Use of Whole-Genome Sequencing to Diagnose a Cryptic Fusion Oncogene

John S. Welch, MD, PhD
Peter Westervelt, MD, PhD
Li Ding, PhD
David E. Larson, PhD
Jeffery M. Klco, MD, PhD
Shashikant Kulkarni, PhD
John Wallis, PhD
Ken Chen, PhD
Jacqueline E. Payton, MD, PhD
Robert S. Fulton, MS
Joelle Veizer, BS
Heather Schmidt, BS
Tammi L. Vickery, BS
Sharon Heath
Mark A. Watson, MD, PhD
Michael H. Tomasson, MD
Daniel C. Link, MD
Timothy A. Graubert, MD
John F. DiPersio, MD, PhD
Elaine R. Mardis, PhD
Timothy J. Ley, MD
Richard K. Wilson, PhD

Context Whole-genome sequencing is becoming increasingly available for research purposes, but it has not yet been routinely used for clinical diagnosis.

Objective To determine whether whole-genome sequencing can identify cryptic, actionable mutations in a clinically relevant time frame.

Design, Setting, and Patient We were referred a difficult diagnostic case of acute promyelocytic leukemia with no pathogenic X-RARA fusion identified by routine metaphase cytogenetics or interphase fluorescence in situ hybridization (FISH). The case patient was enrolled in an institutional review board–approved protocol, with consent specifically tailored to the implications of whole-genome sequencing. The protocol uses a “movable firewall” that maintains patient anonymity within the entire research team but allows the research team to communicate medically relevant information to the treating physician.

Main Outcome Measures Clinical relevance of whole-genome sequencing and time to communicate validated results to the treating physician.

Results Massively parallel paired-end sequencing allowed identification of a cytogenetically cryptic event: a 77-kilobase segment from chromosome 15 was inserted en bloc into the second intron of the RARA gene on chromosome 17, resulting in a classic bcr3 PML-RARA fusion gene. Reverse transcription polymerase chain reaction sequencing subsequently validated the expression of the fusion transcript. Novel FISH probes identified 2 additional cases of t(15;17)–negative acute promyelocytic leukemia that had cytogenetically invisible insertions. Whole-genome sequencing and validation were completed in 7 weeks and changed the treatment plan for the patient.

Conclusion Whole-genome sequencing can identify cytogenetically invisible oncogenes in a clinically relevant time frame.

JAMA. 2011;305(15):1577-1584 www.jama.com

©2011 American Medical Association. All rights reserved.
WHOLE-GENOME SEQUENCING TO DIAGNOSE CRYPTIC ONCOGENE

sion was referred to our institution for consideration of allogeneic stem cell transplantation. She had initially presented with hypofibrinogenemia, disseminated intravascular coagulopathy, and pancytopenia (white blood cell count, 1300/µL; hemoglobin, 11.6 g/dL; platelets, 72 × 10³/µL). Her bone marrow contained 61% atypical promyelocytes with invaginated nuclei (including bilobed forms) and dense primary granules (Figure 1). She started induction chemotherapy with ATRA, cytarabine, and idarubicin. However, her metaphase cytogenetics (46, XX, del(9)(q12q32), del(12)(q12q21), −6, −16, add(16)(p13.2), +2 mar[13/20 cells]) (Figure 2) revealed a complex pattern, which is associated with less than 15% long-term survival and is treated with allogeneic transplantation during first remission whenever possible.14,15

Interphase fluorescence in situ hybridization (FISH) suggested a possible fusion between chromosomes 15 and 17 on der(17) but was most consistent with a RARA-PML fusion, not the pathogenic PML-RARA fusion characteristic of M3 AML (Figure 3). RT-PCR to detect a PML-RARA fusion transcript was not performed at the referring institution.

