Genomic disruption of the histone methyltransferase *SETD2* in chronic lymphocytic leukemia

Helen Parker [1], Matthew JJ Rose-Zerilli [1]*, Marta Larrayoz [1]*, Ruth Clifford [2], Jennifer Edelmann [3], Stuart Blakemore [1], Jane Gibson [4], Jun Wang [5], Viktor Ljungström [6], Tracy Chaplin [5], Ali Roghanian [1], Zadie Davis [7], Anton Parker [7], Eugen Tausch [3], Stavroula Ntoufa [8], Sara Ramos [2], Pauline Robbe [2], Reem Alsolami [2], Andrew J Steele [1], Graham Packham [1], Ana E. Rodríguez-Vicente [9], Lee Brown [1], Feargal McNicholl [10], Francesco Forconi [1], Andrew Pettitt [11], Peter Hillmen [12], Martin Dyer [13], Mark S Cragg [1], Bryan D Young [5], Claude Chelala [5], Richard Rosenquist [6], Kostas Stamatopoulos [8], Stephan Stilgenbauer [3], Samantha Knight [2], Anna Schuh [2],David G Oscier [1,7] and Jonathan C Strefford [1]*

[1] Cancer Research UK Centre and Experimental Cancer Medicine Centre, Academic Unit of Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK,

[2] Oxford National Institute for Health Research Biomedical Research Centre/Molecular Diagnostic Centre, University of Oxford, Oxford, UK,

[3] Department of Internal Medicine III, Ulm University, Ulm, Germany,

[4] Centre for Biological Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, UK,

[5] Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary's, University of London, London, UK,

[6] Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden,

[7] Department of Molecular Pathology, Royal Bournemouth Hospital, Bournemouth, UK,

[8] Institute of Applied Biosciences, Center for Research and Technology Hellas, Thessaloniki, Greece.

[9] Department of Haematology, University Hospital of Salamanca-Biomedical Research Institute of Salamanca; IBMCC, Comprehensive Cancer Center Research, University of Salamanca-CSIC, Salamanca, Spain

[10] Altnagelvin Area Hospital, Western Health and Social Care Trust, Londonderry, UK,

[11] Department of Molecular and Clinical Cancer Medicine, Royal Liverpool and Broadgreen University Hospitals NHS Trust, Liverpool L7 8XP, UK,

[12] Department of Haematology, St James's University Hospital, Leeds, UK,

[13] College of Medicine, Biological Sciences and Psychology, University of Leicester, UK,

* These authors contributed equally to this manuscript

Correspondence to: Professor Jonathan C Strefford, Cancer Genomics Group, Cancer Sciences Division, Somers Cancer Research Building, Southampton General Hospital, Tremona Road, Southampton SO16 6YD. Tel: 44 23 8079 5246. E-mail: JCS@soton.ac.uk

Conflict of interests: The authors state that that there are no conflicts of interests.

Text word count: 3623

Abstract word count: 166

Number figures/tables: 4/2

Number references: 43

Number supplementary figures/tables: 5/6

Abstract

Histone methyltransferases (HMTs) are important epigenetic regulators of gene transcription and are disrupted at the genomic level in a spectrum of human tumors including hematological malignancies. Using high-resolution SNP-arrays, we identified recurrent deletions of the *SETD2* locus in 3% (8/261) of chronic lymphocytic leukemia (CLL) patients. Further validation in two independent cohorts showed that *SETD2* deletions were associated with loss of *TP53*, genomic complexity and chromothripsis. With next generation sequencing we detected mutations of *SETD2* in an additional 3.8% of patients (23/602). In most cases, *SETD2*-deletions or mutations were often observed as a clonal event and always as a mono-allelic lesion, leading to reduced mRNA expression in *SETD2*-disrupted cases. Patients with *SETD2* abnormalities and wild-type *TP53* and *ATM* from five clinical trials employing chemotherapy or chemo-immunotherapy, had reduced progression-free and overall survival compared to cases wild-type for all three genes. Consistent with its postulated role as a tumor suppressor, our data highlights *SETD2* aberration as a recurrent, early loss-of-function event in CLL pathobiology linked to aggressive disease.

Introduction

The transfer of methyl groups from S-adenosyl methionine to lysine or arginine residues on histone proteins, catalyzed by histone methyltransferases (HMTs), is an important regulator of gene transcription. Accordingly, HMTs are disrupted by various mechanisms including chromosomal translocations, genomic loss and/or point mutations in both solid and hematological malignancies.[1] Among the increasing number of HMT aberrations identified in human malignancies, recurrent loss and/or inactivating mutations of the tumor suppressor gene *SETD2*, were initially identified in clear cell renal cell carcinoma (ccRCC)[2] and subsequently in other solid tumors e.g. high grade gliomas.[3] Moreover, *SETD2* mutations have been reported in a subset of patients with acute lymphoblastic leukemia[4] and acute myeloid leukemia, especially those with rearrangements in the another HMT gene, *MLL*.[5] SETD2 is the only enzyme that catalyzes the trimethylation of lysine 36 on histone 3 (H3K36me3), one of the major chromatin marks associated with active transcription. Recent studies have linked SETD2 to the maintenance of genomic integrity, through coordination of homologous recombination repair after double strand breaks. The loss of SETD2 impairs DNA repair and enhances genomic instability, supporting its tumor suppressor role. [6-9]

Chronic lymphocytic leukemia (CLL) is characterized by remarkable clinical heterogeneity such that some patients pursue an indolent course while others require early treatment. Considerable effort has focused on understanding the genetic diversity that underpins this clinical heterogeneity. High-resolution genomic arrays and next-generation sequencing have identified recurring novel regions of genomic copy-number aberrations (CNAs) like del(13q), del(11q), trisomy 12 and del(17p) and recurrent driver mutations in genes such as *TP53*, *ATM*, *SF3B1* and *NOTCH1*, respectively (reviewed in [10]). Mutations frequently involve genes encoding proteins with important roles in cell signalling, cell cycle control, DNA repair and RNA-splicing and processing; however the reported incidence of mutations in chromatin modifiers is lower than in many other haematological malignancies.

