LabAnimal

ARTICLES

https://doi.org/10.1038/s41684-022-00998-x

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

Diana Giannuzzi[1], Laura Marconato[2], Antonella Fanelli[3], Luca Licenziato[3], Raffaella De Maria[3], Andrea Rinaldi[4], Luca Rotta[5], Nicole Rouquet[6], Giovanni Birolo[7], Piero Fariselli[7], Afua A. Mensah[4], Francesco Bertoni[4,8] ✉ and Luca Aresu[3] ✉

Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoid neoplasm in dogs and in humans. It is characterized by a remarkable degree of clinical heterogeneity that is not completely elucidated by molecular data. This poses a major barrier to understanding the disease and its response to therapy, or when treating dogs with DLBCL within clinical trials. We performed an integrated analysis of exome ($n = 77$) and RNA sequencing ($n = 43$) data in a cohort of 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 (*TRAF3*, *SETD2*, *POT1*, *TP53*, *MYC*, *FBXW7*, *DDX3X* and *TBL1XR1*) differed. We developed a prognostic model integrating exonic variants and clinical and transcriptomic features to predict the 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.

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 in this model represent powerful elements for translational studies[1]. Among the many lymphoma subtypes, canine diffuse large B-cell lymphoma (cDLBCL) is the most common, accounting for approximately 50–60% of hematological malignancies in this species[2].

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 outcome[3]. Proposed cDLBCL prognostic classification systems are based on bone marrow infiltration, substage, mitotic rate and histologic features (centroblastic and immunoblastic) without consideration of the mechanisms underlying tumorigenesis[4,5]. 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 with normal B cells, cDLBCL present active NF-κB signaling induced by antigen engagement of the B-cell receptor[6]. Additionally, upregulation of several Toll-like receptors suggests a pathogenesis similar to human activated B-cell-like DLBCL (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[7].

Recent preliminary studies have started to describe the mutational spectrum of canine lymphomas, providing a comprehensive catalog of somatic mutations in coding regions[8]. One 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 *TRAF3* and *MAP3K14* were frequently mutated[9]. Notably, *FBXW7* mutations occurring in a specific codon that is recurrently mutated in several human cancers were identified[9]. Despite the large number of cases included in that study, tumors were not appropriately classified according to World Health Organization criteria and survival data were not reported, limiting the clinical relevance. Also, even if a recent canine pan-cancer study revealed that mutations are preferentially cancer-dependent rather than breed-dependent, some genes might still be breed-specific, partially masking the heterogeneous genetic landscape of canine lymphoma[8].

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 multiinstitutional 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, because treatment and outcome are strongly influenced by the owner. Thirdly, even if genetic alterations are defined, their functional impact and clinical validation are often

[1]Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padua, Padua, Italy. [2]Department of Veterinary Medical Science, University of Bologna, Ozzano dell'Emilia, Bologna, Italy. [3]Department of Veterinary Sciences, University of Turin, Grugliasco, Turin, Italy. [4]Institute of Oncology Research, Faculty of Biomedical Sciences, USI, Bellinzona, Switzerland. [5]Department of Experimental Oncology, European Institute of Oncology (IEO), Milan, Italy. [6]Hastim, Toulouse, France. [7]Department of Medical Sciences, University of Turin, Turin, Italy. [8]Oncology Institute of Southern Switzerland, Ente Ospedaliero Cantonale, Bellinzona, Switzerland. ✉e-mail: francesco.bertoni@ior.usi.ch; luca.aresu@unito.it

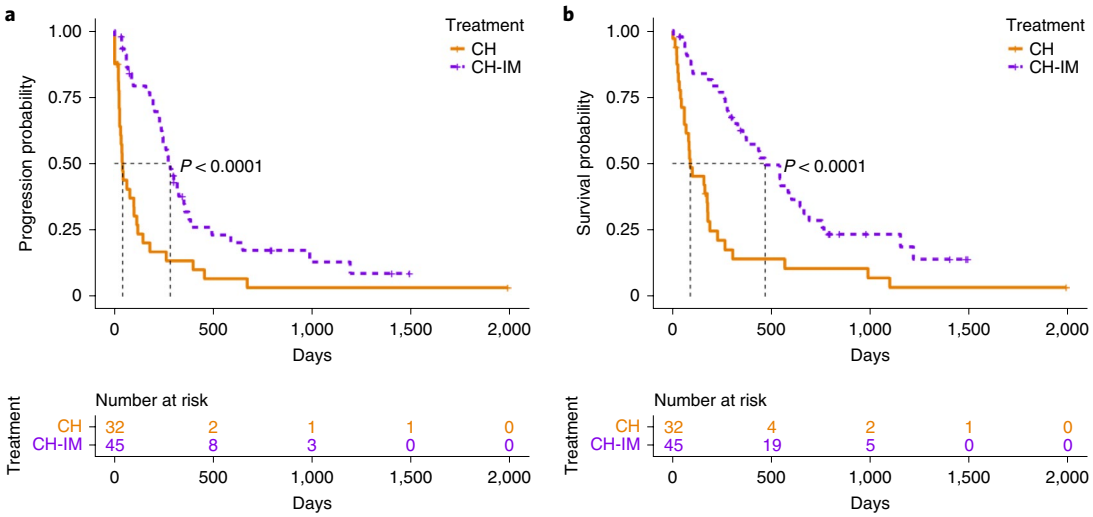

Fig. 1 | KM curves for 77 dogs with DLBCL according to treatment. a,**b**, KM curves of TTP and LSS according to treatment, showing shorter TTP ($P < 0.0001$) and LSS ($P < 0.0001$) for dogs treated with chemotherapy alone (CH) as compared with those treated with chemo-immunotherapy (CH-IM). KM curves were compared by means of log-rank test, and significance was set at $P < 0.05$.

unknown, thus preventing the identification of new therapeutic targets and prognostic markers.

To address these issues and to prospectively inform clinical trials of cDLBCL, we performed a comprehensive multiomics profiling of de novo diagnosed cDLBCL with the goal of clarifying the genetic changes within this tumor. Genetic data obtained from WES were correlated with clinicopathological features. Finally, an integrated model comprising mutations, copy number aberrations (CNAs) and transcriptome was designed to predict overall survival and tumor-free interval. Further, *TP53* mutations were validated in an independent cohort of cDLBCL.

Results

Study population and clinical characteristics. We enrolled 77 de novo cDLBCL cases with matched normal tissues for WES. A subset of the 77 dogs ($n = 43$) were also analyzed for RNA sequencing (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 1 (1.3%) dog had stage III disease. Forty-nine (63.6%) dogs were asymptomatic at presentation (substage a), while 28 (36.4%) showed clinical signs (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 lactate dehydrogenase (LDH) and 23 (29.9%) had received steroids before being diagnosed with lymphoma. Treatment significantly affected both time to progression (TTP) ($P < 0.0001$) and lymphoma-specific survival (LSS) ($P < 0.0001$) (Fig. 1a,b). Indeed, dogs treated with chemo-immunotherapy ($n = 45$; 58.4%) showed a better outcome compared with dogs receiving chemotherapy only ($n = 32$; 41.6%). In dogs receiving chemo-immunotherapy, shorter LSS was significantly associated with substage ($P = 0.04$) and bone marrow infiltration ($P = 0.015$). In dogs receiving chemotherapy only, peripheral blood infiltration correlated with both TTP ($P = 0.021$) and LSS ($P = 0.021$), whereas bone marrow infiltration correlated with LSS only ($P = 0.028$) (Supplementary Data 2).

Landscape of somatic mutations in cDLBCL. On the basis of 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, 10.3–28.7% 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 sorting intolerant from tolerant (SIFT), 1,866 missense variants 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 in only 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* (refs. [9,10]), *SETD2* (refs. [11,12]), *POT1* (refs. [9,13]) and *TP53* (ref. [8])), but others have never been reported in dogs before (*H3C8*, *DIAPH2* and *EHD3*).

