Deep sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia

Running Title: Deep sequencing method for MRD detection

Authors: Malek Faham,1* Jianbiao Zheng,1 Martin Moorhead,1 Victoria E. H. Carlton,1 Patricia Stow,2 Elaine Coustan-Smith,3 Ching-Hon Pui,2 Dario Campana3

Affiliations:
1Sequenta, Inc., South San Francisco, CA; 2Department of Oncology, St Jude Children’s Research Hospital, Memphis, TN; 3Department of Pediatrics, National University of Singapore, Singapore

*Address correspondence to:
Malek Faham, M.D., Ph.D.
400 East Jamie Court, Suite 301, South San Francisco, CA 94080
Phone: (650) 243-3900; Fax (650) 243-3901; Email: malek.faham@sequentainc.com

Primary scientific category: Lymphoid Neoplasia
ABSTRACT

Persistence of minimal residual disease (MRD) during therapy is the strongest adverse prognostic factor in acute lymphoblastic leukemia (ALL). We developed a high-throughput sequencing method which universally amplifies antigen-receptor gene segments. It identifies all clonal gene rearrangements (i.e., leukemia-specific sequences) at diagnosis, allowing monitoring of disease progression and clonal evolution during therapy. In spike-in experiments, the assay specifically detected 1 leukemic cell among greater than 1 million leukocytes. We compared it to gold-standard MRD assays, i.e., multiparameter flow cytometry and allele-specific oligonucleotide polymerase chain reaction (ASO-PCR), using diagnostic and follow-up samples from 106 patients with ALL. Sequencing detected MRD in all 28 samples positive by flow cytometry, and in 35 of the 36 positive by ASO-PCR; it also revealed MRD in 10 and 3 additional samples that were negative by flow cytometry and ASO-PCR, respectively. This new method allows monitoring of treatment response in ALL and other lymphoid malignancies with great sensitivity and precision. The clinicaltrials.gov identifier number for the Total XV study is NCT00137111.
INTRODUCTION

The clinical management of patients with acute lymphoblastic leukemia (ALL) relies on accurate prediction of relapse hazard to determine the intensity of therapy and avoid over- or under-treatment.\(^1\,^2\) Traditional prognostic factors include presenting clinical and biological features, such as age, blast count at diagnosis, immunophenotype and genetic abnormalities.\(^1\,^2\) Based on a large body of evidence involving thousands of patients, the measurement of residual leukemia levels (“minimal residual disease”; MRD) during therapy has now emerged as the most important predictor of outcome in ALL.\(^3\,^4\) As a result, risk-classifications based on MRD assessment are now a critical component of ALL clinical treatment protocols.

Current methodologies to monitor MRD in ALL include flow cytometric detection of aberrant immunophenotypes, which can detect 1 leukemic cell among 10,000 (0.01%) normal bone marrow or peripheral blood cells, and allele-specific oligonucleotide polymerase chain reaction (ASO-PCR) amplification of immunoglobulin and T-cell receptor genes, which has a sensitivity of 0.001%.\(^3\,^5\) Although these methods have proven to be reliable in a clinical setting, they have limitations. Flow cytometry requires a high level of expertise to interpret results proficiently. ASO-PCR requires the development of reagents and assay conditions for each individual patient, which is laborious and time-consuming. Moreover, these methods have limited or no capacity to monitor the evolution of different leukemic subclones during treatment, with the potential of false-negative results. Finally, patients who achieve MRD-negative status by standard criteria but have very low levels of persistent leukemia have a higher risk of relapse than those with no detectable MRD\(^6\), suggesting that improvements in sensitivity of MRD monitoring methods might improve precision in predicting relapse.
We developed a novel, sequencing-based method to identify cells with specific molecular signatures. The method employs consensus primers to universally amplify rearranged immunoglobulin and T-cell receptor gene segments in a sample and relies on high-throughput sequencing and specifically designed algorithms to identify clonal gene rearrangements in diagnostic samples and quantify these rearrangements in follow-up MRD samples. In this study, we assessed the suitability of this method to monitor MRD in ALL. We determined its sensitivity and specificity, delineated the extent of genetic diversity (including clonal evolution) present at diagnosis, and compared its capacity to measure MRD to that of flow cytometry and ASO-PCR in follow-up samples from more than 100 patients with ALL.
MATERIALS AND METHODS

