Supplemental Methods

ECOG-5103

DNA from an initial 2209 patients was genotyped by Illumina Genotyping Services using the HumanOmni1-Quad array (1.14 million SNPs). Subsequently, an additional 1222 patients were also genotyped by Illumina Genotyping Services using the Human OmniExpress array (730,525 SNPs). 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 obtained through imputation in the second subset. 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[18] and reference data from 11 HapMap phase III populations to identify clusters using the first two eigenvectors computed using all SNPs **(Supplementary Figure 1**). Samples clustering with the European American (EA) reference set were analyzed separately from those of African descent (AA).

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.810,891 and 872,021 SNPs were used in EA and AA samples respectively.

ECOG-1199

Top SNPs from Grade 3-4 GWAS in ECOG-5103 were evaluated in ECOG-1199. Based on the number of cases and controls in ECOG-1199, and the estimated effect size from ECOG-5103, 51 SNPs with p-value < $1x10^{-4}$ and linkage disequilibrium (LD) support in the EA cohort and from unique regions of the genome (**Supplementary Table 1)** were considered candidates. Only one SNP from each LD block was selected to optimize coverage as well as to minimize corrections for multiple comparisons. 51 SNPs were genotyped using a QuantStudio™ 12K Flex OpenArray® AccuFill™ System (Thermo Fisher Scientific) platform operated in a CLIA/CAP certified laboratory environment. SNPs were removed if the call rate was <95%, the MAF <3% or the HWE p-value $<10^{-4}$, resulting in a set of 30 top SNPs from the EA ECOG-5103 cohort taken forward for analysis in ECOG-1199 (**Supplementary Table 4**). Ancestry informative markers were not available, and thus analyses were performed separately for those of self-defined Caucasian race.