In this study, we report the identification of recurrent deletions and mutations of the *SETD2* gene in large, well-characterized CLL cohorts. *SETD2* lesions appear to represent early events in CLL pathogenesis, often co-existing with, but preceding *TP53* abnormalities. They are associated with genomic complexity and chromothripsis, and identify a subgroup of patients with poor outcome.

Methods

Patients

We studied samples taken from 1006 CLL patients either at entry into one of five clinical trials or from a cohort of untreated patients with progressive disease managed at the Royal Bournemouth Hospital. Four randomized trials (ADMIRE, ARCTIC, UK CLL4 [11], GCSG CLL8 [12]) compared chemo or chemo-immunotherapy regimens in fit previously untreated patients while the fifth trial (SCSG CLL2O) enrolled ultra-high risk patients who were either refractory to a purine analogue or were previously untreated with a 17p deletion. Further details of the clinical trials are provided in **Table S1**. All patients were diagnosed using standard morphologic and immunophenotypic criteria. Informed consent was obtained from all patients in accordance with the Helsinki declaration, and this study was approved by national or regional research ethics committees.

Patients were grouped into three cohorts (discovery [n=261], extension [n=635] and ultra high-risk [n=110]); details of the cohort composition and *SETD2* analysis are summarised in **Table 1, Supplementary methods** and **Supplementary figure 1**. DNA was extracted from CLL B cell samples (all with >80% tumour purity) and from matched germ-line DNA for *SETD2*-mutated cases as outlined in **Supplementary methods**. The assessment of established biomarkers was performed as previously described.[13] 572 and 602 samples were screened for *SETD2* loss and mutation, respectively, with 168 cases screened for both loss and mutation.

Genome-wide microarray-based copy number analysis

DNA from 261 discovery and 110 ultra-high risk cases was amplified, labelled and hybridized to the Affymetrix SNP6.0 platform, aligned onto the human genome sequence (GRCh37) and analysed in Partek Genomics Suite (Partek Inc, Missouri, USA) as reported previously.[14-18] DNA from 201 pre-treatment extension cases (ADMIRE and ARCTIC) was hybridized to the Illumina HumanOmni1-Quad and HumanOmniS-8 platforms according to manufacturer's protocols. [19,20] Further experimental details are provides in the **Supplementary methods**.

Targeted re-sequencing and whole exome sequencing

93 CLL samples from the discovery cohort (and five matched germ-line controls) were processed and analysed for mutations in *SETD2* (all exons) and a number of clinically relevant genes with a bespoke Haloplex Target Enrichment system (Agilent Technologies) (**Supplementary Methods Table 1**) and processed and analysed as previously reported.[21] An additional 231 cases from our pre-treatment extension cohort were screened for *SETD2* mutations using a TruSeq Custom Amplicon panel (Illumina

Inc. San Diego, CA, USA) as previously described.[20,22] All the variants identified by both platforms were annotated against dbSNP (build 135) and functional prediction was also performed using SIFT and Polyphen2 analysis. Somatically-acquired *SETD2* mutations (n=4) were also identified in the recent whole exome sequencing (WES) study of 278 matched tumor and germ-line cases from the GCSG CLL8 study [23]. Additional experimental details are provided in the **Supplementary methods**.

Sanger validation

Variants in *SETD2* were subjected to validation by conventional Sanger-based sequencing of PCR products obtained from tumor [n=11] and where possible, paired normal genomic DNA [n=5]. The expression of *SETD2* mutations at mRNA level was also tested in samples with available material [n=4]. Primers for DNA or mRNA validation are listed in **Table S2**.

Quantitative RT-PCR

Total RNA was isolated from purified CLL cells of 36 patient samples using RNeasy columns (Qiagen) and reversed transcribed using the Improm™II RT-PCR kit (Promega, UK) according to the manufacturer's instruction. Primers and probes for the housekeeping genes (18s) and target genes (*CCDC12, NBEAL2, KIF9, KLHL18, SETD2*) were selected using the Universal Probe Library (Roche Applied Science, UK) (**Table S3**). Two independent assays were designed to ascertain expression of 3' and 5' *SETD2*. Normal B-cell mRNA was use to normalize the expression of each gene by delta-delta CT method as previously described.[24]

Statistical analysis

Statistical analysis was performed with SPSS v22. Differences between samples were analysed by U-Mann Whitney test. Progression-free survival (PFS) and overall survival (OS) were calculated for clinical trial samples from randomization. Survival analysis was performed by Kaplan-Meier and log-rank analysis. Significant differences were considered with P-values lower than 0.05.

Results

Recurrent deletions of 3p are a feature of CLL

We identified 1024 acquired CNAs (mean 3.9, range 0–45) in our discovery cohort (**Table S4**). Deletions of chromosome 3p [del(3p)] were observed in 8 patients (3%), ranged from 0.45-81 Mb in size (**Table S5**), and identified a well delineated MDR between genomic location 46.96-47.39 Mb, containing the genes *CCDC12, NBEAL2, SETD2, KIF9* and *KLHL18* (**Figure 1A**). We compared the expression of these genes by qRT-PCR in 3p deleted [n=6] versus non-3p deleted patients [n=8] (**Figure 1C**). We were not able to detect the expression of *KIF9* mRNA in CLL or normal B-cells. Within the MDR, the HMT gene *SETD2* was significantly under-expressed, measuring by two different assays targeting the 3' or 5' region of the mRNA ($p<0.0001$ for both assays).

We then aimed to confirm the presence of 3p deletions and refine the MDR in our extension cohorts. Firstly, we identified nine del(3p) cases (4.5%) in our extension pre-treatment cohort, permitting the MDR to be refined to the *SETD2* and *KIF9* loci (47.12-47.36 Mb, **Figure 1A** and **Table S5**). Across our discovery and pre-treatment extension cohorts, *SETD2* deletions were present in 17/461 cases (3.7%), significantly associated with deletions and/or mutations of *TP53* (p=0.003) and genomic complexity (≥3 deletions[25], p=0.04)(**Figure 1B**). GISTIC 2.0 analysis [26], an algorithm for identifying statistically significant regions of CNA above an estimated background rate (FDR q-value <0.25), showed that in 39 *TP53* deleted cases (del(3p), n=15), the *SETD2* region on 3p21.31, was deleted at a significant frequency (q-value=0.001), ranked third after del(13q) and del(17p) (**Figure S2**).

