

Genetic Association of Deafness in Three Piebald Dogs

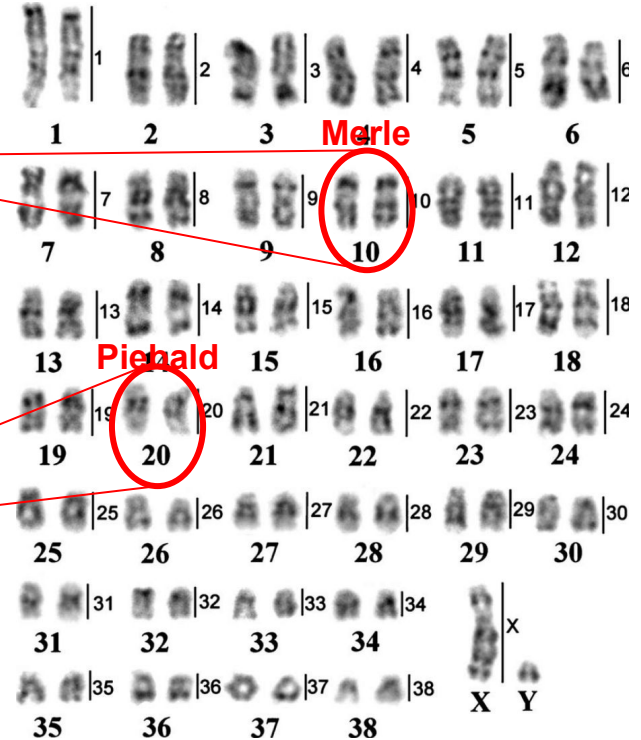


Soyeon Kim, Neel Srejan, Michelle Meng

Background

Piebald, Merle = having a pattern of two different colors of hair

↑ Piebald, Merle Gene = ↑ **Deafness**



Dataset

503 dogs: Dalmatians, Australian Cattle Dogs, English Setters from UK and North America

- 1) deafness.fam - ID and sex of each dog
- 2) deafness.bim - chromosome, name, bp, alleles for each SNP
- 3) deafness.bed - genotypes for each dog at each SNP
- 4) deafness_pheno.txt - phenotype file with dog ID, sex, breed, location, hearing phenotype, and sibling-pair phenotype

| | dogID | sex | breed | location | BAER_test_phenotype | sib_pair_phenotype |
|----------|--------------|------------|-----------------------|-----------------|----------------------------|---------------------------|
| 0 | 10557 | 1 | australian_cattle_dog | UK | hearing | NaN |
| 1 | 10558 | 1 | australian_cattle_dog | UK | bilaterally_deaf | 2.0 |
| 2 | 10560 | 2 | australian_cattle_dog | UK | hearing | 1.0 |
| 3 | 11934 | 2 | australian_cattle_dog | UK | unilaterally_deaf | NaN |

GWAS using PLINK

1. Quantify phenotype → .phen file

| dogID | sex | breed | location | BAER_test_phenotype | | familyID | dogID | BAER_test_phenotype |
|-------|-----|-----------------------|----------|---------------------|---|----------|-------|---------------------|
| 10557 | 1 | australian_cattle_dog | UK | hearing | → | 10557 | 10557 | 0 |
| 10558 | 1 | australian_cattle_dog | UK | bilaterally_deaf | | 10558 | 10558 | 1 |

2. Separate breeds → ID file for each breed/origin

UKdalmatianIDs, NAdalmatianIDs, EnglishSetterIDs, AustralianCattleDogIDs

3. PLINK linear regression on each breed/origin → .assoc.linear files
4. .assoc.linear → QQ plot & Manhattan plot

Principal Component Analysis (PCA)

