

e-Blood

RNA-seq analysis of 2 closely related leukemia clones that differ in their self-renewal capacity

Brian T. Wilhelm,^{1,2} Mathieu Briau,^{1,3} Pamela Austin,^{1,2} Amélie Faubert,¹ Geneviève Boucher,^{1,3} Pierre Chagnon,^{1,4} Kristin Hope,^{1,2} Simon Girard,^{1,2} Nadine Mayotte,^{1,2} Josette-Renee Landry,^{1,4} Josée Hébert,^{1,5} and Guy Sauvageau^{1,2,5}

¹Institute for Research in Immunology and Cancer, University of Montreal, Montreal, QC; ²Laboratory for Molecular Genetics of Stem Cells, Montreal, QC; ³Bioinformatics Platform, Montreal, QC; ⁴Genomics Platform, Montreal, QC; and ⁵Leukemia Cell Bank of Quebec and Division of Hematology, Maisonneuve-Rosemont Hospital, Montreal, QC

The molecular mechanisms regulating self-renewal of leukemia stem cells remain poorly understood. Here we report the generation of 2 closely related leukemias created through the retroviral overexpression of *Meis1* and *Hoxa9*. Despite their apparent common origin, these clonal leukemias exhibit enormous differences in stem cell frequency (from 1 in 1.4, FLA2; to 1 in 347, FLB1), suggesting that one of these leukemias undergoes nearly unlimited self-renewal divisions. Using next-generation RNA-sequencing, we characterized the transcriptomes of

these phenotypically similar, but biologically distinct, leukemias, identifying hundreds of differentially expressed genes and a large number of structural differences (eg, alternative splicing and promoter usage). Focusing on ligand-receptor pairs, we observed high expression levels of *Sdf1-Cxcr4*; *Jagged2-Notch2/1*; *Osm-Gp130*; *Scf-cKit*; and *Bmp15-Tgfb1/2*. Interestingly, the integrin beta 2-like gene (*Itgb2l*) is both highly expressed and differentially expressed between our 2 leukemias (~ 14-fold higher in FLA2 than FLB1). In addition, gene

ontology analysis indicated G-protein-coupled receptor had a much higher proportion of differential expression (22%) compared with other classes (~ 5%), suggesting a potential role regulating subtle changes in cellular behavior. These results provide the first comprehensive transcriptome analysis of a leukemia stem cell and document an unexpected level of transcriptome variation between phenotypically similar leukemic cells. (*Blood*. 2011;117(2):e27-e38)

Introduction

The ability of a distinct population of cells, termed stem cells, to maintain pluripotency is the foundation for the development of complex multicellular life. Within mammals, such multipotent cells are required to maintain a homeostatic blood supply, with a full complement of various terminally differentiated cell types.¹ Despite the importance of this process, the molecular mechanisms underlying the maintenance of stem cells is still poorly understood. It is clear, however, that in many cases the neoplastic transformation of blood cells often involves either the failure or reversal of termination differentiation, resulting in primitive “cancer stem cells” (CSCs), which divide in an unregulated manner.²⁻⁴ CSCs are best described in humans, in which the rare so-called leukemia stem cells (leukemic hematopoietic stem cells [L-HSCs]) can be prospectively isolated and shown to transmit the disease when introduced into immunocompromised mice.⁵ Cells that do not share this phenotype often represent the bulk of the leukemic clone but fail to transmit the disease on transplantation.

In syngeneic mouse models of leukemia, L-HSC frequencies have been found to vary considerably and, in some cases, to be further enriched in fractions of the bulk population that express markers typically associated with normal HSCs. Neering et al,⁶ for example, reported that BCR-ABL + NUP98-HOXA9-induced mouse acute myeloid leukemia (AML) show a very low frequency of L-HSCs, which can be enriched to 1 in 7 cells by sorting for the

lineage-negative Sca1⁺ FLT3⁺ population. In other murine AML models initiated by different oncogenic lesions, L-HSCs can be observed over a much wider range of frequencies, for example, 0.003% for Pten^{-/-} AML,⁷ 1% in E2A-PBX1 + Hoxa9 AML,⁸ and 2.3% to 71.4% for several Hoxa9 + Meis1-induced leukemias (current article).

Genes that are functionally significant for L-HSC expansion may represent the ultimate therapeutic targets for the disease. Such genes may be specific to the molecular alterations that initiate or sustain these diseases. For example, Meis1, a cofactor to Hoxa9 in AML, was recently reported as an important determinant of L-HSC frequency in MLL-AF10- and MLL-AF4-induced leukemias.⁹ Similarly, shRNA targeting of the myocyte enhancer factor 2 gene c, expressed at higher levels in L-HSCs of MLL-AF9 leukemias, decreased the frequency of L-HSCs by more than 50%.¹⁰ The polycomb group gene *Bmi1* is also required for L-HSC self-renewal in Hoxa9 + Meis1-induced leukemias.¹⁰

In this paper, we describe the generation of a series of Hoxa9 + Meis1 AML derived from purified fetal liver (FL) cells. As detailed below, all these normal karyotype leukemias are remarkably similar in several aspects, including their frequency of L-HSCs (typically between 1 in 30-44) except for one leukemia (FLA2) in which almost each cell is an L-HSC. Given the possibility that differences in CSC determinates may be reflected in

Submitted July 4, 2010; accepted October 6, 2010. Prepublished online as *Blood* First Edition paper, October 27, 2010; DOI 10.1182/blood-2010-07-293332.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

© 2011 by The American Society of Hematology

the transcriptomes of these 2 leukemias, we performed RNA-seq^{11,12} to characterize the respective cell types using a SOLiD next generation sequencer, in addition to analyzing the same RNA samples on Nimblegen expression microarrays. We generated 9 or 10 Gbp (Gigabase pairs) worth of mapped, strand-specific, sequence per cell type from 2 biologic replicates (individually grafted mice). Detailed analysis of these sequencing data revealed that a large number of genes are differentially expressed, including virtually all genes seen to be differentially regulated by microarray analysis. The RNA-seq data also revealed a surprisingly large number of structural variations in the transcriptomes of the 2 leukemias that hold the potential to substantially alter the proteome content between the leukemias. This study clearly demonstrates that even very closely related leukemias may differ substantially at the level of transcriptome diversity.

Methods

Mice

Donor C57BL/6:Pep3b (Ly5.1) and recipient C57BL/6 (Ly5.2) mice were bred and maintained as described.¹³ Experimental procedures were revised and approved by the University of Montreal animal ethics committee (Comité de Déontologie de l'Expérimentation sur les Animaux de l'Université de Montréal).

Isolation of murine hematopoietic cell subpopulations

Cells were first purified for Sca1 using magnetic activated cell sorting as per the manufacturer's instructions (Miltenyi Biotec). Further antibody staining and sorting to purify the Kit⁺, Lineage⁻ (CD45R⁻, CD3⁻, Gr-1⁻, Ter119⁻) population were performed as described.¹⁴

Retroviral infection and transplantation

Infection and transplantation procedures were performed as described.¹⁵ Briefly, *Hoxa9-ires-Meis1-pgk-Neo* retroviral vectors were constructed, and high-titer viral producers were generated in the GPE-86 packaging line. FL cells obtained from (PepC3)F1 (Ly5.1) mice were cocultivated on irradiated viral producer cells using infection medium consisting of Dulbecco modified Eagle medium, 15% fetal calf serum, 100 ng/mL IL-11, 50 ng/mL Fms-like tyrosine kinase 3 ligand, 100 ng/mL stem cell factor, 10 μg/mL ciprofloxacin (Serologicals), and 6 μg/mL polybrene (hexadimethrine bromide, Sigma-Aldrich).

