The TiHoCL panel for canine lymphoma: a feasibility study integrating functional genomics and network biology approaches for comparative oncology targeted NGS panel design

Silvia Fibi-Smetana[1], Camila Inglis[2,3], Daniela Schuster[4,5], Nina Eberle[2], José Luis Granados-Soler[2,6], Wen Liu[3], Saskia Krohn[3], Christian Junghanss[3], Ingo Nolte[2], Leila Taher[1,3,4,5]* and Hugo Murua Escobar[3]*

[1]Institute of Biomedical Informatics, Graz University of Technology, Graz, Austria, [2]Small Animal Clinic, University of Veterinary Medicine Hannover Foundation, Hannover, Germany, [3]Clinic for Hematology, Oncology and Palliative Care, Rostock University Medical Center, University of Rostock, Rostock, Germany, [4]Division of Bioinformatics, Department of Biology, Friedrich-Alexander-University, Erlangen, Germany, [5]Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center, University of Rostock, Rostock, Germany, [6]UQVETS Small Animal Hospital, School of Veterinary Science, The University of Queensland, Gatton, QLD, Australia

Targeted next-generation sequencing (NGS) enables the identification of genomic variants in cancer patients with high sensitivity at relatively low costs, and has thus opened the era to personalized human oncology. Veterinary medicine tends to adopt new technologies at a slower pace compared to human medicine due to lower funding, nonetheless it embraces technological advancements over time. Hence, it is reasonable to assume that targeted NGS will be incorporated into routine veterinary practice in the foreseeable future. Many animal diseases have well-researched human counterparts and hence, insights gained from the latter might, in principle, be harnessed to elucidate the former. Here, we present the TiHoCL targeted NGS panel as a proof of concept, exemplifying how functional genomics and network approaches can be effectively used to leverage the wealth of information available for human diseases in the development of targeted sequencing panels for veterinary medicine. Specifically, the TiHoCL targeted NGS panel is a molecular tool for characterizing and stratifying canine lymphoma (CL) patients designed based on human non-Hodgkin lymphoma (NHL) research outputs. While various single nucleotide polymorphisms (SNPs) have been associated with high risk of developing NHL, poor prognosis and resistance to treatment in NHL patients, little is known about the genetics of CL. Thus, the ~100 SNPs featured in the TiHoCL targeted NGS panel were selected using functional genomics and network approaches following a literature and database search that shielded ~500 SNPs associated with, in nearly all cases, human hematologic malignancies. The TiHoCL targeted NGS panel underwent technical validation and preliminary functional assessment by sequencing DNA samples isolated from blood of 29 lymphoma dogs using an Ion Torrent™ PGM System achieving good sequencing run metrics. Our design framework holds new possibilities for the design of similar molecular tools applied to other diseases for which limited knowledge is available and will improve drug target discovery and patient care.

1 Introduction

Driven by plummeting costs, next-generation sequencing (NGS) has revolutionized biomedical research, playing an instrumental role in advancing our understanding of the molecular basis of various diseases (1). NGS has numerous applications, ranging from whole-genome (re) sequencing to targeted sequencing for variant identification or confirmation. In contrast to whole-genome sequencing, targeted NGS focuses on a specific set of genomic loci that are likely to be involved in the phenotype of interest, delivering higher coverage levels at a more affordable cost, and making it amenable to samples containing small DNA amounts. Furthermore, targeted NGS produces substantially smaller datasets, which are easier to manage and analyze (2). These attributes make targeted NGS particularly well suited to detect and characterize specific tumor cell sub-populations, such as subclones harboring drug resistant variants (3), and hence, very attractive for clinical oncology (4, 5). Indeed, in human medicine, targeted NGS has become a common tool to diagnose and monitor cancer, as well as to select therapeutic agents and quantify treatment resistance.

Due to various factors, including financial considerations and the diversity of animal species and breeds, veterinary medicine tends to adopt technological advances more slowly than human medicine. Nonetheless, there are a number of cases in which the use of targeted NGS has already proved to be cost-effective. For example, targeted NGS panels are routinely used in the clinic to detect ovine and equine pathogens (6, 7). And given the growing significance of the well-being of companion animals in our society, it is inevitable that cutting-edge technologies such as targeted NGS will be progressively adopted to improve their health and quality of life.

Targeted NGS requires a certain knowledge of the genetic basis of the disease of interest, in particular, of the variants or mutations that are associated with it. This can pose challenges for diseases that lack extensive research, as it is generally the case for animal diseases. However, biological processes and genetic mechanisms are frequently conserved across species, making findings in human medicine relevant to veterinary medicine –and vice versa–, and facilitating research and technology transfer between the two fields (8). Consistently, veterinary medicine often resorts to comparative approaches to gain insights into health and disease. Herein, we illustrate how knowledge can be leveraged across species and/or diseases to develop a targeted sequencing panel for canine lymphoma (CL).

CL is a spontaneous disease that closely resembles human Non-Hodgkin lymphoma (NHL), a heterogeneous group of lymphoid malignancies, with different cells of origin and biological behaviors. Specifically, CL and NHL present similar clinical, histological, cytogenetic, and molecular features (9, 10). Furthermore, CL and NHL are classified according to analogous histologic systems and treated with the same chemotherapeutic agents. In particular, the standard backbone treatment for CL and NHL is a chemotherapy combination known as CHOP [cyclophosphamide, hydroxydaunorubicin, oncovin/

vincristine, prednisone or prednisolone; (11, 12)]. Groundbreaking CD20 antibody strategies, as used in humans, are unfortunately not available in dogs, thus conventional chemotherapy remains the key strategy for CL. Naturally, although CL and NHL are similar in many regards, they also exhibit important differences. Most notably, while NHL patients often respond well to treatment and over 50% of them are cured after initial therapy (9, 10, 13–17), CL is considered incurable. Indeed, even though over 80% of dogs achieve complete remission upon induction chemotherapy, most of them relapse within 12 months (9, 10, 13–15). Moreover, relapsed patients frequently become refractory to therapy (18) and ultimately die of their disease, with a median overall survival time of only 10–14 months (13).

As in NHL and other canine cancers, increased risk, poor prognosis, and resistance to treatment in CL are very likely related to the genetic background of the host (19–26). Candidate gene, linkage and genome-wide association studies (GWASs) have revealed multiple genetic variants associated with NHL (25, 27–30). Specifically, genetic alterations associated with NHL include both single nucleotide variants and large rearrangements that place genes under the control of promoters and enhancers that are typically only active in the lymphocytes, leading to dysregulation of genes in pathways involving immune function, cell cycle, apoptosis, DNA repair, and carcinogen metabolism (31–33). Although genetic studies are still incipient for CL, evidence suggests that many of the same pathways dysregulated in NHL also exhibit perturbations in CL (31, 32). Nevertheless, how to use this information for patient stratification remains a matter of investigation (34).

Here we present the TiHoCL targeted sequencing panel, which comprises approximately 100 canine loci chosen using a combination of comparative genomics and network-based approaches. Our panel was designed starting from ~600 single nucleotide polymorphisms (SNPs) associated with lymphoma risk and prognosis in humans and only three in dogs, and was successfully technically and functionally assessed on 29 DNA samples extracted from whole blood of CL patients using the Ion Torrent™ Personal Genome Machine (PGM)™ technology. The TiHoCL targeted sequencing panel is a proof of principle to demonstrate how we can use knowledge from human research to develop clinical tools for other species. Facilitated by declining sequencing costs, the broad application of NGS panels like the TiHoCL targeted panel opens the door to the discovery of new drug targets and improved patient care.

2 Materials and methods

2.1 Identification of lymphoma-associated canine and human SNPs

First, we mined the Online Mendelian Inheritance in Man (OMIM) (35), the GWAS catalog (36), and PubMed for single

nucleotide polymorphisms (SNPs). We accessed OMIM through their API; text searches for "lymphoma" were restricted to those containing allelic variants using the search field "av_exists." The GWAS catalog was queried using the web interface; we searched for SNPs using the keyword "lymphoma" and required a p-value for the SNP-disease/trait association $<1\times10^{-5}$. Finally, we searched PubMed using the Entrez Programming Utilities (E-Utilities) (37) with the PubMed filters: "lymphoma[Title/Abstract] AND ("allelic variation" OR polymorphism) AND humans[MeSH]" (for human SNPs), and "lymphoma[Title/Abstract] AND ("allelic variation" OR polymorphism) AND dogs[Title/Abstract]" (for canine SNPs). The abstracts resulting from this query were searched for Reference SNP (rs) numbers. Next, we manually searched PubMed for articles published between the years of 2005 and 2015 with the keywords "lymphoma[MeSH]" and "genetic polymorphism[MeSH]" or "single nucleotide polymorphism[MeSH]." Any human or canine SNPs mentioned in the full text of these articles that were significantly associated (p-value$<10^{-5}$) with hematological malignancy risk, prognosis, outcome and resistance to treatment were selected as relevant.

We merged the SNPs resulting from all searches, removing duplicates and excluding those SNPs absent in dbSNP build 147 (38).

2.2 Identification of canine orthologous loci for human SNPs

To identify the canine orthologous locus of each human SNP we first retrieved the orthologous sequences of a 100-bp window centered at the human (GRCh38/hg38) SNP in the dog (CanFam3.1/canFam3), gorilla (gorGor4.1/gorGor4), mouse (GRCm38/mm10), and cat (ICGSCFelis_catus_8.0/felCat8) genomes with the LiftOver tool from the UCSC Genome Browser (39). SNPs in genomic windows for which no orthologous canine sequence was retrieved or for which the orthologous canine sequence exceeded 1,000-bp were excluded from further analyses. Finally, multiple sequence alignments were computed with T-Coffee v11.00.8cbe486 (40) to assess the evolutionary conservation of each SNP and flanking nucleotides. SNPs aligned to a gap in the canine genome were excluded from further analyses.