In concordance with previous studies[10], *TRAF3* was the most frequently mutated gene in our cohort (53% of the dogs). 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 variants (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% each) (Fig. 4b and Supplementary Table 1). *POT1* mutations were identified in 28.6% of cases. Missense mutations represented

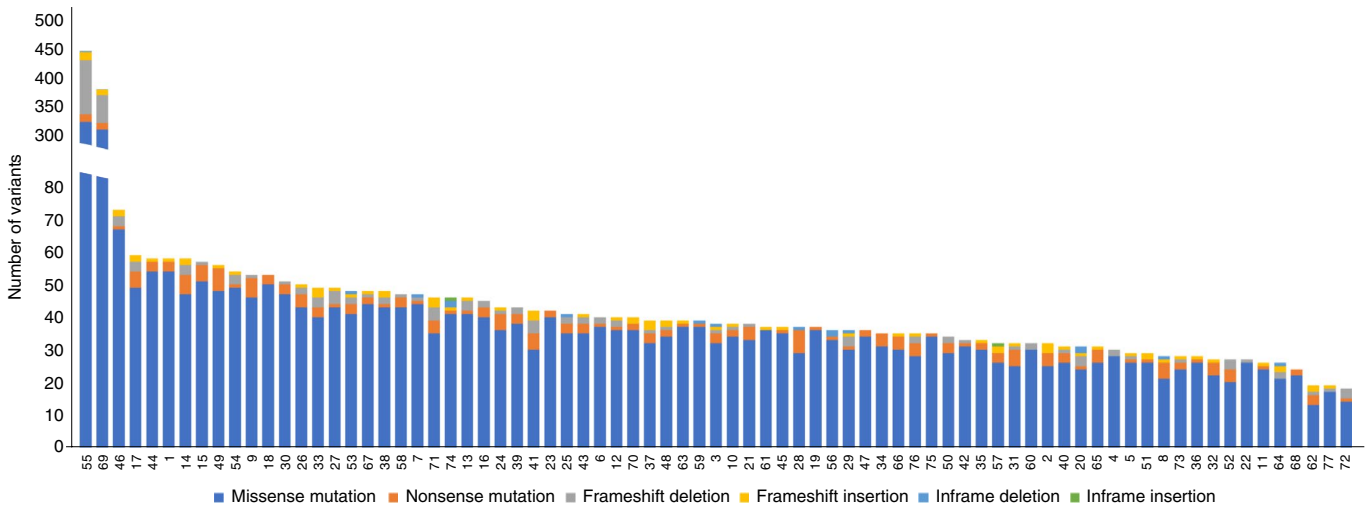

Fig. 2 | Mutational landscape of cDLBCL determined by WES. Distribution of protein-coding somatic short variants across 77 cDLBCLs. Each bar represents a single dog (sample ID number is also indicated below), and variants are sorted in descending order. Different mutation types are indicated by different colors (complete data are supplied in Supplementary Data 3).

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 *POT1* 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 in one case a frameshift deletion and a nonsense variant were identified on the same allele.

We filtered the protein-coding somatic variants for known cancer genes using the Catalogue of Somatic Mutations in Cancer (COSMIC)[14] 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 human DLBCL as retrieved from COSMIC, we found that 1,902 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[15] among the top cancer genes. Significant pairwise interactions (false discovery rate (FDR) 0.1) are depicted in Fig. 5. Notably, *PLEC* mutations exclusively co-occurred with *SETD2*, while *MAP3K14* mutations were mutually exclusive with *TRAF3*, apart from one dog.

The tumor mutational burden (TMB) ranged from 0.12 to 2.94 (mean 0.32; Supplementary Data 1) and was significantly higher in tumors presenting *SETD2* and *TP53* mutations compared with wild-type (WT) tumors ($P < 0.05$), in agreement with what has been reported by others in both canine and human tumors bearing these mutations[8,16,17]. To identify the potential contributions of the mutational processes within tumor exomes, we applied a Bayesian treatment of NMF 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 CNAs occurring in >5% of cases[18]. In addition to activation of the NF-κB 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 Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm was used. A total of 78 significant regions were identified, including 24 gains mainly involving CFA 5, 6, 9 13, 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% of the dogs) and CFA 31 (30% of the dogs), followed by focal gains in CFA 5 and 27. Losses of CFA 14 were identified in 15 dogs (19%). Significant associations with TTP and LSS were detected for 17 and 20 CNAs, respectively (FDR <0.05) (Supplementary Data 6). In particular, the frequently observed gain of the whole CFA 13 was significantly associated with improved TTP in dogs treated with chemo-immunotherapy ($P = 0.027$) (Supplementary Fig. 2 and Supplementary Data 7).

The potential associations between CNAs and somatic mutations were investigated. We found that focal losses in CFA 8 and large broad gains in CFA 13 were strongly associated with *TRAF3*, *SETD2* and *TP53* alterations (Supplementary Figs. 3, 4 and 9). Gains in CFA 31 were associated with *TP53* mutations (Fig. 9) and losses in CFA 14 with *POT1* mutations (Supplementary Fig. 5).

Recurrently mutated genes associate with clinicopathological features of cDLBCL. We assessed whether there were associations between the most recurrent mutations and clinical features (Supplementary Data 8). Within the cohort, dogs with *MYC* and *DDX3X* mutations had a higher percentage of peripheral blood infiltration ($P = 0.03$ and $P = 0.021$, respectively). Also, *DDX3X* mutations were significantly associated with bone marrow infiltration ($P = 0.028$) and a higher LDH level ($P = 0.036$). In line with these results, all dogs with *DDX3X* mutations had stage V disease ($P = 0.002$). When all recurrent mutations were considered collectively, no significant associations were observed by multivariate analysis.

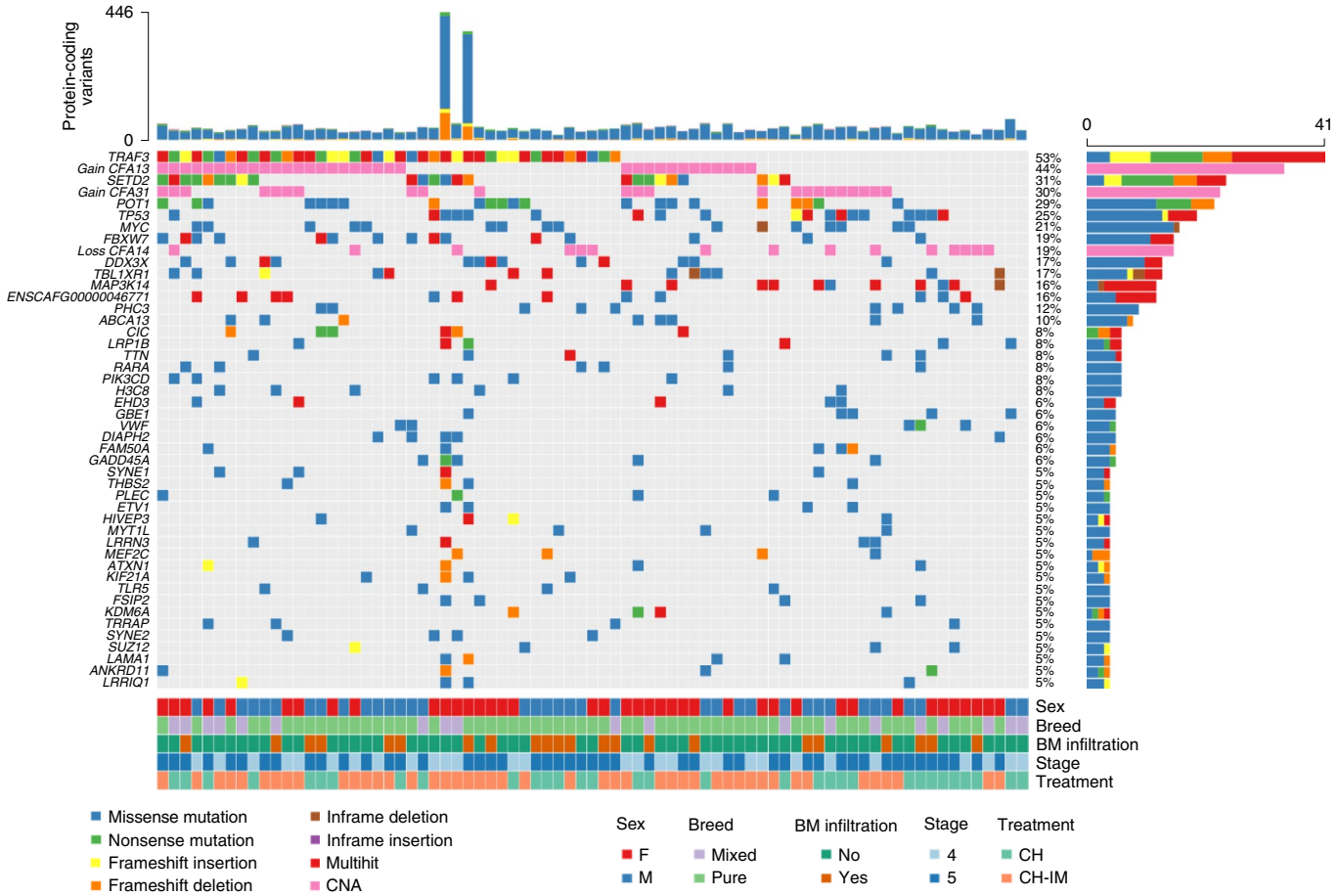

Fig. 3 | Oncoplot of recurrently mutated genes and CNAs in cDLBCL. Recurrently mutated genes (≥5% of patients) harboring somatic short variants (SNVs and indels) and the most frequent and extensive somatic CNAs identified by WES in 77 cDLBCLs are depicted. Genes are represented in descending order according to the frequency of mutation. Different mutation types are identified with different colors. Bar plots on the top show the number of protein-coding variants (0–446) for each dog. Clinicopathological data including sex, breed, bone marrow (BM) infiltration, stage and treatment are shown at the bottom.

