THE KENNEL CLUB GENETICS CENTRE

Flatcoated Retriever research update report – March 2022

Histiocytic Sarcoma project

In January 2021 researchers in the United States published the details of two regions of the genome (called loci) that are associated with histiocytic sarcoma (HS) in Flatcoated Retrievers (FCRs) (Evans et al. 2021). This means that some dogs carry genetic variants in these regions that increase their risk of developing HS. Although the precise variants that confer risk have not been confirmed yet, identifying regions of the genome where risk variants are located is an important first step in understanding the genetic basis of HS better.

One of the loci is on chromosome 5, and overlaps with susceptibility loci for two hematopoietic cancers, hemangiosarcoma and B-cell lymphoma, in the closely related golden retriever breed. The other locus is on chromosome 19 and is unique to the FCR. Together the authors report that these two loci account for about 35% of all disease risk which, if confirmed, is an extremely high value.

Researchers from the Kennel Club Genetics Centre (KCGC) have agreed to collaborate with the paper's authors, Dr Jacquelyn Evan and Dr Elaine Ostrander to achieve two objectives:

1. Confirm that the two loci described in the paper are indeed associated with HS risk in an independent cohort of FCRs. This is standard procedure for research studies investigating the genetics of genetically complex diseases such as cancer. DNA from a new cohort of FCRs with and without HS (details below) will be genotyped for markers within the two risk loci and statistical analysis will be used to determine whether markers on the risk *versions* of chromosome 5 and chromosome 19 are found in cases more often that controls. These risk *versions* are known as risk *alleles*.
2. Calculate the frequency of the risk alleles in FCRs from the UK. Assuming the association is confirmed (1 above) then knowing the frequency of the risk alleles will be essential before we can begin to consider developing a breeding tool based on either or both of these risk loci. If the risk alleles are very common within the UK FCR population it will be much more difficult to apply selective pressure against them, and reduce their frequency, than if they are less common.

Study cohort

Researchers from the KCGC are working with Jane Dobson and Anna Hollis from the Queen's Veterinary Hospital at the University of Cambridge to undertake this project. Between us we have identified a cohort of around 30 new HS cases and over 70 potential controls that have not yet been genotyped for the two risk loci described by Evans et al. These samples are from dogs with a confirmed diagnosis of HS that were patients of Dr Dobson or whose DNA was submitted directly to the KCGC, or who were reported by their owners to be free from HS over the age of 10. As part of this study, we will shortly be contacting the owners of the potential controls to confirm that these dogs did not in fact develop HS during their lifetimes. The KCGC will extract DNA from all of the above samples and send it to Dr Jacquelyn Evans, who worked within Dr Elaine Ostrander at the time the original findings were published but who has now established her own laboratory at Cornell University. Jacquelyn will genotype all the new cases and controls for markers within the risk alleles

to (i) confirm association with HS and (ii) determine the frequency of the risk alleles within the new cohort of FCRs.

Timeline

The DNA extraction and follow up of controls will start imminently and it is expected that the genotyping will take place over the summer, with preliminary results anticipated by early Autumn.

Acknowledgements

We would like to thank Luna's Legacy for providing vital funds to support this research. We would like to thank all the Flatcoated retriever owners who have contributed information, samples from their dogs and funds over the years – we wouldn't be in a position to do any of our research without you.

For all enquiries, please contact Cathryn Mellersh (cm935@cam.ac.uk) or Bryan McLaughlin (bryan.mclaughlin@thekennelclub.org.uk
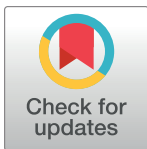
End

RESEARCH ARTICLE

Multi-omics approach identifies germline regulatory variants associated with hematopoietic malignancies in retriever dog breeds

Jacquelyn M. Evans[1], **Heidi G. Parker**[1], **Gerard R. Rutteman**[2], **Jocelyn Plassais**[1], **Guy C. M. Grinwis**[3], **Alexander C. Harris**[1], **Susan E. Lana**[4], **Elaine A. Ostrander**[1] *

1 Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, 2 Department of Clinical Sciences, division Internal Medicine of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands, 3 Department Biomedical Health Sciences, division Pathology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands, 4 College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, Colorado, United States of America

* eostrand@mail.nih.gov

Abstract

Histiocytic sarcoma is an aggressive hematopoietic malignancy of mature tissue histiocytes with a poorly understood etiology in humans. A histologically and clinically similar counterpart affects flat-coated retrievers (FCRs) at unusually high frequency, with 20% developing the lethal disease. The similar clinical presentation combined with the closed population structure of dogs, leading to high genetic homogeneity, makes dogs an excellent model for genetic studies of cancer susceptibility. To determine the genetic risk factors underlying histiocytic sarcoma in FCRs, we conducted multiple genome-wide association studies (GWASs), identifying two loci that confer significant risk on canine chromosomes (CFA) 5 (P_{wald} = 4.83x10^{-9}) and 19 (P_{wald} = 2.25x10^{-7}). We subsequently undertook a multi-omics approach that has been largely unexplored in the canine model to interrogate these regions, generating whole genome, transcriptome, and chromatin immunoprecipitation sequencing. These data highlight the PI3K pathway gene *PIK3R6* on CFA5, and proximal candidate regulatory variants that are strongly associated with histiocytic sarcoma and predicted to impact transcription factor binding. The CFA5 association colocalizes with susceptibility loci for two hematopoietic malignancies, hemangiosarcoma and B-cell lymphoma, in the closely related golden retriever breed, revealing the risk contribution this single locus makes to multiple hematological cancers. By comparison, the CFA19 locus is unique to the FCR and harbors risk alleles associated with upregulation of *TNFAIP6*, which itself affects cell migration and metastasis. Together, these loci explain ~35% of disease risk, an exceptionally high value that demonstrates the advantages of domestic dogs for complex trait mapping and genetic studies of cancer susceptibility.

Author summary

We have identified two regions of the canine genome that explain a striking 35% of risk
for developing histiocytic sarcoma in FCRs. The disease is uniformly lethal, affects 20% of
FCRs, and parallels a cancer of the same name in humans. Both regions harbor genes
involved in cell migration and cancer-related pathways. The first includes variants in reg-
ulatory regions at the tumor suppressor *PIK3R6* locus that are strongly associated with
histiocytic sarcoma and likely confer risk for other hematopoietic cancers. FCRs with risk
alleles at the second locus demonstrate increased expression of *TNFAIP6*, which correlates
with poor prognosis in multiple human cancers. In identifying genomic differences
between affected and unaffected dogs, we advance our understanding of both canine and
human health biology and set the stage for the development of diagnostic and therapeutic
strategies.

Introduction

Histiocytic sarcoma is a rare, aggressive cancer of dendritic cells and macrophages that
accounts for < 1% of hematopoietic malignancies in humans [1,2]. Tumors arise as the pri-
mary neoplasm or concurrently with other hematological malignancies, such as lymphoma or
chronic lymphocytic leukemia, through transdifferentiation or a common neoplastic precursor
[3]. The disease is diagnosed most frequently in adulthood and may present as localized or dis-
seminated, with tumors in multiple sites, including the spleen, liver, lymph nodes, gastrointes-
tinal tract, and skin [2]. The neoplastic cells are typically large and round-polygonal in shape;
spindle cells may also be present [2]. Treatment response is poor, and most patients succumb
to the disease within two years [2]. Limited biological samples for this rare cancer have hin-
dered large-scale genetic studies.

A histologically and clinically similar disease, also termed histiocytic sarcoma, occurs spon-
taneously in dogs [4,5]. Although rare across breeds as a whole, histiocytic sarcoma is common
in flat-coated retrievers (FCRs) and Bernese mountain dogs, affecting ~20% and 25% of dogs
in each breed with near uniform fatality [6–8]. The canine disease also presents as localized in
periarticular tissue or disseminated in the viscera, with the former more common in the FCR
[9] and the latter typical in the Bernese mountain dog [7].

Dogs are a well-described, naturally-occurring model for many human cancers including
sarcomas and hematological cancers, such as osteosarcoma, lymphoma, and leukemia [10–12].
Most breeds were developed within the last 200 years [13], and population bottlenecks coupled
with strong selection for morphological and behavioral traits has created a unique population
structure characterized by reduced genetic diversity and long haplotype blocks within breeds,
but also substantial across-breed variation [14,15]. Detrimental alleles have become enriched
as a consequence of breed formation processes, leading to disease predisposition. These char-
acteristics facilitate complex trait mapping studies in dogs, which require roughly two orders
of magnitude fewer markers compared to human GWASs and as little as 200 individuals [13–
15].

The FCR and Bernese mountain dog are ideal for gaining insight into the etiology of histio-
cytic sarcoma, as the unusually strong breed prevalence suggests a highly penetrant heritable
component. While the two breeds do not share recent common ancestors and they were bred
for distinct physical attributes [16], the similarity of disease progression and outcomes, as his-
tiocytic sarcoma always progresses to metastatic disease with full lethality, is the same between
the breeds. The overall rarity of the disease, and its total absence in the majority of domestic

dog breeds, argues for at least partially overlapping genetic risk factors, likely reflecting the explosion of modern breeds in Western Europe <200 years ago. However, as the primary differences between affected dogs appear to be in initial disease presentation, it is likely that at least partially independent genetic mechanisms underly histiocytic sarcoma susceptibility in each. While our previous GWAS in Bernese mountain dogs successfully identified a locus on canine chromosome (CFA) 11 [17], the FCR genetic predisposition has been hitherto unexplored.

Here, we conducted multiple GWASs in FCRs, identifying two loci on CFA5 and 19 that confer risk for histiocytic sarcoma. We advance beyond previous cancer and complex disease studies in the dog, analyzing whole genome, transcriptome, and ChIP sequencing data to identify putative regulatory variants associated with histiocytic sarcoma susceptibility in FCRs that may also confer risk for hematopoietic cancers in other breeds. We thus leverage the advantages of the canine model system to further our understanding of the biology of a rare human cancer with implications for both canine and human health.

