Table of Contents

[Abstract 2](#_Toc55198804)

[Introduction 2](#_Toc55198805)

[Introduction to FCR 2](#_Toc55198806)

[Introduction to Berner and gene test 2](#_Toc55198807)

[Introduction to methods 2](#_Toc55198808)

[Results 3](#_Toc55198809)

[Annotation 3](#_Toc55198810)

[Candidate genes 3](#_Toc55198811)

[Chronic Myeloid Monocytic Leukemia 3](#_Toc55198812)

[Histiocytic disease 3](#_Toc55198813)

[20 most germline mutated genes in cancer 3](#_Toc55198814)

[Renal Dysplasia 4](#_Toc55198815)

[Patella luxation 4](#_Toc55198816)

[20 most somatic mutated genes in cancer 4](#_Toc55198817)

[Fixation index 4](#_Toc55198818)

[Pathway analysis 5](#_Toc55198819)

[Conclusion 5](#_Toc55198820)

[Methods 5](#_Toc55198821)

[Sample collection 5](#_Toc55198822)

[Library construction and sequencing 5](#_Toc55198823)

[Alignment and filtering 5](#_Toc55198824)

[Variant calling 5](#_Toc55198825)

[Statistical methods 6](#_Toc55198826)

[Annotation 6](#_Toc55198827)

[Candidate genes 6](#_Toc55198828)

[F statistics 6](#_Toc55198829)

[Pathway analysis 6](#_Toc55198830)

[Copy number variation 6](#_Toc55198831)

Investigation of the Germline Variation in Swedish 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 was performed with PLINK 1.9*1*.*

# Introduction

The Flat-Coated Retriever is a happy and carefree dog and has been described as a canine Peter Pan2. It is admired for its great working traits and unique looks.

The breed is thought to be based on several other breeds, notably Setters mixed with the St. Johns Water Dog from Newfoundland, but also Spaniels and many others2,3. The breed was popular among wealthy sports hunters in the early 20th century but declined as Labrador – and Golden Retrievers gained popularity. This population decline was especially pronounced during World War II which represents a genetic bottleneck for the breed.

In the 1990s UK breeders had noticed a high frequency of cancer in the breed and the Flat Coated Retriever Breed Society raised money to fund several studies of the breed, including a cohort study of 174 Flat-Coated Retrievers4,5. They found that not only did the breed have a generally high frequency of cancers, the latter study found that at least 42% of the Flat-Coated Retrievers died from cancer and 16 % of all the dogs died of tumors meeting the criteria for histiocytic sarcoma4 and other studies have found similar results [REF].

## Introduction to Berner and gene test

Another breed with an increased risk of developing histiocytic sarcoma is the Bernese Mountain Dog.

In 2011 the Swedish Flat-Coated Retriever club began a collaboration with the Kerstin Lindblad-Toh group at Uppsala University in order to investigate the health status of the Swedish Flat-Coated Retrievers in order to improve the health of the breed. To investigate the genetic makeup of the breed, whole genome sequencing was performed on samples from 19 Swedish Flat-Coated Retrievers.

## Introduction to methods

One of the most used methods to describe genetic variations between populations are Wrights F-statistics6,7. FST which is the most used measure of genetic differentiation is directly related to the variance of allele frequencies between populations. And can have values 0-1. Thus, if the difference of allele frequencies between populations is large, the FST will likewise be high whereas similar populations will have a low difference in allele frequencies and thus a low FST6. In practice, this means that areas of the genome where there have been a selective advantage on either of populations investigated, the difference between allele frequencies in the populations will be higher than what is expected due to simple genetic drift, and the FST for these locations is high6. Due to the descriptive nature of FST, P-values are not an appropriate way of calculating statistical significance and instead the cut-off is based on normalizing the FST values and only focusing on single nucleotide polymorphisms (SNPs) where this value is more than 5 standard deviations away from the mean.

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 any areas with increased inter-population differences in allele frequencies by F-statistics when comparing to 160 Swedish and American dogs.