2.3 Genomic annotation of canine loci

Canine loci were annotated with the "annotatePeaks.pl" command from the HOMER v4.11.1 suite (41) using canFam3 as genome.

2.4 Gene annotation

Gene annotation was obtained from Ensembl [release 90, (42)] in GTF format.[1]

2.5 Identification of canine loci overlapping with UTRs

Gene annotation was filtered for features covering 50 bp or less. Bedtools intersect v2.27.1 (43) was used to identify overlaps with the remaining gene annotation features. Loci overlapping with annotated UTRs that did not overlap with coding exons in any transcript of the same gene or that overlapped with coding exons for less than two thirds of their length were reported as overlapping with UTRs.

2.6 Identification of canine loci for which the nearest upstream or downstream transcription start site is that of a gene annotated as a transcription factor

The nearest upstream and downstream TSS to each locus was identified based on the gene annotation with bedtools closest. The corresponding genes were classified according to the protein class that they encode (if any) using the online web server of the PANTHER Protein Classification System Database v11.0 (44).[2]

2.7 Identification of differentially expressed genes in CL patients

Differential expression analysis was performed on a total of 67 samples from three Gene Expression Omnibus [GEO, (45)] datasets: GSE112474, GSE30881, and GSE41917. GSE112474 is an RNA-seq dataset containing 12 B-cell, one T-cell, three intestinal lymphoma and four healthy lymph node samples. Intestinal lymphoma samples were considered as T-cell lymphoma samples (46). Differential expression analysis of RNA-seq read counts was performed using the R/Bioconductor package DESeq2 (47). B-cell and T-cell lymphoma libraries were compared to healthy lymph nodes separately. Genes with a p-value$\leq1\times10^{-5}$ were considered differentially expressed. GSE30881 and GSE41917 are gene expression microarray (Affymetrix Canine Genome 2.0) datasets. GSE30881 contains 23 diffuse large b-cell lymphoma lymph node samples and 10 healthy lymph node samples; GSE41917 includes seven B-cell, three T-cell lymphoma and four healthy lymph node samples. CEL files were processed and analyzed with the "simpleaffy" (48) and "affy" (49) R/Bioconductor packages. The robust multi-array average algorithm from the affy package was used for background correction and quantile normalization. Each disease sample group was independently compared to the healthy lymph node sample group. Variance estimators were computed with the empirical Bayes method in the "limma" (50) R/Bioconductor packages. Results were adjusted for multiple testing using the false discovery rate (FDR). Genes with a log_2 fold-change above 2 and an FDR≤0.05 were considered differentially expressed. The two microarray datasets were analyzed independently. Finally, the RNA-seq and microarray DEGs were pooled, removing duplicates. DEGs that could not be assigned to

---

1 https://ftp.ensembl.org/pub/release-90/gtf/canis_familiaris/Canis_familiaris.CanFam3.1.90.gtf.gz, last accessed July 20, 2023.

2 http://pantherdb.org/, last accessed July 20, 2023.

the chromosomes 1 to 38 or X were filtered out. A total 1,169 DEGs were identified in at least one of the comparisons.

2.8 Key regulators of gene expression in CL patients

The 1,000-bp sequences upstream of the TSS of the aforementioned 1,159 DEGs were searched for a maximum of 10 enriched motifs using MEME v4.11.2 (51) with parameters "-dna -nmotifs 5 -maxsize 500,000." Identified motifs were compared to the JASPAR 2016 CORE vertebrate collection of motifs (52) with TOMTOM v4.11.2 (53) with parameters "-min-overlap 5 -dist pearson -evalue -thresh 10.0" to determine which transcription factor (TF) is likely to bind each motif. The analysis was performed separately for each sample group. A total of 21 TFs were identified as such.

2.9 Network analysis of DEGs in CL patients

The Search Tool for the Retrieval of Interacting Genes/Proteins [STRING, (54)] database v10.0 was used to investigate the interaction partners of the DEGs. Individual canine Ensembl gene identifiers were queried by means of the application programming interface (API). For example, the following URL was used to retrieve the interaction partners of "ENSCAFG00000000068": http://version10. string-db.org/api/psi-mi-tab/interactionsList?identifiers=ENSCAFG 00000000068&limit=500&required_score=500&species=9615 (last accessed July 20, 2023). The URL indicates the database, the access type ("api"), the output format ("PSI-MI TAB"), the list of requests (interaction partners for any of the query items), the gene of interest, the maximum number of network nodes (proteins) that are to be returned ("limit"), the threshold of significance to include an interaction ("required_score"), and the species ("9,615," the taxonomy identifier for *Canis lupus familiaris*). After querying, we searched the output for interactions involving the Ensembl protein identifier(s) of the corresponding gene for which the score derived from curated data ("dscore") was greater than 0.5.

2.10 Canine haplotype block map

Genotyping data for 60,968 SNPs in 600 dogs and 10 wolves (55–62) was kindly provided by Adam Boyko (Cornell University). The PLINK tool set v1.90 (63) was used to estimate the canine haplotype block structure. Because of the missing phenotype data, the option "no-phenoreq," which removes the phenotype restriction, was added to the default settings. The coordinates of the haplotype blocks were converted from the assembly version CanFam2 to the assembly version CanFam3.1 using UCSC's LiftOver tool.

2.11 Primer design and synthesis

Primers for generation of amplicons were designed based on sequences with a maximum length of 120 bp centered at the selected SNPs using the Ion AmpliSeq Designer v5.63 (Thermo Fisher Scientific, Waltham, MA, United States).

2.12 Canine lymphoma patient selection

Dogs referred to the Small Animal Clinic of the University of Veterinary Medicine Hannover for diagnostic investigation and treatment for CL between 2008 and 2014 were prospectively considered for enrollment in this study. Diagnosis was conducted based on cytological or histological evaluation of lymph nodes or extranodal lesions such as liver, spleen, and bone marrow. Patients underwent complete staging, consisting of history and physical examination, complete blood cell count, serum biochemistry profile, thoracic radiographs, abdominal ultrasound, cytological evaluation of liver and spleen regardless of their sonographic appearance, and bone marrow aspiration. Clinical staging was based on the World Health Organization (WHO) for canine lymphoma (64). Immunophenotype was determined by flow cytometry. The presence of the CD21 antigen and absence of CD3 antigen was considered diagnostic for B-cell lymphoma. All patients received standard treatment, namely CHOP followed by CCNU (lomustine), with or without L-asparaginase (see (65) for details). Inclusion criteria was a confirmed diagnosis of multicentric B-cell lymphoma. and absence of concomitant diseases that limited study compliance.

2.13 Clinical assessment

Response duration was assessed according to the Veterinary Cooperative Oncology Group (VCOG) consensus document (66). We reported response duration as progression-free survival (PFS), overall survival (OS). PFS was defined as the time from treatment initiation to first relapse; OS was defined as the time from treatment initiation to death from any cause. At the time of sequencing, 22 dogs had died from lymphoma-related causes, and seven were lost to follow-up.

2.14 Sample collection and DNA isolation

Before the first chemotherapeutic administration, blood samples of 29 dogs were collected for routine hematological and biochemical analyses. An aliquot of 200 μL retrieved from residual whole blood of each patient was preserved in EDTA and frozen in −80°C. DNA was extracted using the NucleoSpin Blood Kit (Macherey-Nagel) according to the manufacturer's instructions. DNA quantification was carried out through Qubit 2.0 fluorometry (Thermo Fisher Scientific, Waltham, United States).

2.15 Library preparation and targeted NGS

From each sample, 10 ng of genomic DNA was used to prepare Ion Torrent sequencing libraries. Initial multiplex PCR was carried out at 2 min at 99°C, 18 cycles of 15 s at 99°C, 4 min at 60°C, and held at 10°C. Library quantification was performed via qPCR using Ion Library Quantitation Kit. Library nanosphere coupling and amplification was performed according to standard IonTorrent procedures using Ion OneTouch 2 and Ion OneTouch ES. Single-end sequencing was performed on an Ion Torrent™ Personal Genome Machine™ (PGM™) System using 318 v2 chips and 400 flows (Thermo Fisher Scientific, Waltham, United States).

2.16 Variant calling

The Torrent Suite software v5.4 (Thermo Fisher Scientific, Waltham, United States) was used to demultiplex reads, map the reads to the dog reference genome, and generate run metrics. Specifically, read mapping was performed against the dog reference genome [CanFam3.1[3], (67)] using the Torrent Mapping Alignment Program (TMAP) with default cost values for mismatches (2) and indels (3) and minimum reference similarity (80%). Mapping quality (MAPQ) scores are reported in the Phred-scale.

