Biology 3.0: The Single-Cell (R)evolution

Understanding how cells make choices is one of the great challenges in biology. Whether it is the first cell divisions of embryonic development or activation in response to an immune stimulus, modern medicine has a great interest in how these decisions are made. This is particularly the case in regenerative medicine, where the aim is to create specific cell types for therapeutic applications. While the promise of cell therapies is enormous, most protocols don't make it to clinic, fail in trials, or are not commercially viable, due to issues with efficacy, safety and scalability. New solutions are required to improve these outcomes.

The genes, and therefore proteins, that a cell expresses dictate its fate and function. A fundamental question in biology is therefore how different cell types arise with diverse gene expression, despite carrying the same DNA. It is now well understood that around 10% of the human genome codes for transcription factors (TFs), proteins that control cell fate by regulating which genes are expressed. Using TFs to direct cell fate for therapeutic purposes is therefore a key area of study in regenerative medicine, with proof-of-concept experiments being performed as far back as the 1980s showing that fibroblasts can be converted to muscle cells by overexpressing the TF, MYOD1. While much of the work to date has used educated guesswork to predict factors for conversions, algorithms now exist to improve existing conversions, including CellNet[1] from the Daley, Morrison and Cahan labs in the US, and even to directly predict the TFs most likely to produce a successful conversion, such as the algorithm developed by Rackham *et al.*[2].

Both predicting TFs for conversions and assessing the impact of reprogrammed cells require a good understanding of the target cell types. Over the past decade, powerful methods like single-cell RNA sequencing (scRNAseq) have evolved to allow us to capture biological information at the level of individual cells, the fundamental decision-making units of life. These methods have huge potential to impact cell and gene therapies by better defining cell populations of interest and addressing issues of safety and efficacy in cell therapies.

Shifting the Paradigm from Populations to Single Cells

Cell types have been defined gradually over time, first using morphology and later with more sophisticated methods like microscopy and flow cytometry, but they are very limited in the number of parameters that can be analysed at once. Additionally, many techniques lack sensitivity, requiring thousands to millions of cells to give a reliable readout. In reality, cell identity and function are the result of a complex network of pathways and gene expression patterns that vary across time and even across cells of supposedly the same type. Taking an average view of a small number of parameters in large numbers of cells therefore masks important information. As such, many therapies have failed due to an incomplete understanding of the target cell type. So how can we tackle this problem?

DNA is a key molecule of interest, not only because it carries the blueprint for life but because it is easy to work with, easily amplified, and can be sequenced rapidly and accurately since the advent of next-generation sequencing. RNA can be reverse transcribed to cDNA and processed in the same way using RNAseq, giving a direct readout of cellular activity through the set of genes expressed. With the sequencing of the human genome came the realisation that the majority of human DNA doesn't encode proteins but has a critical role in regulating gene expression and so controlling cell fate, bringing with it a wealth of methods for studying such regulatory function, an area called epigenomics. Since then, technological developments have enabled these methods to be scaled down to understand the DNA landscape and RNA expression of individual cells (scRNAseq; Figure 1). The biological processes that can be studied by sequencing are expanding, seemingly limited only by our creativity in what we can attach a DNA molecule to. Protocols have additionally emerged to capture protein expression through DNA-tagged antibodies (CITEseq), to study antigen specificity of immune cells, and to combine CRISPR-based screening with expression data to increase experimental throughput. This unprecedented access to biology is changing how we approach science, the questions we can ask and our understanding of the fundamental mechanisms of cell biology, and so the way we develop and validate therapies (Figure 2).

Figure 1. Since the first scRNAseq study in 2009, many methods have been published and the number of cells that can be sequenced has increased dramatically

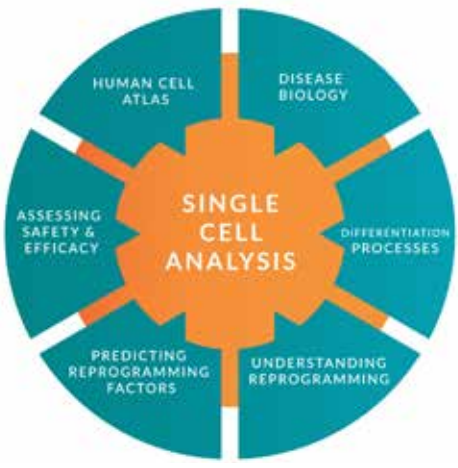

Figure 2. Single-cell analysis can be applied to many aspects of cell and gene therapy, from understanding normal tissues to assessing therapeutic efficacy

