

# Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia

Xose S. Puente<sup>1</sup>, Magda Pinyol<sup>2</sup>, Víctor Quesada<sup>1</sup>, Laura Conde<sup>3</sup>, Gonzalo R. Ordóñez<sup>1</sup>, Neus Villamor<sup>3</sup>, Georgia Escaramis<sup>4</sup>, Pedro Jares<sup>3</sup>, Silvia Beà<sup>3</sup>, Marcos González-Díaz<sup>5</sup>, Laia Bassaganyas<sup>4</sup>, Tycho Baumann<sup>6</sup>, Manel Juan<sup>7</sup>, Mónica López-Guerra<sup>3</sup>, Dolors Colomer<sup>3</sup>, José M. C. Tubío<sup>4,8</sup>, Cristina López<sup>3</sup>, Alba Navarro<sup>3</sup>, Cristian Tornador<sup>4</sup>, Marta Aymerich<sup>3</sup>, María Rozman<sup>3</sup>, Jesús M. Hernández<sup>5</sup>, Diana A. Puente<sup>1</sup>, José M. P. Freije<sup>1</sup>, Gloria Velasco<sup>1</sup>, Ana Gutiérrez-Fernández<sup>1</sup>, Dolors Costa<sup>3</sup>, Anna Carrió<sup>3</sup>, Sara Guijarro<sup>3</sup>, Anna Enjuanes<sup>3</sup>, Lluís Hernández<sup>3</sup>, Jordi Yagüe<sup>7</sup>, Pilar Nicolás<sup>9</sup>, Carlos M. Romeo-Casabona<sup>9</sup>, Heinz Himmelbauer<sup>10</sup>, Ester Castillo<sup>10</sup>, Juliane C. Dohm<sup>10</sup>, Silvia de Sanjosé<sup>11</sup>, Miguel A. Piris<sup>12</sup>, Enrique de Alava<sup>5</sup>, Jesús San Miguel<sup>5</sup>, Romina Royo<sup>13</sup>, Josep L. Gelpl<sup>13</sup>, David Torrents<sup>13</sup>, Modesto Orozco<sup>13</sup>, David G. Pisano<sup>14</sup>, Alfonso Valencia<sup>14</sup>, Roderic Guigó<sup>15</sup>, Mónica Bayés<sup>16</sup>, Simon Heath<sup>16</sup>, Marta Gut<sup>16</sup>, Peter Klatt<sup>17</sup>, John Marshall<sup>18</sup>, Keiran Raine<sup>18</sup>, Lucy A. Stebbings<sup>18</sup>, P. Andrew Futreal<sup>18</sup>, Michael R. Stratton<sup>18</sup>, Peter J. Campbell<sup>18</sup>, Ivo Gut<sup>16</sup>, Armando López-Guillermo<sup>6</sup>, Xavier Estivill<sup>4</sup>, Emili Montserrat<sup>6</sup>, Carlos López-Otín<sup>1\*</sup> & Elías Campo<sup>3\*</sup>

**Chronic lymphocytic leukaemia (CLL), the most frequent leukaemia in adults in Western countries, is a heterogeneous disease with variable clinical presentation and evolution<sup>1,2</sup>. Two major molecular subtypes can be distinguished, characterized respectively by a high or low number of somatic hypermutations in the variable region of immunoglobulin genes<sup>3,4</sup>. The molecular changes leading to the pathogenesis of the disease are still poorly understood. Here we performed whole-genome sequencing of four cases of CLL and identified 46 somatic mutations that potentially affect gene function. Further analysis of these mutations in 363 patients with CLL identified four genes that are recurrently mutated: notch 1 (*NOTCH1*), exportin 1 (*XPO1*), myeloid differentiation primary response gene 88 (*MYD88*) and kelch-like 6 (*KLHL6*). Mutations in *MYD88* and *KLHL6* are predominant in cases of CLL with mutated immunoglobulin genes, whereas *NOTCH1* and *XPO1* mutations are mainly detected in patients with unmutated immunoglobulins. The patterns of somatic mutation, supported by functional and clinical analyses, strongly indicate that the recurrent *NOTCH1*, *MYD88* and *XPO1* mutations are oncogenic changes that contribute to the clinical evolution of the disease. To our knowledge, this is the first comprehensive analysis of CLL combining whole-genome sequencing with clinical characteristics and clinical outcomes. It highlights the usefulness of this approach for the identification of clinically relevant mutations in cancer.**

