Table of Contents

[Abstract 2](#_Toc45471957)

[Introduction 2](#_Toc45471958)

[Results 2](#_Toc45471959)

[iDog 2](#_Toc45471960)

[Genes associated with disease 3](#_Toc45471961)

[SNPs with a Z(F](#_Toc45471962)[ST](#_Toc45471962)[) ≥ 5 3](#_Toc45471962)

[Missense and deleterious mutations 4](#_Toc45471963)

[Deleterious mutations based on SIFT](#_Toc45471964)[4](#_Toc45471964) [4](#_Toc45471964)

[160 dogs 4](#_Toc45471965)

[Discussion 4](#_Toc45471966)

[Methods 5](#_Toc45471967)

[Sample collection 5](#_Toc45471968)

[Library construction and sequencing 5](#_Toc45471969)

[Alignment and filtering 5](#_Toc45471970)

[Variant calling 5](#_Toc45471971)

[Statistical methods 5](#_Toc45471972)

[Candidate genes 5](#_Toc45471973)

[F statistics 5](#_Toc45471974)

[Pathway analysis 6](#_Toc45471975)

[Copy number variation 6](#_Toc45471976)

A study of germline mutations in Flat-Coated Retrievers

Abstract

***Introduction:*** *Flat-coated retrievers are prone to develop several different diseases, most notably histiocytic sarcoma. Interestingly this tendency is not seen in Golden Retrievers, a closely related breed.*

***Aim:*** *This study aims characterize germline variation in Flat-coated Retrievers and compare it to Golden Retrievers and other European breeds by F-statistics and pooled heterozygosity.*

***Methods:*** *Whole-genome sequencing of 19 Flat-coated Retrievers was performed and analyzed with a modified version of the K9-pipeline developed at Uppsala University. The pipeline uses bwa mem and haplotypeCaller. The vcf was then filtered with several filters (Table 1) and F-statistics and pooled heterogozity was performed with PLINK 1.91 with the settings seen in Table 2.*

# Introduction

Flat-coated retrievers are prone to develop several different diseases, most notably histiocytic sarcoma and cancer more generally [ref]. Interestingly the tendency is not seen in Golden Retrievers, a closely related breed. This study aims to characterize germline variation in Flat-coated Retrievers and compare it to Golden Retrievers and other European breeds by F-statistics and pooled heterozygosity.

Dogs have been bred for many centuries. This selective breeding has led to accumulation of different variants, some desirable, such as coat color, some deleterious, such as an increased risk of histiocytic sarcoma in some breeds.

Histiocytic diseases are a group of rare diseases in humans, but with a very poor prognosis in most cases. Two dog breeds, namely Bernese Mountain dogs and Flat-Coated Retrievers are predisposed to different types of histiocytic sarcoma, making them ideal models for the disease. It is estimated that around 36% of all neoplasms in Flat-Coated Retrievers are histiocytic sarcoma (Boercamp 2013).

Earlier studies have shown that somatic mutations in several genes are commonly seen in relation to histiocytic sarcoma.

This study aims to investigate which germline variants are fixed in 1) Flat-Coated Retrievers and 2) Flat-Coated Retrievers diagnosed with HS compared to 160 Swedish and American dogs of different breeds. In addition, comparison to a subset of Golden Retrievers will be done, as this is the closest related breed and which do not tend to develop HS.

Talk about Wrights etc2,3

# Results

## iDog

Results for the case-group Flat-coated Retrievers. The control group is the whole iDog dataset, unless otherwise specified. See TABLE X for details.

