

An Exploratory Analysis into Gene eQTL Profiles

-Metabolic Tissue and NAFLD
Missing Heritability

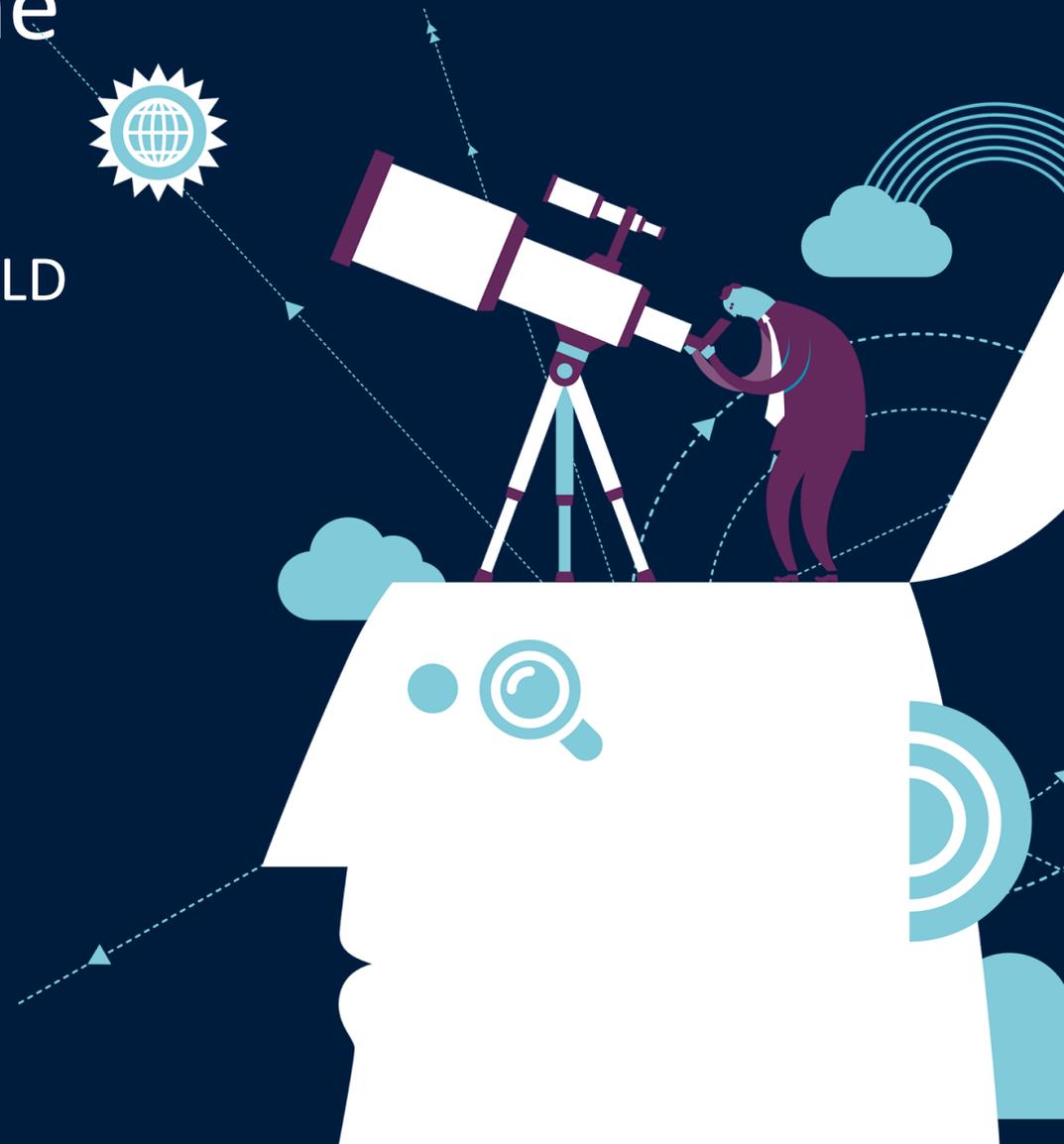
Sonja Boman i6138089

Bachelor Thesis Biomedical Science

Supervisors:

Dr. Martina Kutmon & Dr. Lars Eijssen

03/07



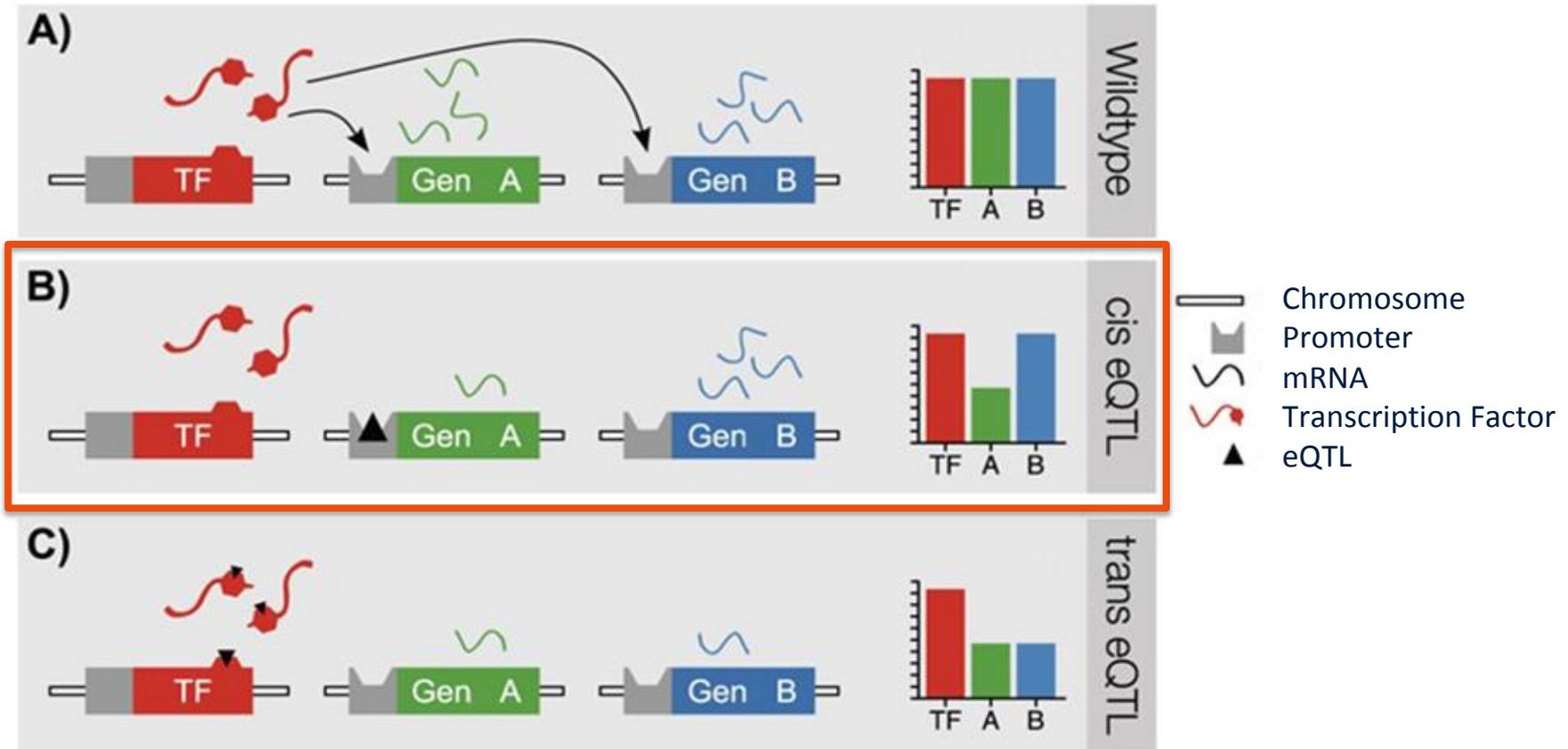
Content

- Introduction
 - eQTL
 - Research Focus
- Materials and Methods
 - GTEx dataset
 - General workflow
- Results and Discussion
 - eQTL profiles of metabolic tissues
 - Investigation into NAFLD missing heritability
- Conclusion and Future Directions



Introduction into eQTLs

Expressive quantitative trait loci



Research Aim

1. Uncover eQTL profiles of metabolic tissues

- RQ: What are the similarities and differences in eQTL regulation between metabolic tissues?
 - Liver
 - Adipose Tissue
 - Skeletal muscle

2. Demonstrate application of eQTL studies

- RQ: Can eQTL analysis contribute to address missing heritability in multi-factorial diseases?
 - Case study on missing heritability in NAFLD

Materials and Methods



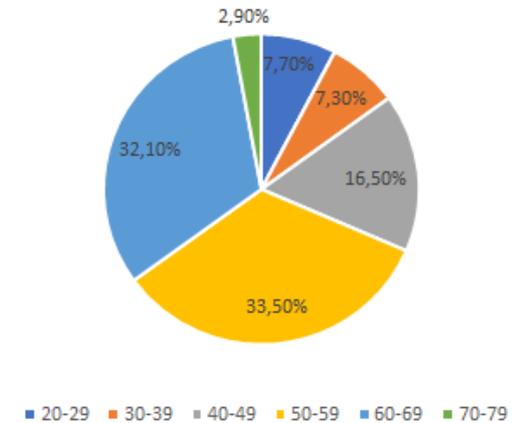
GTEx Data and Data Retrieval

- V7 release
 - 53 tissues from 714 donor
- Data retrieval: Preprocessed eQTL data

Tissue	Donors with Genotyping
Liver	153
Adipose Tissue	358
Skeletal Muscle	491

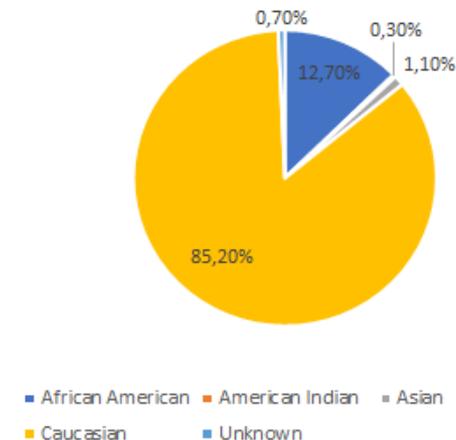
- Pre-processed further:
 - Only protein coding genes
 - Only genes with MAF>0.1

