



UNIVERSITY OF LEEDS

This is a repository copy of *An ultra-deep sequencing strategy to detect sub-clonal TP53 mutations in presentation chronic lymphocytic leukemia cases using multiple polymerases.*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/96412/>

Version: Accepted Version

---

**Article:**

Worrillow, L, Baskaran, P, Care, MA et al. (9 more authors) (2016) An ultra-deep sequencing strategy to detect sub-clonal TP53 mutations in presentation chronic lymphocytic leukemia cases using multiple polymerases. *Oncogene*, 35 (40). pp. 5328-5336. ISSN 0950-9232

<https://doi.org/10.1038/onc.2016.73>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

**An ultra- deep sequencing strategy to detect sub-clonal *TP53* mutations in presentation chronic lymphocytic leukemia cases using multiple polymerases**

Lisa Worrillow,<sup>1</sup> Praveen Baskaran,<sup>2</sup> Matthew A. Care,<sup>1</sup> Abraham Varghese,<sup>3</sup> Talha Munir,<sup>3</sup> Paul A. Evans,<sup>3</sup> Sheila J. O'Connor,<sup>3</sup> Andrew Rawstron,<sup>3</sup> Lee Hazelwood,<sup>4</sup> Reuben M. Tooze,<sup>1</sup> Peter Hillmen,<sup>1</sup> Darren J. Newton<sup>1</sup>

<sup>1</sup>Section of Experimental Haematology, Leeds Institute of Cancer and Pathology, University of Leeds, United Kingdom; <sup>2</sup>Department of Evolutionary Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany; <sup>3</sup>Haematological Malignancy Diagnostic Services, Leeds Teaching Hospitals National Health Services Trust, Leeds, United Kingdom; <sup>4</sup>Cancer Research UK, Ling Ka Shing Centre, University of Cambridge, Cambridge, United Kingdom

Corresponding author: [D.J.Newton@leeds.ac.uk](mailto:D.J.Newton@leeds.ac.uk)

Telephone: +44113 3438439

Fax: +44113 2068177

**Conflict-of-interest disclosure:** The authors declare no competing financial interests.

**Running Title:** Deep Sequencing of Mutant *TP53* in CLL

**Key Words:** *TP53* mutation, deep sequencing, chronic lymphocytic leukemia

## Abstract

Chronic lymphocytic leukemia (CLL) is the most common clonal B-cell disorder characterised by clonal diversity, a relapsing and remitting course, and in its aggressive forms remains largely incurable. Current frontline regimes include agents such as fludarabine, which act primarily via the DNA damage response pathway. Key to this is the transcription factor p53. Mutations in the *TP53* gene, altering p53 functionality, are associated with genetic instability, and are present in aggressive CLL. Furthermore, the emergence of clonal *TP53* mutations in relapsed CLL, refractory to DNA damaging therapy, suggests that accurate detection of sub-clonal *TP53* mutations prior to and during treatment may be indicative of early relapse. In this study we describe a novel deep sequencing workflow using multiple polymerases to generate sequencing libraries (MuPol-Seq), facilitating accurate detection of *TP53* mutations at a frequency as low as 0.3%, in presentation CLL cases tested. As these mutations were mostly clustered within the regions of *TP53* encoding DNA binding domains, essential for DNA contact and structural architecture, they are likely to be of prognostic relevance in disease progression. The workflow described here has the potential to be implemented routinely to identify rare mutations across a range of diseases.

## Introduction

Cancer cells evolve over time into genetically diverse populations carrying clonal and sub-clonal somatically acquired mutations<sup>1,2-4</sup>. The emergence of a dominant, somatically mutated clone occurs when the mutation offers a survival or proliferative advantage over 'competitor' cells under the conditions in their microenvironment<sup>5, 6</sup>. Exploring the cellular heterogeneity of cancers may therefore identify sub-clonal mutations that drive clonal expansion under different conditions.

Chronic lymphocytic leukemia (CLL) is a heterogeneous disorder, characterised by clonally expanded CD5 positive B-cells carrying a range of genetic abnormalities some of which segregate with disease severity<sup>3, 7, 8</sup>. As with many other forms of cancer, an important abnormality associated with poor overall outcome and refractoriness to DNA damaging therapy, is dysfunctional p53<sup>9-16</sup>.

The transcription factor p53, encoded by the *TP53* gene on chromosome 17p13, regulates the cell-cycle and apoptosis in response to cellular stress. Cellular stress in the form of therapy that irreversibly damages DNA (such as the purine analogue fludarabine routinely used to treat CLL), activates p53 to induce the expression of targets that stall cell-cycle progression<sup>17, 18</sup>. This stalling allows time for DNA repair, and if the repair process is unsuccessful, initiation of apoptosis<sup>17, 18</sup>. In keeping with the two-hit model of cancer<sup>19</sup>, the mutations of one *TP53* allele are generally found in the context of monoallelic loss of the remaining wild-type allele at the 17p13 locus<sup>12</sup>. The reported frequency of *TP53* mutations in CLL ranges from approximately 10% without treatment, rising to 50% in relapsed disease, refractory to DNA damaging therapy<sup>12, 15, 16</sup>. As recent next-generation sequencing studies suggest, the rise in frequency of clonal *TP53* mutated populations following DNA damaging therapy most likely reflects an expansion of low abundance, *TP53* mutated sub-clones often undetectable at presentation by conventional sequencing<sup>2, 4, 16, 20</sup>. The majority of *TP53* mutations reported in CLL lie within the p53 DNA binding domain (exons 4 - 10)<sup>21, 22</sup>, which can alter its capacity to bind to target DNA sequences and induce transcription; such abnormalities would offer a distinct survival advantage when a cell is exposed to agents that irreversibly damage DNA<sup>8, 15, 16, 23, 24</sup>.

Given the importance of p53 in conferring therapeutic response to DNA damaging agents, it is not surprising that detection of *TP53* abnormalities at presentation has gained importance in informing treatment choices. Without an intact *TP53* gene, standard chemotherapy options are supplanted by alternative therapies which do not rely on DNA damage to induce cytotoxicity such as the monoclonal antibody therapies ofatumumab (anti-CD20)<sup>25</sup> and alemtuzumab (anti-CD52)<sup>26, 27</sup>, and newer small molecule inhibitors including idelalisib (inhibitor of the delta isoform of phosphatidylinositol 3-kinase)<sup>28</sup>, and ibrutinib (inhibitor of Bruton's Tyrosine Kinase)<sup>29, 30</sup>. However, diagnostic recognition of *TP53* abnormalities that can influence patient treatment is mostly restricted to fluorescence *in situ* hybridisation (FISH) identification of 17p deletion, and conventional Sanger sequencing of the *TP53* mutational hotspot (exons 4 – 10), both of which lack the sensitivity to detect sub-clonally mutated *TP53* populations. Alternative approaches such as high resolution melting still require a conventional sequencing step and also lack sensitivity. Although the microarray based p53 Amplichip (Roche), improves on

sensitivity with a detection limit as low as 2%<sup>31</sup>, it is limited in the type of mutation that it can detect.

Deep sequencing platforms have the potential to solve the issue of sensitivity, and have identified sub-clones with a mutation frequency as low as 0.3% in CLL<sup>2-4, 15, 16</sup>. However, the depth of deep sequencing also leads to reporting of both systematic sequencing and amplification related errors, and distinguishing true variants from these events can be problematic<sup>32-34</sup>. Most bioinformatics approaches predict true variants from errors without taking the individual error rate at specific base positions along a sequence into consideration, with verification of low frequency events often limited to repetition of sequencing over two independent runs<sup>15, 16</sup>.

