

# Clinical-grade Validation of Whole Genome Sequencing Reveals Robust Detection of Low-Frequency Variants and Copy Number Alterations in CLL

Jenny Klintman<sup>1,2\*</sup>, Katerina Barmpouti<sup>1\*</sup>, Samantha JL Knight<sup>3</sup>, Pauline Robbe<sup>1,4</sup>, H       Dreau<sup>1</sup>, Ruth Clifford<sup>1,4,5</sup>, Kate Ridout<sup>2</sup>, Adam Burns<sup>1,2</sup>, Adele Timbs<sup>1</sup>, David Bruce<sup>1,2</sup>, Pavlos Antoniou<sup>6</sup>, Alona Sosinsky<sup>6</sup>, Jennifer Becq<sup>7</sup>, David Bentley<sup>7</sup>, Peter Hillmen<sup>8</sup>, Jenny Taylor<sup>3</sup>, Mark Caulfield<sup>6#</sup> and **Anna Schuh**<sup>1,2#</sup>

<sup>1</sup>*Molecular Diagnostic Centre, Oxford University Hospitals, Oxford, United Kingdom*

<sup>2</sup>*Department of Oncology, University of Oxford, Oxford, United Kingdom*

<sup>3</sup>*NIHR Biomedical Research Centre, Oxford and Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK*

<sup>4</sup>*Nuffield Department of Laboratory Sciences, University of Oxford, Oxford, United Kingdom*

<sup>5</sup>*Haematology Department, University Hospital Limerick, Limerick, Ireland*

<sup>6</sup>*Genomics England, London, UK*

<sup>7</sup>*Illumina, Chesterford, UK*

<sup>8</sup>*St James' Institute of Oncology, St James' University Hospital, Leeds, United Kingdom;*

\*These authors contributed equally to this article.

#These authors contributed equally to this article.

## Introduction

In chronic lymphocytic leukaemia (CLL) deletions and/or mutations of *TP53*, including subclonal mutations are clinically actionable <sup>1</sup>, and associated with short overall survival and poor response to conventional chemo immunotherapy <sup>2</sup>. Routine diagnostic screening for TP53 alterations is recommended using a combination of Fluorescent in situ hybridisation (FISH) and Sanger Sequencing <sup>2</sup>. Patients with del17p/TP53 mutations should be considered for targeted therapy, inclusion into clinical trials or consolidation with allogeneic stem cell transplantation <sup>3</sup>.

Chromosomal aberrations have long been known to impact on CLL prognosis <sup>4</sup>. Genome wide sequencing efforts have revealed the complex genomic architecture of CLL <sup>5</sup> and identified a growing number of driver genes <sup>6</sup>, and also mutation hotspots in non-coding regulatory regions <sup>7</sup>, associated with poor prognosis, for which targeted therapies are under development (i.e. *NOTCH1*, *SF3B1* and *ATM*).

Promising results of the feasibility and clinical utility of using whole genome sequencing (WGS) for clinical diagnosis of congenital diseases have been reported <sup>8,9</sup>, but the application of WGS in clinical diagnostics of blood cancers has not been evaluated. The 100,000 Genomes Project in the United Kingdom aims to develop a future-proof diagnostics platform by introducing WGS into clinical practice <sup>10</sup>. Here, we present a clinical-grade validation of results obtained from WGS as part of the CLL Genomics England Pilot. WGS was used for detection of single nucleotide variants (SNVs), small insertions and deletions (InDels) and CNAs (copy number alterations). These were validated using a

clinically accredited targeted gene panel or gold standard FISH technique and genome-wide high-resolution arrays, respectively.

## **Materials and Methods**

### *Patient Material*

Tumour (peripheral blood) and germline (GL) (saliva) samples from 64 patients with CLL recruited into the ADMIRE (n=33) and ARCTIC (n=31) studies and who were consented for WGS, were randomly selected (total n=250). (REC 09/H1306/54) <sup>11,12</sup>.