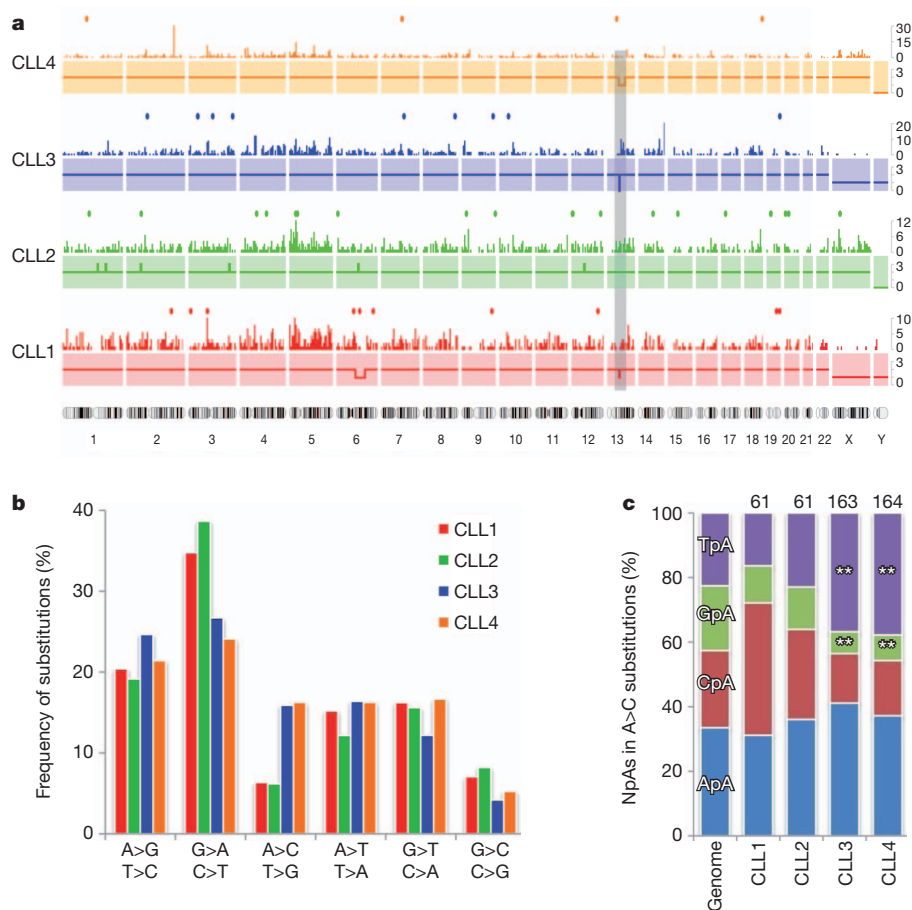
To gain insights into the molecular alterations that cause CLL, we performed whole-genome sequencing of four cases representative of different forms of the disease: two cases, CLL1 and CLL2, with no mutations in the immunoglobulin genes (*IGHV*-unmutated) and two cases, CLL3 and CLL4, with mutations in these genes (*IGHV*-mutated) (Supplementary Table 1 and Supplementary Information). We used a combination of whole-genome sequencing and exome sequencing, as well as long-insert paired-end libraries, to detect variants in chromosomal structure (Supplementary Fig. 1 and Supplementary Tables 2–5).

We obtained more than 99.7% concordance between whole-genome sequencing calls and genotyping data, indicating that the coverage and parameters used were sufficient to detect most of the sequence variants in these samples (Supplementary Information). We detected about 1,000 somatic mutations per tumour in non-repetitive regions (Fig. 1a, Supplementary Fig. 2 and Supplementary Table 6). These numbers of somatic mutations were lower than the numbers in melanoma and lung carcinoma<sup>5,6</sup>, but in agreement with previous estimates of less than one mutation per megabase (Mb) for leukaemias<sup>7</sup>. The most common substitution was the transition G>A/C>T, usually occurring in a CpG context (Fig. 1b and Supplementary Fig. 2). We also detected marked differences in the mutation pattern between CLL samples and these differences were associated with tumour subtype (Fig. 1b). Thus, *IGHV*-mutated cases showed a higher proportion of A>C/T>G mutations than cases with unmutated *IGHV* ( $16 \pm 0.2\%$  versus  $6.2 \pm 0.1\%$ ). The base preceding the adenine in A to C transversions showed an over-representation of thymine, when compared to the prevalence expected from its representation in non-repetitive sequences in the wild-type genome ( $P < 0.001$ , Fig. 1c), and there were fewer A to C substitutions at GpA dinucleotides than would be expected by chance ( $P < 0.001$ ). These differences between CLL subtypes might reflect the molecular mechanisms implicated in their respective development. The pattern and context of mutations are consistent with their being introduced by the error-prone polymerase  $\eta$  during somatic hypermutation in immunoglobulin genes<sup>8</sup>. This indicates that polymerase  $\eta$  could contribute to the high frequency of A > T to C > G transversions in cases with *IGHV*-mutated. It also extends the differences observed between these two CLL subtypes to the genomic level.

