www.sciencemag.org/content/347/6220/1260419/suppl/DC1

Supplementary Materials for

Tissue-based map of the human proteome

Mathias Uhlén,* 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

*Corresponding author. E-mail: mathias.uhlen@scilifelab.se

Published 23 January 2015, *Science* **347**, 1260419 (2014)
DOI: 10.1126/science.1260419

This PDF file includes

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/346/6220/1260419/suppl/DC1)

Table S1: Expression profiles of all genes in 32 tissues including specificity classification, predictions of secreted and transmembrane regions, and protein evidence.

Table S2: Tissue-enriched genes with full expression profiles for 32 tissues.

Table S3: Group-enriched genes with full expression profiles for 32 tissues.

Table S4: Tissue-enhanced genes with full expression profiles for 32 tissues.

Table S5: "Missing" proteins, not detected at the protein-level or at the RNA-level.

Table S6: Top 20 enriched GO terms for enriched genes from each of the 13 representative tissues/tissue groups, showing enrichment factors and p-values corrected for multiple-testing.

Table S7: Prediction of secreted and transmembrane proteins for all transcripts.

Table S8: Transcription factors with full expression profiles for 32 tissues, with specificity classification.

Table S9: Druggable proteins with full expression profiles for 32 tissues, with specificity classification and drug type annotation.

Table S10: Genes involved in cancer with full expression profiles for 32 tissues, with specificity classification.

Table S11: FPKM values for all protein-coding genes in 44 cell lines.

Table S12: List of 852 genes encoding both membrane and secreted isoforms.

Table S13: The content of the tissue-specific GEMs reconstructed based on RNA-seq data.

Table S14: List of the metabolic tasks that are used for making sure that the models can have no futile cycles.

Table S15: List of 256 metabolic tasks that are known to occur in human. This complete list of metabolic tasks was tested for each tissue-specific GEM.

Table S16: Results from simulations to test the performance of the defined tasks of each tissue-specific GEM.

Table S17: List of the metabolic tasks that are known to occur in all human cell/tissue types. This complete list of metabolic tasks was used as input to the tINIT algorithm for the reconstruction of tissue-specific GEMs.

Table S18: FPKM values for all 122 tissue samples (including all individuals) for the 32 tissues.

Supplementary Materials

Materials and Methods

Sample preparation

Human tissue samples used for protein and mRNA expression analyses were collected and handled in accordance with Swedish laws and regulation and obtained from the Department of Pathology, Uppsala University Hospital, Uppsala, Sweden as part of the sample collection governed by the Uppsala Biobank (http://www.uppsalabiobank.uu.se/en/). All human tissue samples used in the present study were anonymized in accordance with approval and advisory report from the Uppsala Ethical Review Board (Reference # 2002-577, 2005-338 and 2007-159 (protein) and # 2011-473 (RNA)). Included cell lines were derived from DSMZ, ATCC or academic research groups (kindly provided by cell line founders). Information regarding sex and age of the donor, tissue origin and source can be found at http://www.proteinatlas.org/about/cellines.

Protein profiling (tissue microarrays and immunohistochemistry)

Generation of tissue microarrays (TMAs), immunohistochemical staining and slide scanning were performed as previously described (41). In brief, formalin fixed and paraffin embedded (FFPE) tissue samples were collected from the pathology archives based on hematoxylin-eosin (HE) stained tissue sections showing representative normal histology for each tissue type. Representative cores (1 mm diameter) were sampled from the FFPE blocks and assembled into two TMAs, containing in total normal tissue samples from 144 individuals. Duplicate samples from 46 cell lines, 10 leukemia blood cell samples and two samples of normal PBMC were used to create cell microarrays (CMA). All cells were fixed in 4% paraformaldehyde and dispersed in agarose prior to paraffin embedding and immunohistochemical staining. TMA and CMA blocks were cut in 4 μm thick sections using waterfall microtomes (Microm HM 355S, Thermo Fisher Scientific, Freemont, CA, USA), dried in RT overnight and baked in 50°C for 12-24 hours prior to immunohistochemical staining. Automated immunohistochemistry was performed using Autostainer 480® instruments (Lab Vision, Freemont, CA, USA), followed by slide scanning using Scanscope XT (Aperio, Vista, CA, USA). The corresponding high resolution images of immunohistochemically stained TMA sections were evaluated and annotated by certified pathologists (Lab SurgPath, Mumbai, India). The CMA images were scored using an automated image analysis system.

Transcript profiling (RNA-seq)

Tissues samples were embedded in Optimal Cutting Temperature (O.C.T.) compound and stored at –80°C. An HE stained frozen section (4 μm) was prepared from each sample using a cryostat and the CryoJane® Tape-Transfer System (Instrumedics, St. Louis, MO, USA). Each slide was

examined by a pathologist to ensure sampling of representative normal tissue. Three sections (10 µm) were cut from each frozen tissue block and collected into a tube for subsequent RNA extraction. The tissue was homogenized mechanically using a 3 mm steel grinding ball (VWR). Total RNA was extracted from cell lines and tissue samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The extracted RNA samples were analyzed using either an Experion automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA) with the standard-sensitivity RNA chip or an Agilent 2100 Bioanalyzer system (Agilent Biotechnologies, Palo Alto, USA) with the RNA 6000 Nano Labchip Kit. Only samples of high-quality RNA (RNA Integrity Number ≥7.5) were used in the following mRNA sample preparation for sequencing.

Analysis of data

Sequencing of a total number of 122 samples from 32 tissues and organs was performed as previously described, using Illumina Hiseq2000 and Hiseq2500, and the standard Illumina RNA-seq protocol (15). In brief, processed reads were mapped to the human genome (GRCh37) using Tophat v2.0.8b (42). To obtain quantification scores for all human genes and transcripts across all 122 samples, FPKM (fragments per kilobase of exon model per million mapped reads) values were calculated using Cufflinks v2.1.1 (43). The gene models from Ensembl build 75 (44) were used in Cufflinks and the total number of protein-coding genes was 20,356. Since 12 genes were missing from the Cufflinks output, the final number of genes with RNA-seq data was 20,344. The average FPKM value of all individual samples for each tissue was used to estimate the gene expression level. A cutoff value of 1 FPKM was used as a limit for detection across all tissues.

RNA-based classification of genes

Each of the 20,344 genes with mapped RNA-seq data was classified into one of six categories based on the FPKM levels in 32 tissues: (1) "Not detected": FPKM<1 in all tissues; (2) "Tissue enriched" – at least a 5-fold higher FPKM level in one tissue compared to all other tissues; (3) "Group enriched" – 5-fold higher average FPKM value in a group of 2-7 tissues compared to all other tissues; (4) "Expressed in all tissues" – detected in all 32 tissues with FPKM>1; (5) "Tissue enhanced" – at least a 5-fold higher FPKM level in one tissue compared to the average value of all 32 tissues; (6) "Mixed" – the remaining genes detected in 1-31 tissues with FPKM>1 and in none of the above categories.

Gene ontology analysis

Enriched gene ontology terms in sets of enriched genes were determined using the DAVID Bioinformatics Resource v 6.7 (45).

Protein and RNA-evidence

Evidence for the existence of protein (Protein evidence) was taken from four different sources, the Human Protein Atlas, UniProt and two recent mass spectrometry based proteogenomics

studies (9, 10). Genes were considered to have evidence of protein presence from the Human Protein Atlas only if the antibody reliability was annotated as “medium” or “high”, from Uniprot if the protein had the “Evidence at protein level” attribute and from proteogenomics if either of the two studies mapped at least two non-discriminating peptides to the protein. The remaining genes, lacking protein evidence from any source, was divided into genes with RNA-evidence, i.e. genes with “Evidence at transcript level” according to UniProt or an FPKM > 1 in at least one of our 32 tissues (RNA evidence), and genes lacking any evidence (No evidence).

Prediction methods for membrane protein topology

The membrane protein analysis was performed as previously described (46) using seven methods for membrane protein topology based on different underlying algorithms, such as hidden Markov models (HMMs), neural networks and support vector machines (SVMs): MEMSAT3 (47), MEMSAT-SVM (48), Phobius version 1.01 (49), SCAMPI multi-sequence-version (50), SPOCTOPUS (25), TMHMM (51) and THUMBUP (52). The resulting predicted transmembrane (TM) regions from these selected methods were used to construct a majority-decision based method (MDM) (46) to discriminate transmembrane proteins from soluble proteins and determine the potential TM positions in the protein sequence. Briefly, the proteins predicted as transmembrane by the MDM consists of all proteins where there is at least one predicted transmembrane region overlapping by four out of the seven methods.

Prediction of the human secretome

For the secretome analysis, three methods for the prediction of signal peptides (SP) were used: SignalP4.0 (23), Phobius (24) and SPOCTOPUS (25). SignalP4.0 is an update of the widely used SignalP3.0 method (53), and is focused on the prediction of signal peptides, whereas the other two algorithms combine the prediction of transmembrane (TM) regions with the prediction of SPs. All three methods, however, are trained to discriminate between the two types of protein features and are based on Neural Networks and Hidden Markov models. Stand-alone applications were acquired for each of the methods and run on a computer cluster as previously described for SPOCTOPUS and Phobius (46). For SignalP4.0, all predicted amino acid cleavage positions <11 were considered false positives and removed from the results in accordance to instructions from the authors.

Similarly to the construction of the MDM for membrane protein prediction, a majority-decision approach was used to predict the human secretome using a new method called MDSEC. All human proteins for which at least two out of the three methods predicted a signal peptide were considered potentially secreted by MDSEC. However, as certain types of membrane proteins also contain an N-terminal signal peptide, the proteins with both a predicted SP and at least one TM region predicted by the MDM were classified as membrane-spanning and removed from the set of secreted proteins. Thus, the resulting list of predicted secreted proteins consists of all protein with a predicted signal peptide by two out of three methods and without a predicted MDM region. As this method requires a predicted signal peptide, proteins secreted using the

unconventional secreted pathway (54) are not taken into account into this set of predicted secreted proteins.

Classification of the human proteome

The combined results from the prediction of the human membrane proteome and the secretome were used to perform an overall classification of each human protein into one of three classes: secreted, membrane-spanning or soluble (not containing any SP or TM regions). Subsequently, the distribution of proteins from these three classes across all human protein-coding genes could be analyzed. All 20,356 genes were categorized based on the classifications of each protein isoform of the gene, which resulted in seven possible combinations: soluble, soluble/secreted, soluble/membrane, secreted, membrane, secreted/membrane and soluble/secreted/membrane. To reduce the complexity, four major categories were constructed from these by keeping the soluble category but combining the soluble/secreted and the secreted groups ("secreted"), combining the soluble/membrane and the membrane groups ("membrane") and by combining the secreted/membrane and soluble/secreted/membrane groups ("membrane and secreted isoforms").

