BovReg
Understanding cattle genomes

Comprehensive BovReg eQTL Nextflow workflow and application on BovReg data sets

Praveen Krishna Chitneedi, Gabriel C. M. Moreira, Jose Espinosa-Carrasco, Changxi Li, Graham Plastow, Daniel Fischer, Carole Charlier, Dominique Rocha, Amanda J Chamberlain, Christa Kuehn

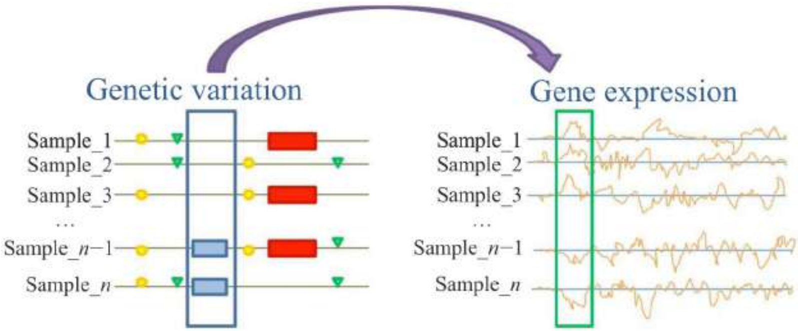
BovReg Final Conference - Brussels
(14-15 February 2024)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668

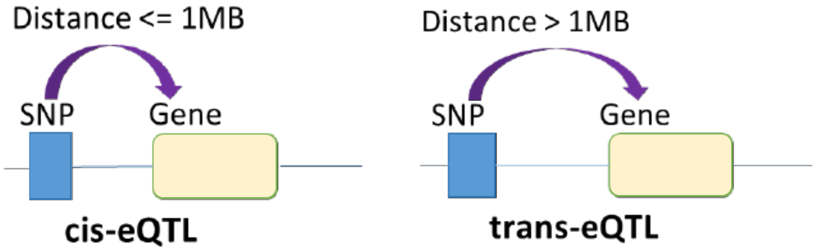
Introduction

eQTL (Expression Quantitative Trait Loci) association study: Identifies the genetic variants that affect the gene expression

eQTLs explains about 70% of heritability due to DNA variants changing gene expression and RNA splicing via cis and trans effects. *Xiang et al., 2023, Cell Genomics*



Tian, Lu, et al 2014



Distance <= 1MB

SNP Gene

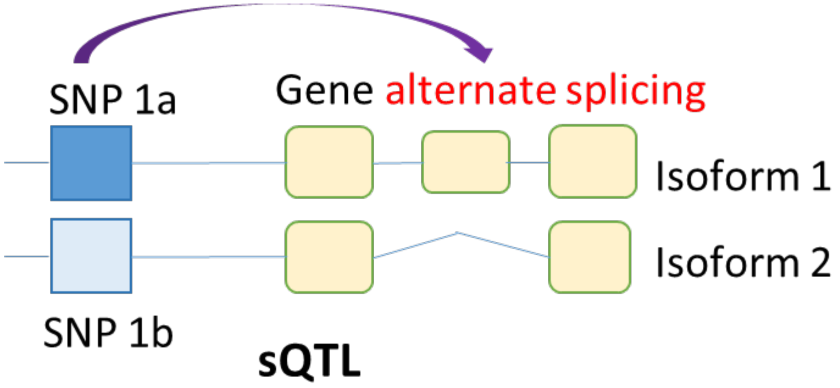
cis-eQTL

Distance > 1MB

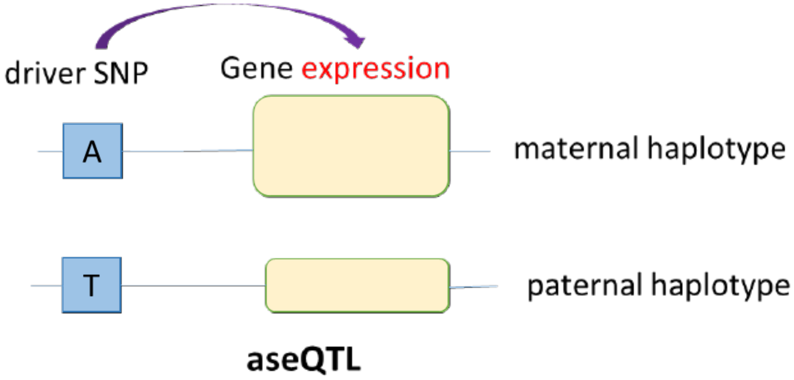
SNP Gene

trans-eQTL

BovReg WP4.3 - Introduction



sQTL: The presence of the variant at the locus affects the splicing of gene



aseQTL: The ratio of expression of a gene differs from 1:1 from two parental haplotypes by the presence of heterozygous SNP at a locus

Association between millions of genetic variants and thousands of genes typically measured in hundreds of individuals