A Single-cell Atlas of the Human Body in Health and Disease

Regenerative medicine approaches typically involve directing the differentiation of stem cells through to adult cell types. This requires an understanding of the normal hierarchy of stem cells through to mature cell types. However, much of our knowledge in this area derives from trial and error, and from studying artificial rather than native cell systems. Now, scRNAseq allows for an unbiased approach to understanding cell types and tissue architecture, taking all cells in a tissue and determining which genes they express. A key example is in haematology, where much of our understanding of differentiation comes from experiments in which donor cells are transplanted into mice, which have been treated to destroy their own blood system. How the donor cells contribute to the different cellular lineages of the blood is then analysed to infer the potential of the original cells. While undoubtedly useful, this is a time-consuming and complicated analysis of a highly perturbed system that puts a great deal of stress on the cells. In contrast, single-cell analyses enable snapshots of dynamic tissues to be taken, for example looking at blood-cell production in the bone marrow. Using subtle changes in gene expression to order cells in "developmental pseudotime", it is possible to reconstruct a typical cellular hierarchy in an unperturbed system.

As a result of the potential in this field, there is a great push to make a single-cell map of the human body through the Human Cell Atlas (HCA) project, a worldwide consortium of researchers. In 2018, researchers from an HCA-affiliated lab at the Broad Institute discovered a previously unknown and rare type of cell in the lung, the pulmonary ionocyte, that is responsible for most of the lung's expression of *CFTR*, a crucial gene mutated in people with cystic fibrosis. This discovery therefore has the potential to dramatically change how cystic fibrosis is treated, particularly using targeted gene therapy.

However, the structure of a tissue and the interactions between cells are critical factors in overall function. This information is lost in conventional scRNAseq methods where tissues are disrupted to release the cells. For example, the distribution of pulmonary ionocytes throughout the lung may be an important factor in disease and therapy, but is currently unknown. Protein expression can be measured in tissues at single-cell resolution using mass cytometry, but with limits on sample and protein throughput. Spatial transcriptomics is also possible using slides arrayed with spatially barcoded oligonucleotides to map RNA expression, although not yet at single-cell resolution due to technical constraints. Despite the drawbacks, coupling the depth of scRNAseq from dissociated cells with the spatial information for a subset of genes allows us to build a comprehensive picture of tissue structure, including cell-cell interactions and where ligand-receptor pairs are expressed. This will aid our understanding of the signals and substrates required to make and maintain cells *in vitro*.

Not only are these methods altering our understanding of blood development, and so our strategies for directed differentiation, but they also have great potential for identifying the pathways driving unwanted or pathological phenotypes. Companies could use these data to identify target pathways to drive diseased cells back into a normal phenotype. For example, a recent study published in *Molecular Cell*[3] generated a single-cell resolution map of the liver and showed that fibrosis associated with non-alcoholic steatohepatitis (NASH), a severe fatty liver disease, results in part from secretion of "stellakines" from hepatic stellate cells. The ability to restore these cells to a normal phenotype could limit or reverse the fibrosis associated with NASH, for which there is currently no cure. Likewise, osteoarthritis (OA) is a chronic joint disease resulting in the irreversible breakdown of cartilage and bone, affecting hundreds of millions of people worldwide at a great cost to healthcare providers and with limited treatment options. Single-cell analysis of cells from OA patients undergoing knee replacements identified multiple populations of cartilage-resident chondrocytes related to disease progression[4]. Understanding the differences between these and healthy cell populations could improve our ability to modulate the disease or generate replacement chondrocytes.

Addressing Issues of Safety by Analysing One Cell at the Time

As well as generating particular cell types, many cell and gene therapy methods aim to alter a cell's genetic code to have a desired outcome. This includes chimeric antigen receptor (CAR)-T cell therapies in which a patient's own immune cells are engineered to more effectively target tumour cells – exemplified by Novartis' KYMRIAH® – and correction of faulty genes in blood stem cells, which can be used to treat severe immunodeficiency, for example. Another key target is the development of "universal cells" that evade a patient's immune response by altering or removing molecules that recognise pathogens, allowing an "off-the-shelf" supply of cells for therapy. These editing events often occur at low frequency, making it difficult to isolate and study the edited cells and the effects on them. Likewise, reprogramming one cell type to another is challenging and often only occurs in a subset of cells, although this process can be improved using a systematic approach for selecting reprogramming factors.

