlas.org](http://www.proteinatlas.org)), offering the possibility to explore the tissue-elevated proteomes in tissues and organs and to analyze tissue profiles for specific protein classes. Comprehensive lists of proteins expressed at elevated levels in the different tissues have been compiled to provide a spatial context with localization of the proteins in the subcompartments of each tissue and organ down to the single-cell level. ■

The list of author affiliations is available in the full article online.

\*Corresponding author. E-mail: mathias.uhlen@scilifelab.se  
Cite this article as M. Uhlen et al., *Science* **347**, 1260419 (2015). DOI: 10.1126/science.1260419

## RESEARCH ARTICLE

## PROTEOMICS

# Tissue-based map of the human proteome

Mathias Uhlen,<sup>1,2,3\*</sup> Linn Fagerberg,<sup>1</sup> Björn M. Hallström,<sup>1,2</sup> Cecilia Lindskog,<sup>4</sup> Per Oksvold,<sup>1</sup> Adil Mardinoglu,<sup>5</sup> Åsa Sivertsson,<sup>1</sup> Caroline Kampf,<sup>4</sup> Evelina Sjöstedt,<sup>1,4</sup> Anna Asplund,<sup>4</sup> IngMarie Olsson,<sup>4</sup> Karolina Edlund,<sup>6</sup> Emma Lundberg,<sup>1</sup> Sanjay Navani,<sup>7</sup> Cristina Al-Khalili Szigyarto,<sup>2</sup> Jacob Odeberg,<sup>1</sup> Dijana Djureinovic,<sup>4</sup> Jenny Ottosson Takanen,<sup>2</sup> Sophia Hober,<sup>2</sup> Tove Alm,<sup>1</sup> Per-Henrik Edqvist,<sup>4</sup> Holger Berling,<sup>2</sup> Hanna Tegel,<sup>2</sup> Jan Mulder,<sup>8</sup> Johan Rockberg,<sup>2</sup> Peter Nilsson,<sup>1</sup> Jochen M. Schwenk,<sup>1</sup> Marica Hamsten,<sup>2</sup> Kalle von Feilitzen,<sup>1</sup> Mattias Forsberg,<sup>1</sup> Lukas Persson,<sup>1</sup> Fredric Johansson,<sup>1</sup> Martin Zwahlen,<sup>1</sup> Gunnar von Heijne,<sup>9</sup> Jens Nielsen,<sup>3,5</sup> Fredrik Pontén<sup>4</sup>

Resolving the molecular details of proteome variation in the different tissues and organs of the human body will greatly increase our knowledge of human biology and disease. Here, we present a map of the human tissue proteome based on an integrated omics approach that involves quantitative transcriptomics at the tissue and organ level, combined with tissue microarray-based immunohistochemistry, to achieve spatial localization of proteins down to the single-cell level. Our tissue-based analysis detected more than 90% of the putative protein-coding genes. We used this approach to explore the human secretome, the membrane proteome, the druggable proteome, the cancer proteome, and the metabolic functions in 32 different tissues and organs. All the data are integrated in an interactive Web-based database that allows exploration of individual proteins, as well as navigation of global expression patterns, in all major tissues and organs in the human body.

There is much interest in annotating all human genes at the level of DNA (*1, 2*), RNA (*3, 4*), and proteins (*5, 6*), with the ultimate goal of defining structure, function, localization, expression, and interactions of all proteins. This has resulted in large-scale projects, such as ENCODE (*7*) and the Human Proteome Project (*8*), aimed to integrate results from many research groups and technical platforms to reach a detailed understanding of each of the ~20,000 human protein-coding genes predicted from the human genome and their corresponding protein isoforms. Recently, drafts of the human proteome based on proteogenomics efforts have been described (*9, 10*), focusing on recent advances in mass spectrometry that allow comprehensive analyses using both isotope-labeled analysis sys-

tems (*11*) and deep proteomics methods (*12*) or genome-wide targeted proteomics efforts (*13*).

A complement to these efforts is the Human Protein Atlas program (*14*), which is exploring the human proteome using gene-centric and genome-wide antibody-based profiling on tissue microarrays. This allows for spatial pathology-based annotation of protein expression, in combination with deep-sequencing transcriptomics of the same tissue types. The strategy is based on the quantitative assessment of transcript expression in complex tissue homogenates, involving a mixture of cell types combined with the precise localization of the corresponding proteins down to the single-cell level, using immunohistochemistry. Recently, we performed a transcriptomics study of 27 different tissues using this approach (*15*), followed by subsequent in-depth studies of the global proteome in a number of these tissues and organs, such as liver (*16*), testis (*17*), and the gastrointestinal (GI) tract (*18*). Here, we have used this approach and extended the analysis to 32 tissue types, representing all major tissues and organs in the human body, to create a genome-wide map of the human tissue-based proteome, with a focus on the analysis of the tissue-elevated proteins and all secreted and membrane proteins. Particular emphasis has been placed on analyses of proteins targeted by pharmaceutical drugs (*19*) and proteins implicated in cancer (*20*). We used the data to generate comprehensive metabolic maps for all 32 tissue types in order to identify differences in metabolism between tissues. In

addition, new transcriptomics data from 36 human cell lines allowed us to compare the proteomes between cell lines and normal cells derived from the same tissue types. Finally, the protein isoforms generated by differential splicing between different tissues were studied with a focus on splice variants with predicted differential subcellular localization. All data are presented in an interactive database ([www.proteinatlas.org](http://www.proteinatlas.org)).

## Results

### Classification of all human protein-coding genes

Samples representing all major tissues and organs (*n* = 44) in the human body were analyzed (Fig. 1A) by using 20,456 antibodies generated “in-house,” as well as 3572 antibodies provided by external suppliers. The antibodies have been used to produce more than 13 million tissue-based immunohistochemistry images, with each image annotated on the single-cell level for all sampled tissues by pathologists. The analysis was complemented with RNA sequencing (RNAseq) data for 32 out of the 44 tissue types. We investigated global expression profiles using hierarchical clustering based on the correlation between 122 biological replicates from the 32 organs and tissues (Fig. 1B and fig. S1). The results reveal testis and brain as outliers and a clear connectivity between the samples from the GI tract (stomach, duodenum, small intestine, colon, and rectum), the hematopoietic tissues (bone marrow, lymph node, spleen, tonsil, and appendix) and the two striated muscle samples (cardiac and skeletal muscle). A principal component analysis (fig. S2A) confirms a close resemblance between cardiac and skeletal muscle but also suggests similarities in global expression between pancreas and salivary gland, as well as differences between the primary lymphoid tissue (bone marrow) and the secondary lymphoid tissues, such as tonsil and spleen.

The transcriptomics study allowed us to refine the classification performed earlier (*15*) of all the 20,344 putative protein-coding genes with RNA-seq data into categories based on their expression across all 32 tissue types (Fig. 1C, Table 1, and tables S1 to S4). Indirectly, this also provides an estimate of the relative protein levels corresponding to each gene, because proteogenomics analyses have shown that the translation rate, in most cases, is constant for a specific transcript across different human cells and tissues at both a cellular level (*21*) and a tissue level (*9*). Although it is still a matter of scientific debate (*22*) whether protein degradation rates could, in some cases, vary for an individual protein in different tissues, an overall concurrence between mRNA and protein levels for a given gene product across various tissues is generally expected (*9, 21*). A large fraction (44%) of the protein-coding genes were detected in all analyzed tissues, and these ubiquitously expressed genes include known “housekeeping” genes encoding mitochondrial proteins and proteins involved in overall cell structure, translation, transcription, and replication. Of all the protein coding genes, 34% showed an elevated expression in at least one of the analyzed tissues,

<sup>1</sup>Science for Life Laboratory, KTH—Royal Institute of Technology, SE-171 21 Stockholm, Sweden. <sup>2</sup>Department of Proteomics, KTH—Royal Institute of Technology, SE-106 91 Stockholm, Sweden. <sup>3</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark. <sup>4</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden. <sup>5</sup>Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. <sup>6</sup>Leibniz Research Centre for Working Environment and Human Factors (IfADo) at Dortmund TU, D-44139 Dortmund, Germany. <sup>7</sup>Lab Surgpath, Mumbai, India. <sup>8</sup>Science for Life Laboratory, Department of Neuroscience, Karolinska Institute, SE-171 77 Stockholm, Sweden. <sup>9</sup>Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden.  
\*Corresponding author. E-mail: [mathias.uhlen@scilifelab.se](mailto:mathias.uhlen@scilifelab.se)

