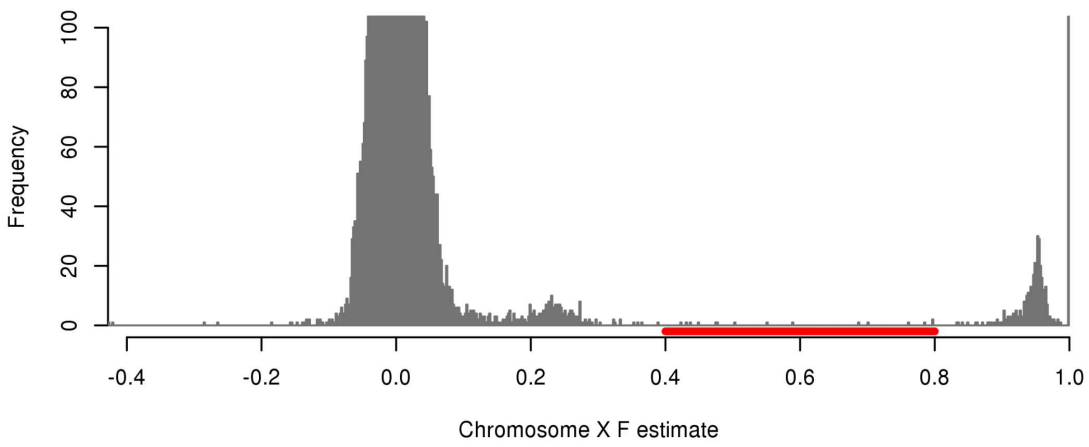
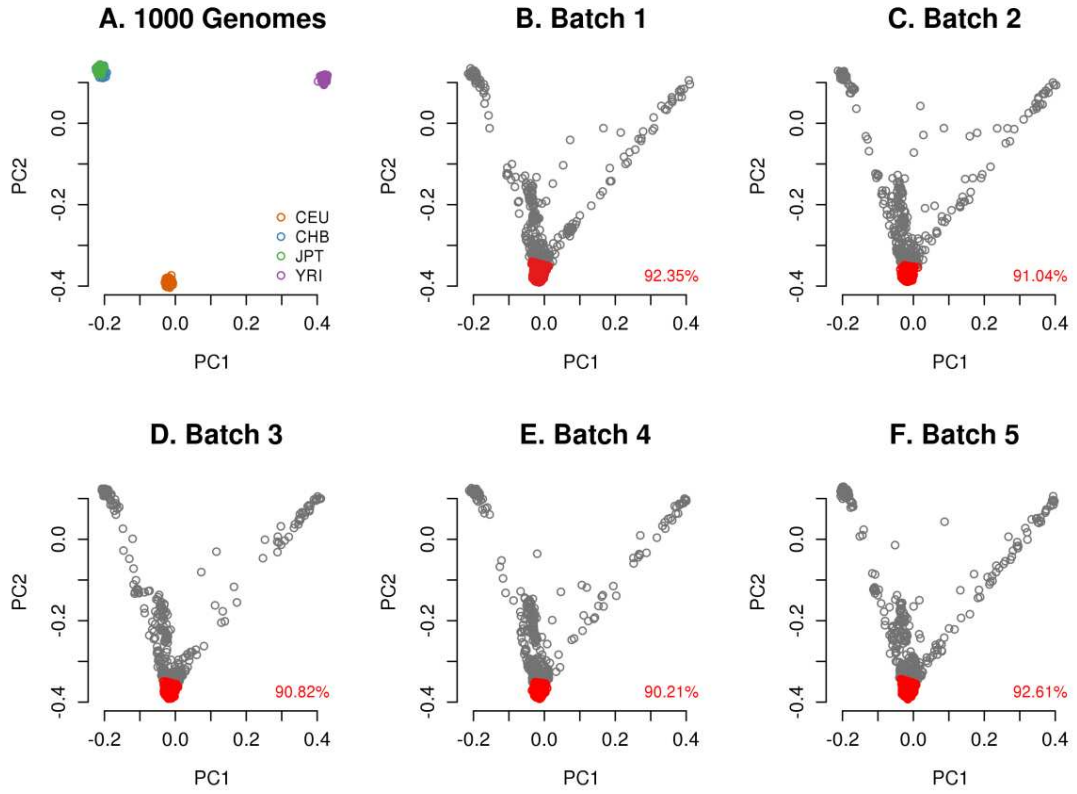
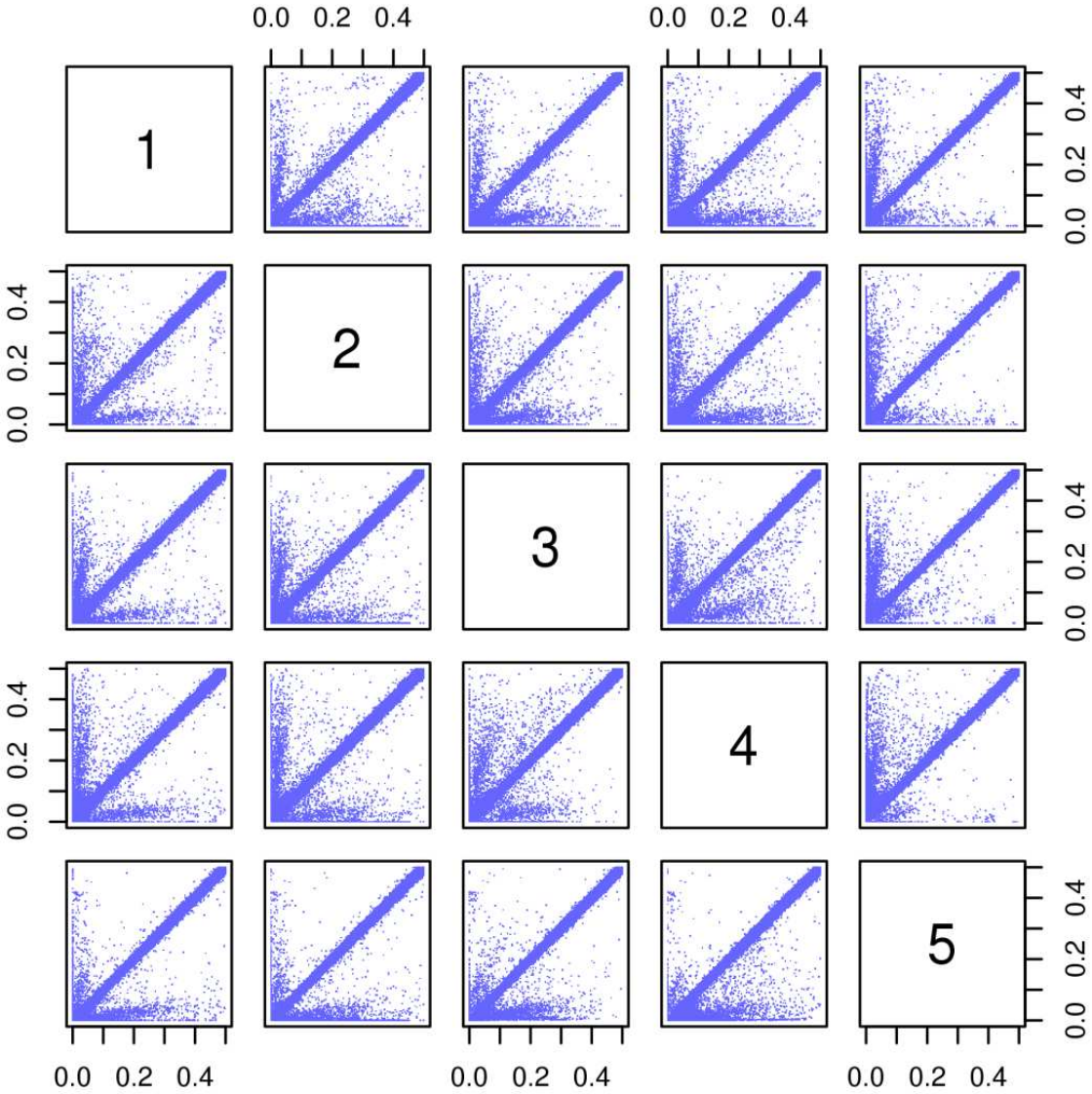