Age Distribution of Donors



- 85% above 40 years old

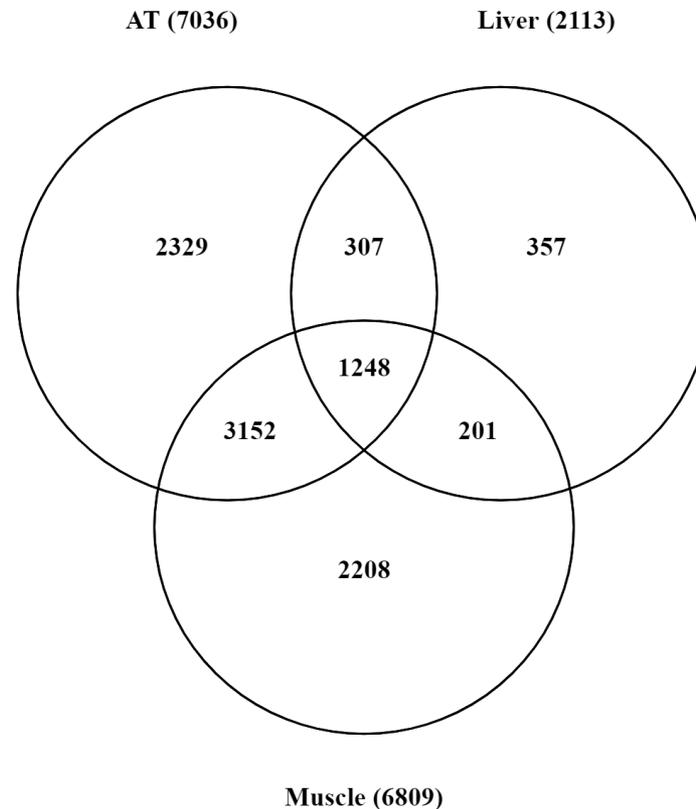
Ethnic Distribution of Donors



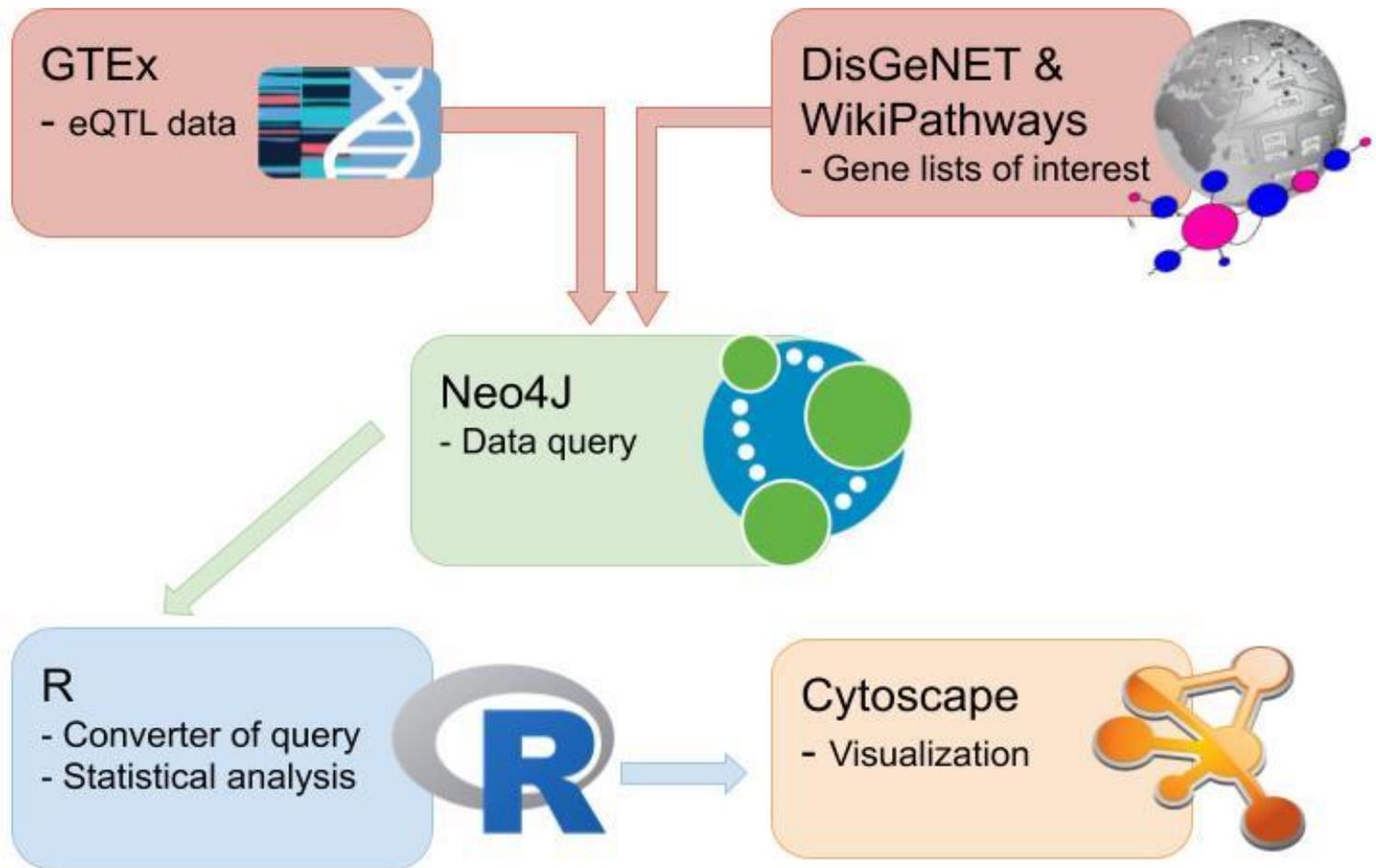
- 85.2% Caucasian

GTEx – Huge Amount of Data

Tissue	Gene Count	eQTL Count	Interaction Count
Liver	2,113	137,261	161,183
Adipose Tissue	7,036	575,672	823,421
Skeletal Muscle	6,809	595,532	869,406



Workflow



Results: General analysis

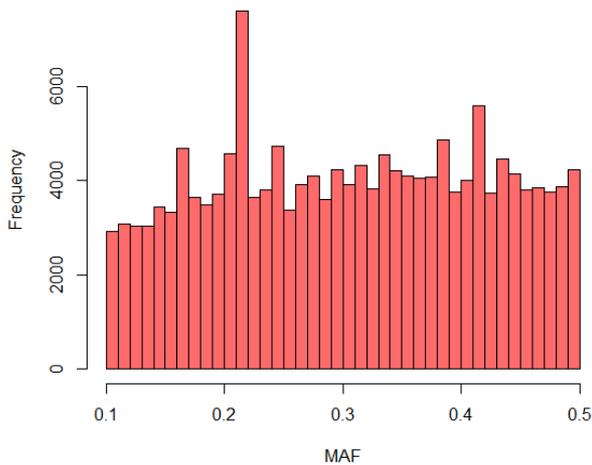


MAF distribution

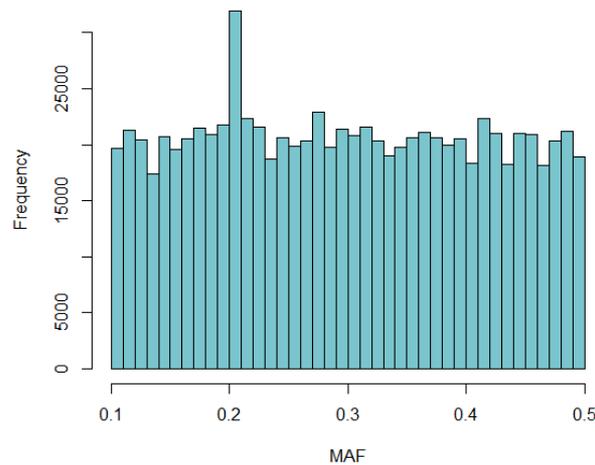
Q: What is the minor allele frequency of eQTLs?

A: MAF 0.2 towers over other values for unexplained reasons.