1. Installed EIGENSTRAT (EIGENSOFT v7.2.1 package)
2. Run smartpca individually for each of the 3 breeds
 - a. Edit .fam file phenotype column to specify breed

```
11237 11237 0 0 2 -9
11246 11246 0 0 1 -9
11602 11602 0 0 2 -9
14166 14166 0 0 2 -9
15863 15863 0 0 2 -9
17551 17551 0 0 1 -9
ACD118 ACD118 0 0 2 -9
```



```
11237 11237 0 0 2 dalmatian
11246 11246 0 0 1 dalmatian
11602 11602 0 0 2 dalmatian
14166 14166 0 0 2 english_setter
15863 15863 0 0 2 english_setter
17551 17551 0 0 1 australian_cattle_dog
ACD118 ACD118 0 0 2 australian_cattle_dog
```

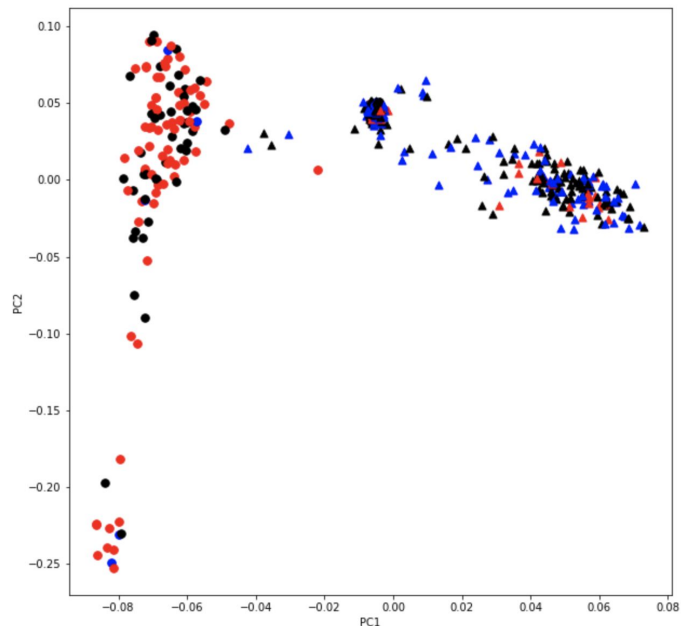
- b. Construct parameter file:

```
genotypename: deafness.bed
snpname:      deafness.pedsnp
indivname:    deafness_new_fullStr.pedind
evecoutname:  breed_out.evec
evaloutname:  breed_out.eval
poplistname:  breed_poplist.txt
altnormstyle: NO
```

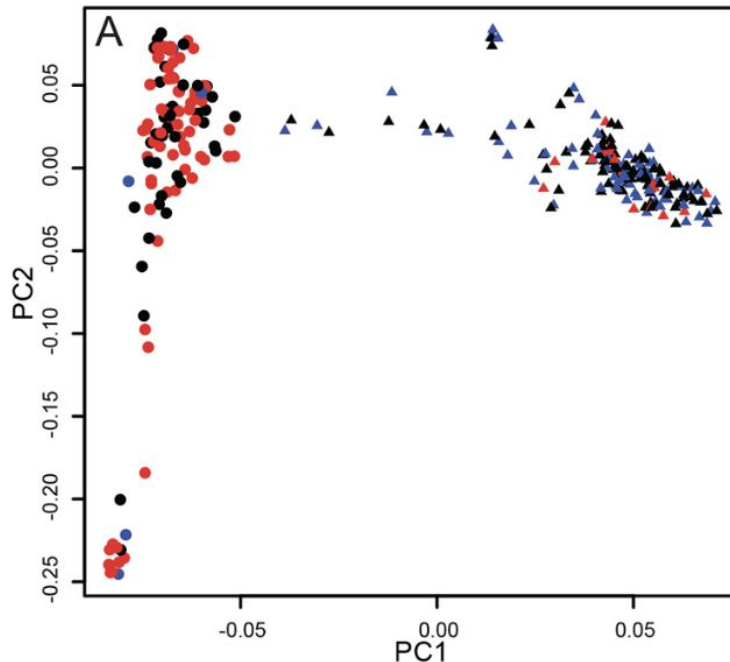
3. Output respective eigenvector and eigenvalue files
4. Use eigenvector file to plot PC1 vs. PC2 using Matplotlib in Jupyter Notebook