Cell culture and characterization

Methylcellulose cultures. Progenitor assays were plated and scored as described.¹³

CRU assay. Stem cell quantification assays were performed essentially as described.¹⁶

Inverse PCR, comparative genomic hybridization, and spectral karyotyping. These techniques were performed as described.¹⁷

Flow cytometry studies

Flow cytometric analyses were performed using an LSRII cytometer (BD Biosciences). Monoclonal antibodies were specific for the following: Sca-1 (D7), Mac1 (M1/70), Gr-1 (RB6-8C5), CD41 (MWR30), CD19 (1D3), CD43 (S7), CD4 (H129.19), and CD8 (53-6.7) from BD Biosciences Pharmingen; Kit (2B8), CD48 (HM48-1), CD34 (RAM34), CD71 (R17217), and CD45R (RA3-6B2) from eBioscience; and CD150 (TC15-12F12.2) and TER-119 from BioLegend. A mixture of monoclonal antibodies against CD45R, TER-119, and Gr-1 was used as a lineage marker (Lineage). Side population staining was performed as previously described.¹⁸ The peak excitation wavelength for oxidized 2-dichlorofluorescein was 488 nm and emission was 525 nm.

Collection of RNA FLA2/FLB1 cells

FLA2/FLB1 cells were thawed and then washed twice (175g for 10 minutes) in 2% fetal bovine serum (Invitrogen) in phosphate-buffered saline. A total of 2×10^5 cells/mouse was then injected into the tail vein of mice sublethally irradiated (700 cGy) C57BL/6 mice. Mice were killed approximately 4 weeks after injection, femur and tibia bones were removed, and bone marrow was flushed with 3 mL of phosphate-buffered saline. Bone marrow aspirates were centrifuged (400g for 2 minutes) and aspirated before being resuspended in 10 mL of Trizol (Invitrogen). RNA was extracted according to the manufacturer's protocols before being treated by RNase-free DNase (Ambion) for 30 minutes at 37°C to degrade any contaminating genomic DNA. After confirming RNA quality and integrity by Bioanalyzer (Agilent Technologies), rRNA was then specifically depleted from samples using the RiboMinus kit (Invitrogen) and polyadenylated transcripts were enriched using the Oligotex Direct mRNA kit (QIAGEN) according to their respective manufacturer's protocols.

Microarray analysis

Total RNA was reverse-transcribed (RT) using a Superscript double-stranded cDNA synthesis kit (Invitrogen). A total of 1 μg of either purified cDNA was labeled with Cy3- or Cy5-conjugated nonamers (TriLink BioTechnologies) using Klenow fragment (3'→5' exo-) according to Nimblegen recommendations. Labeled samples were competitively hybridized on catalog mouse expression arrays (MM8_60mer expr, Roche-Nimblegen) before being washed and scanned according to Nimblegen recommendations. Signals for biologic replicates for each cell type were averaged, and RMA (Robust Multichip Average) normalized of raw expression data and gene expression calls was performed using NimbleScan, Version 2.5 software (Roche-Nimblegen). Raw and processed microarray data were submitted to ArrayExpress under accession number E-TABM-947.

SOLiD sequencing and sequence analysis

cDNA was sequenced according to the manufacturer's protocols for the SOLiD Total RNA-Seq kit for whole transcriptome (AB Life Sciences). Briefly, 10 μg of total RNA from each cell type was depleted of rRNA and enriched for PolyA transcripts as for microarray work above, and 1 μg of this RNA was then fragmented using RNase III. Ligation of the adaptor mix A and reverse transcription were performed following the manufacturer's protocol. Libraries were size selected for fragments between 150 and 300 bp, amplified for 15 cycles of polymerase chain reaction (PCR), and purified using PureLink PCR micro kit (Invitrogen). Library concentrations were determined by quantitative PCR using a standard curve of template at known concentrations (DH10B), and approximately 0.25 ng of the FLA2/FLB1 libraries was used for each full emulsion PCR (emPCR) reaction (4 emPCR/sample). Approximately 300 millions of beads were deposited on a full slide and sequenced using the Opti Fragment Library Sequencing kit-Master Mix 50 on a SOLiD machine (Version 3).

Results

Hoxa9 + Meis1 generated leukemias with different L-HSC frequencies

A subpopulation (Kit⁺Lin⁻Sca1⁺ [KLS]) highly enriched for stem cells and multipotent progenitors was isolated from FL and evaluated for frequency of colony-forming cells (CFCs) and competitive repopulation units (CRU). A fraction of these cells were distributed at various doses (Figure 1A) in 24-well plates coated with viral producer cells transfected with the murine stem cell virus *Hoxa9-ires-Meis1-pgk-Neo* retroviral plasmid, previously documented to provide a full oncogenic complement to primitive bone marrow cells.¹⁰ After infection, the contents of each

