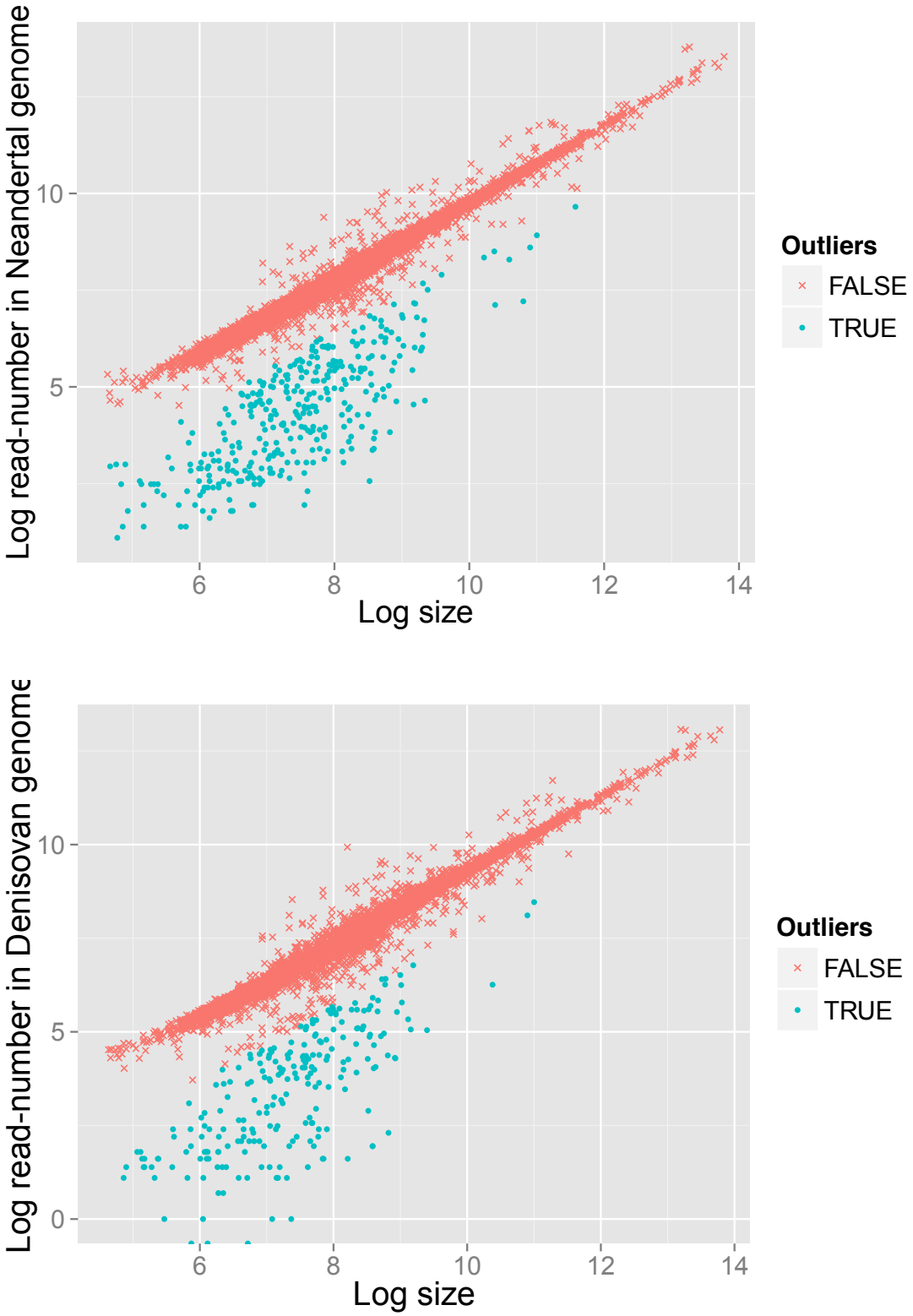

Figure S1. Discovery of putative deletions in Neandertal and Denisovan genome

(A) Correlation between number of Neandertal sequence reads mapping to a given interval and the size of this interval. The presence of low-read-depth outliers indicated in cyan dots reveals the putative Neandertal deletions. Both the number of reads and the size is shown in logarithmic scale. (B) The same graph for Denisovan reads.

Figure S2. GC content of shared and Human-specific deletions represented in box-plots and compared to each other. Significance is tested with a Student's t-test.

