

Haematologica
HAEMATOL/2019/234716
Version 3

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, Viktor Ljungström, Anna Enjuanes, Diego Cortese, Aron Skafason, Eugen Tausch, Katerina Stano Kozubik, Ferran Nadeu, Marine Armand, Jikta Malcikova, Tatjana Pandzic, Jade Forster, Zadie Davis, David Oscier, Davide Rossi, Paolo Ghia, Jonathan C. Strefford, Sarka Pospisilova, Stephan Stilgenbauer, Frederic Davi, Elias Campo, Kostas Stamatopoulos, and Richard Rosenquist

Collaborative Groups: European Research Initiative on CLL (ERIC)

Disclosures: LAS has received honoraria from Abbvie, Gilead and Janssen. RR has received honoraria from AbbVie, Janssen, Illumina and Roche. JCS has received funding from Roche. KS has received honoraria and research funding from AbbVie and Janssen. DR has received research grants from Abbvie, Cellectia, Gilead, Janssen and honoraria from Abbvie, AstraZeneca, Gilead, Janssen and Loxo. SS has received research grants and honoraria from AbbVie, AstraZeneca, Celgene, Gilead, GSK, Hoffmann La-Roche and Janssen. ET has received honoraria from Abbvie and Roche. EC has received funding from Gilead and honoraria from Jansen, Takeda and Celgene.

Contributions: LAS performed research, analysed the data and wrote the manuscript. VL, AN, DC, AS, EU, KSK, FN, MA, JM and TP performed research and analysed the data. JF, ZD, DO, DR, PG, JCS, SP, SS, FD, and EC performed research, analysed the data and/or provided samples and clinical data. KS and RR directed the study and wrote the paper.

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 Sutton¹, Viktor Ljungström^{1,2}, Anna Enjuanes³, Diego Cortese¹, Aron Skafason¹, Eugen Tausch⁴, Katerina Stano Kozubik⁵, Ferran Nadeu³, Marine Armand⁶, Jitka Malcikova⁵, Tatjana Pandzic², Jade Forster⁷, Zadie Davis⁸, David Oscier⁸, Davide Rossi⁹, Paolo Ghia¹⁰, Jonathan C. Strefford⁷, Sarka Pospisilova⁵, Stephan Stilgenbauer⁴, Frederic Davi⁶, Elias Campo³, Kostas Stamatopoulos^{1,11} and Richard Rosenquist^{1, 12}, on behalf of the European Research Initiative on CLL (ERIC).

¹Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

²Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden

³Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain & Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Hospital Clínic of Barcelona, Barcelona, Universitat de Barcelona, Barcelona, Spain

⁴Department of Internal Medicine III, Ulm University, Ulm, Germany

⁵Center of Molecular Medicine, CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic

⁶AP-HP, Hopital Pitie-Salpetriere, Department of Hematology, Sorbonne Université, Paris, France

⁷Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

⁸Department of Haematology, Royal Bournemouth Hospital, Bournemouth, United Kingdom

⁹Hematology Department, Oncology Institute of Southern Switzerland and Institute of Oncology Research, Bellinzona, Switzerland

¹⁰Division of Experimental Oncology, Università Vita-Salute San Raffaele and IRCCS San Raffaele Scientific Institute, Milan, Italy

¹¹Institute of Applied Biosciences, Centre for Research and Technology, Thessaloniki, Greece

¹²Clinical Genetics, Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden

Running title: Multicenter validation of targeted NGS in CLL

Corresponding author:

richard.rosenquist@ki.se; Phone: +46 8 5177 3928

Word count abstract: 251

Number of text pages: 24

Number of figures: 3

Number of tables: 0

Supplemental file 1 (supplemental tables: 15)

Supplemental file 2 (supplemental methods and figures)

Key words: chronic lymphocytic leukemia, next-generation sequencing, targeted gene panels, mutations, unique molecular identifiers, clinical diagnostics

Grant numbers and sources of support: LAS and RR have received funding from the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Karolinska University Hospital, and Radiumhemmets Forskningsfonder, Stockholm. PG has received funding from the Associazione Italiana per la Ricerca sul Cancro – AIRC, Milano, Italy (Investigator Grant #20246 and Special Program on Metastatic Disease – 5 per mille #21198), and ERA NET TRANSCAN-2 Joint Transnational Call for Proposals: JTC 2016 (project #179 NOVEL), project code (MIS) 5041673. JCS has received funding from Bloodwise (11052, 12036), the Kay Kendall

Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581) and the Bournemouth Leukaemia Fund. KS has received funding from the Hellenic Precision Medicine Network in Oncology and the KRIPIS action supported by the General Secretariat for Research and Technology of Greece. SP has received funding from the Czech Science Foundation GA19-15737S and was also supported by the Ministry of Health of the Czech Republic, Grant No. NV19-03-00091. ET and SS were supported by the DFG (SFB1074, project B2) and the EU (FIRE CLL). ET also received funding from Else Kröner-Fresenius-Stiftung (2010_Kolleg24), EC (01KT1601, FIRE CLL), BMBF (031L0076C PRECISe) and Deutsche Forschungsgemeinschaft (SFB 1074 projects B1, B2). EC has received funding from the Instituto de Salud Carlos III (PMP15/00007), CIBERONC, “La Caixa” Foundation (grant CLLEvolution-HR17-00221). We acknowledge the CF Genomics CEITEC MU supported by the NCMG research infrastructure (LM2015091 funded by MEYS CR) for their support with obtaining scientific data presented in this paper. For the Swedish center, sequencing was performed at Clinical Genomics, Uppsala, and the SNP&SEQ Technology Platform, SciLifeLab at Uppsala University, a national infrastructure supported by the Swedish Research Council (Council for Research Infrastructures [VRRFI]) and the Knut and Alice Wallenberg Foundation.

Disclosures: LAS has received honoraria from Abbvie, Gilead and Janssen. RR has received honoraria from AbbVie, Janssen, Illumina and Roche. JCS has received funding from Roche. KS has received honoraria and research funding from AbbVie and Janssen. DR has received research grants from Abbvie, Cellectia, Gilead, Janssen and honoraria from Abbvie, AstraZeneca, Gilead, Janssen and Loxo. SS has received research grants and honoraria from AbbVie, AstraZeneca, Celgene, Gilead, GSK, Hoffmann La-Roche and Janssen. ET has received honoraria from Abbvie and Roche. EC has received funding from Gilead and honoraria from Jansen, Takeda and Celgene.