In this study we describe MuPol-Seq, a novel deep sequencing workflow in which error-prone aspects of sample preparation and the sequencing process are controlled for using different forms of polymerases to prepare libraries, combined with a strict bioinformatics pipeline which considers the positional error rate at each base. By analysing deep sequencing from separate libraries generated from the same sample but using enzymatically different, high fidelity polymerases that incur error biases at different rates, our novel approach offers an improved confidence when calling true mutants (present in both libraries), from background errors (detectable in only one library). By testing both libraries on two separate sequencing runs, the possibility of detecting systematic errors imposed by the sequencing process itself is also improved. As a key aim in this study was to produce a deep sequencing strategy that could be introduced into routine diagnostics, a low cost, highly multiplexed workflow was implemented. We applied this approach to distinguish low abundance *TP53* variations that may predict an individual response to DNA damaging therapy from background errors in presentation CLL cases, and evaluated the use of digital droplet PCR (ddPCR), a highly sensitive technique capable of detecting rare mutant-alleles,<sup>35-38</sup> to validate low abundance mutant base calls. In addition, in one case where a relapse sample was available we were able to show that *TP53* mutations present at CLL diagnosis were also present at relapse by ddPCR following treatment with fludarabine. This is supportive of a contributory role of *TP53* mutants in resistance to DNA damaging drugs, and highlights the need to investigate the clinical relevance of these sub-clones further using reliable, methodical approaches.

## Results

### *Variant identification using the MuPol-Seq pipeline*

Libraries were generated from each of the 35 CLL cases, using Long and Accurate *Thermus aquaticus* (LA *Taq*) and pfu-based Phusion, both high fidelity enzymatically different forms of polymerase (Figure 1). Of the 35 CLL cases sequenced using the 2 forms of polymerase over 2 separate sequencing lanes as out-lined in the MuPol-Seq strategy (Figure 1), sequencing data was acceptable for 33 cases with reads from only 2 cases discarded due to poor quality. By using two forms of polymerase over two sequencing lanes, it was possible to eliminate systematic errors (i.e. errors that only occur in only one library, or in one sequencing lane). The average read depth at any base position across the 8 kb region of *TP53* was 11,000 reads. The overall distribution of sequences generated with each of the 40 tags used in the pooled preparations, was similar indicating a high degree of consistency in library generation (Supplementary Figure S2).

The frequency of mismatches for all samples repeated on each lane and with both polymerases is shown in Figure 2. Note that the frequency of mismatched bases at each base position was taken into consideration when determining errors from true events. The errors are centred round a model mean of 0.003 and maximal error of 0.01. Ratio's greater than 0.01 are not expected to be a result of random errors within our sequencing protocol and should be either sub-clonal variants within the system or errors of PCR amplification. These effects can be observed more clearly when the repeats on each lane and with both forms of polymerase, are plotted together (Supplementary Figure S3).

True variants were called where an observed frequency at a position was significantly increased from normal ( $p$ -value  $<0.05$  adjusted for False Discovery Rate (FDR)), in each sequencing lane and using both forms of polymerase (Supplementary Table S1).

A series of variant and wild-type allele dilutions for the g.7574775 C>T single nucleotide polymorphism (SNP) were also sequenced and analysed identically to the CLL patient genomic samples. Data generated for the SNP g.7574775 C>T allele shows that the frequency of variant reads are proportional to variant allele input (Supplementary Table S2). However, this does not necessarily equate to overall

sensitivity to detect variant alleles at frequencies defined at input, as this type of experiment would not take positional variation into account.

*Frequency and characteristics of sub-clonal TP53 mutations detectable at CLL presentation*

Using the described bioinformatics criteria and outlined strategy to accurately call true mutational events (Figure 1), 26% of cases (n=9) presented with a *TP53* variant detectable by deep sequencing using the Illumina platforms. Of these, only 2 cases would have been identified as *TP53* mutation carriers by Sanger sequencing which is generally accepted as having the sensitivity to detect mutated alleles at a frequency of  $\geq 20\%$ . All *TP53* mutations detectable using our strategy could be found in the IARC *TP53* database further supporting their validity (<http://www-p53.iarc.fr>).

When assessing the number of mutations per case, it was evident that the majority of cases in this series presented with  $>1$  mutation within the *TP53* coding region (6 out of 9; Table 1; Table 2), the average frequency being 2 mutated bases per case. In total 13 mutations, 12 representing minor clones, were identified spanning exons 4 to 10 clustering in the DNA binding domain encompassing the major hotspot codons 220, 237, 245, 248, 249, and 273 each mutated in at least 1 case (Table 2; Figure 3). In particular, variants at codon 237, 245, 248, 249 and 273 all located within exons 7 and 8, occurred at higher frequency than variants at other codons (2%, 7.2%, 19.3%, 36.7% and 59.5% compared to  $<2\%$ , respectively). The majority of mutations within this hotspot were missense, with transitions more prevalent than transversions in the dataset overall. In particular, AG>TC transitions were over-represented (Table 2), and this is consistent with similar studies analysing *TP53* mutations in presentation CLL using alternative sequencing approaches. In cases where the number of mutations detected in *TP53* exceeded 2, the codon positions were mostly outside of the hotspot, and predominantly silent. These same positions, in particular codon 80 and codon 90, were overrepresented in the patient cohort occurring in 3 or more CLL cases.

*Frequency of TP53 mutations, including one relapse case, and their validation by ddPCR and TaqMan real-time allelic discrimination*

Overall, the frequency of mutations in this study ranged from 0.3% to 59.5%, and mutations with an occurrence of 0.6%, 7% and 19% were verified independently using TaqMan real-time PCR based on allelic discrimination (Supplementary Figure S4; Table 5), and ddPCR (Figure 4; Table 5). Wild-type genomic DNA from a case with no detectable *TP53* mutations was employed as a control and patient samples carrying Y220C, R248W, or R273C, were analysed for mutation frequency. As shown in Figures 4 and Supplementary Figure 4, the mutant alleles could be separated from wild-type alleles using both methods, most clearly by ddPCR.

In addition, mutations occurring at a frequency of approximately 20% or more (R248W, R249S, and G245V in cases 9, 2 and 26 respectively), were also verified by conventional Sanger sequencing (Figure 5). The results visualised by Sanger sequencing corresponded with the variant allele frequency observed in each of these cases by MuPol-Seq (Figure 5; Table 5).

In one case (patient 9), a follow-up sample was available at relapse prior to further therapy. Mutations detectable in *TP53* at presentation, were identified at a similar frequency at relapse by ddPCR in the hotspot region (variant allele frequency of 1% and 10.8% for Y220C and R248W respectively; Figure 4). This patient had previously received fludarabine-based therapy. Interestingly, the frequency of *TP53* mutation Y220C increased from 0.43% to 1%, whereas the frequency of *TP53* mutation R248W decreased from 24% to 10.8% when analysed by ddPCR (Figure 4.). Cytogenetic analysis by FISH, and further confirmation of mutational status was not possible at relapse as only a limited amount of archived DNA sample was available.

*The occurrence of sub-clonal mutations in cases featuring other poor prognostic markers*

*TP53* mutation state was compared to known markers of poor prognosis: *IGHV* mutational status, 17p13 loss and 11q23 loss (Table 1). Of the 9 cases with mutated *TP53*, 7 were in the *IGHV* unmutated category. Of the 2 cases with a mutated *IGHV*

region (<98% homology with germ-line *IGHV*), 1 (patient 16) carried a number of very low frequency mutations outside of the major hotspot region, whilst the remaining case (patient 33), carried a missense mutation in the hotspot region at a frequency of 2%.

