Flip-flop genomics: Charting inversions in the human population

Sophie Lanciano, Gael Cristofari

Flip-flop genomics: charting inversions in the human population

Sophie Lanciano[1] and Gael Cristofari[1,*]

[1] University Cote d'Azur, Inserm, CNRS, Institute for Research on Cancer and Aging of Nice (IRCAN), Nice, France

* Correspondence and lead contact: gael.cristofari@univ-cotedazur.fr

ABSTRACT

Detecting large genomic inversions has long been challenging. In a new study, Porubsky *et al.* resolve these complex rearrangements in 41 individuals and discover wide regions that undergo recurrent inversions, some of which even toggle back and forth (Porubsky et al., 2022). Many of these regions are associated with genomic disorders.

MAIN TEXT

Concomitant with the development of high throughput sequencing technologies, a wide variety of genetic variants have been detected in human genomes, ranging from single-nucleotide variants (SNVs), and small (< 50 base pairs) insertions or deletions (indels) to large (kilobase to megabase range) structural variants (SVs) (**Figure 1A**). SVs are extremely heterogeneous in nature and size and include copy number variations (large deletions or insertions, transposable element insertions, duplications and other amplifications) and translocations. These alterations have been associated with a broad range of genetic diseases, including neurodevelopmental disorders. Thus detecting them is essential to interpret genomic data (Carvalho and Lupski, 2016; Ho et al., 2020). In this issue of *Cell*, Porubsky *et al.* tackle another type of SVs, inversions, which have been largely overlooked until now (Porubsky et al., 2022).

While SNVs and indels are relatively easy to identify from short-read sequencing data, the identification of larger genetic rearrangements can be more challenging (**Figure 1A**). In theory, the boundaries of an inversion could be easily detected by paired-end short-read sequencing using discordant read pairs with abnormal orientation between reads of the same pair, and split reads that span the inversion breakpoint. Unfortunately, in practice, inversions are frequently embedded in repeated sequences such as transposable elements or segmental duplications, and are hardly accessible using common short-read strategies (Ho et al., 2020). It is estimated that more than 50% of human inversions are flanked by segmental duplications (Chaisson et al., 2019). The latter are large blocks of DNA (up to hundreds of kilobases) that are present at least twice in a genome. An unexpectedly high number of these duplicated regions are present in primate genomes as compared to other mammals. The current model proposes that non-allelic homologous recombination (NAHR)

between such repeats, when oriented in opposite direction, promotes the apparition of recurrent and internal inversions (Carvalho and Lupski, 2016), also referred as “inversion toggling” (Zody et al., 2008) (**Figure 1B**).

Emerging technologies that overcome the limitations of short-read sequencing improve the detection of SVs in repetitive regions (Ho et al., 2020) (**Figure 1A**). For instance, long reads (tens of kilobases) offered by Pacific Bioscience (PacBio) or Oxford Nanopore Technologies (ONT) disambiguate mapping location in repeated regions and can directly detect SV breakpoints. These long reads sometime span the entire SV sequence, including potential internal rearrangements. Collectively, these information can provide structural insights into phased haplotypes (Ho et al., 2020). However, long read sequencing has a few drawbacks. Beside an error rate that limits the accurate detection of SNVs and small indels, it is not fully adapted for the detection of very large genomic alterations, such as balanced inversions embedded in duplications of several hundred kilobases or in the order of megabases. As a means to provide a comprehensive characterization of structural variants in human and other primates, including complex inversion events, Porubsky *et al*. take advantage of single-cell strand sequencing (Strand-seq), a technique based on selective but shallow sequencing of the template strand in isolated daughter cells upon cell division, which can inform on the orientation of DNA at chromosome-length scale (Sanders et al., 2016).

This approach, applied to a human diversity panel of 41 individuals, with ancestry from several continents, provides the most complete genome-wide survey of inversions (> 50 bp) in the human population, revealing more than 700 inversion events, two third of which were validated by at least one orthogonal method, such as long-read sequencing or optical mapping. As expected, large balanced inversions are mainly located in repeat-rich regions, either segmental duplications or retrotransposons, supporting their role in human genome plasticity. Another common form of inversions are short inversions (<2 kb) frequently found at the 5’ end of L1 retrotransposon insertions (**Figure 1A**) and resulting from an alternative integration pathway during their mobilization, known as twin priming (Ostertag and Kazazian, 2001). The precise analysis of the internal breakpoints and 5’ junctions reinforces the idea that microhomology-mediated end joining (MMEJ) plays a preponderant role in the resolution of L1 retrotransposition (Zingler et al., 2005).