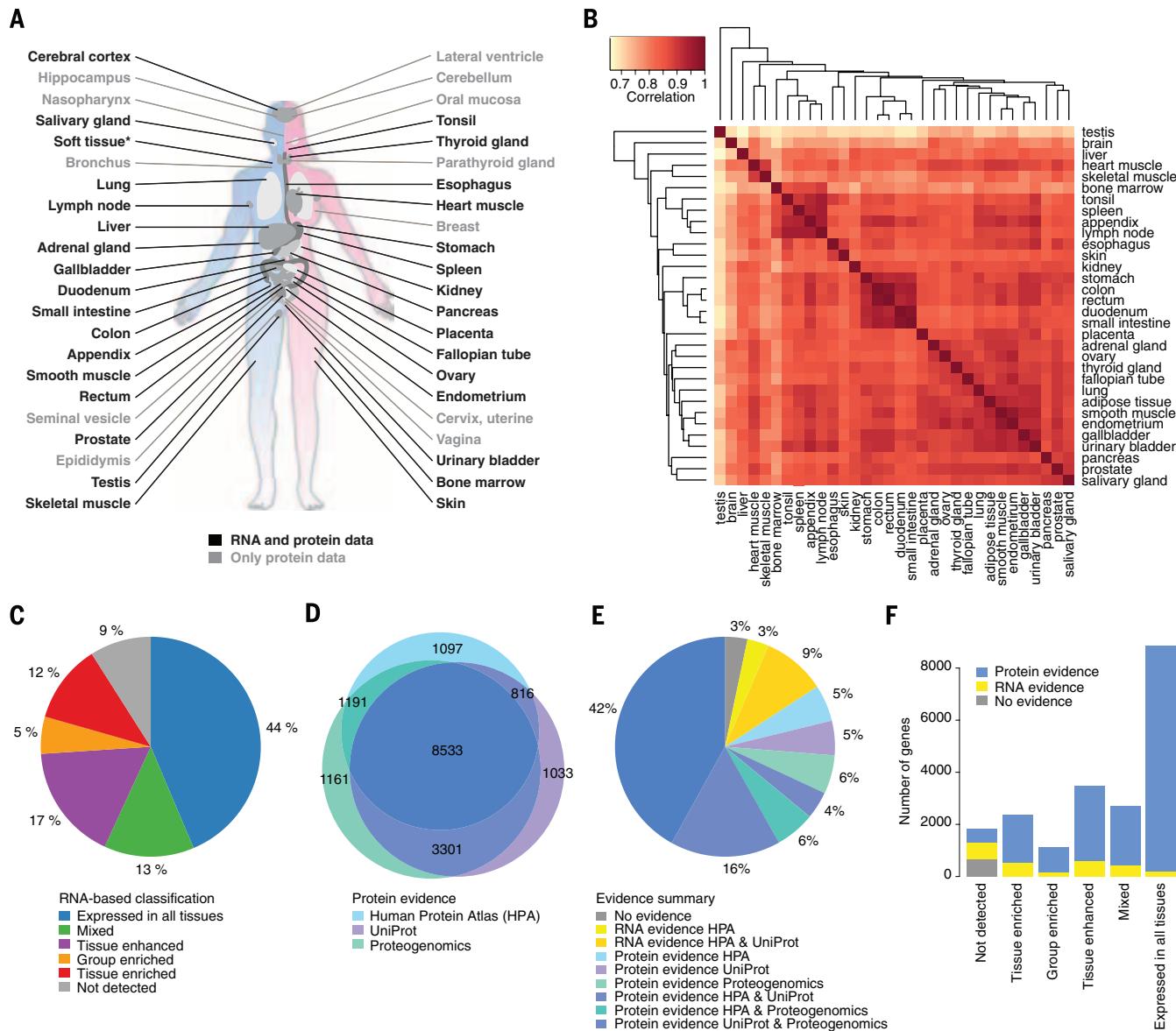
and these were further subdivided into (i) enriched genes with mRNA levels in one tissue type at least five times the maximum levels of all other analyzed tissues, (ii) group-enriched genes with enriched expression in a small number of tissues, and (iii) enhanced genes with only a moderately elevated expression. The use of the word “tissue-specific” has been avoided because this definition depends on arbitrary cut-off levels, and many proteins described in the literature as “tissue-specific” are here shown to be expressed in several tissues. This is exemplified by albumin, which we, as expected, identified as enriched in

liver but also found at high levels, albeit much lower than for liver, in kidney and pancreas.

### Evidence for the human protein-coding genes

We have determined the number of genes for which evidence is available at a protein level by combining our antibody-based data with the manual annotation of literature by the UniProt consortium (5) and the results from the recent mass spectrometry-based proteogenomics analyses (9, 10, 12). The analysis shows that there are 17,132 protein-coding genes with proteins identi-

fied from at least one of the three efforts and 13,841 genes with experimental evidence from at least two of the efforts (Fig. 1D). Furthermore, there is evidence, at the RNA level, for 2546 additional genes based on either our data or annotations by UniProt. Although proteins not yet detected by one of the three methods should be further investigated to establish them as true human proteins, it is noteworthy that out of the 20,356 putative protein-coding genes (in Ensembl release 75) there are only 677 genes (3.3%) for which there is no experimental evidence (table S5). Many of these genes were removed in the



**Fig. 1. Classification and protein evidence of the human protein-coding genes.** (A) The tissues analyzed in this study, including tissues studied both by RNAseq and antibody-based profiling and those analyzed only by antibody-based profiling. For details see table S1. (B) Heat map showing the pairwise correlation between all 32 tissues based on transcript expression levels of 20,344 genes. The average FPKM values for each gene and tissue are used in the analysis. For correlation results of all individual samples, see fig. S1. (C) The number of genes

classified in each expression category according to the definition stated in Table 1. (D) Venn diagram showing the overlap between protein evidence on the basis of three sources: Human Protein Atlas, UniProt, and Proteogenomics. (E) The distribution of genes classified as having protein evidence, evidence only at the transcript level, and genes without any experimental evidence. (F) The number of genes with protein evidence, RNA evidence, and no evidence stratified according to their transcriptomics-based classification into six categories.

later update of Ensembl (release 76) (fig. S2B), and others have been suggested to be noncoding genes on the basis of the lack of correlation in gene family age and cross-species conservation studies. Thus, it is possible that most of these “missing genes” will be removed from the list of protein-coding genes in the future. These genes and the genes with evidence only at the RNA level are obvious targets for more in-depth functional protein studies. A summary of the supporting data is shown in Fig. 1E. Few (2%) of

the ubiquitously expressed genes lack protein evidence (Fig. 1F); however, protein evidence is lacking for 18% of the genes identified here by RNA analysis as elevated (tissue enriched, group enriched, or enhanced). Examples of genes with no previous evidence on the protein level according to UniProt, but now confirmed using antibody-based profiling and proteogenomics (9, 10), are chromosome 2 open reading frame 57 (C2orf57), shown here with an enriched expression in testis localized to the sperm (Fig. 2A), and chromo-

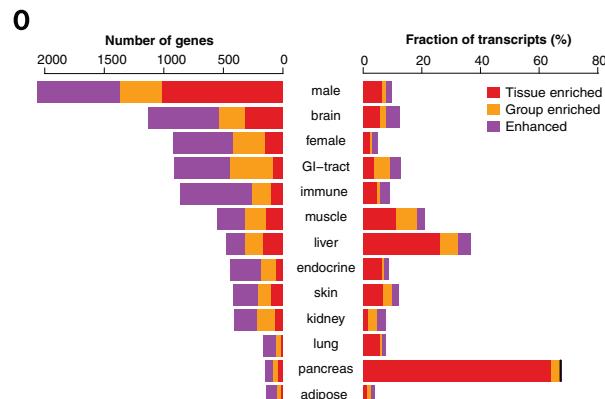
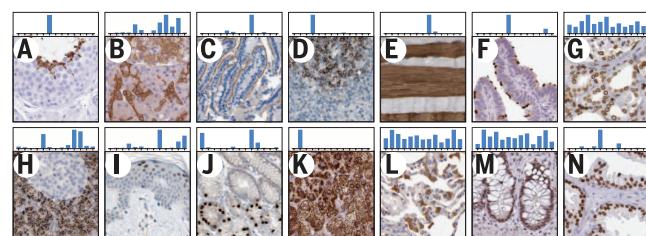
some 8 open reading frame 47 (C8orf47), with expression in a subset of endocrine islet cells and ductal cells of the exocrine pancreas (Fig. 2B).