ABSTRACT

Next-generation sequencing (NGS) has transitioned from research to clinical routine, yet the comparability of different technologies for mutation profiling remains an open question. We performed a European multicenter (n=6) evaluation of three amplicon-based NGS assays targeting 11 genes recurrently mutated in chronic lymphocytic leukemia. Each assay was assessed by two centers using 48 pre-characterized chronic lymphocytic leukemia samples; libraries were sequenced on the Illumina MiSeq instrument and bioinformatics analyses were centralized. Across all centers the median percentage of target reads $\geq 100\times$ ranged from 94.2-99.8%. To rule out assay-specific technical variability, we first assessed variant calling at the individual assay level i.e. pairwise analysis of variants detected amongst partner centers. After filtering for variants present in the paired normal sample and removal of PCR/sequencing artefacts, the panels achieved 96.2% (Multiplicom), 97.7% (TruSeq) and 90% (HaloPlex) concordance at a VAF $>0.5\%$. Reproducibility was assessed by looking at the inter-laboratory variation in detecting mutations and 107/115 (93% concordance) of mutations were detected by all 6 centers, while the remaining 8/115 (7%) variants were undetected by a single center and 6/8 of these variants concerned minor subclonal mutations (VAF $<5\%$). We sought to investigate low-frequency mutations further by using a high-sensitivity assay containing unique molecular identifiers, which confirmed the presence of several minor subclonal mutations. Thus, while amplicon-based approaches can be adopted for somatic mutation detection with VAFs $>5\%$, after rigorous validation, the use of unique molecular identifiers may be necessary to reach a higher sensitivity and ensure consistent and accurate detection of low-frequency variants.

INTRODUCTION

Chronic lymphocytic leukemia (CLL), a clinically and biologically heterogeneous B cell malignancy, was one of the first cancers for which the genomic landscape was uncovered using next-generation sequencing (NGS) technologies, in particular whole-exome/genome sequencing (WES/WGS).¹⁻⁴ These initial studies led to the discovery of a large number of recurrently mutated genes affecting diverse cellular pathways and processes contributing to the pathobiology of CLL. That said, only a limited number of genes carry mutations at a frequency >5% (i.e. *TP53*, *ATM*, *SF3B1* and *NOTCH1*), with the majority of gene mutations detected in only a minor proportion of CLL patients.^{5, 6} To date, more than 2,000 genes have been reported as mutated in CLL, with mutations occurring in classic tumor suppressor genes (e.g. *TP53* and *ATM*), signaling pathways [e.g. the Toll-like receptor (*MYD88*), NF- κ B (*BIRC3*, *NFKBIE*), and NOTCH (*NOTCH1*, *FBXW7*) pathways] as well as genes involved in essential cellular processes such as RNA processing (e.g. *SF3B1*, *RPS15*, *XPO1*).⁵⁻¹⁰

In addition to mutations or defects in *TP53*, which have long been associated with a poor prognosis and progressive disease, mutations in several of these recently analyzed genes (with the exception of *MYD88*) have been linked to an aggressive clinical course with a significantly shorter time to treatment and a poor outcome to chemo(immuno)therapy (e.g. *ATM*, *BIRC3*, *EGR2*, *NFKBIE*, *NOTCH1*, *RPS15*, *SF3B1*, *XPO1*).⁸⁻²⁰ In recent years it has also been reported that patients carrying minor clones harboring *TP53* mutations (i.e. <10% variant allele frequency, VAF) may have an outcome equally as poor as patients with *TP53* mutations with VAFs >10%, the approximate detection limit of Sanger sequencing.²¹⁻²⁴ Employing ultra-deep sequencing technologies these minor clones have been detected at frequencies as low as 0.1% and may expand as the disease progresses and/or at treatment relapse. Minor clones carrying other gene mutations may also affect outcome, however there are few studies published to date, hence precluding firm conclusions from being drawn.²³ That said, based on currently available data it appears that both the mutational complexity and subclonal diversity strongly influence the evolution of CLL.^{6, 25-29}

Owing to the increasing number of clinically relevant gene mutations identified in CLL, the shift from Sanger sequencing to high-throughput technologies is essential and targeted amplicon-based gene panels are a promising option.^{30, 31} These assays have numerous attractive features including the ability to custom design and screen a large number of genes (complete coding sequence or hotspots) and samples simultaneously, ultimately leading to a reduced cost per sample and a higher throughput. Another appealing facet of targeted gene panels is that the sequencing capacity is efficiently utilized, resulting in a higher coverage of the regions of interest (ROI) and hence allowing for more sensitive detection of low-frequency variants; thus, they are ideally suited for routine diagnostics and monitoring procedures. However, despite the necessity for mutation detection using NGS assays, with numerous technologies available, including both commercial and laboratory developed tests (LDTs), important issues that commonly arise concern the specific technique to choose, and the sensitivity, specificity and reproducibility of individual methodologies. These concerns are magnified within the diagnostic setting as the test results may impact on clinical decision-making and the therapeutic stratification of patients.

The European Research Initiative on CLL (ERIC) conducted this multi-center study to better understand the comparability of various gene panel assays by assessing various analytical parameters such as coverage, sensitivity and reproducibility; with the overall aim of highlighting the critical parameters users should take into consideration when introducing targeted NGS into the laboratory. In brief, we selected three amplicon-based assays (HaloPlex, TruSeq and Multiplicom), each protocol was tested by two centers, and all sequencing was performed on the Illumina MiSeq system. Our cohort comprised 48 well-characterized CLL cases and all centers sequenced the same samples. We observed high concordance between the various technologies and test centers when considering gene mutations with VAFs >5%. Although low-frequency variants were detected by all techniques, greater diversity was observed for mutations with a VAF between 1-5%.

METHODS

Patient material

Genomic DNA [gDNA; tumor and germline (buccal swabs or CD19-depleted peripheral blood mononuclear cells)] from 48 CLL patients was used as the analytical material. Samples were sourced from archival material that had previously been analyzed using established molecular techniques to interrogate either a single gene or numerous targets. Cases were selected based on the available mutational data such that 45/48 cases contained a previously identified somatic variant in at least one of the genes included in the panel designs (Supplemental Figure S1). Quality control was performed centrally before distribution of the samples to the six participating institutes. With the exception of the coordinating center, details regarding pre-characterized mutations were not disclosed to participating laboratories. All cases were diagnosed according to iwCLL guidelines and displayed a typical CLL immunophenotype.³² Informed consent was obtained in accordance with the Declaration of Helsinki and ethical approval was granted by local review committees.

