

# Mutations driving CLL and their evolution in progression and relapse

Dan A. Landau<sup>1,2,3,4\*</sup>, Eugen Tausch<sup>5\*</sup>, Amaro N. Taylor-Weiner<sup>1\*</sup>, Chip Stewart<sup>1</sup>, Johannes G. Reiter<sup>1,2,6,7</sup>, Jasmin Bahlo<sup>8</sup>, Sandra Kluth<sup>8</sup>, Ivana Bozic<sup>7,9</sup>, Mike Lawrence<sup>1</sup>, Sebastian Böttcher<sup>10</sup>, Scott L. Carter<sup>1,11</sup>, Kristian Cibulskis<sup>1</sup>, Daniel Mertens<sup>5,12</sup>, Carrie L. Sougnez<sup>1</sup>, Mara Rosenberg<sup>1</sup>, Julian M. Hess<sup>1</sup>, Jennifer Edelmann<sup>5</sup>, Sabrina Kless<sup>5</sup>, Michael Kneba<sup>10</sup>, Matthias Ritgen<sup>10</sup>, Anna Fink<sup>8</sup>, Kirsten Fischer<sup>8</sup>, Stacey Gabriel<sup>1</sup>, Eric S. Lander<sup>1</sup>, Martin A. Nowak<sup>7,9,13</sup>, Hartmut Döhner<sup>5</sup>, Michael Hallek<sup>8,14§</sup>, Donna Neuberg<sup>15§</sup>, Gad Getz<sup>1,16§</sup>, Stephan Stilgenbauer<sup>5§</sup> & Catherine J. Wu<sup>1,2,3,4§</sup>

Which genetic alterations drive tumorigenesis and how they evolve over the course of disease and therapy are central questions in cancer biology. Here we identify 44 recurrently mutated genes and 11 recurrent somatic copy number variations through whole-exome sequencing of 538 chronic lymphocytic leukaemia (CLL) and matched germline DNA samples, 278 of which were collected in a prospective clinical trial. These include previously unrecognized putative cancer drivers (*RPS15*, *IKZF3*), and collectively identify RNA processing and export, MYC activity, and MAPK signalling as central pathways involved in CLL. Clonality analysis of this large data set further enabled reconstruction of temporal relationships between driver events. Direct comparison between matched pre-treatment and relapse samples from 59 patients demonstrated highly frequent clonal evolution. Thus, large sequencing data sets of clinically informative samples enable the discovery of novel genes associated with cancer, the network of relationships between the driver events, and their impact on disease relapse and clinical outcome.

In recent years, unbiased massively parallel sequencing of whole exomes (WES) in chronic lymphocytic leukaemia (CLL) has yielded fresh insights into the genetic basis of this disease<sup>1–4</sup>. Two important constraints have limited previous WES analyses. First, cohort size is critical for statistical inference of cancer drivers<sup>5</sup>, and previous CLL WES series<sup>3</sup> had a power of only 68%, 23% and 7% to detect putative CLL genes mutated in 5%, 3% and 2% of patients, respectively (<http://www.tumorportal.org/power>)<sup>5</sup>. Limited cohort size has also curtailed the ability to effectively learn the relationships between CLL driver events, such as their co-occurrence and the temporal order of their acquisition. Second, the composition of the cohort of previous WES studies has limited the ability to accurately determine the impact of drivers and clonal heterogeneity on clinical outcome, since they included samples collected at variable times from subjects exposed to a variety of therapies.

To overcome these challenges, we analysed WES data from 538 CLLs, including 278 pre-treatment samples collected from subjects enrolled on the phase III CLL8 study<sup>6</sup>. This trial established the combination of fludarabine (F), cyclophosphamide (C) and rituximab (R) as the current standard-of-care first-line treatment for patients of good physical fitness, with a median of >6 years of follow-up. Here we report the discovery of novel genes associated with CLL, the comprehensive genetic characterization of samples from patients before exposure to a uniform and contemporary treatment, and the uncovering of features contributing to relapse from this therapy.

## Unbiased candidate CLL gene discovery

We performed WES of CLL and matched germline samples, collected from 278 subjects enrolled on the CLL8 trial, with mean read depth of 95.0 and 95.7, respectively (Supplementary Tables 1 and 2). Consistent with previous CLL WES studies, we detected a mean  $\pm$  s.d. rate of  $21.5 \pm 7.9$  silent and non-silent single nucleotide variants (sSNVs) and somatic insertions and deletions (sIndels) per exome (Supplementary Tables 2 and 3)<sup>1,3</sup>.

We inferred candidate cancer-associated genes in CLL through implementation of MutSig2CV<sup>5,7</sup>. To maximize statistical sensitivity for driver detection<sup>5</sup>, we combined the CLL8 cohort with two previously reported and non-overlapping WES cohorts<sup>1,3</sup>, thereby increasing the size of the cohort to 538 CLLs. This cohort size is expected to saturate candidate CLL gene discovery for genes mutated in 5% of patients, and provides 94% and 61% power to detect genes mutated in 3% and 2% of patients, respectively<sup>5</sup>.

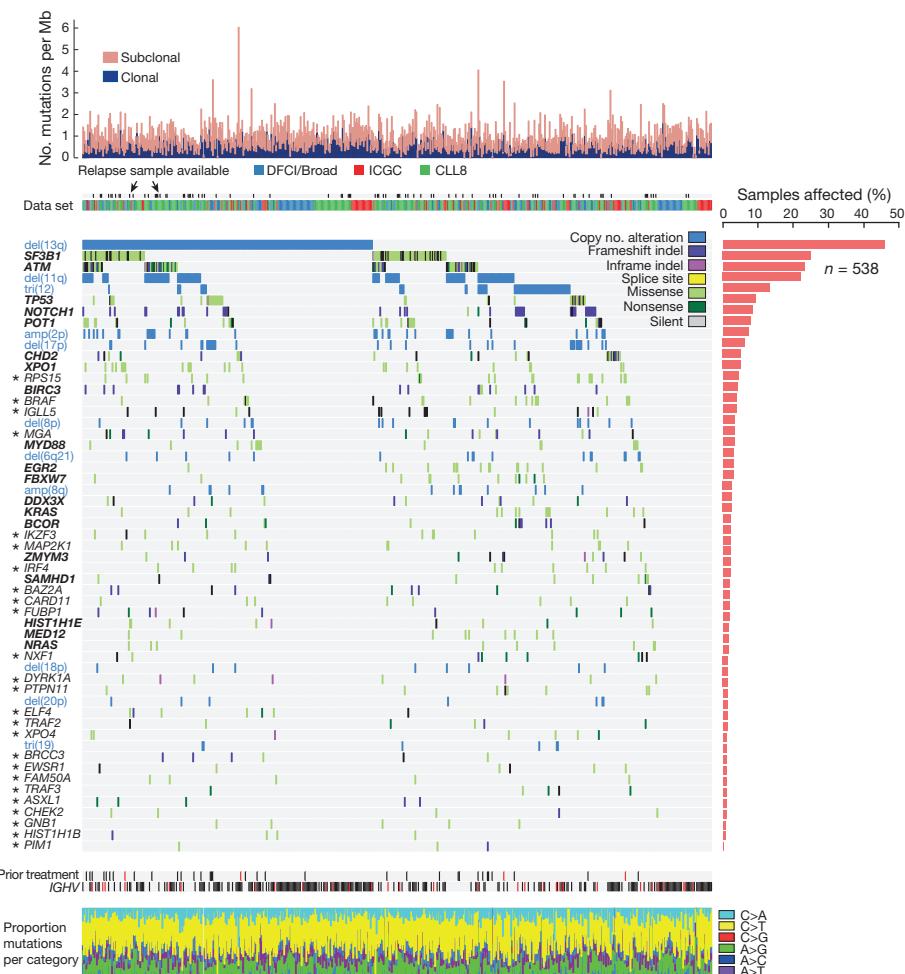
We detected 44 putative CLL driver genes, including 18 CLL mutated drivers that we previously identified<sup>3</sup>, as well as 26 additional putative CLL genes (Figs 1 and 2 and Extended Data Figs 1 and 2). In total, 33.5% of CLLs harboured a mutation in at least one of these 26 additional genes. Targeted DNA sequencing as well as variant allele expression by RNA-seq demonstrated high rates of orthogonal validation (Extended Data Fig. 3).

Of the newly identified putative cancer-associated genes, some were previously suggested as CLL drivers in studies using other

<sup>1</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>4</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Department of Internal Medicine III, Ulm University, Ulm 89081, Germany. <sup>6</sup>IST Austria (Institute of Science and Technology Austria), Klosterneuburg 3400, Austria. <sup>7</sup>Program for Evolutionary Dynamics, Harvard University, Cambridge 02138, Massachusetts, USA. <sup>8</sup>Department of Internal Medicine and Center of Integrated Oncology Cologne Bonn, University Hospital, Cologne 50937, Germany. <sup>9</sup>Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>10</sup>Department of Internal Medicine II, University Hospital of Schleswig-Holstein, Campus Kiel, Kiel 24105, Germany. <sup>11</sup>Joint Center for Cancer Precision Medicine, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>12</sup>Mechanisms of Leukemogenesis, German Cancer Research Center (DKFZ), Heidelberg 69121, Germany. <sup>13</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>14</sup>Cologne Cluster of Excellence in Cellular Stress Responses in Aging-associated Diseases (CECAD), Cologne 50931, Germany. <sup>15</sup>Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>16</sup>Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02129, USA.

\*These authors contributed equally to this work.

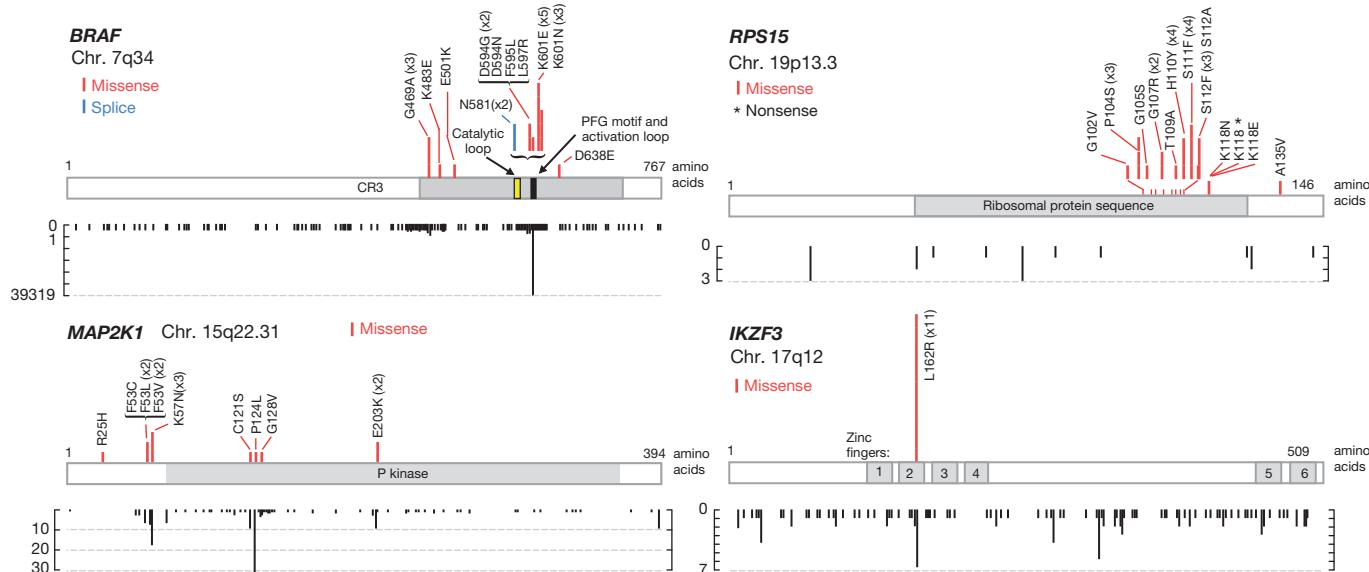
§These authors jointly supervised this work.



