


# SCIENTIFIC REPORTS



OPEN

## DNA methylation profiling reveals common signatures of tumorigenesis and defines epigenetic prognostic subtypes of canine Diffuse Large B-cell Lymphoma

Serena Ferraresso<sup>1</sup>, Arianna Aricò<sup>1</sup>, Tiziana Sanavia<sup>2</sup>, Silvia Da Ros<sup>1</sup>, Massimo Milan<sup>1</sup>, Luciano Cascione<sup>3,4</sup>, Stefano Comazzi<sup>5</sup>, Valeria Martini<sup>5</sup>, Mery Giantin<sup>1</sup>, Barbara Di Camillo<sup>6</sup>, Sandro Mazzariol<sup>1</sup>, Diana Giannuzzi<sup>1</sup>, Laura Marconato<sup>7</sup> & Luca Aresu<sup>1</sup>

Epigenetic deregulation is a hallmark of cancer characterized by frequent acquisition of new DNA methylation in CpG islands. To gain insight into the methylation changes of canine DLBCL, we investigated the DNA methylome in primary DLBCLs in comparison with control lymph nodes by genome-wide CpG microarray. We identified 1,194 target loci showing different methylation levels in tumors compared with controls. The hypermethylated CpG loci included promoter, 5'-UTRs, upstream and exonic regions. Interestingly, targets of polycomb repressive complex in stem cells were mostly affected suggesting that DLBCL shares a stem cell-like epigenetic pattern. Functional analysis highlighted biological processes strongly related to embryonic development, tissue morphogenesis and cellular differentiation, including HOX, BMP and WNT. In addition, the analysis of epigenetic patterns and genome-wide methylation variability identified cDLBCL subgroups. Some of these epigenetic subtypes showed a concordance with the clinical outcome supporting the hypothesis that the accumulation of aberrant epigenetic changes results in a more aggressive behavior of the tumor. Collectively, our results suggest an important role of DNA methylation in DLBCL where aberrancies in transcription factors were frequently observed, suggesting an involvement during tumorigenesis. These findings warrant further investigation to improve cDLBCL prognostic classification and provide new insights on tumor aggressiveness.

Canine diffuse large B-cell lymphoma (cDLBCL) is the most frequent malignancy of B-lymphocytes in dog and comprises approximately 60–70% of all cases. Dose-intense chemotherapy, bone marrow transplantation and immunotherapy have emerged as the treatments of choice, but current therapeutic strategies are associated with short survival and high relapse rates<sup>1</sup>.

Recently, many efforts have been devoted to characterize, in a comprehensive way, the biological bases of cDLBCL pathogenesis. Two different comparative gene expression studies<sup>2,3</sup> demonstrated that cDLBCL shares similar features with its human counterpart, in particular highlighting the interplay among specific molecular

<sup>1</sup>Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro (PD), Italy. <sup>2</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Lymphoma & Genomics Research Program, Institute of Oncology Research, Bellinzona, Switzerland. <sup>4</sup>Oncology Institute of Southern Switzerland, Bellinzona, Switzerland. <sup>5</sup>Department of Veterinary Medicine, Università degli Studi di Milano, Milano (MI), Italy. <sup>6</sup>Department of Information Engineering, University of Padova, Padova (PD), Italy. <sup>7</sup>Centro Oncologico Veterinario, Bologna, Italy. Serena Ferraresso and Arianna Aricò contributed equally to this work. Correspondence and requests for materials should be addressed to L.A. (email: [luca.aresu@unipd.it](mailto:luca.aresu@unipd.it))

pathways (i.e. NF- $\kappa$ B, PI3K/AKT, Notch and JAK/STAT), which may have potential therapeutic implications<sup>2,4</sup>. In addition, recurrent copy number variations were identified by array comparative genomic hybridization (aCGH) including gains in chr13, syntenic to the region in human chromosome 8 containing *MYC* oncogene, and chr31<sup>5–7</sup>.

However, none of these gene signatures have been translated into clinic, suggesting the need for more robust molecular and prognostic studies. In medical research, evidence showed that genetic and genomics alone are not sufficient to explain the biological variability of cancer. The ability of epigenetic mechanisms to drive cells with the same genome towards different phenotypic identities represents one of the best promises in cancer research<sup>8</sup>. Normal epigenetic processes are usually disrupted during the initiation and progression of cancer and aberrant DNA methylation within CpG islands (CpGIs) is the first critical mark of epigenetic modifications affecting regulatory genomic regions in tumor cells. Hypermethylation of gene promoter regions commonly leads to transcriptional silencing while DNA methylation changes in CpG-poor regions (i.e. genic/intergenic) can play a critical role in the regulation of gene activity and genomic stability<sup>9</sup>.

In human, large-scale genomic studies have shown that perturbations of epigenetic patterning are frequent events in B-cell lymphoma<sup>10</sup>. The characterization of lymphoma methylation signatures and the understanding of how their changes contribute to cancer phenotype have paved the way to new therapeutic approaches in this tumor. Thus, nothing is known about the DNA methylome of canine B-cell lymphoma yet. The epigenetic changes driving B-cell lymphoma have been scarcely investigated in dog and current studies on aberrant methylation patterns focused only on single genes<sup>11–15</sup>. Here, genome-wide characterization of cDLBCL epigenome was performed through a design of a CpG microarray platform targeting more than 40,000 CpG regions and coding sequences (CDS) distributed across the entire dog genome. Then, we hypothesized that DNA methylation patterning in cDLBCLs can provide clues about gene deregulation by identifying aberrantly methylated genes and explaining the different clinical behavior of cDLBCL.

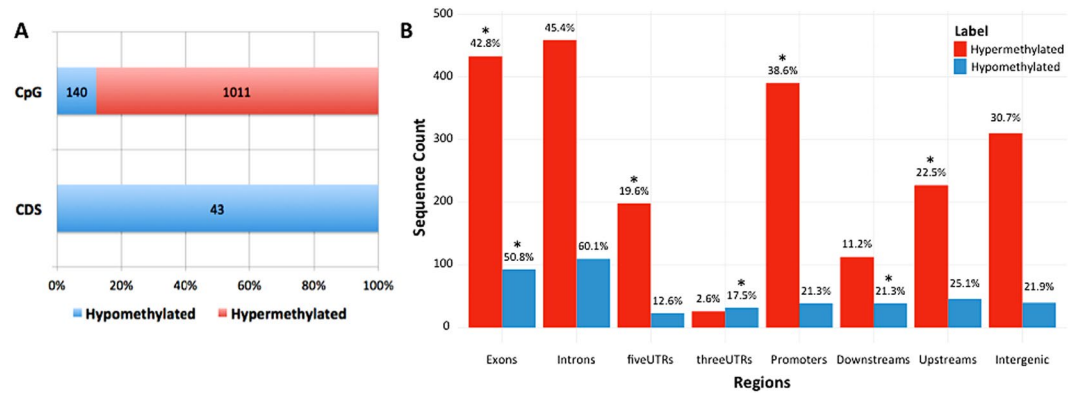
## Results

**Clinical characteristics of DLBCL dogs.** Investigated dogs were composed by 32 (82.1%) purebred and 7 (17.9%) crossbred dogs. Among purebred dogs, German shepherds ( $n = 5$ , 15.6%), Doberman dogs ( $n = 3$ , 9.3%) and Golden retrievers ( $n = 3$ , 9.3%) were the most common. There were 22 (56.4%) females and 17 (43.6%) males. Median age was 7 years (mean, 7.5; range, 3–13 years), and median weight was 25.7 kg (mean, 28.4 kg; range, 6.1–69 kg). Regarding clinical stage, 3 (7.7%) dogs had stage III disease, 16 (41%) dogs had stage IV disease, and 20 (51.3%) dogs had stage V disease. Among dogs with stage V disease, 15 (75%) had bone marrow involvement, 3 (15%) had lung involvement, 1 (5%) had cutaneous involvement, and 1 (5%) had peripheral blood involvement. At the time of diagnosis, 27 (69.2%) dogs were asymptomatic (substage a), whereas 12 (30.8%) dogs showed clinical signs (substage b). Overall, 15 (38.5%) dogs received prednisone at the dose of 0.5–1 mg/kg before being referred. All dogs were treated with the same dose-intense chemotherapeutic (CH) protocol, consisting of L-asparaginase (week 1), Vincristine (weeks 2, 3, 4 and 13), cyclophosphamide (weeks 2 and 13), doxorubicin (weeks 7 and 16), lomustine (weeks 10 and 19), and prednisone (weeks 1 through 20). 23 dogs also received an intradermal injection of an autologous vaccine (VAX) on weeks 4, 5, 6, 7, 12, 16, 20 and 24. 20 (51.3%) dogs relapsed while being treated, whereas in 14 (35.9%) dogs lymphoma recurred after the end of treatment. 5 (12.8%) dogs never relapsed and were still in first complete remission at data analysis closure. Median TTP for all dogs was 162 days (range 1–1174). Median LSS was 281 days (range 12–1175 days), with 1 dog being alive at data analysis closure. Complete clinical features of dogs are reported in Supplementary File 1 (Table S1).

**Identification of loci differentially methylated between cDLBCL and control lymph nodes.** Both raw and normalized methylation data are available at the Gene Expression Omnibus (GEO) repository under accession number GSE94913 (AN: data are kept private until manuscript acceptance, however, for those interested, a reviewer access link has been provided to the editor). Analysis of differential methylated regions (DMRs) focused on 29,513 sequences (29,298 CpG regions and 215 CDS) after filtering out sequences with low methylation. Mann-Whitney Wilcoxon test identified 1,011 hypermethylated CpG regions and 183 hypomethylated sequences (140 CpG and 43 CDS regions) in cDLBCLs ( $n = 37$ ) versus control ( $n = 7$ ) lymph nodes (Fig. 1A and Table S2). These sequences were annotated to genomic regions associated to 823 genes. Interestingly, hyper-methylated sequences were enriched for promoter, 5'-UTRs, upstream and exonic regions (Bonferroni-adjusted  $p$ -values  $< 0.006$ ), whereas hypo-methylated sequences were enriched for exonic, 3'-UTRs and downstream regions (Bonferroni-adjusted  $p$ -values  $< 0.0005$ , Fig. 1B).

**Differentially methylated genes are involved in key pathways of development and morphogenesis.** Functional analysis of differentially methylated genes identified 22 Biological Process (BP), 5 Molecular Function (MF), 2 Cellular Component (CC) Gene Ontology (GO) terms and 3 KEGG pathways as significantly enriched (Tables 1 and S4 for the list of annotated genes). Overall, 19 out of 22 GO\_BP enriched terms were involved in the development/morphogenesis of anatomical structures, including 10 terms directly linked to embryogenesis (e.g. proximal/distal pattern formation, embryonic forelimb/hindlimb morphogenesis, spinal cord association neuron differentiation, embryonic digestive tract morphogenesis, anterior/posterior axis specification) and 9 related to specific tissues development (e.g. muscle organ development, ureter development, neuron differentiation, positive regulation of chondrocyte differentiation). All these terms were mainly represented by genes playing key role in regulating organogenesis (SHH, BMPs, GREM1), body patterning (HOX gene family) and tissues differentiation (FGFR2, FGF18, SOX9).

All GO\_MF and GO\_CC enriched terms were involved in transcription regulation, as well as one of the enriched GO\_BP terms (i.e. positive regulation of transcription from RNA polymerase II promoter). These terms were represented by several transcription factors belonging to different families such as HOX, AP-2, IRX and



**Figure 1.** Genomic distribution of differentially methylated sequences in cDLBCLs versus control lymph nodes. **(A)** Percentages of hyper- and hypo-methylated probes across CpG regions and CDS. Corresponding sequence counts are reported in each barplot. **(B)** Distribution of hyper- and hypo-methylated features across different genomic locations. Percentages with respect to the corresponding total number of hyper- and hypo-methylated sequences are reported. Sequence counts allow repetitions, since a sequence can overlap more than one genomic location. Asterisks (\*) indicate enriched genomic locations, according to Fisher's Exact test (Supplementary File 1).

Zinc-finger transcription factors. Among the 3 significantly enriched KEGG pathways, “Pathways in cancer” (KEGG ID 05200) showed highest significance (Bonferroni-adjusted  $p$ -value  $\leq 0.001$ ) and included 31 differentially methylated genes (Table S4) involved in key mechanisms such as control of apoptosis (p53, HRK), proliferation (PDGFA, FGFs, RARB, SHH, CCNE1, WNT, SHH) and angiogenesis (NOS2, SLC2A1). The 209 differentially methylated genes belonging to at least one enriched GO\_BP, GO\_MF term or KEGG pathway (Table S4) were mapped to the protein-protein interaction (PPI) annotations from STRING database<sup>16</sup>: 181 genes shared at least one interaction, 98 of them either experimentally validated or database-curated (Fig. 2).