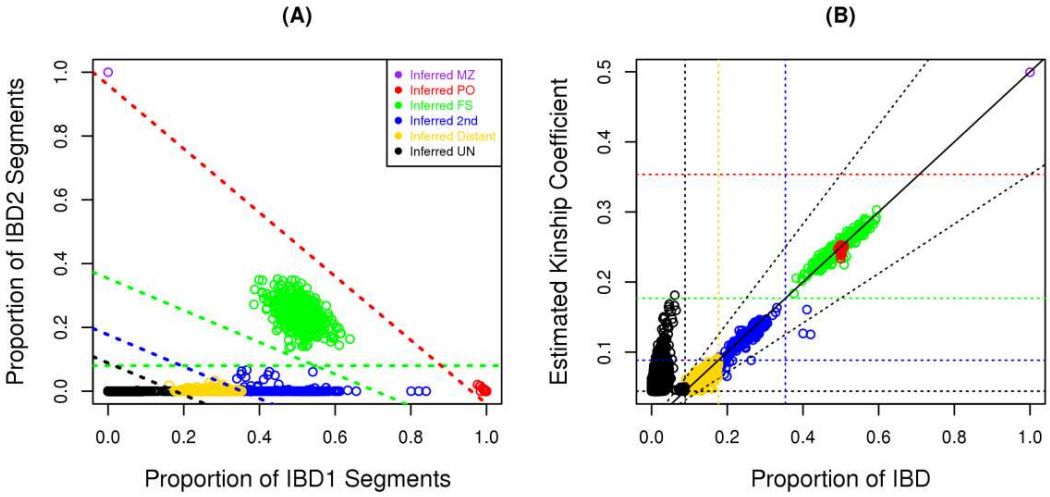
Supplementary Figure S1: Distribution of chromosome X F estimates for CLSA genotyped participants (y-axis truncated). Individuals with chromosome X F estimates within the range of 0.4 to 0.8 (red) are considered to have undefined chromosomal sex.



