

News and Commentary

Chronic lymphocytic leukemia: looking into the dark side of the genome

R Valdés-Mas^{1,2}, J Gutiérrez-Abril^{1,2}, XS Puente¹ and C López-Otín^{*1}

Cell Death and Differentiation (2016) 23, 7–9; doi:10.1038/cdd.2015.155; published online 27 November 2015

The International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) have recently reported in *Nature* their analysis of more than 1000 chronic lymphocytic leukemia (CLL) genomes.^{1,2} CLL, the most common leukemia in adults of Western countries, is a neoplasia characterized by an abnormal accumulation of B lymphocytes in blood, lymph nodes and bone marrow. This disease has traditionally been classified into two molecular groups according to whether the variable region of immunoglobulin heavy-chain genes of tumor B cells has undergone somatic hypermutation in the germline center (IGHV-mut) or not (IGHV-unmut). The former is associated with a favorable clinical course, whereas the latter represents the most aggressive form of the disease. Clinical heterogeneity and high prevalence are the defining features of CLL, but the driving events responsible for its initiation and development are poorly understood.^{3,4} On this basis, ICGC and TCGA projects were set up in 2008 to identify the genetic alterations that trigger this neoplasia.

Both consortia have taken advantage of the most recent next-generation sequencing techniques, allowing the accumulation of an important body of information about the genomic alterations underlying CLL leukemogenesis. The first genomic studies stemming from these initiatives identified recurrently mutated genes such as *NOTCH1*, *SF3B1*, *TP53*, *BIRC3*, *POT1* and *CHD2*.^{5–10} Notably, these studies also revealed a marked genetic heterogeneity of CLL, with most genes mutated at frequencies lower than 5%. These new analyses presented by the ICGC and TCGA consortia, with cohort sizes of 506 and 538 CLL samples, are expected to saturate candidate CLL gene discovery for genes mutated in 5% of patients, providing 94 and 61% accuracy to detect genes mutated in 3 and 2% of patients, respectively. These new genomic analyses have resulted in the identification of 25 novel driver genes that were detected in both studies, including changes in *RPS15*, *IKZF3*, *ZNF292*, *ZMYM3* and *ARID1A*. Therefore, it may be assumed that the set of genes implicated in the pathogenesis of CLL is essentially complete.

In addition, the analysis of 150 CLL tumors by whole-genome sequencing has allowed the interrogation of regions outside the coding regions. The scientific interest in these regions has spiked after the ENCODE project concluded in

2012 that up to 80% of the human genome could have functional relevance.¹¹ However, most genomic studies to date have focused on the identification of functional mutations affecting protein-coding regions, which represent only 2% of the genome. Interestingly, the novel ICGC-CLL genomic study¹ has uncovered recurrent non-coding mutations in different functional elements (Figure 1). The most frequent alteration of this kind has been detected in the 3' UTR of *NOTCH1*, which produces an alternative splicing event within the last exon of this gene. This event results in the synthesis of a shorter protein, lacking the C-terminal PEST domain, which increases the stability and activity of NOTCH1. Strikingly, this non-coding mutation causes an equivalent effect on the protein and a similar poor clinical prognosis in patients as the frequent frameshift (p.P2514Rfs*4) detected in the coding region of *NOTCH1*. This new study has also found intergenic regions that accumulate a high mutational density. The most striking example has been identified in chromosome 9p13, containing an active enhancer that regulates the expression of the relatively close gene *PAX5*. Mutations in this region reduce the expression of *PAX5*, which encodes a transcription factor implicated in B-cell differentiation. Together, these findings confirm the need to explore the so-called dark side of the genome to detect new alterations implicated in the development of tumors that may be used as potential target therapies. Hopefully, strategies for the treatment of patients targeting specific coding mutations might be applied in the same way to patients with equivalent non-coding alterations in their tumoral genomes.