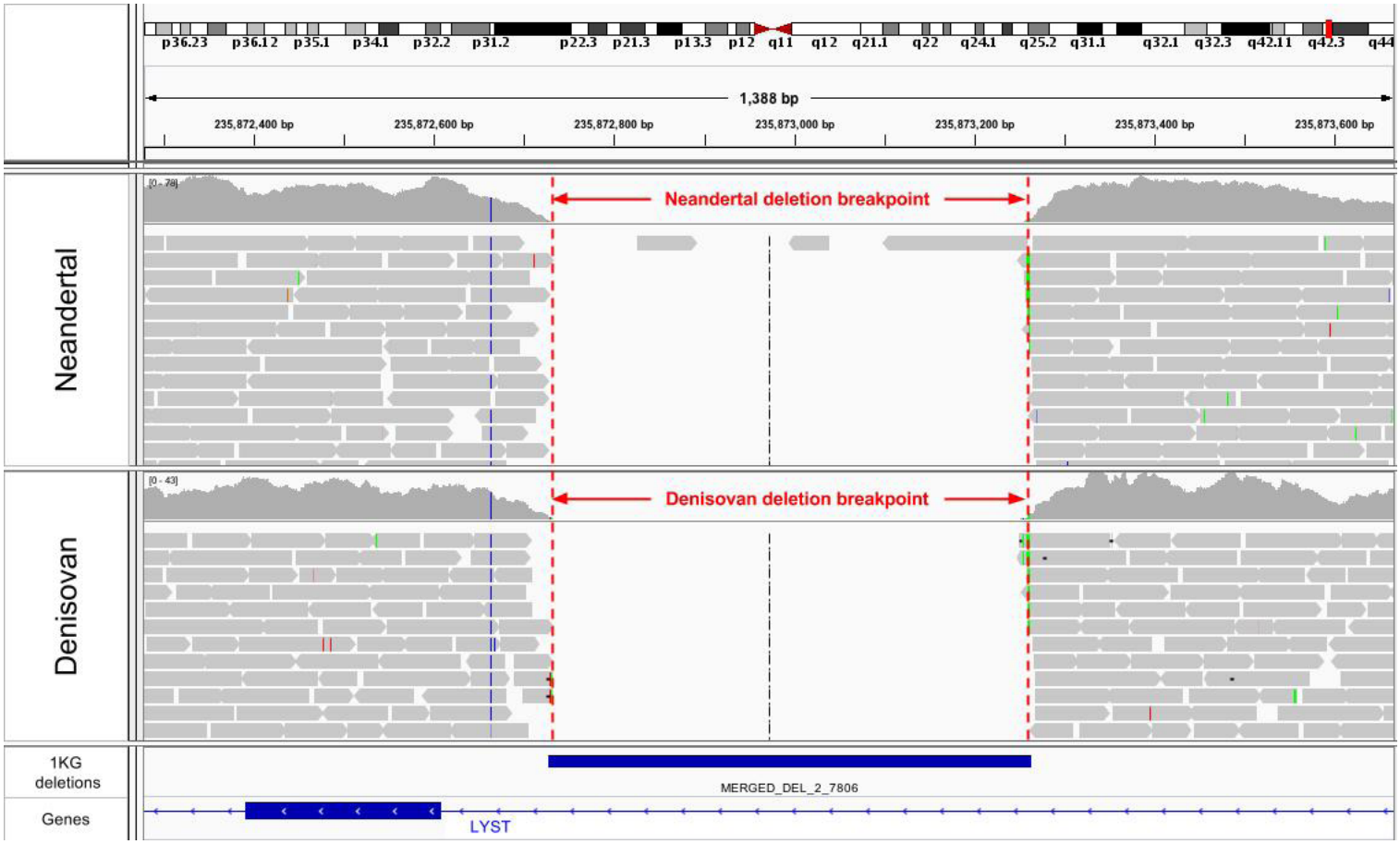

Figure S3A. Examples of manual inspection of identified deletions

IGV (Integrative Genomic Viewer) view of deletion variants. The panels from bottom to top designates: 1) The RefSeq Gene track. The concatenated blue arrows represent intronic regions and the thicker blue box represent exons. 2) 1KG deletions. The blue boxes show human variable deletions. 3) The Denisovan sequencing reads aligned to human reference genome and coverage. 4) The Neandertal sequencing reads aligned to human reference genome and coverage. 5) A ruler indicating the region focused on the chromosome. 6) The ideogram of the chromosome analysed. The red box indicates the portion of chromosome being displayed.
This deletion, which is determined as shared by Neandertal and Denisovan genomes, is located on chr1, overlapping the intron region of LYST (lysosomal trafficking regulator) gene. The edge of Neandertal and Denisovan “reads-free” zone is clear-cut, and concordant with the coordinate of this deletion. These observations suggest that this deletion is indeed shared by Neandertal and Denisovan genomes and likely ancestral, rather than recurrent.