### *Whole Genome Sequencing (WGS)*

We performed WGS on the HiSeqX (Illumina), as previously described <sup>13</sup>. Sequencing reads were aligned to the Human Genome Assembly GRCh37 using the ISAAC Whole Genome Sequencing v4 and we used Strelka 2.4.7 for detection of somatic SNV and InDels, and Manta 0.28.0 and Canvas 1.3.1 for detection of somatic CNAs (Illumina). Variant filtering was based on predicted functional consequence or pathogenicity by *in silico* algorithms SIFT and PolyPhen, COSMIC annotation and a variant allele frequency (VAF) of  $\geq 4\%$ . Variants discrepant against TGS outcome were visually inspected using Integrated Genomics Viewer (IGV) <sup>14</sup>.

### *Targeted Resequencing (TGS)*

TGS was performed as previously described <sup>15</sup>.

### *Genome-wide SNP Platform hybridizations and analysis (Array)*

Illumina HumanOmni2.5-8 Beadships were used according to the manufacturer's instructions. Data was analysed as previously described <sup>16</sup>.

### *Fluorescent in situ hybridisation (FISH)*

As previously described <sup>11,12</sup>.

### *Statistical Analysis*

Concordance of VAFs reported by WGS and TGS was evaluated with Pearson's correlation test. An outlier was defined as 1.5 x upper quartiles of mean % coefficient of variance of paired VAF <sup>17</sup>. Kaplan and Meier survival analysis with log-rank test was applied for time to progression (TTP) versus genomic complexity using GraphPad Prism V7.0a (GraphPad Software Inc., La Jolla, CA, USA). A *p*-value <0.05 was considered statistically significant.

## Results and Discussion

WGS was performed on paired tumour and GL control samples. An average reading depth of 104.6x (SD 11.8x) for tumour samples and 33.5x (SD 3.2x) for paired GL controls was reached; a coverage previously shown to ensure adequate sequencing sensitivity (down to 10% VAF)<sup>18</sup>. Eleven CLL driver genes were evaluated (Figure 1), and 71 variants (VAF of  $\geq 4\%$ ) were detected by WGS in 41 subjects (n=41/64; 64.1%). Without visual inspection, 100% concordance between the technologies was shown for *EGR2*, *RPS15* and *XPO1* and in total for 55 variants (n=55/82; 67.1%). Reasons for discrepancies are shown in Figure 1a. Discordant variants were investigated using IGV<sup>14</sup> resulting in an increase of concordant calls to 96.3% (79/82), clearly proving the WGS coverage to be comparable to TGS (Supplementary Table 1). Reassuringly, evaluation of the VAF reported by WGS and TGS showed a high concordance for both sub-clonal variants and low-frequency *TP53* variants ( $\geq 2\%$  VAF); far lower than previously reported for WGS (Figure 1b-c)<sup>19,20</sup>.

WGS identified 52 CNAs of previously reported clinical significance in CLL. Verification by FISH showed a concordance of 86.7% (n=26/30; Fig 2a) and Array showed a concordance with WGS of 92.9% (n=52/56; Fig 2b), confirming adequate sensitivity for CNA (losses and gains) detection by WGS (Supplementary Table 2).

Genomic complexity (GC) has been associated with adverse prognosis in CLL<sup>21</sup>, even in the era of targeted therapy<sup>22</sup>. GC assessment by conventional karyotyping is technically difficult, labour intensive and costly. Array-based studies have reported a significantly worse prognosis in CLL patients with GC defined by at least 3 or 4 CNAs<sup>16,21</sup>. We evaluated GC by expanding our validation to genome-wide CNA detection (including acquired CNAs >50kB) by Array and WGS (n=60 subjects). Using Array, a total of 228 CNAs was detected versus 192 by WGS corresponding to an overall concordance of 84.2% (n=192/228). GC (defined as detection of  $\geq 4$  CNAs,<sup>23</sup>) was concordantly detected by the techniques in 94.1% of the patients (Figure 2c). GC determined by Array was significantly associated with shorter TTP ( $p=0.029$ ) and the same trend was shown for GC determined by WGS ( $p=0.07$ ) with discrepant outcome in one subject (Figure 2d). However, of note is an overrepresentation of *TP53*/del17p lesions in cases with GC (n=6/7, 85.7%) potentially explaining the association with TTP.

