The genomic landscape of canine diffuse large B-cell lymphoma identifies distinct subtypes with clinical and therapeutic implications

Diana Giannuzzi  
University of Padua

Laura Marconato  
Antonella Fanelli  
University of Turin

Luca Licenziato  
University of Turin

Raffaella De Maria  
University of Turin

Andrea Rinaldi  
Institute of Oncology Research

Luca Rotta  
IEO, European Institute of Oncology, IRCCS

Nicole Rouquet  
Urodelia

Giovanni Birollo  
University of Turin

Piero Fariselli  
University of Turin

Afua Mensah  
Università della Svizzera italiana

Francesco Bertoni  
Institute of Oncology Research

Luca Aresu  
luca.aresu@unito.it  
University of Turin  https://orcid.org/0000-0002-7893-1740

Article

Keywords: Diffuse large B-cell lymphoma, lymphoma, canine genome

Posted Date: September 21st, 2021
Abstract

Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoid neoplasm in dogs, being characterized by a remarkable degree of clinical heterogeneity not completely explained by molecular data. This poses a major barrier to understanding the disease and response to therapy or when using dogs with DLBCL in clinical trials.

We performed an integrated analysis of genome, exome and RNA sequencing in a cohort of 77 canine DLBCL to define the genetic landscape of this tumor. A wide range of signaling pathways and cellular processes were found in common with human DLBCL, but the frequencies of the most recurrently mutated genes were different overall. We developed a prognostic model integrating clinical, exonic variants and transcriptomic features to predict outcome in dogs with DLBCL. These results comprehensively define the genetic drivers of canine DLBCL and can be prospectively utilized to identify new therapeutic opportunities.

Introduction

Lymphoma in domestic dogs is considered a representative and highly predictive spontaneous model for human disease. In particular, the complex genetics interplay, the intact immune system, the environmental exposures and the increasing incidence represent powerful elements for translational studies. Among the many lymphoma subtypes, canine diffuse large B-cell lymphoma (cDLBCL) is the most common, accounting for approximately 60–70% of hematological malignancies in this species.

Current survival rates for cDLBCL after chemotherapy or chemo-immunotherapy are usually disappointing and dogs show markedly different clinical courses and treatment responses, demonstrating a heterogeneous clinical behavior and a difficulty in anticipating the outcome. Proposed cDLBCL prognostic classification systems are based on bone marrow infiltration, substage, mitotic rate and histological features (centroblastic and immunoblastic) without consideration of the mechanisms underlying tumorigenesis. Transcriptomics have shed some light on the pathogenesis of cDLBCL, revealing similarities with its human counterpart, but also important differences that should be considered in veterinary and comparative clinical trials. Compared to normal B-cells, cDLBCL present active NF-kB signaling induced by antigen engagement of the B-cell receptor. Additionally, up-regulation of several Toll-like receptors suggests a pathogenesis similar to human activated B-cell (ABC) DLBCL and the activation of immune-related signatures is correlated with an inferior outcome. Indeed, dogs with a shorter overall survival and tumor-free interval show a higher expression of transcripts coding for proteins involved in JAK/STAT signaling, microenvironment, immune system and p53 pathway.

Recent studies have described the mutational spectrum of canine lymphomas and provided a comprehensive catalog of somatic mutations in coding regions. However, whole genome sequencing (WGS) has not yet been used to investigate canine DLBCL. One large study based on whole exome sequencing (WES) investigated canine B-cell lymphomas obtained from three predisposed breeds (Boxer,
Golden Retriever and Cocker Spaniel) and found that both \textit{TRAF3} and \textit{MAP3K14} were frequently mutated. Notably, \textit{FBXW7} mutations occurring in a specific codon recurrently mutated in several human cancers were identified. Despite the large number of cases included in that study the mutations described may result breed-specific, thereby masking the true heterogeneous mutational landscape of canine lymphoma. Also, since tumors were not appropriately classified according to WHO criteria and clinico-pathological features were not reported, the clinical significance of these results are unknown.

In human medicine the integration of next generation sequencing technologies in clinical practice holds great promise for personalized medicine but correlations between genotype and phenotype are critical for the interpretation of these analyses. Veterinary oncology has only recently been modeling the same approach but the process has been hampered by several difficulties. Firstly, large multi-institutional molecular studies comprising datasets of fully characterized canine tumors often suffer from a lack of funding. Secondly, understanding the link between molecular aberrations and prognosis is challenging since treatment and outcome are strongly influenced by the owner. Thirdly, even if genetic alterations are defined, their functional impact and clinical validation are often unknown, thus preventing the identification of new therapeutic targets and the prediction of prognosis.

To address these issues and to prospectively inform clinical trials of canine DLBCL, we performed a comprehensive multiomics profiling of \textit{de novo} diagnosed cDLBCL with the goal of clarifying the genetic changes within this tumor. Genetic data obtained from WES and WGS were correlated with clinico-pathological features. Finally, an integrated model comprising mutations, copy number aberrations (CNAs) and transcriptome was designed to predict overall survival and tumor-free interval. Further, clinically significant mutations were validated in an independent cohort of cDLBCL.

**Results**

**Study population and clinical characteristics**

We enrolled 77 \textit{de novo} cDLBCL cases with matched normal tissues for WES and 10 cDLBCLs for WGS. A subset of the 77 dogs (n=43) were also analyzed for RNA-seq.

Complete patient demographics and clinical presentation are described in Supplementary Data 1. Mixed-breed dogs (n=19; 24.7%), German shepherds (n=9; 11.7%), Rottweilers (n=7; 9.1%) and Golden retrievers (n=4; 5.2%) were the most common breeds. There were 40 (51.9%) females and 37 (48.1%) males. At diagnosis, median age was 7 years (range 3-15 years) and median weight was 30.8 kg (range, 4.5-81.3 kg). Regarding clinical stage, 48 (62.3%) dogs had stage V disease, 28 (36.4%) had stage IV and only one (1.3%) dog had stage III disease. Forty-nine (63.6%) dogs were asymptomatic at presentation (substage a), while 28 (36.4%) showed symptoms (substage b). Overall, 22 (28.6%) dogs had bone marrow infiltration, with a median of 3% neoplastic cells (range, 1-50%). Peripheral blood was infiltrated in 43 (55.8%) dogs, with a median of 4% neoplastic cells (range, 1-74%). At presentation, 33 (42.9%) dogs had an increased level of serum LDH and 23 (29.9%) had received steroids before being diagnosed with
lymphoma. Treatment significantly affected both time to progression (TTP) and lymphoma specific survival (LSS) (Fig. 1a, b). Indeed, dogs treated with chemo-immunotherapy (n=45; 58.4%) showed a better outcome compared to dogs receiving chemotherapy only (n=32; 41.6%). In these dogs, shorter LSS was significantly associated with substage and bone marrow infiltration. In dogs receiving chemotherapy only, peripheral blood infiltration correlated with both TTP and LSS, whereas bone marrow infiltration correlated with LSS only (Supplementary Data 2).

**Landscape of somatic mutations in cDLBCL**

Based on WES, the median sequencing depth of targeted regions was 265 (range 140-394) for tumors and 246 (range 110-625) for normal samples, with a mapping rate of 99%. Collectively, the total number of short somatic variants identified across tumors ranged from 93 to 2,899, with an average of 282. Of these variants, 10.3% to 28.7% were annotated as protein coding variants with an average of 18.4%, including 4.9% insertions and deletions (indels) and 95.1% single nucleotide variants (SNVs) (Fig. 2). Among the latter, 68.5% were missense (range, 47.3%-84.3%) (Supplementary Fig. 1). By SIFT, 1,866 were classified as deleterious and 1,220 as tolerated. The full list of the nucleotide variants is reported in Supplementary Data 3.

**cDLBCL is characterized by recurrent mutations in specific protein-coding genes**

A total of 2,831 protein coding genes showed a nonsynonymous somatic variant in at least one tumor for a total of 3,769 protein coding variants, and 2,368 genes were mutated only in one sample. More importantly, eight genes (TRAF3, SETD2, POT1, TP53, MYC, FBXW7, DDX3X, and TBL1XR1) were recurrently mutated in at least 15% of the dogs. The top 43 most frequently mutated genes are shown in Fig. 3. Several candidate cancer genes were previously identified as genetic drivers in canine cancers (TRAF3, SETD2, POT1, TP53), but others have never been reported in dogs (TBL1XR1, H3C8, DIAPH2, EHD3).