We classified the somatic mutations into three different classes according to their potential functional effect (Supplementary Information). We also searched for small insertions and deletions (indels) in coding regions: we found and validated five somatic indels, which caused frameshifts in protein-coding regions (Supplementary Table 7).

<sup>1</sup>Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain. <sup>2</sup>Unidad de Genómica, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain. <sup>3</sup>Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, IDIBAPS, 08036 Barcelona, Spain. <sup>4</sup>Genes and Disease Programme, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), 08003 Barcelona, Spain. <sup>5</sup>Servicio de Hematología, Hospital Universitario, Centro de Investigación del Cáncer-IBMC (USAL-CSIC), Universidad de Salamanca, 37007 Salamanca, Spain. <sup>6</sup>Servicio de Hematología, Hospital Clínic, IDIBAPS, Universidad de Barcelona, 08036 Barcelona, Spain. <sup>7</sup>Servicio de Inmunología, Hospital Clínic, IDIBAPS, 08036 Barcelona, Spain. <sup>8</sup>Hospital Clínico Universitario de Santiago de Compostela, 15706 Santiago de Compostela, Spain. <sup>9</sup>Cátedra Inter-Universitaria de Derecho y Genoma Humano, Universidad de Deusto, Universidad del País Vasco, 48007 Bilbao, Spain. <sup>10</sup>Ultrassequencing Unit, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), 08003 Barcelona, Spain. <sup>11</sup>Unidad de Infecciones y Cáncer, Institut Català de Oncologia-IDIBELL, 08907 Hospitalet de Llobregat, Spain. <sup>12</sup>Molecular Pathology Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. <sup>13</sup>Programa Conjunto de Biología Computacional, Barcelona Supercomputing Center (BSC), Institut de Recerca Biomèdica (IRB), Spanish National Bioinformatics Institute, Universitat de Barcelona, 08028 Barcelona, Spain. <sup>14</sup>Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Spanish National Bioinformatics Institute, 28029 Madrid, Spain. <sup>15</sup>Genomic Bioinformatics Programme, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), Spanish National Bioinformatics Institute, 08003 Barcelona, Spain. <sup>16</sup>Centro Nacional de Análisis Genómico, Parc Científic de Barcelona, 08028 Barcelona, Spain. <sup>17</sup>Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain. <sup>18</sup>Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK.

\*These authors contributed equally to this work.



**Figure 1 | Profile of somatic mutations in four CLL genomes.** **a**, Distribution of somatic alterations. For each tumour genome, copy number (solid lines), density of mutations per 5-Mb window (bars) and protein-coding mutations (dots) are shown. The shaded rectangle indicates the location of the 13q14 deletion that was present in three of the four CLL cases. Chromosome numbers

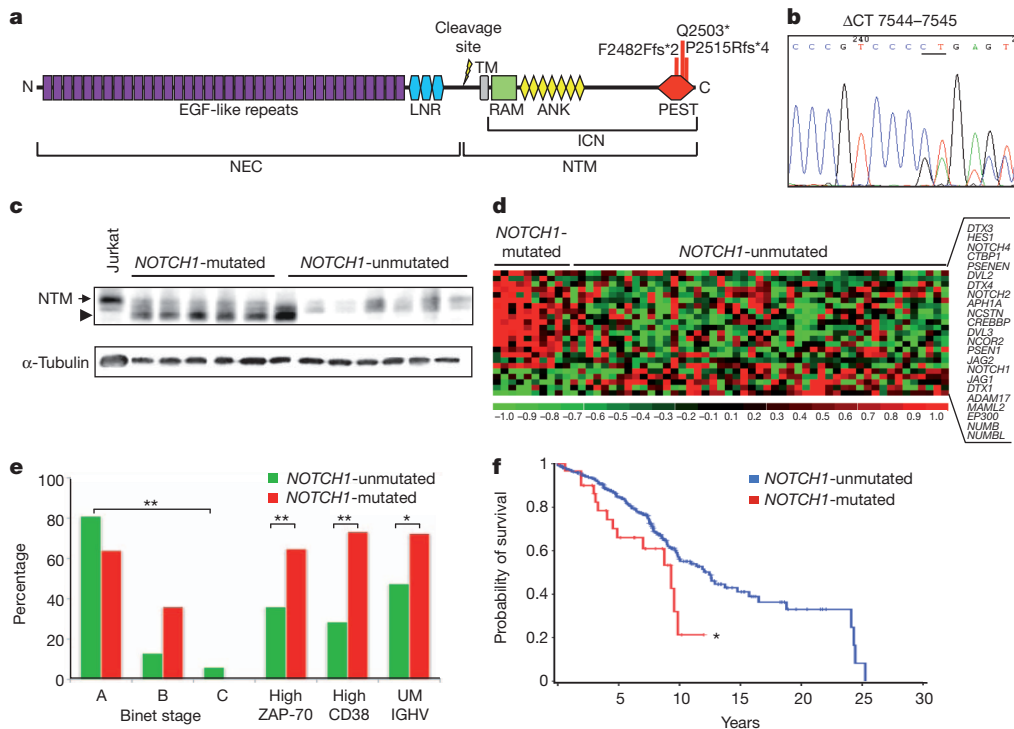
are listed below the four profiles. **b**, Frequency of substitutions in each CLL tumour for the six possible classes of mutation. **c**, Distribution of the four possible NpA dinucleotides for the A to C transversion in each tumour genome, compared with the expected distribution across the genome. The total number of A to C substitutions per case is indicated at the top (\*\*,  $P < 0.001$ ).