Interestingly, *SETD2* deletions without concomitant *TP53/ATM* abnormalities [n=6] also exhibited significantly more genomic complexity than wildtype patients (p=0.01. **Figure 1D**). Two *SETD2*-deleted cases showed evidence of chromosome 3 chromothripsis (based on >10 CNAs per chromosome[18]) (**Figure 1A-B**). In the ultra-high risk cohort, *SETD2*-deletions were detected in 9% of cases [10/110], and were significantly enriched compared to the pre-treatment cohort (p=0.009). All ten had loss of *TP53* and five had concomitant chromosome 3 chromothripsis (**Figure 1B**). To further establish the significance of our *SETD2* deletion in cases with chromosome 3 chromothripsis, we mapped all recurrently 3p deletions in these cases. This analysis showed that whilst additional regions of recurrent deletion were observed on 3p, the only regions shared across all patients included the *SETD2* locus (**Table S6**).

We analysed *SETD2* expression in an extended cohort of patients with 3p deletions [n=16], and again the expression was diminished in these patients compared to wild-type patients (P=0.0068; **Figure S3**). In order to study the clonal nature of the *SETD2* deletions, we assigned each genomic CNA with a

relative copy-number value by normalizing CNA intensity values from array features. We excluded regions with gain and sex chromosome CNAs from the analysis. The cut off for normal copy number was established between 1.7 and 2.3. We could infer that the 3p deletion was in the dominant clonal population in 11/18 (61%) cases with data available for analysis (**Figure S4**).

SETD2 mutations in CLL

To identify somatic gene mutations, we initially employed targeted re-sequencing of 93 discovery cohort cases and identified 122 non-silent mutations (non-synonymous n=80, frameshift indel n=20, splicing n=9, nonframeshift indel n=6, stopgain n=6, stoploss n=1) targeting 37 genes in 71/93 cases (mean 1.8, range 1-4). Sanger sequencing confirmed 93.6% of the tested variants [n=79], whilst the remaining unconfirmed variants were present at low read depth (n=3) or in a low percentage of mutant reads (n=2). We found *ATM* [n=14], *TP53* [n=14], *NOTCH1* [n=20] and *SF3B1* [n=15] mutations at a frequency expected for the studied cohort, which aligns with published data and demonstrates the validity of the re-sequencing platform. We identified non-synonymous *SETD2* mutations in four (4.3%) discovery cases (p.D99G, p.Q1545K, p.W1306*, p.E1955Q) (**Figure 2A-B**). Sanger sequencing validated that all of the *SETD2* mutations were present in tumor DNA. We assessed the *SETD2* background mutation rate, expression level and replication timing data from Lawrence et al [27] and found that the gene shares no properties associated with false-positive candidate cancer genes. To determine whether our *SETD2* mutation rate was indicative of over-representation in our CLL tumours we calculated the nonsynonymous - synonymous ratio (Ka/Ks ratio) for all genes on our discovery cohort targeted sequencing gene panel. Genes with a Ka/Ks ratio greater than 1 are under positive selection, meaning that at least some of the mutations are advantageous [28]. When ranked by Ka/Ks ratio the *SETD2* mutation rate is ranked sixth after *ATM*, *TP53*, *SF3B1, NOTCH1* and *FAT4* (Table S7). These ratios are frequency dependant (*SETD2* mutations are rare events), but it is important to state that no mutated positions were targets of synonymous mutation. We obtained matched germ-line DNA from three patients and confirmed that the mutations were somatically acquired (p.D99G, p.W1306*, p.E1955Q) (**Figure S5A**).

To corroborate this preliminary observation, we investigated 231 cases of our pre-treatment extension cohort by TruSeq amplicon-based sequencing. We identified an additional nine (3.9%) *SETD2* mutations (p.A50T, p.L89F, p.P167L, p.N535S, p.E670K, p.M1742L, p.M1889T (x2), p.I2295M) (**Figure 2A-B**). Sanger sequencing confirmed each *SETD2* variant in the tumor material and in two cases with germ-line material available, the variants were somatically acquired. Assessment of WES data of the CLL8 study[23] samples included in our pre-treatment extension cohort, revealed the presence of somatically-acquired *SETD2* mutations in 4/278 cases (1.4%), namely, p.EEEELQSQQ1919fs, p.L1804fs,

p.VLEYC1576del, p.V1190M. None of these *SETD2* mutations (**Table 2**) are annotated in COSMIC[29]. During the preparation of this manuscript, a study performed by Puente et al. in 506 CLL patients also described both *SETD2* mutations (0.8% of cases) and deletions in 3p (2% of cases) whose MDR encompassed *SETD2* [30] whilst Landau et al identified *SETD2* mutations in 8/538 (1.5%) cases [23].

In total, across our cohorts there were 15 somatically-acquired *SETD2* variants (15/602; 2.5%). An additional eight variants that could not be examined in germ-line material were either absent (n=3), reported to have a very low prevalence (n=5) in 1000 Genomes project or have a sub-clonal variant allele frequency (%VAF <0.45, (n=1); **Table 2**). Therefore, whilst these eight variants are predicted to be functionally deleterious, we cannot exclude that the minority may be rare germ-line variants as they exhibit clonal variant allele frequencies in the tumor material.

We were able to confirm the expression of the *SETD2* mutations at mRNA level in four of our patients with available material (p.D99G, p.Q1545K, p.E1955Q, p.E670K) (**Figure S5A**), and qRT-PCR analysis of three *SETD2* mutated samples showed that *SETD2* mRNA expression was reduced compared to wild-type patients (P=0.035; **Figure S3**).

We performed integrative analysis of 93 cases from our discovery cohort with Haloplex re-sequencing and SNP6.0 copy number data available, by employing the ABSOLUTE algorithm.[31] This approach estimates the cancer cell fraction (CCF) harboring a given mutation by correcting for sample purity and local copy number changes. Mutations were classified as clonal if the CCF was >0.95 with a probability >0.5, and sub-clonal otherwise.[32] In additional cases with proven-somatic *SETD2* mutations (n=4) and paired copy number data from our pre-treatment validation cohorts, we performed this estimation by manually correcting for tumor sample purity and local copy-number. Our analysis demonstrated the expected sub-clonal distribution of established gene mutations, such as *TP53*, *ATM*, *SF3B1* and *NOTCH1*. Interestingly, all our somatically acquired *SETD2* mutations exhibited a clonal CCF, suggesting that these mutations may be early events in the evolution of CLL (**Figure 2C-D & Figure S5B-C**), although further studies are required to confirm this observation.