These findings led to a diagnostic conundrum, and ATRA was discontinued. Persistent AML was observed on day 14. The patient entered a complete remission following induction with cytarabine, idarubicin, and etoposide. She was then referred to our institution for consideration of allogeneic stem cell transplantation. At that time, her bone marrow biopsy revealed no morphologic evidence of AML and had normal metaphase cytogenetics, normal interphase FISH results, and no evidence of PML-RARA by RT-PCR. HLA typing identified 1 matched sibling. This case posed a diagnostic dilemma with prognostic and therapeutic consequences: does the patient have APL, or does she have AML with unfavorable-risk cytogenetics?

Because her leukemic cytomorphology was consistent with APL, we empirically recommended 2 cycles of arsenic trioxide consolidation, which she received.16

Little material from her original leukemia remained for subsequent evaluation, and no clinical samples were available for FISH or RT-PCR. However, 2 vials of bone marrow cells had been cryopreserved under a research protocol at her referring institution. DNA and RNA were generated from these respective samples (the RNA sample was severely degraded). We obtained appropriate consent for whole-genome sequencing and completed this analysis using paired-end reads. Our primary goal was to determine if whole-genome sequencing could identify an actionable mutation (eg, a cryptic X-RARA rearrangement) in a clinically relevant time frame (6-8 weeks).

METHODS

A “movable-firewall” within our research protocol allows for the communication of clinically relevant findings to the patient’s physician and to the patient while strictly maintaining patient anonymity among all research personnel. Deidentified samples are entered into a tissue bank. Clinical information (eg, age, sex, disease, treatment, outcome) is maintained in association with deidentified codes only. A list associating deidentified codes with personal patient information (name, date of birth, treating physician) is maintained in a locked safe; a single protocol administrator has access to this list, and the research team can communicate medically relevant information to the administrator. The administrator communicates this information to the treating physician, who is responsible for informing the patient of the whole-genome sequencing results and their clinical implications.

After obtaining explicit consent for whole-genome sequencing with an institutional review board–approved protocol, DNA libraries were generated from 1 cryovial of the original bone marrow aspirate and from a skin punch biopsy obtained during remission (matched normal cells). We generated 187.1 and 200.1 billion base pairs of DNA sequence from each of the respective samples, with an average read depth of 43.7× and 46.8×, respectively. Library generation, sequence production, and data analysis were performed as previously described.17-20 Adequate genome-wide coverage (>99.5% diploid coverage) was ensured by assessing the coverage of known single-nucleotide polymorphisms in the patient’s genome, as defined with data collected from the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, California).

All of the high-quality single-nucleotide variants (SNVs) found in tumor and skin samples from this patient are available in the database of genotypes and phenotypes (dbGaP) of the National Center for Biotechnology Information (phs000159.v3.p2).

Reagents and methods for PCR validation, RT-PCR, and FISH analyses, and the National Center for Biotechnology Information Entrez Gene identification numbers of all genes relevant to this article, are described in the eMethods section available at http://www.jama.com.

RESULTS

Validated whole-genome sequencing results were completed and reported to the patient’s physician 7 weeks after obtaining the DNA samples.

The timeline of data production, analysis, and validation was as follows: day 1, DNA samples logged in at Washington University Genome Institute; day 5, libraries completed and sequencing begun; day 18, sequence completed; day 22, alignment to reference sequence completed; day 24, prediction of SNVs completed; day 25, structural variants predicted by BreakDancer20; day 52, insertional fusion completely validated by PCR and results transmitted to the treating physician.