GSEA confirmed these results with highly enriched GeneSets related to tissue development and morphogenesis (Table S5). Focusing on lymphoid-specific gene expression signatures collected in Staudt's SignatureDB (<https://lymphochip.nih.gov/signaturedb/>)<sup>17</sup>, 61 gene sets were found significantly enriched (FDR < 25%), within those 46 showed NOM  $p$ -value < 0.05. In particular, two gene signatures specific for DLBCL<sup>18</sup> and predictive for survival outcome (STROMAL-1 and STROMAL-2\_DLBCL\_SURVIVAL\_PREDICTOR) resulted significant (Table S5).

**Associations between methylation levels and clinical features in cDLBCL.** The associations between tumor methylation levels and clinical features (Supplementary File 1 - Table S1) were investigated through statistical pairwise comparisons. For each comparison, samples were grouped as follows: (i) stage (III-IV vs. V); (ii) substage (a vs. b); (iii) extra-nodal sites infiltration (yes vs. no); (iv) treatment (CH vs. CH + VAX); (v) steroid administration before diagnosis (yes vs. no); (vi) relapse (yes vs. no OR before the end of therapy vs. after the end of therapy + no relapse). Mann-Whitney Wilcoxon test did not detect any association, whereas F-test identified 87 sequences showing a significant differential variability in at least one clinical factor (Table S6). In particular, 4 genes (U6, GRB10, CCDC73, ZFAT) and one miRNA (ENSCAFT00000040944) showed increased methylation variability in dogs relapsing before the end of the therapy whereas 7 genes (OBSCN, THSD1, RNFT2, C6orf201, GRB10, CEP170B, ETV6) were associated to bone marrow infiltration. Only 3 CpG regions were found significant combining the clinico-pathological features by multivariate linear regression model. These were related to exonic regions of EIF2D (stage, LSS) and RNASEH1 (age, therapy and TTP), and the intronic regions of KCNAB2 (age, stage, substage, therapy and TTP). Interestingly, among the 7,526 sequences characterized by a nominal  $p$ -value  $\leq 0.01$  (Table S7), most of the 20 genes associated to a worse clinical behavior (i.e. relapse, TTP and LSS) are involved in regulation of apoptosis and cell cycle (CCND2, SMG7, BCL2L1, BAG1).

**DNA methylation identifies three DLBCL subgroups with different overall survival.** Beyond the classification provided by the available clinical factors, we investigated whether methylation profiles were able to provide a different stratification of the cDLBCL samples. The principal component analysis (PCA) explained nearly 25% of the variations in the methylation profiles across cDLBCLs with the first two components (PCs). In total, 138 CpG sequences resulted significantly correlated to the first PC, while no sequences were selected for the second PC. Hierarchical clustering (HCL) on the methylation levels of these CpGs (Fig. 3) identified 3 different subgroups that apparently did not reflect any statistically significant association with the clinical features listed in Table S1. Interestingly, dogs in Cluster#3 showed a higher median LSS (534 days) than Cluster#2 and Cluster#1 (296 and 172 days, respectively). Moreover, Kaplan-Meier curves for LSS confirmed a significant difference in terms of survival between Cluster#3 and Cluster#1 ( $p$ -value = 0.02, Fig. 4). Finally, clustering across the 138 selected sequences identified two distinct groups of genes (Fig. 3). The former, including 4 CpG sequences associated to FAM181A, PDE4C, PARVA and GRID1, showed higher methylation levels in Cluster#1 compared with Cluster#2 and #3. Conversely, the latter group, including all the other CpGs, showed an opposite behavior.

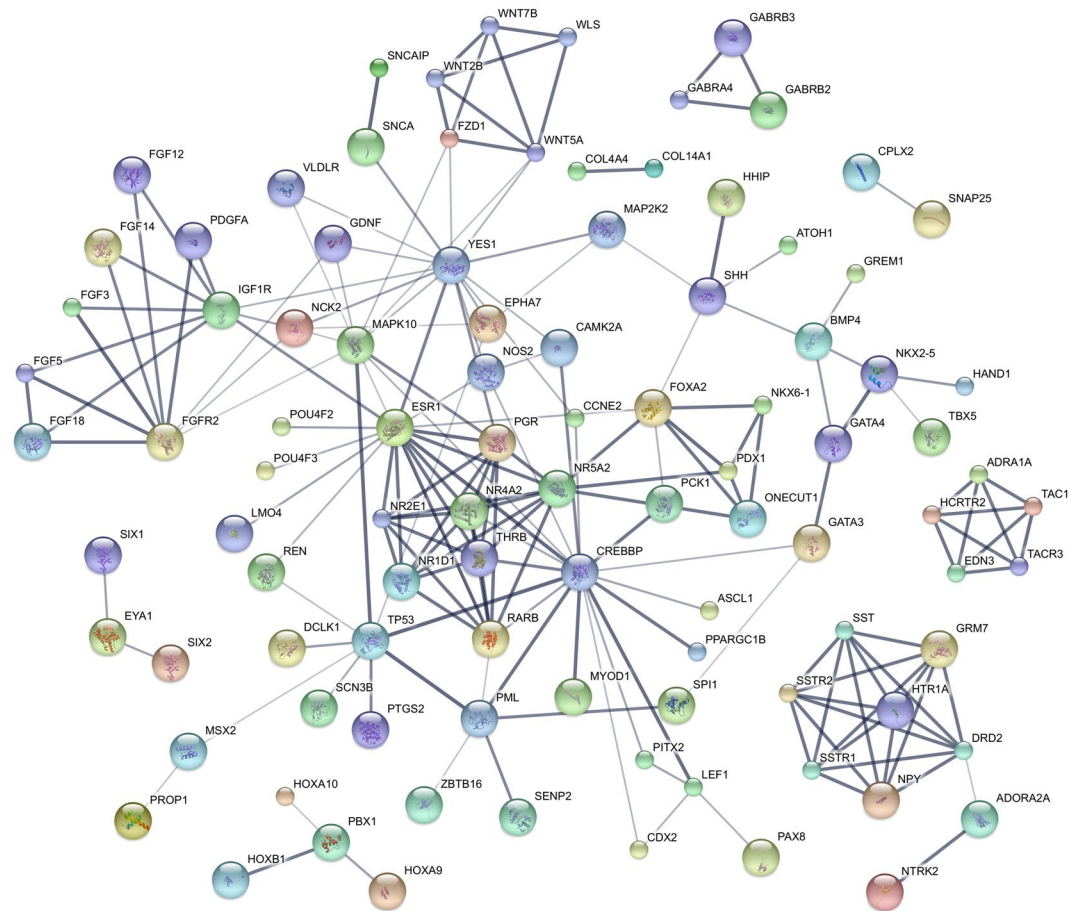
Category	ID	Term	Bonferroni-adjusted pvalue
GO_BP	GO:0009954	proximal/distal pattern formation	1.1E-09
	GO:0048665	neuron fate specification	2.8E-08
	GO:0035115	embryonic forelimb morphogenesis	2.1E-07
	GO:0009952	anterior/posterior pattern specification	5.9E-07
	GO:0021522	spinal cord motor neuron differentiation	8.0E-05
	GO:0072189	ureter development	8.2E-05
	GO:0001764	neuron migration	9.8E-05
	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	1.3E-04
	GO:0045666	positive regulation of neuron differentiation	6.7E-04
	GO:0045665	negative regulation of neuron differentiation	1.1E-03
	GO:0007267	cell-cell signaling	1.1E-03
	GO:0035116	embryonic hindlimb morphogenesis	3.2E-03
	GO:0001759	organ induction	5.3E-03
	GO:0003148	outflow tract septum morphogenesis	9.6E-03
	GO:0002053	positive regulation of mesenchymal cell proliferation	1.5E-02
	GO:0007517	muscle organ development	2.2E-02
	GO:0009948	anterior/posterior axis specification	2.9E-02
	GO:0048557	embryonic digestive tract morphogenesis	3.2E-02
	GO:0048664	neuron fate determination	3.2E-02
	GO:0021527	spinal cord association neuron differentiation	3.2E-02
GO:0032332	positive regulation of chondrocyte differentiation	3.2E-02	
GO:0048701	embryonic cranial skeleton morphogenesis	3.9E-02	
GO_MF	GO:0043565	sequence-specific DNA binding	1.6E-10
	GO:0001077	transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	2.2E-10
	GO:0000978	RNA polymerase II core promoter proximal region sequence-specific DNA binding	1.1E-06
	GO:0003682	chromatin binding	1.3E-06
	GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	2.1E-02
GO_CC	GO:0005667	transcription factor complex	9.0E-06
	GO:0005634	nucleus	6.6E-03
KEGG	5200	Pathways in cancer	3.2E-04
	4080	Neuroactive ligand-receptor interaction	2.3E-03
	4950	Maturity onset diabetes of the young	3.2E-03

**Table 1.** Significantly enriched GO terms and KEGG pathways.

**The magnitude of methylation disruption reveals prognostic relevance.** The stratification of cDLBCLs considering degree and direction of relative methylation difference between cDLBCL and the mean of control lymph nodes, defined as methylation disruption (see Methods), shows how common hyper- and hypo-methylation events among samples are able to identify distinct cDLBCL subgroups according to methylation variability profiles (MVPs, details in Methods and Supplementary File 1). The first two PCs from PCA on methylation changes explained more than 80% of the variation (Supplementary File 1 - Figure S9). HCL based on the MVPs of CpGs correlating to the first PC identified 3 different cDLBCL subgroups, where the previously identified Cluster#3 was confirmed also by MVPs analysis (Cluster#C in Figure S9).

Focusing on groups of cDLBCLs characterized by methylation disruption driven by the most variable hyper- and hypo-methylation events, consensus HCL was performed on subsets of sequences ranging between 250 and 20,000, sorted by decreasing median absolute deviations (MADs) on MVPs. Interestingly, up to the first 2,000 sequences with the highest MADs the clustering was able to identify a specific subgroup of 6 cDLBCLs characterized by the highest variability of differential methylation changes between cDLBCLs and control lymph nodes (Supplementary File 1 - Figure S10); 5 of these dogs showed short-term LSS (LSS < 180 days, p-value < 0.03, Fisher's Exact Test).

**Technical and functional validation of microarray results.** A technical validation of microarray platform by methylation-specific PCR was performed on 5 differentially methylated genes (FGFR2, HOXD10, RASAL3, CYP1B1 and ITIH5) involved in key biological pathways of cancer development (e.g. MAP/ERK, Homeobox signaling and FGF signaling pathways). The methylation levels were assessed in 13 cDLBCLs and 5 control lymph nodes. The promoters of the tested genes resulted hyper-methylated and a statistical significance for HOXD10, RASAL3, CYP1B1 and ITIH5 was found (p < 0.01), thus confirming the reliability of the microarray platform (details in Supplementary File 1).



**Figure 2.** PPI sub-networks from differentially methylated genes belonging to the enriched GO terms and KEGG pathways. The thickness of network edges correlates with the confidential score provided by STRING database: the thicker is the edge, the higher is the confidence score of the interaction.

A functional validation of microarray data was also performed selecting 3 hypermethylated genes (*CADM1*, *CDH11* and *ABCB1*). In this respect, the mRNA restoration after the treatment of a canine B-cell lymphoma cell line (*CLBL1*)<sup>19</sup> with two hypomethylating agents (azacytidine and decitabine) was evaluated through quantitative Real Time PCR. The three transcripts resulted highly expressed in control lymph nodes and scarcely expressed in *CLBL1* cells. After azacytidine treatment a significantly restoration of expression for the three genes was found, conversely decitabine affected only *ABCB1* expression (Supplementary File 1).