These discoveries suggest that the highly variable clinical course of this leukemia is a consequence of its genetic and epigenetic heterogeneity, hampering the development of novel therapies. To address this question, an integrative analysis of the mutational landscape has been developed to discover the common pathways—where all the different CLL driver genes are grouped—that are altered in this pathology. This analysis has shown that a total of eight biochemical pathways are mainly involved in CLL, including BCR signaling, cell cycle regulation, DNA damage response, chromatin remodeling, NF- κ B signaling, NOTCH1 signaling, RNA metabolism and apoptosis. In particular, an imbalance between apoptosis and proliferative signals provokes the

¹Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología (IUOPA), Universidad de Oviedo, Oviedo, Spain

*Corresponding author: C López-Otín, Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología (IUOPA), Universidad de Oviedo, Oviedo 33006, Spain. Tel: +34 985104201; Fax: +34 985103564; E-mail: clo@uniovi.es

²These authors contributed equally to this work.

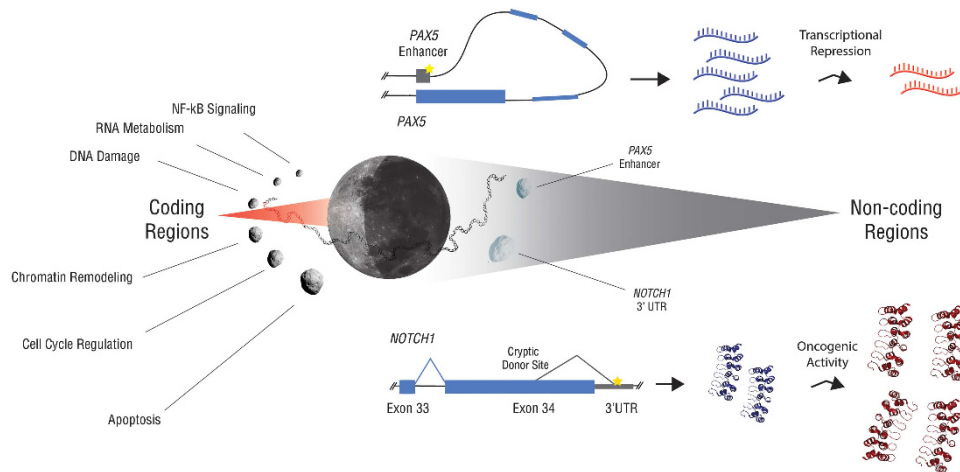


Figure 1 Identification of recurrent non-coding mutations in the whole-genome analysis of CLL. The genomic studies have determined that the different driver genes mutated in CLL can be grouped in eight different biochemical pathways. These pathways are depicted in the figure as small asteroids orbiting around the moon. Whole-exome sequencing has allowed the study of coding regions, representing a small portion of the total human genome (described as the portion of DNA in the bright side of the moon). Whole-genome sequencing has recently allowed the identification of variants in the non-coding regions (present in the dark side of the scheme). The discovery of recurrent mutations in the dark side of the genome (represented as yellow stars) has resulted in the finding of novel mechanisms implicated in tumorigenesis. (Upper part) Tumors with mutations located in an enhancer of *PAX5* show a significant lower expression of this transcription factor. (Lower part) Mutations in the 3' UTR of *NOTCH1* provoke an aberrant splicing event. By this mechanism, the resulting *NOTCH1* protein loses part of the C-terminal domain, which contains a PEST sequence. Removal of this region generates an oncogenic *NOTCH1* isoform

characteristic expansion and accumulation of B cells in this neoplasia. Thus, altered mechanisms and excess of proliferative signals in CLL cells, delivered by the tumor microenvironment, limit the activation of the apoptotic program. In addition, NF-κB and B-cell signaling pathways are constitutively activated, which leads to a widespread overexpression of antiapoptotic genes.¹²

