PLOS ONE

RESEARCH ARTICLE

Detection of new drivers of frequent B-cell lymphoid neoplasms using an integrated analysis of whole genomes

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

1 Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Galicia, Spain, **2** Department of Hematology, Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), SERGAS, Santiago de Compostela, Galicia, Spain, **3** University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain, **4** Department of Pathology, Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), SERGAS, Santiago de Compostela, Galicia, Spain

* adrian.mosquera@live.com

Citation: Mosquera Orgueira A, Ferreiro Ferro R, Díaz Arias JÁ, Aliste Santos C, Antelo Rodríguez B, Bao Pérez L, et al. (2021) Detection of new drivers of frequent B-cell lymphoid neoplasms using an integrated analysis of whole genomes. PLoS ONE 16(5): e0248886. https://doi.org/10.1371/journal.pone.0248886

Editor: Obul Reddy Bandapalli, German Cancer Research Center (DKFZ), GERMANY

Received: September 22, 2020

Accepted: January 19, 2021

Published: May 4, 2021

Data Availability Statement: All data is accessible from the ICGC repository (https://dcc.icgc.org/releases). CLL data was downloaded from https://dcc.icgc.org/releases/current/Projects/CLLE-ES; and lymphoma data was downloaded from https://dcc.icgc.org/releases/current/Projects/MALY-DE.

Funding: Publication costs related to this manuscript have been covered by a grant provided by the "Fundación Galega de Hematoloxía e Hemoterapia".

Abstract

B-cell lymphoproliferative disorders exhibit a diverse spectrum of diagnostic entities with heterogeneous behaviour. Multiple efforts have focused on the determination of the genomic drivers of B-cell lymphoma subtypes. In the meantime, the aggregation of diverse tumors in pan-cancer genomic studies has become a useful tool to detect new driver genes, while enabling the comparison of mutational patterns across tumors. Here we present an integrated analysis of 354 B-cell lymphoid disorders. 112 recurrently mutated genes were discovered, of which *KMT2D*, *CREBBP*, *IGLL5* and *BCL2* were the most frequent, and 31 genes were putative new drivers. Mutations in *CREBBP*, *TNFRSF14* and *KMT2D* predominated in follicular lymphoma, whereas those in *BTG2*, *HTA-A* and *PIM1* were more frequent in diffuse large B-cell lymphoma. Additionally, we discovered 31 significantly mutated protein networks, reinforcing the role of genes such as *CREBBP*, *EEF1A1*, *STAT6*, *GNA13* and *TP53*, but also pointing towards a myriad of infrequent players in lymphomagenesis. Finally, we report aberrant expression of oncogenes and tumor suppressors associated with novel noncoding mutations (*DTX1* and *S1PR2*), and new recurrent copy number aberrations affecting immune check-point regulators (*CD83*, *PVR*) and B-cell specific genes (*TNFRSF13C*). Our analysis expands the number of mutational drivers of B-cell lymphoid neoplasms, and identifies several differential somatic events between disease subtypes.

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

Competing interests: The authors declare no conflicts of interest.

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].

Next-generation sequencing (NGS) technologies have tried to deconvolute the genomic complexity of lymphoid tumors. This information has led to an improved classification of lymphoid neoplasms, mainly thanks to the characterization of the biological heterogeneity within lymphoma subtypes. A good example is that of the gene expression-based classification of DLBCL in two different clinico-biological groups by its cell-of-origin status: either germinal center B cell-like or activated B cell-like [2]. Various groups have also identified new DLBCL subtypes based on their mutational profiles, also observing a correlation between some of these mutational patterns with cell-of-origin status [3, 4]. In the same line, cumulative evidence indicates that co-occurring mutations are drivers of treatment refractoriness and clonal evolution in follicular lymphoma [5, 6]. In the same line, significant advances in the deconvolution of the genomic landscapes of both CLL and Burkitt lymphoma have been made in the past years [7–11], providing new disease-specific drivers, hypermutation events and predictors of adverse outcome. Some key findings include the predominance of *ID3* mutations in Burkitt lymphoma, but not in other *IGH-MYC* rearranged lymphomas [12], as well as the role of aberrant somatic hypermutation (aSHM) in Epstein-Barr positive Burkitt Lymphomas [11]. Such somatic hypermutation in the *IGHV* locus of CLL tumors is also important, as it defines two important types of leukemia which exhibit broadly different clinical and mutational backgrounds [8]. Additionally, some of the mutational drivers of CLL are also important mediators of drug resistance, such as in the case of rituximab-resistance observed in *NOTCH1*-mutated CLLs [13].

An additional line of complexity is conformed by the limited comprehension of the contribution of regulatory mutations to the pathogenesis of cancer. Existing research points towards the deregulation of important driver genes by noncoding mutations in lymphomas. For example, *Batmanov et al.* (2017) discovered regulatory mutations that control *BCL2* and *BCL6* expression in follicular lymphoma [14]; *Arthur et al.* (2018) identified aberrant expression of *NFKBIZ* in DLBCL caused by functional noncoding mutations in the 3' untranslated region of the gene [15], and *Puente et al.* (2015) characterized enhancer mutations that deregulate *PAX5* expression in CLL [8]. Considering the extensive heterogeneity of these disorders, we anticipate that the analysis of larger and diverse patient cohorts will enable the identification of new regulatory driver regions of B-cell tumors.

Although increasing NGS data in cancer is available, the detection of driver mutations continues to be a bottleneck in the development of this technology. Differences in clonality, sample purity, sequencing coverage and quality are challenging for most variant callers. These are addressed using different methods, leading to remarkable disparities in results between algorithms [16]. *Hoffman et al.* [17] compared 10 variant callers on simulated data, reporting considerable differences in sensitivity and precision depending on coverage and variant allele frequency. Concordantly, *Cai et al.* [18] analyzed a set of cancer samples with four different algorithms and observed that only 20.7% of variants were detected by $\geq$2 callers. Therefore, numerous pathogenic variants in large sequencing projects could have passed unnoticed. Furthermore, cancer genomes suffer from the *"long tail"* phenomenon, whereby a few driver genes are recurrently mutated and most mutations are distributed across a vast number of genes [19]. Both the enhanced statistical power and the capacity to analyze divergent molecular mechanisms across tumor types are the main reasons that motivate the increasingly common aggregation of tumors in pan-cancer genomic studies [20–22].

In this report we present an integrated genomic analysis of diverse mature B-cell lymphoid neoplasms using whole genome sequencing (WGS) data produced by the *International Cancer Genome Consortium* (ICGC) [23]. Our results expand the catalog of B-cell lymphoma driver genes, identify novel putative drivers based on functionally connected subnetworks and characterize new structural aberrations and regulatory mutations that modify the expression of several oncogenes and tumor suppressors.

Methods

1. Data source and analysis

WGS data from the *CLLE-ES* and *MALY-DE* projects produced by the *ICGC* were analyzed. This cohort included 132 CLL, 36 Burkitt lymphoma, 85 DLBCL, 97 follicular lymphoma and 4 unspecified B-cell lymphoma cases. RNAseq expression data for a subgroup of the samples was also available.

Tumor-normal matched whole genomes were processed using the bcbio-nextgen pipeline, which provides best practices for NGS data analysis [24]. GRCh37 was used as the reference genome. Low complexity regions, areas with abnormally high coverage, sequences with single nucleotide stretches >50bp and loci with alternative or unplaced contigs in the reference genome were not analyzed. Some polymorphic regions in noncoding regions are prone to be classified as mutation hotspots due to artifacts or biases in the sequencing process (mainly in low coverage regions), and suspicious elements were manually discarded from downstream analysis. Single nucleotide and indel mutation detection was performed with *vardict-java* version 1.5.8 [25], *varscan* version 2.4.3 [26], *mutect2* implemented in GATK version 3.8 [27] and *freebayes* version 1.1.0.46 [28] using default bcbio-nextgen parameters. Events with a minimum sequencing depth (DP) of 10 and a genotype quality (GQ) of 20 Phred in both tumor and normal samples were selected. A mutation was called when detected by ≥2 callers. Mutations were annotated to the *1000G* [29], *gnomAD* [30] and *ExAc* [31] databases. Tumor mutations reported as polymorphisms with a minimum allele frequency > 0.001 in any population were discarded. For copy number aberration (CNA) detections, the *CNVkit* version 0.9.6a0 [32] algorithm was used (ploidy-adjusted and with default parameters). We initially used the circular binary segmentation algorithm, observing important hypersegmentation that could lead to an increase in false positives. Therefore, we finally used the *HaarSeg* segmentation method, retrieving a vast majority of cases with segment counts in the range of 100–200. Events detected in centromeric and telomeric regions were discarded. Similarly, we also removed events within low 100bp-read mappability regions according to *UCSC* tracks.

2. Detection of mutation drivers in coding regions

Three methods were used to detect driver genes using nonsynonymous coding mutation data: *MutSigCV* [33], *dNdScv* [34] and *OncodriveFML* [35]. MutSigCV version 1.3.5 and dNdSCV were run with default parameters. OncodriveFML was run using CADD 1.3 scores. Significance threshold was set at FDR of 10% for all methods.