Results

CFA5 and CFA19 confer risk for histiocytic sarcoma

A GWAS including 177 FCR histiocytic sarcoma cases and 132 FCR controls (Table 1) was performed using 108 084 SNPs. Principal component analysis revealed stratification between FCRs of European vs. North American origin (Fig 1). GWAS was performed in GEMMA [18] using a kinship matrix and linear mixed model to correct for population structure with a genomic inflation factor (λ) of 0.97. A single association exceeding Bonferroni significance (4.26×10^{-7}) was identified on CFA5 with $P_{wald} = 4.83 \times 10^{-9}$ (Fig 2A and S1 Table). The top 27 markers are in high linkage disequilibrium (LD; r2$\geq$0.8) and span a 4.3 Mb region around the lead SNP (CFA5:33001550; Fig 2B). A shared haplotype was identified among 90% of cases, with recombination events defining a narrower 1.2 Mb FCR risk haplotype at CFA5:32389061–33633274 (Fig 2C and S2 Table), which harbors over 40 genes. The risk haplotype is also present in 64% of control dogs. There is considerable LD among markers at this locus ($r^2 \geq 0.6$ 28–37 Mb) with the broader GWAS signal extending to 28 Mb, and many cases continue to share a common haplotype throughout the region.

The CFA5 risk haplotype is present in the heterozygous state in 53% of cases and 43% of controls. To determine whether additional loci differentiate these groups, we performed a

Table 1. Description of samples used in each analysis.

Analysis	Samples
FCR histiocytic sarcoma GWAS #1	177 FCR cases, 132 FCR controls
FCR histiocytic sarcoma GWAS #2—heterozygotes for CFA5 risk	94 FCR cases, 43 FCR controls (subset of GWAS #1)
RNA-seq analyses	11 FCR blood samples
H3K4me1 and H3K4me3 ChIP-seq	7 canine blood samples
Whole genome sequence filtering	3 FCR histiocytic sarcoma cases, 1 FCR control, 4 golden retriever B-cell lymphoma cases[a]
Candidate variant genotyping—histiocytic sarcoma association	Minimum of 79 cases and 69 controls from the GWAS cohort
Variant genotyping—retrievers with hematopoietic cancer	FCR lymphoma cases (n = 20); golden retriever B-cell lymphoma (n = 9) and histiocytic sarcoma (n = 21) cases

[a]Publicly-available genomes from [19]

https://doi.org/10.1371/journal.pgen.1009543.t001
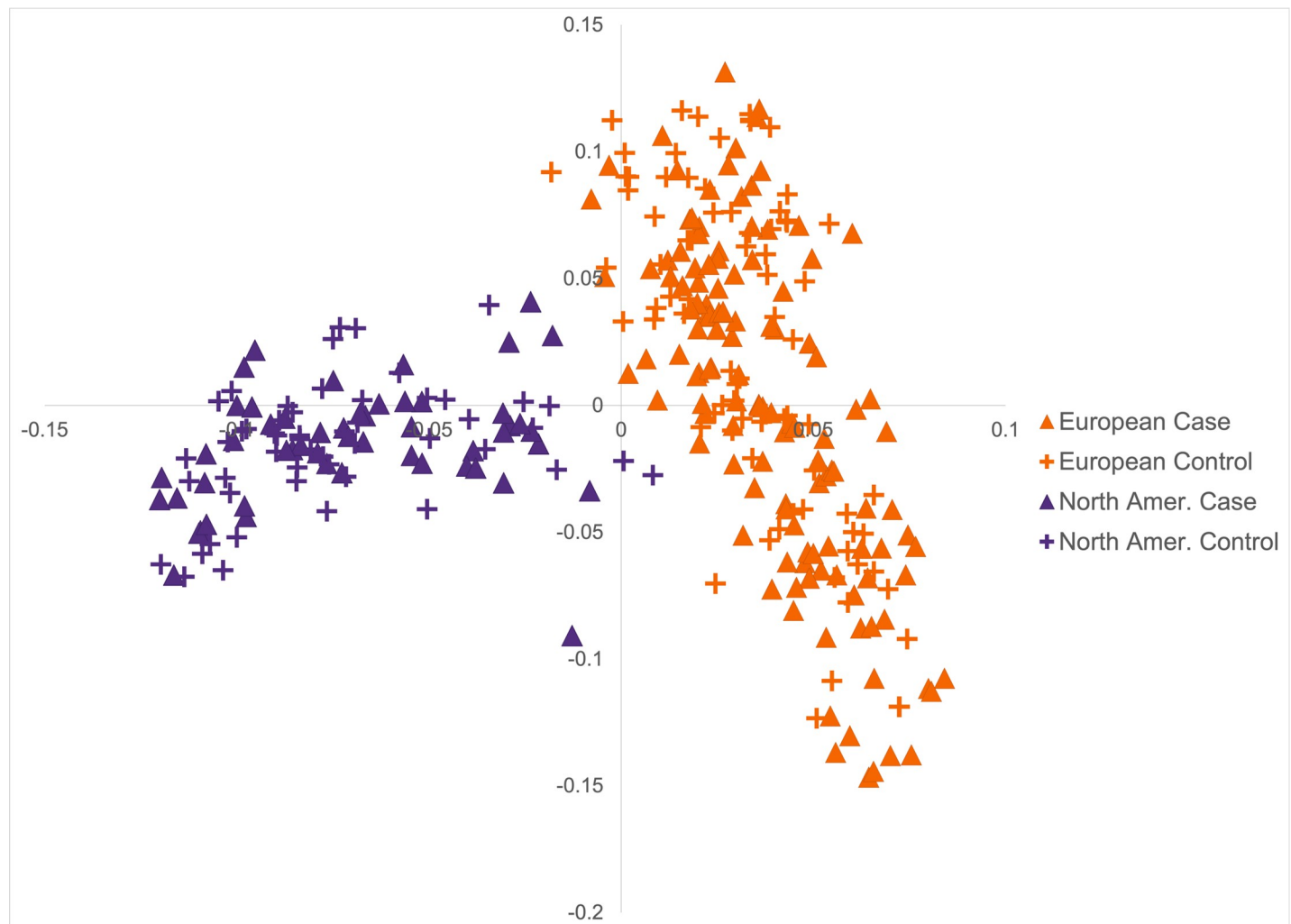
Fig 1. Principal components analysis of FCR GWAS. Principal components 1 (13.6% variance) and 2 (6.6% variance) are plotted on the x and y-axes, respectively. The European and North American FCRs (n = 309) form subpopulations with cases and controls distributed throughout both groups.