We identified 46 mutations that changed the protein-coding sequences of 45 genes in the four patients analysed (Supplementary Table 7). None of these nucleotide substitutions had been previously linked to CLL and among the five indel mutations, only one, in *NOTCH1* (p.P2515Rfs\*4), had been previously found in various lymphoid malignancies, including CLL<sup>9,10</sup>. To determine whether any of these 45 genes was mutated in more than one CLL case, we analysed an initial validation set of 169 CLL patients. We focused on the 26 genes that are expressed at the RNA level in CLL cells (Supplementary Table 7) because mutations in expressed genes are more likely to have a biological effect than those in non-expressed genes. We used a pooled-sequencing strategy that led us to identify four genes with at least one additional mutation in the validation series: these were *NOTCH1*, *MYD88*, *XPO1* and *KLHL6* (Table 1 and Supplementary Information).

Analysis of additional CLL cases revealed that the deletion of a CT dinucleotide in *NOTCH1* (p.P2515Rfs\*4) was found in 29 of 255 patients and two additional mutations in the same region were also found (p.Q2503\* and p.F2482Ffs\*2) (Fig. 2a, b). Accordingly, *NOTCH1* is mutated in 12% of CLL patients (Supplementary Table 8). These mutations generate a premature stop codon, resulting in a *NOTCH1* protein lacking the C-terminal domain, which contains a PEST sequence (a sequence rich in proline, glutamic acid, serine and threonine) (Fig. 2a). Removal of this region results in the accumulation of an active protein isoform in the mutated CLL cells (Fig. 2c and Supplementary Fig. 3). *NOTCH1* is constitutively expressed in CLL<sup>11</sup>, but the *NOTCH1* mutations identified herein generate a more stable and active isoform of the protein. Gene expression analysis of ten *NOTCH1*-mutated and 49 unmutated CLL cases revealed a high

**Table 1 | Genes recurrently mutated in chronic lymphocytic leukaemia**

Gene	Protein	Mutation	Mutated cases / total	Overall frequency (%)	Frequency in <i>IGHV</i> -unmutated (%)	Frequency in <i>IGHV</i> -mutated (%)
<i>NOTCH1</i>	Notch 1	P2515Rfs*4	29/255	12.2	20.4	7
		Q2503*	1/255			
		F2482Ffs*2	1/255			
<i>MYD88</i>	Myeloid differentiation primary response gene 88	L265P	9/310	2.9	0.8	5.6
<i>XPO1</i>	Exportin 1	E571K	3/165	2.4	4.6	0
		E571G	1/165			
<i>KLHL6</i>	Kelch-like 6	F49L/L65P	3/160	1.8	0	4.5
		L90F				
		L58P/T64A/Q81P				



**Figure 2 | Mutational and functional analysis of NOTCH1 in CLL.**

**a**, Schematic representation of human NOTCH1, showing the main domains and locations of the three different somatic mutations identified in CLL. NEC, NOTCH1 extracellular subunit; NTM, NOTCH1 transmembrane subunit; ICN, intracellular domain of NOTCH1; LNR, Lin-12 NOTCH repeats; RAM, RAM domain; ANK, ankyrin repeat domain; PEST, PEST domain. **b**, Electropherogram showing the heterozygous CT deletion recurrently identified in CLL. **c**, Western blot showing NOTCH1 protein levels in CLL cases with or without the NOTCH1 p.P2515Rfs\*4 mutation, and in Jurkat cells

