STEM CELLS

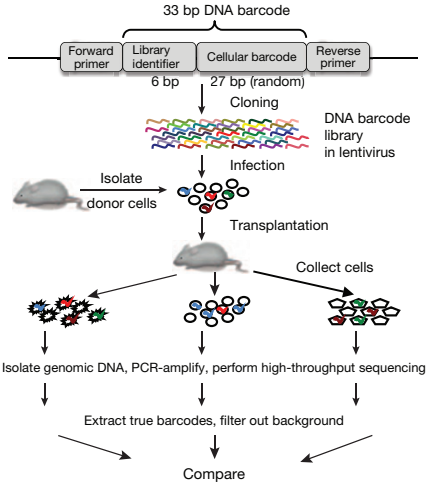
One stem cell at a time

A combination of methods permits clonal tracking of mouse hematopoietic stem cells.

The problem with stem cells, even those in a highly regenerative tissue such as blood, is that they are rare, dynamic and heterogenous. To understand how these cells build and maintain tissue, therefore, it is critical to study the output of single cells. This has been achieved in the hematopoietic system by single-cell transplantation, in which a single purified and transplanted hematopoietic stem cell (HSC) gives rise to all the blood cells in a mouse previously irradiated to eliminate its endogenous HSCs. But these experiments are tedious and require a relatively large number of mice. What is more, only a limited number of individual HSCs can reasonably be examined. Finally, single-cell transplantation is much more suitable for a system such as blood, rather than for studying stem cells in solid organs.

An alternative approach has been to use methods such as viral tagging to uniquely mark single stem cells. If the experiment is designed such that every starting cell sustains a single viral integration event, then the unique genomic integration sites mark both single cells and the clones emanating from them. Rong Lu, Irv Weissman and colleagues at Stanford University now combine this approach with DNA barcoding and high-throughput sequencing to accurately and efficiently track the clonal output of single HSCs in the mouse.

"Most biomedical research investigates specific cell types within a heterogenous cell matrix," says Lu. "We decided to develop techniques to accelerate single-cell studies." The researchers began with a FACS-purified population of mouse HSCs and infected them with lentiviruses carrying a library of molecular barcodes. They chose the titer of the infecting virus to optimize the number of cells infected while still ensuring that most cells received a single viral integration. Each virus carried a 33-base-pair (bp) barcode, with a unique 27-bp sequence and a 6-bp 'identifier' sequence that was identical for all barcodes in the library. Then the researchers transplanted infected cells into irradiated recipient mice and, several months later, examined the distribution of barcodes in blood cell types of interest.



Clonal tracking of mouse HSCs. Barcoded viral tags integrated into HSCs are analyzed with high-throughput sequencing to determine clonal contributions to different blood cell populations. Adapted from *Nature Biotechnology*.

As each barcode sequence initially marks a single cell, the distribution of barcodes in different populations of blood cells after the hematopoietic system was regenerated gives a view of clonal contribution to these populations. But for these high-precision,

GENOMICS

PICKING UP THE PIECES

An improved algorithm for genomic analysis allows scientists to build remarkably accurate and complete genomic sequences from single bacterial cells.

Imagine putting together a jigsaw puzzle with no reference picture. Now imagine doing the same task with many duplicates of some pieces and others that are unique or missing entirely.

Scientists face a similar challenge in attempting genomic analysis of newly discovered bacterial species. However, a new computational strategy devised by scientists at the J. Craig Venter Institute (JCVI) and the University of California at San Diego has proven adept at turning such confusing mountains of fragments into a coherent whole.

Unlike laboratory-friendly bugs such as *Escherichia coli*, most bacterial species are intractable to cultivation, and scientists must make do with limited genetic material. Several years ago, Roger Lasken and colleagues helped usher in a new era of bacterial genomics with a technique known as multiple displacement amplification (MDA). MDA efficiently generates a plethora of amplified sequence fragments from the DNA contents of a single cell; where these fragments overlap, scientists can assemble 'contigs' that span larger genomic segments.

Conventional sequence-assembly algorithms can find MDA-derived data difficult to work with, however. "You get very biased amplification," says Lasken, who is now at the JCVI. "You may have thousands of copies of one part of the sequence and few or no copies of other parts." The latter poses a tougher problem for assembly, and most existing algorithms simply skip genomic segments for which the number of overlapping reads is too limiting for confident contig building. However, this wastes potentially valuable data. Lasken therefore teamed up with University of California at San Diego bioinformatician Pavel Pevzner to design a more efficient algorithm for MDA data analysis.

Instead of starting with a strict threshold for sequence coverage, Pevzner's algorithm begins assembly with a much lower cutoff—enabling potential inclusion of low-coverage regions—and then raises the bar for inclusion at subsequent stages of analysis, as groups of smaller contigs are being assembled into larger contigs. "The real breakthrough is on the informatics side," says Lasken. "Instead of losing those rare reads, this software makes use of them and gets a much more complete assembly."

As proof of concept, they demonstrated that this approach could generate assemblies from individual *E. coli* and *Staphylococcus aureus* bacteria that include a higher proportion of complete

high-throughput experiments to yield meaningful information, technical considerations are paramount.