Reads with mapping quality (MAPQ) below 90 were excluded using SAMtools view v1.10 (68). Duplicate marking was not deemed suitable for amplicon sequencing and thus omitted. Base calling was performed on the remaining reads with Freebayes v0.9.21.7 (69), GATK's Haplotypecaller v4.1.7.0 (70), and BCFtools mpileup and call v1.9 (68). Each sample was processed separately. Reads used as input for Freebayes were realigned with ABRA2 v2.24 (71). Freebayes was run with default parameters; variants called across the entire genome were filtered for those within the amplicon regions with vcftools v0.1.16 (72). BCFtools mpileup was executed with parameters "-d 100,000 -L 100000" to ensure that all available reads were used for variant calling and that regions with high coverage were not skipped during indel calling; in addition, the "-r" parameter was used to restrict variant calling to amplicon regions. The output of mpileup served as input for BCFtools call, which was executed with the multiallelic and rare-variant calling model and parameter "-v" to output variant sites only (i.e., sites at which the sample had a non-reference allele). The reference genome dictionary and index required by GATK Haplotypecaller were generated with picard CreateSequenceDictionary v2.18.29 (73) and GATK IndexFeatureFile, respectively. GATK BaseRecalibrator and ApplyBQSR were used for base quality recalibration; all variants in the Ensembl Variation database[4] were used with the parameter "--known-sites." Finally, GATK Haplotypecaller was run restricting the analysis to the panel amplicons (with the "--intervals" parameter) and with parameter "--max-reads-per-alignment-start" set to 0 to disable read downsampling. The variants called with each of the three variant callers were then filtered with BCFtools view so that only variants with a quality score ≥20 were kept for further analyses and normalized with BCFtools norm to left-align and normalize indels. Finally, BCFtools isec was used with parameters "--collapse both" (to collapse SNPs and indels) and "--nfiles +1"to identify the variants called by 1, 2, and 3 variant callers in each of the samples. Variants called by all three variant callers in at least one sample are subsequently referred to as "pooled." Only sites that were not called variants by any of the three variant callers were considered homozygous for the reference allele.

Pooled variants were annotated with the Variant Effector Predictor (VEP) v109.3 (72, 74).

2.17 Survival analyses

Survival analysis was performed for PFS and OS using the R packages "survminer" v0.4.9 (75), "survival" v3.5–5 (76), and "MASS"

v7.3–53 (77). For OS, the seven patients lost to follow-up at the time of sequencing were censored at the latest date of available records.

Univariable Cox regression (78) was performed using the cophx() function of the "survival" package to analyze associations with the covariates "WHO clinical stage," "sex," "neuter status," "chemotherapeutic protocol," and "age at diagnosis." All variables were treated as categorical; "age at diagnosis" was encoded as 1 for "adult" (2–7 years) and 0 for "senior" (8–11 years).

Pooled variants homozygous or heterozygous for the alternate allele in 11 to 18 dogs and homozygous for the reference allele in all other dogs of the study cohort were subjected to survival analysis following the approach from Collett (79). Briefly, we first performed a univariable Cox regression for each variant using the cophx() function of the "survival" package. Next, we included variants with a p-value ≤ 0.5 in a multivariable Cox regression model, and conducted stepwise backward selection with the stepAIC() function of the "MASS" package to select a combination of variants with lowest Akaike Information Criterion (AIC) value. All variants with a p-value ≤ 0.1 in the resulting model were used for multivariable Cox regression with stepwise forward selection of any variables with a p-value ≤ 0.5 in the univariate Cox regression. Variants with a p-value ≤ 0.1 in the resulting model were used for stepwise bidirectional –forward and backward– regression. Variants with a p-value ≤ 0.05 in the final model were considered significant.

3 Results

3.1 Database mining unveiled 482 putative lymphoma-associated canine loci

To identify putative lymphoma-associated loci in the canine genome, we first mined the Online Mendelian Inheritance in Man (OMIM) database (35), the GWAS catalog (36), and PubMed titles and articles for human or canine SNPs (Methods). This Initial search yielded 89, 173, and 228 human SNPs, respectively. Furthermore, we manually reviewed all full text articles about NHL and CL indexed in PubMed, identifying 219 human and three canine SNPs (Methods). In total, this accounted for 592 non-redundant human (Figure 1A) and three canine SNPs. These three canine SNPs were the only SNPs predisposing for canine B-cell lymphoma that could be identified by a GWAS involving 41 cases and 172 controls (80).

To assess the evolutionary conservation of the 592 human SNPs and identify their orthologous canine loci, we computed multiple sequence alignments between 100 bp-long sequences centered at each human SNP and their orthologous sequences in the dog and three additional species' genomes (Methods). Thereby, we found canine orthologs for 524 (89%) human sequences, but discarded one that diverged considerably in length (Methods) and one on pseudo-chromosome "chrUn" (contigs that could not be assigned to any chromosome). From the remaining 522 canine sequences, 43 contained insertions or deletions at the site orthologous to the SNP in the human sequence, and thus were not considered evolutionary conserved. Excluding these 43 sequences yielded 479 putative lymphoma associated canine loci.

Overall, taking into account the three canine SNPs reported in the literature, we discovered 482 putative lymphoma associated canine loci. Of these, 420 (87%) were non-protein-coding.
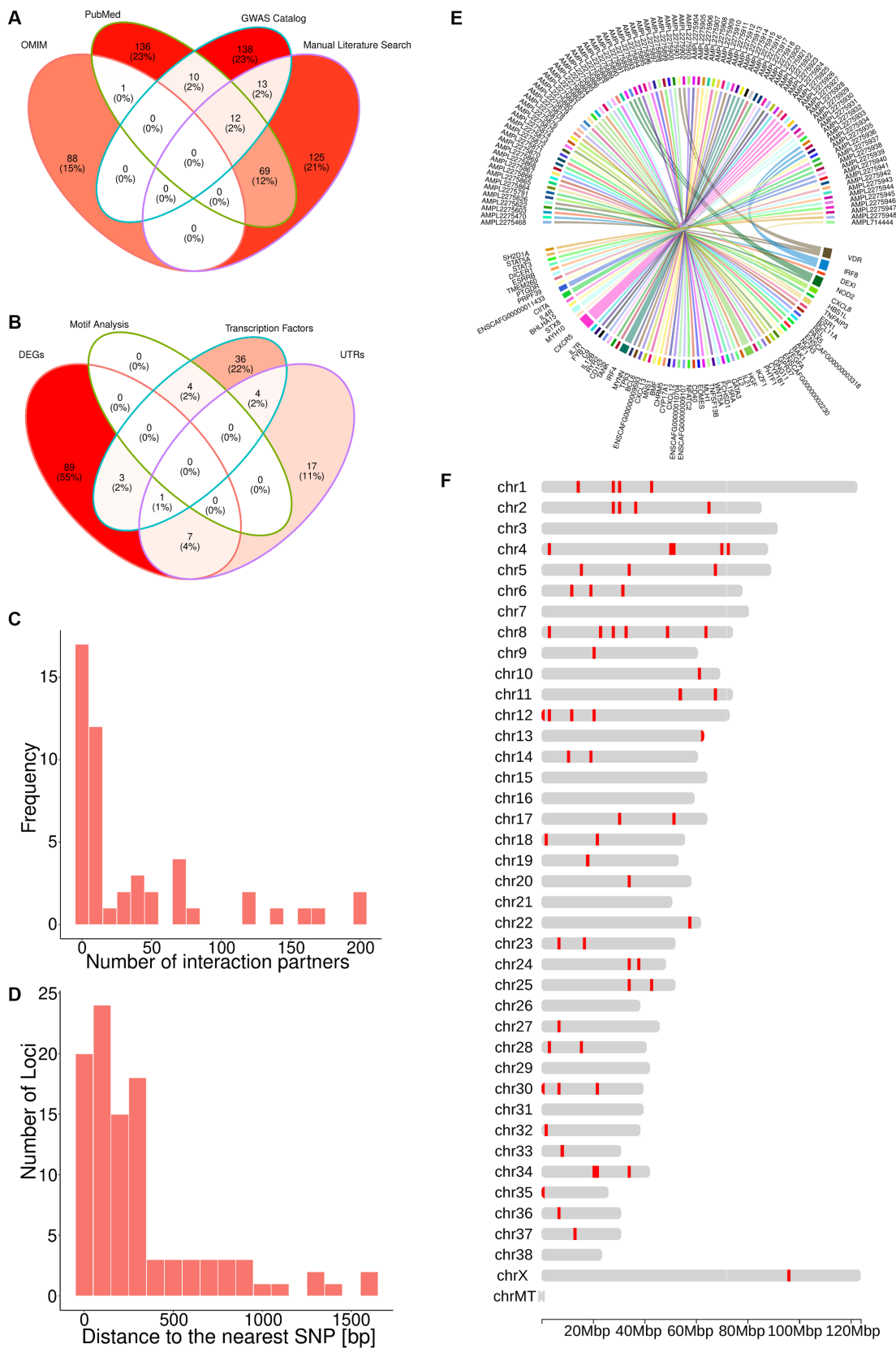
FIGURE 1
The TiHoCL targeted sequencing panel relies on SNPs associated with NHL to detect variants in CL. **(A)** Venn diagram showing the sources of the 592 human lymphoma-associated SNPs; 89 SNPs were retrieved from the OMIM, 173 from the GWAS Catalog, 228 from PubMed, and 219 through manual literature curation. **(B)** Venn diagram displaying how the SNPs from **(A)** were selected based on the different filtering strategies. The brightness of the filling color in **(A,B)** corresponds to the number of SNPs. **(C)** Number of interaction partners for the DEGs. **(D)** Distance between selected 98 potential lymphoma-associated loci and the nearest known SNPs in the canine genome, which was also a filtering criteria for the SNPs. **(E)** Circos plot showing the nearest gene for each target. **(F)** Distribution of the 93 targets across the dog genome.

3.2 Functional genomics and network approaches identified high-priority loci for CL patient management

Since the aim was to develop a fast and relatively inexpensive SNP genotyping panel for diagnosis and monitoring of CL patients, we applied a combination of functional genomics and network approaches to single out approximately 100 most promising putative lymphoma-associated canine loci out of the aforementioned 482.

First, we selected the 29 loci located within untranslated regions (UTRs) and 48 loci for which the nearest upstream or downstream TSS was that of a gene annotated as a TF (Methods). Variants in UTRs may alter transcription binding sites, splicing sites and polyadenylation of mRNAs (80–83). Similarly, variants within or in the neighborhood of TF genes may lead to aberrant TF expression and, in turn, gene dysregulation of almost all known cellular processes related to tumorigenesis (84), which could explain the widespread gene expression dysregulation observed in lymphoma (85, 86).