Figure S3B. Example of external verification of Neandertal deletions using Neandertal Exome data

IGV (Integrative Genomic Viewer) view of deletion variants. The panels from bottom to top designates: 1) 1KG deletions. The blue box show position of a human variable deletion. 2) The RefSeq Gene track. The concatenated blue arrows represent intronic regions and the thicker blue box represent exons. 3) The Altai whole genome sequencing reads aligned to human reference genome and their read-depth across the genome. 4) The aligned reads and read-depth for all three exome sequences. 5) A ruler indicating the region focused on the chromosome. 6) The ideogram of the chromosome analysed. The red line indicates the portion of chromosome being displayed.

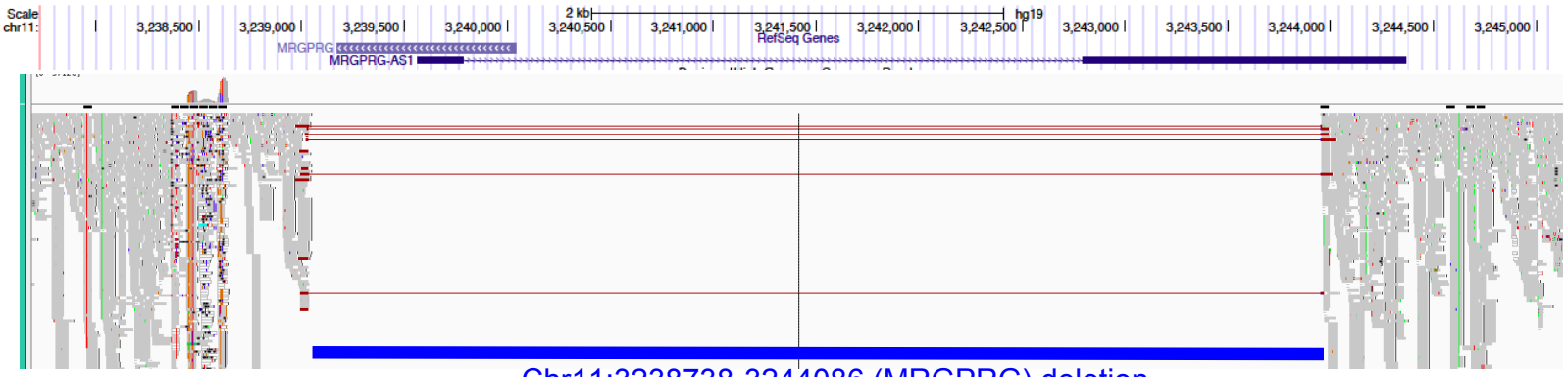

Figure S3C. Snapshot of an example for split-read support

Here is the example of a human deletion also deleted in archaic hominins. The top part shows the chromosomal location. Below is the RefSeq genes, with coding exons depicted with thick blue lines, non-coding but transcribed regions are thinner lines and intronic regions are thinnest blue lines. The gray sticks show the remapped reads from the Denisovan. The red lines spanning the deleted region shows connects two parts of a split-sequence. Please note that there are more split-read support than seen here as multiple lines are overlapping and not-visible. Below is the location of human deletion as reported by 1000 Genomes project as depicted by thick blue line.

