Article

Detection of rare germline variants in the genomes of B cell neoplasms

Adrián Mosquera Orgueira,[1,2,3] Miguel Cid López[1,2,3], Andrés Peleteiro Raíndo,[1,2,3] José Ángel Díaz Arias[1,2], Beatriz Antelo Rodríguez[1,3], Laura Bao Pérez,[2] Natalia Alonso Vence[1,2], Ángeles Bendaña López,[1,2,3] Aitor Abuin Blanco[2], Paula Melero Valentín[2], Roi Ferreiro Ferro[2], Carlos Aliste Santos[3], Máximo Francisco Fraga Rodríguez,[1,3,4], Marta Sonia González Pérez[1,2], Manuel Mateo Pérez Encinas[1,2,3] and José Luis Bello López[1,2,4]

[1]Health Research Institute of Santiago de Compostela (IDIS)
[2]Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Department of Hematology, SERGAS
[3]Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Department of Pathology, SERGAS
[4]University of Santiago de Compostela.
*Correspondence: adrian.mosquera.orgeira@sergas.es; Tel.: +34981950191

Simple Summary: The global importance of rare variants in tumorigenesis has been addressed by pan-cancer analysis, revealing significant enrichments of protein truncating variants in genes such as *ATM*, *BRCA1/2*, *BRIP1* and *MSH6*. Germline variants can influence treatment response and contribute to the development of treatment-related second neoplasms, especially in childhood leukemia. We aimed to analyze the genomes of patients with B-cell lymphoproliferative disorders for the discovery of genes enriched in rare pathogenic variants. We discovered a significant enrichment of 26 genes in germline protein truncating variants (PTVs), affecting cell signaling (*MET*, *JAK2*, *ANGPT2*), energy metabolism (*ACO1*) and nucleic acid metabolism and repair pathways (*NT5E*, *DCK*). Additionally, we detected rare and likely pathogenic variants associated with tumor subtype, disease prognosis and potential druggability, indicating a relevant role of these events in the variability of cancer phenotypes.

Abstract:

Growing evidence has revealed the implication of germline variation in cancer predisposition and prognostication. Here, we describe an analysis of putatively disruptive rare variants across the genomes of 726 patients with B-cell lymphoid neoplasms. We discovered a significant enrichment of 26 genes in germline protein truncating variants (PTVs), affecting cell signaling (*MET*, *JAK2*, *ANGPT2*), energy metabolism (*ACO1*) and nucleic acid metabolism and repair pathways (*NT5E*, *DCK*). Interestingly, some of these variants were restricted to either chronic lymphocytic leukemia (CLL) (i.e., *ANGPT2* and *AKR1C3*) or B-cell lymphoma cases (*PNMT*, *TPT1* and *IGHMBP2*). Additionally, we detected 1,675 likely disrupting variants in genes associated with cancer, of which 44.75% were novel events and 7.88% were PTVs. Among these, the most frequently affected genes were *ATM*, *BIRC6*, *CLTCL1A* and *TSC2*. Homozygous or compound heterozygous variants were detected in 28 cases; and coexisting somatic events were observed in 17 patients, some of which affected key lymphoma drivers such as *ATM*, *KMT2D* and *MYC*. Finally, we observed that variants in the helicase gene *WRN* were independently associated with shorter survival in CLL. Our study results support an important role for rare germline variation in the pathogenesis, clinical presentation and disease outcome of B-cell lymphoid neoplasms.

Keywords: germline, rare variant, cancer, lymphoid, B-cell, lymphoma, CLL, driver, prognosis.

Introduction

B-cell lymphoid neoplasms are the most frequent hematological tumors, and they exhibit a diverse spectrum of entities with heterogeneous clinical behaviour. B-cell lymphoid neoplasms are classically classified in either aggressive lymphomas (DLBCL, Burkitt lymphoma, grade III follicular lymphoma and mantle cell lymphomas), or indolent lymphomas (chronic lymphocytic leukemia (CLL), grade I/II follicular lymphoma, marginal zone lymphoma, lymphoplasmacytic lymphoma…). By frequency, diffuse large B-cell lymphoma (DLBCL) is the most frequent lymphoid neoplasm, accounting for 25% of all cases of non-Hodgkin lymphoma (NHL), closely followed by CLL (19% of NHLs) and follicular lymphoma (12% of NHLs) [1].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

Next-generation sequencing (NGS) technologies have deconvoluted the genomic complexity of B-cell lymphoid tumors to a great extent, revealing the most frequent molecular drivers of disease and the interplay among them. NHL cases show familial predisposition, and much of the heritability of these diseases is still unexplained.[2] Genome-wide association analysis (GWAS) have identified the existence of polymorphisms significantly associated with risk of CLL,[3] DLBCL[4] and follicular lymphoma [5]. Similarly, some polymorphisms are also related with the outcome of B-cell lymphomas [6-8] and CLL [9]; and it has also been proved that some variants cooperate with somatic events in shaping clinical outcomes of cancer patients [10]. Another source of germline variation consists of rare variants (allele frequency < 0.1-1%). The global importance of such rare variants in tumorigenesis has been addressed by pan-cancer analysis, revealing significant enrichments of protein truncating variants in genes such as *ATM*, *BRCA1/2*, *BRIP1* and *MSH6* [11]. Indeed, some of these variants predispose to cancer development through the acquisition of second somatic hits [12], such as point mutations or loss-of-heterozygosity (LOH) [13]. Additionally, germline variation can influence treatment response and contribute to the development of treatment-related second neoplasms, especially in childhood leukemia [14]. Many such rare variants in cancer related genes have been associated with particular cancer subtypes [15-17], but until now little attention has been focused on the genome-wide frequency, pathogenicity and clinical implications of rare variants in lymphoid malignancies. Rare variants in *ATM* and *CDK1* variants have been associated with CLL risk in genome-wide analysis [18], whereas evidence for the implication of infrequent events in other genes come from familial studies or single-gene analysis [19-21].

In this report we performed an exploratory analysis of the frequency and distribution of rare and putatively pathogenic germline variants in the genome of several mature B cell lymphoid neoplasms using high-throughput sequencing data produced by the *International Cancer Genome Consortium* (ICGC) [22]. Our results indicate the existence of multiple genes affected by highly pathogenic germline variants in the genome of these patients, some of which seem to condition phenotypic expression and patient survival.

2. Materials and Methods

2.1 Data source

We processed germline next-generation sequencing data obtained from 726 patients with B-cell lymphoid malignancies that were included in the *International Cancer Genome Consortium*. Briefly, 504 cases pertained to the Spanish Chronic Lymphocytic Leukemia project, and 222 were retrieved from the German Malignant Lymphoma project. Overall, there were 504 chronic lymphocytic leukemia (CLL) or small lymphocytic lymphoma (SLL) cases (including 54 monoclonal B cell lymphocytosis cases), 97 follicular lymphoma cases, 85 diffuse large b-cell lymphoma (DLBCL) cases, 36 Burkitt lymphoma cases and 4 unclassified B-cell lymphoma cases. CLL control samples were derived from non-tumoral leukocytes (<2% tumor contamination), whereas lymphoma controls originated from whole blood or buffy coats checked for negative clonality analysis.

2.2 Germline variant identification and annotation

Most CLL germline samples (440 out of 502) were processed using exome-sequencing kits (*Agilent SureSelect Human All Exon V4 and V4+UTRs*), whereas whole genome sequencing was done on a group of 262 CLL cases and all B-cell lymphoma cases included in the MALY-DE project. We restricted our analysis to protein-coding regions covered by the exome-sequencing kits. Variants were detected using the optimized bcbio-nextgen (version 1.1.5) pipeline [23], and the GRCh37.75 assembly was used as reference. Four different variant callers were used: freebayes (version 1.1.0.46) [24], GATK-Haplotype (GATK version 2.8) [25], Platypus (version 0.8.1.2) [26] and Samtools (version 1.9) [27], with default parameters. Homopolymers and regions with low complexity, alternative contigs or abnormally high coverage were discarded. Similarly, we used 100bp mappability tracks in the UCSC database to filter out variants in low mappability regions. Finally, a variant was called if detected by a minimum of 2 callers and if it had a minimum genotype quality of 30 Phred and a minimum coverage of 10x. Finally, we filtered events with variant allele fraction (VAF) < 30% in order to limit possible contamination of the controls with tumor cells. Variants were annotated using dbSNP [28], 1000 Genomes [29], ExAc [30] and gnomAD [31]. Only variants with a major allele frequency (MAF) below 0.5% in any ethnic

population were retained. Thereafter, we selected 1) all protein truncating variants (PTVs): start lost, stop lost, nonsense, frameshift, splice acceptor and splice donor variants, and 2) missense variants with pathogenicity CADD v.14 [32] scores > 20 Phred (i.e., variants in the top 1% of predicted pathogenicity). Finally, we restricted our analysis to those genes involved in carcinogenesis, and particularly in lymphomagenesis. We collected the following types of genes: 1) 162 genes involved in mendelian inherited cancer syndromes [33], 2) 723 genes included in the *Cancer Gene Census* [34], 3) 135 genes included in the TARGET database ("a database of genes that, when somatically altered in cancer, are directly linked to a clinical action" [35]), 4) 59 recurrently mutated genes in CLL (*Landau et al.*[36] and *Puente et al.* [37]), 5) 150 recurrently mutated genes in DLBCL (*Reddy et al.* [38]) and 6) 72 recurrently mutated genes in Burkitt lymphoma (*Panea et al.* [39]). The final list contained 906 non-redundant genes. Visual analysis of all frameshift insertions and deletions was performed using IGV. Ancestry analysis was performed using Peddy [40], which predicts ancestry using a machine learning model trained on individuals of diverse ancestries from the 1000 Genomes Project reference panel. Only 3 of the patients were of non-european ancestry (1 African, 1 south Asian and 1 east Asian). Genes affected by 5 or more variants were annotated to the top 0.5% and 1% genes in FLAG database [41] in order to highlight potentially spurious discoveries. Finally, survival analysis was performed with cox regression. Multiple testing correction was performed with the FDR method.

