SOMITE

Somite Therapeutics

Direct versus directed differentiation in generating cells for replacement therapy

Last edited: Jan. 14th, 2024

I. Introduction

The field of developmental biology, once the curiosity-driven pursuit of understanding how embryos grow and form, is now providing new and exciting opportunities to impact human health. One important example of this is harnessing knowledge of how different cell types are produced within the embryo, to generate cells and tissues in vitro that can be therapeutically delivered to patients to treat various injuries, diseases, and degenerative conditions. There are two basic approaches that can be used for this, with the confusingly similar designations “direct differentiation” (or direct programming) and “directed differentiation” (or, less commonly, guided differentiation). Before explaining the difference between direct and directed differentiation, and exploring the advantages and disadvantages of each, we start with an allegory to give an intuitive sense of the difference between them:

A scout troop of young boys and girls are on a hike through the woods. The path is fairly straight-forward, they simply have to stay on the path until the one major fork, where they are told to take the left fork and stay on it until they reach the cabin, where their scout leaders will be waiting with a warm fire and ready cooked meal. Somehow, however, the children miss the fork and stay on the right path instead of taking the left. It is miles down the path and nearing dusk when they finally realize their mistake. They are tired, and it is getting cold and dark. What should they do? They essentially have two options. They can back-track to the fork, take the correct path, and follow the prescribed route to the cabin. It will take a long time, but if they pay attention to the signposts and map, they will eventually get there, none the worse for wear. Alternatively, some in the group argue they can get there much faster by utilizing GPS and taking a shortcut, going off-trail through the woods and navigating directly to the cabin. But they do not really know the terrain they will encounter between the trail and the cabin, and cutting through it risks negative consequences (twisted ankles, poison ivy, etc.) That is not to say the second option is not worth considering, there may well be times when it is a great option for

reaching their goal. The better choice will depend on how long the paths are, how rugged the off-trail terrain, how much daylight remains, etc. But in deciding whether to follow the established path or to take a shortcut, it is important to understand the potential advantages and drawbacks to each.

2. Differentiation during development

Pathways are not just for scout troops: The changes cells go through during embryogenesis can be conceptualized as following a "developmental pathway", in which the cells change over time, in a stereotyped stepwise manner, from early embryonic progenitors to mature tissues. The path the cells follow involves a series of binary decisions between alternative cell states, progressively limiting the options of the fates the cells can ultimately adopt. As the cells move from one state to the next there are corresponding changes in gene activity. Indeed, it is the set of genes that are actively transcribed in a cell that define the type of cell it is. For each given cell state, it is important that the set of active genes be stably maintained as the cells multiply and tissues grow (until the next cell state transition). To that end, gene activity is established and reinforced by "epigenetic marks", alterations to the genome (such as DNA methylation and histone modification) that do not mutate the nucleotide sequence, but alter how the genes are expressed. These epigenetic marks are inherited as cells divide. As cells change state, they accumulate new epigenetic marks, including altering some previous marks, but to a large extent they build on the epigenetic changes they previously acquired. As epigenetic marks constrain the genes that will be expressed in a cell, the set of genes active in a particular cell is heavily influenced by its developmental history.

The differentiation pathways cells take during embryogenesis are carefully orchestrated by the signaling molecules in their environment. The combination of the stepwise progression of cell fate restriction (safeguarded by epigenetic marks) with a complex set of positively and negatively acting environmental cues, provides a high level of fidelity to the process, such that all the cells destined to become a certain tissue in an embryo (eg cartilage) make it to that fate, and the mature tissue is not diminished by partially differentiated cells (eg lingering prechondrogenic mesenchyme). Similarly, the binary decision tree (again reinforced epigenetically) assures that the cells achieve distinct and discrete cell fates (eg cartilage cells have the correct properties of cartilage, not - for example - a cell phenotype halfway between cartilage and muscle).

3. Directed differentiation

If you were to draw a roadmap of the differentiation pathways that lead to all the cell types in the adult body, the common starting point would be a pluripotent early embryonic stem cell (ESC). In landmark studies, Martin Evans and Matt Kaufman (1981), and independently Gail Martin (1981), devised protocols for isolating and maintaining ESCs in culture. In an equally important discovery, Kazutashi Takahashi and Shinya Yamanaka (2006) devised a process (through cellular "reprogramming"; more on this below) to generate ESC-like cells from other cell types (so-called "induced pluripotent stem cells", or iPSCs.) The availability of ESCs and iPSCs