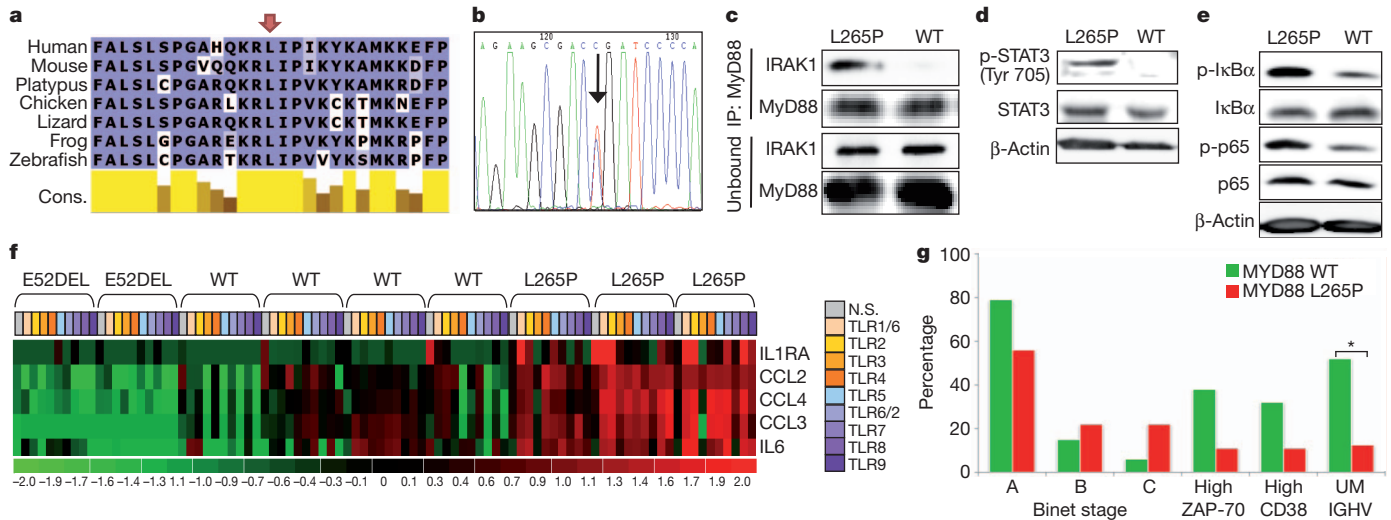
as a control. The arrow indicates the band corresponding to the NTM; the large arrowhead indicates the smaller band corresponding to the mutant form. **d**, Heat map showing the 23 genes of the NOTCH1 pathway that are differentially expressed in NOTCH1-mutated versus non-mutated CLL. **e**, Distribution of disease stage (Binet), ZAP-70 expression status, CD38 expression status and IGHV mutational status (UM, unmutated IGHV) in patients with or without mutations in NOTCH1 (\*,  $P < 0.02$ ; \*\*,  $P < 0.01$ ). **f**, Actuarial probability of overall survival of CLL patients with mutated or unmutated NOTCH1 (\*,  $P = 0.03$ ).

number of differentially expressed genes ( $n = 542$ , false discovery rate  $< 0.05$ ; Supplementary Table 9). Likewise, in a gene-set analysis, we found that there was significant differential expression of the NOTCH1 signalling pathway<sup>12</sup> and two metabolic pathways (oxidative phosphorylation and glycolysis/gluconeogenesis). This is consistent with the NOTCH1-mediated activation of multiple biosynthetic routes in T acute lymphoblastic leukaemia<sup>13</sup>. When the differential expression of individual genes from the NOTCH1 pathway was analysed, 23 of the 46 genes assigned to this pathway<sup>12</sup> showed a significant differential expression ( $P < 0.05$ ) in NOTCH1-mutated CLL (Fig. 2d). NOTCH1-mutated patients had a more advanced clinical stage at diagnosis, more adverse biological features and an overall survival that was significantly shorter than those with NOTCH1 unmutated (10-yr overall survival: 21% versus 56%,  $P = 0.03$ ; Fig. 2e, f). NOTCH1-mutated CLL also underwent transformation into diffuse large B-cell lymphoma more frequently than NOTCH1-unmutated CLL (7 of 31 cases, 23%, versus 3 of 224 cases, 1.3%;  $P < 0.001$ ). The same IGHV clonal rearrangement and NOTCH1 mutation were found in the CLL and corresponding transformed diffuse large B-cell lymphoma of the four cases studied, indicating a clonal relationship of both components.

A recurrent mutation (p.L265P) in the MYD88 gene (Fig. 3a, b) was also identified in 9 of 310 CLL patients (2.9%). During revision of this manuscript, the same mutation has been identified in different lymphomas<sup>14</sup>, highlighting its relevance in the pathogenesis of lymphoid neoplasias. This protein participates in the signalling pathways of interleukin-1 and Toll-like receptors during the immune response<sup>15</sup>. MyD88 immunoprecipitation from CLL cells with the p.L265P mutation resulted in the co-immunoprecipitation of large amounts of IRAK1, in contrast to cells lacking this mutation (Fig. 3c). Other effectors of this signalling pathway, including STAT3, IκBα and NF-κB p65