2.3 Burden test against public controls

We used *Testing Rare vAriants using Public Data* (TRAPD) software in order to compare the enrichment of our cohort of patients in PTVs against 15,708 public controls from the gnomAD v2 whole genome sequencing dataset [42]. Importantly, none of these controls originated from cancer studies. PTV variants with a maximum allele frequency (popmax) < 0.5% were selected. Multiple testing correction was performed with the FDR method.

2.4 Compound heterozygotes and germline-somatic double hit event detection

In order to identify putatively compound heterozygous genes, we selected concurrent rare heterozygous and putatively damaging variants affecting the same gene in the same individual. All variants in linkage disequilibrium ($R^2 > 0.2$) were discarded from this analysis according to 1000 Genomes data, as they could pertain to the sample haplotype.

Second-hit somatic mutations were detected by comparing germline variants with somatic mutations for the same set of individuals present in the ICGC database.

2.5 Myeloid clonal hematopoiesis filtering

Potentially mosaic somatic mutations in the blood controls due to myeloid clonal hematopoiesis of undetermined significance (CHIP) could exist. In order to assess this issue we initially identified a list of 22 recurrently mutated genes in clonal hematopoiesis that had at least one putatively rare germline variant in the final dataset [43-45]. Among these genes, we analyzed if the variants were present in both the control and tumor (lymphoid) compartment, and those mutations that were not found (or found at very low VAF) in the tumoral department were catalogued as likely myeloid CHIP events.

3. Results

3.1 Rare variants overview

1,665 rare germline variants with likely disruptive activity (CADD scores > 20 or protein truncating) were detected in 559 cancer-related genes across 693 (95.45%) patients (**Figure 1**, **Supplementary Table 1**). Overall, the frequency of these rare and likely disrupting mutations in cancer-related genes was superior to those found in non-cancer related genes (4.25×10^{-3} vs 3.61×10^{-3} mutations per gene & patient). Most of these were missense variants (1,559 events, 93.01%).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

Interestingly, we only detected 10 likely somatic mosaic mutations among myeloid-CHIP related genes, which affected *TET2*, *DNMT3A*, *ASXL2*, *BCORL1* and *PPM1D* (**Supplementary Table 2**). These variants were removed from downstream analysis.

Overall, 113 patients (15.56%) harbored 126 PTVs in 103 different loci, which included frameshift, splice donor, splice acceptor, nonsense, stop loss and start loss variants (**Supplementary Table 3**). The frequency of PTVs in this gene list was notoriously superior to that observed in the remaining genes (2.11×10^{-3} vs 7.33×10^{-4} mutations per gene & patient), suggesting an enrichment in loss of function mutations among cancer-related genes. The most frequently affected genes were *ATM* (5 cases), *SETDB1* (5 cases in a single locus), *ISX* (4 cases) and *POLQ* (4 cases).

Some of the missense variants showed a remarkable increased frequency compared in patients with lymphoid neoplasia compared with the non-Finnish European (NFE) gnomAD database. This was the case of the variants rs199502695 in *PRPF40B* (4 cases, 71.17 times more frequent), rs191413750 in *DOCK8* (5 cases, 55.55 times more frequent), rs377188372 in *N4BP2* (4 cases, 34.66 times more frequent) and rs146946726 in *MLLT10* (6 cases, 8.10 times more frequent).

227 different variants have also been described as pathogenic or likely pathogenic somatic mutations in cancer (**Supplementary Table 4**). Remarkably, 2 out of 3 variants in *NOTCH1* are flagged as pathogenic somatic mutations in COSMIC. This overlap also occurred in *BCL6* (2 out of 5 variants), *PTCH1* (2 out of 4 variants), *ATM* (3 out of 14 variants), *CNOT3* (1 out of 2 variants), *DNMT1* (1 out of 2 variants), *FGFR2* (1 out of 2 variants), *JAK3* (1 out of 2 variants), *MTOR* (1 out of 2 variants) and *MDM4* (1 out of 2 variants). Furthermore, this phenomenon occurred in some cancer drivers affected by a single rare variant, such as *CCND2*, *CHIC2*, *CDKN1B*, *CREBBP*, *EZH2*, *FGFR3*, *JAK2*, *PRF1*, *RUNX1*, *SIRPA*, *SUFU*, *TRIP11* and *YWHAE*.

Finally, 11 variants in homozygosis were observed, one of which (c.1642C>T in *ZCCHC8*) was present in 2 different patients (**Table 1**). Similarly, 15 patients harbored two variants in the same gene, probably in the form of compound heterozygotes (**Table 1**). Interestingly, these compound heterozygotes were observed twice in *FAT1* and *ZFHX3*. Moreover, one homozygous nonsense variant and a compound heterozygote were detected in the gene *GLI1*, and one homozygous missense variant plus a compound heterozygote was detected in *MYH9*.

3.2 Rare variants affecting lymphoma driver genes

636 occurrences in 459 different rare variants were detected across 143 driver genes of lymphomagenesis extracted from the literature. These events affected 415 patients (57.16%) (**Supplementary Table 5**). The most commonly mutated genes were *ATM* (25 cases, **Figure 2 A and B**), *BIRC6* (24 cases), *SPEN* (15 cases), *ZNF292* (13 cases), *MGA* (12 cases), *BAZ2A* (12 cases), *NCOR1* (11 cases), *GNA13* (10 cases) and *WDR7* (10 cases) (**Table 2**).

Various variants also affected other drivers of lymphomagenesis, such as *ARID1A* (9 cases), *CHD1* (9 cases), *MECOM* (9 cases), *NOTCH1* (7 cases), *SETD2* (6 cases), *ARID1B* (5 cases), *BLC6* (5 cases), *CTCF* (5 cases), *EP300* (5 cases), *JAK3* (5 cases), *NOTCH2* (5 cases), *MET* (5 cases), *MYC* (5 cases), *TCF3* (5 cases) and *CHD2* (4 cases). Finally, infrequent variants in *TRAF2* were detected in 3 patients; whereas those of *CNOT3*, *ID3*, *IKZF3*, *MTOR*, *POT1* and *STAT5B* occurred in 2 patients each; and those of *ASXL1*, *BRAF*, *CARD11*, *CCND2*, *CCND3*, *CREBBP*, *DTX1*, *ETV6*, *EZH2*, *KRAS*, *MCL1* and *TCL1A* were detected in just one case each. Notably, both variants in *SAMDH1* were PTV (a frameshift and a nonsense event) (**Supplementary Table 6**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

3.3 Rare variants affecting genes involved in cancer syndromes with germline inheritance

84 genes associated with inherited cancer syndromes were affected by a total of 372 occurrences of 225 different rare variants (**Supplementary Table 7**), of which 19 were PTV and affected 22 patients (3%). 131 variants were observed in genes linked with autosomal dominant syndromic cancer, affecting 168 patients. Among these, the most frequently mutated genes were *TSC2* (22 cases), linked to tuberous sclerosis, *APC* (16 cases), linked to hereditary colon cancer, and the DNA polymerase *POLE* (16 cases), involved in predisposition to multiple cancers (**Table 2**). Similarly, 94 variants in 32 genes linked to autosomal recessive cancer were observed, which affected 149 patients. The most commonly affected among these were *ATM* (25 cases), *NBN* (12 cases), *BLM* (12 cases), *DOCK8* (12 cases) and *WRN* (12 cases) (**Table 2**).

Some of these variants were labelled as pathogenic in *ClinVar* (**Supplementary Table 6**). This included a missense variant in *MITF* (rs149617956, 7 cases), a missense variant in *GBA* (rs76763715, 6 cases), a missense variant in *MUTYH* (rs34612342, 2 cases), two missense variants in *SERPINA1* (rs61761869 and rs28931570, 3 cases), a missense variant in *NBN* (rs61754966, 1 case) and a frameshift deletion in *BRCA2* (rs397507591, 1 case). Likely pathogenic variants were detected in *APC* (missense variant, rs139196838, 1 case), *BRIP1* (frameshift insertion, rs878855150, 1 case) and *MET* (missense variant, rs34589476, 1 case).

Interestingly, 95 undescribed variants were detected, and these were particularly frequent in *ATM* (4 missense variants, 1 nonsense variants and 2 frameshift deletions, including a 28 base pair deletion), *EXT1* (2 missense variants and 1 splice donor variants), *MET* (2 missense variants and 1 frameshift deletion), *MITF* (1 missense variants in 2 patients and 1 missense variants in 1 patient), *DOCK8* (3 missense variants), *MSH6* (2 missense variants and 1 nonsense variant), *SMARCA2* (3 missense variants), *SOS1* (3 missense variants), *TRIM37* (3 missense variants) and *WRN* (1 missense, 1 nonsense and 1 splice gain variant).