Our results show that WGS provides technically accurate detection of SNVs, InDels and CNAs. Discrepancies were in the majority of cases explained by false positives in TGS and a lack of consensus of the bioinformatics pipelines used for TGS, WGS and Array, a problem previously highlighted by others<sup>18</sup> that calls for stringent standardisation of bio-informatics algorithms and filters. Concordance rates between the two techniques improved from 67.1% to 96.3% after visual inspection of all discordant variants, arguing for a requirement to complement automated sequencing analysis with skilled evaluation by a trained specialist in order to get clinically valid results. TGS cancer panels have proven to be both time- and cost-effective in routine care and clinical trials<sup>24</sup>. However, the increased availability of targeted therapies calls for an ever increasingly comprehensive panel design

to cover a larger number of targets and types of mutations including low frequency variants<sup>25</sup>, making clinical grade validation that is compliant with accreditation standards cumbersome and cost- and time-consuming<sup>19</sup>. Moreover, an increasing body of evidence suggests genomic complexity as important prognostic factors<sup>7,21,23</sup> that cannot be revealed by targeted approaches. In addition, the overall cost of large panels or WES has become comparable to WGS<sup>19</sup>, not least when considering that separate analyses for chromosomal aberrations will not be needed when using WGS<sup>8</sup>. Finally, WGS provides a future-proof approach to diagnostics as the results can be used for biomarker discovery.

In conclusion, we present the first clinical-grade validation of results obtained from WGS as part of the CLL Genomics England Pilot. We show that WGS confidently detects all types of clinically relevant mutations. Our results confirm that WGS can provide clinically useful and comprehensive genomic characterisation for clinical trials, drug discovery and ultimately precision medicine.

## **Acknowledgement**

This work was supported by the NIHR Biomedical Research Centre, Oxford with funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. In addition, the publication presents independent research commissioned by the Health Innovation Challenge Fund (HICF-1009-026), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this publication are those of the authors and not necessarily those of the Wellcome Trust or the Department of Health. The project was also supported by the Wellcome Trust (090532/Z/09/Z), (076113), (085475). JK receives post-doctoral grants from The Royal Swedish Academy of Sciences, the Swedish Medical Association, Lund University Sweden, the Royal Physiographic Society of Lund, Sweden and the Swedish National Health Services (ALF).

## **Authorship Contributions**

JK, KB and AS designed the research, analyzed the data and JK wrote the paper; JK, KB, SJLK, PR, HD, RC, KR, AB, AT, DB, PA, AS, JB, DB, PH, JT, MC and AS performed the research and/or contributed patient samples and associated data. All authors read and agreed to the final version of the manuscript.

## **Disclosure of Conflict of Interest**

## References

1. Zenz T, Eichhorst B, Busch R, et al. TP53 Mutation and Survival in Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology*. 2010;28(29):4473-4479.
2. Pospisilova S, Gonzalez D, Malcikova J, et al. ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia. *Leukemia*. 2012;26(7):1458-1461.
3. Eichhorst B, Robak T, Montserrat E, et al. appendix 6: Chronic lymphocytic leukaemia: eUpdate published online September 2016 (<http://www.esmo.org/Guidelines/Haematological-Malignancies>). *Ann Oncol*. 2016;27(suppl 5):v143-v144.
4. Döhner H, Stilgenbauer S, Benner A, et al. Genomic Aberrations And Survival In Chronic Lymphocytic Leukemia. *The New England Journal of Medicine*. 2000;343(26):1910-1916.
5. 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.
6. Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152(4):714-726.
7. Puente XS, Bea S, Valdes-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519-524.
8. Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med*. 2017.
9. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet*. 2015;47(7):717-726.
10. The 100,000 Genomes Project Protocol v3, Genomics England. <https://www.genomicsengland.co.uk/>.
11. Howard DR, Munir T, McParland L, et al. Results of the randomized phase IIB ARCTIC trial of low-dose rituximab in previously untreated CLL. *Leukemia*. 2017.
12. Munir T, Howard DR, McParland L, et al. Results of the randomized phase IIB ADMIRE trial of FCR with or without mitoxantrone in previously untreated CLL. *Leukemia*. 2017.
13. Robbe P, Ridout P, Becq J, et al. Identifying High-Risk CLL to Predict Early Relapse after FCR Based Treatment Using Whole Genome Sequencing: First Results from the Genomics England CLL Pilot. 58th American Society of Hematology Congress. Vol. 128. San Diego, CA, USA The American Society of Hematology; 2016:2022.
14. Robinson J, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology*. 2011;29(1):24-26.
15. Guieze R, Robbe P, Clifford R, et al. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood*. 2015;126(18):2110-2117.
16. Knight SJ, Yau C, Clifford R, et al. Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. *Leukemia*. 2012;26(7):1564-1575.
17. Renze J. "Outlier." From MathWorld--A Wolfram Web Resource, created by Eric W. Weisstein. ; 2017.