SETD2 aberrations are associated with inferior progression free- and overall survival

Finally, we analysed the impact of *SETD2* abnormalities (deletion or somatically-acquired mutation) on progression-free (PFS) and overall survival (OS) in front-line trial patients. We observed a significantly shorter PFS in cases with *SETD2* abnormalities that were wild-type for *TP53/ATM* [n=7], compared to cases wild-type for *TP53/ATM/SETD2* [n=62] (PFS: 30 vs. 48 months; p=0.003) (**Figure 2E**). The same patients with *SETD2* abnormalities [n=7] also had a shorter OS than wild-type patients [n=62] (OS: 34 vs. 92 months; p<0.001) (**Figure 2F**). Whilst these data suggest that *SETD2* aberration

may be clinically relevant, further investigation in larger cohorts is needed to understand their full impact on survival.

Discussion

This study was based on an initial high-resolution SNP6.0 array analysis of 261 untreated patients with progressive CLL which identified a recurrent deletion of the short arm of chromosome 3 in 3% of cases (n=8). The MDR included the *CCDC12, NBEAL2, SETD2, KIF9* and *KLHL18* genes, of which *SETD2* was the most significantly under-expressed in tumor cells. We then identified clonal, somatically-acquired *SETD2* mutations in 4.3% of this cohort; no mutated case had a concomitant *SETD2* deletion.

The *SETD2* gene encodes a 230 kDa protein that is non-redundantly responsible for all trimethylation of lysine 36 on histone H3 (H3K36me3),[33,34] a mark that is associated with actively transcribed regions and is involved in transcriptional elongation and splicing. [35] In addition, recent studies have linked this epigenetic histone mark to other important cellular processes such as the regulation of mismatch repair, efficient homologous recombination and the maintenance of genomic stability.[7-9] *In vitro* inhibition of Setd2 decreases global levels of H3K36me3 and impairs the recruitment of the mismatch recognition protein hMutSα onto chromatin, thereby preventing appropriate DNA mismatch repair. Cells lacking the Setd2 protein display microsatellite instability and have elevated levels of spontaneous mutations.[7,36-38] Inactivating *SETD2* mutations were first described in ccRCC[2,6], subsequently in other solid tumors such as paediatric high-grade gliomas and most recently in a subset of patients with acute lymphoid and myeloid leukemias. [2,5,39,40] *SETD2* mutations in ccRCC are frequently associated with 3p deletions resulting in loss of both *SETD2* and VHL genes, while in acute leukemias, *SETD2* mutations may be bi-allelic but 3p loss is rare. *SETD2* genomic abnormalities are associated with decreased H3K36me3 levels, a distinctive DNA methylation signature[6] and chemoresistance in paediatric acute lymphoblastic leukemia [41]. In MLL-rearranged cells from acute leukemic patients, Setd2 knockdown is implicated in disease initiation and progression by promoting the self-renewal capacity of leukemic stem cells.

In view of the role of *SETD2* disruption in tumorigenesis and the identification of *SETD2* abnormalities in our discovery cohort, we then accrued samples from other patient cohorts, including the GCLLSG CLL8 cohort in which 3p deletions had also been detected [18], to confirm the incidence of *SETD2* disruption and evaluate its biological and clinical consequences in CLL. Previously untreated patients sampled at randomization to chemo or chemo-immunotherapy trials had a similar incidence of 3p deletions (4.5%) to that seen in the discovery cohort while a higher incidence of loss (9%) was found in the ultra-high risk cohort. The inclusion of additional cohorts enabled a smaller MDR to be defined, including *SETD2* and *KIF9*, implicating *SETD2* as the key deleted gene. The incidence of *SETD2* mutations was comparable in all cohorts tested. The recent studies by Puente et al and Landau et al, published during the preparation of this manuscript confirms the rare but recurrent nature of *SETD2*

abnormalities [23,30]. *SETD2* deletions were not over-represented by analysis of WES generated copy number data in the work by Landau and the mutation frequencies of both studies were lower than those in our study. The different frequencies reported in these two studies could be explained by cohort composition, as our study included ultra-high risk CLL and patients randomized to clinical trials.

As we found *SETD2* mRNA expression to be down regulated in cases with either *SETD2* deletion or mutations and as we did not observe bi-allelic *SETD2* abnormalities, we assessed whether *SETD2* may also be deregulated by DNA methylation. We reviewed the methylation status of the *SETD2* gene body and promoter regions (15 and 9 CpG probes, respectively) in the study of Kulis and co-workers[42] and found no differential methylation levels between unmutated CLL or mutated CLL or major cytogenetic sub-types (including del(17p) cases). Furthermore, *SETD2* mRNA expression was not found to be correlated with methylation status. Together, this suggests that DNA methylation does not play a substantial role in regulating *SETD2* expression in B-CLL cells, as previously noted in acute leukemia. In addition, when we analysed *SETD2* expression in CLL studies deposited in the Oncomine database,[43] we observed a heterogeneous pattern. Significantly reduced levels are evident in the minority of patients,[44] which given our observed association between *SETD2* deletion and expression could imply gene deletion in those Oncomine samples with low mRNA expression.

Across all cohorts, *SETD2* deletion was found in both *IGHV* mutated and unmutated cases but was strongly associated with *TP53* loss and mutation, likely accounting for its higher incidence in the ultra -high risk cohort. We also noted an association with genomic complexity even in cases lacking a *TP53* or *ATM* abnormality, consistent with the role of *SETD2* in maintaining genomic stability. Moreover, we identify several *SETD2* deletions that appeared to be the result of chromothripsis. The somatically-acquired *SETD2* mutations showed a comparable genomic distribution to those previously described in other tumors and were predicted to have deleterious functional consequences. Furthermore, their association with significantly reduced mRNA expression in those cases analysed, suggest that they either directly affect mRNA expression, or co-exist with other defects in transcriptional control at this locus. Interestingly, we did not observe a statistically significant association between *SETD2* mutations and *TP53* abnormalities or genomic complexity, the implication of which may be differing functional consequences of mono-allelic loss and mutation.

