Cell Regeneration 6 (2017) 1–7

Contents lists available at ScienceDirect

Cell Regeneration

journal homepage: http://www.keaipublishing.com/en/journals/cell-regeneration/

Review

# Genomic and molecular control of cell type and cell type conversions

Xiuling Fu, Fangfang He, Yuhao Li, Allahverdi Shahveranov, Andrew Paul Hutchins*

*Department of Biology, Southern University of Science and Technology of China, Shenzhen 518055, China*

ARTICLE INFO

*Article history:*
Received 7 August 2017
Received in revised form
6 September 2017
Accepted 18 September 2017
Available online 22 November 2017

*Keywords:*
Cell type
Transcription factor
Epigenome
Transdifferentiation

ABSTRACT

Organisms are made of a limited number of cell types that combine to form higher order tissues and organs. Cell types have traditionally been defined by their morphologies or biological activity, yet the underlying molecular controls of cell type remain unclear. The onset of single cell technologies, and more recently genomics (particularly single cell genomics), has substantially increased the understanding of the concept of cell type, but has also increased the complexity of this understanding. These new technologies have added a new genome wide molecular dimension to the description of cell type, with genome-wide expression and epigenetic data acting as a cell type 'fingerprint' to describe the cell state. Using these genomic fingerprints cell types are being increasingly defined based on specific genomic and molecular criteria, without necessarily a distinct biological function. In this review, we will discuss the molecular definitions of cell types and cell type control, and particularly how endogenous and exogenous transcription factors can control cell types and cell type conversions.



## 1. Defining cell type

The cells of an organism are made up of a limited number of 'cell types' that are reused in different tissues and combine to form organs and systems. For example macrophages, phagocytic immune cells, are found throughout the body,[1] as are connective fibroblast cells.[2] However, defining a cell type, especially now that single cell technologies are revealing ever more heterogeneity between cells,[3] is challenging, and it remains unclear how many cell types there are, or exactly how fine the differences are that demarcate two cell types. There have been several estimates for the total number of cell types in an organism, with numbers ranging from between hundreds to thousands of distinct human cell or subcell types. Classical taxonomic approaches estimated the number of cell types in a selection of chordata as between 99 and 122,[4] and around 200 cell types in humans.[5] Systematic attempts to count cell types, using a variety of techniques, particularly newer gene expression data, generally come to a much higher number of cell types. CELLPEDIA is a human annotated database of cell type, based mainly on taxonomy, gene expression data and text mining of publications, it suggests 2260 taxonomic categories for cell types.[6] CELLPEDIA also uses tissue location to define cell type, which may inflate the total. However, the same 'cell types', isolated from different tissue locations, can show radically different gene expression patterns,[1,2] hence tissue location can also be an important determinant of cell type. CellFinder takes a different approach, using a mixture of database amalgamation, text mining, and human annotation, it comes to a total of 1058 human cell types,[7] and readily concedes there are many more cell types to discover. Cell Ontology (CL) describes 2200 'classes' of cell or sub-cell type, and, like the related CellFinder and LifeMap databases uses cell type definitions to map the cell types into a hierarchical model of development.[8,9] These newer studies put the total number of cell types considerably higher than previous estimates, and the true number of cell types seems to be increasing as researchers develop new tools to more accurately map gene expression and the epigenetic status of cells.

## 2. Identification of different cell types in the immune system

An illustrative example of how improvements in technology can drive the discovery of cell types is the proliferation of new immune cell types along the T cell lineage (Fig. 1). Initially defined by morphology alone, T cells were indistinguishable from B cells, and were labelled simply as 'lymphocytes', i.e. cells that occupy lymphoid tissue, but they had no known function.[10] Later, they

* Corresponding author.
*E-mail address:* andrewh@sustc.edu.cn (A.P. Hutchins).

Peer review under responsibility of Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences.

https://doi.org/10.1016/j.cr.2017.09.001

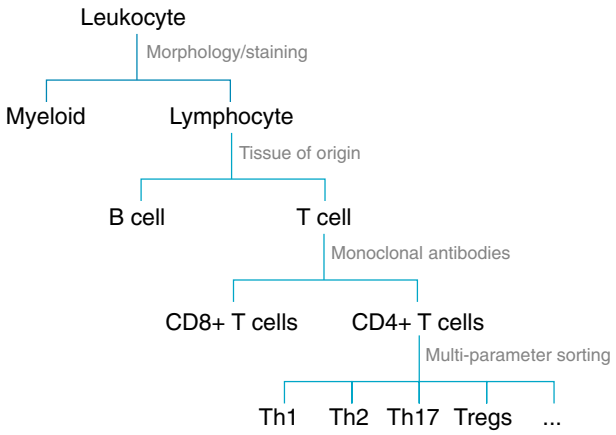


**Fig. 1.** Gradual refinement of the definition of CD4+ T helper cell types. Schematic of the refinement of the CD4+ T helper cells, from the original cell type 'leukocyte', through to a plethora of distinct Th (T helper) cell types. The technology/technique used to separate the cell types is indicated in grey at the branch point.

