Genome-Wide Human SNP Array 6.0 (Affymetrix)

Acute Lymphoblastic Leukemia Phases I & II (ALL P1, ALL P2) – Copy Number

*Protocol performed at St. Jude Children's Research Hospital.

DNA was extracted using QIAGEN QIAamp DNA Mini Kit according to manufacturer's protocol.

Nucleic acid labeling, hybridization and array scanning protocols were used according to Affymetrix manufacturer's protocol for Affymetrix Mapping 250k or Affymetrix Genomewide SNP6 arrays at St. Jude's Children's Research Hospital.

Normalization data transformation protocols were carried out at St. Jude's Children's Research Hospital as follows: 250K genotypes were generated using the BRLMM (Bayesian Robust Linear Model with Mahalanobis) algorithm implemented in GTYPE (Genotyping Analysis Software, Affymetrix). SNP6 genotypes were generated using the birdseed v2 algorithm in Genotyping Console (Affymetrix). Samples that failed standard quality control metrics (contrast quality control) were excluded. To generate copy number data, data were analyzed using a extensively used and validated algorithm developed at St Jude Children's Research Hospital. Affymetrix SNP array CEL files (Level 1 data) and SNP call files (either .CHP or .TXT files; Level 2 data) were imported into dChip and probe level values summarized[1]. Data were exported and normalized using the reference normalization algorithm[2]. This algorithm uses user supplied or computationally detected diploid chromosomes to guide normalization of the entire array on a sample-by-sample basis, and optimizes normalization of complex cancer samples while eliminating batch effects. This procedure generates the .cnmz file that includes both the summarize, normalized prove intensities and the genotype data.

Optimally normalized data were subjected to paired circular binary segmentation[3] with thresholds set to detect copy number segments of >2.3 or <0.7 copies and at least 5 markers (250K data) or 8 markers (SNP6 data). Raw copy number segmentation results inspected and curated in dChip.

References

1. Lin M, *et al.* (2004) dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-40 (PMID: 14871870)
2. Pounds S, *et al.* (2009). Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics* **25**, 315-21 (PMID: 19052058)
3. Venkatraman ES, *et al*. (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-63 (PMID: 17234643)