Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease OPEN

M J J Rose-Zerilli, J Gibson, J Wang, W Tapper, Z Davis, H Parker, M Larrayoz, H McCarthy, R Walewska, J Forster, A Gardiner, A J Steele, C Chelala, S Ennis, A Collins, C Oakes, D G Oscier, J C Strefford

Cite this article as: M J J Rose-Zerilli, J Gibson, J Wang, W Tapper, Z Davis, H Parker, M Larrayoz, H McCarthy, R Walewska, J Forster, A Gardiner, A J Steele, C Chelala, S Ennis, A Collins, C Oakes, D G Oscier, J C Strefford, Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease, *Leukemia* accepted article preview 5 February 2016; doi: 10.1038/leu.2016.10.

Received 14 September 2015; revised 21 December 2015; accepted 4 January 2016; Accepted article preview online 5 February 2016

Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease

Running title: Exome sequencing of progressive M-CLL

Matthew JJ Rose-Zerilli [1*], Jane Gibson [1*], Jun Wang [2], William Tapper [3], Zadie Davis[4], Helen Parker [1], Marta Larrayoz [1], Helen McCarthy [4], Renata Walewska [4], Jade Forster [1], Anne Gardiner [4], Andrew J Steele [1], Claude Chelala [2], Sarah Ennis[3], Andrew Collins [3], Christopher Oakes [5], David G Oscier [1,4] and Jonathan C Strefford[1]

[1]Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[2]Bioinformatics Unit, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[3]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[4]Department of Haematology, Royal Bournemouth Hospital, Bournemouth, United Kingdom

[5] Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, USA

* These authors share first authorship

Correspondence to: Professor Jonathan C Strefford, Cancer Genomics Group, Academic Unit of Cancer Sciences, Somers Cancer research Building, Southampton General Hospital, Tremona Road, Southampton SO16 6YD. Tel: +44(0)23 8120 5246. E-mail: JCS@soton.ac.uk

Text word count: 3999/4000

Abstract word count: 199/200

Number figures/tables: 4/1

Number references: 44/100

Number supplementary figures/tables: 4/9

Key Points

- Disease progression in *IGHV-M* CLL with 'good-risk' cytogenetics is frequently associated with co-evolution of 'poor risk' driver mutations and DNA methylation changes.
- Drug resistance in *IGHV-M* CLL may be consequent upon the emergence of an *IGHV-U* clone

Abstract

The biological features of *IGHV-M* CLL responsible for disease progression are still poorly understood. We undertook a longitudinal study close to diagnosis, pre-treatment and post relapse in thirteen patients presenting with cMBL or Stage A disease and good risk biomarkers (*IGHV-M* genes, no del(17p) or del(11q) and low CD38 expression) who nevertheless developed progressive disease, of whom ten have required therapy. Using cytogenetics, FISH, genome-wide DNA methylation and copy number analysis together with whole exome, targeted deep- and Sanger sequencing, at diagnosis we identified mutations in established CLL driver genes in nine (69%), non-coding mutations (*PAX5* enhancer region) in three, and genomic complexity in two patients. Branching evolutionary trajectories predominated (n=9/13), revealing intra-tumoural epi- and genetic heterogeneity and sub-clonal competition prior to therapy. Of the patients subsequently requiring treatment, two had sub-clonal *TP53* mutations that would not be detected by standard methodologies, three qualified for the very-low risk category defined by integrated mutational and cytogenetic analysis and yet had established or putative driver mutations and one patient developed progressive, therapy-refractory disease associated with the emergence of an *IGHV-U* clone. These data suggest that extended genomic and immunogenetic screening may have clinical utility in patients with apparent good risk disease.

Introduction

Clinical heterogeneity within CLL, especially in the majority of patients presenting with a low tumor burden, provides a continuing impetus for the discovery of prognostic biomarkers.

Immunogenetic features such as *IGHV* mutation status and stereotypy, immunophenotypic markers, genomic abnormalities and serum markers have prognostic significance. A recently described prognostic index incorporating gender, age, performance status, *IGHV* mutation status, deletions of 11q and 17p, serum B2 microglobulin and thymidine kinase distinguished four risk categories with differing 5-year overall and progression-free survivals (1).

Candidate gene approaches and next generation sequencing have led to the discovery of mutations in many genes, including *TP53, ATM, NOTCH1, SF3B1, BIRC3, SAMHD1, EGR2* with prognostic and/or predictive significance, even when first detected as small sub-clones in the case of *TP53* mutation (2-10). A recent whole genome study demonstrated the adverse prognostic significance of multiple driver mutations and implicated novel non-coding mutation (11). Mutations in an intergenic region on 9p13 correlated with reduced *PAX5* expression and 3'UTR *NOTCH1* mutations associated with a poor outcome comparable to cases with an exon 34 *NOTCH1* mutation. Retrospective analyses of non-trial cohorts show that integration of a restricted set of mutations with copy number data refines and enhances the prognostic significance of the latter and suggest that mutations may be incorporated into future prognostic indices (12). Furthermore, copy number array and next generation sequencing data inferred from a single time point or from sequential studies, have demonstrated intra-clonal heterogeneity in CLL, the prognostic significance of sub-clonal mutations and the selective pressure of therapy in determining clonal evolution (13, 14). Recent epigenetic data has identified three CLL subtypes that correlate with B-cell maturity and possess distinct patterns of somatic instability, degree of *IGHV* mutation, mutation risk profiles and clinical outcomes (15-18). Despite this progress there remain patients who would be classified as 'low-risk' based on biomarkers who nevertheless have progressive disease.

To obtain more information about the genomic and epigenomic landscape and clinical significance of abnormalities in *IGHV-M* CLL (M-CLL), we performed a longitudinal study at or close to diagnosis, pre-treatment and post-relapse in thirteen patients. These patients presented with Binet Stage A disease (n=10) or cMBL and good-risk biomarkers (*IGHV-M* genes, no del(17p) or del(11q) and low CD38 expression), ten of whom subsequently required treatment. Using a combination of DNA methylation, copy number analysis, whole exomic sequencing (WES), targeted deep sequencing (TDR) of recurrently mutated CLL driver genes, screening of non-coding mutation and immunogenetic analysis, we identified the presence or acquisition of clonal or sub-clonal driver mutations and DNA methylation changes in eight cases and the emergence of a new

immunogenetic clone in one case.

Methods

Patient data, copy number and methylation analysis

We studied 13 patients diagnosed at the Royal Bournemouth Hospital between 1992-2007 as cMBL or Binet Stage A, Rai stage 0 CLL according to the 2008 IWCLL/NCI guidelines (19). This study was approved by the local REC and informed consent obtained according to the Declaration of Helsinki. *IGHV* sequencing, CD38, cytogenetic and FISH analyses were performed as described (20-23) and only cases with mutated *IGHV* genes (excluding major stereotypes), low CD38 expression and lacking 11q or 17p deletion and the availability of stored material were included. Germline DNA (GL) was obtained from saliva (DNAgenotek). CD19+ B-cells were taken at a median of 1 year (0-7.3 years) from diagnosis when patients had cMBL or Stage A disease (time-point 1 (TP1)). The three patients (pts 1-3) who did not require treatment remained as Stage A with a rising (n=2) or stable lymphocyte count. All three were sampled again (TP2) at a median of 7 years (6-10) from TP1 and one was sampled at further time-point (TP3) 3 years after TP2. In the ten patients requiring treatment, a further sample was taken at a median of 4 months, (range: 0-42) pre-treatment (time-point 2 (TP2)). 6/10 patients who relapsed after first line treatment had a sample taken post relapse (time-point 3 (TP3)) and 2/6 were also sampled at relapse following subsequent treatments (TP4 and 5). For 13 sample-trios (GL, TP1, TP2), DNA regions of copy number alteration (CNA) and differential methylation were identified using SNP6 arrays (Affymetrix) and 450K arrays (Illumina) respectively, as described (7, 24).

Sequencing

WES libraries were prepared from 13 sample-trios (GL, TP1, TP2) as described (25). TDR used Haloplex (Agilent) as described (26), to capture SNVs identified by WES and 22 genes (exons and 5' & 3'-UTRs) that are frequently mutated in CLL (**Table S1**) in all tumour samples. TDR libraries were sequenced at high depth (average x4000) to detect mutation down to the 1% level. For each mutation detected by TDR, variant allele frequencies (VAF) were adjusted for tumour purity estimated as %CD19+ cells. Clonal or sub-clonal mutations were further classified according to (26). We subjected the TDR data to **SciClone** analysis (27) to define the clonal dynamics of mutation clusters into three types: **1)** Static: clusters remain the same over time. **2)** Expanding: all mutations in a cluster increase over time. **3)** Evolving: new mutations in later samples or one or more mutations in a cluster increase over time. **Phylosub** was used for tumour phylogeny analysis to predict the most likely order of mutation events and classify either linear or branching evolution patterns (28). *PAX5* enhancer region was screened as described in (11). We defined mutated genes into those recurrently mutated from previous CLL studies, non-coding mutation and genes mutated in other hematological malignancies (All excluding copy number changes). **Supplementary methods are available on-line.**

Results

Genomic landscape of progressive M-CLL

Clinical features, treatment regimens and the genomic landscape at multiple time points are summarized in **Table 1** and **Figure 2**.

