GENCODE v19
human genome annotation

Human reference
genome build 37

1000 Genomes Phase 3 whole
exome sequencing data

Generate exon junction libraries
(true and decoy libraries)

Alignment (BWA)

Extract all unmapped reads

Report exon junction-read
alignment, controlling false call
rate ≤ 0.05

Identify genes having novel
retroduplication(s), with
supporting exon junctions

1000 Genomes Phase 3 whole
genome sequencing data

Detect insertion sites for identified
genes,
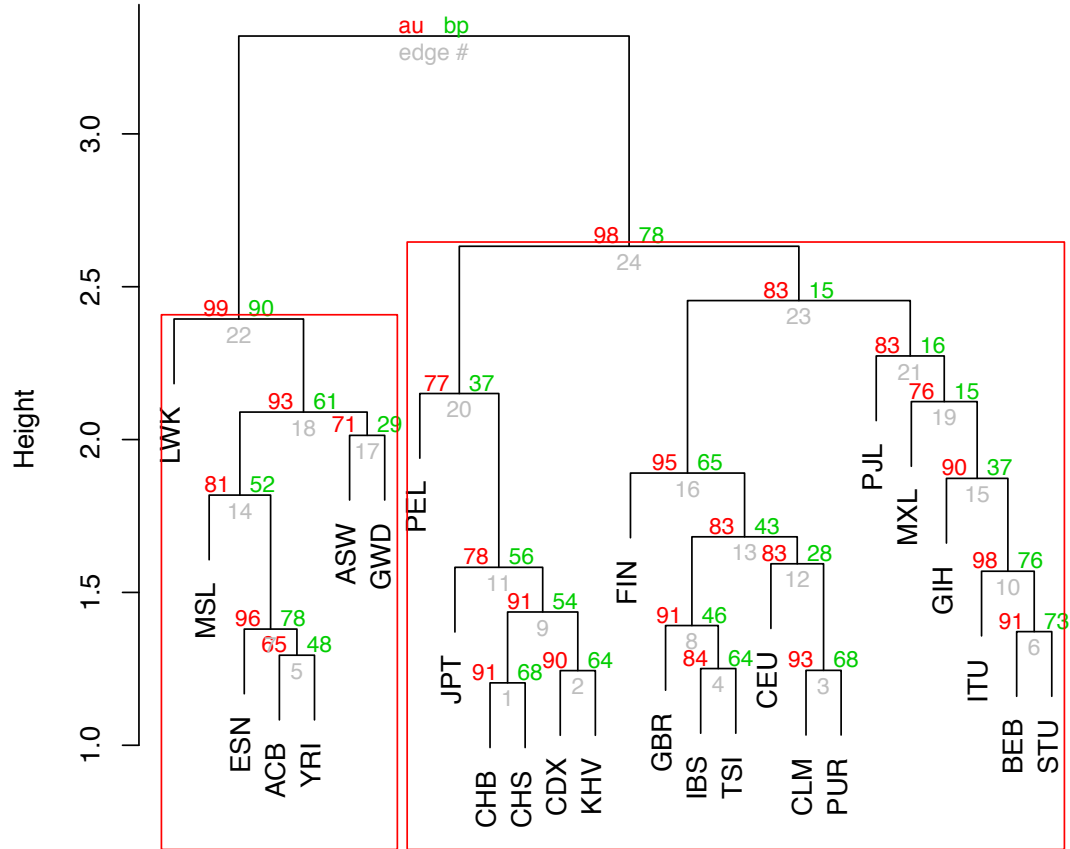using information of paired-end reads

Figure 1 – Simplified flowchart of the retroduplication caller pipeline.

