Stem cells and the evolving notion of cellular identity

George Q. Daley

Stem cells are but one class of the myriad types of cells within an organism. With potential to self-renew and capacity to differentiate, stem cells play essential roles at multiple stages of development. In the early embryo, pluripotent stem cells represent progenitors for all tissues while later in development, tissue-restricted stem cells give rise to cells with highly specialized functions. As best understood in the blood, skin and gut, stem cells are the seeds that sustain tissue homeostasis and regeneration, while in other tissues like the muscle, liver, kidney and lung, various stem or progenitor cells play facultative roles in tissue repair and response to injury. Here, I will provide a brief perspective on the evolving notion of cellular identity and how reprogramming and transcription factor-mediated conversions of one cell type into another have fundamentally altered our assumptions about the stability of cell identity, with profound long-term implications for biomedical research and regenerative medicine.

1. Background and history: from cells to stem cells

In his groundbreaking seventeenth century treatise ‘Micrographia: or, Some physiological descriptions of minute bodies made by magnifying glasses’ (London: J. Martyn and J. Allestry, 1665) Robert Hooke described his visualization of the tiny walled units of a thinly sliced film of cork, which he likened to a honeycomb, and dubbed them ‘cells’. Some 350 years hence, cells are well known as capable of living free as whole organisms or as building blocks of living tissues within highly complex beings, and as remarkably malleable and resilient fundamental units of life.

The twentieth century, arguably the century of genetics, began with the elucidation of genes as units of heredity, mid-way through witnessed the discovery of the double helix, and concluded with the completion of the rough draft of the human genome sequence. Although the genome sequence provides the genetic code, the principle research challenge of the twenty-first century is learning how this code is read and interpreted by cells—how the information in the single-celled human zygote is translated to produce the trillion-odd specialized units within the human body. Developmental biology addresses many of the questions inherent to formulating the body plan, but deciphering the expression of the genetic code is increasingly the provenance of stem cell biology, which seeks to understand epigenetic regulation and cell fate. Just as the gene is the principle sub-unit of the genome, the stem cell is the chief unit of organization for tissue formation and homeostasis.

The concept of developmental hierarchies of cells, with a stem cell at the top of this hierarchy and multiple differentiated offspring derived from it, first emerged in the late nineteenth century and has become an experimental mainstay of the last 65 years, owing to the pioneering definitions of self-renewing multipotential progenitors of all lineages of the blood by Till and McCulloch in a series of pioneering papers [1,2]. In delineating the stem cell for the blood-forming tissue the haematopoietic stem cell (HSC), Till and McCulloch set the stage for understanding the developmental hierarchies of additional tissues and organs—the skin, gut, lung, sperm and others—as well as the...
emergence of whole organisms from the pluripotent stem cells of the early embryo. To ascribe their unique properties, a stem cell must be defined clonally—at the single-cell level—for its ability to self-renew and differentiate, and consequently stem cells represent but a tiny fraction of cells within the tissues of adult organisms.

2. Developmental potency of stem cells

Stem cells are defined by their range of potential fates—a concept termed ‘developmental potency’. The zygote is the most developmentally expansive cell, yet is rarely considered a stem cell in mammals because it cleaves into blastomeres of equal developmental potency for at most three cell divisions, and therefore manifests very limited self-renewal potential. The zygote and early blastomeres form all the tissues of the embryo proper as well as the supportive extraembryonic tissues—the fetal membranes and placenta—a unique capacity called totipotency. To date, no one has succeeded in propagating the zygote or blastomeres continuously in cell culture. Embryonic stem cells (ESCs), by contrast, can be isolated from the inner cell mass of blastocysts and cultured as immortal cell lines. When murine ESCs are aggregated with early blastomeres or returned to the blastocoel cavity of the blastocyst by microinjection, they chimaerize all tissues within the embryo proper but typically not the placenta, which reflects their developmental segregation from the trophodermal lineage, which has its own master cell—the trophoblastic stem cell. ESCs are considered pluripotent. The developmental potency of human ESCs, which for ethical reasons cannot be determined by embryo chimaerism, is instead assessed by observing the spectrum of tissue differentiation that results after subcutaneous injection of cells into immune-compromised mice. Pristine ESCs produce teratomas, encapsulated tumours consisting of disorganized masses of differentiated tissues. Because elements of the three primitive embryonic germ layers—ectoderm, mesoderm and endoderm—coexist in these tumours, human ESCs share pluripotency with mouse ESCs.