Clinical samples

Bone marrow samples were collected at diagnosis and during treatment from 110 children with newly diagnosed B-lineage ALL, representing 22.1% of the 498 patients enrolled in the Total XV study at St. Jude Children’s Research Hospital. The sample selection for this study was based on the demonstrated presence of an immunoglobulin heavy chain locus (IGH, IgH) clonal rearrangement at diagnosis, ASO-PCR MRD data in a follow-up sample, and abundant surplus DNA available. This study was conducted in accordance with Declaration of Helsinki principles. The study was approved by the St. Jude Institutional Review Board, and written informed consent was obtained from the parents or guardians of each child, and assent from the patients as appropriate.

MRD measurements by ASO-PCR and flow cytometry

ASO-PCR MRD studies were performed as previously described, with MRD quantification performed by limiting-dilution (n = 34) or real time PCR (n = 76). Briefly, we made 10-fold serial dilutions of diagnostic DNA in pooled peripheral blood DNA extracted from mononuclear cells of 4-6 healthy donors, and analyzed each dilution in duplicate (in quadruplicate for the single copy dilution). For the limiting dilution method, we used either a 2-round (either semi-nested or with the same patient-specific and consensus primer in both rounds) or a 1-round PCR assay with an initial touchdown phase, and quantitated MRD using Poisson statistics; real-time PCR analysis was performed by a single PCR amplification with sequence-specific TaqMan hydrolysis probes on ABI PRISM Sequence Detection System 7700 (Applied Biosystems).
MRD studies by flow cytometry were performed using a combination of markers identified at diagnosis, as previously described.\textsuperscript{8}

\textit{MRD measurements by Sequenta LymphoSIGHT\textsuperscript{TM} method}

Genomic DNA, isolated using AllPrep DNA mini and/or micro kits (Qiagen), was amplified using locus-specific primer sets for \textit{IgH}\textsuperscript{\textregistered} complete (V\textsubscript{H}DJ\textsubscript{H}), \textit{IgH}\textsuperscript{\textregistered} incomplete (DJ\textsubscript{H}), and T-cell receptors (TCR) including, \textit{TRB}\textsuperscript{\textregistered}, \textit{TRD}\textsuperscript{\textregistered}, and \textit{TRG}\textsuperscript{\textregistered}, designed to allow for the amplification of all known alleles of the germline \textit{IgH} and TCR sequences (see Supplementary Methods for a description of the primer design, and amplification and sequencing reactions).\textsuperscript{9} A clonotype was defined when at least two identical sequencing reads were obtained (see Supplementary Methods).

The frequency of each clonotype in a sample was determined by calculating the number of sequencing reads for each clonotype divided by the total number of passed sequencing reads in the sample. To define leukemic gene rearrangements in samples obtained at diagnosis, we used a frequency threshold of 5\%, i.e., any clonotype present at a frequency of >5\% was regarded as originating from the leukemic clone. In preliminary studies, the frequency of individual clonotypes among normal B-cell populations was consistently below this threshold.

We used the following criteria to identify any clonotypes present in the same diagnostic sample that might have evolved from the leukemic clone through \textit{V}\textsubscript{H} replacement: (1) identical J and D segments, (2) identical J segment deletion size, (3) identical D segment deletion size (the side next to the J segment), (4) random N base insertions between the J and D segments, and (5) different V segments.
The leukemia-derived sequences identified at diagnosis were used as a target to assess the presence of MRD in follow-up samples. For MRD quantitation, we generated multiple sequencing reads for each rearranged B-cell in the reaction. For example, in cells containing an IgH rearrangement (i.e., B-cells), the MRD assay was designed to achieve approximately 10X coverage per B-cell.

