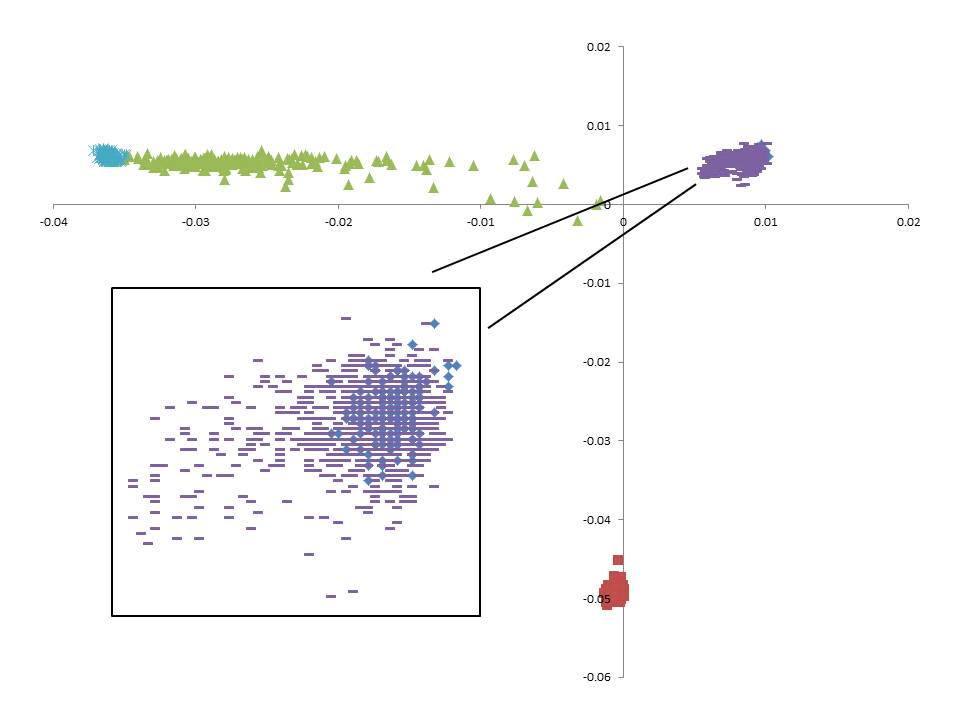
**Supplement Material**

Due to the availability of germline DNA, genotyping was performed at two independent time points in non-overlapping study subsets. DNA from an initial 2209 patients was genotyped by Illumina Genotyping Services using the HumanOmni1-Quad (>1 million SNPs) array. An additional 1222 patients were also genotyped by Illumina Genotyping Services using the Human OmniExpress (730,525 SNPs) array. Both sample sets used the Illumina BeadChip microarray platform for genotyping and the Illumina GenomeStudio software for initial genotyping calls. Of note, SNPs on the OmniExpress were a subset of those on the HumanOmni1-Quad. Those SNPs not on the HumanOmni1-Quad were imputed in the second sample. Briefly, the 1000 Genomes Phase I integrated variant set served as reference using the IMPUTE2 **(**<https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#references) software>**.** All SNPs were mapped to the human genome version GRCh37.3. Prior to imputation, SNPs with missing rate greater than 5%, Hardy Weinberg Equilibrium (HWE) p-values less than 0.0001, or minor allele frequency (MAF) less than 3% were excluded. A principal component analysis was performed using Eigenstrat[20](#_ENREF_20),[21](#_ENREF_21) and reference data from 11 HapMap phase III populations to identify clusters using the first two eigenvectors computed using all SNPs **(see figure below**).Samples clustering with the European American (EA) reference set were retained for further analysis.

Genotypes from both sets were imputed to the level of the 1000 Genome Project. Imputation was performed with the 1000 Genomes Phase I integrated variant set as reference haplotypes, using the IMPUTE2 software, and all SNPs were mapped to the human genome version GRCh37.3. SNPs having low imputation quality (information score <0.30) were removed from further analysis.810,907 SNPs were used in EA samples.