were discriminated based on their tissue of origin; bursa of Fabricius-derived lymphocytes (bone marrow-derived in mammals) became B cells, and thymus-derived lymphocytes became T cells.[11,12] However, B and T cells only became recognized as distinct cell types in the 1960s as B cells were definitively identified as the source of the humoral (i.e. antibody) immune response,[11] whilst T cells were initially recognized as 'B cell helpers' a few years later.[13] The widespread adoption of monoclonal antibody technology led to a burst of activity in defining further T cell types. The cluster of differentiation (CD) antibodies,[14] are a set of defined monoclonal antibodies against a variety of cell surface targets. Two CD antibodies can separate T cells into two distinct cell types: CD4+ T helper cells and CD8+ T cytotoxic cells. T helper cells play a supporting role in immune responses, whilst T cytotoxic cells perform cytotoxic killing of virus-infected cells, importantly, their cell morphology is basically identical and they can only be discriminated by their biological activity and cell surface markers. Further application of monoclonal antibodies and careful flow cytometry experiments divided T helper cells into a wide range of other T helper cell types.[15] For example, naïve T helper cells, that have not encountered their antigen are defined by the absence of CD25,[16] whilst experienced (those that have encountered their antigen) T helper (Th) cells differentiate into four major types, namely, Th1, Th2, Th17, and Tregs (regulatory T cells), along with many more less well characterized T helper cell types.[15,17] Importantly, these cell types are not just finer definitions of subpopulations, but each T helper cell type has a distinct biological function. The four best characterized T helper cell types are Th1, Th2, Th17 and Treg cells, which are important in responding to intracellular pathogens, helminth infection, extracellular pathogens, and maintaining self-tolerance, respectively.[18] However, many more T helper cell types have been discovered (e.g. Th9, Th3, TR1, Th22, Tfh, Thab, nTreg, etc.),[15,19] these new T helper cell types have less clear biological roles, but take part in a range of specific activities, including airway inflammation, allergic reactions, B cell responses and immune-related diseases, amongst other roles.[20]

T and B cells were originally defined based on the organ they were first purified from, and the tissue of origin can have a strong influence on cell type. For example, gene expression microarrays of macrophages purified from different tissues showed greater overall variation in gene expression patterns, when compared to other immune cells,[1,21] or compared to just other lymphoid cells.[22] Dendritic cells (DCs), antigen-presenting cells of the immune system, highlight the opposite problem of separating cell types. DCs and macrophages are challenging to experimentally separate accurately,[23] as they share many of the same cell surface markers. Consequently, there is argument about the difference between macrophages and DCs, and a model has been put forward that suggests DCs and macrophages are a 'spectrum' cell type, with phagocytic cells (macrophages) on one end and antigen-presenting cells (DCs) on the other, with several cell types sitting in the middle of the spectrum, each possessing more or less macrophage or DC character.[24,25] Molecular characterization suggests that, from the perspective of gene expression at least, macrophages and DCs can be distinguished based on a unique gene expression signature,[21,26] and macrophages and DCs respond differently to inflammatory stimuli.[23] Yet, arguments over the differences between these cell types remains.[24,27–29]

## 3. Heterogeneity in embryonic stem cells; defining cell type by biological function

One of the better studied cell types are mouse embryonic stem cells (mESCs), which are derived from early embryos, and maintain the ability to regenerate a full mouse.[30] Although mESCs have many similarities with the inner cell mass (ICM) of the early blastocyst, particularly in the activity of key transcription factors such as OCT4, SOX2, KLF4 and NANOG,[31] there remains debate about their exact origin and cell type,[32] as when the ICM converts to mESCs the cells undergo many gene expression changes.[26,33] mESCs as a cell culture were thought to be relatively homogenous, yet careful study of mESCs revealed small numbers of cells in a typical cell culture with altered gene expression profiles.[34,35] In mESCs, the expression level of the essential pluripotency gene *Nanog*[36,37] naturally fluctuates, and about 5–20% of mESCs express very low levels.[37–39] In culture, mESCs cycle *Nanog* on and off, which helps prime mESCs to differentiate,[39] and so these cells have a distinct phenotype and arguably cell type. *Nanog* is by no means the only example of heterogeneity in mESCs. STELLA, a marker of primordial germ cells, is expressed in 20–30% of mESCs, and those cells with STELLA more closely resemble the ICM, whilst those without STELLA express developmentally later epiblast-specific genes.[40] Indeed, there are multiple cell types contained within a typical mESC culture, including small numbers of cells with radically different biological function. Normally, mESCs very rarely contribute to extraembryonic tissues, such as the trophectoderm (placenta) or primitive endoderm.[30,41] However, mESC cultures contain about 15% of cells that are *Hhex*+ (a homeobox protein that specifically marks endoderm), and these cells can contribute to extraembryonic tissues in mouse chimeras.[41] Although the *Hhex*+ and *Hhex*-mESC's gene expression signature is nearly identical,[26] they have different biological potential, and so can be considered a distinct cell type. One caveat is that these *Hhex*+ cells still contribute to the epiblast and embryo proper, so it is not a pure population of cells. A rarer subset of cells within mESC cultures express the endogenous retrovirus MERVL. MERVL is specifically expressed at the 2 cell stage of embryonic development,[42,43] and using a MERVL-Tomato reporter, the ~2% of mESCs that express MERVL can contribute to extraembryonic tissues,[43] although again, the MERVL+ cells can also contribute to the embryo proper, and the cells can interconvert between MERVL+ and MERVL- cells,[43] suggesting instability in their cell type. It was initially thought that these MERVL expressing cells closely resemble the 2 cell (2C) stage of the embryo, where MERVLs are also specifically expressed,[43] however, recent single cell RNA-seq data suggests these 2C-like cells may more closely resemble the blastocyst, so their ultimate identity remains unclear.[35] Ultimately, the relationship between all of these heterogeneous cell types or sub-cell types within mESC cultures remains unclear. For example, despite their capability of both 2C-like and