## Discussion

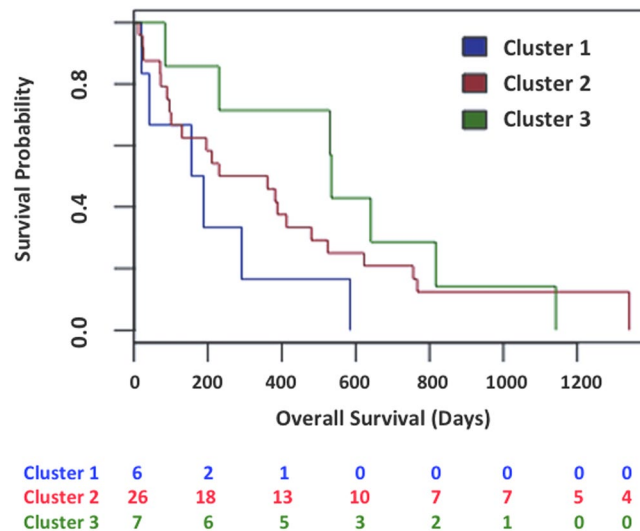
Origin of B-cell lymphoma in dog results from inherited mutations in the germ line or changes in DNA sequences during life<sup>20</sup>. Recently, extensive gene expression profiling identified molecular signatures of cDLBCL and defined two distinct subgroups with prognostic significance<sup>2,3</sup>. However, the biology of this tumor is still not entirely explained by genomic events and transcriptional programs, and much less is known about epigenetic changes<sup>11</sup>. Therefore, genome-wide DNA methylome has been investigated in this study using for the first time a canine DNA CpG microarray. Results revealed that cDLBCLs are characterized by a widespread aberrant methylation affecting 1,194 regions, corresponding to 823 genes. The hyper-methylated sequences were enriched in upstream ( $\leq 10$  kb) or promoter regions, while the hypo-methylated sequences were preferentially located in gene bodies and downstream regions. This was quite expected considering that CpGs are highly susceptible to DNA methyltransferases in cancer, determining gene silencing. Conversely, CpG-poor regions undergo to a global decrease of genomic DNA methylation affecting genome stability, transcriptional elongation, and RNA splicing<sup>9, 21–23</sup>.

In order to indirectly confirm the relationship between methylation and transcription in cDLBCL, we integrated our data with the cDLBCL gene expression dataset published by Mudaliar *et al.*<sup>2</sup>. A total of 107 significant genes were found in common between the two platforms and 98% of the hypermethylated upstream/promoter regions resulted downregulated. The same was observed for hyper-methylated CpGs in gene bodies. By contrast, hyper-methylation of the downstream regions and CpGs hypo-methylation did not show any consistent relationship with gene expression. Hypo-methylated CpGs-poor CDS regions represented the only exception, being downregulated in tumors.



**Figure 3.** Heatmap of the 138 CpG sequences highly correlated with the first principal component. Methylation levels were centered and scaled by sequence.

The biological processes significantly enriched in cDLBCLs appear to be strongly related to embryonic development, tissue morphogenesis and cellular differentiation. DMRs included genes encoding proteins with key roles in development and cell fate determination in all types of cells, such as HOX, BMP, WNT, and SOX. Interestingly, inappropriate or deregulated expression of HOX genes has been implicated in several human cancers and associated to promoter methylation<sup>24–26</sup>. Furthermore, the expression of HOX genes during maturation of hematopoietic cells results tightly regulated and modifications of this mechanism seem to contribute to neoplastic transformation. The hyper-methylation of several members of HOX gene family in our study suggests a



**Figure 4.** Survival outcomes in patient cohort. Kaplan-Meier curves for LSS according to cDLBCL subgroups defined by PCA and hierarchical clustering on methylation patterns.

possible contribution to cDLBCL pathogenesis and points to the importance of certain signaling pathways across human and dog.

Wnt pathway is one of the “usual suspects” in cancer biology. The cellular processes modulated by Wnts range from stem cell self-renewal to cell motility, and are mediated by transcriptional activation as well as through direct effects on cytoplasmic targets<sup>27</sup>. In the present study, several components of this pathway (i.e. WNT2B, WNT5A, WNT7B, FZD1 and LEF1) were found hyper-methylated in cDLBCL. This signaling cascade is multifaceted and somehow enigmatic since it has been widely reported that some members can play tumor-promoting or suppressing role depending on the cell type or availability of key receptors<sup>28,29</sup>. An emblematic example is given by Wnt5a, reported overexpressed in several human cancers<sup>27,29</sup>, that in the presence of specific FZ isoforms, could promote tumor growth by activating the cancer-promoting canonical Wnt signaling pathway. However, in hematological malignancies, including B-cell lymphoma, WNT5A acts as a tumor suppressor able to antagonize the WNT/ $\beta$ -catenin signaling<sup>30–32</sup> and it is found silenced by tumor-specific methylation. Hyper-methylation of Wnt5a promoter was observed in the present study, supporting its role in inhibiting B-cell proliferation, while the significance of FZD1 and LEF1 epigenetic silencing remains unclear. To date, both genes were reported activated in several cancers and associated to chemoresistance and poor prognosis<sup>33,34</sup>.

BMPs frequently inhibit cell differentiation and proliferation and are involved in cancerogenesis in different manners. Indeed, there are controversial *in vitro* and *in vivo* studies regarding the role of BMPs in promoting tumorigenesis and metastasis<sup>35</sup> and showing biological behavior associated with cancer origin<sup>36</sup>. However, several studies reported that BMPs can influence the hematopoietic system and regulate development of hematopoietic stem cells. Interestingly, B- and T-cell lymphopoiesis is inhibited by inducing the activation of Smad 1/5/8<sup>37</sup>. In the present study, three BMPs members belonging to the Transforming growth factor beta (TGF- $\beta$ ) superfamily, namely BMP3, BMP4 and BMP7, were found hyper-methylated in cDLBCL, supporting the role of BMP families as tumor suppressor genes. BMP7 was demonstrated to be methylated both in human DLBCL and follicular lymphoma indicating its physiological relevance<sup>38</sup>.

In our study, we identified a high proportion of Polycomb Group (PcG)-target genes (243/823, 29.5%) differentially methylated. The genes included WNTs, Hedgehog, BMPs, PAXs, FGFs, and FOX factors. Polycomb group (PcG) proteins represent a global silencing system involved in development control and they are able to regulate the transition from proliferation to differentiation, contributing to stem-cell maintenance, and inhibit inappropriate activation of differentiation programs. Recent studies have provided evidence that PcG-target genes are frequently hyper-methylated in several tumors, including lymphoma, and our findings confirm the same trend for cDLBCL. As proposed by Martín-Subero *et al.*<sup>39</sup>, this scenario allows a double interpretation. The most consistent relies on the effect of PcG-target genes aberrant methylation in tumor precursor cells with stem cell-like features initially conferring a growth advantage and abnormal proliferation. In turn, this uncontrolled expansion and the consequent genome instability would promote the gain of further oncogenic mutations (i.e. chromosomal aberration), which foster malignancy progression<sup>22,40,41</sup>. This concept supports the “cancer stem cell theory” which states that primordial cancer-initiating cells, exhibiting self-renewal capacity and multilineage potential, may expand and form the biological origin for the rest of the tumor<sup>42,43</sup>. The second hypothesis relies on the capacity of the neoplastic cells to secondary acquire a stem cell-like epigenetic pattern through gene deregulation caused by chromosomal aberrations<sup>39</sup>. The mechanisms regulating the interplay between DNA promoter methylation and PcG-driven silencing, if any, are still unknown. DNA methylation and histone methylation (i.e. H3K27me3) directed by the PcG family protein EZH2 are mutually exclusive in normal cells, but this relationship seems not maintained in cDLBCL<sup>22,40</sup>. In human, mutations in the SET domain of EZH2 are associated to aberrant

epigenetic events both in DLBCL and follicular lymphoma<sup>44, 45</sup>, whereas this is the first evidence supporting a role of PcG proteins in dog and further investigation are necessary to study the mechanism of EZH2 in cDLBCL.

In human DLBCL, intra-tumor and inter-patient variability in promoter DNA methylation, as well as specific methylation states have been reported to be associated to tumor clinical behaviour<sup>46–48</sup>. Here, the F-test was applied for our data and the analysis yielded 4 genes (U6, GRB10, CCDC73 and ZFAT) with the highest ability to predict survival. Both GRB10 and ZFAT are known to play a role in regulating hematopoietic stem cell self-renewal and haematopoiesis and several studies revealed an association with cancer progression<sup>49–51</sup>. However, their contribution in B-cell lymphoma remains poorly understood. One limitation of our study is the number of cases that might affect the statistical power of testing differential methylation for the clinical features. Further studies with a higher caseload are needed to assess the associations between specific gene methylation and clinical behaviour.

A second aim of this study was to explore whether epigenetic profiles might help to classify cDLBCLs into biologically relevant subgroups. Analysis of methylation levels identified 3 potential subgroups described by 138 CpG sequences which mainly contributed to methylome-wide variations among the cDLBCLs and characterized by biological pathways that are mainly related to nervous system development, cell differentiation and morphogenesis (data not shown). A clear correspondence with the canine ABC and GCB-like DLBCLs, characterized by a differential expression of NF- $\kappa$ B and B-cell receptor pathways, was not established<sup>3</sup>. A comparative approach was attempted by considering the canine-specific geneset, reported by Richards *et al.*<sup>3</sup>, whose expression was able to distinguish cBCL into “ABC-like” and “CCB-like” subtypes. Out of 787 genes (1,180 probes), 597 (1,464 features) were represented in the dog CH3 microarray and the corresponding methylation levels were employed as dataset to perform HCL of cDLBCLs. No distinct separation into two subgroups was achieved; the same evidence was obtained when limiting the dataset to those regions (n = 17) already found differentially methylated between cDLBCL and control lymph nodes (data not shown).

Interestingly, one of the cDLBCL subgroups defined by our methylation profiles (Cluster#3) is associated to long-term survival with respect to the other samples. Comparable results were obtained by stratifying cDLBCLs considering the magnitude of methylation changes with respect to control lymph nodes. This latter analysis reproduces the methodological approach described in Chambwe *et al.*, where 6 subgroups of DLBCL in human with prognostic relevance were identified, “refining” the ABC- and GCB-DLBCL classification<sup>46</sup>. The main advantage of this approach is exploiting the methylation profiles of control samples as a reference in order to perform PCA and clustering focusing on the hyper- and hypo-methylation events in each sample. In the present study, 3 cDLBCL subgroups with different magnitude of DNA methylation changes were identified, confirming the Cluster#3 characterized by long-term survival (Cluster#C). Since many sequences were characterized by low differential methylation levels, which might confound the detection of small cDLBCL subgroups defined by a limited number of sequences showing high variations of methylation disruption, a consensus HCL was performed on subsets of sequences sorted by decreasing MADs of the MVPs. Interestingly, a cluster of 6 dogs based on about 2,000 sequences characterized by the highest methylation variability changes was significantly associated to poor prognosis (LSS < 180 days, p < 0.03). The consensus clustering showed that 6 dogs characterized by poor prognosis robustly clustered together and not with the other samples of the two other clusters by performing HCL several times on subset of samples (Figure S10C). In addition, methylation changes observed in these dogs were characterized by a higher number of hyper-methylation events compared to the other samples, as shown by the average of density plots of their methylation changes (Figure S10A) and the highest Methylation Variability Scores (Figure S10B) defined in Chambwe *et al.*<sup>46</sup>, (i.e. quantitative measures reflecting the magnitude of methylation disruption). This finding supports the hypothesis that the progressive accumulation of aberrant epigenetic changes might confer aggressiveness to the tumor. Future studies with a higher number of cases will be definitely necessary to better assess the contribution of epigenetic profiles in both biological and clinical stratification of cDLBCL. In this context, investigating the level of methylation variability of cDLBCL might be a starting point to highlight the contribution of DNA methylation in the clonal evolution of this tumor.<sup>6</sup>

In conclusion, we profiled genome-wide DNA methylation in cDLBCLs using DNA methylation array and we technically and functionally validated our results by methylation-specific PCR and the treatment of a canine B-cell lymphoma cell line with hypomethylating agents, providing insights on epigenetic switching and heterogeneity of this tumor. Findings collected, herein, suggest that cDLBCL aberrantly has tumorigenic and stem cell-like signatures, highlighting some methylation-based cDLBCL subgroups showing prognostic relevance. In future, new brand high-resolution techniques will be required to better define the contribution of methylation in cDLBCL associated with functional studies of the aberrant methylated genes and the identification of putative tumor biomarkers to predict the clinical outcome.