BovReg WP4.3 - Introduction

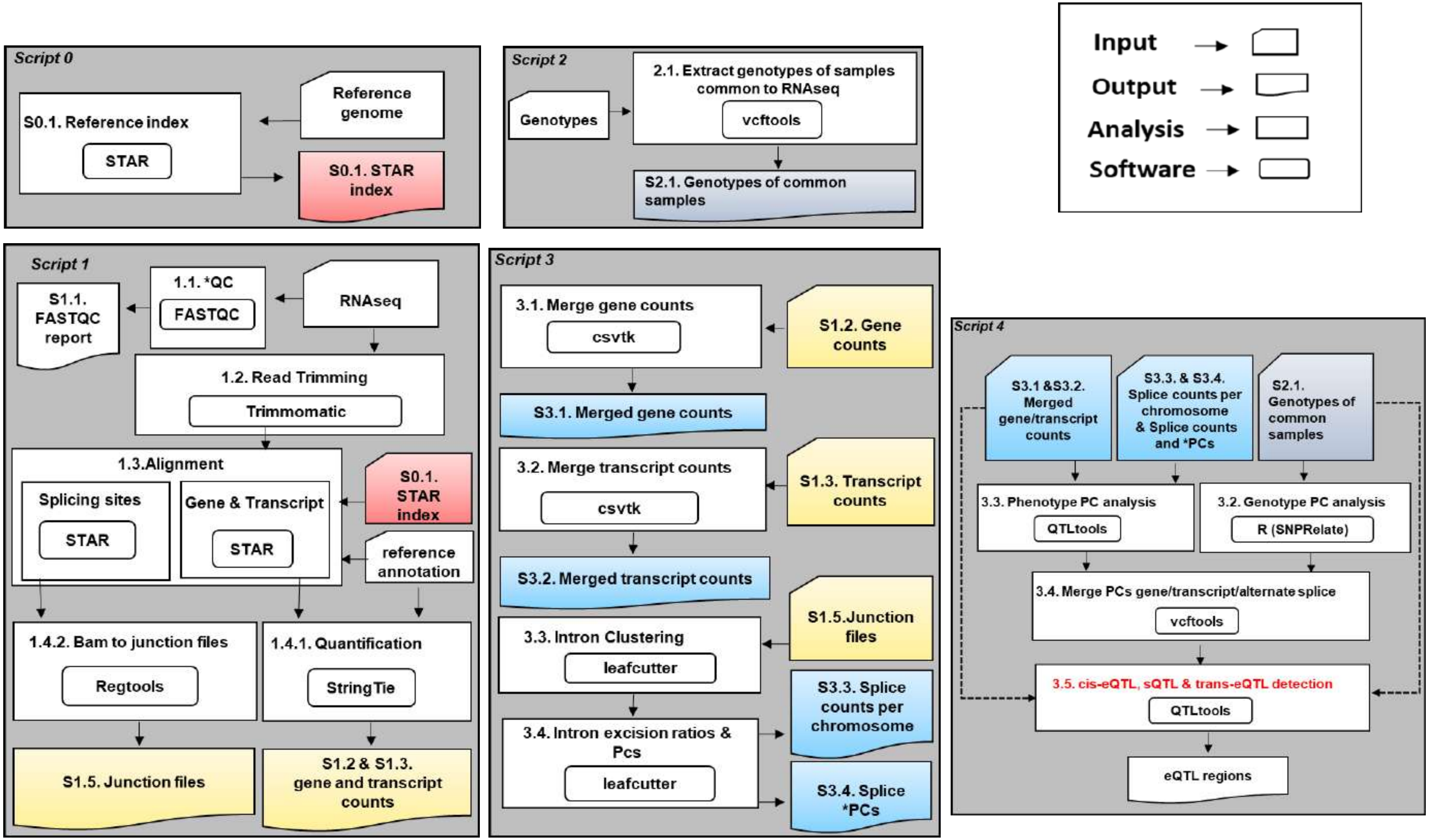
```
nextflow.enable.dsl=2

process sayHello {
  input:
    val cheers
  output:
    stdout

  """
  echo $cheers
  """
}

workflow {
  channel.of('Ciao','Hello','Hola') | sayHello | view
}
```

Nextflow is a workflow management software which enables the writing of scalable and reproducible scientific workflows developed by seqera labs using Groovy programming language.

Features

- Integrate software packages.
- Environment management systems such as Docker, Singularity, and Conda.
- Allows existing pipelines written in common scripting languages, such as R and Python, to be seamlessly coupled together.
- The nextflow DSL allows to run workflows on cloud or high-performance computing (HPC) infrastructures.

eQTL Bioinformatics workflow



Input → ☐
Output → ☐
Analysis → ☐
Software → ☐

Script 0

S0.1. Reference index
STAR
Reference genome
S0.1. STAR index

Script 2

Genotypes
2.1. Extract genotypes of samples common to RNAseq
vcftools
S2.1. Genotypes of common samples