3.4 Rare variants in genes of the Cancer Gene Census and TARGET databases

327 occurrences of 208 rare variants in 95 different genes linked to therapy were identified. These affected 247 patients (34.02%) (**Supplementary Table 8**). The most recurrently affected genes were *ATM* (25 cases), *TSC2* (22 cases), *APC* (16 cases), *ROS1* (16 cases) and *JAK2* (11 cases) (**Table 2**). Among *Cancer Gene Census* genes, 1,346 events were detected in 947 different loci (**Supplementary Table 9**), being the most recurrent ones those in *CLTCL1* (24 cases), *CSMD3* (21 cases), *MYH9* (20 cases), *ANK1* (19 cases),*TPR* (19 cases), *MYH11* (18 cases), *PTPN13* (18 cases) and *POLQ* (17 cases) (**Table 2**). Some genes were enriched in PTV variants, particularly *POLQ* (4 out of 10 different variants), *AKAP9* (3 out of 10), *TSHR* (2 out of 3) and *ISX* (2 out of 2). Furthermore, 1 pathogenic (rs113994096, 5 patients) and 1 likely pathogenic (rs138929605, 1 patient) missense variants in the DNA polymerase *POLG* were also discovered.

3.5 Differential distribution of rare variants and association with patient survival

We did not identify any gene significantly enriched in rare variants in CLL *vs* B-cell lymphoma cases (Fisher test, FDR<5%). Nevertheless, we discovered that some variants were only detected in one subgroup. For example, the missense variant rs1800729 in *TSC2* was exclusively present in CLL (8 cases), and the missense variant rs139075637 in *POLE* was exclusively present in non-CLL B lymphoid tumors (7 cases). Further analysis needs to be performed in order to confirm these findings and rule-out population substructure biases.

Thereafter, we tested if rare variants could be associated with adverse patient outcomes. Due to the heterogeneity of the dataset and sample size limitations, we restricted our analysis to CLL cases, and considered variants present in at least 1% of cases. Interestingly, rare variants in the DNA helicase *WRN* (8 cases) were significantly associated with shorter

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

overall survival (cox p-value 1.16 x 10^{-4}, q-value 0.01, HR [2.35, 14.59], **Figure 3 B**). Indeed, such association was independent of age at diagnosis and CLL/MBL status (p-value 1.97 x 10^{-7}, HR [5.03, 35.48]). Moreover, these variants were also linked to shorter time to first treatment (cox p-value 6.15 x 10^{-4}, HR [1.85, 9.48], **Figure 3 A**), which remained significant after adjusting for age at diagnosis and CLL/MBL status (p-value 1.69 x 10^{-3}, HR [1.64, 8.48]). These patients tended to harbor high-risk karyotype anomalies in the tumor cells: 11q deletion (3 cases, one as an isolated anomaly, one co-occurring with 13q deletion and one co-occurring with 3 other karyotype anomalies), 17p deletion (1 case, co-occurring with a 18p deletion), 8q deletion (1 case, co-occurring with 21q gain), and 6q deletion (1 case, co-occurring with 13q deletion).

Curiously, *ATM* germline variants appeared not to be associated with survival in CLL. We reasoned that this could be due to the inclusion of missense variants in the model, since *ATM* is a gene with great variability in the population. Therefore, we restricted the analysis to patients with truncating variants in *ATM* (4 cases), and discovered that these few patients had a significantly shorter overall survival (p-value 0.02, [HR 1.28,21.53]).

3.6 Association of rare germline variants with somatic mutations

Concurrent rare and likely disruptive germline variants and somatic mutations were detected in 17 cases (**Table 3**). Co-occurring mutations in CLL affected *GNA13*, *KMT2D*, *LRP1B*, *MUC16* and *SPEN*. Additionally, co-occurring mutations in B-cell lymphomas were found in *CSMD3* (grade I follicular lymphoma), *EP300* (DLBCL), *FAT1* (DLBCL), *HIST1H1E* (grade I follicular lymphoma), *KMT2D* (DLBCL), *MCL1* (grade IIIa follicular lymphoma), *MSH6* (grade IIIa follicular lymphoma), *MYC* (DLBCL), *PIM1* (grade I follicular lymphoma), *RNF213* (DLBCL) and *SIN3A* (grade IIIb follicular lymphoma). Additionally, we observed a germline mutation in *ATM* co-occurring with a 11q copy neutral loss of heterozygosity that induced loss of the reference allele in a CLL patient.

3.7 Burden test of protein truncating variants using public controls

Germline PTVs were selected for association with risk of B-cell neoplasms using the burden test against public whole genome sequencing controls. We selected PTVs because these are the most potentially pathogenic variants. As a result, 25 genes were significantly enriched in PTVs among patients affected by B-cell lymphoid neoplasms (q-value < 0.05), and one additional gene (*LPIN3*) showed a trend towards association (q-value 6.83 x 10^{-2}) (**Table 4**). Importantly, inflation statistics were low (λ=1.05; **Supplementary Figure 1**). Overall, we identified 56 different events occurring in 205 cases (**Supplementary Table 10**). The most significantly genes were *PPIL1* (q-value 1.49 x 10^{-9}), *JAK2* (q-value 3.01 x 10^{-9}), *NT5E/CD73* (q-value 3.02 x 10^{-7}), *TPT1* (q-value 1.52 x 10^{-5}) and *TLR4* (q-value 1.93 x 10^{-5}) (**Figure 4**).

Most of these genes participate in important cellular pathways associated with cancer pathogenesis. *ANGPT2*, *JAK2*, *MET*, *PRKCA* and *TLR4* are members of the PI3K-Akt signalling pathway; *DCK* and *NT4E/CD73* regulate nucleotide metabolism; *GPRC5A*, *ESR1*, *LRG5*, *PTPRG* and *STK32C* encode signalling proteins involved in multiple cellular processes; *PPIL1*, *RIOK1*, and *UPF2* mediate mRNA modification and degradation; and *ACO1*, *AKR1C3*, *GSTA4*, *LPIN3* and *UAP1* participate in oncogenic metabolism rewiring. Additionally, *TPT1* regulates cellular growth and proliferation, whereas *IGHMBP2* encodes a member of the helicase superfamily that binds to a specific DNA sequence from the immunoglobulin mu chain switch region.

3.8 Differential distribution of PTVs discovered in burden test and association with patient survival

PTVs affecting *TPT1*, *PNMT* and *IGHMBP2* were significantly enriched in patients affected by lymphomas compared to CLL cases (q value < 0.05, fisher test), whereas a tendency for an enrichment of *AKR1C3* variants in the CLL patients

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

was observed (q-value 0.05, fisher test, **Supplementary Table 11**). Interestingly, PTVs in *AKR1C3*, *ANGPT2*, *DCG*, *FBXO44*, *GSTA4*, *MET*, *PRKCA*, and *UPF2* were exclusively observed in patients affected by CLL.

None of these events were significantly associated with overall survival (q-value <0.05, cox regression), but we identified a trend towards significance between PTVs affecting *ANGPT2* and shorter time to first treatment (p-value 5.08 x 10^{-3}, q-value 7.62 x 10-2, HR [1.47, 8.74], age and MBL/CLL status-adjusted p-value 0.01, HR [1.31, 7.78]; cox regression).

3.9. Figures, Tables and Schemes

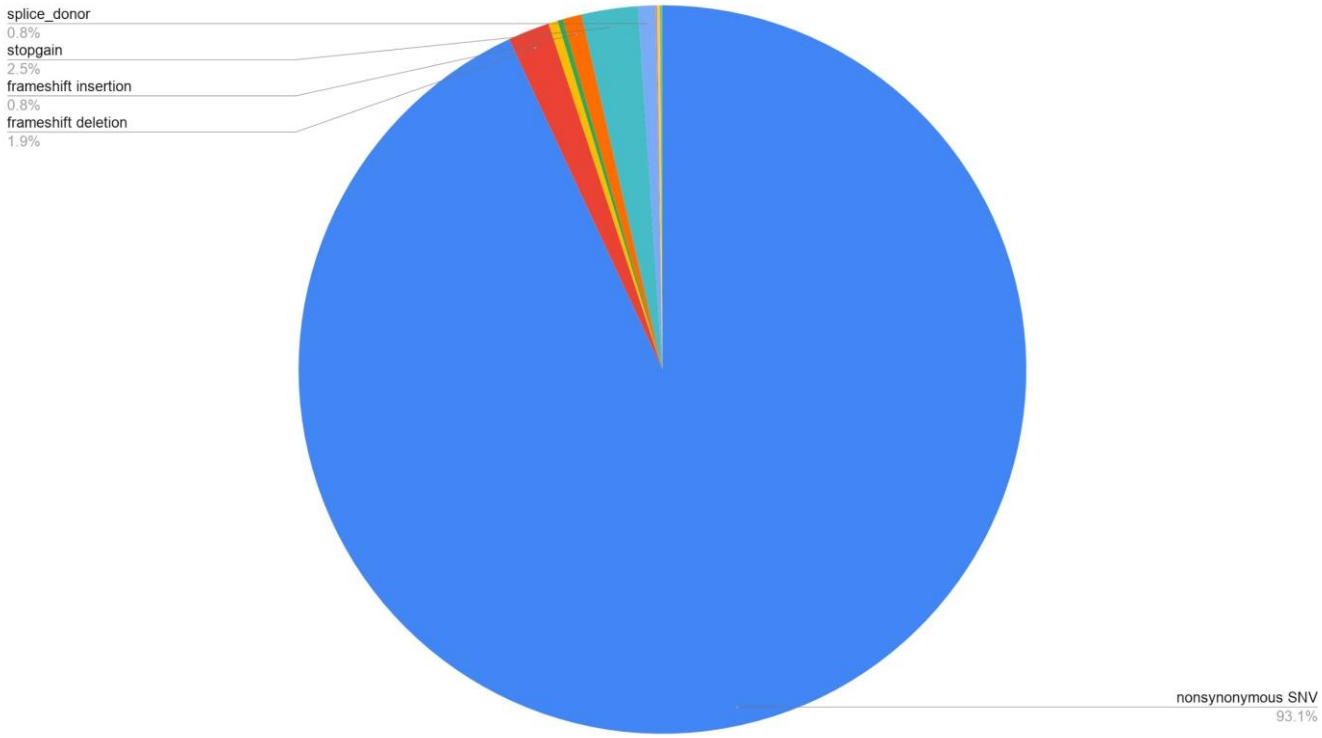

Figure 1. Distribution of the selected variant types present in the cohort.