Hierarchical HotNet [36] was used to infer networks of functionally connected mutated genes. The following protein-protein interaction networks were used: *Hint+Hi2012*, *Irefindex9* and *Multinet*. Mutation frequency and log-transformed *MutSigCV* p-values were used as input scores. Heat scores were permuted 100 times for each network. Hierarchies were constructed and processed with default parameters. The deviation of observed dendrogram distribution from the random expectation at different similarity thresholds was calculated, and significance threshold was set to p-value <0.05. Finally, a consensus network (G_2) was created from the resulting significant subnetworks.

3. Noncoding region annotation and mutation enrichment analysis

Annotations corresponding to promoter regions, 5'UTR, 3'UTR and lincRNAs were retrieved from *Genecode* version 18 [37]. Enhancer regions were obtained from the *GeneHancer* database [38], and those supported by two or more sources of evidence ("elite" enhancers) were selected. Transcription start sites (TSS) were defined as the 100bp-region upstream of the point of transcription. Regulatory regions within telomeric and centromeric positions were discarded.

LARVA [39] was used to identify areas with evidence of positive selection of mutations. *LARVA* models the mutation counts of each target region as a β-binomial distribution in order to handle overdispersion. *LARVA* also includes replication timing information in order to estimate local mutation rate, and provides a β-binomial distribution adjusted for replication timing which is used to compute p-values. Significance threshold was set to FDR<10%. As we used *LARVA* including tumor classification data, the mutation background estimation was calculated for each tumor subtype.

Regions targeted by aSHM were retrieved from literature analysis [40–42], and were used to annotate the list of significantly mutated noncoding regions identified by *LARVA*. In the case of genes without annotation, we used Signal [43] to test for local enrichments in the mutational signature of aSHM.

4. Recurrent focal CNA detection

Gistic2.0 [44] was used to identify recurrent CNA. Focal CNA were defined as those spanning a maximum of 25% of an arm's length. Deletions were called in regions with tumor/normal log ratios < -0.3, and amplifications in regions with ratios >0.3. Evenly spaced pseudomarkers were automatically created by the algorithm, and regions were considered only if they spanned 10 or more pseudomarkers. Sex chromosomes were not analyzed. Arm-level peel off was enabled, and residual q-values were calculated after removing segments shared with higher peaks. Significance threshold was set to FDR of 10%.

5. Gene expression analysis and association with CNA and regulatory mutations

RNA-seq data from tumor samples were transformed to FPKM counts and then rank normalized. The Wilcoxon-Rank sum test was used to detect changes in gene expression between mutated and wild-type cases. Changes in expression of the nearest gene were analyzed. When multiple regulatory regions mapped the same gene, p-values were adjusted for multiple testing using the FDR method (significance threshold of 5%). In the case of significant associations, we used non-linear Kernel regression adjusted for disease subtype in order to rule out independence from diagnostic subtype [45]. In the case of CNA, association with expression of the affected genes was analyzed using Pearson's correlation. P-values were adjusted for multiple testing using the FDR method (significance threshold of 5%).

6. Differential distribution of somatic events and pathways analysis

Differential mutation analysis between the different disease subtypes was performed using Fisher's exact test (significance FDR threshold of 5%).

WebGestalt [46] was used to analyze enrichment of gene networks in biological pathways. The *KEGG* database was used as reference, and a significance threshold of FDR 5% was chosen.

Results

1. Mutation landscape of B-cell lymphoid malignancies

5,743,241 mutations were detected in 354 B-cell malignancy samples. A minor proportion of these affected protein-coding regions (1.34%), of which 71.64% were non-synonymous. Among these, missense mutations predominated (87.64%), followed by nonsense mutations (6.79%) and splice-site mutations (2.76%). The vast majority of mutations were either intergenic (40.28%) or intronic (45.25%). Mutation rate in the cohort was 2.51 mutations/Mb. This mutation rate was different for the different B-cell neoplasms. Mutation rate was 0.48 mutations/Mb in CLL, 0.75 in Burkitt Lymphoma, 3.14 mutations/Mb in follicular lymphoma and 5.69 mutations/Mb in DLBCL.

2. Identification of significantly mutated genes

dNdSCV, *OncodriveFML* and *MutSigCV* detected 88, 52 and 46 recurrently mutated genes, respectively (FDR <10%) (**S1**–**S3 Tables**). Overall, 112 genes were detected as significantly mutated by any of the methods (**S4 Table**). The most frequently mutated were *KMT2D* (27%), *CREBBP* (26%), *IGLL5* (22%), *BCL2* (17%), *TP53* (13%), *ARID1A* (12%) and *TNFRSF14* (12%) (**Fig 1**). 31 genes were not previously described as recurrently mutated in any of the lymphoid malignancies analyzed (**Fig 2**), and these affected 43.22% of patients. The most frequent were *FAM230A* (6%), *LTB* (6%), *FAM186A* (6%), *ZFP36L1* (6%), *PABPC3* (5%) and *ZC3H12A* (5%). Among these, only *LTB* and *ZFP36L1* have been previously described as targets of aSHM.

Missense mutations were the most frequent at the exome level, but some of these novel drivers were predominantly affected by other types of mutations. Nonsense mutations were more frequent in *MAGI3*, *RFX7*, *TRAF6*, *VMA21* and *WDR93*. The genes *HIST2H3D* and *HLA-C* were prone to suffer multiple types of mutations in the same patient, whereas frameshift deletions predominated in *LAPTM5*. Finally, the following genes were targeted by multiple mutation types: *ANGPT1*, *ID2*, *IRF1*, *MAGI3*, *SYNCRIP* and *ZFP36L1*.

3. Detection of low-frequency drivers by functionally altered subnetwork analysis

31 significantly mutated protein subnetworks were detected, involving 313 different genes (**Fig 3 and S5 Table**). 8 networks were mutated in >10% patients. The widest network (Network 1) was composed of 153 genes, among which *CREBBP*, *EEF1A1*, *GNA13*, *STAT6* and *TP53* were the main hubs. This network was enriched in pathways such as *"B cell receptor signalling pathway"*, *"Hepatitis B"* and *"NFKB signalling pathway"* (**S6 Table and S1 Fig**). The second widest network (Network 2) was composed of 22 genes centered around *BCL2*. As expected, this network was notoriously enriched in *"Apoptosis"* pathway genes (**S6 Table and S2 Fig**). The third and fourth biggest networks (Networks 3 and 4) were composed of 20 genes each. A significant enrichment in *"Cell Adhesion Molecules"* pathway genes characterized Network 4 (FDR 2.65 x 10^{-9}).

Multiple of the remaining subnetworks include genes involved in oncogenesis. For example, Network 8 is composed of 5 genes of the *"Notch signalling"* pathway. Network 12 contains B-cell markers (*CD19* and *CD22*) as well as complement proteins (*CR1*/*CD35* and *CR2*/*CD21*). Network 20 contains genes of the toll-like receptor pathway (*MYD88* and *TLR3*), and Networks 25 and 26 are integrated by interleukins and their respective receptors (*IL10* and *IL10RA*; *IL13* and *ILR13RA2*). Furthermore, other less explored routes in lymphomagenesis emerged as significantly mutated. These included cell signalling proteins (*MAML3*, *PRKCB*, *PTPRN2* and *RASGRP3*), gene expression and cell-cycle regulators (*GLI1*, *MAD1L1*, *MNT*,