Loss of one allele of *TP53* through deletions of chromosome 17p often coincides with mutation of the remaining allele and is an indicator of overall poor survival as well as poor response to DNA damaging drugs. Indeed, 4 of the 9 cases with *TP53* mutations were also shown to be 17p- by FISH. In these cases, the mutations were focussed around the primary mutational hotspots (codons 220, 237, 245, 248, 249 and 273), in exons 5 to 8. The patients with large mutated clones (2, 9 and 26), above the level of sensitivity of FISH, all showed 17p deletion (Figure 6). A mutated *TP53* clone was detectable in only one case of 11q-, containing the *ATM* DNA repair gene, and this was outside of the major hotspot region (Table 1; Table 2).

## Discussion

The capability of deep resequencing to distinguish low abundance, disease-relevant mutations without separation from a non-mutated background is advantageous not only in research, but also in a clinical setting as a diagnostic tool. In the context of leukemia, sampling of circulating tumour cells has the potential to reduce the need for invasive, often expensive procedures such as bone marrow or solid tumour biopsy and, as described in this study, can be achieved using a minimum quantity of genomic DNA. Multiplexing using unique sample identifiers increases workload capacity and reduces cost. However, the sensitivity of deep sequencing also reports low frequency background errors generated by DNA polymerase during library preparation and the Illumina sequencing process itself<sup>32-34</sup>. Finding a way to distinguish low abundance mutants from systematic errors is therefore a key factor in the successful routine implementation of deep sequencing.

In this study we addressed this issue using a strategy in which resequencing libraries were derived from amplicons generated using two enzymatically different DNA polymerases, LA *Taq* and Phusion, run on two independent sequencing runs, coupled with a robust bioinformatics pipeline. Although both enzymes incur an error

bias, predominantly towards transitions, *Taq* polymerase has a propensity to incorporate cytidine with template thymine, and at a higher frequency than that of Phusion<sup>34</sup>. This bias allows some distinction of true variants from polymerase errors given that true variants are more likely to be present in sequences generated from both forms of polymerase, eliminating errors that are unique to a single library generated from either LA *Taq* polymerase or Phusion alone. Furthermore, in an attempt to identify or eliminate errors induced during the Illumina sequencing run itself, GC-rich regions being particularly susceptible, data from two independent sequencing runs was analysed following the removal of error-prone homopolymeric tracts (mostly intronic), with only identical variants appearing in both runs classed as true events<sup>34</sup>.

Although other deep sequencing studies have attempted to validate low frequency *TP53* events<sup>15, 16</sup>, our approach is distinguished by the incorporation of two forms of polymerase in the library preparation linked to sensitive validation using two approaches. Like Sanger sequencing the first approach, TaqMan based allelic discrimination, amplifies both wild-type and variant alleles in the same PCR reaction, but the ability to discriminate mutant alleles is enhanced by the presence of two distinctly labelled fluorescent probes, one labelled probe complementary to mutant-alleles and the second differently labelled probe complementary to wild-type alleles. An advancement of the TaqMan based two-coloured probe approach, is ddPCR in which wild-type and mutant-alleles within an individual sample are partitioned into hundreds of droplets prior to PCR. This approach effectively increases the concentration of rare mutant-alleles within the limited number droplets in which they are contained, and in this study convincingly distinguished low abundance *TP53* mutant-alleles from more frequent wild-type alleles. Given that the number of cases of CLL carrying a *TP53* mutation in this study was relatively low (26%), and the number of coinciding *TP53* mutations within each of these cases only averaged at 2 mutations per case, (a lower frequency than other deep sequencing studies examining *TP53* clonality in CLL<sup>15</sup>) this suggests that systematic errors are being filtered out using the described strategy with genuine events reported.

As somatic *TP53* mutations are recognised as indicators of poor prognosis in a number of cancers<sup>39</sup>, we wanted to test whether our novel deep sequencing approach could identify low abundance mutations in early-stage disease. This may

provide information on disease progression, for example in response to p53-dependent therapy, using CLL as a model of disease. Our study concurred with other reports in that a proportion of CLL cases carry sub-clonal *TP53* mutations at diagnosis undetectable by conventional sequencing techniques. The majority of these are missense transitions, correlating mostly with unmutated *IGHV*, a marker of a more aggressive disease<sup>15, 16</sup>. Sub-clonal mutations are important in cancer progression and may particularly contribute to the evolution of drug-resistant clones under the selective pressure of a therapy targeting the mutated pathway. In this study we were able to show (albeit in only one case), that the same *TP53* mutations present at diagnosis emerge at relapse following treatment with DNA damaging fludarabine supporting *TP53* mutants as contributors of drug-resistance. The ability to accurately detect low frequency events therefore offers a distinct clinical advantage over currently implemented diagnostic techniques. Specifically, only 3 of the 9 presentation CLL cases with mutated *TP53* detected in this study carried mutations at a frequency, confidently detectable by Sanger sequencing. Furthermore *TP53* deletion identifiable by FISH, the parameter most commonly used to measure *TP53* abnormalities routinely, was only identifiable in 4 cases, two of which were the cases with coincident high mutation frequency of the remaining allele. Thus the majority of patients in this study would have been reported as *TP53* wild-type. This supports the need for the implementation and further evaluation of sensitive sequencing approaches such as this to inform clinical decision making. This becomes increasingly important as alternative drugs, such as idelalisib<sup>40</sup> and ibrutinib<sup>41</sup>, have been approved for previously untreated patients with CLL who have either deletion of 17p or a *TP53* mutation.

Interestingly, of the cases where 17p deletion correlated with variant *TP53*, at least one dominating *TP53* clonal event was present at a significantly higher frequency than sub-clonal events in other cases within our series. Moreover, these high frequency events were specifically at residues 237, 245, 248, 249 and 273, encoding regions of the p53 DNA binding domain essential for DNA contact (in particular, codons 248 and 273), and structural architecture (codons 245 and 249)<sup>42, 43</sup>. These sites correlate with known *TP53* clustering patterns not only observed in CLL but also in other forms of cancer<sup>8, 43</sup>. Our report suggests a link between 17p deletion and high frequency *TP53* mutational events at specific codons in CLL before

the added selective pressure of DNA-damaging treatment, albeit in a small number of cases. This is perhaps not surprising given that p53 binds to sequence specific response elements through its DNA binding domain controlling expression of cell-cycle and apoptotic proteins. Missense mutations within this domain are therefore likely to disable these pathways, increasing overall cellular instability<sup>8, 42, 43</sup>. Furthermore, it is possible that this phenotype may confer resistance to DNA damaging therapy. Given that mutations in the p53 DNA binding domain are likely to offer a survival advantage under the selective pressure of fludarabine, it was surprising to find approximately half the sub-clonal frequency of R248W present in case 9 at relapse when compared to the presentation sample, when analysed by ddPCR. Interestingly, this coincided with a doubling in frequency of the Y220C DNA binding domain mutation also by ddPCR, in the same relapse and presentation samples. One possibility is that both abnormalities represent separate, competing sub-clones with the *TP53* Y220C mutation conferring greater therapeutic resistance than the R248W *TP53* mutation. Further, longitudinal analysis of this case, should additional material become available, could provide more detail of this unexpected finding.

Unlike other studies, the majority of CLL cases in this study carrying *TP53* mutations (5 out of 9), did not display deleted 17p. Given the bi-allelic nature of cancer-related genetic alterations, and the low frequency of mutational events in these 5 cases (<2%), this may simply reflect the lack of sensitivity of FISH to detect abnormal 17p when the occurrence of loss is in <5% of B-cells. Alternatively, it is possible that inactivating mechanisms undetectable by sequencing or FISH, such as copy neutral loss of heterozygosity are present<sup>12, 23, 44</sup>. A further possibility that is difficult to eliminate by sequencing, is that bi-allelic sub-clonal *TP53* mutations in these 5 cases exist in a proportion of the cellular population. However, given that the mutations predominantly observed in these cases were detectable outside of the mutational hotspot linked with *TP53* dysfunction, it is possible that they are of no clinical relevance. This is supported by the low variant allele frequencies with the lack of a dominating sub-clone when compared to the mutational spectrum observed in 17p deletion carriers (Figure 6), with the majority of silent *TP53* mutations detectable by deep sequencing being observed with the 17p intact group (Table 2).

