Validation of Human Leukemia Mouse Model Using RNA-Sequencing Gene Expression Profiling

Noriko Satake[1], Clifford Tepper[2], Phuong Tran[3]

[1]Department of Pediatrics, UC Davis   [2]Genomics Shared Resource, UC Davis Cancer Center   [3]UC Davis School of Medicine

Background

Leukemia can be maintained and expanded as experimental model systems by serial transplantations in immunocomprised mice. Before being used for novel drug testing, these mouse models need to be validated by demonstrating that they carry the same leukemia as the original leukemia from patients.

There are several methods to validate leukemia mouse models such as T-cell Receptor (TCR) Spectratyping, flow cytometry, microarrays, etc. However, none of these methods can provide absolute levels of transcript expression or precise, high-resolution genomic changes like Gene expression analysis using next-generation sequencing (NGS)-based RNA-sequencing (RNA-Seq) can . Specifically, RNA-Seq provides the below advantages:

- A more precise measurement of transcripts and their isoforms than other methods
- Broader dynamic range for broader dynamic range on gene expression measurement.
- Easier detection of rare and low-abundance transcripts and novel transcripts

Our research has generated mouse models from patients with T-cell Acute Lymphoblastic Leukemia (ALL). In this study, we performed molecular characterization of the patient T-cell ALL and the engrafted mouse leukemia.

Objectives

1. Determine whether the pair of the original patient's T-cell ALL cells and its engrafted 3rd generation NOD/SCID/IL2R-null mouse leukemia cells are similar at the molecular level using correlation and scatter plot analysis based on genome-wide RNA-Seq gene expression analysis.
2. Determine if the xenograft cells still carry high-frequency gene expression signatures in the patient's T-cell ALL. These are genetic alterations utilized to define distinct molecular groups of T-ALL with specific gene expression signatures.
3. Determine if the xenograft still carries the patient's specific ALL surface marker genes.
4. Identify common biochemical processes from the patient tumor that remained intact or altered in the xenograft during the serial passage of transplantation.

Methods

- Total RNA samples from the patient and xenograft were submitted for NGS at the UC Davis Cancer Center's Genomics Shared Resource. RNA-Seq libraries were prepared and sequenced using Solexa sequencing with an Illumina HiSeq 2000 sequencing system. Each sample yielded approximately 40 million 50-base sequence reads. After base calling and quality scoring accomplished by HiSeq Control System with Real Time Analysis (HCS 1.5/RTA 1.13) and CASAVA 1.8 software (Illumina), the raw RNA-Seq reads data was generated as FASTQ files.
- A standard TopHat-Cufflinks workflow was used to map RNA-Seq reads data, perform transcript assembly, and estimate gene/transcript expression levels expressed as numeric FPKM (Fragments Per Kilobase of transcript per Million mapped reads).
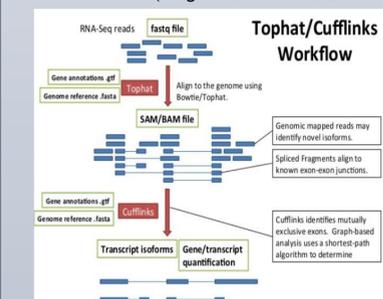
Figure 1. Standard TopHat-Cufflinks Workflow
(Source: "Introduction to RNASeq Data Analysis and Experimental Design" slideshow posted on SlideShares.net)

- RNA-Seq reads (FASTQ format) were aligned and mapped to the reference human genome assembly via TopHat/Bowtie software.
- Gene- and transcript-level expression was quantified with Cufflinks software to produce the output file of normalized gene expression as FPKM values.

- Correlation and scatter plot analysis were done to show conservation of integrity between the original T-cell ALL patient leukemia and the xenografted leukemia.
- The expression level of the patient's high frequency gene expression signatures for T-cell ALL was compared with that of the xenograft to determine if xenograft tumor remained T-cell ALL.
- Correlation analysis and gene-level expression of surface markers were done to determine if xenograft's cells retained marker characteristics of the patient's T-cells.
- Functional analyses utilizing the online DAVID Bioinformatics Database were done on the most similarly expressed genes (i.e. within 15% of gene expression level difference) and the most differentially expressed genes (i.e. having at least 2-fold difference in gene expression level) to find common biochemical processes retained and altered in the xenograft, respectively.