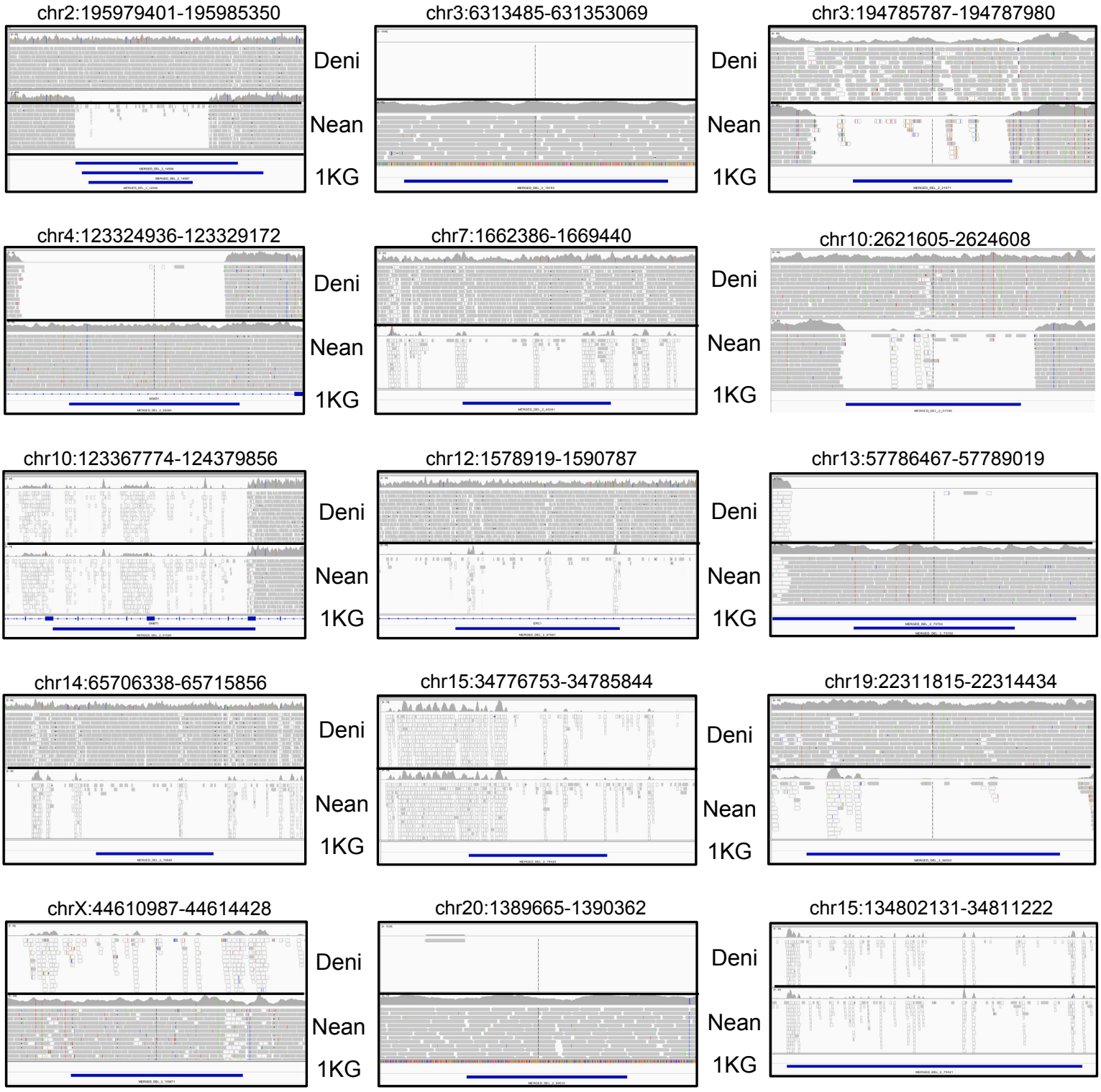

Figure S3D. Snapshots of recurrent deletions

IGV snapshots of all 15 recurrent deletions. The bottom section with the blue lines indicate the positions of human deletions. Above it are the Neandertal and Denisovan reads, respectively. The genomic coordinates of the analyzed region is shown above each graph. Please note the apparent disconcordance between the breakpoints of human and archaic hominin deletions in each case.

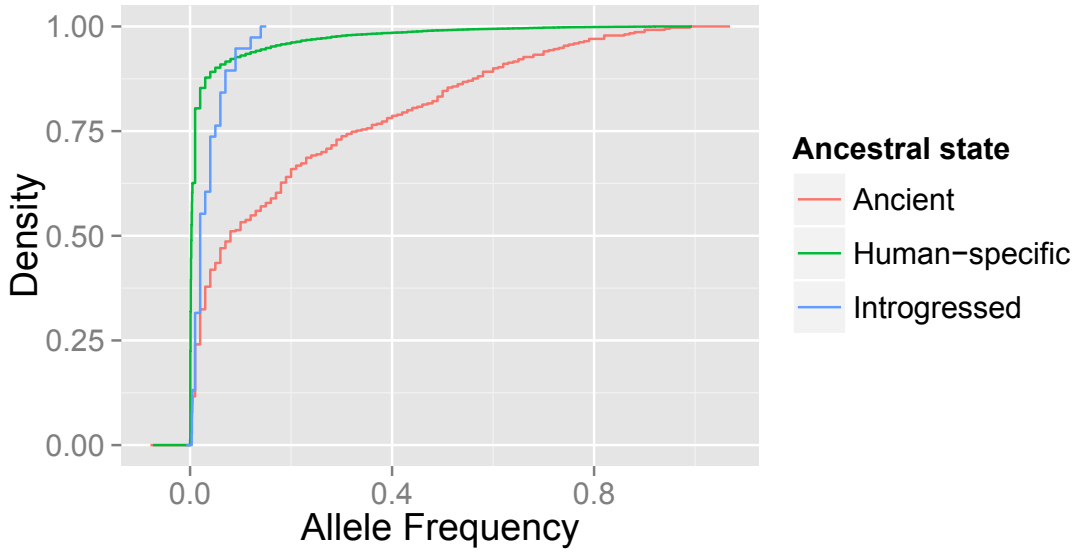

Figure S4A. The cumulative fraction plot for the allele frequency of ancient, introgressed and Human-specific deletions