In dogs treated with chemotherapy only, *TP53* mutations were associated with a shorter TTP (mut, 32 days; WT, 98 days; $P=0.03$) and LSS (mut, 60 days; WT, 176 days; $P=0.008$) (Fig. 10a,b and Supplementary Data 2). Similarly, dogs carrying *MYC* mutation and treated with chemotherapy only were characterized by shorter TTP and LSS ($P=0.01$ and $P=0.04$, respectively) (Fig. 10c,d and Supplementary Data 2). *TP53* mutations were also associated with a shorter median LSS in dogs treated with chemo-immunotherapy ($P=0.03$) (Fig. 10f and Supplementary Data 2). In animals treated with chemo-immunotherapy, another gene, *POT1*, showed prognostic relevance, with mutations associated with a shorter LSS (mut, 330 days; wt, 547 days; $P=0.02$) (Fig. 10e and Supplementary Data 2). The co-occurrence of *TP53* and *FBXW7* mutations resulted in shorter TTP compared with dogs with only *TP53* mutation ($P=0.04$) (Supplementary Table 2).

Finally, TMB was significantly associated with inferior TTP ($P=0.03$) and LSS ($P=0.03$) in dogs treated with chemo-immunotherapy (Supplementary Fig. 6 and Supplementary Data 2).

Somatic mutations are associated with specific transcriptional signatures in cDLBCL. Using data from dogs with both WES and RNA-seq, we evaluated whether the most frequent mutations (*TRAF3, SETD2, POT1, TP53, MYC, FBXW7* and *DDX3X*) and TMB were associated with specific gene expression signatures (Supplementary Table 3). *DDX3X* mutations were characterized by high expression of transcripts involved in the translation initiation complex and in human Burkitt lymphoma (Supplementary Data 9). *MYC* mutations were characterized by signatures associated with tumor microenvironment and apoptosis (Supplementary Data 9).

We have previously shown that cDLBCL can be divided into two main clusters, of which the cluster characterized by high expression of T-cell and macrophage markers shows an inferior outcome[7]. When we applied the same approach here, the poor outcome cluster was associated with high TMB (Wilcoxon rank-sum test, $P<0.05$) and low frequency of *TRAF3*mut cases (13% versus 75%).

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 exons 4–8 in a second group of 56 dogs affected by DLBCL, whose clinicopathological features are reported in Supplementary Data 10. Fifteen dogs harbored mutations in *TP53*, all of which occurred at different nucleotide positions, except in two dogs, which 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 they were also 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 11).

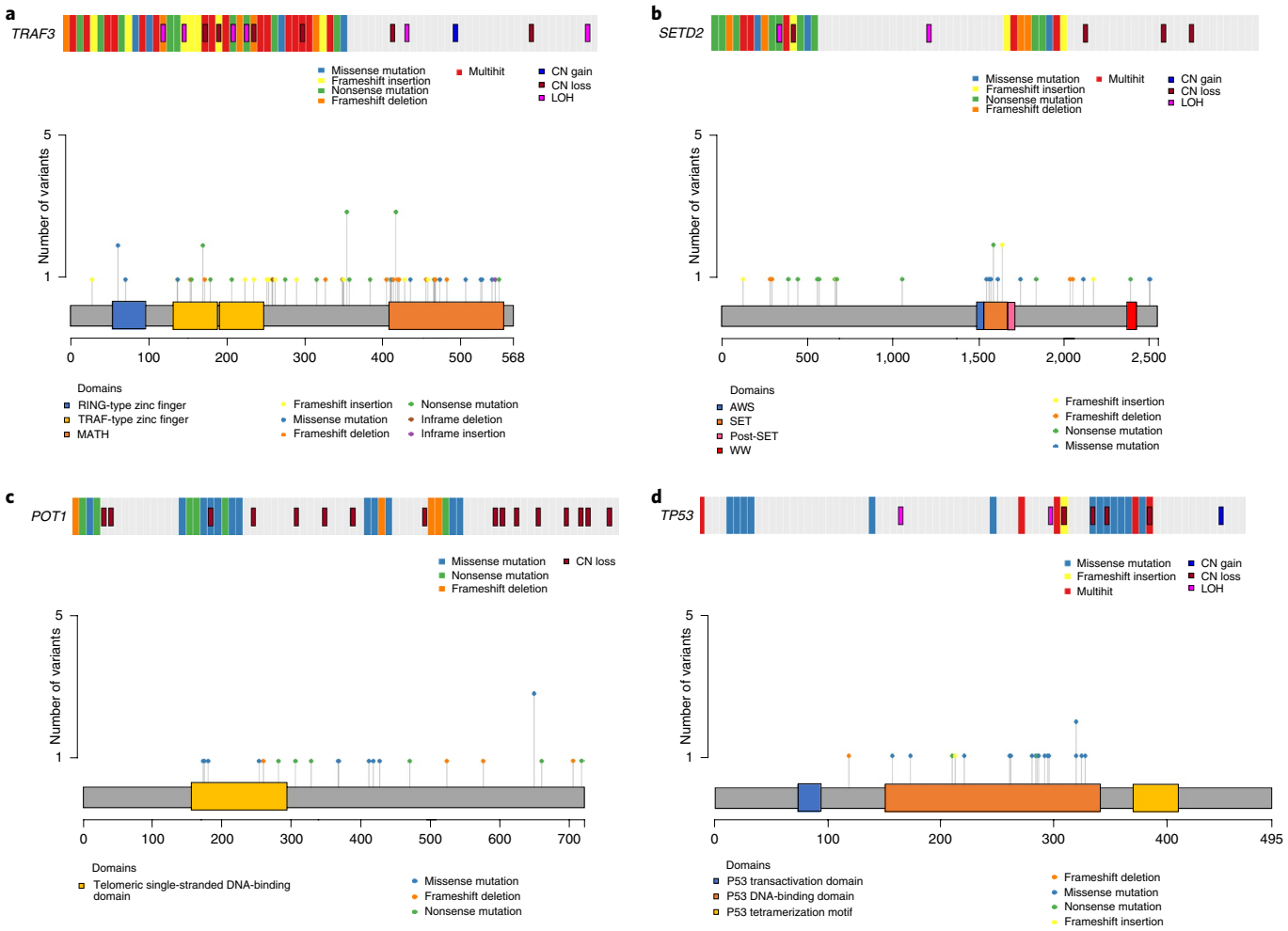

Fig. 4 | Somatic CNAs and short mutations in the top-four mutated genes. a–d, Somatic CNAs and short variants per individual (top) and localization of SNVs and indels at the protein level (bottom) for *TRAF3* (**a**), *SETD2* (**b**), *POT1* (**c**) and *TP53* (**d**) genes. In the lower panels, the height of each bar represents the frequency at which each mutation occurred. Functional domains for each protein are also depicted. AWS, associated with SET domains; CN, copy number; SET, SET domain; WW, WW domain.

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 12). Indeed, *TP53*mut dogs had a significantly shorter TTP ($P<0.0001$) and LSS ($P<0.0001$) compared with *TP53*WT dogs (Supplementary Fig. 7a,b).

Integration of omics and clinicopathological features predicts survival in cDLBCL. We developed a multivariate supervised learning approach for defining the association of survival with clinicopathological variables, genetic features and gene expression data. Among all, the Cox model with elastic net regularization outperformed the random forest models and was considered for further steps (Supplementary Fig. 8a,b). The most predictive features were identified taking advantage of the Least Absolute Shrinkage and Selection Operator (LASSO) shrinkage and combined to generate survival prediction models for LSS and TTP. cDLBCLs from the first cohort of 77 dogs were divided into two subgroups based on their observed risk (long and short survivors) using the median survival time as threshold: 95 and 177 days for TTP and LSS, respectively. Performance was evaluated in cross-validation to avoid overfitting in training and model selection (Methods). The best-performing scores for LSS (area under the receiver operating characteristic (AUROC) 0.95)) and TTP (AUROC 0.87) were obtained with the following variables: age, bone marrow infiltration (%), treatment (chemotherapy versus chemo-immunotherapy), *TP53* genetic status (mut versus WT) and *STAP2* and *G3BP2* gene expression as logCPM (Supplementary Fig. 8c,d). By excluding *STAP2* and *G3BP2* expression data, AUROC for LSS and TTP dropped to 0.90 and 0.79, respectively (Supplementary Fig. 8a,b). Furthermore, when *TP53* genetic data were removed from the analysis, AUROC for LSS and TTP dropped further to 0.83 and 0.74, respectively (Supplementary Fig. 8a,b).