Script 1

S1.1. FASTQC report
1.1. *QC
FASTQC
RNAseq
1.2. Read Trimming
Trimmomatic
1.3.Alignment
Splicing sites
STAR
Gene & Transcript
STAR
S0.1. STAR index
reference annotation
1.4.2. Bam to junction files
Regtools
1.4.1. Quantification
StringTie
S1.5. Junction files
S1.2 & S1.3. gene and transcript counts

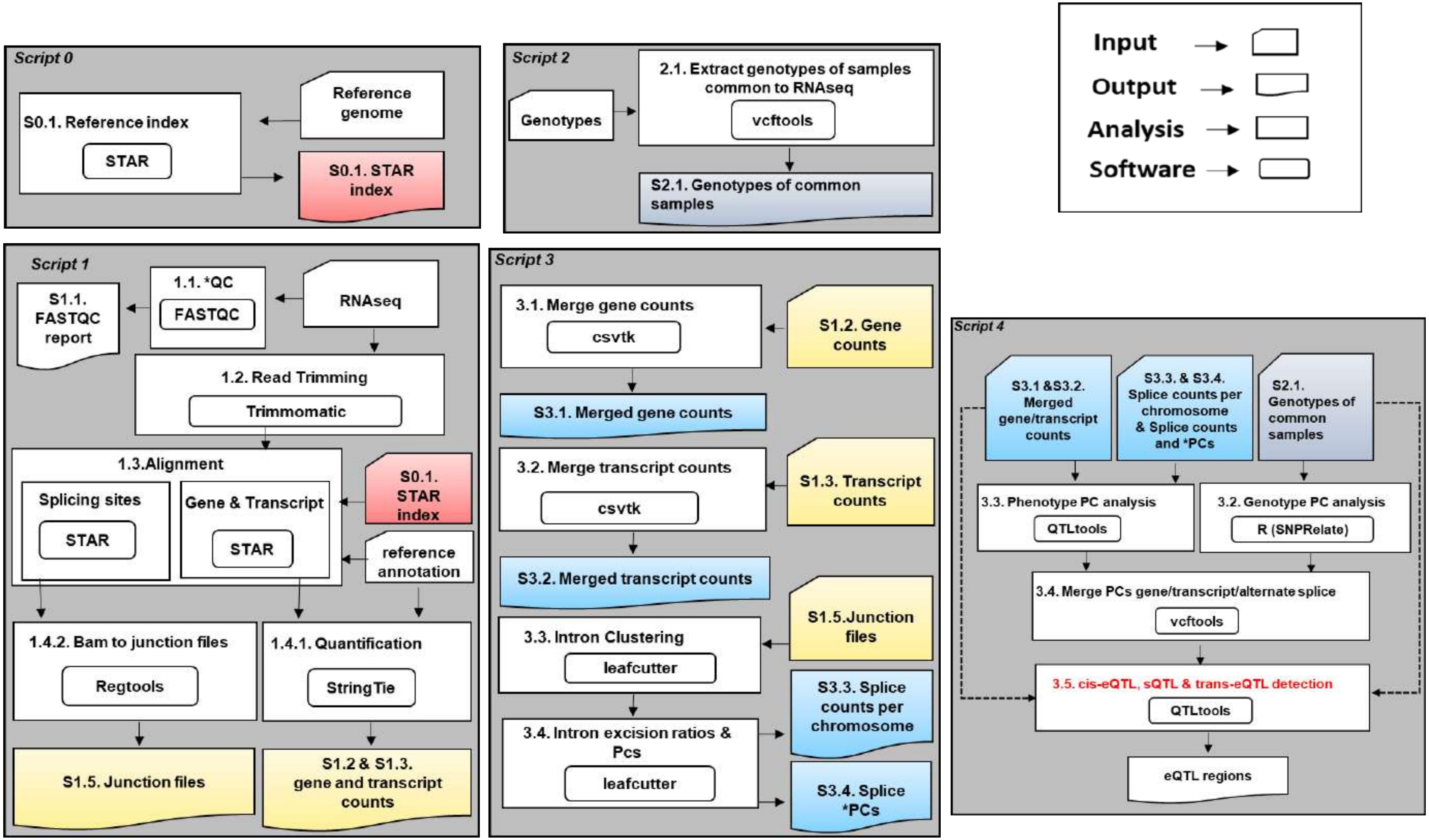
Script 3

3.1. Merge gene counts
csvtk
S1.2. Gene counts
S3.1. Merged gene counts
3.2. Merge transcript counts
csvtk
S1.3. Transcript counts
S3.2. Merged transcript counts
3.3. Intron Clustering
leafcutter
S1.5.Junction files
3.4. Intron excision ratios & Pcs
leafcutter
S3.3. Splice counts per chromosome
S3.4. Splice *PCs

Script 4

S3.1 &S3.2. Merged gene/transcript counts
S3.3. & S3.4. Splice counts per chromosome & Splice counts and *PCs
S2.1. Genotypes of common samples
3.3. Phenotype PC analysis
QTLtools
3.2. Genotype PC analysis
R (SNPRelate)
3.4. Merge PCs gene/transcript/alternate splice
vcftools
3.5. cis-eQTL, sQTL & trans-eQTL detection
QTLtools
eQTL regions

eQTL Bioinformatics workflow



Input
Output
Analysis
Software

Script 0

S0.1. Reference index
STAR
Reference genome
S0.1. STAR index

Script 2

Genotypes
2.1. Extract genotypes of samples common to RNAseq
vcftools
S2.1. Genotypes of common samples