18. Alioto TS, Buchhalter I, Derdak S, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat Commun*. 2015;6:10001.
19. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet*. 2013;14(4):295-300.
20. Frampton MG, Fichtenholtz A, Otto G, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology*. 2013;31(11):1023-1031.
21. Ouillet P, Collins R, Shakhan S, et al. Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. *Blood*. 2011;118(11):3051-3061.
22. Anderson MA, Tam C, Lew TE, et al. Clinicopathological features and outcomes of progression of CLL on the BCL2 inhibitor venetoclax. *Blood*. 2017;129(25):3362-3370.
23. Yu L, Kim HT, Kasar S, et al. Survival of Del17p CLL Depends on Genomic Complexity and Somatic Mutation. *Clin Cancer Res*. 2017;23(3):735-745.
24. Hamblin A, Wordsworth S, Fermont JM, et al. Clinical applicability and cost of a 46-gene panel for genomic analysis of solid tumours: Retrospective validation and prospective audit in the UK National Health Service. *PLoS Med*. 2017;14(2):e1002230.
25. Konig K, Peifer M, Fassunke J, et al. Implementation of Amplicon Parallel Sequencing Leads to Improvement of Diagnosis and Therapy of Lung Cancer Patients. *J Thorac Oncol*. 2015;10(7):1049-1057.

## Figure Legends

**Figure 1 SNV and InDel analysis.** Comparison of detection of SNV and InDels by WGS versus TGS. Analyses were based on eleven CLL driver genes listed in Figure 1b (MED12 and MYD88 also included in gene list, but were not concordantly detected and therefore not in Fig 1b).

a) Reasons for discrepancies between the outcome of the WGS versus TGS platforms *prior to* IGV analysis. After IGV control three variants remained discrepant: two because they were not covered by the TGS design and one due to low total reads by WGS (i.e. below filtering threshold). Additional information to be found in Supplementary Table 1.

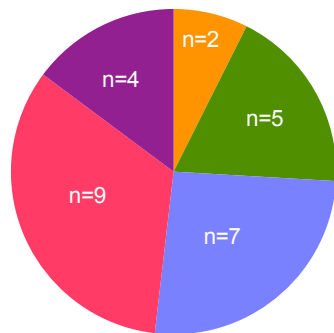
b) Correlation of VAFs (%) reported by TGS versus WGS for all variants with complete VAF data (n=79).

c) Correlation of VAFs (%) reported by TGS versus WGS for low-frequency *TP53* variants (n=7).

**Figure 2 CNA analyses.** Detection of CNAs of previously reported clinical significance are shown in Figures 2a and 2b. a) Frequency (n) and concordance (n, %) comparing detection by WGS and FISH, and b) Frequency (n) and concordance (n, %) comparing detection by WGS and Array. Detection of genomic complexity was evaluated taking all CNAs above 50 kb into consideration. c) Number of subjects detected with genomic complexity (GC) defined by a cut-off of  $\geq 4$  CNAs<sup>23</sup>. d) Association of GC to time to progression determined by Array (upper graph) versus WGS (lower graph). Additional information to be found in Supplementary Table 2.