Using massively parallel DNA sequencing with paired-end reads, we identified 2 sets of breakpoints between chromosomes 15 and 17, which occur in the LOXL1/PML locus and
RARA locus, respectively (schematically described in Figure 4). PCR amplification across each predicted breakpoint validated the en bloc insertion of a 77-kilobase (kb) segment of chromosome 15 (containing parts of the LOXL1 and PML genes) into intron 2 of RARA, the invariant site of RARA-associated translocations (Figure 5A). This insertion generates 3 novel fusion transcripts: bcr3 PML-RARA, LOXL1-PML, and RARA-LOXL1. Expression of bcr3 PML-RARA was validated by RT-PCR using 3 different primer pairs (using the degraded RNA from the original banked AML sample), including the Clinical Laboratory Improvement Act (CLIA-CAP)–certified InVivoscribe PML/RARα Mix2b kit (Invivoscribe Technologies, San Diego, California) (Figure 5B, eFigure 1A, and data not shown). The RARA-LOXL1 fusion was out of frame and is predicted to encode a 67–amino acid protein (eFigure 1B and data not shown). The LOXL1-PML fusion leads to altered splicing and a premature stop codon prior to the PML junction and is predicted to encode a 573–amino acid protein (eFigure 1C).

In addition, we identified and resolved the breakpoints associated with all abnormalities observed with metaphase cytogenetics, including del(9), del(12), and add(16)(p32.2); the latter was in fact a translocation t(16;22)(p13.3; q13.31) (Figure 2, eFigure 2, and data not shown). Two other large deletions not found by conventional cytogenetics were detected by whole-genome sequencing: del(14) and del(19). The latter was also identified in the skin sample, proving that it is an inherited copy number variant (Figure 5A). The predicted deletions of chromosomes 6 and 16 were not detected by whole-genome sequencing. Instead, we identified a 61-megabase inv(6)(p22.3; q13.31) and a translocation t(6;16)(q22.3;p13.3) (Figure 2). We further identified and validated 12 SNVs within protein-coding sequences (Figure 6). SNV allele frequency was consistent with the presence of 2 distinct leukemic clones in the bone marrow, recapitulating the metaphase cytogenetics: 8 of 12 SNVs had a variant allele frequency of 35% to 51%, and 4 of 12 SNVs had a variant allele frequency of 13% to 21% (Figures 2 and 6). The significance of these somatic mutations for...
disease pathogenesis is currently unknown. We designed a new set of fosmid-based FISH probes (each 30-40 kilobases in size) for the detection of insertional fusions that target the minimal PML translocation region (the promoter/enhancer and exons 1-3, which is roughly 30 kilobases) (FIGURE 7). We searched the Washington University Department of Pathology database for AML cases diagnosed during the last 5 years. We identified 11 cases with features suggestive of APL (including any promyelocytic morphology, characteristic CD33+CD34+HLA-DR- immunophenotype, and variable-to-strong myeloperoxidase staining by enzyme cytochemistry) but that lacked normal dual-fusion patterns by FISH. We found that 2 of these specimens contained PML-RARA fusions resulting from cryptic insertions: one was associated with an insertion of PML into the RARA locus (ins[17;15]) and the other with insertion of RARA into the PML locus (ins[15;17]) (FIGURE 8). Both cases (as well as the proband) had RT-PCR confirmation of a PML-RARA fusion (bcr1 and bcr3 isoforms), and all had features typical of APL (TABLE).

**Figure 4.** Ins(17;15) and Resulting PML-RARA Fusion Identified by Whole-Genome Sequencing

A Breakpoints in chromosomes 15 and 17 resulting in PML-RARA fusion

B Binding sites of primers for validation of 77-kb insertion from chromosome 15 to chromosome 17
Figure 5. PCR and Reverse Transcription (RT) PCR Validation of PML-RARA Expression From Case Patient

A, Polymerase chain reaction (PCR) of genomic DNA from the case patient’s skin (normal) and leukemia (tumor) using primers that span the junction of RARA-LOXL1 (P1/P2), PML-RARA (P3/P4), del(12), del(14), LOXL1-PML (P5/P6), and del(19). Note amplification across fusion breakpoints in the leukemia sample but not in the skin sample for all but del(19). DNA ladder: 2176, 1766, 1230, 1033, 653, 517, 453, 394, 298, 234, and 154 base-pairs (bp). B, RNA prepared from cryopreserved leukemia cells from case patient was amplified with a forward primer in PML exon 3 and a reverse primer in RARA exon 3. DNA ladder: 1500, 800, 500, 300, 200, 150, 100, and 50 bp.