*Hhex*+ cells to contribute to extraembryonic tissues, it is unclear how they are related to each other, along with other potential cell types revealed by single cell genomics.[35]

Ultimately mESCs are an *in vitro* artifact, a 'trapped' version of the blastocyst ICM that can grow indefinitely, but still maintain pluripotency. It is possible to capture many additional embryonic cell types, of which some appear to represent earlier timepoints in the developmental process. One such cell type are 'Extended pluripotent stem cells' (EPCs), that can contribute to extraembryonic tissues, and have distinct gene expression compared to mESCs.[44] Other embryonic cell types appear to be developmentally later than mESCs, such as Epiblast stem cells (EpiSCs), that more closely resemble the developing epiblast and have a primitive endoderm-like gene expression signature,[45,46] and lack Esrrb activity.[47] The similar but distinct EpiLCs (epiblast-like cells), lack the primitive endoderm gene expression signature found in EpiSCs, and are instead biased towards a primordial germ cell fate.[48,49] Finally, region-selective EpiSCs (rsEpiSCs) are biased to colonize just the posterior part of the developing embryo, suggesting an even later developmental phenotype than EpiSCs.[50] These and other embryonic cell types indicate that at specific stages, with the right conditions, transient cell types can be captured and maintained *in vitro*.[51]

## 4. A continuum of cell states in the transitions between cell types

It is challenging to define at what point two cell types are distinct, and where two cells are simply at one end of a continuum. Single cell data suggests that cells can transit through stages where the cell type-signatures of both origin and destination cells are simultaneously present. For example, in developing lung, some cells express markers for both alveolar type 1 and 2 cells simultaneously,[52] and early in the embryo some cells simultaneously express genes for the primitive endoderm and epiblast.[53] Consequently, identifying cell types in developmental processes is challenging. Potentially there is a continuum of expression as cells pass through developmental stages, and at any one point along that process the cell is not stable and may collapse into a more stable and distinct cell type (Fig. 2). Single cell mass spectrometry of cell surface markers in developing human B cells revealed a continuous spectrum of B cell development stages, rather than specific barriers,[54] something similar was seen for *in vitro* differentiation of cells to neurons,[55] and in *in vitro* transdifferentiation of cells to myoblasts.[56] This calls into question the existence of cell types

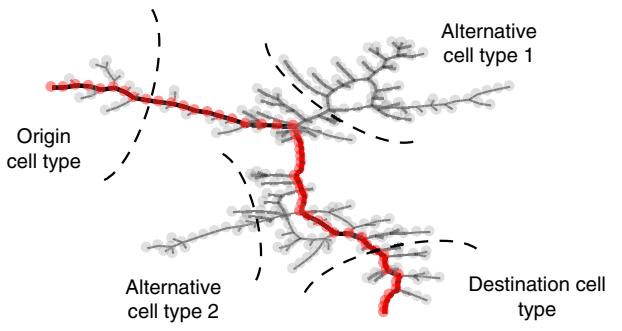


**Fig. 2.** Cells traverse pathways from origin cell types to destination cell types. A hypothetical map of cell fate conversion between an origin cell type and a destination cell type. Each node in the network is a new cellular state, and each edge is a transition between a cell state. Only parts of the network can form stable cell types, and many branching pathways exist. As the cells differentiate they move through intermediate stages, each step with a slightly different gene regulatory network underlying the cell state. When the cell reaches its destination, it becomes locked into that cell type, and can no longer traverse the intermediate states. Figures were drawn using glbase.[104]

during development and, instead of development proceeding in jumps across energy barriers to local energy minima (or distinct cell types), cells develop in a continuous manner with intermediate stages where cells can continue to choose their developmental outcome (Fig. 2). Crucially, as cells differentiate to alternate cell types they lose developmental potential, and consequently most, if not all, adult cells cannot transdifferentiate.[57] There appear to be many epigenetic blocks that lock cells into a specific cell type and limit the cells capability to dedifferentiate and transdifferentiate.[58] A major candidate for the control of cell type is transcriptional control, which may act to lock cells into a cell type.

## 5. Transcriptional control of cell type

Cell type is thought to be controlled through the activity of transcription factors (TFs), that respond to either internal or external cellular cues.[59] TFs bind to DNA and regulate gene expression, and interact with local chromatin to control cell type. Although a comprehensive model describing exactly how TFs perform these feats remains frustratingly elusive.[59,60]

TFs can be expressed in both a cell type-specific and cell type-independent manner. Many, about 60%, of TFs are cell type-specific.[61] Cell type-specific TFs can function as 'master regulators', a class of TF that can specify cell type in the absence of any other activity. The prototypical example is MyoD (*Myod1*), which when overexpressed converts fibroblasts to myoblasts,[62] activating an entire gene expression program in the absence of specific external cues (Fig. 3A).