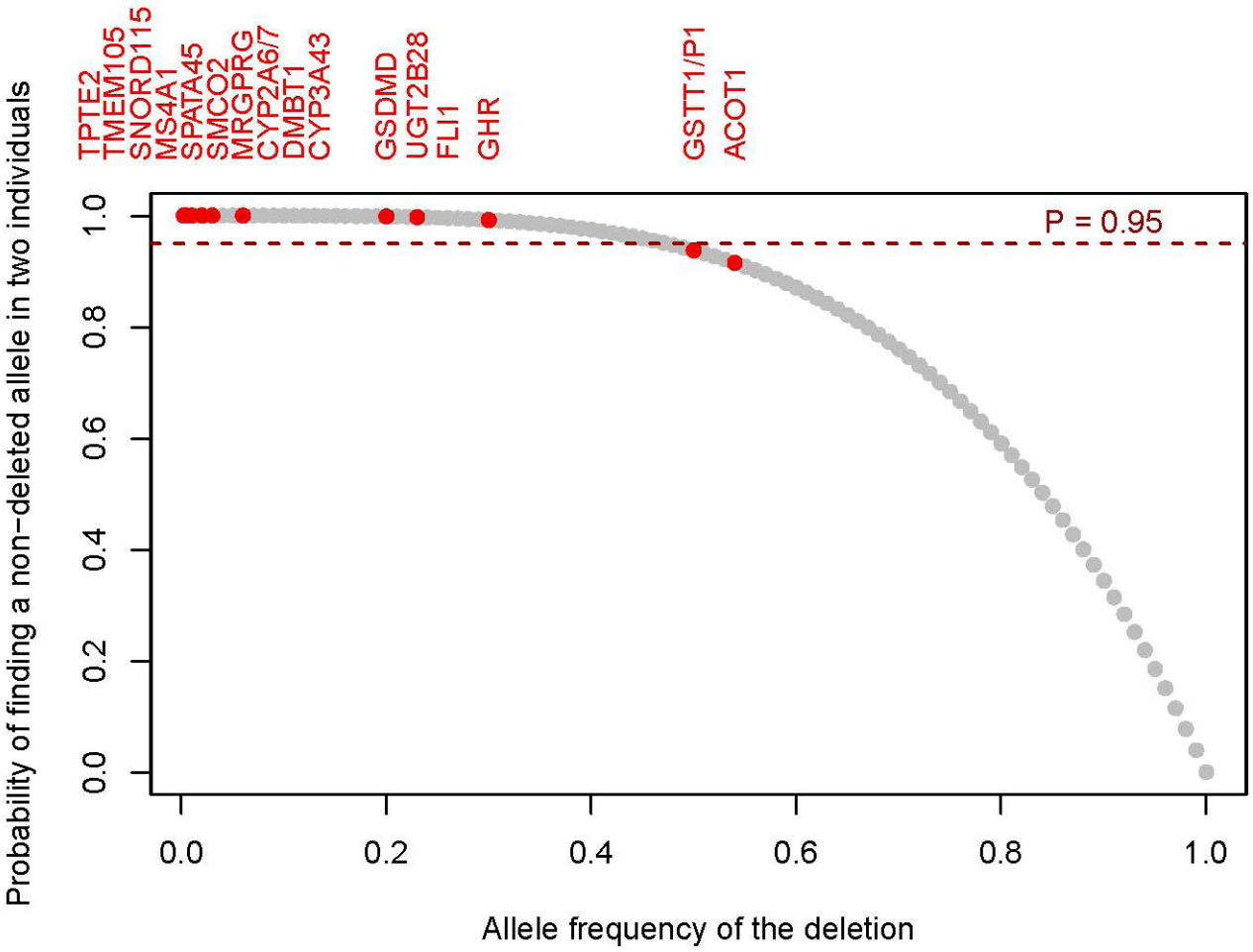

Figure S4B. The probability of finding a non-deleted allele based on the allele frequency of the deletion.

We found that all the 16 exonic human deletions that we find to be deleted in the Altai-Neandertal genome exonic deletions also homozygously deleted in the two additional Neandertal genomes based on their exome sequencing data (Figure S3B). Based on this observation, we calculated the probability distribution of detecting a non-deleted allele among four haplotypes. Y-axis indicates the probability of observing a non-deleted allele among four haplotypes. X-axis indicates the allele frequency of the deletion. The gray dots show the probability distribution. The red dots indicate the probability of finding a non-deleted allele among the four Neandertal haplotypes for specific exonic deletions, assuming the Neandertals to have the same allele frequency for these deletion variants as that observed among humans. According to this analysis, we suggest that the allele frequency of these deletion variants were extremely high or fixed in Neandertal population, otherwise we would have detected at least one non-deletion allele from four haplotypes.

