Supplementary Data

Supplementary Information (S1) Microarray Data Generation and Analysis**.** Raw data was imported into Partek Genomics Suite to generate relevant CN, LOH, and Gene Expression data. CN values were created by logging (base 2) the ~900K raw probe fluorescent values, comparing against a 270 HapMap Sample baseline CN dataset, and converting this information into the corresponding CN values (a value of 2 being a normal diploid CN value). Genetic CN events were reported as amplifications (CN values >2.3) and deletions (CN values <1.7) using Partek's Genomic Segmentation algorithm. The resulting CN event was required to have a minimum of 10 correlated neighboring probes with a p-value threshold of 0.001 for the significance of the difference from two neighboring probes.

SNP genotype data was created by analyzing the ~500K SNP probes through Affymetrix's Genotyping Console using the BRLMM genotyping algorithm to create heterozygous and homozygous calls for each SNP probe. The data file is imported into Partek and compared against a 270 HapMap genotype sample dataset for LOH analysis which generates a list of potential regions with a loss in heterozygosity using a threshold approaching "zero" value (approximately less than 0.07) based upon a heterozygous rate of the probes in the region. Normal regions have a heterozygous rate around 0.3.

To remove statistical variation gene expression values were created by normalizing the arrays using a Quantile Normalization in Partek. The raw probe fluorescent values were logged (Base 2) and contrasted by generating Log_2 ratios from the test sample compared to a normal group using an ANOVA (analysis of variance) 1-way contrast in Partek. This creates a differential expression table where the $\log_2$ ratios were considered significant if the change from normal was >2 or <-2, and reported in a fold change table. This fold change table was restricted to transcripts that had a p-value significance of less than 0.01. The resulting data sets from CN, LOH, and gene expression were integrated in Partek and analyzed to compare the effect of CN and LOH data on the overall gene expression. The data sets containing the respective gene identifiers and corresponding expression values were uploaded into IPA to identify functional network and pathway relevance. For functional analysis, the most significant differentially expressed genes from the dataset that met the fold change cutoff of +/-2, p-value cutoff of 0.01, and were associated with biological functions and/or diseases in Ingenuity's Pathways Knowledge Base were considered. Fischer's exact test was used to calculate a p-value determining the probability that each biological function and disease assigned to the data set is due to the effect of the condition, as opposed to the biological function designation occurring by random chance. Similarly, pathway analysis identified the pathways from the Ingenuity Pathways Analysis Library of Canonical Pathways, derived from Ingenuity Pathways Knowledge Base, that were most significant to the data set. Specific networks of these focus genes were then algorithmically generated based on their connectivity.