https://doi.org/10.1371/journal.pgen.1009543.g001

GWAS using only cases and controls heterozygous for the CFA5 risk haplotype (94 vs. 43; Fig 3 and Table 1), thereby neutralizing the effect of the CFA5 locus. To reduce the possibility that our control group contained dogs who could eventually develop histiocytic sarcoma, we applied a more stringent minimum age at collection for controls (11 years), discarding samples from dogs in the lowest age quartile. This provided further separation between controls and cases, as 75% of cases were diagnosed at <10 years of age, while preserving as much power as possible for the GWAS (S1 Fig and S3 Table). A single locus at 52 Mb on CFA19 (P_{wald} = $2.25\text{x}10^{-7}$) exceeded Bonferroni significance ($4.67\text{x}10^{-7}$) and was confirmed after permutations ($P_{\text{permutations}}$ = 0.014). This approach produced a more robust association compared to that which includes CFA5 genotypes as a covariate in the total GWAS cohort (CFA19:52487724 P_{wald} = $4.25\text{x}10^{-5}$; S4 Table). A 741 kb critical interval is demarcated by the flanking SNPs in highest LD with the lead SNP (CFA19:52487724, $r^2 \geq 0.6$), encompassing just three genes. Ninety-nine of the 177 cases in the total GWAS cohort (n = 309) had periarticular tumors and 77 had tumors in other locations at the time of diagnosis. The risk allele at the CFA19 locus was more common among periarticular cases (P_{Fisher} = 0.015, OR = 2.78, 95%CI = 1.21–6.37).
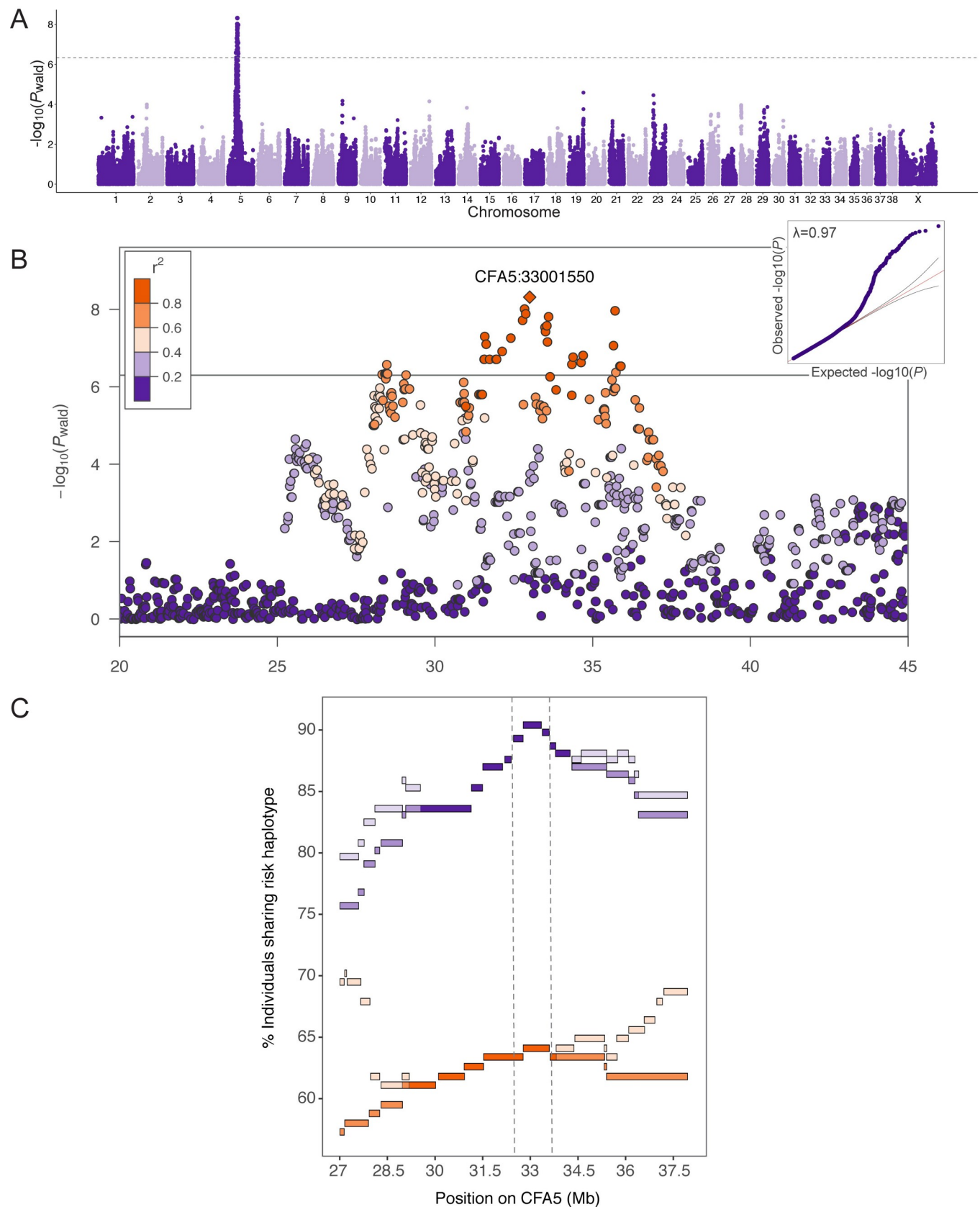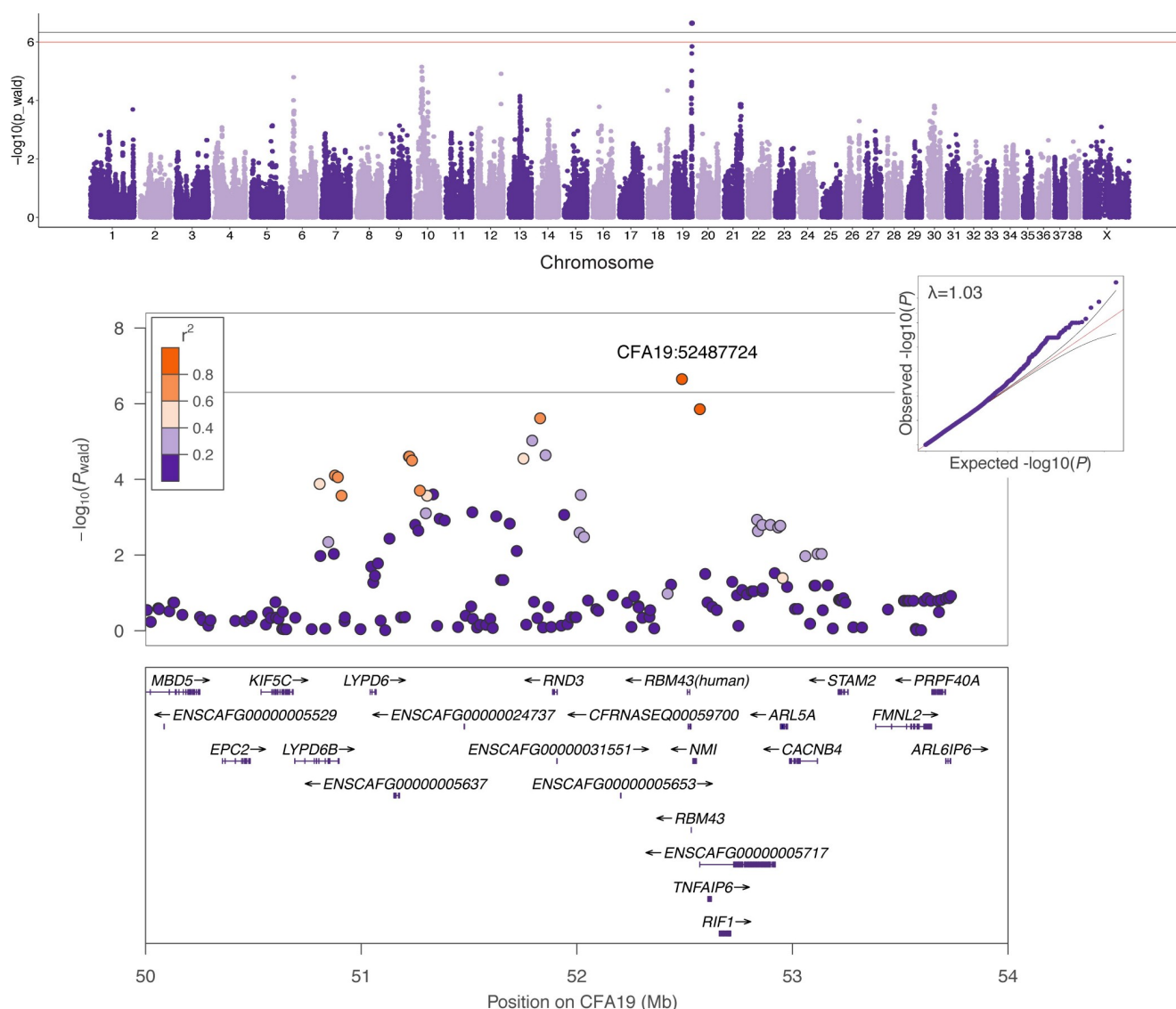
Fig 2. Genome-wide association study results for 177 histiocytic sarcoma FCR cases and 132 controls. A) Manhattan plot of -log$_{10}$P-values (y-axis) for 108084 Illumina SNPs plotted against chromosome position in CanFam3.1 (x-axis). The Bonferroni threshold is plotted on the y-axis in gray (-log$_{10}$P = 6.33). B) Regional Manhattan plot of the CFA5 association with SNPs color-coded according to pairwise LD (r^2) with the lead SNP. C) Length of risk haplotype sharing among cases (purple) and controls (orange) is plotted on the x-axis with the percentage of dogs sharing on the y-axis. Continuous loss of haplotype sharing is tracked in darker purple/orange, while the lighter shades mark points at which some individuals re-gain the common risk haplotype.

In the total cohort (n = 309), SNP genotypes from all autosomes explain 27% ± 14% P_{LRT} = 0.0034) of the risk for developing histiocytic sarcoma. The CFA5:25-40Mb locus alone explains 22% ± 13% P_{LRT} = 1.15x10^{-5}, while CFA19:50.5-53Mb explains 8% ± 5% P_{LRT} = 2.05x10^{-4}. Together, these loci account for 35–37% ± 13% of the phenotypic variance (P_{LRT} = 1.44x10^{-8}). When considering CFA5 and CFA19 genotypes in combination, 39% of dogs who are

Fig 3. Genome-wide association study results for 94 histiocytic sarcoma FCR cases vs. 43 controls. All cases and controls were heterozygous for the CFA5 risk haplotype. Manhattan plot of -log$_{10}$P-values (y-axis) for 107102 SNPs by chromosome position in CanFam3.1 (x-axis) is shown at the top. The Bonferroni and 5% permutations thresholds are plotted as gray and red lines, respectively. QQ plot with genomic inflation factor (λ) and regional Manhattan plot of CFA19 locus, showing pairwise LD (r^2) relative to the lead SNP are below with genes in the region plotted at the bottom.

Table 2. Genotypic combinations for CFA5 and CFA19 risk loci in FCR cases and controls.

Genotype[a]	Cases (n = 177)	Controls (n = 98)[b]	Penetrance (%)[c]
aabb	0	1	-
aaBb	3	7	30
aaBB[d]	14	29	33
Aabb	0	5	-
AaBb	11	17	39
AaBB[d]	83	21	80
AAbb	1	0	-
AABb	11	6	65
AABB[d]	54	12	82

[a]"A" represents the risk haplotype at CFA5; "a" is any non-risk haplotype. "B" represents the risk allele at the CFA19 lead SNP; "b" is non-risk.

[b]Controls were ≥11 years of age at sample collection.

[c]The proportion of dogs with each genotype who are cases.

[d]These genotype combinations were present in FCRs used for WGS variant filtration.

https://doi.org/10.1371/journal.pgen.1009543.t002

heterozygous at both loci are cases, whereas 80% of dogs heterozygous at CFA5 and homozygous at CFA19 are cases (Table 2). Thus, when CFA19 data are included, we observe greatly improved separation of cases and controls relative to analysis with CFA5 genotypes alone.

Multiple hematopoietic malignancies are associated with CFA5 locus

The CFA5 region colocalizes with previously-identified associations for two common hematological malignancies in golden retrievers: hemangiosarcoma (29 Mb) and B-cell lymphoma (34 Mb; Fig 4A) [19]. Although distinct cancers, histiocytic sarcoma, B-cell lymphoma, and hemangiosarcoma all arise from cells in the hematopoietic stem cell pathway: dendritic cells and macrophages, B lymphocytes, and hematopoietic precursor cells, respectively [4,8,20,21]. FCRs and golden retrievers are closely related breeds, sharing an immediate common ancestor among the retriever phylogenetic clade [16]. To search for shared risk haplotypes at this locus, we examined published genotypes [19] from golden retrievers diagnosed with hemangiosarcoma or B-cell lymphoma. Using the same haplotype analysis applied to FCRs (See Methods), we defined a 1.4 Mb B-cell lymphoma risk haplotype (CFA5:33001663–34362236) encompassing the lead golden retriever GWAS SNP (CFA5:34117726) for this cancer (S2 Table). This haplotype overlaps the FCR risk haplotype for 631 kb (Fig 4B), and the interval is strongly associated with hematopoietic cancer in both breeds, with a combined P-value of 4.17×10^{-10} compared to 3.43×10^{-7} and 2.00×10^{-4} in FCRs and golden retrievers alone, respectively. Direct overlap between golden retriever and FCR haplotypes was not observed at the 29 Mb hemangiosarcoma risk locus.