### The tissue-elevated proteome

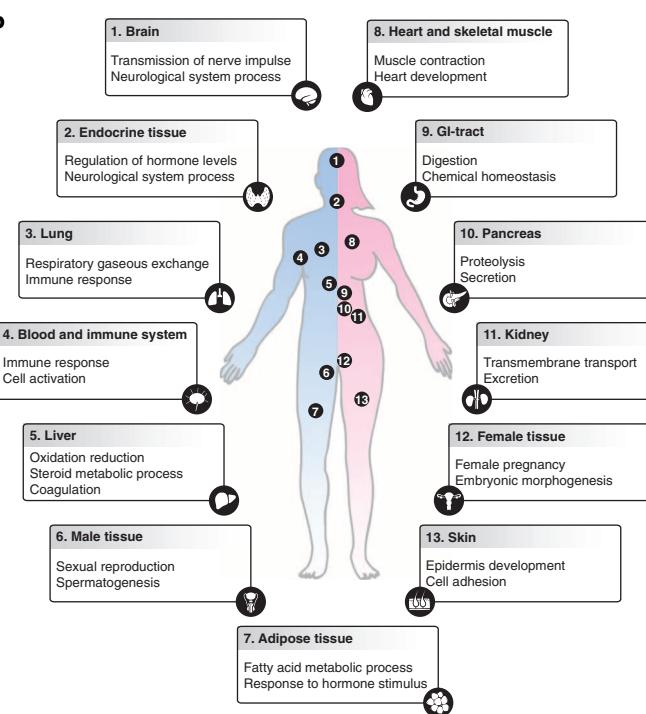
A network plot shows the number of tissue-enriched genes for each tissue type, as well as the number of genes enriched in different groups of tissues and organs (fig. S4). An analysis of selected tissues and organs (Fig. 2O) reveals a large number of elevated genes in male tissue, brain,

**Table 1. Classification of all human protein-coding genes based on transcript expression levels in 32 tissues.**

Category	Description	No. of genes	Fraction of genes (%)
Tissue enriched	mRNA levels in a particular tissue at least five times those in all other tissues	2,355	12
Group enriched	mRNA levels at least five times those in a group of 2–7 tissues	1,109	5
Tissue enhanced	mRNA levels in a particular tissue at least five times average levels in all tissues	3,478	17
Expressed in all	Detected in all tissues (FPKM > 1)	8,874	44
Mixed	Detected in fewer than 32 tissues but not elevated in any tissue	2,696	13
Not detected	FPKM < 1 in all tissues	1,832	9
Total	Total number of genes analyzed with RNAseq	20,344	100
Total elevated	Total number of tissue-enriched, group-enriched, and tissue-enhanced genes	6,942	34



**Fig. 2. Tissue microarray-based protein expression, and analysis of tissue-elevated genes in the different organ systems. (A to N)** Tissue expression and localization for a selection of human proteins. Larger images corresponding to (A) to (N) of the figure are shown in fig. S3. The levels of the corresponding mRNA (FPKM) are displayed as bars for each of the 13 organ systems analyzed (from left: brain, endocrine tissue, lung, blood and immune system, liver, male tissue, adipose tissue, heart and skeletal muscle, GI tract, pancreas, kidney, female tissue, and skin). Examples include testis with C2orf57 expression in sperm (A), pancreas with cytoplasmic C8orf47 expression in both a subset of endocrine cells and ductal cells (B), duodenum with CDHR2 expression in microvilli (C), lymph node with cytoplasmic FCRLA expression in germinal center cells (D), skeletal muscle with cytoplasmic MYL3 expression in slow muscle fibers (E), fallopian tube with ROPNL expression in cilia (F), kidney with SUN2



expression in all nuclear membranes (G), pancreas with GATM expression in mitochondria throughout the exocrine compartment (H), skin with GRHL1 expression in nuclei of the upper epidermal layer (I), stomach with nuclear PAX6 expression in endocrine cells (J), adrenal gland with cytoplasmic expression of CYP11B1 in cortical cells (K), lung with cytoplasmic COMT expression in a subset of pneumocytes and macrophages (L), colon with nuclear ATF1 expression in glandular cells (M), and prostate with nuclear FOXA1 expression in glandular cells (N). (O) The number of elevated genes in the 13 organ systems, as described in (P), and the fraction of all transcripts (FPKM) encoded by these elevated genes for each of these organ systems. (P) An analysis of major GO terms for each tissue on the basis of the tissue-elevated genes in 13 selected tissues or groups of tissues, as described in supplementary methods. For more details of the GO analysis, see table S6.

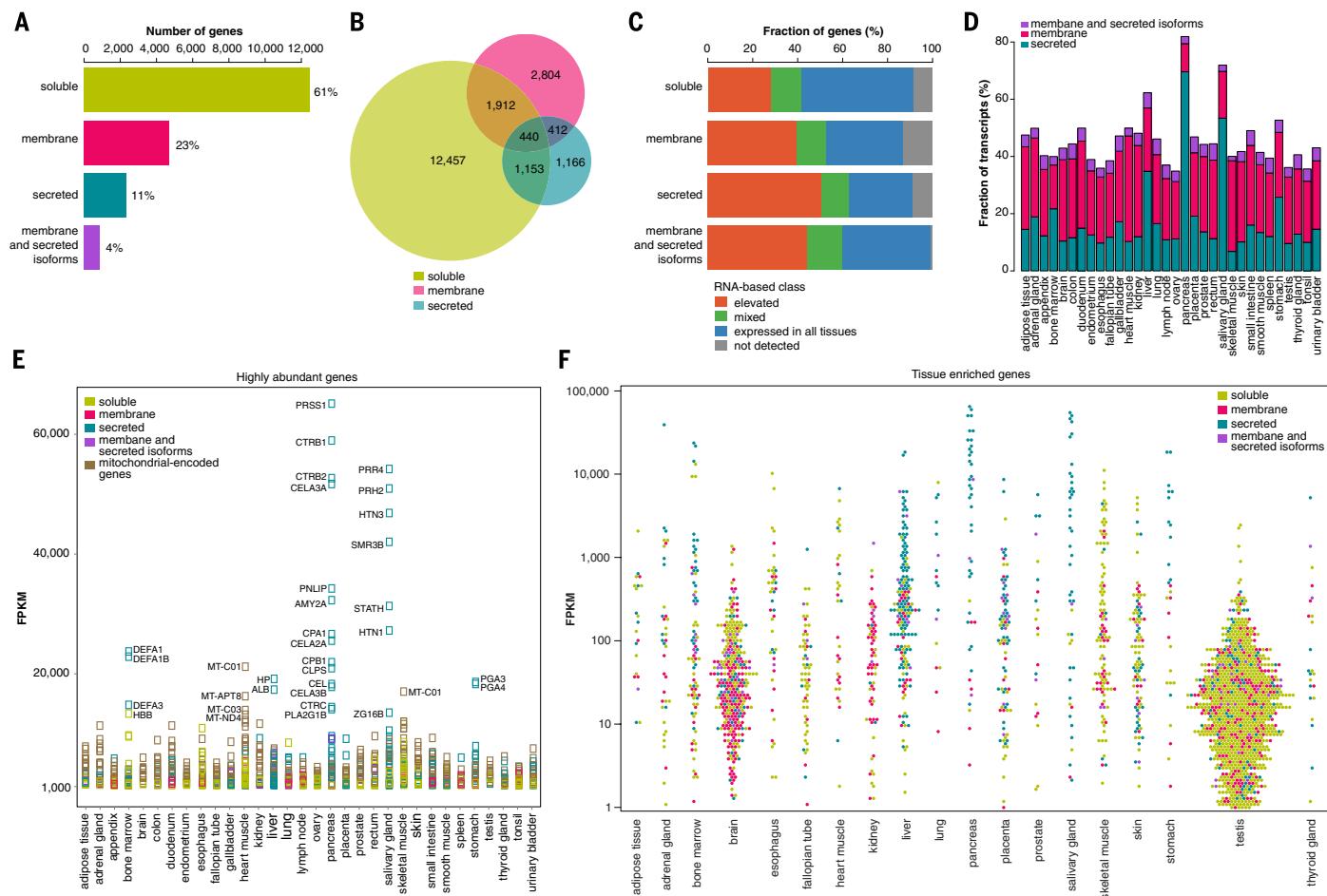
and liver and relatively few in lung, pancreas, and fat (adipose tissue). The transcriptomics analysis also allowed us to determine the fraction of elevated transcripts in each tissue (Fig. 2O). For most tissues, only ~10% of the transcripts are encoded by tissue-elevated genes, with the exception of pancreas and liver, where elevated genes encode 70% and 35% of the transcripts, respectively.

