

Dataset Paper

DNA Microarray Assay Helps to Identify Functional Genes Specific for Leukemia Stem Cells

Haojian Zhang¹ and Shaoguang Li²

¹ Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, MA 02215, USA

² Division of Hematology/Oncology, Department of Medicine, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA

Correspondence should be addressed to Shaoguang Li; shaoguang.li@umassmed.edu

Received 20 May 2013; Accepted 7 July 2013

Academic Editors: M. Jernas, N. Li, and K. Valerie

Copyright © 2013 H. Zhang and S. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chronic myeloid leukemia (CML) is a myeloproliferative disease derived from an abnormal hematopoietic stem cell (HSC) and is consistently associated with the formation of Philadelphia (Ph) chromosome. Tyrosine kinase inhibitors (TKIs) are highly effective in treating chronic phase CML but do not eliminate leukemia stem cells (LSCs), which are believed to be related to disease relapse. Therefore, one major challenge in the current CML research is to understand the biology of LSCs and to identify the molecular difference between LSCs and its normal stem cell counterparts. Comparing the gene expression profiles between LSCs and normal HSCs by DNA microarray assay is a systematic and unbiased approach to address this issue. In this paper, we present a DNA microarray dataset for CML LSCs and normal HSCs to show that the microarray assay will benefit the current and future studies of the biology of CML stem cells.

1. Introduction

Chronic myeloid leukemia (CML) is a clonal myeloproliferative disorder that originates from an abnormal hematopoietic stem cell (HSC) harboring the Philadelphia chromosome (Ph) [1–3]. The Ph is caused by a reciprocal translocation between chromosomes 9 and 22 ($t(9; 22)(q34; q11)$), resulting in the formation of the chimeric BCR-ABL protein [1, 4]. As a result, the BCR-ABL oncogene is formed. BCR-ABL is a constitutively active tyrosine kinase, which has provided the molecular basis for designing therapeutic drugs. Imatinib is the first tyrosine kinase inhibitor (TKI) found to be highly effective in CML treatment through inhibiting BCR-ABL kinase activity [5–7]. Targeted therapy with TKIs induces a complete hematologic and cytogenetic response in more than 90% of chronic phase CML patients [8]. However, TKIs cannot efficiently eradicate leukemia stem cells (LSCs) of CML due to the insensitivity of LSCs to TKIs [9, 10]. Therefore, developing new strategies to target LSC is necessary and critical for curing CML, and the success of this approach requires a full understanding of the biology of LSCs.

Although increasing evidence demonstrated that CML stem cells utilize signaling pathways that are independent of BCR-ABL kinase activity for their maintenance and survival [11, 12], the underlying molecular mechanisms remain unclear. To address these challenging issues, we performed a DNA microarray assay to compare the gene expression profiles between LSCs and normal HSCs. Here, we present the entire microarray dataset in open access format to demonstrate the usefulness of the microarray data in studying LSCs in CML.

2. Methodology

Our previous study has shown that BCR-ABL-expressing Lin⁻Sca-1⁺c-Kit⁺ (LSK) cells function as LSCs [10]. In order to isolate LSCs, eight- to twelve-week-old C57BL/6J mice were used for the induction of CML by BCR-ABL transduction of bone marrow cells in our bone marrow transduction/transplantation mouse model [13–15]. As the parallel control experiment, non-BCR-ABL-expressing (transduced with the empty GFP vector) bone marrow cells were also transplanted into recipient mice, and GFP+LSK cells derived

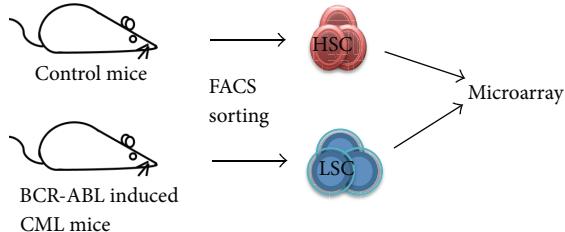


FIGURE 1: Scheme of microarray assay.