A)

Cluster dendrogram with AU/BP values (%)

Distance: manhattan
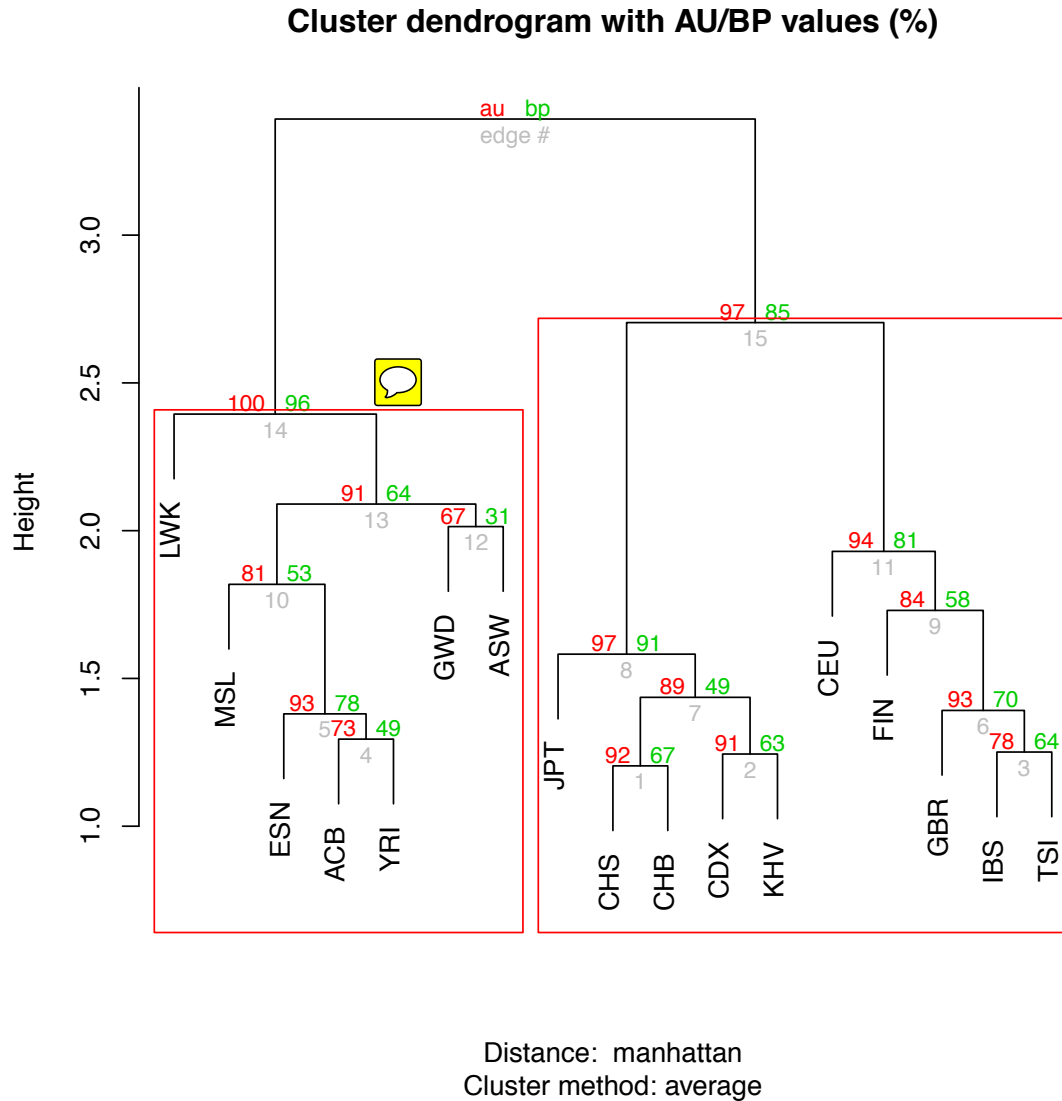Cluster method: average

B)

Figure **2** – Phylogenetic trees built based on novel retroduplications. A – The phylogenetic tree of all 26 populations enrolled in the 1000 Genome Project. B – The phylogenetic tree of 17 non-mixed population. Red number: AU (approximately unbiased) probability value. Green number: BP value, i.e. the frequency of the cluster appearing in bootstrap replicates. In the rectangulars, we highlight clusters with AU ≥ 0.95. Bootstrap resampling was performed 1000 times for generating the trees shown in A and B.