Experimental Hematology

Experimental Hematology 2014;42:598–608

Cellular barcoding: A technical appraisal

Shalin H. Naik[a,b], Ton N. Schumacher[c], and Leïla Perié[c,d]

[a]*Molecular Medicine Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia;* [b]*Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia;* [c]*Division of Immunology, Netherlands Cancer Institute, Amsterdam, The Netherlands;* [d]*Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands*

(Received 15 April 2014; revised 1 May 2014; accepted 2 May 2014)

Cellular barcoding involves the tagging of individual cells of interest with unique genetic heritable identifiers or barcodes and is emerging as a powerful tool to address individual cell fates on a large scale. However, as with many new technologies, diverse technical and analytical challenges have emerged. Here, we review those challenges and highlight both the power and limitations of cellular barcoding. We then illustrate the contribution of cellular barcoding to the understanding of hematopoiesis and outline the future potential of this technology. © **2014 ISEH - International Society for Experimental Hematology. Published by Elsevier Inc.**

When Rudolph Virchow wrote in 1858 "omnis cellula e cellula" (every cell from a pre-existing cell), cell theory was established [1]. Along with the work of Louis Pasteur and others, theories of spontaneous generation were discarded, initiating the search into how complex life originates from a single cell.

Pre-occupation with the single cell has waxed and waned over the years; however, the study of how individual stem and progenitor cells make fate decisions to generate complex tissues is currently at the forefront of biology. Although much progress has been made in lower organisms, perennial questions surrounding single cell fate in higher-order animals still dominate. Do all progenitors contribute equally in cell numbers? At what stage is diversity generated? Is the diversity the result of intrinsic or extrinsic processes? Do these processes involve stochastic or deterministic regulation? What are the factors responsible for the generation of diversity?

One of the most well studied systems addressing such questions is the hematopoietic system. With some variations on the theme, the current paradigm states that hematopoietic stem cells divide to give rise to multipotent progenitors, which then broadly restrict into lymphoid, myeloid and megakaryocyte/erythroid progenitors en route to the more differentiated sublineages [2,3]. This sequence can be seen in the often-drawn branching diagrams of hematopoiesis and assumes that progenitors lose multipotency as hematopoiesis proceeds with division and differentiation. Evidence to back this assumption includes the ability of a single stem cell to reconstitute the entire hematopoietic system [4] and the ability of downstream progenitors to make some but not all subtypes in vivo [5].

Over the last decades, our understanding of hematopoiesis has evolved with leaps that often coincided with changes in assays and technology. Originally, Till and McCulloch investigated the ability of transferred progenitors to generate, from a single cell, colony-forming units in the spleens of irradiated recipients [6]. The time at which colonies were harvested from recipients and the typology of cells re-transferred into new recipients allowed researchers to identify short-, intermediate-, and long-term reconstituting cells in the bone marrow, thus establishing the concept of a stem cell. The advent of in vitro soft agar colony-forming assays, developed by Don Metcalf, recapitulated some of the in vivo findings and led to the discovery of colony-stimulating factors and progenitors able to generate differentiated cells of many lineages [7]. Later, the ability to stain for specific cell surface markers and sort cells by flow cytometry revolutionized the dissection of intermediate progenitor stages [5]. Subsequently, the molecular mechanisms governing these processes were identified using knockdowns, knockouts, and reporter mice. Combined, these technologic progressions have generated our current models of hematopoiesis.

Some of the questions listed earlier regarding individual cell fate cannot be fully addressed by these assays. Even if in vitro assays can track individual cell fate, they can bias cell differentiation and may only partially reproduce the complexity of lineage fates. In other words, although these

Offprint requests to: Shalin H. Naik, PhD, Molecular Medicine Division, Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia; E-mail: shalinhnaik@gmail.com and Ton N. Schumacher, Division of Immunology, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands; E-mail: t.schumacher@nki.nl and Leïla Perié, Theoretical Biology and Bioinformatics, Utrecht University, 3584 CH Utrecht, The Netherlands; E-mail: l.perie@nki.nl

0301-472X/$ - see front matter.
http://dx.doi.org/10.1016/j.exphem.2014.05.003

assays reveal what a single cell can produce under the experimental conditions chosen, they do not reveal what the output of a given progenitor would have been in vivo. Finally, in vivo population-based assays, by definition, miss levels of complexity of individual cell commitment. Thus, methods that are able to track single cell fate in vivo are ultimately required to establish true lineage fate.

