

POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia

Andrew J Ramsay^{1,4}, Víctor Quesada^{1,4}, Miguel Foronda^{2,4}, Laura Conde³, Alejandra Martínez-Trillos³, Neus Villamor³, David Rodríguez¹, Agnieszka Kwarciak¹, Cecilia Garabaya¹, Mercedes Gallardo², Mónica López-Guerra³, Armando López-Guillermo³, Xose S Puente¹, María A Blasco², Elías Campo³ & Carlos López-Otín¹

Chronic lymphocytic leukemia (CLL) is the most frequent leukemia in adults^{1–3}. We have analyzed exome sequencing data from 127 individuals with CLL and Sanger sequencing data from 214 additional affected individuals, identifying recurrent somatic mutations in *POT1* (encoding protection of telomeres 1) in 3.5% of the cases, with the frequency reaching 9% when only individuals without *IGHV@* mutations were considered. *POT1* encodes a component of the shelterin complex and is the first member of this telomeric structure found to be mutated in human cancer. Somatic mutation of *POT1* primarily occurs in gene regions encoding the two oligonucleotide-/oligosaccharide-binding (OB) folds and affects key residues required to bind telomeric DNA. *POT1*-mutated CLL cells have numerous telomeric and chromosomal abnormalities that suggest that *POT1* mutations favor the acquisition of the malignant features of CLL cells. The identification of *POT1* as a new frequently mutated gene in CLL may facilitate novel approaches for the clinical management of this disease.

A long-standing barrier to understanding leukemogenesis has been an inability to accurately identify the genetic alterations that underlie this disease⁴. To gain further insight into the genetic basis of CLL, we have analyzed the exomes of matched tumor and normal samples from the peripheral blood of 127 untreated individuals with CLL, identifying tumoral somatic mutations using the previously described *Sidrón* algorithm^{5,6}. Several of the driver genes, such as *SF3B1*, *NOTCH1* and *MYD88*, have previously been assessed for their functional and clinical roles in CLL^{5–8}. We also observed that *POT1* is a frequent genetic target for somatic mutation in CLL, and, given its prominent role in telomere capping and telomerase recruitment^{9,10}, we hypothesized that mutations in this gene might be a source of genomic instability and represent a new tumor-driving factor. In fact, when the number of mutations, gene size and codon composition were taken into account, *POT1* was the gene that was second most frequently affected by point

mutations ($P = 4.2 \times 10^{-6}$), with only *SF3B1* having a higher frequency ($P = 8.6 \times 10^{-11}$). *POT1* was found to be mutated in approximately 5% of CLL cases (Supplementary Table 1) and, conspicuously, was exclusively altered in samples with wild-type *IGHV@*, suggesting that inactivation of *POT1* could represent a distinct molecular alteration for the development of this aggressive CLL subtype. We next extended our sequence analysis of *POT1* by performing Sanger sequencing in a validation set consisting of 214 matched tumor and normal samples from individuals with CLL (Supplementary Table 2), with the finding of 5 additional somatic mutations. RNA sequencing (RNA-seq) analysis of available cases confirmed the expression of mutant *POT1* alleles in CLL tumor cells (data not shown). Therefore, we have identified 12 somatic point mutations in 341 CLL cases (3.5%; Table 1). This finding also implies that 9% of individuals with wild-type *IGHV@* have a *POT1* mutation, which is a comparable mutational frequency to that of *SF3B1* (11%) in individuals of this CLL subtype⁷. Our findings make *POT1* one of the most frequently mutated genes reported so far in CLL along with *NOTCH1*, *SF3B1*, *TP53* and *ATM*^{5–8}. Notably, 3 of the 12 mutations found in *POT1* are predicted to result in a truncated protein. The remaining somatic mutations were all nonsynonymous and were predominantly predicted to be deleterious by a consensus of five bioinformatics tools¹¹, supporting the idea that *POT1* mutations constitute a driver event in CLL (Table 1).

Shelterin proteins (namely TRF1, TRF2, TIN2, TPP1, RAP1 and POT1) are critical regulators of telomere integrity by modulating telomere capping, replication and telomere extension by telomerase^{12–14}. Structurally, POT1 is the only shelterin that contains two N-terminal OB domains, which confer to this protein affinity for the single-stranded DNA (ssDNA) sequence TTAGGG, in addition to a C terminus that binds TPP1 and anchors POT1 to the shelterin complex. Notably, 9 of the 12 *POT1* somatic mutations we have detected in CLL samples reside within gene regions encoding the OB folds (Fig. 1a). Comparative sequence analysis of the OB fold somatic alterations showed that most mutated residues are phylogenetically conserved and suggests that they have an essential role in

¹Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias (IUOPA) Universidad de Oviedo, Oviedo, Spain.

²Telomeres and Telomerase Group, Molecular Oncology Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. ³Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. ⁴These authors contributed equally to this work. Correspondence should be addressed to C.L.-O. (clo@uniovi.es), E.C. (ecampo@clinic.ub.es) or M.A.B. (mblasco@cnio.es).