Next, we examined the loci in the neighborhood of genes found to be differentially expressed in lymphoma in several transcriptomic studies (Methods). We found four loci located in the Vitamin D Receptor gene, which encodes a putative key regulator of gene expression in lymphoma. Because genetic variation in this gene has been associated with lymphomagenesis (87–94), these four loci were also selected for the panel. Further, the nearest TSS to 100 of the loci corresponded to a DEG. Eleven of these loci were also in the neighborhood of a TF gene or within a UTR and, hence, had already been selected for the panel (Figure 1B). From the remaining loci, we selected 29 near protein-coding genes that possessed more than 10 interaction partners (Methods). Genes with a large number of interaction partners are referred to as network "hubs" (Figure 1C), and hubs are typically deemed important, because their perturbations can have major consequences for the regulatory networks to which they contribute (95).

In total, the aforementioned approaches yielded 101 loci in the canine genome representing the orthologous sites of human SNPs. Importantly, these loci are not necessarily polymorphic in the canine genome. In order to associate them with known canine SNPs, we constructed and analyzed a canine haplotype block map (Methods). We observed that the median block size over all individuals was 5,725 bp and that all loci of interest were located at a distance smaller than the median haplotype block size from a known SNP in the canine genome, with a median distance of 195 bp and a maximum of 1,616 bp (Figure 1D). This implies that the loci of interest and their nearest SNPs are very likely to be in linkage disequilibrium, and hence, we chose the latter to represent the former in our panel. Since some of the loci were located in the proximity of the same SNP, the total number of SNPs associated with the loci of interest was 100.

3.3 The TiHoCL sequencing panel targets 93 SNPs in the canine genome

The 100 canine SNPs derived from the multi-step genomic analyses presented above together with the three SNPs that had been reported in the literature to be associated with CL were used for designing the TiHoCL targeted sequencing panel. Using the Ion Ampliseq Designer (Methods) we were able to design primers for 93 of the 103 SNPs (Figures 1E,F). The resulting custom, single pool,

multiplexed, PCR-based, NGS library panel comprised 93 target amplicons with an insert size between 220 and 335 bp in length (median = 324 bp), each covering one of the aforementioned 93 SNPs, and had a total size of ~30 kb (Supplementary Table S1). We further refer to the target amplicons as *targets*.

3.4 TiHoCL targeted sequencing panel identifies variants in CL samples

We technically validated the TiHoCL targeted sequencing panel using DNA extracted from a cohort of 29 dogs presenting B-cell multicentric lymphoma (Supplementary Table S2). The most encountered breeds were mixed-breed dogs (6/29, 21%), Bernese Mountain Dog (5/29, 17%), Golden Retriever (2/29, 7%), and Rottweiler (2/29, 7%), but other 14 additional breeds were found (14/29, 48%). There were 18 males and 11 female dogs; half of them were neutered, and half were intact. The age at diagnosis ranged from two to 11 years (mean [SD]: 6.45 [2.21] years). Weight was recorded for 25 dogs; they weighed between 6.8 and 51 kg. According to the WHO clinical stage classification, 2 dogs had stage III disease, 16 dogs stage IV disease, and 11 dogs stage V disease. Six dogs were treated with a 12-week CHOP protocol followed by three doses of CCNU (lomustine), while 23 dogs received the same protocol with the addition of L-asparaginase in the first week. A group of five patients relapsed before the induction phase was finished, and all 29 patients relapsed before the end of the study period (24 months). Univariable Cox regression revealed no association between WHO clinical stage classification, sex, neuter status, treatment protocol, and age at diagnosis and PFS or OS (Supplementary Table S3).

Sequencing run metrics showed good data quality for all samples (Supplementary Table S4). Uniformity of base coverage across samples ranged from 89 to 96%. After quality filtering, the average read length per sample ranged between 306 and 318 bp with an average of 312 bp (Figure 2A). Depending on the sample, the number of reads mapped to targets ranged from 188,712 to 582,628. This corresponds to a percentage of reads on target between 98.50 and 99.99%, confirming adequate target design, library preparation, and sequencing. The median target coverage exceeded 1,000X, with an average over all samples of 2,612X (Figure 2B). Similarly, the median sample coverage was greater than 500X, with only nine exceptions, with an average over all targets of 2,641X (Figure 2C). This is in line with the minimum coverage generally recommended for clinical oncology panels (96). In addition, 77 to 88% of the bases in the targets indicated no strand bias. Because of the satisfying run metrics we continued with functional validation.

To assess the level of polymorphism of the targets we performed variant calling (Methods). To obtain a reliable list of potential variants, we used three different variant callers: Freebayes, GATK Haplotypecaller, and BCFtools mpileup and call, and solely considered variants called by all three variant callers for downstream analysis. The three variant callers were chosen because they are partly based on different approaches and are broadly used in the field (97–100).

