Gene Chip® Human Genome U133 Plus 2.0 Array (Affymetrix)
Wilm's Tumor (WT) – Gene Expression

*Protocol performed at Ann and Robert H. Lurie Children's Hospital.

Gene expression analysis was performed with the Affymetrix U133+2 chip (Affymetrix, Santa Clara, CA, USA), according to the manufacturer's protocol using the Gene-Chip Operating Software and normalized using robust multichip average normalization. Unsupervised analysis was performed using Non-negative Matrix Factorization Consensus Version 5[1]. GSEA Version 2.0.14(link is external)[2] was run using 1,000 permutations and phenotype permutation. Significant enrichment was defined as those lists with >50 genes, an FDR < 10%, and a p-value < 5%.

Specifically:
RNA quality was assessed by a bioanalyzer and RNA samples were required to have a RIN > 7. Total RNA was provided to Lurie Children's Hospital Research Center at a concentration of 150 ng/ul (2 ug total) in sets of 16 samples. One WT sample for which sufficient column-purified RNA was available was selected to serve as a control sample (PAJMLZ). Each set of 16 samples received from the BPC included the WT control sample, which was therefore repeated throughout all steps of this procedure in order to ensure consistency among all steps.

250 ng of total RNA was labeled by using the Affymetrix GeneChip 3' IVT Express Kit at Lurie Children's Hospital Research Center. All procedures, including 1st strand reverse transcription, 2nd strand synthesis, in vitro transcription of aRNA, aRNA purification, quantitation, and fragmentation were performed according to the manufacturer's protocol.

Nucleic acid hybridization to the array was performed at Lurie Children's Hospital Research Center by using the AffyMetrix GeneChip Hybridization, Wash and Stain Kit per the manufacturer's instructions.

The arrays were scanned at Lurie Children's Hospital Research Center by using the Gene-Chip Operating Software (GCOS). Each .dat file was visually inspected for large scratches and/or misalignment of the grid. Gene-Chip Operating Software (GCOS) was used to generate .chp files (Level 2 data), which represent the consolidation of all individual probes within a probeset, from .cel files (Level 1 data). From .chp files, GCOS was used to generate .rpt files (Level 3 files), which show probe intensity values and QC values. All samples were inspected for several parameters. Background < 45 (actual range: 28.19–43.18). Noise (Raw Q) < 1.35 (actual range: 0.670–1.30). Scaling Factor < 65% (actual range: 11.487–52.965). % Present call > 35% (actual range: 38.4–57.7). 3'/5' GAPDH < 3.92 (actual range: 0.95–3.48). Samples with parameters outside of these limits were rerun starting at the step of RNA labeling. All .cel files (Level 1 data) were imported into the Broad Institute's GenePattern server and Robust Multichip Average (RMA) normalization was performed using the ExpressionFileCreator module. Data were exported as a single .txt file (Level 2 data) containing probeset information for each individual tumor within a single spreadsheet. Several analytic quality control steps were performed.

Principle component analysis (PCA) was performed to ensure that none of the samples were outliers. Pair-wise correlation coefficient analysis was performed using the data from the WT control sample that was included in each individual batch of samples. The normalized averages of the expression levels from each WT control run showed a correlation coefficient > 98%, indicating a high level of consistency. Six probesets corresponding to five genes were identified that closely correlated with gender (four male genes [RPS4Y1, DDX3Y, SMCY, and EIF1AY] and one female gene [XIST]). All samples were classified as male or female according to the expression patterns of these genes and the results were checked against the known gender of the patient. No discrepancies were detected.

For analyses, 9/10 replicates for PAJMLZ were removed from the RMA gene expression file. A collapsed data file was created by using the Broad Insitute's GenePattern CollapseDataset module with the default parameters and the maximum probe collapse method.

SAM was used to compare gene expression in 51 tumors: favorable histology WT (FHWT) sequenced at CGI with the MLLT1 variant (5) vs the remainder of FHWT sequenced at CGI that do not have the MLLT1 variant (46). Gene expression data is not available for 1 FHWT with the MLLT1 variant. SAM was run using the Level 2 gene expression data. First, probesets that had absent "A" calls for 95% (48) or more samples were filtered out, resulting in the retention of 39913 probesets for analysis. The data were log transformed prior to running SAM. Two class unpaired analysis was run using 200 permutations; probesets with $q < 0.05$ were retained.

References:

1. Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl Acad. Sci. USA.* **101**, 4164–4169 (PMID: 15016911)
2. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. **102**, 15545-15550 (PMID: 16199517)