

GENE EXPRESSION

Spatial characterization of proteomes

“
a spatial atlas
of expression
for thousands
of genes at
single-cell
resolution”

A new integrative genomic resource describes proteomes and transcriptomes across diverse tissues and organs of the human body, providing a spatial atlas of expression for thousands of genes at single-cell resolution.

To build their resource, Uhlén *et al.* used samples from 44 human tissues (representing all the major tissue and organ types), each of which was probed with ~24,000 different antibodies. The resultant >13 million immunohistological images provided information on expression levels and spatial patterns from ~17,000 protein-coding genes (including multiple protein isoforms for some genes). Such a resource provides complementary information to previous proteogenomic maps based on mass spectrometry data: although there might be some protein cross-reactivity with antibody-based methods, the retention of cellular context in the immunohistological images allows spatial resolution to single-cell and subcellular scales. To supplement their proteomic data, the investigators also gathered high-throughput RNA sequencing (RNA-seq) data for 32 of the tissue types.

Combining data sets allowed the investigators to refine estimates of the number of protein-coding human genes. They found 17,132 genes with at least some evidence of a protein product (from their data or previous data) and a further 2,546 genes expressed at the RNA level with no current evidence of a protein product.

Genes were classified on the basis of their expression patterns across tissues. 9,000 genes that were expressed across all tissues analysed were assigned as ‘housekeeping’ genes. By contrast, genes were also classified into three categories that reflected increasing degrees of elevated expression in particular tissues. Of note, the authors intentionally avoided the term ‘tissue specific’ because their data support wider expression across tissues for many genes referred to as tissue specific in the literature.

Uhlén *et al.* also analysed the spatial expression patterns of particular functional groups of genes, and a striking theme was the ubiquitous expression across all tissues for a substantial proportion of these genes. For example, 30% of genes encoding current pharmaceutical targets, 60% of cancer driver genes, 41% of genes encoding transcription factors and 60% of metabolic genes were expressed in all tissue types. As one example, catechol-O-methyltransferase (COMT) is involved in neurotransmitter degradation and in the metabolism of drugs for Parkinson’s disease, yet it is expressed ubiquitously. Such results have implications for attempts to therapeutically target proteins underlying tissue-restricted pathologies, as side effects in other tissues may be more widespread than expected if the target proteins are ubiquitously expressed.

The investigators also characterized regulatory roles for alternative splicing in protein localization and showed that, for a given gene, different tissues can favour the expression of different transcript isoforms that are predicted to switch the encoded proteins between membrane-bound versus secreted isoforms. Although these findings were based on the differential expression of alternative mRNA isoforms, further use of immunohistochemistry or related *in situ* approaches might allow direct analyses of the subcellular localizations of the resultant proteins.

The data from this study are available at <http://www.proteinatlas.org> as a resource for the community and to aid future studies of proteomes and individual proteins.

Darren J. Burgess

ORIGINAL RESEARCH PAPER Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015)

