SEGMENTAL DUPLICATIONS (SD) IN A COMPLETE HUMAN GENOME

On June 17, 2021, we published a post on "The complete sequence of a human genome", pre-published in BioRxiv. The paper is now out in a special issue of Science[1], which includes additional articles on the subject. From a cytogenetic point of view, the one on segmental duplications[2], by Eichler's group, is relevant.
SD and Copy Number Variations (CNV) are the major source of gene evolution and genome variations within and between species, especially the apes. Their precise identification and sequencing were a major technical problem when using the short reads, as Illumina technology does. Third generation sequencing, as illustrated in the above-mentioned post, has completely overcome the problem and now a sequence of the entire human genome, from telomere to telomere (T2T), has been generated.
SDs were estimated to represent approximately 5% of the human genome. The percentage has now risen to 7%. Their precise definition in the 6 individuals studied allowed substantial progress in the understanding of expression and regulation through the methylation of the duplicated genes. For example, the resolved structure of lipoprotein A, including the expanded kringle IV repeat domain, showed that reduced copies of the latter domain are among the strongest genetic associations with cardiovascular disease.
Please note: the cytogenetics part of this work was carried out by an ECA cytogeneticist (Mario Ventura, co-author).

1. https://www.science.org/doi/10.1126/science.abj6987
2. https://www.science.org/doi/10.1126/science.abj6965