In concordance with previous studies, TRAF3 was the most frequently mutated gene in our cohort (53%). A total of 62 mutations were identified and 16 dogs carried multiple aberrations. Of these mutations, 41.9% and 29% were frameshift and nonsense variants, respectively, while 25.8% were missense (Supplementary Table 1). Exon 11 (ENSCAFT00000028719.4) was the most affected with a total of 31 mutations (Fig. 4a). The second most frequently mutated gene was SETD2. We identified 29 somatic mutations in 24 dogs (31%). Nonsense variants were the most frequent (37.9%), followed by frameshift and missense variants (31%) (Fig. 4b and Supplementary Table 1). POT1 mutations were identified in 28.6% of cases. Missense mutations represented 54.5% of variants, while nonsense and frameshift mutations were 27.3% and 18.2%, respectively (Fig. 4c and Supplementary Table 1). Interestingly, three dogs carried the same mutation (G653R). Finally, 24 mutations in TP53 were detected in 19 dogs (24.7%)
and 19 were missense (79.2%), 3 frameshift (12.5%) and 2 nonsense variants (8.3%) (Fig. 4d and Supplementary Table 1). All but one mutation affected the p53 DNA-binding domain. Five dogs presented two TP53 variants and one of them also carried a frameshift deletion and a nonsense variant on the same allele.

We filtered the protein coding somatic variants for known cancer genes using the COSMIC dataset (v92, 723 genes) and a total of 172 cancer genes with at least one protein coding variant were retrieved (Supplementary Data 4). By comparing the genetic alterations in cDLBCLs with several human DLBCL datasets, 1902 genetic alterations in driver genes were shared, including MYC and TP53 (Supplementary Data 4).

To explore patterns of co-mutation and mutual exclusivity, we examined pairwise overlaps using Fisher’s test and mutual exclusion among the top cancer genes. Significant pairwise interactions (FDR=0.1) are depicted in Fig. 5. Notably, PLEC mutations exclusively co-occurred with SETD2, while MAP3K14 mutations were mutually exclusive with TRAF3.

The tumor mutational burden (TMB) ranged from 0.12 to 2.94 (mean, 0.32) and was significantly higher in tumors presenting SETD2 and TP53 mutations compared to wild type (wt) tumors (p<0.05). To identify the potential contributions of the mutational processes within tumor exomes, we applied a Bayesian treatment of non-negative matrix factorization approach. The analysis revealed that the predominant mutational process in all the tumors was Signature 1A which is the product of cytosine deamination at CpG sites due to ageing (Fig. 6).

To construct a comprehensive view of the common genetic alterations underlying cDLBCLs, we grouped genetic aberrations targeting specific oncogenic signaling pathways, including genes and copy number aberrations (CNAs) affected in >5% of cases. In addition to activation of the NF-kB signaling pathway (80% of the dogs), we identified chromatin remodeling and histone modifications (73%), and TNF signaling pathways (66%). The cell cycle was also frequently altered (64%), including alterations in genes with known roles in the G1/S checkpoint (amplifications of MYC and mutations of FBXW7) and the G2/M checkpoint (mutations of TP53) (Fig. 7).

**CNAs in cDLBCL are associated with clinical outcome**

Regions of somatic CNA were defined using WES segmentation data. The proportion of the tumor genome showing chromosomal aberrations ranged from 0.1 to 19.9% per genome (mean 6.8%) and the median number of CNAs was 35 (range, 1-487) (Fig. 8).

To reliably detect recurrent changes in tumors, the GISTIC algorithm was used. A total of 78 significant regions were identified including 24 gains mainly involving CFA 5, 6, 9 13, and 17, 27 and 31, as well as 54 losses mainly involving CFA 1, 6, 17, 26, and 38 (Supplementary Data 5). In line with previous
observations, the most extensive and highly recurrent CNAs were retrieved in CFA 13 overlapping the MYC locus (44%) and CFA 31 (30%), followed by focal gains in CFA 5 and 27. Losses of CFA 14 were identified in 15 dogs (19%). A total of 17 and 20 CNAs were found significantly associated with TTP and LSS, respectively (FDR<0.05) (Supplementary Data 6). In particular, the gain in the whole CFA13 (Supplementary Fig.2 and Supplementary Data 7) was significantly associated with TTP in dogs treated with chemo-immunotherapy.

The potential associations between CNAs and somatic mutations were investigated. We found that focal losses in CFA8 and large broad gains in CFA 13 were strongly associated with TRAF3, SETD2, and TP53 alterations. Moreover, gains in CFA 31 were only associated with TP53 mutations and losses in CFA14 with POT1 mutations (Fig. 9).

The SNV and CNA profiles for each sample were used to analyze the corresponding clonal architecture and revealed the subclonal structure of cDLBCL. To minimize the effect of CNAs and regions with loss of heterozygosity, we considered SNVs in neutral copy number regions. The number of clusters ranged from one to seven (Supplementary Fig. 3) and the number of deleterious variants increased as the number of subclones increased (σ=0.245, p <0.001).

We compared the number of subclones with the clinical features and found a negative correlation between the number of subclones and TTP in dogs treated with chemotherapy only (σ=0.149, p=0.020).

**Recurrently mutated genes associate with clinico-pathological features of cDLBCL**

Purebred dogs were more frequently affected by POT1 mutations than mixed breed dogs. Further, dogs with TP53 and SETD2 mutations (19 and 24 dogs, respectively) were significantly older compared to dogs with wt genes (p=0.013 and p=0.001, respectively). In contrast, TRAF3 and PHC3 mutations occurred more frequently in younger dogs. SETD2 mutations significantly affected females (71% of SETD2mut dogs, p= 0.027) and were inversely correlated to bone marrow infiltration (12% in mutated dogs compared to 36% in wt ones, p=0.039). Among the dogs carrying SETD2 mutations, mixed breed dogs were prevalent, and the median weight of mutated dogs was lower than the wt group. Dogs with MYC and DDX3X mutations had a higher percentage of peripheral blood infiltration. Also, DDX3X mutations were significantly associated with bone marrow infiltration and a higher LDH level. In line with these results, all dogs with DDX3X mutations had stage V disease. Supplementary Tables 2 and 3 show the significant associations obtained by comparing mutations to clinico-pathological features.

The potential association of mutations with clinical outcome has never been explored in dogs with DLBCL. We first examined the association of specific mutations with survival in all cDLBCL cases. In dogs treated with chemotherapy only, TP53 mutations were correlated with a shorter TTP and LSS. The median TTP for TP53 mutant and wt dogs was 32 and 98 days, respectively, while the median LSS for TP53 mutant and wt dogs was 60 and 176 days, respectively (Fig. 1c, d and Supplementary Data 2).
Similarly, dogs carrying MYC mutation were characterized by a shorter TTP and LSS (Fig. 1e, f and Supplementary Data 2).

In dogs treated with chemo-immunotherapy, POT1 mutations were correlated with a shorter LSS. The median LSS for dogs with mutant and POT1 wt was 330 and 547 days, respectively (Fig. 1g and Supplementary Data 2). Likewise, dogs carrying TP53 mutation were associated with a shorter median LSS (Fig. 1h and Supplementary Data 2). Also, mutations in both genes were not associated with any other clinical parameters. Interestingly, the co-occurrence of TP53 and FBXW7 mutations resulted in shorter TTP compared to dogs with only TP53 mutation (Supplementary Table 4).

Finally, TMB was significantly associated with tumor-free interval but not with overall survival in dogs treated with chemo-immunotherapy. Also, the median TTP was significantly shorter in dogs with TMB >0.28 compared to those with a TMB<0.21 (Supplementary Fig. 4a and Supplementary Data 2), and the median LSS was shorter in dogs with TMB>0.28 compared to those with a TMB<0.21 (Supplementary Fig. 4b and Supplementary Data 2).