However, a single master regulator for each cell type seems to be a relatively rare phenomenon, and often the same TF can act in multiple cell types. For example, knocking down *Gata3* in mouse embryos leads to a failure to establish mature blastocysts, likely due to a trophectoderm defect,[63] as when *Gata3* is overexpressed in mESCs it drives them to a trophectoderm cell fate.[64] Yet, despite its importance in the early embryo, *Gata3* is also a critical factor in the specification CD4+ Th2 cells.[65] A further difficulty with the idea of master regulators is tremendous degeneracy in the DNA sequences that individual TFs use to bind to DNA. For example, the homeodomain TFs all bind to a similar version of the same sequence of DNA,[66] despite the involvement of homeodomain proteins in a wide range of developmental processes. This is not restricted to just one family of TFs, as almost all TF families bind to very similar DNA motifs,[67] leading to the vexing issue of finding cell type-specific activity between different TFs that bind to the very similar sequences of DNA. One solution is for pairs (or more) of TFs to combine together to specify a developmental process. For example, the combination of OCT4-SOX2 is critical for pluripotency,[68] but OCT4-SOX17, binding to a slightly different DNA motif acts to specify primitive endoderm,[69] whilst another OCT/POU-family containing complex, BRN2-SOX2, specifies neural progenitors[70] (Fig. 3B). Complex cell type specific assembly of TF complexes is not limited to OCT/POU-SOX factor pairs, as GATA1, GATA2 and PU.1 can assemble on a variety of specific DNA motifs to direct erythroid and neutrophil cell fates.[71] TF–TF pairing appears to be widespread; a systematic analysis of genome-wide TF binding discovered 603 potential constrained TF–TF pairs,[72–74] suggesting a combinatorial code that adds complexity to regulate the diversity of cell types and biological processes.

TFs that have a cell type-independent pattern of expression might not seem a promising area to explore for cell type-specific control, but these TFs can also exert specificity in the correct setting. Around 30% of TFs are cell type-independent at both the RNA,[75] and protein level.[61] It might seem that these TFs are involved in basal cell activities, and indeed many are,[61] but cryptically, many cell type-independent TFs can have highly cell type-

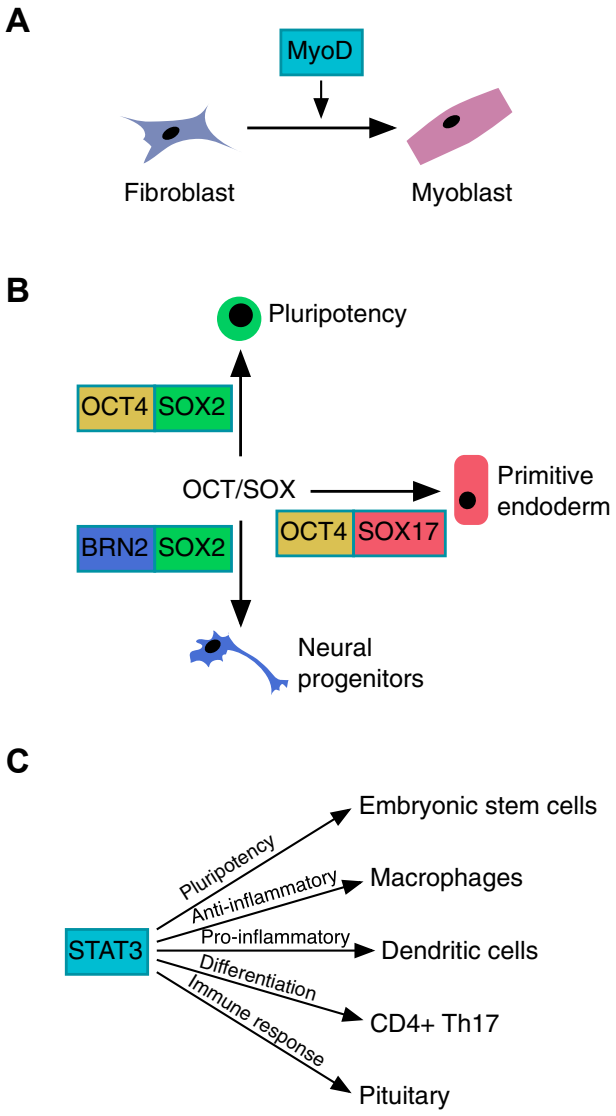


**Fig. 3.** Mechanisms of transcription factor mediated cell type determination. (A) A single 'master' transcription factor can activate an entire gene expression program to alter cell type. In this example, the transcription factor MyoD can convert fibroblasts into myoblasts. (B) Different combinations of transcription factor pairs can specify alternate cell lineages. In this example one OCT (OCT4/BRN2) factor pairs with a SOX (SOX2/SOX17) factor to maintain/specify one of three cell fates: embryonic stem cells (pluripotency), a primitive endoderm (yolk sac), or neural progenitor cell fate. (C) The same transcription factor can have multiple biological roles in different cell types. The example shown here is the transcription factor STAT3, which is expressed in most tissues but has widely divergent biological functions in different cell types.

specific function. For example, STAT3, despite being expressed almost uniformly in cells and tissues,[76] specifies pluripotency in mESCs, an anti-inflammatory response in macrophages, a pro-inflammatory response in dendritic cells, and has a critical role in T helper17 cells type differentiation, amongst many other cell type-specific roles[76–78] (Fig. 3C). Ultimately, many TFs have widely overlapping functions in multiple cell types, and as yet, no comprehensive model of TF cell type control exists.[60]