# Results

Population

## Annotation

Table 2: Consequences found for all positions – see supplemental table XX for the full list

|  |  |  |
| --- | --- | --- |
| **Consequences** | **#** | **%** |
| Intergenic variant | 2481987 | 46% |
| Intron variant | 1923827 | 36% |
| Intron variant, Non-coding transcript variant | 414206 | 8% |
| Upstream gene variant | 272213 | 5% |
| Downstream gene variant | 240308 | 4% |
| Other | 66328 | 1% |

## 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 the *ATR* gene. This mutation 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 genes likely asscociated with renal dysplasia, however, most of these had similar MAF in both cases and controls. 2 deleterious mutations were found, 1 in the gene *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 the gene *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)

## Fixation index

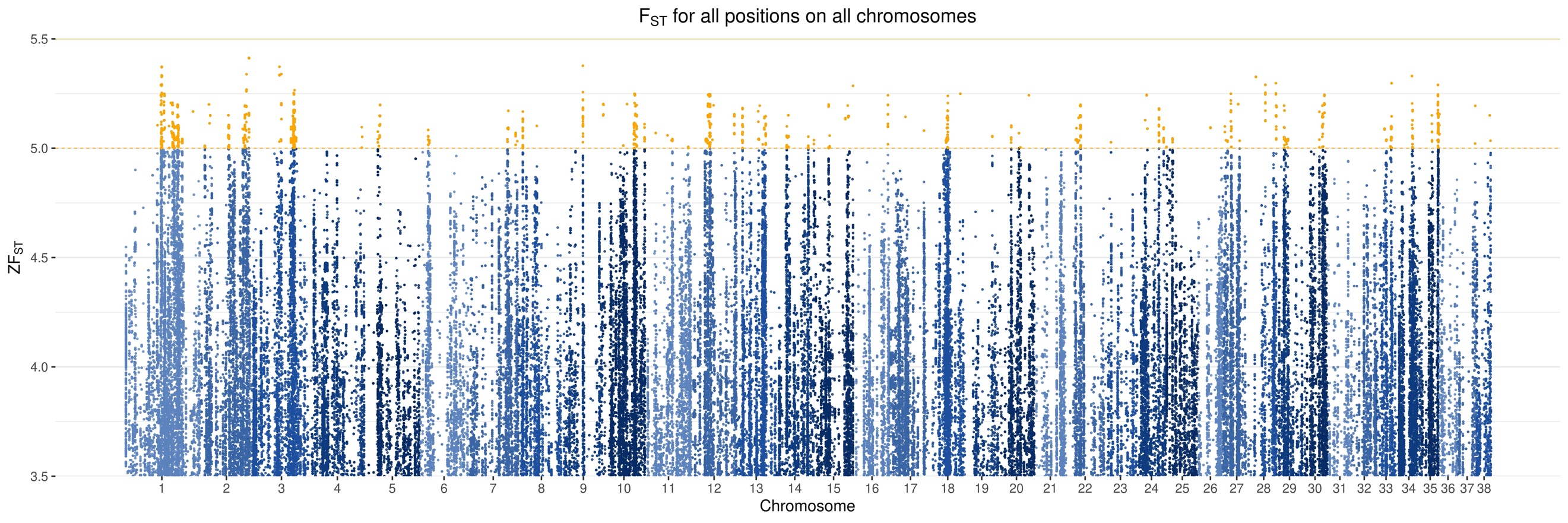


Figure 1: Z-transformed Z(FST) for Flat-coated Retrievers with 160 dogs as the control group

