

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.

## **RESEARCH ARTICLE**



**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).

## ARTICLE RESEARCH



A ELF4

B EWSR1

C FAM50A

D FUBP1

E GNB1

F HIST1H1B

G HIST1H1E

H PTPN11

TRAF2

**J MGA** 

K MAP2K1

175

]\_

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.



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