We employed WES and TDR to identify somatically-acquired mutation in tumour samples from 13 cases with mean coverage of 77x (min-max: 43-127) and 3681x (2142-5268), with more than 86% of all bases covered at >20x and >200x, respectively. (**Figure 1, Tables S2 and S3**). Of the filtered WES variants (**Table S4**), TDR confirmed the presence of 224/312 (72%) SNVs and 7/9 (78%) indels (when present at both TP1 and TP2), respectively (**Table S5 and Supplementary methods**). We used the TDR variants to study temporal clonal evolution and demonstrated that our TP1 and TP2 samples harbored a similar mutation burden, with on average 17 (min-max: 9-26) and 19 (8-29), respectively. After adjusting for tumour purity, we observed no difference in the mean number of clonal (6 v 6) or sub-clonal mutations (10 v 12) in either untreated time-point (TP1 v TP2). All reported variant allele frequencies (%VAF) are adjusted for tumour purity.

Focusing on genes previously shown to be recurrently mutated in CLL; at TP2, clonal mutations were detected in *MYD88* (p.L265P) (pt-2), and *CHD2* (pt-1) among the 3 untreated patients and in *ATM* (pt-4), *DDX3X* (pt-13), *NOTCH1* (pts-6,13), *SF3B1* (pts-6,8), *TP53* (pts-8,9), *NFKBIE* (pt-5), *SPEN* (pt-9), *ZMYM3* (pt-6), *KLHL6* (pt-10), *BIRC3* (pt-13) and *IRF4* (pt-13) among the 10 patients who received treatment. Only five of these mutations were clonal (**Figure 2**). In addition, three patients (pts-3,11,12) exhibited missense mutations (damaging by Polyphen-2) in genes known to have a role in other hematological tumors, *LTF* (pt-3), *ITGA6* (pt-11) and a frame-shift in *TNFAIP3* (pt-12) and all were present at sub-clonal levels (11-42 %VAF). Only one case (pt-7) lacked any recurrently mutated driver mutation documented in CLL or other hematological malignancies. However, WES did identify ten mutated genes from which two candidates emerged, namely missense mutations in *ZBTB7C* a kidney cancer-related gene that interacts with p53 (ref: 29) and *S1PR4* a receptor expressed in hematopoietic cells that interacts with *MAPK3* (ERK1), placing it in the B-cell receptor pathway (30).

The CLL driver mutations (with the exception of *BIRC3* in pt-13) were detectable (by the presence of one or more mutated reads) at TP1 supporting the hypothesis that identification of mutations at diagnosis may identify individuals later requiring therapy. The CLL driver gene mutations with VAFs <1% (*IRF4*, *NOTCH1, SF3B1 and TP53;* in pts-6,8,9,13) at TP1 were ascertained by manual curation of the TDR sequencing reads (**Table S6**) after being originally detected in later tumour time points with a higher VAF, suggesting a larger sub-clonal population at progression. Pileup of reads across all samples provided statistical confidence for calling *TP53* mutation in patient-8 and 9 below a 1% VAF (P=0.013 & <0.001; **Table S6**). Droplet-digital PCR analysis of

patient-13 confirmed the presence of the *NOTCH1* mutation at TP1 (**Figure S1**). Together, this would suggest sequencing depths much greater than x4000 will be required to robustly identify all sub-clonal mutations, for example, patient-8 had a *TP53* mutation (29 %VAF) at TP2, detectable at TP1 in 9/15581 reads (0.1 %VAF), equating to the presence of one mutant cell in ~1000 CLL B-cells. In patient-13, the *BIRC3* mutation at TP2 was not identified at TP1 (0/3624 reads) and conversely a clonal *DDX3X* mutation at TP1 was detected as a small sub-clone at TP2. This lead to a re-appraisal of this case which is discussed in detail later. At relapse, we identified mutations in *SF3B1* (pt-6) and *TP53* (pt-9) with VAF's of 17 and 3.3% respectively which had VAF's of <1% pre-treatment.

We screened for non-coding mutations (11). *PAX5* enhancer region mutations were detected in three patients, estimated at TP1 to be clonal in one case (pt-4) and sub-clonal in the other cases (pts-2,6) (**Figure 2 and S2**). These mutations co-occurred with other mutations: *MYD88* (pt-2), *ATM* (pt-4) and *NOTCH1, SF3B1* and *ZMYM3* (pt-6). We also detected at TP2, the presence of a sub-clonal mutation (6 %VAF; chr17:56408615:T>C) in the mature sequence of hsa-mir-142, this co-occurred with a *ITGA6* mutation in patient-11 (**Figure S2**). The *NOTCH1* 3'UTR mutations, previously observed solely in cases of U-CLL (11), were absent.

Combining karyotypic, FISH and SNP6 data, at TP1, only two patients (pts-8, 10) had no copy number abnormality or translocation, while the remainder had mono (n=5) or mono + biallelic loss of 13q14. Two patients (pts-12,13) with del13q also had trisomy 12. Two patients (pts-6,9) had a complex genome (≥3 CNAs), defined as previously reported (31). SNP6 confirmed the absence of 11q or 17p deletion (**Table S7**). At TP2, additional abnormalities were detected in 3 patients (pts-4,5,8). Interestingly, the complex karyotypic abnormality in patient 8 was associated with expansion of a sub-clonal *TP53* mutation (from 0.1% at TP1 to 29% at TP2) without *TP53* loss. Three patients (pts-2,4,5) had an unbalanced translocation. Patient-13 had a remarkable change in copy number and is discussed later.

Intra-clonal heterogeneity in progressive M-CLL

Our longitudinal approach provided an opportunity to evaluate intra-clonal heterogeneity both before and following therapy.

SciClone analysis of TP1 and TP2 data from patients prior to therapy, enabled us to make the following observations: **1)** Four cases (pts-5,7,10,12) had a static sub-clonal structure with mutation clusters present at similar VAFs at both time points (pt-12 in **Figure S3A**). **2)** Two cases (pts-2,4) had an expanding population where all mutations in a cluster were more dominant at TP2 (pt-4 in **Figure S3A**). **3)** Seven patients (pts-1,3,6,8,9,11,13), had an evolving genome where new mutations appeared (n=8 in 4 patients (pts-6,8,9,13) and/or one or more mutations in a cluster increased, at TP2 (pt-6 in **Figure S3A**). Six of these new mutations had low read depths (< x4000; ranging: 138-3642) in TP1 samples, suggesting there may be a lack of detection

sensitivity. The three remaining mutations (*DSG4*, *SIM1*, *SLC8A2*) had adequate depth (4218-9043) at TP1, suggesting these mutations are either very rare (in <0.5-1% of cells) or represent acquired mutations at disease progression (TP2). For the six patients with post therapy time points, there were no new mutations and we observed two patients (pts-5,12) exhibiting a static structure with same distribution of sub-clones pre- and post-treatment (pt-12 in **Figure S3B**), while four cases (pt-4,6,9,13) showed expansion (pt-4 in **Figure S3B**) or evolution (pt-6 in **Figure S3B**) following an apparent therapy-related sweep selecting resistant/fitter sub-clones. Interestingly, these four cases had either expanding or evolving mutation clusters before therapy. Results for the remaining patients are provided in **Supplementary Figure S4**.

Phylosub analysis predicted a linear evolutionary path, where progeny replaced ancestral clones, in a minority of patients (n=4) in whom SciClone analysis had identified either static (pts-7,12) or expanding (pts-2,4) mutation clusters. Phylosub also predicted that the *ATM*, *MYD88, S1PR4, TNFAIP3* and *ZBTB7C* mutations in these patients were early evolutionary events (placed in the 1st or 2nd nodes of each tree). Complex branching trajectories were predicted in the remaining 9 patients, including two of the three patients (pts-1,3) with no indication for therapy, and provided the following insights: **1)** The *BCL2*, *CHD2*, *NOTCH1*, *SF3B1* and *TLR4* mutations were all predicted to be early events. **2)** Generally located at branch points, the sub-clones that appear to have good fitness, or are selected for at later tumour time-points, contained CLL drivers (*ITPKB*, *NFKBIE*, *SF3B1*, *TP53*, *ZMYM3*) or genes mutated in other haematoloical malignancies (*ITGA6*, *LTF*) supporting the role of genes in the latter two categories as candidate drivers of progression in those patients. **3)** Convergent evolution was only found in two patients (pts-4,8) who exhibited two mutations in a single gene (*ATM*:p.I2606M/p.Q2733K and *IGLL5:*p.C31Y/p.P50S). All four mutations were clonal and only the *IGLL5* mutations were close enough to be detected on separate overlapping reads, but as the VAFs were similar (54-57%) they could have arisen in the same cell.