The predictive model was validated using the second cohort of 56 dogs. All the clinicopathological features and the TP53 status were included, and the AUROC resulted in 0.83 and 0.82 for LSS and TTP, respectively. This mild drop was probably due to the single treatment effect, reducing the ability of the model to discriminate. Indeed, chemo-immunotherapy was the most predictive clinical feature. 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 clinicopathological data, transcriptomic and genomic features.

Discussion

The application of genetic and transcriptomic analyses has led to an increased understanding of the biology of human DLBCL, paving

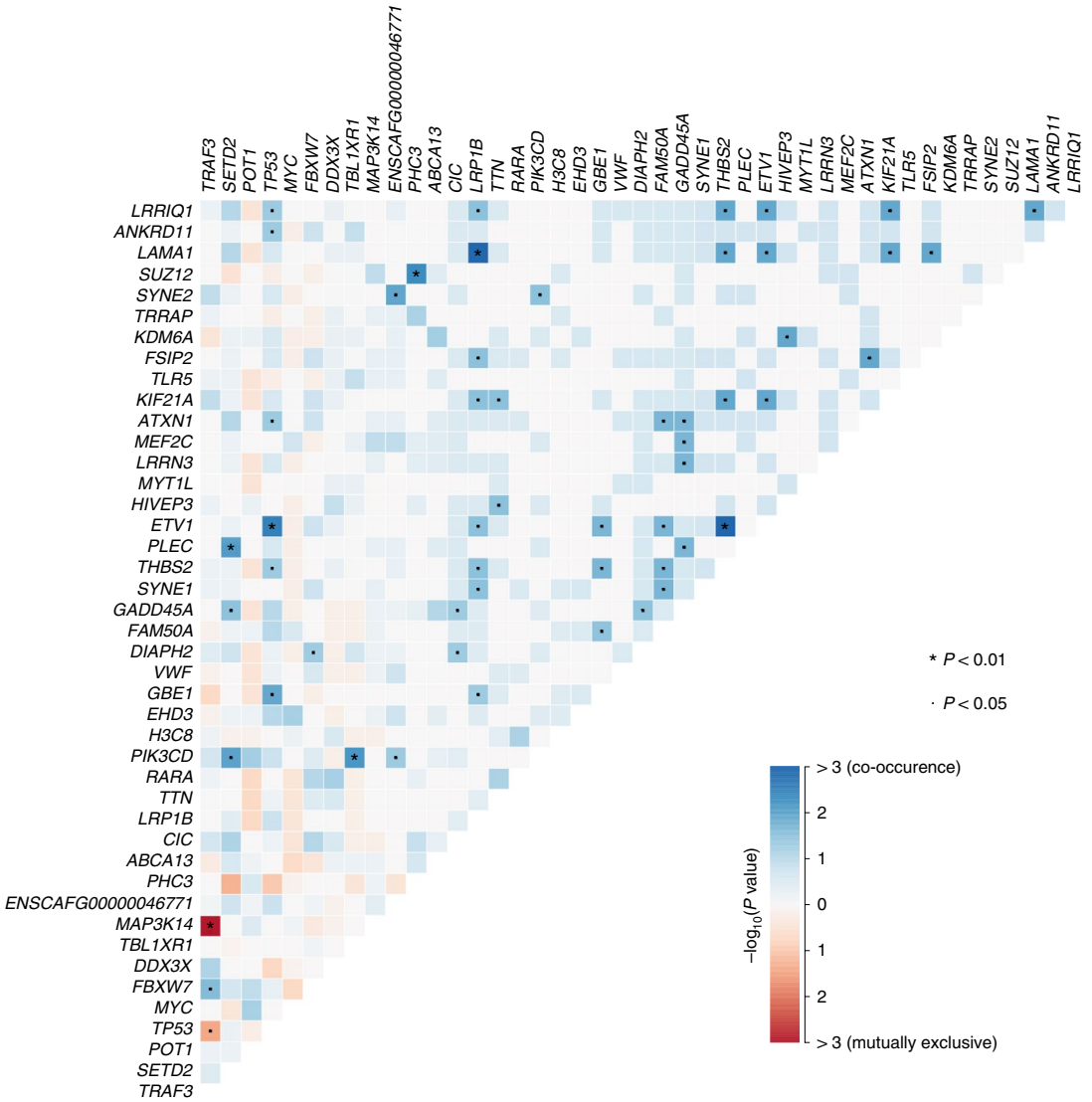

Fig. 5 | Somatic interactions in cDLBCL. Mutually exclusive and co-occurring genes among the 43 recurrently mutated genes are shown. Trend toward co-occurrence and exclusivity are represented with blue and red, respectively (*$P < 0.01$, •$P < 0.05$; Fisher's exact test).

the way to target-specific therapeutic approaches[19]. Here we integrated clinical features, somatic mutations, CNAs and transcriptome in cDLBCL, providing new data on the genetic hallmarks of this tumor type, identifying multiple mutated genes associated with outcome, and 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 part of a complex including TRAF3 itself, TRAF2, BIRC2, BIRC3 and MAP3K14. The disruption of this complex leads to the activation of the noncanonical NF-κB pathway. *TRAF3* inactivation has been described in both canine and human lymphoid neoplasms[9,10,20–24], including in 15% of human DLBCL cases in which it contributes to the activation of the NF-κB signaling cascade[25]. Indeed, in our series, dogs with mutated *TRAF3* presented an active NF-κB transcriptome program, and other genes encoding members of the TRAF3 complex were also 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 that 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[26–28], but also highlight lesions associated with resistance to Bruton's tyrosine kinase inhibitors[24,29], which show antitumor activity in many NF-κB-driven lymphomas, and with implications for the management of dogs with DLBCL and for the use of these animals as models for the human disease.

A total of 31% of dogs harbored mutations in *SETD2*, which was the second most frequently altered gene (Supplementary Fig. 9). SETD2 is a methyltransferase with various substrates, and multiple lines of evidence support its tumor suppressor role[30]. 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[30]. However, SETD2 also has nonhistone substrates, including microtubules and transcription factors (for example, *TP53*), whose aberrant modification can also contribute to cancer development and progression[30]. Mutations in *SETD2* have been reported in 20–40% of canine osteosarcomas[12] as well as in human tumors, including lymphomas[31–35].

