In chronic myeloid leukemia white cells from cytogenetic responders and non-responders to imatinib have very similar gene expression signatures

Background and Objectives. Imatinib induces complete cytogenetic responses (CCR) in the majority of patients with chronic myeloid leukemia (CML) in chronic phase (CP). However, a subgroup of patients is refractory at the cytogenetic level. Clinically, it would be advantageous to identify such patients a priori, since they may benefit from more aggressive therapy.

Design and Methods. To elucidate mechanisms underlying cytogenetic refractoriness, we used Affymetrix oligonucleotide arrays to determine the transcriptional signature associated with cytogenetic refractoriness in unselected white blood or bone marrow cells from 29 patients with CML in first CP prior to treatment with imatinib. Patients with CCR within 9 months were defined as responders (n = 16) and patients lacking a major cytogenetic response (>35% Philadelphia-positive metaphases) after 1 year were defined as non-responders (n = 13).

Results. Differences in gene expression between responders and non-responders were subtle. Stringent statistical analysis with multiple comparison adjustments revealed very few differentially expressed genes. Differentially expressed genes could not be confirmed in an independent test set.

Interpretation and Conclusions. We conclude that transcriptional profiling of unselected white cells is of limited value for identifying genes consistently associated with cytogenetic refractoriness to imatinib.

Key words: CML, imatinib, gene expression profiling, cytogenetic refractoriness.

Haematologica 2005; 90:459-464
©2005 Ferrata Storti Foundation
All patients studied achieved a complete hematologic response to imatinib. Patients were defined as responders (R) to imatinib if they achieved CCR within nine months or less (n=16), while non-responders (NR) were defined as those who had failed to achieve a major cytogenetic response (MCR) within one year of treatment (n=12).

RNA extraction and microarray analysis

RNA samples for the microarray analysis were isolated from total bone marrow (BM) white cells, or peripheral blood (PB) white cells, using guanidine-thiocyanate/phenol-chloroform extraction with RNA purification by RNeasy (Qiagen, Hilden, Germany) (Leipzig cohort) or cesium chloride-gradient purification (Mannheim cohort). RNA was processed by the DNA Core of the IZKF (Interdisciplinary Center for Clinical Research) at the University of Leipzig (Leipzig cohort) or by the Affymetrix Microarray Core at Oregon Health & Science University (Mannheim cohort). In the latter core, each sample of RNA underwent quality assessment on the RNA 6000 LabChip using the 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA). Those samples whose electropherogram showed the presence of discrete 18S and 28S ribosomal RNA peaks and the absence of irregularly-sized low molecular weight RNA species (i.e., degraded RNA) were considered to be of good quality and were used for microarray analysis.

Microarray assays were performed according to the standard protocol described in the Affymetrix GeneChip Expression Analysis Technical Manual, rev. 5 (http://www.affymetrix.com/support/index.affx).

Messenger RNA was amplified and labeled from 5 μg (Leipzig) or 4 μg (Mannheim) of total RNA. An aliquot (200 ng) of the resulting cRNA underwent quality assessment on the RNA 6000 LabChip using the 2100 Bioanalyzer (Agilent). Samples that produced a sufficient cRNA yield and had cRNA electropherograms that showed a size distribution pattern predictive of acceptable microarray assay performance were considered to be of good quality. These samples were fragmented and combined with array hybridization controls consisting of biotinylated cRNAs for four bacterial genes (Affymetrix, Santa Clara, CA, USA) in hybridization buffer. Ten micrograms of cRNA target were then hybridized with the GeneChip HG_U95Av2 array (Affymetrix) and scanned using the GeneArray laser scanner (Affymetrix). The HG_U95Av2 array contains 12,625 transcripts, including 67 control genes. The array image scans were processed with Affymetrix Microarray Suite software, version 5.0 (MAS 5.0). All GeneChip expression arrays contain control probe sets for both spiked and endogenous RNA transcripts (e.g., bacterial genes BioB, BioC, BioD, CreX and species-specific actin and GAPDH). Following image processing and single array analysis of the array pattern with the MAS statistical expression analysis algorithm, six values are examined: background, noise, average signal, percent present and ratio of signal values for probe sets representing the 5’ and 3’ ends of actin and GAPDH transcripts. All assays in this study met standard performance thresholds for background (less than 90) and RNA integrity (housekeeping control gene 3’/5’ ratios of less than 2).

