Lawrence Berkeley National Laboratory

Recent Work

Title

Single-cell genomics for the masses.

Permalink

<https://escholarship.org/uc/item/9n29c8rt>

Journal

Nature biotechnology, 35(7)

ISSN

1087-0156

Author

Tringe, Susannah G

Publication Date

2017-07-01

DOI

10.1038/nbt.3914

Peer reviewed

Single cell genomics for the masses

Susannah G. Tringe, DOE Joint Genome Institute

In this issue of *Nature Biotechnology*, Lan et al. describe a new tool in the toolkit for studying uncultivated microbial communities, enabling orders of magnitude higher single cell genome throughput than previous methods. This is achieved by a complex droplet microfluidics workflow encompassing steps from physical cell isolation through genome sequencing, producing tens of thousands of low-coverage genomes from individual cells.

A decade ago the only option for interrogating the genomes of uncultivated environmental microbes was bulk DNA extraction and sequencing, which for most natural communities yielded highly fragmented data with error-prone functional and phylogenetic assignments. Increased sequencing capacity has dramatically increased (meta)genome coverage and improved assembly and binning algorithms have made “genome resolved metagenomics” a reality, but the informatic challenges are substantial and some populations, even abundant ones, cannot be successfully reconstructed^{1,2}.

In the late 2000s single-cell genomics emerged as a viable alternative to metagenomics for obtaining genome coverage of uncultivated microorganisms^{3,4}. It has since become a routine application at multiple institutions, typically employing DNA stains and fluorescent activated cell sorting with retrospective organism identification by 16S ribosomal RNA sequencing. Improved techniques now enable the flow sorting of hundreds or even thousands of cells from an individual sample, with automated workflows for downstream lysis, amplification, barcoding and sequencing. But the hundreds of genomes obtainable by this approach still only scratch the surface of the diversity present in many environments. While quality has improved substantially, single cell genomes still present a number of challenges in terms of biases in cell isolation and lysis; massive variations in coverage due to amplification; and contamination.

The current manuscript describes a massively parallel droplet based single cell sequencing pipeline. It builds upon work published last year, describing a droplet microfluidic approach to single molecule deep sequencing⁵. Key to that approach was the ability to generate a high-complexity library of barcode droplets by synthesis of random barcodes attached to sequencing adapters, limiting dilution and PCR amplification (see figure). To extend this approach to whole genomes, the team needed to lyse and extract DNA from individually encapsulated cells prior to molecular barcoding and library creation. This was achieved by embedding cells in agarose droplets such that the genomic material was immobilized during enzymatic lysis and, importantly, the subsequent washing steps necessary to remove the enzymes, detergent and cell debris. These agarose-embedded single genomes were then re-encapsulated in droplets pre-loaded with reagents for transposase-mediated fragmentation (“tagmentation”), minimizing cross-droplet contamination. Finally, the fragmented DNA was merged with the droplet barcode library, loaded with PCR reagents and uniquely barcoded Illumina adapters, amplified and barcoded by thermal cycling, and purified for sequencing (see figure). Once generated, the sequences could be clustered by barcode and individually classified by function and taxonomy. In all,

this process required 4 custom microfluidic devices of 3 different designs, with additional washing, PCR and emulsion breaking steps taking place in conventional lab vessels.

Importantly, this workflow lacks the multiple displacement amplification employed in virtually all single cell genome approaches, and thus avoids some of its pitfalls, including chimerism from template switching, branching structures that interfere with library generation, and high coverage biases. However, the resulting genomes are very low coverage: 0.1-1% of each individual genome is covered even in a simple mock community. While the authors suggest that deeper sequencing could produce higher coverage – on average <100 kbp of sequence data were generated per barcode in these pilot experiments - inefficiency in the tagmentation reaction is certain to limit genome recovery. However, if multiple single cells from the same population can be identified and pooled, good genome coverage is achievable.

Given the low individual genome coverage, the method is best suited to investigate co-occurrence of genes or genome features. As a proof of principle, Lan et al. investigated the distribution of antibiotic resistance genes (ARGs) in a coastal seawater community, finding putative ARGs in a nonrandom 0.3% of barcode groups. However their results bear the hallmark of another gremlin that bedevils single cell genomics – contamination. They find beta-lactamase, the most abundant ARG in their data, to be strongly associated not only with *Alteromonas*, a common marine bacterium, but also with *Propionibacterium*, a ubiquitous member of the human skin microbiome. This is not surprising given the extensive handling necessary for these experiments, but highlights the importance of ultra-clean workflows in single cell investigation ⁶. On the other hand, the recovery of all 10 genomes from a mock community, even if only at very low frequency, is encouraging as compared to other single cell methods⁷.

This has potentially exciting implications for host-virus interactions, evolutionary genomics and microbial ecology. Currently, in both genomes reconstructed from metagenomes and single cell genomes, it can be difficult to determine which “anomalous” sequences – such as viruses, plasmids and gene islands that differ in sequence composition from the main genome – truly belong to the genome under investigation and which are merely the result of contamination or misassignment. While this is still true of any individual barcode group obtained by this method, the tens of thousands of cells examined enables a statistical assessment of co-occurrence, heightening confidence in genomic patterns that are observed many times and enabling deeper investigation of phenomena like horizontal gene transfer.

Currently the barcode groups represent very low coverage genomes, but provide invaluable physical linkage data - just as long range mate pair data can connect contigs into scaffolds, shared barcodes scattered across the genome can connect contigs into genome bins. They thereby complement bulk methods like shotgun metagenomics and could augment or obviate the multiple sample interrogation necessary for abundance-based binning. In sum, Lan et al. have brought us one step closer to a time when complex environmental communities can truly be characterized at the granularity of individual cells, and it comes at a time when similar advances are being made in transcriptomics, proteomics and

metabolomics⁸. This unprecedented access to the microbial biosphere will undoubtedly change our understanding of community interactions, allowing each individual to shine.

1. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219 (2016).
2. Sharon, I. et al. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome research* **25**, 534-543 (2015).
3. Podar, M. et al. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Applied and environmental microbiology* **73**, 3205-3214 (2007).
4. Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M. & Lasken, R.S. Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol* **11**, 198-204 (2008).
5. Lan, F., Haliburton, J.R., Yuan, A. & Abate, A.R. Droplet barcoding for massively parallel single-molecule deep sequencing. *Nat Commun* **7**, 11784 (2016).
6. Woyke, T. et al. Decontamination of MDA reagents for single cell whole genome amplification. *PloS one* **6**, e26161 (2011).
7. Clingenpeel, S., Clum, A., Schwientek, P., Rinke, C. & Woyke, T. Reconstructing each cell's genome within complex microbial communities-dream or reality? *Frontiers in microbiology* **5**, 771 (2014).
8. Bock, C., Farlik, M. & Sheffield, N.C. Multi-Omics of Single Cells: Strategies and Applications. *Trends Biotechnol* **34**, 605-608 (2016).

Figure: A workflow for droplet microfluidic single cell genome sequencing