Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting

Lesley-Ann Sutton,[1]* Viktor Ljungström,[1]* Larry Mansouri,[1] Emma Young,[1] Diego Cortese,[1] Veronika Navrkalova,[2] Jitka Malcikova,[2] Alice F. Muggen,[3] Martin Trbusek,[2] Panagiotis Panagiotidis,[4] Frederic Davi,[5] Chrysoula Belessi,[6] Anton W. Langerak,[3] Paolo Ghia,[7-8] Sarka Pospisilova,[2] Kostas Stamatopoulos,[1, 9-10] and Richard Rosenquist[1]

[1]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden; [2]Central European Institute of Technology, Masaryk University, Brno, Czech Republic; [3]Department of Immunology, Erasmus MC, University Medical Center Rotterdam, The Netherlands; [4]First Department of Propaedeutic Medicine, School of Medicine, University of Athens, Greece; [5]Laboratory of Hematology and Universite Pierre et Marie Curie, Hopital Pitie-Salpetriere, Paris, France; [6]Hematology Department, Nikea General Hospital, Pireaus, Greece; [7]Università Vita-Salute San Raffaele, Milan, Italy; [8]Division of Molecular Oncology and Department of Onco-Hematology, IRCCS San Raffaele Scientific Institute, Milan, Italy; [9]Institute of Applied Biosciences, CERTH, Thessaloniki, Greece; [10]Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece

LAS and VL contributed equally to this work.

ABSTRACT

Next-generation sequencing has revealed novel recurrent mutations in chronic lymphocytic leukemia, particularly in patients with aggressive disease. Here, we explored targeted re-sequencing as a novel strategy to assess the mutation status of genes with prognostic potential. To this end, we utilized HaloPlex targeted enrichment technology and designed a panel including nine genes: *ATM*, *BIRC3*, *MYD88*, *NOTCH1*, *SF3B1* and *TP53*, which have been linked to the prognosis of chronic lymphocytic leukemia, and *KLHL6*, *POT1* and *XPO1*, which are less characterized but were found to be recurrently mutated in various sequencing studies. A total of 188 chronic lymphocytic leukemia patients with poor prognostic features (unmutated *IGHV*, n=137; *IGHV3-21* subset #2, n=51) were sequenced on the HiSeq 2000 and data were analyzed using well-established bioinformatics tools. Using a conservative cutoff of 10% for the mutant allele, we found that 114/180 (63%) patients carried at least one mutation, with mutations in *ATM*, *BIRC3*, *NOTCH1*, *SF3B1* and *TP53* accounting for 149/177 (84%) of all mutations. We selected 155 mutations for Sanger validation (variant allele frequency, 10-99%) and 93% (144/155) of mutations were confirmed; notably, all 11 discordant variants had a variant allele frequency between 11-27%, hence at the detection limit of conventional Sanger sequencing. Technical precision was assessed by repeating the entire HaloPlex procedure for 63 patients; concordance was found for 77/82 (94%) mutations. In summary, this study demonstrates that targeted next-generation sequencing is an accurate and reproducible technique potentially suitable for routine screening, eventually as a stand-alone test without the need for confirmation by Sanger sequencing.

Introduction

Sanger sequencing, long considered the 'gold standard' for mutation detection in clinical laboratories, can produce long read-lengths with highly reliable results; however, it has limited throughput and rapidly becomes both cost- and time-prohibitive when larger expanses of the genome require analysis. It is becoming a necessity to implement methodologies that could circumvent these shortcomings because of the increasing number of prognostically relevant and disease-associated genes identified by whole-genome or whole-exome sequencing.[1-4] Targeted next-generation sequencing (NGS), in which a selected fraction of genes is sequenced, is emerging as a much more efficient means to identify genetic variants and is now maturing to the point that it is being considered by many laboratories for routine use.[5-7]

The major advantages of this approach revolve around its capacity to screen a large number of genes and samples simultaneously (currently up to 96 samples). Apart from the ability to multiplex both genes and samples, the profound sequence depth achievable also enables minor subclonal populations to be tracked over time, a phenomenon that was previously underappreciated because of the lack of sensitivity of Sanger sequencing approaches.[8-12] An additional favorable attribute of targeted NGS is the ability to analyze all coding exons within a gene, regardless of its length. This is noteworthy since the size of several prognostic genes, such as *ATM* (62 coding exons), currently hinders comprehensive gene analysis in a clinical setting.