Results

Patient's Clinical Background:

The patient was a deceased pediatric male with a history of T-cell lymphoblastic lymphoma of a mediastinal mass first diagnosed in 10/2010, found to relapse to T-ALL in 3/2012, with the below labs showing pancytopenia, specific T-cell ALL markers and abnormal karyotype with Trisomy 20:

- CBC (3/16/2012): WBC 4.8, RBC 2.6(L), HgB 8.3(L), Platelet 16(L), POLYS (SEGS) 45%, BANDS 8%, LYMPHS 22%, MONOCYTES 3%, EOSINOPHILS 2%, BASOPHILS 1%, METAMYELOCYTES 3%, OTHERS 16%
- Phenotype: CD2, CD3, CD4, CD5, bright CD7, CD8 and CD38. Also express cCD3, but are negative TdT. Variable expression of CD10.
- Karyotype: 47,XY,add(6)(p25),i(7)(q10),add(9)(q34),del(17)(q23),+20[8]. 46,XY[12].

1. Good Correlation Between the Patient and Xenograft Leukemia Sample:

Table 1. Correlation Analysis on Genes Filtered for Having Expression in At Least Two of the Samples (FPKM ≥2)

	Pearson Coefficient	Spearman Coefficient
Correlation on the entire 21,622 genes	0.799488478	0.877581599
Correlation on the 11,003 genes with low-to-high expression level (FPKM >= 2) and at least 1-fold difference from patient to xenograft	0.797985648	0.719141881
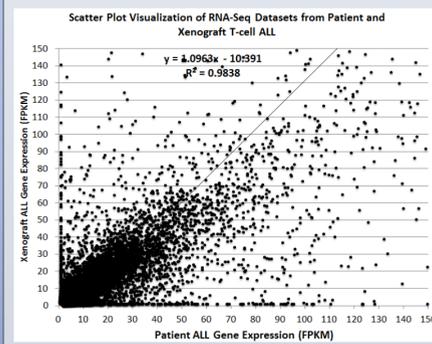
Figure 2. Scatter Plot Visualization of RNA-Seq Datasets from Patient and Xenograft T-cell ALL (FPKM between 2 and 150)

With the R Squared value of 0.98, i.e. very close to 1, the regression line equation shows a high linear correlation between the patient and xenograft gene expression level.

$y = 1.0963x - 10.991$
$R^2 = 0.9838$
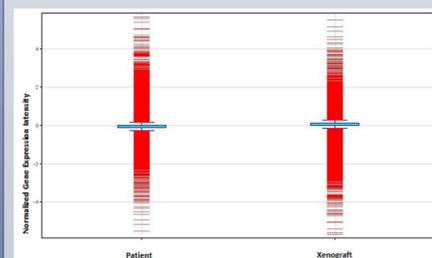
Figure 3. Whisker Box Plot for Gene Expression Level between Patient and Xenograft

The plot shows comparable gene expression level between the patient and xenograft.

- Pearson correlation shows consistent good linear correlation on both gene sets
- Spearman correlation shows high correlation on the entire gene set and a lower but still good correlation on the low-to-high gene expression set.
- Although the difference between Pearson vs. Spearman coefficient may suggest outlier data affecting Pearson analysis, these results suggest a consistently high correlation on gene expression level between patient and xenograft.

2. The Xenograft Exhibiting Most of the High-Frequency Gene Expression Signatures for the Patient's T-cell ALL:

- Studies identify high frequency gene expression signatures for T-cell ALL which are common recurrent cytogenetic and molecular alterations between all molecular T-cell ALL subtypes affecting cell cycle signaling, cell growth and proliferation, chromatin remodeling, T-cell differentiation, and self-renewal [Van Vlierberghe P, et al. J Clin Invest. 2012;122:3398].
- From these lists of high frequency (3% to at least 60%) gene expression signatures for T-cell ALL, the patient was found to express a 15-signature genes consisting of the NOTCH1, FBXW7, WT1, LEF1, BCL11B, RUNX1, PTEN, NRAS, JAK1, IL7R, EZH2, SUZ12, EED, PHF6, and TAL1 gene.
- The Pearson correlation coefficient among these 15 signature gene expression between the patient and xenograft was 0.815, consistent with the good correlation of the entire gene set.