Script 1

S1.1. FASTQC report
1.1. *QC
FASTQC
RNAseq
1.2. Read Trimming
Trimmomatic
1.3.Alignment
Splicing sites
STAR
Gene & Transcript
STAR
S0.1. STAR index
reference annotation
1.4.2. Bam to junction files
Regtools
1.4.1. Quantification
StringTie
S1.5. Junction files
S1.2 & S1.3. gene and transcript counts

Script 3

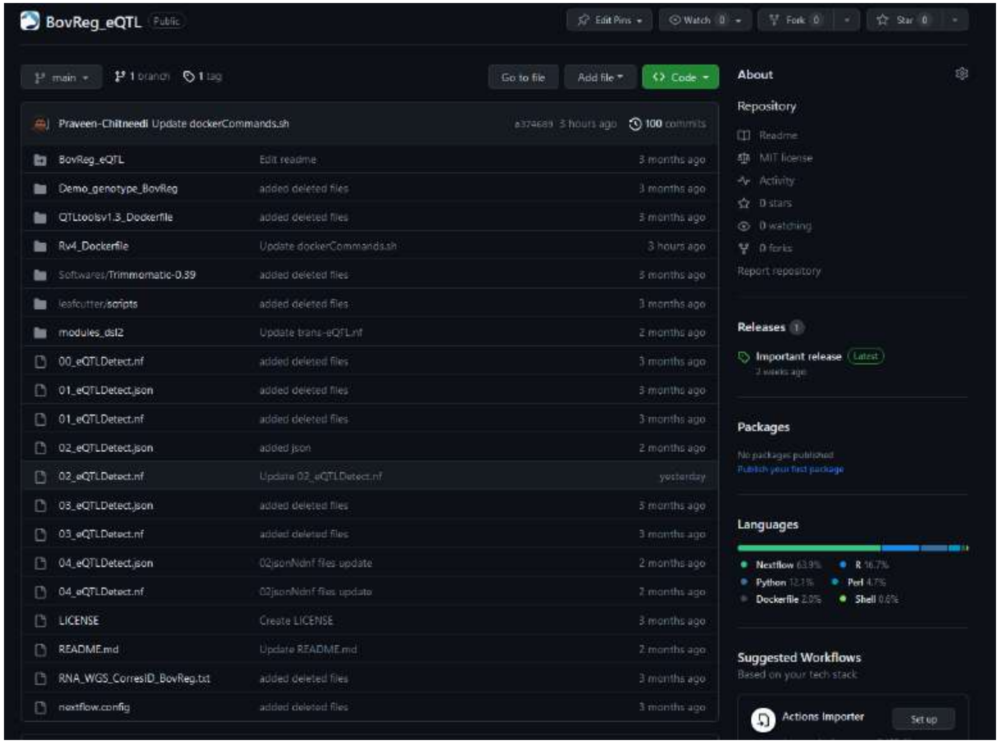
3.1. Merge gene counts
csvtk
S1.2. Gene counts
S3.1. Merged gene counts
3.2. Merge transcript counts
csvtk
S1.3. Transcript counts
S3.2. Merged transcript counts
3.3. Intron Clustering
leafcutter
S1.5.Junction files
S3.3. Splice counts per chromosome
3.4. Intron excision ratios & Pcs
leafcutter
S3.4. Splice *PCs

Script 4

S3.1 &S3.2. Merged gene/transcript counts
S3.3. & S3.4. Splice counts per chromosome & Splice counts and *PCs
S2.1. Genotypes of common samples
3.3. Phenotype PC analysis
QTLtools
3.2. Genotype PC analysis
R (SNPRelate)
3.4. Merge PCs gene/transcript/alternate splice
vcftools
3.5. cis-eQTL, sQTL & trans-eQTL detection
QTLtools
eQTL regions

eQTL Bioinformatics workflow

Source Code



Praveen-Chitneedi Update dockerCommands.sh a374689 3 hours ago 100 commits

BovReg_eQTL	Edit readme	3 months ago
Demo_genotype_BovReg	added deleted files	3 months ago
QTLtoolsv1.3_Dockerfile	added deleted files	3 months ago
Rv4_Dockerfile	Update dockerCommands.sh	3 hours ago
Softwares/Trimmomatic-0.39	added deleted files	3 months ago
leafcutter/scripts	added deleted files	3 months ago
modules_dsl2	Update trans-eQTL.nf	2 months ago
00_eQTLDetect.nf	added deleted files	3 months ago
01_eQTLDetect.json	added deleted files	3 months ago
01_eQTLDetect.nf	added deleted files	3 months ago
02_eQTLDetect.json	added json	2 months ago
02_eQTLDetect.nf	Update 02_eQTLDetect.nf	yesterday
03_eQTLDetect.json	added deleted files	3 months ago
03_eQTLDetect.nf	added deleted files	3 months ago
04_eQTLDetect.json	02jsonNdnf files update	2 months ago
04_eQTLDetect.nf	02jsonNdnf files update	2 months ago
LICENSE	Create LICENSE	3 months ago
README.md	Update README.md	2 months ago
RNA_WGS_CorresID_BovReg.txt	added deleted files	3 months ago
nextflow.config	added deleted files	3 months ago