When human ESCs were first described by Thomson et al. [3], there was an oddity that went largely unnoticed for almost a decade. The human-derived cells grew under distinct conditions, stimulated by fibroblast growth factor instead of LIF, which nurtured the mouse ESCs. Human ESCs appeared not to depend on LIF signalling, nor to activate the downstream pathways driven by the LIF/gp130 receptor complex, implying a distinct physiologic or developmental state [4]. Moreover, the colonies of human ESCs appeared flatter and were more prone to die when passaged as single cells than the mouse ESCs, which grew as domed colonies and could be readily dissociated and passed as single cells. Only when stem cells from the epiblast stage of mouse development were isolated and cultured did the distinctions become clear: human ESCs resembled more closely the epiblast-derived stem cells (EpiSCs) of slightly later stage embryos. Hence, murine EpiSCs appear to be a closer developmental equivalent to human ESCs than murine ESCs [5].

Vigorous recent efforts to isolate human ESCs under conditions that more closely mimic the mouse ESCs have yielded a large number of papers that report slightly different conditions and states of a putative ‘naive’ human stem cell. Such a cell, like the mouse ESC, is closer in its DNA methylation and gene expression patterns to the cells of the inner cell mass of the blastocyst, and hence is termed ‘naive’ [6,7], in contrast to the cells that more closely resemble the epiblast stage of mammalian development, like human ESCs or mouse EpiSCs, which are considered ‘primed’, as if poised to differentiate. Indeed, several groups have claimed to have identified combinations of transgenes or purely chemical means to convert human primed pluripotent stem cells to their naive state. If such naive human ESCs can indeed be introduced as an alternative to classically derived hESCs operating since the first lines of Thomson, there may be advantages in terms of growth rates, ease of passage from plate to plate, and the receptivity to efforts at genome editing. Thus, the search for naive hESCs represents a ‘holy grail’ of stem cell biology and promises many interesting new insights into the metastable or heterogeneous state of pluripotent stem cells [8].

3. Multipotential stem cells

While ESCs garner considerable interest due to their multi-faceted developmental plasticity, and may one day prove invaluable as a source for autologous, personalized transplants of cells for tissue repair or regeneration, the best established existing cell therapies typically involve another class of cells deemed ‘adult’ or ‘somatic’ stem cells. Unlike embryo-derived stem cells, tissue-restricted somatic stem cells are limited in their potency to the cell types of the tissue in which they reside. HSCs generate blood, gut stem cells the lining of the intestines and skin stem cells the epithelial lining of the body. Despite earlier claims of greater plasticity, somatic stem cells are highly constrained and do not without considerable genetic or chemical manipulation trans-differentiate into foreign cell types or tissues, and do not represent ‘repair kits’ for tissues other than their own. The most widely exploited somatic stem cell—HSCs—can be transplanted into patients with various blood diseases, malignant and genetic, after the patient’s diseased blood-forming system has been eradicated by high-dose chemotherapy and irradiation. But blood cells will not readily regenerate or repair other tissues like the heart, brain, liver or kidney after they have been injured or afflicted by disease. Consequently, multipotent stem cells represent highly valuable sources for regenerative therapies within adult tissues, but on account of their lineage-restricted nature, do not provide a source for all potential tissue therapies in the adult. For tissues that do not readily regenerate from stem cell pools in the adult—like the heart, kidney, pancreatic beta cells and much of the brain—only pluripotent stem cells directed to differentiate along specific lineages that recapitulate embryonic development provide hope for future cell replacement therapies.

4. Reverting differentiated somatic cells to pluripotency

During much of history since the time of Hooke, cell biologists have considered development a one way thoroughfare, with the range of a cell’s fates becoming progressively restricted as the zygote progresses from a totipotent state to the pluripotent state of the inner cell mass of the pre-implantation blastocyst, and ultimately devolves into the more limited potential of tissue-restricted cells. Ultimately, most cells in the adult organism attain a fully differentiated state, which remains stable and immutable unless affected by pathologic...
5. Diverting somatic lineages directly to alternate cell fates

Inspired in part by the classical experiments of Weintraub, Lassar and colleagues which exploited ectopic expression of the MyoD TF to convert fibroblasts to muscle cells [17,18], Yamanaka’s TF-based approach to reverting somatic cells to pluripotency also provoked a number of laboratories to test whether TFs might drive conversions of one somatic cell type directly to another, a process called transdifferentiation or direct conversion. Among the first laboratories to demonstrate this phenomenon was that of Doug Melton, who succeeded in converting exocrine pancreatic cells to insulin-producing beta cells by direct injection of mouse pancreas with master regulators of beta-cell fate [19]. Wernig’s group later converted mouse and human fibroblasts to neurons using a similar strategy, only that TFs were engineered into fibroblasts in vitro [20,21]. Subsequently, a flood of papers have reported conversions of one somatic cell type to another via this approach (reviewed in [22]). Indeed, these experiments have further extended the concept that a cell’s molecular composition and thus function are malleable and can be altered for applications in research and potentially for therapeutic purposes. But whether such cells function equivalently to their native counterparts has remained a lingering question among stem cell biologists.