A functional Gene Ontology (GO) analysis for 13 tissues or groups of tissues is summarized in Fig. 2P (see table S6 for details), and the terms identified are consistent with the function of the respective tissues. The largest number of enriched genes is found in the testis ( $n = 999$ ), with many of the corresponding testis-specific proteins involved in the reproductive process and spermatogenesis. It is not unlikely that many of these genes will show a shared expression with oocytes in the female ovaries, which are difficult to analyze because of the different kinetics of

germ cell development, including first rounds of meiosis at the embryonic stages during female life. The tissue with the second largest number of enriched genes is the brain ( $n = 318$ ). The number of genes with expression restricted to neuronal tissue is relatively small, but it is likely that more enriched genes would be added to the list if additional regions, such as the various specialized regions of the brain, were sampled. Genes elevated in liver encode secreted plasma and bile proteins, detoxification proteins, and proteins associated with metabolic processes and glycogen storage, whereas genes elevated in adipose tissue encode proteins involved in lipid metabolic processes, secretion, and transport. Genes elevated in skin encode proteins associated with functions related to the barrier function (squamous cell differentiation and cornification), skin pigmentation, and hair development. In the GI tract, elevated genes predominantly encode proteins

involved in nutrient breakdown, transport, and metabolism; host protection; and tissue morphology maintenance.

As expected, many of the genes enriched in groups of tissues are common for the GI tract and the hematopoietic tissues, respectively, as exemplified on the protein level by cadherin-related family member 2 (CDHR2), expressed in the microvilli of duodenum and small intestine (Fig. 2C), and Fc receptor-like A (FCRLA), expressed in lymph node, tonsil, appendix, and spleen (Fig. 2D). A large number of group-enriched genes involved in contraction are observed in striated (cardiac and skeletal) muscle, as exemplified by the fiber type-specific expression of myosin light chain 3 (MYL3) (Fig. 2E), whereas many genes shared between testis and the fallopian tube, as well as testis and lung, are involved in cell motility, as exemplified by rophilin-associated tail protein-like (ROPN1L), which is expressed in



**Fig. 3. Prediction and analysis of the human secreted and membrane-spanning proteins.** (A) The number and fraction of all human genes ( $n = 20,356$ ) classified into the categories soluble, membrane-spanning, and secreted, as well as genes with isoforms belonging to two or all three categories. (B) Venn diagram showing the number of genes in each of the three main subcellular location categories: membrane, secreted, and soluble. The overlap between the categories gives the number of genes with isoforms belonging to two or all three categories. (C) The fraction of genes in the various protein expression classes for the soluble, secreted, and membrane-spanning proteins, as well as genes with both secreted and

membrane-spanning isoforms. (D) The fraction of transcripts based on FPKM values from each of the three secreted or membrane-spanning categories across the 32 analyzed tissues. (E) The 370 most-abundant genes ( $\text{FPKM} > 1000$ ) in the different tissues, stratified according to their predicted localization on the basis of (C), as well as an additional category of the 13 genes encoded by the mitochondrial genome. The gene names for a selection of the most abundant genes are shown. (F) The transcript levels (FPKM) on a  $\log_{10}$  scale for all genes identified as tissue-enriched are shown for a few selected tissues, with each gene stratified according to predicted localization.

sperm (testis), ciliated cells in respiratory epithelia (lung), and ciliated cells in the fallopian tube (Fig. 2F).

### The human secretome and membrane proteome

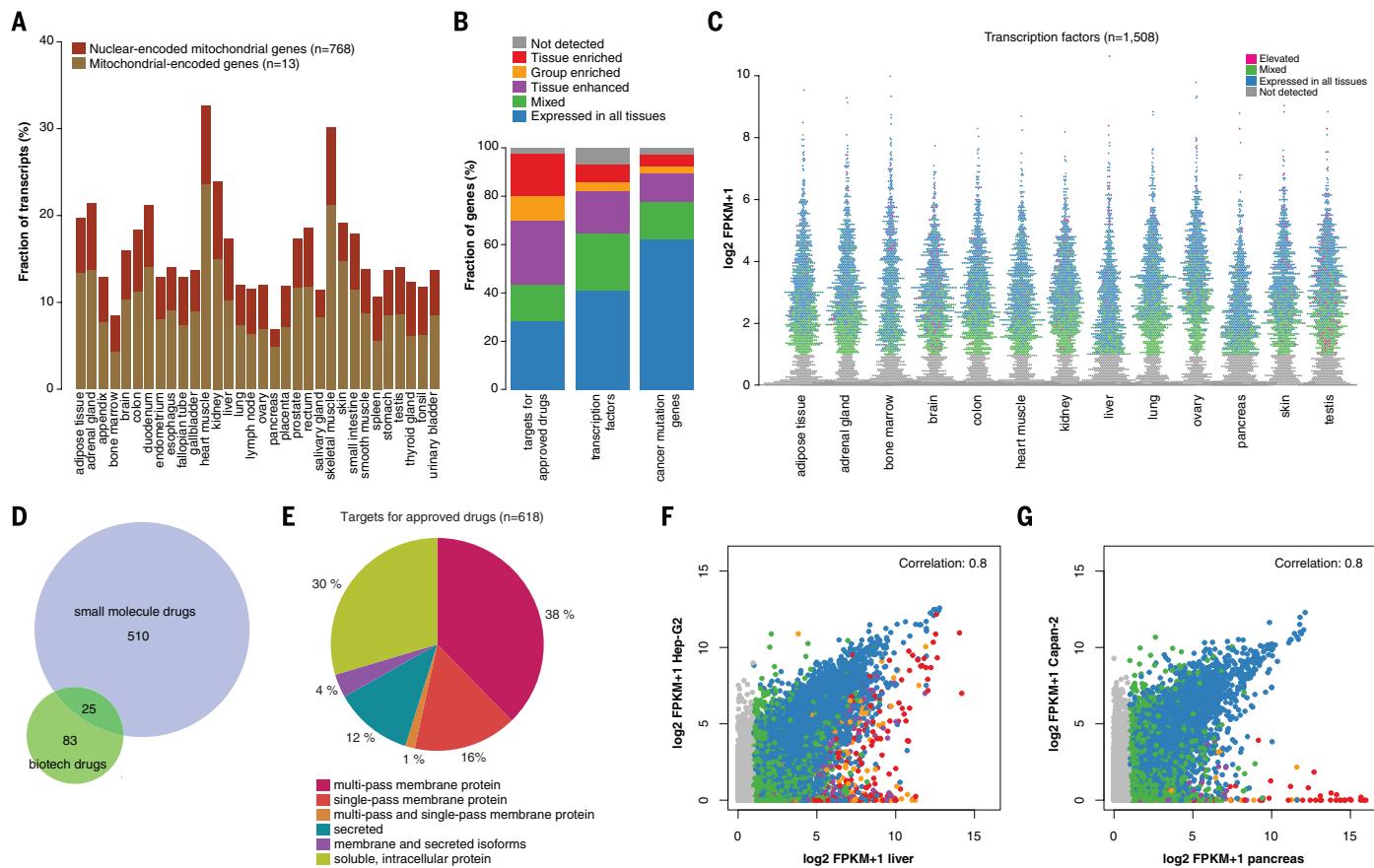
Both secreted and membrane-bound proteins play crucial roles in many physiological and pathological processes. Important secreted proteins include cytokines, coagulation factors, hormones, and growth factors, whereas membrane proteins include ion channels or molecular transporters, enzymes, receptors, and anchors for other proteins. Here, we performed a whole-proteome scan to predict the complete set of human secreted proteins (“secretome”) using three methods for signal-peptide prediction: SignalP4.0 (23), Phobius (24), and SPECTOPUS (25). In addition, the human membrane proteome was predicted using seven membrane–protein topology prediction methods as described (21), which resulted in a majority decision-based method (MDM). For each protein-coding gene, all protein isoforms were annotated for predicted localization: secreted,

membrane spanning, or soluble (intracellular proteins without a predicted signal peptide or membrane-spanning region) (table S1). Some of the proteins predicted to be membrane-spanning are intracellular, e.g., in the Golgi or mitochondrial membranes, and some of the proteins predicted to be secreted could potentially be retained in a compartment belonging to the secretory pathway, such as the endoplasmic reticulum (ER), or remain attached to the outer face of the cell membrane by a GPI anchor. About 3000 human genes are predicted to encode secreted proteins, with another 5500 encoding membrane-bound proteins (Fig. 3A). In the interactive database ([www.proteinatlas.org](http://www.proteinatlas.org)), many of the secreted proteins are detected at the RNA level in tissues, but no protein expression is observed in the antibody-based analysis in the same tissue—most likely because the steady-state levels of proteins in the cell during the secretion process are too low to be detected.

