

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                    |

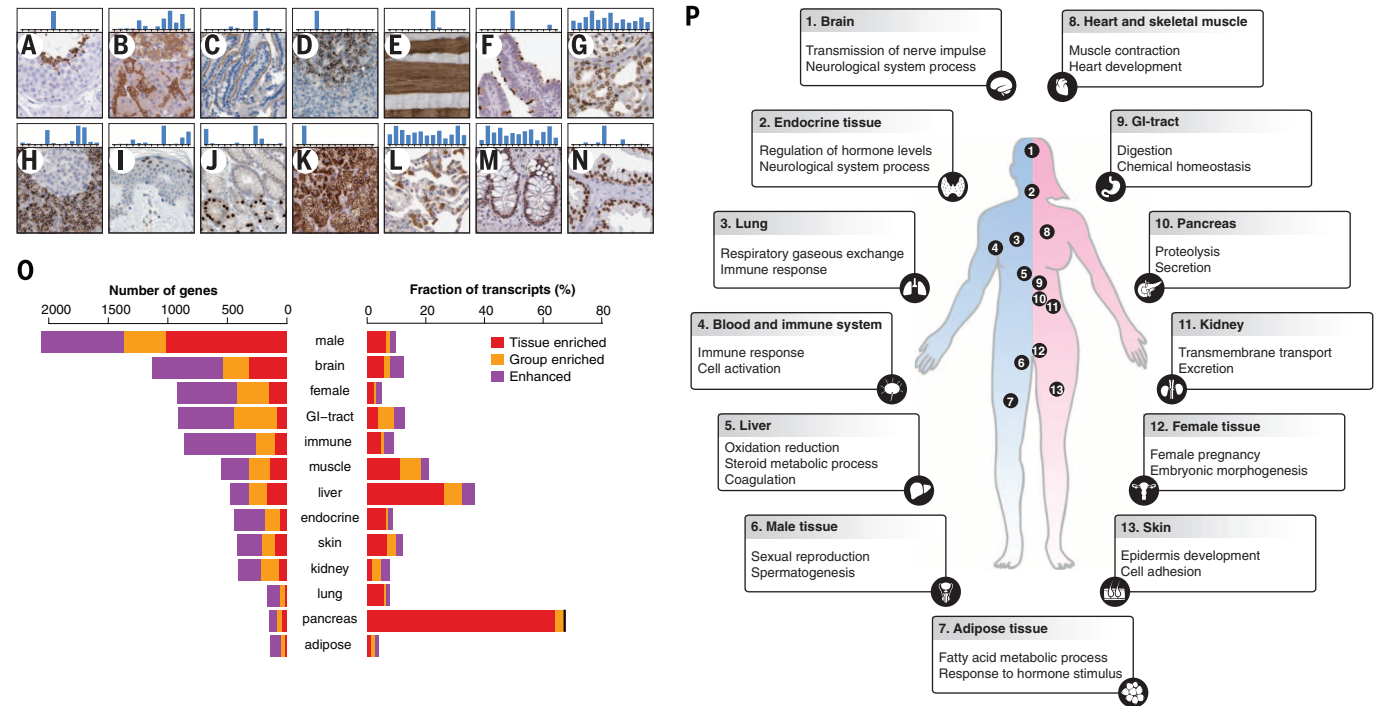


Fig. 2. Tissue microarray-based protein expression, and analysis of tissue-elevated genes in the different organ systems. (A to N) Tissue expression and localization for a selection of human proteins. Larger images corresponding to (A) to (N) of the figure are shown in fig. S3. The levels of the corresponding 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 ROPN1L 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.