PCA for Dalmatians

Reproduction



Original



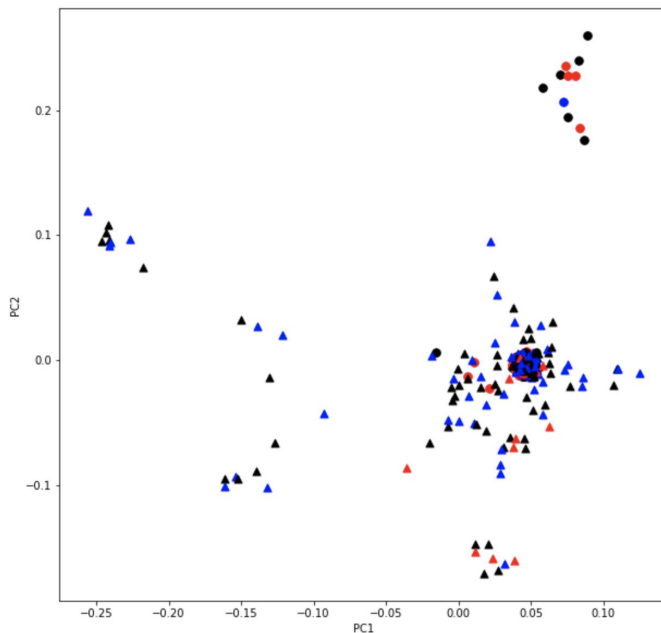
North America ▲
UK ●
Bilaterally Deaf
Unilaterally Deaf
Control

Genetic difference between North American vs. UK populations, shown by separation on PC1 with two distinct clusters.

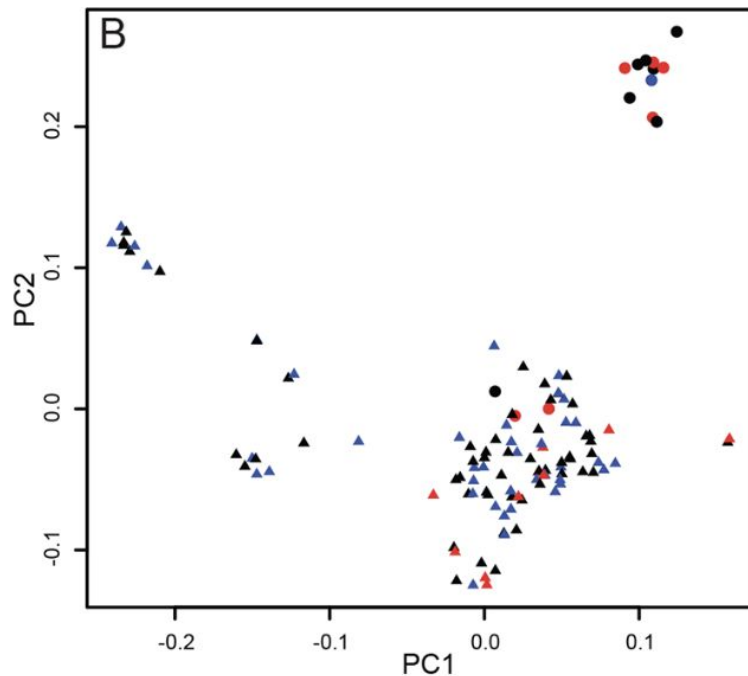
The North American samples were predominantly controls and unilaterally deaf, while the UK samples were predominantly bilaterally deaf and controls.