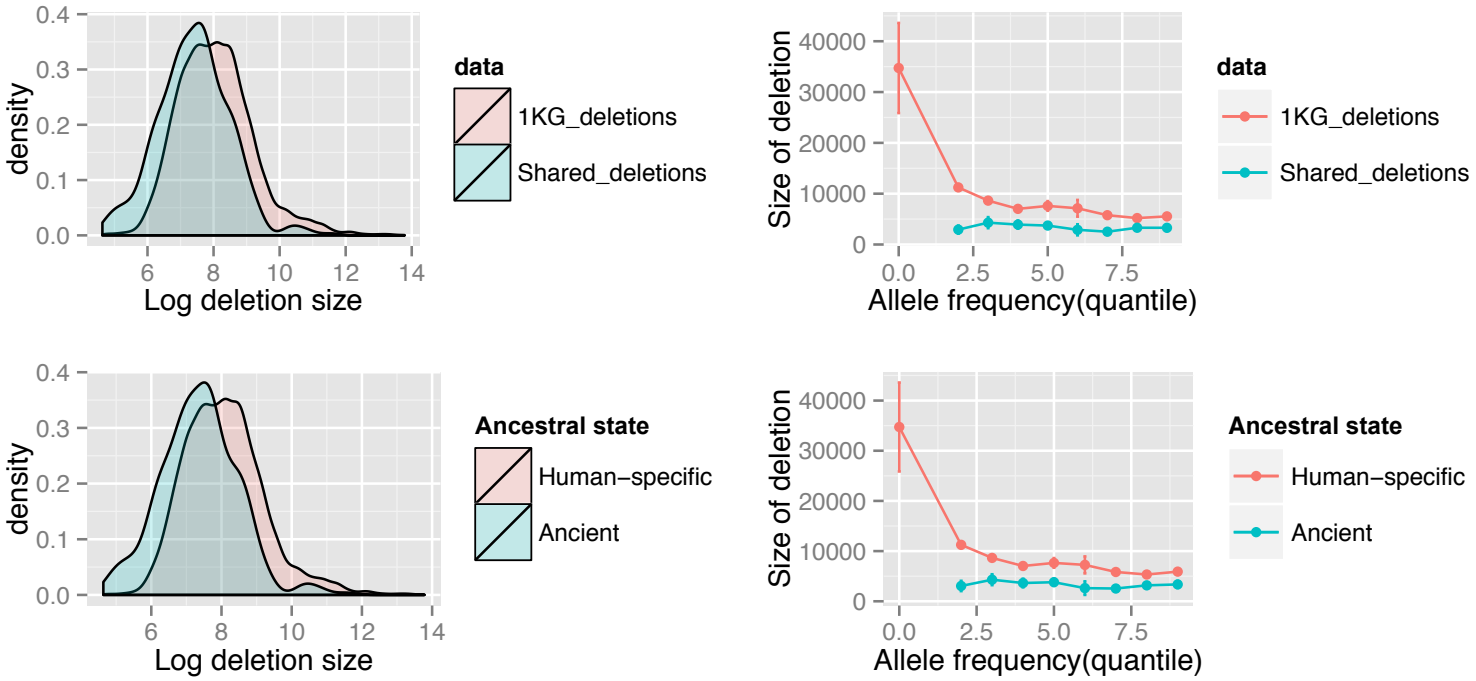

Figure S5. The size spectrum of human-specific, shared and ancient deletions

In the upper-left graph, we are showing a density plot for the deletion size distribution of All 1KG deletions, comparing to the size distribution of deletions shared with archaic hominins. In the lower-left, a similar graph is showing the density of size distribution between Human-specific and ancient deletions. The right panel graph shows the size distribution in allele frequency quantiles for the same category of deletions. Note that shared and deletions are significantly smaller than 1KG and Human-specific deletions, respectively, regardless of the allele frequency quantile.