Selection of protein classes

A list of mitochondrial proteins, containing 13 genes encoded by the mitochondria genome as well as a large number of nuclear-encoded genes (n=766), was obtained from the MitoProteome database (55). In addition, a list of ribosomal proteins (n=152) as well as a list of cell differentiation (CD) molecules (56), commonly described as CD markers, (n=379) were obtained from UniProt (57). All resulting UniProt accessions for the three lists were mapped to Ensembl gene identifiers to allow for an analysis of the transcript abundance of these genes and categorization into the membrane-spanning and secreted classes described above.

Reconstruction of tissue-specific genome-scale metabolic models (GEMs)

Tissue-specific models were reconstructed based on the RNA-seq data and a recently developed task-driven model reconstruction (tINIT) algorithm (37). The tINIT algorithm employs defined metabolic tasks for imposing constraints on the functionality of the reconstructed models. In this context, 56 metabolic functions were defined that are common for all human cell/tissue types and categorized into tasks as energy and redox, internal conversions, substrate utilization and biosynthesis of products (Table S17). These metabolic tasks were used as an input to the tINIT algorithm for the reconstruction of tissue-specific models that can perform all the defined metabolic tasks and at the same time are consistent with the RNA-seq data.

Figure S1. Pair-wise correlation between RNAseq samples. Y-axis displays all 122 samples from a total of 32 tissues. X-axis displays the pair-wise Spearman correlation between each sample compared to all 121 other tissue samples. Each circle represents a tissue sample and the color represents the type of tissue. The correlations with samples from the same tissue type are shown in red color.

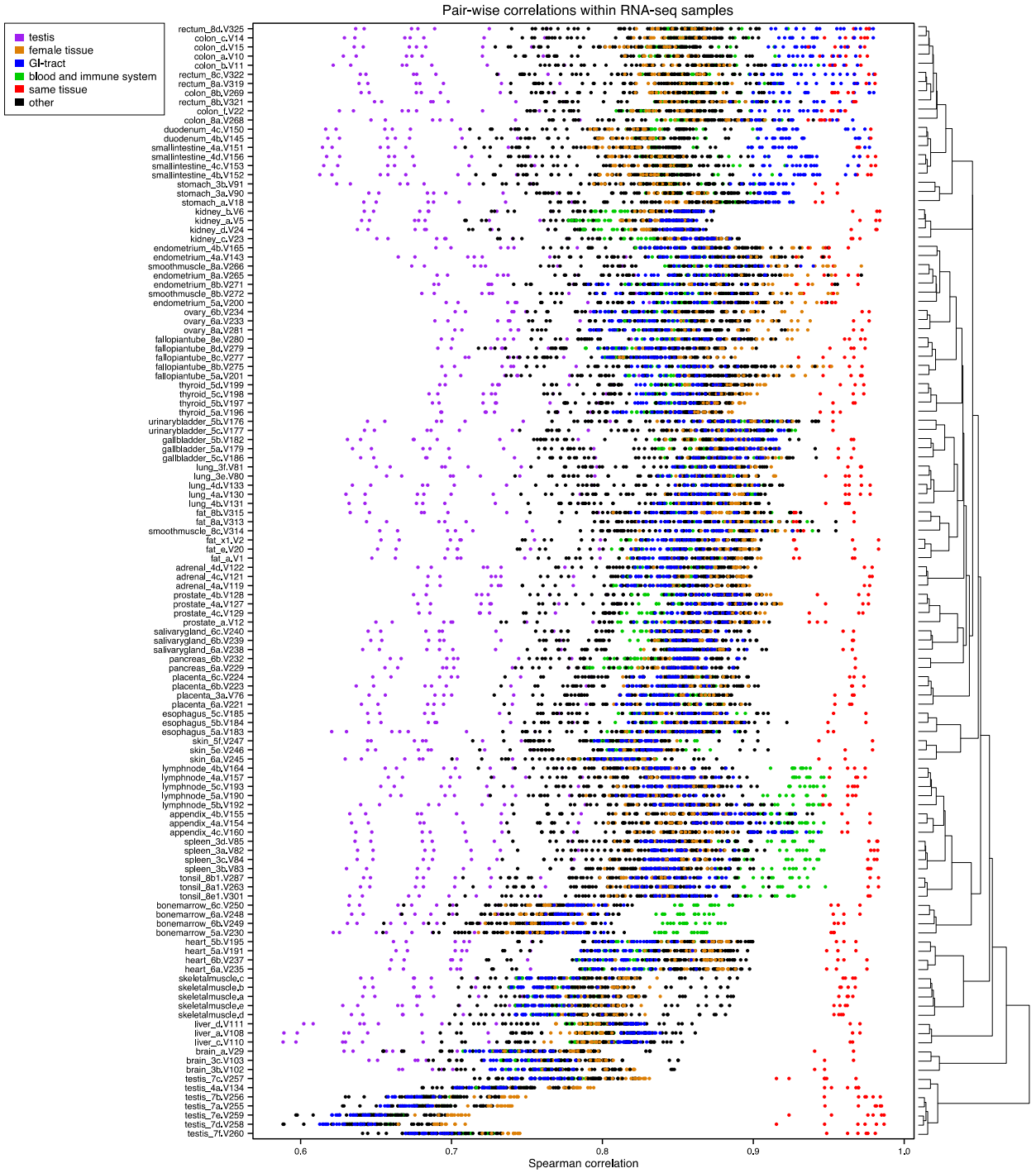


Figure S2. (A) Principal component analysis (PCA) of all 32 tissues. The first two components are shown and each tissue is represented by a circle colored according to tissue type. **(B)** Categorization of the 677 "missing" (with no experimental evidence) protein-coding genes annotated in Ensembl version 75, based on data from Ensembl version 76, CCDS and a publication by Ezkurdia et al (12). In Ensembl version 76, 51% of the genes (n=343) are still annotated as protein-coding (PC) (in dark blue and blue), while only 73 of these are also recognized by CCDS (dark blue). 8% of the genes (in green) are annotated as non-coding or as different types of pseudo-genes, and 41% (in purple) have been removed. In the publication by Ezkurdia et al where they investigated the correlation between detection of genes in proteomics experiments and gene family age and cross-species conservation, a set of possibly non-coding genes was described and their distribution is represented by the dashed pattern in the pie chart. This data indicates that the fraction of "true" protein-coding genes among these "missing" genes is around 30% (dark blue and blue parts without pattern), and that at least 15% (green and purple parts with pattern) does not represent protein-coding genes.