Received 29 October 2012; accepted 22 February 2013; published online 17 March 2013; doi:10.1038/ng.2584

Table 1 Somatic mutations affecting *POT1* in individuals with CLL

Case	Mutation	Exon	Amino-acid change	Mutation type	Prediction
044	g.124537268A>T	4–5	–	Splicing	–
184	g.124537227T>A	5	p.Met1Leu	Missense	Deleterious
013	g.124532338A>T	6	p.Tyr36Asn	Missense	Deleterious
157	g.124511022A>C	7	p.Tyr66*	Nonsense	–
099	g.124503682T>C	8	p.Lys90Glu	Missense	Neutral
742	g.124503669T>C	8	p.Gln94Arg	Missense	Deleterious
195	g.124503669T>C	8	p.Gln94Arg	Missense	Deleterious
006	g.124499045T>C	9	p.Tyr223Cys	Missense	Deleterious
647	g.124493146G>T	10	p.Ser250*	Nonsense	–
072	g.124493098T>A	10	p.His266Leu	Missense	Deleterious
085	g.124493080C>A	10	p.Gly272Val	Missense	Deleterious
284	g.124465325A>C	18	p.Cys591Trp	Missense	Deleterious

protein function (Fig. 1b). The availability of a crystal structure for the N-terminal portion of human POT1 (ref. 15) allowed us to next interrogate the potential structural implications of the OB fold alterations. The observed *POT1* mutations in individuals with CLL either prematurely terminate the resulting protein or affect residues in the POT1-telomere interaction interface, thereby likely destabilizing this interaction and diminishing the ability of POT1 to bind telomeric ssDNA (Fig. 1c). Thus, the mutation affecting Tyr223 would abolish the interaction of this aromatic residue with the terminal guanine base of telomeric DNA, whereas mutations affecting Tyr36, Lys90, Gln94 and His266 likely disrupt electrostatic interactions with different regions of the telomeric sequence (Fig. 1c). This finding suggests that the somatic mutations reported in this work destabilize the binding of POT1 to telomeric DNA. A Monte Carlo simulation confirmed that the clustering of point substitutions in residues close to the POT1-DNA interaction interface is not likely to have occurred by chance ($P < 1 \times 10^{-7}$). Further genomic database analysis has identified additional *POT1* mutations in individuals with multiple myeloma and breast, stomach, squamous cell, hepatocellular, lung and ovarian carcinomas as well as in an individual with CLL from an independent genome study on this leukemia^{8,16} (Fig. 1a). Most of the *POT1* mutations that are annotated in databases affect the OB fold of POT1 (19 of 25 total mutations), confirming that this region is the main target of the

mutational events in this protein in different types of human cancer.

As the identified *POT1* mutations are heterozygous, with tumor cells retaining one wild-type allele, we hypothesized that the mutated POT1 proteins function in a dominant-negative manner, analogous to that previously reported for the Δ OB POT1 mutant^{10,17}. To further assess this possibility, we created heterologous HT1080 cell lines using retroviral gene transfer to express exogenous Myc-tagged wild-type and mutant (Tyr36Asn and Tyr223Cys) POT1 proteins in the presence of endogenous POT1. Immunofluorescence microscopy analysis of HT1080 cells showed

that the wild-type and mutant Myc-tagged POT1 proteins had a punctate nuclear staining pattern and colocalized with the telomere-binding protein TRF1 (Fig. 2a), showing that the mutations affecting the OB fold do not interrupt telomeric localization. Telomeric association of the Myc-tagged mutant POT1 proteins was further confirmed by chromatin immunoprecipitation (ChIP) assays that detected telomeric DNA in immunoprecipitations with antibody to Myc in the corresponding HT1080 cells (Fig. 2b,c). In addition, ChIP assays also confirmed that the shelterin components RAP1, TRF1 and TRF2 remained localized to telomeres in the presence of the mutant POT1 proteins (Fig. 2b,c). Having established the telomeric localization of the mutant POT1 proteins, we next sought to determine whether the p.Tyr36Asn and p.Tyr223Cys alterations of the OB domains affected the ability of POT1 to bind to the 3' end of the G-rich telomeric overhang. Constructs containing cDNA for wild-type and mutant POT1 were translated *in vitro*, and equivalent quantities of the resulting POT1 proteins (Supplementary Fig. 1a) were incubated with radiolabeled telomeric ssDNA. Visualization of the protein-DNA complexes by electrophoretic mobility shift assay (EMSA) confirmed that wild-type POT1 efficiently bound telomeric ssDNA, whereas, in stark contrast, POT1 proteins containing the p.Tyr36Asn or p.Tyr223Cys substitution were unable to bind the telomeric sequence (Fig. 2d). Consistent with this finding, a similar abolishment of telomere binding