Overall, this report describes an ultra-deep resequencing workflow that distinguishes low abundance *TP53* mutations from background generated from known sources. This approach can accurately detect low abundance mutations present at CLL diagnosis some of which will alter functionality of p53 and may therefore drive clonal selection either alone or in response to DNA damaging therapy. Given the low costs of library preparation, the robustness of the bioinformatics pipeline, the requirement of only low amounts of genetic material from non-separated cells, and the ability to multiplex a number of samples, this system is applicable routinely. Future evaluation is required to determine whether the detected low frequency mutations can prospectively predict patient response to therapy. However given the known central role of p53 in chemotherapy responses this is a reasonable prediction.

## Materials and methods

### *Patient samples*

Peripheral blood was collected prospectively from a cohort of 35 newly diagnosed CLL cases (Table 1), and 4 normal donors. The aim was to collect cases representative of the major, prognostically relevant groups in CLL (i.e. *IGHV* mutated and unmutated cases, and cases with or without loss of 17p, 13q, and 11q). Information on cytogenetics (generated by routine FISH), and *IGHV* mutational status<sup>45</sup>, was obtained from the Haematological Malignancy Diagnostic Services (HMDS), Leeds Teaching Hospitals. Likewise, flow cytometry analysis was available through HMDS, and the majority of the cohort presented with >70% CLL cells as assessed by CD5/CD19 positivity, distinguishable from other B-cell disorders using a B-cell panel of flow cytometry fluorescent markers (CD10, CD103, CD11c, CD185, CD196, CD19, CD20, CD200, CD22, CD23, CD25, CD27, CD305, CD31, CD38, CD39, CD43, CD45, CD49d, CD5, CD79b, CD81, CD95, IgD, IgG, IgM, Kappa, Lambda), alleviating the need for leukemic cell separation. An additional sample was available for case 9; a DNA extraction from a peripheral blood sample taken at relapse approximately 18 months after presentation sample. This patient had received fludarabine-based therapy. Ethical approval was gained prior to this study, and patient samples were collected in accordance with the Declaration of Helsinki (REC 04/Q1205/178).

Mononuclear cells were prepared from peripheral blood using the standard Ficoll-Histopaque gradient centrifugation protocol (Lymphoprep, Axis-Shield), and DNA was extracted from mononuclear cell pellets using the Qiamp Blood minikit in accordance with manufacturer's instructions (Qiagen).

#### *Preparation of wild-type/variant dilution series*

In order to show that our strategy can detect a variation proportional to the allele frequency, a case containing a homozygous variant single nucleotide polymorphism (SNP g.7574775 C>T), was diluted serially with a wild-type homozygous case for the same allele. Each dilution was then treated as an individual case, and was prepared, sequenced and analysed identically to patient samples.

#### *TP53 PCR, library preparation, and deep sequencing*

A novel strategy (Figure 1), was adopted by which each DNA sample was amplified with two different classes of DNA polymerase to generate either an 8 kb product, using Long and Accurate (LA) *Taq* polymerase (Takara) or three 3 kb products using pfu-based Phusion High Fidelity polymerase (Finnzymes) from genomic DNA representing the *TP53* gene including coding exons 2 to 11, introns and the 3'untranslated region (Figure 1; Supplementary Table S3). PCR products from amplification with LA *Taq* and Phusion were agarose gel purified using a Zymoclean gel extraction kit (Zymed). To generate sequencing libraries, the purified amplicon was sonicated (Bioruptor, Diagenode) to generate fragments of 200-300bp followed by column purification (Zymogen). DNA was end-repaired for 30 minutes at room temperature using Klenow DNA polymerase, T4 DNA polymerase, and T4 DNA Polynucleotide Kinase (New England Biolabs), before addition of a dA-tail using Klenow 3'-5' exonuclease (New England Biolabs) for 30 minutes at 37°C. Index-tagged Y-shaped adapters (using a design based on Bentley et al, 2008<sup>20</sup>; Supplementary Table 3), were then ligated using Quick ligase (New England Biolabs) for 30 minutes at room temperature followed by a 10 minute 70°C step to inactivate the enzyme. Products were column purified in-between each procedure, and size selection was then performed by excision of the 250-350bp tagged product

from a 2% GTG agarose gel (Lonza). The purified, tagged amplicon library was enriched and prepared for sequencing using Phusion Flash (Finnzymes), in a 14-cycle reaction (Supplementary Table S1). The size distribution of the final product was checked on an Agilent Bioanalyser (Agilent Technologies), and quantified using a picogreen based kit (Quant-iT, Invitrogen). Equal amounts of each uniquely tagged product were mixed to generate a multiplexed pool of up to 40 uniquely indexed samples which were run on two sequencing lanes to remove between run variations (Figure 1), and sequenced using either the GAllx or the Hi-seq platforms (Illumina).

### *Bioinformatics analysis*

Following sequencing, data was collected for analysis, and sorted into individual paired end fastq files for each sample and enzyme using bespoke Perl scripts. All sample fastq files were subject to quality control inspection using Fastqc ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)), trimmed when the base quality fell below a phred quality score of 30 and discarded when the read was less than 20 base pairs, using Prinseq lite software<sup>46</sup>, and adapterRemoval<sup>47</sup>. Reads passing quality control were aligned to the *TP53* gene sequence in human genome version 19 (Genome Reference Consortium GRCh37), using the Burrow-Wheeler-Alignment method<sup>48</sup>.

The number of mismatched bases corresponding to the reference base ( $n$ ) and the total number of base calls at a given position ( $N$ ) were obtained using SAMtools: pileup<sup>49</sup>, and bespoke scripting in Perl. A mismatch ratio ( $R=n/N$ ) was then calculated if the read depth exceeded 500 reads.

In order to find genomic positions that possess a significant detectable increased number of mismatches, i.e. possible candidate somatic mutations, mismatch ratios were compared to a background distribution of mismatch ratios expected from normal samples, predominantly incorporating the reference base<sup>50</sup>.

Sequences from the CLL cases under investigation were compared to sequences generated from normal samples by assuming the mismatches can be modelled as a Poisson distribution<sup>50</sup>. In brief, the distribution describes the probability of obtaining  $n$  mismatches in our sample of interest given a mean number of mismatches observed

in our normal population ( $\lambda$ ).  $\lambda$  is calculated as the mean mismatch ratio multiplied by read depth at that site in our sample of interest.

Potential mutated genomic positions in a given sample were identified provided the mismatch ratio was greater than that of background distribution of samples, and passed the *P*-value cutoff (0.05) after multiple testing correction using the False Discovery Rate (FDR)<sup>51</sup>. However, significant candidates would only be called when this was repeated across all replicates (i.e. in both lanes of the sequencing run), and observed in sequence runs generated from both forms of polymerase (LA *Taq* and Phusion High Fidelity, Figure 1; Supplementary Table S1). The International Agency on Research for Cancer (IARC) *TP53* database was used to annotate mutations based on their codon position (<http://www-p53.iarc.fr>). Additional information on the bioinformatics approach is given in the supplemental data.

