RESEARCH
HIGHLIGHTS

⤷ **TECHNOLOGY**

A new tool for analysing structural variation

A new large-scale, high-throughput method — paired-end mapping (PEM) — for identifying structural variants in the human genome has been described. It provides insights into how these variants arise and suggests that structural variation is even more prevalent in the human genome than previously anticipated.

Two approaches have dominated studies of structural genomic variation: array comparative genomic hybridization (CGH) and fosmid paired-end sequencing; however, the former maps structural variants at a resolution that is insufficient for detecting the actual breakpoints, whereas the latter is laborious. Motivated by this problem and the fact that there are few methods for efficient detection of structural variants of <10 kb, the authors developed PEM. It involves generating paired ends of 3-kb fragments, which are then sequenced using the 454 technology. Significant differences between the paired-end reads and the corresponding reference genomic regions reveal the presence of structural variation. Deletions, inversions and mated and unmated insertions of ≥3 kb, and simple insertions of 2–3 kb, can be detected in this way.

The authors tested PEM on two individuals, both of whom had previously been studied for structural variation. Variants were validated in several ways, including by PCR and comparison with the Database of Genomic Variants. Overall, ~ 1,300 structural variants were identified, implying that structural variation is more significant than single-nucleotide variation when considering the number of bases that are affected. Most PEM-identified variants are small (65% are <10 kb) and several had not been detected previously. Notably, 45% of the variants are shared between these two individuals; considering that one is of African and the other of presumed European ancestry, this finding indicates that many structural variants are common and presumably ancient.

454 sequencing of the breakpoints indicated that many structural variants were associated with segmental duplications and short- to medium-sized repetitive elements such as short and long interspersed nuclear elements (SINEs and LINEs). Contrary to previous reports, no enrichment of *Alu* elements was seen near breakpoints. On the basis of further manual analysis of breakpoint sequences, the authors estimate that 56% of insertion and deletion polymorphisms (indels) arise as a result of non-homologous end joining; a surprise given that many structural variants are associated with segmental duplications. A further 30% of variants occur as a result of retrotransposition, most of which is due to LINE elements. Non-allelic homologous recombination seems to result in structural variation only relatively rarely, but where it does it mainly occurs between LINEs, long terminal repeat (LTR) elements or SINEs. In 4 of the 14 inversions analysed the authors found evidence for homologous recombination between inverted repeats as the underlying mechanism for the variation.

Despite a number of advantages, as with other methods PEM cannot easily identify structural variants that lie in regions of multiple copies of highly similar and long repeats. However, the authors are confident that the method can be refined to become the ultimate tool for analysing structural variation.

Magdalena Skipper

> ❝ …many structural variants are common and presumably ancient. ❞