**Figure 1 | The landscape of putative driver gene mutations and recurrent somatic copy number variations in CLL.** Somatic mutation information is shown across the 55 putative driver genes and recurrent somatic copy number alterations (rows) for 538 primary patient samples (from CLL8 (green), Spanish ICGC (red) and DFCI/Broad (blue)) that underwent WES (columns). Blue labels, recurrent somatic CNA; bold labels, putative CLL cancer genes previously identified in ref. 3; asterisked labels, additional cancer-associated genes identified in this study. Samples were annotated for *IGHV* status (black, mutated; white unmutated; red, unknown), and for exposure to therapy before sampling (black, previous therapy; white, no previous therapy; red, unknown previous treatment status).

detection platforms. For example, the suppressor of MYC *MGA* ( $n = 17$ , 3.2%), which we detected as recurrently inactivated by insertions and nonsense mutations, was previously found to be inactivated through deletions<sup>8</sup> and truncating mutations<sup>8,9</sup> in high-risk CLL

(Extended Data Fig. 4). A gene set enrichment analysis of matched RNA-seq data revealed downregulation of genes that are suppressed upon MYC activation in B cells<sup>10</sup> (Supplementary Table 4). In addition to *MGA*, we report two additional candidate driver genes that



**Figure 2 | Selected novel, putative driver gene maps.** Individual gene mutation maps for select putative drivers, showing mutation subtype (for example, missense), position and evidence of mutational hotspots, based on

COSMIC database information (remaining gene maps shown in Extended Data Fig. 4).  $y$  axis counts at the bottom of the maps reflect the number of identified mutations in the COSMIC database.

probably modulate MYC activity (*PTPN11* (ref. 11) ( $n = 7$ , 1.3%) and *FUBP1* (ref. 12) ( $n = 9$ , 1.7%)), highlighting MYC-related proteins as drivers of CLL.

Another cellular process affected by novel CLL drivers is the MAPK–ERK pathway, with 8.7% of patients harbouring at least one mutation in CLL genes in this pathway. These included mutations in RAS genes (*NRAS*,  $n = 9$  and *KRAS*,  $n = 14$ , totalling 4.1%); *BRAF* ( $n = 21$ , 3.7%); or the novel putative driver *MAP2K1* ( $n = 12$ , 2%). This finding suggests that further therapeutic exploration of MAPK–ERK pathway inhibitors in CLL would be beneficial. Notably, *BRAF* mutations in CLL did not involve the canonical hotspot (V600E) seen in other malignancies<sup>5,13,14</sup>, but rather clustered heavily around the activation segment of the kinase domain (Fig. 2). This may be indicative of a different mechanism of activity<sup>15,16</sup>, and has clinical implications, as *BRAF* inhibitors are thought to be less effective for non-canonical *BRAF* mutations<sup>17,18</sup>.

In addition to highlighting novel cellular processes and pathways affected in CLL, many of the 26 additional CLL genes more densely annotated pathways or functional categories previously identified in CLL<sup>19</sup>, including RNA processing and export (*FUBP1*, *XPO4*, *EWSR1* and *NXF1*), DNA damage (*CHEK2*, *BRCC3*, *ELF4* (ref. 20) and *DYRK1A* (ref. 21)), chromatin modification (*ASXL1*, *HIST1H1B*, *BAZ2B* and *IKZF3*) and B-cell-activity-related pathways (*TRAF2*, *TRAFA3* and *CARD11*).

We discovered a number of putative CLL drivers previously unrecognized in human cancer. In a first example, we found that *RPS15* was recurrently mutated ( $n = 23$ , 4.3%), with mutations localized to the carboxy-terminal region (Fig. 2) at highly conserved sites (median conservation score of 94 out of 100). This component of the S40 ribosomal subunit has not been extensively studied in cancer, although rare mutations have been identified in Diamond–Blackfan anaemia<sup>22</sup>. A gene set enrichment analysis revealed upregulation of gene sets related to adverse outcome in CLL as well as immune response gene sets (Supplementary Table 4). In another example of a previously unrecognized cancer gene, we identified recurrent L162R substitutions ( $n = 11$ , 2.0%) in *IKZF3*, targeting a highly conserved amino acid (93 out of 100 conservation score). This gene is a key transcription factor in B-cell development<sup>23</sup>, and its upregulation has been associated with adverse outcome<sup>24,25</sup>.

In addition to sSNVs and sIndels, we characterized somatic copy number alterations (CNAs) directly from the WES data (Extended Data Fig. 5 and Supplementary Tables 5 and 6). When we accounted for all 55 identified driver events—including non-silent sSNVs and sIndels in putative CLL genes ( $n = 44$ ), and recurrent somatic CNAs ( $n = 11$ )—91.1% of CLLs contained at least one driver. Moreover, 65.4% of CLLs now harboured at least 2 drivers, and 44.4% at least 3 drivers, compared with 55.9% and 31.8% were we to exclude the 26 additional CLL genes.

## Drivers and CLL characteristics

The larger cohort size also provided statistical power to examine associations between genetic alterations and key CLL features. First, we examined whether mutations differed between *IGHV* mutated and unmutated subtypes, the two main subtypes of CLL. In agreement with the relative clinical aggressiveness of *IGHV* unmutated CLL, most drivers were found in a higher proportion in this subtype (Extended Data Fig. 6a). Only three driver genes were enriched in the *IGHV* mutated CLL (del(13q), *MYD88* and *CHD2*), suggesting a role for these specific alterations within the oncogenic process of this subtype.

Second, since therapy could lead to selection of particular driver events, we examined the 33 samples (6.2%, none enrolled on CLL8) that had received therapy before sampling. Previous treatment was associated with enrichment in *TP53* and *BIRC3* mutations del(17p) and del(11q), as previously indicated<sup>26</sup>, as well as in mutated *DDX3X* and *MAP2K1*, suggesting their selection by therapeutic interventions (Extended Data Fig. 6b).

Third, we examined whether coherent patterns of co-occurrence of driver events were evident, limiting our analysis to the 31 drivers with >10 affected patients. Of 465 possible pairs, 11 combinations had statistically significant high or low co-occurrence (Extended Data Fig. 6c, d). As expected, a high degree of co-occurrence was found between mutated *TP53* and del(17p), and between mutated *ATM* and del(11q). Both mutated *ATM* and del(11q) significantly co-occurred with amp(2p), and associations between the presence of tri(12) with mutated *BIRC3* and with mutated *BCOR* were also found. A significantly low rate of co-occurrence was seen between del(13q) and tri(12).

Fourth, we examined the temporal sequence of driver acquisition in the evolutionary history of CLL. To do this, we computed the cancer-cell fraction (CCF) of each mutation across the 538 samples, and identified mutations as either clonal or subclonal<sup>27</sup> (58.1% of mutations classified as subclonal). Both clonal and subclonal sSNVs were similarly dominated by C>T transitions at CpG sites (Extended Data Fig. 7).

We first classified driver events probably acquired earlier or later in the disease course based on the proportion of cases in which the driver was found as clonal (Fig. 3a). This large data set further enabled the inference of temporal relationships between pairs of drivers. We systematically identified instances in which a clonal driver was found together with a subclonal driver within the same sample, as these pairs reflect the acquisition of one lesion (clonal) followed by another (subclonal), providing a temporal ‘edge’ leading from the former to the latter<sup>28,29</sup>. For each driver, we calculated the relative enrichment of out-going edges compared to in-going edges to define early, late and intermediary drivers (Supplementary Table 7). For 23 pairs connected by at least 5 edges, we further established the temporal relationship between the two drivers in each pair, and thereby constructed a temporal map of the evolutionary trajectories of CLL (Supplementary Table 8 and Fig. 3b). This network highlights somatic CNAs as the earliest events with two distinct points of departure involving del(13q) and tri(12). It further demonstrates an early convergence towards del(11q) and substantial diversity in late drivers. Finally, this analysis suggests that in the case of the tumour suppressor genes *ATM* and *BIRC3*, copy loss precedes sSNVs and sIndels in biallelic inactivation.

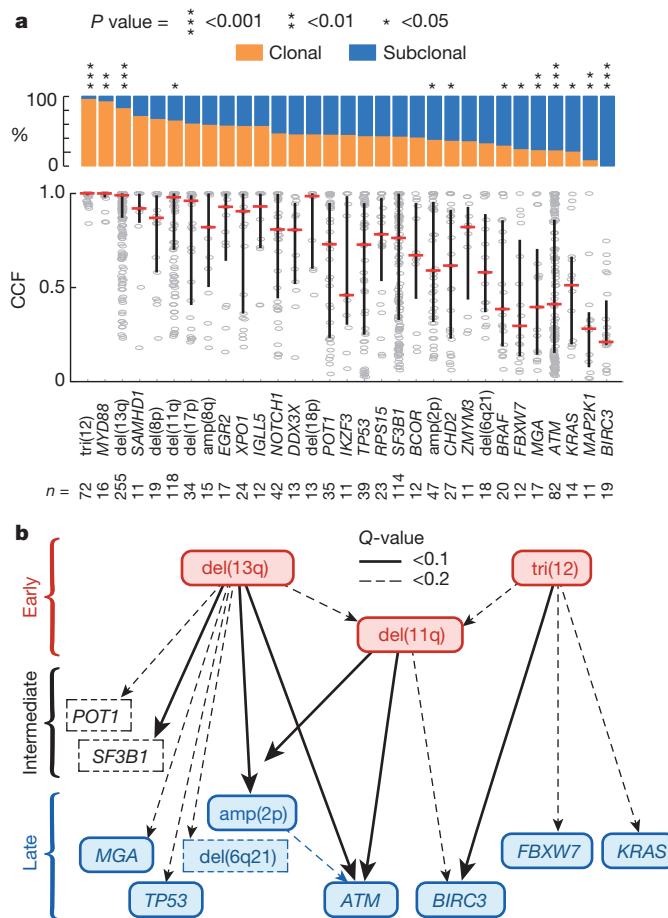
## Impact on clinical outcome

We examined whether the presence of any of the drivers detected in at least 10 of the 278 pre-treatment CLL8 samples was associated with impact on clinical outcome (Fig. 4a and Extended Data Figs 8 and 9; the genomics analysis team was blinded to the clinical outcome data). Previous investigations suggested an impact for 7 CLL genes (*SF3B1*, *ATM*, *TP53*, *XPO1*, *EGR2*, *POT1* and *BIRC3*)<sup>30–33</sup>. We found shorter progression-free survival (PFS) associated only with *TP53* and *SF3B1* mutations. Of the newly identified recurrent lesions evaluated (*MGA*, *BRAF* and *RPS15*), we observed a shorter PFS with mutated *RPS15* (Bonferroni  $P = 0.024$ ).

