Comparative analysis of targeted next-generation sequencing panels for the detection of gene mutations in chronic lymphocytic leukemia: an ERIC multi-center study

Lesley-Ann Sutton,[1] Viktor Ljungström,[1,2] Anna Enjuanes,[3] Diego Cortese,[1] Aron Skaftason,[1] Eugen Tausch,[4] Katerina Stano Kozubik,[5] Ferran Nadeu,[3] Marine Armand,[6] Jitka Malcikova,[5] Tatjana Pandzic,[2] Jade Forster,[7] Zadie Davis,[8] David Oscier,[8] Davide Rossi,[9] Paolo Ghia,[10] Jonathan C. Strefford,[7] Sarka Pospisilova,[5] Stephan Stilgenbauer,[4] Frederic Davi,[6] Elias Campo,[3] Kostas Stamatopoulos[1,11] and Richard Rosenquist[1,12] on behalf of the European Research Initiative on CLL (ERIC)

[1]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; [2]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; [3]Institut d'Investigacions Biomèdiques August Pi iSunyer (IDIBAPS), Barcelona, Spain; Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain and Hospital Clínic of Barcelona, Universitat de Barcelona, Barcelona, Spain; [4]Department of Internal Medicine III, Ulm University, Ulm, Germany; [5]Center of Molecular Medicine, CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic; [6]AP-HP, Hopital Pitie-Salpetriere, Department of Hematology, Sorbonne Université, Paris, France; [7]Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK; [8]Department of Hematology, Royal Bournemouth Hospital, Bournemouth, UK; [9]Hematology Department, Oncology Institute of Southern Switzerland and Institute of Oncology Research, Bellinzona, Switzerland; [10]Division of Experimental Oncology, Università Vita-Salute San Raffaele and IRCCS San Raffaele Scientific Institute, Milan, Italy; [11]Institute of Applied Biosciences, Center for Research and Technology, Thessaloniki, Greece and [12]Clinical Genetics, Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden

 doi:10.3324/haematol.2019.234716

Received: February 11, 2020.

Accepted: March 25, 2020.

Pre-published: April 9, 2020.

Correspondence: *RICHARD ROSENQUIST* - richard.rosenquist@ki.se

Extended Supplementary Materials & Methods

Patient material

Six laboratories, located in the Czech Republic, France, Germany, Spain, Sweden and the United Kingdom, participated in this inter-laboratory NGS study. Forty-eight CLL cases were selected based on mutational data that was ascertained prior to this study using Sanger sequencing, MLPA, targeted NGS or WES. The majority of cases (n=45) contained a previously identified somatic variant in at least one of the genes included in the panel designs (Supplemental Figure S1 & Supplemental Tables S1). All sample and data handling were in accordance with national and international guidelines.

Target enrichment and library construction

The specifics of each amplicon-based targeted NGS assay used in this study are detailed below.

i) HaloPlex Target Enrichment System (Agilent Technologies, Santa Clara, CA)

Probes were designed using Agilent's SureDesign tool (https://earray.chem.agilent.com/suredesign/home.htm). Our design included 3095 amplicons covering 41,962 bases; *in silico* target coverage was 99.9%. 225ng of gDNA was used as input material and the assay was performed following the manufacturer's instructions. In brief, gDNA was fragmented using 16 restriction enzymes and biotin-labeled probes were hybridized to the target DNA. Illumina sequencing barcodes, including index sequences, were incorporated into the targeted fragments. The target DNA and hybridized molecules were captured by streptavidin-labeled magnetic beads and any nicks in the captured, circularized probe-target DNA hybrids were ligated. The circularized DNA was amplified using universal primers and purified. Target-enrichment was validated and quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and samples were pooled in equimolar amounts prior to sequencing.

ii) Illumina TruSeq Custom Amplicon (TSCA) (Illumina, San Diego, CA)

Oligo capture probes flanking each ROI were designed using the Illumina DesignStudio software. The panel comprised 203 amplicons with a total target size of 27,276 bp; *in silico* target coverage was 100%. An input of 250ng gDNA was used for library construction. In brief, the TSCA assay involves hybridizing 2 probes up and downstream of the ROI in unfragmented gDNA. An extension reaction occurs across the ROI, followed by ligation to unite the 2 probes and yield a library of new template molecules with common ends. Each oligo capture probe contains unique, target-specific sequences as well as a universal adapter sequence that is used in the subsequent PCR amplification. Sample-specific indexes were added to each library by PCR and bead-based normalization facilitated volumetric pooling of the libraries prior to sequencing.