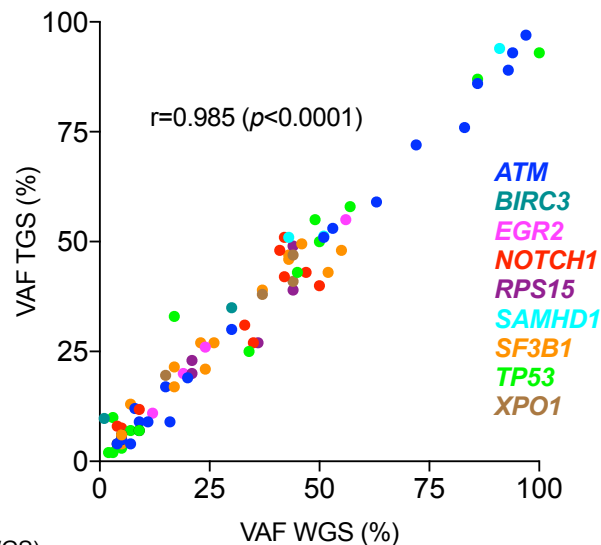
**Figure 1**

a) Reasons for discrepancies between WGS and TGS platforms



- Mutation position not covered by targeted panel
- Germline (i.e. filtered out in normal-tumour subtraction (WGS))
- Not called by TGS bioinformatics pipeline
- Number of reads below calling or filtering threshold (WGS)
- Number of reads below calling or filtering threshold (TGS)

b) All variants, n=79



c) Low-frequency *TP53* variants (n=7)

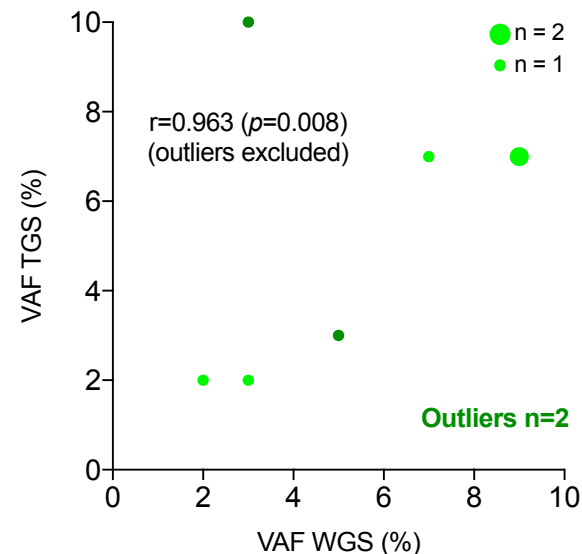
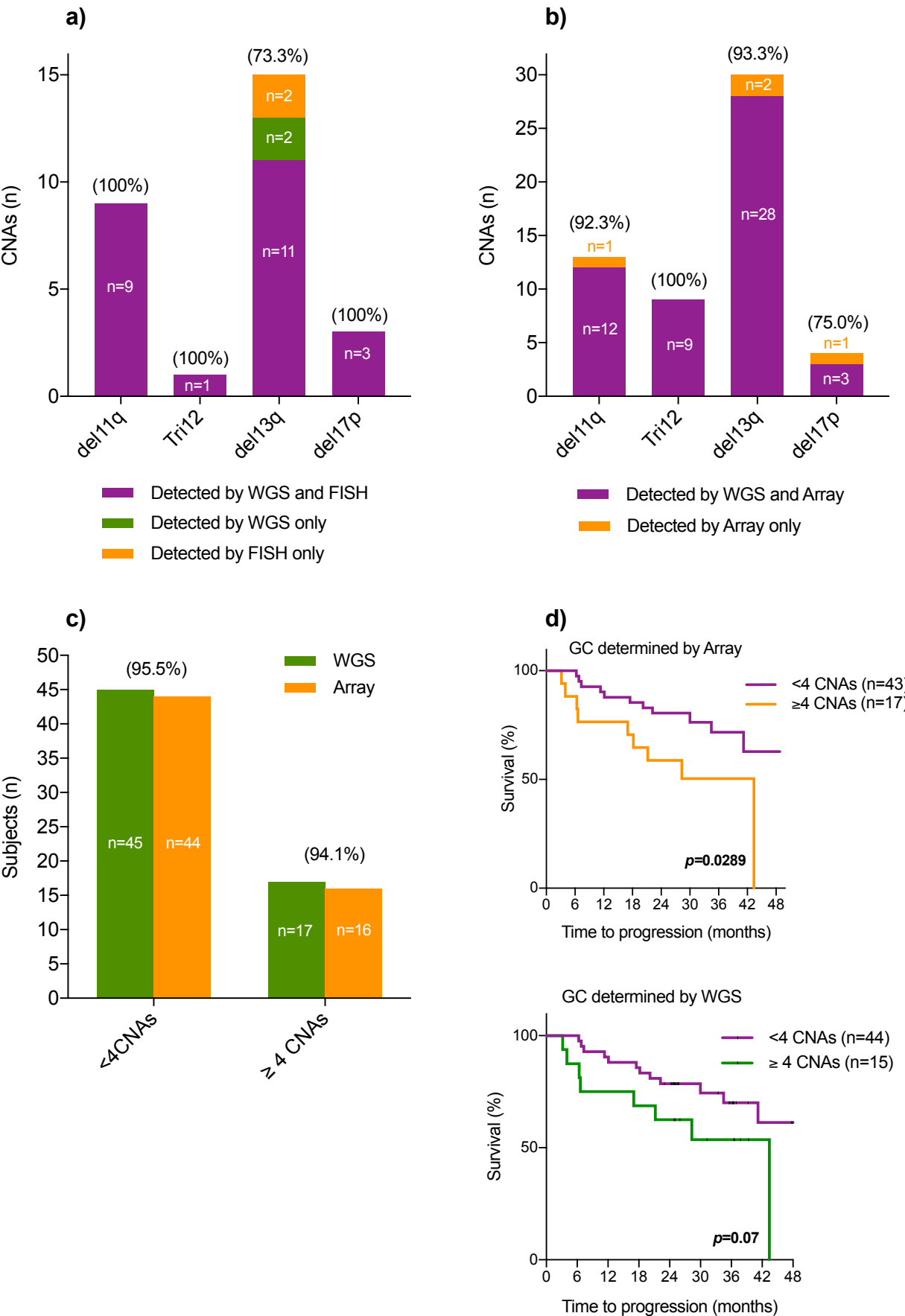