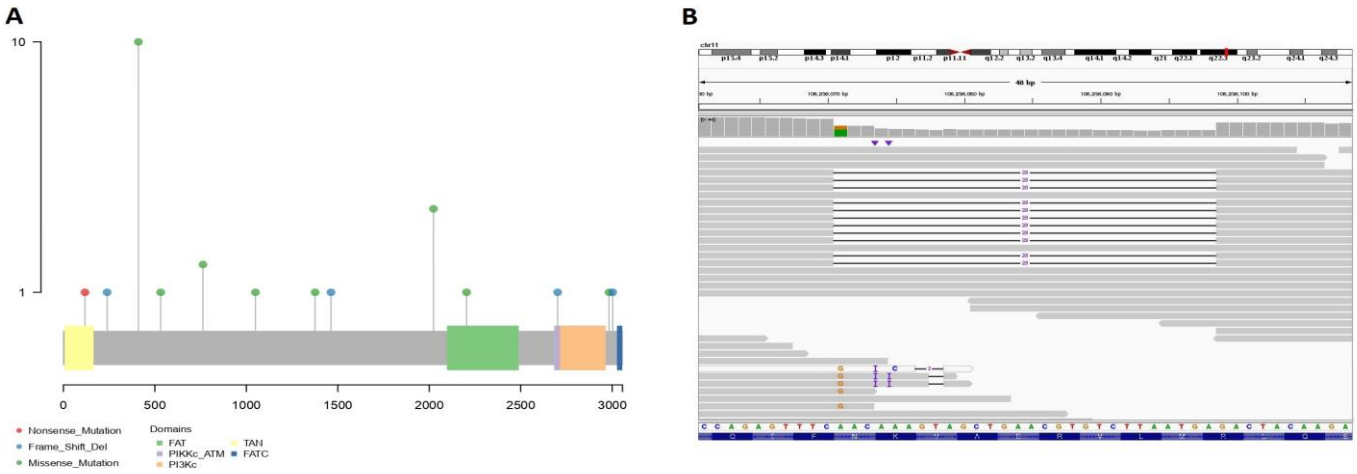

Figure 2. A) Lollipop plot of the rare and predictively disruptive germline variants detected in the *ATM* gene. **B)** Representation of a 28bp frameshift deletion present in one patient.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

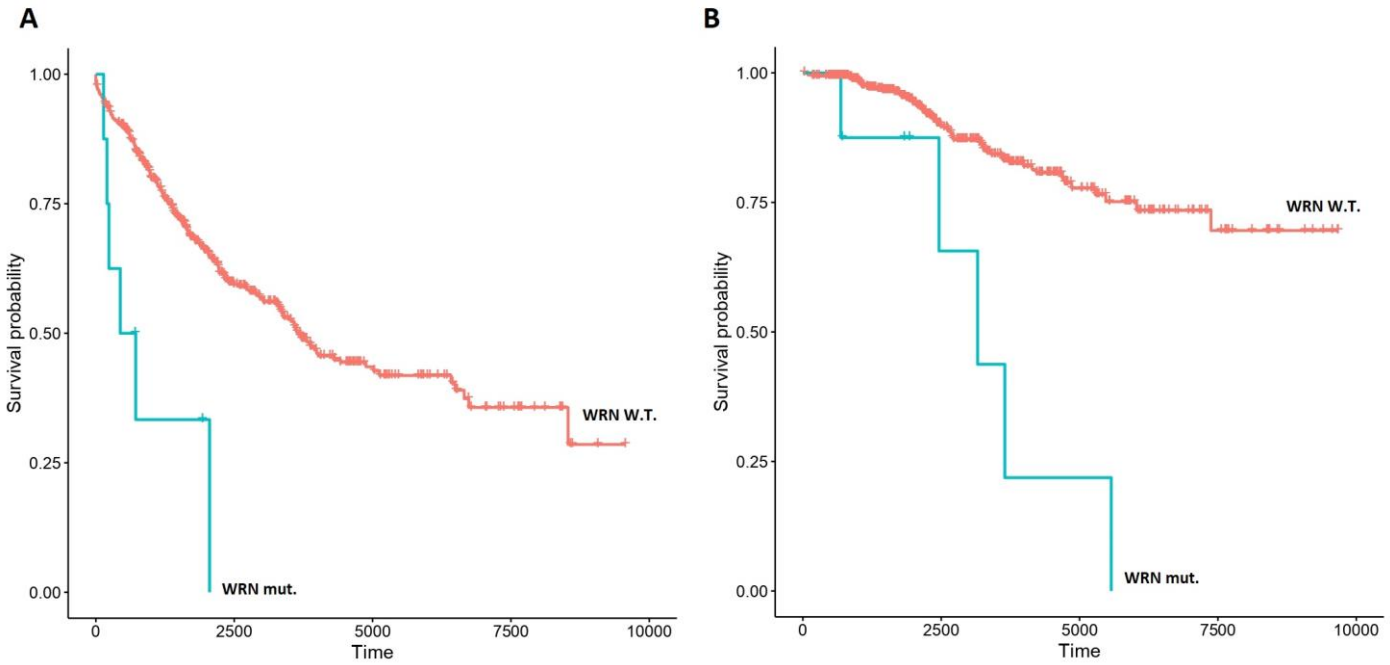

Figure 3. Kaplan-Meyer plots representing the association of rare variants in *WRN* with time to first treatment (A) and overall survival (B) in CLL.

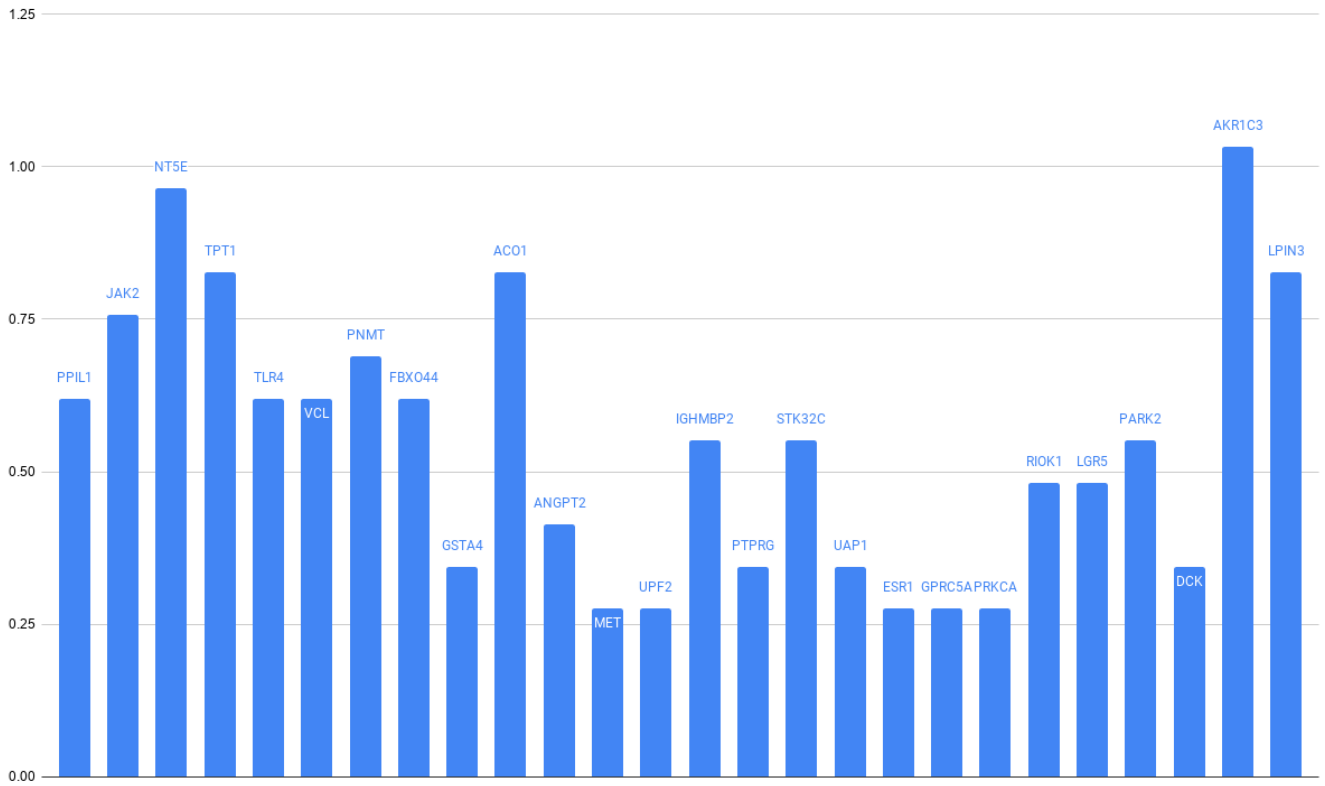

Figure 4. Frequency of PTVs within genes identified in the burden test.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

Table 1. List of all homozygous and compound heterozygous rare and putatively functional variants detected across 726 patients with B-cell lymphoid neoplasms.