Phylosub and SciClone analysis of patients-4,6,9 and 12 is displayed in **Figure 3.** These four patients had three (pts-4,6,12) or four (pt-9) tumour time-points presenting static, expanding or evolving mutation clusters with predicted linear (pt-4,12) or branching (pts-6,9) evolution trajectories before and after therapy. At TP2 in patient-4 we observe expanding populations (nodes C-E) replacing the ancestral population, suggesting that these later mutations are associated with disease progression. Following treatment with bendamustine plus rituximab (BR) (TP3), we observed no reduction in population frequencies. Conversely, in patient-12, we observed a reduction back to baseline (TP1) for nodes C-E, suggesting a similar sensitivity of the descendant clones to Chlorambucil-Rituximab therapy; this patient remains in remission. Interestingly, in patient-6 the *ITPKB*, *SF3B1* and *ZMYM3* mutations were predicted to be in distinct populations (nodes I, J and E, respectively) and following chlorambucil treatment at TP3 the population frequencies of both the I (*ITPKB*) and J (*SF3B1*) nodes increased in comparison to other nodes. In patient-9 following two rounds of therapy (chlorambucil and

BR) we observed 75% del(17p) loss by FISH and a sub-clonal *TP53* mutation (p.Y234C, 3.3 %VAF), at TP4. A complex karyotype was also observed at TP3. The difference between del(17p) FISH clone size and %VAF of the *TP53* mutation would suggest evolutionary independent events, with a rare *TP53* mutated clone detectable at presentation and later acquisition of del(17p) loss in another *TP53*-wildtype clone. Following exposure to chemo-immunotherapy, the resistant *TP53* aberrant clones accumulate and dominate the tumour. Phylosub results for the remaining patients are provided in **Supplementary Figure S4**.

The emergence of an *IGHV-U* immunogenetic clone can drive progression

Substantial differences in the clone size of CNA's and mutations between TP1 (diagnosis) and TP2 (+8 years) in patient-13, and the detection of a *BIRC3* mutation only at TP2, led to a review of karyotypic, FISH, SNP6 and mutational data and targeted re-sequencing analysis of samples taken after treatment with chlorambucil, BR and ofatumumab, all of which were ineffective. This showed a remarkable temporal shift in genomic aberrations supporting a dominant population at diagnosis containing a deletion of 13q, loss of chromosome Y and a unique set of mutations, including *DDX3X, HEPH, RARB* and *TEC.* These were gradually replaced by a 47,XY, trisomy 12, population with mutations in 18 genes including *BIRC3*, *NOTCH1* and *IRF4* (**Figure 4 and Table S8**). As the dominant mutations present at TP2 are more frequently or exclusively associated with *IGHV-U* genes, we reanalyzed the *IGHV* status at TP2, and identified a dominant *IGHV5-10-1*01* (100% identity to germ-line) clone in addition to the *IGHV3-48* clone with 92% germ-line identity present at diagnosis. *IGHV* analysis of six intermediate samples between TP1 and 2 detected the *IGHV-U* clone as far back as 4 years post diagnosis (**Table S9**). Importantly, both the *NOTCH1* mutated and trisomy 12 sub-clones were detectable at diagnosis using TDR and FISH respectively, demonstrating the presence of the *IGHV5-10-1*01* clone at diagnosis but at a level which was undetectable using standard immunogenetic assays. This patient's tumour eventually transformed to a diffuse large B-cell lymphoma, using the *IGHV5-10-1*01* clone. Unfortunately, TDR of the Richter's node biopsy was not successful.

Additional *IGHV* sequencing on the other cases failed to identify any additional patients with evolution of an *IGHV-U* clone (**Table S9**).

DNA methylation subtyping and co-evolution of epigenetic changes

We performed clustering analysis of the TP1 methylation data together with a reference sample set where the three epigenetic subtypes were defined previously (Oakes, in press) (32) and determined that 12/13 patients belonged to the high-programmed CLL DNA methylation subtype, consistent with the selection of our patients based on the presence of mutated-*IGHV* (**Figure S5A**). In all except a single patient, adjacent clustering of the TP2 data confirmed the clonal relationship between tumour time-points. The exception was patient 13 which clustered in the low-programmed CLL subtype at TP2, consistent with the emergence of the *IGHV-U* clone.

Patients with limited genetic evolution had relatively few differences in overall methylation, where as those that exhibited either expansion or evolution of genetic sub-clones showed higher proportions of altered CpG methylation (**Figure S5B**).

Discussion

Patients with *IGHV-M* genes, defined as <98% identity to the germ-line sequence, have a better outcome than those with *IGHV-U* genes (33, 34). While stable cMBL and Stage A CLL are strongly enriched for cases with M-CLL, 37-39% of patients entered into the UKCLL4 and CLL8 trials of first line therapy had M-CLL (35, 36) and the key biological features responsible for progression are still poorly understood (11). M-CLL is biologically heterogeneous and studies have shown that CD38, CD49d and ZAP70 expression, serum markers, stereotypic subset-2 (37), telomere length, del(11q) and del(17p), and genomic abnormalities, influence time to first treatment or outcome following therapy (38-43). In this study, we performed longitudinal genomic and epigenomic characterization before and after therapy in a cohort of M-CLL cases presenting with Stage A disease or cMBL. All cases lacked established biomarkers associated with progression: namely, high CD38 expression, del(11q), del(17p). Nevertheless, 10/13 cases subsequently required treatment.

Prior to treatment we found mutations in genes that are recurrently mutated in CLL, (*ATM*, *BIRC3*, *CHD2*, *DDX3X*, *IRF4*, *ITPKB*, *KLHL6*, *MYD88, NOTCH1*, *NFKBIE, SF3B1*, *SPEN*, *TP53 and ZMYM3*) in nine patients (69%) with a mean of one mutation per case (min-max: 0-4). This mutation incidence is consistent with a recent whole genomic and exomic study, where 83% of M-CLL cases had a driver mutation (11). Nine patients had one or more clonal mutations detected pre-treatment and the majority of the sub-clonal mutations were detectable at the earliest time-point at or soon after diagnosis. While many of the above genes are associated with disease progression and/or resistance to treatment, the clinical significance of others (*ITPKB, KLHL6 and SPEN*) is less certain. A further three patients had mutations in other genes (*LTF*, *ITGA6* and *TNFAIP3*) implicated in other haematological malignancies (COSMIC v 73) (Ref: 44).

We also screened for non-coding mutation in the *PAX5* enhancer region. These mutations co-existed with *MYD88*, *ATM, NOTCH1*, *SF3B1* and *ZMYM3* mutations in contrast to the original description where they were either the sole recurrent mutation or occurred in conjunction with 13q loss (11).

SciClone and Phylosub analysis provided novel insight into the extent of intra-clonal heterogeneity. Clonal expansion or evolution, predominantly in a branching pattern, was found in nine cases pre-treatment, indicating that sub-clonal competition occurs in the absence of selective pressure through therapy, with the resistant/fitter sub-clones dominating the tumour at post therapy time-points. A recent WES study comparing matched pre-treatment and relapse samples demonstrated that clonal evolution was the rule after therapy and the resistant clone could be detected before treatment in ~30% of cases (14). Six cases developed isolated splenomegaly, two had splenomegaly and lymphadenopathy and two had lymphadenopathy prior to treatment. Further spatial-temporal studies will be required to determine the site(s) of clonal evolution. Phylosub analysis also demonstrated the selection of sub-clones containing either *SF3B1*, *TP53, ITGA6*, *ITPKB*,

LTF, *NFKBIE*, or *ZMYM3* mutations, supporting their role as candidate drivers of progression.

One unexpected finding was the emergence of an *IGHV-U* clone in one patient who after six years of stable disease, developed progressive, therapy-refractory disease, culminating in a clonally-related lymphomatous transformation. Each clone had a unique spectrum of gene mutations and epigenetic profiles consistent with two distinct and competing leukemic clones originating from a different pool of lymphocyte progenitors. Bi, or more rarely multi-clonal *IGHV* rearrangements have been documented. From a cohort of 1147 cases, Plevova et al, identified seven cases with both mutated and unmutated clones in which serial studies showed diminution of an *IGHV-M* clone with persistence of a co-existing *IGHV-U* clone, resulting in re-classification to U-CLL (45). Clinically, this was associated with progressive lymphocytosis, disease progression and in some cases, the selection of a *TP53*-defect post therapy. Our case is unusual in that the unmutated clone was not detectable until 4 years after diagnosis using standard methodologies for *IGHV* sequencing, even though a sub-clonal trisomy 12 and *NOTCH1* mutation, associated with the unmutated clone were detected at diagnosis using more sensitive techniques.

Previous whole genome longitudinal studies of copy number (46-48) and/or genomic mutations (49-52) in CLL have included a higher percentage of cases with U-CLL than M-CLL and with progressive rather than stable disease. However, a picture has emerged of clonal evolution which is usually branched rather than linear and is more frequent in cases with progressive disease who have required treatment for which the majority of driver mutations can be detected at initial testing, often as small sub-clones. Our study confirms many of these findings but also highlights the high frequency of CLL driver mutations in a progressive cohort miss-labelled as 'good-risk' and the extent of clonal evolution prior to therapy. DNA methylation analysis revealed that co-evolution of genetic and epigenetic changes is a prominent feature and that this exists regardless of *IGHV* subtype and mutational risk assessment, supporting the perspective that evolution is an important predictor of disease progression (13, 14). A recent study (53) also found the highest number of differentially-methylated CpGs were in cases with genetically evolving and expanding sub-clones.

