A worldwide map of human's structural variants

Francesco Montinaro[1], Cristian Capelli[2]

[1]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia
[2]Department of Zoology, University of Oxford, Oxford, UK

The completion of the first draft of the human genome in 2001 paved the way to the investigation of the biological impact of genetic variants [1,2]. While the initial focus was mostly on easy to score markers, SNPs and STRs, it was clear since the beginning that substantial variation was hidden in the structural architecture of the genome. Deletions, duplications, insertions and inversions are in fact classes of structural variants responsible for a large degree of variability among individuals [3,4]. However, a full appreciation of the significance of structural variation in shaping human genomic diversity was prevented for many years by the narrow set of surveyed populations, small sample sizes, and methodological limitations.

Addressing all these issues, Almarri and colleagues [5] have now assembled an extensive catalogue of human structural variants, collected from almost 1,000 individuals in fifty-four worldwide populations. The study identified more than 120,000 structural polymorphisms, of which 80% were observed for the first time. Many of these new variants were found in the newly investigated populations, highlighting the importance of study panels including individuals from under-represented human groups. Most of these novel regions are small, just a few hundred bp long, but many reside within or in the proximity of exons, the portion of DNA which codes for proteins and have therefore the potential of being functionally relevant. Moreover, the intersection of novel sequencing technology, which enables better characterisation of highly complex genomic regions, with advanced bioinformatic analysis led to the discovery of almost 2 million base pairs of the genome which were not reported in the reference human genome.

All the analysed classes of structural variation show a substantial degree of stratification, with populations from the same continent sharing more of the same variants than groups from different areas. Although most of this differentiation is due to random changes in allele frequencies, a closer inspection revealed a few genomic features which might be favoured by selection. As an example, individuals living in the lowlands of Papua New Guinea show a frequent deletion at the *HBA2* gene, which codes for one of the subunits of haemoglobin, the protein responsible for the transport of oxygen in the blood. Interestingly, the deletion is at low frequency in neighbouring populations living in the highlands; it is possible that the deletion is protective against malaria, which is common in the lowland but absent in the highlands. Other population-specific variants possibly under selection were identified in hunter-gatherer groups from Central Africa and Brazil.

The identified structural differences included also multiallelic copy number variants, in which the number of copies of a specific genomic region is variable among individuals. Of interest were a number of so-called “run-away duplications”, low copy number variants that expanded in number only in specific groups of populations. East Asian populations presented as many as 18 copies of a region upstream the olfactory receptor gene *OR7D2*, while Oceanians were characterised by a more pronounced expansion in the gene for sulfotransferase SULT1A1, a drug metabolising enzymes. Such regional variations raise the possibility that many of these expansions were favoured by local selective pressures.

One of the main advances in the field of human evolution has been the discovery of genetic fragments inherited from archaic hominins in individuals of non-African origin (although a small fraction is present also in Africa; [6–8]) . Exploring the structural variants shared with Denisovan and Neanderthal but not with Africans, the investigators were able to uncover thirteen deletions/duplications inherited from our archaic cousins and showing extreme population differentiation. Some Denisovan variants were observed only in Oceanians or were shared among Oceanians and East Asians, confirming the occurrence of multiple admixture events between *H. sapiens* and these archaic hominins. Notably, a Denisovan-derived short deletion found at very high frequency in Oceanian populations was associated with approximately 1% of autism cases. Similarly, a deletion inherited from Neanderthals might have implications in the efficacy of therapeutic treatments for lymphomas, a specific type of leukaemia and autoimmune diseases. It is possible that at least some of the introgressed variation conferred an evolutionary advantage. Since many of the archaic-derived structural variants might have a role in immune response, it is possible that when humans expanded out of Africa and occupied new areas useful genetic features were “borrowed” from archaic hominins, who were better adapted to local pathogens.

The catalogue assembled by Almarri et al represents a fundamental advancement for the understanding of the formation and distribution of the genetic variation among individuals and populations. However, despite this unprecedented effort, populations from several geographic regions remain largely uncharacterised. Including more samples from such regions will be essential to fully characterize the genomic complexity present in our species, which in turn will allow for a universal development of translational studies and clinical applications [9].

Nevertheless, the newly generated record of variation provides an important set of variants that can be now explored in other populations. More importantly, these genomic features can now be investigated for their functional relevance and potential association with complex phenotypes, many medically relevant. Hopefully, the combination of *in vitro* and *in silico* approaches here implemented will now lead to the routine detection of this type of variation in humans and other species [10], making possible to access this mostly untapped source of genetic variation and reveal in full its evolutionary significance.

Figure 1. Structural variants showing extreme continental differentiation. The map shows the geographic origins of the 54 analysed populations. Symbols along chromosomes show the approximate genomic location of some of the most differentiated structural variants highlighted in Almarri et al. 2020, with colours matching the continental location in the map. Circles refer to variants putatively inherited from Neanderthals or Denisovans.
Image Credits: F Montinaro, A. Corliano’.

References

1. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. Science *291*, 1304–1351.

2. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

3. Sudmant, P.H., Mallick, S., Nelson, B.J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B.P., Baker, C., Nordenfelt, S., Bamshad, M., *et al.* (2015). Global diversity, population stratification, and selection of human copy-number variation. Science *349*, aab3761.

4. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.-Y., *et al.* (2015). An integrated map of structural variation in 2,504 human genomes. Nature *526*, 75–81.

5. Almarri, M.A., Bergström, A., Prado-Martinez, J., Yang, F., Fu, B., Dunham, A.S., Chen, Y., Hurles, M.E., Tyler-Smith, C., and Xue, Y. (2020). Population Structure, Stratification, and Introgression of Human Structural Variation. Cell. Available at: http://dx.doi.org/10.1016/j.cell.2020.05.024.

6. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., *et al.* (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.

7. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., *et al.* (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

8. Chen, L., Wolf, A.B., Fu, W., Li, L., and Akey, J.M. (2020). Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. Cell *180*, 677–687.e16.

9. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. Cell *177*, 26–31.

10. Kronenberg, Z.N., Fiddes, I.T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O.S., Underwood, J.G., Nelson, B.J., Chaisson, M.J.P., Dougherty, M.L., *et al.* (2018). High-resolution comparative analysis of great ape genomes. Science *360*. Available at: http://dx.doi.org/10.1126/science.aar6343.