First, the researchers used simulations based on experimental measurements of barcode library size, diversity and the number of barcodes received per cell, and determined the number of cells that should be infected to ensure a >95% probability of one barcode representing one cell. Second, it was necessary to filter out sequencing errors. For this, Lu and colleagues used the 6-bp library identifier sequence to identify true barcodes. In addition, they developed an algorithm to filter out low-copy-number reads that were likely to be due to background noise.

Lu and colleagues observed 30–50 unique barcodes per mouse in their experiments and distinguished two populations of HSCs with distinct contributions to different blood lineages: one subset of cells was biased to B cells and T cells, the other to granulocytes and B cells. HSC lineage bias is consistent with what has been previously seen with single-cell transplantation experiments, but the combination of techniques used in this work means that many more HSC clones could be monitored with an order of magnitude fewer mice.

So far, the researchers generated 18 barcode libraries, which can in principle be combined in multiplex experiments to increase the scale and efficiency even more. The approach should apply to any cell type that can be infected with a lentivirus, although the complete ablation of endogenous stem cells may prove challenging in systems that regenerate less rapidly than the hematopoietic one. For now, and yet again in adult stem cell biology, blood leads the way.

Natalie de Souza

RESEARCH PAPERS

Lu, R. *et al.* Tracking single hematopoietic cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928–933 (2011).

genes and operons with a lower error rate and fewer misassembled contigs relative to existing algorithms.

Their approach also performed well with individual cells of a previously uncharacterized marine bacterium, generating a genome assembly with larger, higher-quality contigs compared with those produced by older algorithms. Most of the expected metabolic genes appeared to be represented in their assembly, suggesting a high degree of completeness, and the researchers made preliminary deductions about the physiology of this bacterium based on some of the pathways that they identified.

Lasken's team is continuing to improve MDA while Pevzner and colleagues work toward a more streamlined analytical process. "You could conceivably go from finding an organism to having its assembled genome in a week," says Lasken. Such power will undoubtedly prove extremely useful as he and his colleagues at the JCVI continue their efforts to catalog and characterize the numerous bacterial species that make their home in the gut, mouth and other reservoirs of the human body. "We have a huge number of bacteria we know almost nothing about," says Lasken. "If we could even get 5 or 10% of their genome, it could be tremendously interesting."

Michael Eisenstein

RESEARCH PAPERS

Chitsaz, H. *et al.* Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* **29**, 915–921 (2011).

NEWS IN BRIEF

SEQUENCING

The best way to capture exons

Sequencing only the exons in a genome narrows the search for functional variation. Clark *et al.* compared tools that capture the exome by hybridizing mRNA to oligonucleotide baits in solution. Nimblegen baits covered less but overlapped at high density, giving excellent sensitivity for the least sequencing. Illumina and Agilent baits covered more bases at low density, facilitating the discovery of more variants upon deeper sequencing.

Clark, M.J. *et al. Nat. Biotechnol.* **29**, 908–914 (2011).

CHEMICAL BIOLOGY

Incorporating multiple unnatural amino acids

Probes can be site-specifically introduced into proteins by reassigning the stop codon UAG to encode an unnatural amino acid. But incorporation efficiency is low because the stop codon continues to serve as a stop signal. Johnson *et al.* fully reassign the UAG codon in *Escherichia coli* by knocking out the essential release factor 1 and fixing the expression of release factor 2. This greatly improves incorporation of unnatural amino acids, allowing incorporation at multiple sites.

Johnson, D.B.F. *et al. Nat. Chem. Biol.* **7,** 779–786 (2011).

BIOINFORMATICS

High-dimensional single-cell data analysis

Despite recent great technical advances in flow cytometry (allowing up to 17 single-cell parameters to be measured) and mass cytometry (allowing up to 30 or more parameters to be detected), methods for analyzing such high-dimensional single-cell data have lagged behind. Qiu *et al.* now describe an analysis method called spanning-tree progression analysis of density-normalized events (SPADE). SPADE enables the visualization of cellular progressions and hierarchies in a branched-tree structure.

Qiu, P. *et al. Nat. Biotechnol.* **29,** 886–891 (2011).

MICROSCOPY

Deep imaging with STED

To study synapse function in a physiological setting, it is necessary to image deep in the brain. Urban *et al.* describe adaptation to a stimulated emission depletion (STED) super-resolution to facilitate actin imaging inside synapses 120 micrometers below the tissue surface. They equipped their STED microscope with a high-numerical-aperture glycerol-immersion objective lens with a correction collar, which greatly decreased aberrations resulting from light scattering.

Urban, N.T. *et al. Biophys. J.* **101**, 1277–1284 (2011).

BIOCHEMISTRY

Human adipose lipid turnover

Arner *et al.* describe an approach to study lipid turnover in human adipocytes. As a result of above-ground nuclear bomb tests in the 1950s and 1960s, atmospheric ^{14}C levels greatly increased, but since a test ban treaty in 1963, ^{14}C levels have exponentially decreased at a known rate. Comparing ^{14}C incorporation into lipids with atmospheric ^{14}C levels, Arner *et al.* determined that adipocyte lipid turnover is relatively slow, on the order of six times in 10 years.

Arner, P. *et al. Nature* **478**, 110–113 (2011).