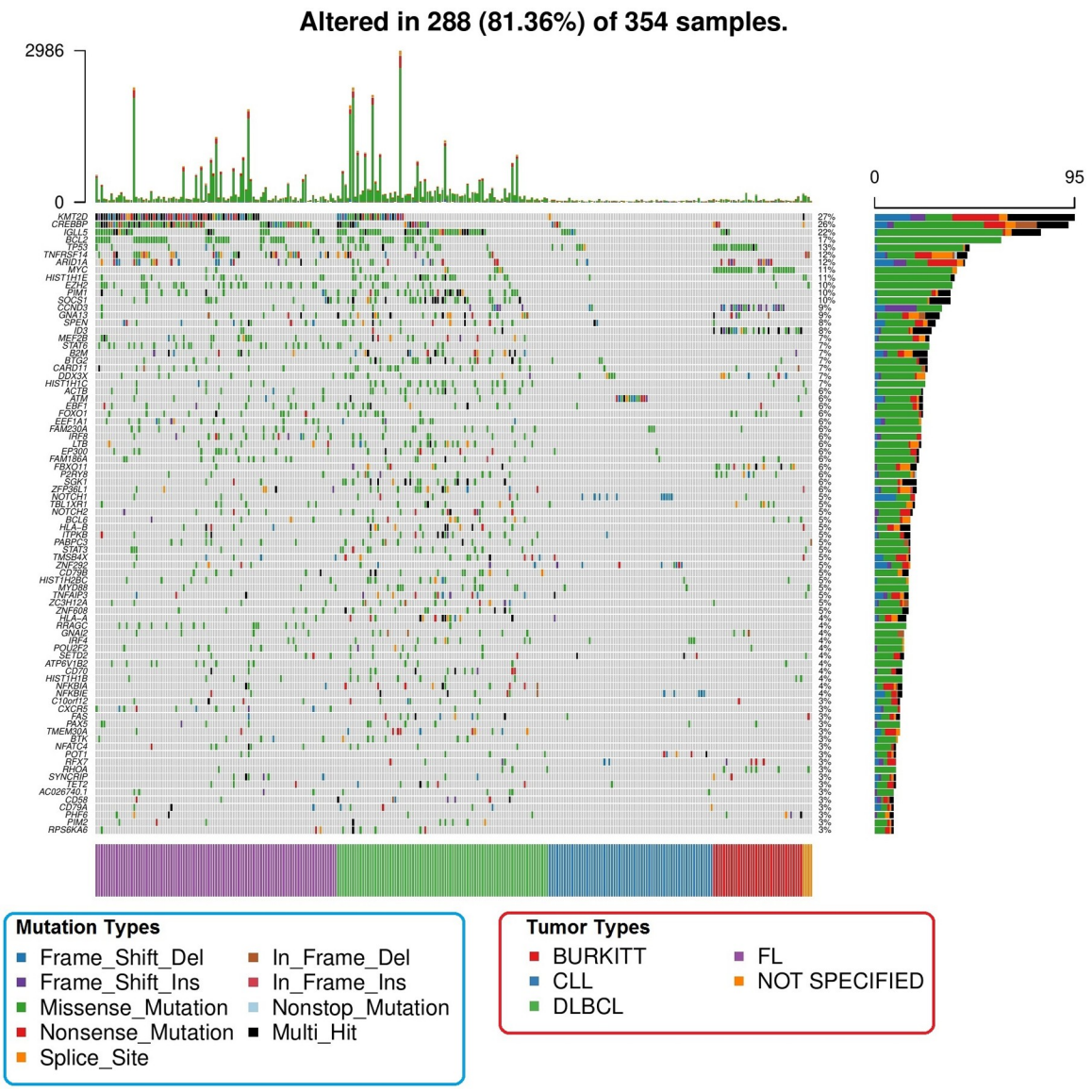

Fig 1. Representation of the most frequent drivers (frequency > 3%) across all the samples. Tumor subtypes are color-coded in the lower bar, mutation distribution is represented in the right-side bars, and per sample mutation number is represented in the top bars.

https://doi.org/10.1371/journal.pone.0248886.g001

PDZD2, *XBP1*), surface receptors and signal transduction proteins (*CD81*, *DRD3*, *GRIN1*, *GRIA2*, *GRIK3*, *LPHN2*, *PTCH1*, *PTPRE*, *PTPRN*, *SMO* . . .), angiogenesis regulators (*BAI2*), ion transport genes (*KCNB1*, *KCNC1*, *KCNC2*, *KCNG2*, *PKD1*, *PKD2* . . .), cytoskeleton and cell adhesion molecules (*ANK3*, *COL5A1*, *COL7A1*, *DSC3*, *ITGB7*, *ITGA5*, *FLNA*, *FN1*, *PKP4*, *SHANK1*, *SPTBN4*), growth factors (*IGF2* and *IGF2R*), immunity genes (*B2M*, *CD1D*, *HLA-G*, *KIR2DL4*, *TAP2*,), vesicle trafficking proteins (*GGA1*, *GGA3*, *M6PR*, *PLIN3* and *SORT1*) and extracellular enzymes (*CFD*, *CTRC*, *KLK2*, *SERPINA3*, *SERPINB6*, *SERPINF2*).

4. Regions enriched in non-coding DNA mutations

Significant enrichments in 180 regulatory elements mapping to 73 different genes were discovered. These involved 54 promoters, 53 UTRs, 33 enhancers, 21 TSS and 19 lincRNAs (S7

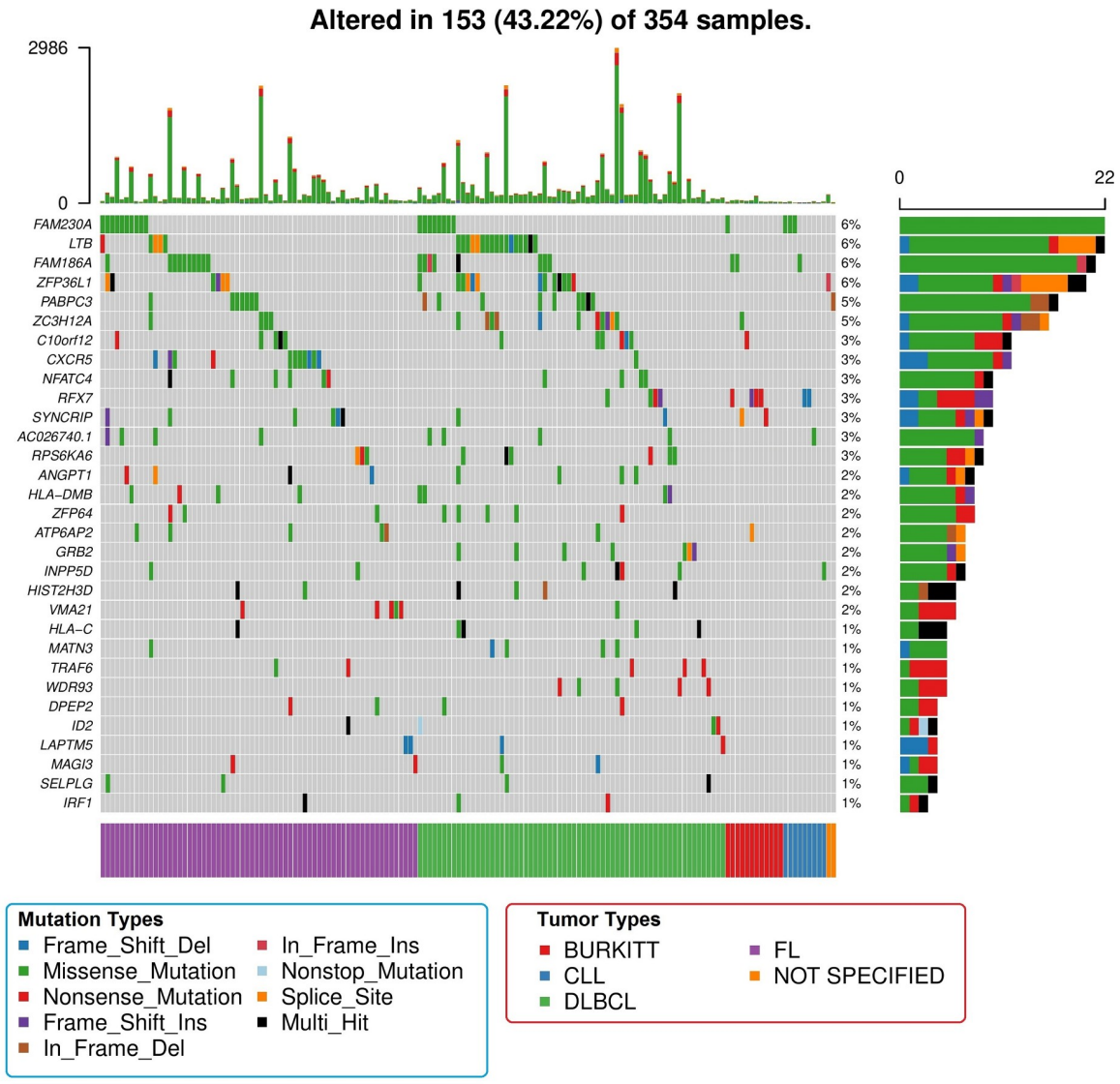

Fig 2. Representation of the new drivers discovered in this analysis. Image details are identical to those of the previous image.

https://doi.org/10.1371/journal.pone.0248886.g002

Table). 122 of these regions overlapped genes targeted by aSHM in lymphomas, and signature analysis indicated that another 4 regions were also likely targets of aSHM. On the contrary, 54 regions affecting 25 different genes were not affected by aSHM by signature analysis. Pathway analysis revealed significant overlaps between these genes and the following pathways: *"Transcriptional misregulation in cancer"* (q-value 8.55 x 10^{-3}), *"Pathways in cancer"* (q-value 8.55 x 10^{-3}), *"MicroRNAs in cancer"* (q-value 1.37 x 10^{-2}), *"JAK-STAT signaling pathway"* (q-value 1.41 x 10^{-2}), *"Toxoplasmosis"* (q-value 1.82 x 10^{-2}) and *"Apoptosis"* (q-value 3.53 x 10^{-2}).