PCA for Australian Cattle Dog

Reproduction



Original



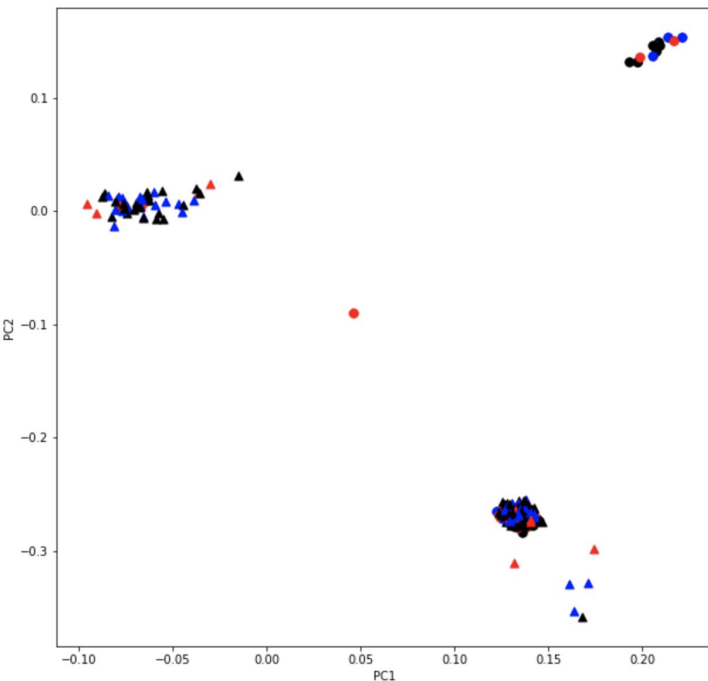
North America ▲
UK ●
Bilaterally Deaf
Unilaterally Deaf
Control

Sample geographic origin separates on PC2, with the exception of three UK samples

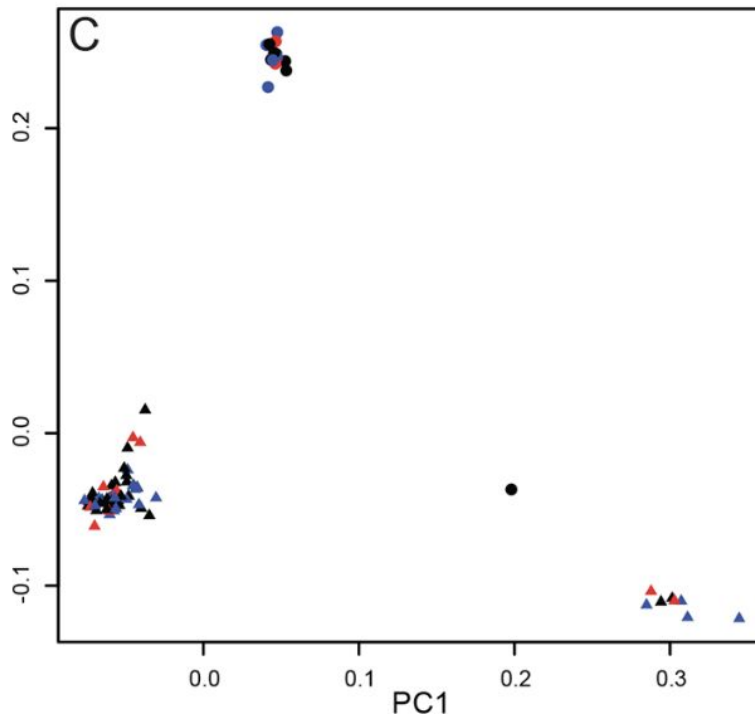
No phenotype separation

PCA for English Setter

Reproduction



Original



North America ▲
UK ●
Bilaterally Deaf
Unilaterally Deaf
Control

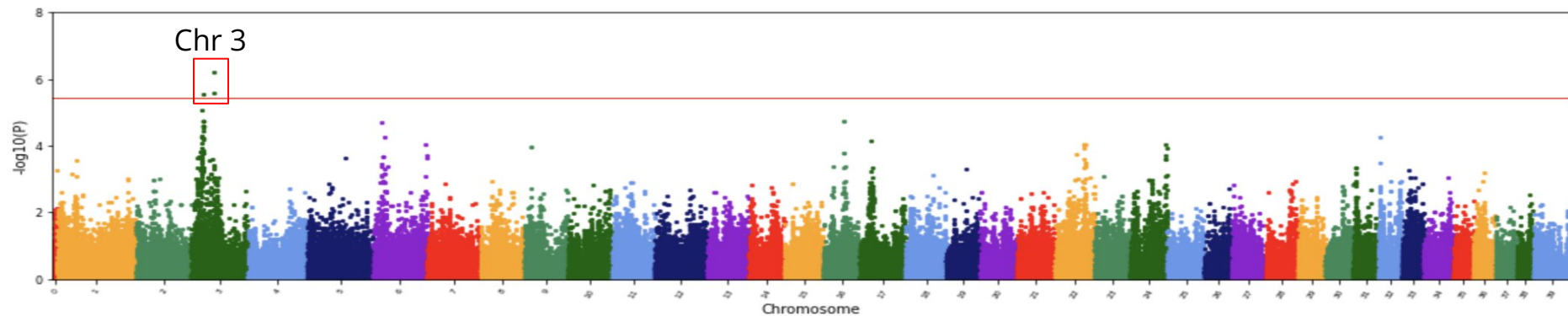
Some geographic structure, but no biased phenotypic distribution from the different locations

