nature cell biology

Towards quantitative cell biology

Prisca Liberali and Lucas Pelkmans

Cell biologists must decide whether to embrace the maturing field of systems biology. We argue that a fusion of the two is urgently needed to strengthen both fields.

As researchers working in both cell biology and systems biology, and making attempts to connect the two fields, we realise that the majority of cell biologists are not familiar with the merits of systems biology. Instead, they often seem more familiar with its pitfalls. This is unfortunate, as cell biologists are in a prime position to harness the power of the technologies that systems biology has brought forth. We do not merely suggest the application of these new technologies to classical cell biological questions, but rather that the fundamental approaches of systems biology, which are unbiased, large-scale, quantitative and multivariate, are integrated into the core of molecular cell biology in the future. This should be seen as a complementary and much-needed extension of traditional approaches in cell biology.

Systems biology has produced staggering amounts of complex data. Although the fields of functional genomics, proteomics and metabolomics are mining and integrating these data, molecular cell biology lags behind. Most importantly, in the last few years, systems biology has embraced the single-cell revolution, applying functional genomics, proteomics and metabolomics at the single-cell level — thereby allowing the analysis of collectives of molecules and the structures they form simultaneously within single cells. Ironically, the single cell has traditionally been the domain of cell biology, but the experiments were not large-scale, quantitative and multivariate. As a consequence, classical approaches in molecular cell biology can be particularly sensitive to experimental bias. Most experiments study only a small fraction or an average of the possible states of a cell biological process. This occurs in cell lysates, homogenized mixes of cellular extracts from millions of cells, representing an average of a complex mixture of states, and also in microscopy, where often only a small number of single cells or subcellular structures are analysed. To reduce such bias, an experiment must be large-scale in the sense that it consists of a large number of samplings. Although this can occasionally be achieved manually, it is clear that this is an ideal task for computers. This not only allows an individual to easily obtain large numbers of samplings, but it also decreases bias. This requires methods that yield results in formalized and quantitative formats (since these are machine-readable) for many molecular, morphological and spatial properties of subcellular structures, single cells and cell populations. With such multivariate measurements, correlation analysis and causal inference can be applied to learn the functional molecular causality

underlying cellular activities across different scales. Embracing these quantitative approaches will allow cell biologists to provide detailed information about their own experiments in a more formalized and structured manner. This can provide a much-needed extension of functional annotation and interaction databases such as DAVID and STRING, or manually curated databases such as KEGG pathways, as these lack accurate information about molecular machines and processes involved in subcellular compartmentalization, membrane trafficking and cytoskeletal regulation. This is necessary, as it is clear that a cornerstone of future scientific endeavours will be to devise strategies to obtain useful information from such data sets.

These quantitative and unbiased approaches hold many benefits for cell biologists. It may allow them to identify particular properties of single cells or subcellular regions that enrich for an otherwise rare phenotype, allowing the design of new experiments that favour the appearance of that specific phenotype. Most importantly, many fundamental properties of a system will only become visible when the spectrum of cellular states is sufficiently sampled. This is, for instance, illustrated in the analysis of quantile assembly of caveolae, heterogeneous dynamics of clathrin-coated pits, or patterns of cell-to-cell variability in virus infection. Also, it is highly relevant for the interpretation of perturbations. Although each single cell may display a normal level of activity corresponding to its state, altered proportions of cellular states in a population (for example, dense and small cells versus sparse and large cells) may lead to the incorrect conclusion that the activity itself is perturbed. Such an indirect perturbation is fundamentally different from a direct perturbation of the cellular activity. Similarly, perturbing a key factor in early endosome function may affect a cellular activity by causing changes in plasma membrane lipid composition on which the cellular activity depends, not because the cellular activity is directly controlled by endocytosis. Although further experiments must always be performed to test for such indirect effects, the reality is that confounding factors may often not be apparent, and thus remain untested, when an unbiased multivariate approach is not used.

By far the most important reason for building these approaches into the core of cell biology is that it allows the field to adopt the language and scientific rigour of the exact sciences, without sacrificing the characteristics that distinguish it from biophysics and bioinformatics. This transformation is necessary for cell biology to maintain its important central position in the rapidly evolving molecular life sciences. Implementing image analysis algorithms, applying multivariate statistics and deriving data-driven models must become second nature for a cell biologist.

Prisca Liberali and Lucas Pelkmans are at the Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.
e-mail: Lucas.pelkmans@imls.uzh.ch

1233