A large fraction (72%) of human genes encode multiple splice variants with different protein sequences. In Fig. 3B, all genes have been classified

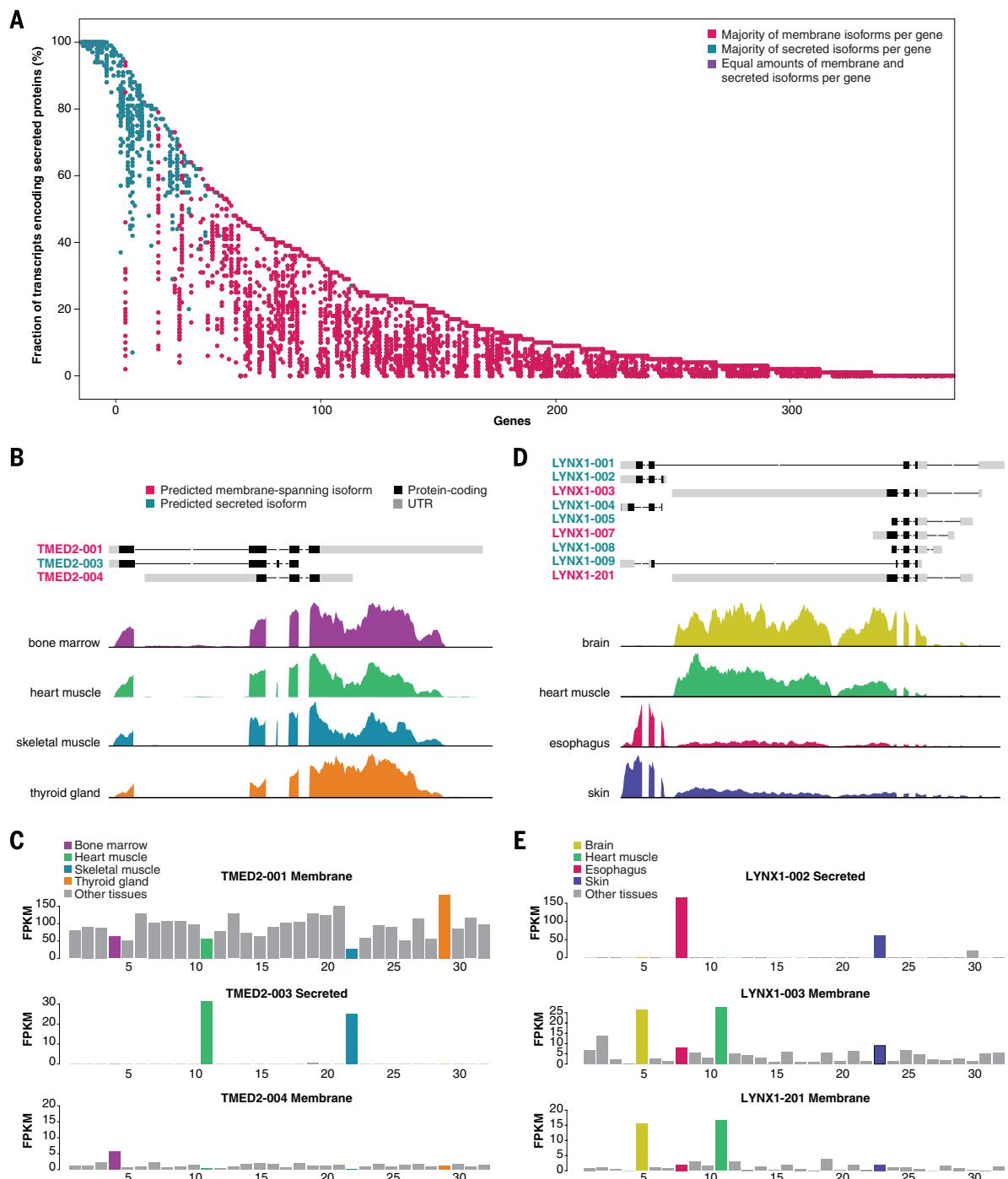
according to the presence of protein isoforms that are intracellular, membrane-spanning, and/or secreted. Note that two-thirds of the genes encoding secreted proteins have at least one splice variant with alternative localization. All protein isoforms ( $n = 94,856$ ) with their predicted localization based on the three signal-peptide-prediction methods, as well as the number of predicted transmembrane segments, are listed in table S7. An analysis across the 32 tissues (Fig. 3C) supports earlier suggestions (21, 26) that a larger fraction of tissue-enriched proteins are secreted or membrane-spanning proteins than are intracellular proteins.

Furthermore, we investigated the fraction of the transcriptome that codes for each class of proteins across the 32 tissues (Fig. 3D and fig. S4). In most cases, the secreted proteins account for between 10 and 20% of the transcripts. In contrast, more than 70% of the transcripts from the pancreas and ~60% from the salivary gland encode secreted proteins. This demonstrates the extreme specialization of these two tissues for production of secreted proteins into the duodenum



**Fig. 4. The human transcriptome in different tissues and organs.** (A) The fraction of transcripts encoded by mitochondrial genes for each of the different tissues and organs, subdivided by genes encoded by the mitochondrial genome and chromosomes, respectively. (B) The fraction of genes classified according to tissue expression pattern and analyzed for all targets of approved drugs ( $n = 618$ ), all transcription factors ( $n = 1508$ ), and proteins implicated in cancer ( $n = 525$ ). (C) The transcript levels (FPKM) for all genes encoding transcription factors in some selected tissues, color-coded according to their global expression category. (D)

The number of pharmaceutical drugs approved by FDA, according to Drugbank (19), that are chemical (small-molecule) or biotech drugs. (E) The number of pharmaceutical drugs approved by FDA (19) stratified according to the predicted localization of the target protein. (F) Pairwise comparison showing all genes expressed in liver tissue and the liver cell line Hep-G2, color-coded according to protein expression category as shown in (B). (G) Pairwise comparison showing all genes expressed in pancreas tissue and the pancreas cell line Capan-2, color-coded according to protein expression category as shown in (B).



**Fig. 5. Differential splicing analysis of transcripts.** (A) Dot plot of genes with multiple isoforms, where at least one isoform is classified as membrane-spanning and another classified as secreted. The x axis shows 366 genes expressed at >5 FPKM in one or more tissues; the y axis shows the sum of FPKM values for all secreted isoforms divided by the total sum of FPKM values for each tissue expressed at >5 FPKM. For each gene, the number of tissues where the secreted transcripts are more abundant than the membrane-spanning transcripts is calculated to define a majority fraction-type as membrane (red), secreted (blue), or equal number for both categories (purple). Each tissue is represented by a circle, and the color is the same across all tissues for the same gene. (B) Example of differential splicing for the gene *TMED2*, with two isoforms predicted as membrane-spanning and one isoform predicted as secreted. The exon-intron structure (with pure intronic sites removed), as well as the location of the untranslated regions (UTR) of three splice variants of *TMED2*, are shown

on top. Normalized read coverage plots for cardiac muscle, skeletal muscle, thyroid gland, and bone marrow highlight the differential use of exons in the selected tissues. (C) Transcript abundance (FPKM values) plotted across all 32 tissues for each isoform. The predicted membrane-spanning transcript (top) is expressed in all tissues, with thyroid gland as the most abundant tissue; a secreted isoform (middle) is only detected in cardiac and skeletal muscle; and a second membrane-spanning isoform (bottom) is expressed at very low levels, with bone marrow as most abundant. (D) Examples of differential splicing for the gene *LYNX1*, with three isoforms predicted as membrane-spanning and six isoforms predicted as secreted from the visualization used in (B). (E) Transcript abundance (FPKM values) for three isoforms of *LYNX1* detected at >5 FPKM. The secreted isoform (top) is expressed at high levels in esophagus and skin, whereas the two membrane-spanning isoforms (middle and bottom) are most abundant in brain and cardiac muscle.

and oral cavity, respectively. About 40% of the transcripts in liver encode secreted proteins. Other tissues with relatively high levels of transcripts encoding secreted proteins include gallbladder, bone marrow, placenta, and different parts of the GI tract, such as stomach, duodenum, and small intestine.

The most abundant genes, normalized as fragments per kilobase of exon per million fragments mapped (FPKM) with a value >1000, in the different tissues are shown in Fig. 3E, and the prediction of the localization of the corresponding proteins reveals that many (53%) are secreted proteins. Among the predicted membrane-spanning proteins, 13 proteins encoded in the mitochondrial genome are the most highly expressed. In Fig. 3F, tissue-enriched genes are shown stratified according to their predicted subcellular localization. Many of the tissue-enriched genes in testis are intracellular, whereas a large number of the tissue-enriched genes in brain and kidney are membrane-bound. In contrast, in many other tissues, such as pancreas, salivary gland, liver, stomach, and bone marrow, most tissue-enriched genes are secreted (fig. S5).