Case ID	Zygosity	Gene	Variant Type
4126692	Homozygote	*AKAP9*	missense
147	Compound Heterozygote	*ARID1B*	missense
1416	Compound Heterozygote	*ATM*	missense & frameshift deletion
1196	Compound Heterozygote	*CBFA2T3*	missense
4167381	Compound Heterozygote	*EPPK1*	missense
1298	Homozygote	*ERBB3*	missense
544	Compound Heterozygote	*FAT1*	missense
1078	Compound Heterozygote	*FAT1*	missense
1309	Homozygote	*GLI1*	nonsense
4190784	Compound Heterozygote	*GLI1*	missense
1260	Compound Heterozygote	*IL6ST*	missense
308	Homozygote	*MYH9*	missense
1565	Compound Heterozygote	*MYH9*	missense
325	Homozygote	*MYO5A*	missense
772	Compound Heterozygote	*NCOR1*	missense
4115001	Homozygote	*NTRK3*	missense
1568	Homozygote	*PDZRN3*	missense
4159421	Compound Heterozygote	*PIM1*	missense
284	Compound Heterozygote	*RNF213*	missense
7	Homozygote	*SFRP4*	missense
1437	Homozygote	*SYNE1*	missense
396	Homozygote	*TSC1*	missense
63	Compound Heterozygote	*WRN*	missense
1052;1295	Homozygote	*ZCCHC8*	missense
757	Compound Heterozygote	*ZFHX3*	missense
1191	Compound Heterozygote	*ZFHX3*	missense

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

Table 2. List of the most recurrently affected genes by rare and predictively disruptive germline variants. Genes in the top 0.5% of the FLAG database are indicated.

Gene	N. cases	FLAG Top 100
FAT3	32	Yes
SYNE1	31	Yes
FAT1	29	Yes
ATM	25	No
BIRC6	24	No
CLTCL1	24	No
ZFHX3	23	Yes
TSC2	22	No
CSMD3	21	No
MYH9	20	No
TPR	19	No
LRP1B	19	Yes
ANK1	19	No
EPPK1	19	Yes
KMT2D	19	Yes
PTPN13	18	No
MYH11	18	No
POLQ	17	No
KMT2C	17	Yes
MUC4	16	No
APC	16	No
ROS1	16	No
POLE	16	No
NBEA	16	No
SPEN	15	No
SETDB1	15	No
LPP	15	No
FAT4	15	Yes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

Table 3. Cases of co-occurring somatic mutations and rare germline variants in the same gene. Marked with an asterisk is an event where a rare and likely disruptive germline variant in *ATM* coexisted with a LOH at 11q that deleted the wild-type allele.

Gene	Case ID	Diagnosis
*ATM**	155	CLL
CSMD3	4111337	Follicular Lymphoma
EP300	4122063	DLBCL
FAT1	4136095	Follicular Lymphoma
GNA13	381	CLL
HIST1H1E	4144951	Follicular Lymphoma
KMT2D	372	CLL
KMT2D	4175941	DLBCL
LRP1B	122	CLL
MCL1	4159421	Follicular Lymphoma
MSH6	4109808	DLBCL
MUC16	1267	CLL
MYC	4107559	DLBCL
PIM1	4102009	DLBCL
RNF213	4109808	DLBCL
SIN3A	4139696	Follicular Lymphoma
SPEN	830	CLL

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

Table 4. Results of the burden test performed with PTVs.

GENE	CASE ALLELE COUNTS	CONTROL ALLELE COUNTS	P-VAL	FDR
PPIL1	9	0	6.11E-13	1.49E-09
JAK2	11	4	1.35E-12	3.01E-09
NT5E	14	22	1.46E-10	3.02E-07
TPT1	12	16	7.91E-09	1.52E-05
TLR4	9	8	1.08E-08	1.93E-05
VCL	9	12	1.11E-07	0.0001
PNMT	10	19	2.48E-07	0.0003
FBXO44	9	14	2.85E-07	0.0004
GSTA4	5	1	9.60E-07	0.0012
ACO1	12	33	2.34E-06	0.0028
ANGPT2	6	5	2.78E-06	0.0032
MET	4	0	3.78E-06	0.0040
UPF2	4	0	3.78E-06	0.0040
IGHMBP2	8	16	5.47E-06	0.0056
PTPRG	5	3	8.32E-06	0.0082
STK32C	8	19	1.47E-05	0.0135
UAP1	5	4	1.80E-05	0.0143
ESR1	4	1	1.82E-05	0.0143
GPRC5A	4	1	1.82E-05	0.0143
PRKCA	4	1	1.82E-05	0.0143
RIOK1	7	16	4.20E-05	0.0322
LGR5	7	17	5.71E-05	0.0413
PARK2	8	24	5.71E-05	0.0413
DCK	5	6	6.14E-05	0.0433
AKR1C3	15	92	7.07E-05	0.0486
LPIN3	12	63	0.0001	0.0683

4. Discussion

Approximately 8% of cancer patients are affected by pathogenic germline variants which confer a strong hereditary component [46]. Interestingly, growing evidence indicates that such variants can modulate cancer evolution and prognosis. For example, truncating variants in genes of the angiogenesis and DNA repair pathways predispose to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0330.v1

development of metastatic disease in prostate cancer [15]. Therefore, we reasoned that the analysis of such variants in patients affected by B-cell lymphoid neoplasms could shed new clues about their pathogenesis and prognostication. Indeed, our results indicate an increased frequency of protein truncating rare variants in 26 genes. Notably, most of these are clearly vinculated to oncogenesis. For example, the list included 5 members of the oncogenic PI3K-Akt pathway [47], such as the oncogenes *JAK2* [48] and *MET* [49]; and the TP53-regulator *TPT1*, which promotes p53 degradation in a MDM2-dependent manner [50]. Another group of affected genes play a role in the metabolic rewiring associated with oncogenesis, such as enzyme aconitase 1 (ACO1), an enzyme that participates in the tricarboxylic acid cycle upstream of IDH [51]. Additionally, we detected a significant number of events in genes that regulate DNA metabolism, such as the oncogene ecto-5'-nucleotidase (*NT5E*/*CD73*) that catalyzes AMP breakdown to adenosine [52] and the gene deoxycytidine kinase (*DCK*) that is required for the phosphorylation of several deoxyribonucleosides and their nucleoside analogs [53]. Curiously, some of these disruptive variants showed differential distribution between CLL and B-cell lymphoma patients. The most significantly enriched genes in B-cell lymphoma patients were *IGHMBP2* (a helicase gene implicated in DNA repair [54]) and *PNMT* (an enzyme involved in catecholamine biosynthesis associated with cancer predisposition [55]). On the contrary, truncating mutations in other genes were restricted to CLL patients, such as those of *AKR1C3* (an oncogenic enzyme catalyzing the conversion of aldehydes and ketones [56]) and *FBXO44* (a mediator of BRCA1 proteasomal degradation [57]). Finally, we also detected a trend towards an association of truncating variants in *ANGPT2* with shorter time to first treatment in CLL. Not surprisingly, the expression of this member of the PI3K-Akt pathway has been previously associated with CLL clinical evolution [58]. Altogether, these results support a role for rare germline variants in the pathogenesis of B-cell lymphoid neoplasms, and they also anticipate their importance as drivers of clinical presentation.

In a different approach, we focused our research on the detection of rare and likely disruptive mutations (both PTVs and non-PTVs) in a set of genes involved in cancer pathways, and particularly in lymphoid neoplasms. The collective high frequency of these rare germline variants in cancer genes supposes a challenge for personalized genomics, as many of these are probably non-functional whereas others play a pathogenic or prognostic role. We identified recurrent highly pathogenic variants affecting important drivers of hematological cancer (*ATM* [59]), epigenetic regulators (*ISX* [60] & *SETDB1* [61]) and mediators of DNA replication (*POLQ* [62]). Recurrent variants were also observed in drug targets, and particularly in the crizotinib targets *ALK*, *MET* and *ROS1*, as well as the everolimus target *TSC2*, which suggest new therapeutic strategies for these patients [63-65]. Additionally, several variants were previously catalogued as pathogenic (such as the E318K variant in the transcription factor *MITF* [66]); others affected strong mediators of inherited predisposition to lymphomas (i.e., *DOCK8*, *EXT1*, *MSH6* and *SOS1* [67-70]); and others have been flagged as pathogenic somatic mutations in cancer, such as *NOTCH1* R912W [71, 72]; and *CNOT3* E20K [73, 74]. Importantly, we observed that variants in the DNA helicase *WRN* were significantly associated with shorter overall survival and time to first treatment in CLL. *WRN* mutated CLL cases tended to harbor high-risk karyotypic anomalies, suggesting an increased genomic instability [75] mediated by altered DNA repair mechanisms [76].