COMMENT
The usefulness of massively parallel DNA sequencing has improved considerably with the introduction of paired-end reads, which allow for better mapping efficiency and more accurate identification of junctional breakpoints associated with structural variants (translocations, insertions, and deletions).

In this report, we describe the use of paired-end–read whole-genome sequencing for real-time oncologic diagnosis and describe the genomic details of an oncogenic fusion gene created by an insertional event. Within 7 weeks, we completed the process of library generation, massively parallel sequencing, analysis, and validation of a novel insertional fusion that created a classic PML-RARA bcr3 variant. These findings altered the medical care of this patient, who received ATRA consolidation instead of an allogeneic stem cell transplant. The patient remains in first remission 15 months after her presentation.

RT-PCR was not performed at the referring institution, and results were negative when evaluated at Washington University when the patient was in remission. We did not initially perform RT-PCR with the RNA generated from the cryopreserved sample, because the RNA was severely degraded. Further, because this patient’s complex cytogenetics predicted an unfavorable prognosis, it was essential to determine whether the patient had a recognized PML-RARA fusion gene (because t(15;17) supersedes other cytogenetic findings and predicts a favorable outcome in patients treated with ATRA). Fortunately, the sequencing of the patient’s tumor genome resolved the conundrum, allowing the RT-PCR results to serve as confirmatory proof of the fusion event (despite RNA degradation) and providing a novel mechanism for its formation, which assured us that the diagnosis was correct.

Alternative laboratory approaches could have been used to detect potential pathogenic RARA rearrangements (eg, nested PCR from a linker-ligated library, long-distance PCR,21 BAC clone screening, targeted 3730 sequencing of the 48.5 kb RARA locus). However, such techniques are labor intensive and can require large amounts of starting material, personalized design, and iterative troubleshooting. Moreover, many of these techniques have success rates not adequate for clinical practice. In contrast, whole-genome sequencing with paired-end libraries can be accomplished with as little as 10 ng of starting DNA, is amenable to an automated “pipeline” strategy, and can consistently deconvolute SNVs, small insertions and deletions, structural variants, and clonality. Further, this approach requires no custom reagents and no foreknowledge of genomic regions that must be assessed for diagnostic accuracy.

This study also confirms that reciprocal RARA-PML fusions are not re-
quired for the development of APL. \textit{RARA-PML} has been proposed to participate in APL pathogenesis, although it is identified in only approximately 67% of APL cases by RT-PCR.\textsuperscript{22} Complex rearrangements and deletions can lead to unusual \textit{RARA-PML} transcripts of uncertain significance.\textsuperscript{22-25} Furthermore, \textit{bcr3 RARA-PML} does not independently lead to leukemia in a murine leukemia model.\textsuperscript{26,27} In this patient, the reciprocal fusion \textit{RARA-PML} was absent, and the alternative \textit{RARA-LOXL1} was fused out of frame.

Whole-genome sequencing resolved not only the \textit{PML-RARA} insertion event but also all other abnormalities observed during routine cytogenetic testing. Loss of chromosome 6 and 16 were not detected with whole-genome sequencing; rather, we identified an additional inversion and translocation, \textit{inv(6)(p22.3;q14.1)} and \textit{t(6;16)(q22.31;p13.3)}. This suggests that genetic information on chromosomes 6 and 16 was actually present within the 2 marker chromosomes but that chromosomal banding patterns were disrupted by a 3-way translocation.