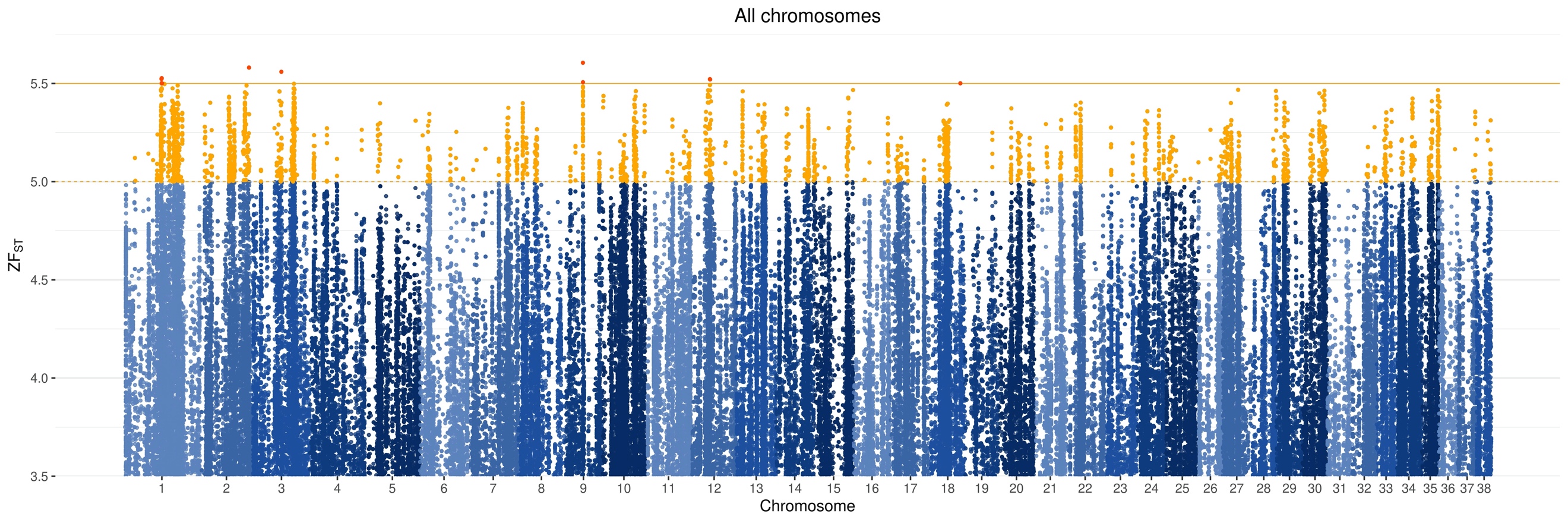


Figure 1: Manhattan plot over Z(FST) when comparing Flat-coated Retrievers with 160 + iDog as controls



### Genes associated with disease

7 SNPs with a Z(FST) ≥ 5 was found in the gene CHD7 and 1 was found in the gene AKAP9 (Table 3). All variants were intron variants compared with the 160 dogs + the iDog dataset

Table 1: SNPs in disease associated genes with a FST ≥ 5 when compared to the 160 dogs + the iDog dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Disease | Type | Gene | Location | FST | Z(FST) | MAF | | | Extra | | Observations | |
|  |  |  |  |  |  | FCR | Control |  | | FCR | | iDog |
| Renal dysplasia | Intron variant | CHD7 | chr29:11175393 | 0.90 | 5.1 | 0.74 | 0.03 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 316 |
| Renal dysplasia | Intron variant | CHD7 | chr29:11175957 | 0.90 | 5.1 | 0.74 | 0.03 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 318 |
| Renal dysplasia | Intron variant | CHD7 | chr29:11189179 | 0.91 | 5.1 | 0.74 | 0.02 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 318 |
| Renal dysplasia | Intron variant | CHD7 | chr29:11190237 | 0.90 | 5.0 | 0.74 | 0.03 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 316 |
| Renal dysplasia | Intron variant | CHD7 | chr29:11194658 | 0.90 | 5.0 | 0.74 | 0.02 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 314 |
| Renal dysplasia | Intron variant | CHD7 | chr29:11199937 | 0.90 | 5.1 | 0.74 | 0.02 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 320 |
| Renal dysplasia | Intron variant | CHD7 | chr29:11269773 | 0.90 | 5.0 | 0.74 | 0.02 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 318 |
| Top-20 somatic  mutations | Intron variant | AKAP9 | chr14:17683459 | 0.91 | 5.1 | 0.84 | 0.02 | IMPACT=MODIFIER  STRAND=1 | | 38 | | 318 |