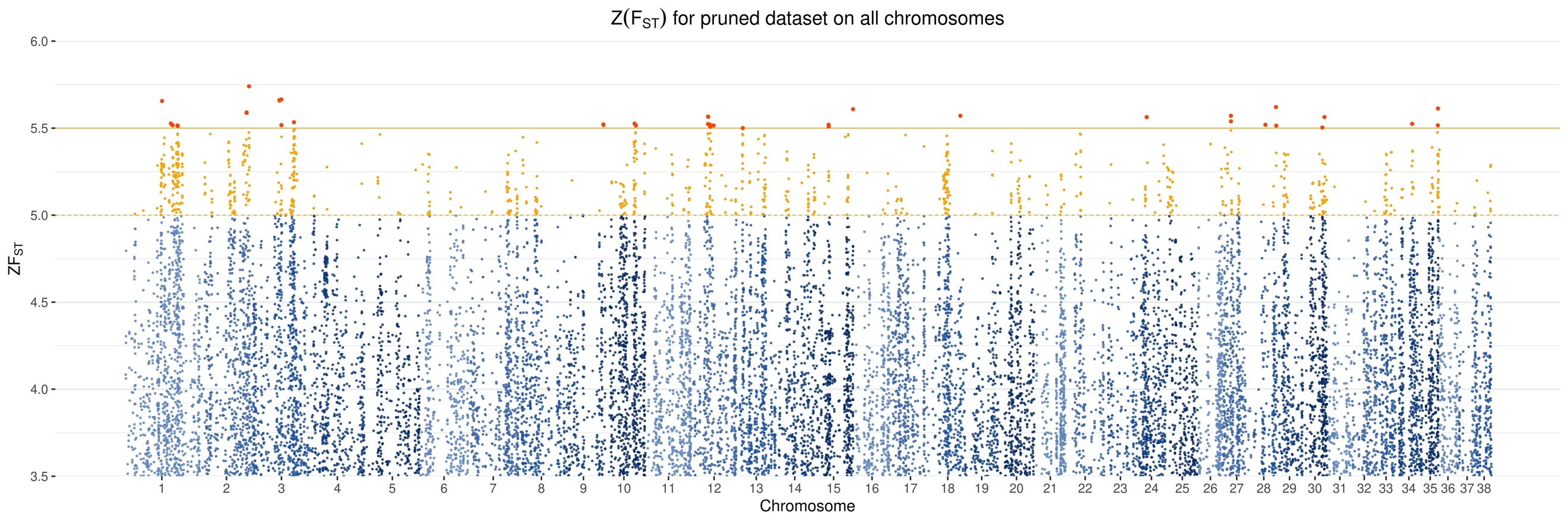


Figure 2: Z-transformed Z(FST) for Flat-coated Retrievers with 160 dogs as the control group

## Pathway analysis

The pathway analysis showed an enrichment for genes in the Histamine H1 receptor mediated signaling pathway (P04385) with a 13-fold enrichment of this pathway (pFDR=0.046).

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.

# Conclusion

Several SNPs were found to be fixated in the Flat-coated Retrievers analyzed, however, due to the small sample size more comprehensive studies are needed in order to confirm the results of this study.

# 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 (ref?). Duplicates were marked with picard 2.10.6 (ref?). 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 |
|  | > 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

### Annotation

The dataset was annotated with VEP 9910 --flag\_pick and only results with PICK=1 is reported in this article.

### 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

Other selected gene variants

### F statistics

To evaluate if any SNPs were fixated in this population.

Fixation index was calculated in 3 different ways; for all positions in the vcf-file, for a pruned dataset, and as a windowed mode. The first 2 is calculated with Plink 1.91 --fst with the “case-control” setting the windowed mode is calculated with vcftools 0.1.15. In all cases a missing genotype rate of <=0.05 was used.

#### Pruned dataset

The complete dataset was pruned based on linkage disequilibrium (plink --indep-pairwise 50 10 0.5), which are the standard settings. A lower r2 might be a better choice, but it was decided to keep the pruning less stringent e.g. to minimize the data-loss.

### Pathway analysis

A binomial overrepresentation test was run against the Panther database11 and the p-values corrected with the False Discovery Rate method. A full log can be found in table Y.

### Copy number variation

Copy number variation was evaluated using cnvkit

1. Shaun Purcell, C. C. PLINK 1.9.

2. Laughton, N. *A Review of the Flat-coated Retriever*. (Pelham, 1980).

3. Peterson, K. A Brief History of the Flat-Coated Retriever. *Show-Sight Magazine* 165–172 (2012).

4. Dobson, J., Hoather, T., McKinley, T. J. & Wood, J. L. N. Mortality in a cohort of flat-coated retrievers in the UK. *Vet. Comp. Oncol.* **7**, 115–121 (2009).

5. Morris, J. S., Bostock, D. E., McInnes, E. F., Hoather, T. M. & Dobson, J. M. Histopathological survey of neoplasms in flat-coated retrievers, 1990 to 1998. *Vet. Rec.* **147**, 291–5 (2000).

6. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F ST GenAlEx (integrated software for analysis of genetic data with an interface to Excel). *Nat. Rev. Genet.* **10**, 639–650 (2015).

7. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370 (1984).

8. Shearin, A. L. *et al.* The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. *Cancer Epidemiol. Biomarkers Prev.* **21**, 1019–1027 (2012).

9. Van der Auwera, G. A. *et al.* From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-11.10.33 (2013).

10. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

11. Mi H, T. P. PANTHER pathway: an ontology-based pathway database cou-pled with data analysis tools. *Methods Mol Biol* 563:123–40 (2009).