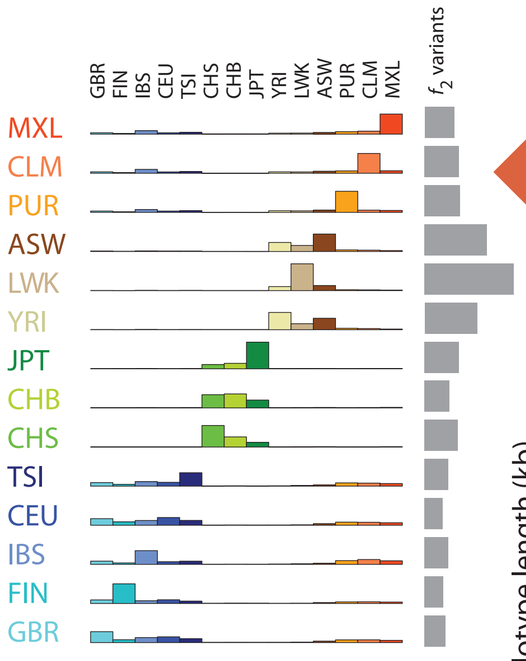
The Wellcome Trust Centre for
Human Genetics

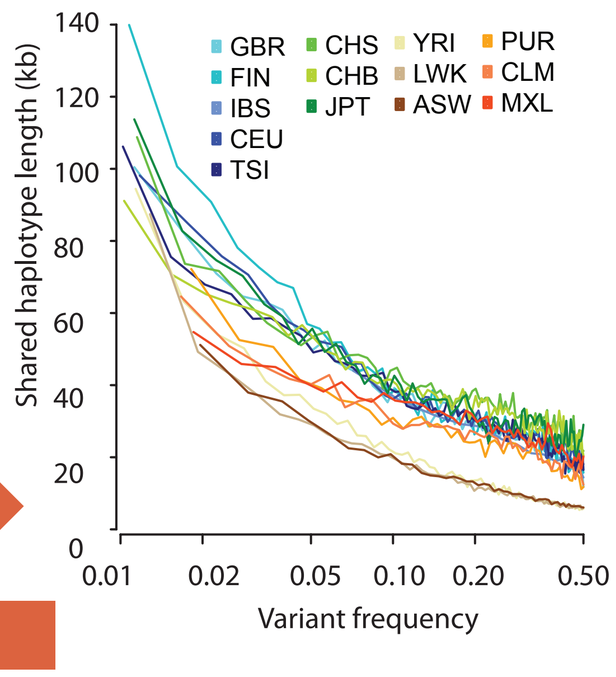
UNIVERSITY OF OXFORD

HapMap and 1000 Genomes: High-resolution maps of human genetic variation

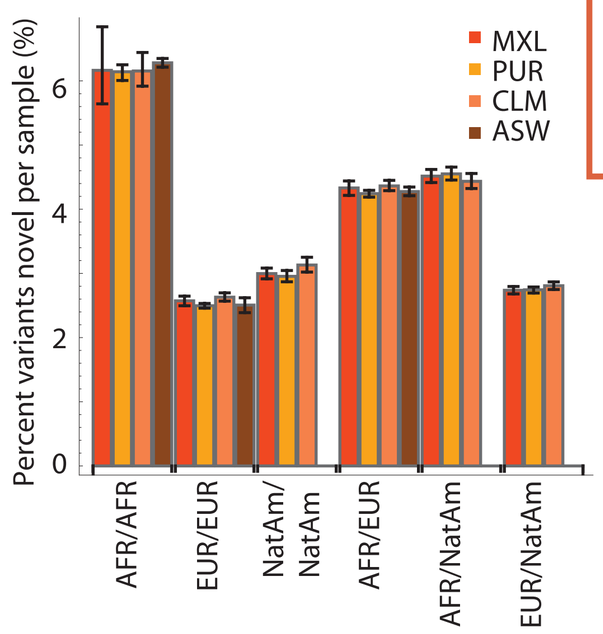


Allele sharing within and between populations: Sharing of f2 variants, those found exactly twice across the entire sample, within and between populations. Each row represents the distribution across populations for the origin of samples sharing an f2 variant with the target population (indicated by the left-hand side). The grey bars represent the average number of f2 variants carried by a randomly chosen genome in each population