From a clinical perspective, a key question is whether the additional information that genomic and epigenomic screening provides in this group is likely to improve patient outcome. While it would be unwise to draw general conclusions from this small study, it does offer three examples where screening could have clinical utility. Firstly, 2/13 cases had small *TP53* mutated clones early in the disease with evidence of clonal selection post therapy. These cases had no *TP53* loss detectable by FISH when the mutant clone was initially detected. Secondly, 6/11 cases with 13q loss fell into the 'very-low' risk category defined by Rossi at al, in which predominantly *IGHV-M* cases with isolated 13q loss, lacking mutations in *TP53, BIRC3, NOTCH1, SF3B1* and *MYD88* had a prolonged TTFT and an expected OS similar to the matched general population (12). Two of our six cases with isolated 13q loss had a progressive lymphocytosis with no indication for treatment and a

mutation in *CHD2* or *LTF* while three of the four who required treatment, had mutations in *NFKBIE, TNFAIP3* or *ITGA6*, suggesting that more extensive screening may aid in the differentiation of cases destined to have stable or progressive disease. Finally, the emergence of an *IGHV-U* clone, not evident at diagnosis, which eventually lead to the patient's death is a rare event but supports either repeat immunogenetic analysis in cases with unexplained progressive or therapy-refractory disease, or the use of more sensitive assays capable of detecting small clones. In summary this study does not evaluate the role of other factors such as cell signalling as an explanation for disease progression, but does support the role for sequential genomic/epigenomic screening as a means of identifying potential driver mutations and predicting progressive disease.

Acknowledgements

This study was funded by The Kay Kendall Leukaemia Fund (KKL 584) and Wessex Medical Research (Innovation fund 2011 R06). We thank the patients for supplying tissue and the infrastructure support from a CR-UK centre grant (C34999/A18087), Leukaemia and Lymphoma Research grant (12021) and ECMC grant (C24563/A15581).

Authorship Contributions

Contribution: This work was funded by grants awarded to J.C.S and M.J.RZ. M.J.RZ, J.F, H.P, ML and A.P performed the experimental work; A.G and A.P performed the molecular diagnostic assays. M.J.RZ, J.G and A.C conducted the exome analyses; C.O analyzed the methylation data; D.G.O contributed patient samples and data. D.G.O, M.J.RZ and J.C.S. initiated and designed the study; M.J.RZ, J.G, D.G.O and J.C.S. wrote the manuscript with contributions from A.C and C.O; and all authors critically reviewed the final manuscript.

Disclosure of Conflicts of Interest

The authors declare no competing financial interests.

Supplementary information is available at Leukemia's website

References

1. Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. Blood 2014;124(1):49-62.

2. Skowronska A, Austen B, Powell JE, Weston V, Oscier DG, Dyer MJ, et al. ATM germline heterozygosity does not play a role in chronic lymphocytic leukemia initiation but influences rapid disease progression through loss of the remaining ATM allele. Haematologica 2012;97(1):142-6.

3. Clifford R, Louis T, Robbe P, Ackroyd S, Burns A, Timbs AT, et al. SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. Blood 2014;123(7):1021-31.

4. Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Famà R, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. Blood 2014;123(14):2139-47.

5. Lionetti M, Fabris S, Cutrona G, Agnelli L, Ciardullo C, Matis S, et al. High-throughput sequencing for the identification of NOTCH1 mutations in early stage chronic lymphocytic leukaemia: biological and clinical implications. Br J Haematol 2014;165(5):629-39.

6. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011 Dec 29;365(26):2497-506.

7. Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. Leukemia. 2014 Jan;28(1):108-17.

8. Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood. 2014 May 22;123(21):3247-54.

9. Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia. 2015 Feb;29(2):329-36.

10. Guièze R, Robbe P, Clifford R, de Guibert S, Pereira B, Timbs A, et al. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. Blood. 2015 Oct 29;126(18):2110-7.

11. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015 Jul 22

12. Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. Blood. 2013 Feb 21;121(8):1403-12.

13. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Cell. 2013 Feb 14;152(4):714-26.

14. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. Nature. 2015 Oct 22;526(7574):525-30.

15. Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet. 2012 Nov;44(11):1236-42.

16. Queirós AC, Villamor N, Clot G, Martinez-Trillos A, Kulis M, Navarro A, et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. Leukemia. 2015 Mar;29(3):598-605.

17. Oakes CC, Claus R, Gu L, Assenov Y, Hüllein J, Zucknick M, et al. Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. Cancer Discov. 2014 Mar;4(3):348-61.

18. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell. 2014 Dec 8;26(6):813-25.

19 Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner 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(12):5446-56.

20. van Dongen JJ, Langerak AW, Brüggemann 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(12):2257-317.

21. Best OG, Ibbotson RE, Parker AE, Davis ZA, Orchard JA, Oscier DG. ZAP-70 by flow cytometry: a comparison of different antibodies, anticoagulants, and methods of analysis. Cytometry B Clin Cytom 2006;70(4):235-41.

22. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 1999;94(6):1848-54.

23. Smoley SA, Van Dyke DL, Kay NE, Heerema NA, Dell' Aquila ML, Dal Cin P, et al. Standardization of fluorescence in situ hybridization studies on chronic lymphocytic leukemia (CLL) blood and marrow cells by the

CLL Research Consortium. Cancer Genet Cytogenet 2010;203(2):141-8.

24. Parker H, Rose-Zerilli M, Parker A, Chaplin T, Chen X, Wade R, et al. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. Leukemia 2011;25(3):489-97.

25. Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, Forster J, et al. Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. PLoS One 2013;8(12):e83244.

26. Parry M, Rose-Zerilli MJJ, Ljungström V, Gibson J, Wang J, Walewska R, et al. Genetics and Prognostication in Splenic Marginal Zone Lymphoma: Revelations from Deep Sequencing. Clin Cancer Res. 2015 Sep 15;21(18):4174-83.

27. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol 2014;10(8):e1003665.

28. Jiao W, Vembu S, Deshwar AG, Stein L, Morris Q. Inferring clonal evolution of tumors from single nucleotide somatic mutations. BMC Bioinformatics 2014;15:35.

29. Jeon BN1, Kim MK, Choi WI, Koh DI, Hong SY, Kim KS, et al. KR-POK interacts with p53 and represses its ability to activate transcription of p21WAF1/CDKN1A. Cancer Res. 2012 Mar 1;72(5):1137-48.

30. Seitz G, Yildirim S, Boehmler AM, Kanz L, Möhle R. Sphingosine 1-Phosphate (S1P) Induces Migration and ERK/MAP-Kinase-Dependent Proliferation in Chronic Lymphocytic Leukemia (B-CLL) Due to Expression of the G Protein-Coupled Receptors S1P1/4. Blood (ASH Annual Meeting Abstracts) 2005 106: Abstract 4996

31. Ouillette P, Fossum S, Parkin B, Ding L, Bockenstedt P, Al-Zoubi A, et al. Aggressive chronic lymphocytic leukemia with elevated genomic complexity is associated with multiple gene defects in the response to DNA double-strand breaks. Clin Cancer Res. 2010 Feb 1;16(3):835-47

32. Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet 2015, in press.

33. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999 Sep 15;94(6):1840-7.

34. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999 Sep 15;94(6):1848-54.

35. Catovsky D, Richards S, Matutes E, Oscier D, Dyer MJ, Bezares RF, et al. Assessment of fludarabine plus

cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. Lancet. 2007 Jul 21;370(9583):230-9.

36. Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood 2014;123(21):3247-54.

37. Strefford JC, Sutton LA, Baliakas P, Agathangelidis A, Malčíková J, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. Leukemia. 2013 Nov;27(11):2196-9.

38. Del Giudice I, Morilla A, Osuji N, Matutes E, Morilla R, Burford A, et al. Zeta-chain associated protein 70 and CD38 combined predict the time to first treatment in patients with chronic lymphocytic leukemia. Cancer. 2005 Nov 15;104(10):2124-32.

39. Bomben R, Dal-Bo M, Benedetti D, Capello D, Forconi F, Marconi D, et al. Expression of mutated IGHV3-23 genes in chronic lymphocytic leukemia identifies a disease subset with peculiar clinical and biological features. Clin Cancer Res. 2010 Jan 15;16(2):620-8.

40. Lin TT, Norris K, Heppel NH, Pratt G, Allan JM, Allsup DJ, et al. Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. Br J Haematol. 2014 Oct;167(2):214-23.

41. Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJJ, Parker A, Lin TT, et al. Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. Leukemia. 2015 Aug 10.

42. Skowronska A, Parker A, Ahmed G, Oldreive C, Davis Z, Richards S, et al. Biallelic ATM inactivation significantly reduces survival in patients treated on the United Kingdom Leukemia Research Fund Chronic Lymphocytic Leukemia 4 trial. J Clin Oncol. 2012 Dec 20;30(36):4524-32.