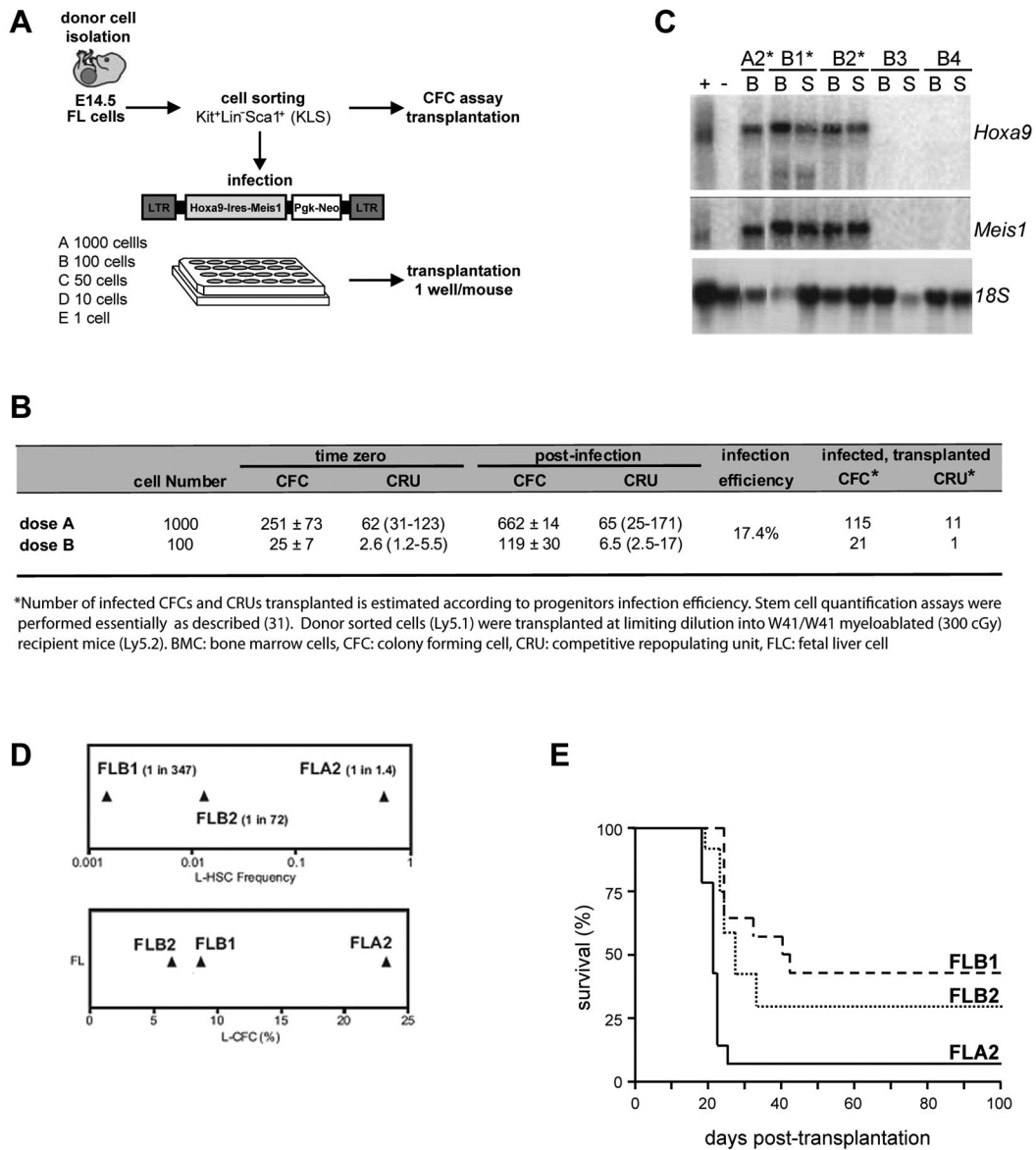


Figure 1. Generation of a stem cell leukemia (FLA2). Overview of the experimental strategy. (A) E14.5 FL cells were isolated from C57BL/6:Pep3b (Ly5.1) mice and sorted for KLS cells. A fraction of sorted cells was plated at limiting dilution in 24-well plates containing murine stem cell virus *Hoxa9-ires-Meis1a-pgk-Neo* retroviral producers. Transduced cells were transplanted into congenic mice at a dose of one well per mouse. (B) Summary of CFC and transplanted CRU infection efficiency. Numbers of infected CFCs and CRUs were estimated according to progenitor infection percentage. (C) Northern blot analysis of *Hoxa9* and *Meis1* mRNA levels in leukemia (*) and healthy mice (B3 and B4) with 18S RNA levels were used as loading control. + indicates GP + E86 viral producers of retrovirus; and -, noninfected GP + E86 cells. (D) Leukemia colony-forming cell (L-CFC; bottom panel) and L-HSC frequencies (top panel) determined in primary recipients of FLA2, FLB2, and FLB1 cultures as measured by limiting dilution analyses. (E) Kaplan-Meier-like plot comparing survivals of secondary mice transplanted at limiting dilutions (ie, one L-HSC per recipient, corresponding to 1 cell for FLA2 and several hundred for FLB1 and FLB2) where the survival curves for FLA2 and FLB1 are significantly different ($P = .0001$).

well were either transferred into methylcellulose culture to assess expansion of progenitors and gene transfer efficiency, or directly transplanted into sublethally irradiated recipients to evaluate the stem cell content (Figure 1A). The higher cell dose transplanted (1000 KLS start cells, corresponding to 115 infected CFCs and 21 infected CRUs), as well as lower quantities of infected cells (100 KLS start cells, corresponding to 21 infected CFCs and one infected CRU) resulted in typical *Hoxa9* + *Meis1*-induced myeloid leukemia occurring within the previously documented time frame¹⁹ (Figure 1B; and data not shown). An intact *Hoxa9* + *Meis1* provirus was observed in all leukemias analyzed, and both oncogenes were expressed at similar levels (Figure 1C). Of importance, leukemias did not occur in mice reconstituted with cell doses containing only transduced CFCs, and leukemia was only

observed in mice that received cell fractions containing at least one stem cell (data not shown). These results suggest that *Hoxa9* + *Meis1* transformed only a subset of primitive hematopoietic cells.

To directly assess the L-HSC compartment, limiting dilution assays were performed in secondary mice, which showed that FL-derived leukemias have wide-ranging L-HSC frequencies, from 1 per 1.4 cells (FLA2 leukemia) to 1 per 347 (FLB1 leukemia) (Figure 1D top panel). Importantly, these L-HSC frequencies remained stable during successive transplantations and on multiple freeze/thaw procedures. Of note, leukemia aggressiveness also correlated with L-HSC frequency, and the shortest time from the injection of one L-HSC to clinical AML was obtained with FLA2 leukemia (Figure 1E; and data not shown).

Table 1. RNA-seq statistics

Statistic	FLA2 replicate 1	FLA2 replicate 2	FLB1 replicate 1	FLB1 replicate 2
Total raw reads	88312565	282300223	166168360	179984287
Mapped reads	51317865	194814974	99469306	122342813
Reads mapped to splice junctions	536577	6625312	1491603	4140951
% mapped to junction	1.05	3.40	1.50	3.38
Aggregate total raw reads	370612788		346152647	
Aggregate mapped reads	246132839		221812119	
Aggregate sequence (Gbp)	10.299		9.193	

FLA2 cells are similar to other *Hoxa9* + *Meis1*-induced leukemias

A unique *Hoxa9* + *Meis1* retroviral insertion site in FLA2 cells was mapped to chromosome 2 by inverse PCR (supplemental Figure 1, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). However quantitative RT-PCR analysis revealed that none of the 20 genes within a 2-Mb distance from the integrated provirus was deregulated in FLA2 cells (data not shown). G banding, spectral karyotyping, and comparative genomic hybridization analyses confirmed that the specific biologic property of FLA2 cells was not consequent to coarse chromosomal abnormalities (supplemental Figure 2). Morphologic analysis, including dissemination characteristics (eg, spleen infiltration), also confirmed the high similarity between all generated leukemias (supplemental Figure 3A; and data not shown). Despite a striking difference in L-HSC frequency, cell surface markers revealed that FLA2 cells were similar to all other leukemias generated within the same experiment, each of which exhibited a homogeneous myeloid cell surface phenotype (Kit⁺Sca1⁻CD150⁻CD48⁺CD34⁺CD71⁺Mac1⁺GR1⁺; supplemental Figure 3B). FLA2 and FLB1 leukemias did not reveal any significant difference in the cell cycle phase distribution (data not shown); however, we did find that FLA2 leukemia contained a slightly elevated proportion of side population cells (supplemental Figure 3C), a fraction known in normal bone marrow to be enriched for cells with long-term reconstitution ability.²⁰

Generation of transcriptome data for leukemic cell types by RNA-seq and microarrays

To analyze differences in the transcriptome content of the 2 cell types, RNA was collected and analyzed from FLA2 and FLB1 mice. Cells of each type were grafted into syngeneic mice, which were killed after 4 weeks when overgrowth of leukemic cells in bone marrow was close to 90% (verified by cytopins, data not shown). RNA from each biologic replicate (2 mice for each cell type) was collected and split for analysis by RNA-seq and by expression microarray. Sequencing of rRNA depleted, polyA⁺ enriched cDNA on the AB SOLiD, Version 3.0 machine generated 370 and 346 million reads for FLA2 and FLB1 replicates, respectively, of which approximately 60% could be mapped in both cases (Table 1). Similar to other studies, we have found that 1% to 3% of mapped reads span exon-exon junctions.²¹ In looking at the position of the mapped reads, we observed good specificity for the majority of reads to fall within annotated gene features, followed by introns (probably the result of isolation of unspliced mRNA) and intergenic regions (supplemental Figure 4). It is also clear that some of the signal in the last 2 classes probably comes from novel transcribed elements. Based on the discovery plot generated from our FLA2 (supplemental Figure 5), the depth of our sequencing is

sufficiently deep enough to have discovered any directed transcription occurring.

Biologic replicates of leukemic cells show good concordance irrespective of assay method

To compare RNAseq data to microarray data, gene models were constructed by downloading the all_mrna table from University of California Santa Cruz (UCSC),²² which contains all non-EST (expressed sequence tag) mRNAs, which have been sequenced. These mRNAs were compared with themselves to find all splicing variants within a given loci. Microarray and RNA-seq data were assigned to specific genes, and the concordance of signals between biologic replicates was analyzed (Figure 2). All RNA samples showed highly consistent, and significant, correlation coefficients regardless of whether they were sequenced or hybridized on microarrays. Although the FLA2 samples did show somewhat lower correlations, the fact that this was consistently observed suggests that there was a slight biologic difference between the 2 FLA2 replicates. Given the consistency of the results, the microarray scores were averaged between replicates and the RNA-seq data pooled to simplify downstream analysis.

RNA-seq allows more sensitive monitoring of transcriptional changes in cells than microarrays

Expression histograms were generated using all genes in common to both platforms, which could be uniquely identified by a single Entrez gene ID to classify genes according to their distribution within all signals (Figure 3). As expected, the dynamic range of RNA-seq was far greater than that of the microarrays, with a greater than 5000-fold difference in their respective range of measurements.

Given the arbitrary nature of thresholds chosen for calling a gene “expressed,” we elected to set 2 thresholds, one for nonexpression and one for marginal expression, as has been done in previous studies.²³ For the microarray data, 6131 probes corresponding to random GC (guanine-cytosine content) content controls were used to assess background signals, as these probes should not hybridize to anything within the samples. The threshold for expression by microarray was set at twice the level of signal from the average of the GC control probes in each channel, whereas the limit for marginal expression was set as 4 times the average GC control signal. Using these thresholds, there were 8480 and 9031 genes expressed and 1646 and 1758 genes marginally expressed of the total of 17 262 genes (analyzed in common between microarray and RNA-seq experiments) in FLA2 and FLB1 cells, respectively.