A fraction of the patients (58%) had matched RNAseq data available. We tested association between regulatory mutations and expression of the adjacent genes. Strong underexpression of *DTX1* was associated with mutations in its promoter, which includes the *GH12J113056* enhancer region (promoter q-value, 2.90 x 10^{-4}; enhancer q-value, 4.80 x 10^{-3}, Fig 4). In the same line, mutations in *S1PR2* enhancer were also significantly associated with *S1PR2* underexpression (p-value 2.46 x 10^{-4}, Fig 4). On the contrary, mutations in the *PAX5/ZCCHC7* enhancer were

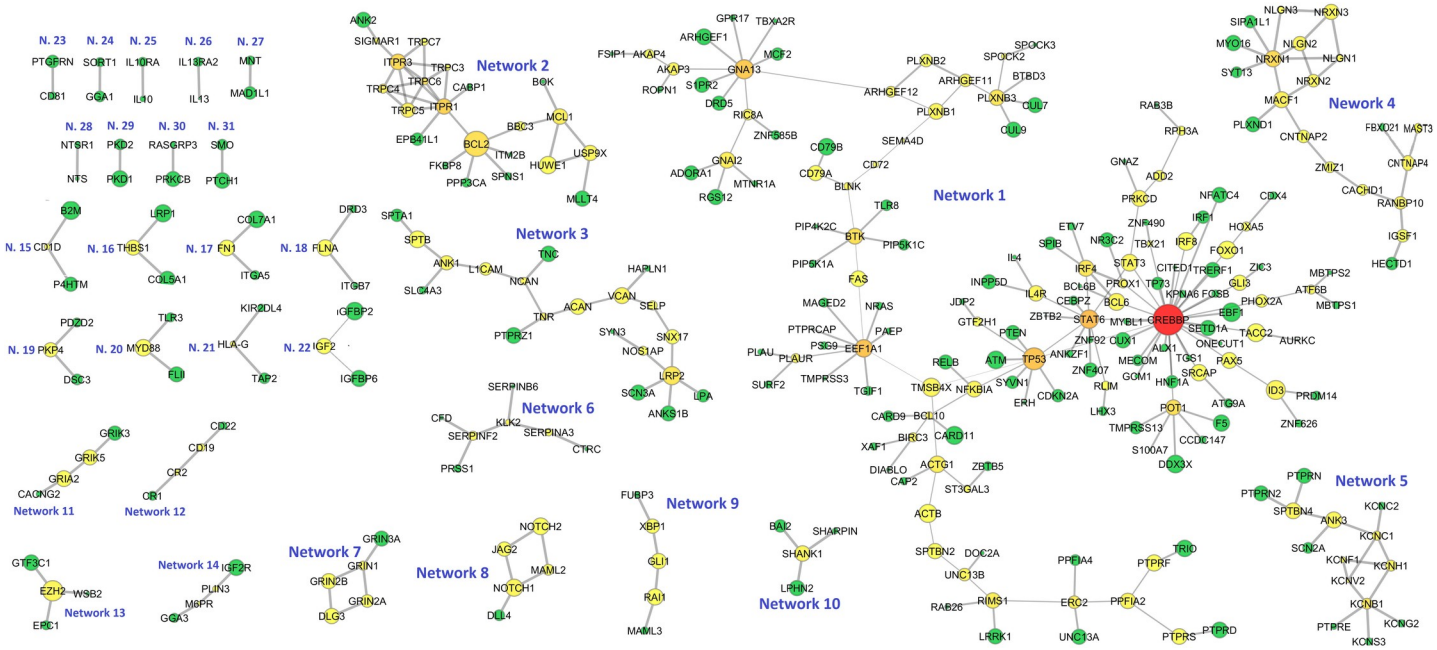

Fig 3. Representation of all significantly mutated protein subnetworks according to Hierarchical HotNet results. Node size is proportional to mutation frequency, node color is proportional to node degree (red: higher degree values, green: lower degree values) and edge width is proportional to betweenness centrality.

https://doi.org/10.1371/journal.pone.0248886.g003

significantly associated with higher expression of *ZCCHC7* (q-value 3.31 x 10^{-2}) but not with *PAX5* expression (q-value 0.93). Using non-linear kernel regression adjusted for diagnostic subtype, we could confirm independent associations for *DTX1* and *GH12J11305* mutations (p-value 2.51x 10^{-3}) and *S1PR2* and its enhancer (p-value 5.01 x 10^{-3}), but not for the association of *ZCCHC7* expression with mutations in the *PAX5/ZCCHC7* enhancer (p-value 0.17).

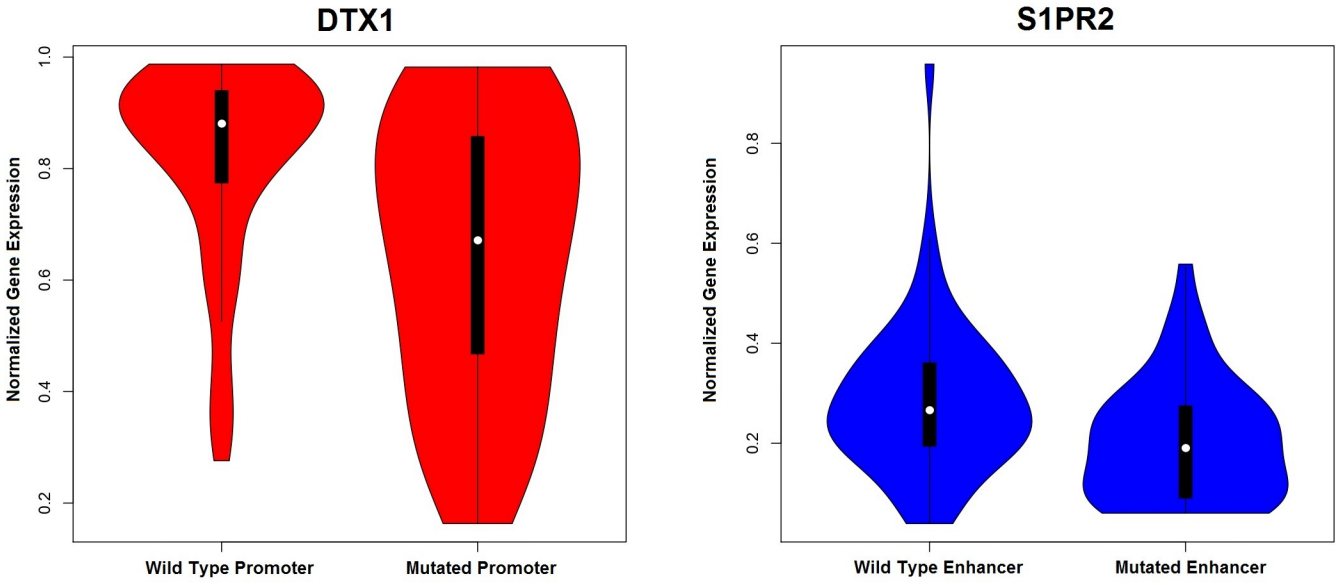

Fig 4. Violin plots representing the distribution of gene expression between mutated and unmutated *DTX1* promoters (left image) and *S1PR2* enhancers (right image).

https://doi.org/10.1371/journal.pone.0248886.g004

5. Genomic regions targeted by recurrent CNA

29 regions were significantly affected by focal CNA (FDR <10%), with a median affected region width of 1,327,586 bp (**S8 Table and S3 and S4 Figs**). We detected 2 recurrently amplified loci and 27 recurrently deleted regions in 17 different chromosomes. Aside from immunoglobulin gene deletions, the most significant deletions were located in 13q14.2 (*DLEU1*, residual q-value 1.55 x 10^{-14}), 3p12.3 (adjacent to *ROBO1*/*ROBO2* locus, q-value 2.90 x x10^{-7}), 16q21 (*CDH5* and *CDH11* loci, residual q-value 1.06 x 10^{-5}), 1p36.32 (*TNFRSF14*, residual q-value 7.56 x 10^{-5}), 6q13 (*LINC00472* locus, residual q-value 8.93 x 10^{-5}) and 13q33.3 (*TNFSF13B* locus, residual q-value 1.02 x 10^{-4}). On the contrary, recurrent amplifications of the non-coding locus 9q34.11 (residual q-value 6.85 x 10^{-3}) and the gene-rich region 13q31.3 (residual q-value 0.06) were identified. Other focal deletions affected genes involved in immune pathways (*IL5RA*) and oncogenic pathways, such as the tumor-suppressor *APAF1* (12q23.1), the immune check-point regulator *PVR* (19q13.31) and the cell cycle regulators *CDKN2A* and *CDKN2B* (9p21.3).

A fraction of the patients (58%) had matched RNAseq data available. Significant positive Pearson's correlations between tumor/normal log2 rations and local gene expression were discovered in 16 CNAs (q-value <0.1; **S9 Table**). This included positive associations between losses in 1p36.32, 3p26.2, 4q35.2, 7p22.3 and 18q23 with the expression of the cancer-related genes *CDKN11B*, *CRBN*, *IRF2*, *PRKAR1B* and *NFATC*.