subunit, showed higher phosphorylation in MYD88-mutated than in unmutated CLL cells (Fig. 3d, e) and there was an increased DNA-binding activity of NF-κB in MYD88-mutated cells (Supplementary Fig. 4). These data support the hypothesis that the MYD88 p.L265P mutation constitutes an activating mutation of this novel proto-oncogene<sup>14,16</sup>. Stimulation of interleukin-1 receptor or Toll-like receptors in MYD88-mutated CLL cells induced the secretion of 5-fold to 150-fold higher levels of interleukin 1 receptor antagonist (IL1RN, also known as IL1RA), interleukin 6 and chemokine (C-C motif) ligands 2, 3 and 4 (CCL2, CCL3 and CCL4), when compared to the secretion of these cytokines by MYD88-unmutated CLLs. Cytokine secretion was elevated in MYD88-mutated cells in response to stimulation of at least four of the eight TLRs tested. No response was observed in lymphocytes carrying the inactivating MYD88 mutation E52DEL (Fig. 3f and Supplementary Fig. 5). The high production of these cytokines has been implicated in the recruitment of macrophages and T lymphocytes by CLL cells, creating a favourable niche for their survival<sup>17</sup>. Moreover, activation of Toll-like receptors in CLL cells promotes the proliferation of tumour cells and protects them from spontaneous apoptosis<sup>18</sup>. Patients with MYD88-mutated CLL were diagnosed at a younger age than those with wild-type MYD88 (median 43 yr, range 38–63, versus median 63 yr, range 27–94;  $P < 0.001$ ) and the disease presented with a more advanced clinical stage (Fig. 3g), although no differences were observed in progression or survival rates. Notably, almost all patients with the MYD88 p.L265P mutation (seven of the eight evaluated) belonged to the IGHV-mutated group.

We also identified four cases with mutations in the same codon of the exportin 1 gene (XPO1; p.E571K and p.E571G). Exportin 1 is implicated in the nuclear export of proteins and mRNAs in yeast, including members of the MAP kinase pathway<sup>19</sup>. The fact that the



**Figure 3 | Mutational and functional analysis of *MYD88* in CLL.** **a**, Multiple sequence alignment of *MyD88* around the mutated residue (arrow) in different species. Cons., degree of conservation. **b**, Electropherogram showing the recurrent heterozygous p.L265P *MYD88* mutation (arrow) detected in CLL. **c**, Cell extracts from a *MYD88*-mutated CLL (L265P) and a *MYD88*-unmutated CLL (WT) were immunoprecipitated with anti-*MyD88* antibody. The immunoprecipitated and unbound fractions were analysed by western blot using anti-IRAK1 and anti-*MyD88* antibodies. **d**, Western blot analysis of phosphorylated STAT3 (p-STAT3 (Tyr 705)) and total STAT3 in cell extracts from *MYD88*-mutated or unmutated CLL tumour cells.  $\beta$ -Actin was used as a control to show equal loading. **e**, Western blots showing phosphorylated I $\kappa$ B $\alpha$  (p-I $\kappa$ B $\alpha$ ), total I $\kappa$ B $\alpha$ , phosphorylated p65 subunit of NF- $\kappa$ B (p-p65) and total p65 subunit of NF- $\kappa$ B in cell extracts from *MYD88*-mutated or unmutated CLL

same residue is mutated in four CLL cases and is part of a highly conserved region (Supplementary Fig. 6) indicates that the mutation affects XPO1 activity. Notably, all four cases with mutations in *XPO1* belonged to the *IGHV*-unmutated subtype and two of them also had the p.P2515Rfs\*4 mutation in *NOTCH1*, indicating that both mutations could have synergic effects in CLL development.

We identified three patients carrying a total of six mutations (F49L/L65P, L90F and L58P/T64A/Q81P) in the gene encoding kelch-like protein 6 (*KLHL6*), which is implicated in the formation of the germinal centre during B cell maturation<sup>20</sup>. All six mutations were clustered between residues 49 and 90 (Supplementary Fig. 7). The presence of several point mutations in *cis*, located near the transcriptional start site of a gene that is highly expressed in the germinal centre, is a characteristic feature of somatic hypermutation. In fact, all three patients had CLL with mutated *IGHV*. Although somatic hypermutation occurs mainly in *IGHV* regions, other proto-oncogenes, including *BCL6*, *MYC* and *PIM1*, are mutated by somatic hypermutation in different lymphomas<sup>21</sup>. However, only *BCL6* has been previously shown to be hypermutated by this mechanism in CLL<sup>21</sup>. Our data show that *KLHL6* is probably also a target of somatic hypermutation in *IGHV*-mutated patients, although its precise contribution to the oncogenic process in CLL remains to be determined.