**Aberrations in TRAF3, SETD2, POT1, TP53, MYC and DDX3X define specific transcriptional signatures in cDLBCL**

Gene expression profiling is now used to classify human DLBCL subtypes with clinical relevance, whereas in dogs attempts to apply a similar approach have failed. Conversely, signatures defined by T-cell mediated immune response, apoptosis, JAK/STAT signaling, microenvironment, inflammatory response and p53 pathway are enriched in cDLBCL with a worse prognosis. Here, we applied a clustering analysis using these prognostic signatures to separate tumors and identify genetic alterations associated with the two groups. RNA-seq data were available for 43 dogs and using previously published nomenclature 21 tumors were classified as cold DLBCL and 22 as hot DLBCL. The outcome was worse in the second group, but not significantly. Also, the two groups shared the vast majority of cancer genes and CNA at statistically indistinguishable frequencies. However, cold cDLBCL had frequent mutations in TRAF3 (75%), conversely dogs belonging to hot DLBCL showed rare mutations in TRAF3 (13%). Gene Set Enriched Analysis (GSEA) indirectly confirmed this result revealing an enrichment of gene sets related to immune and inflammatory response and T-cell activation in TRAF3 wt tumors compared to TRAF3 mut (Supplementary Data 8).

We also investigated the downstream transcriptomic changes in cDLBCLs characterized by mutations in the most frequent candidate cancer genes. Differential analysis was performed comparing dogs with mutated and wt genes. A total of 867, 136, 92, 51, and 26 genes were differentially expressed in DDX3X, SETD2, TP53, MYC, and POT1 comparisons, respectively (Supplementary Table 5). By GSEA an enrichment of gene signatures involved in the regulation of factors of the translation initiation complex and human Burkitt lymphoma were identified in dogs bearing DDX3X mutations (Supplementary Data 9). Dogs with MYC mutations were characterized by signatures associated with tumor microenvironment...
and apoptosis (Supplementary Data 10). Finally, the expression levels of the immune signatures characterizing hot cDLBCLs were positively associated with TMB (Wilcoxon rank-sum test, p < 0.05).

Integration of -omics and clinico-pathological features predicts survival in cDLBCL

We developed a multivariate supervised learning approach for defining the association of survival with clinico-pathological variables, genetic features and gene expression data. Among all, the Cox model outperformed the random forest models and was considered for further steps (Supplementary Fig. 5a, b). The features were combined to generate a prediction score and cDLBCLs were divided into two subgroups based on their predicted risk (long and short survival). The best performing scores for LSS (AUROC= 0.95) and TTP (AUROC=0.87) were obtained with the following variables: age, bone marrow infiltration (%), treatment (chemotherapy vs chemo-immunotherapy), TP53 genetic status (mut vs wt), and STAP2 and G3BP2 gene expression as logCPM (Supplementary Fig. 5c, d). By excluding STAP2 and G3BP2 expression data AUROC for LSS and TTP dropped to 0.90 and 0.79, respectively (Supplementary Fig. 5a, b). Furthermore, when TP53 genetic data was removed from the analysis AUROC for LSS and TTP dropped further to 0.83 and 0.74, respectively (Supplementary Fig. 5a, b). To prospectively use these data in clinical practice we have developed an interactive webtool (http://compbiomed.hpc4ai.unito.it/canine-dlbcl) to predict survival and tumor relapse using clinico-pathological data, transcriptomic and genomic features.

CNAs identified by WGS are associated with clinical outcome

We further performed WGS in 10 selected dogs to investigate a possible correlation between genetic profiling and clinical outcome. All dogs were treated with chemo-immunotherapy and were divided into two groups, namely “bad responders” (n=5) and “good responders” (n=5).

TTP and LSS were significantly different between the two groups (Supplementary Fig. 6a, b). The median sequencing depth was 20 (range 15-26) for cDLBCLs and 12 (range 7-18) for normal samples. Based on WGS, a median of 35 somatic coding mutations (range: 27–343) was identified in “bad responders” and 32 (range: 17–47) in good responders. Mutation signatures were analyzed, and three prevalent signatures were uncovered in “bad” and “good” responders (Supplementary Fig. 7a, b).

Although differential exposure analysis revealed no significant difference between the two groups, the signature in “good responders” that mainly contributed to the mutation counts matched COSMIC signature 1A, whereas this was true in only one “bad responder”.

In both groups two other signatures equally contributed to mutation spectra and resembled COSMIC signature 9 that is attributed to polymerase η and is implicated in somatic hypermutation in lymphoid cells, and signature 17, that has been previously described in human B-cell lymphomas. In addition,
PCSK5 and ORI11 were exclusively mutated in “bad responders”, whereas TASOR2, ENSCAF00000038553 and ENSCAF0000041925 were only mutated in “good responders” (Supplementary Data 11). CNA analysis identified two losses in CFA8 exclusively in “good responders” and one gain in CFA 13 in “bad responders”.

Confirmation of the negative prognostic value of mutated TP53 in an independent cohort of cDLBCL

We confirmed TP53 mutations by Sanger sequencing and further examined the exons 4-8 in a second group of 56 dogs affected by DLBCL, whose clinico-pathological features are reported in Supplementary Data 12. Fifteen dogs were mutated with 14 different mutations. Indeed, two dogs shared the same mutation. Eleven mutations were novel, while three were already reported in dbSNP albeit with unknown frequency. In three dogs, variants were classified as germline since retrieved in matched normal tissue. Among the somatic mutations, we classified nine missense, one frameshift deletion, one frameshift insertion, and one splice-acceptor variant. All missense mutations were predicted as deleterious by SIFT (Supplementary Data 13). Mutations in TP53 were associated with age and significantly enriched in dogs diagnosed with stage IV disease (p=0.001). The prognostic relevance of TP53 mutations was confirmed (Supplementary Data 14). Indeed, TP53mut dogs had a significantly shorter TTP and LSS compared to TP53wt dogs (Supplementary Fig. 8a, b).

Discussion

The application of genetic and transcriptomic analyses has led to an increased understanding of the biology of human DLBCL and allowed the design of target specific therapeutic approaches. Here, we integrated clinical features, somatic mutations, CNAs, and transcriptome in cDLBCL providing new data on the genetic hallmarks of this tumor and identifying multiple mutated genes associated with outcome thereby providing potential therapeutic targets.

TRAF3, mutated in 53% of the dogs, was the most frequently affected gene in our series. The gene product is a part of a complex including TRAF3 itself, TRAF2, BIRC2, BIRC3, and MAP3K14. The disruption of this complex leads to the activation of the non-canonical NF-κB pathway. TRAF3 inactivation has been described in both canine and human lymphoid neoplasms8,10−15, including in 15% of human DLBCL cases in which its contributes to activation of the NF-κB signaling cascade16. Indeed, in our series, dogs with mutated TRAF3 presented an active NF-κB transcriptome program and additional genes encoding members of the complex were recurrently mutated, especially MAP3K14. The latter codes for a kinase (also known as NIK, NF-κB inducing kinase), which phosphorylates NFKB2 (p100), causing its proteasomal processing and the formation of p52-containing NF-κB dimers which translocate into the nucleus to transactivate target genes. MAP3K14 mutations were largely mutually exclusive with TRAF3 mutations and overall, 80% of cDLBCL presented genetic lesions compatible with NF-κB activation, emphasizing the importance of this pathway in cDLBCL pathogenesis. These data provide potential
therapeutic targets\textsuperscript{17–19} but also highlight lesions associated with resistance to BTK inhibitors\textsuperscript{14,20} with implications for the management of dogs with cDLBCL and for the use of these animals as models for the human disease.

A total of 31\% of dogs harbored mutations in \textit{SETD2}, which was the second most frequently altered gene. \textit{SETD2} is a methyltransferase with various substrates and multiple lines of evidence support its tumor suppressor role\textsuperscript{21}. Its most well-characterized function is the trimethylation of lysine 36 on histone H3 (H3K36me3). Loss of this histone mark can lead to deregulated methylation of intragenic regions, RNA splicing, homologous recombination and mismatch repair. However, \textit{SETD2} also has non-histone substrates, including microtubules and transcription factors (for example, \textit{TP53}), whose aberrant modification can also contribute to cancer development and progression. Mutations in \textit{SETD2} have been reported in 20–40\% of canine osteosarcomas\textsuperscript{22} as well as in human tumors, including lymphomas\textsuperscript{23–27}. Differently from our observations in cDLBCL, \textit{SETD2} mutations are present in less than 10\% of human DLBCL\textsuperscript{26,27}, but more common in T-cell lymphomas\textsuperscript{23–25}. \textit{SETD2} was not the only gene encoding proteins involved in chromatin remodeling and transcription regulation. Over 70\% of cDLBCL contained at least one mutated gene of this class, including histone 3 members (\textit{H3C8} or \textit{H3C12}, 10\%), \textit{KDM6A}, \textit{SUZ12} (5\%), \textit{KDM2A}, \textit{KDM3B}, and \textit{EZH2} (3\%) (Fig. 10).