To determine the absolute measure of the total leukemia-derived molecules present in the follow-up sample, we added a known quantity of reference IgH sequence into the reaction and counted the associated sequencing reads. The known quantity of reference IgH sequence was derived from a pool of plasmids containing 3 unique IgH clonotypes, quantitated using standard RT-PCR methods. The resulting factor (number of molecules per sequence read) was then applied to the leukemia associated clonal rearrangement reads to obtain an absolute measure of the total leukemia-derived molecules in the reaction. A similar calculation can be performed to assess the total number of rearranged IgH molecules, or B-lineage cells, in the reaction. Finally, we calculated the total leukocytes in the reaction by measuring the total DNA in the reaction using standard picogreen methods and RT-PCR using genomic markers such as β actin DNA.

These metrics were combined to calculate a final MRD measurement, which is the number of leukemia-derived molecules divided by the total leukocytes in the sample. In cases where there are multiple index clones, we utilize the highest frequency index clone to calculate the percentage of MRD in follow-up samples. This quantitation scheme is illustrated schematically in Figure 1b.
Assessment of Sequenta LymphoSIGHT platform technical performance

Diagnostic samples from 12 ALL patients were used for technical assessment of the LymphoSIGHT™ platform. The selected samples had a >95% clonal rearrangement frequency among B-lineage cells. Additionally, the samples were almost exclusively comprised of leukemic B-lineage cells, based on equivalence between the total number of rearranged IgH molecules and leukocytes in each sample.

Serial dilutions of the 12 selected samples were prepared in duplicate in normal peripheral blood mononuclear cells, in a range between 1 in 1 million to 1 in 1,000 cells. Dilutions were then subjected to amplification and sequencing in two replicate experiments. Correction factors for sample purity and dilution preparations were applied, and RT-PCR based measurements of the sample concentration were utilized for increased accuracy. We compared expected and measured frequencies in log space. For correlation and slope, the lowest dilution was excluded as some values were 0.
RESULTS

Sensitivity, precision and quantitative capacity of the immune repertoire sequencing assay

Figure 1 provides an overview of the Sequenta LymphoSIGHT™ method. Using universal primer sets, we first amplify immunoglobulin heavy chain (IgH) and T cell receptor (TCR) sequences from genomic DNA in a two-stage PCR. The amplified product is sequenced to obtain a high number of reads (e.g., 10^6 reads) (Figure 1a). Typically the number of reads exceeds the number of starting molecules allowing every starting molecule to be sampled. The sequence reads are analyzed to determine similar sequences that form a clonotype. Following clonotype determination, a standard quantitation scheme is used to calculate MRD metrics (see Figure 1b, Methods), including clonotype frequency, number of leukemia-associated molecules in the reaction, number of total leukocytes in the reaction, and MRD level.

We assessed the technical performance of the method using diagnostic samples from 12 of the 110 ALL patients included in the study, carrying 13 leukemic IgH clonotypes. Serial dilutions of leukemic cells in normal peripheral blood mononucleated cells, ranging from <1 in 1 million to >1 in 1,000 cells, were prepared and analyzed in duplicate. The high precision of the assay was demonstrated by the low average relative standard deviations (range, 4.1% to 7.6%) at clonotype frequencies at or above 3x10^-5 (Supplementary Table 1). Thus, the assay is highly quantitative for frequencies above 10^-5. Random error increased at clonotype frequencies below 10^-5, as expected from Poisson statistics with low number of input molecules in the reaction (Supplementary Table 1). For each clonotype, the assay showed high r^2 values with a range of 0.977 and 0.996 (mean 0.988, median 0.991) between each of the expected and the measured clonotype frequencies (Figure 2, Supplementary Table 2). The slopes ranged from 0.878 to 1.14
(mean 1.00, median 0.977), illustrating the quantitative nature of the assay over at least 3 orders of magnitude.

The immune repertoire sequencing assay unequivocally detected leukemic signatures in all dilutions with expected concentration of at least one leukemic cell in 1 million leukocytes (a sensitivity of \(10^{-6}\)) (Figure 2). We also tested the sensitivity of the assay to detect levels of leukemic cells below 1 in 1 million leukocytes (Figure 2). We detected an average expected concentration of \(7.4 \times 10^{-7}\) and expected number of molecules of 5. We did not detect leukemia signals in two cases with low expected concentrations (\(4.2 \times 10^{-7}\) and \(6.7 \times 10^{-7}\) with 2 and 4 expected molecules, respectively), a result likely caused by insufficient leukemia-derived molecules in the reaction and consistent with Poisson sampling.