Germline-germline or germline-somatic *"double-hit"* events were identified in cancer driver genes. Germline-germline *"double-hit"* events were detected in 28 cases (3.85% of cases), and curiously 6 genes were affected in more than one patient, including the Hedgehog signalling gene *GLI1* [77] and the homeobox tumor suppressor *ZFHX3* [78]. Additionally, germline-somatic *"double hit"* events occurred in 17 cases (2.34%). Notably, this phenomenon affected common drivers of lymphomagenesis such as the oncogenes *MYC* and *PIM1* [79, 80], the tumor suppressors *ATM*, *FAT1*, *KMT2D* and *MSH6* [81-83], the histone acetyltransferase *EP300* [84], the histone gene *HIST1H1E* [85], the transcriptional regulator *SIN3A* [86], the NOTCH pathway member *SPEN* [87] and the apoptotic proteins *GNA13* and *MCL1* [88, 89].

This study has several limitations. First, some background heterogeneity could exist between Spanish CLL and German lymphoma populations, although we believe this should be minimal. Secondly, many relevant oncogenes and tumor suppressors were very rarely mutated, and the interpretation of these variants in terms of survival will need the sequencing of thousands of cases. Additionally, the presence of mosaic somatic mutations in the controls due to clonal hematopoiesis could have led to some false positives. In this line, we observed that only a minority of variants in genes associated with CHIP were likely somatic events, but nevertheless our results should be taken with caution among this group of genes. Finally, another limitation arises from the heterogeneity and limited sample size of the B-cell lymphoma dataset, which dissuaded us from making a survival analysis in such cases.

5. Conclusions

Our results indicate the existence of multiple genes affected by highly pathogenic germline variants in the genomes of patients with B-cell neoplasms, including a significant enrichment of 26 genes in protein-truncating variants. Additionally, the differential distribution of some of these variants suggests a contribution to the phenotypic variability of B-cell neoplasms. Furthermore, the association of some variants with shorter survival, along with the disruptive nature of some others, point towards new functional, prognostic and therapeutic implications. Finally, the elevated number of rare and likely pathogenic variants in cancer genes supposes a challenge for personalized genomics, and future analysis integrating more layers of biological information and other types of cancers are envisaged in order to clarify their benign or pathogenic role.

Supplementary Materials

Supplementary Figure 1. Quantile-quantile plot of the burden test.

Supplementary Table 1. Annotated list of all rare and putatively functional variants detected across 726 patients with B-cell lymphoid neoplasms.

Supplementary Table 2. Myeloid CHIP-related genes affected by rare variants in this study. The number of known variants and the number of likely somatic mosaic events detected is indicated in the corresponding columns.

Supplementary Table 3. Annotated list of all rare and putatively functional protein-truncating variants across 726 patients with B-cell lymphoid neoplasms.

Supplementary Table 4. List of all filtered rare germline events overlapping known somatic mutations in the COSMIC database.

Supplementary Table 5. List of all rare and putatively functional variants in known drivers of B-cell lymphoid tumors detected across 726 patients with B-cell lymphoid neoplasms.

Supplementary Table 6. List of all filtered rare germline events included in the ClinVar database.

Supplementary Table 7. List of all rare and putatively functional variants in genes linked to syndromic cancer detected across 726 patients with B-cell lymphoid neoplasms.

Supplementary Table 8. List of all rare and putatively functional variants in genes of the TARGET database across 726 patients with B-cell lymphoid neoplasms.

Supplementary Table 9. List of all rare and putatively functional variants in genes of the Cancer Gene Census database across 726 patients with B-cell lymphoid neoplasms.

Supplementary Table 10. PTVs that affect all genes significantly enriched in patients with B-cell lymphoid neoplasms.

Supplementary Table 11. Differential distribution of PTVs in genes identified through burden test between CLL and non-CLL cases.

Author Contributions: A.M.O. had the idea and performed the research. A.M.O., M.C.L., A.P.R., J.A.D.A., L.B.P., M.S.G.P. and N.A.V. wrote the paper. A.B.L., A.A.B., P.M.V., R.F.F. and C.A.S. evaluated and suggested changes to the manuscript. M.F.F.R., M.S.G.P., M.M.P.E. and J.L.B.L. approved the final version of the manuscript. All authors have read and agreed to the published version of the manuscript."

Funding: "This research received no external funding. Article processing charges will be paid with funding from the *Fundación Galega de Hematoloxía e Hemoterapia*".

Institutional Review Board Statement: "Not applicable."

Informed Consent Statement: "Not applicable."

Data Availability Statement: Data for this study was downloaded from the International Cancer Genome Consortium repository.

Acknowledgments: We'd like to thank the *Supercomputing Center of Galicia* (CESGA), the *International Cancer Genome Consortium* and the *Fundación Galega de Hematoloxía e Hemoterapia* for their support in this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Teras LR et al. 2016 US lymphoid malignancy statistics by World Health Organization subtypes. CA Cancer J Clin. 2016 Nov 12;66(6):443-459. doi: 10.3322/caac.21357. Epub 2016 Sep 12. PubMed PMID: 27618563.

2. Slager SL et al.. Genetic susceptibility to chronic lymphocytic leukemia. *Semin Hematol*. 2013;50(4):296-302. doi:10.1053/j.seminhematol.2013.09.007

3. Law PJ et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. Nat Commun. 2017 Feb 6;8:14175. doi: 10.1038/ncomms14175. PubMed PMID: 28165464; PubMed Central PMCID: PMC5303820.

4. Kleinstern G et al. Inherited variants at 3q13.33 and 3p24.1 are associated with risk of diffuse large B-cell lymphoma and implicate immune pathways. Hum Mol Genet. 2020 Jan 1;29(1):70-79. doi: 10.1093/hmg/ddz228. PubMed PMID: 31600786; PubMed Central PMCID: PMC7001601.

5. Skibola CF et al. Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. Am J Hum Genet. 2014 Oct 2;95(4):462-71. doi: 10.1016/j.ajhg.2014.09.004. PubMed PMID: 25279986; PubMed Central PMCID: PMC4185120.

6. Bashash M et al. Genetic polymorphism at BCL2 as a predictor for rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone efficacy in patients with diffuse large B-cell lymphoma. Haematologica. 2017 May;102(5):e199-e202. doi: 10.3324/haematol.2016.159087. Epub 2017 Feb 2. PubMed PMID: 28154089; PubMed Central PMCID: PMC5477624.

7. Bielska M et al. Polymorphism in IKZF1 gene affects clinical outcome in diffuse large B-cell lymphoma. Int J Hematol. 2017 Dec;106(6):794-800. doi: 10.1007/s12185-017-2315-0. Epub 2017 Sep 6. PubMed PMID: 28879630.

8. Ghesquieres H et al. Genome-Wide Association Study of Event-Free Survival in Diffuse Large B-Cell Lymphoma Treated With Immunochemotherapy. J Clin Oncol. 2015 Nov 20;33(33):3930-7. doi: 10.1200/JCO.2014.60.2573. Epub 2015 Oct 12. PubMed PMID: 26460308; PubMed Central PMCID: PMC4879713.

9. Mosquera Orgueira A et al. The association of germline variants with chronic lymphocytic leukemia outcome suggests the implication of novel genes and pathways in clinical evolution. BMC Cancer. 2019 May 29;19(1):515. doi: 10.1186/s12885-019-5628-y. PubMed PMID: 31142279; PubMed Central PMCID: PMC6542042.

10. Musa J et al. Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes. Nat Commun. 2019 Sep 11;10(1):4128. doi: 10.1038/s41467-019-12071-2. PubMed PMID: 31511524; PubMed Central PMCID: PMC6739408.

11. Lu C et al. Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun. 2015 Dec 22;6:10086. doi: 10.1038/ncomms10086. PubMed PMID: 26689913; PubMed Central PMCID: PMC4703835.

12. Brown AL et al. RUNX1-mutated families show phenotype heterogeneity and a somatic mutation profile unique to germline predisposed AML. Blood Adv. 2020 Mar 24;4(6):1131-1144. doi: 10.1182/bloodadvances.2019000901. PubMed PMID: 32208489; PubMed Central PMCID: PMC7094007.

13. Park S et al. Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. Nat Commun. 2018 Jul 4;9(1):2601. doi: 10.1038/s41467-018-04900-7. PubMed PMID: 29973584; PubMed Central PMCID: PMC6031629.

14. Pui CH et al. Somatic and germline genomics in paediatric acute lymphoblastic leukaemia. Nat Rev Clin Oncol. 2019 Apr;16(4):227-240. doi: 10.1038/s41571-018-0136-6. Review. PubMed PMID: 30546053.

15. Mijuskovic M et al. Rare germline variants in DNA repair genes and the angiogenesis pathway predispose prostate cancer patients to develop metastatic disease. Br J Cancer. 2018 Jul;119(1):96-104. doi: 10.1038/s41416-018-0141-7. Epub 2018 Jun 19. Erratum in: Br J Cancer. 2019 Apr;120(8):867. PubMed PMID: 29915322; PubMed Central PMCID: PMC6035259.

16. Goldstein AM et al. Rare germline variants in known melanoma susceptibility genes in familial melanoma. Hum Mol Genet. 2017 Dec 15;26(24):4886-4895. doi: 10.1093/hmg/ddx368. PubMed PMID: 29036293; PubMed Central PMCID: PMC5886297.