The presence of a detectable pre-treatment subclonal driver has been previously associated with shorter remissions in patients treated with heterogeneous therapies<sup>3</sup>. In the CLL8 cohort, we again found that the presence of a pre-treatment subclonal driver was associated with a significantly shorter PFS (hazard ratio (HR) 1.6 (95% confidence interval (CI) 1.2–2.2),  $P = 0.004$ ). This association remained significant in both the FC (fludarabine and cyclophosphamide) and FCR (fludarabine, cyclophosphamide and rituximab) treatment arms (Fig. 4b), with a non-significant trend when *IGHV* mutation status was added to a multivariable model in addition to the treatment arm (1.3 (0.9–1.9),  $P = 0.102$ ).

## Clonal evolution at disease relapse

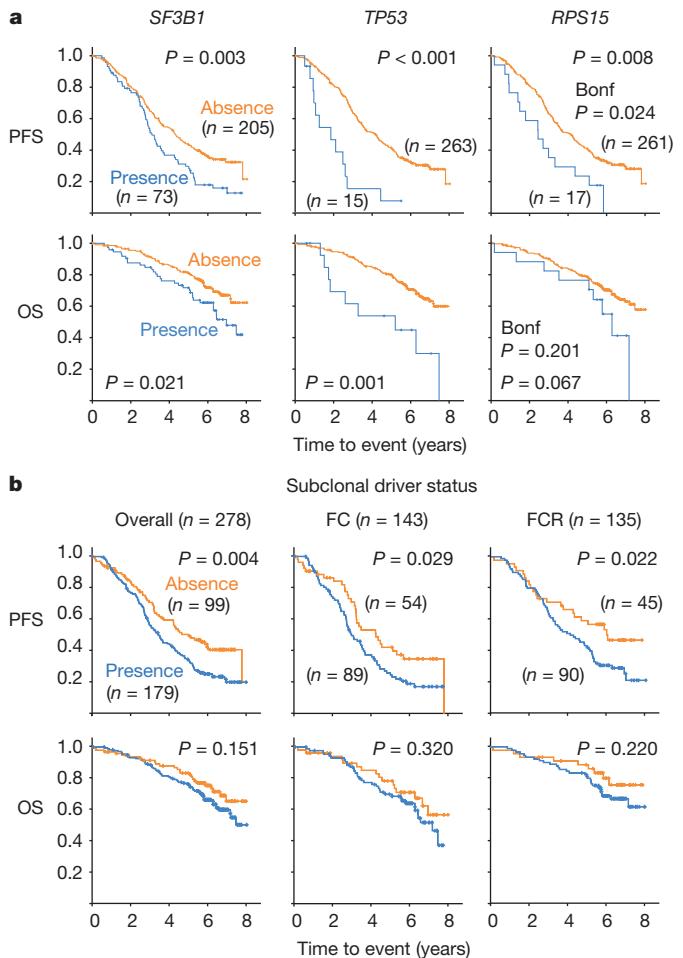
To define clonal evolution in disease relapse, we performed WES on matched samples collected at the time of relapse from 59 of 278 CLL8



**Figure 3 | Inferred evolutionary history of CLL.** **a**, The proportion in which a recurrent driver is found as clonal or subclonal across the 538 samples is provided (top), along with the individual cancer cell fraction (CCF) values for each sample affected by a driver (tested for each driver with a Fisher's exact test, comparing to the cumulative proportions of clonal and subclonal drivers excluding the driver evaluated). Median CCF values are shown (bottom, bars represent the median and interquartile range for each driver). **b**, Temporally direct edges are drawn when two drivers are found in the same sample, one in clonal and the other in subclonal frequency. These edges are used to infer the temporal sequences in CLL evolution, leading from early, through intermediate to late drivers. Note that only driver pairs with at least five connecting edges were tested for statistical significance and only drivers connected by at least one statistically significant edge are displayed (see Supplementary Methods and Supplementary Tables 6 and 7).

subjects (Supplementary Tables 9 and 10). We observed large clonal shifts between pre-treatment and relapse samples in the majority of cases (57 of 59), thus demonstrating that CLL evolution after therapy is the rule rather than the exception (Fig. 5a). The relapse clone was already detectable in pre-treatment WES in 18 of 59 (30%) cases, demonstrating that the study of pre-treatment diversity anticipates the future evolutionary trajectories of the relapsed disease<sup>34</sup>. By targeted deep sequencing, we screened for relapse drivers in 11 of the 41 of pre-treatment samples in which WES did not detect the relapse driver. In 7 of these 11 CLLs, at least one relapse driver was detected in the pre-treatment sample (Supplementary Table 10).

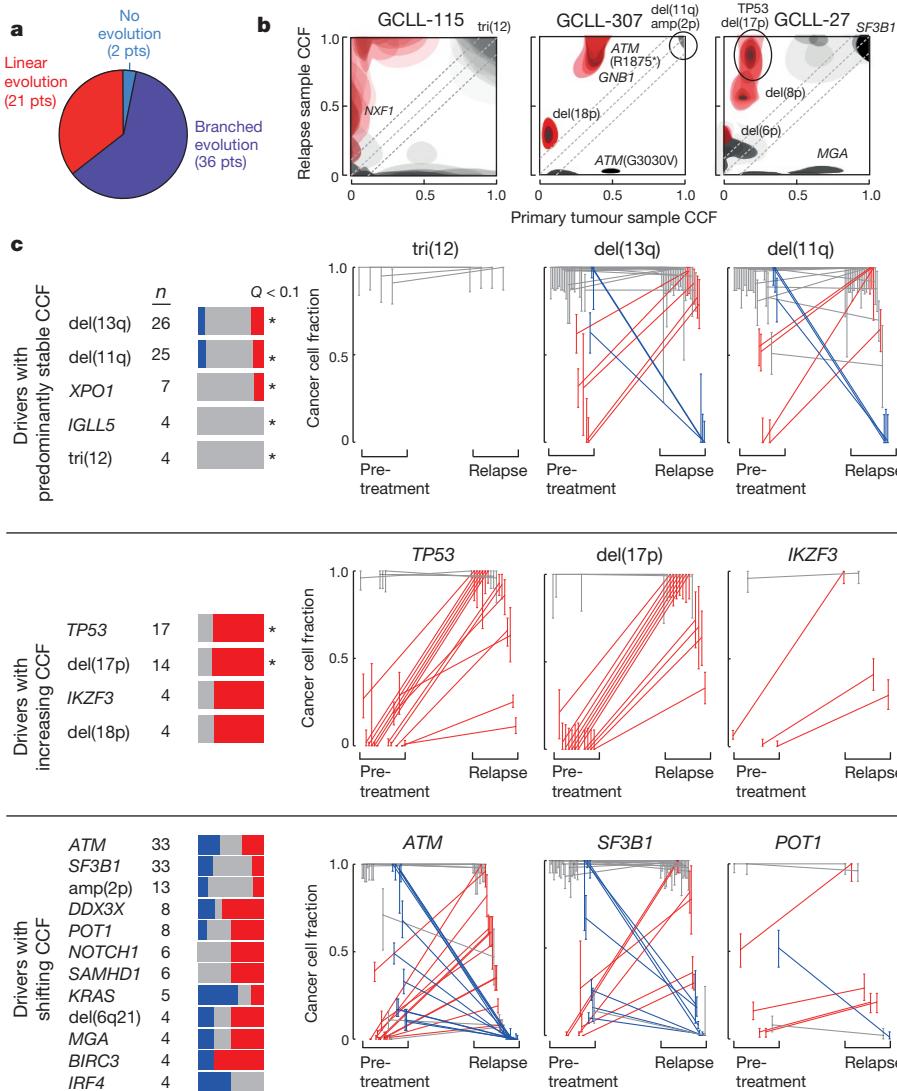
We further compared the pre-treatment and relapse CCF for each driver, and observed three general patterns. First, tri(12), del(13q) and del(11q), suggested as early drivers (Fig. 3b), tended to remain stably clonal despite marked, often branched, evolution (Fig. 5b (CLL cases GCLL-115 and GCLL-307), Fig. 5c, top row, and Extended Data Fig. 10). This confirms that these are indeed early events probably shared by the entire malignant population. Second, TP53 mutations and del(17p) demonstrated increases in CCF upon relapse, suggesting



**Figure 4 | Associations of CLL drivers with clinical outcome.** **a**, Kaplan-Meier analysis (with logrank P values) for putative drivers with associated impact on progression-free survival (PFS) or overall survival (OS) probabilities in the cohort of 278 patients that were treated as part of the CLL8 trial. For candidate CLL genes tested here for the first time regarding impact on outcome, a Bonferroni P value is also shown. **b**, Presence of a subclonal driver is associated with a lower probability of PFS, in both the FC and FCR arms, and a trend towards shorter OS.

a fitness advantage under therapeutic selection (Fig. 5b (GCLL-27) and Fig. 5c, middle row). The novel driver IKZF3 increased in CCF in 3 of 4 relapse cases (and remained clonal in the fourth), supporting the suggestion that these mutations probably enhance fitness. Third, mutations in SF3B1 and ATM, identified as temporally intermediate or late drivers, seemed just as likely to decline in CCF as they were to increase (Fig. 5c, bottom row). These results suggest that within this therapeutic context such mutations do not provide the same strength of fitness advantage compared to TP53 disruption. In addition, we observed nine instances each of multiple distinct alleles of ATM and SF3B1 mutations within the same CLL (for example, GCLL-307 in Fig. 5b), indicating convergent evolution of these late-occurring CLL drivers.