Results (cont.)
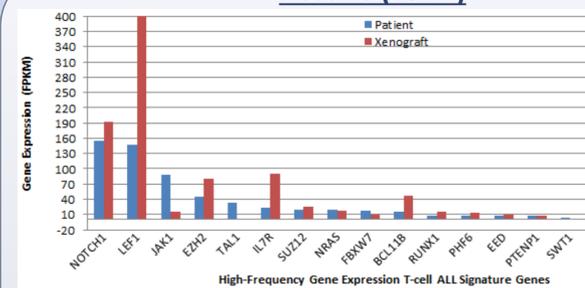
Figure 4. Comparison between Patient and Xenograft on High-Frequency Signature T-cell ALL Gene Expression

- Though the JAK1 and TAL1 signature genes were suppressed in xenograft, most high-frequency gene expression T-cell ALL signature genes (13 of 15 genes) were expressed in similar or slightly elevated level in xenograft.
- JAK1 gene involving in T-cell differentiation and transcription factor oncogenes became 5.6-fold suppressed in xenograft, suggesting perhaps the xenograft utilized an alternate signaling pathway.
- TAL1 gene involves in T-cell differentiation and transcription factor oncogenes and became 32-fold suppressed in xenograft, implying normal TAL1 functioning. This may be explained by a mutated sequence of the patient's aberrant looping region flanking TAL1 locus in the xenograft therefore voiding the patient's TAL1 activation [Zhou Y1, et al. Blood. 2013;122:4199].

3. Good Gene Expression Conservation of Patient's Surface Markers on Xenograft:

- The patient's specific surface ALL marker gene expression for the CD2, CD3, CD4, CD5, CD7, CD8, and CD38 genes with variable CD10 expression and the lack of the TdT gene expression were analyzed and compared.
- Pearson correlation coefficient was 0.840, slightly better than overall correlation.
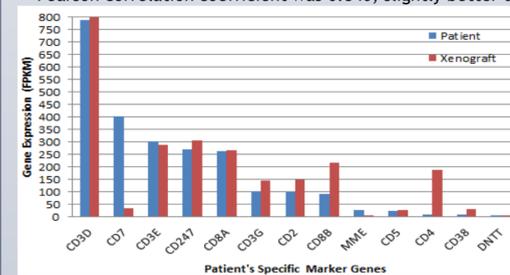
Figure 5 . Comparison between the Patient versus Xenograft for Patient's Specific Marker Gene Expression

- Most of the patient's surface ALL marker genes (10 of 13 genes) show similar or higher expression level in xenograft.
- The TdT (aka DNTT) gene was expressed at comparably low level (patient 3.3 FPKM and xenograft 2.7 FPKM)
- CD10 (aka MME or CALLA) gene, a common ALL marker involving apoptotic capability, was expressed at moderate level in patient and low level in xenograft.
- CD7, a co-stimulatory receptor for T-cell activation via intracellular signaling, was 12-fold reduced in the xenograft expression. Studies show CD7 expression can be variable, e.g. increased in one ALL group [Zhang Y, et al. Zhonghua Bing Li Xue Za Zhi. 2015;44:57. Chinese] or at low or absent level in another ALL group [Sarma A, et al. Cancer Biomark. 2015;15:501].
- CD4 expression involving T-cell regulation was dramatically 18-fold up-regulated in the xenograft, possibly indicative of selection for a mature cellular phenotype.

4. Statistics on Similarly and Differentially Expressed Genes:

Table 2. Statistics of Gene Sets

Gene Set	# of Genes	# Common Biochem Processes
Unique genes	21,622	Too large for search on DAVID
Genes with low-to-high expression (FPKM >=2) and having >= 1 fold of gene expression difference in xenograft	11,003	Too large for search on DAVID
Genes with low-to-high expression (FPKM >=2) and having >= 2 folds of gene expression difference in xenograft	3,044	Too large for search on DAVID
Top Similarly Expressed Genes with low-to-high expression (FPKM >=2) and within 15% of gene expression difference in xenograft	2,548	800
Top Differentially Down-regulated Xenograft Genes with moderate-to-high expression (FPKM >=20) and Down-regulated in xenograft and >= 2 folds of gene expression difference in xenograft as compared to that of patient	606	197
Top Differentially Up-regulated Xenograft Genes with moderate-to-high expression (FPKM >=20) and Down-regulated in xenograft and >= 2 folds of gene expression difference in xenograft as compared to that of patient	140	37

The Top Similarly Expressed Genes and the Top Differentially Up-regulated Xenograft Genes have very high total genes count over common biochemical process ratio. The opposite was shown in the Top Differentially Down-regulated Xenograft Genes.

Results (cont.)