In addition to these four genes, we identified a series of large genomic alterations that were previously reported<sup>2</sup>. They included the deletion, in three cases, of the 13q14 region<sup>22</sup>, and a 40-Mb deletion in chromosome 6q14–q22 (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 5). Finally, in one patient we detected a p.P281R mutation in the cyclin D2 gene (*CCND2*), which resulted in the accumulation of cyclin D2 in tumour cells (Supplementary Fig. 8). This finding, together with the high conservation of this residue and the identification of mutations in the equivalent residue of cyclin D1 (*CCND1*) in endometrial cancer<sup>23</sup>, indicates that this *CCND2* mutation could be a driver contributing to the development of CLL in this patient. The finding illustrates the putative relevance of non-recurrent mutations for the pathogenesis of CLL.

tumour cells. **f**, Heat map representing the cytokine levels secreted by B cells from eight individuals after Toll-like receptor stimulation. Only the five cytokines that showed the most significant differences between *MYD88*-mutated and *MYD88*-unmutated CLL are shown. 'E52DEL' indicates B cells from two patients with an inactivating *MYD88* mutation, 'WT' corresponds to tumour cells from CLL patients without *MYD88* mutation and 'L265P' indicates tumour cells from patients carrying a mutated *MYD88*. The stimulation experiments for each of the Toll-like receptors (TLRs) are represented in different colours. NS, no stimulus. **g**, Distribution of disease stage (Binet), ZAP-70 expression status, CD38 expression status and *IGHV* mutational status (UM, unmutated *IGHV*) in patients according to the presence or absence of p.L265P *MYD88* mutation (\*,  $P < 0.03$ ).

The International Cancer Genome Consortium project was founded on the concept that sequencing of cancer genomes could reshape our understanding of cancer biology, with direct implications for clinical translation<sup>24</sup>. Our study of four CLL genomes underscores this transformative potential, although additional studies will be necessary to translate these findings to the clinic. We have identified four recurrently mutated genes and provided novel insights into the mechanisms by which leukaemic cells recruit, instruct and coordinate a tumour microenvironment. Currently, the biological identification of different subgroups of CLL is based on markers such as *IGHV* mutational status, cytogenetics, ZAP-70 expression or CD38 expression, which are not fundamental agents in the leukaemic process. The classification of patients based on genomic drivers of the disease is conceptually appealing, as shown by our demonstration that *NOTCH1* and *MYD88* mutations identify distinct subgroups of patients with particular clinical and biological features. Furthermore, we provide functional evidence that both *NOTCH1* and *MYD88* mutations are activating events and potential therapeutic targets. The potential to personalize therapeutic choices for patients on the basis of the genomic architecture of their cancers is the long-term aspiration for studies such as this, combining whole-genome sequencing, functional studies and clinical analysis of patients with cancer.

## METHODS SUMMARY

Four patients with CLL, who had given informed consent for sample collection and analysis, were studied. Tumour samples were obtained before treatment and tumour cells were separated from non-tumour cells by immunomagnetic depletion of T cells, natural killer cells, monocytes and granulocytes (Supplementary Information). Tumour cell purity was  $\geq 98\%$  as assessed by flow cytometry. Normal blood cells from the same patient were obtained after treatment, resulting in no detectable, or less than 0.05%, tumour cell contamination, as assessed by flow cytometry. Additional samples from 363 patients were obtained for clinical validation. Protocols for long-insert and short-insert library construction and for massively parallel paired-end sequencing have been described elsewhere (ref. 25 and Supplementary Information). Genotyping and copy number analysis were performed using the Affymetrix SNP6.0, Agilent 1M and Illumina OmniQuad arrays

on the same cases used for whole-genome sequencing. For the validation of candidate genes in a set of 169 additional CLL patients, we used a combination of PCR amplification and Illumina sequencing in pooled samples, resulting in efficient identification of germline and somatic mutations (Supplementary Information). Sequencing data were aligned to the human reference genome (GRCh37) using Burrows–Wheeler alignment (BWA)<sup>26</sup> and somatic substitutions were identified using Sidrón, a probabilistic binomial model that uses genotyping data to calibrate sequencing error per sample. Functional analyses of the identified mutations were performed using cryopreserved primary tumour cells. For gene expression analysis, RNA was purified from tumour cells and analysed using the HU133 plus 2.0 GeneChip (Affymetrix). For immunoprecipitation and western blotting, CLL cell extracts were prepared and detected using the indicated antibodies (Supplementary Information). For Toll-like receptor stimulation of CLL cells, the Human TLR1–9 agonist kit (InvivoGen) was used.

Received 16 November 2010; accepted 6 April 2011.

Published online 5 June 2011.