Figure 2



## **Supplementary Method Description**

### *Whole Genome Sequencing (WGS)*

DNA from tumour and germline (GL) samples was prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina Inc., San Diego, CA, USA). Sequencing was performed using the Illumina HiSeq X platform with a read depth threshold of 30x for GL samples and 75x for tumour samples (Illumina Inc., San Diego, CA, USA). For sequence alignment and variant calling analysis the ISAAC Whole Genome Sequencing v4<sup>1</sup> was used and tumour-normal subtraction was performed using Strelka-2.4.7<sup>2</sup>. Sequencing reads were aligned to the Human Genome Assembly GRCh37 (Ensembl database version 81<sup>3</sup>). SNVs and InDels passing a predefined quality score were annotated using Nirvana-1.3.5.2 against Ensembl database v81 and COSMIC database v73<sup>3,4</sup>. Somatic structural variations were analysed with Manta-0.28.0<sup>5</sup> and somatic CNA analysis was undertaken using Canvas-1.3.1<sup>6</sup> and Manta-0.28.0<sup>5</sup>. Variant filtering was based on predicted functional consequence, predicted pathogenicity by *in silico* algorithms SIFT and PolyPhen, COSMIC annotation and a variant allele frequency (VAF) of  $\geq 4\%$ . For the purpose of this study only variants in genes included in a pre-defined gene list of 11 CLL driver genes were selected for further analysis and for comparisons towards the outcome of TGS (Figure 1). Somatic variants discrepantly called by WGS versus TGS visual inspection using Integrative Genomics Viewer (IGV) was undertaken<sup>7,8</sup>.