6. Differential events between B-cell lymphoma subtypes

We studied the distribution of the significant genomic events across the different B-cell lymphoproliferative subtypes (**S10 Table**). The greatest number of significant disparities (FDR < 5%) was discovered between CLL *versus* DLBCL and CLL *versus* follicular lymphoma. We detected 77 differential events between CLL and DLBCL, 34 differential events between CLL and follicular lymphoma, 11 differential events between CLL and Burkitt lymphoma, 17 differential events between DLBCL and Burkitt lymphoma, 11 differential events between follicular lymphoma and DLBCL and 9 differential events between follicular lymphoma and Burkitt lymphoma. Overall, 76.43% of all differential events were between CLL and any of the other lymphomas.

Although most differential events were less common in CLL, *IGH* deletions were highly enriched in CLL compared with the remaining lymphomas, and *IGL* deletions were more frequent in CLL compared to follicular lymphoma. Additionally, 11p15.5 deletions were more frequent in CLL than in follicular lymphoma or DLBCL, and 11q22.3 deletions were significantly more frequent in CLL than in DLBCL. In the same line, non-coding mutations in the *IGH* locus were significantly more frequent in CLL than in DLBCL or follicular lymphoma, and those in the *IGL* loci predominated in CLL over DLBCL. Furthermore, non-coding mutations in *RP11-789C2.1* (4q28.3) and in the first intron of *BACH2* (6q15) were significantly increased in CLL compared with DLBCL. On the contrary, coding mutations in 97 driver genes were significantly depleted in CLL compared with the remaining disease subgroups, likely reflecting the lower mutational burden on CLL. Additionally, 10 structural aberrations (9 deletions and 1 amplification) were significantly less frequent in CLL than in DLBCL or follicular lymphoma. The most significant were particularly predominant in DLBCL cases, and these were 6q26 loss (q-value 5.36 x 10^{-5}), 16q21 loss (q-value 4.56 x 10^{-4}) and 13q13.3 gain (q-value 4.56 x 10^{-4}).

As expected, mutations in *MYC*, *ID3* and *CCND3* were more frequent in Burkitt Lymphoma than in DLBCL or follicular lymphoma, and additionally we also detected the significant enrichments of Burkitt Lymphomas in *TP53* and *FBXO11* mutations. On the contrary,

mutations in KMT2D and BCL2 were more prevalent among follicular lymphoma and DLBCL than in Burkitt Lymphoma. 1p36.32 deletion was absent in Burkitt lymphoma, but noncoding mutations in *RP11-44H4.1* (3q27.3) were enriched in Burkitt lymphoma compared to DLBCL. Finally, the comparison of follicular lymphoma with DLBCL revealed significant differences in the mutational frequency of 11 genes. Of these, mutations in *CREBBP*, *KMT2D*, *TNFRSF14* and *RRAGC* were enriched in follicular lymphoma, and those of *BTG2*, *HLA-A*, *PIM1*, *IGLL5*, *SOCS1*, *CD83* and *SGK1* were enriched in DLBCL.

Discussion

Different analyses of B-cell lymphoproliferative disorders have deconvoluted part of the complex genomic landscape of these neoplasms. Despite extensive evidence in other fields [20, 34], this is the first combined analysis of whole genomes of B-cell lymphoid tumors performed to our knowledge. In this work, we detected 112 recurrently mutated genes across the genomes of different B-cell lymphoid malignancies, of which 31 (27.7%) were not previously described in any of the analyzed tumor subtypes. Among these, some of the most frequently mutated (*FAM230A*, *FAM186A* and *PABPC3*) are barely characterized genes with testis-biased expression. On the contrary, many others play roles in pathways linked with lymphomagenesis. For example, *GRB2* and *INPP5D* participate in the B-cell receptor pathway [47, 48]; *LTB* and *TRAF6* regulate NFKB pathway activity [49, 50], and *ANGPT1* and *RPS6KA6* play a role in the MAPK pathway [51, 52]. Functional evidence supports the implication of the transcription factor *RFX7* [47–53] and the zinc finger protein *ZFP36L1* [54] in oncogenesis. Several other genes are members of the family of known lymphoma drivers, such as *CXCR5* [55], *HIST2H3D* [56], *ID2* [57] and *IRF1* [58]. Additionally, 180 regulatory regions were significantly enriched in mutations, with a significant contribution of aSHM target loci (67.7% of cases). Importantly, we could detect regulatory mutations accompanied by aberrant underexpression of the tumor suppressors *DTX1* [59, 60] and *S1PR2* [61, 62].

31 significantly mutated subnetworks involving 313 genes were discovered. In comparison with single gene approaches, this perspective provides a more complete landscape of mutational processes in B-cell lymphomas, and points towards the existence of new altered proteomic subnetworks in lymphomas. As a result, the genes *CREBBP*, *BCL2*, *EEF1A1*, *GNA13*, *STAT6* and *TP53*, and the pathways *"B-cell receptor"*, *"Apoptosis"*, *"Notch signalling"*, *"Polycomb Repressive Complex"* and *"Toll-like receptor"* emerge as master players of lymphomagenesis. Nevertheless, our results also support the implication of a myriad of novel players in the pathogenesis of B-cell lymphoid disorders, such as cell signalling proteins, cell-cycle regulators, ion transporters, cytoskeleton proteins, vesicle trafficking factors, extracellular enzymes and immunity genes. For some of these, the association with lymphomagenesis is well established, as in the case of *CR2*/*CD21* [63], *CD81* [64], *GLI1* [65], *SMO* [66] and the self-activating autocrine loops of *IL10* and *IL13* with their receptors [67, 68]. For other genes, likely associations can be inferred from its function, such as the cell-cycle checkpoint protein *MAD1L1* [69] and the angiogenesis inhibitor *BAI2* [70]. Finally, another group of genes belongs to emerging pathways in cancer whose function in B-cell lymphomas awaits further elucidation, such as glutamate receptors [71], ion channels [72] and microvesicles [73].

Focal recurrent structural alterations were detected in 29 loci. These events tended to affect known drivers of lymphomagenesis, and some were previously described, such as 13q14.2 deletions (*RB1* gene) [74], *CDKN2A*/*CDKN2B* [75], *TP58* [76], *DLEU1* (13q13.2) [8], and *ILR5RA* [77]. On the contrary, other novel deletions were either described in other tumors, such as *CDH11* in retinoblastoma [78] or *ROBO1* in breast cancer [79]; or affected genes vinculated with cancer pathways such as the tumor suppressor & non-coding RNA *LINC00472*

[80], the B-cell specific maturation regulator *TNFRSF13C* (*BAFF receptor*) [81], the immune checkpoint *PVR* [82], the receptor tyrosine kinase-coupled signaling regulator *SIRPB1* [83] or the proapoptotic gene *APAF1* [84]. Furthermore, recurrent amplifications in a noncoding region within 9q34.11 were also detected, whose function needs to be further clarified.

Finally, some clues are provided about the distribution of these mutational events across B-cell tumor subtypes. A relative depletion of CLL in mutations affecting common drivers was found, in line with the lower mutational burden of this tumor. Additionally, several differences in mutation frequencies were also detected between follicular lymphoma, DLBCL and Burkitt lymphoma. As expected, mutations in frequent drivers such as *CREBBP*, *KMT2D*, *TNFRSF14* and *RRAGC* were more frequent among follicular lymphomas, those of *MYC*, *CCND3* and *ID3* prevailed among Burkitt lymphoma and those of *BTG2*, *PIM1*, *SGK1* and *SOCS1* were more frequent among DLBCL. Some of these findings are concordant with reported frequencies of driver genes across distinct lymphoma subtypes [85–87], whereas others provide new clues about the pathogenesis and possible drug targets in these tumors. For example, the increased mutational burden of *TP53* and the BCL6-regulator gene *FBXO11* among Burkitt lymphomas suggest an increased deregulation of these pathways in this disease [88, 89]. Additionally, the skewed mutational profile of the immune check-point regulator *CD83* [90] towards DLBCL tumors might have both biological and therapeutic implications.

This work, as many others, has some limitations. For example, although the included B-cell disorders represent a majority of patients in real practice, other frequent B-cell malignancies need to be taken into account in the future. Furthermore, protein network analysis is still limited by incomplete annotation of the protein interactome and by the type of input scores that can be used as input. Additionally, it should be noted that data produced from different research groups can be affected by batch effects, which is the reason why we used a uniform and optimized pipeline for the analysis.

In conclusion, we present an integrated overview of the genomic drivers of some of the most frequent B-cell lymphoproliferative disorders. Our results shed new light about the pathogenic mutations and structural aberrations in coding and noncoding regulatory regions of the genome of B-cell lymphoproliferative disorders, and pinpoint towards disease-specific mutational events that might be useful both for therapeutic and diagnostic purposes.