That said, for NGS technology to be successfully integrated into clinical practice, several criteria must be fulfilled. Firstly, analytical accuracy must be both high and well-defined within the clinically relevant genes or regions. Second, the NGS workflow should provide good coverage i.e. yield sufficient

 doi:10.3324/haematol.2014.109777
The online version of this article has a Supplementary Appendix.
Manuscript received on April 24, 2014. Manuscript accepted on December 3, 2014.
Correspondence: richard.rosenquist@igp.uu.se

data to ensure that the vast majority of targeted bases are at an appropriate depth to permit high-quality variant calling. Finally, the method should be highly robust and reproducible; in this respect, as the workflow becomes more streamlined, automation should reduce the technical complexity. Given the immense amount of data generated by NGS, issues regarding bioinformatics are also pertinent and careful consideration is required on matters such as data curation, analysis and interpretation, overall management, and storage, the main objective being a system that would integrate with the NGS data analysis pipeline to ensure rapid and consistent calling of clinically relevant mutations.

Large-scale sequencing studies have recently revealed a number of novel recurrent mutations in chronic lymphocytic leukemia (CLL),[1-4] the prime examples being mutations within *NOTCH1*, *SF3B1* and *BIRC3*, which are found at higher frequencies in patients with more aggressive disease and a poor clinical outcome.[13-21] Prompted by these findings, we sought to explore whether targeted NGS could be a superior method for mutational screening of genes with prognostic potential in CLL, and hence be used as a sensitive and accurate routine tool for diagnostic and prognostic purposes. To this end, we designed a HaloPlex gene panel with genes selected for inclusion comprising: (i) those with established prognostic and predictive significance e.g. *TP53*; (ii) those that appear to have a specific significance e.g. *ATM*, *NOTCH1*, *SF3B1*, *MYD88* and *BIRC3*, the clinical value of which is currently being assessed; and (iii) those found to be recurrently mutated at a lower frequency in CLL but whose significance remains unknown e.g. *KLHL6*, *POT1* and *XPO1*.[4,12-25] We demonstrate that the sensitivity and specificity of targeted NGS of a disease-specific subset of genes equaled (if not exceeded) that of Sanger sequencing and could, therefore, conceivably be used in clinical practice, eventually as a stand-alone test.

Methods

Patients' material

One hundred and eighty-eight CLL patients with poor prognostic features were included in this study (73% *IGHV* unmutated, 27% *IGHV3-21* subset #2; *Online Supplementary Table S1*). All samples came from an international, multicenter cohort of patients who were diagnosed according to the International Workshop on CLL criteria and displayed a typical CLL immunophenotype.[26] Informed consent was collected according to the Helsinki Declaration and ethical approval was granted by local ethical review committees.

Targeted enrichment and next-generation sequencing

Nine genes of prognostic interest in CLL were submitted for HaloPlex (Agilent Technologies, Santa Clara, CA) probe design to the SureDesign service (https://earray.chem.agilent.com/suredesign/home.htm). Our design included 2,309 amplicons covering 41,962 bases; all coding exons within selected genes were targeted and comprised 202 regions, with 99.1% of target bases covered by at least one probe. The target regions were captured using the HaloPlex Target Enrichment kit according to the manufacturer's instructions. Paired-end sequencing (100 bp reads) was performed across two lanes of the HiSeq2000 instrument (Illumina, San Diego, CA, USA) and the mean read depth within the regions of interest was ~1500 reads/base.

Data analysis and variant calling

Illumina sequencing adapters were removed using Cutadapt (v.0.9.5) and trimmed reads were aligned to the hg19 human reference genome (February 2009 assembly) using MosaikAligner version 2.1.33 (http://arxiv.org/abs/1309.1149)[27] with the following parameter settings: -hs 15, -ms 7, -mhp 100, -act 35, -minp 0.95 and –mmp 0.05, allowing for a maximum of 5% mismatches per read. Tumor variants were identified using VarScan 2 in mpileup2cns mode with detection thresholds of 20x coverage, variant allele frequency 0.1, a minimum of 10 reads supporting the variant, and an average base quality of 15 at the variant position.[28] In addition to the 188 CLL cases, 18 samples from normal tissue (buccal swabs or sorted T cells) were also included in the sequencing run and bioinformatically processed. Variants detected within the normal tissues facilitated the detection of recurrent variants which likely arose during library preparation or as a sequencing error and were subsequently removed from the tumor call files. Variants had to meet the following conditions to be included in downstream analysis: (i) be located within an exonic or splicing region; (ii) be non-synonymous; (iii) not be listed in the European 1,000 genomes variant database; (iv) not be listed in dbSNP137 unless also listed in the Catalogue of Somatic Mutations in Cancer (COSMIC)65 database; and (v) have a variant allele frequency >0.1 (Figure 1A).