## Methods

**Dogs and samples.** The study cohort included 40 dogs affected by newly diagnosed, multicentric DLBCL that underwent complete and standardized staging work-up and that were treated with chemotherapy or chemo-immunotherapy<sup>52</sup>. The diagnosis of DLBCL was obtained by histopathological and immunohistochemical analysis (CD20 and CD79) of one enlarged peripheral lymph node surgically removed at initial presentation. A portion of the tumor was preserved frozen in RNAlater<sup>®</sup> solution (Life Technologies, Carlsbad, CA) under sterile conditions. Medical records of all dogs were reviewed to obtain relevant clinical information, including signalment, breed, sex, age, hematological and biochemical abnormalities, clinical stage, substage and treatment (Supplementary File 1-Table S1). Time to progression (TTP) was measured as the interval between initiation of treatment and progressive disease (PD). Dogs not experiencing PD at the end of the study or dogs lost to follow-up before PD were censored for TTP analysis. Lymphoma-specific survival (LSS) was measured as the interval between initiation of treatment and lymphoma-related death. 8 dogs with no cancer disease and defined clinically healthy by a complete blood exam and physical examination underwent lymphadenectomy and lymph nodes were used as controls. Samples stored in RNA-later were analyzed to selectively extract DNA from regions compatible with lymphoid follicles.



The study was approved by Committee of the University of Padova (CPDA148778/14 - protocol 20086MSFH3) and a mandatory written consent from all dog's owners was obtained. All the experiments were performed in accordance with relevant guidelines and regulations.

**Dog CH3 microarray design.** In order to assess methylation profiles of dogs affected by DLBCL, a canine CpG microarray platform was developed (GEO accession: GPL23069). Probe design was carried out by the Agilent bioinformatic support team using proprietary prediction algorithms to locate CpG Islands on *C. familiaris* draft genome as deposited on Ensembl database (CanFam 3.1) and to design high quality oligo-probes. Microarray probes were selected in order to provide the highest possible coverage of dog genome. CDS regions and CpG islands were given top priority. Chromosome X was excluded from analysis and probe design. A total of 170,000 probes (60mers, sense orientation) were designed on both CpG and CDS regions. In details, 102,000 probes were designed targeting a total of 36,807 CpG regions while 68,000 probes were directed against 672 CDS; average base pare tiling was 90 bp. Microarray probes were synthesized *in situ* using the Agilent non-contact ink-jet technology with a 4 × 180 K format. Each array included Agilent's default positive and negative controls.

**Sample processing and data normalization.** Detailed description of sample processing, data quality assessment and normalization is reported in Supplementary File 1. Briefly, for both cDLBCLs and control lymph nodes, DNA methylation was measured by two-color competitive hybridization between the methylated fraction and a not-enriched aliquot of the same DNA. Loess normalization was applied to correct for the Cy3/Cy5 dye bias for each dye. After quality control (QC) of the resulting log2 ratios, between-samples Quantile normalization was then applied to the QC-passed arrays (37 cDLBCLs and 7 controls).

**Data Analysis.** Data analyses were performed using R statistical computing software (<http://www.r-project.org>). Details are reported in Supplementary File 1. The median of the probe signal was calculated to retrieve the methylation levels of the target 672 CDS and 36,807 CpG regions. DMRs between cDLBCLs and lymph nodes were identified by Mann–Whitney Wilcoxon test, filtering out sequences showing enriched methylation in less than 25% of the two groups (i.e. cDLBCLs and control lymph nodes). Both Mann–Whitney Wilcoxon and F-test for differential variability were then applied to study associations between gene methylation levels and clinical features. Possible combinations of multiple clinical features were also investigated by multivariate linear regression model. For all the statistical tests, Bonferroni-adjusted p-values < 0.01 were considered significant. Functional characterization of DMRs was performed by different state-of-the-art enrichment approaches as topGO and Gene Set Enrichment Analysis (GSEA).

The ability of methylation profiles in defining new tumor stratifications was first investigated by applying PCA. Target sequences showing high correlation (i.e. above 0.85) with the main principal components (PCs) were selected by using one-way analysis of variance<sup>53</sup>. Focusing on these sequences, HCL using Euclidean distance and Ward linkage was applied to identify cDLBCLs subgroups characterized by different methylation profiles and clusters of CpG/CDS regions characterized by a specific methylation pattern across cDLBCLs.

PCA and HCL were applied to MVPs, representing the density functions of the differential methylation levels between each cDLBCL and the median methylation level calculated across control lymph nodes. Clustering was performed according to a distance defined by the area bounded by the MVP-based density curves of sample pairs (Supplementary File 1). Moreover, consensus clustering on subsets of sequences sorted by decreasing MAD was applied to stratify according to sequences characterized by highly variable methylation disruption across cDLBCL samples. To evaluate the robustness of the results, the consensus clustering provides quantitative evidence for determining the number and membership of possible clusters by randomly subsampling 1,000 times the samples. Clustering performance from subsampling are summarized by a consensus matrix with values ranging between 0 (never clustered together) and 1 (always clustered together). In addition, each cDLBCL sample was characterized by a Methylation Variability Score (MVS) representing the difference of the area under the density curves between the MVP of each cDLBCL sample and the median MVP from the controls: the higher the MVS, the greater the methylation disruption<sup>46</sup>. Further details on the analysis of methylation disruption and the consensus clustering are reported in Supplementary File 1. Finally, Fisher's exact test was performed on the whole CpG-probe set  $\beta$  values treated as categorical data to identify the differentially methylated probes. The probes were classified as “methylated” ( $\beta$  value  $\geq 0.5$ ) or “unmethylated” ( $\beta$  value < 0.5) (for details see Table S3).

## References

1. Aresu, L. Canine Lymphoma, More Than a Morphological Diagnosis: What We Have Learned about Diffuse Large B-Cell Lymphoma. *Front Vet Sci.* **3**, 77 (2016).
2. Mudaliar, M. A. *et al.* Comparative gene expression profiling identifies common molecular signatures of NF- $\kappa$ B activation in canine and human diffuse large B cell lymphoma (DLBCL). *PLoS One* **8**(9), e72591 (2013).
3. 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**(16), 5029–39 (2013).
4. Seelig, D. M. *et al.* Constitutive activation of alternative nuclear factor kappa B pathway in canine diffuse large B-cell lymphoma contributes to tumor cell survival and is a target of new adjuvant therapies. *Leuk Lymphoma* **8**, 1–9 (2016).
5. Tagawa, H. *et al.* Comparison of genome profiles for identification of distinct subgroups of diffuse large B-cell lymphoma. *Blood* **106**(5), 1770–7 (2005).
6. Aricò, A. *et al.* Array-based comparative genomic hybridization analysis reveals chromosomal copy number aberrations associated with clinical outcome in canine diffuse large B-cell lymphoma. *PLoS One* **9**(11), e111817 (2014).
7. Thomas, R. *et al.* Refining tumor-associated aneuploidy through ‘genomic recoding’ of recurrent DNA copy number aberrations in 150 canine non-Hodgkin lymphomas. *Leuk Lymphoma* **52**(7), 1321–35 (2011).
8. Toh, T. B., Lim, J. J. & Chow, E. K. Epigenetics in cancer stem cells. *Mol Cancer* **16**(1), 29 (2017).
9. Stirzaker, C., Taberlay, P. C., Statham, A. L. & Clark, S. J. Mining cancer methylomes: prospects and challenges. *Trends Genet.* **30**(2), 75–84 (2014).

10. Pan, H. *et al.* Epigenomic evolution in diffuse large B-cell lymphomas. *Nat Commun.* **6**, 6921 (2015).
11. Ferrarresso, S. *et al.* Epigenetic Silencing of TFPI-2 in Canine Diffuse Large B-Cell Lymphoma. *PLoS ONE* **9**(4), e92707 (2014).
12. Sato, M. *et al.* Hypermethylation of the death-associated protein kinase CpG island in canine B-cell lymphoid tumors. *Vet Immunol Immunopathol.* **161**(3–4), 222–31 (2014).
13. Fujiwara-Igarashi, A. *et al.* Inhibition of p16 tumor suppressor gene expression via promoter hypermethylation in canine lymphoid tumor cells. *Res Vet Sci.* **97**(1), 60–3 (2014).
14. Bryan, J. N. *et al.* Hypermethylation of the DLC1 CpG island does not alter gene expression in canine lymphoma. *BMC Genet.* **13**(10), 73 (2009).
15. Tomiyasu, H. *et al.* Evaluation of DNA methylation profiles of the CpG island of the ABCB1 gene in dogs with lymphoma. *Am J Vet Res.* **75**(9), 835–41 (2014).
16. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–52 (2015).
17. Shaffer, A. L. *et al.* A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol Rev.* **210**, 67–85 (2006).
18. Lenz, G. *et al.* Lymphoma/Leukemia Molecular Profiling Project. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med.* **359**(22), 2313–23 (2008).
19. Rütgen, B. C. *et al.* Establishment and characterization of a novel canine B-cell line derived from a spontaneously occurring diffuse large cell lymphoma. *Leuk Res.* **34**(7), 932–8 (2010).
20. Elvers, I. *et al.* Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* **25**(11), 1634–45 (2015).
21. Choi, J. K., Bae, J. B., Lyu, J., Kim, T. Y. & Kim, Y. J. Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol.* **10**(9), R89 (2009).
22. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer.* **11**(10), 726–34 (2011).
23. Fleischer, T. *et al.* Genome-wide DNA methylation profiles in progression to *in situ* and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol.* **15**(8), 435 (2014).
24. Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* **10**(5), 361–71 (2010).
25. Barber, B. A. & Rastegar, M. Epigenetic control of Hox genes during neurogenesis, development, and disease. *Ann Anat.* **192**(5), 261–74 (2010).
26. Skvarova Kramarzova, K. *et al.* Homeobox gene expression in acute myeloid leukemia is linked to typical underlying molecular aberrations. *J Hematol Oncol.* **24**(7), 94 (2014).
27. Sedgwick, A. E. & D'Souza-Schorey, C. Wnt Signaling in Cell Motility and Invasion: Drawing Parallels between Development and Cancer. *Cancers (Basel).* **29**, 8(9) (2016).
28. Serman, L., Nikuseva Martic, T., Serman, A. & Vranic, S. Epigenetic alterations of the Wnt signaling pathway in cancer: a mini review. *Bosn J Basic Med Sci.* **14**(4), 191–4 (2014).
29. McDonald, S. L. & Silver, A. The opposing roles of Wnt-5a in cancer. *Br J Cancer.* **101**(2), 209–214 (2009).
30. Staal, F. J. & Clevers, H. C. WNT signalling and haematopoiesis: a WNT-WNT situation. *Nat Rev Immunol.* **5**(1), 21–30 (2005).
31. Blanc, E., Roux, G. L., Benard, J. & Raguenez, G. Low expression of Wnt-5a gene is associated with high-risk neuroblastoma. *Oncogene* **24**, 1277–1283 (2005).
32. Bisson, J. A., Mills, B., Paul Helt, J. C., Zwaka, T. P. & Cohen, E. D. Wnt5a and Wnt11 inhibit the canonical Wnt pathway and promote cardiac progenitor development via the Caspase-dependent degradation of AKT. *Dev Biol.* **398**(1), 80–96 (2015).
33. Wu, W. *et al.* High LEF1 expression predicts adverse prognosis in chronic lymphocytic leukemia and may be targeted by ethacrynic acid. *Oncotarget.* **7**(16), 21631–43 (2016).
34. Flahaut, M. *et al.* The Wnt receptor FZD1 mediates chemoresistance in neuroblastoma through activation of the Wnt/beta-catenin pathway. *Oncogene.* **28**(23), 2245–56 (2009).
35. Thawani, J. P. *et al.* Bone morphogenetic proteins and cancer: review of the literature. *Neurosurgery.* **66**(2), 233–46 (2010).
36. Zhang, L. *et al.* BMP signaling and its paradoxical effects in tumorigenesis and dissemination. *Oncotarget.* **7**(47), 78206–78218 (2016).
37. Huse, K. *et al.* Role of Smad proteins in resistance to BMP-induced growth inhibition in B-cell lymphoma. *PLoS One.* **7**(10), e46117 (2012).
38. Bethge, N. *et al.* A gene panel, including LRP12, is frequently hypermethylated in major types of B-cell lymphoma. *PLoS One.* **9**(9), e104249 (2014).
39. Martin-Subero, J. I. *et al.* New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood.* **113**(11), 2488–97 (2009).
40. Velichutina, I. *et al.* EZH2 mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood.* **116**(24), 5247–55 (2010).
41. Marchesi, I. & Bagella, L. Targeting Enhancer of Zeste Homolog 2 as a promising strategy for cancer treatment. *World J Clin Oncol.* **7**(2), 135–48 (2016).
42. Oren, O. & Smith, B. D. Eliminating Cancer Stem Cells by Targeting Embryonic Signaling Pathways. *Stem Cell Rev* (2016).
43. Frank, N. Y., Schatton, T. & Frank, M. H. The therapeutic promise of the cancer stem cell concept. *J Clin Invest.* **120**(1), 41–50 (2010).
44. Jiang, Y., Hatzi, K. & Shaknovich, R. Mechanisms of epigenetic deregulation in lymphoid neoplasms. *Blood.* **121**(21), 4271–9 (2013).
45. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet.* **42**(2), 181–5 (2010).
46. Chambwe, N. *et al.* Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. *Blood* **123**(11), 1699–708 (2014).
47. Shaknovich, R. *et al.* DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood.* **18**, 116(20), e81–9 (2010).
48. Shaknovich, R., De, S. & Michor, F. Epigenetic diversity in hematopoietic neoplasms. *Biochim Biophys Acta* **1846**(2), 477–84 (2014).
49. Mroue, R., Huang, B., Braunstein, S., Firestone, A. J. & Nakamura, J. L. Monoallelic loss of the imprinted gene Grb10 promotes tumor formation in irradiated Nf1 +/- mice. *PLoS Genet.* **11**(5), e1005235 (2015).
50. Yan, X. *et al.* Deletion of the Imprinted Gene Grb10 Promotes Hematopoietic Stem Cell Self-Renewal and Regeneration. *Cell Rep.* **17**(6), 1584–1594 (2016).
51. Tsunoda, T. & Shirasawa, S. Roles of ZFAT in haematopoiesis, angiogenesis and cancer development. *Anticancer Res.* **33**(7), 2833–7 (2013).
52. Marconato, L. *et al.* Randomized, placebo-controlled, double-blinded chemioimmunotherapy clinical trial in a pet dog model of diffuse large B-cell lymphoma. *Clin Cancer Res.* **20**(3), 668–77 (2014).
53. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J Stat Softw.* **25**, 1–18 (2008).