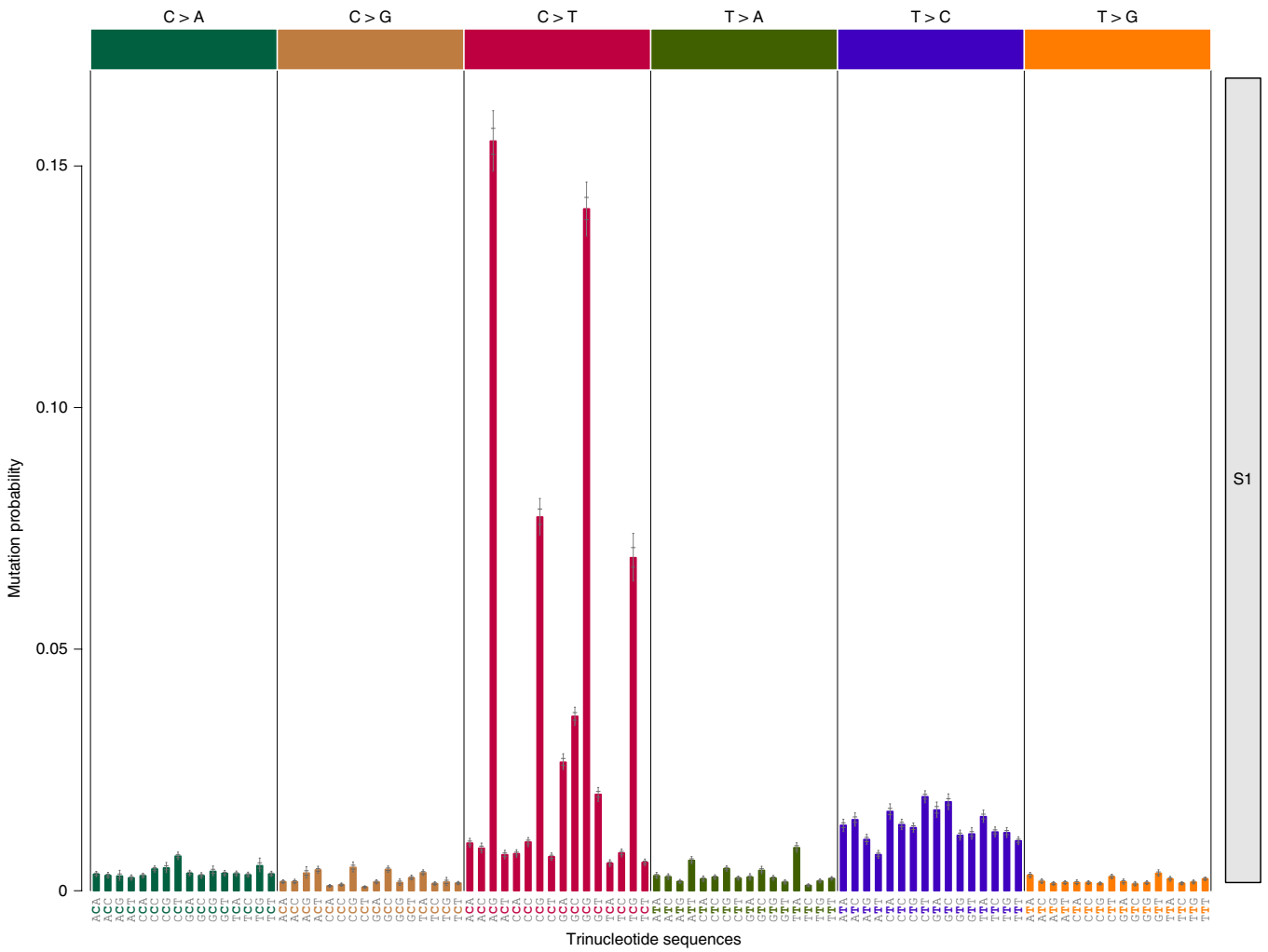

Fig. 6 | Mutational signature analysis in cDLBCL. Mutation signature analysis was performed using nonsynonymous and synonymous substitutions for all 77 tumors. A single mutational signature (S1), corresponding to signature 1A, was identified as predominant. The plot shows the distribution of the six types of substitution (in 96 different trinucleotide contexts) defined by the pyrimidine as inferred from the NMF algorithm. Each subgraph within a signature represents one substitution. The bars within each subgraph include the nucleotides on either side of the mutation location in the reference genome. The error bars represent ±standard error (SE) of the coefficients calculated over the replications of the extraction process. Graphics have been created using the signeR package within R software v3.6.3 (www.r-project.org).

Differently from our observations in cDLBCL, *SETD2* mutations are present in less than 10% of human DLBCL[34,35], while more common in T-cell lymphomas[31–33]. *SETD2* was not the only mutated 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 (*H3C8* or *H3C12*, 10%), *KDM6A*, *SUZ12* (5%), *KDM2A*, *KDM3B* and *EZH2* (3%) (Supplementary Fig. 9).

The mutations in *H3C8* and *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[36]. The mutation inhibits the activity of the polycomb repressive complex 2 (PRC2), composed of the K27 histone methyltransferase EZH2 (enhancer of zeste homolog 2) and the core accessory proteins EED, SUZ12 and RbAp48 (ref. [37]) (Supplementary Fig. 9): we also observed recurrent inactivating heterozygous mutations in two of the PRC2 proteins (*EZH2*, 3%; *SUZ12*, 5%). As methionine cannot be methylated by 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[37]. However, perhaps owing 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[37–39]. The removal of di- and trimethyl groups from H3K27 is done by two histone demethylases, one of which, UTX, is encoded by *KDM6A*. *KDM6A* was mutated in 5% cDLBCL (Supplementary Fig. 9). UTX forms a complex with H3K4 methyltransferases MLL2 (KMT2D)/MLL3 (KMT2C)[37]. The inactivation of the two methyltransferases is observed in 20–30% of human DLBCL[40,41], and it determines a diminished global H3K4 methylation with deregulation of CD40, Toll-like and B-cell receptor signaling pathways[42,43].

In our series, we did not find any mutation in the acetyltransferases CREBBP and EP300, which activate transcription via acetylation of histone H3 lysine 27 (H3K27Ac) and are recurrently inactivated in human lymphomas[40,44]. Besides impairing PRC2 activity, H3K27M might also affect acetyltransferase activity, leading to aberrant gene expression and enhancer dysfunction[45], and this could at least partially mimic the effects of *CREBBP* and *EP300* mutations.

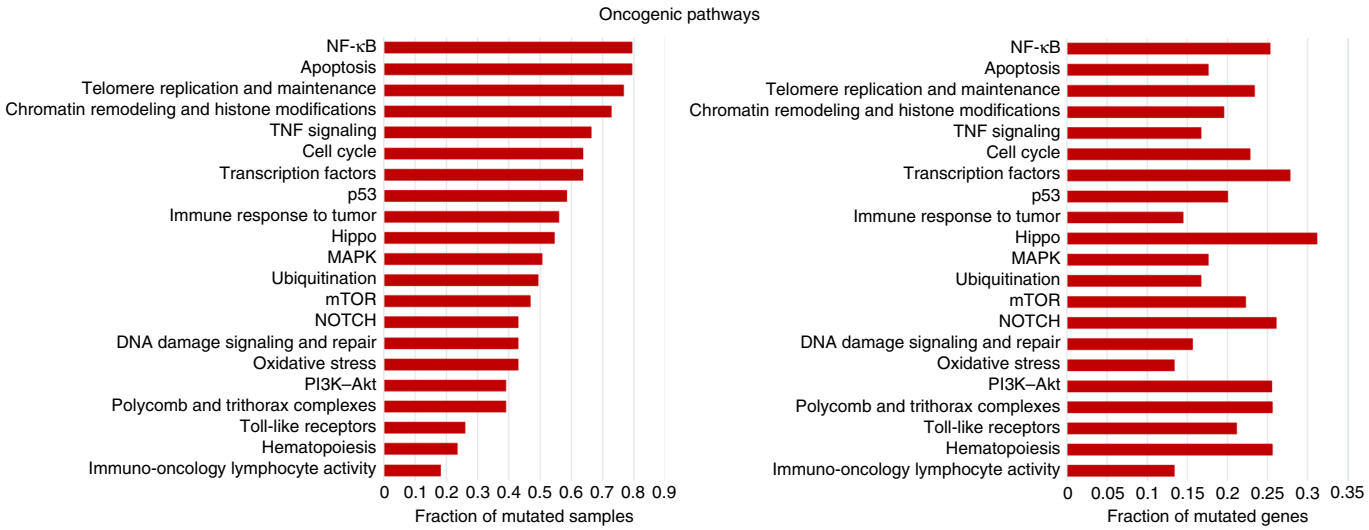

Fig. 7 | Oncogenic pathways affected by SNVs and CNAs in cDLBCL. The percentage of cases with deregulated pathway (left) and the fraction of genes belonging to the pathway that are affected by genetic events across the whole series of cases (right) are represented. The genes belonging to each pathway were defined on the basis of GSEA gene set collection (HALLMARK and C2, C6, C7 and C8). When genes were found in multiple pathways, the most biologically relevant was considered to avoid redundancies.

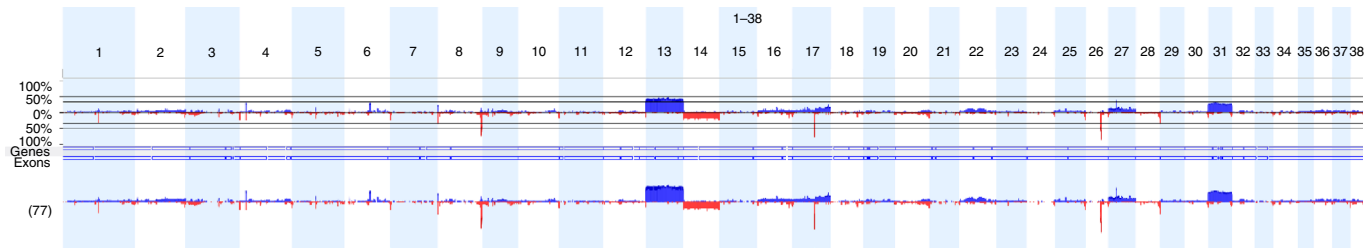

Fig. 8 | CNA profile of cDLBCL. CNAs obtained by WES data of all dogs with a *G*-score >1.0 and a significance of $Q < 0.05$ according to the GISTIC algorithm are plotted together: gains and losses are identified by blue and red bars, respectively. On the *x* axis, chromosomes 1–38 are reported, while 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 highest-frequency CNAs.

KDM2A (JHDM1A/FBXL11) and *KDM3B* (JMJD1B/ JMJD1B) genes, each one mutated in 3% of cDLBCL, encode 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[46] (Supplementary Fig. 9). In cancer, they are thought to function in a tumor type-specific fashion[46] and rare mutations are reported in human B-cell lymphomas[14].

Previously reported in a single case of canine B-cell lymphoma[10], *TBL1XR1* was mutated in 17% of our cDLBCLs. The 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[44,47]. *TBL1XR1* mutations were associated with a trend for a poor prognosis also in our series of cDLBCL, but the low frequency of mutations prevented statistical significance being reached. TBL1XR1 is a core component of the SMRT/NCOR1 transcriptional repressor complexes, thus also contributing to transcription regulation (Supplementary Fig. 9).