Validation of mutations by Sanger sequencing

Selected variants were bi-directionally sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit and an ABI 3730 DNA Analyzer (Life Technologies, Carlsbad, CA, USA).

Assessing the reproducibility of targeted next-generation sequencing

To estimate technical precision and reproducibility, we repeated the procedure with 63 CLL cases that had been included in the initial sequencing run and were found to carry at least one mutation. A new gene panel including six of the prognostic genes, i.e. *BIRC3*, *NOTCH1*, *POT1*, *SF3B1*, *TP53* and *XPO1*, was designed and the entire experimental setup was independently replicated, i.e. new libraries were prepared (using the same genomic DNA) and re-sequenced on the HiSeq2000. Data analysis and variant calling were performed as detailed above and runs were compared.

Results and Discussion

To determine the feasibility of targeted NGS in a clinical setting, we constructed a gene panel including nine prognostic genes and applied the technology to 188 CLL patients with a poor prognosis (cases with unmutated *IGHV* genes or with B-cell receptors expressing the *IGHV3-21* gene, typical of clinically aggressive stereotyped subset #2[29-32]) in order to increase the possibility of finding mutations within these genes. Our aim was to assess: (i) the sequencing quality of targeted NGS, in terms of representation and coverage; (ii) the sequence reliability, measured in terms of sensitivity and specificity, as compared to Sanger sequencing; and (iii) the reproducibility of this targeted NGS methodology, using a new gene panel incorporating six of the previously analyzed prognostic genes. Since our study was focused on the technical applicability of targeted NGS to clinical diagnostics (and also due to the availability of material) we did not include constitutive DNA in our sequencing runs.

Our initial panel design covered 99.1% of the targeted coding regions and sequencing yielded an average read