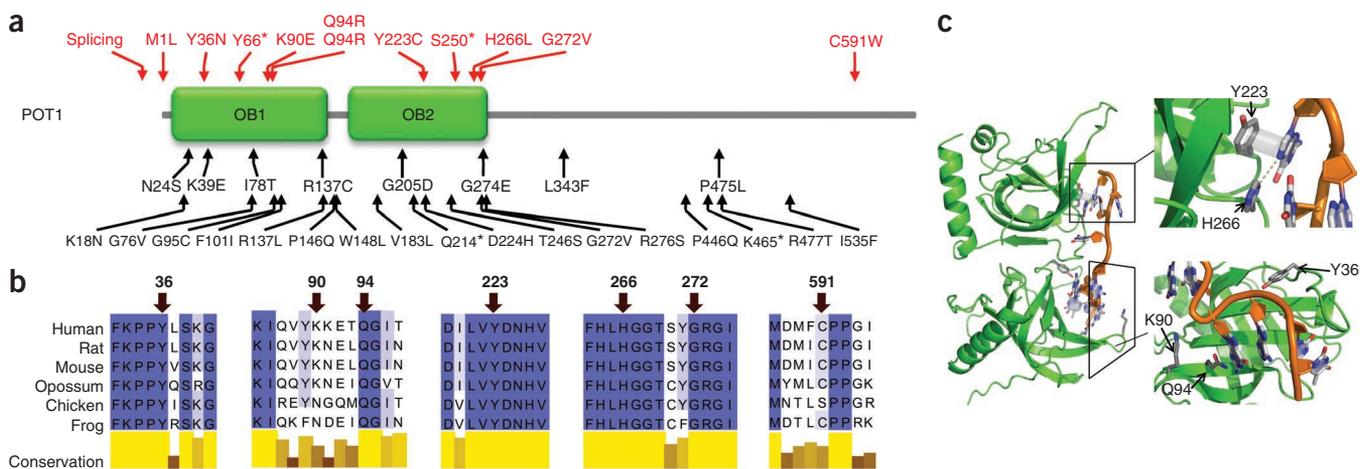
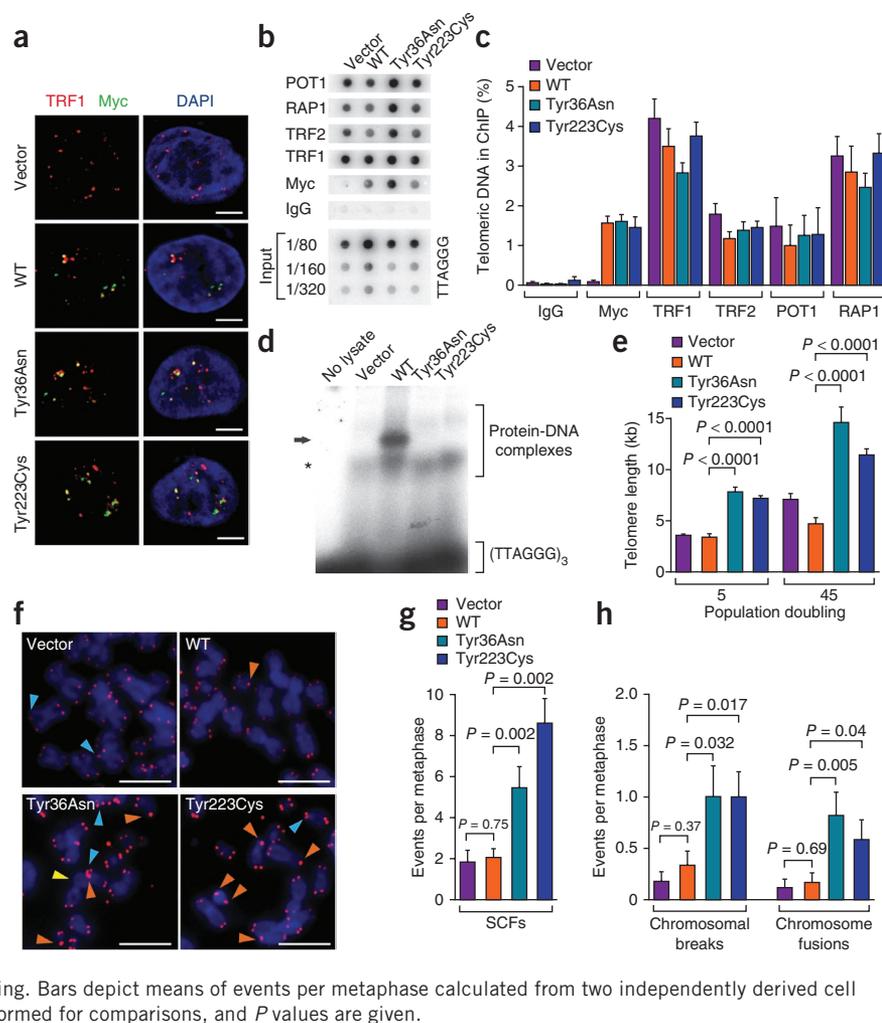


Figure 1 Structural impact of *POT1* somatic mutations. (a) Impact of mutations on the primary structure of the POT1 protein. The alterations identified in this study (red) are depicted on a representation of the protein, with the positions of the OB domains indicated (green boxes). Below, additional somatic alterations in other tumors and CLL studies are shown in black. (b) Phylogenetic features of the mutated residues. Protein sequence alignments around the non-truncating mutated residues (arrows) in evolutionarily diverse species are shown. Completely and partially conserved residues are highlighted in dark and light purple, respectively. (c) Ribbon representation of the POT1 OB domains (green) bound to a telomere-like oligonucleotide (orange). Detailed views of several mutated amino acids identified in individuals with CLL are shown to the right.

Figure 2 Biological properties of wild-type POT1 and the Tyr36Asn and Tyr223Cys POT1 mutants. **(a)** Localization of Myc-tagged wild-type (WT) and mutant POT1 proteins to telomeres in transiently transfected HT1080 cells. Scale bars, 5 μ m. **(b)** Telomeric ChIP assays in HT1080 cells transiently infected with the indicated retroviral constructs using the antibodies to the indicated proteins. Input controls were hybridized with a TTAGGG repeat probe. IgG, immunoglobulin G control. **(c)** Quantification of four ChIP experiments in **b** representing the percentage of telomeric DNA isolated with each antibody. Bars depict mean values \pm s.e.m. **(d)** Tyr36Asn and Tyr223Cys POT1 mutants are unable to bind the G-strand overhang *in vitro*. EMSA of [³²P]-labeled oligonucleotides (TTAGGG)₃ in the presence of the indicated POT1 proteins. **(e)** The p.Tyr36Asn and p.Tyr223Cys alterations in POT1 result in aberrant telomere length homeostasis. Mean telomere lengths were calculated by HTQFISH on metaphase spreads in HT1080 cells infected with the indicated retroviral constructs at population doublings 5 and 45. Bars show mean telomere lengths \pm s.e.m. Two-tailed Student's *t* tests were performed, and *P* values are given. **(f)** Telomeric FISH on the metaphase spreads of HT1080 cells infected with the indicated retroviral constructs at population doubling 45 (red, TTAGGG-Cy3; blue, DAPI). Arrowheads: orange, SCFs; yellow, chromosome-type end-to-end fusions; light blue, MTSS. Scale bars, 5 μ m. **(g,h)** Quantification of the numbers of SCFs (**g**) and chromosomal breaks and fusions (**h**) per metaphase in HT1080 cells infected with the indicated retroviral constructs at population doubling 45 as determined by telomeric FISH staining. Bars depict means of events per metaphase calculated from two independently derived cell lines \pm s.e.m. Two-tailed Student's *t* tests were performed for comparisons, and *P* values are given.



capacity was seen when the p.Phe62Ala substitution was introduced into the OB domain of the POT1 mouse homolog, Pot1a¹⁸. These experiments show that mutant POT1 proteins are able to localize to telomeres with other shelterin components but are unable to bind telomeric G-strand overhangs.