In our study, both *SETD2* deletions and mutations often appeared to be clonal and may precede *TP53* abnormalities in at least some cases. Setd2 has been shown to directly regulate the transcription of a subset of genes via cooperation with the transcription factor p53,[45] and the link between *SETD2* and *TP53* is an interesting association worthy of functional validation. It is possible that the *SETD2* alterations present in our CLL cases may contribute to further inactivation of p53-mediated checkpoint

control, a situation that has been proposed in ccRCC.[8] The low frequency of *SETD2* disruption and the association with *TP53* abnormalities hinder an accurate assessment of its clinical consequences. Nevertheless, we observed a shorter PFS and OS in patients with *SETD2* but no *TP53* or *ATM* abnormalities compared to cases wild type for all three genes. In support of this preliminary clinical observation, it has been shown that 3p deletions in head and neck squamous carcinoma (HNSCC) are associated with reduced survival. [46] Furthermore, the authors showed that the co-existence of a *TP53* abnormality with del(3p) decreased survival further, an observation that we could not confirm in our cohort.

In summary, our current study provides the first comprehensive analysis of CNAs and mutations targeting the *SETD2* gene in a large cohort of patients with CLL. We find somatic deletions and mutations in ~7% of CLL patients requiring treatment. These associate with *TP53* dysfunction, genomic complexity and chromothripsis and may be early clonal events. Functional studies are now warranted to elucidate the exact biological importance of *SETD2* in CLL pathogenesis, but our data adds to a growing body of evidence suggesting a role for H3K36me3 in tumorigenesis that may be exploited for the development of novel therapeutic approaches.

Acknowledgements

The authors gratefully acknowledge all patients who contributed to this study. This work was funded by Bloodwise (11052, 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund. SS is supported by the Else Kröner-Fresenius-Stiftung (2012_A146), and Deutsche Forschungsgemeinschaft (SFB 1074 projects B1, B2). The LRF CLL4 trial was funded by a core grant from Leukaemia and Lymphoma Research. D.G. and D.C. acknowledge the support by The Royal Marsden Hospital and The Institute of Cancer Research National Institute of Health Research Biomedical Research Center. RR is supported by the Swedish Cancer Society, the Swedish Research Council, Science for Life Laboratory, Uppsala University, Uppsala University Hospital, and the Lion's Cancer Research Foundation, Uppsala.

Author Contributions

H.P., M.J.J.R.-Z., M.L., R.C., J.E., S.B., T.C., A.R., A.P., S.N., M.C., R.A., S.R., S.K., A.R., B.Y. and L.B. performed the experimental work; A.P. performed the molecular diagnostic assays; M.J.J.R.-Z., J.G., J.W, P.R., V.L.. and S.K. conducted the statistical and bioinformatic analyses; F.N., F.F., A.P., P.H., M.D., S.N., K.S., S.S., R.R., A.S. and D.O. contributed patient samples and data; J.C.S. initiated and designed

the study; H.P. , M.J.J.R.-Z., D.G.O, M.L and J.C.S. wrote the paper with contributions from R.C., M.L., G.P., A.J.S., R.R., K.S., A.S. .; and all authors critically reviewed the final paper.

Supplementary information is available at Leukemia's website

Conflict of interests: The authors state that that there are no conflicts of interests.

References

1. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* 2011; **11**:726-734.
2. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* 2010; **21**:360-3.
3. Fontebasso AM, Schwartzentruber J, Khuong-Quang DA, Liu XY, Sturm D, Korshunov A, et al. Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta Neuropathol* 2013; **125**:659-69.
4. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 2012; **481**:157-63.
5. Zhu X, He F, Zeng H, Ling S, Chen A, Wang Y, et al. Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat Genet* 2014; **46**:287-93.
6. Duns G, van den Berg E, van Duivenbode I, Osinga J, Hollema H, Hofstra RM, et al. Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer Res* 2010; **70**:4287-91.
7. Pfister SX, Ahrabi S, Zalmas LP, Sarkar S, Aymard F, Bachrati CZ, et al. SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell Rep* 2014; **7**:2006-18.
8. Carvalho S, Vítor AC, Sridhara SC, Martins FB, Raposo AC, Desterro JM, et al. SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *Elife* 2014:e02482.
9. Kanu N, Grönroos E, Martinez P, Burrell RA, YiGoh X, Bartkova J, et al. SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. *Oncogene* 2015:[Epub ahead of print].
10. Guièze R, Wu CJ. Genomic and epigenomic heterogeneity in chronic lymphocytic leukemia. *Blood* 2015:[Epub ahead of print].
11. 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; **370**:230-9.
12. Hallek M, Fischer K, Fingerle-Rowson G, Fink AM, Busch R, Mayer J, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* 2010; **376**:1164-74.
13. Oscier D, Wade R, Davis Z, Morilla A, Best G, Richards S, et al. Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica* 2010; **95**:1705-12.
14. 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**:e83244.
15. 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**:489-97.
16. Rose-Zerilli M, Forster J, Parker H, Parker A, Rodriguez A, Chaplin T, et al. ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia, data from the UK LRF CLL4 trial. *Haematologica* 2014; **99**:736-42.
17. Edelmann J, Saub J, Ibach S, Holzmann K, Tausch E, Bloehdorn J, et al. High Resolution Genomic Profiling of Primary "Ultra High Risk" and Refractory Chronic Lymphocytic Leukemia: Results from the CLL2O Trial *Blood (ASH Annual Meeting Abstracts)* 2014:3288.

18. Edelmann J, Holzmann K, Miller F, Winkler D, Bühler A, Zenz T, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* 2012; **120**:4783-94.

19. 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; **26**:1564-75.

20. 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**:1021-31.

21. 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; **21**:4174-83.

22. Robbe P, Clifford R, Timbs A, Burns A, Titsias M, Cabes M, et al. Comprehensive genome-wide analysis of CLL samples from UK 1st line and relapsed/refractory clinical trials. . *EHA Annual Meeting Abstracts* 2013:EHA18ABSSUB-4559.