About

Repository

Readme
MIT license
Activity
0 stars
0 watching
0 forks
Report repository

Releases 1

Important release Latest
2 weeks ago

Packages

No packages published
Publish your first package

Languages

- Nextflow 63.9%
- R 16.7%
- Python 12.1%
- Perl 4.7%
- Dockerfile 2.0%
- Shell 0.6%

Suggested Workflows
Based on your tech stack

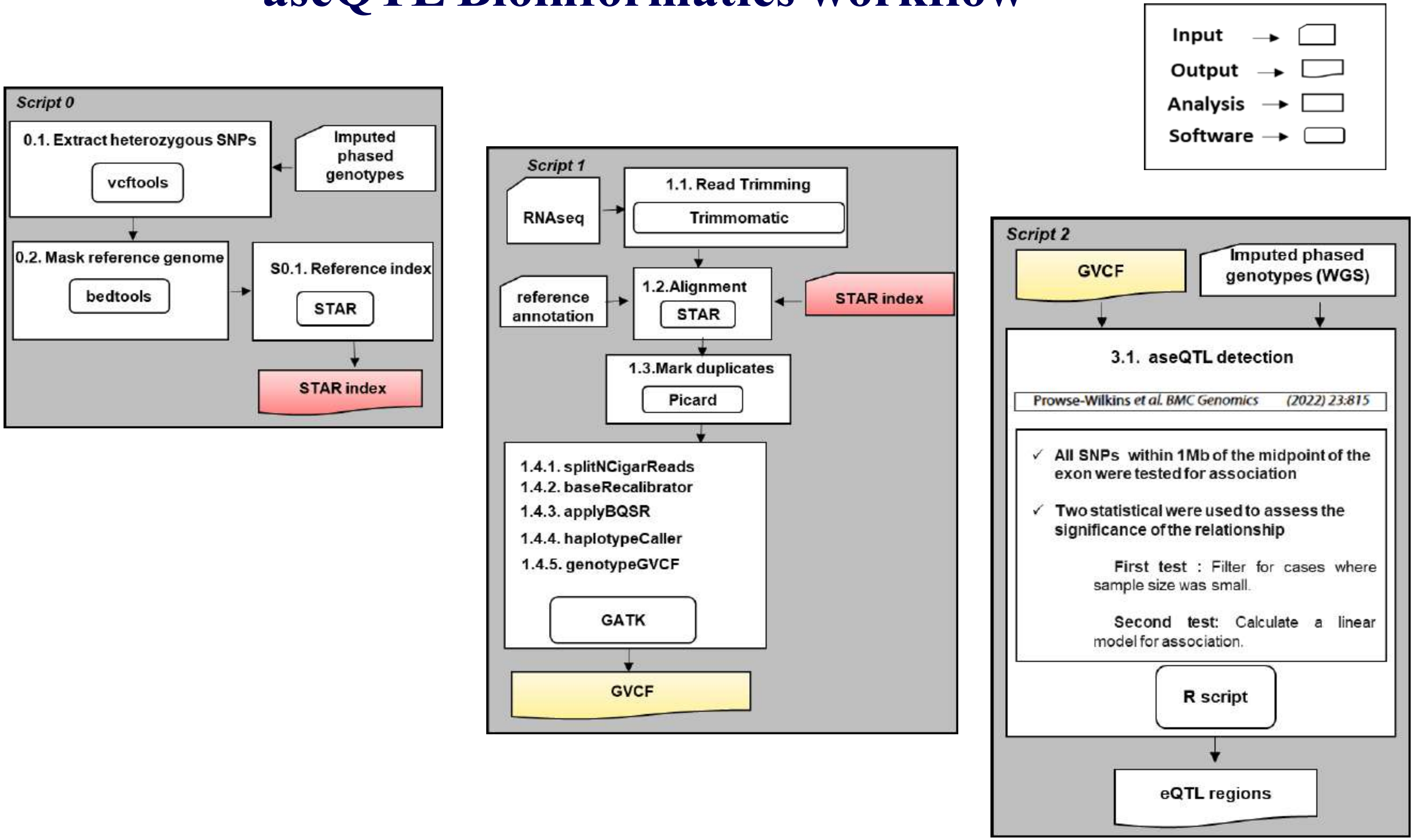
Actions Importer

https://github.com/BovReg/BovReg_eQTL

Publication

- The eQTL BovReg workflow paper was submitted to NAR Genomics and Bioinformatics
- Currently under review

aseQTL Bioinformatics workflow



Input
Output
Analysis
Software

Script 0

0.1. Extract heterozygous SNPs
vcftools
Imputed phased genotypes
0.2. Mask reference genome
bedtools
S0.1. Reference index
STAR
STAR index