#### *Validation by TaqMan Allelic Discrimination, digital droplet PCR, and Sanger Sequencing*

A proportion of variants detectable by our sequencing and bioinformatics approach were subjected to TaqMan based allelic discrimination (Applied Biosystems), ddPCR (QX200, Bio-Rad), and Sanger sequencing for validation. It was anticipated that ddPCR would be superior to TaqMan based allelic discrimination, as previous studies already show accurate detection of rare mutations at variant allele frequencies of 0.01% and below<sup>35-38</sup>. Sanger sequencing was adopted to confirm mutations present by MuPol-Seq at a frequency close to, or within the capabilities of Sanger sequencing detection and visualisation ( $\geq 20\%$ ).

#### *TaqMan Allelic Discrimination*

A separate TaqMan allelic discrimination mastermix was prepared for 3 of the mutant *TP53* alleles (R248W, R273C, and Y220C) detected by Mu-Pol sequencing for which assays could be custom designed (Applied Biosystems). Each mastermix, containing 1x TaqMan universal PCR mastermix (Applied Biosystems), 1x Assay-by-Design mix (-a custom designed mix of mutant-specific FAM-labelled probe, wild-type specific VIC-labelled probe, forward and reverse primers, Applied Biosystems; Supplementary Table S3), mixed with replicate serial dilutions (0 to 60ng) of genomic

DNA from *TP53* variant cases or wild-type cases (detectable by deep sequencing), or 'no template' control (water) in a 96-well optical plate, as described in the manufacturer's instructions (Applied Biosystems). Each plate was run on the Applied Biosystems 7700 thermocycler (thermocycle conditions, Supplementary Table S3), and real-time quantitative plots and end-point reads were generated using Applied Biosystems 7700 software in order to assess the presence/absence of mutated alleles.

#### *Digital Droplet PCR*

Digital droplet PCR is a technique based on the dilution of mutants away from wild-type alleles by the partitioning of the genomic template into nanoliter droplets. In this study, droplets were formed from the same replicate serial dilutions of genomic DNA representing wild-type and variant *TP53* alleles and 1 x Assay-by Design dye-labelled probe/primer mixes as described for TaqMan Allelic discrimination, together with 2x ddPCR supermix for probes (Bio-Rad). In addition, ddPCR was also used to investigate the presence of *TP53* mutations detectable at CLL diagnosis in case 9 (Y220C and R248W), in the available relapse sample from the same case. Droplets were generated from each mix using droplet generation oil, droplet cartridges and gaskets, and a droplet generator (all products from Bio-Rad), following manufacturer's guidelines. After transferring droplets to a PCR compatible plate (Eppendorf), amplifications were performed (thermocycling conditions shown in Supplementary Table S3), before loading each PCR plate into a QX200 droplet reader (Bio-Rad). The reader was operated using manufacturer's guidelines and data was analysed using QX200 Quantalife Software (Bio-Rad). Thresholds were set manually based on clustering as described elsewhere<sup>38</sup>.

#### *Sanger Sequencing*

Genomic DNA from cases 26, 2 and 9 was amplified by standard PCR using the thermocycling conditions, and the 3.6Kb primer set (spanning the exon 7 region containing R254V, R249S, and R248W) shown in Supplementary Table S3. Each amplicon was run on a 1% agarose gel, and bands were excised and purified as described previously. Purified amplicons were sequenced using standard conditions by Source Biosciences (Nottingham), using an additional internal primer preceding

exon 7 (5'-CGACAGAGAGCGAGATTCCATC-3'). Sequences were aligned using Chromas Lite software version 2.1.1 (Technelysium).

**Conflict-of-interest disclosure:** The authors declare no competing financial interests.

### **Acknowledgements**

This project was funded by the Kay Kendall Leukaemia Fund ([www.kklf.org.uk](http://www.kklf.org.uk)). We would also like to thank the staff in the Haematological Malignancy Diagnostic Services Unit at Leeds Teaching Hospital who provided information on diagnostic patient data and samples for the study, and the sequencing facility at the University of Leeds who provided a technical service running the Illumina next-generation sequencing systems.

### **Authorship**

Contribution: L.W. prepared the manuscript, and performed the validation experiments. P.B., M.A.C. and L.H carried out the bioinformatics, and P.B. and L.H co-wrote the bioinformatics sections. A.V. and T.M obtained and prepared patient samples. P.A.E. performed mutational analysis of *IGHV* and S.J.O. carried out FISH cytogenetic analysis. A.R. performed flow analysis, and assisted in study design. R.M.T. helped design the study, and reviewed the article. P.H. and D.J.N. conceived and designed the study, and D.J.N. wrote part of the manuscript and carried out the laboratory work.

**Supplementary Information** accompanies the paper on the Oncogene website (<http://www.nature.com/onc>).

### **References**

- 1 Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012; 481: 306-313.
- 2 Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS *et al*. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 2013; 152: 714-726.
- 3 Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R *et al*. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* 2012; 120: 4191-4196.
- 4 Ojha J, Ayres J, Secreto C, Tschumper R, Rabe K, Van Dyke D *et al*. Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood* 2014.
- 5 Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I *et al*. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2008; 105: 13081-13086.
- 6 Wang J, Khiabani H, Rossi D, Fabbri G, Gattei V, Forconi F *et al*. Tumor evolutionary directed graphs and the history of chronic lymphocytic leukemia. *Elife* 2014; 3.
- 7 Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H *et al*. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood* 2008; 111: 5446-5456.

- 8 Zenz T, Eichhorst B, Busch R, Denzel T, Habe S, Winkler D *et al.* TP53 Mutation and Survival in Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology* 2010; 28: 4473-4479.
- 9 Zenz T, Benner A, Dohner H, Stilgenbauer S. Chronic lymphocytic leukemia and treatment resistance in cancer: the role of the p53 pathway. *Cell Cycle* 2008; 7: 3810-3814.
- 10 Dohner H, Fischer K, Bentz M, Hansen K, Benner A, Cabot G *et al.* p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. *Blood* 1995; 85: 1580-1589.
- 11 Dicker F, Herholz H, Schnittger S, Nakao A, Patten N, Wu L *et al.* The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype. *Leukemia* 2009; 23: 117-124.
- 12 Zenz T, Habe S, Denzel T, Mohr J, Winkler D, Buhler A *et al.* Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the contribution of 17p deletion, TP53 mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial. *Blood* 2009; 114: 2589-2597.
- 13 Delgado J, Espinet B, Oliveira AC, Abrisqueta P, de la Serna J, Collado R *et al.* Chronic lymphocytic leukaemia with 17p deletion: a retrospective analysis of prognostic factors and therapy results. *British journal of haematology* 2012; 157: 67-74.
- 14 Wawrzyniak E, Kotkowska A, Blonski JZ, Siemieniuk-Rys M, Ziolkowska E, Giannopoulos K *et al.* Clonal evolution in CLL patients as detected by FISH

- versus chromosome banding analysis, and its clinical significance. *European journal of haematology* 2014; 92: 91-101.
- 15 Malcikova J, Stano-Kozubik K, Tichy B, Kantorova B, Pavlova S, Tom N *et al.* Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. *Leukemia* 2014.
  - 16 Rossi D, Khiabani H, Spina V, Ciardullo C, Brusca A, Fama R *et al.* Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* 2014; 123: 2139-2147.
  - 17 Balint EE, Vousden KH. Activation and activities of the p53 tumour suppressor protein. *British journal of cancer* 2001; 85: 1813-1823.
  - 18 Tian K, Rajendran R, Doddanajiah M, Krstic-Demonacos M, Schwartz JM. Dynamics of DNA damage induced pathways to cancer. *PloS one* 2013; 8: e72303.
  - 19 Knudson AG, Jr. Overview: genes that predispose to cancer. *Mutation research* 1991; 247: 185-190.
  - 20 Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; 456: 53-59.
  - 21 Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P *et al.* Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Human mutation* 2007; 28: 622-629.