iii) Multiplicom CLL MASTR Plus (Agilent Technologies, Santa Clara, CA)

This pre-designed assay uses multiple PCRs for target amplification of DNA and generates 251 amplicons/sample ranging in length from 261-437 bp covering a total of 67,020 bases; *in silico* target coverage was 100%. Based on the DNA quality, a minimum of 120ng of gDNA was used as input divided across 6 separate multiplex PCRs. The PCR products for each sample were combined according to a predefined assay specific mixing scheme. Amplicons were tagged with multiplex identifiers in a second universal PCR, each tagged library was purified with AMPure XP beads (Beckman Coulter, Brea, CA) and quantified using the Qubit dsDNA HS Assay kit (Thermofisher Scientific, Waltham, MA). Purified libraries were pooled in equimolar amounts, denatured and diluted prior to sequencing.

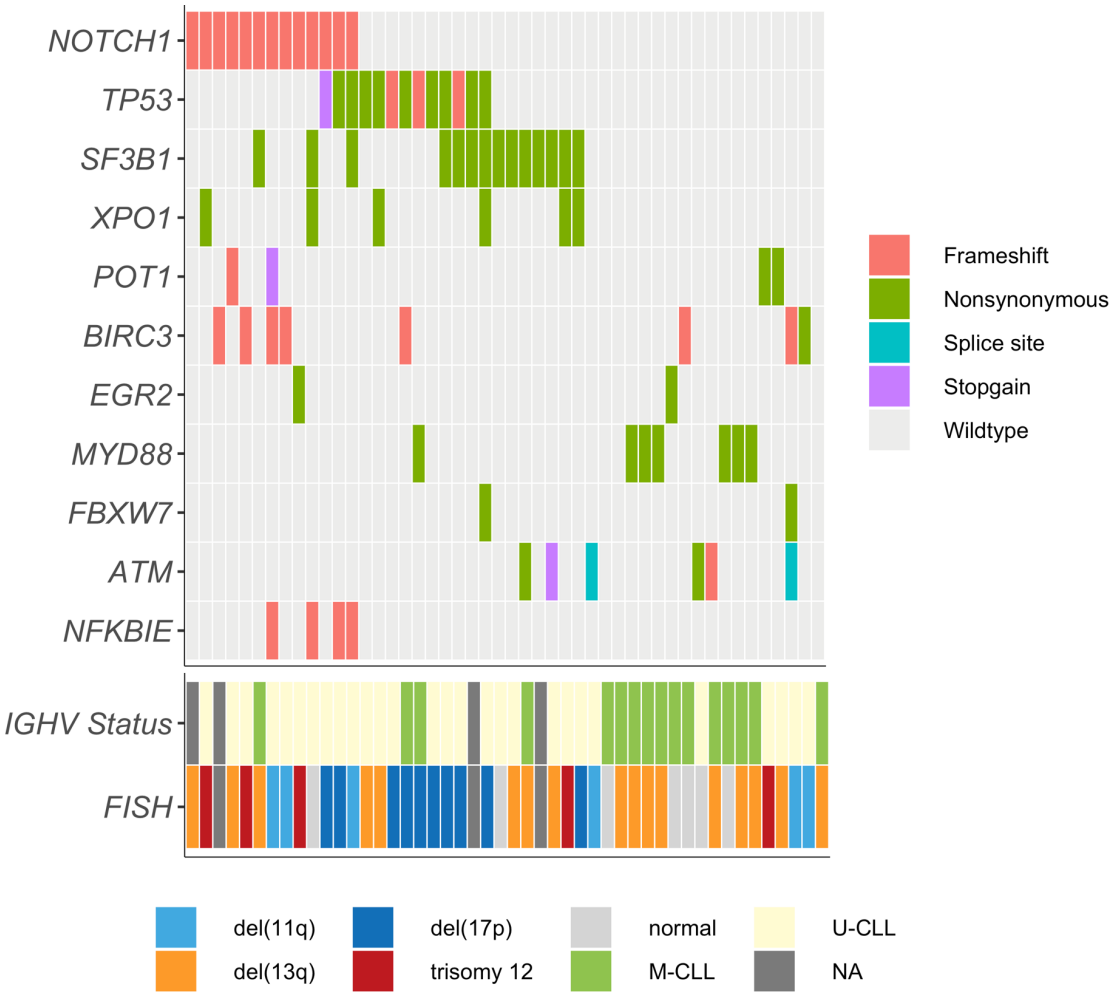
iv) HaloPlex^{HS} capture-based custom-design assay (Agilent Technologies, Santa Clara, CA)