Analysis of all clonotypes together reflects multiple sources of error, including error associated with quantitation of the original leukemic DNA, pipetting error in the dilution preparations, error in measurement of the original sample purity (i.e., samples may not be comprised of 100% leukemic B-cells), and random and systematic error associated with the assay. Despite these potential sources of error, we measured an \(r^2\) value of 0.94 and a slope of 1.01 between the expected and measured clonotype frequencies. Thus, the cell dilution experiments clearly demonstrated the immune repertoire sequencing assay to be precise and quantitative with sensitivity levels at or below 1 leukemic cell per 1 million leukocytes.

**Detection of clonal rearrangements of multiple receptor genes in diagnostic samples**

We used the sequencing assay to detect rearrangements of immune cell receptor genes, i.e., \(IgH@\) complete (\(V_HDJ_H\)), \(IgH@\) incomplete (\(DJ_H\)), \(TRB@\), \(TRD@\) and \(TRG@\), in diagnostic
bone marrow samples from 100 of the 110 ALL patients known to have IgH rearrangements. An additional 10 samples were only assayed for IgH V_{H} DJ_{H} and were not included in this clonal gene rearrangement analysis. As expected, all 100 diagnostic samples demonstrated a high frequency clonal rearrangement for at least one receptor, herein referred to as a “calibrating receptor”. The vast majority (94) had at least 2 calibrating receptors, with 51 having 3 or more.

The IgH complete V_{H} DJ_{H} assay was the most frequent gene rearrangement: at least one IgH V_{H} DJ_{H} clonal rearrangement was detected in 96 of the diagnostic ALL samples (Table 1). TRD@ was the second most informative receptor, and at least one TRD@ clonal rearrangement was detected in 63 of the samples (Table 1). In contrast, TRB@ clonal rearrangements were only identified in 16 samples (Table 1). Concomitant clonal rearrangements of IgH (mature and/or immature) and at least 1 TCR gene were found in 86 of 100 samples collected at diagnosis.

We also assessed the cumulative number of clonal gene rearrangements that were present in each sample as more than one clonal rearrangement could be detected with one receptor. Among the 100 diagnostic samples studied, 98, 80, and 51 had 2 or more, 3 or more, and 4 or more clonal rearrangements across all receptors, respectively (Table 1).

One advantage of our method is that all rearranged genes are captured, enabling the comprehensive delineation of the landscape of IgH genetic diversity and clonal selection that occurred during the leukemogenic process. We analyzed 106 of the 110 diagnostic ALL samples with complete IgH V_{H} DJ_{H} gene rearrangements for evidence of IgH clonal evolution due to ongoing secondary recombination events during disease progression. The evolved clonotypes were identified by virtue of their sequence relatedness to the high frequency clonal rearrangement. The number of evolved clones per ALL sample varied widely from 0 to 6933.
clonotypes. Fifty-six of 106 samples (53%) demonstrated V_HDJ_H to V_HDJ_H evolution, which is consistent with a V_H replacement model. These samples could be categorized into five groups based on the number of evolved clones present, with 39 patients (37%) having 1-50 evolved clones and 17 patients (16%) having a high degree of evolution (>50 evolved clones) (Figure 3). These data show that the sequencing assay captured the extensive evolution that is present in diagnostic samples, a potential cause of false-negative results during MRD monitoring by ASO-PCR focusing on selected gene rearrangements.