17. Kleiblova P et al. Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. Int J Cancer. 2019 Oct 1;145(7):1782-1797. doi: 10.1002/ijc.32385. Epub 2019 May 20. PubMed PMID: 31050813.

18. Tiao G et al. Rare germline variants in ATM are associated with chronic lymphocytic leukemia. Leukemia. 2017;31(10):2244-2247. doi:10.1038/leu.2017.201

19. McMaster ML et al. Germline mutations in Protection of Telomeres 1 in two families with Hodgkin lymphoma. *Br J Haematol*. 2018;181(3):372-377. doi:10.1111/bjh.15203

20. Offit K et al. Rare variants of ATM and risk for Hodgkin's disease and radiation-associated breast cancers. *Clin Cancer Res*. 2002;8(12):3813-3819.

21. Srivastava A et al. Identification of Familial Hodgkin Lymphoma Predisposing Genes Using Whole Genome Sequencing. *Front Bioeng Biotechnol*. 2020;8:179. Published 2020 Mar 6. doi:10.3389/fbioe.2020.00179

22. International Cancer Genome Consortium et al. International network of cancer genome projects. Nature. 2010 Apr 15;464(7291):993-8. doi: 10.1038/nature08987. Erratum in: Nature. 2010 Jun 17;465(7300):966. Himmelbaue, Heinz [corrected to Himmelbauer, Heinz]; Gardiner, Brooke A [corrected to Gardiner, Brooke B]; Cross, Anthony [corrected to Cros, Anthony]. PubMed PMID: 20393554; PubMed Central PMCID: PMC2902243.

23. Valls-Guimera, R. Bcbio-nextgen: Automated, distributed, next-gen sequencing pipeline. EMBnet J. 17, 30, https://doi.org/10.14806/ej.17.B.286 (2012).

24. Erik Garrison, Gabor Marth. Haplotype-based variant detection from short-read sequencing. 2012, arXiv:1207.3907

25. Poplin R et al. Scaling accurate genetic variant discovery to tens of thousands of samples. 2017. bioRxiv

26. Rimmer A et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014 Aug;46(8):912-918. doi: 10.1038/ng.3036. Epub 2014 Jul 13. PubMed PMID: 25017105; PubMed Central PMCID: PMC4753679.

27. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PubMed PMID: 19505943; PubMed Central PMCID: PMC2723002.

28. Sherry ST et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308-11. PubMed PMID: 11125122; PubMed Central PMCID: PMC29783.

29. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central PMCID: PMC4750478.

30. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285-91. doi: 10.1038/nature19057. PubMed PMID: 27535533; PubMed Central PMCID: PMC5018207.

31. Karczewski KJ et al. Genome Aggregation Database (gnomAD) Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. In bioRxiv, 2019 (p. 531210). https://doi.org/10.1101/531210

32. Rentzsch P et al. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019 Jan 8;47(D1):D886-D894. doi: 10.1093/nar/gky1016. PubMed PMID: 30371827; PubMed Central PMCID: PMC6323892.

33. Gröbner SN et al. The landscape of genomic alterations across childhood cancers. Nature. 2018 Mar 15;555(7696):321-327. doi: 10.1038/nature25480. Epub 2018 Feb 28. Erratum in: Nature. 2018 Jul;559(7714):E10. PubMed PMID: 29489754.

34. Sondka Z et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018 Nov;18(11):696-705. doi: 10.1038/s41568-018-0060-1. Review. PubMed PMID: 30293088; PubMed Central PMCID: PMC6450507.

35. Van Allen EM et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat Med. 2014 Jun;20(6):682-8. doi: 10.1038/nm.3559. Epub 2014 May 18. PubMed PMID: 24836576; PubMed Central PMCID: PMC4048335.

36. Landau DA et al. Mutations driving CLL and their evolution in progression and relapse. Nature. 2015 Oct 22;526(7574):525-30. doi: 10.1038/nature15395. Epub 2015 Oct 14. PubMed PMID: 26466571; PubMed Central PMCID: PMC4815041.

37. Puente XS et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015 Oct 22;526(7574):519-24. doi: 10.1038/nature14666. Epub 2015 Jul 22. PubMed PMID: 26200345.

38. Reddy A et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. Cell. 2017 Oct 5;171(2):481-494.e15. doi: 10.1016/j.cell.2017.09.027. PubMed PMID: 28985567; PubMed Central PMCID: PMC5659841.

39. Panea RI et al. The whole-genome landscape of Burkitt lymphoma subtypes. Blood. 2019 Nov 7;134(19):1598-1607. doi: 10.1182/blood.2019001880. PubMed PMID: 31558468; PubMed Central PMCID: PMC6871305.

40. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet. 2017 Mar 2;100(3):406-413. doi: 10.1016/j.ajhg.2017.01.017. Epub 2017 Feb 9. PubMed PMID: 28190455; PubMed Central PMCID: PMC5339084.

41. Shyr C et al. FLAGS, frequently mutated genes in public exomes. BMC Med Genomics. 2014 Dec 3;7:64. doi: 10.1186/s12920-014-0064-y. Erratum in: BMC Med Genomics. 2017 Nov 29;10 (1):69. PubMed PMID: 25466818; PubMed Central PMCID: PMC4267152.

42. Guo MH et al. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. Am J Hum Genet. 2018 Oct 4;103(4):522-534. doi: 10.1016/j.ajhg.2018.08.016. Epub 2018 Sep 27. PMID: 30269813; PMCID: PMC6174288.

43. Steensma DP et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015 Jul 2;126(1):9-16. doi: 10.1182/blood-2015-03-631747. Epub 2015 Apr 30. PMID: 25931582; PMCID: PMC4624443.

44. Zink F et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017 Aug 10;130(6):742-752. doi: 10.1182/blood-2017-02-769869. Epub 2017 May 8. PMID: 28483762; PMCID: PMC5553576.

45. Gibson CJ et al. Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. J Clin Oncol. 2017 May 10;35(14):1598-1605. doi: 10.1200/JCO.2016.71.6712. Epub 2017 Jan 9. PMID: 28068180; PMCID: PMC5455707.

46. Huang KL et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 2018 Apr 5;173(2):355-370.e14. doi: 10.1016/j.cell.2018.03.039. PubMed PMID: 29625052; PubMed Central PMCID: PMC5949147.

47. Martini M et al. PI3K/AKT signaling pathway and cancer: an updated review. Ann Med. 2014 Sep;46(6):372-83. doi: 10.3109/07853890.2014.912836. Epub 2014 Jun 5. PMID: 24897931.

48. Derenzini E, Younes A. Targeting the JAK-STAT pathway in lymphoma: a focus on pacritinib. Expert Opin Investig Drugs. 2013 Jun;22(6):775-85. doi: 10.1517/13543784.2013.775244. Epub 2013 Feb 26. PMID: 23442043.

49. Raghav K et al.. Untying the gordion knot of targeting MET in cancer. Cancer Treat Rev. 2018 May;66:95-103. doi: 10.1016/j.ctrv.2018.04.008. Epub 2018 Apr 26. PMID: 29730462.

50. Amson R et al. TPT1/ TCTP-regulated pathways in phenotypic reprogramming. Trends Cell Biol. 2013 Jan;23(1):37-46. doi: 10.1016/j.tcb.2012.10.002. Epub 2012 Oct 30. PMID: 23122550.

51. Yamamoto M et al. Iron-induced epigenetic abnormalities of mouse bone marrow through aberrant activation of aconitase and isocitrate dehydrogenase. Int J Hematol. 2016 Oct;104(4):491-501. doi: 10.1007/s12185-016-2054-7. Epub 2016 Jul 5. PMID: 27380194.

52. Gao ZW et al. The roles of CD73 in cancer. Biomed Res Int. 2014;2014:460654. doi: 10.1155/2014/460654. Epub 2014 Jul 14. PMID: 25126561; PMCID: PMC4121992.

53. Yang C et al. Deoxycytidine kinase regulates the G2/M checkpoint through interaction with cyclin-dependent kinase 1 in response to DNA damage. Nucleic Acids Res. 2012 Oct;40(19):9621-32. doi: 10.1093/nar/gks707. Epub 2012 Jul 31. PMID: 22850745; PMCID: PMC3479177.

54. Chunli W et al. IGHMBP2 overexpression promotes cell migration and invasion in esophageal squamous carcinoma. Yi Chuan. 2015 Apr;37(4):360-6. Chinese. doi: 10.16288/j.yczz.14-371. PMID: 25881701.

55. Zhong J et al. A Transcriptome-Wide Association Study Identifies Novel Candidate Susceptibility Genes for Pancreatic Cancer. J Natl Cancer Inst. 2020 Oct 1;112(10):1003-1012. doi: 10.1093/jnci/djz246. PMID: 31917448; PMCID: PMC7566474.

56. Byrns MC et al. Inhibitors of type 5 17β-hydroxysteroid dehydrogenase (AKR1C3): overview and structural insights. J Steroid Biochem Mol Biol. 2011 May;125(1-2):95-104. doi: 10.1016/j.jsbmb.2010.11.004. Epub 2010 Nov 16. PMID: 21087665; PMCID: PMC3047600.

57. Lu Y et al. The F-box protein FBXO44 mediates BRCA1 ubiquitination and degradation. J Biol Chem. 2012 Nov 30;287(49):41014-22. doi: 10.1074/jbc.M112.407106. Epub 2012 Oct 18. PMID: 23086937; PMCID: PMC3510804.