Supporting information

S1 Fig. Volcano plot of the most significantly enriched KEGG terms in Network 1 genes.
(PNG)

S2 Fig. Volcano plot of the most significantly enriched KEGG terms in Network 2 genes.
(PNG)

S3 Fig. Deletion plot produced by Gistic2.0.
(JPG)

S4 Fig. Amplification plot produced by Gistic2.0.
(JPG)

S1 Table. Significantly mutated genes detected by dNdScv.
(XLSX)

S2 Table. Significantly mutated genes detected by MutSigCV.
(XLSX)

S3 Table. Significantly mutated genes detected by OncodriveFML.
(XLSX)

S4 Table. List of all significantly mutated driver genes.
(XLSX)

S5 Table. Significant subnetworks revealed by Hierarchical HotNet.
(XLSX)

S6 Table. Overrepresentation analysis of Network 1 and Network 2 genes in KEGG pathways terms.
(XLSX)

S7 Table. List of regulatory mutations significantly enriched in noncoding mutations (FDR <0.1). The last two columns indicate if the region maps to known target gene of aSHM, or if the mutational signature in that loci is consistent with aSHM.
(XLSX)

S8 Table. List of significant and recurrent focal CNA regions (FDR <10%).
(XLSX)

S9 Table. List of significant positive correlations between tumor/normal log2 ratios of recurrent CNA and local gene expression. FDR is reported in case a CNA affected the locus of more than 1 gene.
(XLSX)

S10 Table. Differential analysis of mutation events across the different subtypes of B-cell lymphoid neoplasms. Only events with FDR <5% are shown.
(XLSX)

Acknowledgments

We would like to thank the *International Cancer Genome Consortium* for facilitating the data, and to the *Supercomputing Center of Galicia* (CESGA) for providing informatics support for the analysis.

Author Contributions

Conceptualization: Adrián Mosquera Orgueira.

Data curation: Adrián Mosquera Orgueira.

Formal analysis: Adrián Mosquera Orgueira, And´res Peleteiro Raindo, Miguel Cid López.

Investigation: Adrián Mosquera Orgueira, Roi Ferreiro Ferro, José Ángel Díaz Arias, Carlos Aliste Santos, Beatriz Antelo Rodríguez, Laura Bao Pérez, Natalia Alonso Vence, Ággeles Bendaña López, Aitor Abuin Blanco, Paula Melero Valentín, And´res Peleteiro Raindo, Miguel Cid López.

Methodology: Adrián Mosquera Orgueira, José Ángel Díaz Arias, Carlos Aliste Santos, Beatriz Antelo Rodríguez, Laura Bao Pérez, Paula Melero Valentín, And´res Peleteiro Raindo, Miguel Cid López.

Project administration: Beatriz Antelo Rodríguez.

Software: Adrián Mosquera Orgueira, Aitor Abuin Blanco, Miguel Cid López.

Supervision: José Ángel Díaz Arias, Laura Bao Pérez, Natalia Alonso Vence, José Luis Bello López.

Validation: Adrián Mosquera Orgueira, Manuel Mateo Pérez Encinas, Marta Sonia González Pérez, Máximo Francisco Fraga Rodríguez, José Luis Bello López.

Visualization: Adrián Mosquera Orgueira, José Ángel Díaz Arias, Manuel Mateo Pérez Encinas, Marta Sonia González Pérez, Máximo Francisco Fraga Rodríguez.

Writing – original draft: Adrián Mosquera Orgueira, Roi Ferreiro Ferro, José Ángel Díaz Arias, Carlos Aliste Santos, Beatriz Antelo Rodríguez, Laura Bao Pérez, Natalia Alonso Vence, Ággeles Bendaña López, Aitor Abuin Blanco, And´res Peleteiro Raindo, Miguel Cid López, José Luis Bello López.

Writing – review & editing: Carlos Aliste Santos, Ággeles Bendaña López, Manuel Mateo Pérez Encinas, Marta Sonia González Pérez, Máximo Francisco Fraga Rodríguez, José Luis Bello López.

References

1. Teras LR, DeSantis CE, Cerhan JR et al. 2016 US lymphoid malignancy statistics by World Health Organization subtypes. CA Cancer J Clin. 2016 Nov 12; 66(6):443–459. https://doi.org/10.3322/caac.21357 Epub 2016 Sep 12. PMID: 27618563.
2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000 Feb 3; 403(6769):503–11. https://doi.org/10.1038/35000501 PMID: 10676951.
3. Chapuy B, Stewart C, Dunford AJ et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med. 2018 May; 24(5):679–690. https://doi.org/10.1038/s41591-018-0016-8 Epub 2018 Apr 30. Erratum in: Nat Med. 2018 Aug;24(8):1292. Nat Med. 2018 Aug;24(8):1290–1291. PubMed Central PMCID: PMC6613387. PMID: 29713087
4. Schmitz R, Wright GW, Huang DW et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med. 2018 Apr 12; 378(15):1396–1407. https://doi.org/10.1056/NEJMoa1801445 PMID: 29641966; PubMed Central PMCID: PMC6010183.
5. Loeffler M, Kreuz M, Haake A, et al. Genomic and epigenomic co-evolution in follicular lymphomas. Leukemia. 2015 Feb; 29(2):456–63. https://doi.org/10.1038/leu.2014.209 Epub 2014 Jul 16. PMID: 25027518.
6. Pastore A, Jurinovic V, Kridel R et al. Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: a retrospective analysis of a prospective clinical trial and validation in a population-based registry. Lancet Oncol. 2015 Sep; 16(9):1111–1122. https://doi.org/10.1016/S1470-2045(15)00169-2 Epub 2015 Aug 6. PMID: 26256760.
7. Landau DA, Tausch E, Taylor-Weiner AN et al. Mutations driving CLL and their evolution in progression and relapse. Nature. 2015 Oct 22; 526(7574):525–30. https://doi.org/10.1038/nature15395 Epub 2015 Oct 14. PMID: 26466571; PubMed Central PMCID: PMC4815041.
8. Puente XS, Beà S, Valdés-Mas R et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015 Oct 22; 526(7574):519–24. https://doi.org/10.1038/nature14666 Epub 2015 Jul 22. PMID: 26200345.
9. Panea RI, Love CL, Shingleton JR et al. The whole-genome landscape of Burkitt lymphoma subtypes. Blood. 2019 Nov 7; 134(19):1598–1607. https://doi.org/10.1182/blood.2019001880 PMID: 31558468; PubMed Central PMCID: PMC6871305.
10. López C, Kleinheinz K, Aukema SM et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. Nat Commun. 2019 Mar 29; 10(1):1459. https://doi.org/10.1038/s41467-019-08578-3 PMID: 30926794; PubMed Central PMCID: PMC6440956.
11. Grande BM, Gerhard DS, Jiang A et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. Blood. 2019 Mar 21; 133(12):1313–1324. https://doi.org/10.1182/blood-2018-09-871418 Epub 2019 Jan 7. PMID: 30617194; PMCID: PMC6428665.
12. Richter J, Schlesner M, Hoffmann S et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet. 2012 Dec; 44(12):1316–20. https://doi.org/10.1038/ng.2469 Epub 2012 Nov 11. PMID: 23143595.

13. Stilgenbauer S, Schnaiter A, Paschka P et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood. 2014 May 22; 123(21):3247–54. https://doi.org/10.1182/blood-2014-01-546150 Epub 2014 Mar 20. PMID: 24652989.

14. Batmanov K, Wang W, Bjørås M et al. Integrative whole-genome sequence analysis reveals roles of regulatory mutations in BCL6 and BCL2 in follicular lymphoma. Sci Rep. 2017 Aug 1; 7(1):7040. https://doi.org/10.1038/s41598-017-07226-4 PMID: 28765546; PubMed Central PMCID: PMC5539289.

15. Arthur SE, Jiang A, Grande BM et al. Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. Nat Commun. 2018 Oct 1; 9(1):4001. https://doi.org/10.1038/s41467-018-06354-3 PMID: 30275490; PubMed Central PMCID: PMC6167379.

16. Sandmann S, de Graaf AO, Karimi M et al. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. Sci Rep. 2017 Feb 24; 7:43169. https://doi.org/10.1038/srep43169 PMID: 28233799

17. Hofmann AL, Behr J, Singer J et al. Detailed simulation of cancer exome sequencing data reveals differences and common limitations of variant callers. BMC Bioinformatics. 2017 Jan 3; 18(1):8. https://doi.org/10.1186/s12859-016-1417-7 PMID: 28049408

18. Cai L, Yuan W, Zhang Z et al. In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. Sci Rep. 2016 Nov 22; 6:36540. https://doi.org/10.1038/srep36540 PMID: 27874022

19. Vogelstein B, Papadopoulos N, Velculescu VE et al. Cancer genome landscapes. Science. 2013 Mar 29; 339(6127):1546–58. https://doi.org/10.1126/science.1235122 Review. PMID: 23539594; PubMed Central PMCID: PMC3749880.

20. Priestley P, Baber J, Lolkema MP et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019 Nov; 575(7781):210–216. https://doi.org/10.1038/s41586-019-1689-y Epub 2019 Oct 23. PMID: 31645765; PubMed Central PMCID: PMC6872491.

21. Ma X, Liu Y, Liu Y et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018 Mar 15; 555(7696):371–376. https://doi.org/10.1038/nature25795 Epub 2018 Feb 28. PMID: 29489755; PubMed Central PMCID: PMC5854542.

22. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct; 45(10):1113–20. https://doi.org/10.1038/ng.2764 PMID: 24071849; PubMed Central PMCID: PMC3919969.

