James Diao
MB&B 452
10 February 2017

Section 2: Proteomics/Sequence Alignment

These two readings introduce and the first efforts to document the proteome, or the entire complement of proteins expressed by histologically normal human cells and tissues. These are the efforts of two independent research groups that curated catalogs of human proteins. Both teams broke new ground, identifying novel proteins and translated pseudogenes, among other complexities. Both papers differ from previous work in that they examine a wide diversity of tissue types to provide comprehensive coverage of more than 84% and 92% of the expected proteome (based on the number of coding genes).

Kim et al. analyzed 30 different tissue types (including 17 adult tissues, 7 fetal tissues and 6 types of hematopoietic cells) by fractionation, SDS-PAGE, basic reversed-phase liquid chromatography, followed by high-resolution mass spectrometry. The lab generated all of their own data. This allowed them to easily conduct intra-dataset comparisons, but also validate their results using peptide-based resources such as PeptideAtlas and GPMDB. The team discovered that almost 50% of their discovered peptides were not documented in these databases, promising a large increase in coverage. Lastly, Kim et al. found 193 new proteins that map to "noncoding" DNA regions, especially long intergenic non-coding RNAs (lincRNAs). All of their results were published online through the NCBI database.

Wilhelm et al., on the other hand, compiled 60% of their data from raw mass spectrometry data from public databases and their colleagues. They filled in the remaining 40% using their own data, generated from 60 human tissues, 13 biofluids, and 147 cancer cell lines. The research group aimed to ensure quality control by using high-resolution public data and a uniform processing pipeline. This data was deposited in ProteomicsDB, a searchable public database, for the proteomics community. Upon further analysis, the team found that the ratio of protein-to-mRNA levels is relatively constant globally, meaning that protein levels are often predictable from mRNA levels. Lastly, Wilhelm et al. connected their work to drug-sensitivity data, allowing them to identify proteins that are correlated with drug resistance or sensitivity.

- A draft map of the human proteome. Nature 509,575–581 (29 May 2014)
- Mass-spectrometry-based draft of the human proteome. Nature 509, 582–587 (29 May 2014)