## 6. Exogenous expression of transcription factors can drive conversion of cell type

TFs have been instrumental in the forced conversion of one cell type to another. The earliest use of a TF to drive transdifferentiation was the transfection of *Myod1* (MyoD), to convert cells to myoblasts,[62] and *Cebpa* and *Cebpb* to convert B cells into macrophages.[79] However, the most dramatic demonstration of the power of TFs was the conversion of fibroblasts to mESCs using just four TFs: OCT4, SOX2, KLF4 and c-MYC.[80] Since this breakthrough many other transdifferentiation protocols have been discovered,[81] along with the use of small molecules to convert cell type, for example the conversion of fibroblasts to neurons[82] or fibroblasts to mESCs.[83] Intriguingly, many of the small molecules used in these protocols directly interfere with epigenetic control, such as DZNep (methylation inhibitor), VPA (histone deacetylase inhibitor), or Tranylcypromine (histone demethylase and monoamine oxidase inhibitor), indicating that epigenetic control is a major factor in the determination of cell type.[84]

Transdifferentiation protocols mediated by TFs nonetheless remain relatively few,[81] and there are many target cell types we would like to *in vitro* differentiate, but cannot. Consequently, there has been a lot of activity in designing systematic computational approaches to predict candidates and improve existing approaches. Many approaches attempt to identify cell type-specific genes,[26] as these are relatively easy to identify, and their specific presence in a cell type is often (although by no means always), indicative of function. Computational efforts to identify transdifferentiation factors,[85] has included modelling development onto patterns of gene expression,[26] and approaches to discover 'core' TFs that are both cell type-specific and expressed at high levels.[86] Mogrify used a cell ontology tree to map cell type-specific genes against their developmental pattern and so identify TFs specific to a developmental lineage. Mogrify also includes nearest neighbour protein–protein interactions to overcome limitations in discovering cell type-independent TFs.[87] This technique was very successful in discovering previously known transdifferentiation TFs, and was used to predict and then validate TFs that transdifferentiated keratinocytes into endothelial cells.[87] CellNet describes another approach using vast amounts of microarray data to build cell type-specific gene regulatory networks, and then to apply these networks to predict cell type-specific regulatory modules, and so candidate TFs for transdifferentiation.[88] CellNet set out to solve a common problem in transdifferentiation and differentiation experiments where the differentiated cells fail to completely silence the gene expression program of the originating cell type, and remain immature.[89] CellNet was successfully used to improve the transdifferentiation of B cells to macrophages, and also identified an alternate colon cell fate for cells that were transdifferentiating to hepatocytes.[90] Pairs of TFs often antagonize each other's function, hence pairs of TFs with opposing gene expression in two cell types could be used to predict master regulators of lineages.[91,92] Methods that combine gene expression data with epigenetic data have also been successful in predicting transdifferentiation TFs.[93] Another approach extended the discovery of cell type-regulatory modules by looking at gene expression in other mammalian species, and discovered many primate-specific long non-coding RNAs (lncRNAs) with putative cell type-specific functions.[94] Indeed, lncRNAs are also expressed in a cell type-specific pattern,[26,95] and are good candidates for cell type-specific control.[96,97] However, a comprehensive explanation of how TFs, lincRNAs, and other non-coding RNAs can control cell type, and why transdifferentiation is rare in the adult organism remains unclear. One possible solution to this problem are developmental landscapes and cellular pathways that describe the routes cells can traverse to alter cell type.

## 7. Developmental landscapes

The concept of cell type is a powerful and attractive model to explain how a limited supply of information can encode a wide

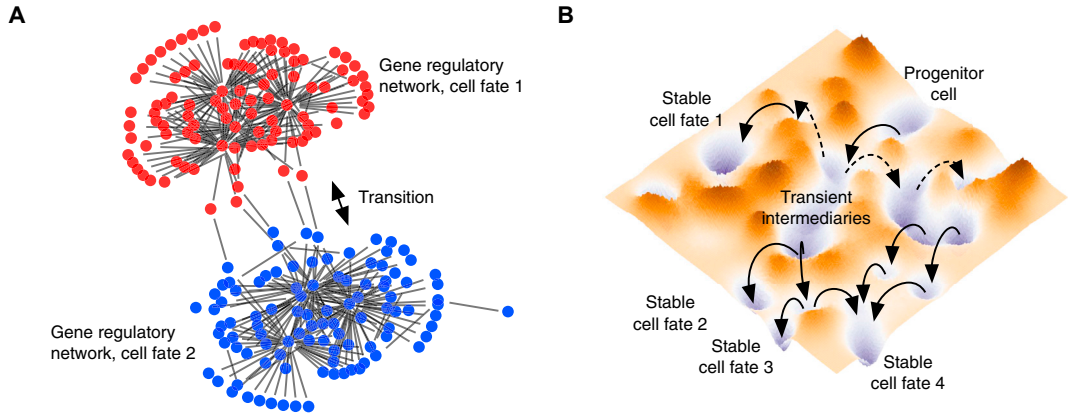


**Fig. 4.** Gene regulatory control of cell type. (A) Two gene regulatory networks are shown for two hypothetical cell states (red and blue). Each node in the network is a gene or protein, and each edge is a regulatory interaction, such as a TF binding to regulate a gene, or a kinase activating or repressing a protein, etc. Sets of genes organize into self-stabilizing networks that maintain cell type, and impede transdifferentiation. (B) One view of cell type is the concept of a 'landscape' for all cell type possibilities. In this idea, cell types exist in a probability space, and cell types are then defined as the local energy minima in which a cell type can stably exist (purple depressions), surrounded by barriers that prevent conversion of cell type (orange hills).