Compared to the reference genome, we detected a total of 188 "pooled" (heterozygous or homozygous) variants (Figure 3A, Supplementary Table S5), with each sample exhibiting between 56 and 96 variants (Methods). For 10 of the variants, all 29 samples were homozygous for the same allele (but polymorphic relative to the
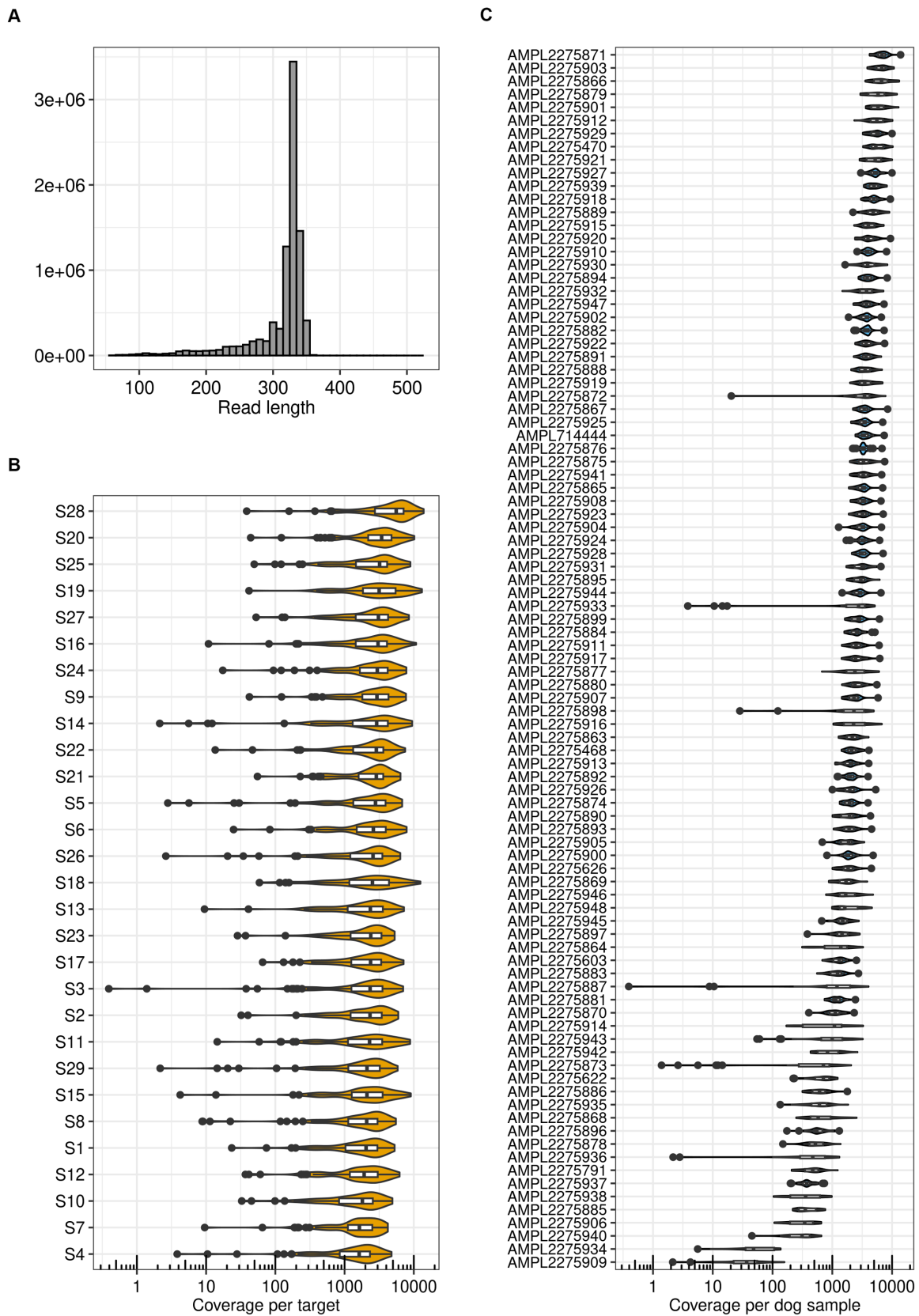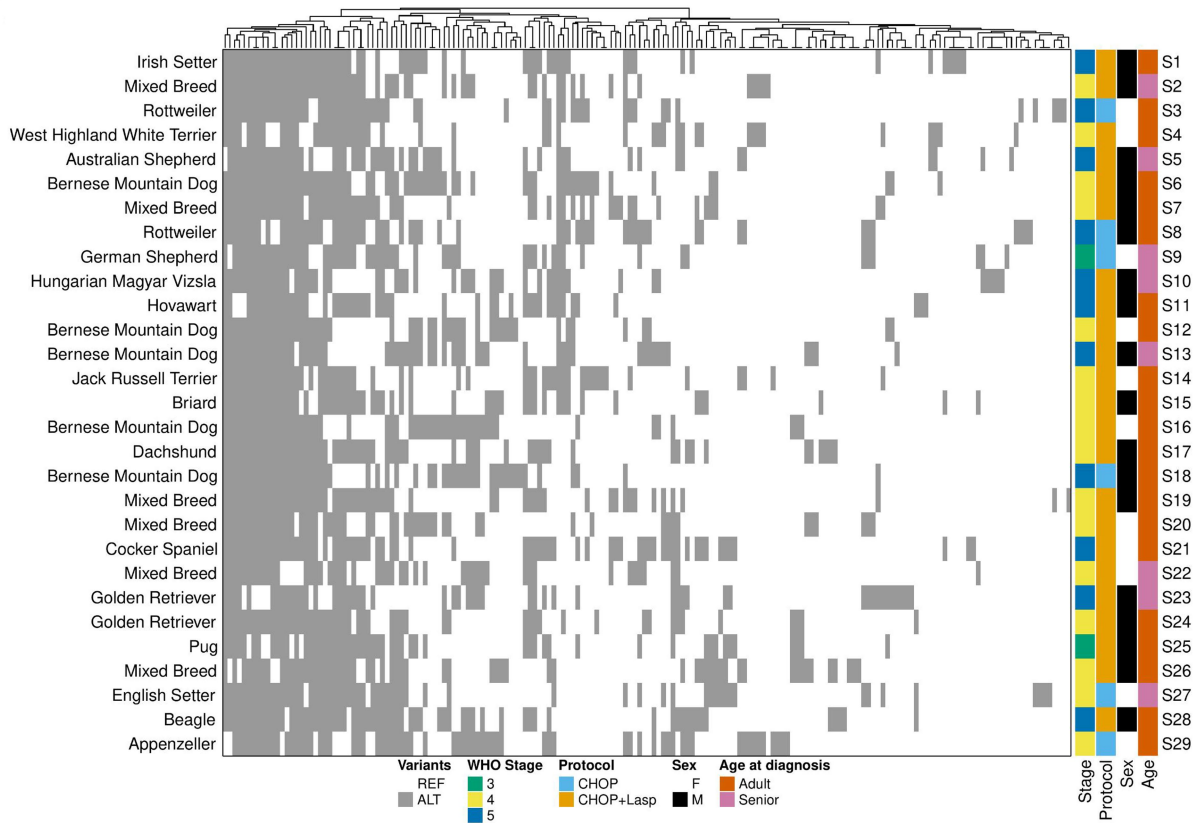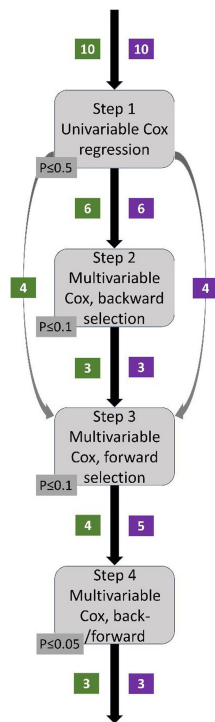
FIGURE 2
TiHoCL targeted sequencing panel achieved good run metrics. **(A)** Distribution of read lengths obtained after mapping quality (MAPQ) filtering with the panel on 29 dog samples. **(B)** Target coverage per sample. The Median coverage was greater than 500X for all samples. **(C)** Coverage of all samples per target. For seven targets the median coverage was smaller than 500X, two of them being outliers with a median coverage smaller than 100X.
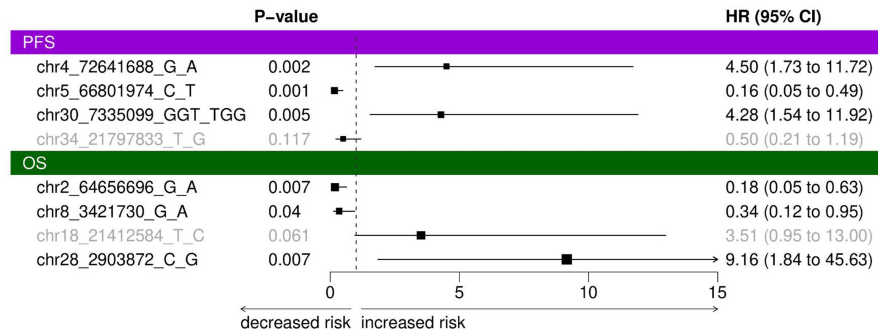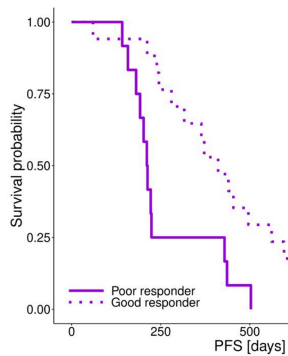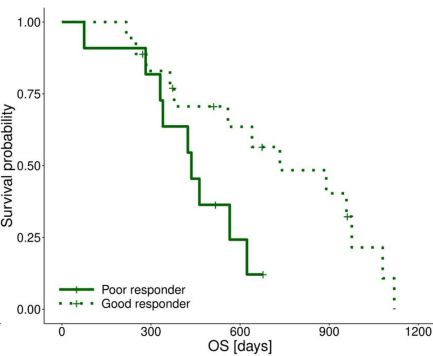
FIGURE 3
TiHoCL targeted sequencing panel enabled detection of natural genetic variation in the study cohort and of variants associated with CL relapse and survival. **(A)** Heatmap of polymorphic "pooled" variants across 29 dog samples. Variants are indicated with respect to the reference genome. REF:

(Continued)

reference genome sequence). Of the 188 pooled variants, the majority (81%) were SNPs, followed by indels (19%). Consistent with the panel design, the 87% were non-protein coding. From the original 93 SNPs used to design the panel, 70 were also polymorphic relative to the reference genome sequence in the samples. Each target exhibited an average of 0.7 variants per sample.

Of the pooled variants, 12 were seen in 11–19 dogs, and considered polymorphic in the cohort. We checked this list for variants showing the same ALT/REF pattern and removed the duplicated ones for the next step which left 10 variants. These variants were then subjected to multistep survival analysis using Cox proportional hazards regression (Figure 3B). The Cox (proportional hazards or PH) model describes the relation between the (cumulative) incidence of an event of interest (in our case, death for OS and relapse for PFS) and a set of covariates (in our case, genetic variants), and is a commonly used approach for analyzing survival time data in medical research. Using univariate Cox regression analysis, we first identified 6 variants weakly associated with overall survival (OS). Then, we applied multivariable Cox regression analysis to establish a prognostic model. The final model included three variants (Figure 3C). Among them, one was associated with increased risk of death and two were associated with decreased risk of death. For PFS, the same procedure resulted in a model including three variants (Figure 3C), of which two were associated with increased risk of relapse and one was associated with decreased risk of relapse. The final models clearly distinguished patients into poor (median OS: 436 days, median PFS: 212 days) and good responders (median OS: 600 days, median PFS: 411; Figures 3D,E).

To summarize, the TiHoCL targeted sequencing panel showed good run metrics and allowed us to characterize known and novel genetic variation in the cohort. Furthermore, we identified eight variants associated with a decreased and increased risk of relapse or death.

3.5 TiHoCL targeted sequencing panel is labor- and cost-saving

The total costs for sequencing on an Ion Torrent PGM platform, including primer pool, library preparation, emulsion PCR and massively parallel sequencing, are approximately 220 € per sample using the maximum capacity of a 318 v2 chip with 16 samples. Initially, the 318 v2 Chip allows 5×10^6 reads, thus considering the uniform target amplification kinetics of the TiHoCL Panel, 100 samples could be sequenced with coverage of 500X. However, as sequencing efficiency is strongly dependent on chip-loading, the sequencing of 16 to 29 samples per chip will be delivering easily 1,000X coverage of all analyzed bases for less than 250€ per sample in 2 to 5 h.

4 Discussion

The widespread use of targeted sequencing panels in human oncology (101) anticipates new directions for veterinary science, wherein similar panels have the potential to revolutionize our understanding and management of cancer in animals. Recent technological developments illustrate the significance and viability of this. For example, Sirivisoot and colleagues (102) used mass spectrometry to assess variation in 41 loci previously reported to drive lymphomagenesis and associated with genes producing targetable proteins, and found that their panel provides useful prognostic information when screening a cohort of 60 dogs with diffuse large B-cell lymphoma. To move the research to the next level, we demonstrate how functional genomics and network approaches can be combined to select SNPs mined from the literature with the aim of designing a targeted panel for CL. This panel was evaluated in a cohort of 29 CL patients, achieving satisfactory sequencing metrics, capturing natural polymorphisms in the canine population, and exhibiting a promising potential to stratify CL patients.

Although our knowledge about the molecular mechanisms underlying CL is limited, CL resembles the much better researched human disease NHL in many regards, and this can be used to design tools that can help guide treatment decisions in CL patients.