For the RNA-seq equivalent, thresholds were set at 1-fold length normalized coverage, and 5-fold length normalized coverage for expression, based on the distribution of signals. At these thresholds, we found there were 12 115 and 11 866 transcripts

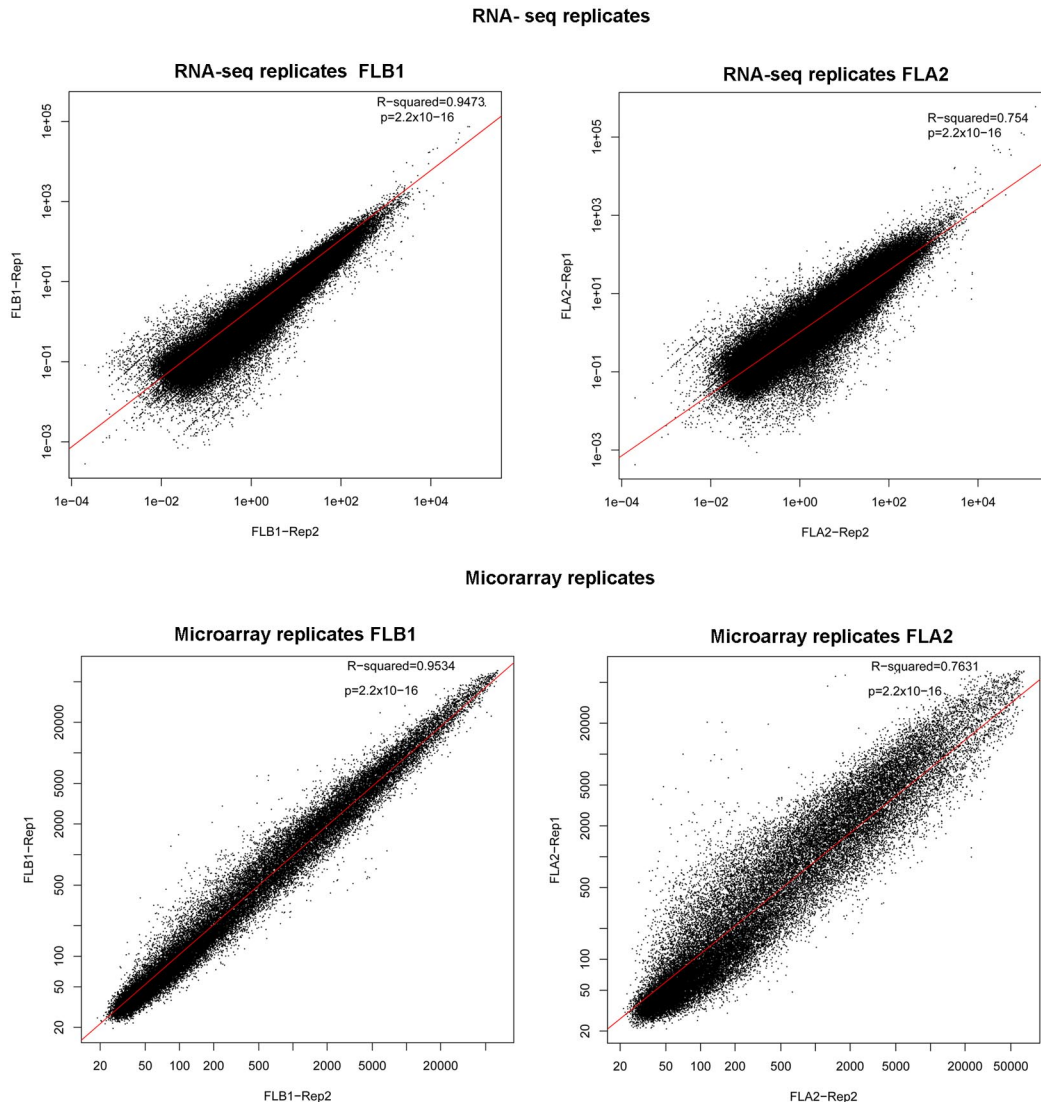


Figure 2. Comparison of biologic replicates by platform. Expression scores of genes for each biologic replicate of RNA-seq (top row) or microarray (bottom row) are shown plotted on log₁₀ scales. Linear models (red line) and concordance scores with *P* values were calculated using the *lm* and *concordance* functions and demonstrate good reproducibility regardless of assay platform used.

expressed and 1724 and 1641 transcripts marginally expressed of the total 17 262. Thus, although the number of marginally expressed genes is similar with both platforms, the greater dynamic range of RNA-seq results in more genes being categorized as being expressed.

The increased sensitivity of RNA-seq is also evident in a scatterplot of signals of each gene from each platform (Figure 3). The threshold for expression for genes by either microarray (red) or RNA-seq (green) is shown by either vertical or horizontal lines, respectively. In both cell types, the pattern of signals is similar with a fairly linear range down to the expression threshold for microarrays. Beyond this point, although all genes are not called as expressed by microarray, 2 populations of genes are visible: those that are expressed by RNA-seq (red dots) and those that are not expressed by either microarray or RNA-seq (green dots).

FLA2/FLB1 cells have differentially expressed genes despite overexpressing the same 2 genes

Unexpectedly, these 2 leukemias contained a large number of differentially expressed genes. At a threshold of 2-fold up- or

down-regulation, 647 and 1072 genes are found to be differentially regulated, and above marginal expression levels, by RNA-seq and microarray, respectively (supplemental Table 1); however, many of these ratios are only marginally over 2-fold. Using a more conservative cutoff of 5-fold reduces the number of differentially expressed genes to 137 and 93, of which the top 50 by RNA-seq are shown (Table 2). Despite the absence of probes for several genes on the microarrays, or marginal expression of others, both methodologies showed good overlap. Therefore, when gene lists using conservative cut-offs are compared, it is evident that there is good correlation between both methods and that genes determined to be differentially regulated by microarray are largely a subset of those determined by RNA-seq (Figure 4A).

Specific analysis of mouse miRNAs, obtained from miRBase,²⁴ showed that, of the 567 annotated elements, remarkably few are highly expressed, with the top 15 miRNAs (2.6%) by expression level accounting for more than 95% of all transcription from miRNAs (supplemental Table 2). Of the miRNAs that are expressed at a level probably biologically significant, only 2 (mmu-mir-2140 and mmu-mir-2133-1) exhibit differential expression