- 22 Olivier M, Petitjean A, Marcel V, Petre A, Mounawar M, Plymoth A *et al.* Recent advances in p53 research: an interdisciplinary perspective. *Cancer gene therapy* 2009; 16: 1-12.
- 23 Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E *et al.* Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011; 29: 2223-2229.
- 24 Jethwa A, Hullein J, Stolz T, Blume C, Sellner L, Jauch A *et al.* Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. *British journal of haematology* 2013; 163: 496-500.
- 25 Gravanis I, Ersboll J, Skovlund E, Abadie E, Marty M, Pignatti F. The European Medicines Agency review of ofatumumab (Arzerra(R)) for the treatment of chronic lymphocytic leukemia in patients refractory to fludarabine and alemtuzumab: summary of the scientific assessment of the European medicines agency committee for medicinal products for human use. *The oncologist* 2010; 15: 1335-1343.
- 26 Lozanski G, Heerema NA, Flinn IW, Smith L, Harbison J, Webb J *et al.* Alemtuzumab is an effective therapy for chronic lymphocytic leukemia with p53 mutations and deletions. *Blood* 2004; 103: 3278-3281.
- 27 Stilgenbauer S, Zenz T, Winkler D, Buhler A, Schlenk RF, Groner S *et al.* Subcutaneous alemtuzumab in fludarabine-refractory chronic lymphocytic leukemia: clinical results and prognostic marker analyses from the CLL2H study of the German Chronic Lymphocytic Leukemia Study Group. *Journal of*

- clinical oncology : official journal of the American Society of Clinical Oncology 2009; 27: 3994-4001.
- 28 Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P *et al.* Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *The New England journal of medicine* 2014; 370: 997-1007.
- 29 Woyach JA, Furman RR, Liu TM, Ozer HG, Zapatka M, Ruppert AS *et al.* Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. *The New England journal of medicine* 2014; 370: 2286-2294.
- 30 Komarova NL, Burger JA, Wodarz D. Evolution of ibrutinib resistance in chronic lymphocytic leukemia (CLL). *Proceedings of the National Academy of Sciences of the United States of America* 2014; 111: 13906-13911.
- 31 Grollman AP, Shibutani S, Moriya M, Miller F, Wu L, Moll U *et al.* Aristolochic acid and the etiology of endemic (Balkan) nephropathy. *Proceedings of the National Academy of Sciences of the United States of America* 2007; 104: 12129-12134.
- 32 Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988; 239: 487-491.
- 33 Nakamura S, Nakaya T, Iida T. Metagenomic analysis of bacterial infections by means of high-throughput DNA sequencing. *Experimental biology and medicine* 2011; 236: 968-971.

- 34 McInerney P, Adams P, Hadi MZ. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular biology international* 2014; 2014: 287430.
- 35 Huggett JF, Cowen S, Foy CA. Considerations for digital PCR as an accurate molecular diagnostic tool. *Clin Chem* 2015; 61: 79-88.
- 36 Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS *et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* 2015; 518: 552-555.
- 37 Chang-Hao Tsao S, Weiss J, Hudson C, Christophi C, Cebon J, Behren A *et al.* Monitoring response to therapy in melanoma by quantifying circulating tumour DNA with droplet digital PCR for BRAF and NRAS mutations. *Sci Rep* 2015; 5: 11198.
- 38 Stieglitz E, Troup CB, Gelston LC, Haliburton J, Chow ED, Yu KB *et al.* Subclonal mutations in SETBP1 confer a poor prognosis in juvenile myelomonocytic leukemia. *Blood* 2015; 125: 516-524.
- 39 Goh AM, Coffill CR, Lane DP. The role of mutant p53 in human cancer. *The Journal of pathology* 2011; 223: 116-126.
- 40 Use CfMPfH.  
[www.ema.europa.eu/docs/en\\_GB/document\\_library/Summary\\_of\\_opinion\\_-\\_Initial\\_authorisation/human/003843/WC500170181.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/003843/WC500170181.pdf), 2014.
- 41 Use CfMPfH.  
[www.ema.europa.eu/docs/en\\_GB/document\\_library/Summary\\_of\\_opinion\\_-\\_Initial\\_authorisation/human/003791/WC500170191.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/003791/WC500170191.pdf), 2014.

- 42 Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. *Annual review of biochemistry* 2008; 77: 557-582.
- 43 Joerger AC, Fersht AR. The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harbor perspectives in biology* 2010; 2: a000919.
- 44 Mohr J, Helfrich H, Fuge M, Eldering E, Buhler A, Winkler D *et al.* DNA damage-induced transcriptional program in CLL: biological and diagnostic implications for functional p53 testing. *Blood* 2011; 117: 1622-1632.
- 45 van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; 17: 2257-2317.
- 46 Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome biology* 2007; 8: R143.
- 47 Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L *et al.* The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* 2012; 28: 2059-2061.
- 48 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760.

- 49 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078-2079.
- 50 Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F *et al*. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Science translational medicine* 2012; 4: 136ra168.
- 51 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 1995; 57: 289-300.

**Table 1. Characteristics of presentation CLL cases and corresponding *TP53* mutations as determined by deep sequencing**

| Case | Age<br>(years) | Sex    | Cytogenetics |         |       | <i>*IGHV</i><br>status | <i>TP53</i><br>mutations        |
|------|----------------|--------|--------------|---------|-------|------------------------|---------------------------------|
|      |                |        | 13q14-       | ↑17p13- | 11q23 |                        |                                 |
| 1    | 86             | male   | >95%         | No      | No    | Unmutated              | -                               |
| 2    | 86             | female | No           | 66%     | No    | Unmutated              | R249S, R273C                    |
| 3    | 45             | male   | 61%          | No      | No    | Unmutated              | -                               |
| 4    | 53             | female | No           | No      | No    | Unmutated              | S90P                            |
| 5    | 69             | male   | 59%          | No      | No    | Unmutated              | nd                              |
| 6    | 39             | female | No           | No      | No    | Unmutated              | -                               |
| 7    | 63             | male   | 84%          | No      | No    | Unmutated              | S90P, R335R                     |
| 8    | 52             | male   | No           | No      | No    | Unmutated              | -                               |
| 9    | 62             | female | No           | >95%    | No    | Unmutated              | Y220C, R248W (P +R)             |
| 10   | 60             | female | 87%          | No      | No    | Unmutated              | T81T, S90P, T155P               |
| 11   | 76             | female | No           | No      | 93%   | Unmutated              | S90P                            |
| 12   | 53             | male   | 95%          | No      | No    | Unmutated              | -                               |
| 13   | 62             | male   | 99%          | No      | No    | Mutated                | -                               |
| 14   | 63             | male   | 22%          | No      | No    | Mutated                | nd                              |
| 15   | 53             | male   | No           | No      | No    | Mutated                | -                               |
| 16   | 56             | male   | 91%          | No      | No    | Mutated                | T81T, S90P, T155P, R333R, R335R |
| 17   | 67             | male   | 95%          | No      | 95%   | Unmutated              | -                               |
| 18   | 63             | male   | No           | 95%     | No    | Unmutated              | -                               |
| 19   | 63             | male   | 88%          | No      | 96%   | Unmutated              | -                               |
| 20   | 46             | male   | No           | No      | 98%   | Unmutated              | -                               |
| 21   | 68             | male   | No           | 72%     | No    | Unmutated              | -                               |
| 22   | 54             | male   | 94%          | No      | 93%   | Unmutated              | -                               |
| 23   | 46             | male   | >95%         | >95%    | No    | Unmutated              | -                               |
| 24   | 68             | male   | 91%          | No      | 90%   | Unmutated              | -                               |
| 25   | 49             | male   | 90%          | No      | 92%   | Unmutated              | -                               |
| 26   | 65             | female | 25%          | 84%     | No    | Unmutated              | T81T, M160V, V197L, G245V       |
| 27   | 67             | nd     | 30%          | 72%     | No    | Unmutated              | -                               |
| 28   | nd             | nd     | No           | No      | 95%   | Unmutated              | -                               |
| 29   | 64             | male   | No           | No      | No    | Unmutated              | -                               |
| 30   | nd             | nd     | No           | No      | No    | Unmutated              | -                               |

|    |    |        |     |     |    |           |       |
|----|----|--------|-----|-----|----|-----------|-------|
| 31 | 57 | male   | nd  | nd  | nd | nd        | -     |
| 32 | 79 | female | nd  | nd  | nd | nd        | -     |
| 33 | 66 | male   | 76% | 51% | No | mutated   | M237I |
| 34 | 53 | male   | No  | No  | No | Unmutated | -     |
| 35 | 56 | female | 96% | No  | No | mutated   | -     |