We then sought to determine whether mutant POT1 localized at telomeres causes telomere elongation in a dominant-negative fashion, as previously described for the Δ OB POT1 mutant^{10,17}. Expression of wild-type and mutant POT1 proteins was maintained over the 45 population doublings examined (**Supplementary Fig. 1b**), and, in agreement with previous reports, exogenous expression of wild-type POT1 did not significantly increase telomere length in HT1080 cells, whereas the mutant POT1 proteins significantly increased telomere length in each of the population doubling points examined (**Fig. 2e** and **Supplementary Fig. 1c,d**) without altering telomerase activity (**Supplementary Fig. 1e,f**). Consistent with the mutations behaving in a dominant-negative manner to facilitate partial and/or full loss of POT1 function, ablation of human *POT1* (ref. 19) and of mouse *POT1* orthologs^{18,20} resulted in a dysfunctional telomeric phenotype characterized by elongated telomeres and unprotected telomere ends that lead to chromosomal alterations. Accordingly, we examined the frequency of chromosomal aberrations in HT1080 cells expressing wild-type and mutant POT1 proteins after 5, 15 and 45 population doublings by telomeric FISH analysis of metaphase spreads. At population doubling 45, HT1080 cells expressing mutant POT1 proteins showed a marked

propensity for the fusion of sister chromatids (>5 fusions per metaphase; **Fig. 2f,g**), with the frequency of incidence increasing significantly with progressive population doublings (**Supplementary Fig. 2a**). As shelterin components are gatekeepers guarding against telomere fragility^{12,13}, we also measured the presence of multitelomeric signals (MTSS), a telomere aberration that has been associated with increased replication fork stalling at telomeres. Notably, HT1080 cells expressing mutant POT1 proteins had a tendency to have higher numbers of MTSSs, suggesting that there was elevated telomere fragility due to the actions of the mutant POT1 proteins (**Supplementary Fig. 2b**). We also observed significantly more chromosomal breaks and fusions in cells expressing the mutant POT1 proteins in comparison to cells expressing wild-type POT1 or an empty vector (**Fig. 2h** and **Supplementary Fig. 2c,d**). In contrast, overexpression of wild-type POT1 led to no change in the frequency of each of the chromosomal events examined (**Fig. 2f,g** and **Supplementary Fig. 2a–d**), showing that the chromosomal alterations were due to the specific activities of mutant POT1 in the heterologous cell lines rather than to a generalized increase in POT1 protein amounts.

Unprotected telomeres can lead to chromosomal instability and the acquisition of genomic aberrations, a key event in tumorigenesis²¹. In individuals with CLL, increased genomic complexity has been correlated with decreased survival^{22,23}. We first assessed telomere lengths in purified tumor cells from CLL cases harboring *POT1* mutations in comparison to those in age-matched control CLL

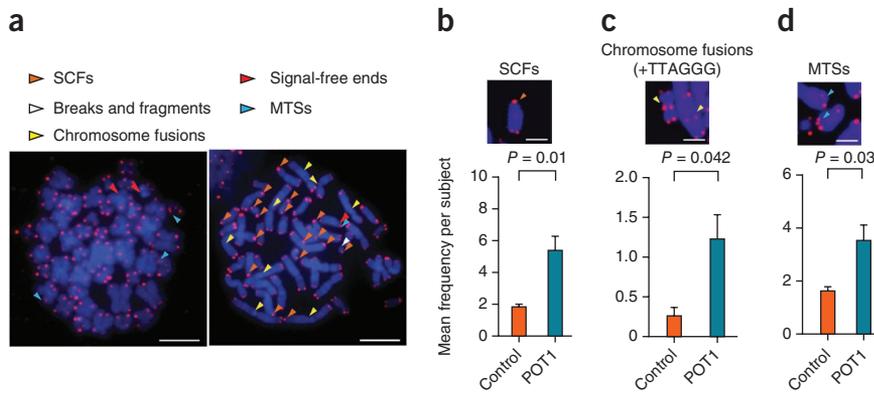


Figure 3 Chromosomal aberrations in cells from individuals with CLL expressing mutated *POT1*. (a) Representative fluorescence microscopy images of telomeric FISH on metaphase spreads from *POT1*-mutated and control (wild-type *POT1*) CLL samples (red, TTAGGG-Cy3; blue, DAPI). Scale bars, 5 μ m. (b–d) Representative images (red, TTAGGG-Cy3; blue, DAPI; scale bars, 2 μ m) (top) and quantitation (bottom) of per-individual chromosomal aberrations are shown for SCFs (b), chromosome-type end-to-end fusions with detectable TTAGGG signal from telomeric FISH (c) and MTSs (d) in CLL cases. Bars depict per-individual means calculated from four controls and six *POT1*-mutated CLL cell lines \pm s.e.m. Two-tailed Student's *t* tests were performed for comparisons, and *P* values are given.