In vivo tracking of single cell fate has been achieved through hematopoietic reconstitution from a single cell [4], the use of retrovirus-tagged progenitors, with clonal output extrapolated from Southern blots [8,9], or through multiplexed expression of fluorophores [10–12]. However, these methods are either restricted to dozens of clones (e.g., single cell transfer, multiplexed fluorophore expression) or have a restricted dynamic range (Southern blot analyses). Some labs have valiantly scaled up single cell transfer experiments [13–15], yet the process remains labor intensive and limited to the detection of more common output patterns. Although these studies represent landmarks in single cell fate tracking, systems that allow higher throughput, better quantitation, and tracking of large numbers of individual cells simultaneously would benefit the field.

A new technology, termed *cellular barcoding,* originally developed by our group [16,17], with more recent versions from our group [18,19] and others [20–23], is emerging as a powerful tool to address individual cell fates on a large scale. The basic principle underlying cellular barcoding involves the tagging of individual cells of interest with unique heritable identifiers or barcodes (Fig. 1). The barcode collection (or "barcode library") is constructed artificially from semirandom, noncoding stretches of DNA and is delivered into the genome of progenitor cells of interest using a lenti- or retroviral vector. Barcode-labeled cells are then transferred into recipient mice and allowed to develop in vivo into the various lineages.

As the barcode is integrated into the genome, each subsequent daughter cell also inherits this genetic tag. In this way, different cell types can be isolated later, and their genomic DNA assessed for its barcode signature (Fig. 1). Originally, the detection of barcodes was achieved using a custom DNA microarray [16,17], but this technique has now been replaced by next-generation sequencing [18,19,20–23]. The latter affords better quantitation than microarray and allows massively parallel processing of samples by the use of index primers for different samples, which are then pooled for sequencing. By comparing the shared and distinct barcodes between cell types, one can establish progenitor fates at the single cell level on a large scale. Using this technology is akin to doing hundreds of single cell assays simultaneously in one mouse and clearly has power in addressing many questions in the field of single cell development.

As with many new technologies, however, diverse technical and analytical challenges have emerged with cellular barcoding. Here, we review those challenges and highlight both its power and limitations, give examples of experimental results, and outline the future potential of this technology.

Experimental procedure

We have summarized a number of frequently asked questions about cellular barcoding experiments in Box 1. Our advice to research groups contemplating the use of cellular barcoding is to pay specific attention to three experimental factors: (1) library size, (2) the transduction step, and (3) technical replicates.

Library size

One of the very first steps in cellular barcoding technology is choosing the size of the library, that is, the number of different DNA barcodes that should be available. The length of the DNA stretch predicts maximal theoretical diversity. In practice, however, library diversity depends on other variables, in particular the cloning of the library into delivery vectors. Both the length of the barcodes and the size of the libraries differ between the current versions of cellular barcoding [18,20–22]. Increasing the size of the barcode library can be beneficial, as it allows analysis of the fate of a larger number of cells within a single animal without "repeat use" of barcodes (see below). On the other hand, only the exact sequence composition of libraries with a lower diversity can accurately be described presently by second-generation sequencing. Such library sequencing allows the generation of a reference file of "true barcodes," making it straightforward to filter out the large number of sequencing and polymerase chain reaction (PCR) errors in experimental data.

Transduction

Different transduction times, ranging from 6 to 48 hours, have also been used. We advocate a short transduction culture of 6 hours to reduce the chance of biasing the fate of the cells. During transduction, two types of events can occur that have consequences on the interpretation of the data: multiple integration and repeat use.

Multiple integrations of barcodes into one cell (Fig. 2) would influence quantification, as they would read as multiple progenitors with the same fate, rather than one cell labeled with two different barcodes. To avoid such an issue, most protocols aim for low transduction efficiency, thereby reducing the probability of multiple integrations. The transduction efficiency has to be monitored for every system and depends on the amenability of progenitors for transduction. For example, Lu et al. established that at their transduction rate of 50%, $>95\%$ of cells had one integration [21]. In our system, we typically aim for 10–15% transduction efficiency to decrease this chance even further. Multiple integrations within one cell can be considered a relatively minor problem, which would lead to overestimation of the amount of data obtained, but would not influence the output patterns inferred (see Box 1).