The HaloPlex^{HS} design comprised 3052 amplicons covering 27,231 bases; *in silico* target coverage was 99.8%. gDNA (50ng/samples) from 38 CLL cases included in the technological comparison (material was unavailable for 10 samples) was used for library preparation (similar work-flow as described above

for the Haloplex target enrichment system with an additional step involving the incorporation of UMIs) and the entire experimental setup was repeated and re-sequenced on the MiSeq instrument.

Interpretation of false-positive findings

A variant was deemed as a false-positive if it was only found in a single test center; a total of 4 false-positives with a VAF ≥5% were detected in 2 centers. More specifically, from the data generated by Center 2, a p.L265P variant in *MYD88* was found at a frequency of 22.8% (349/1583 reads) in sample T13 (Supplemental Table S14). This mutation, a recurrent hotspot mutation in CLL, was absent in sample T13 according to other centers but was found by all centers in sample T12; hence, it is possible that this false-positive call in T13 arose from nucleic acid carry-over and contamination between adjacent samples. The remaining false-positive variants were found in the data generated by center 5 and concerned two mutations in sample T5 [*NOTCH1* p.P2462fs (32%) and *SF3B1* p.K700E (39%)] and a *TP53* nonsynonymous substitution (p.C176Y; 10%) in sample T3. Further investigation revealed that the mutational profile of sample T5, as reported by center 5, was distinct from that evidenced in all other test centers and instead mirrored that of sample T4 (with differences in VAFs explained either by dilution of true variants or contaminating input DNA leading to false-positive variants) (Supplemental Table S14). Finally, although no definitive reason could account for the presence of the *TP53* p.C176Y substitution observed in sample T3 in center 5's dataset, it is conceivable that PCR overamplification during the library preparation led to the false-positive result.



Supplemental Figure 1: Distribution of mutations and genetic aberrations. Brick-plot illustrating: (i) mutations detected with a VAF >10%, (ii) hierarchical ordering of recurrent genetic aberrations in CLL as detected by FISH, and, (iii) immunoglobulin heavy variable gene mutational status. Columns represent all patients (n=48) included in the study. Rows correspond to particular genes/genetic lesions. Color coding indicates the type of mutation or genetic alteration. M-CLL: patients harboring mutated IG genes i.e. <98% germline identity; U-CLL: patients harboring unmutated IG genes i.e. ≥98% germline identity; NA: data not available; normal: cases negative for del(17p), del(11q), trisomy 12 or del(13q).

Supplemental Tables

All supplemental tables below are uploaded as a separate Excel file.

Table S1. Molecular characteristics of patients included in the study

Table S2. Coverage obtained for samples prepared with the Illumina TSCA kit

Table S3. Coverage obtained for samples prepared with the Agilent HaloPlex target enrichment system

Table S4. Coverage obtained for samples prepared with the Multiplicom CLL MASTR kit

Table S5. Samples not fulfilling a 90% coverage threshold

Table S6. Coverage per target region for the 5 samples not fulfilling the 90% coverage threshold using the Illumina TSCA kit

Table S7. Coverage per target region for the 7 samples not fulfilling the 90% coverage threshold using the Agilent HaloPlex kit

Table S8. Pairwise comparison of mutations detected by centers utilizing the Multiplicom assay

Table S9. Mutations detected in the 9 genes covered by all three assays utilized in this study

Table S10. Variants found by only a single center

Table S11. Pairwise comparison of mutations detected by centers utilizing the Illumina assay

Table S12. Pairwise comparison of mutations detected by centers utilizing the HaloPlex assay

Table S13. *EGR2* and *NFKBIE* mutations detected in this study

Table S14. Comparison of the mutational profile for samples T4 & T5 and samples T12 & T13

Table S15. Coverage achieved for samples when prepared with the Agilent HaloPlex^{HS} target enrichment system