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[48] and H3K27M mutants[39], demethylating agents for KDM6A mutants and HDAC3 inhibitors for TBL1XR1 mutants[49]. The latter[50] might also show activity in H3K27M mutants, owing to the possibly reduced acetyltransferase activity of CREBBP and EP300 on H3K27M[45]. LSD1 inhibitors[51,52] and especially KDM5 inhibitors[53] could work in KDM6A mutants, which might have concomitant deregulated MLL2/MLL3 activity[37]. 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[12,54]. Compared with the previous two studies in canine lymphoma[9,10], 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.

The somatic mutations observed in our series of cDLBCL appeared almost exclusively compatible with the spontaneous deamination at CpG sites previously associated with aging, which

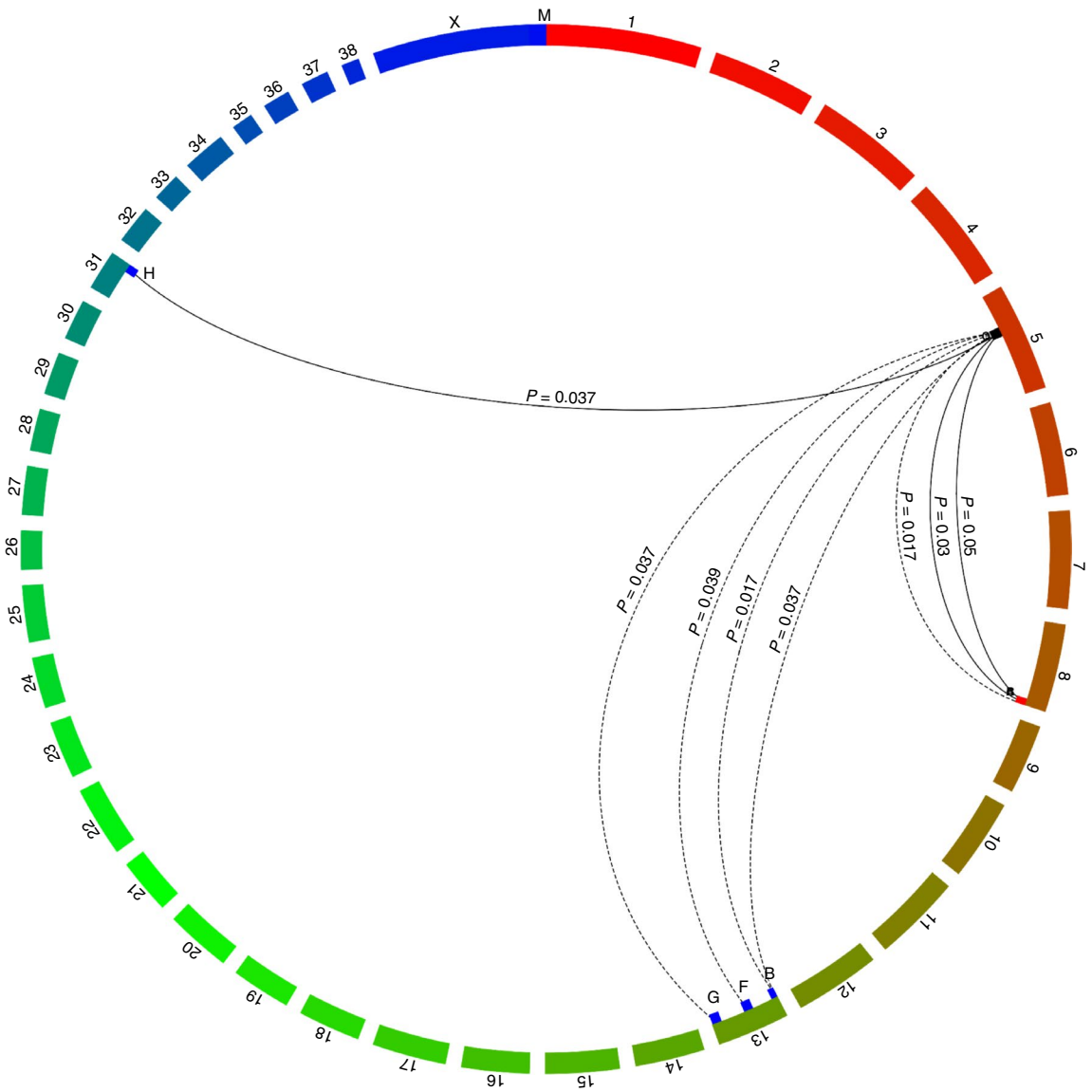

Fig. 9 | Association of *TP53* mutation with CNAs. The circos plot shows the statistically significant associations ($P < 0.05$) between *TP53* mutation (black square on CFA 5) and CNAs, including focal losses in CFA 8 (red square) and gains in CFA 13 and 31 (blue squares).

is also the most frequent mechanism reported in human DLBCL[44]. We did not observe signatures compatible with the activity of activation-induced cytidine deaminase, the second most common mechanism leading to mutations in the human counterpart[44,55]. Considering the so far lack of described chromosomal translocations involving the immunoglobulin genes in cDLBCL, instead frequent in humans, our data suggest potential differences between the two species in the effects caused by the activation-induced cytidine deaminase-mediated somatic hypermutations in the transformation process.

Recent advances in chemo-immunotherapy have offered new options for the treatment of canine cancers[56,57]. The addition of an autologous vaccination (APAVAC) to CHOP-based chemotherapy has dramatically prolonged both remission time and overall survival in dogs with DLBCL[3]. The APAVAC vaccine consists of heat shock proteins purified from the dog's tumor, and their presentation to and recognition by the dogs' immune system provides protection[58]. While chemotherapy still represents the cornerstone of lymphoma treatment, it is usually not curative for dogs with DLBCL. Immunotherapy may circumvent the immune evasion caused by cancer heterogeneity by immunizing the host against a large repertoire of individual tumor-associated antigens[59].

However, not all patients benefit from immunotherapy[3], and predicting a patient's response would reduce financial costs. Here we identified two groups of animals differing in 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[44,47,60], 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

