A more educated guess

Guessing before proving! Need I remind you that it is how all important discoveries have been made?

Henri Poincaré

Almost all of the interesting things that happen in living creatures are accomplished by proteins. Proteins give cells physical structures and mechanical properties, protein-based enzymes drive the reactions of life, protein-based hormones carry signals and information to act on protein-based receptors and, through them, on protein-based computing apparatus in the cytoplasm of cells. Even genes switch on and off only on the command of proteins. Researchers wanting to understand cells, and wanting to understand, for example, how diseased cells differ from normal ones, tend therefore to be most interested in the cells' proteins. For many types of research, it would be really useful to know how much of each type of protein a cell is making. The problem is that making quantitative studies of thousands of proteins at once is difficult, expensive, and requires quite large amounts of tissue.

For this reason, biologists often give up on measuring the proteins directly and instead measure the amounts of the messenger RNAs (mRNAs), which are produced from genes and which direct the production of proteins. Well-established techniques such as micro-arrays and, more recently, RNA-Seq, allow the relative amounts of all of the mRNAs in a cell to be measured with reasonable precision (indeed RNA-Seq offers really quite high precision if there is time and money to pay enough 'reads'). There are huge databases of mRNA expression in tissues during development and disease: the GUDMAP database run from this lab is just one example. The problem is that knowing how much mRNA for a protein is present does not necessarily mean that one knows how much protein is there. We have to guess that there is enough of a relationship that we are not completely fooling ourselves when interpreting exprimental results. This is a problem that is widely acknowledged but not solved.

Until, maybe, now.

Last month, Fredrik Edfors and colleagues published an interesting paper in Molecular Systems

Biology that tackled the problem head-on. They began with nine human cell lines and ethically-obtained human tissues, and subjected samples of each to RNA-Seq analysis to determine the relative abundances of mRNAs encoding different proteins in each sample. They expressed these abundances as copies of mRNA for a specific protein per million mRNAs overall.

The authors then used a very carefully constructed protein assay, which used known amounts of isotope-labelled tracer proteins to calibrate the system, to determine the relative abundances of 55 specific proteins in each of their samples. By carefully assessing how many cells were in each sample, they were able to express these as molecules of protein per cell. These ranged from just a few thousand molecules of some proteins per cell to tens of millions of other proteins in the same cell. The researchers' use of both cell lines and normal human tissues allowed them to observe, in passing, that cell lines can show very odd protein expression profiles compared to normal tissues (no surprise there!).

Having obtained accurate measurements of mRNAs and accurate measures of proteins, Edfors and colleagues were able to bring the two datasets together to ask whether there is any correlation between the amount of mRNA for a protein and the amount of that protein. The answer turned out to be a little complicated. For a given protein, there is a relationship between the amount of mRNA and the amount of protein that held true across the cell lines and tissues. The statistical strength of this relationship was good (Pearson's r = around 0.9, for the technically-minded). But it turns out that the ratio between RNA and protein, though constant over many cell types for one protein, varies a lot (more than ten-fold) between different proteins. The implication is that, for a given protein, differences in the easily-measured RNA can be correlated to differences in protein even in different tissues, which is very encouraging.

The new study does not make all of the problems of measuring mRNA as a proxy for protein go away. Its author were careful to use stable cell lines and adult tissues which can be assumed to be in steady-state. Those of is who study systems that are in rapid change, whether in development of disease, face the problem that timing can really alter the ratio of mRNA and protein. Most dramatically, when a gene is first switched on mRNA is made and protein made from that mRNA only later, so there will be a short period in which there is mRNA but no protein has yet been made. Similarly, when a gene is switched off the mRNA will disappear relatively quickly (hours to days, usually) but proteins can last for weeks. Still, for people who study unchanging adult tissues in

steady-state, this new study can provide the assurance that using mRNA as a proxy for protein level is not unreasonable. At last we have something more than just a guess.

Jamie Davies,
Edinburgh,
November 2016

Links:

The Edfors et al. Paper: http://msb.embopress.org/content/12/10/883

GUDMAP: www.gudmap.org