6. CellNet: a metric for discerning cell identity and enhancing cell engineering

Numerous laboratories within the field of stem cell biology are consumed by attempts to convert various cell types in vitro into distinct altered states: whether directing pluripotent stem cells to differentiate along specific lineages towards specialized terminal cells, directing the conversion of one somatic cell type to another with enforced expression of TFs, or reverting differentiated somatic cells to pluripotency by expression of the Yamanaka factors. Such efforts promise to validate cultures of human cells as improved models of human pathology and physiology, and raise hopes that cell-based target validation, drug-screening and mechanistic research will be strengthened. Of some concern however, are the standards by which cells engineered in vitro are compared to native tissues for both molecular identity and physiology. Most claims that cells engineered in a dish have achieved a given cell fate equivalent to native tissues are typically founded on analysis of a small number of diagnostic markers, on limited functional assays and on global assessment of gene expression analysed by clustering algorithms that provide at best a qualitative suggestion of similarity without an objective metric of identity.

Given the momentous challenge of assessing the quality and fidelity of cells engineered in vitro, and the lack of rigorous metrics for assessing cell states, my laboratory has worked closely with that of James Collins (MIT) to generate a computational platform that defines how closely an engineered cell approaches the identity of a target cell or tissue type, and provides a set of candidate regulators of the cells’ transcriptional state that require further modulation to push the engineered cells ever closer to their target. CellNet is a publicly available network-biology platform (http://cellnet.hms.harvard.edu/) written by Patrick Cahan with input from Hu Li and Edroaldo Lummertz de Rocha that has been tested extensively against both published datasets [23] and used within our own laboratory to enhance the in vitro production of several cell types [24]. At its core, CellNet is founded on the notion that cell- and tissue-specific gene processes like metaplasia or frank malignancy. While today we comprehend that virtually all cells, save lymphocytes, retain the full complement of DNA as all other cells, and become specialized by virtue of expressing certain subsets of genes while silencing others, early thinkers like August Weissman speculated that development proceeded by a series of ‘qualitative divisions’ in which daughter cells differ as each inherits a different set of factors that specify function. Later Hans Spemann proved through a series of painstaking and ingenious micromanipulations of early newt blastomere nuclei that all had equal potential to generate viable organisms, establishing the paradigm-shifting notion of nuclear equivalence. Spemann envisioned but never succeeded in performing his ‘fantastical experiment’—transplanting the nucleus of a highly differentiated cell back to the egg to test whether it would regain and manifest embryonic potential. Reporting success in precisely that fantastical experiment, Briggs and King showed in 1952 that when nuclei were transferred from the blastula stage cells of a frog embryo into a fresh egg, a third resulted in swimming tadpoles, showing that later embryonic stage cells showed nuclear equivalence. Ultimately, John Gurdon proved the full developmental potential of highly differentiated intestinal cells of the frog in generating functional organisms, thereby unequivocally establishing the principle of nuclear equivalence throughout development. Somatic cell nuclear transfer (SCNT) ultimately succeeded in proving the full developmental competency of the nuclear genome of mammalian cells of the sheep [9], and now well over a dozen distinct mammalian species. Indeed, SCNT together with directed differentiation of murine ESCs represents a classical paradigm for customized gene and cell therapy to treat genetic disease, as has been shown in murine models [10]. While nuclear transfer has ultimately succeeded in humans [11], it has proven quite cumbersome, requiring harvesting of fresh oocytes from women willing to donate to research.