array of complex developmental patterns. Development is a highly-ordered process marked by major stages of cell differentiation during gastrulation that establish the three somatic germ lineages of the mesoderm (blood, muscle), endoderm (lung, digestive tract) and ectoderm (skin, brain).[98] As cells take part in development they differentiate to specific cell types and lose developmental competency. What mechanistically underlies these cell type conversions in development is less well understood. What is clear is that combinations of interconnected genes form 'gene regulatory networks' (Fig. 4A), and the genes and proteins regulate each other at multiple levels to maintain a semi-stable cell type. But what ultimately builds Waddington's epigenetic landscapes, or the ocean expanses of Cook's islands is unknown.[99] These two models suggest cell types are like depressions in a topological landscape (Waddington), or islands of cell type stability separated by expanses of unstable sea (Cook's islands)[99] (Fig. 4B). Gene regulatory networks underlying cell types may function as 'attractors' for cell types to cluster around,[100] similar to probabilistic cell state maps that can construct landscapes for cells,[101] or cellular network entropy,[102] or landscapes constructed based on molecular similarity.[87] Ultimately, conceptual ideas such as Waddington's epigenetic landscape, and Cook's islands, still lack robust biological mechanisms that can explain all aspects of cell type control. Specifically, why there is a limited number of cell types at all, why there are so many unstable intermediary states between cell types, why some cell types are stable and some are not, and why there are strict limits placed on cell type transdifferentiation. Nonetheless, it is becoming possible to model the organization of cell type on a large scale,[26,103] even if we cannot as yet understand the process in detail. New technologies will continue to be applied to this problem; its deeper understanding will have important implications for understanding cellular regeneration and ultimately tissue regeneration and how these techniques can be applied to the understanding of development and human disease.

## Acknowledgements

We thank Miguel Esteban for comments on an early draft of this manuscript. This work was supported by the National Natural Science Foundation of China (31471242), Shenzhen Peacock Plan, and the Shenzhen Science and Technology Innovation Committee general program (JCYJ20170307110638890).

## References

1. Gautier EL, Shay T, Miller J, et al. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol*. 2012;13(11):1118–1128.
2. Chang HY, Chi JT, Dudoit S, et al. Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci U S A*. 2002;99(20):12877–12882.
3. Symmons O, Raj A. What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Mol Cell*. 2016;62(5):788–802.
4. Bell G, Mooers AO. Size and complexity among multicellular organisms. *Biol J Linn Soc*. 1997;60(3):345–363.
5. Valentine JW, Collins AG, Meyer CP. Morphological complexity increase in metazoans. *Paleobiology*. 1994;20(2):131–142.
6. Hatano A, Chiba H, Moesa HA, et al. CELLPEDIA: a repository for human cell information for cell studies and differentiation analyses. *Database (Oxf)*. 2011;2011:bar046.
7. Stachelscheid H, Seltmann S, Lekschas F, et al. CellFinder: a cell data repository. *Nucleic Acids Res*. 2014;42:D950–D958.
8. Edgar R, Mazor Y, Rinon A, et al. LifeMap discovery: the embryonic development, stem cells, and regenerative medicine research portal. *PLoS One*. 2013;8(7):e66629.
9. Diehl AD, Meehan TF, Bradford YM, et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J Biomed Semant*. 2016;7(1):44.
10. Gowans JL. The life-history of lymphocytes. *Br Med Bull*. 1959;15(1):50–53.
11. Cooper MD. The early history of B cells. *Nat Rev Immunol*. 2015;15(3):191–197.
12. Miller JF. Events that led to the discovery of T-cell development and function – a personal recollection. *Tissue Antigens*. 2004;63(6):509–517.
13. Crotty S. A brief history of T cell help to B cells. *Nat Rev Immunol*. 2015;15(3):185–189.
14. Clark G, Stockinger H, Balderas R, et al. Nomenclature of CD molecules from the tenth human leucocyte differentiation antigen workshop. *Clin Transl Immunol*. 2016;5(1):e57.
15. Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations. *Annu Rev Immunol*. 2010;28:445–489.
16. Chen W, Jin W, Hardegen N, et al. Conversion of peripheral CD4+CD25− naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. *J Exp Med*. 2003;198(12):1875–1886.
17. Proserpio V, Piccolo A, Haim-Vilmovsky L, et al. Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. *Genome Biol*. 2016;17:103.
18. Li R, Liang J, Ni S, et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell*. 2010;7(1):51–63.
19. Akdis M, Palomares O, van de Veen W, van Splunter M, Akdis CA. TH17 and TH22 cells: a confusion of antimicrobial response with tissue inflammation versus protection. *J Allergy Clin Immunol*. 2012;129(6):1438–1449. quiz 1450–1451.
20. Hirahara K, Nakayama T. CD4+ T-cell subsets in inflammatory diseases: beyond the Th1/Th2 paradigm. *Int Immunol*. 2016;28(4):163–171.
21. Hashimoto D, Miller J, Merad M. Dendritic cell and macrophage heterogeneity in vivo. *Immunity*. 2011;35(3):323–335.
22. Kim CC, Lanier LL. Beyond the transcriptome: completion of act one of the Immunological Genome Project. *Curr Opin Immunol*. 2013;25(5):593–597.