This series also informs us regarding the mutagenesis of the tumour suppressor genes TP53 and ATM, where biallelic inactivation is common. In the case of ATM, we typically find a fixed clonal del(11q22.3) and subclones harbouring sSNVs affecting the other allele that shift in CCF over time (for example, GCLL-307). We confirmed that the breakpoints of somatic CNAs in matched relapse and pre-treatment samples were highly consistent, probably representing the same deletion event. These data suggest that mono-allelic ATM deletion provides a fitness advantage that enables the expansion of the malignant population with subsequent growth of multiple co-existing clones



**Figure 5 | Matched pre-treatment and relapse samples reveal patterns of clonal evolution in relation to therapy.** **a**, The number and proportion of the patterns of clonal evolution of CLLs studied at pre-treatment and at relapse. pts, patients. **b**, Selected plots of 2D clustering of pre-treatment and relapse CCF demonstrating clonal stability of tri(12) (CLL case GCLL-115), concordant increase in CCFs of TP53 and del(17p) (GCLL-27), and clonal shifts in ATM sSNVs in a sample with clonally stable monoallelic deletion of ATM (GCLL-307). Red colouring was added when greater than half of the CCF probability indicated >0.1 increase in CCF at relapse. **c**, Clonal evolution of CLL drivers. Left panel: for each driver with at least 4 instances detected across the 59 CLLs, the proportion of instances where the CCF increased (red), decreased (blue) or remained stable (grey) over time is shown (see Supplementary Methods for details of the statistical analysis). The driver events were distributed to three main groups: predominantly stable events (top); predominantly increasing CCF (middle); and all other patterns (bottom). Right panel: comparison (modal CCF with 95% CI) between pre-treatment and relapse samples for select CLL drivers (see Extended Data Fig. 10 for the remaining driver events from the cohort of 59 CLLs).

that harbour a ‘second hit’ (genetic disruption of the remaining allele). Thus, while a biallelic lesion is clearly selected for (Extended Data Fig. 6c), the longitudinal data support the temporal analysis (Fig. 3b) in which del(11q) precedes ATM mutations, reflecting the higher likelihood of a focal copy number loss compared with a deleterious point mutation<sup>35,36</sup>. In contrast, we consistently observed a concordant rise of del(17p) and TP53 mutations in all 12 CLLs harbouring both of these events, and none of these cases exhibited multiple distinctly evolving TP53 mutated clones. These observations suggest that a true biallelic inactivation of TP53 is required, and indeed, across the 538 CLL samples, the odds ratio for co-occurrence of del(17p) and TP53 mutation was far greater than the odds ratio for co-occurrence of del(11q) and ATM mutation (97.22 versus 10.99, respectively). These observations are in agreement with a recent analysis that suggested that with the exception of a few genes such as TP53, tumour suppressor genes in sporadic cancers are haploinsufficient to begin with, and that the second hit only further builds on this fitness advantage<sup>37</sup>.

## Conclusions

This study of WES in CLL enabled a comprehensive identification of putative cancer-associated genes in CLL, generating novel hypotheses regarding the biology of this disease, and identifying previously unrecognized putative CLL drivers such as RPS15 and IKZF3. The detailed characterization of the compendium of driver lesions in cancer is of

particular importance as we strive to develop personalized medicine, because driver genes may inform prognosis (for example, RPS15 mutations) and identify lesions that may be targeted by therapeutic intervention (for example, MAPK pathway mutations and specifically the unexpected enrichment for non-canonical BRAF mutations). Through the inclusion of samples collected within a landmark clinical trial with mature outcome data, we could further study the impact of genetic alterations in the context of the current standard-of-care front-line therapy. As targeted therapy is rapidly transforming the treatment algorithms for CLL, future studies will be required to re-examine these associations in this context<sup>38</sup>.

An important benefit of the larger cohort size is the enhanced ability to explore relationships between driver lesions based on patterns of their co-occurrence. Focusing on temporal patterns of driver acquisition—based on the distinction between clonal versus subclonal alterations in a cross-sectional analysis—we derived a temporal map for the evolutionary history of CLL. In the context of relapse after first-line fludarabine-based therapy, we note highly frequent clonal evolution, and that the future evolutionary trajectories were already anticipated in the pre-treatment sample in one-third of cases with WES.

This study provides an indication of the potential benefits to be gained by applying novel genomic technologies to growing cohort sizes across leukaemias: the continued discovery of novel candidate cancer genes, the deeper integration of genetic analysis with standardized clinical information (collected within clinical trials) to

inform prognosis and therapy, and the ability to delineate the complex network of relationships between cancer drivers in the history and progression of the malignant process.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 March; accepted 11 August 2015.

Published online 14 October 2015.

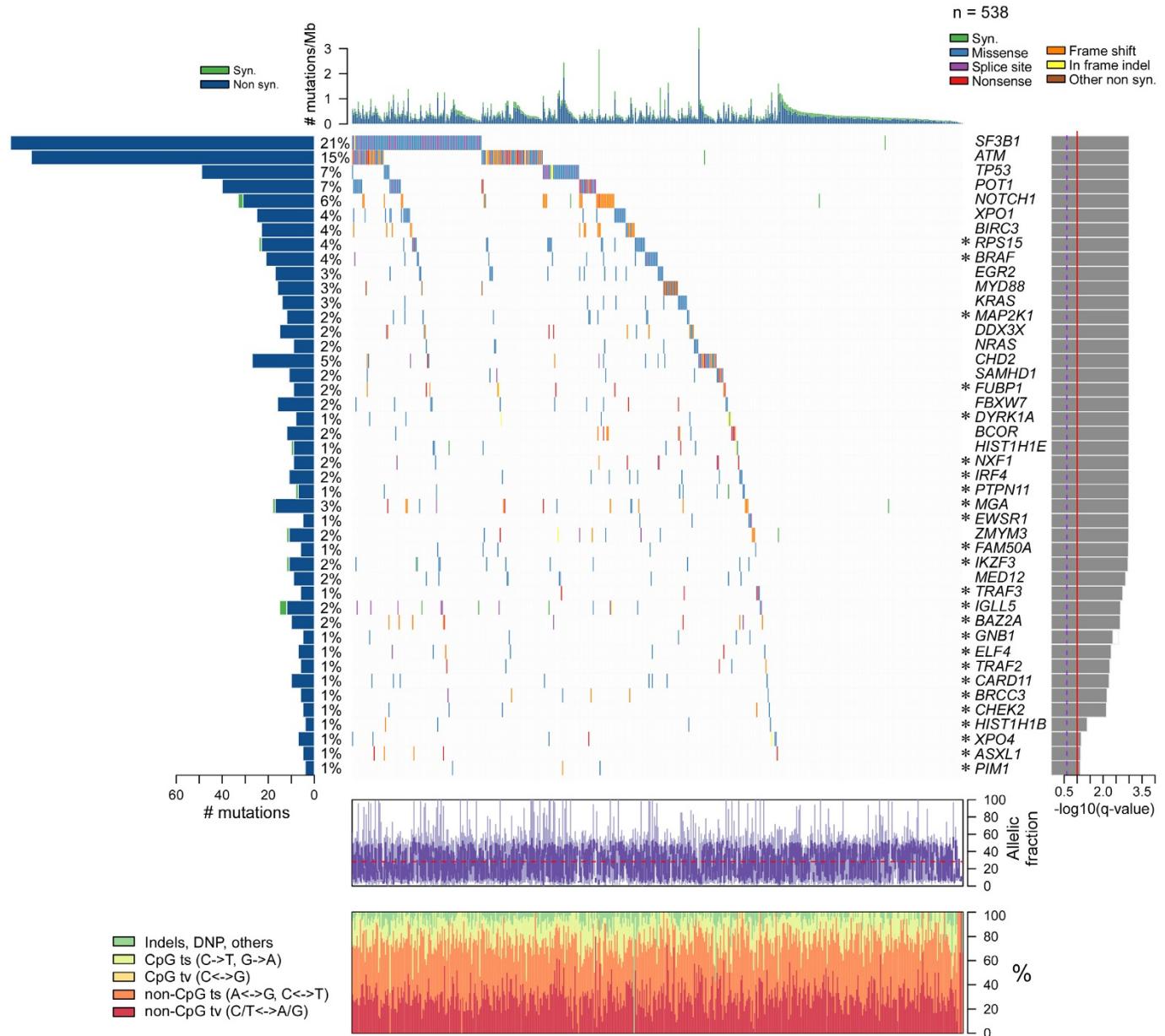
1. Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genet.* **44**, 47–52 (2012).
2. Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
3. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
4. Schuh, A. et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**, 4191–4196 (2012).
5. Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
6. Hallek, M. et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376**, 1164–1174 (2010).
7. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
8. Edelmann, J. et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* **120**, 4783–4794 (2012).
9. De Paoli, L. et al. MGA, a suppressor of MYC, is recurrently inactivated in high risk chronic lymphocytic leukemia. *Leuk. Lymphoma* **54**, 1087–1090 (2013).
10. Schlosser, I. et al. Dissection of transcriptional programmes in response to serum and c-Myc in a human B-cell line. *Oncogene* **24**, 520–524 (2005).
11. Jiang, X. et al. Critical role of SHP2 (PTPN11) signaling in germinal center-derived lymphoma. *Haematologica* **99**, 1834–1845 (2014).
12. Zhang, J. & Chen, Q. M. Far upstream element binding protein 1: a commander of transcription, translation and beyond. *Oncogene* **32**, 2907–2916 (2013).
13. Tacci, E. et al. BRAF mutations in hairy-cell leukemia. *N. Engl. J. Med.* **364**, 2305–2315 (2011).
14. Brastianos, P. K. et al. Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. *Nature Genet.* **46**, 161–165 (2014).
15. Heidorn, S. J. et al. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell* **140**, 209–221 (2010).
16. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
17. Yang, H. et al. RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. *Cancer Res.* **70**, 5518–5527 (2010).
18. Jeburaj, B. M. et al. BRAF mutations in chronic lymphocytic leukemia. *Leuk. Lymphoma* **54**, 1177–1182 (2013).
19. Landau, D. A. & Wu, C. J. Chronic lymphocytic leukemia: molecular heterogeneity revealed by high-throughput genomics. *Genome Med.* **5**, 47 (2013).
20. Sashida, G. et al. ELF4/MEF activates MDM2 expression and blocks oncogene-induced p16 activation to promote transformation. *Mol. Cell. Biol.* **29**, 3687–3699 (2009).
21. Park, J. et al. Dyk1A phosphorylates p53 and inhibits proliferation of embryonic neuronal cells. *J. Biol. Chem.* **285**, 31895–31906 (2010).
22. Gazda, H. T. et al. Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. *Am. J. Hum. Genet.* **83**, 769–780 (2008).
23. Ferreiros-Vidal, I. et al. Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. *Blood* **121**, 1769–1782 (2013).
24. Billot, K. et al. Dereulation of Aiolos expression in chronic lymphocytic leukemia is associated with epigenetic modifications. *Blood* **117**, 1917–1927 (2011).
25. Nückel, H. et al. The IKZF3 (Aiolos) transcription factor is highly upregulated and inversely correlated with clinical progression in chronic lymphocytic leukaemia. *Br. J. Haematol.* **144**, 268–270 (2009).
26. Beà, S. et al. Genetic imbalances in progressed B-cell chronic lymphocytic leukemia and transformed large-cell lymphoma (Richter's syndrome). *Am. J. Pathol.* **161**, 957–968 (2002).
27. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnol.* **30**, 413–421 (2012).
28. Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627 (2013).
29. Wang, J. et al. Tumor evolutionary directed graphs and the history of chronic lymphocytic leukemia. *Elife* **3** (2014).
30. Stilgenbauer, S. et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* **123**, 3247–3254 (2014).
31. Rossi, D. et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* **121**, 1403–1412 (2013).
32. Damm, F. et al. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov.* **4**, 1088–1101 (2014).
33. Winkelmann, N. et al. Low frequency mutations independently predict poor treatment-free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis. *Haematologica* **100**, e237–e239 (2015).
34. Puente, X. S. & Lopez-Otin, C. The evolutionary biography of chronic lymphocytic leukemia. *Nature Genet.* **45**, 229–231 (2013).
35. Nowak, M. A. et al. The role of chromosomal instability in tumor initiation. *Proc. Natl. Acad. Sci. USA* **99**, 16226–16231 (2002).
36. Bozic, I. et al. Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. USA* **107**, 18545–18550 (2010).
37. Davoli, T. et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).
38. Hallek, M. Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. *Am. J. Hematol.* **90**, 446–460 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank all members of the Broad Institute's Biological Samples, Genetic Analysis and Genome Sequencing Platforms, who made this work possible (NHGRI-U54HG003067). We further thank all patients and their physicians for CLL8 trial participation and donation of samples; M. Mendila, N. Valente, S. Zurfluh, M. Wenger and J. Wingate for their support in conception and conduct of the CLL8 trial. D.A.L. is supported by an ACS Postdoctoral Fellowship, ASH Scholar Award, and the Burroughs Wellcome Fund Career Award for Medical Scientists and by the NIH Big Data to Knowledge initiative (BD2K, 1K01ES025431-01). J.G.R. was supported by the European Research Council (ERC) start grant 279307: Graph Games, Austrian Science Fund (FWF) grant no. P23499-N23, and FWF NFN grant no S11407-N23 RISE. S.B. is supported by the German Jose Carreras Leukemia Foundation (project R 06/03v). M.H. is supported by the Deutsche Forschungsgemeinschaft (KFO 286, Project 6). S.S. is supported by the Else Kröner Fresenius-Stiftung (2010\_Kolleg24, 2012\_A146), Virtual Helmholtz Institute (VH-VI-404), CLL Global Research Foundation (Alliance), and Deutsche Forschungsgemeinschaft (SFB 1074 projects B1, B2). C.J.W. acknowledges support from the Blavatnik Family Foundation, AACR (SU2C Innovative Research Grant), and NIH/NCI (1R01CA182461-02, 1R01CA184922-01, 1U10CA180861-01).

