Genome Wide Association analysis (GWAS)

install PLINK 1.9

install vcftools

Install R and RStudio

Make a working directory for the GWAS analysis

Download the sample VCF file and phenotype data

convert VCF into Plink readable format (map,ped) then Plink binary format (fam,bed,bim)

create list of alternative alleles

Run a simple association analysis

Create Manhattan plot

Meta-Analysis of Genome Wide Association Studies (meta_GWAS.html)

Jupyter Notebook and Python for data science. (Jupyter-Notebook-Notes.html)

GitHub (github.html)

Publicly available databases (database_resources.html)

Assessing and Assembling Nanopore data (analyzing_nanopore_data.html)

Annotate with prokka: (analyzing_nanopore_data.html#annotate-with-prokka)

Exploratory RNAseq data analysis using RMarkdown (rmarkdown_rnaseq.html)

Amazing Resources for learning Rmarkdown (rmarkdown_rnaseq.html#amazing-resources-for-learning-rmarkdown)

Differential expression analysis with DESeq2 (deseq2-asthma.html)

Analyzing ChIP-seq data (chip-seq.html)

De novo transcriptome assembly with Trinity (assembly-trinity.html)

Annotating de novo transcriptomes

Genome Wide Association analysis (GWAS)

To test the association of of a genome-wide set of genetic variants with a given trait.

install PLINK 1.9 (https://www.cog-genomics.org/plink/1.9/)

```
cd /usr/local/bin/
sudo wget https://www.cog-genomics.org/static/bin/plir
sudo unzip -o plink_linux_x86_64.zip
sudo rm -f plink_linux_x86_64.zip
cd plink-1.07-x86_64/
echo export PATH=$PATH:$(pwd) >> ~/.bashrc
source ~/.bashrc
```

install vcftools (https://vcftools.github.io/)

```
cd
git clone https://github.com/vcftools/vcftools.git
cd vcftools
./autogen.sh
./configure
make
sudo make install
```

Genome Wide Association analysis (GWAS)

install PLINK 1.9

install vcftools

Install R and RStudio

Make a working directory for the GWAS analysis

Download the sample VCF file and phenotype data

convert VCF into Plink readable format (map,ped) then Plink binary format (fam,bed,bim)

create list of alternative alleles

Run a simple association analysis

Create Manhattan plot

Meta-Analysis of Genome Wide Association Studies (meta_GWAS.html)

Jupyter Notebook and Python for data science. (Jupyter-Notebook-Notes.html)

GitHub (github.html)

Publicly available databases (database_resources.html)

Assessing and Assembling Nanopore data (analyzing_nanopore_data.html)

Annotate with prokka: (analyzing_nanopore_data.html#annotate-with-prokka)

Exploratory RNAseq data analysis using RMarkdown (rmarkdown_rnaseq.html)

Amazing Resources for learning Rmarkdown (rmarkdown_rnaseq.html#amazing-resources-for-learning-rmarkdown)

Differential expression analysis with DESeq2 (deseq2-asthma.html)

Analyzing ChIP-seq data (chip-seq.html)

De novo transcriptome assembly with Trinity (assembly-trinity.html)

Annotating de novo transcriptomes

Install R and RStudio

```
sudo apt-get update && sudo apt-get install -y gdebi-c
sudo gdebi -n rstudio-server-1.0.143-amd64.deb
```

Make a working directory for the GWAS analysis

```
mkdir ~/GWAS && cd ~/GWAS
```

Download the sample VCF file and phenotype data

Genotyping of 476840 SNPs in 53 dogs (24 yellow coat and 29 dark coat)

```
wget https://de.cyverse.org/dl/d/E0A502CC-F806-4857-9(
gunzip pruned_coatColor_maf_geno.vcf.gz
wget https://de.cyverse.org/dl/d/3B5C1853-C092-488C-8(
```

convert VCF into Plink readable format (map,ped) then Plink binary format (fam,bed,bim)

```
vcftools --vcf pruned_coatColor_maf_geno.vcf --plink -
plink --file coatColor --allow-no-sex --dog --make-bec
```

create list of alternative alleles

```
cat pruned_coatColor_maf_geno.vcf | awk 'BEGIN{FS="\t'
```