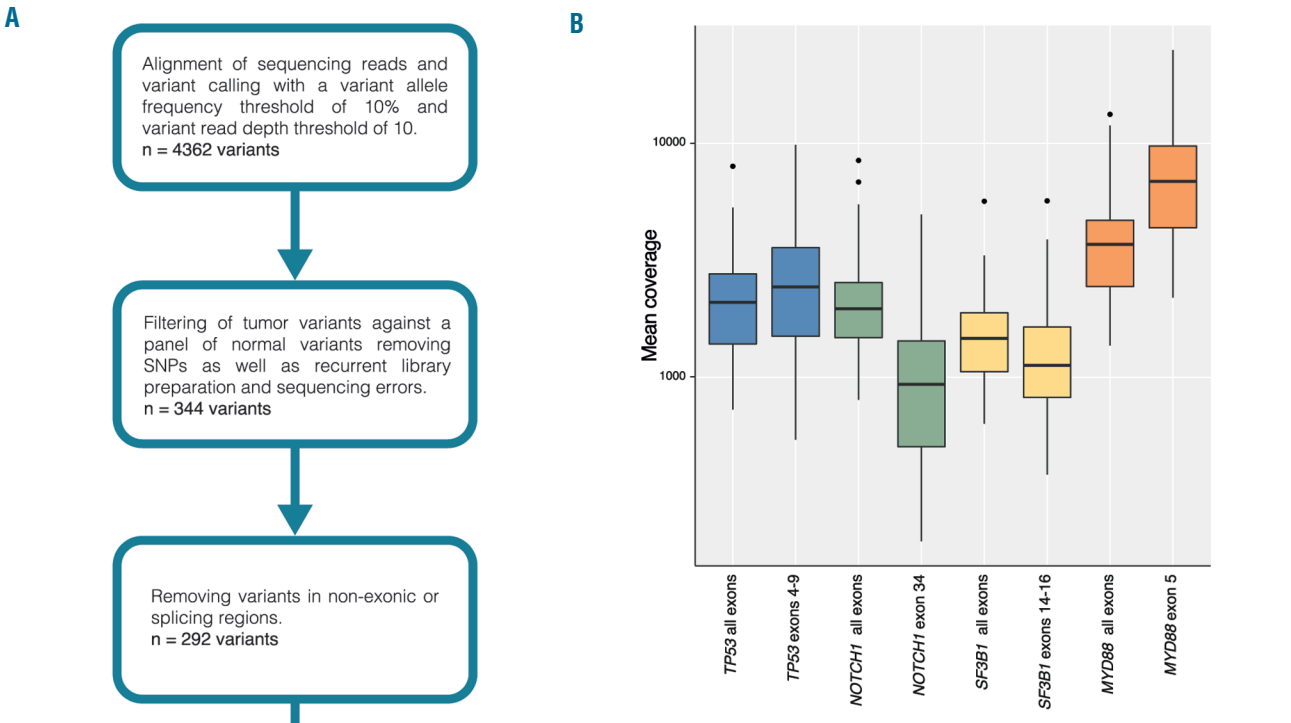

Figure 1. Filtering procedure for data analysis and variant annotation. Following library preparation and sequencing, the next steps include quality assessment, read alignment against a reference genome and variant calling. Filtering criteria were imposed to minimize false-positive rates and included the removal of variants that were supported by less than ten reads, removing variants located in intergenic, intronic or untranslated regions and recurrent library preparation/sequencing errors, as well as removing synonymous variants and those that are present in dbSNP137 and do not have a COSMIC ID. In contrast, we retained 'single nucleotide polymorphisms' (SNPs) with a COSMIC ID on the grounds that they may in fact represent true somatic mutations of pathogenic relevance. This strategy of retaining single nucleotide polymorphisms that have been assigned a COSMIC ID led to the identification of 14 cases with potentially pathogenic variants within *TP53* (n=11), *ATM* (n=2) and *XPO1* (n=1) (*Online Supplementary Table S3*). Ten of these cases were subsequently sequenced by the Sanger method and all were validated. These results demonstrate that for certain, well-characterized genes such as *TP53*, somatic mutations may be inadvertently removed by filtering out all mutations found in single nucleotide polymorphism databases. Figure 1A outlines the filtering procedure implemented in this study. Figure 1B shows the median coverage for all samples in selected genes within all exons or recurrent hotspot exons.

depth of ~1500 reads/base within the regions of interest. While average read depth is a useful metric, uniformity of coverage within the targeted regions must also be considered to ensure high quality variant calling. For this reason we set a stringent cut-off and only included for further analysis those cases in which ≥80% of bases within the targeted regions of interest achieved at least 100x coverage; 96% (180/188) of our samples met this criteria (median: 95%, range: 80-100%). Although choosing to be strict, since any fixed cut-off is inherently arbitrary, it is worth mentioning that of the eight cases excluded, only two were deemed to have failed while the remaining cases varied in the percent of bases achieving 100x coverage (for the 8 excluded cases; median: 63%, range: 5-78%) (*Online Supplementary Table S2*). As an illustrative example of the profound sequence depth, the median coverage for all samples in selected genes with recurrent hotspot regions is shown in Figure 1B.

After a stringent filtering procedure (summarized in Figure 1A), we subsequently found that 114/180 (63%) patients carried at least one mutation; *SF3B1* (37/180; 21%), *ATM* (35/180; 19%), *NOTCH1* (33/180; 18%), *TP53* (25/180; 14%), *XPO1* (14/180, 8%), *POT1* (9/180; 5%), *BIRC3* (4/180; 2%), *MYD88* (2/180; 1%) and *KLHL6* (1/180; 0.6%) (Figure 2A). A total of 177 mutations were detected, comprising 97 unique mutations and 11 recurrent mutations or recurrently mutated codons. Recurrent mutations were observed in five of the nine genes analyzed, i.e. *SF3B1*, *NOTCH1*, *TP53*, *POT1* and *XPO1*, and accounted for 80/177 (45%) of all mutations (*Online Supplementary Table S3*). We detected single nucleotide variants and deletions or insertions, with missense single nucleotide variants being the most prevalent (113/177; 64%) (Figure 2B).