Data analysis

MAS 5.0 was initially used for image analysis, signal quantification and intra-chip normalization (i.e., global scaling). A hierarchical clustering of samples based on MAS 5.0 signal data showed a noticeable difference between the two sites. In order to reduce the site difference, robust multichip analysis (RMA) was implemented in the Bioconductor (http://www.bioconductor.org) was used as an alternative method of signal quantification, pre-processing, and intra- and inter-chip normalization. Significance analysis of microarrays (SAM) was performed on the training set to identify
tify differentially regulated genes associated with a lack of response to imatinib. Genes that met a q value of ≤10% were selected for further validation in the test set. The unadjusted t-test p value and fold change of these genes were computed using the test set. As the test set failed to replicate the findings from the training set, we performed analysis of variance (ANOVA) on all patients (n=28) in the training and test sets in order to identify sources of variation. For each gene, the ANOVA model included the site (Leipzig vs. Mannheim), source of samples (PB vs. BM), the response status (R vs. NR) and two-way and three-way interaction terms. The p values from the ANOVA F statistics for each effect were adjusted for multiple comparisons using the false discovery rate (FDR) of Benjamini and Hochberg.\(^\text{11}\) Data visualization was performed using a hierarchical clustering analysis and multidimensional scaling (MDS). Gene annotation was obtained through the Affymetrix NetAffx site ([http://www.affymetrix.com/analysis/index.affx](http://www.affymetrix.com/analysis/index.affx)).

**Results**

**Gene expression profiling of unselected white cells**

Oligonucleotide microarrays were used to explore the gene expression profiles of unselected white blood or bone marrow cells from 29 patients with CML in CP1. Unsupervised hierarchical clustering of samples, based on log 2 transformed MAS 5.0 signal values, revealed clustering of patients based on the site of sample processing and sample source but not on cytogenetic response (Figure 1). The site effect was substantially attenuated after RMA normalization (Figure 1). With both methods, one patient's sample (67) was found to be an outlier. No reason was identified by review of clinical parameters, image file or sample quality parameters; therefore, this sample was retained in further analysis.

**Significance analysis of microarrays**

Since unsupervised hierarchical clustering failed to separate the patients according to cytogenetic response, it was reasoned that the effects of sample source and processing might have prevented the identification of genes associated with response to imatinib. Employment of SAM revealed 15 genes differentially expressed between R and NR in the training set (Table 2); however, differential expression was not confirmed in the test set (Figure 1).

**Analysis of variance**

Since there was no overlap in the differentially expressed genes between the two patient cohorts, analysis of variance was performed in order to identify sources of variability. All samples were combined, and we performed gene-by-gene ANOVA to evaluate the impact of the site and sample source, in addition to the response status. Using a false discovery rate of 0.10 or less, ANOVA identified genes differentially expressed between the two sites (4421 genes) and between BM and PB samples (439 genes). None of the genes met the statistical significance criteria for response status or for two or three-way interactions between the sites, source and response status.

**Sample size and power considerations**

Since the results indicated that the differences between R and NR were subtle, the observed differences were used to estimate the theoretical number of samples required to identify genes significantly associated with cytogenetic refractoriness. Using gene-specific mean squared errors (MSE) from the ANOVA model, the average sample size required to detect a 1.5-fold change between R and NR, using 80% power and a 1% significance level to adjust for multiple comparisons, was 258 patients (129 responders and 129 non-responders). An even larger sample size would be required to develop a robust test set for validation of a classification algorithm.

**Discussion**

Predicting cytogenetic response prior to imatinib would be clinically useful, since it would allow for early risk stratification. The patients for our study were selected from patient populations treated at two different centers, based on cytogenetic response and availability of sufficient RNA. Most of the non-responders did not achieve even a minor cytogenetic response, whilst 46.7% of responders with available data were in CCR at 3 months. Our patient population therefore represented, for most part, the extremes of the spectrum of responses. Nonetheless, we found that the pre-therapeutic gene expression profiles of unselected white cells were of limited value in identifying genes consistently associated with lack of cytogenetic response to imatinib. Our results fail to confirm two recently published studies. Using cDNA arrays on mononuclear cells from 22 CML patients (18 in CP), Kaneta et al.\(^\text{16}\) reported 71 genes associated with MCR. McLean et al.\(^\text{17}\) analyzed the transcriptional profile of total PB from 66 newly diagnosed patients in CP. From a list of 55 genes with at least a 1.7-fold difference between patients with CCR and patients with >65% Ph-positive metaphases after 13 months of treatment, they selected an optimal list of 31 predictive genes using a leave one out strategy. Remarkably, the discriminating genes did not overlap between the two studies, or with the 15 most significant genes in our training set. Using RMA.

haematologica/the hematology journal | 2005; 90(4) | 461 |
and the published genes (with the exception of eight genes reported by Kaneta et al. for which the probe ID were missing in the published manuscript), we could not distinguish between R and NR in our data set.