opened the door to mass producing a panoply of cell types for therapeutic use. The concept is fairly simple and straight-forward. If the starting cells in your culture are functionally the same as those found in the earliest embryo, and you expose them to the exact same set of signals, in the same sequence, as they would encounter in the embryo, one should be able to recapitulate the cell state transitions that normally occur and generate any cell type one desires (reviewed in Cohen and Melton, 2011). The caveat is, for this to succeed at a high level, one needs to have a deep and detailed understanding of the steps cells go through on a particular differentiation pathway, and of the series of signals that 'direct' them along that path. Thankfully, knowledge derived from 30-plus years of modern developmental biology research provides the insights one needs to produce a number of key cell types (such as muscle satellite cells based on the long series of investigations by the Pourquie lab), and application of high-throughput genomic assays, in combination with machine learning approaches, promise to provide the intricate understanding (the "digital twin") needed to optimize protocols and establish conditions to produce any cell type normally generated in the embryo.

4. Direct differentiation

There is, however, an alternative approach, circumventing the intermediate steps in the normal differentiation process, and not requiring as intricate knowledge of the embryo; instead taking a shortcut and going directly from the stem cell to the cell type of interest. The ability of cells to convert from one cell fate to another has long been known to developmental biologists. Known as "transdifferentiation" or "transdetermination", it can be seen (for example) during amphibian limb regeneration (Brockes and Kumar, 2002) or in transplanted fly imaginal disks (Hadorn, 1968). The phenomenon has more recently been renamed "direct reprogramming", primarily because stem cell biologists have a penchant for giving their rediscoveries new names. Direct reprogramming was first achieved in an intentional way by Hal Weintraub and associates (Davis et al, 1987), showing that driving expression of a single transcription factor, called MyoD, was sufficient to transform a fibroblast into a myoblast. This works because gene regulation within a cell is often hierarchical, with one or a few key transcription factors at the top of a cascade, activating a further sets of transcription factors that ultimately turn on all the downstream genes that give a cell type its unique characteristics. Moreover, there are feedback loops between the genes at the top of the hierarchy such that activating just one (or a few) can trigger the entire cell type-specific genetic program. (For a time, such transcription factors at the top of cell differentiation hierarchies were given the unfortunate name "master regulators", although the term has thankfully fallen into disuse.) A second reason underlying the success of directly reprogramming fibroblasts into myoblasts may be that the two cell types already share certain aspects of their transcriptomes and their epigenetic states. Indeed MyoD is unable, by itself, to convert stem cells to myoblasts.

There is one more subtle, but important semantic distinction that needs to be understood to follow discussions on this topic: When transcription factor(s) are used to directly differentiate a mature cell type into a different differentiated cell type (eg. the example above of transforming a fibroblast into a myoblast, or the direct differentiation of differentiated pancreatic exocrine cells into insulin-producing ß-cells achieved through the concerted expression of three transcription factors; Zhou et al., 2008), the process is termed "direct reprogramming". In

contrast, when the same process of transforming a cell to a desired differentiated phenotype uses a stem cell (ESC or IPSC) as a starting point, it is simply referred to as “direct programming”.

Direct differentiation is a potentially powerful approach, and it has started to be applied in the context of cell replacement therapy efforts. For example, developmental studies have established that the transcription factor Pax7 is critical in establishing satellite cells in the embryo. Rita Perlingeiro and colleagues have recently shown that Pax7 can be harnessed to help drive differentiation of human iPSCs into satellite cells, in conjunction with a complex guided differentiation protocol (Kim et al., 2021). While this study made critical use of a known candidate, algorithms have been developed, integrating gene expression data and knowledge of regulatory networks, to help predict potential regulators of cell type transitions (eg. Okawa et al., 2016). To expand the approach to induce formation of any desired cell type, the laboratory of George Church produced a screenable library of every predicted transcription factor and transcription factor isoform encoded in the human genome. Screening this library, they were able to identify different transcription factors capable of directing differentiation of neurons, oligodendrocytes, fibroblasts or endothelial cells; each with near 100% efficiency in just four days (Ng et al., 2021). This was the foundational IP for the launching of the cell therapy company GCTX. Parallel work from a group headed by Mark Kotter tackled a second stumbling block to actualizing this approach: when delivering a transcription factor-encoding transgene to stem cells, one needs to insert the transgene into a place in the genome where it will not be silenced, while at the same time not disrupt an endogenous gene (a so-called “safe harbor”). Also ideally the transgene would be controllable (ie inducible) by exogenous factors. They developed such a system, and using it identified transcription factors capable of driving direct differentiation into neurons, oligodendrocytes and myogenic cells (Pawlowski et al., 2017), the foundation of another company, Bitbio.