5. Conservation of Homeostasis and Selected Cell-Cycle-Related Processes Common Among the Most Similarly Expressed Genes:

Table 3. Common Biochemical Processes for the Most Similarly Expressed Genes in Patient & Xenograft

Biochemical Processes	Number of Genes	P-Value
Protein localization	202	0.00000
Protein transport	185	0.00000
Establishment of protein localization	185	0.00000
RNA processing	183	0.00000
Macromolecule catabolic process	181	0.00000
Cell cycle	175	0.00000
Proteolysis	173	0.00009
Phosphate metabolic process	167	0.00001
Phosphorus metabolic process	167	0.00001
Cellular macromolecule catabolic process	166	0.00000
Intracellular transport	163	0.00000
Macromolecular complex subunit organization	162	0.00000
Macromolecular complex assembly	152	0.00000
Protein catabolic process	150	0.00012
Phosphorylation	146	0.00004
Proteolysis involved in cellular protein catabolic process	143	0.00000
Cellular protein catabolic process	143	0.00000
Modification-dependent macromolecule catabolic process	138	0.00000
Modification-dependent protein catabolic process	138	0.00000
Translation	136	0.00000
Cell cycle process	129	0.00000
Regulation of cell death	127	0.03684
Regulation of apoptosis	126	0.04029
Regulation of programmed cell death	125	0.03231

- A large number of genes involving in the homeostasis, apoptosis and selected cell-cycle related processes was conserved from the patient to the xenograft.
- This may suggest in order for the T-cell ALL to be preserved in xenograft, conservation of these processes was perhaps essential during serial transplantation passage.

6. Suppression of Immune Response and Cell Death Regulation Processes Common Among the Most Down-regulated Xenograft Genes:

Figure 6. Comparison of Expression Level for the Top 20 Most Down-regulated Xenograft Genes

Table 4. Common Biochemical Processes for the Down-regulated Genes in Xenograft

Biochemical Processes	Number of Genes Involved	P-Value
Intracellular signaling cascade	56	0.00355
Immune response	44	0.00001
Homeostatic process	43	0.00010
Regulation of apoptosis	41	0.00144
Regulation of programmed cell death	41	0.00172
Regulation of cell death	41	0.00184
Immune response	39	0.00002
Defense response	37	0.00012
Cellular homeostasis	34	0.00001
Cell death	33	0.01943
Death	33	0.02079
Response to organic substance	32	0.03242
Biological adhesion	30	0.05589
Cell adhesion	30	0.05678
Response to wounding	29	0.00314

7. Amplified Regulation in Immune Response, Cell Cycle, and Embryonic Development Common Among the Most Up-regulated Xenograft Genes:

Figure 7. Comparison of Expression Level for the Top 20 Most Up-regulated Xenograft Genes

Table 5. Most Common Biochemical Processes for the Up-regulated Genes in Xenograft

Biochemical Processes	Number of Genes Involved	P-Value
Cell cycle	16	0.00013
Immune response	12	0.00501
Regulation of cell cycle	9	0.00139
Cell cycle process	9	0.00045
Cell cycle phase	8	0.01815
Cell proliferation	8	0.02333
Chordate embryonic development	7	0.02079
Embryonic development ending in birth or egg hatching	7	0.02162
Embryonic development via intracellular	7	0.02487
Regulation of kinase activity	7	0.02877
Mitotic cell cycle	7	0.03342
Regulation of transferase activity	7	0.03418
Regulation of phosphorylation	7	0.08288
Regulation of phosphate metabolic process	7	0.09590
Regulation of phosphorus metabolic process	7	0.09590

Conclusion

We demonstrated the xenograft model preserved the patient's molecular and leukemia characteristics via the below findings:

1. The patient and xenograft genome showed similar expression at the molecular level via consistently good correlation in correlation and scatter plot analysis.
2. At the gene-to-gene level, the xenograft cells continued to express at similar or amplified level most of the high-frequency gene expression signatures T-cell ALL.
3. At the gene-to-gene level, the xenograft cells had good gene conservation on most of the patient's ALL surface markers.

We further investigated and found common biochemical processes that the xenograft cells retained or altered mostly involved homeostasis, cell cycle regulation, and immune modulation processes. Further studies are needed to understand the specific underlying biological drive from these results.

Future directions on the project to validate this T-cell ALL mouse model are:
- Use gene expression data to validate the patient's karyotype
- Perform detailed Fold-Change analysis for further study on altered genes

Acknowledgments