Extensive analysis of large balanced inversions shows that at least 40 of them, representing as much as ~0.6% of the genome, are subject to inversion toggling (**Figure 1B**). Although this flip-flop property was already observed throughout the evolution of primate genomes (Porubsky et al., 2020), this new study indicates that the process also acts on a much shorter evolutionary scale. In addition, toggling segments are not homogeneously distributed along chromosomes, but are enriched within the sex chromosomes. The absence of homologous

recombination and X chromosome hemizygosity have long been proposed to promote NAHR and thus inversions associated with repeats (Carvalho and Lupski, 2016). The authors now provide support for this model at the population level. They also document the dynamics of recurrent inversions. First, they estimate the rate of toggling between $3.4x10^{-6}$ and $2.7x10^{-4}$ per locus per generation. Second, by leveraging phased SNVs, they propose an evolutionary scenario for several recurrent inversions, and find evidence of both independent parallel events and serial events (**Figure 1B**). In the future, defining the precise break points of inversions within segmental duplications or retrotransposons could provide a direct way to trace back the history of inversions if these repeats are sufficiently divergent. Indeed, each recombination event is expected to occur at a distinct position.

Finally, the consequences of these inversion hotspots, when located nearby a gene, are multiple. They can lead to the formation of chimeric-transcripts or perturb interaction between genes and regulatory sequences, such as enhancers, by disrupting topologically associating domains (TADs) boundaries (Ho et al., 2020). Strikingly, regions of toggling also overlap with regions associated with neurodevelopmental genomic disorders. Beyond direct effects of inversions, the authors predict haplotypes that could act as pre-mutational states for subsequent morbid rearrangements (**Figure 1B**), which could have important implications for the diagnosis of genomic disorders (Carvalho and Lupski, 2016).

SV-detection has been driven by methodological advances in genome-wide genotyping assays, each technological advance providing access to a new type of variations. Almost 20 years after the initial release of the reference human genome, Porubsky *et al.* make a giant effort to characterize the inversion landscape of the human genome, an important resource for human genetics, and set a new standard for future genomics studies.

ACKNOWLEDGMENTS

Work in the laboratory of GC is supported by the French government, through the Agence Nationale de la Recherche (Idex UCAJEDI, ANR-15-IDEX-01; Labex SIGNALIFE, ANR-11-LABX-0028-01; ImpacTE, ANR-19-CE12-0032; ActiveLINE, ANR-21-CE12-0001), the Fondation pour la Recherche Médicale (FRM, DEQ20180339170), Inserm (GOLD cross-cutting program on genomic variability), CNRS (GDR 3546).

DECLARATION OF INTERESTS

SL declares not competing interests. GC is an unpaid associate editor of the journal *Mobile DNA* (Springer-Nature).

REFERENCES

Carvalho, C.M.B., Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. 17, 224–238.

Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. 10, 1784.

Ho, S.S., Urban, A.E., Mills, R.E. (2020). Structural variation in the sequencing era. Nat. Rev. Genet. 21, 171–189.

Ostertag, E.M., Kazazian, H.H. (2001). Twin Priming: A Proposed Mechanism for the Creation of Inversions in L1 Retrotransposition. Genome Res. 11, 2059–2065.

Porubsky, D., Sanders, A.D., Höps, W., Hsieh, P., Sulovari, A., Li, R., Mercuri, L., Sorensen, M., Murali, S.C., Gordon, D., Cantsilieris, S., et al. (2020). Recurrent inversion toggling and great ape genome evolution. Nat. Genet. 52, 849–858.

Porubsky, D., Höps, W., Ashraf, H., Hsieh, P., Rodriguez-Martin, B., Yilmaz, F., Ebler, J., Hallast, P., Maggiolini, F.A.M., Harvey, W.T., et al. (2022). Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. Cell

Sanders, A.D., Hills, M., Porubský, D., Guryev, V., Falconer, E., Lansdorp, P.M. (2016). Characterizing polymorphic inversions in human genomes by single-cell sequencing. Genome Res. 26, 1575–1587.

Zingler, N., Willhoeft, U., Brose, H.-P., Schoder, V., Jahns, T., Hanschmann, K.-M.O., Morrish, T.A., Löwer, J., Schumann, G.G. (2005). Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. Genome Res. *6*, 780–789

Zody, M.C., Jiang, Z., Fung, H.-C., Antonacci, F., Hillier, L.W., Cardone, M.F., Graves, T.A., Kidd, J.M., Cheng, Z., Abouelleil, A., et al. (2008). Evolutionary toggling of the MAPT 17q21.31 inversion region. Nat. Genet. 40, 1076–1083.