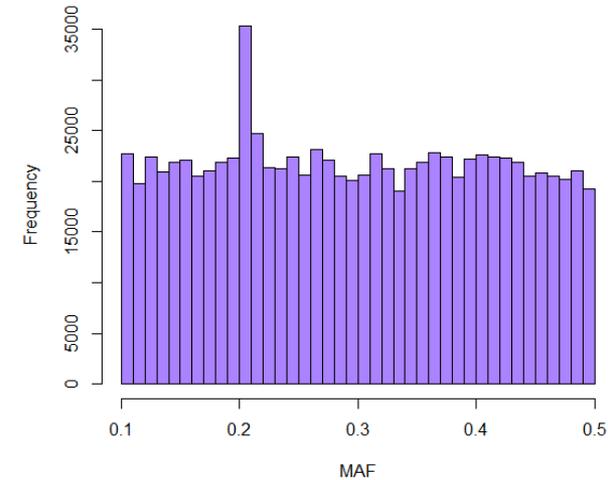
Liver Variants MAF



Adipose Tissue Variants MAF



Muscle Variants MAF

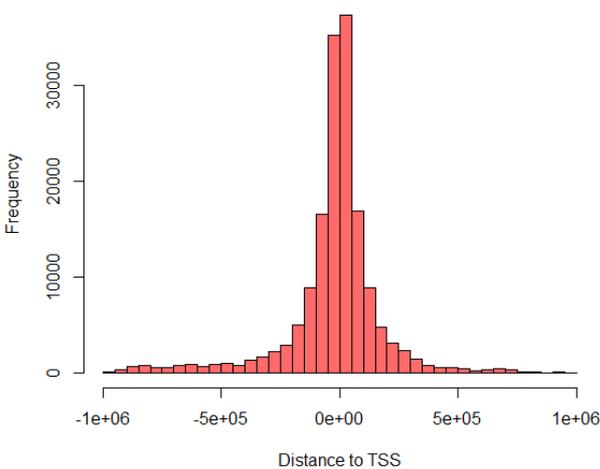


eQTL locations

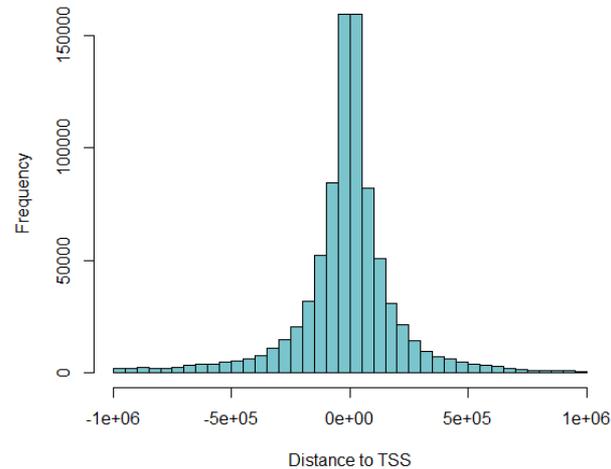
Q: Where are eQTLs located in respect to their eGenes?

A: eQTLs are located increasingly proximate to transcription start site (TSS)

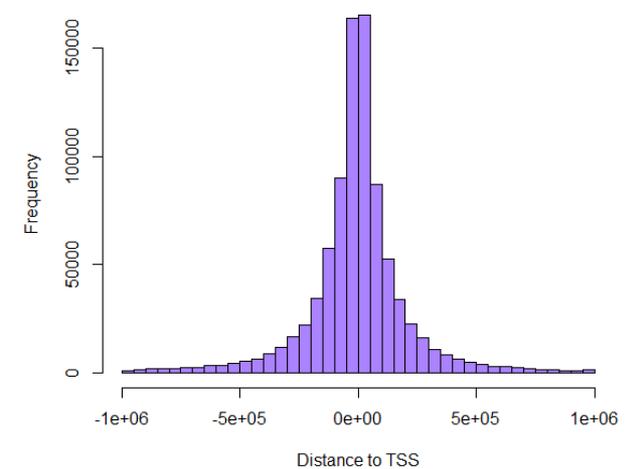
Liver Distance to TSS



Adipose Tissue Distance to TSS



Muscle Distance to TSS

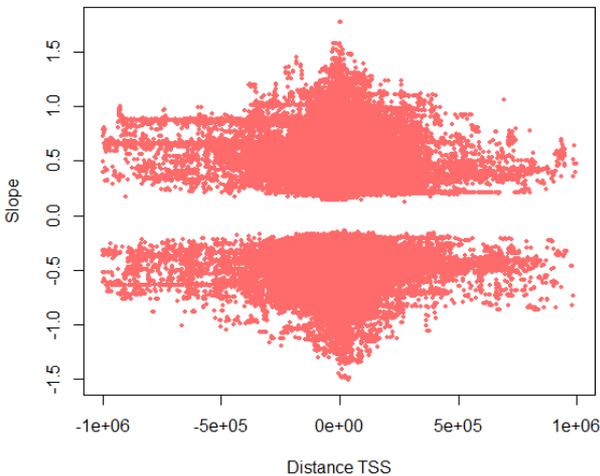


eQTL locations

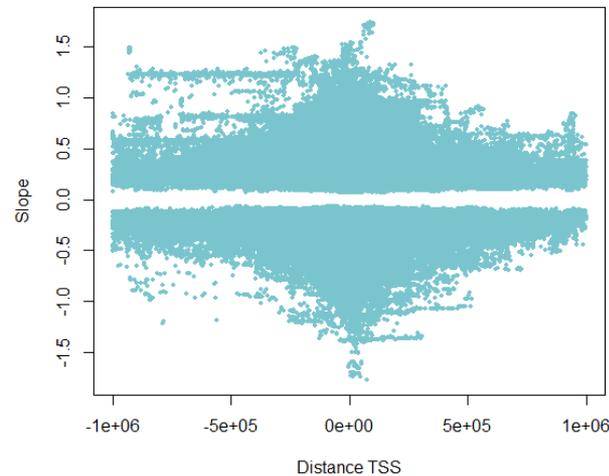
Q: Does proximity to eGenes influence effect size?

A: eQTLs closer to TSS have a larger effect size.