23. 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:[Epub ahead of print].

24. An Q, Wright SL, Konn ZJ, Matheson E, Minto L, Moorman AV, et al. Variable breakpoints target PAX5 in patients with dicentric chromosomes: A model for the basis of unbalanced translocations in cancer. *Proc Natl Acad Sci USA* 2008; **105**:17050-17054.

25. 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; **16**:835-47.

26. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12**:R41.

27. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014; **505**:495-501.

28. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 2000; **15**:496-503.

29. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research* 2011; **39**:D945-50.

30. 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:[Epub ahead of print].

31. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotech* 2012; **30**:413-21.

32. 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**:714-26.

33. Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J* 2008; **27**:406-20.

34. Yoh SM, Lucas JS, Jones KA. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYP/Setd2-mediated histone H3K36 methylation. *Gene Dev* 2008; **22**:3422-34.

35. Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* 2012; **13**:115-26.

36. Li F, Mao G, Tong D, Huang J, Gu L, Yang W, et al. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* 2013; **153**:590-600.

37. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* 2009; **41**:376-81.

38. Sun XJ, Wei J, Wu XY, Hu M, Wang L, Wang HH, et al. Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem* 2005; **280**:35261-71.

39. Huether R, Dong L, Chen X, Wu G, Parker M, Wei L, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun* 2014; **3**:3630.

40. Mar BG, Bullinger LB, McLean KM, Grauman PV, Harris MH, Stevenson K, et al. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. *Nat Commun* 2014; **5**:3469.

41. Lee WP, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One* 2014; **9**:e90581.

42. 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; **44**:1236-42.

43. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 2007; **9**:166-80.

44. Haslinger C, Schweifer N, Stilgenbauer S, Döhner H, Lichter P, Kraut N, et al. Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J Clin Oncol* 2004; **22**:3937-49.

45. Xie P, Tian C, An L, Nie J, Lu K, Xing G, et al. Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. *Cell Signal* 2008; **20**:1671-8.

46. Gross AM, Orosco RK, Shen JP, Egloff AM, Carter H, Hofree M, et al. Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss. *Nat Genet* 2014; **46**:939-43.

Figure Legends

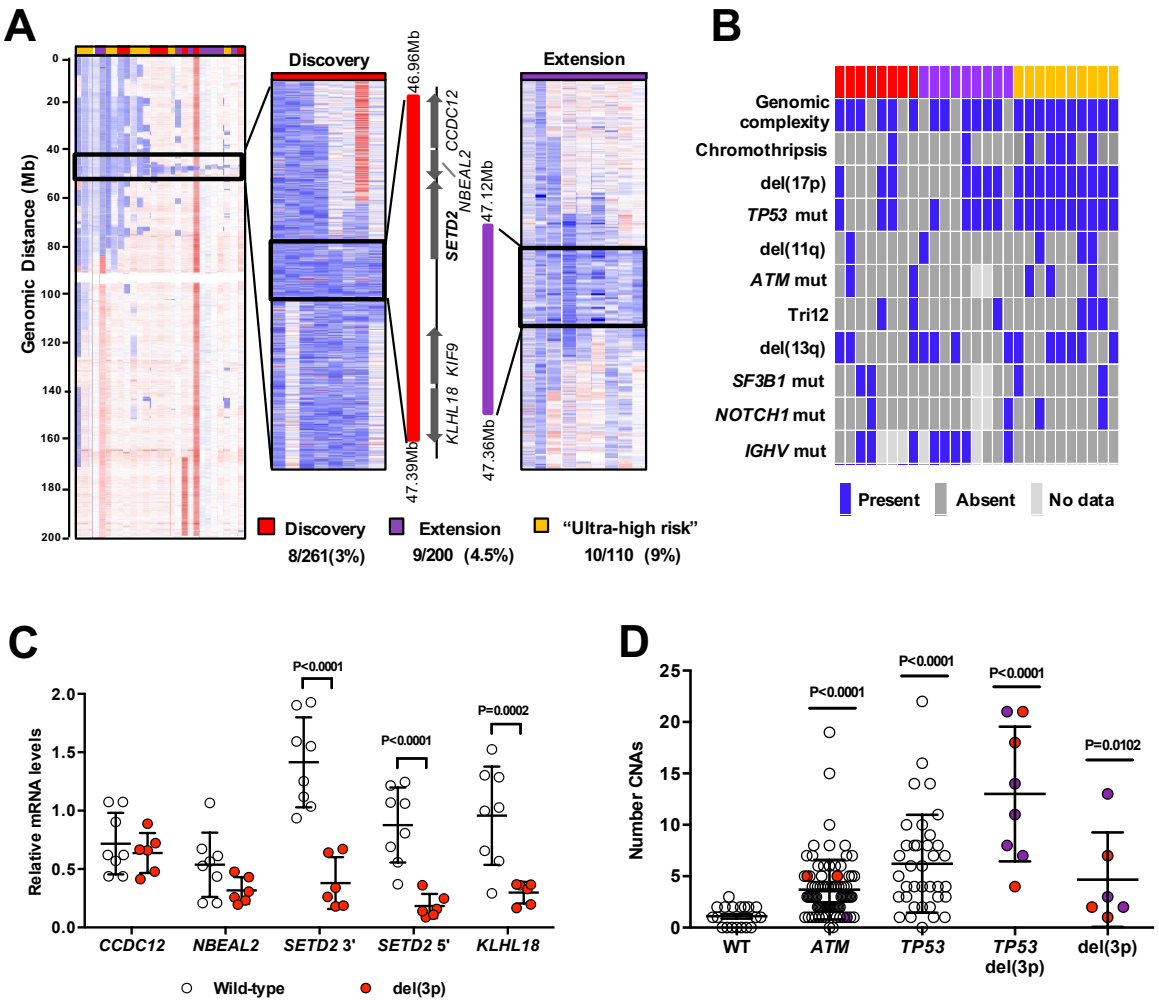
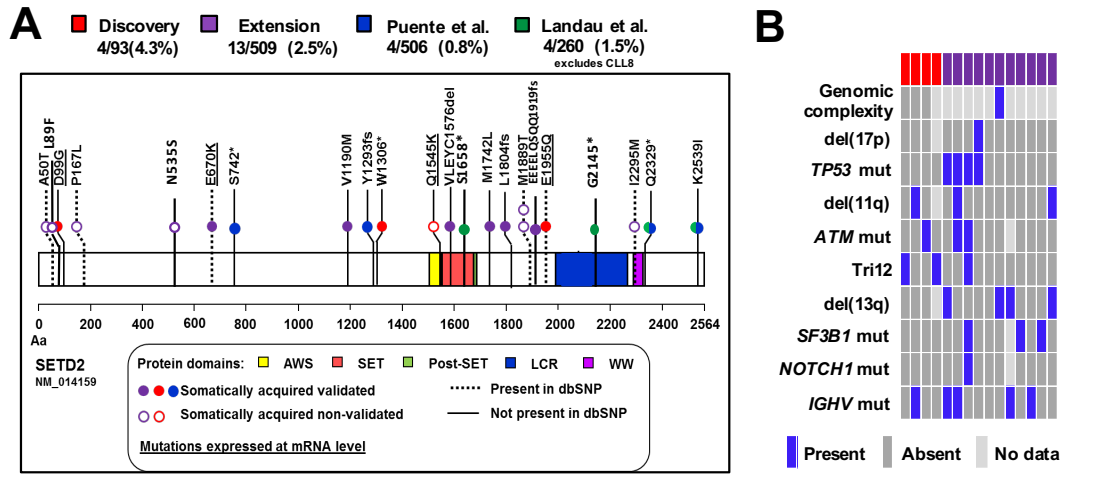
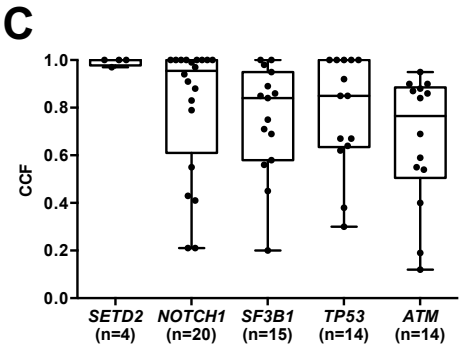