RNA-seq and allele-specific expression

To investigate potential effects of the CFA5 risk haplotype on gene expression, RNA-seq data were generated from RNA isolated from 11 FCR whole blood samples (Table 1). Differential expression analysis was based on the risk haplotype, comparing four dogs who were homozygous for the risk haplotype vs. seven dogs who were heterozygous. The frequency of the risk allele in the FCR control population indicates the difficulty in finding homozygous non-risk individuals; however, this would clearly be beneficial in future expression studies. Forty-three genes and five non-coding RNAs demonstrated significant differential expression. The nearest gene to the CFA5 critical interval, *NLRP1*, was 1.7Mb upstream, suggesting the risk locus may

have distal effects (S5 Table). When comparing gene expression levels in individual samples to the average expression across controls (see Methods, [22]), seven genes and one lncRNA demonstrated significant individual expression (z-score≥|2.5|) among cases. After excluding one heterozygous dog who received chemotherapy one week prior to the blood draw, comparison of the four dogs homozygous for the CFA5 risk haplotype to the remaining six heterozygous dogs revealed an additional 17 genes or non-coding RNAs with significant differential and individual expression (S5 Table).

Because RNA samples were only available from a small number of homozygous and heterozygous individuals and no dogs without risk alleles, the power to detect changes in gene regulation through differential expression analyses was limited. Allele-specific expression (ASE) analysis provides an alternative approach to investigate differential expression utilizing heterozygous individuals. ASE compares expression levels for two alleles at a given coding SNP within an individual, which may result from *cis*-regulation by variants in non-coding regions. This controls for sources of error between individuals, like environmental, technical, or *trans*-regulatory effects [23,24]. We performed an ASE analysis for RNA samples isolated from blood for the seven FCRs heterozygous for the CFA5 risk haplotype. We examined genes within 500 kb on either side of the 631 kb shared risk haplotype, extending our search to include potential long-range enhancer-gene interactions [25]. Variants demonstrating significant ASE in two or more FCRs were identified in seven genes: *CD68*, *MPDU1*, *CHD3*, *BORCS6*, *NDEL1*, and *PIK3R6* (Fig 5 and S6 Table). Both *NudE Neurodevelopment Protein 1 Like 1* (*NDEL1*) and *Phosphoinositide-3-kinase regulatory subunit 6* (*PIK3R6*) lie within the minimal 631 kb shared risk haplotype and contain variants demonstrating significant ASE in at least six of seven FCRs. *NDEL1* functions in neuron migration and neurite outgrowth, microtubule organization, and cell signaling. It has been associated with neurodegenerative disease [26] and may play a role in glioblastoma [27]. *PIK3R6* functions in the PI3K/Akt pathway, which is commonly dysregulated in cancer, in leukocytes [28].

As in human cancers, our data show histiocytic sarcoma is not fully explained by one gene or locus. We next tested effects of the CFA19 risk locus on changes in gene expression. We performed differential expression analysis between four cases homozygous for CFA19 risk and four heterozygous unaffected controls. The CFA5 genotypes were matched between the two groups with one dog homozygous for the CFA5 risk haplotype and three heterozygous dogs in each group. Among the top differentially expressed genes was *TNF alpha induced protein 6* (*TNFAIP6*; $P_{adjusted}$ = 0.024), which lies 37 kb downstream of the GWAS susceptibility critical interval. *TNFAIP6* shows a 10.9-fold increased expression in histiocytic sarcoma cases homozygous for the CFA19 haplotype relative to heterozygous individuals. No other differentially expressed genes were proximal to the CFA19 critical interval (S7 Table). Three of the four cases demonstrate significant increased individual expression (z-score = 2.84–7.61, equivalent to *P*<0.01) relative to all controls (n = 7) at *TNFAIP6* (S7 Table). Comparison of the log2 (TPM) expression at this gene for the four cases vs. seven controls indicates significant differential expression (Wilcoxon *P* = 0.024); exclusion of the case who received chemotherapy increases the *P*-value to 0.067 (S2 Fig).

Variant filtering and ChIP-seq analysis

We next sought to identify potential pathogenic variants within the CFA5 631 kb risk haplotype. Using WGS from four FCRs, three cases and one control, we filtered for variants concordant with the risk haplotype (See Table 2). Because FCRs and golden retrievers diagnosed with hematopoietic cancer shared a 631 kb risk haplotype, we hypothesized that they may also share the pathogenic variant(s) on this haplotype. We thus included published WGS from four
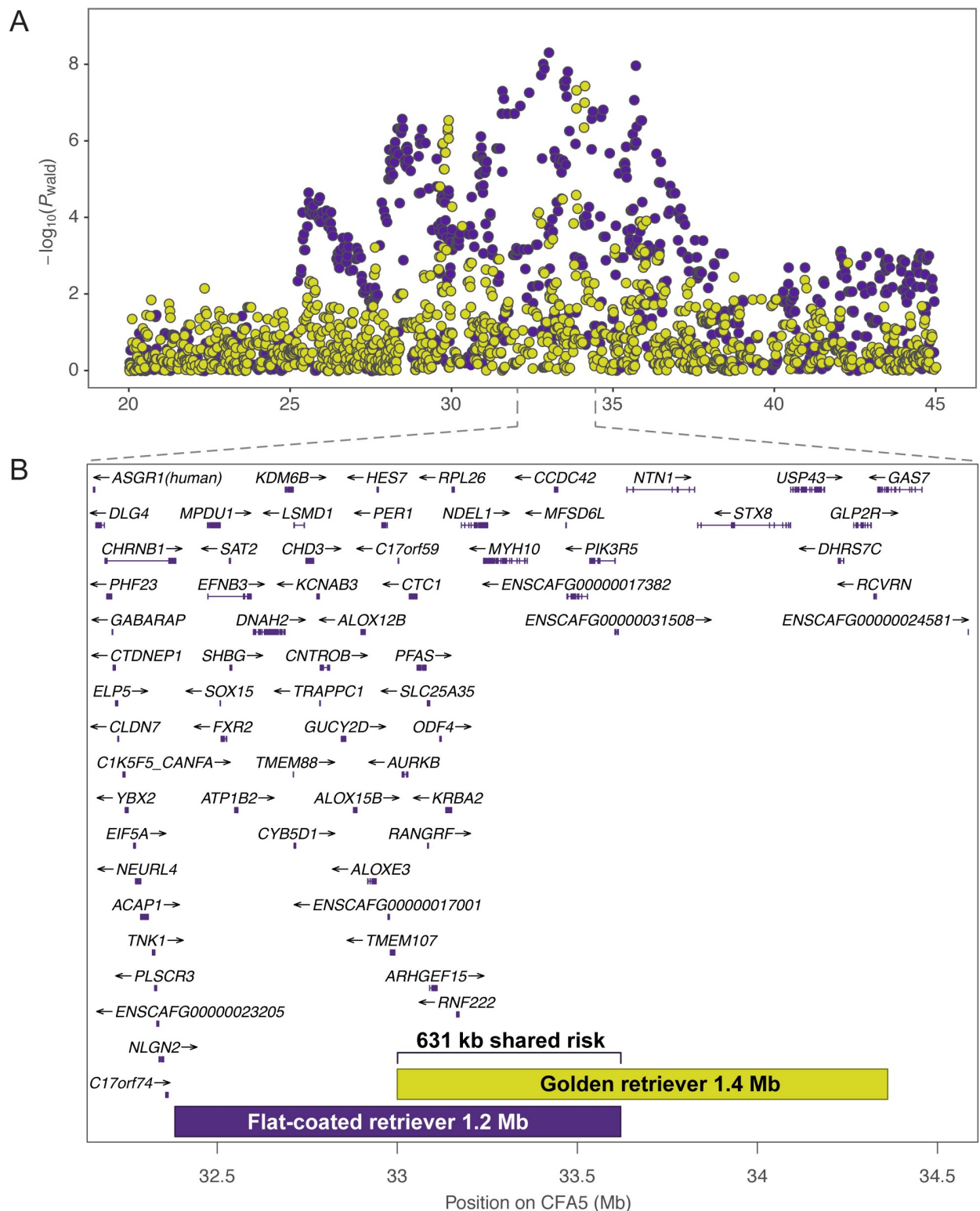
Fig 4. Colocalization of FCR histiocytic sarcoma CFA5 locus with golden retriever hematological malignancy loci. A) Regional Manhattan plot showing FCR histiocytic sarcoma GWAS SNPs in purple. Results of a combined golden retriever GWAS with 142 hemangiosarcoma, 41 B-cell lymphoma, and 172 controls are overlaid in gold. The peaks at 29 Mb and 33 Mb (CanFam3.1) in golden retrievers correspond to hemangiosarcoma and B-cell lymphoma risk, respectively. B) Regions harboring risk haplotypes identified independently in FCRs (purple) with histiocytic sarcoma and golden retrievers (gold) with B-cell lymphoma are plotted with genes in the region. The risk haplotypes overlap for a shared 631 kb span (bracket).