## Acknowledgements

PRAT 2014, Università degli Studi di Padova.

### Author Contributions

Conception and design: L.A., S.F. Acquisition of data: A.A., S.F., L.M., M.G., D.S., L.C. Analysis and interpretation of data (e.g., raw data preprocessing, statistical analysis, computational analysis): T.S., L.A., S.F., B.D., S.M., D.G., A.A., L.M., S.C., V.M., S.D., M.M. Manuscript preparation: L.A., S.F., T.S., A.A., L.M. Study supervision: L.A., S.F.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-11724-w](https://doi.org/10.1038/s41598-017-11724-w)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

## SUPPLEMENTAL INFORMATION

### **DNA methylation profiling reveals common signatures of tumorigenesis and defines epigenetic prognostic subtypes of canine Diffuse Large B-cell Lymphoma**

Serena Ferraresso, Arianna Aricò, Tiziana Sanavia, Silvia Da Ros, Massimo Milan, Luciano Cascione, Stefano Comazzi, Valeria Martini, Mery Giantin, Barbara Di Camillo, Sandro Mazzariol, Diana Giannuzzi, Laura Marconato, Luca Aresu

#### **Table of Contents**

<b>Supplemental Materials and Methods</b>	<b>2</b>
cDLBCL cohort	2
DNA extraction and sonication	3
Enrichment of methylated double-stranded DNA	3
Sample labeling and hybridization	3
Data Quality Control and Preprocessing	3
Details on statistical analyses	4
Annotation and functional analysis	5
Bisulfite conversion	6
Methylation specific PCR (MSP)	6
Gene expression analysis of CLBL1 cells treated with hypomethylating agents	7
<b>Supplemental Results</b>	<b>9</b>
Data preprocessing	9
Microarray data technical validation	9
Microarray data functional validation	11
Supplemental References	12
Supplemental Figures	13
Supplemental Tables	15

## SUPPLEMENTAL MATERIAL AND METHODS

### cDLBCL cohort

Clinical features of dogs affected by DLBCL are reported in Table S1.

**Table S1. Clinical data of DLBCL dogs included in the study**

Dog Number	Age (y)	Sex	Stage	Sub-stage	Extranodal site infiltration	Treatment	Pre-Pred	Relapse pre/post end of therapy
1*	13	M	5	b	BM 13,5%	CH+VAX	yes	post
2	5	M	5	b	BM 12,7%, lung	CH+VAX	no	post
3	8	F	5	b	lung	CH	yes	pre
4	12	F	3	a	no	CH+VAX	no	never
5	5	M	3	a	no	CH+VAX	yes	post
6	3	M	5	a	BM 4,3%	CH	no	pre
7	7	M	4	a	no	CH	yes	pre
8	8	F	4	a	no	CH+VAX	no	never
9	3	M	5	a	PB 10%	CH	no	pre
10*	13	F	5	b	BM 7,2%	CH	no	pre
11	8	F	5	a	BM 3,1%	CH+VAX	no	post
12	6	M	5	b	BM 47,4%	CH	yes	pre
13	10	M	3	a	no	CH	yes	post
14	6	M	5	a	skin	CH+VAX	no	post
15	9	F	5	b	lung	CH+VAX	no	pre
16	5	M	4	a	no	CH+VAX	yes	post
17	10	F	4	a	no	CH+VAX	no	pre
18	5	F	5	b	BM 5,3%	CH+VAX	yes	pre
19	10	M	4	b	no	CH+VAX	no	pre
20	9	F	5	a	BM 55,2%	CH	no	pre
21	13	F	4	a	no	CH+VAX	yes	pre
22	5	M	4	a	no	CH+VAX	no	post
23	10	F	4	a	no	CH+VAX	no	never
24	5	M	5	a	BM 4,7%	CH	no	pre
25	6	F	5	a	BM 5,7%	CH+VAX	no	post
26	10	F	4	a	no	CH+VAX	no	post
27	5	M	5	a	BM 6,6%	CH+VAX	no	post
28	10	F	5	a	BM 14,9%	CH	yes	pre
29	11	F	4	b	no	CH+VAX	yes	post
30	10	M	5	b	lung	CH	yes	pre
31	6	M	4	b	no	CH	yes	pre
32	4	F	4	b	no	CH+VAX	yes	post
33	12	F	4	a	no	CH	no	pre
34	4	F	5	a	BM 11,2%	CH+VAX	no	pre
35	8	F	5	a	BM 5,1%	CH+VAX	no	pre
36	4	F	5	a	BM 5%	CH+VAX	no	never
37	5	F	4	a	no	CH	no	never
38	5	F	4	a	no	CH	yes	post
39	10	M	4	a	no	CH	no	pre
40*	10	M	5	a	BM 5%	CH	yes	pre

**M**=male, **F**= female, **Pre-pred**= steroid administration before diagnosis, **BM**=Bone Marrow Cells; **PB**=Peripheral Blood cells; **CH**=chemio; **VAX**=vaccine. (\*): Sample excluded from statistical analyses

### **DNA extraction and sonication**

Genomic DNA was extracted from lymph nodes using the DNeasy Blood & Tissue Kit (QIAGEN, CA, USA) according to the manufacturer's instructions. DNA concentration and quality were measured by Qubit fluorometer (Life Technologies, CA, USA) and by Agarose gel electrophoresis. An amount of 5 µg of extracted genomic DNA was fragmented by sonication, using Covaris S2 (Covaris, MA, USA), to obtain a fragmented DNA that ranges from 200 to 700 bp in size, diluted in 130 µl. Covaris settings were the following: n° cycles= 2, cycle duration= 60 sec, duty cycle= 10%, cycle/burst= 200 and intensity= 5, with a bath temperature of 5±1°C. An aliquot of 20 µl (total 700-800 ng) was used as reference DNA, not enriched of methylated double-stranded DNA.

### **Enrichment of methylated double-stranded DNA**

The remaining aliquot (110 µl, ~4 µg) of fragmented DNA was enriched of methylated double-stranded DNA by using MethylMiner™ Methylated DNA Enrichment Kit (Life Technologies), following the manufacturer's recommendations. MethylMiner™ uses a biotinylated recombinant fragment of the human MBD2 protein to enrich for fragments of methylated DNA. The methylated fraction of genomic DNA thus obtained was employed for methylation analysis through canine methylation microarray.

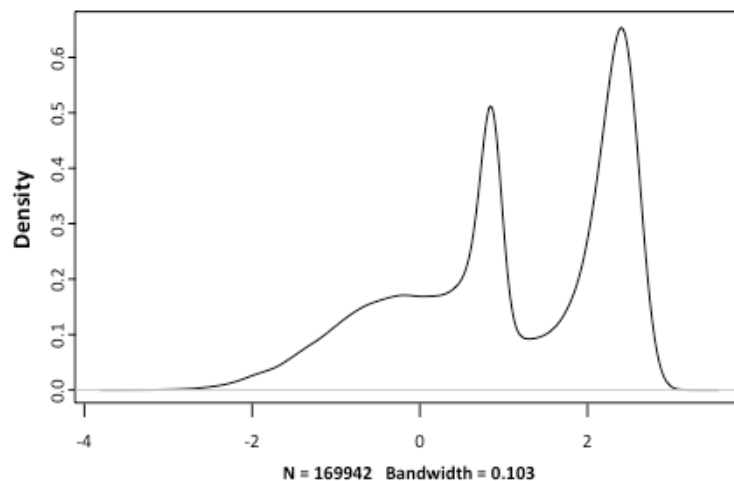
### **Sample labeling and hybridization**

Enriched-fraction and total gDNA (reference) obtained from 48 lymph node samples (40 cDLBCLs and 8 control dogs) were labelled independently with cyanine 5-deoxyuridine triphosphate (dUTP) and cyanine 3-dUTP, respectively. Sample labeling was performed by using SureTag Complete DNA Labeling Kit (Agilent Technologies) following the manufacturer's recommendations. For each sample, equal amount of enriched (Cy5-labeled) and reference (Cy3-labeled) DNAs were co-hybridized to the microarray platform. Arrays were scanned at 3µm resolution using an Agilent G2565CA scanner, and image data were processed using Feature Extraction version 10.7 with CGH-1200-Jun14 protocol (Agilent Technologies).

### **Data Quality Control and Preprocessing**

A total of 58 probes exhibiting signal saturation and one cDLBCL sample not valid according to Methylation Microarray QC metrics (Agilent Technologies) were filtered out. The MedianSignal of the probes was considered for further preprocessing. The ProcessedSignal provided by Agilent Feature Extraction algorithm was not employed since it was characterized by a higher overall variability. Differences on signal variability were observed between Cy3 and Cy5 signals, probably due to the capture/enrichment step performed on DNA of Cy5-labeled samples. To adjust the Cy3/Cy5 dye bias, Loess normalization was applied to each dye using the information between-

array to remove intensity-dependent trends, but without scaling the overall median signal towards zero. Specifically, the Loess curve was estimated keeping the within-array median value calculated across the probes. After dye bias correction, quality assessment of the resulting log<sub>2</sub>-signal ratios was performed using the arrayQualityMetrics package in Bioconductor (<http://www.bioconductor.org>). Principal Component Analysis (PCA) was also considered to evaluate anomalies among the samples. Three samples (i.e. Dog#1, Dog#10 and one control dog) failing quality controls on MA plots, box-plots and between-array distances were excluded from the analysis. On the remaining arrays, between-samples Quantile normalization was applied to the corresponding log<sub>2</sub>-signal ratios. The distribution of the median log<sub>2</sub>-ratios calculated across samples was characterized by two clear peaks (Figure S1): the first one (log<sub>2</sub>-ratios > 2) represented hyper-methylated probes, while the second one (0 < log<sub>2</sub>-ratios < 1) represented probes showing a methylation level similar to the reference. The latter was not centered to zero since the MedianSignal included the background noise that may be different between the dyes. Therefore, in order to subtract the background noise and having this peak centered to zero, the overall signal was finally scaled by a factor equal to 0.85. To estimate the peaks, we used the function findPeaks of R package quantmod.



**Figure S1. Density plot of the median array calculated from the normalized data.**

### **Details on statistical analyses**

F-test: Since cDLBCL is a heterogeneous disease, driven by perturbations of different molecular pathways, and varying from individual to individual, epigenetic instability or the loss of epigenetic control of important genomic domains can lead to increased methylation variability, not always associated to a difference of methylation levels. Recently, it has been found that differential variability between normal and cancer tissues can be very useful for identifying methylation markers of cancer (Hansen et al., 2011). Therefore, for the clinical factors differential variability

was tested using the F-test, one of the most popular approaches for testing the equality of variances.

Multivariate linear regression: Linear combinations of clinical/pathological factors significantly associated to methylation level were investigated through a multivariate linear regression model. Specifically, the factors of the final model were selected with a step-down procedure: all the factors were initially included in the full model considering main effects only, then they were sequentially removed if their removal did not result in a significant change in fitting the data, using F-test.