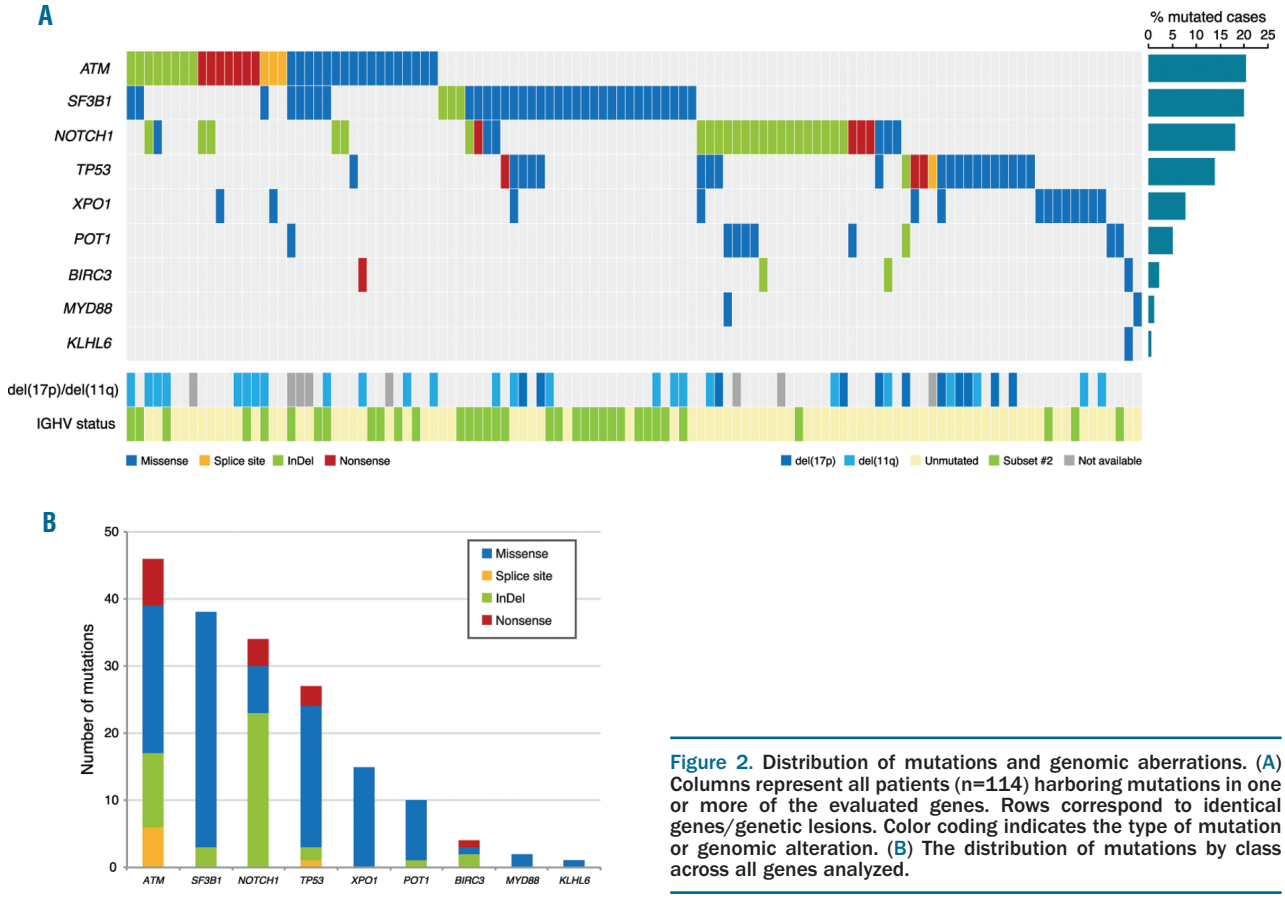

Figure 2. Distribution of mutations and genomic aberrations. (A) Columns represent all patients (n=114) harboring mutations in one or more of the evaluated genes. Rows correspond to identical genes/genetic lesions. Color coding indicates the type of mutation or genomic alteration. (B) The distribution of mutations by class across all genes analyzed.

Of the patients carrying mutations, 45% (51/114) were found to carry more than one mutation. Eighty-two percent of these patients (42/51) harbored mutations in more than one of the evaluated genes, while the remaining 9/51 (18%) cases had multiple mutations in the same gene (Figure 2A), thus explaining the apparent discrepancy between the total number of mutations and the total number of mutated cases for a given gene e.g. the *ATM* gene. Looking at the co-occurrence of mutations, a significant proportion of *ATM* mutated cases also carried mutations in other genes (18/35; 51%). More specifically, *ATM* and *SF3B1* mutations occurred together in eight cases, while *ATM*-mutant cases were also positive for *NOTCH1* in six cases (Figure 2A).

While it is known that some genes are characterized by mutation hotspots, others revealed no consistent pattern of mutations and instead sequence changes were detected throughout the entire coding region. For instance, 97% (37/38) of *SF3B1* mutations were localized to exons 14-16 (one single nucleotide variant was found within exon 18), whereas the 46 mutations detected in the *ATM* gene were found dispersed throughout 30 exons, spanning from exon 4 to exon 63 (Figure 3). This latter finding further illustrates the impractical nature of screening large genes for mutations by Sanger sequencing, and hence underscores the need for large-scale analysis, now achievable through the use of technologies such as HaloPlex. In contrast to genes with known hotspot regions, the precise clinical impact of individual variants outside hotspots or in genes lacking hotspot regions altogether, e.g. the large *ATM* gene, is still a source of uncertainty; although bi-allelic *ATM* inactivation is associated with shorter time to treatment and shorter overall survival,[33-36] further investigation is required in large, prospective clinical cohorts that include matched tumor/germline samples.