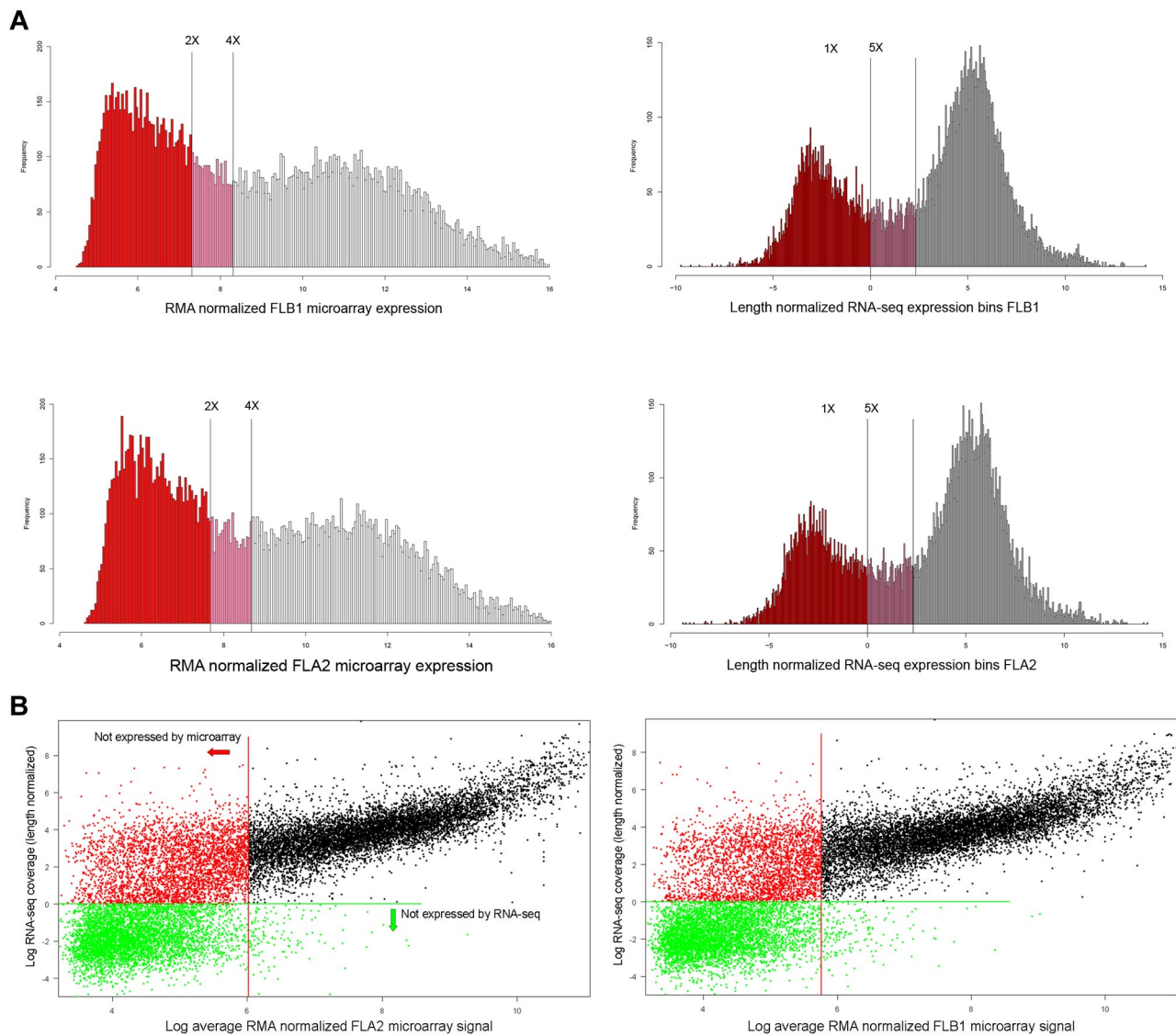


Figure 3. Expression histograms and platform comparisons. (A) Expression histograms based on microarray data (left column) and RNA-seq results (right column). The \log_2 value of the length normalized expression levels of histogram bins is plotted along the x-axis, whereas the number of genes falling into each bins is plotted along the y-axis. Thresholds for “no expression” or “marginal expression” are shown as vertical lines at either 2-fold and 4-fold above GC control probes signals (microarray) or 1-fold and 5-fold length normalized coverage (RNA-seq), respectively. Genes classed as not expressed are colored in red, marginally expressed genes in pink, and expressed as empty bars. (B) Individual gene expression levels (\log_2) are shown plotted by RNA-seq (y-axis) and by microarray (x-axis) values for either FLA2 (top panel) or FLB1 (bottom panel). Genes below marginal expression levels for microarray (2-fold GC controls, Figure 5) are shown left of the red vertical line. Genes below marginal expression levels for RNA-seq (1-fold LN coverage in panel A) are shown below the green horizontal line. The increased sensitivity of RNA-seq expression values is apparent in the 2 populations of genes visible left of the red vertical line, which all represent nonexpressed genes by microarray.

between cell types (fla2:flb1 of 0.47 and 3.17, respectively). Given the lack of experimental validation of targets of these miRNAs and the multiplicity of computationally predicted targets, it is not possible to clearly identify a mechanism by which these miRNAs would contribute to the cellular behavior differences; however, this possibility cannot be ruled out without further investigations.

Differentially expressed genes have a biased distribution

To characterize the genes showing differential regulation, we analyzed the Gene Ontology (GO) annotation²⁵ associated with the 310 genes exhibiting a greater than 3-fold difference in expression by RNA-seq. Analysis using the functional annotation clustering feature of DAVID²⁶ showed that the highest scoring annotation cluster is enriched in genes associated with GO terms for immune

system development, hemopoietic or lymphoid organ development, hemopoiesis, and myeloid cell differentiation (data not shown). The enrichment in GO terms associated with blood development and differentiation suggest that the differential expression of these genes is consistent with the differences in self-renewal capacity exhibited by the cell types.

As an alternative method to ascertain biases in GO term annotation with respect to differentially expressed genes, the GO terms for a wide variety of cellular processes were selected from AMIGO. These classes were then used to determine the percentage of genes expressed in each class (by RNA-seq) and then, of these, the number that were differentially expressed (Table 3; supplemental Table 3). Interestingly, whereas most processes had 50% to 80% of their associated genes being expressed and 2% to 6% of the genes differentially expressed, G-protein-coupled receptor activity was a

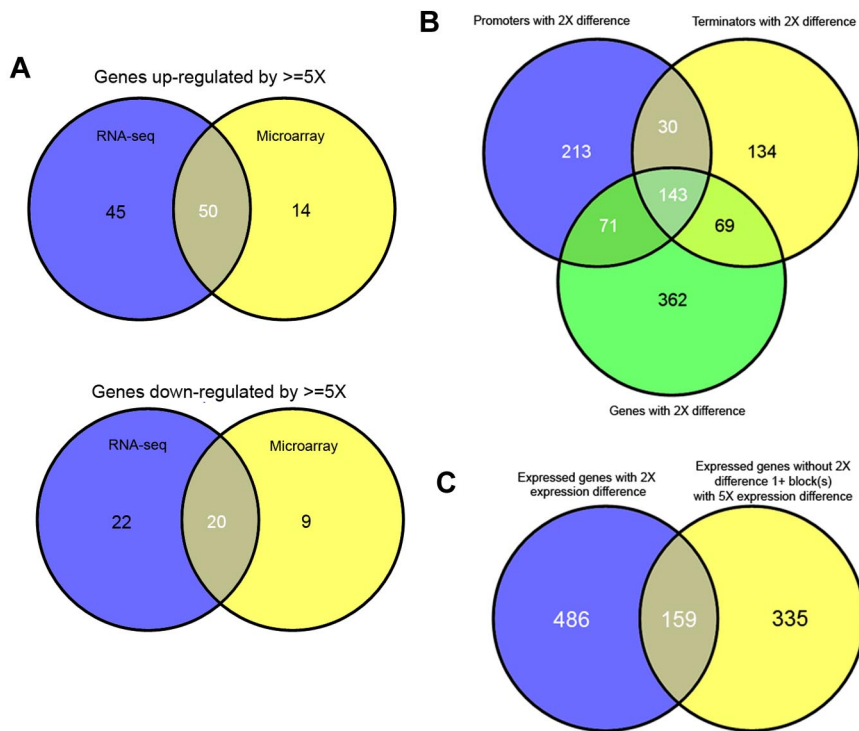


Figure 4. Comparison of structural variations and differentially expression genes within transcriptomes. The overlap between genes determined to be up-regulated (top) or down-regulated (bottom) by RNA-seq (blue) or microarray (yellow) is shown by Venn diagrams (A). Genes up-regulated/down-regulated by microarray generally represent a subset of genes up-regulated by RNA-seq. The overlap between promoters, terminators, and genes, which show a differential usage/expression of more than 2-fold, are shown (B). The overlap between genes with differential expression (> 2 -fold) and blocks (exons or parts of exons), which show differential expression (> 5 -fold) but are contained with genes, which do not show differential expression are shown (C).

notable exception where only approximately 5% of genes were expressed but approximately 22% were differentially expressed (Table 3). This effect is unlikely to simply be the result of size differences between classes given that the G-protein-coupled receptor and developmental process classes have approximately similar numbers of members but a percentage of expressed and differentially expressed genes that differs significantly between the 2.