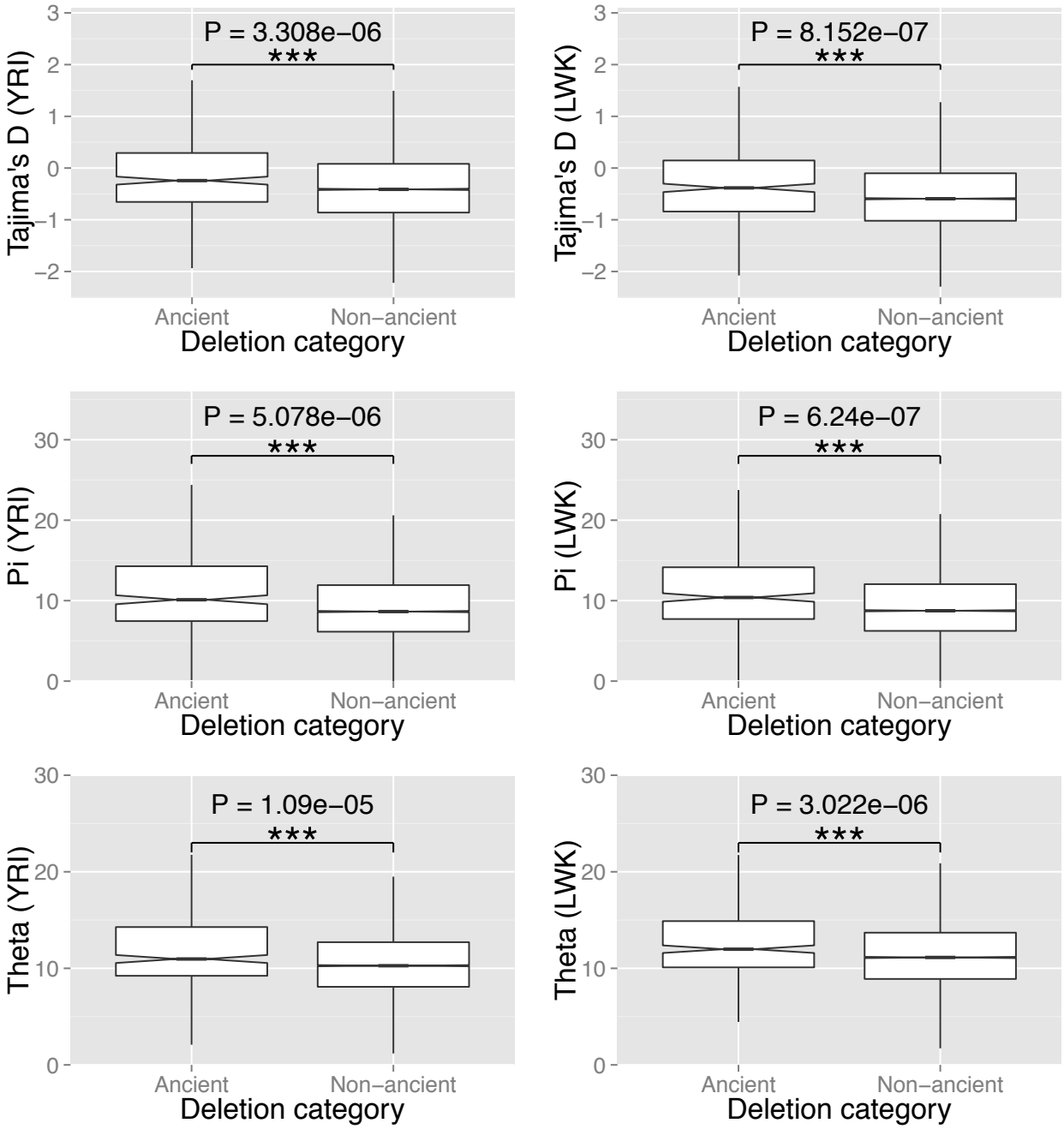

Figure S6. Comparison of population genetics statistics between regions harboring ancient and non-ancient deletions

This figure compares Tajima's D, Pi and Theta between ancient and non-ancient deletions in two African populations (YRI and LWK). The significance is tested by Student's t-test and indicated on the boxplots.