Script 1

RNAseq
1.1. Read Trimming
Trimmomatic
reference annotation
1.2.Alignment
STAR
STAR index
1.3.Mark duplicates
Picard

1.4.1. splitNCigarReads
1.4.2. baseRecalibrator
1.4.3. applyBQSR
1.4.4. haplotypeCaller
1.4.5. genotypeGVCF

GATK
GVCF

Script 2

GVCF
Imputed phased genotypes (WGS)

3.1. aseQTL detection

Prowse-Wilkins et al. *BMC Genomics* (2022) 23:815

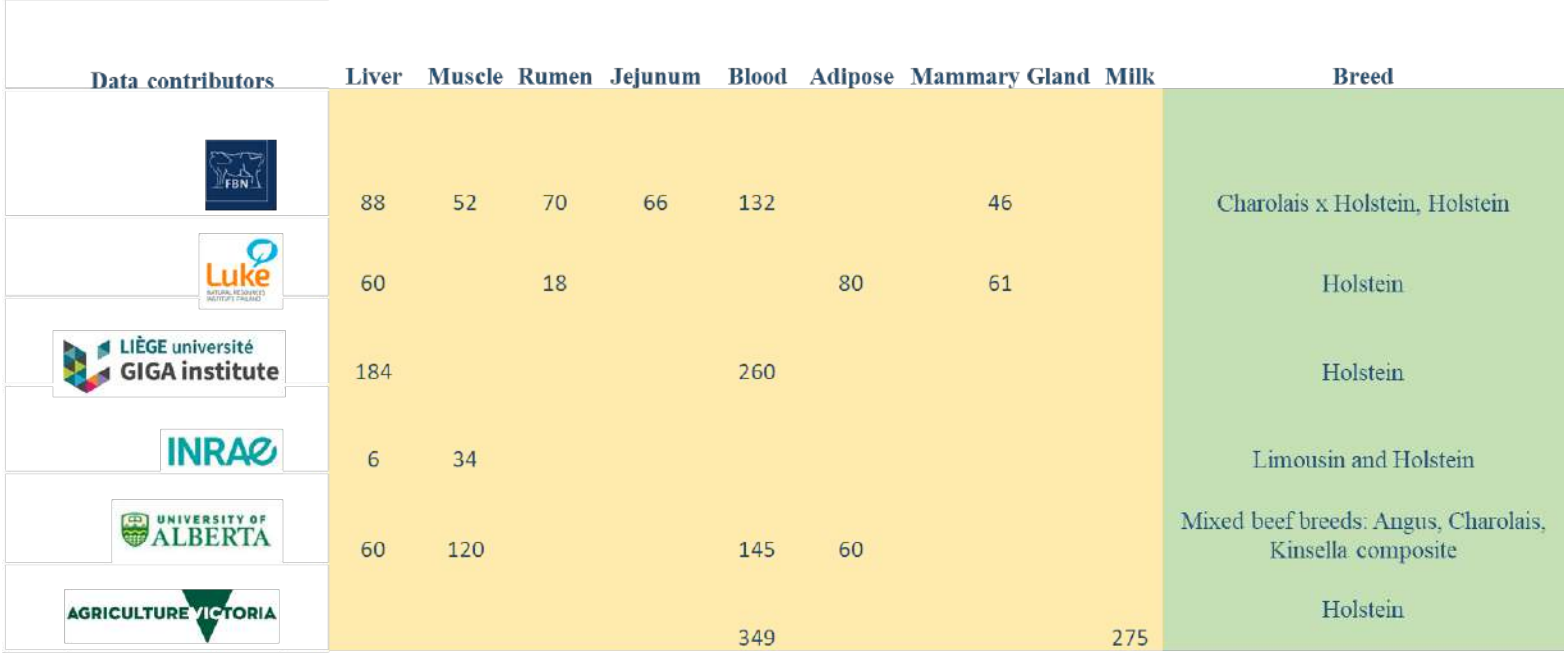
- ✓ All SNPs within 1Mb of the midpoint of the exon were tested for association
- ✓ Two statistical were used to assess the significance of the relationship

First test : Filter for cases where sample size was small.

Second test: Calculate a linear model for association.