Challenges

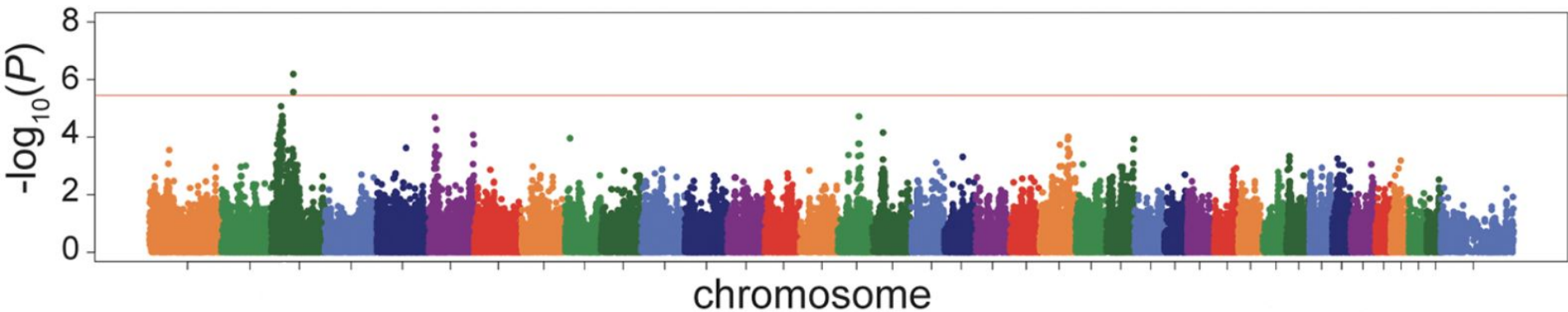
1. Older EIGENSTRAT used by paper not available anymore.
2. No ways to remove SNP's under 5% minimum allele frequency
3. Plink --maf 0.05 vs EIGENSTRAT (no maf filter) outputs different PCA's
4. Not enough time to run GEMMA for GWAS step 2,3
5. Used PLINK for GWAS step 2, 3 and had some overlapping significant SNP's

Manhattan Plot bilaterally deaf vs. control Australian cattle dogs

Reproduction(PLINK)