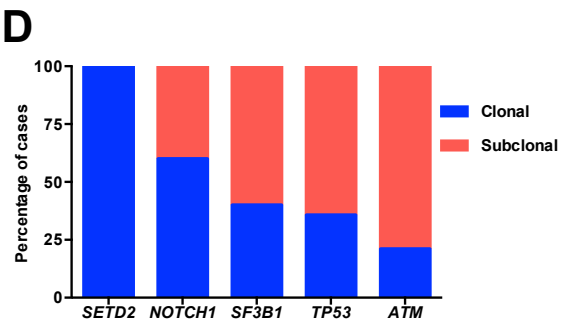
Figure 1. *SETD2* deletions in our discovery, extension and ultra-high risk cohorts

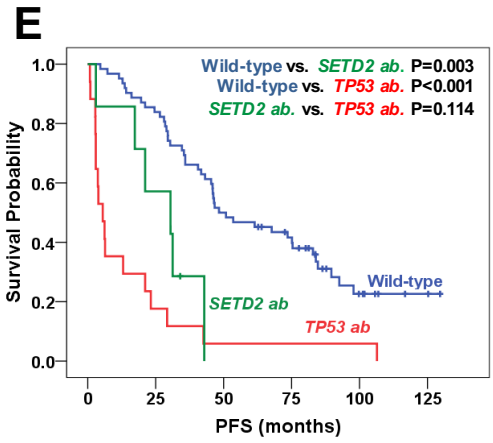
Figure 1A. SNP6.0 data for the del(3p) cases. Genomic location is indicated by the ladder to the left. Each column represents one patient. Loss, gain and normal copy-number are shown as blue, red and white, respectively. The black box indicates the MDR, and is displayed in greater detail for our discovery and extension cohorts. The genes in the MDR with their transcriptional direction are displayed in the middle, with the MDR from the discovery and extension cohorts shown by the red and purple bars, respectively. **1B.** Matrix displaying the biomarkers and genomic features associated with del(3p) cases with the discovery, extension and ultra-high risk cases shown in red, purple and yellow, respectively. **1C.** Real-time PCR expression for the five genes localized in the discovery MDR in cases with or without del(3p). All the samples were negative for KIF9. 18s was employed as housekeeping gene. Expression in normal B-cells was used as a normalization sample. Mean ± SD is represented. **1D.** Scatterplots displaying the number of CNA observed in subgroups of our cohort (excluding ultra-high risk cases). Cases were assigned to a subgroup using a hierarchical model; presence of del(17p) and/or *TP53* mutation, then del(11q) and/or *ATM* mutation, then del(3p) cases with and without *TP53* abnormalities and then wild-type (WT) cases containing no del(17p), del(11q), del(3p) or mutations in *ATM* and *TP53*. Mean ± SD is represented.


A
Discovery
4/93(4.3%)
Extension
13/509 (2.5%)
Puente et al.
4/506 (0.8%)
Landau et al.
4/260 (1.5%)
excludes CLL8
A50T
L89F
D99G
P167L
N535S
E670K
S742*
V1190M
Y1293fs
W1306*
Q1545K
VLE1C1576del
S1658*
M1742L
L1804fs
M1889T
EEEL0SQQ1919fs
E1955Q
G2145*
I2295M
Q2329*
K2539I
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2564
Aa
SETD2
NM_014159
Protein domains:
AWS
SET
Post-SET
LCR
WW
Somatically acquired validated
Somatically acquired non-validated
Present in dbSNP
Not present in dbSNP
Mutations expressed at mRNA level

B
Genomic complexity
del(17p)
TP53 mut
del(11q)
ATM mut
Tri12
del(13q)
SF3B1 mut
NOTCH1 mut
IGHV mut
Present
Absent
No data


C
CCF
1.0
0.8
0.6
0.4
0.2
0.0
SETD2 (n=4)
NOTCH1 (n=20)
SF3B1 (n=15)
TP53 (n=14)
ATM (n=14)