After the initial screening, we selected 155 mutations for validation, with mutant allele frequencies ranging from 10-99% (*Online Supplementary Table S4*). Collectively, samples chosen for validation contained various types of mutations, including missense and splice site mutations, insertions, and deletions, to ensure that our targeted NGS approach had both the capability and sensitivity to accurately detect a spectrum of mutations. Sanger sequencing was used as the comparator method to assess the concordance and detection rate of our NGS method and we were able to confirm 143/155 (92%) mutations. The 11 discrepant mutations not found by Sanger sequencing affected five genes (*ATM*, *BIRC3*, *NOTCH1*, *POT1*, and *SF3B1*), and consisted of five non-synonymous single nucleotide variants and six frame-shift deletions (*Online Supplementary Table S5*). Since the variant allele frequency of these discordant variants ranged from 11-27% it is most probable that for these cases, given the limit of detection of Sanger sequencing, dropout of the variant allele may have led to a homozygous reference call; however, we cannot definitively rule out that these variants were falsely generated during the amplification or sequencing process.

The reproducibility of our procedure was assessed by

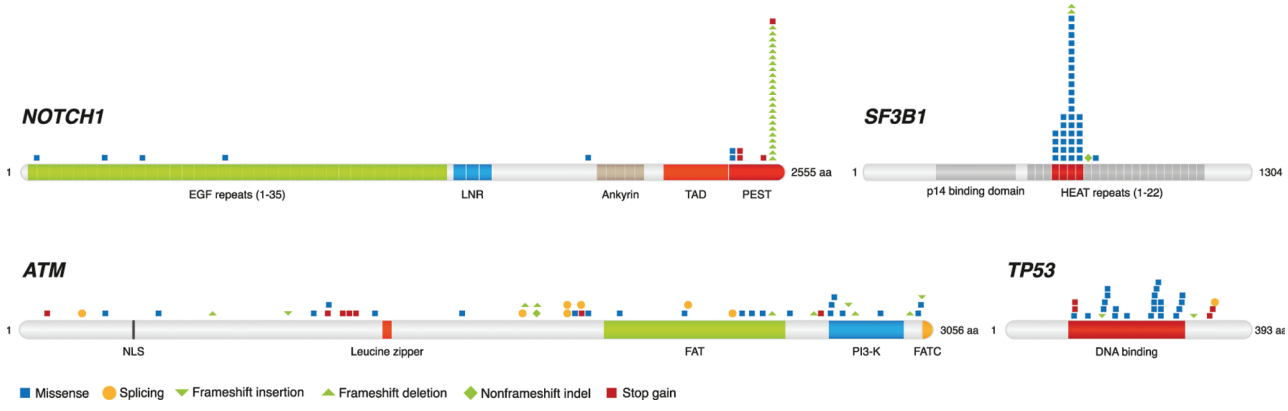

Figure 3. Localization and frequencies of mutations within *ATM*, *NOTCH1*, *SF3B1* and *TP53*. This schematic illustrates how mutations within *NOTCH1*, *SF3B1* and *TP53* predominantly localize to hotspots; recurrent mutations within *SF3B1* are targeted to its highly conserved C-terminal domain, with all but one mutation in this study being detected between the fifth to eighth HEAT domains (exons 14-16). Similarly, a recurrent two base-pair deletion in exon 34 (PEST domain) accounts for the majority of *NOTCH1* mutations observed in CLL. For *TP53*, 24/27 (89%) mutations occurred within the DNA-binding motifs (exons 4-8); remaining mutations were located in exons 9 and 10. In contrast, mutations in the *ATM* gene are unique and distributed throughout the gene's 62 coding exons. To date, no mutational hotspots have been identified in *ATM*.

repeating the experiment using a new panel design, which included six of the aforementioned prognostic genes i.e. *BIRC3*, *NOTCH1*, *POT1*, *SF3B1*, *TP53* and *XPO1*. *KLHL6* and *MYD88* were excluded from the second panel since, given the nature of our cohort, very few cases were found to harbor mutations in these genes using the initial HaloPlex gene panel (one and two cases, respectively). Hence, for validation purposes the addition of these genes would not have added much value. Regarding *ATM*, the very size of the gene limited its appeal for the validation procedure.

The new panel was then tested on 63 CLL patients who were found to carry at least one mutation in the initial sequencing run. By comparing data acquired from each sequencing run sample-by-sample, we found that 77/82 (94%) variants demonstrated concordance and could, thus, be considered as true positives (*Online Supplementary Table S4*). Conversely, no additional variants were identified in the repeat NGS data that were not identified in the initial sequencing run, thus providing us with a 0% false-positive rate. However, five false-negatives, i.e. variants found in the initial sequence analysis but not in the subsequent sequencing run to assess reproducibility, were detected and comprised three non-synonymous single nucleotide variants (in *SF3B1*, *NOTCH1* and *POT1*) and frame-shift insertions in *BIRC3* and *TP53*; four of these mutations had also been confirmed by Sanger sequencing. A final point worthy of note concerns the aforementioned 11 low-frequency variants that were detected in the initial sequencing run but could not be confirmed by Sanger sequencing (*Online Supplementary Table S5*). Three of these variants (variant allele frequency 11-27%) were included and subsequently identified in the validation NGS sequencing run. These findings further strengthen the likelihood that these low frequency variants are indeed true and that the validation was hampered by the detection limit of Sanger sequencing.

