THOUGHT LEADER

Computational Tools for Accelerating Regenerative Medicine

By Aida Moreno-Moral, PhD

High-resolution data, computational power, and novel algorithms help realize robust strategies

Aida Moreno-Moral, PhD

Our bodies have a natural ability to heal through our repair mechanisms. In disease, these mechanisms trigger inflammation and scarring to regrow the lost tissue. However, in many cases, too many functional cells are lost and replaced, instead, with scar tissue, which can result in reduced organ function.

Cell therapy has arisen as a leading approach to tackle tissue regeneration. It involves the direct transplantation of cells to compensate for those lost. In some cases, it is possible to extract healthy cells of interest from the patient or donor and grow them ex vivo for transplantation; however, in most cases, growing cells in sufficient quantities for transplantation is difficult because our cells do not have enough growth capacity, a challenge shared across all cell therapies.

In 2007,[1] a method for converting a mature a human cell type into an induced pluripotent stem cell (iPSC), a cell that can grow and transform into different cell types, was discovered and provided a way to tackle this issue. However, the process of making iPSCs and then correctly differentiating them into the cells of interest is an inefficient and time-consuming process requiring the development of complex differentiation protocols customized for every type of target cell.

These protocols usually require precise concentrations of expensive molecules/chemicals that need to be replaced at specific time points, sometimes over several months. This process is mostly optimized by experimental trial and error and often requires many years of work, which can become very expensive.

Advantages of transdifferentiation

A more efficient solution would be to directly convert between cells of interest, without having to go through the stem cell state, avoiding the risks that this encompasses (for example, permanent activation of cancer genes) and the lengthy differentiation process. This process, known as transdifferentiation, occurs naturally in the body, but in a limited manner. Like induced pluripotency, this process can be also triggered in our

The combination of high-resolution data, computational power, and novel algorithms … is speeding the development of in vivo cell transdifferentiation and current cell therapies.

cells, although it requires the identification of a core set of genetic switches, which can push the cell from one identity to another.

Cell transdifferentiation can also be achieved in vivo, bypassing the need for transplantation altogether. This approach could be used to directly convert scar tissue cells into the cells they replaced, that is, into the original functional cell population that was lost through disease or injury.

Computational tools to drive cell conversion

To achieve cell conversion, we need to identify the set of genes that would transform one cell into another. This can be a difficult problem to solve. Our cells contain over 20,000 protein-coding genes (>50,000 if we also consider non-protein-coding genes), and in any given cell, thousands of genes are expressed at any given time. The number of possible combinations exceeds the number of stars in the galaxy, rendering even high-throughput experimental approaches impractical.

Recently, major breakthroughs were made in cell therapy when computational tools were developed, providing a data-driven solution to this problem. These algorithms[2–5] can predict the set of transcription factors that, when applied to any given cell type, will convert it into another.

An algorithm that can predict factors for cell conversion requires two main kinds of information: accurate profiles of the source and target cells, and data capturing how genes interact and regulate one another. In some cases, simply growing or harvesting enough cells to generate data makes attempting certain conversions challenging. However, the more accurately we can characterize the cell status, the better our predictions will be, and so high-throughput technologies are essential. These techniques can measure thousands of molecules at a time (usually of a single type), including DNA, mRNA levels, epigenetic marks, or protein levels.

Profiling cells: In 2013, *Nature* selected single-cell sequencing as their method of the year,[6] and after this method became mainstream, an international race began to profile every cell in the human body.[7,8] Producing data at this scale requires collaboration between several institutions across the world; however, more data does not necessarily mean better data.

Omics techniques are very sensitive, and we need to be aware of potential sources of variability in our datasets so these can be accounted for. Unfortunately, this creates barriers for the reuse of many smaller public datasets: if the experiments are not designed in a way that ensures the data is comparable

---

At Left. If we are to determine which regulatory genes led a cell to assume a certain identity, we must account for how they caused the cell to activate its unique combination of protein-expressing and non-protein-expressing genes. Doing so for any given cell is difficult because the number of possible combinations is thousands of times greater than the number of stars in our galaxy. Luckily, computational tools can help us identify regulatory genes and advance the development of next-generation regenerative therapies. Pablo Moreno Moral

between experiments, the chances are that our predictions will capture noise instead of biological signals. Large consortia, like the ongoing FANTOM (Functional Annotation of the Mammalian Genome) project, provide a good example of data generated under strictly controlled conditions and have been essential for the development of cell conversion tools.

Modeling genetic programs: On top of having a good profile of many cell types, including both source and target cells, we need a way to build models that can capture how genes interact with each other and regulate themselves. These models allow us to identify the genetic programs that are active in the source and target cells. They also allow us to select a small set of genes that can be targeted to switch cell identity.

Gene regulation requires the activation of different parts of the DNA and interaction of several types of molecules. To build gene regulatory networks, it is often best to integrate different types of omics data capturing complementary aspects of gene regulation, for instance, protein-protein interactions and protein-DNA interactions.

The aim here is to capture the behavior of cells in a comprehensive manner so we can predict how a cell will behave when we attempt to convert it into another. New techniques are constantly being developed that hopefully, one day, can solve some of the ongoing challenges in building this type of model, such as how the regulatory landscape of different types of cells changes with time and upon external stimuli.

Once a cell conversion algorithm is built with these datasets, depending on the algorithm used, we only need profiles of our target and/or source cells of interest, then the algorithm can be used to predict how to transdifferentiate between any two cell types. Furthermore, because these computational tools are designed to find regulatory molecules that can switch genetic programs, they can also be used to help reduce established iPSC differentiation protocol timelines or to convert diseased cells to healthy ones.

Accelerating regenerative therapies

The combination of high-resolution data, computational power and novel algorithms, such as Mogrify®, have made computationally guided cell conversions a reality. This approach is speeding up the development of in vivo cell transdifferentiation and current cell therapies toward shorter, safer, and more robust strategies. We can now develop regenerative medicine strategies previously seen as impossible and tackle regeneration in ways that we have not even imaged. GEN

Aida Moreno-Moral, PhD (aida@mogrify.co.uk), is principal bioinformatician at Mogrify. Website: www.mogrify.co.uk.

References available online.

Article originally published in *GEN* June 2020 Vol. 40 No. 6