There are several potential explanations for the discrepant results. Firstly, different cells were analyzed. RNA extracted from whole blood, as in the study by McLean et al., contains large amounts of globin transcripts, and yields significantly lower present calls than RNA extracted from selected white or mononuclear cells. Secondly, our study included patients in late chronic phase pre-treated with various drugs and the cytogenetic response in such patients may be influenced by factors different from those in newly
Table 2. The 15 most significantly differentially expressed genes identified for responders and non-responders in the training set (q value <0.10), together with the fold change and level of significance of these genes when examined in the test set.

<table>
<thead>
<tr>
<th>Probe Set ID</th>
<th>Gene Title</th>
<th>Public Database Number</th>
<th>Fold Change</th>
<th>Fold Change</th>
<th>Unadjusted T Test</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>40215_at</td>
<td>UDP-glucose ceramide glucosyltransferase</td>
<td>7357</td>
<td>+2.1</td>
<td>-1.4</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>37985_at</td>
<td>lamin B1</td>
<td>4001</td>
<td>+1.9</td>
<td>-1.3</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>38402_at</td>
<td>lysosomal-associated membrane protein 2</td>
<td>3920</td>
<td>+1.8</td>
<td>-1.1</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>2065_s_at</td>
<td>BCL2-associated X protein</td>
<td>581</td>
<td>+1.4</td>
<td>1.0</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>39064_at</td>
<td>5,10- methenyltetrahydrofolate synthetase</td>
<td>10588</td>
<td>+1.4</td>
<td>+1.1</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>1360_at</td>
<td>X-ray repair complementing defective repair in Chinese hamster cells 4</td>
<td>7518</td>
<td>+1.3</td>
<td>1.0</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>138_at</td>
<td>mitogen-activated protein kinase kinase kinase 1</td>
<td>11184</td>
<td>-1.3</td>
<td>+1.1</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>33134_at</td>
<td>adenylate cyclase 3</td>
<td>109</td>
<td>-1.3</td>
<td>-1.1</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>39709_at</td>
<td>selenoprotein W, 1</td>
<td>6415</td>
<td>-1.4</td>
<td>1.0</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>36811_at</td>
<td>lysyl oxidase-like 1</td>
<td>4016</td>
<td>-1.6</td>
<td>-1.2</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>36757_at</td>
<td>histone 1, H3h</td>
<td>8357</td>
<td>-1.6</td>
<td>-1.1</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>41337_at</td>
<td>amino-terminal enhancer of split</td>
<td>166</td>
<td>-1.7</td>
<td>1.0</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>41743_i_at</td>
<td>optineurin</td>
<td>10133</td>
<td>-1.8</td>
<td>+1.2</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>39081_at</td>
<td>no gene title given</td>
<td>4502</td>
<td>-1.9</td>
<td>+1.1</td>
<td>&gt;0.05</td>
<td></td>
</tr>
<tr>
<td>36780_at</td>
<td>clusterin</td>
<td>1191</td>
<td>-2.6</td>
<td>-1.9</td>
<td>&gt;0.05</td>
<td></td>
</tr>
</tbody>
</table>

The fold change (FC) is defined as: ±2^Δ|Δ|, where |Δ| is the absolute value of the difference in mean log (signal) between R and NR groups. The positive sign (+) indicates up-regulation of R compared to NR, while the negative sign (-) indicates down-regulation. Note that a fold change of 1.0 is equivalent to no difference in gene expression.

Diagnosed patients. In addition, since we do not have data on patients’ serum imatinib concentrations, it may be that adequate drug levels were not achieved in some patients. Most importantly, we used an independent control group for validation, and applied rigorous statistical methods for data analysis.

Significant differences were evident between samples from Leipzig and Mannheim, likely reflecting different methods of sample preparation, and between samples derived from PB vs. BM, confirming previously published data. However, regardless of these potential confounding factors, the differences between R and NR were extremely subtle in both cohorts, and associated with very high false discovery rates. In addition, our power calculations indicate that although it would theoretically still be possible to identify a more robust set of predictive genes, this would require a very large cohort of patients.

In addition to these considerations, CML may differ from other hematologic malignancies, such as acute myeloid leukemia or acute lymphoblastic leukemia, in which gene expression signatures have been shown to correlate with response to chemotherapy and survival. In these disorders refractoriness to therapy may be a quality of all blast cells and thus its signature may be detectable in the population. In CML, this may be more comparable to hematologic than to cytogenetic refractoriness. The latter may be mediated by refractory progenitor cells, the transcriptional signature of which may be undetectable in the noise of unselected white cells. Thus, it is conceivable that enriching for progenitor cells (e.g. CD34+ cells) may sufficiently reduce the noise to uncover the transcriptional signature associated with cytogenetic refractoriness. We are currently testing this hypothesis by analyzing expression profiles of CML CD34+ progenitor cells, for correlation with cytogenetic response.