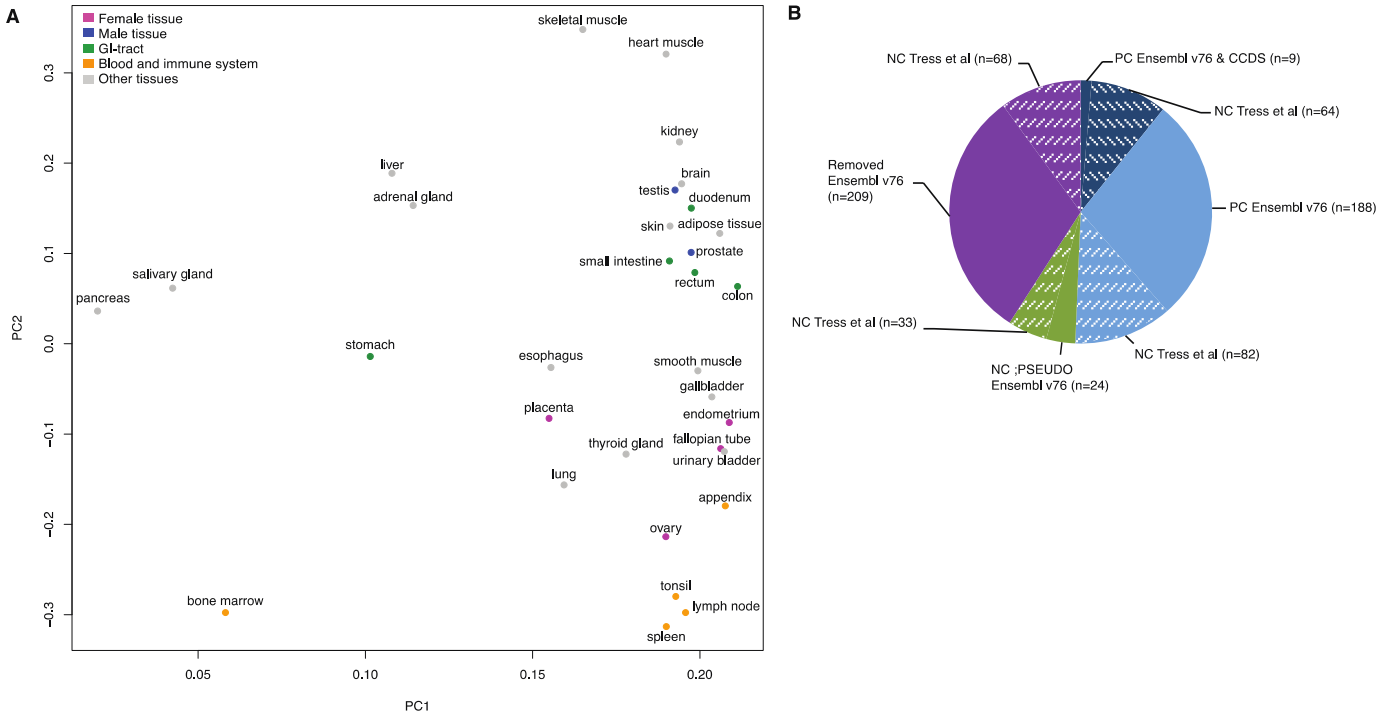
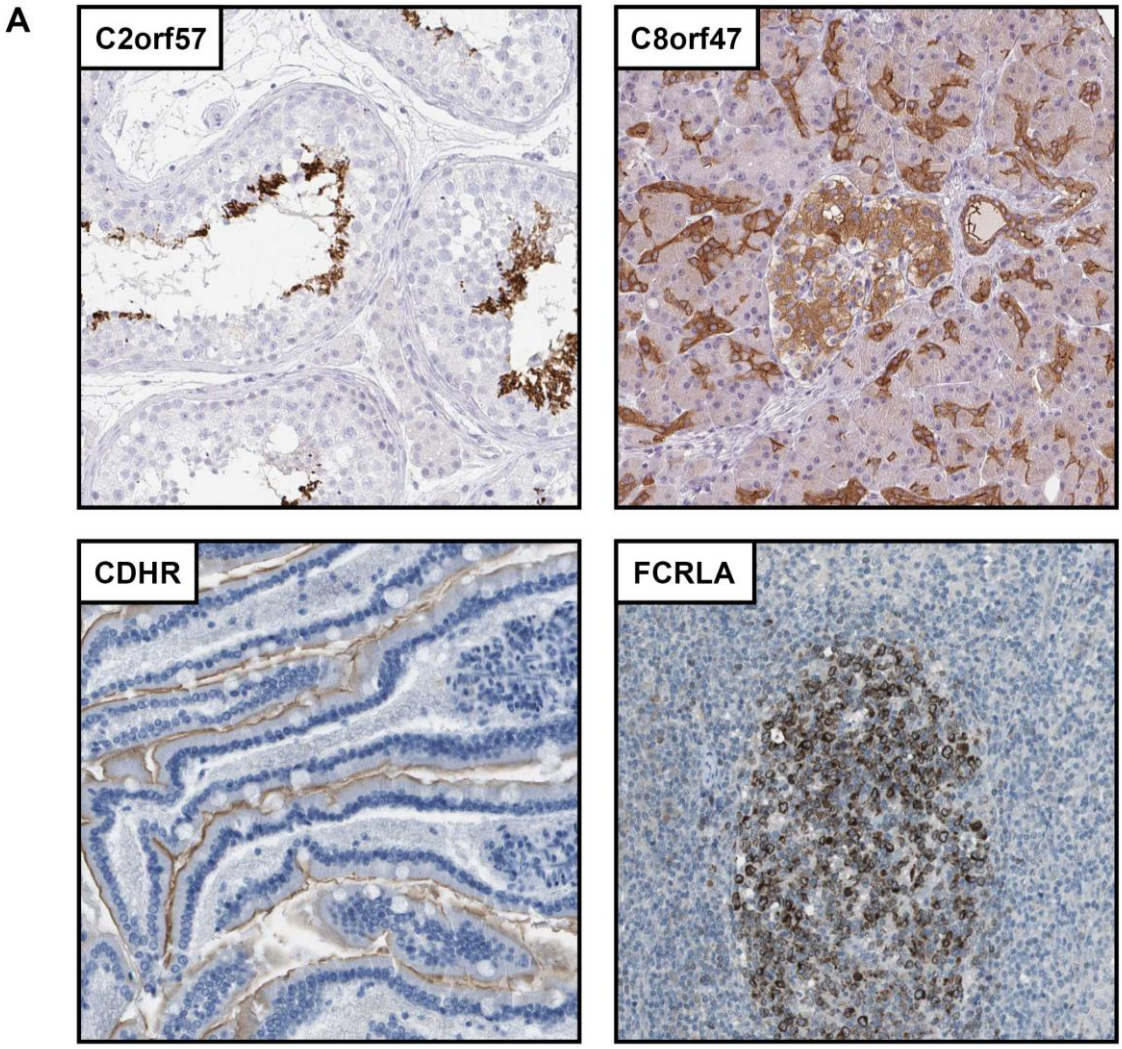
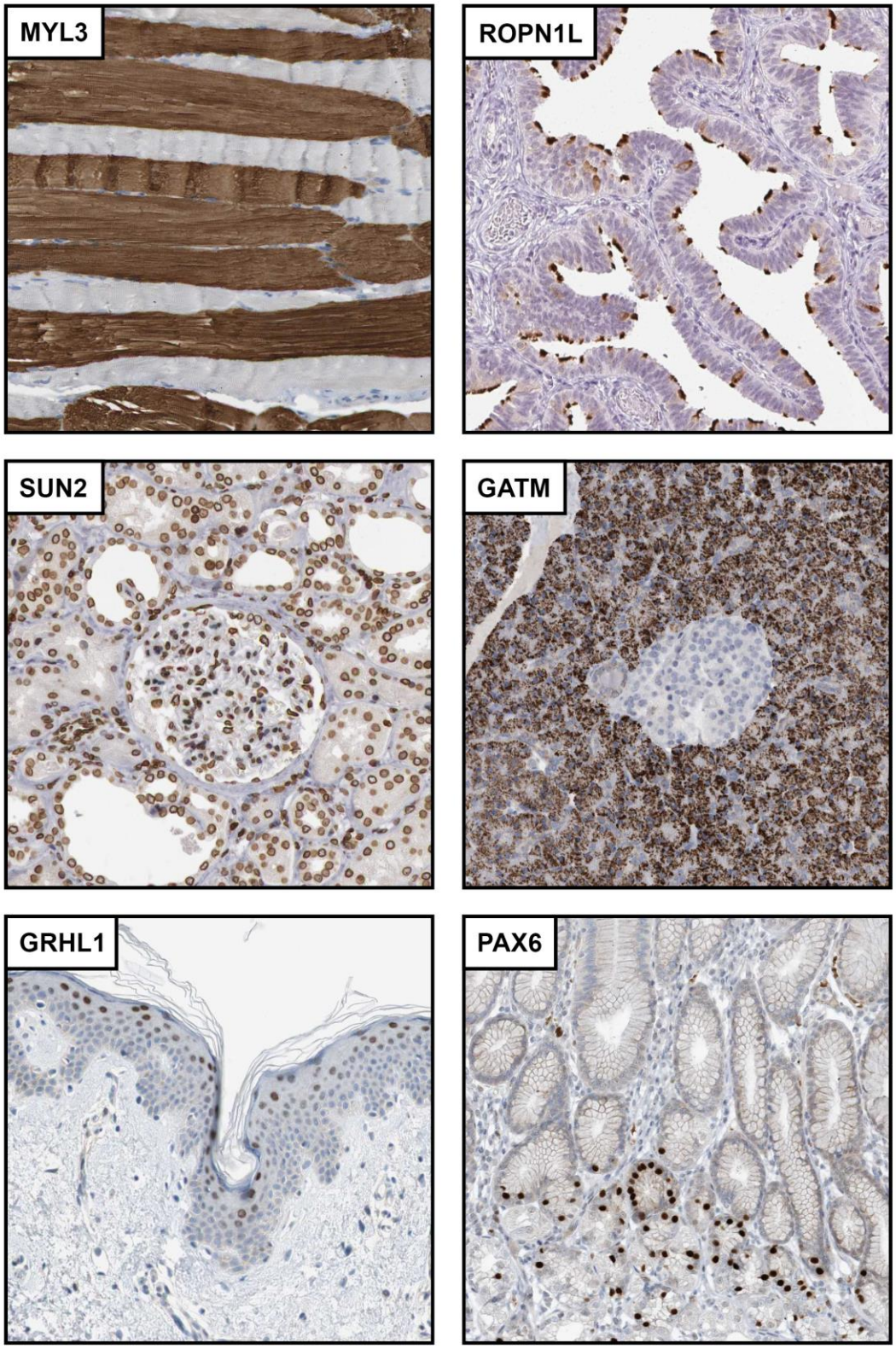


Figure S3. Tissue expression and localization for a selection of human proteins, displaying larger images corresponding to Figure 2A-N. The figure includes: (**A**) C2orf57, C8orf47, CDHR2, FCRLA; (**B**), MYL3, ROPN1L, SUN2, GATM, GRHL1, PAX6; and (**C**) CYP11B1, COMT, ATF1 and FOXA1. See legend to Figure 2 for details.



B



C

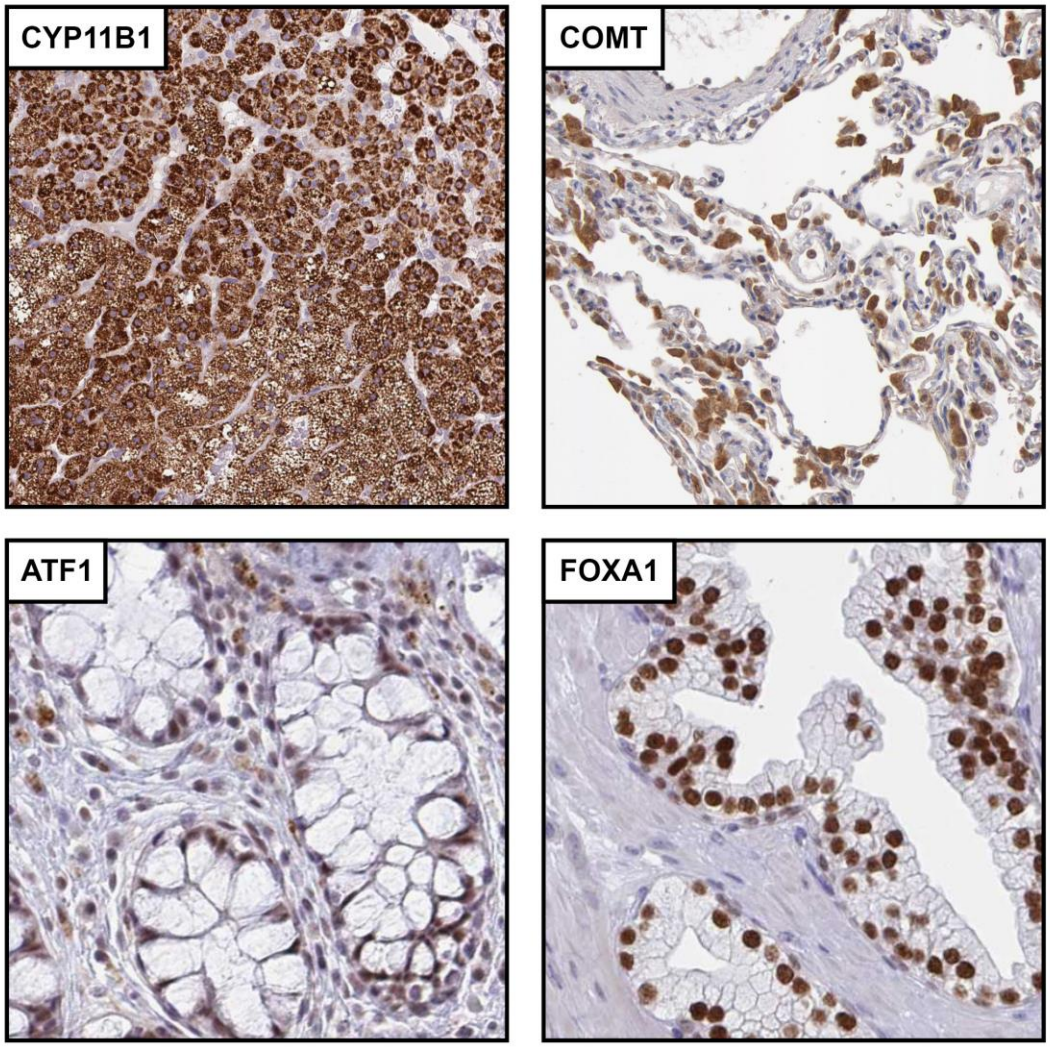


Figure S4. Network plot showing the relationship between tissue enriched and group enriched genes in their respective expressed tissues. Red nodes represent the number of tissue enriched genes in each tissue. Orange nodes show the number of genes that are group enriched in combinations of up to five different tissues and with a minimum of three genes in each combination. This results in a total number of grey tissue nodes of 31 as smooth muscle has no tissue enriched genes and no combinations with more than three group enriched genes. The size of the orange and red nodes is related to the number of genes enriched in a particular combination of tissues.