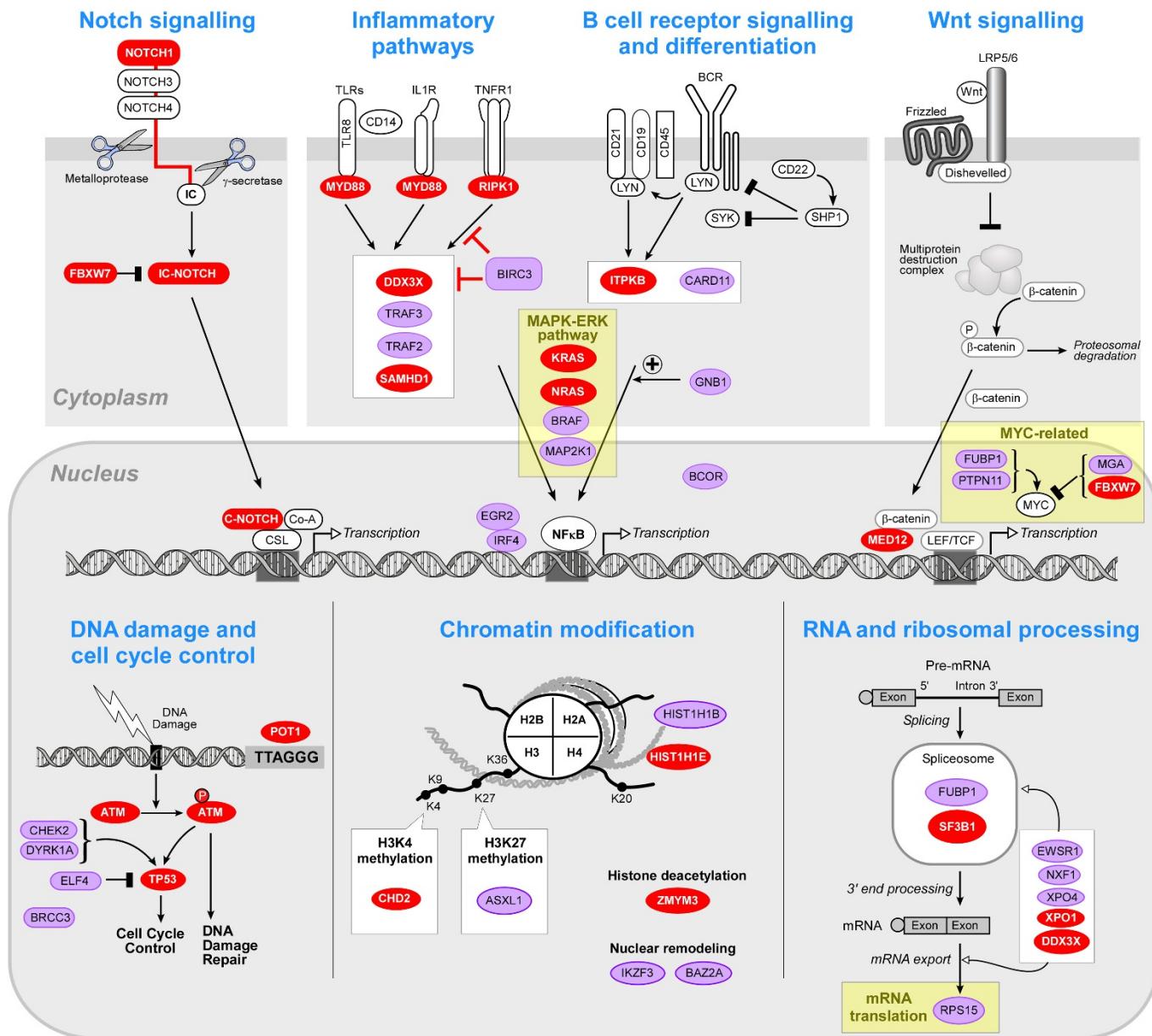
**Author Contributions** All authors contributed extensively to the work presented in this paper. D.A.L., D.N., G.G., E.T., S.S. and C.J.W. contributed to study conception and design. E.T., S.B., J.E., S.K., M.K., M.R., A.F., K.F., H.D., M.H. and S.S. performed patient selection, provided the DNA samples, and prior matched clinical and genetic data sets. C.L.S., S.G. and E.S.L. enabled sample sequencing. D.A.L., A.N.T., C.S., M.L., K.C., M.R., J.M.H., S.L.C. and G.G. contributed to the computational genomics analysis. D.A.L., E.T., J.G.R., J.B., S.K., I.B., D.M., M.A.N., D.N., G.G., S.S. and C.J.W. contributed to additional data analysis as well as manuscript preparation. All authors contributed to the writing of the manuscript.

**Author Information** CLL8 WES data is deposited in dbGaP under accession code phs000922.v1.p1. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.W. (c.wu@partners.org), S.S. (Stephan.Stilgenbauer@uniklinik-ulm.de) or G.G. (gadgetz@broadinstitute.org).



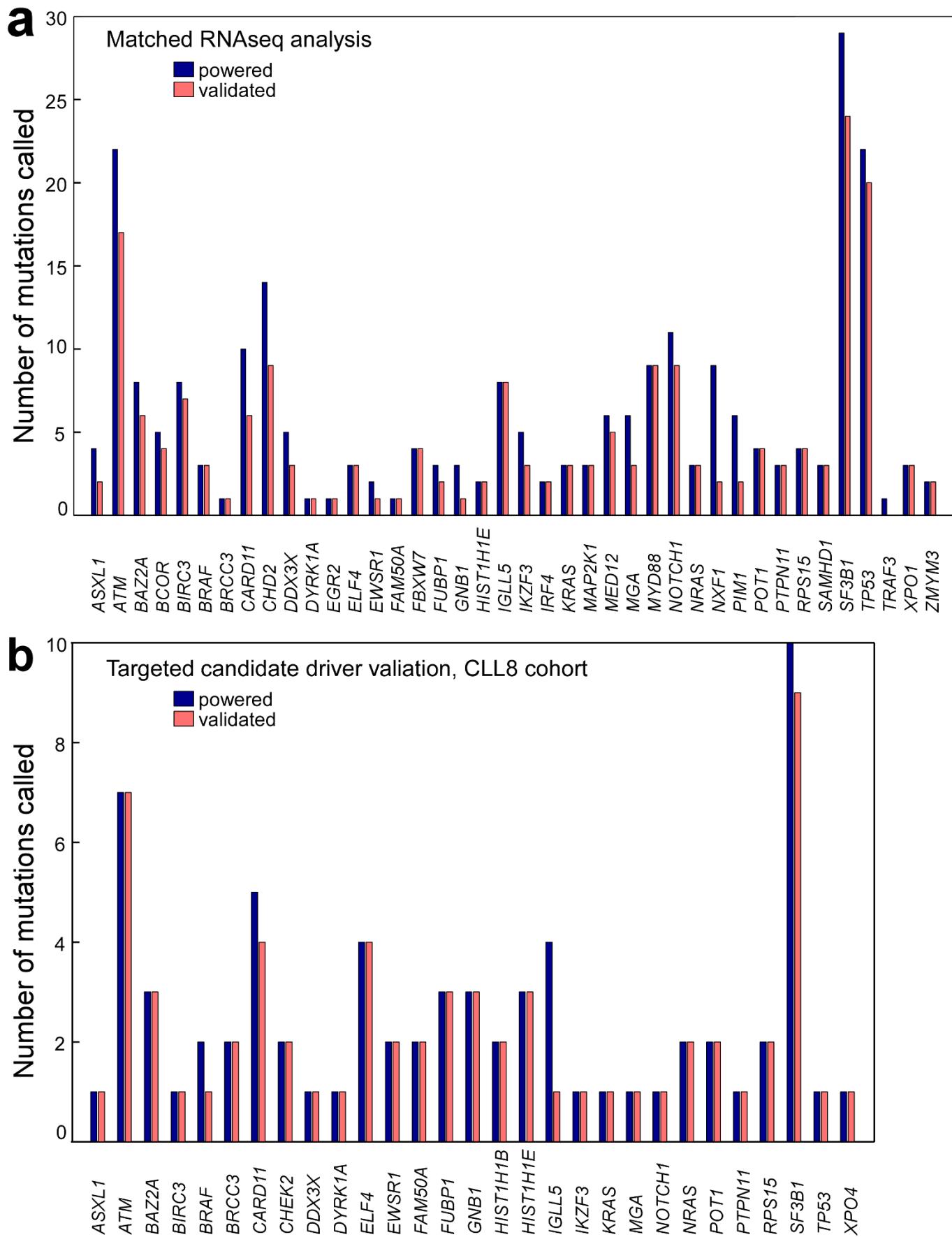
**Extended Data Figure 1 | Candidate CLL cancer genes discovered in the combined cohort of 538 primary CLL samples.** Significantly mutated genes identified in 538 primary CLL. Top panel: the rate of coding mutations (mutations per megabase) per sample. Centre panel: detection of individual genes found to be mutated (sSNVs or sIndels) in each of the 538 patient samples (columns), colour-coded by type of mutation. Only one mutation per gene

is shown if multiple mutations from the same gene were found in a sample. Right panel: Q-values (red,  $Q < 0.1$ ; purple dashed,  $Q < 0.25$ ) and Hugo symbol gene identification. New candidate CLL genes are marked with asterisks. Left panel: the percentages of samples affected with mutations (sSNVs and sIndels) in each gene. Bottom panel: plots showing allelic fractions and the spectrum of mutations (sSNVs and sIndels) for each sample.