cases with wild-type *IGHV@* containing no mutation in *POT1*, *TP53* or *ATM* (Supplementary Table 3), observing no significant differences in telomere lengths or in telomerase activities (Supplementary Fig. 3a,b). The absence of changes in telomere length due to *POT1* mutations potentially derives from the variance in telomere lengths within human populations, a situation that is distinct from those present in the analysis of isogenic cell lines and inbred mouse strains. Next, we examined the integrity of telomeres in the *POT1*-mutated CLL samples through telomeric FISH analysis of metaphase spreads (Fig. 3a). Notably, similar to observations made in HT1080 cells, we found significantly more sister chromatid-type end-to-end fusions (SCFs) in *POT1*-mutated samples than in control samples, with up to five fusions occurring per metaphase ($P = 0.01$) (Fig. 3a,b). We also detected significantly more chromosome-type fusions containing TTAGGG signals in *POT1*-mutated samples than in control samples ($P = 0.042$) (Fig. 3c), but the number of chromosome-type fusions with no detectable TTAGGG signal was not significantly different between the two populations ($P = 0.072$) (Supplementary Fig. 3c), in agreement with the normal telomere lengths observed in *POT1*-mutated samples. As in HT1080 cells, we detected a modest but significant increase in the presence of MTSs in *POT1*-mutated CLL cells compared to the controls (2.2-fold increase, $P = 0.03$; Fig. 3d), reflecting the presence of greater telomere fragility in these individuals.

Collectively, the observation of a high incidence of telomere-containing chromosomal fusions coincident with no change in the frequency of telomere-deficient fusions is consistent with these fusions being produced by *POT1* mutation-mediated telomere uncapping rather than by overt telomere shortening, resembling the chromosomal phenotype of mouse cells with complete ablation of Pot1 activity^{18,20}. This notion is further supported by the fact that we did not find significantly greater numbers of signal-free ends per metaphase in *POT1*-mutated CLL cells or in HT1080 cells expressing mutant *POT1* proteins (Supplementary Figs. 3d and 2e, respectively). We also observed significant increases in the frequency of nuclear aberrations and the presence of unresolved mitotic structures in *POT1*-mutated samples relative to control samples (Supplementary Fig. 3e–g). Notably, the higher incidence of micronuclei in *POT1*-mutated CLL cells suggests a potential link between *POT1* function and the recently identified phenomenon of chromothripsis^{24,25}. Further,

versus 46%; $P = 0.001$) and the expression of ZAP-70 (75% versus 31%; $P = 0.001$), than individuals without *POT1* somatic mutations (Supplementary Table 4).

The International Cancer Genome Consortium was formed on the idea that next-generation sequencing techniques would reshape and extend understanding of cancer biology to generate novel clinical approaches, including for CLL²⁷. The identification of *POT1* as a new frequently mutated gene in CLL and the first shelterin found to be mutated in human cancer will hopefully be useful for improved patient prognosis and will provide additional knowledge for the development of novel therapeutic approaches to treat this disease.

URLs. European Genome-phenome Archive (EGA), <https://www.ebi.ac.uk/ega/>; Picard, <http://picard.sourceforge.net/index.shtml>; Protein Data Bank (PDB), <http://www.rcsb.org/pdb/>.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession code. Sequencing, expression and genotyping array data have been deposited at the European Genome-phenome Archive (EGA), which is hosted at the European Bioinformatics Institute (EBI), under accession EGAS00000000092.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to D.A. Puente, S. Guijarro, S. Martín, C. Capdevila, M. Sánchez and L. Plá for excellent technical assistance and to N. Villahoz and C. Muro for excellent work in the coordination of the CLL Spanish Consortium. We thank T. de Lange (The Rockefeller University) for providing the *POT1* plasmid. We are also very grateful to all individuals with CLL who have participated in this study. This work was funded by the Spanish Ministry of Economy and Competitiveness through the Instituto de Salud Carlos III (ISCIII) and the Red Temática de Investigación del Cáncer (RTICC) del ISCIII. C.L.-O. is an Investigator of the Botín Foundation. Research in the laboratory of M.A.B. is funded by the Spanish Ministry of Economy and Competitiveness Projects SAF2008-05384 and CSD2007-00017, the Madrid Regional Government Project S2010/BMD-2303 (ReCaRe), the European Union Seventh Framework Programme Project FHEALTH-2010-259749 (EuroBATS), The European Research Council (ERC) Project GA 232854 (TEL STEM CELL), the Körber European Science Award from the Körber Foundation,

POT1-mutated CLL cells and HT1080 cells expressing mutant *POT1* proteins showed no difference in telomeric DNA damage relative to control cells (Supplementary Fig. 4a–e), as has been observed with complete ablation of *POT1* proteins from mammalian cells^{18,19,21}. This discrepancy potentially derives from the presence of wild-type *POT1* in both cell systems, which has been implicated in suppressing the telomeric DNA damage response in cells expressing the Δ OB *POT1* isoform²⁶. Finally, the correlation between genomic instability and poor clinical outcomes, previously observed for individuals with CLL, prompted us to assess the clinical features of individuals with *POT1* mutations. Individuals with somatic mutations in *POT1* presented with more advanced disease at diagnosis (Binet B: 13% in cases with wild-type *POT1* versus 42% in cases with mutated *POT1*; $P = 0.02$) and with more adverse biological features, such as the more frequent presence of wild-type *IGHV@* (100%

the Preclinical Research Award from Fundación Lilly (Spain), Fundación Botín (Spain) and the AXA Research Fund.