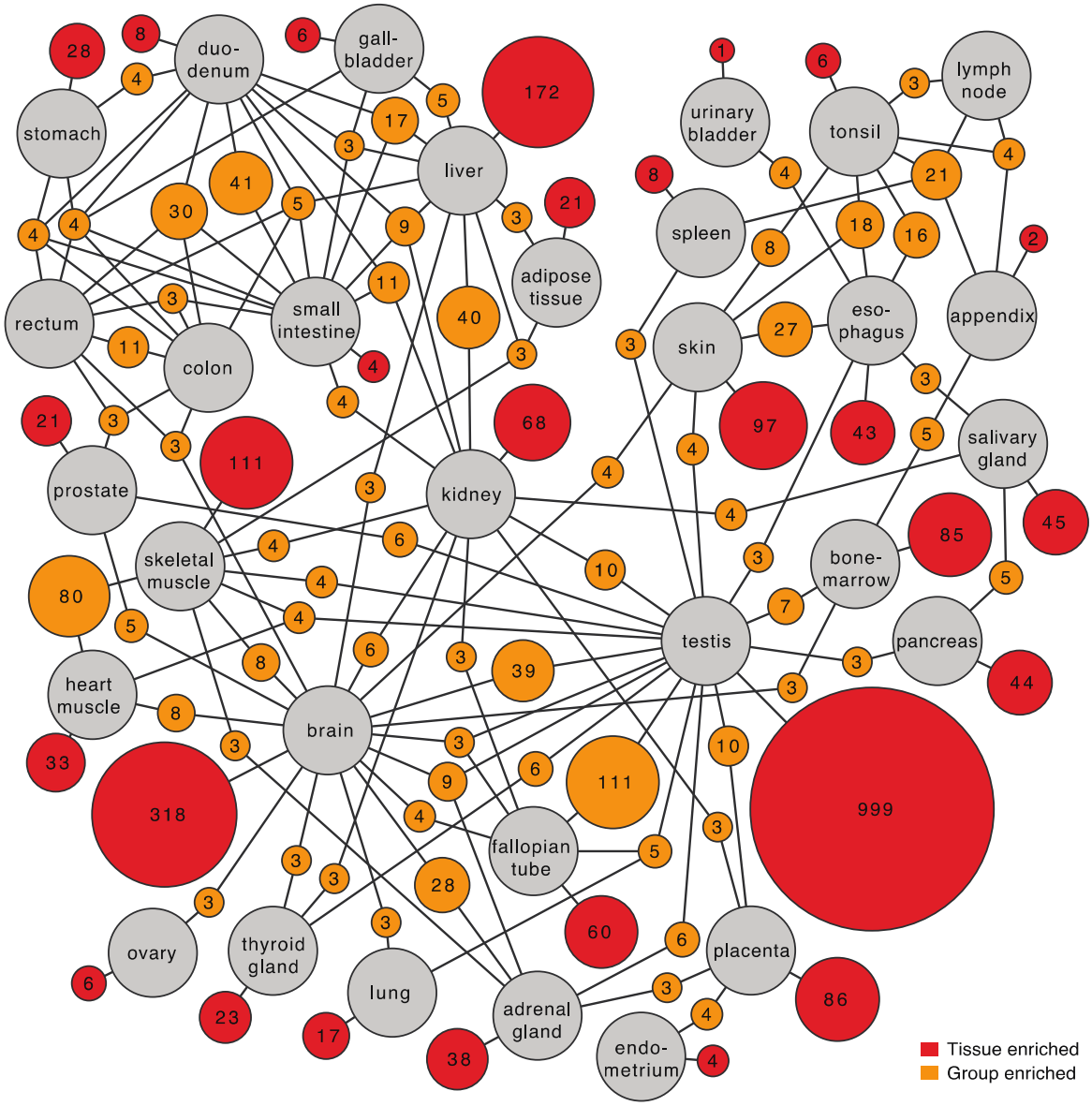


Figure S5. Pie charts showing the fraction of the transcriptome (as fraction of the total sum of FPKM values) in the protein expression classes for the intracellular soluble, secreted, and membrane-spanning proteins, as well as the class with both membrane and secreted isoforms.

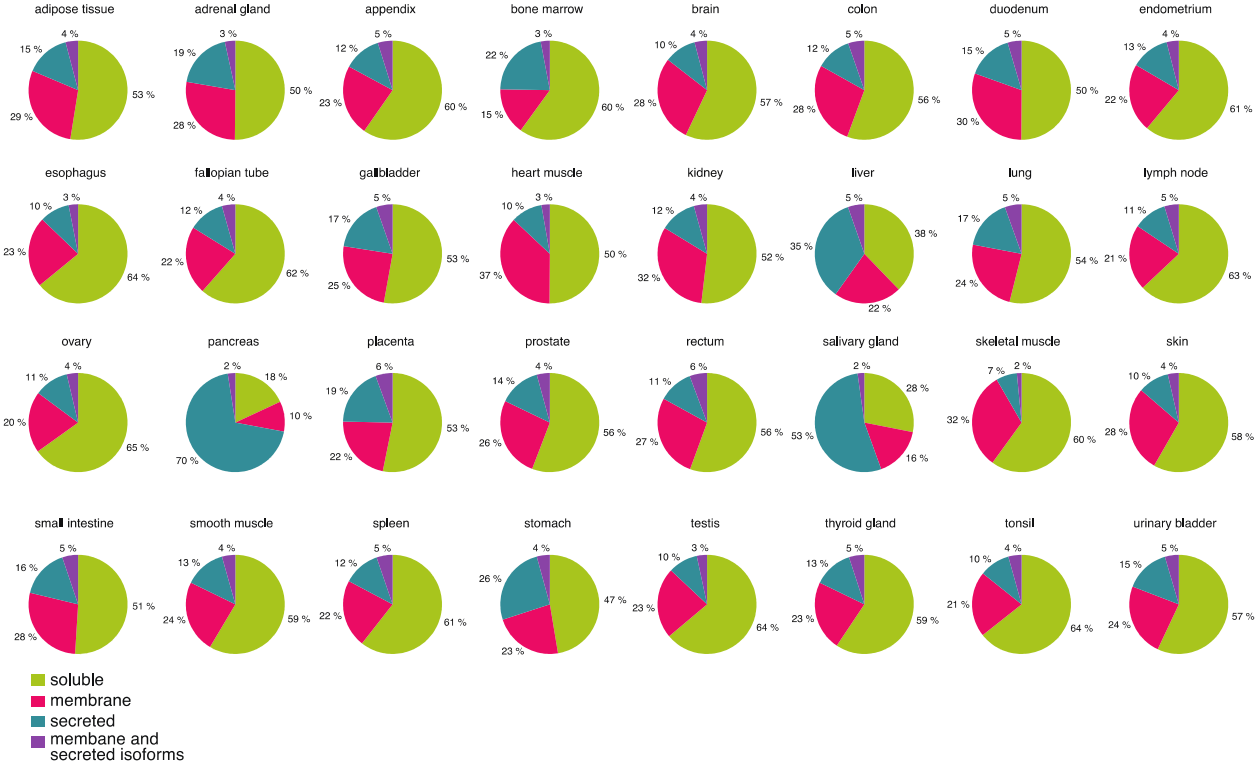


Figure S6. Pairwise comparison of gene expression in normal tissues and cell lines established from transformed and/or immortalised progenitor cells within those tissues, with each gene color-coded based on the transcript expression category for each tissue.

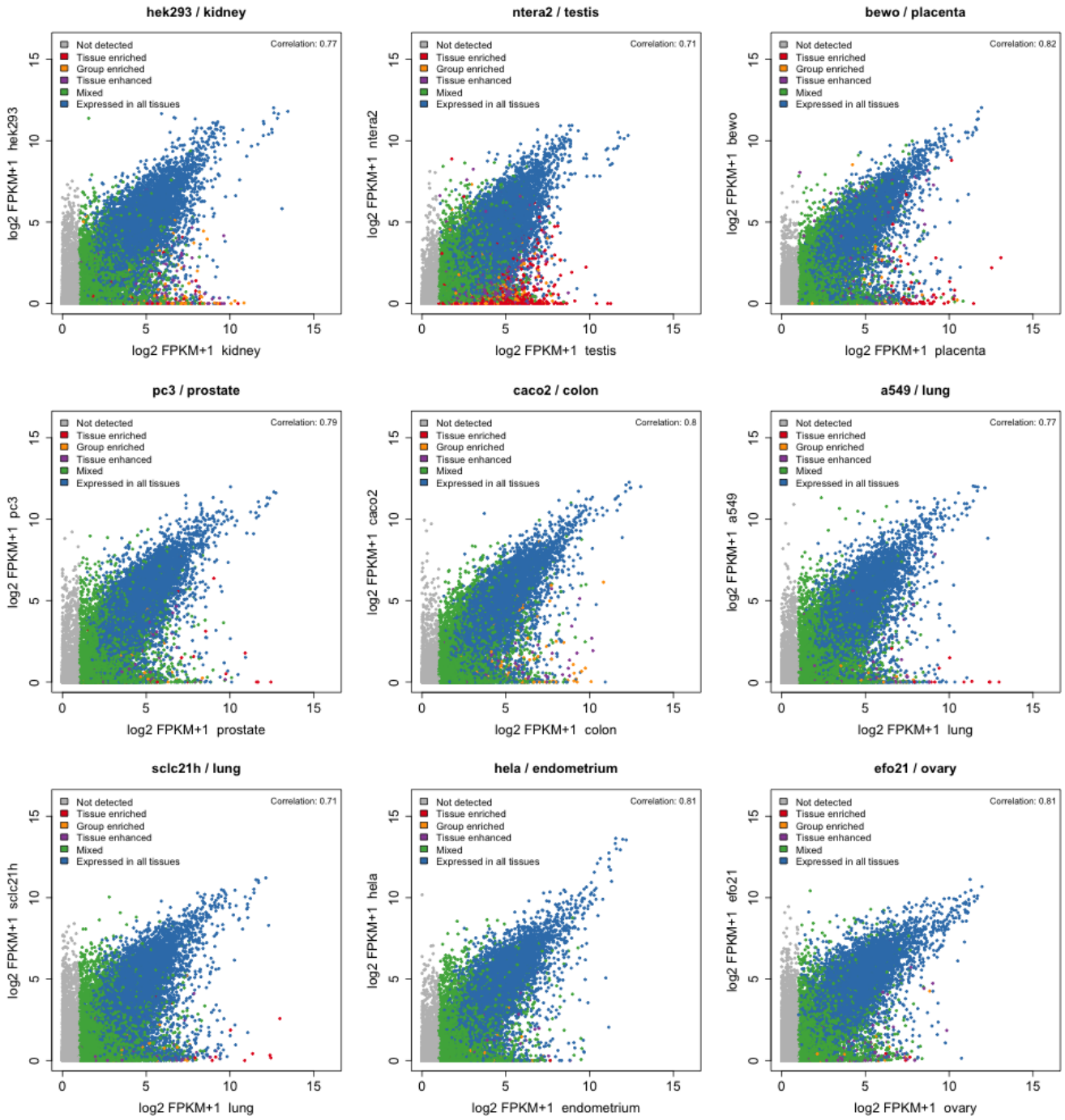


Figure S7. Pairwise comparison of the genes incorporated in each tissue-specific GEM. The total number of common genes between two tissues is shown in the bottom triangle whereas the upper triangle shows the total number of different genes between two tissues.