Supplementary Figure S2: Principal component (PC) plots. (A) Plot of first 2 PC for the analyzed populations from 1000 Genomes. (B-F) Projection of CLSA participants onto 1000 Genomes PC plot for genotype batch 1 to 5 followed by k-means clustering of PC1-4 (grey points). The largest cluster overlaps the 1000 Genomes CEU population (red points and percentage of total in batch is provided).

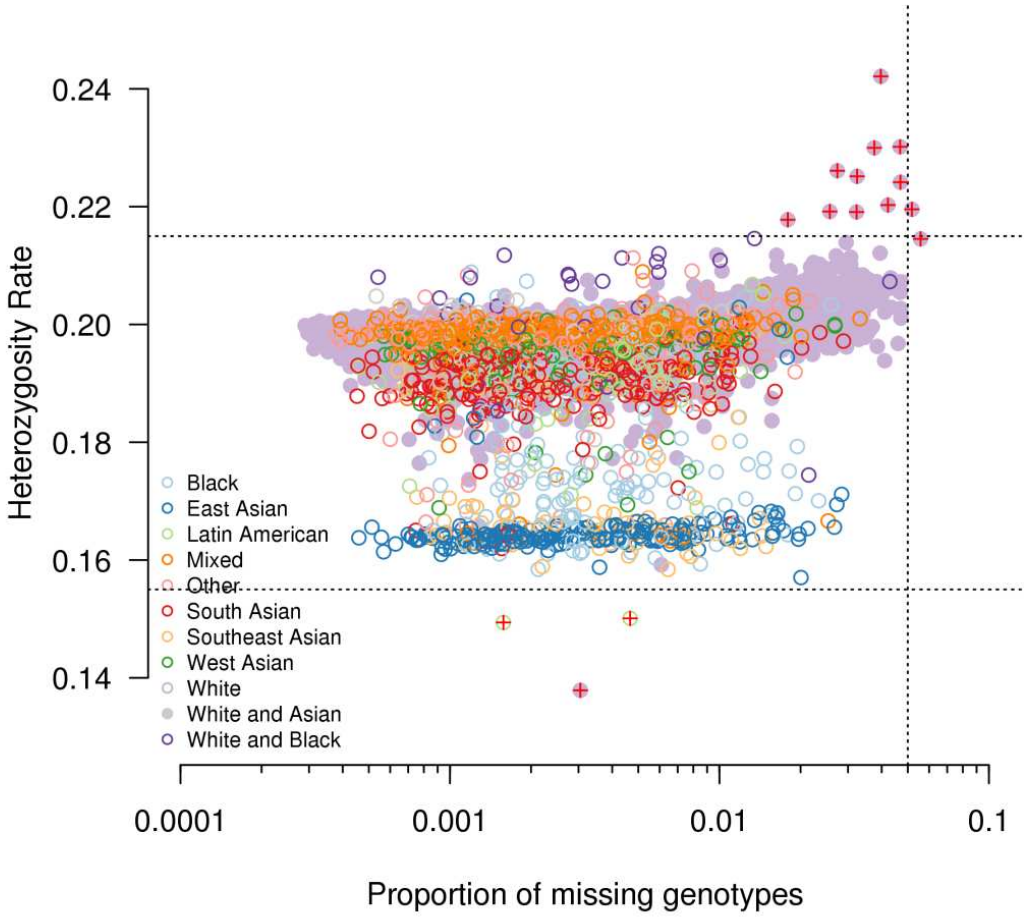


Supplementary Figure S3: Pairwise plot of allele frequency of SNPs from genotype batch 1 to 5.