In the same vein, it has recently been reported that very small leukemic subclones carrying mutations in *TP53* may play an important role in disease progression and have the same unfavorable prognostic impact as clonal *TP53* defects.[11,12] While not the focus of this particular study, we noted that minor subclones i.e. those with a variant allele frequency between 1% and 9.9%, could be identified using HaloPlex technology and their subsequent detection in the validation sequencing run is highly indicative of their status as a true mutation (exemplified in *Online Supplementary Table S6*). Nevertheless, further studies are required to investigate the significance of small subclones in other prognostic genes such as *BIRC3*, *NOTCH1* and *SF3B1*.

While the application of targeted NGS in the clinical arena is encouraging, as with every novel technology, limitations and challenges need to be carefully considered before routine implementation in clinical practice. The present study represents a step in this direction. In particular, we demonstrate that NGS is an accurate and reproducible technique with high potential for routine diagnostics, eventually as a stand-alone test without the need for confirmation by Sanger sequencing as evidenced by the low number of false-negatives. Moreover, the simultaneous sequencing of multiple genes in clinical testing will inevitably result in the detection of novel variants whose clinical significance (if any) is unknown. Several bioinformatics tools are available to aid with the interpretation of mutational data e.g. PolyPhen and SIFT, however, caveats exist regarding the use of predictive software and while such software may be valuable as a screening tool for identifying alleles of high pathogenic potential, none has been clinically validated.[37] However, this is not specifically a new challenge and as larger cohorts are studied and combined with functional analyses, the ability to discriminate pathogenic variants from incidental mutations i.e. those unrelated to the disease, should improve. Of equal importance is the issue of how to handle big biodata in a diagnostically/clinically meaningful manner; our study also provides evidence that it is possible to devise a data analysis pipeline based on published algorithms that achieves

high accuracy and provides automated rule-based filtering and annotation of variants, thus offering bioinformatics solutions tailored to the user's requirements. While several of the genes included in our panel have clinically relevant hotspot regions, the clinical impact of finding mutations outside these regions remains uncertain. In routine practice, only regions of known clinical significance should be reported, whereas for research purposes the entire regions could be explored (preferably and whenever possible together with germline DNA), which eventually will further our understanding as to whether such mutations are clinically significant. Finally, since a number of different enrichment techniques exist for targeted NGS, it will also be important to compare these against HaloPlex, to determine the capacity of each methodology and define the minimal requirements (e.g. coverage, sequence depth, variant frequency) for future clinical testing.

In summary, we have demonstrated the applicability of targeted NGS in CLL, when appropriately validated with deep coverage and proper quality controls, for routine mutational screening. In the near future it will be possible to combine copy-number analysis with mutation analysis, thereby enabling the simultaneous detection of known copy-number aberrations [e.g. del(11q), del(13q), del(17p) and trisomy 12] along with genes selected for inclusion within a targeted panel. This will ultimately provide a more comprehensive view of the genomic landscape of CLL with implications for improved clinical decision-making in the context of the implementation of tailored therapies.

Acknowledgments

This research project was supported by the Nordic Cancer Union, the Swedish Cancer Society, the Swedish Research Council, the Lion's Cancer Research Foundation, Uppsala, and the European Community's 7th Framework Program (FP7/2007-2013) under grant agreement n. 259796 (DiaTools), Associazione Italiana per la Ricerca sul Cancro (AIRC) (Special Program Molecular Clinical Oncology 5 per mille #9965), Milano, Italy; Ricerca Finalizzata 2010 (RF-2010-2318823), Ministero della Salute, Roma, Italy; KRIPIS action, funded by the General Secretariat for Research and Technology of Greece and the European Regional Development Fund of the EU under the O.P. Competitiveness and Entrepreneurship, NSRF 2007-2013, by project MSMT-CR VaVPI CZ.1.05/1.1.00/02.0068 (CEITEC), research grants IGA-MZ-CR NT13493-4/2012 and FNB Sup 14/13 and 7th FP Health project NGS-PTL/2012-2015/n.306242 of European Commission and MSMT-CR (2013-2015, n.7E13008).