The mutations in \textit{H3C8} and \textit{H3C12} occurred in the same hotspot, determining amino acid 27 (or 28 according to COSMIC annotation) conversion from lysine (K) into methionine (M). This mutation has also been observed in human pediatric gliomas, where it defines a specific entity termed diffuse midline glioma, which is an infiltrative midline high-grade glioma with predominantly astrocytic differentiation\textsuperscript{28}. The mutation inhibits the activity of the polycomb repressive complex 2 (PRC2), composed of the K27 histone methyltransferase \textit{EZH2} (enhancer of zeste homolog 2) and the core accessory proteins EED, \textit{SUZ12}, and RbAp48\textsuperscript{29}. Since methionine cannot be methylated by \textit{EZH2}, gliomas bearing the K27M mutation present a global reduction of H3K27me levels, a modification associated with gene silencing, and DNA hypomethylation at many loci. However, perhaps due to a redistribution of the PRC2 complex, H3K27M mutated gliomas still retain a substantial number of genes with the H3K27me3 mark and are dependent on the remaining PRC2 enzymatic activity\textsuperscript{29–31}.

In cDLBCL we also observed recurrent inactivating heterozygous mutations in \textit{EZH2} and \textit{SUZ12}. The removal of di- and trimethyl groups from H3K27 is done by two histone demethylases, one of which, UTX, is encoded by \textit{KDM6A}. \textit{KDM6A} was mutated in 5\% cDLBCL. UTX forms a complex with H3K4 methyltransferases \textit{MLL2} (KMT2D)/\textit{MLL3} (KMT2C)\textsuperscript{29}. The inactivation of the two methyltransferases is observed in 20–30\% of human DLBCL\textsuperscript{32,33} and it determines a diminished global H3K4 methylation with deregulation of CD40, Toll-like, and B-cell receptor signaling pathways\textsuperscript{34,35}.

In our series we didn’t find any mutations in the acetyltransferases \textit{CREBBP} and \textit{EP300}, which activate transcription via acetylation of histone H3 lysine 27 (H3K27Ac). Besides impairing PRC2 activity,
H3K27M might also affect acetyltransferase activity, leading to aberrant gene expression and enhancer dysfunction \(^{36}\), and this could at least partially mimic the effects of *CREBBP* and *EP300* mutations.

KDM2A (JHDM1A/FBXL11) and KDM3B (JMJD1B/ JMJD1B) are two histone lysine demethylases that catalyze the demethylation of transcriptionally repressive mono- and di-methylated histone H3 lysine 36 (H3K36me1/me2) and 9 (H3K9me1/me2), respectively \(^{37}\). Not much is known about the roles of these two enzymes in cancer, which appears to be dependent on tumor type \(^{37}\). However, rare mutations are reported in human B-cell lymphomas \(^{38}\).

We also identified several novel targets of mutation in canine cancers, including in particular *TBL1XR1*. This gene is frequently mutated in human ABC DLBCL, specifically in the genetically defined MCD/Cluster 5 subtypes, which are characterized by frequent extranodal localization, immune-escape lesions, and an unfavorable clinical outcome \(^{39,40}\). *TBL1XR1* mutations were associated with a poor prognosis but the low frequency of dogs with this alteration prevented statistical significance being reached. TBL1XR1 is a core component of the SMRT/NCOR1 transcriptional repressor complexes, thus also contributing to transcription regulation.

Similarly to the NF-κB signaling cascade, the high frequency of alterations in genes involved in chromatin regulation and transcription regulation is of dual relevance to veterinary and comparative medicine. Evidence from preclinical studies indicate the potential use of therapeutic agents including EZH2 inhibitors for KDM6A \(^{41}\) and H3K27M mutants \(^{31}\), demethylating agents for KDM6A mutants and HDAC3 inhibitors for TBL1XR1 mutants \(^{42}\). The latter \(^{43}\) might also show activity in H3K27M mutants, due to the possibly reduced acetyltransferase activity of CREBBP and EP300 on H3K27M \(^{36}\). LSD1 inhibitors \(^{44,45}\) and especially KDM5 inhibitors \(^{46}\) could work in KDM6A mutants, which might have concomitant deregulated MLL2/MLL3 activity \(^{29}\). However, it is important to keep in mind that although cDLBCL and its human counterpart appear phenotypically similar, subtle genetic differences exist between the two.

In our series, we identified recurrently mutated genes found in WES studies of other canine cancers including osteosarcoma, melanoma and lymphoma \(^{22,47}\). Compared to the previous two studies in canine lymphoma, the frequency of mutations for the top four genes was overall higher in the current work. This difference could be attributable to several reasons. First, we included fully characterized cDLBCL, whereas in previous studies, a general diagnosis of B-cell lymphoma was reported and mostly obtained by fine needle aspiration. Second, in our series, normal DNA was available for every dog and used as a match for variant calling. Finally, from a technical point of view, a wider whole exome enrichment kit was used here.

Recent advances in chemo-immunotherapy have offered new options for the treatment of canine cancers. The addition of APAVAC to CHOP-based chemotherapy has dramatically prolonged both remission time and overall survival in dogs with DLBCL. This data was confirmed in our cohort, but here we also identified genetic markers that were able to predict clinical behavior. While immune responses are generally considered to be durable and adaptable, thus conferring a significant advantage over...
chemotherapy as single treatment modality, not all patients will respond to immunotherapy and pulling out the non-responders is challenging. In light of this, predicting a patient’s response to immunotherapy to achieve a better outcome or save treatment costs remains an active area of clinical research. Here, we identified two groups of animals differing for their clinical outcome. Dogs older than 10 years, having bone marrow involvement, and TP53 mutations had a poor outcome and a small benefit from chemo-immunotherapy. Conversely, the absence of such features identified a cDLBCL subset with good prognosis and an important gain in survival if treated with chemo-immunotherapy.

The negative prognostic impact of TP53 mutations has been reported in human DLBCL, but it is still unclear in dogs. In our series, TP53 mutations were frequently found by WES, and these findings were validated in a second group of dogs, maintaining a similar frequency and prognostic significance. Loss of TP53 causes disruption of checkpoint responses to DNA damage and contributes to genomic and chromosomal instability. Here, most of the mutations affected the p53 DNA-binding domain and we can thus hypothesize an inactivating effect, even if this would require an experimental validation. In addition, the co-occurrence of TP53 and FBXW7 mutations resulted in a worse outcome than TP53 mutation alone.

Other than TP53, TRAF3 mutants were associated with a better outcome than TRAF3 wild-type animals; conversely POT1 mutations, previously reported in cDLBCL, were associated with poor outcome in our dogs. POT1 mutations contribute to cancer development in multiple ways, including human chronic lymphocytic leukemia, where the predominant effects of POT1 loss are increased telomere length in cells with telomerase activity and genomic instability contributing to tumorigenesis.

In conclusion, our results suggest that clinical trials testing new targeted agents in cDLBCL should be evaluated in the context of clinical features and genetic aberrations, including mutations affecting TP53, non-canonical NF-κB and chromatin remodeling.