43. 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. J Clin Oncol. 2011 Jun 1;29(16):2223-9.

44. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer. 2004 Jul 19;91(2):355-8.

45. Plevova K, Francova HS, Burckova K, Brychtova Y, Doubek M, Pavlova S, et al. Multiple productive immunoglobulin heavy chain gene rearrangements in chronic lymphocytic leukemia are mostly derived from independent clones. Haematologica. 2014 Feb;99(2):329-38.

46. Gunnarsson R, Mansouri L, Isaksson A, Göransson H, Cahill N, Jansson M, et al. Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. Haematologica. 2011 Aug;96(8):1161-9.

47. Knight SJ, Yau C, Clifford R, Timbs AT, Sadighi Akha E, Dréau HM, 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 Jul;26(7):1564-75.

48. Braggio E, Kay NE, VanWier S, Tschumper RC, Smoley S, Eckel-Passow JE, et al. Longitudinal genome-wide analysis of patients with chronic lymphocytic leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. Leukemia. 2012 Jul;26(7):1698-701.

49. Ouillette P, Saiya-Cork K, Seymour E, Li C, Shedden K, Malek SN. Clonal evolution, genomic drivers, and effects of therapy in chronic lymphocytic leukemia. Clin Cancer Res. 2013 Jun 1;19(11):2893-904.

50. 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 Nov 15;120(20):4191-6.

51. Ojha J, Secreto C, Rabe K, Ayres-Silva J, Tschumper R, Dyke DV, et al. Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. Leukemia. 2014 Dec;28(12):2395-8.

52. 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. 2015 Jan 15;125(3):492-8.

53. Smith EN, Ghia EM, DeBoever CM, Rassenti LZ, Jepsen K, Yoon KA, et al. Genetic and epigenetic profiling of CLL disease progression reveals limited somatic evolution and suggests a relationship to memory-cell development. Blood Cancer J. 2015 Apr; 5(4): e303.

Figure Legends

Figure 1: Study overview

(A) Inclusion criteria for study and definition of disease progression. (B) Tumour time point sampling time line for the 13 patients. Tx = treatment. **(C)** Flow diagram describing genomic analyses and result summaries. WES = Whole Exome Sequencing. TDR = Targeted Deep Re-sequencing. Example data plots for SciClone mutation clustering, Phylosub phylogenetic trees and concentric pie charts (each layer, inner to outer, is a sampling time point) displaying imputed SNV population frequencies at each phylogenetic node. *For indel filtering we accepted a high-false positive WES rate to ensure we could capture all of the 'true' somatically-acquired indel variants by TDR (**Supplementary methods**). When considering indels present in 2 or more tumour time-points (2+TPs) our indel TDR validation rate (78%, 7/9%) was in line with the SNV rate (72%).

Figure 2: Heat-map representation of tumour time-points analysed by WES and targeted deep re-sequencing

From top to bottom: Key to heat-map cell shading. Patient characteristics, light blue cell shading indicates patients with follow-on tumour samples (ie. TP3, TP4, TP5); dark grey cells indicated a positive result. Sub-clonal (light green cells) and clonal mutations (dark green cells) in each patient, grouped into recurrently mutated CLL driver genes, non-coding mutation described in Puente et al (6) and genes mutated in haematological malignancies. Numbers in cells denote tumor purity-adjusted %VAFs from TDR. SC = sub-clonal, C = Clonal from Sanger-seq traces. Presence of multiple productive-*IGH* relating to patient-13, chromosomal translocation or genome complexity is denoted by dark grey cells. SNP6.0 data for TP1 and TP2 samples.

Figure 3: PhyloSub analysis of TDR results in predicted linear and branching clonal evolutionary pattern

By columns **(A)** Two patient examples (pt-12 and 4) of linear evolution path, **(B)** Two patient examples (pt-6 and 9) of complex branching trajectories. **(A) & (B) top panel:** XYZ scatter-graphs displaying the SciClone mutation clustering analysis on TDR datasets from sequential tumour time points TP1 (x-axis; first tumour sample), 2 (z-axis; progression) and 3 (y-axis; post-treatment). Data point symbols denote a distinct mutation cluster and the x=y=z line is displayed as a dashed blue arrow and denotes no change in the tumor purity-adjusted %VAF of mutation clusters between time points (clonal equilibrium). Selected gene symbols are displayed adjacent to its corresponding mutation cluster. **(A) & (B) bottom panel:** Concentric pie charts (each layer, inner to outer, is an early to later sampling time point) displaying imputed SNV population frequencies at each phylogenetic node. Predicted phylogenetic tree structure (best model shown), with population frequencies for each node from Phylo-sub analysis. Blue and red boxes denote large changes SNV population frequencies prior to and after first-line treatment, suggesting ongoing clonal dynamics and selection by therapy, respectively. Selected gene symbols are displayed adjacent to the corresponding segment of the pie chart or phylogenetic node.

Figure 4: Evolution of multiple productive-*IGH* in CLL patient 13

(A) From top to bottom: Five tumour time points with corresponding clinical, cytogenetic and immuno-genetic data. Mutation heat-map representation of five tumour time-points analysed by targeted deep re-sequencing. Numbers in cells denote tumor purity-adjusted %VAFs from TDR. Cell colours are linked to the SNV population nodes/frequencies displayed in part B and C. Lighter shading indicates a sub-clonal mutation. Blue asterisks/del13q14 = M-CLL clone (IGHV3-48; 92% identity to germ-line; del(13q14)) and Red asterisks/Trisomy12 = U-CLL clone (IGHV5-10*01; 100% identity to germ-line; Trisomy 12). **(B & C)** Filled light blue and red boxes denote mutations and cytogenetic abnormalities inferred into the M-CLL and U-CLL clone, respectively. **From left to right:** Concentric pie charts (each layer, inner to outer, is an early to later sampling time point) displaying imputed SNV population frequencies at each phylogenetic node. Predicted phylo-genetic tree structure, with population frequencies for each node from Phylo-sub analysis. Best models are displayed for analyses using all mutations **(B)** or only mutations associated with either the M-CLL or U-CLL clone providing insights into the probable order of mutation **(C)**. Note from TP3 onwards the mutations associated with the M-CLL clone are not detectable by sequencing. Open blue and red boxes denote large changes SNV population frequencies prior to and after first-line treatment, suggesting ongoing clonal dynamics and selection by therapy, respectively.

Supplementary Tables:

Table S1: List of the 22 genes frequently mutated in CLL that were captured and resequenced by TDR

Table S2: Whole exome sequencing quality metrics and coverage statistics

Table S3: Targeted deep sequencing quality metrics and coverage statistics

Table S4: 786 mutations identified by whole exome sequencing

Table S5: Targeted deep sequencing validation of whole exome sequencing mutations

Table S6: Manually-curated established or putative mutations from the TDR data

Table S7: SNP6.0 copy number data

Table S8: Sequential targeted deep sequencing on patient 13

Table S9: Sequential IGHV analysis

Supplementary Figures:

Figure S1: Droplet-digital PCR analysis of the NOTCH1 delCT mutation in Patient-13. (**A-C**) X/Y-scattergraph of mutant positive (blue), mutant-wildtype double-positive (orange), wildtype positive (green) and negative droplets (black) Pinks lines are user-defined thresholds for each droplet cluster (**D**) Fractional abundance of mutant droplets in Germ-line, tumour TP1 and TP3 gDNA template from Patient-13 and a NGS NOTCH1 wildtype CLL control sample (**E**) Table displaying the droplet number for each sample in D and fractional abundance estimates (with 95% Confidence Intervals).

Figure S2: Non-coding mutation

(**Left panel**) Sanger-sequencing verified PAX5 enhancer region somatically-acquired mutations in three patients. Mutated base-pairs highlighted in green. Patients 2 and 6 have sub-clonal mutations at TP1. Patient-4 had a clonal mutation (overlapping base peaks) at TP1. (**Right panel**) A sub-clonal hsa-mir-142 mutation identified in whole exome sequencing (WES) data from patient 11 at disease progression (TP2). Mutated sequencing reads identified by blue C bases (6% VAF, 3/51 WES reads). Human mir-142 RNA sequence and mutation site in brackets. The red arrow indicates the position of the mutated base (on the mature micro RNA hairpin molecule.

Figure S3: SciClone analysis of mutation clusters pre and post first-line therapy

XY scatter-graphs display the SciClone mutation clustering analysis performed on TDR datasets from sequential tumour time points in columns **(A)** before therapy and **(B)** after first-line treatment for three example patients displaying either a static (pt-12; treated with Chlor R. *TEX13B* data point not displayed as it resides on Chr X), an

expanding (pt-4; BR) or an evolving (pt-6; Chlor) form of clonal dynamics prior to and after treatment. Data point symbols denote a distinct mutation cluster and the x=y line is displayed as a dashed arrow and denotes no change in the tumor purity-adjusted %VAF of mutation clusters between time points (clonal equilibrium). Selected gene symbols are displayed adjacent to its corresponding mutation cluster.