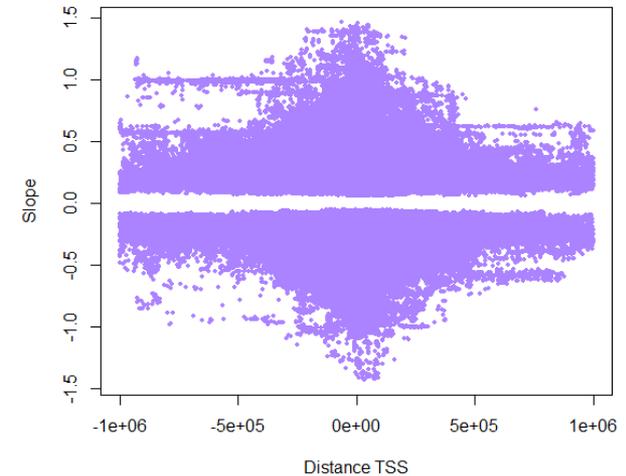
Liver eQTL slope / TSS distance



Adipose Tissue eQTL slope / TSS distance



Muscle eQTL slope / TSS distance

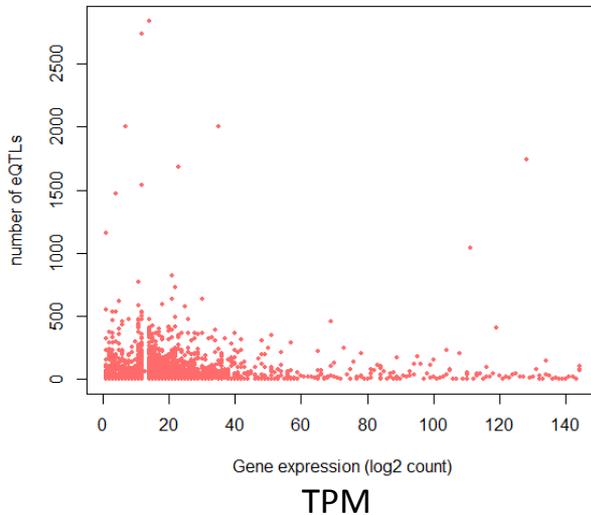


Gene expression vs. eQTL count

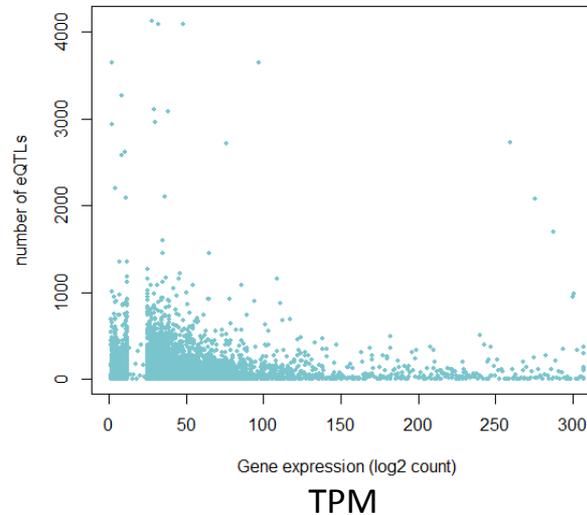
Q: Does gene expression level impact eQTL count?

A: Gene expression level does not impact eQTL count.

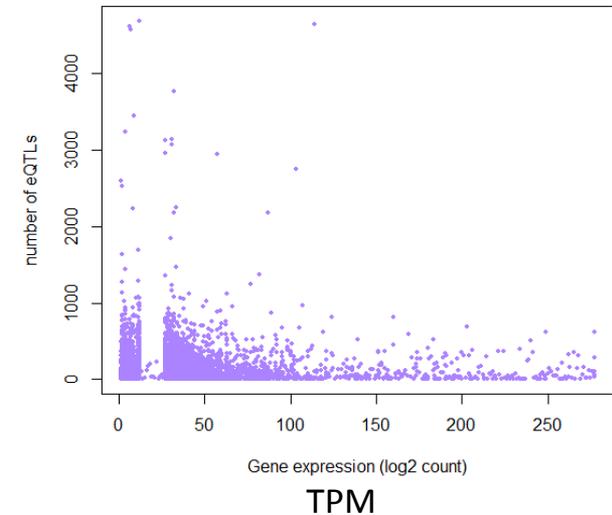
Liver expression vs. eQTL count



Adipose Tissue expression vs. eQTL count



Muscle expression vs. eQTL count

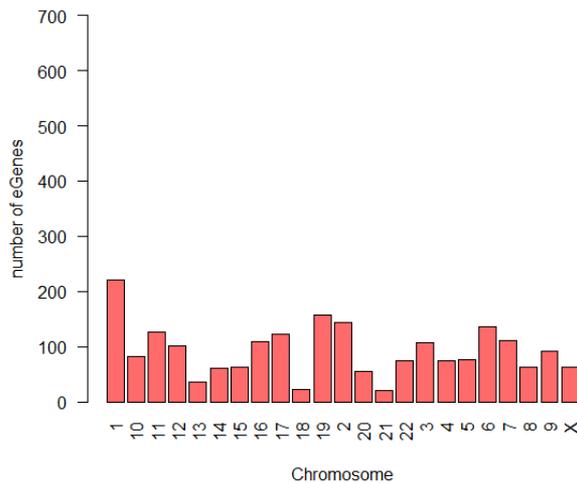


Gene location

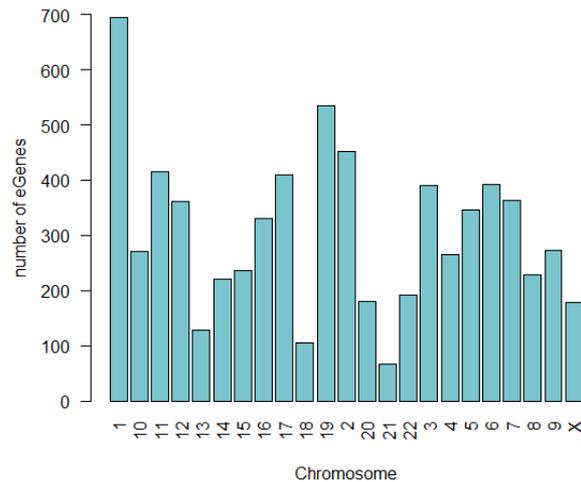
Q: What is chromosomal distribution of genes influenced by eQTLs?

A: Chromosome 1 has highest eGene count, chromosome 21 lowest.

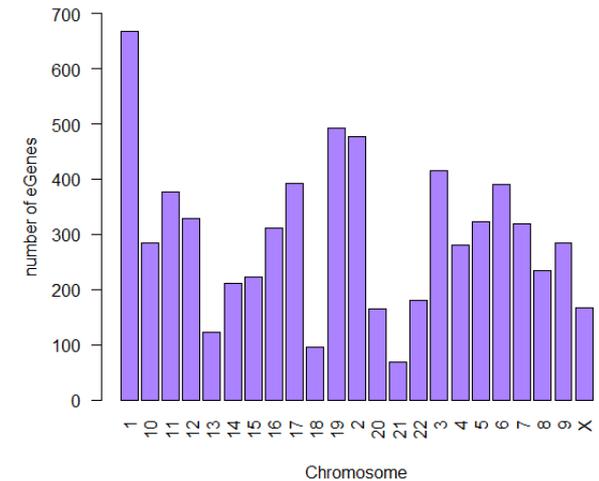
Liver chromosomal location (eGenes)



Adipose Tissue chromosomal location (eGenes)



Muscle chromosomal location (eGenes)