**Methods**

**Animal recruitment and samples acquisition**

Dogs with newly diagnosed, previously untreated, multicentric cDLBCL of any World Health Organization (WHO) clinical stage were included in the study. To be eligible for enrollment, dogs had to undergo a complete staging work-up, consisting of history and physical examination, complete blood cell count with differential, serum biochemistry profile, thoracic radiographs and abdominal ultrasound, cytological evaluation of liver and spleen regardless of the ultrasonographic appearance, and immunophenotype determined by flow cytometry (FC) on a lymph node (LN) aspirate, peripheral blood and bone marrow aspirate. Before the initiation of therapy, all dogs underwent lymphadenectomy to confirm DLBCL histotype by routine histology and immunohistochemistry (CD3, CD20, CD79 and PAX5) and to provide material for vaccine generation. A portion of the neoplastic lymph node was always preserved in RNA-later and stored at -80°C. In addition to tumor samples, skin punch biopsies were obtained from all dogs...
included in the study to provide matched paired normal tissue. Dogs’ owners were required to give written informed consent. Approval for this study was granted by the Ministero dell’Istruzione, dell’ Università e della Ricerca Ethical Board (Number RBSI14EDX9).

Depending on owner’s preference, dogs were treated either with chemotherapy or with chemo-immunotherapy. Unvaccinated dogs received a CHOP-based protocol, including L-asparaginase, vincristine, cyclophosphamide, doxorubicin, lomustine and prednisone.

Regarding dogs receiving chemo-immunotherapy, the detailed method of the APAVAC vaccine preparation and the protocol used has been described elsewhere. Briefly, dogs received L-asparaginase, vincristine, cyclophosphamide, doxorubicin, lomustine, prednisone, and a total of 8 intradermal injections of 0.5 ml vaccine.

Treatment response was classified as complete remission (CR), partial remission (PR), stable disease (SD) or progressive disease (PD). Response was evaluated at each therapeutic session and was required to last for at least 28 days. Follow-up evaluation consisted of monthly physical examination, peripheral LNs size measurement and cytological evaluation during the first year, and every other month thereafter. Relapse was defined as the clinical reappearance and cytological evidence of lymphoma with or without FC confirmation in any anatomical site in dogs having experienced CR. Once relapse was confirmed, a complete restaging work-up was undertaken, and a second round of chemotherapy was offered. TTP was calculated as the interval between initiation of treatment and PD or relapse, whereas LSS was measured as the interval between initiation of treatment and lymphoma-related death. Dogs lost to follow-up or dead for lymphoma-unrelated causes before PD, as well as those still in CR at the end of the study, were censored for TTP analysis. Dogs alive at the end of the study, lost to follow-up or dead due to causes other than lymphoma were censored for LSS analysis.

**RNA/DNA isolation and sequencing**

Total RNA and DNA were extracted from all samples, using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. DNA concentration and quality were measured by Qubit fluorometer (Life Technologies Ltd, Paisley, UK) and Agarose gel electrophoresis. RNA concentration and integrity were measured respectively, in a NanoDrop ND-1000 spectrophotometer and a Bioanalyzer 2010 instrument (Agilent Technologies, Palo Alto, CA, USA). A total of 43 non-normalized libraries for RNA sequencing experiments were prepared by using SureSelect Strand Specific RNA-Seq Library Preparation kit (Agilent Technologies) and single end sequencing (50SE) was carried out on an Illumina HiSeq2500 (Illumina Inc., San Diego, CA, USA). Raw Illumina sequencing data are deposited in the SRA database (GenBank) under accession number SRP137798. All Illumina reads were analyzed with FastaQC software in order to assess sequence quality. A total of 154 high quality whole genome sequencing libraries were prepared using the Illumina-compatible KAPA HyperPlus Library Preparation Kit (Roche Sequencing and Life Science, Wilmington, MA). Exome capture was performed using Roche’s SeqCap EZ Share Prime Developer Kit (Roche Sequencing and Life Science, Wilmington, MA) for non-
human genomes following the SeqCap EZ HyperCap Workflow User’s Guide. The developer’s reagent (06684335001) was used in place of COT-1 and index-specific hybridization enhancing oligos were also used. Exome capture was performed using SeqCap EZ Canine Exome Custom Design (canine 140702_canFam3_exomeplus_BB_EZ_HX1 probe set, Roche Sequencing and Life Science, Wilmington, MA) in which probes were designed on the target enrichment design of 150 megabases developed by Broeckx et al. 52. Libraries were quantified using the Qubit DNA Assay Kit in a Qubit 2.0 Fluorometer (Thermo Fischer, Foster City, CA, USA) and quality was assessed using the Bioanalyzer 2100 instrument. For WGS, ten libraries (fragments ranging from 300 to 400 bp) were then pooled and sequenced on an Illumina NovaSeq 6000 platform in a paired-end (150 PE) mode. The chosen target sequencing coverage was 30x and 250x for WGS and WES, respectively.

**RNA-Seq, WES and WGS reads quality control, alignment and preprocessing**

Data quality control was first performed on raw Illumina reads for RNA-Seq, WES and WGS using FastQC v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/download.html) software. RNA-Seq reads trimming, mapping, gene quantification and differential expression analysis steps were performed using the pipeline previously described in Aresu et al. 7. We calculated the two prognostic transcriptomic subgroups as previously described 53. Briefly, we retrieved immune gene sets for immune cells, tumor infiltrating lymphocytes (TILs), proinflammatory molecules, cytokines and cytokine receptors, regulatory T-cells and immune checkpoints, and performed single sample GSEA to derive the enrichment score of each immune-related term. By applying unsupervised consensus clustering analysis, we separated cDLBCL into two subgroups (high immunity = “hot” and low immunity = “cold”).

For WES and WGS, Trimmomatic v.0.36 was used to select high-quality reads and remove adapter sequences. The paired-end clean reads were aligned against the CanFam3.1 reference genome from UCSC (https://genome-euro.ucsc.edu/cgi-bin/hgGateway?db=canFam3) using bwa-mem software v.0.7.15 with -M and -t 15 options 54. Pre-processing analysis for variant calling-ready bam file per sample (tumor and normal bam) was performed by various steps following the Genome Analysis Toolkit 4 (GATK4) v.4.1.0.0 best practices 55, namely Picard AddOrReplaceReadGroups (http://broadinstitute.github.io/picard), and GATK4 Mark Duplicates, Indel Realignment, and Base Quality Score Recalibration. Alignment quality control metrics were calculated using Picard CollectHSMetrics tools and GATK4 DepthOfCoverage (for WES and WGS, respectively).

**Somatic variant calling**

Somatic SNVs and indels of WES and WGS samples were identified using two callers: GATK4 (v.4.1.0.0) 56, Mutect2 57 and Strelka v.1.0.13/Manta v.1.6.0 58,59 using default arguments according to previous literature 60,61. Mutect2 was run in tumor-matched normal mode. To reduce germline artifacts, a panel of normals (PON) was built using the GATK4 CreateSomaticPanelOfNormals tool by using all matched normal samples (n=77). An additional filter was added to exclude known single-nucleotide polymorphisms as annotated in the dbSNP 146 downloaded from the Dog Genome SNP Database.
Strelka was run together with Manta for best somatic indels performance and for the processing of WES samples, the -exome and -callRegions parameters were added to restrict the calls to the targeted regions only. Resulting raw somatic calls of SNVs and indels from the two callers were then filtered using caller-specific filters and hard filters as defined in the respective developer’s guidelines. Briefly, Mutect2 filters were performed with GATK4 FilterMutectCalls tool and comprised, for instance, the removal of calls under germline risk, calls with low depth, artifacts related to sequencing platforms or repeat regions; hard filters were based on the filtering of variants present in dbSNP and PONs to avoid false positive calls (https://software.broadinstitute.org/gatk/documentation/article?id=11136). For Strelka, the default filters were used (https://github.com/Illumina/strelka/blob/v2.9.x/docs/userGuide/README.md#somatic). A consensus calling was then applied to filtered somatic calls from the two callers (in VCF version 4.1/4.2 format) by overlapping the remaining variants using the bcftools software package’s -isec option 63. Finally, resulting consensus VCF files were annotated with ANNOVAR 64. Mutation Annotation Format (MAF) using vcf2maf utility (https://github.com/mskcc/vcf2maf) was also created. Annotated VCFs and MAFs were used for all the downstream analyses.