Figure S4: SciClone and Phylosub results for the remaining patients

(**Left panel**) XY scatter-graphs display the SciClone mutation clustering analysis performed on TDR datasets from sequential tumour time points TP1 (x-axis; first tumour sample) vs TP2 (y-axis; progression sample). Data point symbols denote a distinct mutation cluster and the x=y line is displayed as a dashed arrow and denotes no change in the tumor purity-adjusted %VAF of mutation clusters between time points (clonal equilibrium). Selected gene symbols are displayed adjacent to its corresponding mutation cluster. The table displays TDR %VAFs for each mutation, read depths, SciClone cluster assignment and probabilities (**Right panel)** Predicted phylogenetic tree structure (best model shown), with tumour time-point (TPn) population frequencies for each node from Phylo-sub analysis.

Figure S5: Programmed epigenetic subtypes and longitudinal stability of DNA methylation patterns

(A) Co-clustering heat-map of the top 1,000 most variable CpGs across all patient samples (black boxes and bracketed numbers) and 127 CLL cases (open boxes) previously assigned to one of the three primary epigenetic subtypes (12) termed: high (red dendrogram/boxes), intermediate (yellow dendrogram/boxes) and low (green dendrogram/boxes) programmed CLLs (HP-, IP-, and LP-CLLs, respectively). 12/13 cases clustered with HP-CLLs (patient-8 resides in IP-CLL group) and both time-points cluster adjacently, except for patient-13 where TP2 sample clusters with the LP-CLLs. Blue = hyper-methylated CpG, White = hypo-methylated CpG.

(B) Stacked bar chart of % of CpGs different between TP1 and TP2 450K DNA methylation datasets. Hyper- and hypo-methylation differences (Blue and white bars, respectively) of greater than 10% (out of a total of 459,625 CpGs analyzed) were tallied per case and were adjusted for differences in tumor cell content and copy number changes between samples. Patients are grouped by their genomic evolution status from SciClone analysis (Static, Expanding and Evolving). Patient-11 has a small % methylation difference probably corresponding to the evolution of only a minor sub-clonal mutation cluster at TP2 (cluster #3; **See supplementary figure S2**).

Table 1: Overview of patient biomarker and clinical data

		Tumour time point 1				Tumour time point 2							
Patient ID	**Age at Diagnosis**	**IGHV (% identity)**	**FISH/Karyo**	**%CD38 expression**	**Lymphocyte Count (x10^9/L)**	**IGHV (% identity)**	**FISH/Karyo**	**%CD38 expression**	**Lymphocyte Count (x10^9/L)**	**LDT**	**1st Treatment (Response)**	**Treatment at relapse (Response)**	**Current Status at last follow up (03/8/15)**
1	52	IGHV4-61 (93%)	del13q +/- (54%)	2	6	IGHV4-61 (93%)	del13q +/-,-/- (10, 85%)	1	142	> 1 yr	-	-	Stable CLL
2	72	IGHV3-73 (91%)	del13q 46, XY, der4(4)t(4;12)(q35;q13)	1	5	IGHV3-73 (91%)	no change	1	25	> 1 yr	-	-	Stable CLL
3	57	IGHV2-70 (93%)	del13q +/- (90%)	1	92	IGHV2-70 (93%)	no change	1	88	> 1yr	-	-	Stable CLL
4	61	IGHV4-59 (89%)	del13q +/- (6%) 45,X -Y, t(7,13)(a11.2;q14)	15	34	IGHV4-59 (89%)	del13q +/-,-/- (76, 8%), 45, X –Y, t(7;13)(q11.2;q14)	50	136	> 1 yr	BR (CR)	-	High risk MDS. Died.
5	79	IGHV4-61 (92%)	del13q +/-,-/- (9/86%) 46, XY, t(6;13)(q26;q14)	1	107	IGHV4-61 (92%)	del13q +/-,-/- (14/86%) 46, XY, t(6;13)(q26;q14)	-	198	> 1 yr	Chlor (CR)	BR (PR)	Stable CLL
6	70	IGHV3-48 (97%)	del13q +/- (88%)	1	20	IGHV3-48 (97%)	del13q +/-,-/- (53, 10%)	2	127	6-12 months	Chlor (GR)	BR (CR)	In remission
7	47	IGHV4-34 (92%)	del13q +/-,-/- (19, 72%)	6	59	IGHV4-34 (92%)	no change	1	81	> 1 yr	Chlor R (PR)	Alemtuz (CR, MRD +ve)	In remission
8	59	IGHV3-23 (96%)	normal	3	21	IGHV3-23 (96%)	del13q +/- (66%) 46, XY, del(9)(q21),t(12;15)(p11;q15)	1	48	> 1 yr	Chlor Of (CR)	-	In remission
9	74	IGHV3-7 (89%)	del13q +/- (54%)	1	17	IGHV3-7 (89%)	del13q +/- (91%) (+ del17p at TP3)	1	185	6-12 months	Chlor (PR)	BR (CR)	On Ibrutinib
10	56	IGHV3-48 (93%)	normal	5	181	IGHV3-48 (93%)	no change	2	139	> 1 yr	Chlor (PR)	Continuum	Stable CLL
11	64	IGHV3-23 (91%)	46 XY. No 13q FISH	1	58	IGHV3-23 (91%)	NT	1	145	> 1 yr	B Of (CR)	-	In remission
12	63	IGHV4-34 (96%)	47 XY,+12. No 13q FISH	1	42	IGHV4-34 (96%)	NT	1	77	6-12 months	Chlor R (CR)	BR (CR, MRD +ve) Continuum	In remission
13	67	IGHV3-48 (92%)	tri 12 (2%) del13q -/- (55%)	9	18	IGHV3-48 (92%) & IGHV5-10-1*01 (100%)	tri 12 (73%) (tri 12 (75%) at TP3)	21	158	> 1 yr	Chlor (NR)	BR (PR), Of (PR)	Richters Syndrome, NR to CHOP Of. Died.

Footnote: bold Patient ID = not treated by last follow up. NT = Not tested. LDT = Lymphocyte Doubling Time (from diagnosis for the first year of follow-up)

A

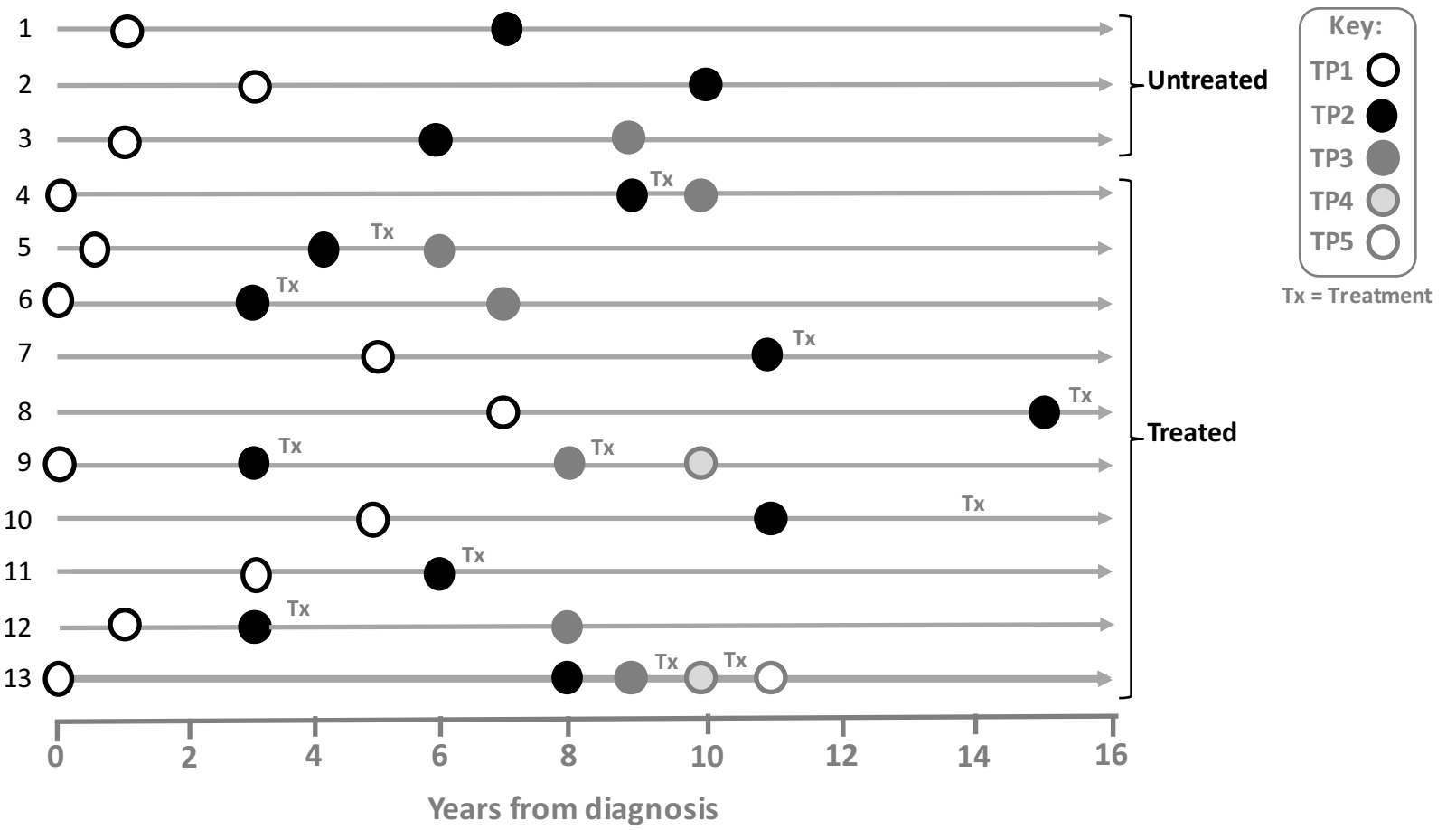
Stage of disease:
Binet Stage A (n=10)
cMBL (n=3)