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med. 2011;208(7):1389-1401.
2. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475(7354):101-105.
3. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011;365 (26):2497-2506.
4. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet. 2012;44(1):47-52.
5. Kohlmann A, Grossmann V, Klein HU, et al. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. J Clin Oncol. 2010;28(24):3858-3865.
6. Kohlmann A, Klein HU, Weissmann S, et al. The Interlaboratory RObustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. Leukemia. 2011;25(12):1840-1848.
7. Mansouri L, Sutton LA, Ljungstrom V, et al. Feasibility of targeted next-generation sequencing of the TP53 and ATM genes in chronic lymphocytic leukemia. Leukemia. 2014;28(3):694-696.
8. Schuh A, Becq J, Humphray S, et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood. 2012;120(20):4191-4196.
9. Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152(4):714-726.
10. Landau DA, Carter SL, Getz G, et al. Clonal evolution in hematological malignancies and therapeutic implications. Leukemia. 2014;28(1):34-43.
11. Rossi D, Khiabanian H, Spina V, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. Blood. 2014;123(14):2139-2147.
12. Malcikova J, Stano-Kozubik K, Tichy B, et al. Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. Leukemia. 2014 Oct 28. doi: 10.1038/leu.2014.297. [Epub ahead of print]
13. Del Giudice I, Rossi D, Chiaretti S, et al. NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive transcriptional profiling and refine the intermediate prognosis of +12 CLL. Haematologica. 2012;97(3):437-441.
14. Rossi D, Fangazio M, Rasi S, et al. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. Blood. 2012; 119(12):2854-2862.
15. Mansouri L, Cahill N, Gunnarsson R, et al. NOTCH1 and SF3B1 mutations can be added to the hierarchical prognostic classification in chronic lymphocytic leukemia. Leukemia. 2013;27(2):512-514.
16. Rossi D, Rasi S, Spina V, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. Blood. 2013;121(8): 1403-1412.
17. Schnaiter A, Paschka P, Rossi M, et al. NOTCH1, SF3B1, and TP53 mutations in fludarabine-refractory CLL patients treated with alemtuzumab: results from the CLL2H trial of the GCLLSG. Blood. 2013;122(7): 1266-1270.
18. Strefford JC, Sutton LA, Baliakas P, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. Leukemia. 2013;27 (11):2196-2199.
19. Cortese D, Sutton LA, Cahill N, et al. On the way towards a 'CLL prognostic index': focus on TP53, BIRC3, SF3B1, NOTCH1 and MYD88 in a population-based cohort. Leukemia. 2014;28(3):710-713.
20. Stilgenbauer S, Schnaiter A, Paschka P, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood. 2014;123(21):3247-3254.
21. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia. 2014 Jun 19. doi: 10.1038/leu.2014.196. [Epub ahead of print]
22. Zenz T, Eichhorst B, Busch R, et al. TP53 mutation and survival in chronic lymphocytic leukemia. J Clin Oncol. 2010;28(29):4473-4479.
23. Oscier DG, Rose-Zerilli MJ, Winkelmann N, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood. 2013;121(3):468-475.
24. Trbusek M, Malcikova J. TP53 aberrations in chronic lymphocytic leukemia. Adv Exp Med Biol. 2013;792:109-31.
25. Jeromin S, Weissmann S, Haferlach C, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated

CLL patients. Leukemia. 2014;28(1):108-117.
26. Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood. 2008; 111(12):5446-5456.
27. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. 2011;17(1):10-12.
28. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-576.
29. Ghia P, Stamatopoulos K, Belessi C, et al. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene. Blood. 2005;105(4):1678-1685.
30. Thorselius M, Krober A, Murray F, et al. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. Blood. 2006;107(7):2889-2894.
31. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood. 2007;109(1):259-270.
32. Bomben R, Dal Bo M, Capello D, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. Br J Haematol. 2009;144(4): 492-506.
33. Stankovic T, Stewart GS, Fegan C, et al. Ataxia telangiectasia mutated-deficient B-cell chronic lymphocytic leukemia occurs in pregerminal center cells and results in defective damage response and unrepaired chromosome damage. Blood. 2002;99(1):300-309.
34. Austen B, Powell JE, Alvi A, et al. Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. Blood. 2005;106(9):3175-3182.
35. Austen B, Skowronska A, Baker C, et al. Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. J Clin Oncol. 2007;25(34):5448-5457.
36. Skowronska A, Austen B, Powell JE, et al. ATM germline heterozygosity does not play a role in chronic lymphocytic leukemia initiation but influences rapid disease progression through loss of the remaining ATM allele. Haematologica. 2012;97(1):142-146.
37. Flanagan SE, Patch AM, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers. 2010;14(4):533-537.