Analysis of methylation disruption: Starting from methylation data in cDLBCL samples and control lymph nodes a matrix X was defined describing the methylation changes for each sequence  $j$  and each cDLBCL sample  $i$  as  $x_{ij} = y_{ij} - z_j$ , which is the methylation difference between the sample  $i$  and the median methylation  $z_j$  calculated across the 7 control lymph nodes at sequence  $j$ . PCA analysis was performed on  $x_{ij}$  values as preliminary analysis of the variability in methylation changes. The methylation variability profile for each cDLBCL sample  $i$  (MVP) was then defined as the density function  $f_i(x)$  across all the regions represented on the array. The function was estimated using the density() function in R with bandwidth parameter 0.1<sup>33</sup>.

To define a distance matrix for the clustering, the squared L2-distance between the MVP density functions were calculated for all pairs of patient samples. This distance represents the squared difference in the area under the curve between two samples and is approximated using the Trapezoidal rule (Supplementary Material in Chambwe et al.<sup>33</sup>).

Consensus clustering was then performed on this matrix applying Ward's linkage, using R package ConsensusClusterPlus (Wilkerson et al., 2010). Specifically, HCL was performed 1,000 times on resampled subsets of the cDLBCL samples (using 80% of samples as subset) and evaluated the number of clusters  $k=2,3,\dots,15$ . We note that the relative change in area under the cumulative distribution functions of the consensus matrix (described in Methods) for each  $k$  is maximum at 3, indicating the best separation of the clusters.

Finally, to provide also a quantitative measure of the magnitude of methylation disruption observed in each sample, Methylation Variability Score of cDLBCL sample  $i$  was defined as the deviation of each cDLBCL MVP describe by  $f_i(x)$  to that of the expected MVP of a control lymph node, described by the mean density function  $\bar{g}(x)$ <sup>33</sup>:

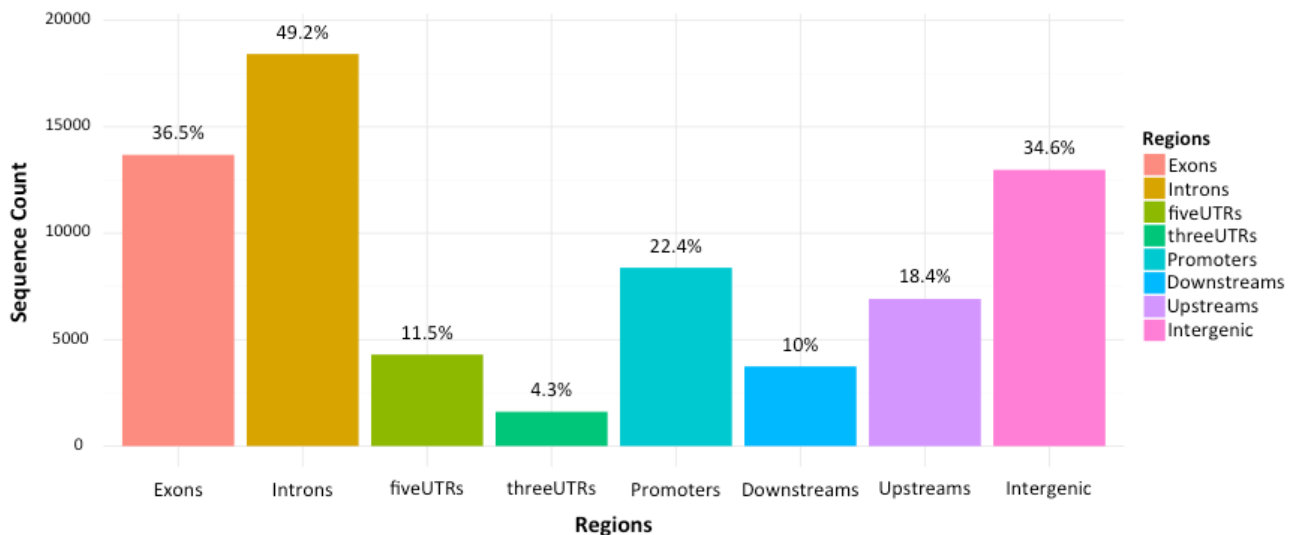
$$MVS_i = \int [f_i(x) - \bar{g}(x)]^2 dx$$

### **Annotation and functional analysis**

In order to improve the biological interpretation of the significant sequences, CanFam3 annotations from both RefSeq and Ensembl retrieved from UCSC table browser were associated to each sequence. In particular, we first checked whether each sequence overlaps at least one of the



following genomic locations: 5'-UTR, 3'-UTR, exonic, intronic, promoter/upstream (i.e. 2k/10k bases upstream from transcription start site, respectively), downstream (i.e. 2k bases from transcription start site). If the sequence did not overlap any of these locations (i.e. it is an intergenic region), the nearest gene was associated, assuming possible distal regulatory effects on the associated gene.



**Figure S2. Distribution of target sequences (CpG and CDS) across the dataset.** Percentages with respect to the total number of sequences in the chip are reported. The sum of these percentages is not equal to 100% since each sequence can overlap more than one genomic region.

In order to identify enriched genomic locations with respect to the selection of the differentially methylated sequences, Fisher's Exact test was performed on the number of the selected sequences with respect to the total number of sequences available in the microarray platform.

Finally, the biological terms from Gene Ontology (GO) and KEGG pathways able to significantly characterize the selected sequences were identified by performing an enrichment analysis. topGO R package with default options (Alexa et al., 2006) and Fisher's Exact test were applied on GO terms and KEGG pathways respectively, considering as significant the terms/pathways with adjusted Bonferroni p-value <0.05. Functional annotations were retrieved from R packages org.Cf.eg.db, GO.db and KEGG.db. Focusing on the selected sequences belonging to the enriched terms, the corresponding protein-protein interactions (PPIs) derived from STRING database were considered for further interpretation of the obtained results (Szklarczyk et al., 2015).

Gene set enrichment analysis (GSEA) was performed on the entire dataset using the Gene Set Enrichment Analysis v2.0.13 software (Subramanian et al. 2005) downloaded from the Broad Institute ([www.broadinstitute.org/gsea](http://www.broadinstitute.org/gsea)). GSEA analysis was performed by using gene symbols retrieved by blastx against UniProt database. For enrichment analysis Gene sets were downloaded

from the C2-CP C4-CM and C6 collections in MsigDB v3.1 (Molecular Signature Database). In addition, more specific lymphoid gene sets were retrieved from Staudt's SignatureDB (<https://lymphochip.nih.gov/signaturedb/>, Shaffer et al. 2006). Pathway Enrichment analysis was performed on each collection independently, T-test metric was employed for gene ranking, and 1,000 permutations were applied for p-value assignment.

### **Bisulfite conversion**

Genomic DNA from 13 cDLBCLs (Dog number:1, 6, 14, 15, 18, 19, 20, 24, 30, 32, 33, 34, 38) and five lymph node samples (Ctrl#1, Ctrl#2, Ctr#4, Ctr#7, Ctrl#8) was quantified using NanoDrop1000 Spectrophotometer (Thermo Scientific). For each sample, 500 ng of genomic DNA were bisulfite treated using the MethylCode™ Bisulfite Conversion Kit (Invitrogen™, Carlsbad, California) following manufacturer's specifications. Bisulfite-converted DNA was then employed as template for MSP.

### **Methylation specific PCR (MSP)**

A technical validation of microarray platform by methylation-specific PCR (Hernández et al. 2013) was performed on 5 differentially methylated genes (FGFR2, HOXD10, RASAL3, CYP1B1 and ITIH5). On the CpG islands of these genes (Table S2), Methylation-specific primers were designed by means of Methyl Primer Express software (Applied Biosystems, Foster City, CA). For each gene, two primer sets were designed: i) METH primers designed to amplify the DNA if methylated (scenario in which the cytosines in CpG dinucleotides are methylated and are not be bisulfite converted into uracil); ii) NO-METH primers designed to amplify the same DNA if not methylated (scenario in which all cytosines are supposed to be bisulfite converted into uracil).

All the MSPs were carried out using 5 ng of bisulfite-converted gDNA; 600-600 nM primer pair was used for all genes apart from CYP1B1 (300-600nM METH, 600-300nM NO-METH), ITIH5 METH (300-600nM), RASAL3 METH and NO-METH (50-50 nM). Real time amplification was carried out using the Master Mix SYBR® Green PCR Master Mix Applied Biosystems and Stratagene Mx3000P Agilent Technologies. For CYP1B1 and ITIH5 genes the Ct values were acquired at 76 and 74 °C, respectively, to eliminate primer dimer contribution to the amplification plot. Negative controls (with no bisulfite-converted gDNA or water as template) were run in every plate for each assay. The quantification of methylation level for each target gene was carried out by calculating the ratio of methylated to unmethylated primers pairs as  $\Delta CT (=CT_{Meth} - CT_{NoMeth})$  as described by Zeschnigk et al. (2004).

Statistical analyses were performed using a commercially available statistical software program (SPSS v20.0). Data were analysed using a non-parametric statistical method because of the limited

number of cases. Sample methylation levels, were evaluated for significant differences between controls and cDLBCLs using the Mann-Whitney test.

**Table S2. Primer pairs employed for MSP**

Gene		Methylation primer 5'-3'	No Methylation primer 5'-3'
<b>FGFR2</b>	Forward	GTTATACGGGGGCGTTGAC	TGTGGTTATATGGGGGTGTTGAT
	Reverse	GCGAAAACCAAATACCGAATACG	ACTCCTTCACAAAACCAAATACCA
<b>HOXD10</b>	Forward	GGTCGGTTGTTTGTAGCGC	GTTGGGTTGTTTGTAGTGT
	Reverse	CTCGAAATCACGTACTCCG	CCTCCTCACAAATCACATACTCCA
<b>ITIH5</b>	Forward	AGAATTTTCGGGGATGCGGATC	TGTAGAATTTTGGGGATGTGGATT
	Reverse	CAACTATCCACGACGTCCTCG	AAACAACCTATCCACAACATCCTCA
<b>RASAL3</b>	Forward	CGTTGGAGTTCGCGTTGTTT	GGGTGTTGGAGTTTGTGTTGTTT
	Reverse	CACCCTACTCCCCGAAACG	ACCAACCTCTAATCACTCAAATCCA
<b>CYP1B1</b>	Forward	GGTTAGAGGTCGGTAGGTTGC	GTGGTTAGAGGTTGGTAGGTTGT
	Reverse	AAACGCTACTCTACGCTCCG	AAATTCCCACACACCTATCAAAAACA

### Gene expression analysis of CLBL1 cells treated with hypomethylating agents

A functional validation of microarray data was performed evaluating the mRNA expression restoration of 3 hypermethylated genes after the treatment of a canine B-cell lymphoma cell line (CLBL1: Rütgen et al., 2010), with hypomethylating agents. To this purpose azacytidine (AZA) and decitabine (DEC) were used. Among the hypermethylated genes, CADM1, CDH11 and ABCB1 were selected.

The CLBL1 cell line was maintained in T25 or T75 flasks as previously reported (Rütgen et al., 2010). Cells were seeded at a concentration of  $3 \times 10^5$  cells/well in a 6-well flat bottom plate (Sarstedt Italia, Verona, Italy) and incubated for 72 h with AZA and DEC (Sigma-Aldrich, Milan, Italy) at the final concentration corresponding to their IC<sub>50</sub> values (3.42 and 0.13  $\mu$ M, respectively), determined by Alamar Blue test (Promega, Madison, USA). Due to its chemical instability, AZA dilution was freshly prepared every 24 h and added onto each well. Four independent experiments were performed.

At the end of the treatment, cells were collected and washed with PBS. Then, total RNA was extracted using the RNeasy® Mini Kit (Qiagen®, Hilden, Germany) and quantified with NanoDrop 1000 Spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA). One  $\mu$ g of total RNA was reverse-transcribed using the High Capacity cDNA Reverse Transcription kit (Life Technologies, Carlsbad, California, United States), according to the manufacturer's instructions.

For each target transcript, gene-specific primers encompassing one intron were designed (see Table S3). Two internal control genes (ICGs: GOLGA1 and CCZ1) previously published in Giantin et al. (2013) and Giantin et al. (2016) were selected.