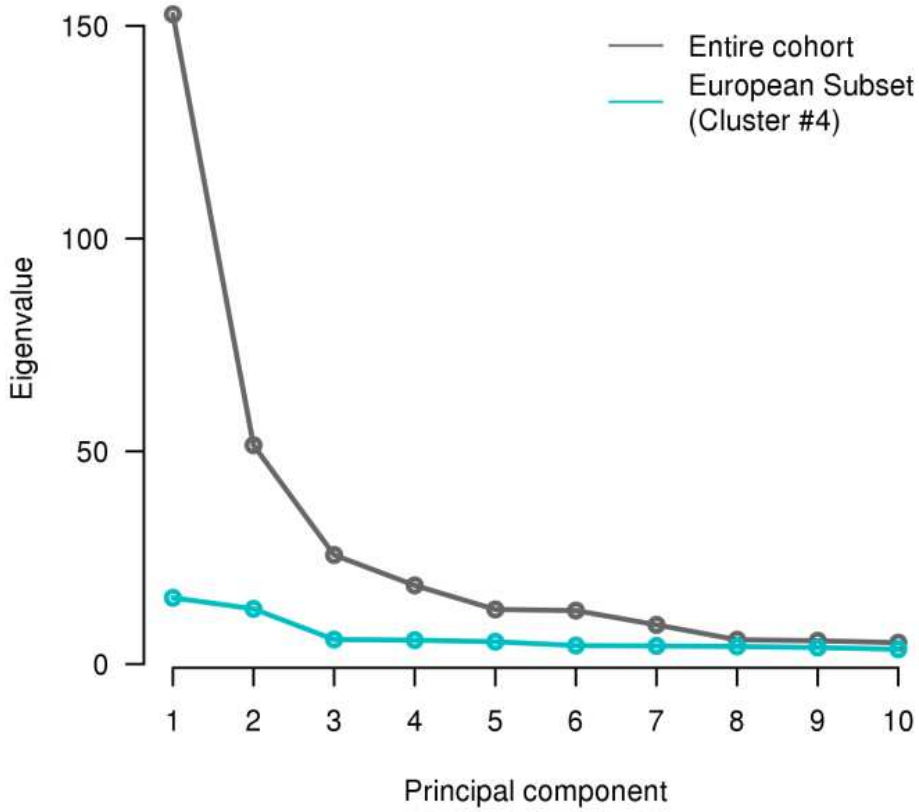


Supplementary Figure S4: Inference of familial relatedness using KING.