\* where 'mutated' is <98% homology with germ-line; † 17p13 deletion is monoallelic unless stated; - no mutations detectable by deep sequencing; nd, Data not available, as failed Mu-Pol Seq quality control.(P+R), mutations shown were present in case 9 at presentation and relapse.

**Table 2. Description of cases, characteristic *TP53* mutations, and variant allele frequency (VAF) of each *TP53* mutation per case, as determined by multiple polymerase deep sequencing (*P*-value cut off <0.05, False Discovery Rate corrected)**

| Cases (n=9) | Allele description (Wt>Vt) |     | Amino acid change | Exon | Effect of Variant | VAF (%) |
|-------------|----------------------------|-----|-------------------|------|-------------------|---------|
| 2           | transition                 | G>A | R273C             | 8    | missense          | 7.2     |
|             | transversion               | C>A | R249S             | 7    | missense          | 36.7    |
| 4           | transition                 | A>G | S90P              | 4    | missense          | 1.4     |
| 7           | transversion               | A>C | R335R             | 10   | silent            | 0.5     |
|             | transition                 | A>G | S90P              | 4    | missense          | 1.1     |
| 9           | transition                 | G>A | R248W             | 7    | missense          | 19.3    |
|             | transition                 | T>C | Y220C             | 6    | missense          | 0.6     |
| 10          | transversion               | T>G | T155P             | 5    | missense          | 1.1     |
|             | transition                 | A>G | S90P              | 4    | missense          | 1.0     |
|             | transition                 | T>C | T81T              | 4    | silent            | 1.4     |
| 11          | transition                 | A>G | S90P              | 4    | missense          | 0.9     |
| 16          | transversion               | A>C | R335R             | 10   | silent            | 0.8     |
|             | transition                 | A>G | R333R             | 10   | silent            | 0.6     |
|             | transversion               | T>G | T155P             | 5    | missense          | 1.4     |
|             | transition                 | A>G | S90P              | 4    | missense          | 1.1     |
|             | transition                 | T>C | T81T              | 4    | silent            | 1.8     |
| 26          | transversion               | C>A | G245V             | 7    | missense          | 59.5    |
|             | transversion               | C>G | V197L             | 6    | missense          | 0.3     |
|             | transition                 | T>C | M160V             | 5    | missense          | 0.6     |
|             | transition                 | T>C | T81T              | 4    | silent            | 1.4     |
| 33          | transition                 | C>T | M237I             | 7    | missense          | 2.0     |

**Table 3. Summary of *TP53* mutations detectable by deep sequencing, Sanger sequencing, TaqMan based allelic discrimination, and digital droplet PCR.**

| Case      | Variant | Mu-Pol Seq VAF | <sup>†</sup> ddPCR VAF (SD) | Sanger   | End-point |
|-----------|---------|----------------|-----------------------------|----------|-----------|
| <b>2</b>  | R273C   | 7.2            | 9.9(±0.83)                  | -        | Detected  |
|           | R249S   | 36.7           | -                           | Detected | -         |
| <b>9P</b> | R248W   | 19.3           | 24(±1.9)                    | Detected | Detected  |
|           | Y220C   | 0.6            | 0.48(±0.09)                 | -        | Detected  |
| <b>9R</b> | R248W   | -              | 10.8 (±0.25)                | -        | -         |
|           | Y220C   | -              | 1 (±0.14)                   | -        | -         |
| <b>26</b> | G245V   | 59.5           | -                           | Detected | -         |

<sup>†</sup> digital droplet PCR, Standard Deviation (SD) of variant allele frequency's (VAF), based on replicates. 9P, presentation sample for case 9, whereas 9R is the relapse sample for the same case.

**Figure 1. Schematic representation of the technical strategy to identify minor *TP53* variants from background errors using multiple polymerases to prepare libraries prior to deep sequencing**

**Figure 2. Frequency distribution of mismatch ratios averaged over chronic lymphocytic leukemia cases across all loci and non-reference bases.**

**Figure 3. Schematic representation of the p53 protein showing functional domains with corresponding amino acid positions. Symbols shown above representation show the positions and nature of the *TP53* mutations detected by deep sequencing.**

**Figure 4. Quantitative digital droplet PCR (A) and end point TaqMan based allelic discrimination PCR to detect *TP53* variants Y220C (upper plots) and R248W (lower plots) in patient 9. (A) Variant-containing droplets (FAM-labelled, shown as blue dots) are quantified by comparison with wild-type droplets (VIC-labelled, shown as green dots) (B) TaqMan based PCR demonstrating clear discrimination of variant-containing wells (FAM-labelled probe, shown as red symbols) from wild type-containing wells (VIC-labelled**

probe, shown as blue symbols) for patient 9 dilution series. In both, vertical axis is FAM channel fluorescence and horizontal axis is VIC channel fluorescence.

**Figure 5. Sanger sequencing of *TP53* mutations detectable by deep sequencing at a frequency of (A) 59.5% in case 26 (G245V, C>A), (B) 36.7% in case 2 (R249S, C>A), and (C) 19.3% in case 9 (R248W, G>A).**

**Figure 6. Frequency and amino acid position of *TP53* mutations detectable by deep sequencing in 17p deleted (lower plot), and 17p wild-type (upper plot) CLL patients.**

Figure 1.

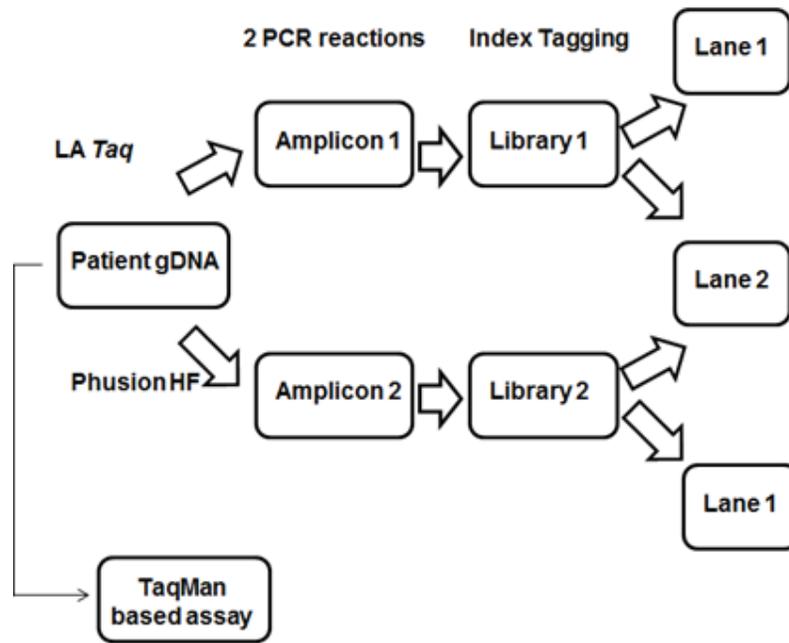


Figure 2.