### *Targeted Resequencing (TGS)*

A TruSeq Custom Amplicon (TSCA, Illumina, Inc., San Diego, CA, USA) panel was designed, targeting mutational hotspots in nine CLL driver genes. For samples in which an EGR2 and/or a RPS15 variant was detected with WGS, re-sequencing was performed using an expanded targeted panel including eleven driver genes (see Figure 1). Libraries were created from 250ng of DNA and 2x150bp paired-end sequencing performed using the MiSeq platform (Illumina, Inc., San Diego, CA, USA). Initial alignment and variant calling analysis was performed with the TSCA Workflow of MiSeq Reporter v.2.4 (Illumina) using the default settings. A second alignment followed by variant calling using the packages Stampy (v. 1.0.22)<sup>9</sup> and Platypus (v.0.5.1)<sup>10</sup> was used to detect additional InDels undetected by MiSeq Reporter. High confidence functional sequence variants were identified by annotating and filtering on quality, predicted functional consequence, predicted pathogenic by *in silico* algorithms SIFT, PolyPhen previously published in CLL literature, or in COSMIC database (v73), or absent from 1000 genomes project<sup>4,11,12</sup>. All mutations with a variant allele frequency (VAF)  $\geq 5\%$  and below 1% VAF for TP53 variants and with adequate coverage were included for analysis. All variants were confirmed by visual inspection in IGV<sup>7,8</sup>.

### *Genome-wide SNP Platform hybridizations and analysis*

Illumina HumanOmni2.5-8 Beadchip hybridisations were performed according to manufacturer's protocols. The data were processed using GenomeStudioV2009.2 (Illumina, Inc., San Diego, CA, USA) and analysed using Nexus 6.1 Discovery Edition (BioDiscovery, Inc., El Segundo, CA, USA), as

previously described. All copy number events and regions of cnLOH were confirmed visually within the Nexus software<sup>13</sup>.

### **References for Supplementary Method Description**

1. Racz C, Petrovski R, Saunders CT, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*. 2013;29(16):2041-2043.
2. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.
3. Ensembl Genome Browser.
4. The Catalogue Of Somatic Mutations In Cancer (COSMIC).
5. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220-1222.
6. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics*. 2016;32(15):2375–2377.
7. Robinson J, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology*. 2011;29(1):24-26.
8. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-192.
9. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res*. 2011;21(6):936-939.
10. Rimmer A, Phan H, Mathieson I, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet*. 2014;46(8):912-918.
11. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
12. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013;34(1):57-65.
13. Knight SJ, Yau C, Clifford R, et al. Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. *Leukemia*. 2012;26(7):1564-1575.

### **Legends for Supplementary Tables**

**Supplementary Table 1** Single nucleotide variants (SNVs), insertions and deletions (InDels) detected by Whole genome sequencing (WGS) and Targeted sequencing (TGS). Table shows number of detected variants *before* and *after* visual inspection. Comparisons between the techniques were made based on a gene list of 11 previously reported CLL driver genes (see Figure 1). Lower part of the table lists the discrepant variants and reasons for discrepancies. Visual inspection using IGV was undertaken for all discrepant calls resulting in an increased concordance rate from 67.1% to 96.3%. Three variants remained discordant after visual inspection: two because they were not covered by the TGS design and one due to low total reads by WGS.

**Supplementary Table 2** Samples from a total of 60 subjects were previously analysed using Genome-wide SNP Array hybridizations (Array). Calls from the Array data set were intersected with the WGS dataset. In total 228 CNAs were detected with Array and 192 with WGS. Table shows types of CNAs and concordance between the techniques