AUTHOR CONTRIBUTIONS

V.Q., A.J.R., A.K. and X.S.P. developed the bioinformatics algorithms and performed the analysis of sequence data. A.J.R., M.F., D.R., C.G. and M.G. performed functional studies. L.C., A.M.-T., N.V., M.L.-G. and A.L.-G. performed clinical analysis. A.J.R., V.Q., M.A.B., E.C. and C.L.-O. conceived and directed the research and wrote the manuscript, which all authors have approved.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- 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. *Nat. Rev. Cancer* **10**, 37–50 (2010).
- Pekarsky, Y., Zanesi, N. & Croce, C.M. Molecular basis of CLL. *Semin. Cancer Biol.* **20**, 370–376 (2010).
- Cramer, P. & Hallek, M. Hematological cancer in 2011: new therapeutic targets and treatment strategies. *Nat. Rev. Clin. Oncol.* **9**, 72–74 (2012).
- Puente, X.S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
- Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2012).
- Fabbri, G. *et al.* Analysis of the chronic lymphocytic leukemia coding genome: role of *NOTCH1* mutational activation. *J. Exp. Med.* **208**, 1389–1401 (2011).
- Wang, L. *et al.* *SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
- Baumann, P. & Cech, T.R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).
- Loayza, D. & De Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013–1018 (2003).
- González-Pérez, A. & Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88**, 440–449 (2011).
- Tejera, A.M. *et al.* TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Dev. Cell* **18**, 775–789 (2010).
- Martínez, P. *et al.* Increased telomere fragility and fusions resulting from *TRF1* deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev.* **23**, 2060–2075 (2009).
- Sfeir, A., Kabir, S., van Overbeek, M., Celli, G.B. & de Lange, T. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science* **327**, 1657–1661 (2010).
- Lei, M., Podell, E.R. & Cech, T.R. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct. Mol. Biol.* **11**, 1223–1229 (2004).
- Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
- Kendellen, M.F., Barrientos, K.S. & Counter, C.M. POT1 association with TRF2 regulates telomere length. *Mol. Cell Biol.* **29**, 5611–5619 (2009).
- Wu, L. *et al.* *Pot1* deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126**, 49–62 (2006).
- Hockemeyer, D., Sfeir, A.J., Shay, J.W., Wright, W.E. & de Lange, T. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. *EMBO J.* **24**, 2667–2678 (2005).
- Hockemeyer, D., Daniels, J.P., Takai, H. & de Lange, T. Recent expansion of the telomeric complex in rodents: two distinct POT1 proteins protect mouse telomeres. *Cell* **126**, 63–77 (2006).
- Martínez, P. & Blasco, M.A. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nat. Rev. Cancer* **11**, 161–176 (2011).
- Roos, G. *et al.* Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. *Blood* **111**, 2246–2252 (2008).
- Lin, T.T. *et al.* Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. *Blood* **116**, 1899–1907 (2010).
- Stephens, P.J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).
- Barrientos, K.S. *et al.* Distinct functions of POT1 at telomeres. *Mol. Cell Biol.* **28**, 5251–5264 (2008).
- Hudson, T.J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).

ONLINE METHODS

Subjects. All individuals gave informed consent for their participation in the study following the guidelines of the International Cancer Genome Consortium (ICGC). Clinical characteristics of the 341 subjects are shown in **Supplementary Table 1**.

Collection and preparation of samples. The tumor samples used for exome sequencing were obtained from fresh or cryopreserved mononuclear cells. To purify the CLL fraction, samples were incubated with a cocktail of magnetically labeled antibodies directed against T cells (CD2 and CD3), natural killer (NK) cells (CD11b and CD56), monocytes (CD11b and CD13) and granulocytes (CD11b, CD14 and CD15), with adjustment for the percentage of each contaminating population (AutoMACS, Miltenyi Biotec). The degree of contamination by non-CLL cells in the CLL fraction was assessed by immunophenotyping and flow cytometry. Whole blood was sedimented with 2% dextran, and the leukocyte fraction was obtained. DNA was extracted from purified samples using a FlexiGene DNA kit (Qiagen). The quality of purified DNA was assessed by SYBR Green staining on agarose gels, and DNA was quantified using a Nanodrop ND-100 spectrophotometer. The tumor samples for exome sequencing contained $\geq 95\%$ neoplastic cells, and the contamination from neoplastic cells in normal DNA was $< 2\%$. The samples for the CLL validation cohort contained $> 30\%$ neoplastic cells.

Exome enrichment. We sheered 3 μg of genomic DNA from each sample and used it for the construction of a paired-end sequencing library as described in the paired-end sequencing sample preparation protocol provided by Illumina. Enrichment of exon sequences was then performed for each library using the SureSelect Human All Exon 50Mb kit (Agilent Technologies) following the manufacturer's instructions. Exon-enriched DNA was pulled down by magnetic beads coated with streptavidin (Invitrogen), washed and eluted, and 18 additional cycles of amplification of the captured library were performed. Exon enrichment was validated by RT-PCR on a 7300 Real-Time PCR System (Applied Biosystems) using a set of two pairs of primers to amplify exons and one pair of primers to amplify an intron. Enriched libraries were sequenced in one lane of an Illumina Genome Analyzer IIx sequencer using the standard protocol.

Read mapping and processing. For exome sequencing, reads from each library were mapped to the human reference genome (GRCh37) using Burrows-Wheeler Aligner (BWA) with the sampe option, and a BAM file was generated using SAMtools. Optical or PCR duplicates were removed using Picard. Bases were initially called using the Mapping and Assembly with Qualities (MAQ) consensus model implemented in SAMtools. For the identification of somatic substitutions, we used the *Sidron* algorithm, which has previously been described^{5,6}. This algorithm was adapted to account for sites with very high coverage resulting from exome enrichment. Thus, we added new cutoff values for the S parameter (30 for coverage higher than 50 and 50 for coverage higher than 100).