Tumor Mutational Burden and Mutational Signatures analysis

Tumor mutation burden (TMB) was calculated by dividing the total number of single nucleotide variants by the amount of the genome covered by exome sequencing and the result was expressed as the number of SNV per megabase (Mb). Mutational signatures were predicted both in WES and WGS. Briefly, SNVs mutations were analyzed in the context of the flanking nucleotides (96 possible trinucleotide combinations) using a full Bayesian treatment to the non-negative matrix factorization (NMF) model implemented in the SigneR 65 R package. The number of mutational processes per Mb and their signatures have been evaluated for each sample, and the identified ones were compared with validated signatures in human cancer 66,67. Furthermore, WGS “bad” and “good responders” were compared in a differential exposure analysis to highlight signatures that were differentially active among the two groups (p < 0.05).

Copy Number Analysis

CNA analysis was performed on WES and WGS data using the NEXUS Copy Number v.10.0 software (Biodiscovery Inc., El Segundo, CA). CNAs and loss of heterozygosity (LOH) were identified starting from pre-processed BAM files using a matched approach (“BAM ngCGH-matched”) and against the CanFam3.1 reference genome. Then, a FASST2 segmentation algorithm with a significance threshold of 1.0e-12 was applied. The dbSNP was also given to the software to assess germline events. In the case of WGS, the bin size was reduced from 1000 (default for WES) to 500. Aberrations were defined as a minimum of 3 consecutive segments and a maximum probe spacing of 1000 kbp between adjacent probes before breaking a segment. Segments were classified with log2 tumor/reference ratio value of >0.5 as high gains, 0.5 to 0.2 as gains, -0.25 to -1.5 as losses, <-1.5 as big losses, and the heterozygous imbalance threshold was set to 0.4. Furthermore, independently from the automatically generated CNA
calls, each sample was visually inspected for using Nexus Copy Number software v.10.0 (Biodiscovery Inc., El Segundo, CA). Recurrent CNAs were determined within NEXUS using the GISTIC algorithm with a G-score cut-off of G > 1.0 and a significance of Q < 0.05. CNA frequency comparisons amongst groups (i.e. clinical features, presence/absence of somatic mutations) were performed with a two-tailed Fisher's exact test ($p < 0.05$).

**Intra-Tumor Heterogeneity analysis**

To verify the presence of multiple subclonal populations of cells tumor samples were inferred using sciClone R package applying default parameters according to the developer's manual. Copy number segments were extracted from WES samples using circular binary segmentation (CBS) method implemented in CNVkit, whereas mutations were extracted from Strelka/Mutect2 intersection VCF files. LOH regions inferred from NEXUS Copy Number were excluded.

**Functional annotation analysis**

The analysis comparing mutational processes and transcriptome to evaluate enriched gene sets in WT/mutated samples was performed using the GSEA in a preranked mode, with a threshold of significance of FDR < 5%. The curated canonical KEGG pathways (c2.cp.kegg.v7.4), GO terms (c5.all.v7.4), oncogenic (c6.all.v7.4) and immunologic (c7.all.v7.4) signatures, and Hallmarks (h.all.v7.4) catalogs, were from Molecular Signatures Database (MsigDB) were used.

**Prediction models**

Two predictive models were tested and implemented in the scikit-survival module (version 0.15.0) in Python (version 3.8.9): Cox's proportional hazard's model with elastic net penalty and random survival forest. Separate models were built for LSS and TTP. Performance of the models was evaluated by eight-fold cross-validation with two holdout sets, repeated twenty times on random permutations of the samples. One holdout set was needed to select the optimal model by hyperparameter grid search and the other was used to evaluate the final performance of the optimal model. Performance was evaluated by the area under the receiver operating characteristic (AUROC) for classifying dogs with survival time lower than the median survival in the dataset. Given the limited number of samples and the much larger number of available features and the observed poor performances of models fitted using too many features, we decided to eschew automatic feature selection methods. Thus, various datasets with different features were tested to find those that were most predictive and combine them in a single predictive model. Categorical clinical features were one-hot-encoded producing 41 numeric and binary features.

We tested a dataset of 77 samples with 2832 genetic mutations (binary) and clinical features and a dataset of 43 samples with clinical features and the top 100 most significantly differentially expressed between tumors and controls. Features whose regression coefficient were consistently different from zero
among the cross-validation fittings of the best performing models were selected and put together in the final dataset.

**PCR amplification and Sanger sequencing**

Sanger sequencing was performed to validate the protein coding somatic mutations identified by WES in *TP53* gene. Primer pairs were specifically designed using Primer3 (https://primer3.ut.ee/) and Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) tools (Supplementary Table 6). PCR was performed in a final volume of 20 µl using HotStarTaq DNA Polymerase kit (QIAGEN, Hilden, Germany) and 50 ng of gDNA with the following cycling conditions: initial denaturation at 95°C for 5 min, 33 cycles at 95°C for 30 s, at 58°C for 30 s and at 72°C for 40 s, with final extension at 72°C for 10 min. PCR products were purified using the ExoSAP-IT PCR Product Cleanup Reagent (Applied Biosystems, Foster City, CA, USA). Purified products were sequenced in the forward or reverse direction using the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) following manufacturer's instructions and analyzed on a SeqStudio Genetic Analyzer (Applied Biosystems). Sequencing electropherograms were manually inspected using Chromas v2.6.6 software. Mutations were identified by comparing sequences obtained from each tumor to the canFam3 reference genome and were classified as somatic when absent in matched normal tissue.

**Statistical analysis**

To explore the associations between clinico-pathological variables and mutated genes and identify different classes of clinico-pathological variables in mutated or wt individuals, we built classification trees using the recursive partitioning algorithm implemented in the *party* R package. Survival analysis was conducted using survival and survminer R packages. The following clinico-pathological variables were tested for their influence on TTP and LSS by means of univariate and multivariate Cox proportional hazard model: treatment (chemotherapy vs chemo-immunotherapy), breed (pure vs mixed), sex (female vs male), age (<10 years vs ≥10 years), weight (<10kgs vs ≥10kgs), stage (IV, V), substage (a, b), peripheral blood infiltration (%), bone marrow infiltration (%), presence of bone marrow infiltration (yes vs no), LDH activity (normal vs increased), pretreatment with steroids (yes vs no) and TMB (<0.21 vs >0.28). The influence of the mutational status of the 43 most frequently mutated genes (mut or wt) on both TTP and LSS was also evaluated. Variables with a p-value ≤0.200 were included in the multivariate analysis. For categorical variables, Kaplan-Meier curves were drawn and compared by means of log-rank test.

**Data availability statement**

Raw Illumina reads of RNA-Seq, WES and WGS are publicly available in SRA Archive with reference number SRP137798 and PRJNA752630.

**Declarations**
Author contributions

L.A. and F.B. designed the study, interpreted the data and wrote the manuscript; N.R. contributed to study design; L.M. provided samples and clinical data and contributed to manuscript revision; D.G., A.F., L.L., R.D.M., A.R., L.R. and A.A.M. carried out experiments; D.G. conducted bioinformatic analysis and contributed to data interpretation; D.G., A.F. and L.L. performed the analysis and visualization of data and conducted statistical analysis; G.B. and P.F. developed the in silico prediction models; all authors contributed to manuscript revision and approved the final draft.

Additional informations

Supplementary Information accompanies this paper.

Competing interests

The authors declare no competing interests.

References

1. Villarnovo, D., McCleary-Wheeler, A. L. & Richards, K. L. Barking up the right tree: advancing our understanding and treatment of lymphoma with a spontaneous canine model. *Curr Opin Hematol* 24, 359–366, doi:10.1097/MOH.0000000000000357 (2017).

2. Aresu, L. Canine Lymphoma, More Than a Morphological Diagnosis: What We Have Learned about Diffuse Large B-Cell Lymphoma. *Front Vet Sci* 3, 77, doi:10.3389/fvets.2016.00077 (2016).

3. Marconato, L. et al. Opportunities and challenges of active immunotherapy in dogs with B-cell lymphoma: a 5-year experience in two veterinary oncology centers. *J Immunother Cancer* 7, 146, doi:10.1186/s40425-019-0624-y (2019).

4. Valli, V. E., Kass, P. H., San Myint, M. & Scott, F. Canine lymphomas: association of classification type, disease stage, tumor subtype, mitotic rate, and treatment with survival. *Vet Pathol* 50, 738–748, doi:10.1177/0300985813478210 (2013).