Our panel design primarily relied on knowledge from the literature and the OMIM and GWAS catalog databases. Automatic text mining resulted in 467 SNPs. Manual inspection of the full text of selected PubMed articles produced 125 additional SNPs. In this manner, we identified a total of 592 human and three canine SNPs. These SNPs were then filtered for the most promising candidates using a combination of functional genomics and network biology approaches to build a CL panel including approximately 100 loci. Including a relatively small number of loci in the panel design enabled us to subject samples to deep sequencing—which is considered superior over technologies such as whole-genome sequencing, whole-exome sequencing, and PCR when analyzing multiple mutational spots in parallel (103, 104)—at an affordable cost. On these grounds we specifically prioritized loci within UTRs, near TFs, or near genes that are differentially expressed in CL patients and code for proteins that act as "hubs" in relevant regulatory networks. The TiHoCL panel was technically and preliminary functionally assessed on a cohort of 29 CL patients. Most (90%) amplicons performed well, reaching a coverage of 500X. Furthermore, the panel detected 56 to 96 variants in each CL patient, successfully capturing some of the natural genetic variation within the cohort and paving the way for personalized CL therapy. Highly encouraging, a stringent multistep approach to variable selection and survival analysis on a set of 12 variants detected

in more than 10 patients identified four variants associated with decreased or increased risk of death, and another four, of relapse. These variants merit further investigation for validation as prognostic biomarkers.

Two main limitations should be acknowledged. First, our panel was based on a relatively small number of SNPs extracted from the literature and specialized databases, which only partially represent the scientific knowledge about CL. In particular, we restricted ourselves to two manually-curated, high-quality databases to increase our chances of including high-confident SNPs in the panel design, but those are not the only –or the most comprehensive– databases that can be used as source of disease-associated SNPs. For future panel designs it could be worth considering databases like ClinVar (105), which is focused on disease associated SNPs, COSMIC (105, 106), which is focused on SNPs for cancer research and diagnostics, or dbVar (107), a database for human structural variation to cover also larger rearrangements. In addition, our search did not distinguish between germline and somatic variation, and thus, the SNPs finally included in the panel were not optimized for patient stratification and management, which is the ultimate goal of the TiHoCL panel. Indeed, such restriction appeared excessive in light of the scarce genetic knowledge available for CL. Second, functional assessment of the panel was performed on a cohort of only 29 dogs, which may not represent the full spectrum of CL genetic subtypes and canine breeds. Although our results provide preliminary evidence about the functional capabilities of the TiHoCL panel, we acknowledge that the variants identified as being associated with risk of relapse or death warrant further examination. The decision on the cohort size and diversity was made considering the exploratory nature of our study and the constraints imposed by cost, time and case incidence. While focusing on specific, lymphoma-prone breeds would help reduce genetic background heterogeneity and population structure, leading to more reliable results, this approach could miss important genetic interactions, and would have limited utility for mixed-breed dogs. Breed diversity is also supported by a study of Elvers and colleagues (108) indicating that breeds predisposed to B-cell lymphoma have commonly mutated genes and pathways. A larger cohort would increase the statistical power of the study, and improve the reliability and robustness of our findings. In summary, knowledge transfer to clinical practice would entail leveraging the most current information for panel design and an extensive functional validation of the thereby enhanced TiHoCL panel on a larger and more diverse cohort.

Our design and assessment framework is a proof of principle, representing a foundation upon which other targeted sequencing panels could be built. The version of the TiHoCL targeted sequencing panel presented here can be seen as the first stage in an iterative development model. Our framework streamlines the establishment of refined and updated versions of the TiHoCL panel. As genomic data for lymphoma accumulates, new loci could be periodically incorporated into the panel, especially those arising from progressively more accessible whole-genome sequencing studies, while systematically removing loci with limited potential. With each iteration, the panel's performance is set to increase, enabling novel opportunities for enhancing CL patient management. Our frameworks flexibility ensures its enduring relevance in the constantly evolving landscape of cancer research and propels the field of precision veterinary medicine to new heights.

Because human diseases are generally more thoroughly researched and documented, most genetic databases, including OMIM and the GWAS catalog, are limited to human variants. Moreover, most of the SNPs obtained from the literature were human SNPs. Therefore, the TiHoCL targeted sequencing panel for CL was primarily built based on knowledge acquired for a different, but related disease, namely NHL. This concept is transferable to other cancer types or diseases that share features with diseases in other species, like sarcomas in dogs or urothelial neoplasms in sea lions (109). In contrast to the situation in human medicine, advanced molecular analysis tools for veterinary patients are scarce, limiting clinical options. The reasons for this are manyfold, but economic factors should not be underestimated. Thus, many more resources are invested in human than in dog cancer research. In addition, while many costs in human medicine are often covered by national health care systems, veterinarian costs are normally covered by the animal's owner. Of course, our design framework could also be beneficial for human medicine and have an impact on translational medicine. Clinical trials in dogs with spontaneous cancers must be conducted ethically and in compliance with relevant animal welfare regulations, but offer several advantages compared to studies in humans, including reduced liability and faster approval timelines. CL is a unique spontaneous model for NHL, and therefore, can be used to understand NHL progression and drug resistance processes, as well as to develop new treatments (9, 10, 14). Furthermore, while certain diseases might be rare in humans and hence, difficult to study, they could occur frequently in dogs, providing a unique perspective for research.

The broad application of panels similar to our TiHoCL targeted sequencing panel for CL not only has the potential of improving the health and welfare of all species suffering from cancer, but also contribute to our understanding of the underlying pathology.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA1017182.

Ethics statement

The animal studies were approved by the Small Animal Clinic of the University of Veterinary Medicine Hannover, Foundation (Hanover, Germany) in accordance with the German Animal Welfare Guidelines, approved by the Ethics Committee of the State of Lower Saxony, Germany (No. 14/1700). None of the dogs were euthanized due to reasons of sample collection. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

SF-S: Writing – original draft, Data curation, Formal Analysis, Methodology, Software, Visualization, Writing – review & editing. CI: Data curation, Writing – original draft, Writing – review & editing. DS: Data curation, Formal Analysis, Software, Writing – review & editing.

NE: Data curation, Writing – review & editing. JG-S: Data curation, Writing – review & editing. WL: Data curation, Writing – review & editing. SK: Investigation, Writing – review & editing. CJ: Funding acquisition, Project administration, Writing – review & editing. IN: Funding acquisition, Methodology, Project administration, Writing – review & editing. LT: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. HM: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1301536/full#supplementary-material

References

1. Di Resta C, Ferrari M. Next generation sequencing: from research area to clinical practice. *EJIFCC*. (2018) 29:215–20.

2. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP, et al. Targeted next-generation sequencing can replace sanger sequencing in clinical diagnostics. *Hum Mutat*. (2013) 34:1035–42. doi: 10.1002/humu.22332

3. Gerstung M, Beisel C, Rechsteiner M, Wild P, Schraml P, Moch H, et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat Commun*. (2012) 3:811. doi: 10.1038/ncomms1814

4. Garcia EP, Minkovsky A, Jia Y, Ducar MD, Shivdasani P, Gong X, et al. Validation of OncoPanel: a targeted next-generation sequencing assay for the detection of somatic variants in Cancer. *Arch Pathol Lab Med*. (2017) 141:751–8. doi: 10.5858/arpa.2016-0527-OA

5. Deshpande A, Lang W, McDowell T, Sivakumar S, Zhang J, Wang J, et al. Strategies for identification of somatic variants using the ion torrent deep targeted sequencing platform. *BMC Bioinformatics*. (2018) 19:5. doi: 10.1186/s12859-017-1991-3

6. Suminda GGD, Bhandari S, Won Y, Goutam U, Kanth Pulicherla K, Son YO, et al. High-throughput sequencing technologies in the detection of livestock pathogens, diagnosis, and zoonotic surveillance. *Comput Struct Biotechnol J*. (2022) 20:5378–92. doi: 10.1016/j.csbj.2022.09.028

7. Petersen JL, Coleman SJ. Next-generation sequencing in equine genomics. *Vet Clin North Am Equine Pract*. (2020) 36:195–209. doi: 10.1016/j.cveq.2020.03.002

8. Lustgarten JL, Zehnder A, Shipman W, Gancher E, Webb TL. Veterinary informatics: forging the future between veterinary medicine, human medicine, and one health initiatives-a joint paper by the Association for Veterinary Informatics (AVI) and the CTSA one health Alliance (COHA). *JAMIA Open*. (2020) 3:306–17. doi: 10.1093/jamiaopen/ooaa005

9. Richards KL, Suter SE. Man's best friend: what can pet dogs teach us about non-Hodgkin's lymphoma? *Immunol Rev*. (2015) 263:173–91. doi: 10.1111/imr.12238

10. Ito D, Frantz AM, Modiano JF. Canine lymphoma as a comparative model for human non-Hodgkin lymphoma: recent progress and applications. *Vet Immunol Immunopathol*. (2014) 159:192–201. doi: 10.1016/j.vetimm.2014.02.016

11. Linschoten M, Kamphuis JAM, van Rhenen A, Bosman LP, Cramer MJ, Doevendans PA, et al. Cardiovascular adverse events in patients with non-Hodgkin lymphoma treated with first-line cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) or CHOP with rituximab (R-CHOP): a systematic review and meta-analysis. *Lancet Haematol*. (2020) 7:e295–308. doi: 10.1016/S2352-3026(20)30031-4

12. Mee MW, Faulkner S, Wood GA, Woods JP, Bienzle D, Coomber BL. RNA-Seq analysis of gene expression in 25 cases of canine lymphoma undergoing CHOP chemotherapy. *BMC Res Notes*. (2022) 15:111. doi: 10.1186/s13104-022-06003-5

13. Zandvliet M. Canine lymphoma: a review. *Vet Q*. (2016) 36:76–104. doi: 10.1080/01652176.2016.1152633

14. Marconato L, Gelain ME, Comazzi S. The dog as a possible animal model for human non-Hodgkin lymphoma: a review. *Hematol Oncol*. (2013) 31:1–9. doi: 10.1002/hon.2017