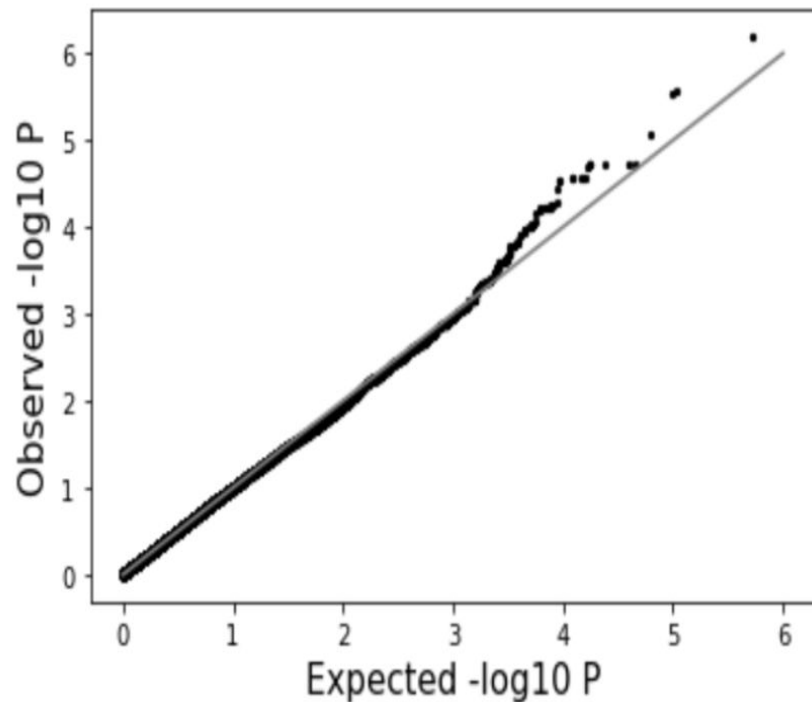


Original(GEMMA)

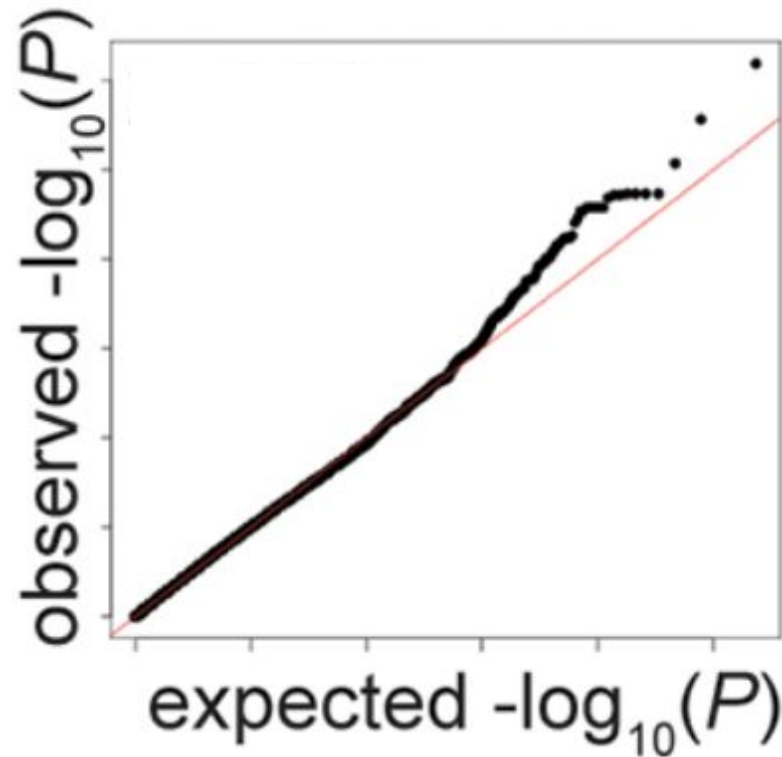


QQ plot bilaterally deaf vs. control Australian cattle dogs

Reproduction(PLINK)



Original(GEMMA)



Significant SNP's: original paper(GEMMA)