Target enrichment and library construction

Three amplicon-based targeted NGS assays were used in this study. Two assays, the HaloPlex Target Enrichment System (Agilent Technologies, Santa Clara, CA) and the Illumina TruSeq Custom Amplicon (TSCA) (Illumina, San Diego, CA), were customized and targeted 11 genes: full coding sequence (*ATM*, *BIRC3*, *EGR2*, *FBXW7*, *MYD88*, *NFKBIE*, *POT1* and *TP53*) or hotspot regions [*NOTCH1* (exon 34), *SF3B1* (exons 14-16 and 18) and *XPO1* (exons 15-16)]. The third assay, the Multiplicom CLL Multiplex MASTR Plus (Agilent Technologies, Santa Clara, CA), is a commercially designed panel targeting the full coding sequence of nine of the genes listed above; *NFKBIE* and *EGR2* are not included in the design. Intron-exon boundaries were covered by all assays to enable the detection of splice-site mutations. A HaloPlex^{HS} capture-based custom-design assay incorporating unique molecular identifiers (UMIs) was used to validate and more accurately quantify variants present at low frequencies within the 11 genes detailed above. The specifics of each assay are detailed in the online supplementary material.

Next-generation sequencing and data analysis

Cluster generation and paired-end sequencing was performed on the MiSeq instrument (Illumina, San Diego, CA). All sequencing data were centrally analyzed using a custom bioinformatics pipeline. Illumina sequencing adapters were removed using TrimGalore version 0.6.0 and trimmed reads were aligned to the human reference genome hg19/NCBI GRCh37 using BWA mem version 0.7.12 with standard parameters. Variants were identified using VarScan2 version 2.3.7 in mpileup2cns mode with a minimum average quality of 30 and annotated with SnpEff and SnpSift. Bioinformatic analyses of the NGS dataset containing unique molecular identifiers (UMIs) was performed as follows: SurecallTrimmer v4.0.1 (Agilent Technologies, Santa Clara, CA) was used to remove adaptor sequences, mask HaloPlex enzyme footprints and trim low-quality bases. Processed reads were aligned to the hg19 human reference genome with BWA v.0.7.12 and duplicate reads were marked using the Agilent LocatIt tool (v4.0.1). Variants were called with Pisces v.5.2.7.47 using a VAF cutoff of 0.5% and annotated with SnpEff and SnpSift. Variants had to meet the following conditions to be included in downstream analyses: (i) located within an exonic or splicing region; (ii) be non-synonymous; and (iii) not listed in the European 1000 genomes variant database.

RESULTS

Gene Panel Coverage

The gene panel designs for Illumina TSCA, HaloPlex and Multiplicom covered 100%, 99.9% and 100% of the targeted coding regions, respectively. Sequencing 48 CLL samples on the MiSeq platform yielded a median coverage of 3834x and 1540x (range, 2991-7761 and 161-3068, respectively) for the two centers utilizing the Illumina TSCA panel while a median coverage of 1062x and 1953x (range, 334-2427 and 338-7496, respectively) was achieved in the centers performing the analysis with the HaloPlex enrichment system (3 samples failed to generate any reads during the sequencing run in center 5) (Figure 1; Supplemental Tables S2 and S3). For the Multiplicom CLL MASTR kit, 24 samples were sequenced over two runs and this reduction in samples/run combined with the smaller size of the panel (*NFKB1E* and *EGR2* were not included in this predesigned panel) was reflected in the increase in median coverage obtained; 5469x and 3426x (range, 1432-13589 and 2244-4194, respectively) (Figure 1; Supplemental Table S4).

Variability was observed in the depth of coverage generated for individual samples analyzed within the same center and also between centers, however the % of bases covered at 100x was comparable between samples prepared using the same technology at partner sites (Figure 1, Supplemental Tables S2-S4). More specifically, when considering the % of target bases obtaining at least 100x coverage, the two centers utilizing the Multiplicom technology achieved a median of 99.8% and 99.6% (range, 99.4-99.8% and 99.5-99.7%), the Illumina TSCA test centers obtained median values of 99.7% and 96.5% (range, 99.6-100% and 70.5-98%) while the centers evaluating the HaloPlex system recorded median values of 94.2% and 97.6% (range, 64.8-97.2% and 90.2-99.4%) (Figure 1; Supplemental Tables S2-S4). Although overall high sequencing coverage was obtained, a few samples in individual centers had less than 90% of bases within the targeted ROIs achieving at least 100x coverage (Figure 1; Supplemental Tables S2-S5). Detailed information on the coverage per target region for the samples not fulfilling a 90% coverage threshold are provided in Supplemental Tables S6-S7.

Pairwise analysis of variant calls

Although the gene content of both panels was comparable, with the noted exception of the absence of the *NFKB1E* and *EGR2* gene hotspots from the Multiplicom assay, the number and precise location of probes varied between the assay. To rule out any inherent assay-specific technical variability or probe bias we first assessed variant calling at the individual assay level i.e. pairwise analysis of variants detected amongst centers utilizing the same assay. We initially limited this analysis to the 9 genes analyzed by all techniques (*NFKB1E* and *EGR2* will be discussed separately below). After filtering for variants present in the 1000 genomes project, the paired normal sample and synonymous variants and removal of PCR/sequencing artefacts, variants with a VAF > 0.5% were compared. Application of a strict VAF cut-off at this stage of the analysis could be misleading as variants borderline of the threshold would not be taken into consideration and be deemed as discordant. Instead, concordance was defined as the ability to detect a variant irrespective of the VAF, while accuracy was defined as the degree of agreement between the VAFs detected by the same panel. To aid in this analysis, variants were assigned to one of 5 groups: group 1, variant detected in both centers with a VAF >5%; group 2, variant was detected in one center with a VAF >5% and in the partner center with a VAF <5%; group 3, variant was detected in both centers with a VAF <5%; group 4, variant was detected in only one of the partner centers and with a VAF >5%; and, group 5, variant was detected in only one of the partner centers and with a VAF <5%.

Looking first at the mutational data generated by the Multiplicom assay, a total of 133 variants passed the quality filters and these could be grouped as follows: group 1 (n=104), group 2 (n=8), group 3 (n=16), group 4 (n=1) and group 5 (n=4) (Figure 2; Supplemental Table S8). Hence, variant calling between the two centers utilizing the Multiplicom assay reached a concordance of 128/133 (96.2%) with 5 variants detected by only one center (4/5 variants had a VAF <5% with the remaining variant having a VAF of 22.8%) (Figure 2; Supplemental Tables S8-S10). This latter higher frequency variant concerned a p.L265P substitution within *MYD88* and likely constitutes a false-positive

stemming from the wet-lab sample handling (further detailed in Supplemental Materials). Overall, none of the four variants with a VAF <5% reported by a single institute using the Multiplicom assay were found by any other center utilizing the Illumina or HaloPlex methodologies; further supporting the notion that they are false positive results (Supplemental Table S10).