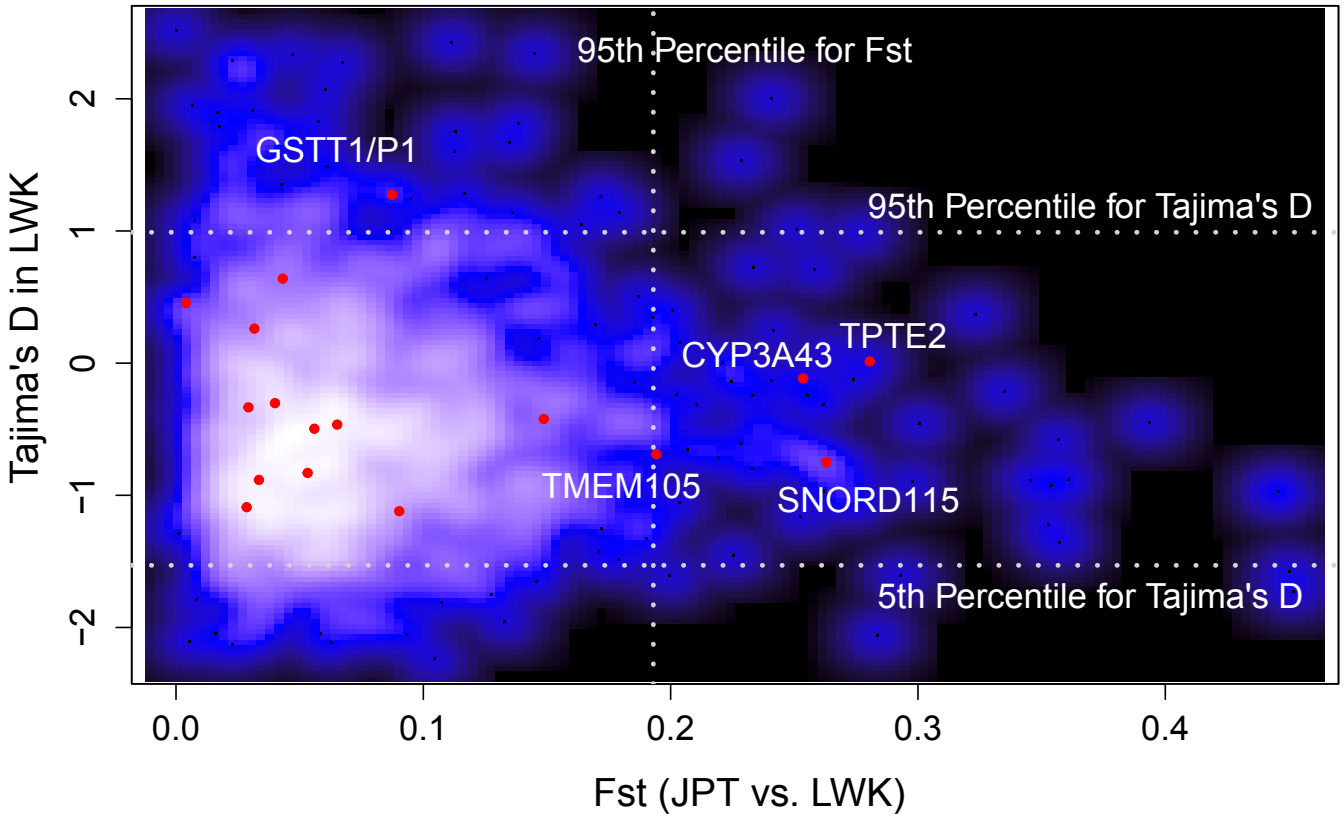

Figure S7. Comparison of Tajima's D and FST values

This figure shows the Tajima's D in LWK population on the y-axis and the FST between JPT and LWK populations on the x-axis. Only values for exonic deletions are depicted. The white cloud indicate thousands of data-points, purple cloud indicates hundreds of data-points. The individual blue clouds indicate less than 10 data points. The red dots are exonic deletions that are shared with archaic hominins. The dotted lines show the 95th and 5th percentile points for the x- and y-axes distributions.

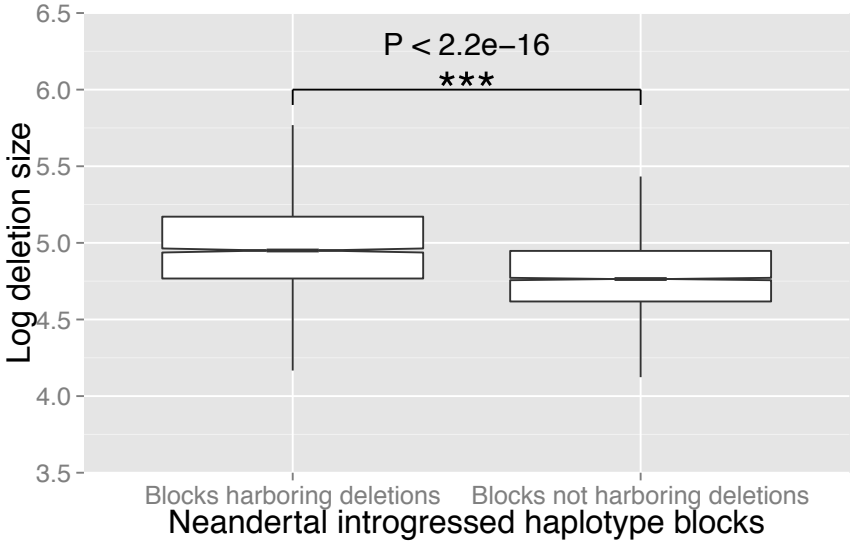

Figure S8. Sizes of introgressed regions where introgressed deletions were found

The boxplot compares the sizes of all Neandertal introgressed blocks to the Neandertal introgressed harboring introgressed deletions. The significance was assessed using a Student's t test.