### The housekeeping proteome

Transcriptomics analysis shows that close to 9000 genes (table S1) are expressed in all analyzed tissues, which suggests that the gene products are needed in all cells to maintain basic cellular structure and function. These housekeeping proteins include ribosomal proteins involved

in protein synthesis, enzymes essential for cell metabolism and gene expression, and mitochondrial proteins needed for energy generation, as well as proteins responsible for the structural integrity of the cell. Most of these proteins are expressed at similar levels throughout the human body, as exemplified in kidney by the expression of the nuclear membrane protein SUN2 present in all cells (Fig. 2G), whereas a few proteins show great variability in expression levels—for example, the mitochondrial protein glycine amino transferase (GATM), with high expression in exocrine pancreas (Fig. 2H), kidney, and liver but relatively low expression levels in all other tissues. An interesting class of proteins is encoded by mitochondrial genes, and in Fig. 4A, the transcriptional load of these genes is shown across different tissues. The highest fractions of transcripts encoding mitochondrial proteins are found in cardiac muscle (32% of all transcripts) and skeletal muscle (28%), which demonstrates the importance of energy metabolism for striated muscle tissue.

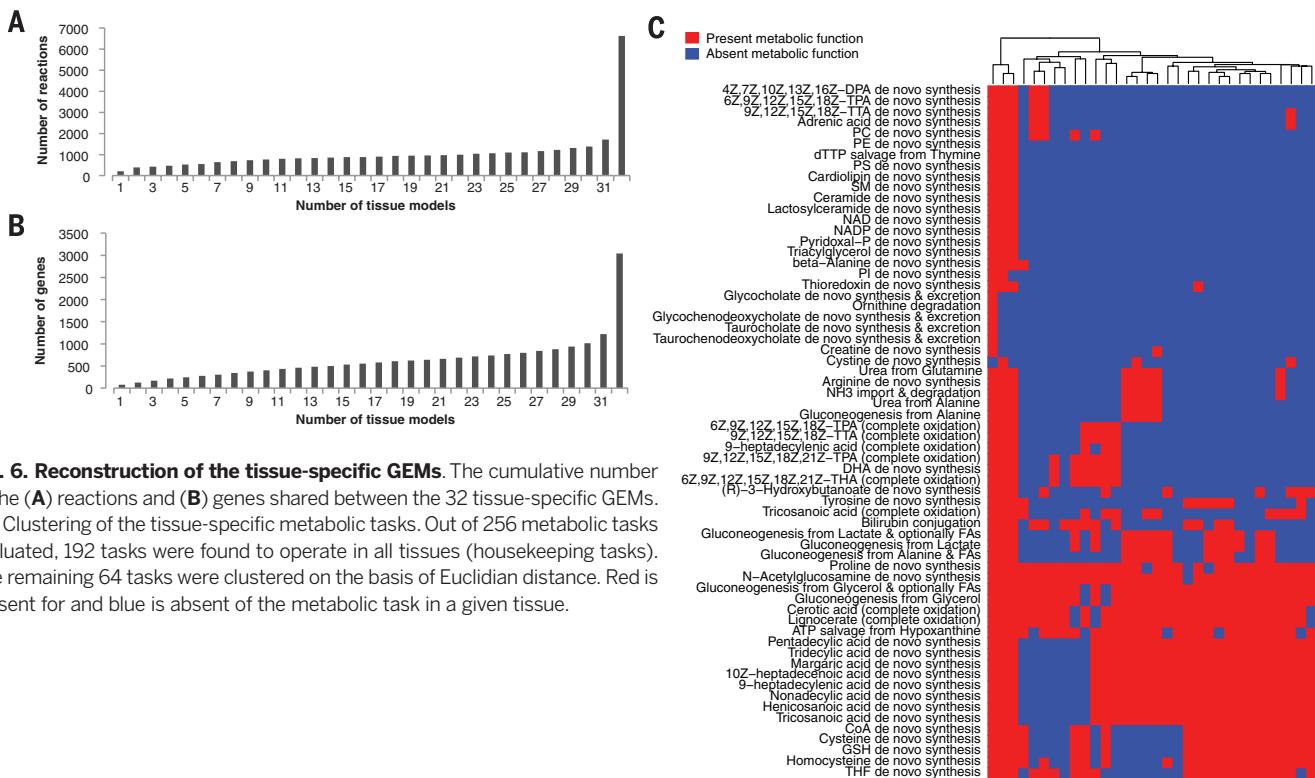
### The regulatory proteome

Transcription factors, of which ~1,500 have been identified in humans (27), comprise an important class of regulatory proteins as they function as on/off switches for gene expression. The fraction of transcription factor genes classified according to tissue specificity is shown in Fig. 4B, which suggests a tissue distribution similar to that of the complete proteome, with as many as 41% of the genes expressed in all tissues and only 29%

identified as elevated (enriched, group enriched, or enhanced). Many of the more-abundantly expressed transcription factors are found in all tissues (Fig. 4C). However, there are examples of abundant transcription factors that belong to the tissue-elevated categories, such as (i) grainyhead-like 1 (GRHL1) with enhanced expression in esophagus and skin (squamous epithelia) and selective localization to the uppermost nucleated epidermal keratinocytes (Fig. 2I) and (ii) paired box 6 (PAX6) involved in eye and brain development and differentiation of pancreatic islet cells, with group-enriched expression in brain, pancreas, and stomach, selectively localized to a subset of glandular cells in the stomach mucosa (Fig. 2J) and to islet cells in the pancreas. The tissue-enriched transcription factors identified here (table S8) will enable new insights into the regulatory pattern of the different tissues.

### The druggable proteome

Most pharmaceutical drugs act by targeting proteins and modulating their activity. Target proteins belong to four main families: enzymes, transporters, ion channels, and receptors. The U.S. Food and Drug Administration (FDA) has approved drugs targeting human proteins from 618 genes, according to Drugbank (19), with most drugs acting on signal transduction proteins that convert extracellular signals into intracellular responses. Antibody-based drugs are usually unable to penetrate the plasma membrane, and therefore, they target cell surface proteins, such



as receptors, whereas small-molecule drugs can diffuse into cells and act also on intracellular targets. An analysis of the proteins encoded from the 618 genes shows that 535 proteins are targeted by small chemical molecules, whereas 108 proteins are targeted by biotech drugs (Fig. 4D). The predicted subcellular localization (Fig. 4E) shows that 59% of the targets are predicted membrane proteins and that 16% are secreted, including those with both secreted and membrane-bound isoforms. The genes corresponding to these drug targets were classified according to tissue specificity, and the results (Fig. 4B and table S9) show a bias for tissue-elevated proteins (enriched, group enriched, or enhanced), although as many as 30% of the approved drugs target proteins expressed in all analyzed tissues. One example of a target with enriched expression is cytochrome P450 11B1 (CYP11B1), which is involved in the conversion of progesterone to cortisol in the adrenal gland (Fig. 2K), whereas a ubiquitously expressed protein is the catechol-O-methyltransferase (COMT), which is associated with degradation of neurotransmitters and is important in the metabolism of drugs used in treatment of Parkinson's disease. COMT displays cytoplasmic expression in all analyzed tissues, including lung (Fig. 2L). The ubiquitous expression may have implications for treatments using these proteins as drug targets.

### The cancer proteome

Genes implicated in cancer are often essential for orderly growth, survival, and basic cell functions in normal cells and tissues, whereas overexpression, loss of expression, or expression of a mutated protein contributes to dysfunction and tumor growth. The number of genes implicated in cancer is dependent on definitions; however, 259 genes have been shown to be mutated across 21 tumor types (28); 290 genes have been reported as cancer driver genes across 12 tumor types (29); and 525 genes have been implicated in malignant transformation, according to a catalog of somatic mutations in cancer (COSMIC) (20). Expression analysis based on our transcriptomics data shows that a majority (60%) of these last-mentioned genes (Fig. 4B and table S10) is expressed in all tissues, with only a fraction of genes expressed in a tissue- or group-enriched manner. Examples are the activating transcription factor 1 (ATF1) (Fig. 2M), a protein expressed in all tissues with known translocations in sarcomas, and the forkhead box A1 (FOXA1) (Fig. 2N), a protein with enhanced expression where somatic mutations in a subset of prostate cancers have been reported (30). The lack of tissue specificity for many of these genes is not surprising because many of the corresponding proteins are involved in normal growth regulation and cell cycle control, but it also emphasizes the possible adverse effects of treatment with drugs targeting proteins expressed in all tissues.

### Tissue versus cell lines

Human biology and diseases are often explored using cell lines as model systems. We compared the body-wide expression in human tissues with

expression in cancer cell lines derived from corresponding tissue types. The transcriptomes for 11 cell lines were described earlier (31), whereas the transcriptomes for an additional 36 cell lines were generated as part of this study (see table S11). Genome-wide expression patterns comparing normal tissues with corresponding human cell lines are shown in fig. S6, as exemplified by the liver cancer-derived cell line Hep-G2 (Fig. 4F), and the pancreas cancer-derived cell line Capan-2 (Fig. 4G). Many of the tissue-enriched genes identified in normal tissues are down-regulated or completely “turned off” in the corresponding cell lines, and in contrast, the housekeeping proteins are expressed at the same level in both tissues and corresponding cell lines. These results support earlier studies (32) suggesting that cell lines are “dedifferentiated,” with shared characteristics and lack of tissue-specific features due to down-regulation of tissue-enriched genes. This implies that conclusions from cell line studies should only be conferred on the corresponding tissue with caution.