Sanger sequencing. PCR products were cleaned using ExoSap IT (USB Corporation) and sequenced using ABI Prism BigDye Terminator v3.1 technology (Applied Biosystems) with 5 pmol of each primer (**Supplementary Table 5**). Sequencing reactions were run on an ABI-3730 automated sequencer (Applied Biosystems). All sequences were manually examined.

Characterization of mutations. The amino-acid sequences of POT1 from evolutionarily diverse species were gathered from NCBI and aligned with CLUSTALX2 (ref. 28). Alignment was depicted with Jalview v2.7 (ref. 29). The structure of the OB domains of POT1 was obtained from the accession 3KJO in the Protein Data Bank (PDB) and was rendered with PyMol v0.99. To statistically assess mutational selection of residues that cluster within the OB folds, we compiled a list of all residues closer than 3.5 Å to the telomeric DNA in the crystal structure of POT1 (PDB 3KJP). This defined an interface between both molecules comprising 24 residues (residues 31, 33, 36, 39–42, 48, 60, 62, 87, 89, 94, 159, 161, 223, 224, 243, 245, 266, 267, 270, 271 and 273). We then performed a Monte Carlo simulation with 10 million sets of 8 random nonsynonymous mutations. All had less than 6 residues in this interface (only 1 set had 5).

Retroviral expression. The pLPC myc human *POT1* plasmid was a gift of T. de Lange (Addgene plasmid 12387)¹⁰. The *POT1* variants encoding p.Tyr36Asn and p.Tyr223Cys changes were generated by site-directed mutagenesis of the wild-type cDNA using the QuikChange XL site-directed mutagenesis kit (Stratagene) and the corresponding oligonucleotides (**Supplementary Table 6**). Retroviruses were packaged in HEK 293T cells using a vesicular stomatitis virus G (VSVG)-based package. HT1080 cells were seeded in 6-well plates at 20–30% confluency 24 h before infection. The following day, 1 ml of viral supernatant was added to the growth medium supplemented with 5 $\mu\text{g}/\text{ml}$ polybrene (Millipore), and plates were centrifuged at 200g for 1 h at room temperature and incubated for 8 h. This protocol was repeated twice, and cells were allowed to recover for 24 h in growth medium before undergoing selection with puromycin for 3 d. Cells then underwent serial passaging and were collected at the indicated population doubling points.

Confocal microscopy. HT1080 cells expressing Myc-tagged wild-type and mutant (Tyr36Asn and Tyr223Cys) POT1 were plated on sterile glass coverslips and allowed to adhere overnight. After 24 h, cells were fixed with 2% formaldehyde and were permeabilized by incubation with 1% Triton X-100 in PBS. Cells were washed, blocked with 0.2% cold-water-fish gelatin (Sigma) and 0.5% BSA in PBS and incubated for 4 h at room temperature with 9E10 antibody to Myc (Abcam; 1:1,000 dilution) in combination with an antibody to TRF1 (Abcam, ab10579; 1:1,000 dilution) in blocking buffer. For immunofluorescence detection of TIF proteins, HT1080 cells and peripheral blood mononuclear cells (PBMCs) were similarly fixed and permeabilized before incubation with antibodies to 53BP1 (Novus Biologicals, NB100-304; 1:500 dilution) and TRF1. After multiple washes, cells were incubated with Alexa Fluor 488- and Alexa Fluor 568-conjugated goat secondary antibodies to rabbit IgG (1:250 dilutions) for 1 h at room temperature. Nuclei were stained by incubating cells with 4,6-diamidino-2-phenylindole (DAPI), after which coverslips were mounted on slides, and cells were imaged with a Leica SP2 confocal microscope (Leica Microsystems). Images were processed using ImageJ and displayed using CorelDraw.

Protein blot analysis. Whole-cell lysates were collected in a buffer containing 1% Triton X-100, 50 mM Tris-HCl (pH 7.4), 150 mM NaCl and protease inhibitor cocktail (Roche Applied Science), and protein concentrations were determined by bicinchoninic acid assay (BCA; Pierce). Equal amounts of lysates (20 μg) were separated by SDS-PAGE and transferred to a nitrocellulose membrane that was blocked with 5% nonfat dry milk. Membranes were incubated overnight at 4 °C with 9E10 antibody to Myc (Abcam, ab32; 1:1,000 dilution), antibody to phosphorylated CHK1 (Cell Signaling Technology, 2341; 1:1,000 dilution) or antibody to β -actin (Sigma, A5441; 1:10,000 dilution) and then washed before incubation with species-appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies for 1 h at room temperature. Membranes were analyzed using an LAS-3000 mini imaging system (Fujifilm).

ChIP assays. For ChIP analysis, 3×10^6 cells were used per condition. Formaldehyde was added directly to the tissue culture medium to a final concentration of 1%, and cells were incubated for 15 min at room temperature (25 °C) on a shaking platform. Cross-linking was then stopped by the addition of glycine to a final concentration of 0.125 M. Cross-linked cells were washed twice with cold PBS, collected and lysed at a density of 20×10^6 cells/ml for 10 min at 4 °C in a buffer of 1% SDS, 50 mM Tris-HCl (pH 8.0) and 10 mM EDTA containing protease inhibitors. Lysates were sonicated to obtain chromatin fragments of < 1 kb in size and centrifuged for 15 min in a microcentrifuge at room temperature. Chromatin was diluted 1:10 with a buffer composed of 1.1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris-HCl (pH 8.0) containing protease inhibitors and precleared with a salmon sperm DNA-Protein A–50% agarose slurry (Upstate). Chromatin fragments were incubated with 10 μg of one of the following antibodies at 4 °C overnight on a rotating platform: 9E10 antibody to Myc, antibody to TRF1 (Abcam, ab1423), antibody to TRF2 (Millipore, 05-521), antibody to POT1 (Abcam, ab21382) and antibody to RAP1 (Bethyl Laboratories, A300-306A). Salmon sperm DNA-protein A agarose beads (60 μl) were then added, and incubation was continued for 1 h. Immunoprecipitation pellets were washed with 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl

