Published in final edited form as:

Nat Biotechnol. 2015 March ; 33(3): 260–261. doi:10.1038/nbt.3168.

Singling out blood development

Eva M Fast and Len I Zon

Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA

Abstract

Analysis of gene expression in thousands of single cells generates a model of the blood regulatory network.

The more we learn about the intricacies of embryonic development, the more it seems that populations of apparently similar cells are in fact heterogeneous, with individual cells developing at different rates than their neighbors. Efforts to measure these differences have been hampered by technological limitations, and most studies of transcription in embryos have been carried out on pools of cells, making it difficult to tease apart the gene regulatory networks that control developmental processes. In this issue, Moignard *et al.*¹ describe an approach for inferring the regulatory interactions of blood development by assessing gene expression in thousands of single cells and computationally reducing these multidimensional datasets to direct interactions between genes. This powerful analysis not only provides a global glimpse of blood development but may serve as a blueprint for future modeling studies based on single-cell data.

Moignard *et al.*¹ are among the first to assay cellular differentiation using single-cell expression analysis over an *in vivo* time course^{2,3}. They begin by capturing 3,934 blood precursor cells from mouse embryos at four successive developmental stages, making this the most comprehensive single-cell expression study of organ development to date (Fig. 1a). The cells are isolated by fluorescence-activated cell sorting using the mesodermal marker Flkl (Kdr) and the blood-specific marker Runx1. Capturing the entire population with blood-forming potential is a prerequisite for building a comprehensive model of blood development.

Next, the expression levels of 42 transcription factors and marker genes related to blood and endothelium are measured in each of the 3,934 cells by qRT-PCR. It is not surprising that the resulting multidimensional data set of >150,000 expression scores requires extensive and sophisticated computational analysis. This analysis yields a network that models genetic interactions during blood development using the Boolean rules AND, OR and NOT, known as an asynchronous Boolean network.

zon@enders.tch.harvard.edu.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

The computational analysis involves two main steps. The first step aims to developmentally link the four Flk1⁺ cell populations despite their being collected at different time points and from different embryos. Specifically, the authors expect that cells destined to become blood or endothelium will follow different developmental trajectories. These two populations of Flk1⁺ cells cannot be resolved by conventional methods, such as hierarchical clustering and principal component analysis. But application of a newly developed computational approach based on diffusion metrics, called diffusion maps, succeeds in ordering all the sorted cells more closely according to developmental time (Fig. 1b).

One possible reason why standard methods fail to cluster the data is that Flk1 is a very general marker. Flk1⁺ cells in the early embryo have the potential to form both blood and endothelial cells but also other mesodermal lineages, such as cardiac tissue. Moignard *et al.*¹ address this possibility by performing RNA-seq analysis on pools of 50 cells from each of the four time points, which allows them to analyze genes that were not included in the predefined qRT-PCR set used for the single-cell analysis. The authors conclude that most of the sorted single cells are indeed destined to become blood, a precondition for the subsequent analysis.

The second step, and the central part of the paper, involves distilling the gene expression states of the putative blood lineage into a regulatory network that models blood development (Fig. 1c). Single-cell expression states are transformed into regulatory interactions between genes through a pioneering analytic method called the single-cell network synthesis (SCNS) toolkit (<http://scns.stemcells.cam.ac.uk/>) (Fig. 1c). Briefly, gene expression is discretized in an on-and-off pattern, and all existing expression states are ordered in a state diagram based on similarity. Incorporating the time variable into these expression states allows the authors to model potential genetic interactions as an asynchronous Boolean network.

It could be argued that Boolean logic cannot capture the full complexity of biological systems. Nevertheless, constructing a whole gene network based on single-cell expression states is an innovative, first-of-its-kind effort. Previous single-cell studies were only able to infer pairwise genetic interactions^{4,5}. The asynchronous Boolean network constructed by Moignard *et al.*¹ not only predicts direct relationships between genes but can also be used to model the consequences of gene perturbation on blood development as a whole.

In an effort to validate their model, Moignard *et al.*¹ experimentally demonstrate both a Boolean interaction between Hox and Sox factors on Erg and confirm the prediction that overexpression of Sox7 inhibits blood formation. Although these results are encouraging, the real value of the asynchronous Boolean network will be to serve as a resource for further genetic investigations.

Single-cell expression profiling adds another layer of detail to transcriptional analysis, but it also increases complexity. For *in vivo* studies, off-the-shelf protocols do not yet exist, and researchers will have to carefully evaluate whether single-cell analysis is needed to address their scientific question. Eventually, all transcriptome analysis may be performed at the single-cell level. For the moment, data processing remains a big hurdle, especially for laboratories that do not have sufficient bioinformatics support. Reports such as the one by

Moignard *et al.*¹ are thus especially valuable because they provide the community with a comprehensive informatics toolkit for handling single-cell data sets.

Cellular heterogeneity is a common theme of this and several other single-cell studies. Often it is not possible to fully determine whether noise in the data is of a technical nature or is inherent to the biological process. A recent paper presents a method to control for the influence of the cell cycle on the variability of gene expression in single-cell studies⁶. Especially during development, frequent cell division can affect the expression of multiple genes and thereby contribute to variability at the single-cell level.

Although Moignard *et al.*¹ report heterogeneity in developmental maturation, it is possible that the qRT-PCR approach they used missed some clustering genes that would have further resolved different developmental lineages. This caveat will soon become obsolete as rapid advances in RNA-seq technologies make multiplexed analyses of large numbers of single cells more feasible⁷. Analyzing single-cell gene expression in a transcriptome-wide manner may allow better control for heterogeneity and reveal novel cellular subpopulations with different developmental potentials.

In the past, gene regulatory networks in development have been assembled from multiple, often time-consuming genetic studies. Moignard *et al.*¹ present an innovative approach to modeling such networks from single-cell expression measurements. One time-course study is thereby able to generate numerous hypotheses about genetic interactions. It will be exciting to see how this technology complements genetic studies and uncovers novel regulatory programs controlling organ development.

References

1. Moignard V et al. Nat. Biotechnol 33, 269–276 (2015). [PubMed: 25664528]
2. Swiers G et al. Nat. Commun 4, 2924 (2013). [PubMed: 24326267]
3. Treutlein B et al. Nature 509, 371–375 (2014). [PubMed: 24739965]
4. Guo G et al. Cell Stem Cell 13, 492–505 (2013). [PubMed: 24035353]
5. Moignard V et al. Nat. Cell Biol 15, 363–372 (2013). [PubMed: 23524953]
6. Buettner F et al. Nat. Biotechnol 33, 155–160 (2015). [PubMed: 25599176]
7. Pollen AA et al. Nat. Biotechnol 32, 1053–1058 (2014). [PubMed: 25086649]