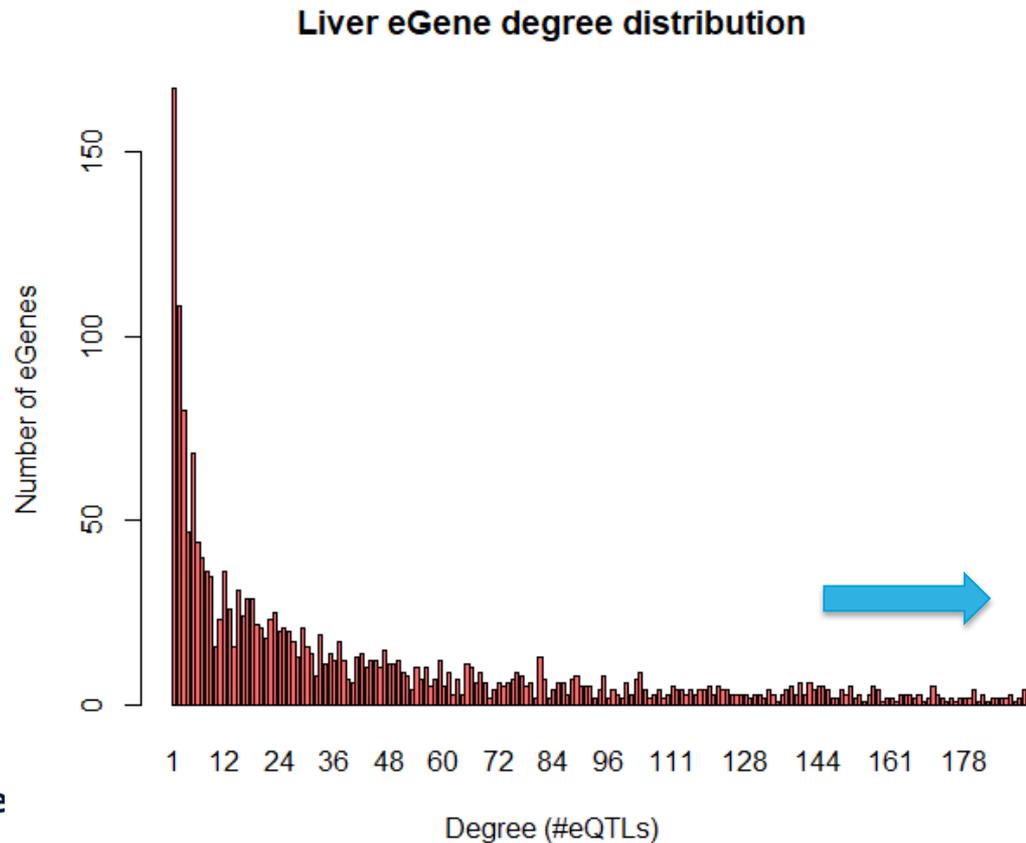
Results

Liver eQTL Analysis



Liver eGenes

- Degree distribution of eGenes ranged between 1-2841



Liver eGene Hubs

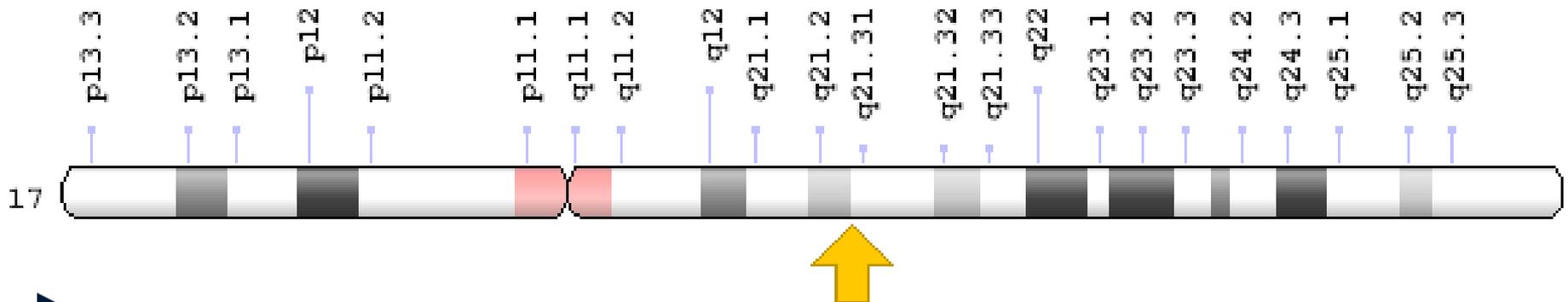
Degree (# eQTLs)	Gene	Position
2841	LRRC37A2	17q21.31
2734	ARL17A	17q21.31
2004	C4A	6p21.33
2002	HLA-DQA2	6p21.32
1737	HLA-C	6p21.33
1682	HLA-DRB5	6p21.32
1541	HLA-DRB1	6P21.32
1468	HLA-DQB2	6p21.32
1158	ZFP57	6p22.1
1040	HLA-A	6p22.1

Metabolic Tissue eGenes Hubs

Liver	Adipose Tissue	Skeletal Muscle
LRRC37A 17q21.31	HLA-DQB2 6p21	HLA-DQB1 6p21
ARL17 17q21.31	HLA-DQA1 6p21	HLA-C 6p21
HLA-DQB2 6p21	HLA-DQB1 6p21	HLA-DQB2 6p21
C4A 6p21	ZFP57 6p21	HLA-DRB5 6p21
HLA-DQA2 6p21	HLA-DRB1 6p21	HLA-DQA1 6p21
HLA-C 6p21	LRRC37A 17q21.31	LRRC37A 17q21.31
HLA-DRB5 6p21	LRRC37A 17q21.31	KANSL1 17q21.31
ZFP57 6p21	KANSL1 17q21.31	LRRC37A 17q21.31
HLA-DQB1 6p21	ARL17 17q21.31	PLEKHM1 17q21.31
HLA-A 6p21	CRHR1 17q21.31	ARL17 17q21.31