The qPCR reaction was performed in a final volume of 10  $\mu$ L, using 12.5 ng of cDNA, the Power SYBR Green PCR Master Mix (Life Technologies, Carlsbad, California, United States) and a

Stratagene Mx3000P thermal cycler (Agilent Technologies, Santa Clara, California, United States). Standard qPCR conditions were used, except for the analysis of CADM1 and CDH11, for which Ct values were acquired at 78°C to eliminate primer dimers contribution to the amplification plot. Different concentrations of forward (F) and reverse (R) primers were tested. The presence of specific amplification products was confirmed by dissociation curve analysis. For each qPCR assay, negative controls (with total RNA or water as template) and positive controls (the cDNA of 6 canine control lymph nodes) were run. Standard curves were obtained using the best performing primer concentration and serial dilutions of control lymph node cDNA. Each dilution was amplified in duplicate. The  $\Delta\Delta C_t$  method (Livak and Schmittgen, 2001) was used for the analysis of gene expression results.

Statistical analysis was performed using GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego, USA). Data were analysed using unpaired t-test. A *P* value < of 0.05 or less was considered as statistically significant.

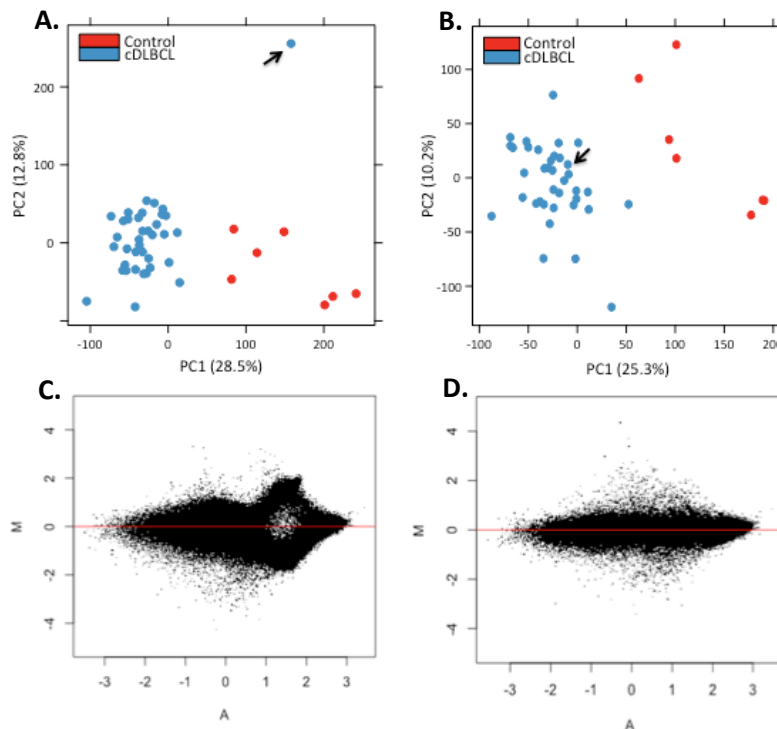
**Table S3. Primer pairs used for gene expression analysis (qPCR) in CLBL1**

<b>Gene</b>		<b>Primer sequence 5'-3'</b>	<b>Primer concentration (nM)</b>
<b>CADM1</b>	Forward	GGTGAGGAGATTGAAGTGAAC TG	50
	Reverse	TCCTCCACCTCCGATTG C	300
<b>CDH11</b>	Forward	CATTAACGACAACCCTCCTGAG	300
	Reverse	CTGGATGACCGACGTTCCC	50
<b>ABC B1</b>	Forward	GACGTTGGGGAGCTTAACAC	300
	Reverse	CGCCAATTCCTTCATTGATT	600

## SUPPLEMENTAL RESULTS

### Data preprocessing

The combining Loess and Quantile normalization pre-process used to normalize the methylation data, was able to overcome some pitfalls of the data distribution. In Fig. S3, the position of Dog#19 is shown both in the PCA and MA plots with respect to the median across the samples, highlighting the difference obtained by the two normalization approaches. In Fig S3 A. and C., only Quantile normalization is applied, whereas in Fig S4 B. and D. Loess approach is combined with Quantile. In the latter, Dog#19 clustered with the cDLBCLs group. Indeed, this sample showed the most evident dye-bias (Figure S4) generated by the fact that Cy5 signal was altered by the experimental enrichment step generating an additional bias compared to the Cy3-labeled reference. This trend was observed in more than half of the samples of the dataset. The Loess normalization step was able to determine the adjustment of this bias using the information from each dye. Furthermore, since Quantile normalization assumes a common distribution of data, the Loess-normalized data were characterized by a more similar between-array distribution compared to the raw data (Figure S5), thus allowing Quantile-normalization to have the best fit.



**Figure S3. Differences between Quantile-only and Loess-plus-Quantile normalization on methylation data.** PCA analysis on cDLBCL and control samples (upper panel) and MA plots on sample Dog#19 (lower panel), showing differences between Quantile normalization applied directly on raw data (left panel) and the same Quantile normalization applied on Loess-normalized data adjusted for the dye-bias. Arrows in the PCA plots indicate the positions of Dog#19.

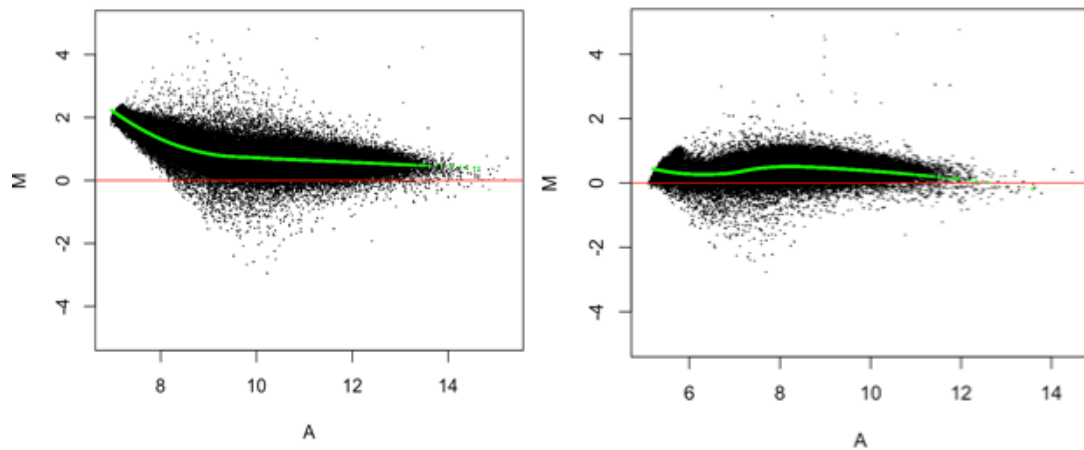


Figure S4. Between-array MA plot of cDLBCL Dog#19, separating Cy5-dye (right panel) and Cy3-dye (left panel) signals.

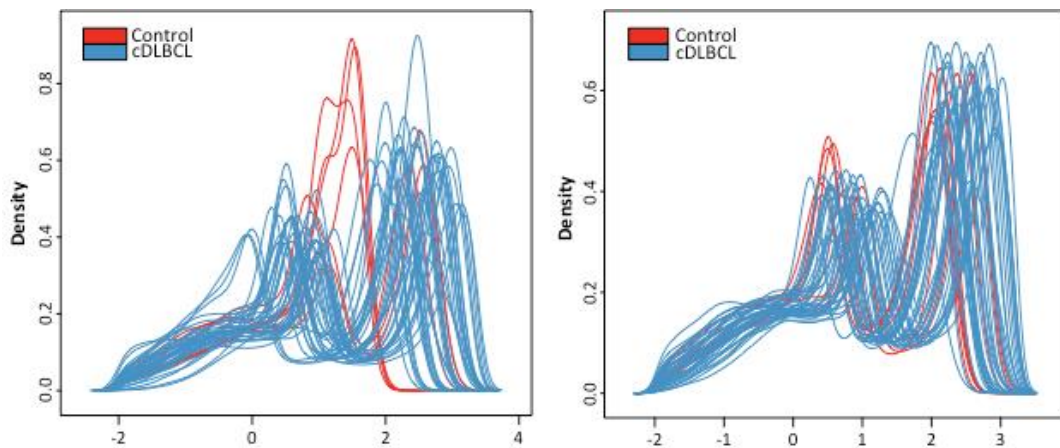


Figure S5. Density plots of the arrays on raw (left panel) and Dye-bias normalized (right panel) data.

### Microarray data technical validation

In order to quantify the ratio of methylated to unmethylated alleles, the  $\Delta CT$  ( $=CT_{Meth} - CT_{NoMeth}$ ) value was determined (Table S3) as described by Zeschnigk et al. (2004). The Mann-Whitney test comparing Meth/No-Meth primer pairs showed a significant hypermethylation between the two groups for HOXD10, RASAL3 ( $p < 0.001$ ), CYP1B1 and ITIH5 ( $p < 0.01$ ), while FGFR2 was only marginally significant ( $p = 0.07$ ).

**Table S3.  $\Delta$ CT values calculated, for each target gene, on cDLBCL and Control samples**

cDLBCL	FGFR2	HOXD10	ITIH5	RASAL3	CYP1B1
15	3.56	-1.94	-6.51	-3.99	3.95
30	2.94	-1.07	-5.45	-3.51	5.73
20	-1.48	-1.45	-5.07	-3.14	-1.02
33	5.04	-0.69	-2.77	-3.08	5.21
14	1.16	-2.61	-4	-4.17	3.5
24	5.45	-1.08	-3.27	-2.96	0.08
12	5.62	-2.65	-	-4.5	4.73
19	-0.16	0.98	0.2	-2.28	1.43
6	1.94	-0.35	-3.36	-4.07	2.84
38	7.25	0	-2.56	-6.36	1.23
32	7.12	-2.03	-2.92	-4.44	0.55
18	5.82	0.27	-3.79	-3.64	3.95
34	9.31	2.37	-3.8	-4.03	3.29
<b>Controls</b>					
Ctrl#7	4.4	3.47	-0.89	-0.34	4.9
Ctrl#4	7.13	4.4	-1.69	-1.07	5.87
Ctrl#8	8.36	4.03	-2.65	-2.64	5.73
Ctrl#2	8.6	5.57	-2.22	-2.57	6.51
Ctrl#1	6.39	3.72	-1.96	-1.49	5.56

**Microarray data functional validation**

All primer pairs for gene expression analysis had an acceptable efficiency (range 90 % ÷ 110 %), and a slope in the range of -3.6/-3.1. The main features of the validated qPCR assays are reported in Table S4.

**Table S4. Main features (slope, efficiency,  $R^2$ , dynamic range) for each qPCR assay.**

Gene	Slope	Efficiency (%)	$R^2$	Dynamic range (Ct)
CADM1	-3.36	98.4	0.99	24.48 – 34.60
CDH11	-3.47	94.1	0.99	22.75 – 36.62
ABCB1	-3.18	106.4	0.99	28.79 – 36.17

The effect of AZA and DEC treatment on CADM1, CDH11 and ABCB1 mRNA expression are summarized in Figure S6-S8, respectively.

Selected genes were all constitutively and highly expressed in the control lymph nodes, while in the B-cell lymphoma cell line (CLBL1) they were almost completely silenced. Following the treatment with AZA, the mRNA expression was significantly restored ( $P < 0.05$  or less). Conversely, DEC affected the re-expression of ABCB1 ( $P < 0.01$ ), while did not exert any effect on CADM1 and CDH11 (data not shown).

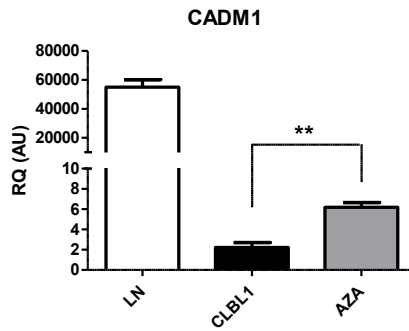


Figure S6: CADM1 mRNA expression in control lymph nodes (LN), CLBL1 cells alone and treated with azacytidine (AZA, 3.42  $\mu$ M). RQ values are expressed in arbitrary units (AU) as means  $\pm$  SEM.

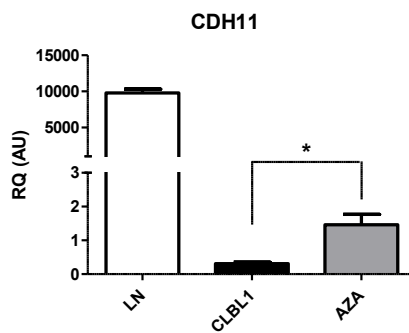


Figure S7: CDH11 mRNA expression in control lymph nodes (LN), CLBL1 cells alone and treated with azacytidine (AZA, 3.42  $\mu$ M). RQ values are expressed in arbitrary units (AU) as means  $\pm$  SEM.