Pairwise comparison of the data generated from the centers utilizing the Illumina assay revealed that 129 variants were detected and could subsequently be assigned to the groups defined above as follows i.e. group 1 (n=100), group 2 (n=6), group 3 (n=19), group 4 (n=0) and group 5 (n=4) (Figure 2; Supplemental Table S9 & S11). The concordance rate was 97.7% (125/128) with 3 variants detected by only one center; all 3 variants had a VAF <5% (Supplemental Tables S9 & S10). The finding of 2/3 of these variants in the centers utilizing the other technologies portends to them being true subclonal mutations present in a minor proportion of leukemic cells (Supplemental Table S9).

A total of 127 variants were found using the HaloPlex assay; however, as 3 samples failed the sequencing run in center 5 (T28, T41 and T48) which collectively harbored 8 variants, these samples (and their variants) were excluded from the pairwise analysis. The remaining 119 variants were distributed amongst the various groups as follows: group 1 (n=89), group 2 (n=5), group 3 (n=13), group 4 (n=7) and group 5 (n=5) (Figure 2; Supplemental Table S9 & S12). The concordance rate was 90% (107/119) with 12 variants detected by only one center; 7/12 variants had a VAF >5% with the remaining 5/12 variants having a VAF <5%. Investigating the variants with a VAF >5% revealed that (i) 3/7 of these variants were found in all other centers indicating that they were false-negatives for center 6; (ii) 2 variants in T5 could be false-positives for center 5 (discussed in more detail below); and, (iii) for the remaining 2 variants (VAFs, 10% and 6.7%), one was not found by any other center while the latter was found at a very low frequency of 0.56% (Supplemental Tables S9 & S10). Notably, the 5/12 variants found with a VAF <5% were all detected by the other methodologies.

Overall, the VAFs reported were in agreement between partner centers and this similarity was evidenced by the narrow range of VAFs for a particular variant, thereby indicating the accuracy of each assay in reporting VAFs (Figure 2).

NFKB1E and *EGR2* were only included in the Illumina and HaloPlex gene panel designs. Within the HaloPlex dataset a total of 6 mutations were reported in the *NFKB1E* gene, five of which were found in both institutes at a VAF >5%, with the remaining variant found either above or below 5% in the partner center (Supplemental Table S13). Thus, the concordance rate for *NFKB1E* mutations detected using the HaloPlex assay was 100%. A total of 7 *NFKB1E* mutations were found by the Illumina assay. However, while 4 were concordant, being found in both centers and at a VAF >5%, the remaining 3 variants were only found by a single center (2 with a VAF >5% and the remainder with a VAF <5%) (Supplemental Table S13). Two of the three variants were found by the centers performing the analysis with the HaloPlex technology, indicating that these mutations are unlikely to be false-positives in the single Illumina center and false-negatives in the second center is the more probable explanation.

Finally, the HaloPlex assay detected a total of 5 variants within the *EGR2* gene; 3 variants were found with a VAF >5% by both centers while 2 variants were found by both centers with a low VAF (<5%) (Supplemental Table S13). Taken collectively, the concordance for *EGR2* mutation detection reached 100% for the HaloPlex assay. In contrast, *EGR2* proved to be the most difficult gene in this study to sequence when using the Illumina assay. Four variants were detected using the Illumina panel, however, only one was found by both centers, which, surprisingly, concerned a low frequency variant (VAFs, 2.3% and 2.2%) (Supplemental Table S13). Notably, 2 of the variants found by Illumina were not present in the HaloPlex datasets. This dropout of amplicons for *EGR2* is supported by the coverage data (Supplemental Table S6). While reaching a definitive conclusion as to why this dropout occurred is not possible, drawing from our previous experience, reagents based on the older TSCA chemistry had a relatively short period within which they performed optimally i.e. while well covered regions maintained a high coverage when using reagents nearing or older than 3 months, the

coverage for problematic regions such as the *NOTCH1* PEST domain, *NFKBIE*, *EGR2* and certain regions within the *TP53* gene decreased dramatically.

Inter-laboratory variation in detecting mutations

We next assessed the reproducibility of targeted NGS by looking at the inter-laboratory variation in detecting mutations. As these assays were not specifically designed to detect variants with a very low VAF i.e. <5%, filtering was set such that a VAF of 5% had to be reached in at least one center for the variant to be included in the concordance counts. We found that 107/115 (93% concordance) mutations were consistently detected by all 6 participating centers (Supplemental Table S9). These variants could be segregated into two groups: (i) 87/107 (81%) variants were found at a frequency >5% in all 6 centers; and, (ii) 20/107 (18.7%) of the detected variants were bordering the 5% cut-off, and, with the exception of sample T5, their VAFs spanned a narrow range (1-12%) (Supplemental Table S9). This is illustrated in Figure 3 for variants detected within *TP53*, *SF3B1*, and *NOTCH1*. The remaining 8/115 (7%) variants were not detected by a single center i.e. the variants were successfully found in the sequencing data from 5/6 of the participating centers. Two of these variants went undetected in centers utilizing the HaloPlex technology and concerned frameshift insertions within the *ATM* gene (p.L1327fs (center 5) and p.V2201fs (center 6)) (Supplemental Table S9). Analysis of the raw sequencing data revealed that amplicon dropout leading to a gap in coverage resulted in these undetected *ATM* frameshift mutations. The final 6 mutations were present at a low frequency and were not found within the following centers sequencing datasets, center 4 (n=1), center 5 (n=2) and center 6 (n=3). These minor subclonal mutations were detected by the partner institute utilizing the same methodology, thus demonstrating the ability of the assay to detect and amplify in this genomic region.

Finally, a variant was deemed as a false-positive if it was only found in a single test center. Overall, four false-positives with a VAF \geq 5% were detected in 2 centers and analysis of these variants led to

the conclusion that it is highly probable that three of these four false-positive calls arose independent of the assay design or sequencing protocol. A full description of these variants is provided in the 'Interpretation of false-positive findings' section in Supplementary materials and in Supplemental Table 14.