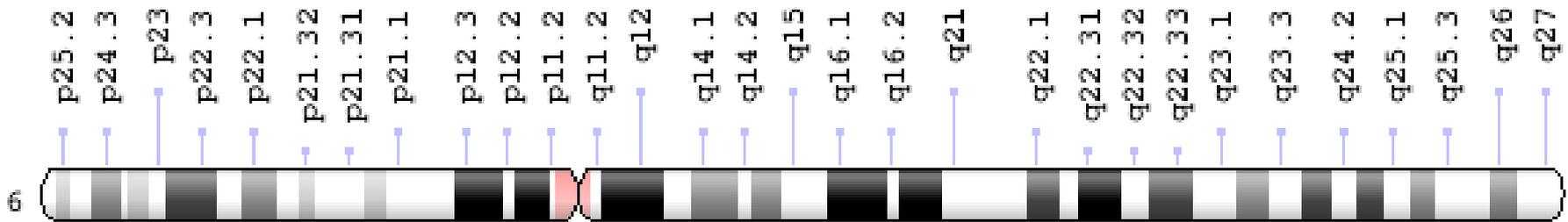
Liver eGenes: 17q21.31

- Locus has 9 genes and 3609 eQTLs
- Functionally genes are part of **cell cycle**
- Band is known for inversion polymorphisms
- Dysregulation associated with neurological conditions and cancers