23. Hutchins AP, Takahashi Y, Miranda-Saavedra D. Genomic analysis of LPS-stimulated myeloid cells identifies a common pro-inflammatory response but divergent IL-10 anti-inflammatory responses. *Sci Rep*. 2015;5:9100.
24. Hume DA, Mabbott N, Raza S, Freeman TC. Can DCs be distinguished from macrophages by molecular signatures? *Nat Immunol*. 2013;14(3):187–189.
25. Hume DA. Macrophages as APC and the dendritic cell myth. *J Immunol*. 2008;181(9):5829–5835.
26. Hutchins AP, Yang Z, Li Y, et al. Models of global gene expression define major domains of cell type and tissue identity. *Nucleic Acids Res*. 2017;45(5):2354–2367.
27. Randolph G, Merad M. Reply to: "Can DCs be distinguished from macrophages by molecular signatures?". *Nat Immunol*. 2013;14(3):189–190.
28. Geissmann F, Gordon S, Hume DA, Mowat AM, Randolph GJ. Unravelling mononuclear phagocyte heterogeneity. *Nat Rev Immunol*. 2010;10(6):453–460.
29. Jaitin DA, Weiner A, Yofe I, et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell*. 2016;167(7):1883–1896.
30. Beddington RS, Robertson EJ. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development*. 1989;105(4):733–737.
31. Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008;133(6):1106–1117.
32. Nichols J, Smith A. The origin and identity of embryonic stem cells. *Development*. 2011;138(1):3–8.
33. Tang F, Barbacioru C, Bao S, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell*. 2010;6(5):468–478.
34. Torres-Padilla ME, Chambers I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development*. 2014;141(11):2173–2181.
35. Kolodziejczyk AA, Kim JK, Tsang JC, et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*. 2015;17(4):471–485.
36. Mitsui K, Tokuzawa Y, Itoh H, et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*. 2003;113(5):631–642.
37. Chambers I, Silva J, Colby D, et al. Nanog safeguards pluripotency and mediates germline development. *Nature*. 2007;450(7173):1230–1234.
38. Singh AM, Hamazaki T, Hankowski KE, Terada N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells*. 2007;25(10):2534–2542.
39. Kalmar T, Lim C, Hayward P, et al. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol*. 2009;7(7):e1000149.
40. Hayashi K, de Sousa Lopes SMC, Tang F, Lao K, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*. 2008;3(4):391–401.
41. Morgani SM, Canham MA, Nichols J, et al. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep*. 2013;3(6):1945–1957.
42. Hutchins AP, Pei D. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci Bull (Beijing)*. 2015;60(20):1722–1733.
43. Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. 2012;487(7405):57–63.
44. Yang Y, Liu B, Xu J, et al. Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell*. 2017;169(2):243–257.
45. Brons IG, Smithers LE, Trotter MW, et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. 2007;448(7150):191–195.
46. Tesar PJ, Chenoweth JG, Brook FA, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 2007;448(7150):196–199.
47. Hutchins AP, Choo SH, Mistri TK, et al. Co-motif discovery identifies an Esrrb-Sox2-DNA ternary complex as a mediator of transcriptional differences between mouse embryonic and epiblast stem cells. *Stem Cells*. 2013;31(2):269–281.
48. Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*. 2011;146(4):519–532.
49. Nakaki F, Hayashi K, Ohta H, Kurimoto K, Yabuta Y, Saitou M. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature*. 2013;501(7466):222–226.
50. Wu J, Okamura D, Li M, et al. An alternative pluripotent state confers interspecies chimaeric competency. *Nature*. 2015;521(7552):316–321.
51. Wu J, Izpisua Belmonte JC. Dynamic pluripotent stem cell states and their applications. *Cell Stem Cell*. 2015;17(5):509–525.
52. Treutlein B, Brownfield DG, Wu AR, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*. 2014;509(7500):371–375.
53. Guo G, Huss M, Tong GQ, et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell*. 2010;18(4):675–685.
54. Bendall SC, Davis KL, Amir el AD, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*. 2014;157(3):714–725.
55. Treutlein B, Lee QY, Camp JG, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature*. 2016;534(7607):391–395.
56. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381–386.
57. Wagers AJ, Weissman IL. Plasticity of adult stem cells. *Cell*. 2004;116(5):639–648.
58. Sarig R, Tzahor E. The cancer paradigms of mammalian regeneration: can mammals regenerate as amphibians? *Carcinogenesis*. 2017;38(4):359–366.
59. Brent MR. Past roadblocks and new opportunities in transcription factor network mapping. *Trends Genet*. 2016;32(11):736–750.
60. Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordan R, Rohs R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci*. 2014;39(9):381–399.
61. Zhou Q, Liu M, Xia X, et al. A mouse tissue transcription factor atlas. *Nat Commun*. 2017;8:15089.
62. Tapscott SJ, Davis RL, Thayer MJ, Cheng PF, Weintraub H, Lassar AB. MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science*. 1988;242(4877):405–411.
63. Home P, Ray S, Dutta D, Bronshteyn I, Larson M, Paul S. GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. *J Biol Chem*. 2009;284(42):28729–28737.
64. Ralston A, Cox BJ, Nishioka N, et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development*. 2010;137(3):395–403.
65. Zhu J, Min B, Hu-Li J, et al. Conditional deletion of Gata3 shows its essential function in T(H)1–T(H)2 responses. *Nat Immunol*. 2004;5(11):1157–1165.
66. Badis G, Berger MF, Philippakis AA, et al. Diversity and complexity in DNA recognition by transcription factors. *Science*. 2009;324(5935):1720–1723.
67. Jolma A, Yan J, Whitington T, et al. DNA-binding specificities of human transcription factors. *Cell*. 2013;152(1–2):327–339.
68. Rodda DJ, Chew JL, Lim LH, et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem*. 2005;280(26):24731–24737.
69. Aksoy I, Jauch R, Chen J, et al. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J*. 2013;32(7):938–953.
70. Lodato MA, Ng CW, Wamstad JA, et al. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet*. 2013;9(2):e1003288.
71. May G, Soneji S, Tipping AJ, et al. Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. *Cell Stem Cell*. 2013;13(6):754–768.
72. Jankowski A, Szczurek E, Jauch R, Tiuryn J, Prabhakar S. Comprehensive prediction in 78 human cell lines reveals rigidity and compactness of transcription factor dimers. *Genome Res*. 2013;23(8):1307–1318.
73. Ravasi T, Suzuki H, Cannistraci CV, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*. 2010;140(5):744–752.
74. Gerstein MB, Kundaje A, Hariharan M, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature*. 2012;489(7414):91–100.
75. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 2009;10(4):252–263.
76. Hutchins AP, Diez D, Miranda-Saavedra D. Genomic and computational approaches to dissect the mechanisms of STAT3's universal and cell type-specific functions. *JAKSTAT*. 2013;2(4):e25097.
77. Hutchins AP, Diez D, Takahashi Y, et al. Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions. *Nucleic Acids Res*. 2013;41(4):2155–2170.
78. Vahedi G, Takahashi H, Nakayamada S, et al. STATs shape the active enhancer landscape of T cell populations. *Cell*. 2012;151(5):981–993.
79. Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages. *Cell*. 2004;117(5):663–676.
80. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126(4):663–676.
81. Xu J, Du Y, Deng H. Direct lineage reprogramming: strategies, mechanisms, and applications. *Cell Stem Cell*. 2015;16(2):119–134.
82. Li X, Zuo X, Jing J, et al. Small-molecule-driven direct reprogramming of mouse fibroblasts into functional neurons. *Cell Stem Cell*. 2015;17(2):195–203.
83. Hou P, Li Y, Zhang X, et al. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science*. 2013;341(6146):651–654.
84. Ruzov A, Tsenkina Y, Serio A, et al. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res*. 2011;21(9):1332–1342.
85. Bian Q, Cahan P. Computational tools for stem cell biology. *Trends Biotechnol*. 2016;34(12):993–1009.
86. D'Alessio AC, Fan ZP, Wert KJ, et al. A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Rep*. 2015;5(5):763–775.