from these mice were used as normal hematopoietic stem cells. At day 13 after bone marrow transplantation, 30 CML mice were treated with imatinib and another 30 CML mice were treated with placebo. Similarly, 30 control mice were treated with imatinib and another 30 control mice were treated with a placebo (Table 1). The drug was given orally in a volume of 0.25 mL by gavage (100 mg/kg; total 6 doses; 1 dose per 4 hours). Water was used as placebo. These mice were sacrificed at day 14 after bone marrow transplantation to collect bone marrow cells for the isolation of total RNA. Red blood cells (RBCs) were depleted using RBC lysis buffer (containing NH_4Cl , KHCO_3 , and EDTA). To stain the cells with antibodies, bone marrow cells were suspended in staining medium (Hank's Balanced Salt Solution (HBSS) with 2% heat-inactivated calf serum) and incubated with biotin-labeled lineage antibody cocktail containing a mixture of antibodies against CD3, CD4, CD8, B220, Gr-1, Mac-1, and Ter119 at 4°C for 15 min for identifying lineage-positive cells for elimination. After washing, the fluorochrome-labeled secondary antibody (APC-Cy7-conjugated streptavidin) for recognizing biotin and PE-conjugated c-Kit and APC-conjugated Sca-1 antibodies were added to the cells for 15 min at 4°C in the dark. All these antibodies were purchased from eBioscience, USA. After washing, Fluorescence-Activated Cell Sorter (FACS) was run for sorting LSCs and normal HSCs. About 20 000 cells were directly sorted into RLT buffer (RNeasy Micro Kit, Qiagen, CA, USA) (Figure 1).

The total RNA was extracted using the RNeasy Micro Kit (Qiagen, CA, USA), and the quality of RNA was assessed using a 2100 Bioanalyzer instrument and RNA 6000 Pico LabChip assay (Agilent Technologies, Palo Alto, CA, USA). The ribosomal RNA ratios (28S/18S) are around 1.4, the RNA integrity numbers (RINs) of all samples are above 8, and all the RNA samples have an OD 260/280 ratio of 1.8 to 2. Utilizing the GeneChip Whole Transcript Sense Target Labeling Assay kit (Affymetrix, Santa Clara, CA, USA), 20 ng of total RNA undergoes reverse transcription with random hexamers tagged with T7 sequence. The double-stranded cDNA generated was then amplified by T7 RNA polymerase to produce cRNA. Second cycle first strand cDNA synthesis was performed by incorporating dUTP which was later used as sites where fragmentation was created by utilizing a uracil DNA glycosylase and apurinic/apirimidinic endonuclease 1 enzyme mix. The fragmented cDNA was then labeled by terminal transferase attaching a biotin molecule using Affymetrix proprietary DNA labeling reagent. Approximately 2.0 μg of fragmented and biotin-labeled cDNA was

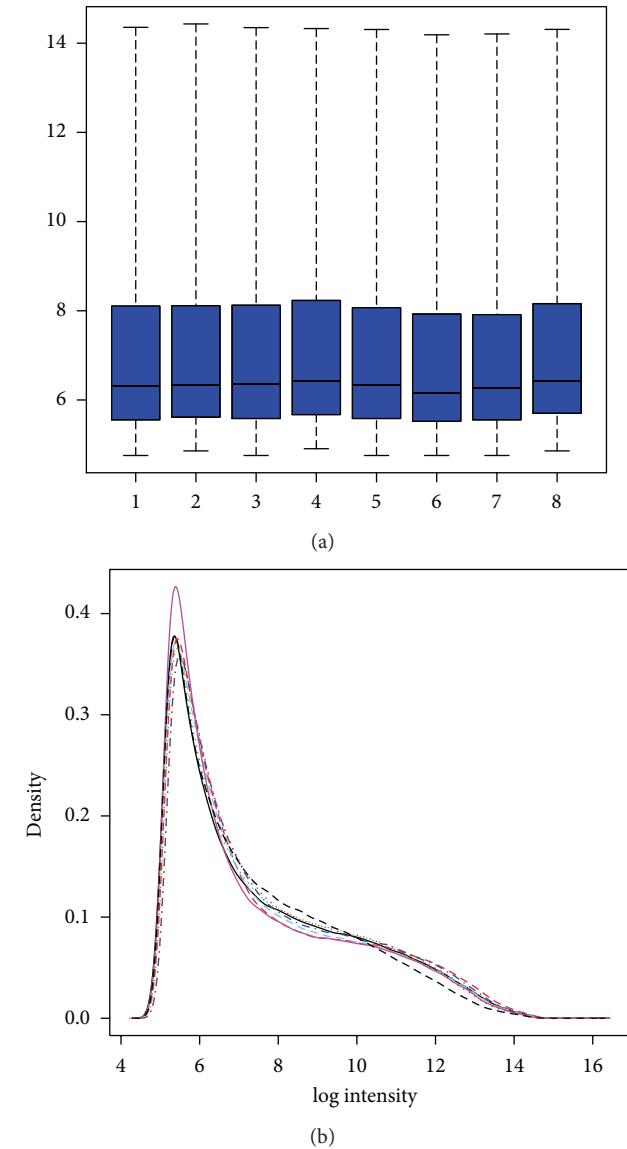


FIGURE 2: Quality control result of microarray data. (a) Boxplot of data intensities. (b) Density histogram of data intensities.

hybridized onto a Mouse Gene ST 1.0 Array (Affymetrix, Santa Clara, CA, USA) for 16 hours at 45°C . Posthybridization staining and washing were performed according to manufacturer's protocols using the Fluidics Station 450 instrument (Affymetrix). Finally, the arrays were scanned with a GeneChip Scanner 3000. Images were acquired, and cel files were generated and then used for analysis.