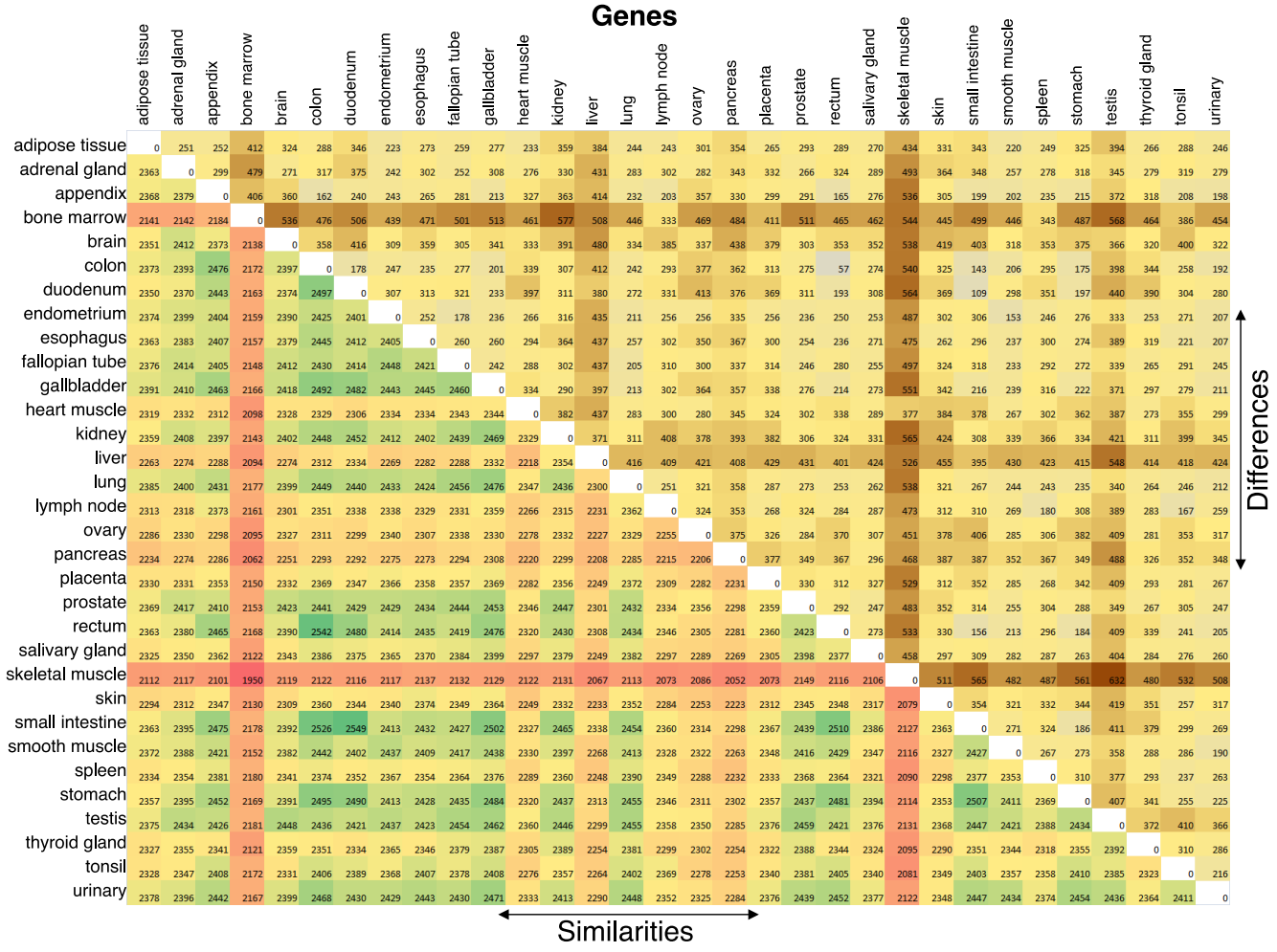
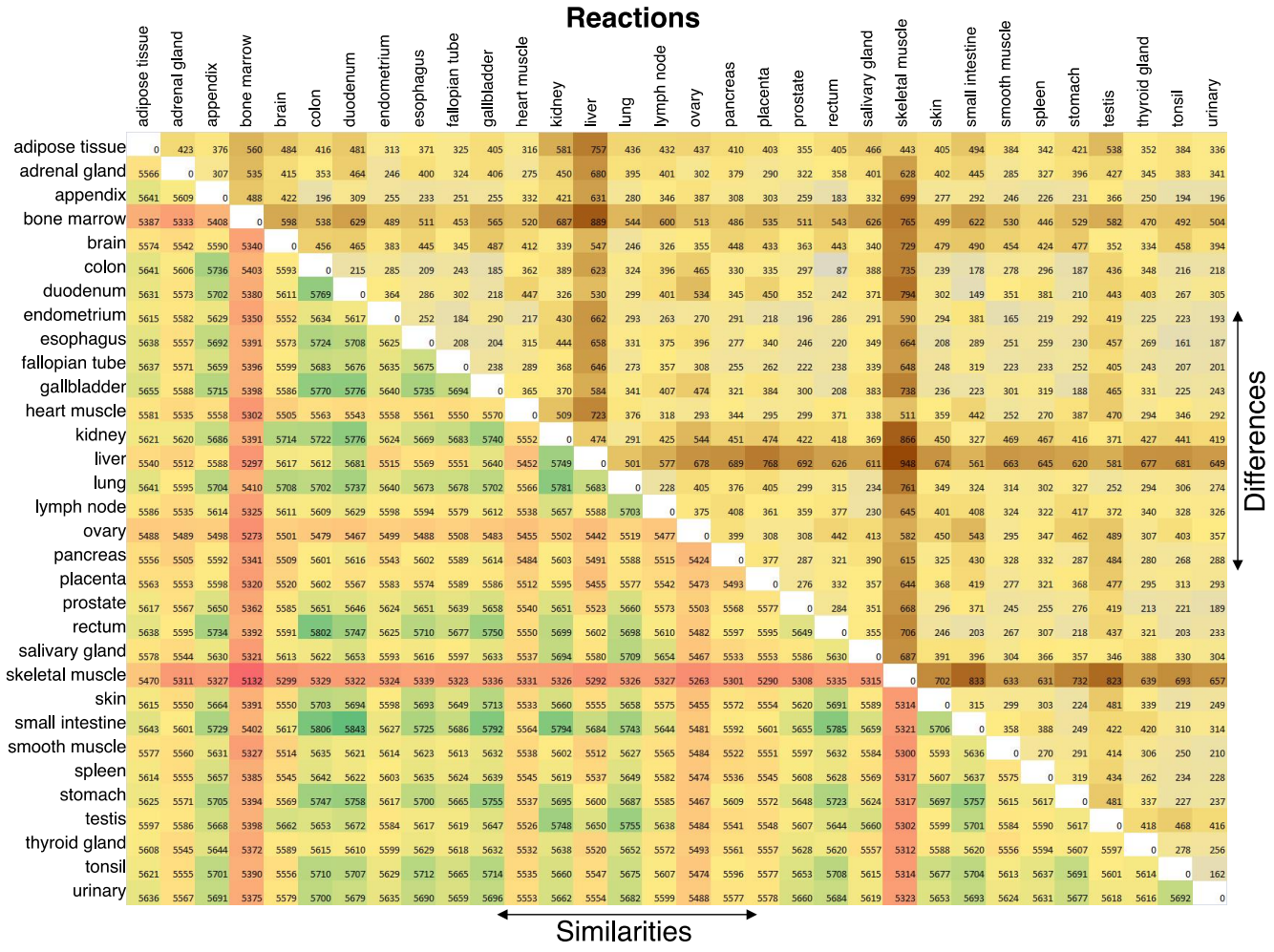


Figure S8. Pairwise comparison of the reactions incorporated in each tissue-specific GEM. The total number of common reactions between two tissues is shown in the bottom triangle whereas the upper triangle shows the total number of different reactions between two tissues.



Supplementary table legends

All tables are available as separate Excel files.

Table S1: Expression profiles of all genes in 32 tissues including specificity classification, predictions of secreted and transmembrane regions, and protein evidence.

Table S2: Table of tissue-enriched genes with full expression profiles for 32 tissues.

Table S3: Table of group-enriched genes with full expression profiles for 32 tissues.

Table S4: Table of tissue-enhanced genes with full expression profiles for 32 tissues.

Table S5: Table of "missing" proteins, not detected at the protein-level or at the RNA-level.

Table S6: Table of top 20 enriched GO terms for enriched genes from each of the 13 representative tissues/tissue groups, showing enrichment factors and p-values corrected for multiple-testing.

Table S7: Prediction of secreted and transmembrane proteins for all transcripts.

Table S8: Table of transcription factors with full expression profiles for 32 tissues, with specificity classification.

Table S9: Table of druggable proteins with full expression profiles for 32 tissues, with specificity classification and drug type annotation.

Table S10: Table of genes involved in cancer with full expression profiles for 32 tissues, with specificity classification.

Table S11: Table of FPKM values for all protein-coding genes in 44 cell lines.

Table S12: List of 852 genes encoding both membrane and secreted isoforms.

Table S13: The content of the tissue-specific GEMs reconstructed based on RNA-seq data.

Table S14: List of the metabolic tasks that are used for making sure that the models can have no futile cycles.

Table S15: List of 256 metabolic tasks that are known to occur in human. This complete list of metabolic tasks was tested for each tissue-specific GEM.

Table S16: Results from simulations to test the performance of the defined tasks of each tissue-specific GEM.

Table S17: List of the metabolic tasks that are known to occur in all human cell/tissue types. This complete list of metabolic tasks was used as input to the tINIT algorithm for the reconstruction of tissue-specific GEMs.

Table S18: Full table of FPKM values for all 122 tissue samples (including all individuals) for the 32 tissues.

