Genome-Wide Human SNP Array 6.0 (Affymetrix)
Acute Myeloid Leukemia (AML) – Copy Number

*Protocols performed at the Fred Hutchinson Cancer Research Center.

All genotyping was performed according to manufacturer's protocol. Briefly, two identical aliquots containing 250ng of DNA were digested with specific restrictions enzymes in separate reactions; one reaction contained Nsp1 and the other Sty1. Immediately following digestion, each sample was ligated with adaptors containing a complementary sequence to the overhang generated at digestion. Following ligation, each sample was subjected to PCR amplification using standard reagents. Following PCR, each sample was assayed on a 2% agarose gel to ensure that a DNA smear of appropriate size was produced. The Nsp and Sty amplifications were combined, purified and quantitated. All samples with at least 180 micrograms total DNA were allowed to continue to fragmentation using the enzymatic reaction Affymetrix Fragmentation reagent. The fragmented DNA was assayed on a 4% agarose gel to ensure that the size of the DNA collapsed to less than 75nt. Following fragmentation, the DNA was end-labeled with terminal deoxy transferase and Affymetrix DNA labeling reagent.

Nucleic acid hybridization was performed according to the manufacturer's protocol for the AffyMetrix 6.0 SNP array. Each sample was then resuspended in hybridization buffer and hybridized to the Affymetrix 6.0 array for 16 hours. Following hybridization, the arrays were washed on the Affymetrix Fluidics station and scanned on the GeneChip scanner.

The array scanning protocol was performed according to the manufacturer's protocol for the AffyMetrix 6.0 SNP array.

All data were processed using the standard analysis suite provided by Affymetrix. The QC call rate is developed for each sample using a subset of SNPs and the DM algorithm. A QC call rate of greater than 87% is a passing score for Affymetrix, the average call rate for this dataset was 99.4%. Samples passing the QC call rate are then clustered using the Birdseed algorithm. Individual data files (CEL files) were uploaded to Partek Genomics Suite (St. Louis, MO). Using a paired analysis (each patient's remission samples was used as the reference), copy number was calculated for each probeset and is indicated in the level 2 file, TARGET_AML_level2_paired_CN_log2_format.txt.

To find areas of the genome amplified or deleted, the Partek segmentation algorithm was applied to the level 2 dataset.

The unfiltered copy number segmentation file, TARGET_AML_CN_level3_unfiltered_Diagnostic.txt, contains all segments for each patient, both changed and unchanged, with no filtering parameters applied. To reduce the number of false positives and filter out segments of the genomes that are unchanged, TARGET_AML_CN_level3_filtered_Diagnostic.txt, contains only segments with <1.7 or >2.3 copy

number, have >99 markers and a p-value <0.05. In addition, segments from the Y chromosome and mitochondrial genome were removed.