| Breed | N | CFA | bp | SNP ID | af all (deaf/controls) | pve ^a | beta | P-value |
|--------------------------|----------------------|-----|-------------|-------------------------|------------------------|------------------|--------|-------------------------|
| Dalmatians North America | 20 deaf, 91 controls | 30 | 37,235,914 | BICF2P1106247 | 0.104 (0.300/0.060) | 0.167 | -0.353 | $7.25 \times 10^{-6*}$ |
| | | 30 | 33,816,254 | BICF2P113616 | 0.333 (0.625/0.269) | 0.155 | -0.222 | $1.60 \times 10^{-5*}$ |
| | | 23 | 48,506,877 | BICF2G630365393 | 0.441 (0.725/0.379) | 0.150 | -0.220 | $2.28 \times 10^{-5*}$ |
| | | 30 | 22,647,163 | BICF2G630405064 | 0.068 (0.200/0.038) | 0.130 | -0.408 | $8.93 \times 10^{-5*}$ |
| | | 37 | 27,255,309 | BICF2G630132623 | 0.243 (0.450/0.198) | 0.122 | -0.245 | $1.54 \times 10^{-4*}$ |
| Dalmatians UK | 72 deaf, 43 controls | 38 | 21,626,523 | BICF2G63068103 | 0.152 (0.083/0.267) | 0.127 | 0.350 | $8.22 \times 10^{-5*}$ |
| Australian cattle dogs | 16 deaf, 61 controls | 3 | 37,793,043 | BICF2G630338450 | 0.299 (0.656/0.205) | 0.277 | -0.313 | $6.46 \times 10^{-7**}$ |
| | | 3 | 17,067,881 | BICF2G630703558 | 0.117 (0.344/0.057) | 0.229 | -0.408 | $8.45 \times 10^{-6*}$ |
| | | 16 | 36,220,138 | BICF2P1229299 | 0.091 (0.281/0.041) | 0.213 | -0.453 | $1.91 \times 10^{-5*}$ |
| | | 6 | 10,527,823 | BICF2S23125774 | 0.240 (0.500/0.172) | 0.212 | -0.330 | $2.05 \times 10^{-5*}$ |
| | | 17 | 18,275,241 | chr17_18275241 | 0.110 (0.313/0.057) | 0.187 | -0.377 | $6.96 \times 10^{-5*}$ |
| | | 6 | 75,622,113 | BICF2P481353 | 0.071 (0.219/0.033) | 0.184 | -0.505 | $8.37 \times 10^{-5*}$ |
| | | 22 | 48,747,165 | BICF2G630335709 | 0.494 (0.188/0.574) | 0.181 | 0.239 | $9.66 \times 10^{-5*}$ |
| | | 9 | 8,460,580 | BICF2S23511312 | 0.130 (0.313/0.082) | 0.178 | -0.395 | $1.09 \times 10^{-4*}$ |
| | | 24 | 47,255,337 | TIGRP2P322787_rs9139922 | 0.136 (0.344/0.082) | 0.176 | -0.344 | $1.19 \times 10^{-4*}$ |
| English setters | 11 deaf, 39 controls | 39 | 111,315,267 | BICF2G6304357 | 0.220 (0.421/0.154) | 0.192 | -0.276 | 1.20×10^{-3} |

Significant SNP's: reproduced table(PLINK)


| Breed | CHR | SNP ID | BP | BETA | P | Rank* |
|--------------------------|-----|-----------------|----------|---------------|--------------|-------|
| Dalmatians North America | 30 | BICF2P1106247 | 37235914 | -0.79066 | 0.000007 | 1 |
| | 30 | BICF2P113616 | 33816254 | -0.4430 | 0.000016 | 3 |
| | 23 | BICF2G630365393 | 48506877 | -0.4407 | 0.000023 | 4 |
| Dalmatians UK | 38 | BICF2G63068103 | 21626523 | 0.6994 | 0.000082 | 1 |
| Australian cattle dog | 3 | BICF2G630338450 | 37793043 | -6.266000e-01 | 6.461000e-07 | 1 |
| | 3 | BICF2G630703558 | 17067881 | -8.161000e-01 | 8.448000e-06 | 4 |
| | 16 | BICF2P1229299 | 36220138 | -9.054000e-01 | 1.908000e-05 | 6 |
| | 6 | BICF2S23125774 | 10527823 | -6.591000e-01 | 2.050000e-05 | 11 |
| | 17 | chr17_18275241 | 18275241 | -7.500000e-01 | 6.950000e-05 | 29 |
| English Setters | N/A | N/A | N/A | N/A | N/A | N/A |

Future Pipeline

Pipeline from paper: Eigenstrat + GEMMA = PCA, Manhattan, QQ plots and SNP table

Pipeline from project: Eigenstrat + PLINK = PCA, Manhattan, QQ plots and SNP table

Future pipeline: As seen above PLINK as a standalone tool is capable of reproducing all the figures in the paper, thus our proposed future pipeline is to have a two sided pipeline in which one side is PLINK and the other is EIGENSTRAT + GEMMA. We can combine the results to have a wider range of significant SNP's, doesn't limit us to the capabilities of a chosen software tool.



THANK YOU!
Any Questions?

