SUPPLEMENTARY APPENDIX

The side population enriches for leukemia-propagating cell activity and Wnt pathway expression in zebrafish acute lymphoblastic leukemia

Jeremy T. Baeten,[1] Michael R. Waarts,[1] Margaret M. Pruitt,[1] Wen-Ching Chan,[2] Jorge Andrade[2] and Jill L.O. de Jong[1]*

[1]University of Chicago, Biological Sciences Division, Department of Pediatrics, Chicago and [2]University of Chicago, Center for Research Informatics, Chicago, IL, USA

 doi:10.3324/haematol.2018.206417

Received: September 7, 2018.

Accepted: January 9, 2019.

Pre-published: January 10, 2019.

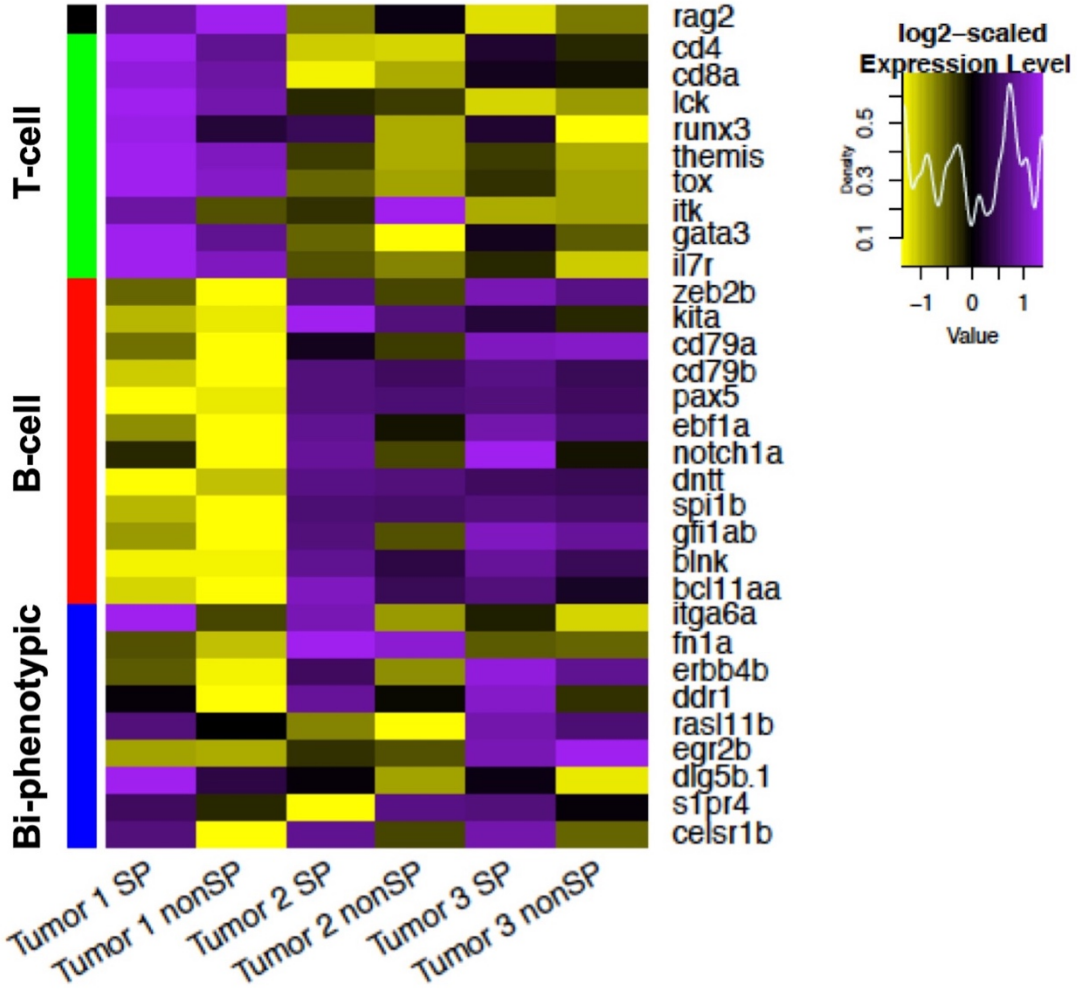
Correspondence: *JILL L.O. DE JONG* - jdejong@peds.bsd.uchicago.edu