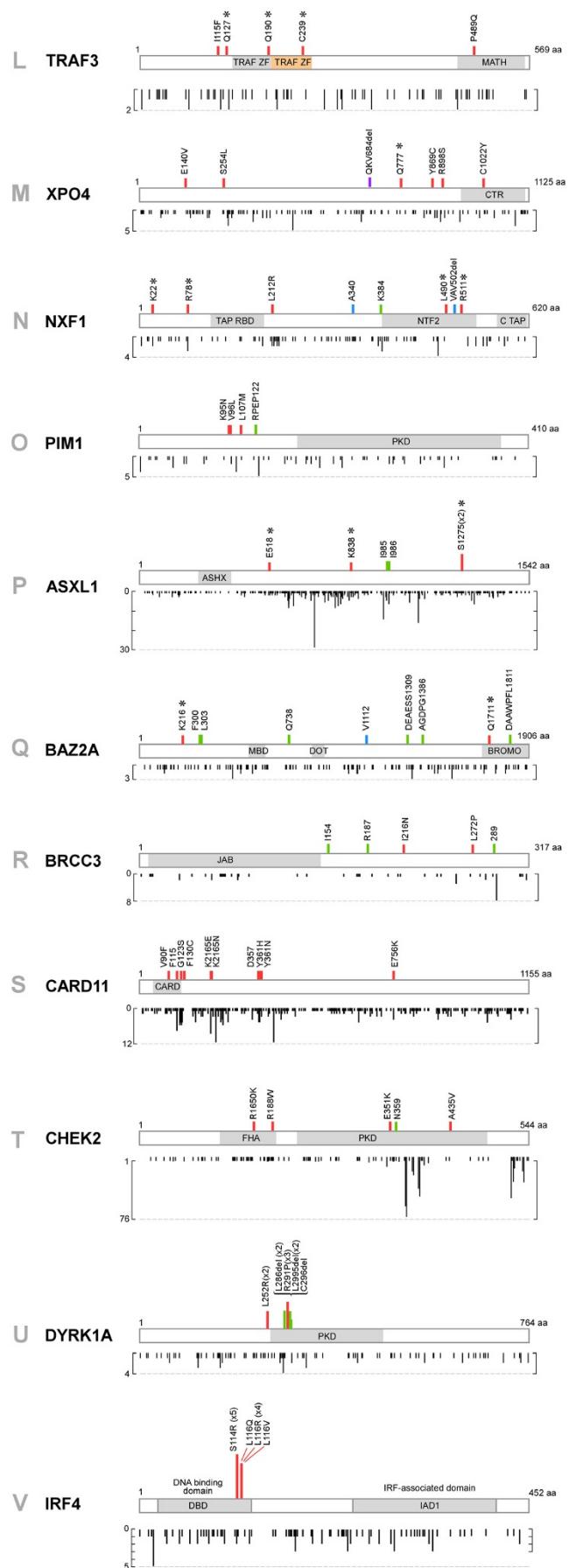
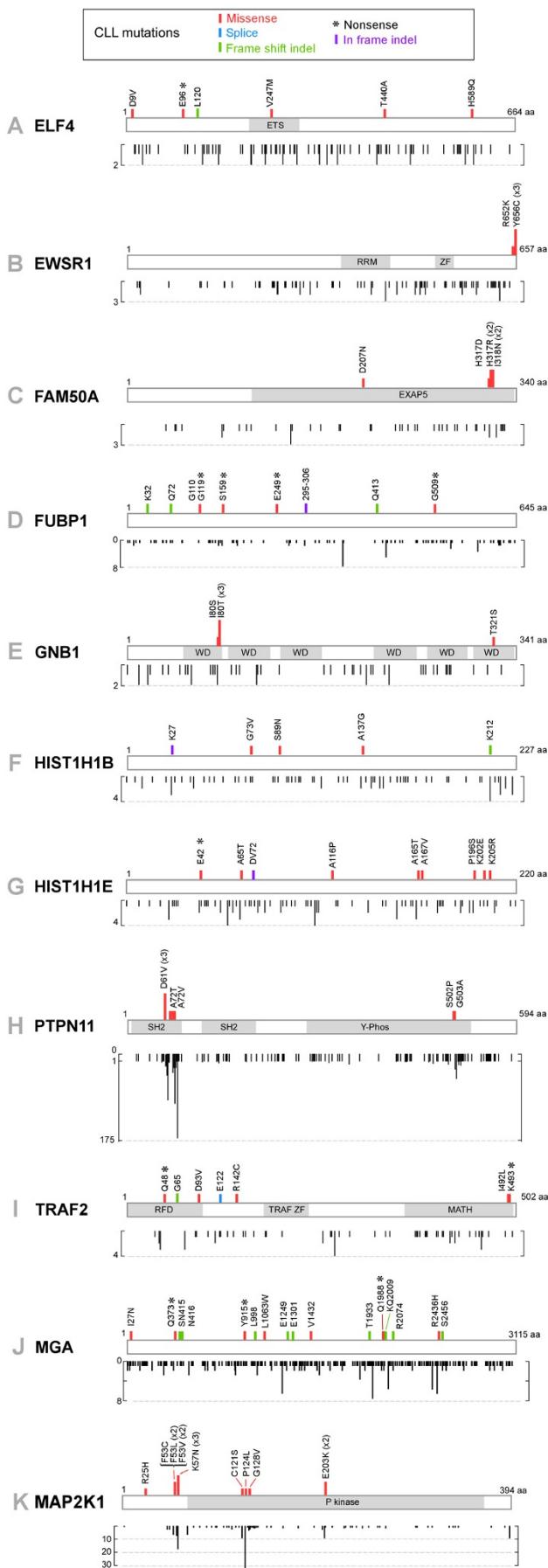
**Extended Data Figure 2 | Cellular networks and processes affected by putative CLL drivers.** Putative CLL cancer genes cluster in pathways that are central to CLL biology such as Notch signalling, inflammatory response and B-cell receptor signalling. In addition, proteins that participate in central cellular processes such as DNA damage repair, chromatin modification and

mRNA processing, export and translation are also recurrently affected. New CLL subpathways highlighted by the current driver discovery effort are shown in yellow boxes. Red circles indicate putative driver genes previously identified<sup>3</sup>; purple circles indicate genes newly identified in the current study.



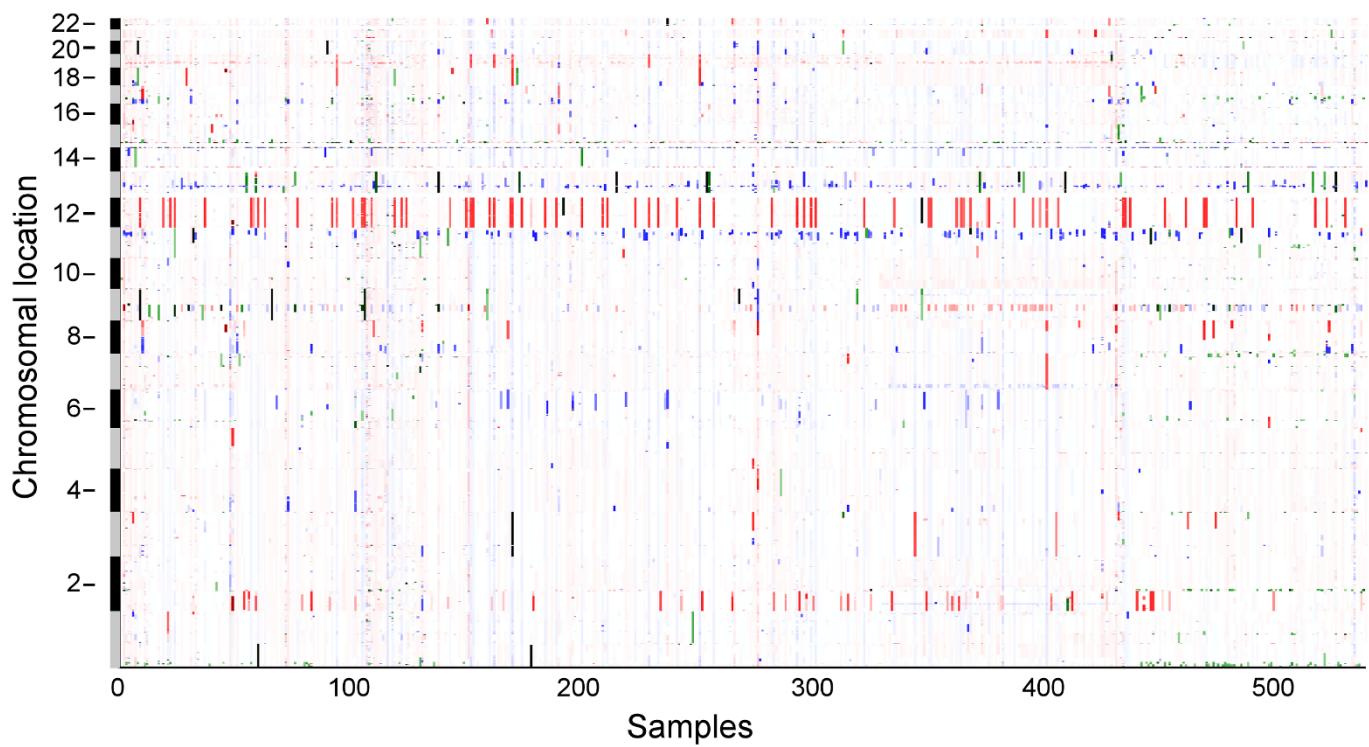
**Extended Data Figure 3 | RNA-seq expression data for candidate CLL genes and targeted candidate driver validation.** **a**, Matched RNA-seq and WES data were available for 156 CLLs (103 CLLs previously reported<sup>3</sup> and 53 CLLs from the ICGC studies<sup>1</sup>). From the WES of these 156 cases, we identified 318 driver mutations (sSNVs and sIndels). For each site, we quantified the number of alternative reads corresponding to the somatic mutation in matched RNA-seq data. We subsequently counted the number of instances in which a mutation was detected ('detected') and compared it to the number of instances in which mutation detection had >90% power based on the allelic

fraction in the WES and the read depth in the RNA-seq data ('powered'). Overall, we detected 78.1% of putative CLL gene mutations at sites that had >90% power for detection in RNA-seq data. **b**, Targeted orthogonal validation (Access Array System, Fluidigm) was performed for 71 mutations (sSNVs and sIndels) in putative CLL genes, affecting 47 CLLs from the CLL8 cohort (selected on the basis of sample availability). With a mean depth of coverage of 7,472×, 65 of the 71 mutations (91.55%) validated, with a higher variant allele fraction compared with normal sample DNA (binomial  $P < 0.01$ ).