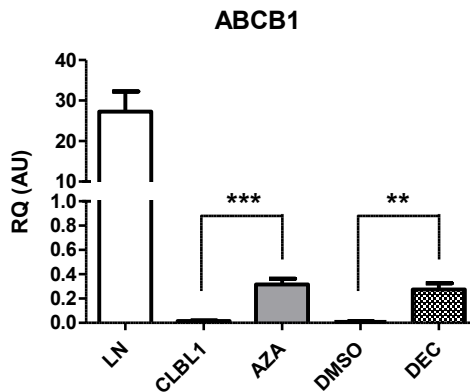


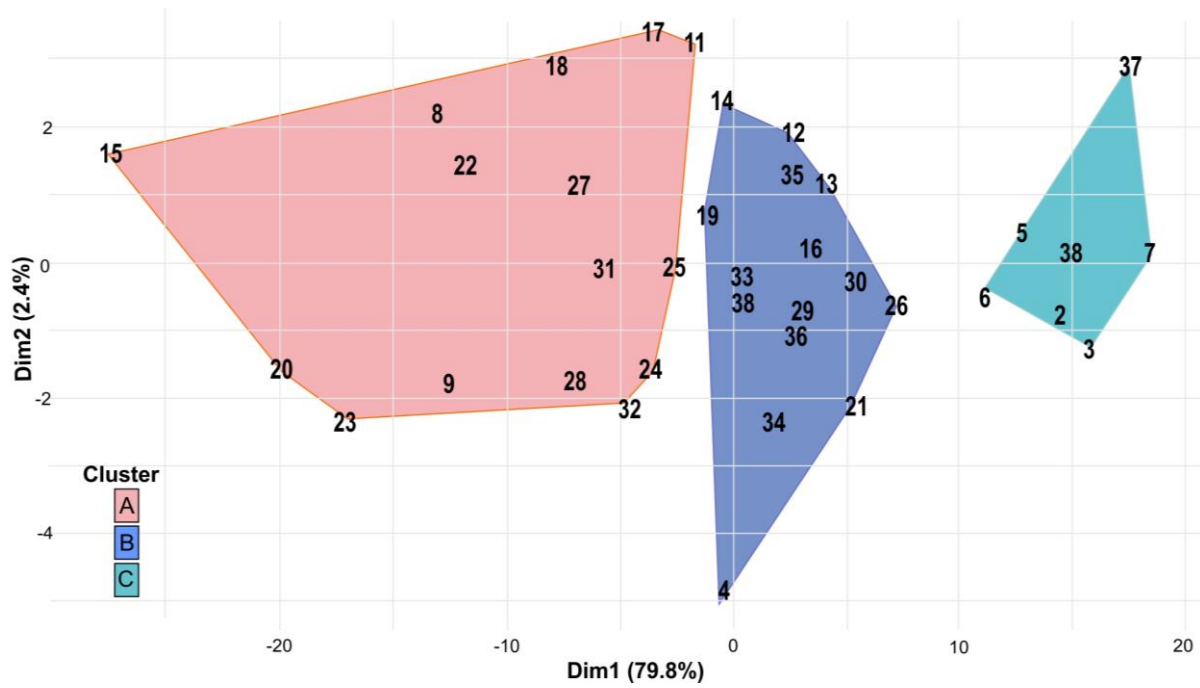
Figure S8: ABCB1 mRNA expression in control lymph nodes, CLBL1 cells alone and treated with the vehicle (DMSO 0.1%), azacytidine (AZA, 3.42  $\mu$ M) and decitabine (DEC, 0.13  $\mu$ M). RQ values are expressed in arbitrary units (AU) as means  $\pm$  SEM.



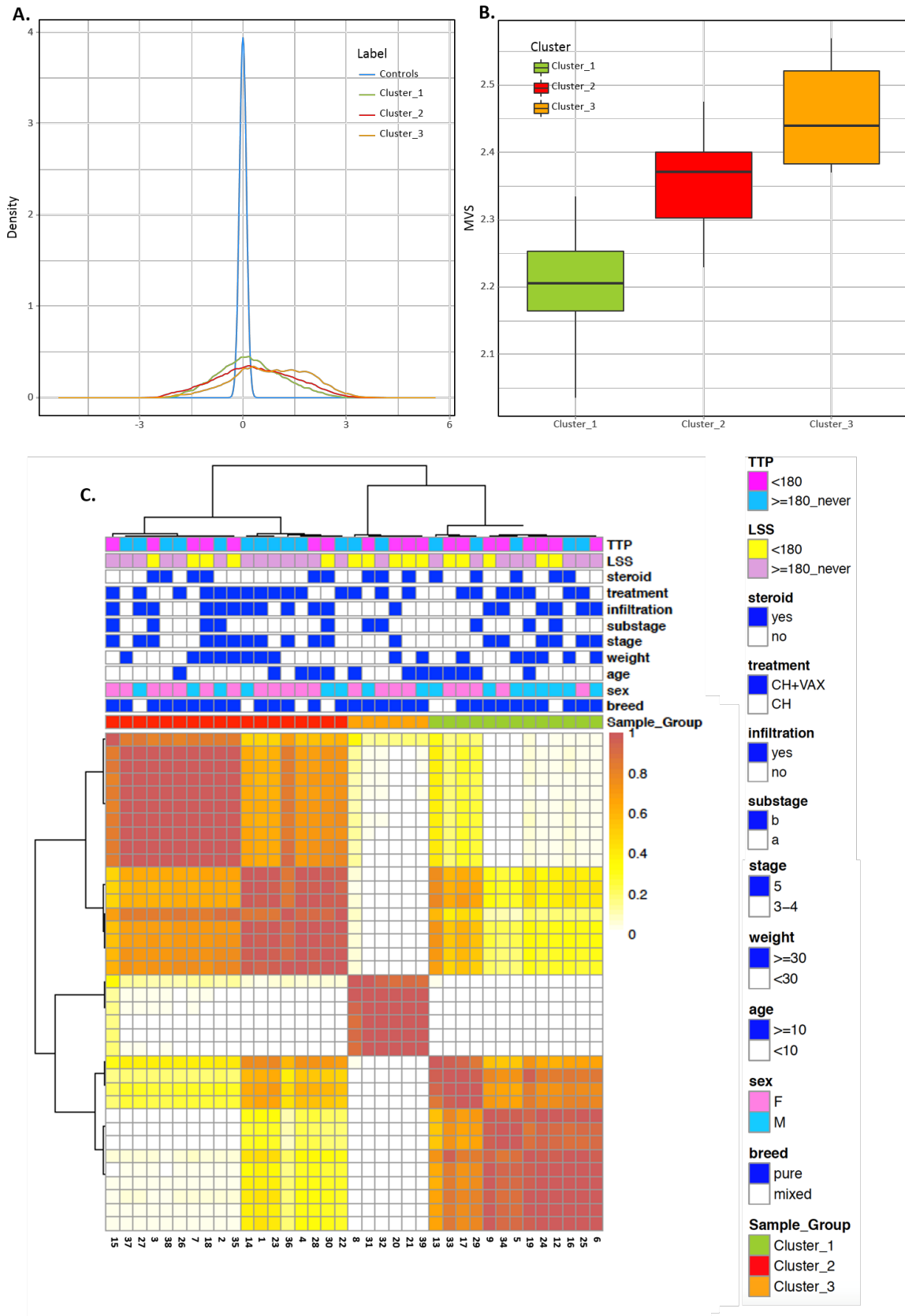
## SUPPLEMENTAL REFERENCES

- Alexa A, Rahnenfuhrer J, and Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating go graph structure. *Bioinformatics* **2006**; 22:1600–1607.
- Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* **2011**; 43: 768-775.
- Giantin M, Vascellari M, Lopparelli RM, Ariani P, Vercelli A, Morello EM *et al.* Expression of the aryl hydrocarbon receptor pathway and cyclooxygenase-2 in dog tumors. *Res Vet Sci.* **2013**; 94:90-9.
- Giantin M, Baratto C, Marconato L, Vascellari M, Mutinelli F, Dacasto M *et al.* Transcriptomic analysis identified up-regulation of a solute carrier transporter and UDP glucuronosyltransferases in dogs with aggressive cutaneous mast cell tumours. *Vet J.* **2016**; 212:36-43.
- Hernández H.G., Tse M.Y., Pang S. C., Arboleda H., Forero D.A. Optimizing methodologies for PCR-based DNA methylation analysis. *BioTechniques* **2013**; 55:181-197
- Livak KJ and Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method *Methods* **2001**; 25:402–408.
- Rütgen BC, Hammer SE, Gerner W, Christian M, de Arespachoga AG, Willmann M *et al.* Establishment and characterization of a novel canine B-cell line derived from a spontaneously occurring diffuse large cell lymphoma *Leuk Res* **2010**; 34:932–938.
- Shaffer AL, Wright G, Yang L, Powell J, Ngo V, Lamy L *et al.* A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol Rev* **2006**; 210:67-85.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **2005**; 102:15545–15550.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **2015**; 43:D447-52
- Wilkerson MD and Haynes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**; 26(12):1572-1573.
- Zeschnigk M, Böhringer S, Price EA, Onadim Z, Masshöfer L, Lohmann DR. A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. *Nucleic Acids Res* **2004**; 32(16):e125.

## SUPPLEMENTAL FIGURES



**Figure S9. PCA plot on cDLBCL samples based on MVPs.** Colors correspond to the clusters identified by the sequences highly correlated with the first principal component.



**Figure S10: Consensus hierarchical clustering on the first 2,000 sequences showing the highest median absolute deviation of the MVPs across cDLBCLs** **A.** Density function of MVPs of cDLBCL clusters compared to those of controls, **B.** Boxplot of the MVPs by cluster. **C.** Heatmap for consensus matrix (K=3)

## SUPPLEMENTAL TABLES

*Excel file Supplementary Table S2*

**Table S2.xls Differentially hyper- and hypo-methylated sequences on cDLBCL samples with respect to normal lymph nodes.** Columns H-I report the median methylation level across cDLBCL and normal samples, respectively. Columns K-R report the overlapping Refseq (K-N) and Ensembl (O-R) transcripts, considering for each transcript the following genomic locations: exon, intron, 5'-UTR, 3'-UTR, “proxUP” (i.e. until 2kb upstream from transcription start site), “upstr” (i.e. from 2kb to 10kb upstream from transcription start site), “proxDOWN” (i.e. until 2kb downstream from transcription start site). Columns S-V: nearest Refseq or Ensembl transcripts calculating the distance from the transcription start site (TSS).

*Excel file Supplementary Table S3*

**Table S3.xls** List of significant probes/genes obtained after categorical division of the Beta values in two classes. Column A reports the exact genomic location of the probe and Column B reports the gene symbol.

*Excel file Supplementary Table S4*

**Table S4.xls Significantly enriched GO terms and KEGG pathways for the differentially hyper- and hypo-methylated sequences on cDLBCL samples with respect to normal lymph nodes.**

*Excel file Supplementary Table S5*

**Table S5.xls Significantly enriched gene signatures highlighted by Gene Set Enrichment Analysis (GSEA) on the entire dataset of probes.**

*Excel file Supplementary Table S6*

**Table S6.xls Sequences showing differential methylation variability on one or more clinical factors across the cDLBCL samples.** Columns G-N report the overlapping Refseq (G-J) and Ensembl (K-N) transcripts, considering for each transcript the following genomic locations: exon, intron, 5'-UTR, 3'-UTR, “proxUP” (i.e. until 2kb upstream from transcription start site), “upstr” (i.e. from 2kb to 10kb upstream from transcription start site), “proxDOWN” (i.e. until 2kb downstream from transcription start site). Columns O-R: nearest Refseq or Ensembl transcripts calculating the distance from the transcription start site (TSS).

*Excel file Supplementary Table S7*

**Table S7.xls Results from multivariate linear regression analysis, investigating different combination of clinical factors across the cDLBCL samples.** Columns H-R: clinical factors considered for the analysis; “1” indicates the presence of that factor in the linear regression model. Columns S-Z report the overlapping Refseq (S-V) and Ensembl (W-Z) transcripts, considering for each transcript the following genomic locations: exon, intron, 5’-UTR, 3’-UTR, “proxUP” (i.e. until 2kb upstream from transcription start site), “upstr” (i.e. from 2kb to 10kb upstream from transcription start site), “proxDOWN” (i.e. until 2kb downstream from transcription start site). Columns AA-AD: nearest Refseq or Ensembl transcripts calculating the distance from the transcription start site (TSS).