58. Martinelli S et al. ANGPT2 promoter methylation is strongly associated with gene expression and prognosis in chronic lymphocytic leukemia. Epigenetics. 2013 Jul;8(7):720-9. doi: 10.4161/epi.24947. Epub 2013 May 14. PMID: 23803577; PMCID: PMC3781191.

59. Blakemore SJ et al. Clinical significance of TP53, BIRC3, ATM and MAPK-ERK genes in chronic lymphocytic leukaemia: data from the randomised UK LRF CLL4 trial. Leukemia. 2020 Feb 3. doi: 10.1038/s41375-020-0723-2. [Epub ahead of print] PubMed PMID: 32015491.

60. Hsu SH et al.. Proinflammatory homeobox gene, ISX, regulates tumor growth and survival in hepatocellular carcinoma. Cancer Res. 2013 Jan 15;73(2):508-18. doi: 10.1158/0008-5472.CAN-12-2795. Epub 2012 Dec 5. PubMed PMID: 23221382.

61. Ropa J et al. SETDB1 mediated histone H3 lysine 9 methylation suppresses MLL-fusion target expression and leukemic transformation. Haematologica. 2019 Sep 26. pii: haematol.2019.223883. doi: 10.3324/haematol.2019.223883. [Epub ahead of print] PubMed PMID: 31558673.

62. Wang Z et al. DNA polymerase θ (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse. J Biol Chem. 2019 Mar 15;294(11):3909-3919. doi: 10.1074/jbc.RA118.005188. Epub 2019 Jan 17. PubMed PMID: 30655289; PubMed Central PMCID: PMC6422074.

63. Shaw AT et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med. 2014 Nov 20;371(21):1963-71. doi: 10.1056/NEJMoa1406766. Epub 2014 Sep 27. PubMed PMID: 25264305; PubMed Central PMCID: PMC4264527.

64. Perini GF et al. Clinical response to everolimus in a patient with Hodgkin's lymphoma harboring a TSC2 mutation. Blood Cancer J. 2016 May 13;6:e420. doi: 10.1038/bcj.2016.25. PubMed PMID: 27176796; PubMed Central PMCID: PMC4916294.

65. Maroto P et al. Biallelic TSC2 Mutations in a Patient With Chromophobe Renal Cell Carcinoma Showing Extraordinary Response to Temsirolimus. J Natl Compr Canc Netw. 2018 Apr;16(4):352-358. doi: 10.6004/jnccn.2017.7041. PubMed PMID: 29632054.

66. Yokoyama S et al. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. Nature. 2011 Nov 13;480(7375):99-103. doi: 10.1038/nature10630. PubMed PMID: 22080950; PubMed Central PMCID: PMC3266855.

67. Buchbinder D et al. J Clin Immunol. 2019 Aug;39(6):592-595. doi: 10.1007/s10875-019-00663-y. Epub 2019 Jul 2. PubMed PMID: 31267431.

68. Ripperger T, Schlegelberger B. Acute lymphoblastic leukemia and lymphoma in the context of constitutional mismatch repair deficiency syndrome. Eur J Med Genet. 2016 Mar;59(3):133-42. doi: 10.1016/j.ejmg.2015.12.014. Epub 2015 Dec 30. Review. PubMed PMID: 26743104.

69. Cavé H et al. Acute lymphoblastic leukemia in the context of RASopathies. Eur J Med Genet. 2016 Mar;59(3):173-8. doi: 10.1016/j.ejmg.2016.01.003. Epub 2016 Feb 5. PubMed PMID: 26855057.

70. Nakane T et al. Pre-B-cell acute lymphoblastic leukemia in a boy with hereditary multiple exostoses caused by EXT1 deletion. Pediatr Hematol Oncol. 2014 Oct;31(7):667-9. doi: 10.3109/08880018.2014.935538. Epub 2014 Aug 12. PubMed PMID: 25116165.

71. Vollbrecht C et al. Comprehensive Analysis of Disease-Related Genes in Chronic Lymphocytic Leukemia by Multiplex PCR-Based Next Generation Sequencing. PLoS One. 2015 Jun 8;10(6):e0129544. doi: 10.1371/journal.pone.0129544. eCollection 2015. PubMed PMID: 26053404; PubMed Central PMCID: PMC4459702.

72. Neumann M et al. Mutational spectrum of adult T-ALL. Oncotarget. 2015 Feb 20;6(5):2754-66. PubMed PMID: 25595890; PubMed Central PMCID: PMC4413615.

73. Puente XS et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011 Jun 5;475(7354):101-5. doi: 10.1038/nature10113. PubMed PMID: 21642962; PubMed Central PMCID: PMC3322590.

74. Barbieri CE et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012 May 20;44(6):685-9. doi: 10.1038/ng.2279. PubMed PMID: 22610119; PubMed Central PMCID: PMC3673022.

75. Poot M. Genes, Proteins, and Biological Pathways Preventing Chromothripsis. Methods Mol Biol. 2018;1769:231-251. doi: 10.1007/978-1-4939-7780-2_15. Review. PubMed PMID: 29564828.

76. Chen L et al. WRN, the protein deficient in Werner syndrome, plays a critical structural role in optimizing DNA repair. Aging Cell. 2003 Aug;2(4):191-9. PubMed PMID: 12934712.

77. Skoda AM et al. The role of the Hedgehog signaling pathway in cancer: A comprehensive review. Bosn J Basic Med Sci. 2018 Feb 20;18(1):8-20. doi: 10.17305/bjbms.2018.2756. Review. PubMed PMID: 29274272; PubMed Central PMCID: PMC5826678.

78. Walker CJ et al. Patterns of CTCF and ZFHX3 Mutation and Associated Outcomes in Endometrial Cancer. J Natl Cancer Inst. 2015 Sep 1;107(11). pii: djv249. doi: 10.1093/jnci/djv249. Print 2015 Nov. PubMed PMID: 26330387; PubMed Central PMCID: PMC4643635.

79. Ennishi D et al. Genetic profiling of MYC and BCL2 in diffuse large B-cell lymphoma determines cell-of-origin-specific clinical impact. Blood. 2017 May 18;129(20):2760-2770. doi: 10.1182/blood-2016-11-747022. Epub 2017 Mar 28. PubMed PMID: 28351934.

80. Baron BW et al. PIM1 gene cooperates with human BCL6 gene to promote the development of lymphomas. Proc Natl Acad Sci U S A. 2012 Apr 10;109(15):5735-9. doi: 10.1073/pnas.1201168109. Epub 2012 Mar 26. PubMed PMID: 22451912; PubMed Central PMCID: PMC3326467.

81. Ortega-Molina A et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. Nat Med. 2015 Oct;21(10):1199-208. doi: 10.1038/nm.3943. Epub 2015 Sep 14. PubMed PMID: 26366710; PubMed Central PMCID: PMC4676270.

82. Yi D et al. Germline TP53 and MSH6 mutations implicated in sporadic triple-negative breast cancer (TNBC): a preliminary study. Hum Genomics. 2019 Jan 10;13(1):4. doi: 10.1186/s40246-018-0186-y. PubMed PMID: 30630526; PubMed Central PMCID: PMC632751.

83. Laginestra MA et al. Whole exome sequencing reveals mutations in FAT1 tumor suppressor gene clinically impacting on peripheral T-cell lymphoma not otherwise specified. Mod Pathol. 2020 Feb;33(2):179-187. doi: 10.1038/s41379-019-0279-8. Epub 2019 Apr 25. Erratum in: Mod Pathol. 2019 Sep 26;:. PubMed PMID: 31028364; PubMed Central PMCID: PMC6994417.

84. Meyer SN et al. Unique and Shared Epigenetic Programs of the CREBBP and EP300 Acetyltransferases in Germinal Center B Cells Reveal Targetable Dependencies in Lymphoma. Immunity. 2019 Sep 17;51(3):535-547.e9. doi: 10.1016/j.immuni.2019.08.006. Epub 2019 Sep 10. PubMed PMID: 31519498.

85. González-Rincón J et al. Unraveling transformation of follicular lymphoma to diffuse large B-cell lymphoma. PLoS One. 2019 Feb 25;14(2):e0212813. doi: 10.1371/journal.pone.0212813. eCollection 2019. PubMed PMID: 30802265; PubMed Central PMCID: PMC6388933.

86. Panea RI et al. The whole-genome landscape of Burkitt lymphoma subtypes. Blood. 2019 Nov 7;134(19):1598-1607. doi: 10.1182/blood.2019001880. PubMed PMID: 31558468; PubMed Central PMCID: PMC6871305.

87. Rossi D et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med. 2012 Aug 27; 209 (9): 1537-51. doi: 10.1084 / jem.20120904. Epub 2012 Aug 13. PubMed PMID: 22891273; PubMed Central PMCID: PMC3428941.

88. Healy JA et al. GNA13 loss in germinal center B cells leads to impaired apoptosis and promotes lymphoma in vivo. Blood. 2016 Jun 2;127(22):2723-31. doi: 10.1182/blood-2015-07-659938. Epub 2016 Mar 17. PubMed PMID: 26989201; PubMed Central PMCID: PMC4991089.

89. Fletcher S. MCL-1 inhibitors - where are we now (2019)? Expert Opin Ther Pat. 2019 Nov;29(11):909-919. doi: 10.1080/13543776.2019.1672661. Epub 2019 Oct 14. Review. PubMed PMID: 31566022.