High-sensitivity assay to confirm variant calling

To gain a better understanding of whether the low-frequency mutations were indeed true variants or false-positives arising from the library preparation or sequencing steps, the entire experiment was repeated using a custom-design HaloPlex high-sensitivity assay and gDNA from 38 of the CLL patients originally sequenced system (additional material was unavailable from 10 of the original 48 samples). The HaloPlex^{HS} system follows a similar workflow to the standard HaloPlex assay (described in Supplemental Materials and Methods), however a cardinal feature is the attachment of a unique molecular identifiers (UMI) to individual captured DNA molecules. During downstream analyses the UMI is used to collapse reads originating from the same molecule, thereby improving base calling accuracy and permitting accurate quantification of the mutant allele fraction by excluding PCR amplification bias and improving discrimination of variant nucleotides from background sequencing errors.

A median of 95.3% of bases within the targeted ROIs achieved at least 100x coverage (Supplemental Table S15). Of the 128 somatic variants identified during the comparative analysis, excluding 18 mutations in samples not re-sequenced, 103 variants were confirmed (VAF range, 0.5-100%) (Supplemental Table S9). No additional variants were identified in the repeat NGS data that were not identified previously, while a few low-frequency variants (VAF <5%) could not be verified. The high-sensitivity assay data also enabled us to investigate the variants found in only a single center (described above) and provided further evidence that variants detected by only a single center were indeed false-positives (Supplemental Table 10).

DISCUSSION

As the number of genes with diagnostic, prognostic, or predictive significance increases there is a need for robust assays that detect multiple alterations simultaneously. Whether for research, to provide a better understanding of the molecular mechanisms driving disease pathogenesis and evolution, or within clinical diagnostics, targeted gene panels in combination with NGS have evolved as a pragmatic cost-effective approach. These assays are also uniquely positioned to yield a volume of data that is more manageable and easier to interpret than the complex datasets generated by WES or WGS. As a consequence, targeted gene panels are becoming increasingly popular in the cancer molecular diagnostics arena and systems for implementing NGS to routine clinical practice need to be established. This is essential since, despite its numerous attractive attributes, mutational profiling by NGS can be challenging and involves workflows comprising several distinct parts, i.e. wet-bench components, bioinformatics analyses, and clinical interpretation of the variant calls. In addition, with the abundance of assays and sequencing platforms on the market, all with their own technical and methodological specifications, ensuring test-to-test reproducibility and inter-laboratory reliability is fundamental.

Several regulatory and professional organizations have published guidelines and best-practice recommendations to assist laboratories in the transition from the previous gold-standard methodology of Sanger sequencing for mutation detection to NGS.³³⁻³⁷ Whilst these reports focus on critical aspects of clinical gene sequencing such as documentation of the protocols and reporting the specificity and sensitivity of an assay, comparative analyses of different enrichment techniques and workflows are rare.^{38, 39} Additional inter-laboratory studies focusing on variant calling comparisons between different methodologies are therefore required. To aid in this endeavor, the present study compared the ability of different NGS technologies to accurately detect a spectrum of mutations with varying VAFs within genes of prognostic relevance in CLL, specifically focusing on the sensitivity and reproducibility of targeted gene panels. Aside from the particular gene panel utilized in the

various test centers, we kept parameters as uniform as possible including use of the same sequencing system, same patient samples etc. To avoid confounding results due to individual customization of bioinformatic pipelines at participating institutes i.e. combinations of aligners, variant callers, differences in filtering parameters and QC, bioinformatics was performed centrally.

A first critical step in our analysis was to compare the metrics of sequence coverage and depth. Variability in depth of coverage between centers using different technologies could be attributed to the differing size of the panel designs i.e. despite targeting the same regions, the HaloPlex design includes redundancy to ensure targets are covered even if an amplicon drops out, whereas the Multiplicom technology sequenced fewer samples per run. While incomplete coverage was not a consistent problem for any of the methods utilized, considerable differences in the coverage of *EGR2* and *NFKB1E* genes were observed, particularly when using the TSCA gene panel. Certain regions within these genes failed to amplify and had lower read depth in comparison to other genes targeted indicating that they were intrinsically more difficult to amplify. This may stem from the varying ability of the different probe and primer sets to anneal in these high GC regions, thus leading to a reduction in the efficiency of target capture or amplification. These difficulties may have been exacerbated by the decline in the TSCA chemistry overtime. As a final note on coverage, direct comparison of the sequencing depth obtained from panels designed with/without UMIs is not meaningful since the number of reads does not reflect the actual number of unique template gDNA molecules as many reads will be duplicates generated during PCR. While molecular barcodes do not prevent PCR duplication from occurring they do facilitate their removal and hence the overall coverage can appear lower albeit more accurate.

Next, the ability of disparate NGS methods to detect variants with varying allelic frequencies was compared and we observed a high degree of concordance and accuracy when performing pairwise analysis. Imposing an arbitrary VAF cut-off i.e. 10% or 5%, at the initial stage of the analysis could provide an inaccurate view of concordance as variants borderline of a threshold may appear to be

discrepant, hence we first looked at the agreement between variants irrespective of their VAF and yielded concordance rates of 96.2% (Multiplicom), 97.7% (Illumina) and 90% (HaloPlex). We next assessed the reproducibility of targeted NGS by looking at the inter-laboratory variation and found that 107/115 mutations (93% concordance) were consistently detected by all 6 participating centers. Another encouraging finding was the accuracy or similarity i.e. narrow range, amongst the VAFs recorded from all centers. Notable exceptions to the narrow range of VAFs pertained to *EGR2* and *NFKB1E*, which provided less homogeneous results, especially between different methodologies and, as mentioned above, these differences likely originate from the suboptimal coverage obtained in certain regions of these genes.

There is currently no data that definitively determines what VAF cut-off is clinically significant, indeed this cut-off may vary depending on the particular gene in question, and the clinical relevance of any gene mutation can only be deemed by a clinical trial demonstrating a significant association with outcome. That said, a VAF cut-off of 5 or 10% is currently widely used in clinical routine. In this study, several variants were observed at VAF's bordering the 5-10% threshold and hence would have been removed had a hard filtering step been applied. From a technical perspective, a plausible strategy to prevent clinically significant calls from being removed due to stringent cut-offs is to flag relevant variants as hotspots within the bioinformatic pipeline thus increasing the sensitivity by reducing the VAF threshold for that specific variant. Noted exceptions to this work-around would concern genes without defined hotspots i.e. tumor suppressor genes, within which, in theory, any mutation could be detrimental e.g. *TP53* and *ATM*. To circumnavigate the VAF/cut-off filtering issue in these instances, one solution would be to lower the technical cut-off of the method and avoid setting a hard cut-off for reporting and implementing a review and visualization step, thus removing the possibility of filtering out a potentially relevant mutation.