- Rozman, C. & Montserrat, E. Chronic lymphocytic leukemia. *N. Engl. J. Med.* **333**, 1052–1057 (1995).
- Zenz, T., Mertens, D., Kuppers, R., Dohner, H. & Stilgenbauer, S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature Rev. Cancer* **10**, 37–50 (2010).
- Damle, R. N. *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94**, 1840–1847 (1999).
- Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848–1854 (1999).
- Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
- Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Spencer, J. & Dunn-Walters, D. K. Hypermutation at A-T base pairs: the A nucleotide replacement spectrum is affected by adjacent nucleotides and there is no reverse complementarity of sequences flanking mutated A and T nucleotides. *J. Immunol.* **175**, 5170–5177 (2005).
- Weng, A. P. *et al.* Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
- Sportoletti, P. *et al.* *NOTCH1* PEST domain mutation is an adverse prognostic factor in B-CLL. *Br. J. Haematol.* **151**, 404–406 (2010).
- Rosati, E. *et al.* Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* **113**, 856–865 (2009).
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2010).
- Palomero, T. *et al.* *NOTCH1* directly regulates *c-MYC* and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103**, 18261–18266 (2006).
- Ngo, V. N. *et al.* Oncogenically active *MYD88* mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
- O'Neill, L. A. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Rev. Immunol.* **7**, 353–364 (2007).
- Coste, I. *et al.* Dual function of MyD88 in RAS signaling and inflammation, leading to mouse and human cell transformation. *J. Clin. Invest.* **120**, 3663–3667 (2010).
- Burger, J. A. *et al.* High-level expression of the T-cell chemokines CCL3 and CCL4 by chronic lymphocytic leukemia B cells in nurse-like cell cocultures and after BCR stimulation. *Blood* **113**, 3050–3058 (2009).
- Muzio, M. *et al.* Expression and function of toll like receptors in chronic lymphocytic leukaemia cells. *Br. J. Haematol.* **144**, 507–516 (2009).
- Ferrigno, P., Posas, F., Koepp, D., Saito, H. & Silver, P. A. Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin  $\beta$  homologs NMD5 and XPO1. *EMBO J.* **17**, 5606–5614 (1998).
- Kroll, J. *et al.* The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal center formation. *Mol. Cell. Biol.* **25**, 8531–8540 (2005).
- Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
- Aqeilan, R. I., Calin, G. A. & Croce, C. M. miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ.* **17**, 215–220 (2010).
- Moreno-Bueno, G. *et al.* Cyclin D1 gene (*CCND1*) mutations in endometrial cancer. *Oncogene* **22**, 6115–6118 (2003).
- Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was funded by the Spanish Ministry of Science and Innovation (MICINN) through the Instituto de Salud Carlos III (ISCIII) and Red Temática de Investigación del Cáncer (RTICC) del ISCIII. C.L.-O. is an Investigator of the Botín Foundation and D.T., of the ICREA program. We thank E. Santos for his support of this project, A. Carracedo and J. Benítez for genotyping studies, C. Fortuny for the supply of samples and N. Villahoz and M. C. Muro for their work in the coordination of the CLL-ICGC Consortium. We are also grateful to all patients with CLL who participated in this study.

**Author Contributions** X.S.P., V.Q., G.R.O., J.M.P.F., G.V., and A.G.-F. developed the bioinformatic algorithms for analysis of sequence data, implemented SMIPS for clinical validation and performed functional studies. M.P., L.C., P.J., M.J., M.L.-G., D.Colomer, S.G., C.L., A.N. and J.Y. were responsible for downstream validation analysis and functional studies. S.B., A.C., G.E., L.B., J.M.C.T., J.C.D. and C.T. studied structural variants. M.G., M.B., S.H., D.A.P., H.H., E.C., J.M., K.R. and L.A.S. were responsible for generating libraries, performing exome capture and running sequencers. P.N., C.M.R.-C. and M.A. prepared and supervised the bioethical requirements. S.S. designed the epidemiologic study. P.K. gave conceptual advice and revised the manuscript. N.V., T.B., D.Costa, A.C., A.E., L.H., M.A.P., E.A., J.S.M., E.M., A.L.-G. and E.C. performed clinical and biological studies. M.R., N.V., J.M.H. and M.G.-D. were the pathologists who reviewed and confirmed the diagnoses. R.R., J.L.G., M.O., D.T., D.G.P., A.V. and R.G. were in charge of bioinformatics data management. I.G., P.A.F., M.R.S. and P.J.C. coordinated the sequencing efforts and performed primary data analysis. X.S.P., X.E., A.L.-G., C.L.-O. and E.C. directed the research and wrote the manuscript, which all authors have approved.

**Author Information** Sequencing, expression and genotyping array data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>), which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00000000092. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to E.C. ([ecampo@clinic.ub.es](mailto:ecampo@clinic.ub.es)) or C.L.-O. ([clo@uniovi.es](mailto:clo@uniovi.es)).