(pH 8.0) and 150 mM NaCl (one wash); 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.0) and 500 mM NaCl (one wash); 0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl (pH 8.0) (one wash); and 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA (two washes). Chromatin was eluted from the beads twice by incubation with 250 μ l of 1% SDS and 0.1 M NaHCO₃ for 15 min at room temperature with rotation. After adding 20 μ l of 5 M NaCl, cross-links were reversed by incubation for 4 h at 65 °C. Samples were supplemented with 20 μ l of 1 M Tris-HCl (pH 6.5), 10 μ l of 0.5 M EDTA, 20 μ g of RNase A and 40 μ g of proteinase K and incubated for 1 h at 45 °C. DNA was then recovered by phenol-chloroform extraction and ethanol precipitation, slot-blotted onto a Hybond N+ membrane and hybridized with a telomeric probe obtained from a plasmid containing 1.6 kb of TTAGGG repeats. We quantified the signal using ImageJ software. For total input DNA samples, aliquots corresponding to a 1:80 dilution of the amount of lysate used in the immunoprecipitations were processed along with the rest of the samples during the cross-link reversal step. We calculated the amount of telomeric DNA immunoprecipitated in each ChIP assay on the basis of the signal relative to the corresponding total input telomeric DNA signal. In all cases, we represented the ChIP values as a percentage of the total input telomeric DNA, thus correcting for differences in the number of telomere repeats.

In vitro translation and G-strand binding assays. Human *POT1* in a T7 expression vector (OriGene) was used to generate the *POT1* variants encoding the Tyr36Asn and Tyr223Cys mutants by site-directed mutagenesis as outlined above. Each of the T7 vectors containing wild-type and mutant human *POT1* cDNA was used for *in vitro* protein expression with the TNT coupled reticulocyte lysate kit (Promega) following the manufacturer's instructions. Briefly, a 50- μ l reaction mixture containing 1 μ g of plasmid DNA, 2 μ l of EasyTag L-[³⁵S]-methionine (1,000 Ci/mmol; PerkinElmer) and 25 μ l of rabbit reticulocyte lysate was incubated at 30 °C for 90 min. A 5- μ l fraction of each reaction was analyzed by SDS-PAGE, and proteins were visualized and relative amounts quantified using an FLA 7000 phosphorimager system (Fujifilm). DNA binding assays were performed as described previously with minor modifications³⁰. In 20- μ l reaction mixtures, 5 μ l of each translation reaction was incubated with 10 nM 5-[³²P]-labeled telomeric oligonucleotide (GGTTAGGGTTAGGGTTAGGG) and 1 μ g of the nonspecific competitor DNA poly(dI-dC) in binding buffer (25 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 1 mM EDTA and 5% glycerol). Reactions were incubated for 10 min at room temperature, and protein-DNA complexes were analyzed by electrophoresis on a 6% polyacrylamide Tris-borate EDTA gel run at 80 V for 3 h. Gels were visualized by exposure to a phosphorimager screen.

Telomere length assays. Using the TeloTAGGG Telomere Length Assay kit (Roche), 2 μ g of DNA for each sample were digested and subjected to Southern blot analysis, and determination of mean terminal restriction fragment (TRF) length was performed according to the manufacturer's instructions. Densitometry arrays of 30–40 quadrants were generated across the entire lane for each sample using Multi-Gauge V3.0 software (Fujifilm). After background subtraction, absorbance (A_i) was obtained in each of these quadrants corresponding to different DNA lengths (L_i). Mean TRF was defined as $\Sigma(A_i)/\Sigma(A_i/L_i)$ for each sample.

TRAP assays. Telomerase activity was measured with a modified telomeric repeat amplification protocol (TRAP) assay as previously described³¹.

Telomeric FISH on metaphase spreads and HTQFISH. For telomeric FISH on metaphase spreads, 3×10^6 PBMCs or HT1080 cells expressing Myc-tagged wild-type and mutant (Tyr36Asn and Tyr223Cys) *POT1* were stimulated with 7.5 ng/ml 12-O-tetradecanoylphorbol-13-acetate (TPA) in RPMI medium supplemented with FBS, penicillin-streptomycin, β -mercaptoethanol, sodium pyruvate, nonessential amino acids and L-glutamine. After 72 h, metaphase spreads were obtained, and telomeric FISH was performed as described³². At least 5–10 metaphase spreads per subject or HT1080 cell line were analyzed, and chromosomal aberrations were quantified and represented as frequency per metaphase. High-throughput quantitative telomeric FISH (HTQFISH) was performed as described³³, seeding at least 150,000 cells in duplicate wells and analyzing >5,000 nuclei per HT1080 cell line or subject sample.

28. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
29. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
30. Baumann, P., Podell, E. & Cech, T.R. Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. *Mol. Cell Biol.* **22**, 8079–8087 (2002).
31. Blasco, M.A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
32. Samper, E. *et al.* Normal telomere length and chromosomal end capping in poly(ADP-ribose) polymerase-deficient mice and primary cells despite increased chromosomal instability. *J. Cell Biol.* **154**, 49–60 (2001).
33. Canela, A., Vera, E., Klatt, P. & Blasco, M.A. High-throughput telomere length quantification by FISH and its application to human population studies. *Proc. Natl. Acad. Sci. USA* **104**, 5300–5305 (2007).