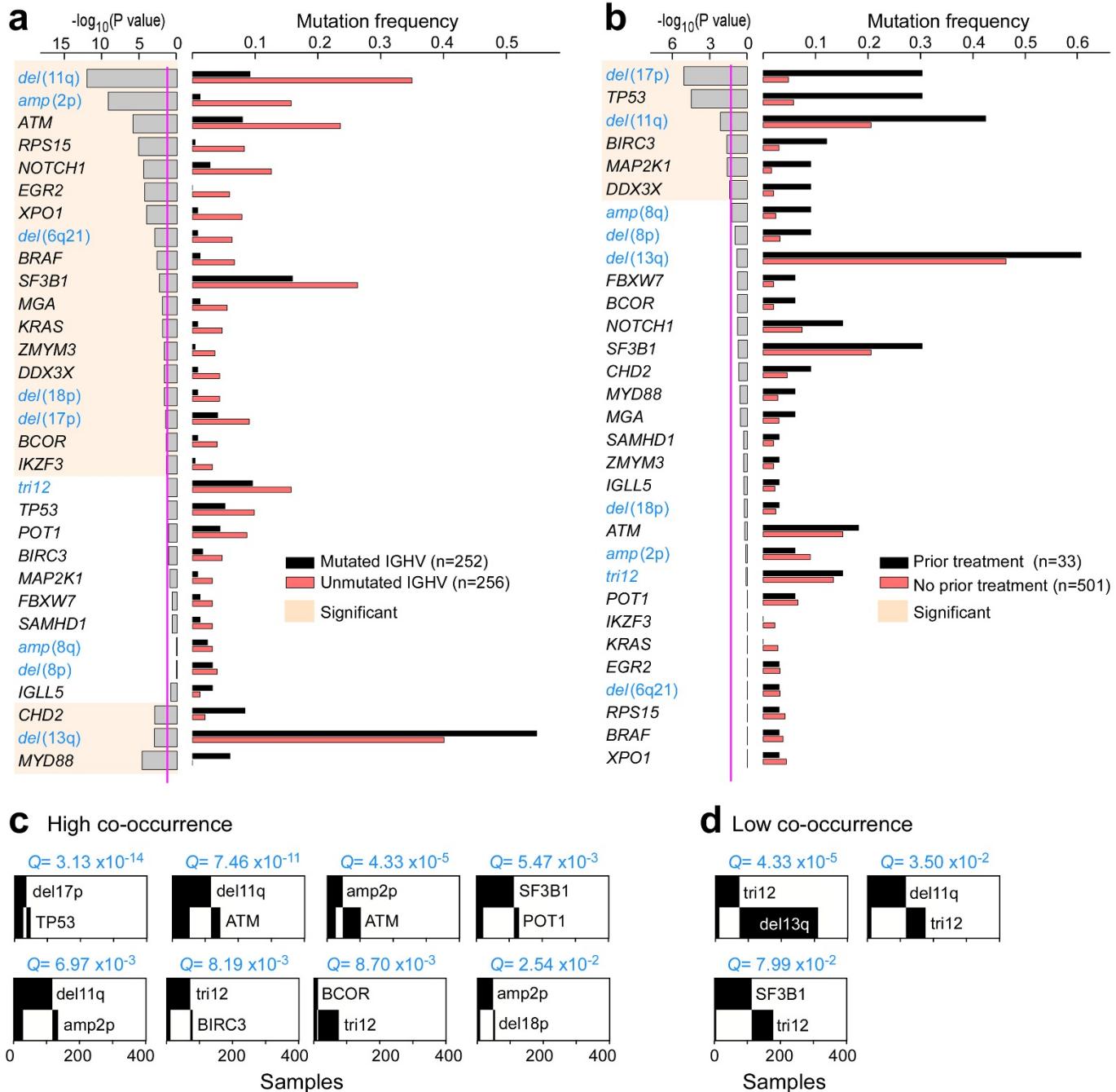


**Extended Data Figure 4 | Gene mutation maps for candidate CLL genes.** a–v, Individual gene mutation maps are shown for all newly identified candidate CLL cancer genes not included in Fig. 2. The plots show mutation subtype (for example, missense, nonsense) and position along the gene.

Red = copy number gain  
Blue = copy number loss  
Green = cnLOH

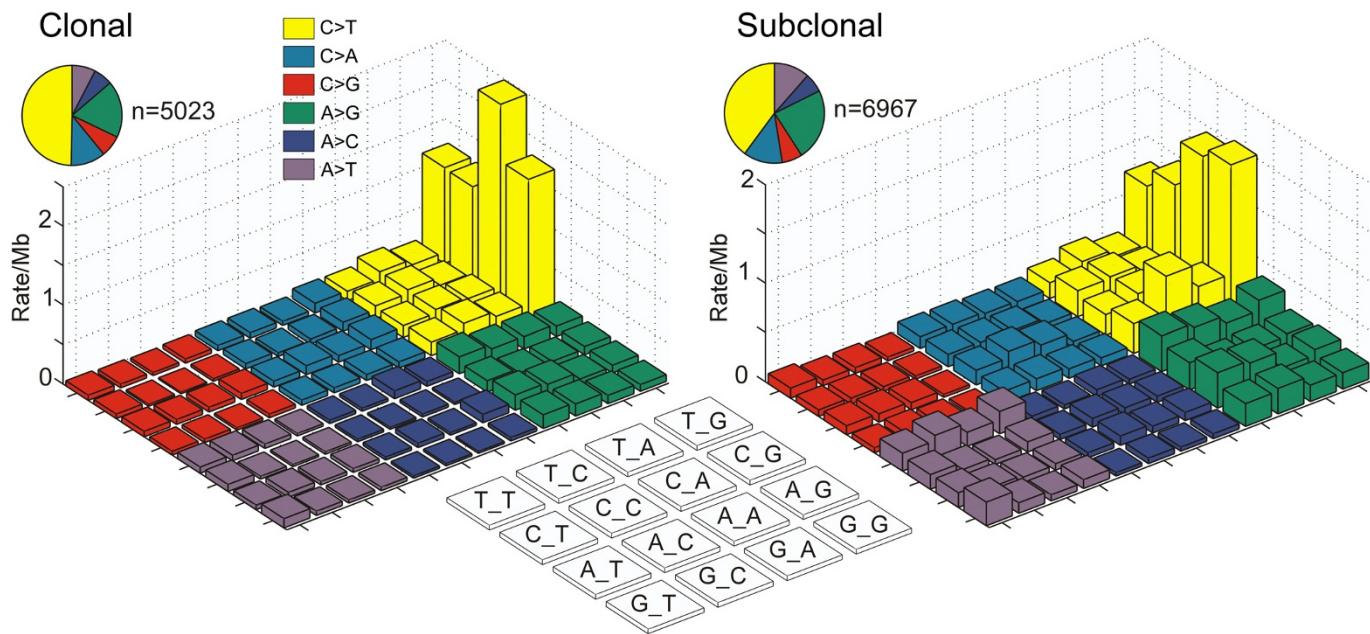


**Extended Data Figure 5 | CLL copy number profiles.** Copy number profile across 538 CLLs detected from WES data from primary samples (see Supplementary Methods).



**Extended Data Figure 6 | Annotation of drivers based on clinical characteristics and co-occurrence patterns.** **a**, Putative drivers affecting greater than 10 patients were assessed for enrichment in *IGHV* mutated versus unmutated CLL subtype (Fisher's exact test, magenta line denotes  $P = 0.05$ ). **b**, Putative drivers affecting greater than 10 patients were assessed for enrichment in samples that received therapy before sampling (Fisher's exact

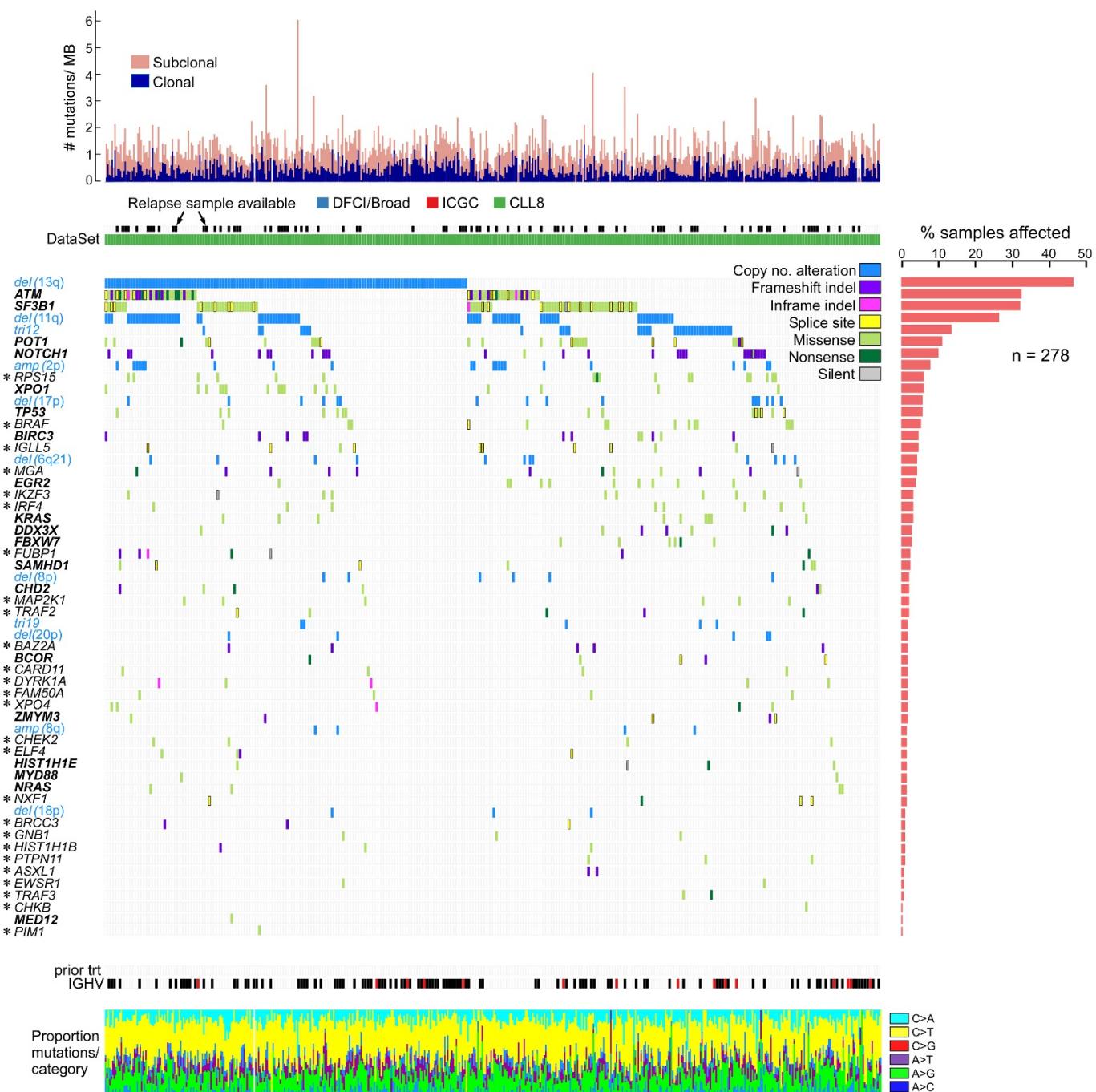
test). Putative drivers affecting greater than 10 patients were tested for co-occurrence. **c, d**, Significantly high (**c**) or low (**d**) co-occurrences are shown ( $Q < 0.1$ , Fisher's exact test with Benjamini Hochberg, false discovery rate, after accounting for prior therapy and *IGHV* mutation status, see Supplementary Methods).