Microarray data analysis was as follows. The average signal intensities for each probeset within arrays were calculated by the *rma* function provided within the *affy* package for *R*. The quality control of the microarray data was shown in Figure 2. The RMA method incorporates convolution background correction and summarization based on multiarray model that fits robustly using the median polish algorithm. Data were quantile-normalized to bring arrays onto a common scale. In order to systematically identify

TABLE 1: Summary of sample replicates in this microarray.

Sample	Replicate number	Total used mice	Column of normalized signal intensity (1 replicate per column)
HSC treated with placebo	2	Total 60 mice; 30 mice per replicate	2, 6
LSC treated with placebo	2	Total 60 mice; 30 mice per replicate	3, 7
HSC treated with Imatinib	2	Total 60 mice; 30 mice per replicate	4, 8
LSC treated with Imatinib	2	Total 60 mice; 30 mice per replicate	5, 9

the critical genes that are specific for LSCs, we used the Shaoguang/Dongguang Li (SDLI) optimization method, which is developed by us [16]. Exploratory hierarchical clustering analysis was conducted for classifying samples across time points. Array preprocessing and fold-change analysis was conducted for each batch separately. The data were sorted to show the pattern of the fold changes for all genes. Expression of all genes was analyzed simultaneously. To find the gene subset that corresponds to LSC function at the minimum/maximum extent, an Orthogonal Array (OA) sampling procedure was performed to construct a multi-subset class predictor with a pyramidal hierarchy [16]. The detailed mathematical information is described as follows.

To sort the data based on chromosome groups, the average change was calculated: average change (j) = $(\sum e(i, j))/n(j)$, where $e(i, j)$ are the expression fold changes and $n(j)$ are the numbers of genes in every chromosome group, $j = 1, 2, \dots, 20$, $i = 1, 2, \dots, n(j)$.

To enhance the signal-to-noise ratio, the dynamics of all the gene expression levels were calculated: average absolute change (j) = $(\sum |e(i, j)|)/n(j)$, where $e(i, j)$ are the expression fold changes and $n(j)$ are the numbers of genes in every chromosome group, $j = 1, 2, \dots, 20$, $i = 1, 2, \dots, n(j)$. Next, the absolute difference between two expression groups was calculated: absolute difference (j) = $(\sum |e_{p190}(i, j) - e_{p210}(i, j)|)/n(j)$, where $e(i, j)$ are the expression fold changes and $n(j)$ are the numbers of genes in every chromosome group, $j = 1, 2, \dots, 20$, $i = 1, 2, \dots, n(j)$.

To find the gene subset that corresponds to the minimum (or maximum) of the objective function to give optimum solutions, an Orthogonal Array (OA) sampling procedure was combined with some search space reduction strategies for constructing a multi-subset class predictor with a pyramidal hierarchy. To discover the optimum solution, we consider a multidimensional continuous function $f(x)$ with multiple global minima and local minima on subset G of R^n .

- (i) Definition of local minima: for a given point $x^* \in G$, if there exists a δ -Neighborhood of x^* , $O(x^*, \delta)$, such that $x \in O(x^*, \delta)$ and $f(x^*) \leq f(x)$, then x^* is called a local minimal point of $f(x)$.
- (ii) Definition of global minima: if for every $x \in G$ the inequality is correct, then x^* is called a global minimum of $f(x)$ on G and the global minima of $f(x)$ on G form a global minimum set.

(iii) Finding of the global minima: for a given constant C_0 such that the level set $H_0 = \{x \mid f(x) < C_0, x \in G\}$ is nonempty, if $\mu(H_0) = 0$, where μ is the Lebesgue measure of H_0 , then C_0 is the minimum of $f(x)$ and H_0 is the global minimum set. Otherwise, assume that $\mu(H_0) > 0$ and C_1 is the mean value of $f(x)$ on H_0 . Then $C_1 = 1/\mu(H_0) \int f(x) d\mu$ and $C_0 \geq C_1 \geq f(x^*)$, and then the level set H_k and mean value C_{k+1} of $f(x)$ on H_k were gradually constructed as follows:

$$H_k = \{x \mid f(x) < C_k, x \in G\},$$

$$C_{k+1} = \frac{1}{\mu(H_k) \int_{H_k} f(x) d\mu}. \quad (1)$$

With the assistance of OA's sampling, a decreasing sequence of mean values $\{C_k\}$ and a sequence of level sets $\{H_k\}$ were obtained.