https://doi.org/10.1371/journal.pgen.1009543.g004
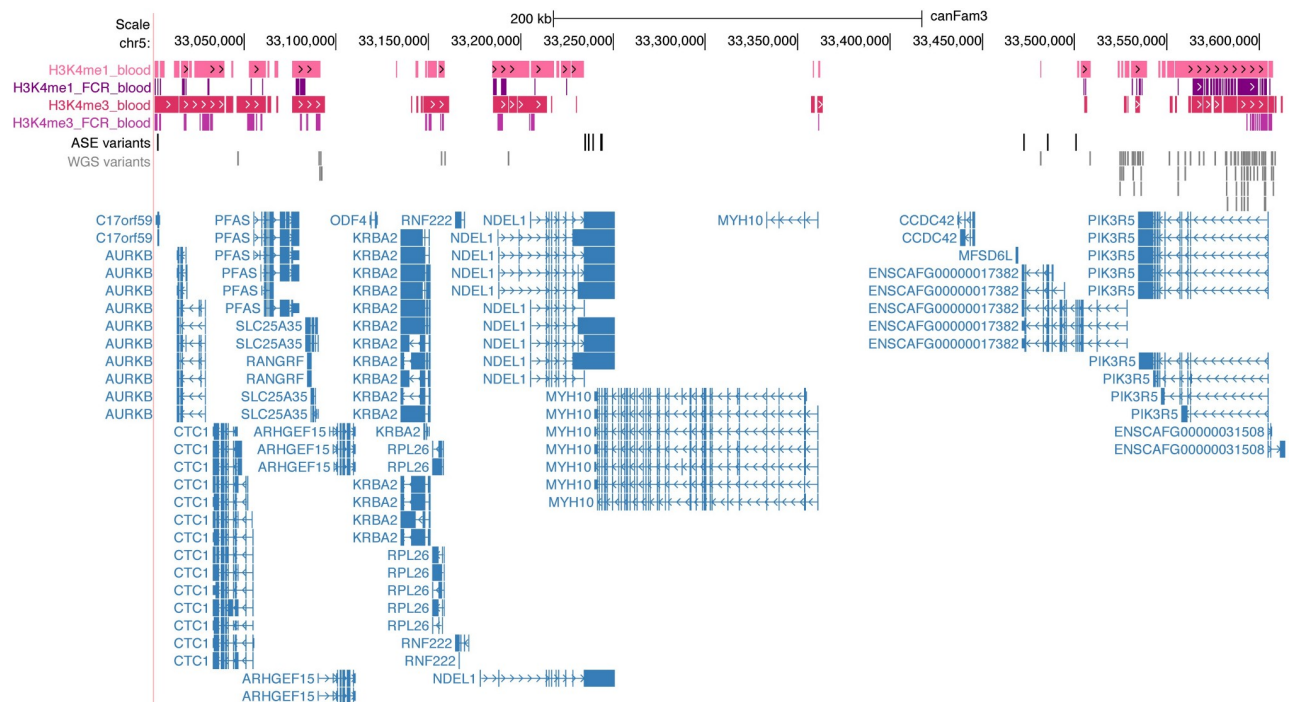
Fig 5. Multi-omics variant analysis. UCSC CanFam3.1 tracks at the 631 kb shared risk haplotype show the blood ChIP-seq regions for H3K4me1 and H3K4me3 for Bernese mountain dogs (pink) and the FCR (purple). ASE variants (black), 98 WGS variants meeting filtering criteria (gray), and the genes in the region are shown below (*ENSCAF00000017382 = PIK3R6*).

https://doi.org/10.1371/journal.pgen.1009543.g005

golden retrievers diagnosed with B-cell lymphoma (three heterozygous for the risk haplotype and one homozygous) for filtering (Table 1). A total of 284 variants matched the segregation pattern of the CFA5 risk haplotype in the four FCR and four golden retriever WGS. A conservative allele frequency threshold of 50% in 1090 genomes from 233 other breed dogs (S8 Table) was applied to eliminate variants common across many breeds, resulting in 218 variants, none of which were unique to FCRs and golden retrievers (S9 Table). No variants were predicted to impact protein sequence or splice sites. Visual inspection of the interval in Integrative Genomics Viewer [29] revealed no structural variants segregating with the risk haplotype. The CanFam3.1 reference genome contains six gaps, totaling approximately 3 kb, within the critical interval, which may mask variants relevant to histiocytic sarcoma susceptibility.

We next investigated potential regulatory variants, which are not fully annotated in the CanFam3.1 reference. To identify promoter and enhancer regions in canine cell types relevant to cancers investigated herein, ChIP-seq data from peripheral blood mononuclear cells from seven dogs were generated for two histone marks, H3K4me1 and H3K4me3, to identify canine promoters and enhancers (See Methods, S10 Table). Publicly available ATAC-seq data identifying open chromatin regions from multiple canine tissues, i.e. spleen, lymph node, and bone marrow [30], were combined with blood ChIP-seq data to define regulatory regions.

Of the selected 218 variants (AF<50%) within the 631 kb critical interval, 98 overlapped with ChIP-seq and/or ATAC-seq regions (Fig 5). As none of the variants were completely unique to the FCR or golden retriever, we considered whether there could be combinations of variants private to the breed. The 98 variants were phased, allowing us to generate haplotypes in retriever and spaniel breeds, the latter of which were included because the retriever and spaniel clades share a recent common ancestor [16], yet the spaniel is not at risk for histiocytic sarcoma. Thus, a comparison of haplotypes in the region between the breeds might highlight

combinations of variants that are neutral polymorphisms versus those that are unique to the retriever and possibly pathogenic. However, no blocks of continuous variants were unique to FCRs and golden retrievers. This does not preclude more distal combinations of variants that may be unique to affected individuals; however, it is likely that the causal mutations are present in other breeds.

Transcription factor binding motif analysis and variant genotyping

To further explore candidate pathogenic variants, we selected regulatory regions from blood ChIP-seq data surrounding *NDEL1* and *PIK3R6*, candidates from ASE analysis, to interrogate variants for possible transcription factor binding motif alterations. ATAC-seq regions over-lapped with blood ChIP-seq and were thus included. Regulatory elements at *PIK3R6* and *PIK3R5* contained 92% of the 98 variants overlapping ChIP-seq within the 631 kb shared hap-lotype (Fig 5 and S9 Table). Five of the 98 WGS variants had significant scores in two tran-scription factor (TF) motif programs (See Methods), suggestive of a difference in binding affinity between the FCR risk and non-risk alleles (S11 Table); all were within *PIK3R5/6* regions. An additional variant (CFA5:33528647), significant in one TF binding affinity pro-gram (FIMO, P_{adj} = 0.0067) and demonstrating a difference in *SP1* and *KLF5* binding affinity between risk and non-risk variants in sTRAP (log(P) = 0.5), was chosen for Sanger sequencing because it is within a human *PIK3R6* regulatory region in the GeneHancer promoter- and enhancer-gene interaction database where *SP1* and *KLF5* are reported to bind [31]. Of the six variants selected for genotyping, one lies within a 12 bp G repeat in a GC-rich region, and we were unable to obtain reliable genotypes for this variant in all dogs (CFA5:33531804). The remaining five variants were significantly associated with histiocytic sarcoma (S12 Table). We calculated Fisher's exact P-values for matched genotypes across 79 case and 69 control FCRs (Table 3). Variants at CFA5:33531780 and 33576022 had the lowest P-values of 4.2×10^{-5} and 4.6×10^{-5}, respectively (lead SNP $P = 1.8 \times 10^{-4}$, Table 3), and were located in ChIP-seq regulatory regions upstream of *PIK3R6*.

The 631 kb shared haplotype delineated here is associated with histiocytic sarcoma in FCRs and B-cell lymphoma in golden retrievers, and we hypothesize that it harbors one or more pathogenic variants contributing to susceptibility for both diseases in each breed. Although less frequently than hemangiosarcoma and B-cell lymphoma, which affect 20% and 6% of the breed respectively [32], golden retrievers also develop histiocytic sarcoma (7% of all tumors in the breed) [4,12,33]. Additional genotyping of the five candidate variants within this region, reveals that they are present in ~75% of FCRs with lymphoma (B-cell n = 3, T-cell n = 4, and unspecified subtype n = 13; Table 1), and are in complete LD with the lead histiocytic sarcoma

Table 3. Variants associated with histiocytic sarcoma and predicted to alter transcription factor binding sites.

Position	Variant	P-value[a]	Odds Ratio	95% Confidence Interval	AF in other breeds[b]	Transcription factor	Haplotype with preferred binding
33528647	C>CCG	6.9×10^{-5}	2.63	1.64–4.21	0.29	KLF5, SP1	risk, risk
33531780	G>A	4.2×10^{-5}	2.71	1.69–4.35	0.45	SP1	risk
33576022	GAGAA>GGAAAGAA	4.6×10^{-5}	2.64	1.65–4.23	0.011	EWSR1-FL1	risk
33587141	G>A	1.2×10^{-4}	2.49	1.56–3.99	0.20	IRF1	risk
33594214	G>A	1.1×10^{-4}	2.56	1.60–4.10	0.23	PLAG1, EWSR1-FL1, RREB1	non-risk, risk, non-risk

[a]Fisher's exact P-values in a matched population of 79 case and 69 control FCRs. Lead SNP in the same cohort $P = 1.8 \times 10^{-4}$, OR = 2.48, 95%CI = 1.55–3.97.

[b]WGS from 1066 breed dogs, excluding FCRs, golden retrievers, and Labrador retrievers.

GWAS SNP ($r^2 = 1$), consistent with our hypothesis. Three of the five variants were also in complete LD with this SNP in golden retrievers with B-cell lymphoma (n = 9) or histiocytic sarcoma (n = 21), i.e. CFA5:33576022, 33587141, and 33594214. The remaining two had r^2 values of 0.32 (CFA5:33528647) and 0.16 (CFA5:33531780; S12 Table), indicating that they are not on the risk haplotype in golden retrievers. In aggregate, these data provide strong support that one or more of the three variants located on the CFA5 risk haplotype are likely to confer susceptibility to histiocytic sarcoma and B-cell lymphoma in both retriever breeds.

