Brad Goodell/PhotoDisc/Getty

➡ GENETIC VARIATION

Mapping inversions in single cells

A novel approach that combines DNA template strand sequencing (Strand-seq) with a newly developed software package, Invert.R, can map and genotype putative inversions genome-wide in an untargeted manner in single cells.

Genetic variation among individuals fundamentally underlies phenotypic diversity and disease susceptibility, but the identification of structural variation has traditionally lagged behind that of single-nucleotide polymorphisms. In particular, copy-number-neutral genomic rearrangements such as inversions, which alter the orientation but not the content of a DNA segment, have remained largely undefined in the human genome, owing to a lack of adequate tools.

Strand-seq is a previously established single-cell sequencing technique that can distinguish parental DNA template strands inherited by daughter cells after mitosis based on the prior incorporation of the thymidine analogue 5-bromo-2′-deoxyuridine (BrdU) during DNA replication. BrdU-positive DNA strands are ablated during genomic library construction to ensure the selective sequencing of only BrdU-negative template strands for each chromosome of a single cell.

Sanders *et al.* used Strand-seq to generate libraries from bone marrow cells of a male donor. Sequence reads corresponding to an inverted DNA segment map to the complementary DNA strand of a reference genome, with respect to flanking sequences, highlighting inverted genomic regions even in highly repetitive regions of the genome. Inversions were defined as a non-reference genotype at a given loci in at least two cells. The authors highlight one Strand-seq library, corresponding to one cell, in which 21 putative inversions could be determined, including two previously characterized inversions on chromosomes 8p23 and 7q11.

To enable high-throughput analyses in multiple cells, the authors developed the R-based bioinformatics pipeline Invert.R, which systematically interrogates each single-cell library to localize and genotype putative inversions based on read alignments in Strand-seq libraries. The software was validated by analysing the inversions on chromosomes 8p23 and 7q11, which showed that the breakpoints located by Invert.R matched those previously established using different orthogonal techniques, such as fluorescence *in situ* hybridization or next-generation sequencing. Moreover, the resolution of breakpoint mapping using Invert.R was similar to that of manual breakpoint predictions. Reducing the sequencing depth of the library reduced the ability of the software to detect smaller inversions, but very low genomic coverage (0.05x) nevertheless still allowed for inversions greater than 25 kb to be called.

Scaling their approach to a sample population, the authors generated 47 Strand-seq libraries from a pool of 353 separate cord blood donors. Overall, 111 polymorphic inversions (corresponding to 34.9 Mb (1.13%) of the genome) were identified in the sampled population, with multiple inversions clustering in highly polymorphic domains on some chromosomes. Substantial intercellular heterogeneity in the structural composition of each genome was revealed by comparing the genotype of polymorphic loci, which suggests that the relatedness between cells of a mixed sample can be determined by the inversions mapped in single Strand-seq libraries.

Finally, Sanders *et al.* generated a genome-wide map of all inversions present within an individual genome. To this end, they analysed 140 Strand-seq libraries from bone marrow cells of a single man, identifying 86 inversions, corresponding to 34.4 Mb (1.11%) of the genome. They compared this inversion profile with that generated from 106 Strand-seq libraries of cord blood cells from a female newborn, in which 60 inversions, comprising 23.3 Mb (0.77%) of the genome, were identified. Several predicted inversions overlapped almost precisely with those previously validated using alternative techniques.

Taken together, this new framework allows a more global approach to the analysis of inversions by facilitating population-based studies through the comparison of multiple genomes in a high-resolution, high-throughput manner.

Linda Koch

ORIGINAL ARTICLE Sanders, A. D. *et al.* Characterizing polymorphic inversions in human genomes by single cell sequencing. *Genome Res.* http://dx.doi.org/10.1101/gr.201160.115 (2016)
FURTHER READING Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016)