23. International Cancer Genome Consortium. International network of cancer genome projects. Nature 464, 993–998, https://doi.org/10.1038/nature08987 (2010). PMID: 20393554

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

25. Lai Z, Markovets A, Adhesmaki M O et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 44, e108, https://doi.org/10.1093/nar/gkw227 (2016). PMID: 27060149

26. Koboldt DC, Zhang Q, Larson DE et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22, 568–576, https://doi.org/10.1101/gr.129684.111 (2012). PMID: 22300766

27. do Valle IF, Giampieri E, Simonetti G et al. Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. BMC Bioinformatics 17, 341, https://doi.org/10.1186/s12859-016-1190-7 (2016). PMID: 28185561

28. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv, 1207.3907 [q-bio.GN] (2012)

29. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature 526 68–74, https://doi.org/10.1038/nature15393 (2015). PMID: 26432245

30. Karczewski, K.J. et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 531210, https://doi.org/10.1101/531210

31. Lek M, Francioli LC, Tiao G et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291, https://doi.org/10.1038/nature19057 (2016). PMID: 27535533

32. Talevich E, Shain AH, Botton T et al. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol. 2016 Apr 21; 12(4):e1004873. https://doi.org/10.1371/journal.pcbi.1004873 PMID: 27100738

33. Lawrence MS, Stojanov P, Polak P et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013 Jul 11; 499(7457):214–218. https://doi.org/10.1038/nature12213 Epub 2013 Jun 16. PMID: 23770567; PubMed Central PMCID: PMC3919509.

34. Martincorena I, Raine KM, Gerstung M et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017 Nov 16; 171(5):1029-1041.e21. https://doi.org/10.1016/j.cell.2017.09.042 Epub 2017 Oct 19. Erratum in: Cell. 2018 Jun 14;173(7):1823. PMID: 29056346

35. Mularoni L, Sabarinathan R, Deu-Pons J et al. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. 2016 Jun 16; 17(1):128. https://doi.org/10.1186/s13059-016-0994-0 PMID: 27311963

36. Reyna MA, Leiserson MDM, Raphael BJ. Hierarchical HotNet: identifying hierarchies of altered subnetworks. Bioinformatics. 2018 Sep 1; 34(17):i972–i980. https://doi.org/10.1093/bioinformatics/bty613 PMID: 30423088

37. Harrow J, Frankish A, Gonzalez JM et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22, 1760–1774, https://doi.org/10.1101/gr.135350.111 (2012). PMID: 22955987

38. Fishilevich S, Nudel R, Rappaport N et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford), https://doi.org/10.1093/database/bax028 (2017). PMID: 28605766

39. Lochovsky L, Zhang J, Fu Y et al. LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. Nucleic Acids Res 43, 8123–8134, https://doi.org/10.1093/nar/gkv803 (2015). PMID: 26304545

40. Khodabakhshi AH, Morin RD, Fejes AP, et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. Oncotarget. 2012; 3(11):1308–1319. https://doi.org/10.18632/oncotarget.653 PMID: 23131835

41. Jiang Y, Soong TD, Wang L, Melnick AM, Elemento O. Genome-wide detection of genes targeted by non-Ig somatic hypermutation in lymphoma. PLoS One. 2012; 7(7):e40332. https://doi.org/10.1371/journal.pone.0040332 Epub 2012 Jul 12. PMID: 22808135; PMCID: PMC3395700.

42. Fukumura K, Kawazu M, Kojima S, Ueno T, Sai E, Soda M, et al. Genomic characterization of primary central nervous system lymphoma. Acta Neuropathol. 2016 Jun; 131(6):865–75. Epub 2016 Jan 12. https://doi.org/10.1007/s00401-016-1536-2 PMID: 26757737

43. Degasperi A, Amarante TD, Czarnecki J et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. Nat Cancer. 2020 Feb; 1(2):249–263. https://doi.org/10.1038/s43018-020-0027-5 Epub 2020 Feb 17. PMID: 32118208; PMCID: PMC7048622.

44. Mermel CH, Schumacher SE, Hill B et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011; 12(4):R41. https://doi.org/10.1186/gb-2011-12-4-r41 PMID: 21527027

45. Hayfield T, Racine JS (2008). "Nonparametric Econometrics: The np Package." Journal of Statistical Software, 27(5). http://www.jstatsoft.org/v27/i05/.

46. Liao Y, Wang J, Jaehnig EJ et al. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019 Jul 2; 47(W1):W199–W205. https://doi.org/10.1093/nar/gkz401 PMID: 31114916

47. Jiang X, Lu X, Zhang Y et al. Interplay between HGAL and Grb2 proteins regulates B-cell receptor signaling. Blood Adv. 2019 Aug 13; 3(15):2286–2297. https://doi.org/10.1182/bloodadvances.2018016162 PMID: 31362927

48. Chen Z, Shojaee S, Buchner M et al. Signalling thresholds and negative B-cell selection in acute lymphoblastic leukaemia. Nature. 2015 May 21; 521(7552):357–61. https://doi.org/10.1038/nature14231 Epub 2015 Mar 23. Erratum in: Nature. 2016 Jun 2;534(7605):138. PMID: 25799995

49. Das R, Coupar J, Clavijo PE et al. Lymphotoxin-β receptor-NIK signaling induces alternative RELB/NF-κB2 activation to promote metastatic gene expression and cell migration in head and neck cancer. Mol Carcinog. 2019 Mar; 58(3):411–425. https://doi.org/10.1002/mc.22938 PMID: 30488488

50. Fang J, Muto T, Kleppe M et al. TRAF6 Mediates Basal Activation of NF-κB Necessary for Hematopoietic Stem Cell Homeostasis. Cell Rep. 2018 Jan 30; 22(5):1250–1262. https://doi.org/10.1016/j.celrep.2018.01.013 PMID: 29386112

51. Hashiramoto A, Sakai C, Yoshida K et al. Angiopoietin 1 directly induces destruction of the rheumatoid joint by cooperative, but independent, signaling via ERK/MAPK and phosphatidylinositol 3-kinase/Akt. Arthritis Rheum. 2007 Jul; 56(7):2170–9. https://doi.org/10.1002/art.22727 PMID: 17599743

52. Rafiee M, Keramati MR, Ayatollahi H et al. Down-Regulation of Ribosomal S6 kinase RPS6KA6 in Acute Myeloid Leukemia Patients. Cell J. 2016 Jul-Sep; 18(2):159–64. Epub 2016 May 30. https://doi.org/10.22074/cellj.2016.4310 PMID: 27540520

53. Weber J, de la Rosa J, Grove CS et al. PiggyBac transposon tools for recessive screening identify B-cell lymphoma drivers in mice. Nat Commun. 2019 Mar 29; 10(1):1415. https://doi.org/10.1038/s41467-019-09180-3 PMID: 30926791

54. Suk FM, Chang CC, Lin RJ et al. ZFP36L1 and ZFP36L2 inhibit cell proliferation in a cyclin D-dependent and p53-independent manner. Sci Rep. 2018 Feb 9; 8(1):2742. https://doi.org/10.1038/s41598-018-21160-z Erratum in: Sci Rep. 2019 Nov 20;9(1):17457. PMID: 29426877

55. Krysiak K, Gomez F, White BS et al. Recurrent somatic mutations affecting B-cell receptor signaling pathway genes in follicular lymphoma. Blood. 2017 Jan 26; 129(4):473–483. https://doi.org/10.1182/blood-2016-07-729954 Epub 2016 Nov 14. PMID: 28064239

56. Morin RD, Mendez-Lago M, Mungall AJ et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011 Jul 27; 476(7360):298–303. https://doi.org/10.1038/nature10351 PMID: 21796119

57. Rohde M, Bonn BR, Zimmermann M et al. Relevance of ID3-TCF3-CCND3 pathway mutations in pediatric aggressive B-cell lymphoma treated according to the non-Hodgkin Lymphoma Berlin-Frankfurt-Münster protocols. Haematologica. 2017 Jun; 102(6):1091–1098. https://doi.org/10.3324/haematol.2016.156885 Epub 2017 Feb 16. PMID: 28209658

58. Ramis-Zaldivar JE, Gonzalez-Farre B, Balagué O et al. IRF4-rearranged Large B-cell lymphoma (LBCL) has a genomic profile distinct to other LBCL in children and young adults. Blood. 2019 Nov 18. pii: blood.2019002699. https://doi.org/10.1182/blood.2019002699 PMID: 31738823

59. Rossi D, Trifonov V, Fangazio M 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. https://doi.org/10.1084/jem.20120904 Epub 2012 Aug 13. PMID: 22891273; PubMed Central PMCID: PMC3428941.

60. Meriranta L, Pasanen A, Louhimo R et al. Deltex-1 mutations predict poor survival in diffuse large B-cell lymphoma. Haematologica. 2017 May; 102(5):e195–e198. https://doi.org/10.3324/haematol.2016.157495 Epub 2017 Feb 9. PubMed Central PMCID: PMC5477623. PMID: 28183850

61. Stelling A, Hashwah H, Bertram K et al. The tumor suppressive TGF-β/SMAD1/S1PR2 signaling axis is recurrently inactivated in diffuse large B-cell lymphoma. Blood. 2018 May 17; 131(20):2235–2246. https://doi.org/10.1182/blood-2017-10-810630 Epub 2018 Apr 3. PMID: 29615404.