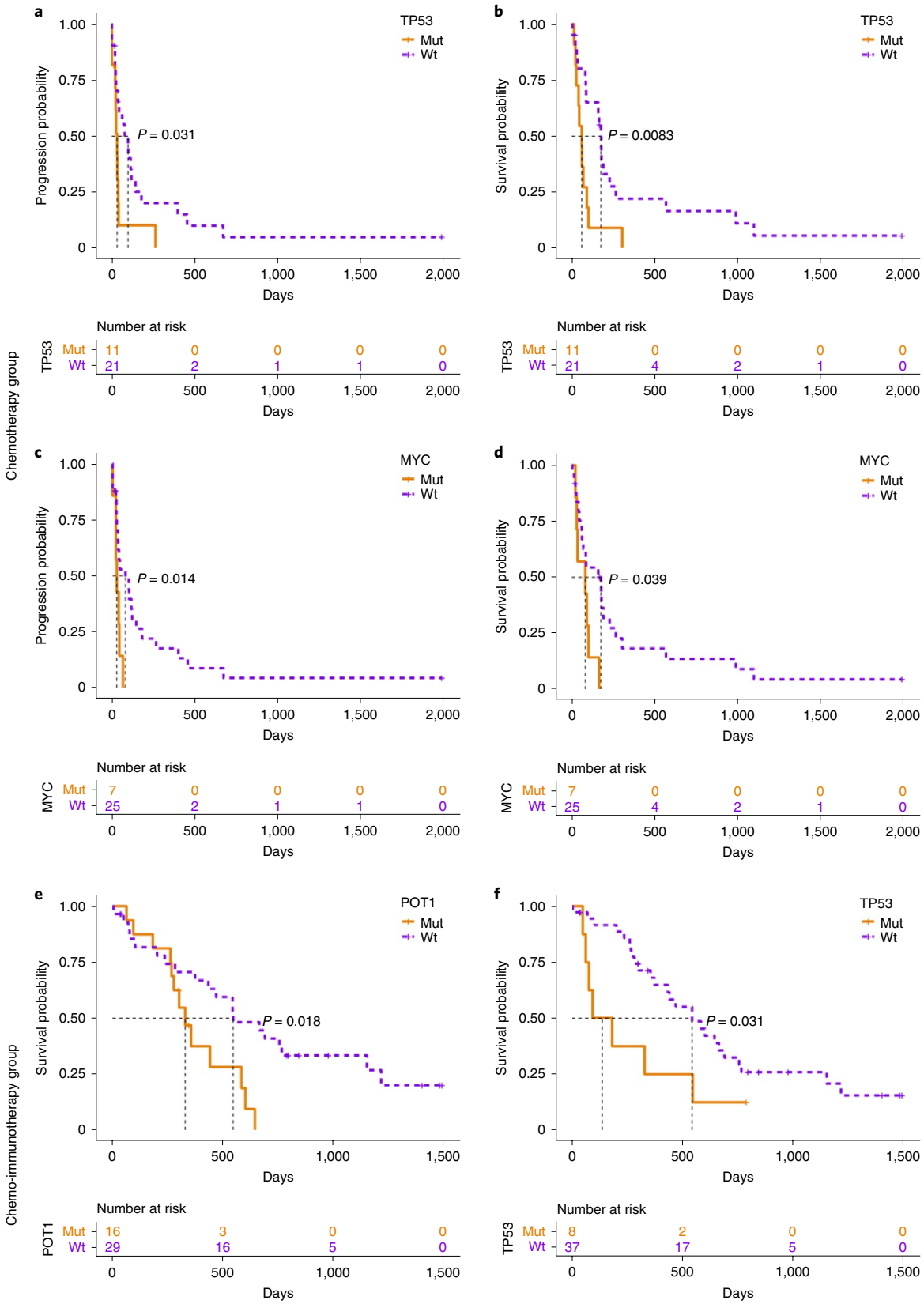

Fig. 10 | KM curves for 77 dogs with DLBCL according to *TP53*, *MYC* and *POT1* mutational status. a–d, KM curves for 32 dogs treated with chemotherapy alone, representing TTP and LSS according to *TP53* (**a** and **b**) and *MYC* mutational status (**c** and **d**); TTP and LSS are shorter in dogs bearing mutations in *TP53* ($P = 0.03$ and $P = 0.008$, respectively) and *MYC* ($P = 0.01$ and $P = 0.04$, respectively) as compared with WT. **e**,**f**, KM curves for 45 dogs treated with chemo-immunotherapy showing LSS according to *POT1* (**e**) and *TP53* mutational status (**f**); LSS is shorter in dogs bearing mutations in *POT1* ($P = 0.02$) and *TP53* ($P = 0.03$) as compared with WT. KM curves were compared by means of log-rank test, and significance was set at $P < 0.05$.

experimental validation. In addition, the co-occurrence of *TP53* and *FBXW7* mutations resulted in a worse outcome than *TP53* mutation alone.

Other than *TP53*, *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, with telomerase activity and genomic instability[61] 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*, noncanonical NF-κB pathway and chromatin remodeling.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41684-022-00998-x.

Received: 9 October 2021; Accepted: 13 May 2022;
Published online: 20 June 2022

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 (2017).
2. Avery, A. C. The genetic and molecular basis for canine models of human leukemia and lymphoma. *Front. Oncol.* **10**, 23 (2020).
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 (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 (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 (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 (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 (2019).
8. Alsaihati, B. A. et al. Canine tumor mutational burden is correlated with TP53 mutation across tumor types and breeds. *Nat. Commun.* **12**, 4670 (2021).
9. Elvers, I. et al. Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* **25**, 1634–1645 (2015).
10. Bushell, K. R. et al. Genetic inactivation of TRAF3 in canine and human B-cell lymphoma. *Blood* **125**, 999–1005 (2015).
11. Gardner, H. L. et al. Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2. *Commun. Biol.* **2**, 266 (2019).
12. Sakthikumar, S. et al. SETD2 is recurrently mutated in whole-exome sequenced canine osteosarcoma. *Cancer Res.* **78**, 3421–3431 (2018).
13. Smith, P. A. D., Waugh, E. M., Crichton, C., Jarrett, R. F. & Morris, J. S. The prevalence and characterisation of TRAF3 and POT1 mutations in canine B-cell lymphoma. *Vet. J.* **266**, 105575 (2020).
14. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
15. Leiserson, M. D., Reyna, M. A. & Raphael, B. J. A weighted exact test for mutually exclusive mutations in cancer. *Bioinformatics* **32**, i736–i745 (2016).
16. Lu, M. et al. Pan-cancer analysis of SETD2 mutation and its association with the efficacy of immunotherapy. *NPJ Precis. Oncol.* **5**, 51 (2021).
17. Li, L., Li, M. & Wang, X. Cancer type-dependent correlations between TP53 mutations and antitumor immunity. *DNA Repair* **88**, 102785 (2020).
18. Sanchez-Vega, F. et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell* **173**, 321–337.e10 (2018).
19. Miao, Y., Medeiros, L. J., Li, Y., Li, J. & Young, K. H. Genetic alterations and their clinical implications in DLBCL. *Nat. Rev. Clin. Oncol.* **16**, 634–652 (2019).
20. Keats, J. J. et al. Promiscuous mutations activate the noncanonical NF-κB pathway in multiple myeloma. *Cancer Cell* **12**, 131–144 (2007).
21. Demchenko, Y. N. et al. Classical and/or alternative NF-κB pathway activation in multiple myeloma. *Blood* **115**, 3541–3552 (2010).
22. Otto, C. et al. Genetic lesions of the TRAF3 and MAP3K14 genes in classical Hodgkin lymphoma. *Br. J. Haematol.* **157**, 702–708 (2012).
23. Rossi, D. et al. Alteration of BIRC3 and multiple other NF-κB pathway genes in splenic marginal zone lymphoma. *Blood* **118**, 4930–4934 (2011).
24. Rahal, R. et al. Pharmacological and genomic profiling identifies NF-κB-targeted treatment strategies for mantle cell lymphoma. *Nat. Med.* **20**, 87–92 (2014).
25. Zhang, B. et al. An oncogenic role for alternative NF-κB signaling in DLBCL revealed upon deregulated BCL6 expression. *Cell Rep.* **11**, 715–726 (2015).
26. Moore, C. R., Edwards, S. K. & Xie, P. Targeting TRAF3 downstream signaling pathways in B cell neoplasms. *J. Cancer Sci. Ther.* **7**, 67–74 (2015).
27. 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 (2019).
28. 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 (2016).
29. Diop, F. et al. Biological and clinical implications of BIRC3 mutations in chronic lymphocytic leukemia. *Haematologica* **105**, 448–456 (2020).
30. Licht, J. D. SETD2: a complex role in blood malignancy. *Blood* **130**, 2576–2578 (2017).
31. Ji, M. M. et al. Histone modifier gene mutations in peripheral T-cell lymphoma not otherwise specified. *Haematologica* **103**, 679–687 (2018).
32. McKinney, M. et al. The genetic basis of hepatosplenic T-cell lymphoma. *Cancer Discov.* **7**, 369–379 (2017).
33. 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 (2020).
34. Zhang, J. et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl Acad. Sci. USA* **110**, 1398–1403 (2013).
35. Li, J. et al. SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget* **7**, 50719–50734 (2016).
36. Fuller, C. E., Jones, D. T. W. & Kieran, M. W. New classification for central nervous system tumors: implications for diagnosis and therapy. *Am. Soc. Clin. Oncol. Educ. Book* **37**, 753–763 (2017).
37. 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 (2014).
38. Sturm, D. et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
39. Mohammad, F. et al. EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* **23**, 483–492 (2017).
40. Pasqualucci, L. et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* **43**, 830–837 (2011).
41. Morin, R. D. et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
42. Zhang, J. et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat. Med.* **21**, 1190–1198 (2015).
43. 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 (2015).
44. 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 (2018).
45. Trovato, M., Patil, V., Gehre, M. & Noh, K. M. Histone variant H3.3 mutations in defining the chromatin function in mammals. *Cells* https://doi.org/10.3390/cells9122716 (2020).
46. 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 (2019).
47. 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.e14 (2020).

48. 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 (2019).
49. Venturutti, L. et al. TBL1XR1 mutations drive extranodal lymphoma by inducing a pro-tumorigenic memory fate. *Cell* **182**, 297–316.e27 (2020).
50. Mondello, P. et al. Selective Inhibition of HDAC3 targets synthetic vulnerabilities and activates immune surveillance in lymphoma. *Cancer Discov.* **10**, 440–459 (2020).
51. 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 (2021).
52. Hatzi, K. et al. Histone demethylase LSD1 is required for germinal center formation and BCL6-driven lymphomagenesis. *Nat. Immunol.* **20**, 86–96 (2019).
53. Heward, J. et al. KDM5 inhibition offers a novel therapeutic strategy for the treatment of KMT2D mutant lymphomas. *Blood* **138**, 370–381 (2021).
54. Wong, K. et al. Cross-species genomic landscape comparison of human mucosal melanoma with canine oral and equine melanoma. *Nat. Commun.* **10**, 353 (2019).
55. Ye, X. et al. Genome-wide mutational signatures revealed distinct developmental paths for human B cell lymphomas. *J. Exp. Med.* https://doi.org/10.1084/jem.20200573 (2021).
56. Marconato, L., Gelain, M. E. & Comazzi, S. The dog as a possible animal model for human non-Hodgkin lymphoma: a review. *Hematol. Oncol.* **31**, 1–9 (2013).
57. Dias, J. N. R. et al. Immunotherapeutic strategies for canine lymphoma: changing the odds against non-Hodgkin lymphoma. *Front. Vet. Sci.* **8**, 621758 (2021).
58. 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 (2014).
59. Tamura, Y., Peng, P., Liu, K., Daou, M. & Srivastava, P. K. Immunotherapy of tumors with autologous tumor-derived heat shock protein preparations. *Science* **278**, 117–120 (1997).
60. 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 (2012).
61. Ramsay, A. J. et al. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45**, 526–530 (2013).