Let

$$\lim_{k \rightarrow \infty} C_k = C^*,$$

$$\lim_{k \rightarrow \infty} H_k = H^*, \quad (2)$$

where C^* is the minimum of $f(x)$ on G and H^* is the global minimum set.

This analytical method allows us to narrow down candidate genes to a list of 10–20 genes for further functional tests.

3. Dataset Description

The dataset associated with this Dataset Paper consists of 3 items which are described as follows.

Dataset Item 1 (Table). Probe set ID and signal intensities of 8 samples. RMA function was used to normalize and calculate the signal intensities. The column HSC 1 presents the normalized signal intensity of HSC sample 1; LSC 1, the normalized signal intensity of LSC sample 1; HSC with Imatinib 1, the normalized signal intensity of HSC with imatinib treatment sample 1; LSC with Imatinib 1, the normalized signal intensity of LSC with imatinib treatment sample 1; HSC 2, the normalized signal intensity of HSC sample 2; LSC 2, the normalized signal intensity of LSC sample 2;

HSC with Imatinib 2, the normalized signal intensity of HSC with imatinib treatment sample 2; LSC with Imatinib 2, the normalized signal intensity of LSC with imatinib treatment sample 2. A brief summary is also shown in Table 1.

- Column 1:* Probe Set ID
- Column 2:* HSC 1
- Column 3:* LSC 1
- Column 4:* HSC with Imatinib 1
- Column 5:* LSC with Imatinib 1
- Column 6:* HSC 2
- Column 7:* LSC 2
- Column 8:* HSC with Imatinib 2
- Column 9:* LSC with Imatinib 2

Dataset Item 2 (Table). An updated annotation file with gene designations and probe set ID numbers based on tables available from the Affymetrix website. This will aid the reader in finding a gene of interest based on gene symbols and gene products through the probe ID number. The column Species Scientific Name presents the genus and species of the organism represented by the probe set; Sequence Source, the database from which the sequence used to design this probe set was taken; Representative Public ID, the accession number of a representative sequence; Gene Symbol, a gene symbol when one is available (from UniGene). Note that for consensus-based probe sets, the representative sequence is only one of several sequences (sequence subclusters) used to build the consensus sequence and it is not directly used to derive the probe sequences. The representative sequence is chosen during array design as a sequence that is best associated with the transcribed region being interrogated by the probe set.

- Column 1:* Probe Set ID
- Column 2:* GenBank Accession Number
- Column 3:* Species Scientific Name
- Column 4:* Sequence Type
- Column 5:* Sequence Source
- Column 6:* Target Description
- Column 7:* Representative Public ID
- Column 8:* Gene Title
- Column 9:* Gene Symbol

Dataset Item 3 (Table). We compared gene expression between LSCs and HSCs. This item shows the list of genes which differently express themselves in LSCs when compared to HSCs. The cutoff for relative fold change is set up to 2. The column Relative Fold Change 1 presents the relative fold change comparing LSC sample 1 with HSC sample 1, and Relative Fold Change 2 presents the relative fold change comparing LSC sample 2 with HSC sample 2.

- Column 1:* Probe Set ID
- Column 2:* Relative Fold Change 1

TABLE 2: Examples of potential key genes in leukemia stem cells.

Gene symbol	Gene title
Alox5	Arachidonate 5-lipoxygenase
Tpsb2	Tryptase beta 2
Blk	B lymphoid kinase
Scd1	Stearoyl-Coenzyme desaturase 1
Aplnr	Apelin receptor
Nov	Nephroblastoma overexpressed gene

Column 3: Relative Fold Change 2

Column 4: Gene Symbol

Column 5: Gene Title

4. Concluding Remarks

We present a detailed description about microarray dataset that might make this dataset very useful for any investigator to compare the gene expression profile of normal HSCs and LSCs with or without imatinib treatment. It will also make easy to identify critical regulators of LSCs of CML. By comparing the gene expression profiles of LSCs with those of HSCs, we identified several genes that might be critical for the maintenance and self-renewal of LSCs. A brief gene list was shown in Table 2. Among these genes, the roles of several genes, including *Alox5* [17], *Blk* [12], and *Scd1* [18], have been functionally validated using serial of genetic approaches. Ongoing and further studies from our laboratory and other groups on the biology of LSCs in CML will be facilitated with this dataset in an open access format.