Biomarker Profile:
IGHV-mutated
No *TP53* abnormality
No del(11q)
Low CD38 expression

Disease Progression:
First-line therapy (n=10)
OR
Rising lymphocyte count (n=3)

Availability of
post-treatment material (n=6)

B



C

Methods

WES/SNP6.0
Exome-seq mean coverage 77x (range 43-127). 86% bases covered at >20x

Samples:
TP1 & 2 tumors
(n=13 & 13)

TDR
All WES variants and exon re-sequencing of 22 CLL genes.
Mean coverage 3681x (range 2142-5268)
95% of all bases covered at >200x
Samples:
All tumor time points
(n= 36)

Bioinformatic analyses
1) **Purity adjustment** - equivalent to ABSOLUTE
2) **SciClone** - mutation clustering
3) **Phylosub** - Population frequencies and predicted phylogenetic trees

Results

SNVs
786 somatically-acquired mutations (missense, nonsense or frame-shift)*

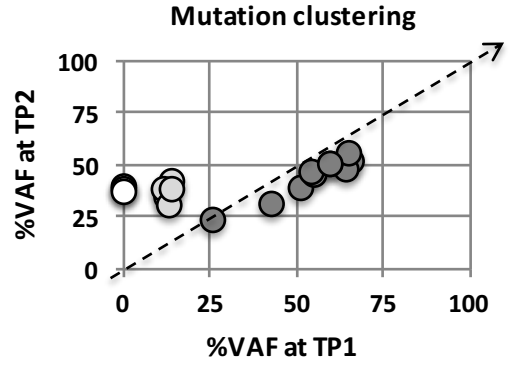
CNAs
85 acquired CNAs
Deletions = 75
Gains = 10

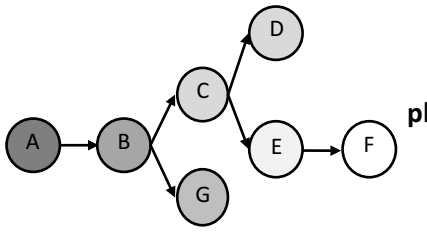
Validation of WES findings

SNVs = 224/312 (72%)

Indels = 27/474 (5.7%)*
Indels (2+TPs) = 7/9(78%)

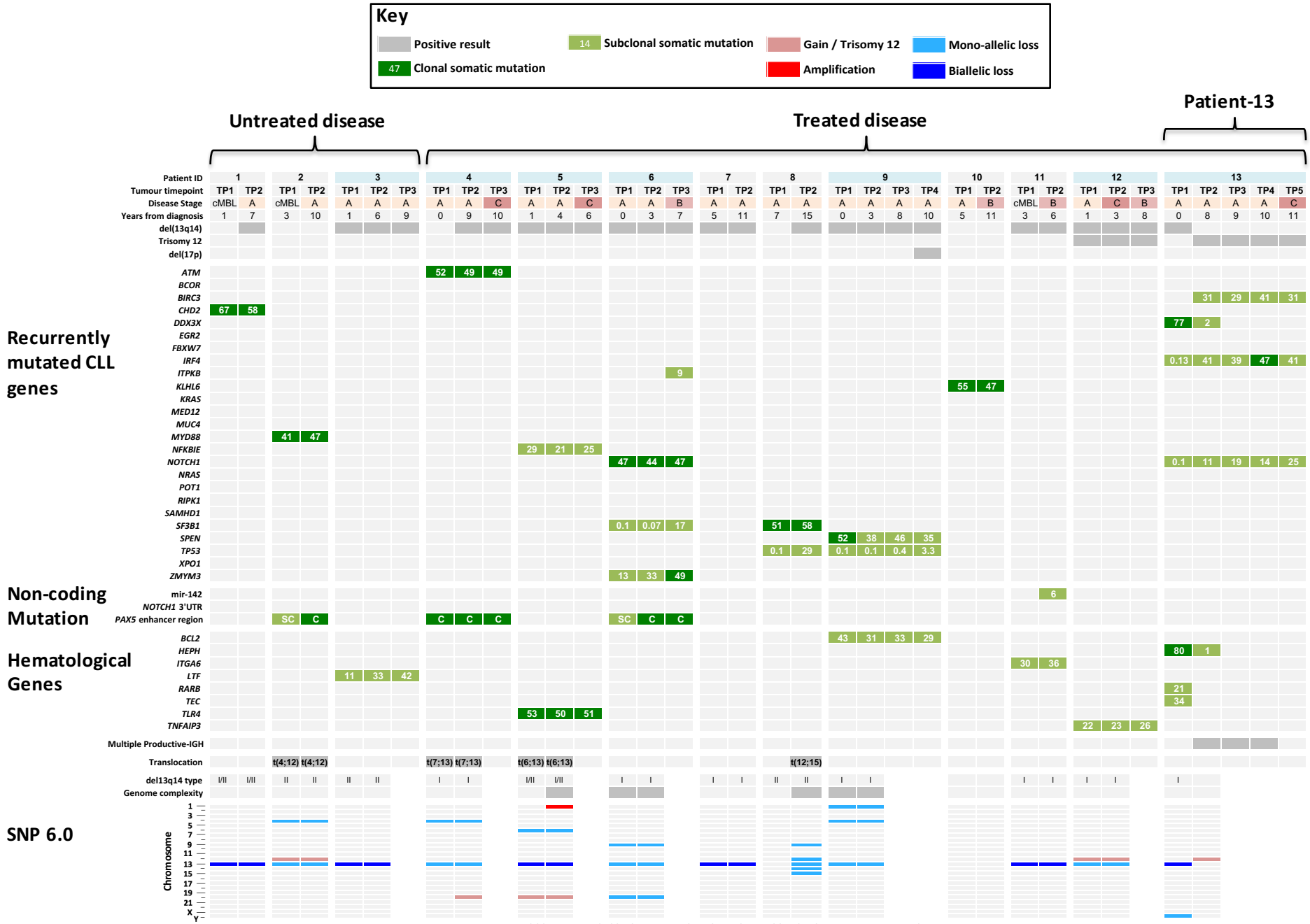
Analysis of temporal clonal dynamics
Sub-clonal population structures before and after 1st line therapy:
Static (n=4)
Expanding (n=2)
Evolving (n=7)
Linear (n=4) or Branching (n=9) evolution





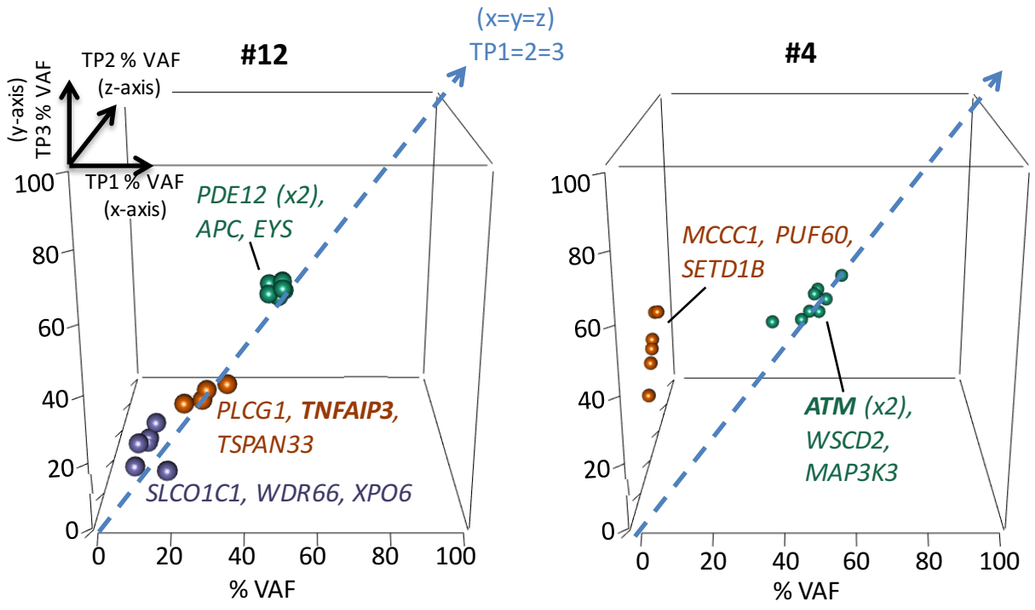
Predicted phylogenetic trees

SNV population Frequencies
(time points ordered from inner to outer circles)