REFERENCES AND NOTES

1. P. Flicek, I. Ahmed, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, C. García-Girón, L. Gordon, T. Hourlier, S. Hunt, T. Juettemann, A. K. Kähäri, S. Keenan, M. Komorowska, E. Kulesha, I. Longden, T. Maurel, W. M. McLaren, M. Muffato, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, G. R. Ritchie, M. Ruffier, M. Schuster, D. Sheppard, D. Sobral, K. Taylor, A. Thormann, S. Trevanion, S. White, S. P. Wilder, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, J. Harrow, J. Herrero, T. J. Hubbard, N. Johnson, R. Kinsella, A. Parker, G. Spudich, A. Yates, A. Zadissa, S. M. Searle, Ensembl 2013. *Nucleic Acids Res.* **41**(Database), D48–D55 (2013). Medline doi:10.1093/nar/gks1236

2. 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). Medline doi:10.1093/nar/gkr1079

3. H. Kawaji, J. Severin, M. Lizio, A. R. Forrest, E. van Nimwegen, M. Rehli, K. Schroder, K. Irvine, H. Suzuki, P. Carninci, Y. Hayashizaki, C. O. Daub, Update of the FANTOM web resource: From mammalian transcriptional landscape to its dynamic regulation. *Nucleic Acids Res.* **39**(Database), D856–D860 (2011). Medline doi:10.1093/nar/gkq1112

4. A. Brazma, H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, P. Kemmeren, G. G. Lara, A. Oezcimen, P. Rocca-Serra, S. A. Sansone, ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**, 68–71 (2003). Medline doi:10.1093/nar/gkg091

5. M. Magrane, U. Consortium, UniProt Knowledgebase: A hub of integrated protein data. *Database (Oxford)* **2011**, bar009 (2011). Medline doi:10.1093/database/bar009

6. P. Gaudet, G. Argoud-Puy, I. Cusin, P. Duek, O. Evalet, A. Gateau, A. Gleizes, M. Pereira, M. Zahn-Zabal, C. Zwahlen, A. Bairoch, L. Lane, neXtProt: Organizing protein knowledge in the context of human proteome projects. *J. Proteome Res.* **12**, 293–298 (2013). Medline doi:10.1021/pr300830v

7. I. Dunham, A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Frietze, J. Harrow, R. Kaul, J. Khatun, B. R. Lajoie, S. G. Landt, B.-K. Lee, F. Pauli, K. R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shoresh, J. M. Simon, L. Song, N. D. Trinklein, R. C. Altshuler, E. Birney, J. B. Brown, C. Cheng, S. Djebali, X. Dong, I. Dunham, J. Ernst, T. S. Furey, M. Gerstein, B. Giardine, M. Greven, R. C. Hardison, R. S. Harris, J. Herrero, M. M. Hoffman, S. Iyer, M. Kellis, J. Khatun, P. Kheradpour, A. Kundaje, T. Lassmann, Q. Li, X. Lin, G. K. Marinov, A. Merkel, A. Mortazavi, S. C. J. Parker, T. E. Reddy, J. Rozowsky, F. Schlesinger, R. E. Thurman, J. Wang, L. D. Ward, T. W. Whitfield, S. P. Wilder, W. Wu, H. S. Xi, K. Y. Yip, J. Zhuang, B. E. Bernstein, E. Birney, I. Dunham, E. D. Green, C. Gunter, M. Snyder, M. J. Pazin, R. F. Lowdon, L. A. L. Dillon, L. B. Adams, C. J. Kelly, J. Zhang, J. R. Wexler, E. D. Green, P. J. Good, E. A. Feingold, B. E. Bernstein, E. Birney, G. E. Crawford, J. Dekker, L. Elnitski, P. J. Farnham, M. Gerstein, M. C. Giddings, T. R. Gingeras, E. D. Green, R. Guigó, R. C. Hardison, T. J. Hubbard, M. Kellis, W. J. Kent, J. D. Lieb, E. H. Margulies, R. M. Myers, M. Snyder, J. A. Stamatoyannopoulos, S. A. Tenenbaum, Z. Weng, K. P. White, B.

Wold, J. Khatun, Y. Yu, J. Wrobel, B. A. Risk, H. P. Gunawardena, H. C. Kuiper, C. W. Maier, L. Xie, X. Chen, M. C. Giddings, B. E. Bernstein, C. B. Epstein, N. Shoresh, J. Ernst, P. Kheradpour, T. S. Mikkelsen, S. Gillespie, A. Goren, O. Ram, X. Zhang, L. Wang, R. Issner, M. J. Coyne, T. Durham, M. Ku, T. Truong, L. D. Ward, R. C. Altshuler, M. L. Eaton, M. Kellis, S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, P. Batut, I. Bell, K. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. P. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, G. Li, O. J. Luo, E. Park, J. B. Preall, K. Presaud, P. Ribeca, B. A. Risk, D. Robyr, X. Ruan, M. Sammeth, K. S. Sandhu, L. Schaeffer, L.-H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, Y. Hayashizaki, J. Harrow, M. Gerstein, T. J. Hubbard, A. Reymond, S. E. Antonarakis, G. J. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, K. R. Rosenbloom, C. A. Sloan, K. Learned, V. S. Malladi, M. C. Wong, G. P. Barber, M. S. Cline, T. R. Dreszer, S. G. Heitner, D. Karolchik, W. J. Kent, V. M. Kirkup, L. R. Meyer, J. C. Long, M. Maddren, B. J. Raney, T. S. Furey, L. Song, L. L. Grasfeder, P. G. Giresi, B.-K. Lee, A. Battenhouse, N. C. Sheffield, J. M. Simon, K. A. Showers, A. Safi, D. London, A. A. Bhinge, C. Shestak, M. R. Schaner, S. Ki Kim, Z. Z. Zhang, P. A. Mieczkowski, J. O. Mieczkowska, Z. Liu, R. M. McDaniell, Y. Ni, N. U. Rashid, M. J. Kim, S. Adar, Z. Zhang, T. Wang, D. Winter, D. Keefe, E. Birney, V. R. Iyer, J. D. Lieb, G. E. Crawford, G. Li, K. S. Sandhu, M. Zheng, P. Wang, O. J. Luo, A. Shahab, M. J. Fullwood, X. Ruan, Y. Ruan, R. M. Myers, F. Pauli, B. A. Williams, J. Gertz, G. K. Marinov, T. E. Reddy, J. Vielmetter, E. Partridge, D. Trout, K. E. Varley, C. Gasper, A. Bansal, S. Pepke, P. Jain, H. Amrhein, K. M. Bowling, M. Anaya, M. K. Cross, B. King, M. A. Muratet, I. Antoshechkin, K. M. Newberry, K. McCue, A. S. Nesmith, K. I. Fisher-Aylor, B. Pusey, G. DeSalvo, S. L. Parker, S. Balasubramanian, N. S. Davis, S. K. Meadows, T. Eggleston, C. Gunter, J. S. Newberry, S. E. Levy, D. M. Absher, A. Mortazavi, W. H. Wong, B. Wold, M. J. Blow, A. Visel, L. A. Pennachio, L. Elnitski, E. H. Margulies, S. C. J. Parker, H. M. Petrykowska, A. Abyzov, B. Aken, D. Barrell, G. Barson, A. Berry, A. Bignell, V. Boychenko, G. Bussotti, J. Chrast, C. Davidson, T. Derrien, G. Despacio-Reyes, M. Diekhans, I. Ezkurdia, A. Frankish, J. Gilbert, J. M. Gonzalez, E. Griffiths, R. Harte, D. A. Hendrix, C. Howald, T. Hunt, I. Jungreis, M. Kay, E. Khurana, F. Kokocinski, J. Leng, M. F. Lin, J. Loveland, Z. Lu, D. Manthravadi, M. Mariotti, J. Mudge, G. Mukherjee, C. Notredame, B. Pei, J. M. Rodriguez, G. Saunders, A. Sboner, S. Searle, C. Sisu, C. Snow, C. Steward, A. Tanzer, E. Tapanari, M. L. Tress, M. J. van Baren, N. Walters, S. Washietl, L. Wilming, A. Zadissa, Z. Zhang, M. Brent, D. Haussler, M. Kellis, A. Valencia, M. Gerstein, A. Reymond, R. Guigó, J. Harrow, T. J. Hubbard, S. G. Landt, S. Frietze, A. Abyzov, N. Addleman, R. P. Alexander, R. K. Auerbach, S. Balasubramanian, K. Bettinger, N. Bhardwaj, A. P. Boyle, A. R. Cao, P. Cayting, A. Charos, Y. Cheng, C. Cheng, C. Eastman, G. Euskirchen, J. D. Fleming, F. Grubert, L. Habegger, M. Hariharan, A. Harmanci, S. Iyengar, V. X. Jin, K. J. Karczewski, M. Kasowski, P. Lacroute, H. Lam, N. Lamarre-Vincent, J. Leng, J. Lian,

M. Lindahl-Allen, R. Min, B. Miotto, H. Monahan, Z. Moqtaderi, X. J. Mu, H. O'Geen, Z. Ouyang, D. Patacsil, B. Pei, D. Raha, L. Ramirez, B. Reed, J. Rozowsky, A. Sboner, M. Shi, C. Sisu, T. Slifer, H. Witt, L. Wu, X. Xu, K.-K. Yan, X. Yang, K. Y. Yip, Z. Zhang, K. Struhl, S. M. Weissman, M. Gerstein, P. J. Farnham, M. Snyder, S. A. Tenenbaum, L. O. Penalva, F. Doyle, S. Karmakar, S. G. Landt, R. R. Bhanvadia, A. Choudhury, M. Domanus, L. Ma, J. Moran, D. Patacsil, T. Slifer, A. Victorsen, X. Yang, M. Snyder, K. P. White, T. Auer, L. Centanin, M. Eichenlaub, F. Gruhl, S. Heermann, B. Hoeckendorf, D. Inoue, T. Kellner, S. Kirchmaier, C. Mueller, R. Reinhardt, L. Schertel, S. Schneider, R. Sinn, B. Wittbrodt, J. Wittbrodt, Z. Weng, T. W. Whitfield, J. Wang, P. J. Collins, S. F. Aldred, N. D. Trinklein, E. C. Partridge, R. M. Myers, J. Dekker, G. Jain, B. R. Lajoie, A. Sanyal, G. Balasundaram, D. L. Bates, R. Byron, T. K. Canfield, M. J. Diegel, D. Dunn, A. K. Ebersol, T. Frum, K. Garg, E. Gist, R. S. Hansen, L. Boatman, E. Haugen, R. Humbert, G. Jain, A. K. Johnson, E. M. Johnson, T. V. Kutyavin, B. R. Lajoie, K. Lee, D. Lotakis, M. T. Maurano, S. J. Neph, F. V. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, E. Rynes, P. Sabo, M. E. Sanchez, R. S. Sandstrom, A. Sanyal, A. O. Shafer, A. B. Stergachis, S. Thomas, R. E. Thurman, B. Vernot, J. Vierstra, S. Vong, H. Wang, M. A. Weaver, Y. Yan, M. Zhang, J. M. Akey, M. Bender, M. O. Dorschner, M. Groudine, M. J. MacCoss, P. Navas, G. Stamatoyannopoulos, R. Kaul, J. Dekker, J. A. Stamatoyannopoulos, I. Dunham, K. Beal, A. Brazma, P. Flicek, J. Herrero, N. Johnson, D. Keefe, M. Lukk, N. M. Luscombe, D. Sobral, J. M. Vaquerizas, S. P. Wilder, S. Batzoglou, A. Sidow, N. Hussami, S. Kyriazopoulou-Panagiotopoulou, M. W. Libbrecht, M. A. Schaub, A. Kundaje, R. C. Hardison, W. Miller, B. Giardine, R. S. Harris, W. Wu, P. J. Bickel, B. Banfai, N. P. Boley, J. B. Brown, H. Huang, Q. Li, J. J. Li, W. S. Noble, J. A. Bilmes, O. J. Buske, M. M. Hoffman, A. D. Sahu, P. V. Kharchenko, P. J. Park, D. Baker, J. Taylor, Z. Weng, S. Iyer, X. Dong, M. Greven, X. Lin, J. Wang, H. S. Xi, J. Zhuang, M. Gerstein, R. P. Alexander, S. Balasubramanian, C. Cheng, A. Harmanci, L. Lochovsky, R. Min, X. J. Mu, J. Rozowsky, K.-K. Yan, K. Y. Yip, E. Birney; ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). Medline doi:10.1038/nature11247