### SNPs with a Z(FST) ≥ 5

In total 3839 SNPs had a Z(FST) ≥ 5, most of them (52%) intergenic (Table 1). Especially interesting was 12 SNPs with a Z(FST) ≥ 5.5, including an intron variant in each of the genes TBC1D32, and NBL1, and 1 long non-coding RNA (Table 2), no SIFT score is available for these.

Table 1: Consequences of mutations with a Z(FST) ≥ 5 when comparing Flat-coated Retrievers with the iDog dataset.

Table 2: List of SNPs with a Z(FST) ≥ 5.5 when comparing the Flat-coated Retrievers with the 160 dogs + the iDog dataset

No SNPs had a Z(FST) ≥ 5 when comparing with the subset of 20 Golden Retrievers. However, SNPs are seemingly fixated relative to both the Golden Retrievers and the iDog dataset (FST FCR vs GR = 1 and Z(FST FCR vs 160+ iDog) ≥ 5). These are found in 2 genes; 10 SNPs in NRXN1 and 5 SNPs in MCTP2 (see Table X for position).

### Missense and deleterious mutations

Of the 9 missense mutations with a Z(FST) ≥ 5. 5 was deemed to be deleterious based on SIFT score. These were located in the genes TRPM6, PCLO, CRP, FBN1, and CC2D2A (See table X for position).

### Deleterious mutations based on SIFT4

In total 2589 SNPs was deemed to be deleterious by SIFT, of these 1848 had a non-low confidence. 19 of these SNPs had a Z(FST) ≥ 4. Missense mutations in 4 genes were deemed deleterious and had a Z(FST) ≥ 5. Namely in TRPM6, PCLO, CC2D2A, and FBN1 (See table X for details).