Applying scRNAseq to these processes allows the cells which have been targeted and the effect on their gene expression to be identified in a way that is otherwise not possible (Figure 3). Understanding these processes, and whether there are common molecular events, such as the reorganisation of the epigenome, that must occur could improve experimental strategies. There are also some concerns over the safety of such modifications, particularly where viruses are used to introduce genes into the genome, as their insertion can have detrimental off-target effects. To tackle this problem, several companies are working together to develop a single-cell viral integration assay to improve the efficiency of viral DNA transfer to target cells and to quantify the amount and location of viral DNA in the host genome. While the target is immunotherapies involving T cells, this could lead to safer and more efficient manipulation of cells for a range of therapeutic applications.

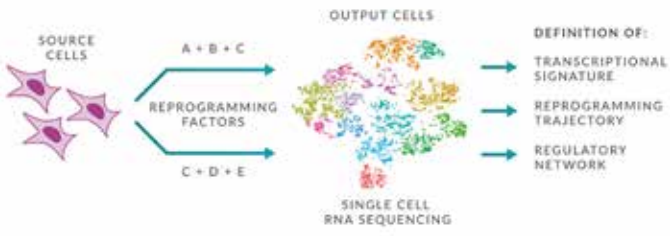

Figure 3. Single-cell analysis can be used to assess the efficacy of transcription factor combinations in reprogramming and transdifferentiation strategies

Downstream of preclinical studies, single-cell approaches could also be used to monitor and improve the long-term stability and efficacy of therapies. For example, performing scRNAseq on a sample of patient blood would enable the contribution of transplanted blood stem cells to the pool of blood cells, or the expression of a modified gene, to be assessed over time.

Many therapies are also costly and limited in throughput because a patient's own cells must be harvested, modified and transplanted. An "off-the-shelf" solution that could be rapidly administered to any patient would clearly be advantageous in many situations but suffers from a lack of ability to expand many cell types and maintain their cellular functions in culture. For example, CAR-T cell therapies may be limited by T cell exhaustion, where the repeated activation of these critical immune cells leads to their dysfunction and death. Several recent studies from Howard Chang's lab at Stanford University investigated the mechanisms underlying T cell exhaustion in very limited samples from patients before and after treatment for a specific cancer, made possible only because of the limited cell input required by single-cell technologies[5,6]. This revealed several genes and regulatory elements on the DNA correlating with the exhausted phenotype. Therefore, assessing the stability of gene expression and the epigenome over time can indicate whether differentiation and reprogramming strategies have achieved a stable cell state and what pathways regulate these processes, and can offer targets to modulate them.

The (R)evolution of Biology

The promise of cell and gene therapies is huge, particularly with approaches to reduce the guesswork involved in producing different cell types. Single-cell technologies require a shift in thinking and in the way that researchers view biological systems and problems but provide access to biology at an unprecedented resolution. A new era of biology has begun in which we can answer questions that we simply didn't have the means to address before. In time, patients will reap the benefits of this enhanced view of human biology.

REFERENCES

1. Cahan, P. et al. CellNet: network biology applied to stem cell engineering. Cell. 158(4), 903-915 (2014).
2. Rackham, O. J. L. et al. A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. 43 (3), 331–338 (2016).
3. Xuelian, X. et al. Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. Mol. Cell. 75 (3), 644–660 (2019).
4. Ji, Q. et al. Single-cell RNAseq analysis reveals the progression of human osteoarthritis. Ann. Rheum. Dis. 0, 1-11 (2018).
5. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotech. 37(8), 925-936 (2019).
6. Yost, K. E. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. 25(8), 1251-1259 (2019).

Dr Vicki Moignard

Dr Vicki Moignard is a Senior Scientist at Mogrify Ltd, a Cambridge- (UK) based company aiming to transform cell therapy by the systematic discovery of novel cell conversions. She earned a PhD from the University of Cambridge, focusing on the application of single-cell gene expression technologies in understanding blood cell development. Most recently she worked as a postdoctoral fellow at GlaxoSmithKline investigating the regulation of immune cell differentiation.

Email: vicki.moignard@mogrify.co.uk

This article has been published in Volume 2 Issue 3 of the *IBI Journal*