R script
eQTL regions

BovReg data for eQTL analyses



Data contributors	Liver	Muscle	Rumen	Jejunum	Blood	Adipose	Mammary Gland	Milk	Breed
FBN	88	52	70	66	132		46		Charolais x Holstein, Holstein
Luke	60		18			80	61		Holstein
LIÈGE université GIGA institute	184				260				Holstein
INRAE	6	34							Limousin and Holstein
UNIVERSITY OF ALBERTA	60	120			145	60			Mixed beef breeds: Angus, Charolais, Kinsella composite
AGRICULTURE VICTORIA					349			275	Holstein

Genotype data imputed to whole genome sequence level available for corresponding RNAseq samples

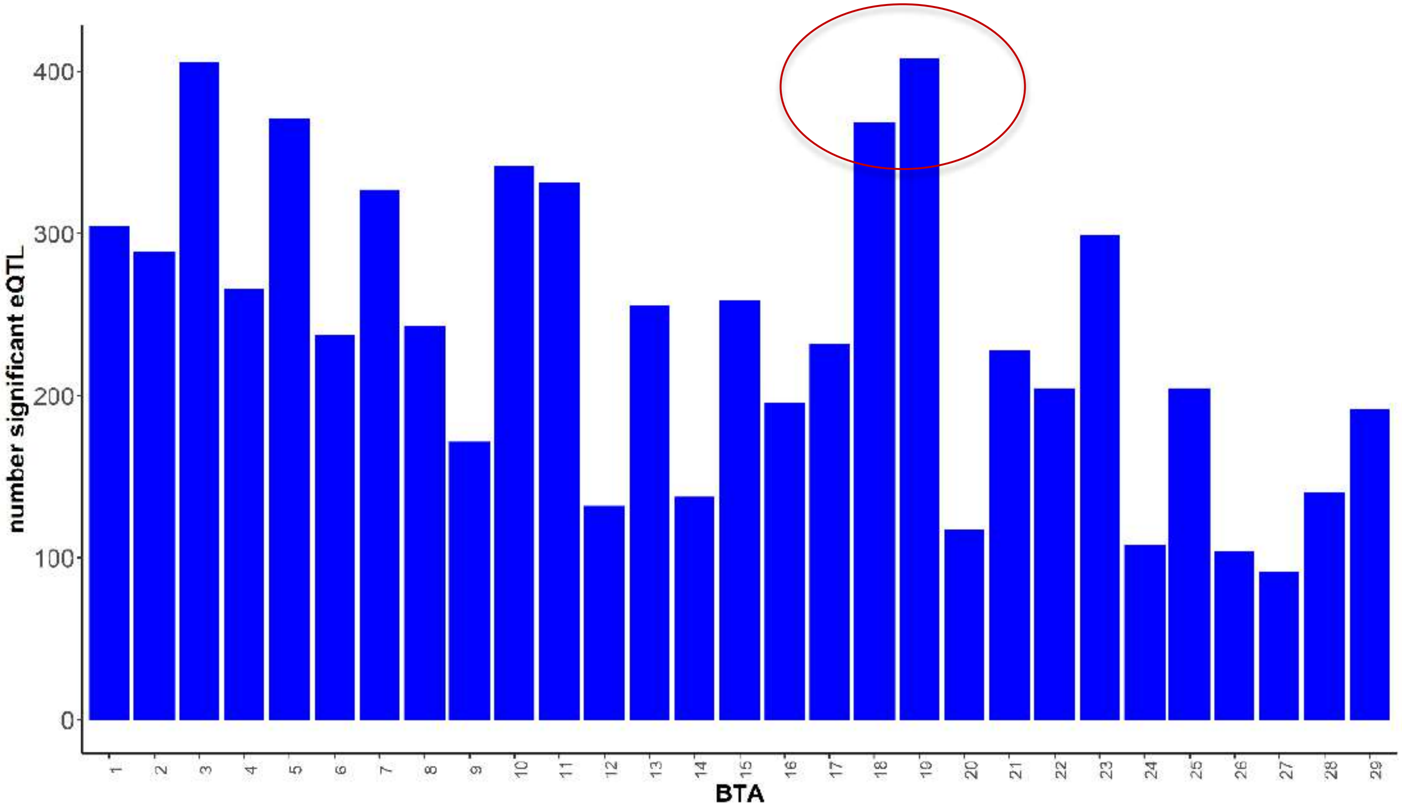
Workflow support

CRG Centre for Genomic Regulation

Results : Cis-eQTL

Expression phenotypes		Liver	Rumen	Jejunum	Muscle	Adipose	Blood	Mammary	Milk
Genes	in annotation	44638	44638	44638	44638	44638	44638	44638	44638
	after filtering out 10% samples	15497	13795	14324	15226	19280	16391	19681	13032
	$P_{nominal} < 0.01$	15123	10731	12304	12539	9072	15573	3701	11142
	$P_{adjusted} < 0.05$	6540	1747	2393	1235	2393	6453	949	490
Transcripts	in annotation	278769	278769	278769	278769	278769	278769	278769	278769
	after filtering out 10% samples	56359	48650	52877	53014	72031	82561	50643	53382
	$P_{nominal} < 0.01$	56129	39975	42393	42767	70971	77358	18232	45887
	$P_{adjusted} < 0.05$	27647	5067	4556	2941	8010	28741	11047	2415
sQTL	$P_{adjusted} < 0.05$	7176	6092	7251	1787	2358	33744	9592	20

Results : Cis-eQTL in liver tissue



Distribution of experiment-wise significant cis gene-level eQTL (padj < 0.05) per chromosome for liver

