Developmental biology

The inner lives of early embryonic cells

Chengxiang Qiu & Jay Shendure

Characterization of the early developmental process called gastrulation has mostly been limited to snapshots at different time points. A model of mouse gastrulation now maps the transitions between cell types continuously in time.

During embryonic development, new cell types emerge with stunning speed and robustness. The process of gastrulation — in which a single layer of cells gives rise to multiple 'germ layers' — is fundamental to the early development of most animals. Although studied for more than 150 years, many aspects of gastrulation remain elusive, not least a comprehensive understanding of the molecular factors governing the specification of the many cell lineages that emerge from this process. Writing in *Cell*, Mittnenzweig *et al.*[1] densely sample gene expression in gastrulating mouse embryos over a 36-hour window and construct a continuous model of cell-lineage specification.

If we think of cells during gastrulation as characters in a silent film (and, indeed, there are beautiful films of gastrulation[2]), how can we understand the internal monologues and ever-changing motivations of the personas on the screen? Only in the past five years or so, with the emergence of technologies that characterize the molecular profiles of individual cells, have we been in a position to fully monitor cells' 'inner lives' throughout gastrulation, as cell lineages develop. One such technology is single-cell RNA sequencing (scRNA-seq), which profiles the messenger RNA contents of individual cells.

Several key questions remain that could be addressed through single-cell techniques. For example, what is the precise timing of cell-type specifications in the developing embryo? Can we find a model that accurately describes these specifications? What are the principal molecular factors involved? And which of these factors 'drive' cell-type specifications, and which 'respond' to them?

In most animals, including mammals, three germ layers result from gastrulation: the ectoderm, the mesoderm and the endoderm. In the mouse, the pre-eminent model system for mammalian development, gastrulation begins about 6.5 days after fertilization (that is, embryonic day 6.5, or E6.5). Although we and others have performed scRNA-seq across early mouse development at what might seem like reasonable temporal resolution (for example, sampling every 6 hours from E6.5 to E8.5 (ref. 3), or every 24 hours from E9.5 to E13.5 (ref. 4)), the pace of change during mouse development is so fast that this might be woefully inadequate. In our film analogy, this would be akin to watching a film but with only a handful of scenes narrated.

In this context, Mittnenzweig *et al.* set out to generate a continuous representation of cell-state dynamics during mouse gastrulation[1]. They applied scRNA-seq to 153 mouse embryos from E6.5 to E8.1, altogether profiling gene expression in about 33,000 individual cells. Because the accuracy of estimating embryo age on the basis of morphological landmarks is limited, the embryos' ages were instead inferred from the molecular data, resulting in each embryo being assigned to one of 13 time points.

To increase the temporal resolution of their representation of development beyond a series of snapshots, Mittnenzweig *et al.* posited that cells within any given embryo are, to some extent at least, at different stages of developmental maturity relative to one another. The authors grouped cells according to their molecular similarity into 461 subsets termed 'metacells', each consisting of cells that were very similar, but that, notably, might have come from different embryos and/or from different time points. The authors then applied an algorithm to estimate the fraction of cells from each metacell at time t that 'flow' to other metacells at time $t + 1$. Crucially, the inferred flows between these metacells are continuous with respect to time, despite the temporally discrete nature of the embryos from which they were derived.

With this continuous model of mouse gastrulation in hand (Fig. 1), Mittnenzweig and colleagues are able to investigate several interesting questions. First, how and when do new cell types emerge during gastrulation, and what are the associated changes in the patterns of gene expression? For example, their model not only predicts that primitive erythroid cells (which give rise to early red blood cells) originate from a region called the primitive streak, but also constrains the timing of that contribution to before E6.7, and places in order the successive waves of expression of different transcription factors associated with this lineage.