8. Y. K. Paik, S. K. Jeong, G. S. Omenn, M. Uhlen, S. Hanash, S. Y. Cho, H. J. Lee, K. Na, E. Y. Choi, F. Yan, F. Zhang, Y. Zhang, M. Snyder, Y. Cheng, R. Chen, G. Marko-Varga, E. W. Deutsch, H. Kim, J. Y. Kwon, R. Aebersold, A. Bairoch, A. D. Taylor, K. Y. Kim, E. Y. Lee, D. Hochstrasser, P. Legrain, W. S. Hancock, The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **30**, 221–223 (2012). Medline doi:10.1038/nbt.2152

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

10. M. S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T.

Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T. C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda, A. Pandey, A draft map of the human proteome. *Nature* **509**, 575–581 (2014). Medline doi:10.1038/nature13302

11. M. Mann, Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7**, 952–958 (2006). Medline doi:10.1038/nrm2067

12. I. Ezkurdia, D. Juan, J. M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia, M. L. Tress, Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* **23**, 5866–5878 (2014). Medline doi:10.1093/hmg/ddu309

13. V. Lange, P. Picotti, B. Domon, R. Aebersold, Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **4**, 222 (2008). Medline doi:10.1038/msb.2008.61

14. M. Uhlen, P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, H. Wernerus, L. Björling, F. Ponten, Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010). Medline doi:10.1038/nbt1210-1248

15. L. Fagerberg, B. M. Hallström, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, A. Asplund, E. Sjöstedt, E. Lundberg, C. A. Szigyarto, M. Skogs, J. O. Takanen, H. Berling, H. Tegel, J. Mulder, P. Nilsson, J. M. Schwenk, C. Lindskog, F. Danielsson, A. Mardinoglu, A. Sivertsson, K. von Feilitzen, M. Forsberg, M. Zwahlen, I. Olsson, S. Navani, M. Huss, J. Nielsen, F. Pontén, M. Uhlén, Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014). Medline doi:10.1074/mcp.M113.035600

16. C. Kampf, A. Mardinoglu, L. Fagerberg, B. M. Hallström, K. Edlund, E. Lundberg, F. Pontén, J. Nielsen, M. Uhlen, The human liver-specific proteome defined by transcriptomics and antibody-based profiling. *FASEB J.* **28**, 2901–2914 (2014). Medline doi:10.1096/fj.14-250555

17. D. Djureinovic, L. Fagerberg, B. Hallström, A. Danielsson, C. Lindskog, M. Uhlén, F. Pontén, The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol. Hum. Reprod.* **20**, 476–488 (2014). Medline doi:10.1093/molehr/gau018

18. G. Gremel, A. Wanders, J. Cedernaes, L. Fagerberg, B. Hallström, K. Edlund, E. Sjöstedt, M. Uhlén, F. Pontén, The human gastrointestinal tract-specific transcriptome and proteome as defined by RNA sequencing and antibody-based profiling. *J. Gastroenterol.* (2014). 10.1007/s00535-014-0958-7 Medline doi:10.1007/s00535-014-0958-7

19. V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y.

Zhou, .D. S. Wishart, DrugBank 4.0: shedding new light on drug metabolism, *Nucleic Acids Res.* **42** (Database), D1091–D1097 (2014). Medline doi:10.1093/nar/gkt1068

20. COSMIC catalogue of somatic mutations in cancer (2014); http://cancer.sanger.ac.uk/cancergenome/projects/census.

21. E. Lundberg, L. Fagerberg, D. Klevebring, I. Matic, T. Geiger, J. Cox, C. Algenäs, J. Lundeberg, M. Mann, M. Uhlen, Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **6**, 450 (2010). Medline doi:10.1038/msb.2010.106

22. Y. Taniguchi, P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010). Medline doi:10.1126/science.1188308

23. 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). Medline doi:10.1038/nmeth.1701

24. 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). Medline doi:10.1093/nar/gkm256

25. H. Viklund, A. Bernsel, M. Skwark, A. Elofsson, SPOCTOPUS: A combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **24**, 2928–2929 (2008). Medline doi:10.1093/bioinformatics/btn550

26. 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). Medline doi:10.1371/journal.pcbi.1000598

27. E. Wingender, T. Schoeps, J. Dönitz, TFClass: An expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* **41** (D1), D165–D170 (2013). Medline doi:10.1093/nar/gks1123

28. D. Tamborero, A. Gonzalez-Perez, C. Perez-Llamas, J. Deu-Pons, C. Kandoth, J. Reimand, M. S. Lawrence, G. Getz, G. D. Bader, L. Ding, N. Lopez-Bigas, Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650 (2013). Medline

29. D. M. Muzny, M. N. Bainbridge, K. Chang, H. H. Dinh, J. A. Drummond, G. Fowler, C. L. Kovar, L. R. Lewis, M. B. Morgan, I. F. Newsham, J. G. Reid, J. Santibanez, E. Shinbrot, L. R. Trevino, Y.-Q. Wu, M. Wang, P. Gunaratne, L. A. Donehower, C. J. Creighton, D. A. Wheeler, R. A. Gibbs, M. S. Lawrence, D. Voet, R. Jing, K. Cibulskis, A. Sivachenko, P. Stojanov, A. McKenna, E. S. Lander, S. Gabriel, G. Getz, L. Ding, R. S. Fulton, D. C. Koboldt, T. Wylie, J. Walker, D. J. Dooling, L. Fulton, K. D. Delehaunty, C. C. Fronick, R. Demeter, E. R. Mardis, R. K. Wilson, A. Chu, H.-J. E. Chun, A. J. Mungall, E. Pleasance, A. Gordon Robertson, D. Stoll, M. Balasundaram, I. Birol, Y. S. N. Butterfield, E. Chuah, R. J. N. Coope, N. Dhalla, R. Guin, C. Hirst, M. Hirst, R. A. Holt, D. Lee, H. I. Li, M. Mayo, R. A. Moore, J. E. Schein, J. R. Slobodan, A. Tam, N. Thiessen, R. Varhol, T. Zeng, Y. Zhao, S. J. M. Jones, M. A. Marra, A. J. Bass, A. H. Ramos, G. Saksena, A. D. Cherniack, S. E. Schumacher, B. Tabak, S. L. Carter, N. H.

Pho, H. Nguyen, R. C. Onofrio, A. Crenshaw, K. Ardlie, R. Beroukhim, W. Winckler, G. Getz, M. Meyerson, A. Protopopov, J. Zhang, A. Hadjipanayis, E. Lee, R. Xi, L. Yang, X. Ren, H. Zhang, N. Sathiamoorthy, S. Shukla, P.-C. Chen, P. Haseley, Y. Xiao, S. Lee, J. Seidman, L. Chin, P. J. Park, R. Kucherlapati, J. Todd Auman, K. A. Hoadley, Y. Du, M. D. Wilkerson, Y. Shi, C. Liquori, S. Meng, L. Li, Y. J. Turman, M. D. Topal, D. Tan, S. Waring, E. Buda, J. Walsh, C. D. Jones, P. A. Mieczkowski, D. Singh, J. Wu, A. Gulabani, P. Dolina, T. Bodenheimer, A. P. Hoyle, J. V. Simons, M. Soloway, L. E. Mose, S. R. Jefferys, S. Balu, B. D. O'Connor, J. F. Prins, D. Y. Chiang, D. Neil Hayes, C. M. Perou, T. Hinoue, D. J. Weisenberger, D. T. Maglinte, F. Pan, B. P. Berman, D. J. Van Den Berg, H. Shen, T. Triche Jr., S. B. Baylin, P. W. Laird, G. Getz, M. Noble, D. Voet, G. Saksena, N. Gehlenborg, D. DiCara, J. Zhang, H. Zhang, C.-J. Wu, S. Yingchun Liu, S. Shukla, M. S. Lawrence, L. Zhou, A. Sivachenko, P. Lin, P. Stojanov, R. Jing, R. W. Park, M.-D. Nazaire, J. Robinson, H. Thorvaldsdottir, J. Mesirov, P. J. Park, L. Chin, V. Thorsson, S. M. Reynolds, B. Bernard, R. Kreisberg, J. Lin, L. Iype, R. Bressler, T. Erkkilä, M. Gundapuneni, Y. Liu, A. Norberg, T. Robinson, D. Yang, W. Zhang, I. Shmulevich, J. J. de Ronde, N. Schultz, E. Cerami, G. Ciriello, A. P. Goldberg, B. Gross, A. Jacobsen, J. Gao, B. Kaczkowski, R. Sinha, B. Arman Aksoy, Y. Antipin, B. Reva, R. Shen, B. S. Taylor, T. A. Chan, M. Ladanyi, C. Sander, R. Akbani, N. Zhang, B. M. Broom, T. Casasent, A. Unruh, C. Wakefield, S. R. Hamilton, R. Craig Cason, K. A. Baggerly, J. N. Weinstein, D. Haussler, C. C. Benz, J. M. Stuart, S. C. Benz, J. Zachary Sanborn, C. J. Vaske, J. Zhu, C. Szeto, G. K. Scott, C. Yau, S. Ng, T. Goldstein, K. Ellrott, E. Collisson, A. E. Cozen, D. Zerbino, C. Wilks, B. Craft, P. Spellman, R. Penny, T. Shelton, M. Hatfield, S. Morris, P. Yena, C. Shelton, M. Sherman, J. Paulauskis, J. M. Gastier-Foster, J. Bowen, N. C. Ramirez, A. Black, R. Pyatt, L. Wise, P. White, M. Bertagnolli, J. Brown, T. A. Chan, G. C. Chu, C. Czerwinski, F. Denstman, R. Dhir, A. Dörner, C. S. Fuchs, J. G. Guillem, M. Iacocca, H. Juhl, A. Kaufman, B. Kohl III, X. Van Le, M. C. Mariano, E. N. Medina, M. Meyers, G. M. Nash, P. B. Paty, N. Petrelli, B. Rabeno, W. G. Richards, D. Solit, P. Swanson, L. Temple, J. E. Tepper, R. Thorp, E. Vakiani, M. R. Weiser, J. E. Willis, G. Witkin, Z. Zeng, M. J. Zinner, C. Zornig, M. A. Jensen, R. Sfeir, A. B. Kahn, A. L. Chu, P. Kothiyal, Z. Wang, E. E. Snyder, J. Pontius, T. D. Pihl, B. Ayala, M. Backus, J. Walton, J. Whitmore, J. Baboud, D. L. Berton, M. C. Nicholls, D. Srinivasan, R. Raman, S. Girshik, P. A. Kigonya, S. Alonso, R. N. Sanbhadti, S. P. Barletta, J. M. Greene, D. A. Pot, K. R. Mills Shaw, L. A. L. Dillon, K. Buetow, T. Davidsen, J. A. Demchok, G. Eley, M. Ferguson, P. Fielding, C. Schaefer, M. Sheth, L. Yang, M. S. Guyer, B. A. Ozenberger, J. D. Palchik, J. Peterson, H. J. Sofia, E. Thomson; Cancer Genome Atlas Network, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012). Medline doi:10.1038/nature11252