Owing to the heterogeneous subclonal composition of tumors together with ongoing clonal evolution, clinical samples are often fraught with somatic mutations that exist at low VAFs.^{26, 28, 29, 40} As evidence accumulates it is becoming apparent that these low-level variants, at least within certain

genes, may contribute to disease pathogenesis or serve as early indicators of resistance to specific therapies.^{21-24, 41-44} Despite data supporting the importance of somatic mutations with low-*VAF*, their detection by conventional NGS is challenging. Even at a high-read depth, NGS shows a marked decrease in the detection of somatic mutations with low *VAFs* and accurate separation of erroneous variants from true low-level variants cannot always be achieved by stringent filtering alone. We performed an additional round of testing using a high sensitivity assay containing UMIs as a comparator method. This decision was made for two reasons: (i) to obtain the most accurate estimate of *VAFs*; and, (ii) to confidently detect low-frequency variants. When exploring the validity of minor variants using the dataset generated by the high sensitivity assay as a validator we found that although low frequency variants were detected by all techniques, greater diversity was observed for mutations with *VAFs* <5%. While these results advocate for the use of amplicon-based approaches for mutation detection within a clinical or research setting, the need for stringent validation should not be underestimated and is essential prior to the implementation of assays into routine. To further help with this transition, ERIC has recently initiated a multi-center project specifically focusing on low-frequency *TP53* mutations with the aim to reach a consensus on whether the current cut-off of 10% for reporting *TP53* mutations should be lowered.⁴⁵

Our study is not without limitations, one of which relates to the fact that the mutational status of all genes to be analyzed was not known prior to their inclusion in the study. While desirable, this was not possible due to the difficulty in trying to obtain samples that had sufficient material to be distributed to all centers. However, this limitation was mitigated by performing an additional sequencing round using a high-sensitivity assay incorporating UMIs independent of the three gene panels being assessed. Another limitation related to the determination of the sensitivity level for each assay included in the study. Although our study was focused on the comparability and reproducibility of results obtained in different centers using various amplicon-based assays, commercial controls harboring variants and allelic frequencies that have been well-characterized, would have been beneficial in determining the reliable limits of detection (which can differ

depending on the specific variant). In addition, read counts from such controls can aid in the assessment of run variability. For the validation and use of NGS for a clinical laboratory service, sensitivity control samples with a range of VAFs for target regions should always be included to ensure that the validated lower limit of detection is maintained while controls harboring specific variants and types of variants verify assay performance.

In conclusion, the data herein provides strong evidence that distinct gene panel designs and workflows have high analytical specificity and sensitivity and are capable of making consistent variant calls. Importantly, this assessment of interlaboratory reliability is not only paramount in the molecular diagnostics setting but also impacts on research where combining data generated from centers globally is crucial to provide homogeneous, reliable results and ultimately improve our understanding of rarer gene mutations thus facilitating stratification of patients based on their molecular profiles. As NGS technologies continue to advance and assays extend beyond variant calling to include copy-number analysis and allow for the robust identification of clinically relevant low-frequency variants, diagnostic and research laboratories need to remain flexible so as to adapt in this dynamic era of precision medicine.

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. 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*. 2011;44(1):47-52.
4. Wang LL, Lawrence MS, Wan YZ, et al. SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *N Engl J Med*. 2011;365(26):2497-2506.
5. Puente XS, Bea S, Valdes-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519-524.
6. Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526(7574):525-530.
7. Sutton LA, Rosenquist R. Deciphering the molecular landscape in chronic lymphocytic leukemia: time frame of disease evolution. *Haematologica*. 2015;100(1):7-16.
8. Mansouri L, Sutton LA, Ljungstrom V, et al. Functional loss of I kappa B epsilon leads to NF-kappa B deregulation in aggressive chronic lymphocytic leukemia. *J Exp Med*. 2015;212(6):833-843.
9. Ljungstrom V, Cortese D, Young E, et al. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood*. 2016;127(8):1007-1016.
10. Young E, Noerenberg D, Mansouri L, et al. EGR2 mutations define a new clinically aggressive subgroup of chronic lymphocytic leukemia. *Leukemia*. 2017;31(7):1547-1554.
11. Zenz T, Eichhorst B, Busch R, et al. TP53 Mutation and Survival in Chronic Lymphocytic Leukemia. *J Clin Oncol*. 2010;28(29):4473-4479.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*. 2015;29(2):329-336.
20. Sutton LA, Young E, Baliakas P, et al. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica*. 2016;101(8):959-967.
21. Rossi D, Khiabani H, Spina V, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood*. 2014;123(14):2139-2147.
22. Malcikova J, Stano-Kozubik K, Tichy B, et al. Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. *Leukemia*. 2015;29(4):877-885.
23. Nadeu F, Delgado J, Royo C, et al. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood*. 2016;127(17):2122-2130.
24. Brieghel C, Kinalis S, Yde CW, et al. Deep targeted sequencing of TP53 in chronic lymphocytic leukemia: clinical impact at diagnosis and at time of treatment. *Haematologica*. 2019;104(4):789-796.
25. 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.