15. Marconato L. The staging and treatment of multicentric high-grade lymphoma in dogs: a review of recent developments and future prospects. *Vet J*. (2011) 188:34–8. doi: 10.1016/j.tvjl.2010.04.027

16. Ansell SM. Non-Hodgkin lymphoma: diagnosis and treatment. *Mayo Clin Proc*. (2015) 90:1152–63. doi: 10.1016/j.mayocp.2015.04.025

17. Zahid U, Akbar F, Amaraneni A, Husnain M, Chan O, Riaz IB, et al. A review of autologous stem cell transplantation in lymphoma. *Curr Hematol Malig Rep*. (2017) 12:217–26. doi: 10.1007/s11899-017-0382-1

18. Zandvliet M, Teske E, Schrickx JA. Multi-drug resistance in a canine lymphoid cell line due to increased P-glycoprotein expression, a potential model for drug-resistant canine lymphoma. *Toxicol In Vitro*. (2014) 28:1498–506. doi: 10.1016/j.tiv.2014.06.004

19. Hedström G, Thunberg U, Amini RM, Zainuddin N, Enblad G, Berglund M. The MDM2 polymorphism SNP309 is associated with clinical characteristics and outcome in diffuse large B-cell lymphoma. *Eur J Haematol*. (2014) 93:500–8. doi: 10.1111/ejh.12388

20. Cordano P, Lake A, Shield L, Taylor GM, Alexander FE, Taylor PRA, et al. Effect of IL-6 promoter polymorphism on incidence and outcome in Hodgkin's lymphoma. *Br J Haematol*. (2005) 128:493–5. doi: 10.1111/j.1365-2141.2004.05353.x

21. Park YH, Sohn SK, Kim JG, Lee MH, Song HS, Kim MK, et al. Interaction between BCL2 and interleukin-10 gene polymorphisms alter outcomes of diffuse large B-cell lymphoma following rituximab plus CHOP chemotherapy. *Clin Cancer Res*. (2009) 15:2107–15. doi: 10.1158/1078-0432.CCR-08-1588

22. Gustafson HL, Yao S, Goldman BH, Lee K, Spier CM, LeBlanc ML, et al. Genetic polymorphisms in oxidative stress-related genes are associated with outcomes following treatment for aggressive B-cell non-Hodgkin lymphoma. *Am J Hematol*. (2014) 89:639–45. doi: 10.1002/ajh.23709

23. Bassig BA, Cerhan JR, Au WY, Kim HN, Sangrajrang S, Hu W, et al. Genetic susceptibility to diffuse large B-cell lymphoma in a pooled study of three eastern Asian populations. *Eur J Haematol*. (2015) 95:442–8. doi: 10.1111/ejh.12513

24. Wang SS, Vajdic CM, Linet MS, Slager SL, Voutsinas J, Nieters A, et al. Associations of non-Hodgkin lymphoma (NHL) risk with autoimmune conditions according to putative NHL loci. *Am J Epidemiol*. (2015) 181:406–21. doi: 10.1093/aje/kwu290

25. Baecklund F, Foo JN, Bracci P, Darabi H, Karlsson R, Hjalgrim H, et al. A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival. *BMC Med Genet*. (2014) 15:113. doi: 10.1186/s12881-014-0113-6

26. He J, Liao XY, Zhu JH, Xue WQ, Shen GP, Huang SY, et al. Association of MTHFR C677T and A1298C polymorphisms with non-Hodgkin lymphoma susceptibility: evidence from a meta-analysis. *Sci Rep*. (2014) 4:6159. doi: 10.1038/srep06159

27. Hernández-Verdin I, Labreche K, Benazra M, Mokhtari K, Hoang-Xuan K, Alentorn A. Tracking the genetic susceptibility background of B-cell non-Hodgkin's lymphomas from genome-wide association studies. *Int J Mol Sci [Internet]*. (2020) 22:122. doi: 10.3390/ijms22010122

28. Ghesquieres H, Slager SL, Jardin F, Veron AS, Asmann YW, Maurer MJ, et al. Genome-wide association study of event-free survival in diffuse large B-cell lymphoma treated with Immunochemotherapy. *J Clin Oncol*. (2015) 33:3930–7. doi: 10.1200/JCO.2014.60.2573

29. Cerhan JR, Wang S, Maurer MJ, Ansell SM, Geyer SM, Cozen W, et al. Prognostic significance of host immune gene polymorphisms in follicular lymphoma survival. *Blood*. (2007) 109:5439–46. doi: 10.1182/blood-2006-11-058040

30. Habermann TM, Wang SS, Maurer MJ, Morton LM, Lynch CF, Ansell SM, et al. Host immune gene polymorphisms in combination with clinical and demographic factors predict late survival in diffuse large B-cell lymphoma patients in the pre-rituximab era. *Blood*. (2008) 112:2694–702. doi: 10.1182/blood-2007-09-111658

31. Armitage JO, Gascoyne RD, Lunning MA, Cavalli F. Non-Hodgkin lymphoma. *Lancet*. (2017) 390:298–310. doi: 10.1016/S0140-6736(16)32407-2

32. Cerhan JR, Slager SL. Familial predisposition and genetic risk factors for lymphoma. *Blood*. (2015) 126:2265–73. doi: 10.1182/blood-2015-04-537498

33. Cerhan JR, Ansell SM, Fredericksen ZS, Kay NE, Liebow M, Call TG, et al. Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma. *Blood*. (2007) 110:4455–63. doi: 10.1182/blood-2007-05-088682

34. Giannuzzi D, Marconato L, Cascione L, Comazzi S, Elgendy R, Pegolo S, et al. Mutational landscape of canine B-cell lymphoma profiled at single nucleotide resolution by RNA-seq. *PLoS One*. (2019) 14:e0215154. doi: 10.1371/journal.pone.0215154

35. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). World Wide Web. Available at: https://omim.org/ (Accessed December 5, 2016).

36. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res*. (2017) 45:D896–901. doi: 10.1093/nar/gkw1133

37. NCBI Resource Coordinators. Database resources of the National Center for biotechnology information. *Nucleic Acids Res*. (2018) 46:D8–D13. doi: 10.1093/nar/gkx1095

38. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res*. (1999) 9:677–9. doi: 10.1101/gr.9.8.677

39. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*. (2013) 14:144–61. doi: 10.1093/bib/bbs038

40. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. (2000) 302:205–17. doi: 10.1006/jmbi.2000.4042

41. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. (2010) 38:576–89. doi: 10.1016/j.molcel.2010.05.004

42. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, et al. Ensembl 2017. *Nucleic Acids Res*. (2017) 45:D635–42. doi: 10.1093/nar/gkw1104

43. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033

44. Thomas R, Smith KC, Ostrander EA, Galibert F, Breen M. Chromosome aberrations in canine multicentric lymphomas detected with comparative genomic hybridisation and a panel of single locus probes. *Br J Cancer*. (2003) 89:1530–7. doi: 10.1038/sj.bjc.6601275

45. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. (2002) 30:207–10. doi: 10.1093/nar/30.1.207

46. Taher L, Beck J, Liu W, Roolf C, Soller JT, Rütgen BC, et al. Comparative high-resolution transcriptome sequencing of lymphoma cell lines and de novo lymphomas reveals cell-line-specific pathway dysregulation. *Sci Rep*. (2018) 8:6279. doi: 10.1038/s41598-018-23207-7

47. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. (2014) 15:550. doi: 10.1186/s13059-014-0550-8

48. Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix quality control and data analysis. *Bioinformatics*. (2005) 21:3683–5. doi: 10.1093/bioinformatics/bti605

49. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy—analysis ofAffymetrix GeneChipdata at the probe level. *Bioinformatics*. (2004) 20:307–15. doi: 10.1093/bioinformatics/btg405

50. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007

51. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. (2009) 37:W202–8. doi: 10.1093/nar/gkp335

52. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. (2016) 44:D110–5. doi: 10.1093/nar/gkv1176

53. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. *Genome Biol*. (2007) 8:R24. doi: 10.1186/gb-2007-8-2-r24

54. 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. doi: 10.1093/nar/gku1003

55. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, et al. Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*. (2009) 181:1493–505. doi: 10.1534/genetics.108.098830

56. Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, Corey L, et al. Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci U S A*. (2009) 106:13903–8. doi: 10.1073/pnas.0902129106

57. Boyko AR. The domestic dog: man's best friend in the genomic era. *Genome Biol*. (2011) 12:216. doi: 10.1186/gb-2011-12-2-216

58. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol*. (2010) 8:e1000451. doi: 10.1371/journal.pbio.1000451

59. vonHoldt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H, et al. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res*. (2011) 21:1294–305. doi: 10.1101/gr.116301.110

60. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. (2010) 464:898–902. doi: 10.1038/nature08837

61. Serres-Armero A, Povolotskaya IS, Quilez J, Ramirez O, Santpere G, Kuderna LFK, et al. Similar genomic proportions of copy number variation within gray wolves and modern dog breeds inferred from whole genome sequencing. *BMC Genomics*. (2017) 18:977. doi: 10.1186/s12864-017-4318-x

62. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. *Nat Commun*. (2016) 7:10460. doi: 10.1038/ncomms10460

63. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. (2015) 4:7. doi: 10.1186/s13742-015-0047-8

64. Owen LN. Classification of tumors in domestic animals [internet]. (1980). Available at: http://apps.who.int/iris/bitstream/handle/10665/68618/VPH_CMO_80.20_eng.pdf;jsessionid=A47651067E1EDA208CAE09C97BA3227B?sequence=1.