### The isoform proteome

Protein isoforms endow the structural space of the human proteome with breadth and complexity (33). Isoforms are produced through alternative splicing, posttranslational modifications, proteolytic cleavage, somatic recombination, or genetic variations in protein-coding regions. We explored genes encoding isoforms with different predicted localization (secreted or membrane spanning) (table S12). A large number of these genes ( $n = 366$ ) are displayed together with the fraction of all transcripts (mRNA molecules) in Fig. 5A, with splice variants that yield secreted proteins. Most of the genes (67%) have more than 80% of the transcripts encoding only one of the two localizations across all 32 tissues, but there are some proteins for which the majority of the transcripts encode a secreted form in one tissue, whereas the majority of the transcripts encode a membrane protein in another tissue. As an example, the expression levels for different isoforms of the poorly understood transmembrane emp24 domain-trafficking protein 2 (TMED2) are shown in Fig. 5, B and C. Cardiac muscle has a tissue-specific expression of the secreted form, whereas the membrane-bound form is detected in all other tissue types, although at variable levels. Similarly, the protein Ly6 or neurotoxin 1 (LYNX1) shows a selective expression of the secreted isoform in the esophagus and the skin, whereas the membrane-bound form is found in other tissue types and is most abundantly expressed in the brain and the cardiac muscle (Fig. 5, D and E). The different localizations of the isoforms are consistent with the predicted functions of the different isoforms. In most cases, one of the isoforms dominates across all tissues, which is also consistent with earlier studies (34). These are starting points to explore the relation between tissue-specific expression and function.

### Tissue-based map of human metabolism

Genome-scale metabolic models (GEMs) provide not only the best representation of the metabolic

capabilities of cell and/or tissue types but also quantitative descriptions of the genotype-phenotype relationship (35). Using the RNAseq data, we reconstructed tissue-specific GEMs for 32 different tissues using the generic metabolic model, HMR2 (36), and generated a map of the complete human metabolism. All models were generated such that they can carry out 56 metabolic tasks identified to be present in all human cell types (37). The numbers of the reactions, metabolites, and genes incorporated into each tissue-specific GEM are presented (table S13), and the models are provided in SBML format at the Human Metabolic Atlas portal (38). In order to confirm that none of the models have futile cycles, we ensured that high-energy compounds cannot be generated from low-energy compounds using metabolic tasks including rep phosphorylation of adenosine triphosphate or the generation of a proton gradient over the membranes (table S14).

A total of 6627 reactions, 3040 genes, and 4847 metabolites were present in at least one of the tissue models, and 4912 reactions, 1822 genes, and 3984 metabolites were present in all models. This shows that about 75% of all metabolic reactions in the human body are operating in all key tissues, which clearly illustrates the central role metabolism is playing for basic cellular function. At a gene level, the consensus expression in all tissues is, however, less (i.e., about 60%), which shows that, even though different tissues have the same metabolic reactions, it is different isoforms of the enzymes that are responsible for catalyzing these reactions. Our analysis is the first genome-wide illustration of this wide variation in enzyme usage for catalyzing the same reaction between human tissues.

We found that only 207 of the reactions (Fig. 6A) and 74 of the genes (Fig. 6B) were unique to any of the tissues, and notable differences between the genes (fig. S7) and reactions (fig. S8) based on pairwise comparisons of the various tissues were observed. Between 57 and 632 genes differed in these comparisons of the tissue models, representing 9 to 21% of the genes shared in all models. Bone marrow has the lowest number of genes and reactions, whereas liver has a large number of genes and reactions not present in any other tissue. Many of the metabolic reactions in liver involve specialized lipid metabolism, e.g., de novo synthesis and secretion of bile acids including glycocholate, taurocholate, glycochenodeoxycholate, and taurochenodeoxycholate, but there are also other metabolic functions specific to liver such as ornithine degradation. To further investigate the metabolic capability of each tissue-specific GEM, we defined 256 metabolic tasks (table S15) that are known to occur in humans. The analysis shows that 192 of these metabolic tasks can be performed in all analyzed tissues, whereas the remaining 64 metabolic tasks were performed by some GEMs and clustering of these 64 metabolic tasks is shown in Fig. 6C (see also table S16). The analysis demonstrates liver as the most metabolically active tissue, followed by adipose and skeletal

muscle. For all the remaining tissues, there are variations in the metabolic activities, but with clustering of activities in tissues with similar function and morphology, e.g., stomach, duodenum, and small intestine.

## Discussion

Here, we present a tissue-based map of the human proteome from analyses of 32 tissues and 47 cell lines, with gene expression data on both the RNA and protein level and with supplementary analyses on the protein level for an additional 12 tissues. An interactive resource is presented as part of the Human Protein Atlas portal ([www.proteinatlas.org](http://www.proteinatlas.org)). This allows exploration of the tissue-elevated proteomes in these tissues and organs and analysis of tissue profiles for specific protein classes, including proteins involved in housekeeping functions in the human body, such as cell growth, energy generation, and metabolic pathways; groups of proteins involved in diseases; and proteins targeted by pharmaceutical drugs. Comprehensive lists of genes expressed at elevated levels in these tissues have been compiled, with quantitative expression profiles provided by the deep-sequencing transcriptomics complemented with immunohistochemistry. This provides localization of the proteins in the subcompartments of each tissue and organ down to the single-cell level. To facilitate integration with other biological resources, all data are available for download and through collaborations cross-linked with efforts such as UniProt (5), NextProt (6), ProteomicsDB (9), Metabolic Atlas (38), and the pan-European ELIXIR project (39). An important short-term objective is to facilitate international efforts (5, 7, 8, 40) to explore the “missing proteins,” with the aim to provide a finite list of human protein-coding genes and to generate firm protein evidence and expression characteristics for all of these genes. In addition, the primary data here can be used to expand the analysis of the isoform proteome to better understand the role of this diverse proteome for the functional biology of humans.