Discussion

Genetic investigation of histiocytic sarcoma has been limited in humans due to its rarity, and identification of underlying genetic risk factors has not been undertaken. In this study, we leveraged the frequency of this cancer in the FCR breed and, in a GWAS of 309 dogs, identified two loci associated with histiocytic sarcoma on CFA5 and 19, respectively, that explain ~35% of risk. The former colocalizes with risk for two other cancers of hematopoietic origin and contains a shared risk haplotype for histiocytic sarcoma and B-cell lymphoma in FCRs and golden retrievers, respectively. We subsequently applied a multi-omics approach and identified ASE in *PIK3R6* and regulatory variants local to this gene that are predicted to impact TF binding. The second locus, on CFA19, has not been previously associated with cancer risk in dogs and increases risk for histiocytic sarcoma in combination with the CFA5 locus. The CFA19 critical interval lies upstream of *TNFAIP6* whose expression is upregulated in blood samples from FCR cases compared to healthy FCRs. This work reveals a shared germline predisposition for multiple hematopoietic cancers in dogs, reflecting clinical observations in the human literature of concurrent or subsequent histiocytic sarcoma with lymphomas thought to result from a common progenitor cell.

Chromosome 5 was recently identified, among other loci, through a histiocytic sarcoma GWAS in Bernese mountain dogs [34]. In that GWAS, the best-associated SNPs were at 30.4 Mb. The authors noted that inclusion of 13 FCR cases and 15 FCR controls shifts their lead SNP to CFA5:33823740 with a slightly lower *P*-value, and they identify a common haplotype at 33.8–34.3 Mb [34]. By comparison, our data strongly indicate that the primary risk haplotype for histiocytic sarcoma in FCRs is at 33.0–33.6 Mb and includes 18 genes. In our data, the maximum haplotype sharing is observed here among FCR cases (Fig 2C), as well as in the closely-related golden retriever breed, where cases diagnosed with B-cell lymphoma share the same risk haplotype (Fig 4). When genotypes for the CFA5 lead SNP were included as a covariate in our FCR histiocytic sarcoma GWAS, we also observed top SNPs proximal to CFA11:44 Mb and CFA2:29 Mb peaks (S4 Table) as described in the published Bernese mountain dog GWAS [34].

Our ASE RNA-seq analysis at CFA5 highlights two genes, the most compelling of which is *PIK3R6*, which encodes a regulatory subunit of the PI3Kγ heterodimer within the PI3K/Akt pathway. PI3Kγ expression is predominantly restricted to leukocytes [28] where it functions in cell migration, angiogenesis, and immune response [35]. *PIK3R6* has a tumor suppressive role; knockdown increases PI3Kγ signaling and the potential for cells to metastasize [35,36]. The PI3K/Akt pathway is commonly activated in cancers [28], and somatic mutations in *PIK3CD* and *PI3KCA*, which encode other PI3K isoforms, have recently been detected in a subset of primary human histiocytic sarcoma tumors as well as secondary tumors and their co-occurring lymphomas [3,37], although mutations in PI3Kγ genes have not yet been reported. *PIK3R6* was downregulated in B-cell lymphoma tumors from golden retrievers that harbored the CFA5:29 Mb hemangiosarcoma risk haplotype in this breed [19]. Our ASE results suggest that variants on the CFA5:33 Mb haplotype may impact *PIK3R6* regulation.

Our ChIP-seq and WGS analyses identified five candidate variants strongly associated with histiocytic sarcoma in FCRs and predicted to impact TF binding sites proximal to *PIK3R6*. While these variants were also detected in FCRs and golden retrievers with lymphoma and golden retrievers with histiocytic sarcoma, supporting the hypothesis that variants within the shared risk haplotype underly susceptibility to both diseases in the retrievers, we note that sample sizes were small. Genotyping in larger cohorts is necessary to confirm the observations made here.

Among the candidate variants identified herein, CFA5:33576022, located upstream of *PIK3R6*, emerged as particularly strong, demonstrating the lowest allele frequency in 232 breeds that are at low to zero risk for developing histiocytic sarcoma (1%; Table 3). The variant consists of a GAA insertion within a GAAA microsatellite region and creates an ETS DNA binding site containing the core ETS 5'-GGAA-3' motif. The ETS family of transcription factors are involved in tumorigenesis in many cancers, including lymphomas, leukemias, and Ewing sarcoma [38,39]. Functional studies will be necessary to elucidate the precise mechanisms by which the associated variants impact pathogenesis. We note there may be additional variants within the critical interval that contribute to risk through other mechanisms, such as post-translational modifications, or variants that function in a combinatorial fashion.

The CFA5 risk haplotype is sufficiently common among FCRs that the broader chromosomal region may have been under selection at some point. Multiple across-breed GWASs have associated the ~29–33 Mb region with skull shape [40–42], specifically muzzle length and breadth, both important factors in the FCR breed standard, i.e., ideal physical and behavioral breed characteristics [43]. The deleterious cancer alleles may have increased in frequency in the FCR as a hitchhiking event with nearby variants that control the desirable skull shape phenotype.

At the second associated locus, CFA19, RNAseq analysis shows that *TNFAIP6* is upregulated in whole blood of cases homozygous for the CFA19 risk allele. This gene is immediately downstream of the critical interval defined by high LD. *TNFAIP6* is a member of the hyaluronan-binding protein family with roles in inflammation, extracellular matrix stability, and cell migration. Importantly, increased *TNFAIP6* expression is correlated with reduced survival for several cancers [44,45], and it is upregulated in two lymphoma tumor types in humans [46]. *TNFAIP6* is also a biomarker in colorectal cancer patients who have increased expression in peripheral blood cells relative to controls [47]. Interestingly, the risk-associated allele at this locus was more common among cases in our study with periarticular tumors versus tumors at other sites at diagnosis.

At the CFA5 locus alone, similar proportions of cases and controls are heterozygous for the risk haplotype. But homozygosity for the CFA19 risk allele accounts for 88% of these cases (Table 2). The presence of some control dogs who are also homozygous for the CFA19 risk allele and heterozygous at CFA5 is likely due to the age cutoff of 11 years for enrolling controls (upper quartile of case age at diagnosis was 10 years; S1 Fig). While a higher minimum age for controls is ideal, most FCRs die by age 12 of cancer or cardiac, renal, or musculoskeletal disease [6]. Late onset cancers have proven difficult to study in humans, e.g. prostate cancer. Our results demonstrate that it is possible to differentiate reliable case and control groups for late onset cancers in breed dogs, even though assignment to a control group will prove imperfect as the dogs age.

The most unexpected result in this study is the fact that together the CFA5 and 19 risk loci, explain ~35% of risk for histiocytic sarcoma in FCRs. This high value is due to the population structure of domestic dogs. Each breed is a closed population, experiencing frequent bottlenecks and strong artificial selection, resulting in small numbers of variants having large effects on traits. These results also highlight the value of the dog model for studies of complex cancers,

particularly those with a heterogeneous phenotype. In the absence of family-based linkage studies, few mechanisms exist in human cancer genetics to identify risk loci for rare, but lethal, cancers.

Candidate genes at both loci, *PIK3R6* and *TNFAIP6*, have tumor suppressive and metastatic roles in other cancers, and mutations in members of the PI3K and tumor necrosis factor pathways have been identified in human histiocytic sarcoma and lymphoma tumors [3,37]. Our results suggest these pathways are also important in risk for developing these tumors, and that FCRs may be a valuable clinical model for the development of therapies for canine and human histiocytic sarcoma as well as other hematopoietic cancers.

Materials & methods

Ethics statement

Samples were collected with written informed owner consent in accordance with Animal Care and Use Committee guidelines at the collecting institution: National Human Genome Research Institute (NHGRI) Animal Care and Use Committee, GFS-05-1; Utrecht Animal Experiments Committee, as required under Dutch legislation, ID 2007.III.08.110; Colorado State University VTH Clinical Review Board, VCS #2019–227; and Departmental Ethics and Welfare Committee (University of Cambridge, protocol CR44).

Sample collection

DNA was isolated by standard phenol-chloroform protocol from whole blood samples of FCRs diagnosed with histiocytic sarcoma via histopathology or cytology, and FCRs age ≥10 years with no history of cancer (S1 Text). Samples originated from North America and Europe (S13 Table). DNA samples from FCRs diagnosed with lymphoma were included for variant genotyping, as well as golden retrievers with lymphoma or histiocytic sarcoma (S13 Table and S1 Text).

Genome-wide association and haplotype definitions

Genotypes were generated on the Illumina (San Diego, CA, USA) Canine HD 170k SNP array for 177 FCR cases and 132 FCR controls (GEO GSE163784). Association analyses were performed with GEMMA [18]. Illumina Canine HD 170k SNP array genotypes were downloaded for golden retrievers diagnosed with B-cell lymphoma (n = 41) or hemangiosarcoma (n = 143), and 172 who were ≥10 years old and cancer-free at collection [19]. Additional details are in S1 Text. Risk haplotypes were defined from cases having at least one copy of associated alleles at lead GWAS SNPs. Centromeric and telomeric boundaries were delimited by SNPs at which ≥3 cases no longer shared at least one copy of the case-associated allele, and where this pattern extended beyond those boundaries for ≥1 kb. All genomic positions are reported in CanFam3.1.