Comparison of MRD results by immune repertoire sequencing with those by standard MRD assays

We assessed MRD using the sequencing assay in follow-up samples collected during therapy from 106 patients with newly diagnosed B-lineage ALL and IgH V_HDJ_H clonal gene rearrangements at diagnosis. In addition to the previously demonstrated presence of IgH gene rearrangements at diagnosis, the main criteria for inclusion in the study was the availability of abundant surplus DNA collected during follow-up. In the Total XV trial, MRD was measured by flow cytometry and ASO-PCR during and at the end of remission induction therapy (days 19, 26 and 46), during continuation therapy (week 7, 17 and 48) and at the end of therapy (week 120 for girls and 146 for boys). Of the 106 samples studied, 45 were collected during induction therapy (n = 18) or at the end of remission induction therapy (n = 27), 44 during continuation therapy and 17 at the end of therapy (Supplementary Table 3), thus providing a full representation of the different types of bone marrow cellularity found during ALL treatment.
We first analyzed concordance between MRD results obtained by our assay and flow cytometry, which was used to monitor MRD in 105 of the 106 cases. Importantly, the sequencing test was performed without knowledge of the results of flow cytometry testing. The two methods gave concordant MRD-positive or MRD-negative results in 95 of 105 (90%) samples (Figure 4a, Supplementary Table 3). In 10 of 105 samples (10%), MRD was positive by sequencing, but undetectable by flow cytometry (Figure 4a, Table 2); 7 of these 10 samples were also positive by ASO-PCR. Most of the discordance between the sequencing assay and flow cytometry can be explained by the sensitivity limitations of flow cytometry. For example, MRD levels ranged between 0.00004% and 0.011% by sequencing in 9 of the 10 discordant samples (Table 2). MRD was undetectable by flow cytometry and measured at 0.96% and 1.2% by sequencing and ASO-PCR, respectively, in the final discordant sample (Table 2).

We next compared the MRD results obtained by immune repertoire sequencing and ASO-PCR. Again, the sequencing test was performed without knowledge of the previously collected results by ASO-PCR. Results were concordant in 102 of 106 follow-up samples (96%): 35 (33%) were MRD-positive and 67 (63%) were MRD-negative by both methods (Figure 4b, Supplementary Table 3). In 3 of 106 samples, MRD was positive by sequencing (0.0014%, 0.0004%, and 0.00004%), but undetectable by ASO-PCR (Figure 4b, Table 2), a result which could be explained by the fact that these levels of MRD would be outside the quantitative range of ASO-PCR. In the remaining sample, a faint signal corresponding to 0.0020% MRD was detected by ASO-PCR (performed by a limiting-dilution assay) but MRD was undetectable by sequencing (Figure 4b, Table 2). The reason for this discrepancy is unclear. The DNA analyzed by the sequencing test in this sample corresponded to 1.07 million input cells. Thus, if the ASO-PCR estimate was correct, DNA from approximately 20 ALL cells
should have been present in the sample; this amount of clonal rearrangements should be easily detectable by the sequencing assay. However, MRD estimates by ASO-PCR can be inaccurate at such extremely low levels, and it is possible that the sample analyzed by sequencing did not contain any leukemic DNA at all. Alternatively, the ASO-PCR signal may have been an artifact. This discrepancy notwithstanding, these comparisons with “gold standard” MRD assays provide strong evidence of the reliability and sensitivity of the sequencing assay when applied to clinical samples collected from patients undergoing chemotherapy.

When present in the detection range, MRD levels measured by flow cytometry are typically accurate, as this method directly counts leukemic cells among normal cells. In the 29 samples with detectable MRD in all three platforms, MRD levels measured by the sequencing assay were highly correlated to those of flow cytometry. Specifically, the MRD level measured by the two platforms showed an \( r^2 \) of 0.83 and a slope of 1.03, supporting the quantitative accuracy of the sequencing assay. ASO-PCR versus flow cytometry MRD results showed an \( r^2 \) of 0.75 and a slope of 1.04.