5. The problem with shortcuts

It must be born in mind that it is still very early days in the development of both direct differentiation and directed differentiation as methods for generating cells of therapeutic benefit. Nonetheless, from where we now stand, there are some important caveats to consider regarding the use of direct differentiation in this context.

First of all, it is by no means certain that all of the cell types one would like to generate can be triggered through forced expression of a single, or a few transcription factors. It is possible, but it is equally plausible that some cell types require multiple independent gene regulatory networks operating in parallel to achieve the desired cell phenotype. It may be instructive in this regard that the two independent transcription factor screens cited above identified overlapping sets of target cell types (neurons and oligodendrocytes being 2 of 4 to come out of the Ng et al screen, and 2 of 3 to come out of the Pawlowski et al screen). There may be a strong bias or even a limitation on cell types achievable by direct differentiation. In contrast, all cell types arise during embryogenesis and hence should be able to be recapitulated by directed differentiation, once the correct set of guiding signals is identified.

A second caveat that needs to be considered regarding direct programming is that, unlike guided differentiation, direct differentiation involves the genetic modification of cells, which always carries risks associated with potential oncogenic transformation. This concern can be partially ameliorated by utilizing "safe harbor" sites for the integration of transgenes into the genome (minimizing the possibility of unintended insertional mutagenesis), and preclinical safety testing of the in vivo potential of the transcription factors being used. Nonetheless, the approach needs to be applied with caution, particularly in settings where some or all of the cells may not be fully programmed to the target phenotype.

A third issue is how closely the phenotypes of directly programmed cells actually match those of the relevant endogenous cell type they are targeting. While the studies cited above assessed expression of multiple markers of the cell types they were addressing, they did not examine the transcriptome of individual cells to see the extent and efficiency of full conversion. In a meta analysis of multiple studies of direct differentiation and directed differentiation, it was repeatedly observed that cells derived by directed differentiation more closely approximated the phenotype of the target cells than those generated by direct differentiation (Cahan et al., 2014). In particular, while the gene network characteristic of the differentiated cell type was generally activated, in many cases of direct differentiation, the gene network of the starting cells was only partially, or not at all downregulated. This is not surprising in light of what we know about epigenetic regulation during differentiation (discussed above). Of course, these differences may, or may not be important functionally, likely dependent on the specific cell type in question, and what one is asking the cells to do therapeutically. The fact that such differences arise in the context of direct differentiation is consistent with what we know about the best characterized directly reprogrammed cell type, the iPSC.

iPSCs (themselves reprogrammed from fibroblasts by the expression of four transcription factors) are an exceedingly powerful tool, and their success is often given as proof of principle for the direct differentiation approach. However, their generation requires a fairly long process (3-4 weeks for human iPSCs) allowing progressive remodeling of epigenetic marks to a configuration approximating ESCs (much longer than the 4 days taken for direct differentiation of other cell types). Moreover, the process of producing fully functional iPSCs is very inefficient, with 0.01-0.1 percent of the cells being fully converted (Ghaedi and Niklason, 2016). Of course the criteria applied to iPSCs are quite stringent, demanding that the reprogrammed cells be truly pluripotent. As complete conversion may not be necessary for some targeted differentiated cell types, but when bulk cultures of "100% reprogrammed cells" are utilized, some caution must be taken regarding the possible rogue activity of the inevitable partially converted cells in the culture.

Finally of less current importance, but perhaps of greatest significance in the long run, generating individual cell types is just the tip of the iceberg, albeit the low hanging fruit. The holy grail for regenerative medicine is not replacement cells, but replacement tissues and organs. And you cannot get there one reprogrammed cell type at a time, or even by generating and mixing multiple cell types together in the right proportions. For functional anatomy, cells need to be in the right relative positions, with correct orientations, proper intercellular

connections, etc. And the most straightforward way to produce such detailed complexity is through recapitulation of the developmental pathways through which such complex tissues normally form in the embryo. This advantage of the directed differentiation approach can be seen in the generation of muscle satellite cells. Satellite cells quickly lose their-self renewing capacity in culture (Charville et al 2015; Montarras et al 2005). However, when satellite cells are produced through directed differentiation, they are generated in juxtaposition with their normal support cells, which act to maintain their potency as the satellite cells are expanded in culture (Chal et al 2016).

In summary, direct differentiation is a potentially powerful approach, and it is great for the field and for patients that both direct and directed differentiation are being actively explored for generating therapeutic cell types. However, in establishing a platform for generating cells for therapeutic use in 2024, directed differentiation is likely to be the more efficacious and more versatile method.