RNA-sequencing and expression analyses

RNA was isolated from 11 FCR whole blood samples (four histiocytic sarcoma cases and seven healthy, aged controls). For one of the four cases (FCR1), blood was drawn one week after beginning Lomustine chemotherapy treatment. The remaining three samples were collected prior to any treatment. Libraries were prepared using Illumina TruSeq Stranded Total RNA kit with Ribo-zero Globin depletion and sequenced to a minimum of 100 million pairs of 150 bp reads per sample on an Illumina NovaSeq6000 (SRA PRJNA685036). ASEReadCounter was used to determine allele counts for allele-specific expression analysis. Variants showing

significant allele-specific expression (chi-square $P \leq 0.05$) in two or more individuals were pri-
oritized. Read counts determined in RSEM were used for differential expression analyses in
DESeq2 with Benjamini-Hochberg correction [48]. Genes with adjusted $P < 0.05$ and log2fold-
change absolute values greater-than or equal-to 1 were considered significantly over- or
under-expressed. To examine individual expression, z-scores were calculated at each gene by
comparing the individual's expression levels to the mean and standard deviation of the control
group after variance stabilizing transformation of the counts in DESeq2 as described previ-
ously [22]. Z-scores of +/- 2.5 indicate a significant change in expression [22]. See S1 Text. All
sequence data generated herein were aligned to the CanFam3.1 reference genome.

ChIP sequence

Peripheral blood mononuclear cells were extracted by Ficoll (Cytiva, Marlborough, MA, USA)
from fresh blood collected from two FCRs and six Bernese mountain dogs. As only two FCR
samples were available for this experiment, we incorporated data from Bernese mountain dogs
collected at the time of FCR sampling to identify general canine regulatory regions in cells of
hematopoietic origin, which have not yet been fully annotated in the canine reference genome.
Immunoprecipitation was performed for two histone marks, H3K4me1 and H3K4me3 (S1
Text; SRA PRJNA685036). ChIPseq regions identified in the FCR overlapped those in the Ber-
nese mountain dog samples. Publicly-available ATAC-seq data from multiple breeds were also
used to define regulatory regions from relevant tissues, including spleen, lymph nodes, and
bone marrow [30]. Megquier et al. [30] observed that patterns of enrichment for ATAC-seq
sites were similar between individuals, regardless of breed, and different across tissues, as
expected from studies in other species [49,50].

Whole genome sequence and variant filtering

Whole genome resequencing (WGS) data were generated (S1 Text), for five FCR histiocytic
sarcoma cases and three healthy FCR controls ≥ 10 years old (SRA PRJNA448733 and
PRJNA685036). Variants were filtered for concordance with associated risk haplotypes in
FCRs and genomes of four golden retrievers diagnosed with lymphoma [19] (S1 Text). Allele
frequencies were calculated from 1090 publicly available genomes from other breed dogs (S8
Table). Integrative Genomics Viewer (IGV), software for the visualization of genome sequence
[29], was used to scan the CFA5 critical interval for large structural variants not called in the
VCF file. Reads are color-coded in IGV to flag aberrant insert size indicating insertions, dele-
tions, or interchromosomal rearrangements or pair-orientation indicating inversions, duplica-
tions, or translocations.

Transcription factor motif analysis

For a given candidate variant, two fasta files were created, one containing the allele on the risk
haplotype, the other the non-risk allele, each including 30bp on either side of the variant site.
We required significant TF binding affinity predictions in two programs in order to increase
confidence in the potential impact of a given WGS variant on TF binding. The first, Find Indi-
vidual Motif Occurrences (FIMO) [51], scans input sequence for known transcription factor
motifs, assigns a log-likelihood ratio score for each motif based on sequence position, converts
these scores to P-values and q-values. Transcription Factor Binding Affinity Prediction
(sTRAP) [52,53] predicts binding affinity of each transcription factor in a given matrix to
wild-type and mutant sequence, compares affinity values between the two, and calculates
which TF has the greatest difference in affinity resulting from the sequence change. Variants

for which the same motif was identified by both programs as significant in either the risk or non-risk allele sequence were prioritized (S1 Text).

Variant genotyping

Genotyping was accomplished by Sanger sequencing or agarose gel electrophoresis. Primer sequences, thermal cycling, and reaction conditions are in S14 Table. PCR products were sequenced as described previously [54].

Supporting information

S1 Text. Additional materials and methods.
(DOCX)

S1 Fig. FCR GWAS cohort age distribution.
(PDF)

S2 Fig. Boxplots of transcripts per million counts for *TNFAIP6*.
(PDF)

S1 Table. FCR GWAS #1 results.
(XLSX)

S2 Table. Haplotype analysis.
(XLSX)

S3 Table. FCR GWAS #2 results.
(XLSX)

S4 Table. FCR GWAS with CFA5 genotypes as a covariate.
(XLSX)

S5 Table. Differential expression analysis results and individual expression z-scores for CFA5.
(XLSX)

S6 Table. Allele-specific expression results for CFA5.
(XLSX)

S7 Table. Differential expression analysis results and individual expression z-scores for CFA19.
(XLSX)

S8 Table. WGS SRA accession numbers.
(XLSX)

S9 Table. WGS variant filtering.
(XLSX)

S10 Table. Summary of ChIP-Seq read counts per sample.
(XLSX)

S11 Table. Predicted effects of risk variants on transcription factor binding affinity.
(XLSX)

S12 Table. Genotypes at candidate variants in retrievers.
(XLSX)

S13 Table. Cohort information.
(XLSX)

S14 Table. Primer sequences and reaction conditions for variant genotyping.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Jacquelyn M. Evans, Heidi G. Parker, Elaine A. Ostrander.

Data curation: Jocelyn Plassais, Alexander C. Harris.

Formal analysis: Jacquelyn M. Evans, Heidi G. Parker, Jocelyn Plassais.

Funding acquisition: Elaine A. Ostrander.

Investigation: Jacquelyn M. Evans, Heidi G. Parker, Gerard R. Rutteman, Jocelyn Plassais, Guy C. M. Grinwis.

Methodology: Jacquelyn M. Evans, Gerard R. Rutteman, Jocelyn Plassais.

Resources: Gerard R. Rutteman, Guy C. M. Grinwis, Susan E. Lana, Elaine A. Ostrander.

Supervision: Heidi G. Parker, Elaine A. Ostrander.

Validation: Jacquelyn M. Evans, Heidi G. Parker.

Visualization: Jacquelyn M. Evans.

Writing – original draft: Jacquelyn M. Evans, Elaine A. Ostrander.

Writing – review & editing: Jacquelyn M. Evans, Heidi G. Parker, Jocelyn Plassais, Elaine A. Ostrander.

References

1. Skala SL, Lucas DR, Dewar R. Histiocytic sarcoma: Review, discussion of transformation from B-cell lymphoma, and differential diagnosis. Arch Pathol Lab Med. 2018; 142(11):1322–9. https://doi.org/10.5858/arpa.2018-0220-RA PMID: 30407858

2. Takahashi E, Nakamura S. Histiocytic sarcoma: An updated literature review based on the 2008 who classification. J Clin Exp Hematop. 2013; 53(1):1–8. https://doi.org/10.3960/jslrt.53.1 PMID: 23801128

3. Egan C, Lack J, Skarshaug S, Pham TA, Abdullaev Z, Xi L, et al. The mutational landscape of histiocytic sarcoma associated with lymphoid malignancy. Mod Pathol. 2020. https://doi.org/10.1038/s41379-020-00673-x PMID: 32929178

4. Kennedy K, Thomas R, Breen M. Canine histiocytic malignancies-challenges and opportunities. Vet Sci. 2016; 3(1). https://doi.org/10.3390/vetsci3010002 PMID: 29056712

5. Erich SA, Constantino-Casas F, Dobson JM, Teske E. Morphological distinction of histiocytic sarcoma from other tumor types in Bernese mountain dogs and flatcoated retrievers. In Vivo. 2018; 32(1):7–17. https://doi.org/10.21873/invivo.11198 PMID: 29275293

6. Dobson J, Hoather T, McKinley TJ, Wood JL. Mortality in a cohort of flat-coated retrievers in the uk. Vet Comp Oncol. 2009; 7(2):115–21. https://doi.org/10.1111/j.1476-5829.2009.00181.x PMID: 19453365

7. Abadie J, Hedan B, Cadieu E, De Brito C, Devauchelle P, Bourgain C, et al. Epidemiology, pathology, and genetics of histiocytic sarcoma in the Bernese mountain dog breed. J Hered. 2009; 100 Suppl 1: S19–27.

8. Affolter VK, Moore PF. Localized and disseminated histiocytic sarcoma of dendritic cell origin in dogs. Vet Pathol. 2002; 39(1):74–83. https://doi.org/10.1354/vp.39-1-74 PMID: 12102221

9. Constantino-Casas F, Mayhew D, Hoather TM, Dobson JM. The clinical presentation and histopatho-logic-immunohistochemical classification of histiocytic sarcomas in the flat coated retriever. Vet Pathol. 2011; 48(3):764–71. https://doi.org/10.1177/0300985810385153 PMID: 20930108

10. Roode SC, Rotroff D, Avery AC, Suter SE, Bienzle D, Schiffman JD, et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved regions for sub-type differentiation. Chromosome Res. 2015; 23(4):681–708. https://doi.org/10.1007/s10577-015-9475-7 PMID: 26037708

11. Davis BW, Ostrander EA. Domestic dogs and cancer research: A breed-based genomics approach. ILAR J. 2014; 55(1):59–68. https://doi.org/10.1093/ilar/ilu017 PMID: 24936030

12. Dobson JM. Breed-predispositions to cancer in pedigree dogs. ISRN Vet Sci. 2013; 2013:941275. https://doi.org/10.1155/2013/941275 PMID: 23738139

13. Karlsson EK, Lindblad-Toh K. Leader of the pack: Gene mapping in dogs and other model organisms. Nat Rev Genet. 2008; 9(9):713–25. https://doi.org/10.1038/nrg2382 PMID: 18714291

14. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-spe-cific linkage disequilibrium in Canis familiaris. Genome Res. 2004; 14(12):2388–96. https://doi.org/10.1101/gr.3147604 PMID: 15545498

15. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005; 438(7069):803–19. https://doi.org/10.1038/nature04338 PMID: 16341006

16. Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, et al. Genomic analy-ses reveal the influence of geographic origin, migration, and hybridization on modern dog breed devel-opment. Cell Rep. 2017; 19(4):697–708. https://doi.org/10.1016/j.celrep.2017.03.079 PMID: 28445722

17. Shearin AL, Hedan B, Cadieu E, Erich SA, Schmidt EV, Faden DL, et al. The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. Cancer Epidemiol Biomarkers Prev. 2012; 21(7):1019–27. https://doi.org/10.1158/1055-9965.EPI-12-0190-T PMID: 22623710

18. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44(7):821–4. https://doi.org/10.1038/ng.2310 PMID: 22706312

19. Tonomura N, Elvers I, Thomas R, Megquier K, Turner-Maier J, Howald C, et al. Genome-wide associa-tion study identifies shared risk loci common to two malignancies in golden retrievers. PLoS Genet. 2015; 11(2):e1004922. https://doi.org/10.1371/journal.pgen.1004922 PMID: 25642983

20. Lamerato-Kozicki AR, Helm KM, Jubala CM, Cutter GC, Modiano JF. Canine hemangiosarcoma origi-nates from hematopoietic precursors with potential for endothelial differentiation. Exp Hematol. 2006; 34(7):870–8. https://doi.org/10.1016/j.exphem.2006.04.013 PMID: 16797414

21. Avery AC. The genetic and molecular basis for canine models of human leukemia and lymphoma. Front Oncol. 2020; 10:23. https://doi.org/10.3389/fonc.2020.00023 PMID: 32038991

22. Parker HG, Dhawan D, Harris AC, Ramos-Vara JA, Davis BW, Knapp DW, et al. RNAseq expression patterns of canine invasive urothelial carcinoma reveal two distinct tumor clusters and shared regions of dysregulation with human bladder tumors. BMC Cancer. 2020; 20(1):251. https://doi.org/10.1186/s12885-020-06737-0 PMID: 32209086

23. Kang EY, Martin LJ, Mangul S, Isvilanonda W, Zou J, Ben-David E, et al. Discovering single nucleotide polymorphisms regulating human gene expression using allele specific expression from RNA-seq data. Genetics. 2016; 204(3):1057–64. https://doi.org/10.1534/genetics.115.177246 PMID: 27765809

24. Almlof JC, Lundmark P, Lundmark A, Ge B, Maouche S, Goring HH, et al. Powerful identification of cis-regulatory SNPs in human primary monocytes using allele-specific gene expression. PLoS One. 2012; 7(12):e52260. https://doi.org/10.1371/journal.pone.0052260 PMID: 23300628

25. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507(7493):455–61. https://doi.org/10.1038/nature12787 PMID: 24670763

26. Chansard M, Hong JH, Park YU, Park SK, Nguyen MD. Ndel1, Nudel (noodle): Flexible in the cell? Cytoskeleton (Hoboken). 2011; 68(10):540–54. https://doi.org/10.1002/cm.20532 PMID: 21948775

27. Jiang Y, Song Y, Wang R, Hu T, Zhang D, Wang Z, et al. NFAT1-mediated regulation of NDEL1 pro-motes growth and invasion of glioma stem-like cells. Cancer Res. 2019; 79(10):2593–603. https://doi.org/10.1158/0008-5472.CAN-18-3297 PMID: 30940662

28. Thorpe LM, Yuzugullu H, Zhao JJ. Pi3k in cancer: Divergent roles of isoforms, modes of activation and therapeutic targeting. Nat Rev Cancer. 2015; 15(1):7–24. https://doi.org/10.1038/nrc3860 PMID: 25533673

29. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. Cancer Res. 2017; 77(21):e31–e4. https://doi.org/10.1158/0008-5472.CAN-17-0337 PMID: 29092934

30. Megquier K, Genereux DP, Hekman J, Swofford R, Turner-Maier J, Johnson J, et al. Barkbase: Epigenomic annotation of canine genomes. Genes (Basel). 2019; 10(6). https://doi.org/10.3390/genes10060433 PMID: 31181663

31. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. Genehancer: Genome-wide integration of enhancers and target genes in genecards. Database (Oxford). 2017;2017. https://doi.org/10.1093/database/bax028 PMID: 28605766

32. Glickman LT GN, Thorpe R. The golden retriever club of America national health survey Golden Retriever Club of America2000 [Available from: https://grca.org/wp-content/uploads/2015/08/healthsurvey.pdf.

33. Kent MS, Burton JH, Dank G, Bannasch DL, Rebhun RB. Association of cancer-related mortality, age and gonadectomy in golden retriever dogs at a veterinary academic center (1989–2016). PLoS One. 2018; 13(2):e0192578. https://doi.org/10.1371/journal.pone.0192578 PMID: 29408871

34. Hédan B, Cadieu E, Rimbault M, Vaysse A, Dufaure de Citres C., Devauchelle P, et al. Identification of common predisposing loci to hematopoietic cancers in four dog breeds. PLoS Genet. 2021; 17(4): e1009395. https://doi.org/10.1371/journal.pgen.1009395 PMID: 33793571

35. Turvey ME, Klingler-Hoffmann M, Hoffmann P, McColl SR. P84 forms a negative regulatory complex with p110gamma to control pi3kgamma signalling during cell migration. Immunol Cell Biol. 2015; 93 (8):735–43. https://doi.org/10.1038/icb.2015.35 PMID: 25753393

36. Brazzatti JA, Klingler-Hoffmann M, Haylock-Jacobs S, Harata-Lee Y, Niu M, Higgins MD, et al. Differential roles for the p101 and p84 regulatory subunits of pi3kgamma in tumor growth and metastasis. Oncogene. 2012; 31(18):2350–61. https://doi.org/10.1038/onc.2011.414 PMID: 21996737

37. Egan C, Nicolae A, Lack J, Chung HJ, Skarshaug S, Pham TA, et al. Genomic profiling of primary histiocytic sarcoma reveals two molecular subgroups. Haematologica. 2020; 105(4):951–60. https://doi.org/10.3324/haematol.2019.230375 PMID: 31439678

38. Testoni M, Chung EY, Priebe V, Bertoni F. The transcription factor ets1 in lymphomas: Friend or foe? Leuk Lymphoma. 2015; 56(7):1975–80. https://doi.org/10.3109/10428194.2014.981670 PMID: 25363344

39. Grunewald TG, Bernard V, Gilardi-Hebenstreit P, Raynal V, Surdez D, Aynaud MM, et al. Chimeric EWSR1-FLI1 regulates the ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. Nat Genet. 2015; 47(9):1073–8. https://doi.org/10.1038/ng.3363 PMID: 26214589

40. 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(8):e1000451. https://doi.org/10.1371/journal.pbio.1000451 PMID: 20711490

41. Mansour TA, Lucot K, Konopelski SE, Dickinson PJ, Sturges BK, Vernau KL, et al. Whole genome variant association across 100 dogs identifies a frame shift mutation in dishevelled 2 which contributes to Robinow-like syndrome in bulldogs and related screw tail dog breeds. PLoS Genet. 2018; 14(12): e1007850. https://doi.org/10.1371/journal.pgen.1007850 PMID: 30521570

42. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, Faden DL, et al. Variation of BMP3 contributes to dog breed skull diversity. PLoS Genet. 2012; 8(8):e1002849. https://doi.org/10.1371/journal.pgen.1002849 PMID: 22876193

43. Flat-Coated Retriever Society of America. Breed standard [Available from: https://fcrsa.org/breed-standard/.

44. Rachidi SM, Qin T, Sun S, Zheng WJ, Li Z. Molecular profiling of multiple human cancers defines an inflammatory cancer-associated molecular pattern and uncovers KPNA2 as a uniform poor prognostic cancer marker. PLoS One. 2013; 8(3):e57911. https://doi.org/10.1371/journal.pone.0057911 PMID: 23536776

45. Shin SB, Jang HR, Xu R, Won JY, Yim H. Active PLK1-driven metastasis is amplified by TGF-beta signaling that forms a positive feedback loop in non-small cell lung cancer. Oncogene. 2020; 39(4):767–85. https://doi.org/10.1038/s41388-019-1023-z PMID: 31548612

46. Hu S, Xu-Monette ZY, Balasubramanyam A, Manyam GC, Visco C, Tzankov A, et al. CD30 expression defines a novel subgroup of diffuse large B-cell lymphoma with favorable prognosis and distinct gene expression signature: A report from the international dlbcl rituximab-chop consortium program study. Blood. 2013; 121(14):2715–24. https://doi.org/10.1182/blood-2012-10-461848 PMID: 23343832

47. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, Bao W, et al. A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer. 2010; 126(5):1177–86. https://doi.org/10.1002/ijc.24910 PMID: 19795455

48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

49. Kingsley NB, Kern C, Creppe C, Hales EN, Zhou H, Kalbfleisch TS, et al. Functionally annotating regulatory elements in the equine genome using histone mark ChIP-seq. Genes (Basel). 2019;11(1). https://doi.org/10.3390/genes11010003 PMID: 31861495

50. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518(7539):317–30. https://doi.org/10.1038/nature14248 PMID: 25693563

51. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics. 2011; 27(7):1017–8. https://doi.org/10.1093/bioinformatics/btr064 PMID: 21330290

52. Manke T, Heinig M, Vingron M. Quantifying the effect of sequence variation on regulatory interactions. Hum Mutat. 2010; 31(4):477–83. https://doi.org/10.1002/humu.21209 PMID: 20127973

53. Thomas-Chollier M, Hufton A, Heinig M, O'Keeffe S, Masri NE, Roider HG, et al. Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat Protoc. 2011; 6(12):1860–9. https://doi.org/10.1038/nprot.2011.409 PMID: 22051799

54. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun. 2019; 10(1):1489. https://doi.org/10.1038/s41467-019-09373-w PMID: 30940804