In a subset of 9 samples that had previously been identified as having non-\( \text{IgH} \) recombinations, we tested the concordance between MRD results using the \( \text{IgH} \ V_H \text{DJ}_H \) sequencing assay and other receptor rearrangement sequencing assays (\( \text{IgH} \ D\text{J}_H \) and TCR assays). In all cases, results were concordant with the previously observed data. Specifically, 7 samples were MRD negative according to both the \( \text{IgH} \ V_H \text{DJ}_H \) and a second receptor, and 2 samples were MRD positive according to the \( \text{IgH} \ V_H \text{DJ}_H \) and a second receptor (Supplementary Table 4).
DISCUSSION
We developed a novel approach to monitoring response to treatment in patients with leukemia. The method employs consensus primers and high-throughput sequencing to universally amplify and sequence all rearranged immunoglobulin and TCR gene segments present in a leukemic clone. It allows the identification of all clonal gene rearrangements in diagnostic samples and to track their evolution during therapy. A previous study demonstrated the utility of high-throughput sequencing for MRD quantitation in chronic lymphocytic leukemia.\textsuperscript{18} Additionally, while this manuscript was in preparation, Wu \textit{et al.} reported the use of high-throughput sequencing to detect MRD in patients with T-lineage ALL; their results compared favorably with those of flow cytometry but no comparison was made with ASO-PCR.\textsuperscript{19} In our study, we showed that the sequencing assay is precise, quantitative, and sensitive when applied to bone marrow samples collected throughout treatment in patients with B-lineage ALL. Our method is quantitative at frequencies above \(10^{-5}\) and the lower limit of detection is below \(10^{-6}\). Moreover, the technical performance data presented here demonstrate that the sequencing assay sensitivity is limited only by the number of input cells and thus can detect residual disease at levels well below 1 in 1 million leukocytes (0.0001\%). This represents at least one to two orders of magnitude higher sensitivity than standard ASO-PCR and flow-cytometric methods, respectively.

Comparisons with “gold standard” MRD assays provide strong evidence of the reliability and sensitivity of the sequencing assay when applied to clinical samples collected from patients undergoing chemotherapy. Our assay offers potential advantages over ASO-PCR, the standard method to target \(IgH\) and TCR genes for MRD studies. First, it allows monitoring of all leukemic rearrangements, regardless of their prevalence at diagnosis. This feature should abrogate the risk
of false-negative MRD results due to clonal evolution during the course of the disease, whereby subclones representing a minority of antigen-receptor gene rearrangements might be neglected at diagnosis but become predominant at relapse.\textsuperscript{14,20-25} We observed that the majority of samples studied at diagnosis already had a high-degree of genetic diversity at the $IgH$ locus, in agreement with the results of another study focusing on the immune repertoire sequencing method to determine the prevalence of clonal evolution in diagnostic samples from patients with B-lineage ALL.\textsuperscript{26} Second, current ASO-PCR assays require the development of customized primer sets and assay conditions for detection of each individual clonal gene rearrangement in each patient, a laborious and time-consuming process which becomes increasingly so if multiple receptors and/or clonal rearrangements are assayed. Because the sequencing assay utilizes a set of universal primers, it can assess multiple clonal rearrangements in every patient without the need for individualized procedures. We currently estimate that the assay will have a turnaround time of approximately 7 days for MRD detection. A schematic diagram of the analysis and workflow when performed prospectively is shown in Figure 5. Finally, there is evidence suggesting that MRD positivity below 0.01\% has prognostic implications\textsuperscript{6} although it remains to be determined whether the enhanced sensitivity of the immune repertoire sequencing approach will provide additional clinical information in the context of current chemotherapy and allogeneic stem cell transplant protocols for ALL. Conceivably, a high sensitivity of MRD detection before transplant could be useful in improving the timing of transplant;\textsuperscript{27} post-transplant, it should allow for early intervention to avoid relapse.

Our approach relies on high-throughput sequencing to identify leukemia-associated gene rearrangements at diagnosis and measure their prevalence during treatment. Although the costs of the sequencing process were historically prohibitive for routine application, they have been
steadily decreasing, with the cost of sequencing falling by a factor of ~20,000 over the past decade. The current sequencing costs for rearranged antigen-receptor genes are comparable to those of flow cytometry and ASO-PCR, and are predicted to decrease substantially with further advances in sequencing technology. High-throughput sequencing data may be complex to interpret but, as shown by our results, the analytic algorithms that we designed allow unequivocal detection and accurate quantitation of leukemia-derived sequences. The prognostic power of MRD during treatment has been amply demonstrated in numerous correlative studies performed in patients with newly diagnosed childhood and adult ALL or relapsed ALL, and in patients undergoing hematopoietic stem cell transplant. Monitoring of MRD has become a central component of the modern managements of patients with ALL, and is beginning to be introduced as a criterion for testing novel anti-leukemic agents. The sensitivity, universal applicability, and capacity to capture clonal evolution of the method described here, together with the results of our comparison with standard MRD assays in clinical samples, strongly support its potential as a next-generation MRD test for ALL. Although we limited our study to the most common form of ALL, i.e. B-lineage, the same approach can also be applied to study T-lineage ALL or other lymphoid malignancies, such as chronic lymphocytic leukemia, non-Hodgkin lymphoma and multiple myeloma.
ACKNOWLEDGMENTS