Methods

Animal recruitment and sample acquisition. Dogs with newly diagnosed, previously untreated, multicentric cDLBCL of any World Health Organization clinical stage were included in the study. To be eligible for enrollment, dogs had to undergo a complete staging workup, 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 on a lymph node 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[58]. 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[58]. Briefly, dogs received L-asparaginase, vincristine, cyclophosphamide, doxorubicin, lomustine, prednisone and a total of eight 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. At the end of treatment, all dogs underwent end-staging including flow cytometry on a peripheral lymph node, peripheral blood and bone marrow, and imaging[5,62]. Follow-up evaluation consisted of monthly physical examination, peripheral lymph node 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 flow cytometry confirmation in any anatomical site in dogs having experienced CR. Once relapse was confirmed, a complete restaging workup 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 from 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 owing to causes other than lymphoma were censored for LSS analysis[63].

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) and agarose gel electrophoresis. RNA concentration and integrity were measured respectively, in a NanoDrop ND-1000 spectrophotometer and a Bioanalyzer 2100 instrument (Agilent Technologies). A total of 43 non-normalized libraries for RNA-seq 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). Raw Illumina sequencing data are deposited in the SRA database (GenBank) under accession number SRP137798. All Illumina reads were analyzed with FastaQC software to assess sequence quality. A total of 154 high-quality WES libraries were prepared using the Illumina-compatible KAPA HyperPlus Library Preparation Kit (Roche Sequencing and Life Science). Exome capture was performed using Roche's SeqCap EZ Share Prime Developer Kit (Roche Sequencing and Life Science) for nonhuman 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) in which probes were designed on the target enrichment design of 150 megabases developed by Broeckx et al.[64]. Libraries were quantified using the Qubit DNA Assay Kit in a Qubit 2.0 Fluorometer (Thermo Fisher), and quality was assessed using the Bioanalyzer 2100 instrument.

RNA-seq and WES reads quality control, alignment and preprocessing. Data quality control was first performed on raw Illumina reads for RNA-Seq and WES using FastQC v0.11.9 (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]. For WES, Trimmomatic v0.36 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 v0.7.15 with -M and -t 15 options[65]. Preprocessing 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) v4.1.0.0 best practices[66], 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'.

Somatic variant calling. Somatic SNVs and indels of WES samples were identified using two callers: GATK4 (v4.1.0.0) (ref. [67]), Mutect2 (ref. [68]) and Strelka v1.0.13/Manta v1.6.0 (refs. [69,70]) using default arguments according to previous literature[71,72]. Mutect2 was run in tumor-matched normal mode. To reduce germline artifacts, a panel of normals 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 (http://dogsd.big.ac.cn/)[73]. 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 panel of normals 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 v4.1/4.2 format) by overlapping the remaining variants using the bcftools software package's -isec option[74]. Finally, resulting consensus VCF files were annotated with ANNOVAR[75]. 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.

TMB and mutational signature analysis. Tumor mutation burden (TMB) was calculated for each sample by dividing the number of protein-coding variants (that is, the number of exonic variants excluding the synonymous ones) 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: 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[76] 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[77,78].

Copy number analysis. CNA analysis was performed on WES data using the NEXUS Copy Number v10.0 software (Biodiscovery). CNAs and loss of heterozygosity (LOH) were identified starting from preprocessed 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 1.0×10^{-12} was applied. The dbSNP was also given to the software to assess germline events. Aberrations were defined as a minimum of three consecutive segments and a maximum probe spacing of 1,000 kbp between adjacent probes before breaking a segment. Segments were classified with $\log_2$ tumor/reference ratio value of >0.5 as high gains, 0.5–0.2 as gains, −0.25 to −1.5 as losses and <−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). Recurrent CNAs were determined within NEXUS using the GISTIC algorithm with a *G*-score cutoff of $G>1.0$ and a significance of $Q<0.05$. CNA frequency comparisons among groups (that is, clinical features and presence/absence of somatic mutations) were performed with a two-tailed Fisher's exact test ($P<0.05$).

Functional annotation analysis. The analysis comparing mutational processes and transcriptome to evaluate enriched gene sets in WT/mut samples was performed using gene set enrichment analysis (GSEA) in a preranked mode, with a threshold of significance of FDR <5% (ref. [79]). 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 from Molecular Signatures Database (MsigDB) were used[7].

Prediction models. Two predictive models were tested and implemented in the scikit-survival module (v0.15.0) in Python (v3.8.9): 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 eightfold cross-validation with two holdout sets, repeated 20 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 AUROC for classifying dogs with survival time lower than the median survival in the dataset. Given the limited number of samples, 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 2,832 genetic mutations (binary) and clinical features and a dataset of 43 samples with clinical features and the top 100 most significantly differentially expressed genes between tumors and controls. Features whose regression coefficient was consistently different from zero among the cross-validation fittings of the best performing models were selected and combined 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 4). PCR was performed in a final volume of 20 µl using HotStarTaq DNA Polymerase kit (QIAGEN, Hilden, Germany) and 50 ng of genomic DNA (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). 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 clinicopathological variables and recurrently mutated genes (at least 5% of cases) and identify different classes of clinicopathological variables in mutated or WT individuals, we built classification trees using the recursive partitioning algorithm implemented in the 'party' R package. The correlation analysis between TMB and genes mutated in at least 15% of cases was performed by means of Student's *t*-test. Bonferroni correction was applied for multiple comparison analyses.

Survival analysis was conducted using 'survival' and 'survminer' R packages. The following clinicopathological variables were tested for their influence on TTP and LSS by means of univariate and multivariate Cox proportional hazard model: treatment (chemotherapy versus chemo-immunotherapy), breed (pure versus mixed), sex (female versus male), age (<10 years versus ≥10 years), weight (<10 kg versus ≥10 kg), stage (IV and V), substage (a and b), peripheral blood infiltration (%), bone marrow infiltration (%), presence of bone marrow infiltration (yes versus no), LDH activity (normal versus increased), pretreatment with steroids (yes versus no) and TMB (<0.21 versus >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, with the exception of TMB, were included in the multivariate analysis. For categorical variables, Kaplan–Meier (KM) curves were drawn and compared by means of log-rank test.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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

References

62. Chalfon, C. et al. Minimal residual disease in lymph nodes after achievement of complete remission predicts time to relapse in dogs with large B-cell lymphoma. *Vet. Comp. Oncol.* **17**, 139–146 (2019).
63. 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 (2010).
64. Broeckx, B. J. et al. Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Sci. Rep.* **5**, 12810 (2015).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Van der Auwera, G. & O'Connor, B. in *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st edn (O'Reilly Media, 2020).
67. McKenna, A. et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
68. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
69. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
70. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
71. 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 (2018).
72. Amin, S. B. et al. Comparative molecular life history of spontaneous canine and human gliomas. *Cancer Cell* **37**, 243–257.e7 (2020).
73. Bai, B. et al. DoGSD: the dog and wolf genome SNP database. *Nucleic Acids Res.* **43**, D777–D783 (2015).
74. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
75. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
76. 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 (2017).
77. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
78. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
79. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Acknowledgements

This study was financially supported by the 'Scientific Independence of young Researchers' (SIR) grant received from the Ministero dell'Università (K9LYMPHOMET, protocol RBSI14EDX9).

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. and L.R. carried out experiments; D.G. conducted bioinformatic analysis and contributed to data interpretation; D.G., A.F. and L.L. performed formal analysis and data visualization and conducted statistical analysis; G.B. and P.F. developed the in silico prediction models; A.A.M. interpreted data and revised the manuscript; all authors contributed to manuscript revision and approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41684-022-00998-x.

Correspondence and requests for materials should be addressed to Francesco Bertoni or Luca Aresu.

Peer review information *Lab Animal* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature research

Corresponding author(s): Luca Aresu

Last updated by author(s): Oct 21, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	*Provide a description of all commercial, open source and custom code used to analyse the data in this study, specifying the version used OR state that no software was used.*

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In this study sample size was not predetermined using a statistical method. The cases included in the study are those for which sufficient material was available for genomic analyses and that met the inclusion criteria indicated in Materials and Methods section.
Data exclusions	All the generated data were used for downstream analyses.
Replication	This kind of experiments do not require technical replication.
Randomization	The treatment groups were not determined following pre-established criteria but exclusively from the owner's preference.
Blinding	Blinding is not applicable to our study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	This study did not involve laboratory animals.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	Approval for this study was granted by the Ministero dell'Istruzione, dell' Università e della Ricerca Ethical Board (Number RBSI14EDX9).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com