26. Landau DA, Carter SL, Stojanov P, et al. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell*. 2013;152(4):714-726.
27. Rasi S, Khiabani H, Ciardullo C, et al. Clinical impact of small subclones harboring NOTCH1, SF3B1 or BIRC3 mutations in chronic lymphocytic leukemia. *Haematologica*. 2016;101(4):e135-e138.
28. Nadeu F, Clot G, Delgado J, et al. Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia*. 2018;32(3):645-653.
29. Leeksa AC, Taylor J, Wu B, et al. Clonal diversity predicts adverse outcome in chronic lymphocytic leukemia. *Leukemia*. 2019;33(2):390-402.
30. 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.
31. Sutton LA, Ljungstrom V, Mansouri L, et al. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica*. 2015;100(3):370-376.
32. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-2760.
33. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol*. 2012;30(11):1033-1036.
34. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*. 2013;15(9):733-747.
35. Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Lab Med*. 2015;139(4):481-493.
36. Jennings LJ, Arcila ME, Corless C, et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2017;19(3):341-365.
37. Roy S, Coldren C, Karunamurthy A, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn*. 2018;20(1):4-27.
38. Haslam K, Catherwood MA, Dobbin E, Sproul A, Langabeer SE, Mills KI. Inter-Laboratory Evaluation of a Next-Generation Sequencing Panel for Acute Myeloid Leukemia. *Mol Diagn Ther*. 2016;20(5):457-461.
39. Hirsch B, Endris V, Lassmann S, et al. Multicenter validation of cancer gene panel-based next-generation sequencing for translational research and molecular diagnostics. *Virchows Arch*. 2018;472(4):557-565.
40. Rossi D, Khiabani H, Rasi S, et al. Small Subclones Harboring NOTCH1, SF3B1 or BIRC3 Mutations Are Clinically Irrelevant in Chronic Lymphocytic Leukemia. *Blood*. 2014;124(21):295.
41. Ahn IE, Underbayev C, Albitar A, et al. Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. *Blood*. 2017;129(11):1469-1479.
42. Blombery P, Anderson MA, Gong JN, et al. Acquisition of the Recurrent Gly101Val Mutation in BCL2 Confers Resistance to Venetoclax in Patients with Progressive Chronic Lymphocytic Leukemia. *Cancer Discov*. 2019;9(3):342-353.
43. Woyach JA, Furman RR, Liu TM, et al. Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. *N Engl J Med*. 2014;370(24):2286-2294.
44. Tausch E, Close W, Dolnik A, et al. Venetoclax resistance and acquired BCL2 mutations in chronic lymphocytic leukemia. *Haematologica*. 2019;104(9):e434-e437.
45. Malcikova J, Tausch E, Rossi D, et al. ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia-update on methodological approaches and results interpretation. *Leukemia*. 2018;32(5):1070-1080.

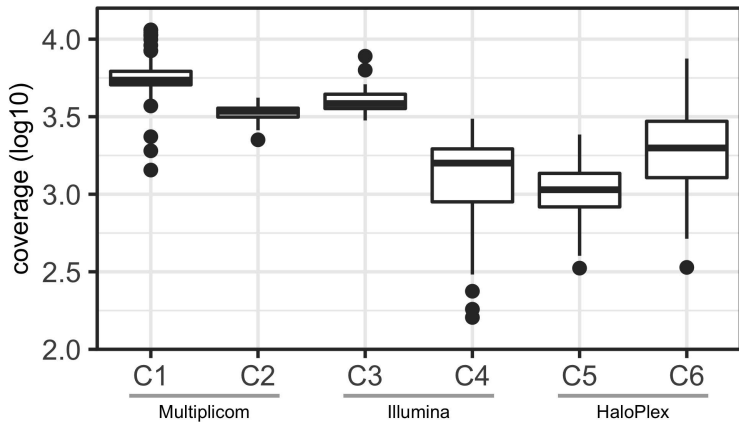
FIGURE LEGENDS

Figure 1: Gene panel coverage. Boxplots illustrating the coverage obtained for all samples and the percent of bases covered at 100x across all six test centers. As illustrated, high coverage was obtained for the majority of samples across all centers, however a few samples had less than 90% of bases within the targeted ROIs obtaining >100x coverage. Notably, the partner center utilizing the same technology obtained high coverage for these particular samples across all ROIs indicating that the low coverage arose from a local issue rather than an inherent technological or design issue. For the Multiplicom CLL MASTR kit, 24 samples were sequenced over two run and this reduction in samples/run combined with the smaller size of the panel was reflected in the increase in median coverage obtained compared to the other assays whereby all 48 samples were sequenced in a single run. C1-C6 represent the six participating centers: C1 & C2 (Multiplicom), C3 & C4 (Illumina TSCA) and C5 & C6 (HaloPlex).

Figure 2: Pairwise comparison of mutations detected by centers utilizing the Multiplicom, Illumina TSCA and HaloPlex assays. Mutation plots depicting the concordance and accuracy (narrow range) of the VAFs of variants found in this study. Variants are listed on the y axis and variants detected by each center are denoted by different shapes. The VAF scale is on the x-axis and a VAF difference of less than 10% between partner centers is marked in green while a VAF difference of greater than 10% between partner centers is marked in blue. VAF: variant allelic frequency.

Figure 3: *TP53*, *SF3B1*, and *NOTCH1* variants detected by the six centers. Heatmaps and boxplots illustrating mutations within *TP53*, *SF3B1*, and *NOTCH1*. Columns represent the 6 test centers (C1-C6) and rows correspond to a mutation found in a specific sample. Color coding indicates the type of mutation whereas the shading depicts the VAF (lighter to darker tones correspond to an increasing VAF). White boxes indicate that the variants were not detected by a particular center. *TP53* splice

site mutations resulted in the following nucleotide changes: c.96+1G>T, c.673-2A>T, c.783-2A>C and c.920-2A>G). VAF: variant allelic frequency. C1 & C2 (Multiplicom), C3 & C4 (Illumina TSCA) and C5 & C6 (HaloPlex).



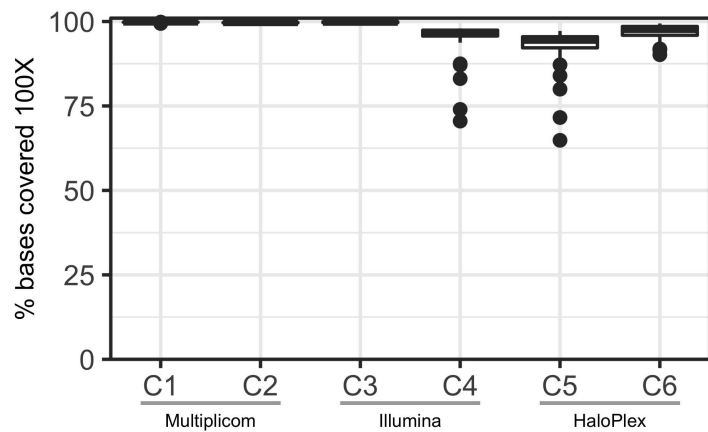
Multiplicom

Illumina

HaloPlex

<i>Median</i>	5469	3426.4	3834.1	1540.1	1062.4	1953.1
<i>Max</i>	13588.7	4193.6	7760.8	3067.8	2426.7	7496
<i>Min</i>	1431.7	2243.9	2990.5	160.5	333.8	337.5