This work was supported in part by grants CA60419 and CA21765 from the National Cancer Institute, by the American Lebanese Syrian Associated Charities (ALSAC), and by research funding from Sequenta, Inc. to M.F., J.Z., M.M. and V.C.

AUTHOR CONTRIBUTIONS

M.F., J.Z., M.M. and V.C. performed the sequencing assays and analyzed and interpreted data; P.S. and E.C.S. provided ASO-PCR and flow cytometry MRD data; C-H.P. led the clinical trial under which the samples were collected for MRD studies; M.F. and D.C. designed the research, analyzed and interpreted data, and wrote the manuscript with the input of all other authors. All authors approved the final version of the manuscript.

DISCLOSURE OF CONFLICTS OF INTEREST

M.F., J.Z., M.M. and V.C. are employees of and stockholders in Sequenta, Inc.
REFERENCES


TABLES

Table 1. Number of patients by number of calibrating clones in each receptor.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGH@-VDJ</td>
<td>4</td>
<td>51</td>
<td>35</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGH@-DJ</td>
<td>67</td>
<td>22</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRB@</td>
<td>84</td>
<td>13</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRD@</td>
<td>37</td>
<td>41</td>
<td>13</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRG@</td>
<td>47</td>
<td>29</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>29</td>
<td>21</td>
<td>16</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2. Summary of discordant MRD results for sequencing, flow cytometry and ASO-PCR methods.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Input Cells</th>
<th>Sequencing</th>
<th>Flow Cytometry</th>
<th>PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>248496</td>
<td>0.00004</td>
<td>UD</td>
<td>UD</td>
</tr>
<tr>
<td>12</td>
<td>133669</td>
<td>0.0027</td>
<td>UD</td>
<td>0.0017</td>
</tr>
<tr>
<td>25</td>
<td>142310</td>
<td>0.0106</td>
<td>UD</td>
<td>0.0005</td>
</tr>
<tr>
<td>30</td>
<td>120307</td>
<td>0.0032</td>
<td>UD</td>
<td>0.0076</td>
</tr>
<tr>
<td>31</td>
<td>1109253</td>
<td>0.0014</td>
<td>UD</td>
<td>UD</td>
</tr>
<tr>
<td>41</td>
<td>626380</td>
<td>0.0005</td>
<td>UD</td>
<td>0.0051</td>
</tr>
<tr>
<td>58</td>
<td>171326</td>
<td>0.0102</td>
<td>UD</td>
<td>0.0120</td>
</tr>
<tr>
<td>73</td>
<td>983289</td>
<td>0.0004</td>
<td>UD</td>
<td>UD</td>
</tr>
<tr>
<td>98</td>
<td>162217</td>
<td>0.0097</td>
<td>UD</td>
<td>0.0038</td>
</tr>
<tr>
<td>29</td>
<td>161142</td>
<td>0.9607</td>
<td>UD</td>
<td>1.2000</td>
</tr>
<tr>
<td>36</td>
<td>1070242</td>
<td>UD</td>
<td>UD</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

UD = Undetectable
FIGURE LEGENDS

Figure 1. Overview of LymphoSIGHT method. a) Schematic of PCR primer strategy and sequencing assay. b) Schematic of MRD quantitation scheme.