65. Simon D, Nolte I, Eberle N, Abbrederis N, Killich M, Hirschberger J. Treatment of dogs with lymphoma using a 12-week, maintenance-free combination chemotherapy protocol. *J Vet Intern Med*. (2006) 20:948–54. doi: 10.1892/0891-6640(2006)20[948:todwlu]2.0.co;2

66. Vail DM, Michels GM, Khanna C, Selting KA, London CA. Veterinary cooperative oncology group. Response evaluation criteria for peripheral nodal lymphoma in dogs (v1.0)--a veterinary cooperative oncology group (VCOG) consensus document. *Vet Comp Oncol*. (2010) 8:28–37. doi: 10.1111/j.1476-5829.2009.00200.x

67. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. *Nucleic Acids Res*. (2022) 50:D988–95. doi: 10.1093/nar/gkab1049

68. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience [Internet]*. (2021) 10:giab008. doi: 10.1093/gigascience/giab008

69. Erik Garrison GM. Haplotype-based variant detection from short-read sequencing [internet]. (2012). Available at: https://arxiv.org/abs/1207.3907.

70. van der Auwera GA, O'Connor BD. *Genomics in the cloud: Using Docker, GATK, and WDL in Terra*. Sebastopol, CA: O'Reilly Media (2020).

71. Mose LE, Perou CM, Parker JS. Improved indel detection in DNA and RNA via realignment with ABRA2. *Bioinformatics*. (2019) 35:2966–73. doi: 10.1093/bioinformatics/btz033

72. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330

73. Picard [Internet]. Available at: http://broadinstitute.github.io/picard/ (Accessed July 20, 2023).

74. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. *Genome Biol*. (2016) 17:122. doi: 10.1186/s13059-016-0974-4

75. Survival [Internet]. Available at: https://cran.r-project.org/web/packages/survival/survival.pdf (Accessed July 20, 2023).

76. Survminer [Internet]. Available at: https://cran.r-project.org/web/packages/survminer/survminer.pdf (Accessed July 20, 2023).

77. MASS [Internet]. Available at: https://www.stats.ox.ac.uk/pub/MASS4/ (Accessed July 20, 2023).

78. Cox DR. Regression models and life-tables. *J R Stat Soc*. (1972) 34:187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x

79. Collett D. *Modelling survival data in medical research*. New York: Chapman & Hall/CRC Press (2023). p. 556.

80. Oleson L, von Moltke LL, Greenblatt DJ, Court MH. Identification of polymorphisms in the 3′-untranslated region of the human pregnane X receptor (PXR) gene associated with variability in cytochrome P450 3A (CYP3A) metabolism. *Xenobiotica*. (2010) 40:146–62. doi: 10.3109/00498250903420243

81. Revathidevi S, Sudesh R, Vaishnavi V, Kaliyanasundaram M, MaryHelen KG, Sukanya G, et al. Screening for the 3′UTR polymorphism of the PXR gene in south Indian breast Cancer patients and its potential role in pharmacogenomics. *Asian Pac J Cancer Prev*. (2016) 17:3971–7.

82. Skeeles LE, Fleming JL, Mahler KL, Toland AE. The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. *PLoS One*. (2013) 8:e58609. doi: 10.1371/journal.pone.0058609

83. Jiang Y, Chen J, Wu J, Hu Z, Qin Z, Liu X, et al. Evaluation of genetic variants in microRNA biosynthesis genes and risk of breast cancer in Chinese women. *Int J Cancer*. (2013) 133:2216–24. doi: 10.1002/ijc.28237

84. Jiang W, Mitra R, Lin CC, Wang Q, Cheng F, Zhao Z. Systematic dissection of dysregulated transcription factor-miRNA feed-forward loops across tumor types. *Brief Bioinform*. (2016) 17:996–1008. doi: 10.1093/bib/bbv107

85. Zhai K, Chang J, Hu J, Wu C, Lin D. Germline variation in the 3′-untranslated region of the POU2AF1 gene is associated with susceptibility to lymphoma. *Mol Carcinog*. (2017) 56:1945–52. doi: 10.1002/mc.22652

86. Yang B, Liu C, Diao L, Wang C, Guo Z. A polymorphism at the microRNA binding site in the 3′ untranslated region of C14orf101 is associated with non-Hodgkin lymphoma overall survival. *Cancer Genet*. (2014) 207:141–6. doi: 10.1016/j.cancergen.2014.03.007

87. Gascoyne DM, Lyne L, Spearman H, Buffa FM, Soilleux EJ, Banham AH. Vitamin D receptor expression in Plasmablastic lymphoma and myeloma cells confers susceptibility to vitamin D. *Endocrinology*. (2017) 158:503–15. doi: 10.1210/en.2016-1802

88. Kulling PM, Olson KC, Olson TL, Feith DJ, Loughran TP Jr. Vitamin D in hematological disorders and malignancies. *Eur J Haematol*. (2017) 98:187–97. doi: 10.1111/ejh.12818

89. Purdue MP, Hartge P, Davis S, Cerhan JR, Colt JS, Cozen W, et al. Sun exposure, vitamin D receptor gene polymorphisms and risk of non-Hodgkin lymphoma. *Cancer Causes Control*. (2007) 18:989–99. doi: 10.1007/s10552-007-9039-z

90. Smedby KE, Eloranta S, Duvefelt K, Melbye M, Humphreys K, Hjalgrim H, et al. Vitamin D receptor genotypes, ultraviolet radiation exposure, and risk of non-Hodgkin lymphoma. *Am J Epidemiol*. (2011) 173:48–54. doi: 10.1093/aje/kwq340

91. Kelly JL, Drake MT, Fredericksen ZS, Asmann YW, Liebow M, Shanafelt TD, et al. Early life sun exposure, vitamin D-related gene variants, and risk of non-Hodgkin lymphoma. *Cancer Causes Control*. (2012) 23:1017–29. doi: 10.1007/s10552-012-9967-0

92. Purdue MP, Lan Q, Kricker A, Vajdic CM, Rothman N, Armstrong BK. Vitamin D receptor gene polymorphisms and risk of non-Hodgkin's lymphoma. *Haematologica*. (2007) 92:1145–6. doi: 10.3324/haematol.11053

93. Pezeshki SMS, Asnafi AA, Khosravi A, Shahjahani M, Azizidoost S, Shahrabi S. Vitamin D and its receptor polymorphisms: new possible prognostic biomarkers in leukemias. *Oncol Rev*. (2018) 12:366. doi: 10.4081/oncol.2018.366

94. Gandini S, Gnagnarella P, Serrano D, Pasquali E, Raimondi S. Vitamin D receptor polymorphisms and cancer. *Adv Exp Med Biol*. (2014) 810:69–105. doi: 10.1007/978-1-4939-0437-2_5

95. Helsen J, Frickel J, Jelier R, Verstrepen KJ. Network hubs affect evolvability. *PLoS Biol*. (2019) 17:e3000111. doi: 10.1371/journal.pbio.3000111

96. Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med*. (2020) 12:91. doi: 10.1186/s13073-020-00791-w

97. Hwang S, Kim E, Lee I, Marcotte EM. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep*. (2015) 5:17875. doi: 10.1038/srep17875

98. Lefouili M, Nam K. The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci Rep*. (2022) 12:11331. doi: 10.1038/s41598-022-15563-2

99. Yao Z, You FM, N'Diaye A, Knox RE, McCartney C, Hiebert CW, et al. Evaluation of variant calling tools for large plant genome re-sequencing. *BMC Bioinformatics*. (2020) 21:360. doi: 10.1186/s12859-020-03704-1

100. Stegemiller MR, Redden RR, Notter DR, Taylor T, Taylor JB, Cockett NE, et al. Using whole genome sequence to compare variant callers and breed differences of US sheep. *Front Genet*. (2022) 13:1060882. doi: 10.3389/fgene.2022.1060882

101. Pei XM, Yeung MHY, Wong ANN, Tsang HF, Yu ACS, Yim AKY, et al. Targeted sequencing approach and its clinical applications for the molecular diagnosis of human diseases. *Cells [Internet]*. (2023) 12:493. doi: 10.3390/cells12030493

102. Sirivisoot S, Kasantikul T, Techangamsuwan S, Radtanakatikanon A, Chen K, Lin TY, et al. Evaluation of 41 single nucleotide polymorphisms in canine diffuse large B-cell lymphomas using MassARRAY. *Sci Rep*. (2022) 12:5120. doi: 10.1038/s41598-022-09112-0

103. Lim ECP, Brett M, Lai AHM, Lee SP, Tan ES, Jamuar SS, et al. Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients. *Hum Genomics*. (2015) 9:33. doi: 10.1186/s40246-015-0055-x

104. Lih CJ, Sims DJ, Harrington RD, Polley EC, Zhao Y, Mehaffey MG, et al. Analytical validation and application of a targeted next-generation sequencing mutation-detection assay for use in treatment assignment in the NCI-MPACT trial. *J Mol Diagn*. (2016) 18:51–67. doi: 10.1016/j.jmoldx.2015.07.006

105. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. (2018) 46:D1062–7. doi: 10.1093/nar/gkx1153

106. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in Cancer. *Nucleic Acids Res*. (2019) 47:D941–7. doi: 10.1093/nar/gky1015

107. Lappalainen I, Lopez J, Skipper L, Hefferon T, Spalding JD, Garner J, et al. DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res*. (2013) 41:D936–41. doi: 10.1093/nar/gks1213

108. Elvers I, Turner-Maier J, Swofford R, Koltookian M, Johnson J, Stewart C, et al. Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res*. (2015) 25:1634–45. doi: 10.1101/gr.194449.115

109. Schiffman JD, Breen M. Comparative oncology: what dogs and other species can teach us about humans with cancer. *Philos Trans R Soc Lond B Biol Sci [Internet]*. (2015) 370:20140231. doi: 10.1098/rstb.2014.0231