30. C. S. Grasso, Y. M. Wu, D. R. Robinson, X. Cao, S. M. Dhanasekaran, A. P. Khan, M. J. Quist, X. Jing, R. J. Lonigro, J. C. Brenner, I. A. Asangani, B. Ateeq, S. Y. Chun, J. Siddiqui, L. Sam, M. Anstett, R. Mehra, J. R. Prensner, N. Palanisamy, G. A. Ryslik, F. Vandin, B. J. Raphael, L. P. Kunju, D. R. Rhodes, K. J. Pienta, A. M. Chinnaiyan, S. A. Tomlins, The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012). Medline doi:10.1038/nature11125

31. F. Danielsson, M. Wiking, D. Mahdessian, M. Skogs, H. Ait Blal, M. Hjelmare, C. Stadler, M. Uhlén, E. Lundberg, 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). Medline doi:10.1021/pr3009308

32. M. Schnabel, S. Marlovits, G. Eckhoff, I. Fichtel, L. Gotzen, V. Vécsei, J. Schlegel, Dedifferentiation-associated changes in morphology and gene expression in primary human articular chondrocytes in cell culture. *Osteoarthritis Cartilage* **10**, 62–70 (2002). Medline doi:10.1053/joca.2001.0482

33. E. Birney, J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, M. S. Kuehn, C. M. Taylor, S. Neph, C. M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. A. Greenbaum, R. M. Andrews, P. Flicek, P. J. Boyle, H. Cao, N. P. Carter, G. K. Clelland, S. Davis, N. Day, P. Dhami, S. C. Dillon, M. O. Dorschner, H. Fiegler, P. G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K. D. James, B. E. Johnson, E. M. Johnson, T. T. Frum, E. R. Rosenzweig, N. Karnani, K. Lee, G. C. Lefebvre, P. A. Navas, F. Neri, S. C. Parker, P. J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F. S. Collins, J. Dekker, J. D. Lieb, T. D. Tullius, G. E. Crawford, S. Sunyaev, W. S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I. L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H. A. Hirsch, E. A. Sekinger, J. Lagarde, J. F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J. S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M. C. Dickson, D. J. Thomas, M. T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K. G. Srinivasan, W. K. Sung, H. S. Ooi, K. P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M. L. Tress, A. Valencia, S. W. Choo, C. Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T. G. Clark, J. B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C. N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J. S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R. M. Myers, J. Rogers, P. F. Stadler, T. M. Lowe, C. L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S. E. Antonarakis, Y. Fu, E. D. Green, U. Karaöz, A. Siepel, J. Taylor, L. A. Liefer, K. A. Wetterstrand, P. J. Good, E. A. Feingold, M. S. Guyer, G. M. Cooper, G. Asimenos, C. N. Dewey, M. Hou, S. Nikolaev, J. I. Montoya-Burgos, A. Löytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N. R. Zhang, I. Holmes, J. C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W. J. Kent, E. A. Stone, S. Batzoglou, N. Goldman, R. C. Hardison, D. Haussler, W. Miller, A. Sidow, N. D. Trinklein, Z. D. Zhang, L. Barrera, R. Stuart, D. C. King, A. Ameur, S. Enroth, M. C. Bieda, J. Kim, A. A. Bhinge, N. Jiang, J. Liu, F. Yao, V. B. Vega, C. W. Lee, P. Ng, A. Shahab, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M. J. Oberley, D. Inman, M. A. Singer, T. A. Richmond, K. J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J. C. Fowler, P. Couttet, A. W. Bruce, O. M. Dovey, P. D. Ellis, C. F. Langford, D. A. Nix, G. Euskirchen, S. Hartman, A. E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C. K. Glass, M. G. Rosenfeld, S. F. Aldred, S. J. Cooper, A. Halees, J. M. Lin, H. P. Shulha, X. Zhang, M. Xu, J. N. Haidar, Y. Yu, Y. Ruan, V. R. Iyer, R. D. Green, C. Wadelius, P. J. Farnham, B. Ren, R. A. Harte, A. S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A. S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R. M. Kuhn, D. Karolchik, L. Armengol, C. P. Bird, P. I. de Bakker, A. D. Kern, N. Lopez-Bigas, J. D. Martin, B. E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I. B.

Hallgrímsdóttir, J. Huppert, M. C. Zody, G. R. Abecasis, X. Estivill, G. G. Bouffard, X. Guan, N. F. Hansen, J. R. Idol, V. V. Maduro, B. Maskeri, J. C. McDowell, M. Park, P. J. Thomas, A. C. Young, R. W. Blakesley, D. M. Muzny, E. Sodergren, D. A. Wheeler, K. C. Worley, H. Jiang, G. M. Weinstock, R. A. Gibbs, T. Graves, R. Fulton, E. R. Mardis, R. K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D. B. Jaffe, J. L. Chang, K. Lindblad-Toh, E. S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu, P. J. de Jong, ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007). Medline doi:10.1038/nature05874

34. M. Gonzàlez-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). Medline doi:10.1186/gb-2013-14-7-r70

35. A. Mardinoglu, J. Nielsen, Systems medicine and metabolic modelling. *J. Intern. Med.* **271**, 142–154 (2012). Medline doi:10.1111/j.1365-2796.2011.02493.x

36. A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen, J. Nielsen, Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014). Medline doi:10.1038/ncomms4083

37. R. Agren, A. Mardinoglu, A. Asplund, C. Kampf, M. Uhlen, J. Nielsen, Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721 (2014). Medline doi:10.1002/msb.145122

38. Human Metabolic Atlas, (2014); http://www.metabolicatlas.org/.

39. L. C. Crosswell, J. M. Thornton, ELIXIR: A distributed infrastructure for European biological data. *Trends Biotechnol.* **30**, 241–242 (2012). Medline doi:10.1016/j.tibtech.2012.02.002

40. J. F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, M. Vidal, Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005). Medline doi:10.1038/nature04209

41. C. Kampf, I. Olsson, U. Ryberg, E. Sjöstedt, F. Pontén, Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *J. Vis. Exp.* **2012**, 3620 (2012). Medline

42. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009). Medline doi:10.1093/bioinformatics/btp120

43. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010). Medline doi:10.1038/nbt.1621

44. P. Flicek, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, E. Kulesha, F. J. Martin, T. Maurel, W. M. McLaren, D. N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S. J. Trevanion, A. Vullo, S. P. Wilder, M. Wilson, A. Zadissa, B. L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T. J. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D. R. Zerbino, S. M. Searle, Ensembl 2014. *Nucleic Acids Res.* **42** (Database), D749–D755 (2014). Medline doi:10.1093/nar/gkt1196

45. Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources v6.7 (2014); http://david.abcc.ncifcrf.gov/.

46. L. Fagerberg, K. Jonasson, G. von Heijne, M. Uhlén, L. Berglund, Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010). Medline doi:10.1002/pmic.200900258

47. D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538–544 (2007). Medline doi:10.1093/bioinformatics/btl677

48. T. Nugent, D. T. Jones, Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 159 (2009). Medline doi:10.1186/1471-2105-10-159

49. L. Käll, A. Krogh, E. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004). Medline doi:10.1016/j.jmb.2004.03.016

50. A. Bernsel, H. Viklund, J. Falk, E. Lindahl, G. von Heijne, A. Elofsson, Prediction of membrane-protein topology from first principles. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7177–7181 (2008). Medline doi:10.1073/pnas.0711151105

51. E. L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, in *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, J. Glasgow, Ed., Montréal, Québec, 28 June to 1 July 1998 (AAAI Press, Menlo Park, CA, 1998).

52. H. Zhou, Y. Zhou, Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model-based method. *Protein Sci.* **12**, 1547–1555 (2003). Medline doi:10.1110/ps.0305103

53. J. D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004). Medline doi:10.1016/j.jmb.2004.05.028

54. W. Nickel, Pathways of unconventional protein secretion. *Curr. Opin. Biotechnol.* **21**, 621–626 (2010). Medline doi:10.1016/j.copbio.2010.06.004

55. D. Cotter, P. Guda, E. Fahy, S. Subramaniam, MitoProteome: Mitochondrial protein sequence database and annotation system. *Nucleic Acids Res.* **32**, D463–D467 (2004). Medline doi:10.1093/nar/gkh048

56. H. Zola, B. Swart, I. Nicholson, B. Aasted, A. Bensussan, L. Boumsell, C. Buckley, G. Clark, K. Drbal, P. Engel, D. Hart, V. Horejsí, C. Isacke, P. Macardle, F. Malavasi, D. Mason, D. Olive, A. Saalmueller, S. F. Schlossman, R. Schwartz-Albiez, P. Simmons, T. F. Tedder, M. Uguccioni, H. Warren, CD molecules 2005: Human cell differentiation molecules. *Blood* **106**, 3123–3126 (2005). Medline doi:10.1182/blood-2005-03-1338

57. UniProt Consortium, Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42** (D1), D191–D198 (2014). Medline doi:10.1093/nar/gkt1140