Second, what are the characteristics of *in vivo* cell-type specification? Do new cell types emerge through a series of rapidly made 'decisions' between two different cell fates, resulting in the sharp, branch-like bifurcations that often appear in textbook flow charts of

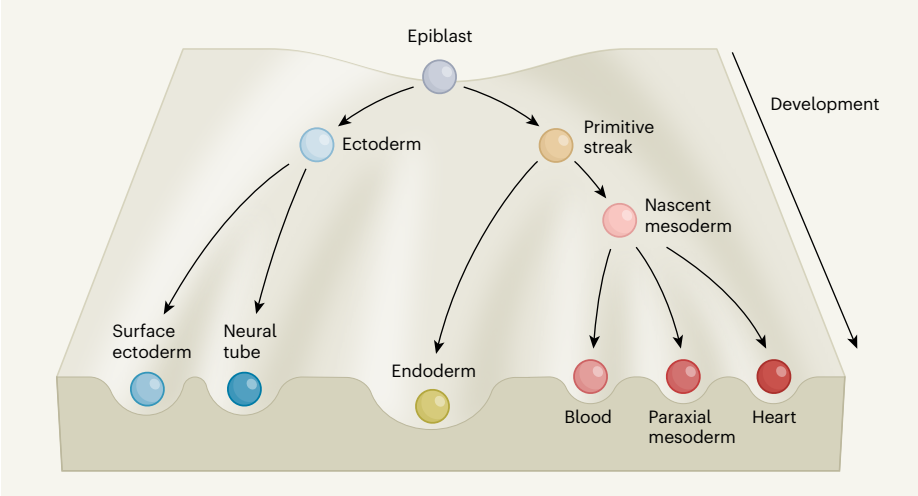

Figure 1 | A continuous-flow model of mouse gastrulation. Mittnenzweig *et al.*[1] profiled the RNA contents of cells from mouse embryos during gastrulation, a process during which a single cell layer called the epiblast transforms into three layers: ectoderm, endoderm and mesoderm. In doing so, the authors generated a model of gastrulation that is continuous over time, showing the transitions between different cell types; a simplified version is shown here. Shallower, basin-like regions are intended to depict gradual, continuous transitions, whereas deeper, canyon-like regions depict more-definitive separations. Both bifurcations and multifurcations of cell lineages are observed.

cell development, or are more-complex patterns observed, such as multifurcations and continuous transitions? Mittnenzweig *et al*. suggest the answer to be 'all of the above'.

For example, the developmental trajectory of cells in the primitive streak bifurcates sharply, such that these cells become either mesodermal or endodermal cells (Fig. 1). By contrast, the differentiation of cells in the nascent mesoderm is inferred to be gradual and continuous, and with more than two destinations. The model also enables the inference of flows that change with time; for instance, before E7.1, epiblast cells overwhelmingly transition to acquiring primitive-streak fates, but shortly after that point, they mostly transition to acquiring ectodermal fates.

Finally, what are the molecular factors that underlie differentiation, and do individual factors act alone or in combination? The authors claim that, with the exception of some lineages (notably, the node, cardiomyocyte and haemato-endothelial lineages), the landscape of gastrulation is predominantly characterized by a dependence on overlapping combinations of factors, as well as on a gradual unfolding of commitment. For example, although cells of the nascent mesoderm progress into a spectrum of fates, these fates are not sharply separated from one another, and there is no clear delineation between the sets of transcription factors that seem to specify each fate. The authors propose that, rather than a series of specific factors governing a stepwise, hierarchical progression of specification, combinations of molecular factors regulate diverse mesodermal fates in a 'fuzzy' and almost probabilistic manner. To highlight the delicacy of this program, the authors carried out experiments in which inferred key regulators were genetically disrupted, which led to delayed differentiation of affected lineages.

Of course, all models have limitations, and this model has its own. First, its resolution is limited by the underlying data, although simply processing more embryos would address this. Second, its metacells and flows are inferred solely from the similarity of the transcriptional profiles of the cells, and so there is a risk of missing or misinterpreting certain bona fide relationships[5]. Particularly rapid changes in the nematode worm *Caenorhabditis elegans,* for instance, elude efforts to reconstruct lineages in 'pseudotime' – that is, ordering cells by their developmental stage rather than their age in real time[6]. Third, the model ignores cells' spatial coordinates within embryos as well as their actual lineage relationships, two crucial aspects of development that are increasingly amenable to measuring and recording, respectively[7,8].