The transcriptomes of FLA2/FLB1 exhibit numerous structural differences

Given the potential to analyze transcript structure using RNA-seq data,²⁷ we undertook to characterize the differential use of alternative promoters, terminators, and exons within the transcriptomes of FLA2/FLB1. To begin with, we computationally defined all single and multiple promoters and terminators based on all mRNAs downloaded from UCSC and measured the usage of these elements based on coverage by cell type (Tables 4,5). Interestingly, a strong bias toward the use of multiple promoters and terminators, where they exist, was observed, with approxi-

mately two-thirds of all being used in preference to a single promoter or terminator.

When we compared all promoters and terminators that had a more than 2-fold differential expression/usage, we found, as expected, that a majority of these features belonged to genes that also had an overall 2-fold difference (Figure 4B). For the 213 promoters and 134 terminators that did not overlap, we could see examples of alternative usage by cell type (supplemental Figure 6A). However, it is also probable that some of these nonoverlapping promoters and terminators result from slight differences in coverage of elements close to the threshold.

We also examined all exons for evidence of differential usage and again could find evidence for exons that showed differential usage when the overall genes did not. A large majority (67%) of exons that exhibited a 5-fold or greater differential usage were located within genes that did not have more than an overall 2-fold differential expression (Figure 4C). Although the same caveat regarding expression differences at or near thresholds applies, the higher threshold for exon usage mitigates this to a large degree. Through manual examination, we can again find clear

Table 3. Differential expression by GO category

GO category	Total in class	Expressed in class, no.	Expressed in class, %	Differentially expressed, no.	Differentially expressed, %
Transcriptional activation	288	136	47.22	7	5.15
Transcriptional repression	227	113	49.78	7	6.19
Translational regulation	82	64	78.05	4	6.25
Chromosome organization	64	49	76.56	2	4.08
Kinase activity	382	261	68.32	11	4.21
GPCR activity	1258	66	5.25	15	22.73
Nuclear export activity	24	21	87.50	1	4.76
Developmental process	1247	554	44.43	52	9.39
Chromatin organization	308	265	86.04	7	2.64

GPCR indicates G-protein-coupled receptor.

Table 4. Statistics of alternative promoter usage

	Single promoters FLA2	Single promoters FLB1	Multiple promoters FLA2	Multiple promoters FLB1
Determined	15699	15699	6046	6046
Nonexpressed	9975	9959	2473	2490
Expressed from single promoters	5724	5740	1049	1026
Expressed from multiple promoters	NA	NA	2524	2530

NA indicates not applicable.

examples of differential splicing without differential expression (supplemental Figure 6B-C).

Identification of SNV differences in cell types

The nucleotide resolution of the RNA-seq data allowed us to identify single nucleotide variations (SNVs) in either cell line with respect to the public reference sequence. To restrict our analysis to positions with sufficient sequence coverage, we required at least 20 reads covering a position in aggregate between the 2 biologic replicates before including it. Comparing commercial inbred mouse strains with the related reference sequence, we could find 1825 SNVs (supplemental Table 4) within mRNAs, which were identical between our cell types but differed synonymously from the reference sequence. In addition to these changes, we could detect 470 nonsynonymous SNVs in either cell relative to the reference sequence. These totals would seem high given the relative similarity of the cells, and their derivation ultimately from the same mouse strain despite their extended passaging *in vivo*; and so, similar to other deep sequencing studies,^{28,29} a large majority of these changes probably represent sequencing errors. To apply more restrictive selection criteria, when only consistent calls between biologic replicates are analyzed (whether heterozygous or homozygous) are used, there are 74 nonsynonymous changes identified. Interestingly, these changes include a mutation in the *Meis1* gene, where the C at genomic position 18911310 in the reference is consistently called a C in FLA2 cells (413 reads), whereas it is consistently called a T in FLB1 cells (781 reads). This SNV is predicted to change an aspartic acid to an asparagine in the HR1 region previously identified to be important for interaction with Pbx1.³⁰ A small percentage of the changes were also predicted to induce premature stop codons in the canonical protein sequence. We attempted to validate 10 of the predicted nonsynonymous changes, through sequencing-selected PCR-amplified regions from the cDNA. We were able to validate 2 of the changes, including the change in *Meis1*, which although somewhat low, agrees with the validation rate in similar studies.^{28,29}

The genomes of FLA2/FLB1 contain numerous unannotated transcribed elements

Although our analysis focused on annotated features derived from cloned mRNAs, we also sought to identify novel transcribed regions using our RNA-seq data. Using a 50-bp sliding window approach, we defined novel elements as starting when they had a minimum coverage level of 3-fold within the window and ending

once the coverage dropped below this limit. Novel elements identified with this approach were required to have a minimum size of 75 bp and a maximum overlap of 25% with annotated repetitive elements to be included. To further reduce the number of false positive elements in the list, we then applied a highly conservative minimum expression level of 16-fold (a log₂ value of 4 on Figure 3) for all elements (supplemental Table 5) and sought to classify all remaining elements into various categories as depicted (Figure 5A).

Most novel elements identified in FLA2 cells (supplemental Figure 7B) were relatively small in size, with a size distribution centered around approximately 120 bp in size (supplemental Figure 8). As a result, we partitioned the 2 largest classes of elements unrelated to known genes (antisense and novel genes) into 2 size groups, those above or below 150 bp in size, to avoid potentially mixing 2 distinct populations of elements. As demonstrated by the summary of elements (Table 6), with the exception of novel untranslated regions (which represent extensions of known mRNAs), more than 75% of novel elements in FLA2 cells are found within the same class in FLB1 cells, suggesting that these elements are genuine unannotated genes/exons. To find additional evidence that these novel elements are the result of directed transcription, we looked to determine whether they overlapped with regions of genome, which show strong levels of evolutionary conservation in the PhastCons Conserved Elements³¹ track of UCSC. When we examined only the novel elements that showed any degree of overlap with PhastCons elements, the average percentage of base pairs covered by PhastCons regions in these elements is very high (~50%-67%). In addition to this evidence, in many (though not all) cases, we were able to identify strong EST support for transcription in the regions as well as Ensembl gene models (supplemental Figure 4).

Given the robust and widespread role for antisense transcription,³² we used our strand-specific RNA-seq data to identify regions of antisense transcriptions within annotated regions. As expected, we found a large number of transcribed novel elements overlapping annotated genes with the majority of these residing within introns (supplemental Table 6). Apart from the number of elements, there does not appear to be any significant difference in the distribution of the large or small antisense elements identified. We could not identify any clear relationship between the expression of the genes and novel antisense elements that overlapped them (data not shown); however, the regulatory relationship between sense/antisense transcripts can be complex and would

Table 5. Statistics of alternative terminator usage

	Single terminators FLA2	Single terminators FLB1	Multiple terminators FLA2	Multiple terminators FLB1
Determined	15947	15947	5798	5798
Nonexpressed	10055	10136	2314	2352
Terminated at single sites	5892	5811	979	914
Terminated at multiple sites	NA	NA	2505	2532

NA indicates not applicable.

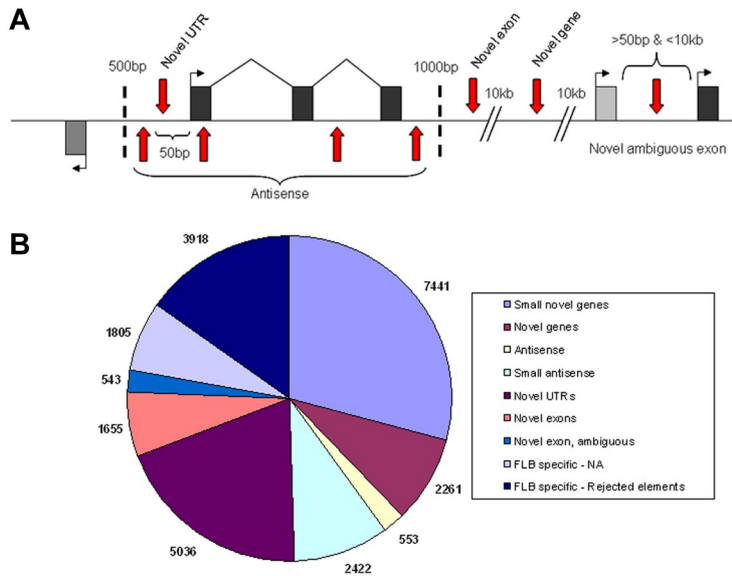


Figure 5. Classification of novel transcribed elements discovered. Exons within model genes are represented as black or gray exons, with the first exons marked by a bent arrow indicating the direction of transcription and vertical red arrows representing the various possible locations of novel transcribed elements (A). Locations relative to the model gene (shown in base pairs) represent the distant limits used for the various categories. The number of novel elements found within each category are shown (B) with small novel genes and small antisense being separated from their parent categories by a size limit of 150 bp.