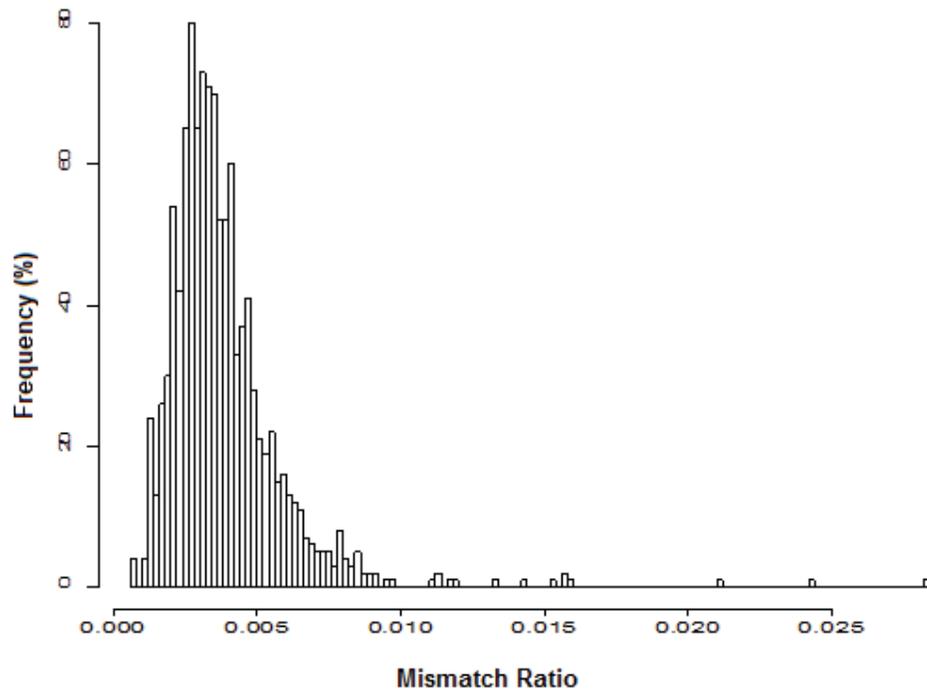
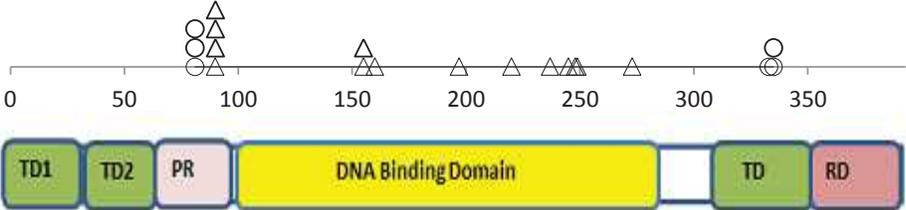


Figure 3.

- Silent mutation
- △ Missense mutation



TD1/2 – transactivation domains; PR – proline-rich regions; TD – Tetramerization domain; RD – regulatory domain; mut - mutation

Figure 4.

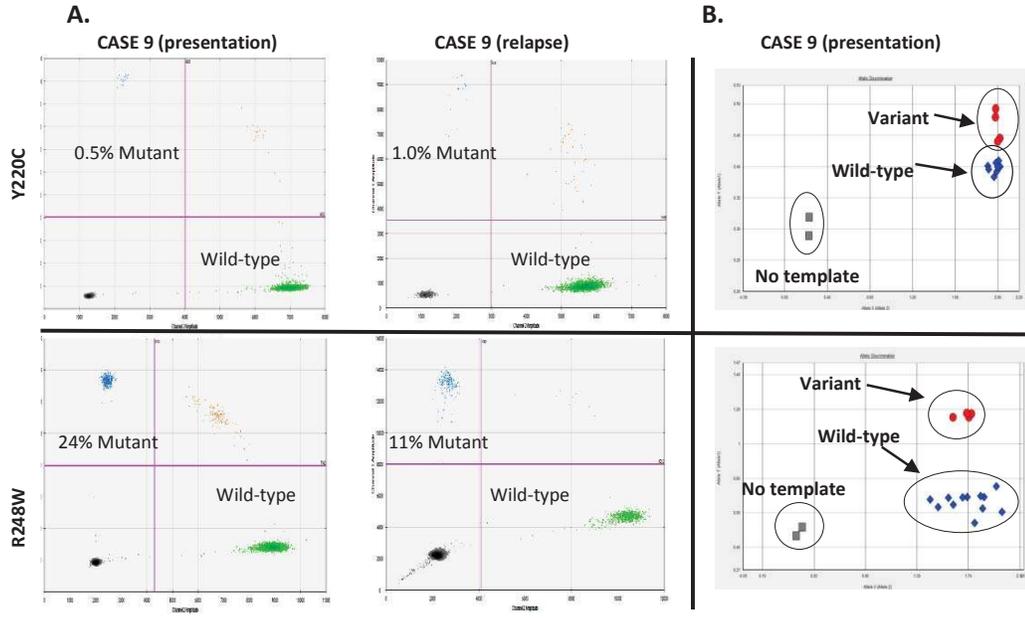


Figure 5.

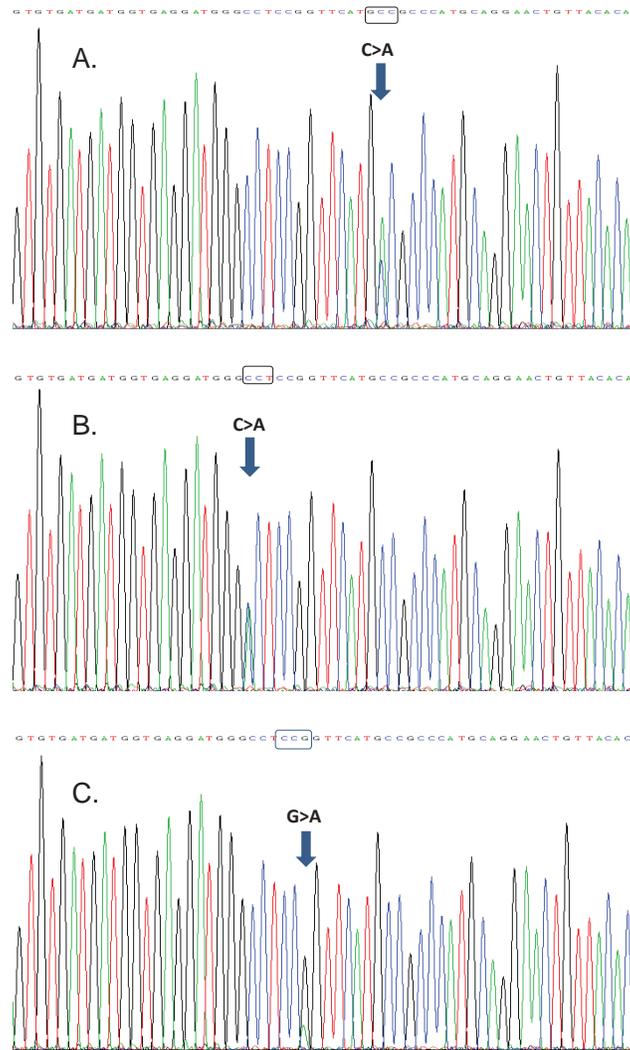


Figure 6.