Absolute values



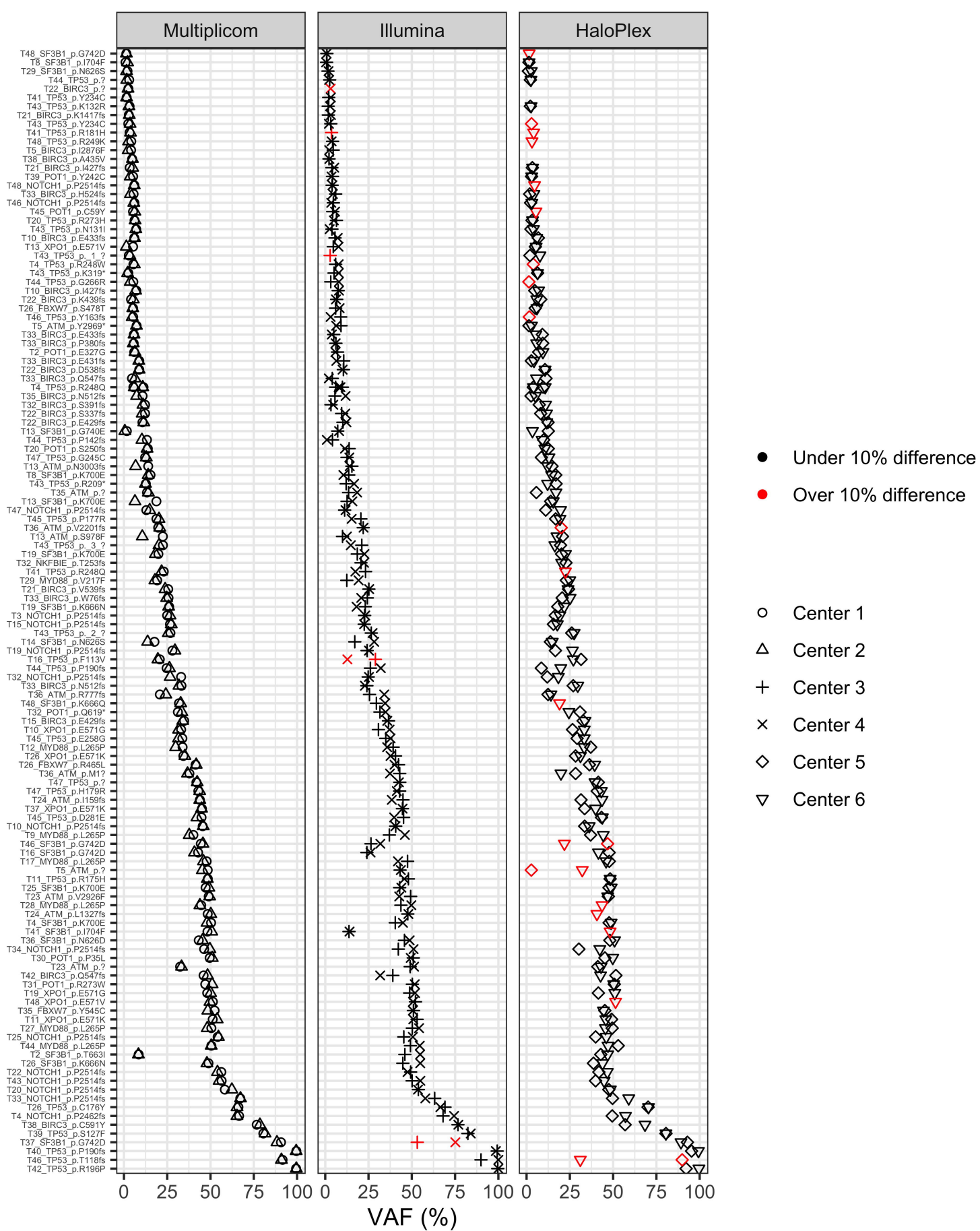
Multiplicom

Illumina

HaloPlex

<i>Median</i>	99.8	99.6	99.7	96.5	94.2	97.6
<i>Max</i>	99.8	99.7	100	98	97.2	99.4
<i>Min</i>	99.4	99.5	99.6	70.5	64.8	90.2

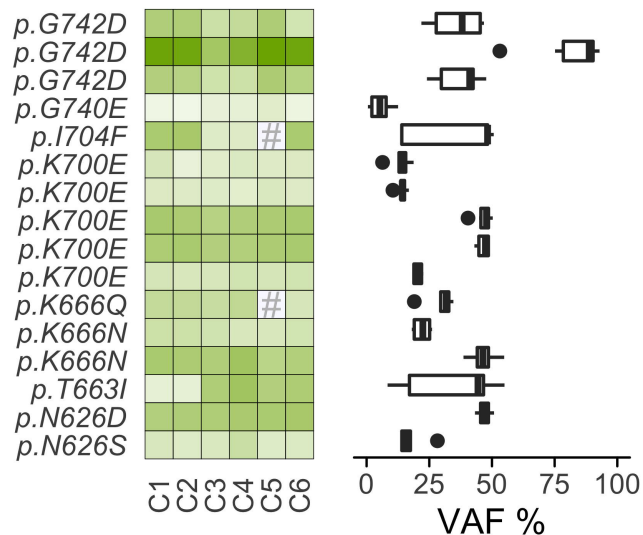
Percentages



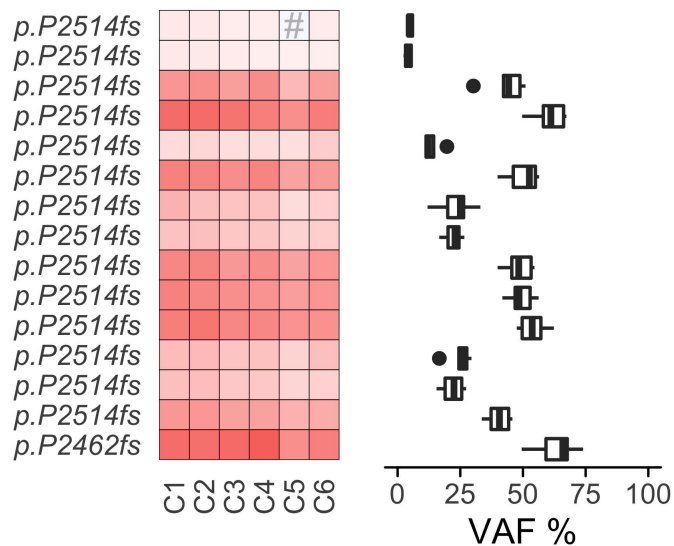
TP53



SF3B1



NOTCH1



■ Frameshift
 ■ Nonsynonymous
 ■ Splice variant
 ■ Stopgain

VAF % 0 100