Notwithstanding these limitations, the model of mouse gastrulation developed by Mittnenzweig *et al*. is impressive, and shows how continuous maps of complex differentiation landscapes might be recovered despite discrete sampling. Together with other work published in the past few years[3,6,9], it represents a substantial step forward on the path to a complete understanding of cells' inner lives during this most important of times in an animal's life[10].

Chengxiang Qiu and **Jay Shendure** are in the Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. **J.S.** is also at the Brotman Baty Institute for Precision Medicine, the Allen Discovery Center for Cell Lineage Tracing, and the Howard Hughes Medical Institute, Seattle.
e-mail: shendure@uw.edu

1. Mittnenzweig, M. *et al. Cell* https://dx.doi.org/10.1016/j.cell.2021.04.004 (2021).
2. McDole, K. M. *et al. Cell* **175**, 859–876 (2018).
3. Pijuan-Sala, B. *et al. Nature* **566**, 490–495 (2019).
4. Cao, J. *et al. Nature* **566**, 496–502 (2019).
5. Wagner, D. E. & Klein, A. M. *Nature Rev. Genet.* **21**, 410–427 (2020).
6. Packer, J. S. *et al. Science* **365**, eaax1971 (2019).
7. McKenna, A. *et al. Science* **353**, aaf7907 (2016).
8. Lohoff, T. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2020.11.20.391896 (2020).
9. Argelaguet, R. *et al. Nature* **576**, 487–491 (2019).
10. Vicente, C. *Development* **142**, 2547–2548 (2015).

This article was published online on 30 April 2021.

Medical research

A molecular connection to Crohn's disease risk

Scott Plevy

Mutations of the *NOD2* gene are risk factors for Crohn's disease. Many aspects of how they contribute to the condition are unknown. The discovery of cell populations that are involved suggests new therapeutic options. **See p.275**

Crohn's disease, a chronic inflammatory bowel disease, affects many people. For example, more than 0.3% of the populations of Canada and Germany have the condition, and its incidence is increasing worldwide[1]. Better therapies are needed, but progress in treating Crohn's disease has been hampered by the lack of understanding of how it arises. On page 275, Nayar *et al*.[2] shed light on a long-standing mystery about one risk factor for Crohn's disease, and their findings have important clinical implications.

Crohn's disease can affect any part of the gut. Most commonly, it affects the ileum region, causing inflammation that frequently results in fibrosis (the deposition of fibrous connective tissue as an injury response). This leads to the narrowing (or stricture) of the lumen of the ileum, which often requires surgical intervention[3]. Crohn's disease provides a useful model of illnesses that are mediated by genes and environmental interactions. In this case, genetic susceptibility underpins the disease-causing inflammatory responses to gut microorganisms.

Genetic variations, called polymorphisms, of the *NOD2* gene are the strongest genetic risk association for Crohn's disease; approximately 20% of all such risk of developing the disease is related to three single nucleotide polymorphisms of this gene[4]. Furthermore, *NOD2* mutations are strong predictive factors for the development of ileum strictures and for the need for surgery in Crohn's disease, which is a widely validated association between the genetic underpinnings of this condition and manifestations of the disease[3].

However, connecting the *NOD2* gene to disease susceptibility presented a paradox. NOD2 is an intracellular receptor (Fig. 1) that recognizes the molecule muramyl dipeptide (MDP) – a ubiquitous component of bacterial cell walls. Before *NOD2* was described as a risk gene for Crohn's disease, NOD2 function was best understood in immune cells that aid the innate branch of immune defences. NOD2 activation in these cells leads to the expression of inflammatory molecules called cytokines, and an abnormally intense inflammatory response can mediate intestinal damage in Crohn's disease[5,6]. One might therefore have expected that *NOD2* mutations known as loss-of-function mutations, which do not generate a fully functional version of the encoded protein, would protect against Crohn's disease. Yet such loss-of-function mutations of *NOD2* were identified as risk factors for the disease. Subsequent research therefore pivoted to focus on a different aspect of NOD2 biology in the intestine, investigating how functional NOD2 maintains homeostasis in the intestine, where the body's largest biomass of immunologically active cells is constantly exposed to MDP from gut microbes, and how *NOD2* mutations perturb this balance and lead to disease[5].

The role of *NOD2* mutations in the emergence