\textbf{Figure 7.} Fosmid Selection and Position of Homology to Chromosomes 15 and 17

\begin{table}
\begin{tabular}{|c|c|}
\hline
\textbf{Chromosome 15} & STOML1 \text\Hline
\hline
\textsuperscript{PML} & Fosmid ABC7-4063300P2, 34 kb \textsuperscript{PML} \textsuperscript{bcr3} \textsuperscript{bcr2} \textsuperscript{bcr1} \text▲
\hline
\end{tabular}
\end{table}

\begin{table}
\begin{tabular}{|c|c|}
\hline
\textbf{Chromosome 17} & CDC6 \text\Hline
\hline
\hline
\textsuperscript{RARA} & Fosmid ABC10-45543500L3, 41 kb \textsuperscript{RARA} \textsuperscript{bcr1} \text▲
\hline
\end{tabular}
\end{table}

Schematic representation of fosmid homology to the \textit{STOML1/PML} locus and to the \textit{CDC6/RARA} locus. ATG indicates coding sequence start; CC, coiled-coil domain; Kb, kilobase; ZF, zinc finger domain. \textit{bcr1}, \textit{bcr2}, and \textit{bcr3} indicate breakpoints in \textit{PML} associated with different-sized \textit{PML-RARA} fusion transcripts.

\textbf{Figure 8.} Interphase and Metaphase Fluorescence In Situ Hybridization (FISH) of Bone Marrow Aspirate Cells of Additional Patients

\begin{figure}
\begin{center}
\includegraphics[width=\textwidth]{figure8.png}
\end{center}
\end{figure}

Fusion events in additional patients 1 (A and B) and 2 (C and D). A, One fusion signal, 2 red signals (chromosomes 15), and 1 green signal (chromosome 17). B, Fusion event on der(17), consistent with \textit{ins(17;15)}. C, One fusion signal, 2 green signals (chromosomes 17), and 1 red signal (chromosome 15). D, Fusion event on der(15). Probes labeled with Spectrum Orange (\textit{STROML1/PML}) and Spectrum Green (\textit{CDC6/RARA}).
Acute promyelocytic leukemia is increasingly deep coverage will become evident. 

Meaningful diagnostic time frames are dependent on the cancer type being assessed. Some cancers (eg, follicular lymphoma and myelodysplastic syndrome) have long diagnostic windows before patients need therapy; others (eg, breast, lung, colon) have moderately long diagnostic windows (those in which definitive chemoradiotherapy is generally made within 6 to 8 weeks (6-week time frame for whole-genome sequencing is therefore clinically relevant). The critical decision in the treatment in AML and ALL is not which induction therapy to use (because uniform approaches to remission-induction are currently used for both diseases) but whether patients should receive consolidation therapy with chemotherapeutic approaches or should receive allogeneic transplantation. This decision is generally made within 6 to 8 weeks of initial presentation; for AML and ALL, a 6-week time frame for whole-genome sequencing is therefore clinically relevant. However, to fully use this potentially transformative technology
to make informed clinical decisions, standards will have to be developed that allow for CLIA-CAP certification of whole-genome sequencing and for direct reporting of relevant results to treating physicians.

Author Contributions: Drs Ding and Larson had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.


Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Westervelt reported receiving lecture fees from Celgene and Novartis; Dr DiPersio reported receiving consulting and lecture fees from Genzyme. No other authors reported disclosed.

Funding/Support: This study was supported by the Washington University Cancer Genome Initiative, by National Institutes of Health (NIH) grant PO1 CA101937, by Barnes-Jewish Hospital Foundation grant 00335-0505-02 (Dr Ley), and by NIH grant U54 HG002579 (Dr Wilson). Dr Welch is a fellow of the Leukemia & Lymphoma Society.

Role of the Sponsor: The funding organizations had no role in design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Online Only Material: The eMethods and eFigures 1-4 are available at http://www.jama.com.

Additional Contributions: We thank the Washington University Cancer Genomics Initiative for its support and Charles W. Caldwell, MD, PhD, and Carl Freter, MD (both at University of Missouri-Columbia School of Medicine), for referring the patient and for collecting, storing, and contributing the cryopreserved bone marrow cells used for this study. Neither of these individuals received compensation for their contributions.