D
Percentage of cases
100
75
50
25
0
SETD2
NOTCH1
SF3B1
TP53
ATM
Clonal
Subclonal


E
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
Wild-type vs. SETD2 ab. P=0.003
Wild-type vs. TP53 ab. P<0.001
SETD2 ab. vs. TP53 ab. P=0.114
Wild-type
SETD2 ab
TP53 ab
0
25
50
75
100
125
PFS (months)

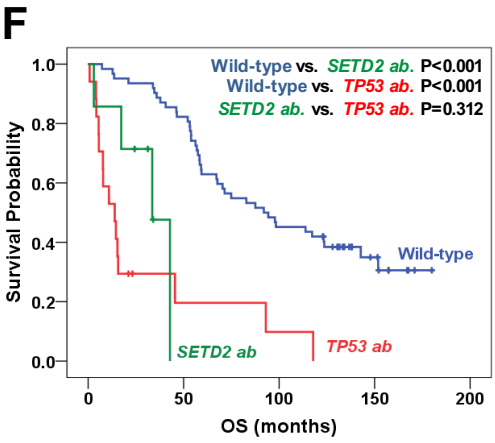

F
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
Wild-type vs. SETD2 ab. P<0.001
Wild-type vs. TP53 ab. P<0.001
SETD2 ab. vs. TP53 ab. P=0.312
Wild-type
SETD2 ab
TP53 ab
0
50
100
150
200
OS (months)

Figure 2. *SETD2* mutations in our discovery and extension cohorts

Figure 2A. Schematic diagram of the Setd2 protein with their key functional domains. Mutations are displayed on the diagram. The colour denotes the cohort, and the filled circles are mutations that have been confirmed as somatically acquired. **2B.** Matrix displaying the biomarkers and genomic features associated with *SETD2* mutated cases in the discovery (red) and extension (purple) cases. **2C.** Analysis of the clonality for *SETD2* and other recurrently mutated genes on CLL. For each case the cancer cell fraction (CCF) is derived manually or with the ABSOLUTE algorithm. Only somatically acquired validated mutations are displayed (cases 69, 100, S21 and 88). The number of mutations (n) for each gene in the analysis is shown (bottom). **2D.** Percentage of cases harboring clonal or subclonal mutations for each of the genes displayed. **2E.** Kaplan-Meier and log-rank analysis for progression-free survival (PFS) in patients carrying *SETD2* abnormalities ("*SETD2* ab") but wild-type for *TP53* or *ATM* deletion and/or mutation compared to those with TP53 abnormalities ("*TP53* ab") and those wild-type for *TP53*, *ATM* and *SETD2* ("Wild-type"). **2F.** Kaplan-Meier and log-rank analysis for overall survival (OS) in the same categories described in E.

Table 1

	Patient ID	*SETD2* Mutation cDNA change	*SETD2* Mutation protein change	Functional prediction (Polyphen2;SIFT)	Somatically acquired validated	dbSNP	MAF 1000 Genomes	Mutation Taster	Conserved
Discovery cohort	69	c.5863G>C	p.E1955Q	P;D	yes	rs761536283	-	P	*M.Musculus*
	100	c.296A>G	p.D99G	-;D	yes	-	-	M	*M.Musculus*
	255	c.4633C>A	p.Q1545K	P;T	ND	-	-	M	*D.Melano*
	S21	c.3918G>A	p.W1306*	D;D	yes	-	-	M	*G.Gallus*
Extension cohort (includes CLL8 cases)	149	c.2008G>A	p.E670K	P;D	yes	rs374976472	-	M	*G.Gallus*
	88	c.5224A>C	p.M1742L	P;D	yes	-	-	M	*D.Rerio*
	4273	c.148G>A	p.A50T	D;D	ND	rs191985301	0.020% (1/5008)	P	*M.Musculus*
	4530	c.6885A>G	p.I2295M	B;D	ND	rs150476239	0.020% (1/5008)	M	*D.Melano*
	4546	c.500C>T	p.P167L	B;-	ND	rs78682369	0.020% (1/5008)	P	*not conserved*
	4715	c.5666T>C	p.M1889T	P;D	ND	rs148097513	0.040% (2/5008)	M	*G.Gallus*
	4172	c.5666T>C	p.M1889T	P;D	ND	rs148097513	0.040% (2/5008)	M	*G.Gallus*
	4426	c.1604A>G	p.N535S	B;T	ND	-	-	P	*M.Musculus*
	4426	c.265C>T	p.L89F	B,D	ND	-	-	P	*M.Musculus*
	#266	c.5755-5781delGAAGAGGAAGAATTGCAGTCACAAC	p.EEEELQSQQ1919fs	-;-	yes	-	-	M	*M.Musculus*
	#278	c.5411_5412delAC	p.L1804fs	-;-	yes	-	-	M	*partly conserved*
	#269	c.4727_4741delTCCTAGAATATTGTG	p.VLEYC1576del	-;-	yes	-	-	M	*M.Musculus*
	#313	c.3568G>A	p.V1190M	B;T	yes	-	-	P	*M.Musculus*
Landau et al# and Puente et al*	#028	c.6433G>T	p.G2145*	-;D	yes	-	-	M	*M.Musculus*
	#065	c.4973C>G	p.S1658*	-;T	yes	-	-	M	*M.Musculus*
	*15	c.2225C>G	p.S742*	-;T	yes	-	-	M	*partly conserved*
	*#141	c.6985C>T	p.Q2329*	-;T	yes	-	-	M	*not conserved*
	*#141	c.7616A>T	p.K2539I	D;D	yes	-	-	M	*M.Musculus*
	*177	c.3876-3877delGT	p.Y1293fs	-;-	yes	-	-	M	*M.Musculus*

Footnote: *ND: Not done, due to lack of germline material. PolyPhen2 prediction (B = Benign; P = Probably Damaging; D = Damaging; - = No prediction). SIFT prediction (D = Damaging; T = Tolerated; - = No prediction). MAF (minimal allele frequency in 1000 Genomes project). MutationTaster2 prediction (P=polymorphism; M=disease causing). Underlined text indicates an AID/APOBEC recognition motif. Mutation annotation was performed against COSMIC v 73 and no overlapping mutations were found. *Cases included in Puente et al and # Landau et al.*