Figure 2. Technical performance of immune repertoire sequencing assay. Diagnostic samples from 12 ALL patients containing 13 leukemic IgH clonotypes were used for technical performance studies. Serial dilutions were prepared in duplicate from normal peripheral blood mononucleated cells (PBMCs), in a range between <1 in 1 million to >1 in 1,000 tumor cells per PBMC. The seven duplicated dilutions were then subjected to amplification and sequencing in two replicate experiments. Expected and observed frequencies were compared on a logarithmic scale (panels a-l). Replicate 1 is represented by circles, while replicate 2 is represented by crosses in all panels. Panel g shows the results from two cancer clones that were present in this tumor.

Figure 3. Clonal evolution mechanisms at the IgH gene locus. Diagnostic samples containing a clonal gene rearrangement, as determined using the VDJ (N=106) assay were categorized into five groups based on the number of evolved clones present in the diagnostic sample. The fraction of all samples having a given number of evolved clones is shown.
Figure 4. Comparison of MRD results obtained by sequencing, flow cytometry and ASO-PCR. MRD results obtained using sequencing method are compared to flow cytometry results for 105 ALL patients (a) and to ASO-PCR results for 106 ALL patients (b). The number of concordant measurements are shown in the lower left and upper right boxes and the number of discordant measurements are shown in the upper left and lower right boxes.

Figure 5. Schematic diagram of sequencing analysis and workflow for prospective sample collection.
Figure 1.

A. PCR AMPLICONS → SEQUENCING LIBRARY → SEQUENCE DATA

5' J PRIMER J D V TAG PRIMER TAG 5'

P5 115bp READ 95bp READ V PRIMER P7

B. QUANTITATE DNA

EXTRACT DNA

ADD KNOWN QUANTITY OF REFERENCE IgH

AMPLIFY IgH MOLECULES

SEQUENCE MOLECULES

\[ \text{D} = \text{Amount of DNA} \]

\[ \text{K} = \text{DNA per Cell} \]

\[ \text{N}_{\text{TOT}} = \frac{D}{K} = \text{Total Number of Leukocytes} \]

\[ \text{S}_{\text{B}} = \text{Number of B Cell Reads} \]

\[ \text{S}_{\text{L}} = \text{Number of Leukemia Reads} \]

\[ \text{S}_{\text{R}} = \text{Number of Reference Reads} \]

FREQUENCY OF LEUKEMIA CLONE AMONG B CELLS: \[ \text{S}_{\text{L}} \div (\text{S}_{\text{L}} + \text{S}_{\text{B}}) \]

NUMBER OF LEUKEMIA MOLECULES PER LEUKOCYTE: \[ \text{S}_{\text{R}} \times \left( \frac{\text{N}_{\text{R}} \times \text{S}_{\text{B}}}{\text{N}_{\text{TOT}}} \right) \]
Figure 2.
Figure 3.

[Graph showing the fraction of samples with (N) evolved clones as a function of the number of evolved clones.]
Figure 4.
Figure 5.

**STEP 1: IDENTIFICATION OF LEUKEMIC CLONE IN DIAGNOSTIC SAMPLES**

<table>
<thead>
<tr>
<th>Sample collection</th>
<th>Clonality test for all receptors</th>
<th>Identify leukemic clone based on frequency in immune cell repertoire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone marrow aspirate</td>
<td>IgH@ V-J</td>
<td></td>
</tr>
<tr>
<td>TRB@</td>
<td>TRD@</td>
<td>TRG@</td>
</tr>
</tbody>
</table>

**STEP 2: DETECTION OF LEUKEMIC CLONE IN FOLLOW-UP SAMPLES**

<table>
<thead>
<tr>
<th>Serial follow-up sample collection</th>
<th>MRD test for positive receptors</th>
<th>Monitor MRD over time in serial follow-up samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone marrow aspirate OR Peripheral blood</td>
<td>IgH@ V-J</td>
<td></td>
</tr>
<tr>
<td>TRB@</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MRD detected 3 months after end of therapy
Deep sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia

Malek Faham, Jianbiao Zheng, Martin Moorhead, Victoria E. H. Carlton, Patricia Stow, Elaine Coustan-Smith, Ching-Hon Pui and Dario Campana