Gene Chip® Human Genome U133 Plus 2.0 Array (Affymetrix)

Acute Lymphoblastic Leukemia Phase I (ALL P1) - Gene Expression

*This protocol was performed at University of New Mexico.

1-3 µg of total RNA was labeled and hybridized to Affymetrix U133_Plus_2 arrays according to the manufacturer's recommendations (Affymetrix). A mask to remove uninformative probe pairs and Affymetrix controls was applied to all the arrays (resulting in the removal of 171 probe sets) and the default Affymetrix MAS 5.0 normalization was used on the remaining 54,504 probe sets. Array experimental quality was assessed using the following parameters, and all arrays met these criteria for inclusion:

- GAPDH more than 5000
- more than 20% expressed genes
- GAPDH 3./5. ratios less than 4
- linear regression R2 values of spiked poly(A) controls more than 0.90.

This gene expression dataset may be accessed via the NCI caArray site or at Gene Expression Omnibus under accession number GSE11877.

Microarray gene expression profiling data were available from an initial 54,504 probe sets after masking and filtering of minimal probe sets and controls (Supplemental data). Three different unsupervised, unbiased methods were used to select genes for standard hierarchical clustering: High Coefficient of Variation (HC) as originally described by Eisen et al.[1], Cancer Outlier Profile Analysis (COPA), and Recognition of Outliers by Sampling Ends (ROSE), a novel method similar to COPA developed in the Richard Harvey laboratory at the University of New Mexico[2]. In HC, the 54,504 probe sets were ordered by their coefficients of variation and the highest 254 probe sets were used for clustering; this method identifies probe sets having an overall high variance relative to mean intensities. COPA selects outlier probe sets, also in an unsupervised fashion, on the basis of their absolute deviation from median at a fixed point (typically the 95th percentile). ROSE was developed as an alternative to COPA, and selects probe sets both on the basis of the size of the outlier group they identify as well as the magnitude of the deviation from expected intensity (ROSE and COPA)[2]. For all 3 probe selection methods, the top 254 probe sets (Harvey et al.; supplemental Table 7A[2]) were clustered using EPCLUST(link is external) (Version 0.9.23 beta, Euclidean distance, average linkage UPGMA). A threshold branch distance was applied, and the largest distinct branches above this threshold containing more than 8 patients were retained and labeled. The HC method was used as the basis of cluster definition and nomenclature, with each of the 8 predominant clusters first identified through HC being assigned a number (H1-H8). All clusters are prefixed by the method of their probe set selection (H indicates HC; C, COPA; and R, ROSE), with COPA and ROSE numbers being assigned based on the similarity of a specific cluster group's membership (patient membership) to that seen in the original H clusters. The top 100 median rank order probe sets for each ROSE cluster are provided in Supplemental data. In the validation cohort (COG CCG 1961), the same initial masking criteria were applied to the raw data, yielding 54 504 probe sets for analysis. Applying ROSE with the same parameters used for the COG P9906 ALL cohort[2], 167 probe sets were

identified for clustering. The selection criteria used for COG P9906 was also used for COPA and HC, and the top 167 probe sets derived from these methods were used for hierarchical clustering (Harvey et al.; supplemental Table 7A[2]).

RNA Sample Preparation Methodology

Gene Expression Profiling Method[2] RNA was isolated from pretreatment diagnostic ALL samples in the 207 patients (131 bone marrow, 76 peripheral blood) using TRIzol (Invitrogen); all samples had more than 80% leukemic blasts.

References:

1. Eisen MB, Spellman PT, Brown PO, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. **95** (25):14863-14868 (PMID: 9843981)
2. Harvey RC, Mullighan CG, Wang X, Dobbin KK, Davidson GS, Bedrick EJ, Chen IM, Atlas SR, Kang H, Ar K, Wilson CS, Wharton W, Murphy M, Devidas M, Carroll AJ, Borowitz MJ, Bowman WP, Downing JR, Relling M, Yang J, Bhojwani D, Carroll WL, Camitta B, Reaman GH, Smith M, Hunger SP, Willman CL. (2010). Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood.* **116** (23), 4874-84 (PMID: 20699438)