In regards to clinical management, the paradigm of 'watch and wait' strategy has been applied for the treatment of CLL patients with asymptomatic disease, treating them with chemotherapy only if symptoms are manifested. In recent years, the development of BCR inhibitors (ibrutinib, idelalisib), the anti-CD20 monoclonal antibodies obinutuzumab and ofatumumab, and the BCL2 inhibitors, which induce apoptosis in tumor cells, has expanded the available treatment possibilities.¹³ A future analysis of the response to these inhibitors of tumors bearing mutations in any of the newly identified CLL driver genes could contribute to the development of novel therapies for the treatment of drug-resistant patients carrying mutations in one of these drivers. Actually, aggressive forms of the disease with a refractory response or patients with relapse after treatment are still frequent. Understanding tumor development and its evolution after different therapies will improve the clinical approaches for CLL.¹⁴ The novel TCGA-CLL genomic study has assessed the genetic evolution over the course of the disease and in response to therapy.² By estimating the fraction of cancer cells, using sequencing data, the authors found that 58% of mutations were present in subclonal tumor populations, reflecting high intratumoral heterogeneity in CLL. The subsequent construction of a temporal map of the evolutionary trajectories of CLL has unveiled that copy number variants, such as *del(13q)*, *tri(12)* and *del(11q)*, are the earliest events of the tumorigenesis. Moreover, the study of pre-treatment and relapse samples of

59 CLL subjects has uncovered new insights into the clonal evolution in disease relapse. Thus, when large clonal shifts were detected, the relapsed clone was observed by whole-exome sequencing of pretreated samples in 30% of the cases. This analysis revealed that *TP53* mutations, *del(17p)* and *IKZF3* confer a fitness advantage under therapeutic selection. The progress in the identification of these genes responsible for tumor relapse could substitute the current 'watch and wait' strategy with the 'Anticipation-Based Chemotherapy' approach.¹⁵ Combining standard chemotherapy with other therapies targeting these drivers could minimize relapses in patients in the not-too-distant future.

The revolution derived from the implementation of next-generation sequencing technology has transformed our knowledge about CLL by deciphering the mutational events responsible for its initiation and progression. These genomic analyses have confirmed the characteristic molecular heterogeneity of this frequent form of leukemia. Accordingly, the determination of the prognostic significance of every identified alteration will require the longitudinal analysis of thousands of CLL patients, whose foundations have been laid down in these studies. Finally, these new works have also shed light on the dark side of the genome, identifying functional mutations outside coding regions, and have extended the evolutionary biography of this disease. These findings may provide the basis for the development of new therapeutic strategies for CLL patients, thereby expanding the range of druggable targets. In 1973, Theodosius Dobzhansky famously said that 'nothing in biology makes sense except in the light of evolution', but maybe nowadays he would say further: 'not even the shadows'.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements. This work was supported by grants from Ministerio de Economía y Competitividad and Instituto de Salud Carlos III (RTICC), Spain. The Instituto Universitario de Oncología is supported by Fundación Bancaria Caja de Ahorros de Asturias. CL-O is an Investigator of the Botín Foundation supported by Banco Santander through its Santander Universities Global Division.

1. Puente XS *et al. Nature* 2015; **526**: 519–524.
2. Landau DA *et al. Nature* 2015; **526**: 525–530.
3. Rodriguez D *et al. Int J Hematol* 2015; **101**: 219–228.
4. Sutton LA, Rosenquist R. *Semin Cancer Biol* 2015; **34**: 22–35.
5. Puente XS *et al. Nature* 2011; **475**: 101–105.
6. Quesada V *et al. Nat Genet* 2012; **44**: 47–52.
7. Wang L *et al. N Engl J Med* 2011; **365**: 2497–2506.
8. Rossi D *et al. Blood* 2012; **119**: 521–529.
9. Ramsay AJ *et al. Nat Genet* 2013; **45**: 526–530.
10. Rodriguez D *et al. Blood* 2015; **126**: 195–202.
11. Consortium EP. *Nature* 2012; **489**: 57–74.
12. Pekarsky Y, Croce CM. *Cell Death Differ* 2015; **22**: 6–11.
13. Byrd JC *et al. J Clin Oncol* 2014; **32**: 3039–3047.
14. Landau DA *et al. Cell* 2013; **152**: 714–726.
15. Puente XS, Lopez-Otin C. *Nat Genet* 2013; **45**: 229–231.