The impractical technical and controversial aspects of human SCNT were bound to limit its impact. Indeed, in a major paradigm-shifting demonstration that transcription factors (TFs) play central roles in specifying cell fates, Yamanaka and colleagues [12] demonstrated that ectopic expression of four TFs normally expressed in ESCs (Oct4, Sox2, KLF4 and c-MYC) were sufficient to reset a somatic cell’s differentiated state back to pluripotency. Within little more than a year, Yamanaka’s group as well as Thomson’s and my own applied TF reprogramming successfully to human cells, achieving the derivation of induced pluripotent stem cells (iPSC; [13–15]). We used the reprogramming technology to assemble the first repository of patient-derived iPSC representing a spectrum of genetic disease, including immune-deficiency, Down’s syndrome and Huntington’s disease, among others [16]. Indeed, Yamanaka-style reprogramming has revolutionized the field of stem cell biology, rendering the notion that differentiated cellular state can be efficiently reprogrammed to pluripotency, ushering in a profound set of possibilities for modelling human disease and developing new drug and cell-based therapies.
regulatory networks (GRNs) are major molecular determinants of cell identity that govern both the steady-state transcriptional programme as well as cellular responses to the environment and to various contextual perturbations like disease and ageing. We reasoned that defining the status of GRNs within cells engineered in vitro would provide a robust metric of cell identity as well as a means of determining which TFs were most dysregulated, thereby providing a plausible set of candidates to be further manipulated in future experiments. Using publicly available microarray expression data and a modified version of the Context Likelihood of Relatedness algorithm [25], Cahan produced a global GRN that defines regulatory relationships among all annotated transcriptional regulators and target genes across most cell types, and then refined these relationships into cell and tissue type-specific GRNs that could be used to build a cell-type classifier. Other metrics included quantitative assessments of GRN status and a network influence score, which ranks TFs according to their effect on the GRNs.

With CellNet in hand, we evaluated the outcomes of 226 experimentally derived cell populations drawn from 56 published studies, and gleaned several overarching lessons. First, virtually all reprogrammed cells—i.e. those reverted to pluripotency by virtue of Yamanaka-style reprogramming—achieved near identity to the gold standard ESC, whether mouse or human. Second, efforts at directed differentiation of pluripotent stem cells (either ESCs or iPSCs) using morphogens and selective culture conditions in vitro achieved on average higher classification scores relative to their target tissues than experiments that endeavour to directly convert one differentiated somatic cell type directly into another via ectopic expression of master regulatory TFs [23]. These analyses indicate that reprogramming is a remarkably robust transition in cell fates, perhaps in part due to the highly selective culture conditions established for pluripotent cell types. However, when making neurons or cardiomyocytes or hepatocytes from pluripotent cells, we have much to learn before we can claim great success in deriving organoids or tissues at multiple stages of development. Furthermore, the current training data derive largely from parenchymal tissues and whole organs, which encompass many cell types, while most engineered cells are aclimated to culture in a dish. Despite these limitations, when primary cells are cultured from whole organs or tissues, purified, and then subjected to analysis via CellNet, the de-differentiation effects of cell culture and the stability of CellNet as a robust classifier of a cell’s identity [23]. CellNet will be augmented in future versions as transcriptional profiles become available for additional target cells and tissues at multiple stages of development. Moreover, as more researchers turn to RNA sequencing as a mode of transcriptional profiling, CellNet will need to be adapted to accept these data (and such software refinement is underway). At the theoretical limit, as RNA sequencing of single cells becomes more robust, diagnostic GRNs might be discoverable for every cell type at every stage of organismal development, enabling algorithms like CellNet to define the transcriptional roadmap for human development. Such a platform would greatly facilitate efforts at directed differentiation of cell types in vitro and accelerate the capacity for stem cells to be translated for research and medical applications.

7. Prospects for the future

The definition of the cell has been refined at ever greater resolution since its first description by Hooke 350 years ago. In recent years, the capacity to reprogram and redirect cells such that they morph from one identity to another has highlighted the highly plastic and malleable nature of the genome and the fluid state of cells. Indeed, cells have assumed a chameleon-like character. Understanding the precise molecular mechanisms that define the formation, stabilization and transitions among GRNs promises to inform future efforts to engineer cells more faithfully in vitro and to diagnose and treat cellular pathology and promote repair and regeneration in vivo. Moreover, understanding GRNs as modules of cellular function offers the possibility that new types of hybrid cells with highly engineered functions might be produced for biomedical...
applications, the stuff of synthetic biology. Why be constrained to making a perfect hepatocyte or struggle to make a beta cell when there might be advantages to engineering a hepatocyte to perform glucose-responsive insulin secretion? It is provocative to consider what the cell will look like in another 350 years.

Competing interests. The author is a member of the scientific advisory boards and receives consulting fees or holds equity in the following biotechnology companies: True North, Solasia, Epizyme, Verastem, Ocata, Raze and MPM Capital. The author is an affiliate member of the Broad Institute, and an investigator of the Howard Hughes Medical Institute and the Manton Center for Orphan Disease Research.

Funding. The author is supported by grants from the US National Institutes of Health: R24-DK092760, RC4-DK090913, and U01-HL100001 Progenitor Cell Biology Consortium; Alex’s Lemonade Stand; Doris Duke Medical Foundation and Ellison Medical Foundation.

References