Results : cis-eQTL comparative study

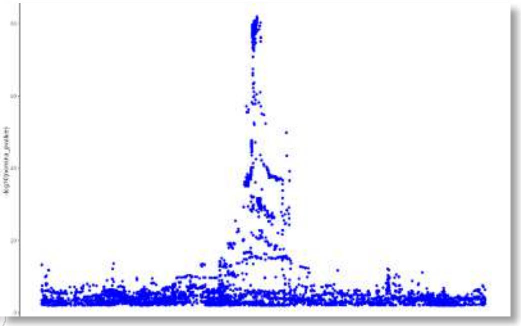
Example BTA6

Chr	Start	End	Strand	Blood	Liver	Mammary	Rumen	Milk	Adipose	Jejunum	Muscle
6	88543235	88706096	+	0	1	0	0	0	0	0	0
6	88720503	88741984	+	0	1	0	0	0	0	0	0

- An association signal at BTA 6:88.6 Mb with peak at 200-Kb region (BTA 6:88.5–88.7 Mb) in GC CNV region (*Young-Lim Lee et al.* 2021).
- Similar to our study GC CNV was predominantly regulation expressed in liver tissue.

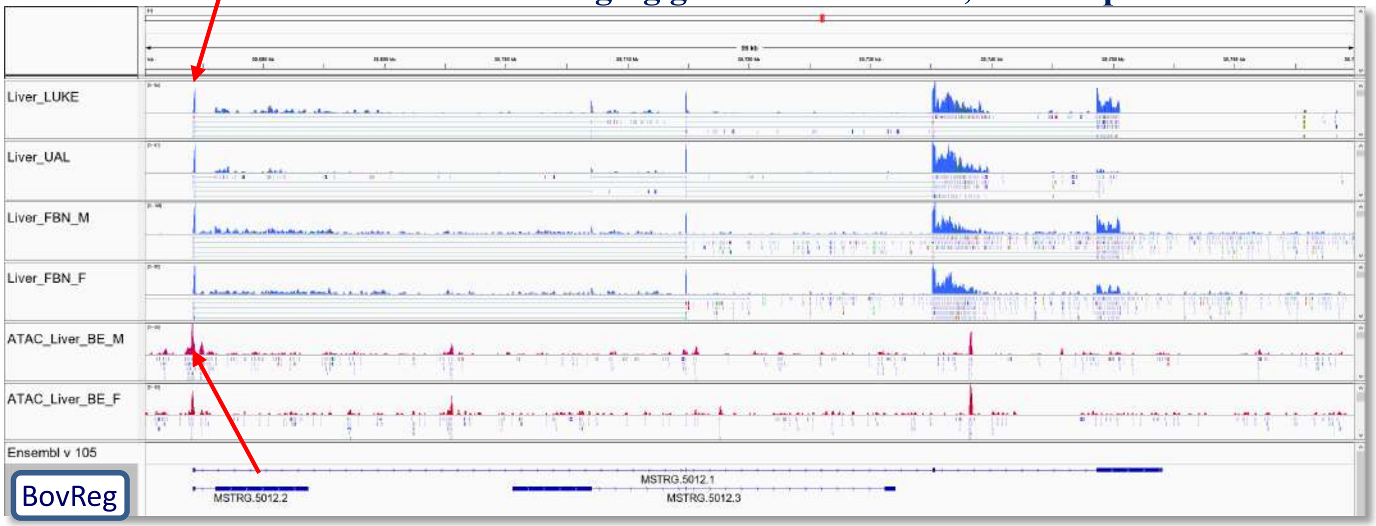
Results : Cis-eQTL with BovReg annotation

MSTRG.5012: a novel, yet undescribed lncRNA with strong eQTL signal



cis eQTL signal for MSTRG.5012

Merging genome annotation, transcript classification



Summary

- We created a Bioinformatics pipeline for eQTL detection using Nextflow and made available in GitHub
- Performed eQTL association studies with BovReg data sets
- With BovReg annotation, we discovered significant eQTLs in previously unannotated transcripts
- These eQTL results were fed into subsequent work packages on BovReg
- Integrate the aseQTL analysis with BovReg eQTL pipeline

BovReg *PARTNERS*

AARHUS UNIVERSITY

CRG Centre for Genomic Regulation

diagenode Innovating Epigenetics Solutions

edinethics

EMBL-EBI

ETH zürich

FBN

FACULDADE DE MEDICINA VETERINÁRIA UNIVERSIDADE DE LISBOA

INSTYTUT GENETYKI I HODOWLI ZWIERZĄT PAN

INRAE

LIÈGE université GIGA Medical Genomics

Université de Limoges

Luke LUONNONVARAKESKUS

MARTIN-LUTHER UNIVERSITÄT HALLE-WITTENBERG

ROSLIN

UNIVERSITY OF ALBERTA

THE UNIVERSITY OF EDINBURGH

Universiteit Utrecht

WAGENINGEN UNIVERSITY & RESEARCH

Thank you for your attention

www.bovregproject.eu

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668