5. Aresu, L. et al. Minimal residual disease detection by flow cytometry and PARR in lymph node, peripheral blood and bone marrow, following treatment of dogs with diffuse large B-cell lymphoma. *Vet J* 200, 318–324, doi:10.1016/j.tvjl.2014.03.006 (2014).

6. Richards, K. L. et al. Gene profiling of canine B-cell lymphoma reveals germinal center and postgerminal center subtypes with different survival times, modeling human DLBCL. *Cancer Res* 73, 5029–5039, doi:10.1158/0008-5472.CAN-12-3546 (2013).

7. Aresu, L. et al. New molecular and therapeutic insights into canine diffuse large B-cell lymphoma elucidates the role of the dog as a model for human disease. *Haematologica* 104, e256-e259, doi:10.3324/haematol.2018.207027 (2019).

8. Elvers, I. et al. Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* 25, 1634–1645, doi:10.1101/gr.194449.115
9. Leiserson, M. D., Reyna, M. A. & Raphael, B. J. A weighted exact test for mutually exclusive mutations in cancer. *Bioinformatics* **32**, i736-i745, doi:10.1093/bioinformatics/btw462 (2016).

10. Keats, J. J. *et al.* Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. *Cancer Cell* **12**, 131–144 (2007).

11. Demchenko, Y. N. *et al.* Classical and/or alternative NF-kappaB pathway activation in multiple myeloma. *Blood* **115**, 3541–3552, doi:10.1182/blood-2009-09-243535 (2010).

12. Otto, C. *et al.* Genetic lesions of the TRAF3 and MAP3K14 genes in classical Hodgkin lymphoma. *Br. J. Haematol.* **157**, 702–708, doi:10.1111/j.1365-2141.2012.09113.x (2012).

13. Rossi, D. *et al.* Alteration of BIRC3 and multiple other NF-kappaB pathway genes in splenic marginal zone lymphoma. *Blood* **118**, 4930–4934, doi:10.1182/blood-2011-06-359166 (2011).

14. Rahal, R. *et al.* Pharmacological and genomic profiling identifies NF-kappaB-targeted treatment strategies for mantle cell lymphoma. *Nat. Med.* **20**, 87–92, doi:10.1038/nm.3435 (2014).

15. Bushell, K. R. *et al.* Genetic inactivation of TRAF3 in canine and human B-cell lymphoma. *Blood* **125**, 999–1005, doi:10.1182/blood-2014-10-602714 (2015).

16. Zhang, B. *et al.* An Oncogenic Role for Alternative NF-κB Signaling in DLBCL Revealed upon Deregulated BCL6 Expression. *Cell Rep.* **11**, 715–726, doi:10.1016/j.celrep.2015.03.059 (2015).

17. Moore, C. R., Edwards, S. K. & Xie, P. Targeting TRAF3 Downstream Signaling Pathways in B cell Neoplasms. *J. Cancer Sci. Ther.* **7**, 67–74, doi:10.4172/1948-5956.1000327 (2015).

18. Whillock, A. L., Mambetsariev, N., Lin, W. W., Stunz, L. L. & Bishop, G. A. TRAF3 regulates the oncogenic proteins Pim2 and c-Myc to restrain survival in normal and malignant B cells. *Sci. Rep.* **9**, 12884, doi:10.1038/s41598-019-49390-9 (2019).

19. Mambetsariev, N., Lin, W. W., Wallis, A. M., Stunz, L. L. & Bishop, G. A. TRAF3 deficiency promotes metabolic reprogramming in B cells. *Sci. Rep.* **6**, 35349, doi:10.1038/srep35349 (2016).

20. Diop, F. *et al.* Biological and clinical implications of BIRC3 mutations in chronic lymphocytic leukemia. *Haematologica* **105**, 448–456, doi:10.3324/haematol.2019.219550 (2020).

21. Licht, J. D. SETD2: a complex role in blood malignancy. *Blood* **130**, 2576–2578, doi:10.1182/blood-2017-10-811927 (2017).

22. Sakthikumar, S. *et al.* SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma. *Cancer Res.* **78**, 3421–3431, doi:10.1158/0008-5472.Can-17-3558 (2018).

23. Ji, M. M. *et al.* Histone modifier gene mutations in peripheral T-cell lymphoma not otherwise specified. *Haematologica* **103**, 679–687, doi:10.3324/haematol.2017.182444 (2018).

24. McKinney, M. *et al.* The Genetic Basis of Hepatosplenic T-cell Lymphoma. *Cancer Discov.* **7**, 369–379, doi:10.1158/2159-8290.Cd-16-0330 (2017).

25. Laginestra, M. A. *et al.* Whole exome sequencing reveals mutations in FAT1 tumor suppressor gene clinically impacting on peripheral T-cell lymphoma not otherwise specified. *Mod. Pathol.* **33**, 179–187, doi:10.1038/s41379-019-0279-8 (2020).
26. Zhang, J. et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 1398–1403, doi:10.1073/pnas.1205299110 (2013).

27. Li, J. et al. SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget*, **7**, 50719–50734, doi:10.18632/oncotarget.9368 (2016).

28. Fuller, C. E., Jones, D. T. W. & Kieran, M. W. New Classification for Central Nervous System Tumors: Implications for Diagnosis and Therapy. *American Society of Clinical Oncology Educational Book*, 753–763, doi:10.1200/edbk_175088 (2017).

29. Ezponda, T. & Licht, J. D. Molecular pathways: deregulation of histone h3 lysine 27 methylation in cancer-different paths, same destination. *Clin. Cancer Res.*, **20**, 5001–5008, doi:10.1158/1078-0432.CCR-13-2499 (2014).

30. Sturm, D. et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell*, **22**, 425–437, doi:10.1016/j.ccr.2012.08.024 (2012).

31. Mohammad, F. et al. EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.*, **23**, 483–492, doi:10.1038/nm.4293 (2017).

32. Pasqualucci, L. et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.*, **43**, 830–837, doi:10.1038/ng.892 (2011).

33. Morin, R. D. et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*, **476**, 298–303, doi:10.1038/nature10351 (2011).

34. Zhang, J. et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat. Med.*, **21**, 1190–1198, doi:10.1038/nm.3940 (2015).

35. Ortega-Molina, A. et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat. Med.*, **21**, 1199–1208, doi:10.1038/nm.3943 (2015).

36. Trovato, M., Patil, V., Gehre, M. & Noh, K. M. Histone Variant H3.3 Mutations in Defining the Chromatin Function in Mammals. *Cells*, **9**, doi:10.3390/cells9122716 (2020).

37. McCann, T. S. et al. Biology and targeting of the Jumonji-domain histone demethylase family in childhood neoplasia: a preclinical overview. *Expert Opin. Ther. Targets*, **23**, 267–280, doi:10.1080/14728222.2019.1580692 (2019).

38. Tate, J. G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.*, **47**, D941-d947, doi:10.1093/nar/gky1015 (2019).

39. Chapuy, B. et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat. Med.*, **24**, 679–690, doi:10.1038/s41591-018-0016-8 (2018).