Median length of haplotype identity between two chromosomes that share variants of a given frequency in each population



The average proportion of variants that are new (compared with the pilot phase of the project) among those found in regions inferred to have different ancestries within ASW, PUR, CLM and MXL populations



Population abbreviations: ASW, people with African ancestry in Southwest United States; CEU, Utah residents with ancestry from Northern and Western Europe; CHB, Han Chinese in Beijing, China; CHS, Han Chinese South, China; CLM, Colombians in Medellin, Colombia; FIN, Finnish in Finland; GBR, British from England and Scotland, UK; IBS, Iberian populations in Spain; LWK, Luhya in Webuye, Kenya; JPT, Japanese in Tokyo, Japan; MXL, people with Mexican ancestry in Los Angeles, California; PUR, Puerto Ricans in Puerto Rico; TSI, Toscani in Italia; YRI, Yoruba in Ibadan, Nigeria. Ancestry-based groups: AFR, African; AMR, Americas; EAS, East Asian; EUR, European

WHAT WAS KNOWN

- A pair of human genomes vary, on average, at one position every 1000 base pairs
- Most of these differences result from common genetic variants (present in at least 1 in 20 individuals)
- However, most genetic variants are rare (found in fewer than 1 in 100 individuals) and often restricted to a particular geographical region
- Genetic variation has a particular structure arising from the interplay between recombination, demographic history and selection
- Characterising the structure of common and rare variation is critical to designing and interpreting studies to map the genetic basis of complex and rare diseases

WHAT WE DID

- As part of international consortia, initially the HapMap Project and later the 1000 Genomes Project, we carried out genome-wide characterisation of common variation and subsequently rare variation through SNP genotyping and whole genome sequencing from thousands of individuals from across the world
- We released the maps of human genomic variation as an open-source resource, with no restrictions on data usage
- We developed tools for manipulating data, standards for data sharing and methods for interpreting and using genetic variation in the study of human disease, as well as gaining many insights into the demographic and evolutionary history of humans

WHAT THIS ADDS

- The final release of the 1000 Genomes project is a validated haplotype map from 2,504 individuals consisting of 88m single nucleotide polymorphisms, 3.6m short insertions and deletions, and more than 60,000 larger structural rearrangements
- We found that individuals from different populations carry different profiles of rare and common variants
- We showed that low-frequency variants show substantial geographic differentiation, which is further increased by the action of purifying selection
- We found that each individual carries hundreds of rare non-coding variants at conserved sites, including transcription-factor-binding sites, multiple loss-of-function variants that knock-out gene function and a handful of variants known to cause genetic disorders when present in two copies
- This publicly-available resource enables analysis of common and low-frequency variants in individuals from diverse, including admixed, populations

REFERENCES

A haplotype map of the human genome.

The International HapMap Consortium. Authors include G A McVean [Project Leader], P Donnelly [Principal Investigator] & L R Cardon [Principal Investigator]

Nature 2005 437, 1299–132.

A second-generation human haplotype map of over 3.1 million SNPs.

The International HapMap Consortium. Authors include G A McVean [Project Leader], P Donnelly [Principal Investigator] & L R Cardon [Principal Investigator]

Nature 2007 449, 851–861.

A map of human genome variation from population-scale sequencing.

The 1000 Genomes Project Consortium [authors include P Donnelly, G A McVean, A Auton, Z Iqbal, G Lunter, J L Marchini & S Myers]

Nature 2010 467, 1061–1073.

An integrated map of genetic variation from 1,092 human genomes.

The 1000 Genomes Project Consortium. Authors include G A McVean, P Donnelly, G Lunter, J L Marchini, S Myers, A Gupta-Hinch, Z Iqbal, I Mathieson, A Rimmer, D K Xifara & A Kerasidou

Nature 2012 491, 56–65.

A global reference for human genetic variation.

The 1000 Genomes Project Consortium.

Nature 2015 526, 68–74.