## REFERENCES AND NOTES

- P. Flieck et al., Ensembl 2013. *Nucleic Acids Res.* **41** (Database), D48–D55 (2013). doi: [10.1093/nar/gks1236](https://doi.org/10.1093/nar/gks1236); pmid: [23203987](https://pubmed.ncbi.nlm.nih.gov/23203987/)
- 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** (Database), D130–D135 (2012). doi: [10.1093/nar/gkr1079](https://doi.org/10.1093/nar/gkr1079); pmid: [22121212](https://pubmed.ncbi.nlm.nih.gov/22121212/)
- H. Kawaji et al., Update of the FANTOM web resource: From mammalian transcriptional landscape to its dynamic regulation. *Nucleic Acids Res.* **39** (Database), D856–D860 (2011). doi: [10.1093/nar/gkq112](https://doi.org/10.1093/nar/gkq112); pmid: [21075797](https://pubmed.ncbi.nlm.nih.gov/21075797/)
- A. Brazma et al., ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**, 68–71 (2003). doi: [10.1093/nar/gkg091](https://doi.org/10.1093/nar/gkg091); pmid: [12519949](https://pubmed.ncbi.nlm.nih.gov/12519949/)
- M. Magrane, U. Consortium, UniProt Knowledgebase: A hub of integrated protein data. *Database (Oxford)* **2011**, bar009 (2011). doi: [10.1093/database/bar009](https://doi.org/10.1093/database/bar009); pmid: [21447597](https://pubmed.ncbi.nlm.nih.gov/21447597/)
- P. Gaudet et al., neXtProt: Organizing protein knowledge in the context of human proteome projects. *J. Proteome Res.* **12**, 293–298 (2013). doi: [10.1021/pr300830v](https://doi.org/10.1021/pr300830v); pmid: [23205526](https://pubmed.ncbi.nlm.nih.gov/23205526/)
- I. Dunham et al., An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247); pmid: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
- Y. K. Paik et al., The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **30**, 221–223 (2012). doi: [10.1038/nbt2152](https://doi.org/10.1038/nbt2152); pmid: [22398612](https://pubmed.ncbi.nlm.nih.gov/22398612/)
- M. Wilhelm et al., Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014); pmid: [24870543](https://pubmed.ncbi.nlm.nih.gov/24870543/)
- M. S. Kim et al., A draft map of the human proteome. *Nature* **509**, 575–581 (2014). doi: [10.1038/nature13302](https://doi.org/10.1038/nature13302); pmid: [24870542](https://pubmed.ncbi.nlm.nih.gov/24870542/)
- M. Mann, Functional and quantitative proteomics using SILAC. *Nat. Rev. Molec. Cell Biol.* **7**, 952–958 (2006). doi: [10.1038/nrm2067](https://doi.org/10.1038/nrm2067); pmid: [17139335](https://pubmed.ncbi.nlm.nih.gov/17139335/)
- I. Ezkurdia et al., Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* **23**, 5866–5878 (2014). doi: [10.1093/hmg/ddu309](https://doi.org/10.1093/hmg/ddu309); pmid: [24939910](https://pubmed.ncbi.nlm.nih.gov/24939910/)
- V. Lange, P. Picotti, B. Domon, R. Aebersold, Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **4**, 222 (2008). doi: [10.1038/msb.2008.61](https://doi.org/10.1038/msb.2008.61); pmid: [18854821](https://pubmed.ncbi.nlm.nih.gov/18854821/)
- M. Uhlen et al., Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010). doi: [10.1038/nbt1210-1248](https://doi.org/10.1038/nbt1210-1248); pmid: [21139605](https://pubmed.ncbi.nlm.nih.gov/21139605/)
- L. Fagerberg et al., Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014). doi: [10.1074/mcp.M113.035600](https://doi.org/10.1074/mcp.M113.035600); pmid: [24309898](https://pubmed.ncbi.nlm.nih.gov/24309898/)
- C. Kampf et al., The human liver-specific proteome defined by transcriptomics and antibody-based profiling. *FASEB J.* **28**, 2901–2914 (2014). doi: [10.1096/fj.14-250555](https://doi.org/10.1096/fj.14-250555); pmid: [24648543](https://pubmed.ncbi.nlm.nih.gov/24648543/)
- D. Djureinovic et al., The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol. Hum. Reprod.* **20**, 476–488 (2014). doi: [10.1093/molehr/gau018](https://doi.org/10.1093/molehr/gau018); pmid: [24598113](https://pubmed.ncbi.nlm.nih.gov/24598113/)
- G. Gremel et al., The human gastrointestinal tract-specific transcriptome and proteome as defined by RNA sequencing and antibody-based profiling. *J. Gastroenterol.* **50**, 46–57 (2014). doi: [10.1007/s00535-014-0958-7](https://doi.org/10.1007/s00535-014-0958-7); pmid: [24789573](https://pubmed.ncbi.nlm.nih.gov/24789573/)
- V. Law et al., DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res.* **42** (D1), D1091–D1097 (2014). doi: [10.1093/nar/gkt1068](https://doi.org/10.1093/nar/gkt1068); pmid: [24203711](https://pubmed.ncbi.nlm.nih.gov/24203711/)
- COSMIC catalogue of somatic mutations in cancer (2014); <http://cancer.sanger.ac.uk/cancergenome/projects/census>.
- E. Lundberg et al., Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **6**, 450 (2010). doi: [10.1038/msb.2010.106](https://doi.org/10.1038/msb.2010.106); pmid: [21179022](https://pubmed.ncbi.nlm.nih.gov/21179022/)
- Y. Taniguchi et al., Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010). doi: [10.1126/science.1188308](https://doi.org/10.1126/science.1188308); pmid: [20671182](https://pubmed.ncbi.nlm.nih.gov/20671182/)
- 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](https://doi.org/10.1038/nmeth.1701); pmid: [21959131](https://pubmed.ncbi.nlm.nih.gov/21959131/)
- L. Käll, A. Krogh, E. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35** (Web server), W429–W432 (2007). doi: [10.1093/nar/gkm256](https://doi.org/10.1093/nar/gkm256); pmid: [17483518](https://pubmed.ncbi.nlm.nih.gov/17483518/)
- H. Viklund, A. Bernal, M. Skwark, A. Elofsson, SPECTOPUS: A combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **24**, 2928–2929 (2008). doi: [10.1093/bioinformatics/btn550](https://doi.org/10.1093/bioinformatics/btn550); pmid: [18945683](https://pubmed.ncbi.nlm.nih.gov/18945683/)
- D. Ramsköld, E. T. Wang, C. B. Burge, R. Sandberg, An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLOS Comput. Biol.* **5**, e1000598 (2009). doi: [10.1371/journal.pcbi.1000598](https://doi.org/10.1371/journal.pcbi.1000598); pmid: [20011106](https://pubmed.ncbi.nlm.nih.gov/20011106/)
- E. Wingender, T. Schoeps, J. Dönitz, TFClass: An expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* **41** (D1), D165–D170 (2013).
- D. Tamborero et al., Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650 (2013). doi: [10.1038/srep02650](https://doi.org/10.1038/srep02650); pmid: [24084849](https://pubmed.ncbi.nlm.nih.gov/24084849/)
- D. M. Muzny et al., Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012). doi: [10.1038/nature11252](https://doi.org/10.1038/nature11252); pmid: [22810696](https://pubmed.ncbi.nlm.nih.gov/22810696/)
- C. S. Grasso et al., The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012). doi: [10.1038/nature11125](https://doi.org/10.1038/nature11125); pmid: [22722839](https://pubmed.ncbi.nlm.nih.gov/22722839/)
- F. Danielson 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](https://doi.org/10.1021/pr3009308); pmid: [23227862](https://pubmed.ncbi.nlm.nih.gov/23227862/)
- M. Schnabel et al., Dedifferentiation-associated changes in morphology and gene expression in primary human articular chondrocytes in cell culture. *Osteoarthritis Cartilage* **10**, 62–70 (2002). doi: [10.1053/joca.2001.0482](https://doi.org/10.1053/joca.2001.0482); pmid: [11795984](https://pubmed.ncbi.nlm.nih.gov/11795984/)
- E. Birney et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007). doi: [10.1038/nature05874](https://doi.org/10.1038/nature05874); pmid: [17571346](https://pubmed.ncbi.nlm.nih.gov/17571346/)
- M. González-Porta, A. Frankish, J. Rung, J. Harrow, A. Brazma, Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol.* **14**, R70 (2013). doi: [10.1186/gb-2013-14-7-r70](https://doi.org/10.1186/gb-2013-14-7-r70); pmid: [23815980](https://pubmed.ncbi.nlm.nih.gov/23815980/)
- A. Mardinoglu, J. Nielsen, Systems medicine and metabolic modelling. *J. Intern. Med.* **271**, 142–154 (2012). doi: [10.1111/j.1365-2796.2011.02493.x](https://doi.org/10.1111/j.1365-2796.2011.02493.x); pmid: [22142312](https://pubmed.ncbi.nlm.nih.gov/22142312/)
- A. Mardinoglu et al., Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014). doi: [10.1038/ncomms4083](https://doi.org/10.1038/ncomms4083); pmid: [24419221](https://pubmed.ncbi.nlm.nih.gov/24419221/)
- R. Agren et al., Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721 (2014). doi: [10.1002/msb.145122](https://doi.org/10.1002/msb.145122); pmid: [24646661](https://pubmed.ncbi.nlm.nih.gov/24646661/)
- Human Metabolic Atlas (2014); [www.metabolicatlas.org/](http://www.metabolicatlas.org/).
- 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](https://doi.org/10.1016/j.tibtech.2012.02.002); pmid: [22417641](https://pubmed.ncbi.nlm.nih.gov/22417641/)
- J. F. Rual et al., Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).

## ACKNOWLEDGMENTS

We acknowledge the entire staff of the Human Protein Atlas program; the Science for Life Laboratory; and the pathology team in Mumbai, India, for valuable contributions. We thank the Department of Pathology at the Uppsala Akademiska Hospital, Uppsala, Sweden, and Uppsala Biobank for kindly providing clinical diagnostics and specimens used in this study. We also acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure (NGI), and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. Funding was provided by the Knut and Alice Wallenberg Foundation. The authors declare that they have no conflict of interest. Correspondence and requests for materials should be addressed to M.U. The mRNA levels of all genes in each tissue sample ( $n = 122$ ) are available in Table S18. The supplementary Excel tables are available in the supplementary material and at [www.proteinatlas.org/about/publicationdata](http://www.proteinatlas.org/about/publicationdata). The raw sequencing data are available at ArrayExpress ([www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2836/](http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2836/)) and BioProject (NIH) ([www.ncbi.nlm.nih.gov/bioproject/PRJNA183192](http://www.ncbi.nlm.nih.gov/bioproject/PRJNA183192)). All Protein Atlas (protein) data are available in structured XML format and can be downloaded from [www.proteinatlas.org/about/download](http://www.proteinatlas.org/about/download).

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/347/6220/1260419/suppl/DC1](http://www.sciencemag.org/content/347/6220/1260419/suppl/DC1)  
Materials and Methods  
Figs. S1 to S8  
Tables S1 to S18  
References (41–61)

25 August 2014; accepted 5 December 2014  
10.1126/science.1260419