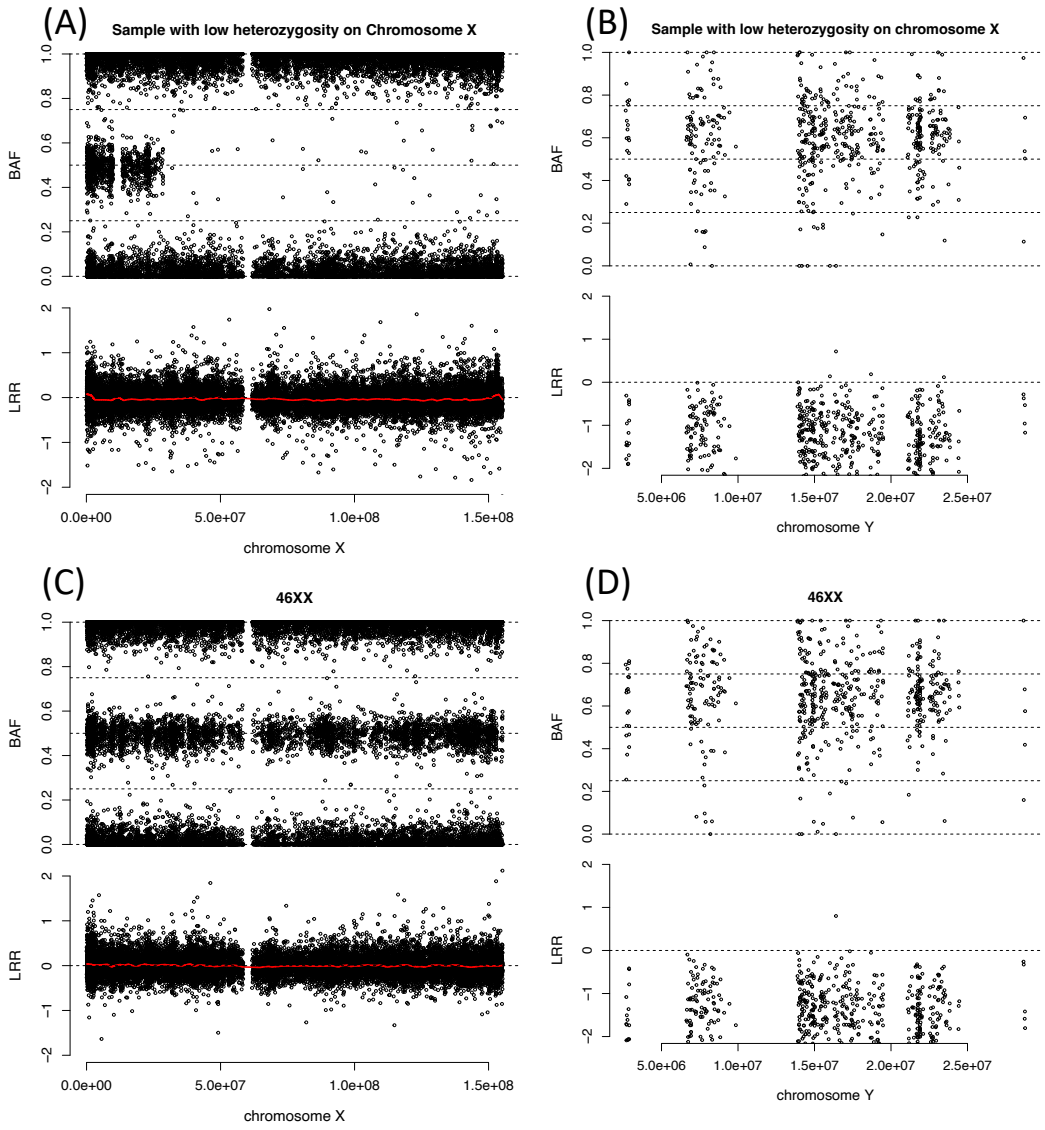
(A) Inference using IBD segments. (B) Inference using proportion IBD and kinship coefficient. Relationships in legend are abbreviated as: MZ=Monozygotic twin, PO=Parent/offspring, FS=Full sibling, 2nd=Second-degree relative, 3rd=Third-degree relative, Distant=Greater than 3rd degree relative, UN=Unrelated. Limits for inferring relationship type are indicated by dashed lines that are color-coded to match those listed in the legend.



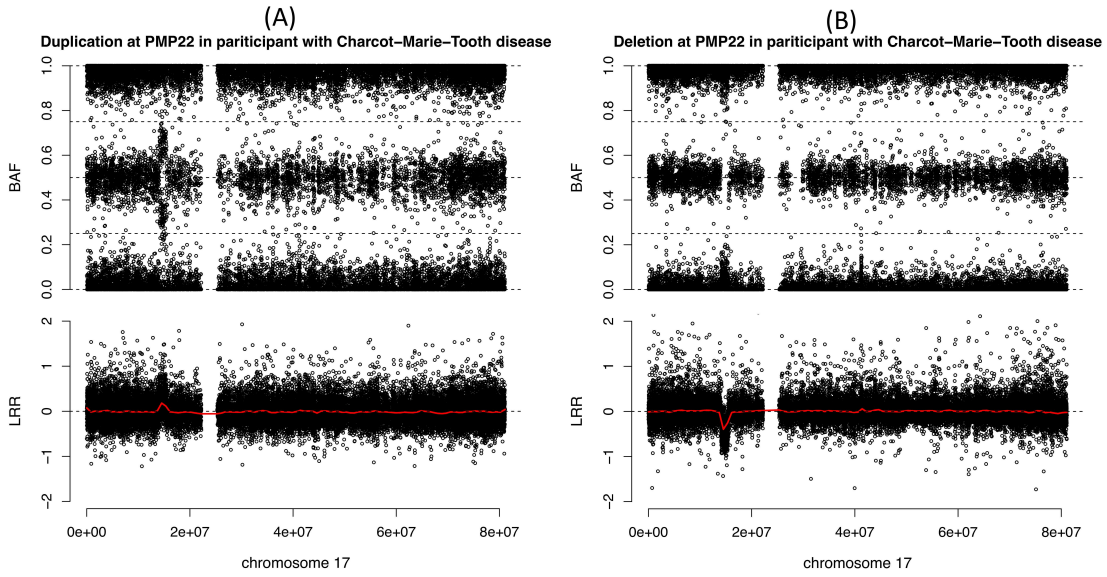
Supplementary Figure S5: Sample-wise heterozygosity versus genotype missingness. Points are color coded according to self-reported ancestry category. Outliers are marked with a red plus sign.