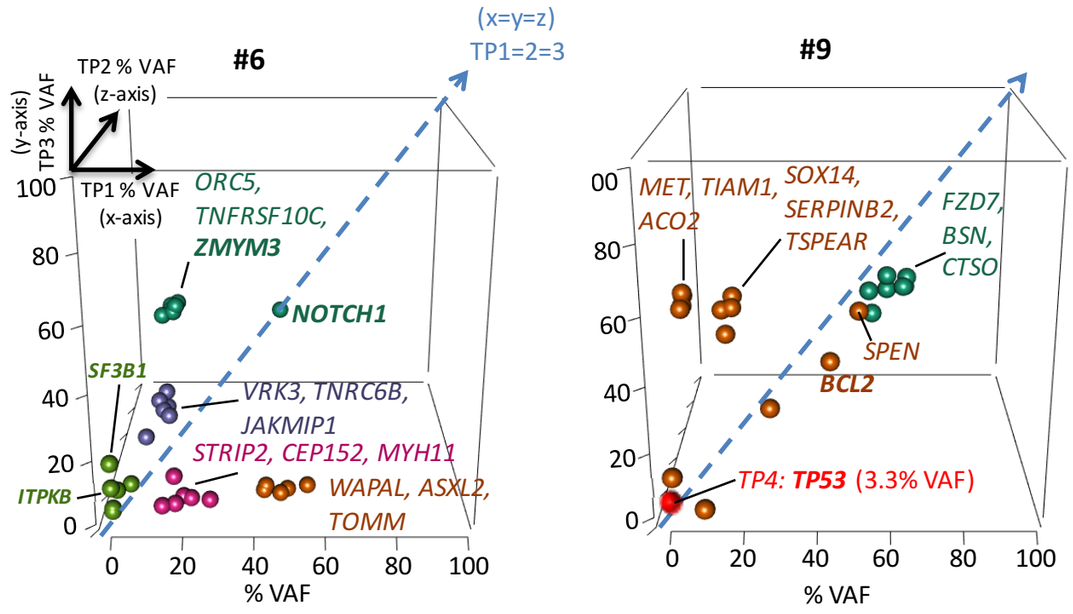


A

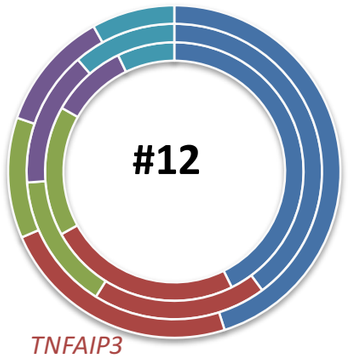
SciClone

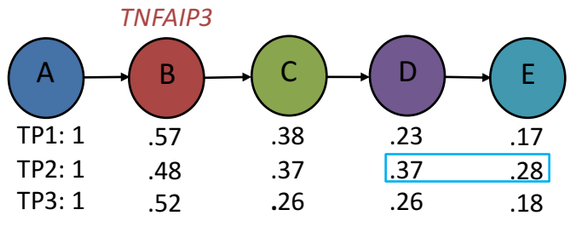


B



Linear Evolution

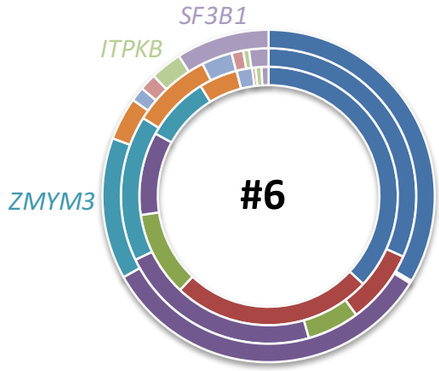


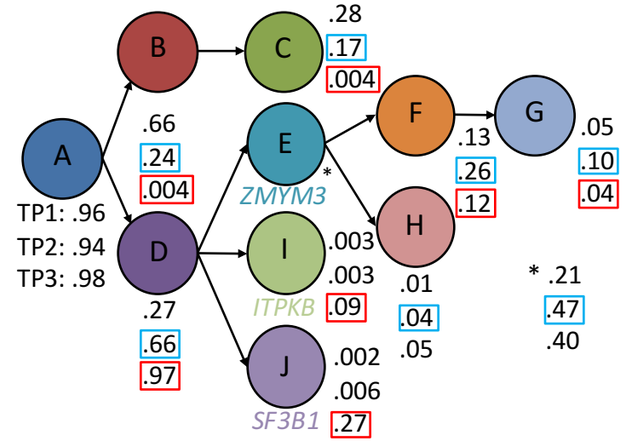


Mutations in each node

(A): *APC; ATL3; EYS; HRH4; OR8K3; PDE12* **(B):** *KIAA1217; TSPAN33; PLCG1; TNFAIP3* **(C):** *XPO6; GBX2; TEX13B* **(D):** *MARK4; CSE1L; DNPEP; WDR66* **(E):** *SLCO1C1*

Branching Evolution

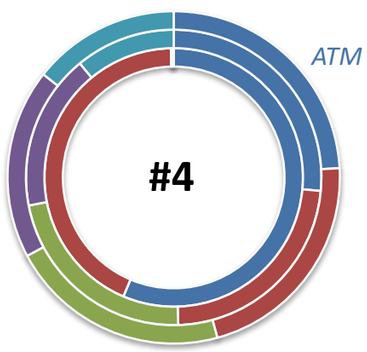


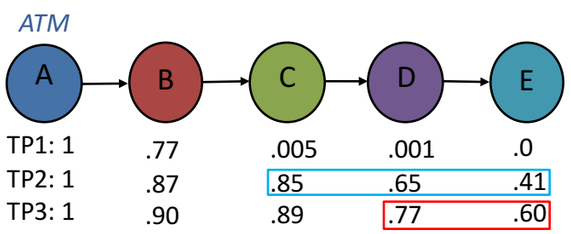


Mutations in each node

(A): *NOTCH1* **(B):** *ASXL2; CEP152; MYH11; PLXNA4; STRIP2; TOMM34; UNKL; WAPAL* **(C):** *SPO11; NCOA2* **(D):** *AKAP1; KIAA1549L; NAV2; TNFRSF10C; ORC5* **(E):** *CEACAM1; EFNB2; JAKMIP1; TDO2; TNRC6B; VRK3; ZMYM3* **(F):** *SPOCK3* **(G):** *KCTD8* **(H):** *RAD50; HTR4* **(I):** *ITPKB* **(J):** *SF3B1*

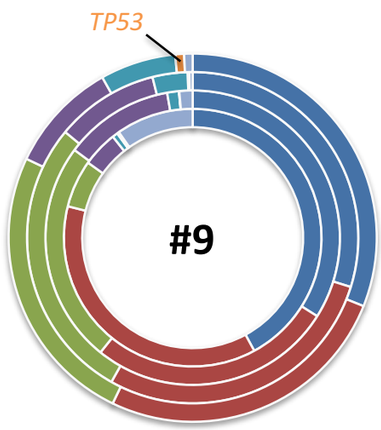
Phylosub

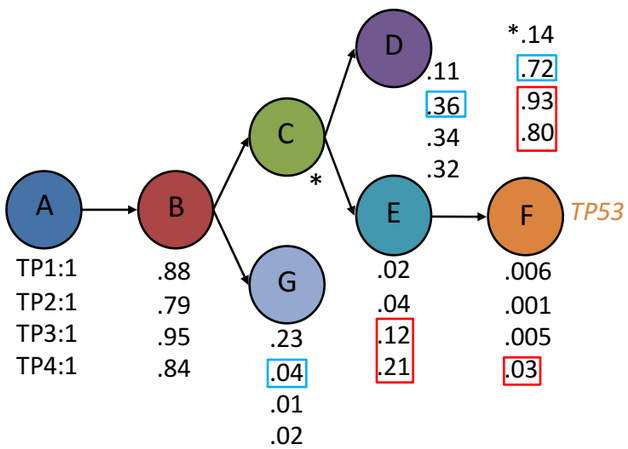




Mutations in each node

(A): *ATM (x2); CIRBP; PLCE1; TNRC18; WSCD2* **(B):** *MAP3K3; PKD1L1* **(C):** *PUF60; KIAA1984* **(D):** *SETD1B; SIM1; NSRP1* **(E):** *MCCC1*





Mutations in each node

(A): *BSN; CTSO; MAU2; OR4A5; SCEL; SH3RF1* **(B):** *BCL2; FZD7; SPEN* **(C):** *ALLC; ASB3,GPR75-ASB3; CEACAM19; DGCR2; MET; RWDD2A; SERPINB2; TIAM1; TSPEAR* **(D):** *ACBD7* **(E):** *ACO2* **(F):** *TP53* **(G):** *SOX14*