Supplementary Methods

RNA-sequencing and analysis

RNA from sorted SP and non-SP tumor cells was isolated using the Trizol reagent (Ambion) and standard extraction[1]. One million cells were sorted from each population, and after initial extraction, the RNA was treated with DNAse I (Zymo) for 15 minutes and then re-extracted in Trizol. Concentrations were measured on a Bioanalyzer with a cutoff of a 9.0 RNA integrity number. Samples were sequenced using an Illumina Hi-Seq4000 instrument (single-read 50bp). The quality of raw sequencing data was assessed using FastQC v0.11.5[2]. Meanwhile, Illumina adapter/primer sequences were detected from sequencing reads. All RNA reads were first mapped to the zebrafish (GRCz10) reference genome using STAR v2.5.2b release with default parameters[3]. Picard v2.8.1 (http://broadinstitute.github.io/picard/) was used to collect mapping metrics. The resulting files from the previous alignment step were taken as input to quantify transcriptional expression using Cufflinks v2.2.1[4] and Rsubread::featureCounts v1.28.1[5]. Afterwards, several methods of differential expression analysis (DEA); including Cuffdiff v2.2.1[4], edgeR v3.20.7[6], DESeq2 v1.18.1[7], and limma v3.45.5[8], were employed to discover differentially expressed (DE) mRNA genes (fold change ≥ 4 or ≤ -4.) between the SP and non-SP groups, based on expression estimation of individual mRNA genes. To obtain the groups with similar expression trend based on the identified DE mRNA genes, several in-house scripts were implemented using R (https://www.r-project.org/) and Python (https://www.python.org/) languages. The identified DE genes were further used as input to functional analysis for the identification of enrichment of signaling pathways using PANTHER pathway analysis[9]. Gene set enrichment analysis (GSEA) was performed using the GSEA 3.0 software and Gene Ontology gene sets publically available from the Broad institute [10, 11]. In order to use these human gene sets, transcript lists were converted to their human orthologs using the Beagle Biological Annotation database curated and maintained by Andre Bernards [12].

Supplementary Figures



Supplementary Figure 1. Rag2-*Myc* transgene produces ALL tumors of T and B-cell lineage.

Heat map using RNA sequencing data of three tumors sorted for SP and non-SP with a selected gene set of putative T-ALL, B-ALL, and bi-phenotypic genes. Expression shown as $\log_2$-scaled.

Supplementary Data Tables

Supplementary Table 1. Limit dilution transplants. Limit dilution transplant data for Figure 1C.

Supplementary Table 2. Clonal evolution transplants. Limit dilution transplant data for Figure 2

Supplementary Table 3. SP sorted transplants. Limit dilution transplant data for Figure 3.

Supplementary Table 4. DEGs within tumors. 272 differentially expressed genes between SP and Non-SP in each individual tumor. Values given are fpkm.

Supplementary Table 5. DEGs across tumors. 761 differentially expressed genes between SP and Non-SP across all tumors. Values given are log2fc.

Supplementary Table 6. PANTHER analysis. Upregulated pathways in SP for both DEG sets. Wnt pathway genes in each set listed to the right of the table.

Supplementary Table 7. GSEA significant pathways. Gene Ontology: Biological Processes gene set enrichment analysis for SP sorted tumors. 464 significantly upregulated and 185 downregulated biological processes gene sets within sorted SP vs nonSP RNAseq datasets. Significance determined by NOM p-value < 0.05 and FDR q-value < 0.250.

Supplementary References

1. Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA Using TRIzol (TRI Reagent). Cold Spring Harbor Protocols. 2010;2010(6):pdb.prot5439.
2. Andrews S. FastQC. 0.11.5 ed: Babraham Bioinformatics, 2010.
3. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
4. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nature biotechnology. 2010;28(5):511-515.
5. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-930.
6. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-140.
7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.
8. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47-e47.
9. Mi H, Huang X, Muruganujan A, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Research. 2017;45(D1):D183-D189.
10. Mootha VK, Lindgren CM, Eriksson K-F, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics. 2003;34(267.
11. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-15550.
12. Bernards A. Ras Superfamily and Interacting Proteins Database. Methods in Enzymology: Academic Press, 2006:1-9.