probably only be definitively revealed by analyzing time-course expression data.³³

Receptor-ligand expression on leukemia stem cells

We analyzed the expression of selected signaling pathways, which have been implicated in stem cell behavior (Figure 6). The components of each pathway are colored according to their expression within a set of 10 bins (with bin 10 containing all genes not expressed) and with differential expression between FLA2 and FLB1 shown by the red and green colors (FLA2/FLB1 ratio of > 2-fold or < 0.5 in red and green, respectively). Highlighting a possible autocrine activity in L-HSC activity, the following ligand-receptor couples were expressed at the highest levels in our leukemias: *Sdf1-Cxcr4*; *Jagged2-Notch2/1*; *Osm-Gp130*; *Scf-cKit*; and *Bmp15-Tgfb1/2*. *Vegf-Vegfr2*; *Wnt6/9b/10b-Frizzled4/5/7/9/10* were expressed but at much lower levels. We also found high expression levels of several integrin genes. Of these, the integrin β 2-like gene (*Itgb2l*) is both highly expressed and differentially expressed between our 2 leukemias (~ 14-fold higher in FLA2 than FLB1). Besides this exception, no difference in expression pattern could be identified for key members of major signaling pathways thought to participate in hematopoietic stem cell self-renewal. Interestingly and in contrast to normal hematopoietic stem cells, components of the Hedgehog and thrombopoietin pathways (eg, *c-mpl*) are not expressed. Although the function of the former remains contro-

versial,³⁴⁻³⁶ *c-mpl* is critical for HSC quiescence and maintenance.³⁷

Discussion

This study presents a base pair resolution RNA-seq and microarray analysis of the transcriptome of 2 closely related leukemic cell types generated by the overexpression of *Meis1* and *HoxA9*. We have identified a substantial number of genes that were differentially expressed between FLA2/FLB1 cells, including some that have been implicated in processes, such as cell cycle progression and differentiation. The *S100a9/8* genes for instance (up-regulated in FLA2 cell) have been suggested to be used by tumors, via the activation of the mitogen-activated protein kinase p38, to facilitate mobility and invasion.³⁸ Still other genes that are highly expressed specifically in FLA2 have also been found to be highly induced when primitive hematopoietic cells overexpress transcription factors that have been demonstrated to induce the expansion of stem cells (E. Denault, B.T.W., F. Barabé, G.S., unpublished observations, April 2010). This results hint that there may be some commonality between sets of genes that regulate self-renewal of normal and cancerous stem cells. In addition, the GO annotations of the genes that are differentially expressed do show a significant bias and seem to favor differential expression of G-protein-coupled receptors as a general class. It suggests the possibility that perhaps other classes of genes, possibly involved in more fundamental

Table 6. Summary of statistics for novel elements

	No. found FLA2	Average LN coverage FLA2 (median)	Found in same class FLB1, no.	Found in same class FLB1, %	Average PhastCoverage	PhastCoverage excluding nonoverlapping, no.	PhastCoverage excluding nonoverlapping, %
Novel gene	2261	270 (34)	1924	85.1	33.7	968	67.0
Small novel gene	7441	79 (28)	5449	73.2	6	576	56.9
Antisense	553	185 (33)	458	82.8	30	221	61.7
Small antisense	2422	84 (28)	1819	75.1	12.2	403	54.7
Novel UTR	5036	16 (5)	3168	62.9	21.6	1408	48.6
Novel exon	1655	81 (28)	1282	77.5	19.1	411	59.8
Novel exon, ambiguous	543	79 (34)	464	85.5	30.2	268	52.3

LN indicates length normalized; and UTR, untranslated region.

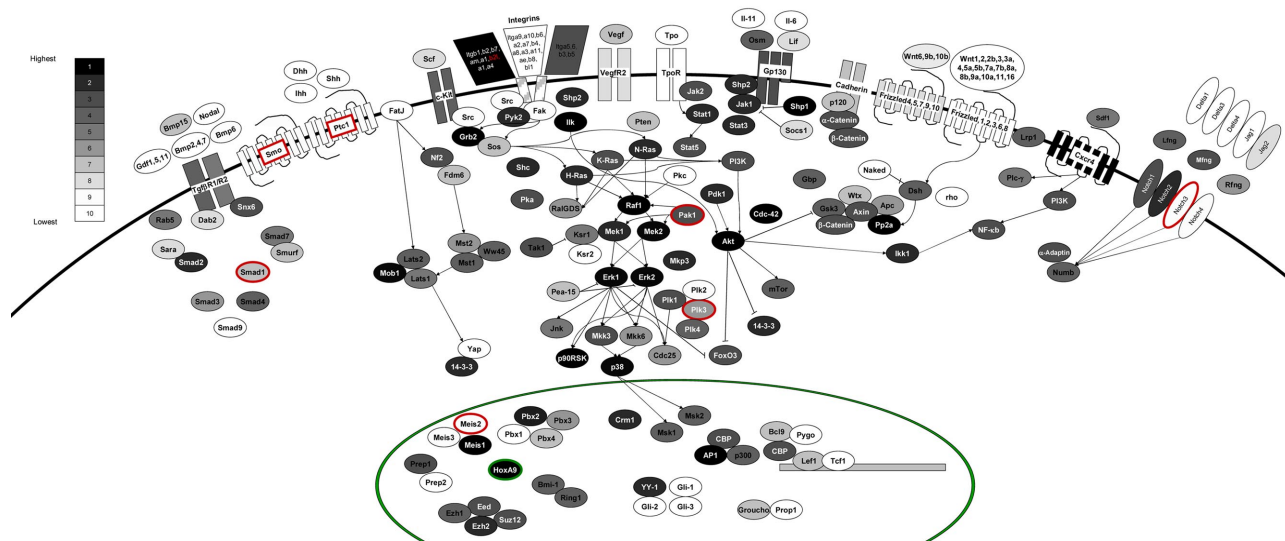


Figure 6. Signaling pathways active in FLA2/FLB1 cells. The components of a variety of intracellular signaling pathways are shown, colored according to their expression level. All expressed genes are divided evenly into bins 1 to 9 (1 being most highly expressed) with approximately 1500 genes per bin and all genes that are not expressed (~ 13k) were placed into bin 10 where expression levels are based on RNA-seq values from FLA2 cells. The color scale corresponding to expression bins is shown in top left of diagram. Genes that are differentially expressed are shown surrounded by a red or green circle (for higher expression in FLA2 or FLB1 cells, respectively) or by red text in the case of *Itgb2* (for higher expression in FLA2). Only partial wiring of known signaling pathways is shown to maintain clarity in the diagram.

aspects of cell biology, are not as suitable “evolutionary targets” for determining or regulating cell fate. Although determination of cell fate is certainly not as clear-cut as differential activation of a single class of genes, we are unaware of such a bias having been previously reported with respect to differential gene expression in broad GO categories.

Interestingly, a large number of the genes that are highly up-regulated in FLA2 cells (compared with FLB1) belong to the previously characterized cystatin gene family,^{39,40} consisting of 10 highly related genes located within a cluster on chromosome 16. These genes have not previously been documented to be involved in stem cell behavior or cancer progression and may not even be required for normal cellular growth.^{39,41} A probable explanation for their up-regulation is that these genes may share a common regulatory element, which is bound by either the overexpressed *Meis1/HoxA9* in these cells or by another transcription factor specifically up-regulated in FLA2 cells, as there is no evidence for a viral integration effect near the region of these genes by quantitative RT-PCR (data not shown) or by RNA-seq (supplemental Figure 1).