FIGURE

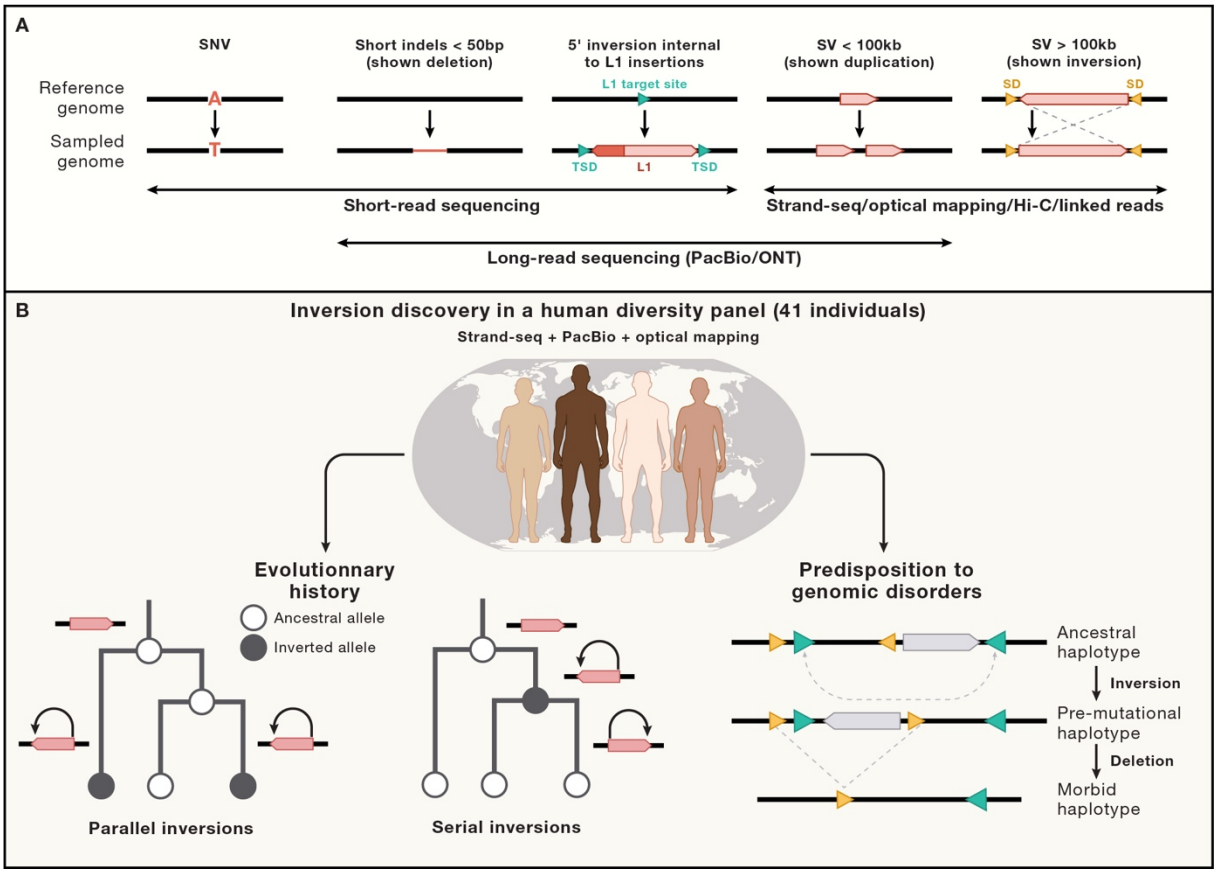

Figure 1 – Resolving inversions provides new insights into human genome dynamics and genomic disorders. (**A**) Exploring the full spectrum of genetic variations (top) requires a variety of genome-wide genotyping assays (bottom) acting at different scales ranging from single base pair to megabases. Variable regions are highlighted in red or pink. For the sake of simplicity, only a deletion is depicted for indels; similarly, only a duplication and an inversion are shown to exemplify SVs. The insertion of transposable elements, such as L1 (or Alu and SVA, not shown) is a particular form of SV. L1 integration is often associated with an inversion at its 5' end (shown in red). SNV: single nucleotide variant; indel: insertion or deletion; L1: long interspersed element 1; TSD: target-site duplication; SV: structural variant. (**B**) The landscape of inversions obtained in a human diversity panel of 41 individuals highlights the evolutionary history of recurrent inversions. Some events were inferred to have occurred independently in several lineages (parallel inversions). Evidence for serial toggling along the same lineage were also observed (serial inversions). The study also identifies pre-mutational states for genomic disorders, *i.e.* alleles for which the inversion event can predispose to subsequent morbid rearrangements (right). A hypothetical situation is shown here, where a functional element (*e.g.* gene, promoter, enhancer, etc) (solid grey arrow) is flanked by a large (green triangles) and a short (yellow triangles) segmental duplications. Upon inversion between the green inverted repeats, one of the yellow repeats is inverted. In

the new configuration, the pre-mutational state, the pair of yellow repeats is now in direct orientation. In turn, they can recombine by non-allelic homologous recombination leading to the deletion of the functional element, the formation of a morbid allele and a genomic disorder. Similarly, inversions can produce “protective haplotypes” that prevent further morbid rearrangement (not shown).