**Extended Data Figure 7 | Mutation spectrum analysis, clonal versus subclonal sSNVs.** The spectrum of mutation is shown for the clonal and subclonal subsets of coding somatic sSNVs across WES of 538 samples. The rate is calculated by dividing the number of trinucleotides with the specified sSNVs by the covered territory containing the specified trinucleotide. Both clonal and subclonal sSNVs were similarly dominated by C > T

transitions at CpG sites. Thus, this mutational process that was previously associated with ageing<sup>39</sup> not only predates oncogenic transformation (since clonal mutations will be highly enriched in mutations that precede the malignant transformation<sup>40</sup>) but also is the dominant mechanism of malignant diversification after transformation in CLL.

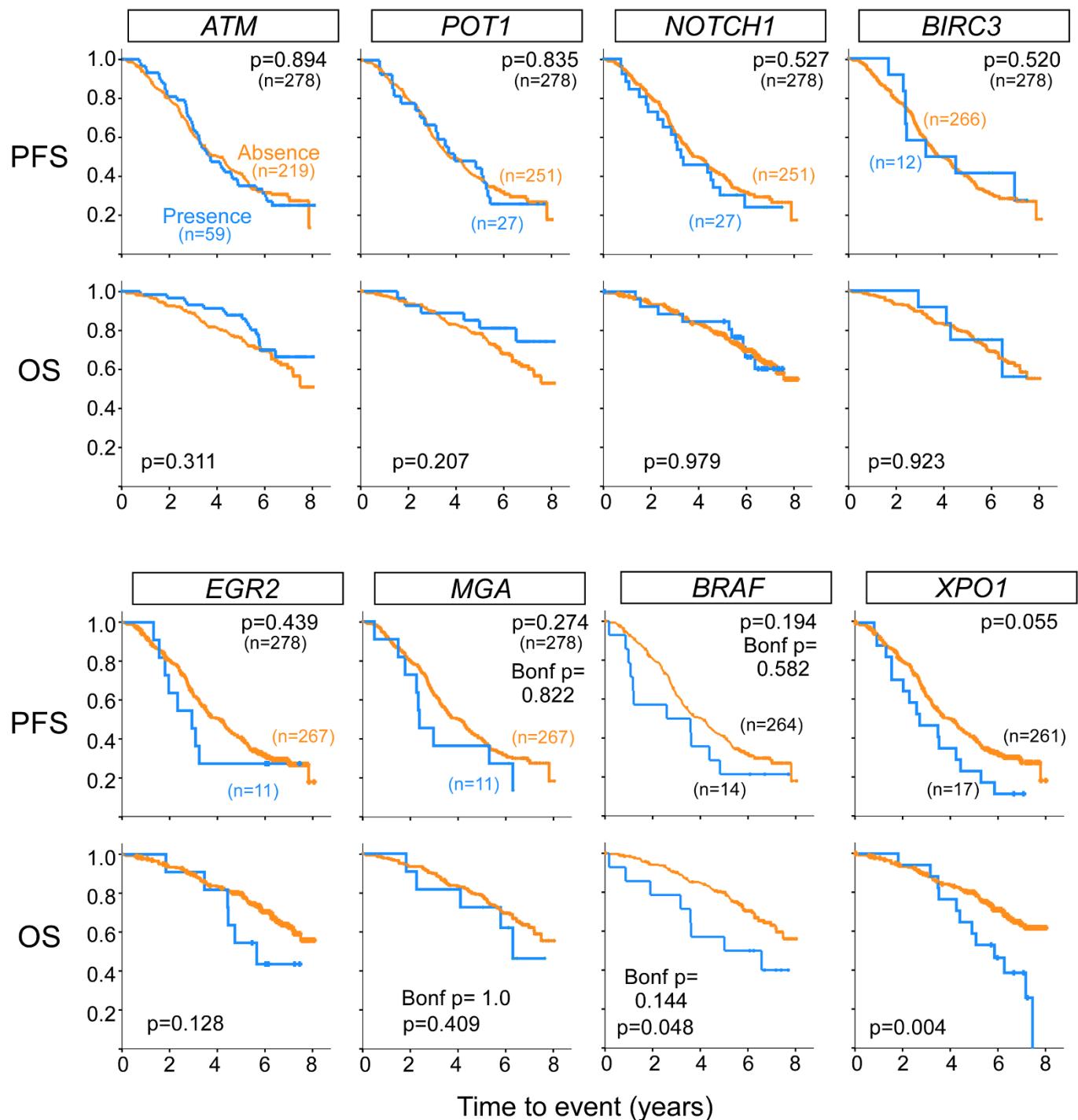
39. Alexandrov, L.B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).  
 40. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).



#### Extended Data Figure 8 | The CLL driver landscape in the CLL8 cohort.

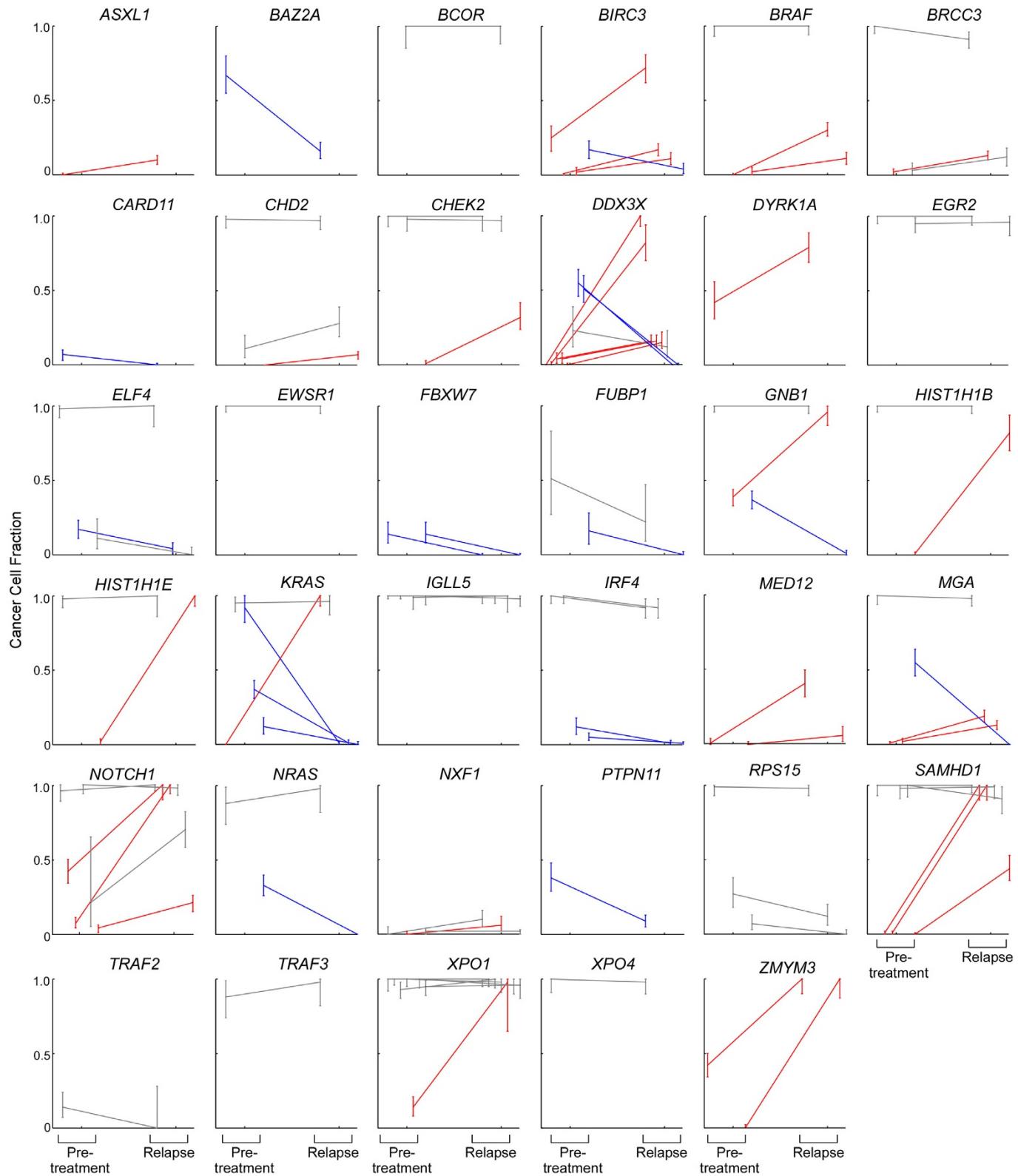
Somatic mutation information shown across the 55 candidate CLL cancer genes and recurrent somatic CNAs (rows) for 278 CLL samples collected from patients enrolled on the CLL8 clinical trial primary that underwent WES

(columns). Recurrent somatic CNA labels are listed in blue, candidate CLL cancer genes are listed in bold if previously identified in Landau *et al.*<sup>3</sup>, and with an asterisk if newly identified in the current study.



**Extended Data Figure 9 | CLL8 patient cohort clinical outcome (from 278 patients) information by CLL cancer gene.** Kaplan–Meier analysis (with logrank  $P$  values) for putative drivers not associated with significant impact on progression-free survival (PFS) or overall survival (OS) in the cohort of 278

patients that were treated as part of the CLL8 trial. For candidate CLL genes tested here for the first time regarding impact on outcome, a Bonferroni  $P$  value is also shown.



**Extended Data Figure 10 | Comparison of pre-treatment and relapse cancer cell fraction (CCF) for non-silent mutations in candidate CLL genes across 59 CLLs.** For each CLL gene mutated across the 59 CLLs that were sampled longitudinally, the modal CCF is compared between the pre-treatment

and relapse samples. CCF increases (red), decreases (blue) or stable CCF (grey) over time are shown (in addition to CLL genes shown in Fig. 5). A significant change in CCF over time (red or blue) was determined if the 95% CI of the CCF in the pre-treatment and relapse samples did not overlap.