62. Bouska A, Zhang W, Gong Q et al. Combined copy number and mutation analysis identifies oncogenic pathways associated with transformation of follicular lymphoma. Leukemia. 2017 Jan; 31(1):83–91. https://doi.org/10.1038/leu.2016.175 Epub 2016 Jun 16. PMID: 27389057; PubMed Central PMCID: PMC5214175.

63. Otsuka M, Yakushijin Y, Hamada M et al. Role of CD21 antigen in diffuse large B-cell lymphoma and its clinical significance. Br J Haematol. 2004 Nov; 127(4):416–24. https://doi.org/10.1111/j.1365-2141.2004.05226.x PMID: 15521918.

64. Vences-Catalán F, Kuo CC, Rajapaksa R et al. CD81 is a novel immunotherapeutic target for B cell lymphoma. J Exp Med. 2019 Jul 1; 216(7):1497–1508. https://doi.org/10.1084/jem.20190186 Epub 2019 May 23. PMID: 31123084; PubMed Central PMCID: PMC6605745.

65. Yoon JW, Gallant M, Lamm ML et al. Noncanonical regulation of the Hedgehog mediator GLI1 by c-MYC in Burkitt lymphoma. Mol Cancer Res. 2013 Jun; 11(6):604–15. https://doi.org/10.1158/1541-7786.MCR-12-0441 Epub 2013 Mar 22. PMID: 23525267.

66. Todorovic Balint M, Jelicic J, Mihaljevic B et al. Gene Mutation Profiles in Primary Diffuse Large B Cell Lymphoma of Central Nervous System: Next Generation Sequencing Analyses. Int J Mol Sci. 2016 May 6; 17(5). pii: E683. https://doi.org/10.3390/ijms17050683 PMID: 27164089; PubMed Central PMCID: PMC4881509.

67. Béguelin W, Sawh S, Chambwe N et al. IL10 receptor is a novel therapeutic target in DLBCLs. Leukemia. 2015 Aug; 29(8):1684–94. https://doi.org/10.1038/leu.2015.57 Epub 2015 Mar 3. PMID: 25733167.

68. Zhang Y, Li C, Zhang M et al. IL-13 and IL-13Rα1 are overexpressed in extranodal natural killer/T cell lymphoma and mediate tumor cell proliferation. Biochem Biophys Res Commun. 2018 Sep 18; 503(4):2715–2720. https://doi.org/10.1016/j.bbrc.2018.08.030 Epub 2018 Aug 11. PMID: 30107911.

69. Tsukasaki K, Miller CW, Greenspun E et al. Mutations in the mitotic check point gene, MAD1L1, in human cancers. Oncogene. 2001 May 31; 20(25):3301–5. https://doi.org/10.1038/sj.onc.1204421 PMID: 11423979.

70. Kee HJ, Koh JT, Kim MY et al. Expression of brain-specific angiogenesis inhibitor 2 (BAI2) in normal and ischemic brain: involvement of BAI2 in the ischemia-induced brain angiogenesis. J Cereb Blood Flow Metab. 2002 Sep; 22(9):1054–67. https://doi.org/10.1097/00004647-200209000-00003 PMID: 12218411.

71. Yu LJ, Wall BA, Wangari-Talbot J et al. Metabotropic glutamate receptors in cancer. Neuropharmacology. 2017 Mar 15; 115:193–202. https://doi.org/10.1016/j.neuropharm.2016.02.011 Epub 2016 Feb 16. Review. PMID: 26896755; PubMed Central PMCID: PMC4987272.

72. Litan A, Langhans SA. Cancer as a channelopathy: ion channels and pumps in tumor development and progression. Front Cell Neurosci. 2015 Mar 17; 9:86. https://doi.org/10.3389/fncel.2015.00086 Review. PMID: 25852478; PubMed Central PMCID: PMC4362317.

73. Whiteside TL. Lymphoma exosomes reprogram the bone marrow. Blood. 2018 Apr 12; 131(15):1635–1636. https://doi.org/10.1182/blood-2018-02-830497 PMID: 29650731; PubMed Central PMCID: PMC5897871.

74. Wada M, Okamura T, Okada M, et al. Frequent chromosome arm 13q deletion in aggressive non-Hodgkin's lymphoma. Leukemia. 1999; 13(5):792–798. https://doi.org/10.1038/sj.leu.2401395 PMID: 10374885

75. Pinyol M, Cobo F, Bea S, et al. p16(INK4a) gene inactivation by deletions, mutations, and hypermethylation is associated with transformed and aggressive variants of non-Hodgkin's lymphomas. Blood. 1998; 91(8):2977–2984. PMID: 9531609

76. Dave BJ, Pickering DL, Hess MM et al. Deletion of cell division cycle 2-like 1 gene locus on 1p36 in non-Hodgkin lymphoma. Cancer Genet Cytogenet. 1999; 108(2):120–126. https://doi.org/10.1016/s0165-4608(98)00138-1 PMID: 9973938

77. Scholtysik R, Kreuz M, Hummel M, et al. Characterization of genomic imbalances in diffuse large B-cell lymphoma by detailed SNP-chip analysis. Int J Cancer. 2015; 136(5):1033–1042. https://doi.org/10.1002/ijc.29072 PMID: 25042405

78. Marchong MN, Chen D, Corson TW, et al. Minimal 16q genomic loss implicates cadherin-11 in retinoblastoma. Mol Cancer Res. 2004; 2(9):495–503. PMID: 15383628

79. Bhattacharya R, Mukherjee N, Dasgupta H, et al. Frequent alterations of SLIT2-ROBO1-CDC42 signalling pathway in breast cancer: clinicopathological correlation. J Genet. 2016; 95(3):551–563. https://doi.org/10.1007/s12041-016-0678-2 PMID: 27659325

80. Ye Y, Yang S, Han Y, et al. Linc00472 suppresses proliferation and promotes apoptosis through elevating PDCD4 expression by sponging miR-196a in colorectal cancer. Aging (Albany NY). 2018; 10(6):1523–1533. https://doi.org/10.18632/aging.101488 PMID: 29930217

81. Smulski CR, Eibel H. BAFF and BAFF-Receptor in B Cell Selection and Survival. Front Immunol. 2018; 9:2285. Published 2018 Oct 8. https://doi.org/10.3389/fimmu.2018.02285 PMID: 30349534

82. Stamm H, Oliveira-Ferrer L, Grossjohann EM, et al. Targeting the TIGIT-PVR immune checkpoint axis as novel therapeutic option in breast cancer. Oncoimmunology. 2019; 8(12):e1674605. Published 2019 Oct 12. https://doi.org/10.1080/2162402X.2019.1674605 PMID: 31741778

83. Song Q, Qin S, Pascal LE, et al. SIRPB1 promotes prostate cancer cell proliferation via Akt activation. Prostate. 2020; 80(4):352–364. https://doi.org/10.1002/pros.23950 PMID: 31905248

84. Gortat A, Sancho M, Mondragón L et al. Apaf1 inhibition promotes cell recovery from apoptosis. Protein Cell. 2015; 6(11):833–843. https://doi.org/10.1007/s13238-015-0200-2 PMID: 26361785

85. Okosun J, Wolfson RL, Wang J et al. Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma [published correction appears in Nat Genet. 2016 May 27;48(6):700]. Nat Genet. 2016; 48(2):183–188. https://doi.org/10.1038/ng.3473 PMID: 26691987

86. Lackraj T, Goswami R, Kridel R. Pathogenesis of follicular lymphoma. Best Pract Res Clin Haematol. 2018; 31(1):2–14. https://doi.org/10.1016/j.beha.2017.10.006 PMID: 29452662

87. Kotsiou E, Okosun J, Besley C, et al. TNFRSF14 aberrations in follicular lymphoma increase clinically significant allogeneic T-cell responses. Blood. 2016; 128(1):72–81. https://doi.org/10.1182/blood-2015-10-679191 PMID: 27103745

88. Schneider C, Kon N, Amadori L, et al. FBXO11 inactivation leads to abnormal germinal-center formation and lymphoproliferative disease. Blood. 2016; 128(5):660–666. https://doi.org/10.1182/blood-2015-11-684357 PMID: 27166359

89. Leventaki V, Rodic V, Tripp SR, et al. TP53 pathway analysis in paediatric Burkitt lymphoma reveals increased MDM4 expression as the only TP53 pathway abnormality detected in a subset of cases. Br J Haematol. 2012; 158(6):763–771. https://doi.org/10.1111/j.1365-2141.2012.09243.x PMID: 22845047

90. Küppers R. CD83 in Hodgkin lymphoma. Haematologica. 2018; 103(4):561–562. https://doi.org/10.3324/haematol.2018.188870 PMID: 29572346