Genome Wide Association analysis (GWAS)

install PLINK 1.9

install vcftools

Install R and RStudio

Make a working directory for the GWAS analysis

Download the sample VCF file and phenotype data

convert VCF into Plink readable format (map,ped) then Plink binary format (fam,bed,bim)

create list of alternative alleles

Run a simple association analysis

Create Manhattan plot

Meta-Analysis of Genome Wide Association Studies (meta_GWAS.html)

Jupyter Notebook and Python for data science. (Jupyter-Notebook-Notes.html)

GitHub (github.html)

Publicly available databases (database_resources.html)

Assessing and Assembling Nanopore data (analyzing_nanopore_data.html)

Annotate with prokka: (analyzing_nanopore_data.html#annotate-with-prokka)

Exploratory RNAseq data analysis using RMarkdown (rmarkdown_rnaseq.html)

Amazing Resources for learning Rmarkdown (rmarkdown_rnaseq.html#amazing-resources-for-learning-rmarkdown)

Differential expression analysis with DESeq2 (deseq2-asthma.html)

Analyzing ChIP-seq data (chip-seq.html)

De novo transcriptome assembly with Trinity (assembly-trinity.html)

Annotating de novo transcriptomes

Run a simple association analysis

–assoc performs a standard case/control association analysis which is a chi-square test of allele frequency.

By default, the minor allele is coded A1 and tested for being the risk allele. The minor allele is expected to be the alternative allele but could happen to be the reference one. –reference-allele allow you to use your list of A1 alleles

–adjust enables correction for multiple analysis and automatically calculates the genomic inflation factor

```
plink --bfile coatColor.binary --make-pheno coatColor.
```

Create Manhattan plot

Install qqman package

```
sudo Rscript -e "install.packages('qqman',  contriburl=c
```

Identify statistical cutoffs

```
unad_cutoff_sug=$(tail -n+2 coatColor.assoc.adjusted | a
unad_cutoff_conf=$(tail -n+2 coatColor.assoc.adjusted |
```

Run the plotting function

```
Rscript -e 'args=(commandArgs(TRUE));library(qqman);'\
'data=read.table("coatColor.assoc", header=TRUE); data=c
'bitmap("coatColor_man.bmp", width=20, height=10);'\
'manhattan(data, p = "P", col = c("blue4", "orange3"),'\
'suggestiveline = 12,'\
'genomewideline = 15,'\
'chrlabs = c(1:38, "X"), annotateTop=TRUE, cex = 1.2);'\
'graphics.off();' $unad_cutoff_sug $unad_cutoff_conf
```

Genome Wide Association analysis (GWAS)

install PLINK 1.9

install vcftools

Install R and RStudio

Make a working directory for the GWAS analysis

Download the sample VCF file and phenotype data

convert VCF into Plink readable format (map,ped) then Plink binary format (fam,bed,bim)

create list of alternative alleles

Run a simple association analysis

Create Manhattan plot

Meta-Analysis of Genome Wide Association Studies (meta_GWAS.html)

Jupyter Notebook and Python for data science. (Jupyter-Notebook-Notes.html)

GitHub (github.html)

Publicly available databases (database_resources.html)

Assessing and Assembling Nanopore data (analyzing_nanopore_data.html)

Annotate with prokka: (analyzing_nanopore_data.html#annotate-with-prokka)

Exploratory RNAseq data analysis using RMarkdown (rmarkdown_rnaseq.html)

Amazing Resources for learning Rmarkdown (rmarkdown_rnaseq.html#amazing-resources-for-learning-rmarkdown)

Differential expression analysis with DESeq2 (deseq2-asthma.html)

Analyzing ChIP-seq data (chip-seq.html)

De novo transcriptome assembly with Trinity (assembly-trinity.html)

Annotating de novo transcriptomes

The top associated mutation is a nonsense SNP in MC1R (c.916C>T) known to control pigment production. The file `coatColor_man.bmp` can be visualized with RStudio or downloaded to your local computer (from local commandline) via:

```
scp username@ipaddress:/home/username/GWAS/coatColor_man.bmp .
```

Variant calling pipeline for a mammalian genome (GATK_pipeline.html)

Meta-Analysis of Genome Wide Association Studies (meta_GWAS.html)