87. Rackham OJ, Firas J, Fang H, et al. A predictive computational framework for direct reprogramming between human cell types. *Nat Genet*. 2016;48(3):331–335.
88. Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ. CellNet: network biology applied to stem cell engineering. *Cell*. 2014;158(4):903–915.
89. Li Q, Hutchins AP, Chen Y, et al. A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes. *Nat Commun*. 2017;8:15166.
90. Morris SA, Cahan P, Li H, et al. Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell*. 2014;158(4):889–902.
91. Heinaniemi M, Nykter M, Kramer R, et al. Gene-pair expression signatures reveal lineage control. *Nat Methods*. 2013;10(6):577–583.
92. Crespo I, Del Sol A. A general strategy for cellular reprogramming: the importance of transcription factor cross-repression. *Stem Cells*. 2013;31(10): 2127–2135.
93. Lang AH, Li H, Collins JJ, Mehta P. Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. *PLoS Comput Biol*. 2014;10(8):e1003734.
94. Yang Y, Yang YT, Yuan J, Lu ZJ, Li JJ. Large-scale mapping of mammalian transcriptomes identifies conserved genes associated with different cell states. *Nucleic Acids Res*. 2017;45(4):1657–1672.
95. Amin V, Harris RA, Onuchic V, et al. Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. *Nat Commun*. 2015;6:6370.
96. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011;477(7364): 295–300.
97. Bao X, Wu H, Zhu X, et al. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res*. 2015;25(1):80–92.
98. Tam PP, Behringer RR. Mouse gastrulation: the formation of a mammalian body plan. *Mech Dev*. 1997;68(1-2):3–25.
99. Sieweke MH. Waddington's valleys and Captain Cook's islands. *Cell Stem Cell*. 2015;16(1):7–8.
100. Ribeiro AS, Kauffman SA. Noisy attractors and ergodic sets in models of gene regulatory networks. *J Theor Biol*. 2007;247(4):743–755.
101. Hanna JH, Saha K, Jaenisch R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell*. 2010;143(4):508–525.
102. Banerji CR, Miranda-Saavedra D, Severini S, et al. Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Sci Rep*. 2013;3:3039.
103. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest AR, Kawaji H, Rehli M, et al. A promoter-level mammalian expression atlas. *Nature*. 2014;507(7493):462–470.
104. Hutchins AP, Jauch R, Dyla M, Miranda-Saavedra D. glbase: a framework for combining, analyzing and displaying heterogeneous genomic and high-throughput sequencing data. *Cell Regen (Lond)*. 2014;3(1):1.