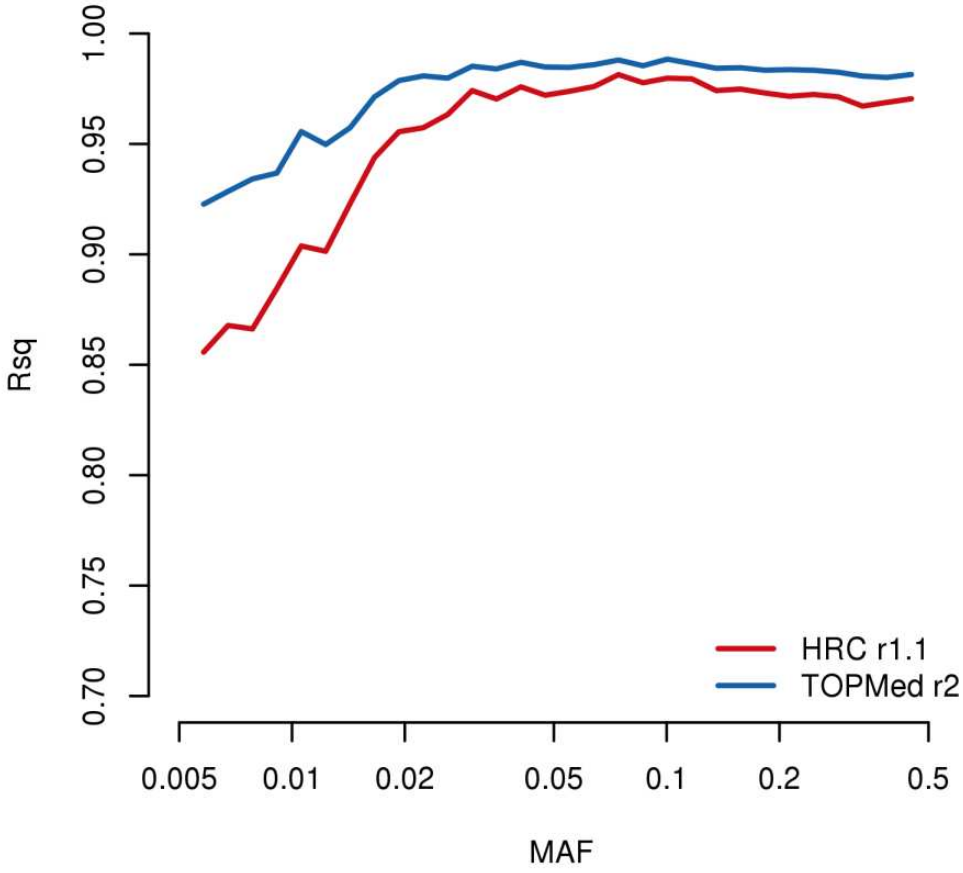
Supplementary Figure S6: Eigenvalues for PCA analysis of the entire cohort (grey) and the European ancestry subset (cluster 4, Robin egg blue), demonstrating a reduction in genetic variance within the European ancestry subset.



Supplementary Figure S7: BAF (TOP) and log2 ratio (BOTTOM) of chromosomes X (A) and Y (B) are shown for sample with low heterozygosity on chromosome X compared to sample with 46,XX (C-D).



Supplementary Figure S8: BAF (TOP) and log2 ratio (BOTTOM) of chromosome 17 are shown for sample with duplication (A) or deletion (B) at *PMP22* locus.



Supplementary Figure S9: Imputation quality of the CLSA cohort using the TOPMed versus Haplotype Reference Consortium (HRC) reference panel stratified by minor allele frequency (MAF) bins (data shown is from chromosome 22).