As shown in Figure 6, the integrin $\beta 2$ -like gene is both highly expressed in our leukemias and shows differential expression (14-fold higher in FLA2). Given the lack of annotation regarding the function of this gene in the context of hematopoiesis, it is not clear whether the differential expression of this gene may have direct relevance to the differences in self-renewal exhibited by the FLA2/FLB1 cells.

Variations in SNV content could also play a role in altering the proteome content. As noted, we located a high confidence, homozygous SNV in the *Meis1* gene in the FLB1 cells, which almost certainly derived from the original retroviral transfection. This mutation is within the HR1 required for interaction with PBX1 domain, the loss of which has been demonstrated to cause a longer latency of leukemia development in mice.³⁰ This apparently incidental point mutation provides a potential principle cause for the transcriptional differences between cell types: if the transcriptional complexes of *Meis1/HoxA9/Pbx1* differ in stability, efficacy, or localization, this might explain why the transcriptomes of the closely related cells are so different. Moreover, such a scenario

would mean that this difference must be “upstream” of whatever changes are induced allowing a higher level of “self renewal” in FLA2 cells, making this system even more interesting as a model for dissecting the mechanisms of L-HSC self-renewal.

As previous studies have noted, transcriptomic characterization by RNA-seq is not only as reproducible as traditional approaches but far more sensitive and generates an extremely rich dataset.^{42,43} Most remarkably, these data demonstrate that even extremely similar cells can have substantial differences in their transcriptomes not simply at the level of differential expression but also at the level of transcript structures. With respect to the question of how such differences might enable the FLA2 cells to reinitiate leukemias with orders of magnitude greater efficiency than FLB1 cells, we cannot pinpoint a single cause in the many differences that we have found. Despite this, we do now have several interesting possibilities (eg, point mutation in *Meis1* in FLB1, differential expression of HSC factors), which we can use to follow-up this initial study to further dissect the molecular mechanism responsible for these self-renewal differences.

Acknowledgments

The authors thank GS laboratory members for valuable comments on this manuscript.

This work was supported by the Canadian Institutes of Health Research (G.S.). B.T.W. was supported by a Cole foundation fellowship.

Authorship

Contribution: B.T.W. and G.S. carried out the experimental design; B.T.W., P.A., A.F., P.C., S.G., N.M., and J.H. carried out the experimental work; B.T.W., M.B., G.B., and J.-R.L. analyzed the data; and B.T.W., K.H., and G.S. wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Guy Sauvageau, Institute for Research in Downtown Station, Montreal, QC, Canada, H3C 3J7; Immunology and Cancer, University of Montreal, CP 6128, e-mail: guy.sauvageau@umontreal.ca.

References

- Zon LI. Intrinsic and extrinsic control of haematopoietic stem-cell self-renewal. *Nature*. 2008; 453(7193):306-313.
- Clarke MF, Dick JE, Dirks PB, et al. Cancer stem cells: perspectives on current status and future directions. AACR Workshop on cancer stem cells. *Cancer Res*. 2006;66(19):9339-9344.
- Huntly BJ, Gilliland DG. Leukaemia stem cells and the evolution of cancer-stem-cell research. *Nat Rev Cancer*. 2005;5(4):311-321.
- Warner JK, Wang JC, Hope KJ, Jin L, Dick JE. Concepts of human leukemic development. *Oncogene*. 2004;23(43):7164-7177.
- Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. 1994;367(6464):645-648.
- Neering SJ, Bushnell T, Sozer S, et al. Leukemia stem cells in a genetically defined murine model of blast-crisis CML. *Blood*. 2007;110(7):2578-2585.
- Yilmaz OH, Valdez R, Theisen BK, et al. Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature*. 2006;441(7092):475-482.
- Kroon E, Thorsteinsdottir U, Mayotte N, Nakamura T, Sauvageau G. NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. *EMBO J*. 2001;20(3):350-361.
- Wong P, Iwasaki M, Somerville TC, So CW, Cleary ML. Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. *Genes Dev*. 2007;21(21):2762-2774.
- Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature*. 2003;423(6937):255-260.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.
- Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*. 2009;48(3):249-257.
- Kroon E, Kros J, Thorsteinsdottir U, Baban S, Buchberg AM, Sauvageau G. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *EMBO J*. 1998;17(13):3714-3725.
- Okada S, Nakauchi H, Nagayoshi K, Nishikawa S, Miura Y, Suda T. In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood*. 1992;80(12):3044-3050.
- Thorsteinsdottir U, Sauvageau G, Humphries RK. Enhanced in vivo regenerative potential of HOXB4-transduced hematopoietic stem cells with regulation of their pool size. *Blood*. 1999;94(8):2605-2612.
- Antonchuk J, Sauvageau G, Humphries RK. HOXB4 overexpression mediates very rapid stem cell regeneration and competitive hematopoietic repopulation. *Exp Hematol*. 2001;29(9):1125-1134.
- Bilodeau M, Girard S, Hebert J, Sauvageau G. A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells. *Nat Methods*. 2007;4(3):263-268.
- Arai F, Hirao A, Ohmura M, et al. Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell*. 2004;118(2):149-161.
- Thorsteinsdottir U, Mamo A, Kroon E, et al. Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. *Blood*. 2002;99(1):121-129.
- Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med*. 1996;183(4):1797-1806.
- Shah SP, Morin RD, Khattraj J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. 2009;461(7265):809-813.
- UCSC. UCSC genome browser. 2010. <http://genome.ucsc.edu/>.
- Cloonan N, Forrest AR, Kolle G, et al. Stem cell transcriptome profiling via massive-scale mRNA-sequencing. *Nat Methods*. 2008;5(7):613-619.
- iRBase. miRBase. 2010. <http://www.mirbase.org/>.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
- Landry JR, Mager DL, Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet*. 2003;19(11):640-648.
- Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukemia genome. *Nature*. 2008;456(7218):66-72.
- Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*. 2009;361(11):1058-1066.
- Mamo A, Kros J, Kroon E, et al. Molecular dissection of Meis1 reveals 2 domains required for leukemia induction and a key role for Hoxa gene activation. *Blood*. 2006;108(2):622-629.
- Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15(8):1034-1050.
- Yelin R, Dahary D, Sorek R, et al. Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol*. 2003;21(4):379-386.
- Katayama S, Tomaru Y, Kasukawa T, et al. Antisense transcription in the mammalian transcriptome. *Science*. 2005;309(5740):1564-1566.
- Dierks C, Beigi R, Guo GR, et al. Expansion of Bcr-Abl-positive leukemic stem cells is dependent on Hedgehog pathway activation. *Cancer Cell*. 2008;14(3):238-249.
- Trowbridge JJ, Scott MP, Bhatia M. Hedgehog modulates cell cycle regulators in stem cells to control hematopoietic regeneration. *Proc Natl Acad Sci U S A*. 2006;103(38):14134-14139.
- Zhao C, Chen A, Jamieson CH, et al. Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. *Nature*. 2009;458(7239):776-779.
- Yoshihara H, Arai F, Hosokawa K, et al. Thrombopoietin/MPL signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. *Cell Stem Cell*. 2007;1(6):685-697.
- Hiratsuka S, Watanabe A, Aburatani H, Maru Y. Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nat Cell Biol*. 2006;8(12):1369-1375.
- Bilodeau M, MacRae T, Gaboury L, et al. Analysis of blood stem cell activity and cystatin gene expression in a mouse model presenting a chromosomal deletion encompassing Csta and Stfa211. *PLoS ONE*. 2009;4(10):e7500.
- Tsui FW, Tsui HW, Mok S, et al. Molecular characterization and mapping of murine genes encoding three members of the stefin family of cysteine proteinase inhibitors. *Genomics*. 1993;15(3):507-514.
- Mouse Genome Informatics. Information obtained from the NIH Mouse Knockout Inventory. *MGI Direct Data Submission*. 2004.
- Asmann YW, Klee EW, Thompson EA, et al. 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genomics*. 2009;10:531.
- Wilhelm BT, Marguerat S, Watt S, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*. 2008;453(7199):1239-1243.