## 160 dogs

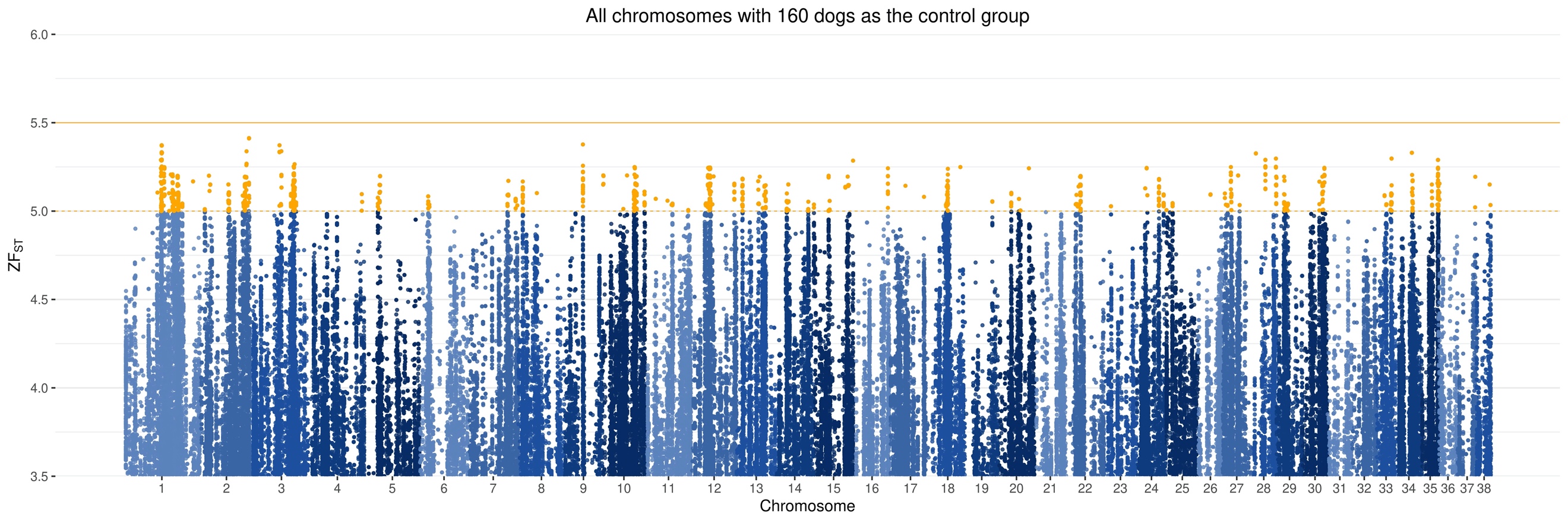


Table 4

|  |  |  |
| --- | --- | --- |
| Type | No | % |
| Intergenic variant | 1138 | 55% |
| Intron variant | 605 | 29% |
| Intron variant/non-coding transcript variant | 156 | 8% |
| Downstream gene variant | 88 | 4% |
| Upstream gene variant | 68 | 3% |
| Non-coding transcript exon variant | 3 | 0% |
| Synonymous variant | 1 | 0% |
| Splice region variant/intron variant | 1 | 0% |
| Missense variant/splice region variant | 1 | 0% |
| Missense variant | 1 | 0% |
| Total | 2062 | 100% |

# Discussion

Other groups have found other associations, Shearing et al. 20125 found a correlation between histiocytic sarcoma and CFA11:44150645 (CFA11:47179346 canFam2), in this study, 17 of the FCR was homozygous for the alternative allele (90%) and 2 was heterozygous, none were homozygous for the reference allele. Whereas only 35% of the control group of 160 dogs was homozygous for alternative allele. However, this SNP did not seem to be fixated in the FCR population investigated in this study Z(FST)=0.9.

# Methods

## Sample collection

## Library construction and sequencing

## Alignment and filtering

The reads were aligned to the CanFam3.1 reference genome with BWA mem 0.7.12. Duplicates were marked with picard 2.10.6. Furthermore, the reads were realigned and recalibrated with GATK 3.5 using GATK best practices6.

## Variant calling

Variants were called with HaplotypeCaller and genotyped using GenotypeGVCFs.

The file was then split and filtered separately for SNPs and INDELs using SelectVariants and VariantFiltration, with the filters seen in Table 1.

|  |  |  |
| --- | --- | --- |
| Filters | SNP | INDELS |
| QD | < 2.0 | < 2.0 |
| FS | > 60.0 | > 200.0 |
| SOR | > 3.0 | > 10.0 |
| MQ | < 40.0 |  |
| QD | < 2.0 |  |
| FS | > 60.0 |  |
| MQRankSum | < -12.5 |  |
| ReadPosRankSum | < -8.0 | < -20 |
| Max maf | 0.99992 |  |

## Statistical methods

### Candidate genes

In order to investigate genes of interest, several gene lists were generated (see (Sup) Table X)

3 based on TCGA data:

20 most commonly mutated genes in CMML

20 most common somatic mutations associated with cancer

20 most common germline mutations associated with cancer

3 based on prior knowledge from the literature

Genes associated with histiocytic disease

Genes associated with patella luxation

Genes associated with renal dysplasia

### F statistics

To evaluate if any SNPs were fixated in this population. F-statistics were calculated with 3 different populations and annotated with VEP 994.

1) 160 dogs provided by Erik Axelsson

2) 20 golden retrievers, a subset of 1)

3) The iDog gvcf containing 127 individuals7.

The FST was calculated using Plink 1.91 --fst with the “case-control” setting. The threshold for significance was set to 5 standard deviations from the mean (Z-transformation).

### Pathway analysis

### Copy number variation

Copy number variation was evaluated using cnvkit