Table of Contents

[Abstract 2](#_Toc51594488)

[Introduction 2](#_Toc51594489)

[Results 2](#_Toc51594490)

[160 dogs 2](#_Toc51594491)

[Candidate genes 2](#_Toc51594492)

[Fixation index 4](#_Toc51594493)

[Discussion 4](#_Toc51594494)

[Methods 4](#_Toc51594495)

[Sample collection 4](#_Toc51594496)

[Library construction and sequencing 4](#_Toc51594497)

[Alignment and filtering 4](#_Toc51594498)

[Variant calling 4](#_Toc51594499)

[Statistical methods 5](#_Toc51594500)

[Candidate genes 5](#_Toc51594501)

[Missense and deleterious mutations 5](#_Toc51594502)

[F statistics 5](#_Toc51594503)

[Pathway analysis 5](#_Toc51594504)

[Copy number variation 5](#_Toc51594505)

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.*

***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

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.

Flat-coated retrievers are prone to develop several different diseases, most notably histiocytic sarcoma and cancer more generally2–4.

This study aims to explore the genetic makeup of Flat-coated Retrievers in regards to phenotype-associated genes and SNPs. In addition, the study will investigate any breed-specific variation by looking for areas of decreased variation in the Flat-coated Retrievers by pooled heterozygosity and looking for any selected areas by F-statistics in comparison to 160 Swedish and American dogs.

Talk about Wrights etc5,6

# Results







## 160 dogs

### Candidate genes

The genes list for each disease can be found in table X, and the mentioned SNPs in table Y.

#### Chronic Myeloid Monocytic Leukemia

12 missense mutations were found and 3 was found to be deleterious. However, only 1 of these was commonly found in the Flat-coated Retriever. This mutation was found in the FLT gene and was found to have a moderate impact (MAFFCR=0.74, MAFcontrol=0.28).

#### Histiocytic disease

15 missense mutations were found, 1 was found to be deleterious, however, it had a low MAF in both FCR and controls (MAFFCR=0.05, MAFcontrol=0.16). 1 mutation in the Spi-C gene was found to be homozygous in all FCR, but was also commonly present in the control group (MAFFCR=1, MAFcontrol=0.82).

#### 20 most germline mutated genes in cancer

23 missense mutations were found and of these, 1 deleterious mutation was found in ATR that was commonly present in the FCR and less so in the control group (MAFFCR=0.92, MAFcontrol=0.28). When looking at the missense mutations that are not deleterious, 3 mutations are present in the BRCA2 gene at a higher frequency, than in the control group (MAFFCR=0.82|0.82|0.71, MAFcontrol=0.34|0.35|0.27), these are likely in linkage disequilibrium with each other. 1 mutation in the POLE gene and 1 mutation in the FAT1 gene is present homozygously in all FCRs, however, it was also commonly present in the control group (MAFFCR=1, MAFcontrol= 0.88 | 0.82).

#### Renal Dysplasia

16 deleterious missense mutations were found in these genes, however, most of these had similar MAF in both cases and controls. 2 deleterious mutations was found, 1 in SALL4 (MAFFCR=0.47, MAFcontrol= 0.007), the other was in the gene FREM2 (MAFFCR=0.82, MAFcontrol= 0.30). In total, 41 missense mutations were found in these genes, one found in DHCR7 had a high MAF in the FCRs, but not in the control (MAFFCR=0.74, MAFcontrol= 0.14).

#### Patella luxation

No deleterious mutations were found in this subset, but 12 missense mutations were found. Only one, in an uncharacterized gene, had a differing MAF between cases and controls (MAFFCR=0.61, MAFcontrol= 0.18).

#### 20 most somatic mutated genes in cancer

33 missense mutations were found and 10 of these were found to be deleterious. Most notably 2 missense mutations in AKAP9 had a high MAF in the FCR and not in the control, 1 was tolerated while the other was not (MAFFCR=0.84 | 0.84, MAFcontrol= 0.05 | 0.10), the fact that he 2 mutations has the exact same MAF and are close together (73 bp) suggest that they are in linkage disequilibrium with each other. A group of mutations in the gene RNF213 was found to have the same high MAF in FCR, and a somewhat lower MAF in the control group (MAFFCR=0.95, MAFcontrol= 0.67-0.70)

Table 4: Consequences of mutations with a Z(FST) ≥ 5 when comparing Flat-coated Retrievers with the 160 dogs

### Fixation index



Figure 2: Z-transformed Z(FST) for Flat-coated Retrievers with 160 dogs as the control group

# Discussion

Other groups have found other associations, Shearing et al. 20128 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. grepl(“

# 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 practices9.

## 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

### Missense and deleterious mutations

To investigate the consequences of the different mutations, all SNPs were annotated with VEP.

### F statistics

To evaluate if any SNPs were fixated in this population. F-statistics were calculated with 2 different populations and annotated with VEP 997.

1) 160 dogs provided by Erik Axelsson

2) 20 golden retrievers, a subset of 1)

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).

Data pruned with prune, R2=0.5

### Pathway analysis

### Copy number variation

Copy number variation was evaluated using cnvkit