40. Wright, G. W. et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell*, **37**, 551–568 e514, doi:10.1016/j.ccell.2020.03.015 (2020).
41. Schulz, W. A., Lang, A., Koch, J. & Greife, A. The histone demethylase UTX/KDM6A in cancer: Progress and puzzles. *Int. J. Cancer* **145**, 614–620, doi:https://doi.org/10.1002/ijc.32116 (2019).
42. Venturutti, L. *et al.* TBL1XR1 Mutations Drive Extranodal Lymphoma by Inducing a Pro-tumorigenic Memory Fate. *Cell* **182**, 297–316.e227, doi:10.1016/j.cell.2020.05.049 (2020).
43. Mondello, P. *et al.* Selective Inhibition of HDAC3 Targets Synthetic Vulnerabilities and Activates Immune Surveillance in Lymphoma. *Cancer Discov.* **10**, 440–459, doi:10.1158/2159-8290.CD-19-0116 (2020).
44. Hollebecque, A. *et al.* Phase I Study of Lysine-Specific Demethylase 1 Inhibitor, CC-90011, in Patients with Advanced Solid Tumors and Relapsed/Refractory Non-Hodgkin Lymphoma. *Clin. Cancer Res.* **27**, 438–446, doi:10.1158/1078-0432.Ccr-20-2380 (2021).
45. Hatzi, K. *et al.* Histone demethylase LSD1 is required for germinal center formation and BCL6-driven lymphomagenesis. *Nat. Immunol.* **20**, 86–96, doi:10.1038/s41590-018-0273-1 (2019).
46. Heward, J. A. *et al.* KDM5 inhibition offers a novel therapeutic strategy for the treatment of KMT2D mutant lymphomas. *Blood*, doi:10.1182/blood.2020008743 (2021).
47. Wong, K. *et al.* Cross-species genomic landscape comparison of human mucosal melanoma with canine oral and equine melanoma. *Nat Commun* **10**, 353, doi:10.1038/s41467-018-08081-1 (2019).
48. Xu-Monette, Z. Y. *et al.* Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study. *Blood* **120**, 3986–3996, doi:10.1182/blood-2012-05-433334 (2012).
49. Ramsay, A. J. *et al.* POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45**, 526–530, doi:10.1038/ng.2584 (2013).
50. Marconato, L. *et al.* Randomized, placebo-controlled, double-blinded chemoimmunotherapy clinical trial in a pet dog model of diffuse large B-cell lymphoma. *Clin Cancer Res* **20**, 668–677, doi:10.1158/1078-0432.CCR-13-2283 (2014).
51. Vail, D. M. *et al.* Response evaluation criteria for peripheral nodal lymphoma in dogs (v1.0)—a Veterinary Cooperative Oncology Group (VCOG) consensus document. *Vet Comp Oncol* **8**, 28–37, doi:10.1111/j.1476-5829.2009.00200.x (2010).
52. Broeckx, B. J. *et al.* Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Sci Rep* **5**, 12810, doi:10.1038/srep12810 (2015).
53. Aresu, L. *et al.* Prognostic Value of PD-L1, PD-1 and CD8A in Canine Diffuse Large B-Cell Lymphoma Detected by RNAscope. *Vet Sci* **8**, doi:10.3390/vetsci8070120 (2021).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, doi:10.1093/bioinformatics/btp324 (2009).
55. Van der Auwera, G. & O’Connor, B. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra.* 1st edn, (O’Reilly Media, 2020).
56. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303, doi:10.1101/gr.107524.110 (2010).
57. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31, 213–219, doi:10.1038/nbt.2514 (2013).
58. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817, doi:10.1093/bioinformatics/bts271 (2012).
59. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222, doi:10.1093/bioinformatics/btv710 (2016).
60. Hendricks, W. P. D. et al. Somatic inactivating PTPRJ mutations and dysregulated pathways identified in canine malignant melanoma by integrated comparative genomic analysis. PLoS Genet 14, e1007589, doi:10.1371/journal.pgen.1007589 (2018).
61. Amin, S. B. et al. Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. Cancer Cell 37, 243–257.e247, doi:10.1016/j.ccell.2020.01.004 (2020).
62. Bai, B. et al. DoGSD: the dog and wolf genome SNP database. Nucleic Acids Res. 43, D777-783, doi:10.1093/nar/gku1174 (2015).
63. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993, doi:10.1093/bioinformatics/btr509 (2011).
64. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164, doi:10.1093/nar/gkq603 (2010).
65. Rosales, R. A., Drummond, R. D., Valieris, R., Dias-Neto, E. & da Silva, I. T. signeR: an empirical Bayesian approach to mutational signature discovery. Bioinformatics 33, 8–16, doi:10.1093/bioinformatics/btw572 (2017).
66. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421, doi:10.1038/nature12477 (2013).
67. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. Nature 578, 94–101, doi:10.1038/s41586-020-1943-3 (2020).
68. Miller, C. A. et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol 10, e1003665, doi:10.1371/journal.pcbi.1003665 (2014).
69. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol 12, e1004873, doi:10.1371/journal.pcbi.1004873 (2016).
70. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550, doi:10.1073/pnas.0506580102 (2005).

Figures
Figure 1

Kaplan-Meier (KM) curves for 77 dogs with DLBCL. a, b KM curves of time to progression (TTP) and lymphoma specific survival (LSS) according to treatment, showing shorter TTP and LSS for dogs treated with chemotherapy alone compared to those treated with chemo-immunotherapy. c-f KM curves for 32 dogs treated with chemotherapy alone representing TTP and LSS according to TP53 (c, d) and MYC mutational status (e, f): TTP and LSS are shorter in dogs bearing mutations in TP53 and MYC compared
to wild-type. g, h KM curves for 45 dogs treated with chemo-immunotherapy showing LSS according to POT1 (g) and TP53 mutational status (h): LSS is shorter in dogs bearing mutations in POT1 and TP53 compared to wild-type.

**Figure 2**

Mutational landscape of cDLBCL determined by WES. Distribution of protein coding somatic short variants across 77 cDLBCLs. Each bar represents a single dog and variants are sorted in descending order. Different mutation types are indicated by different colors (complete data are supplied in Supplementary Data 3).
Figure 3

Oncoplot of recurrently mutated genes and Copy Number Aberrations (CNAs) in cDLBCL. Recurrently mutated genes (≥5% of patients) harboring somatic short variants (SNVs and indels) and the most frequent and extensive somatic CNAs identified by WES in 77 cDLBCLs are depicted. Genes are represented in descending order according to the frequency of mutation. Different mutation types are identified with different colors. Barplots on the top show the number of protein coding variants (0-446) and the number of clusters for subclonal cell populations (0-7) for each dog. Clinico-pathological data including sex, breed, bone marrow (BM) infiltration, stage and treatment are shown at the bottom.
Figure 4

Somatic CNAs and short mutations in the top-four mutated genes. Somatic CNAs and short variants per individual (upper panel) and localization of SNVs and indels at the protein level (lower panel) for TRAF3 (a), SETD2 (b), POT1 (c) and TP53 (d) genes. The height of each bar represents the frequency at which each mutation occurred. Functional domains for each protein are also depicted.
Figure 5

Somatic interactions in cDLBCL. Mutually exclusive and co-occurring set of genes among the 43 recurrently mutated genes are shown. Trend towards co-occurrence and exclusivity are represented with blue and red, respectively.
Figure 6

Mutational signatures in cDLBCL. Mutation signature analysis showing the relative distribution of six types of substitutions in 96 different trinucleotide contexts revealed that Signature 1A was predominant in all tumors. This signature is the result of spontaneous deamination of 5-methylcytosine at CpG islands due to ageing.
Figure 7

Oncogenic pathways affected by short variants and CNAs. Barchart showing signaling pathways hit by somatic variants (left panel) sorted in descending order by fraction of affected samples. For each pathway, the fraction of affected genes is also indicated (right panel).

Figure 8

CNAs profile of cDLBCL. CNAs obtained by WES data of all dogs with a G-score > 1.0 and a significance of Q < 0.05 according to GISTIC algorithm are plotted together: gains and losses are identified by blue and red bars, respectively. On the x-axis chromosomes 1–38 are reported, while the cumulative frequency (%) is represented on the y-axis. Gain of CFA 13 and 31 and loss of CFA 14 are the most extensive and with the highest frequency CNAs.
Figure 9

Association of TP53 mutation with CNAs. The circos plot shows the statistically significant associations (p<0.05) between TP53 mutation (black square on CFA 5) and CNAs including focal losses in CFA 8 (red square) and gains in CFA 13 and 31 (blue squares).
Figure 10

Illustration of the N-terminal histone tail of H3.3 with emphasis on the K36, K27 and K9 amino acid residues. Writers and erasers of the histone code are displayed, including SETD2, KDM2A, EZH2, KDM6A, SUZ12, KDM3B and TBL1XR1. Mutation frequencies in cDLBCL are reported in brackets.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData1.xlsx
- SupplementaryInformation.docx
- SupplementaryData4.xlsx
- SupplementaryData14.xlsx
- SupplementaryData9.xlsx
- SupplementaryData2.xlsx
- SupplementaryData5.xlsx
- SupplementaryData13.xlsx
- SupplementaryData7.xlsx
- SupplementaryData10.xlsx
• SupplementaryData12.xlsx
• SupplementaryData8.xlsx
• SupplementaryData11.xlsx
• SupplementaryData6.xlsx
• SupplementaryData3.xlsx