**Supplementary Table 1**

	WGS n (%)		TGS* n (%)		Concordant variants WGS and TGS n (%)*		Total variants detected by WGS and/or TGS n (%)
	BEFORE visual inspection	AFTER visual inspection	BEFORE visual inspection	AFTER visual inspection	BEFORE visual inspection	AFTER visual inspection	
Total	71 (83.5)	84 (98.8)	69 (84.1)	80 (97.6)	55 (67.1)	79 (96.3)	85
<i>Gene</i>							
ATM	16 (84.2)	19 (100)	14 (73.7)	19 (100)	11 (57.9)	19 (100)	19
BIRC3	3 (75)	4 (100)	2 (50)	4 (100)	1 (25)	4 (100)	4
EGR2	5 (100)	5 (100)	4 (100)*	4 (100)*	4 (100)*	4 (100)*	5
MED12	1 (100)	1 (100)	0 (0)	0 (0)	0 (0)	0 (0)	1
MYD88	1 (100)	1 (100)	0 (0)	0 (0)	0 (0)	0 (0)	1
NOTCH1	7 (70)	10 (100)	10 (100)	10 (100)	7 (70)	10 (100)	10
RPS15	8 (100)	8 (100)	6 (100)*	6 (100)*	6 (100)*	6 (100)*	8
SAMHD1	2 (66.6)	3 (100)	3 (100)	3 (100)	2 (66.6)	3 (100)	3
SF3B1	14 (100)	14 (100)	12 (85.7)	14 (100)	12 (85.7)	14 (100)	14
TP53	10 (62.5)	15 (93.8)	14 (87.5)	16 (100)	8 (50)	15 (93.8)	16
XPO1	4 (100)	4 (100)	4 (100)	4 (100)	4 (100)	4 (100)	4

\* DNA not available for TGS confirmation of EGR2 (n=1) and RPS15 (n=2) variants. N=82 variants available for evaluation of concordance.

<i>Gene</i>	Variant (aa)**	Detected by	Reasons for discordant calls
ATM	Y1475C	TGS	Germline variant. Subtracted by WGS bioinformatic pipeline
ATM	S1691R	TGS	Germline variant. Subtracted by WGS bioinformatic pipeline
ATM	G2765S	TGS	Germline variant. Subtracted by WGS bioinformatic pipeline
ATM	P2842L	WGS	Number of reads below calling or filtering threshold (TGS)
ATM	Y2437N	WGS	Not called by the TGS bioinformatic pipeline
ATM	c.3077+1G>T	WGS	Number of reads below calling or filtering threshold (TGS)
ATM	Y2755LfsTer2	WGS	Not called by the TGS bioinformatic pipeline
ATM	c.3077+1G>T	WGS	Not called by the TGS bioinformatic pipeline
BIRC3	Q547fs	TGS	Number of reads below calling or filtering threshold (WGS)
BIRC3	Q547NfsTer21	WGS	Number of reads below calling or filtering threshold (TGS)
BIRC3	Q552Ter4	WGS	Not called by the TGS bioinformatic pipeline
MED12	Q33K	WGS	Amplicon not covering variant in this sample
MYD88	S219C	WGS	Exon not included in TGS gene panel
NOTCH1	P2514RfsTer	TGS	Number of reads below calling or filtering threshold (WGS)
NOTCH1	P2514RfsTer	TGS	Number of reads below calling or filtering threshold (WGS)
NOTCH1	P2514RfsTer	TGS	Number of reads below calling or filtering threshold (WGS)
SAMHD1	D585N	TGS	Germline variant. Subtracted by WGS bioinformatic pipeline
SF3B1	K666N	WGS	Number of reads below calling or filtering threshold (TGS)
SF3B1	G742N	WGS	Not called by the TGS bioinformatic pipeline
TP53	77_77del	TGS	Number of reads below calling or filtering threshold (WGS)
TP53	V41L	TGS	Number of reads below calling or filtering threshold (WGS)
TP53	R116Q	TGS	Number of reads below calling or filtering threshold (WGS)
TP53	R116G	TGS	Number of reads below calling or filtering threshold (WGS)
TP53	G578T	TGS	Germline variant. Subtracted by WGS bioinformatic pipeline
TP53	C238S	TGS	Number of reads below calling or filtering threshold (WGS)
TP53	Y163C	WGS	Not called by the TGS bioinformatic pipeline
TP53	G279E	WGS	Not called by the TGS bioinformatic pipeline

\*\* Reference genome hg19

***Supplementary Table 2***

Type of copy number (CN) aberration	N (%) detected by Array (N tot=228)	N (%) detected by WGS (N tot=192)	Concordance (%)
CN Gain	68 (29.8)	58 (30.2)	85.3
cn loss of heterozygosity	14 (6.1)	5 (2.6)	35.7
CN Loss	143 (62.7)	126 (65.6)	88.1
Homozygous loss	3 (1.3)	3 (1.6)	100