Dataset Availability

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at <http://dx.doi.org/10.1155/2013/520285>.

Conflict of Interests

The authors declare that there is no conflict of interests.

Acknowledgments

This work was supported by grants from the Leukemia & Lymphoma Society and the National Institutes of Health (R01-CA122142, R01-CA114199) to Shaoguang Li.

References

- [1] S. Wong and O. N. Witte, “The BCR-ABL story: bench to bedside and back,” *Annual Review of Immunology*, vol. 22, pp. 247–306, 2004.
- [2] J. V. Melo and D. J. Barnes, “Chronic myeloid leukaemia as a model of disease evolution in human cancer,” *Nature Reviews Cancer*, vol. 7, no. 6, pp. 441–453, 2007.

- [3] M. Savona and M. Talpaz, "Getting to the stem of chronic myeloid leukaemia," *Nature Reviews Cancer*, vol. 8, no. 5, pp. 341–350, 2008.
- [4] R. Ren, "Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia," *Nature Reviews Cancer*, vol. 5, no. 3, pp. 172–183, 2005.
- [5] B. J. Druker, C. L. Sawyers, H. Kantarjian et al., "Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome," *New England Journal of Medicine*, vol. 344, pp. 1038–1042, 2001.
- [6] B. J. Druker, M. Talpaz, D. J. Resta et al., "Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia," *New England Journal of Medicine*, vol. 344, no. 14, pp. 1031–1037, 2001.
- [7] B. J. Druker, S. Tamura, E. Buchdunger et al., "Effects of a selective inhibitor of the Ab1 tyrosine kinase on the growth of Bcr-Ab1 positive cells," *Nature Medicine*, vol. 2, no. 5, pp. 561–566, 1996.
- [8] B. J. Druker, F. Guilhot, S. G. O'Brien et al., "Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia," *New England Journal of Medicine*, vol. 355, no. 23, pp. 2408–2417, 2006.
- [9] S. M. Graham, H. G. Jørgensen, E. Allan et al., "Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro," *Blood*, vol. 99, no. 1, pp. 319–325, 2002.
- [10] Y. Hu, S. Swerdlow, T. M. Duffy, R. Weinmann, F. Y. Lee, and S. Li, "Targeting multiple kinase pathways in leukemic progenitors and stem cells is essential for improved treatment of Ph+ leukemia in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 45, pp. 16870–16875, 2006.
- [11] A. S. Corbin, A. Agarwal, M. Loriaux, J. Cortes, M. W. Deininger, and B. J. Druker, "Human chronic myeloid leukemia stem cells are insensitive to imatinib despite inhibition of BCR-ABL activity," *Journal of Clinical Investigation*, vol. 121, no. 1, pp. 396–409, 2011.
- [12] H. Zhang, C. Peng, Y. Hu et al., "The Blk pathway functions as a tumor suppressor in chronic myeloid leukemia stem cells," *Nature Genetics*, vol. 44, pp. 861–871, 2012.
- [13] Y. Hu, Y. Liu, S. Pelletier et al., "Requirement of Src kinases Lyn, Hck and Fgr for BCR-ABL1-induced B-lymphoblastic leukemia but not chronic myeloid leukemia," *Nature Genetics*, vol. 36, no. 5, pp. 453–461, 2004.
- [14] C. Peng, J. Brain, Y. Hu et al., "Inhibition of heat shock protein 90 prolongs survival of mice with BCR-ABL-T315I-induced leukemia and suppresses leukemic stem cells," *Blood*, vol. 110, no. 2, pp. 678–685, 2007.
- [15] S. Li, R. L. Ilaria Jr., R. P. Million, G. Q. Daley, and R. A. Van Etten, "The P190, P210, and P230 forms of the BCR/ABL oncogene induce a similar chronic myeloid leukemia-like syndrome in mice but have different lymphoid leukemogenic activity," *Journal of Experimental Medicine*, vol. 189, no. 9, pp. 1399–1412, 1999.
- [16] S. Li and D. Li, *DNA Microarray Technology and Data Analysis in Cancer Research*, World Scientific, 2009.
- [17] Y. Chen, Y. Hu, H. Zhang, C. Peng, and S. Li, "Loss of the Alox5 gene impairs leukemia stem cells and prevents chronic myeloid leukemia," *Nature Genetics*, vol. 41, no. 7, pp. 783–792, 2009.
- [18] H. Zhang, H. Li, N. Ho, D. Li, and S. Li, "Scd1 plays a tumor-suppressive role in survival of leukemia stem cells and the development of chronic myeloid leukemia," *Molecular and Cellular Biology*, vol. 32, pp. 1776–1787, 2012.