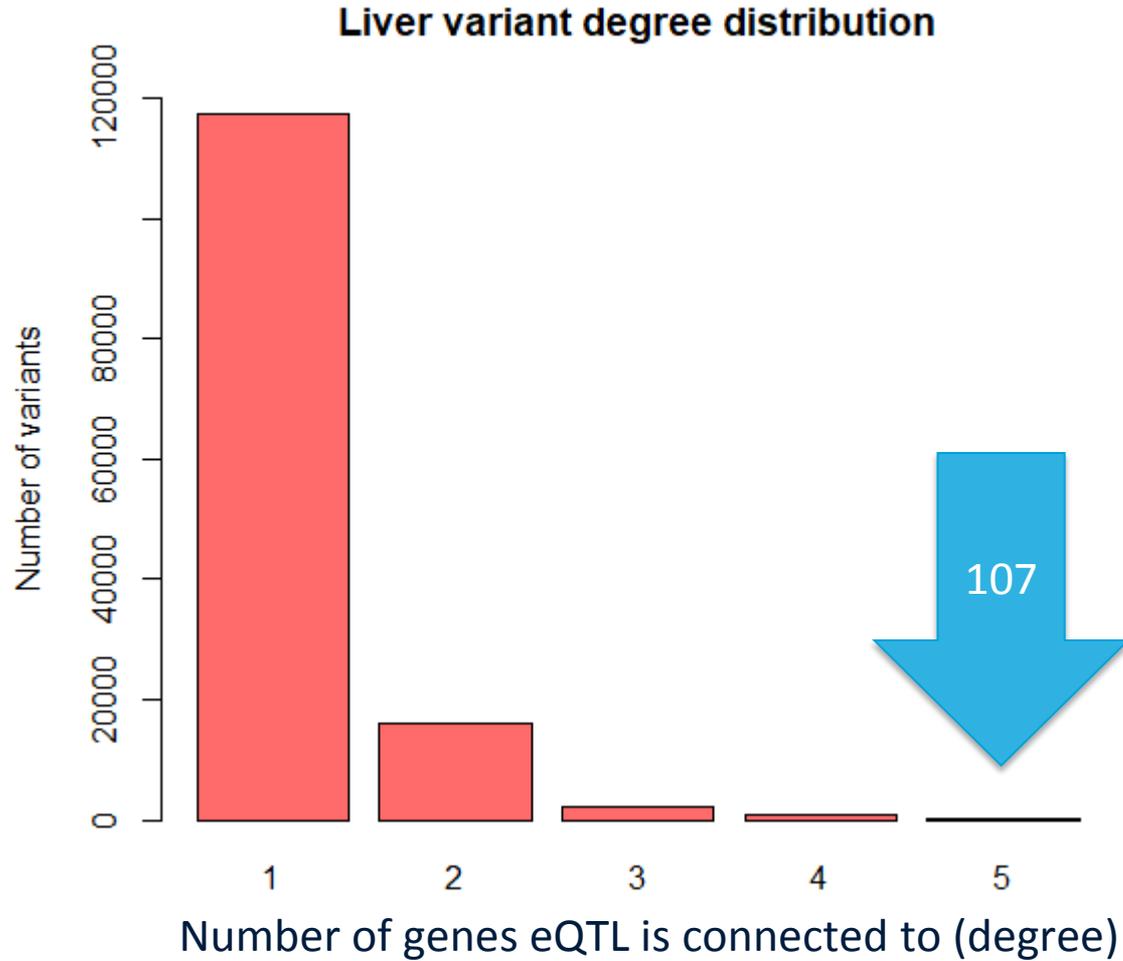


Liver eGenes: 6p21.31-p21.33

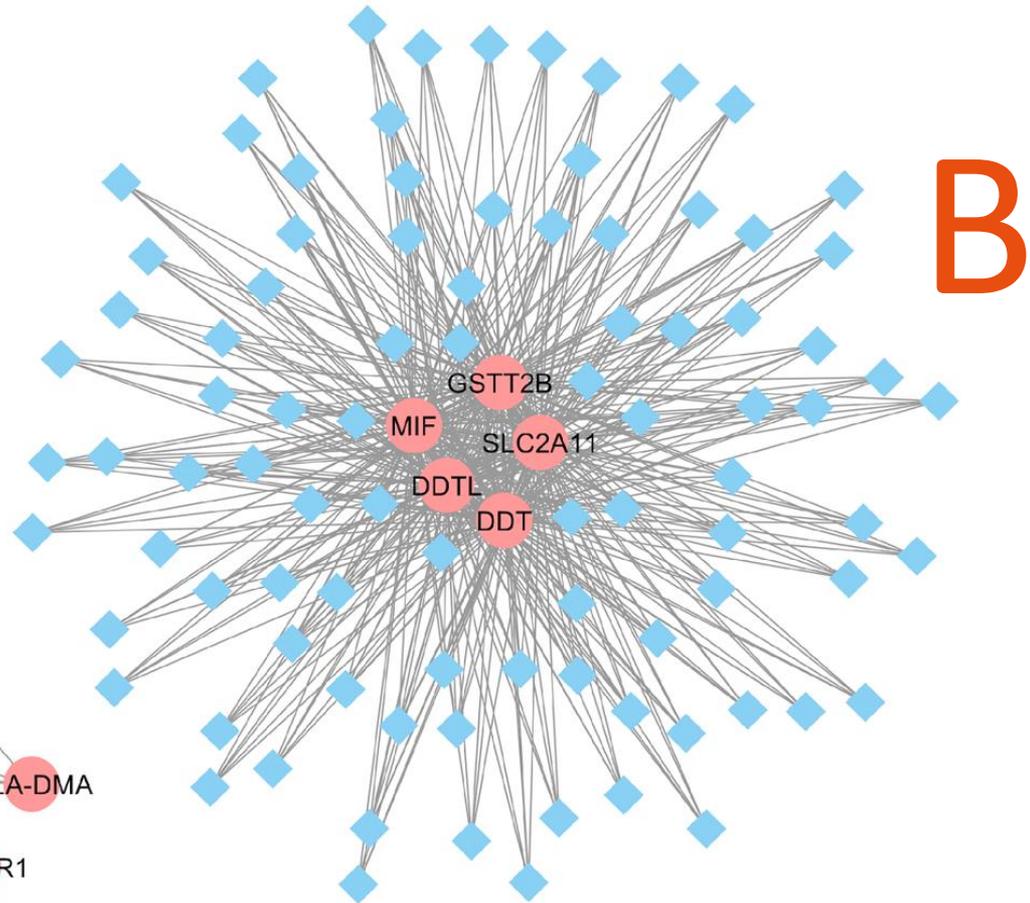
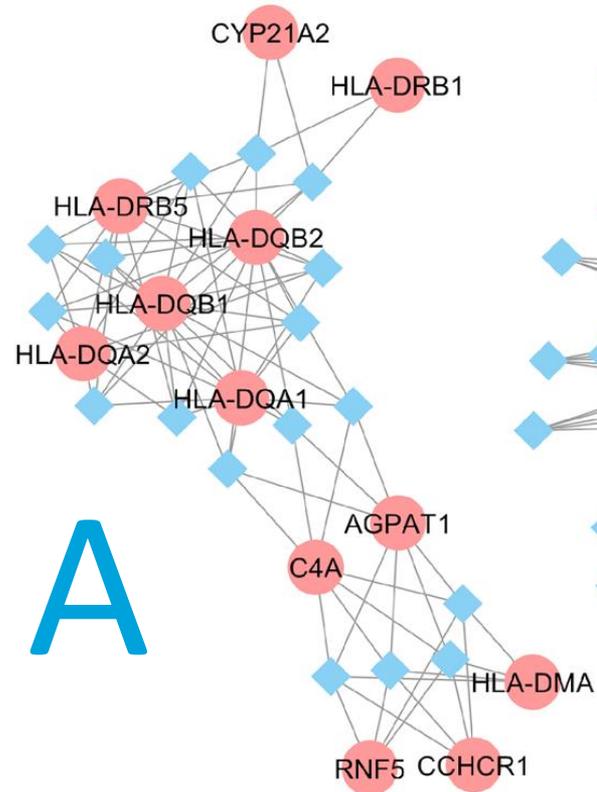
- Human Leukocyte Antigen (HLA)-region
- 24 eGenes and 10,829 eQTLs
- Harbors MHC protein coding genes (HLA-genes) of the **immune system**
- Dysregulation associated with numerous diseases, importantly autoimmune diseases



Liver variants



Liver Variants



Metabolic Tissue Variants

- All metabolic tissues had 2 clusters for eQTLs with highest degree

Tissue	Position cluster A	Position cluster B
Liver	6p21	22q11
Adipose Tissue	6p21	3p21.31
Skeletal Muscle	6p21	3p21.31



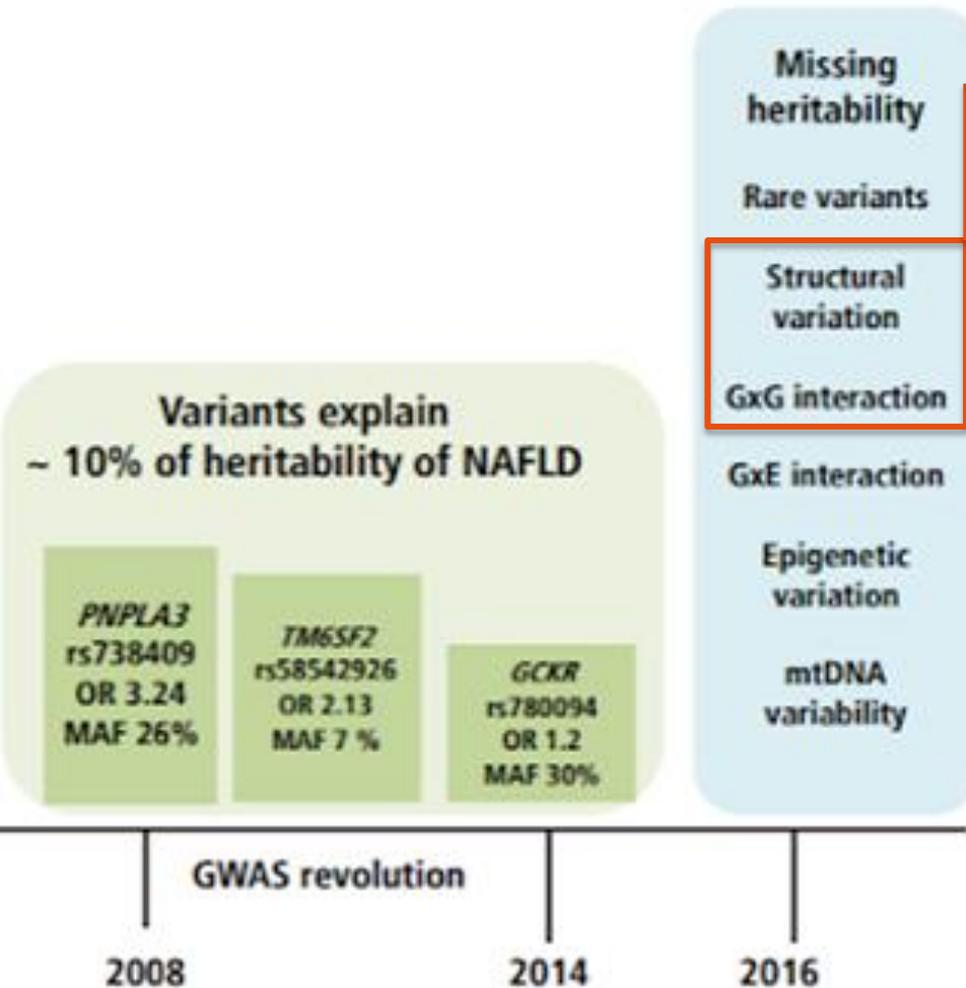
- Cell Cycle
- Immune System
- Metabolism

Results

Missing Heritability: NAFLD



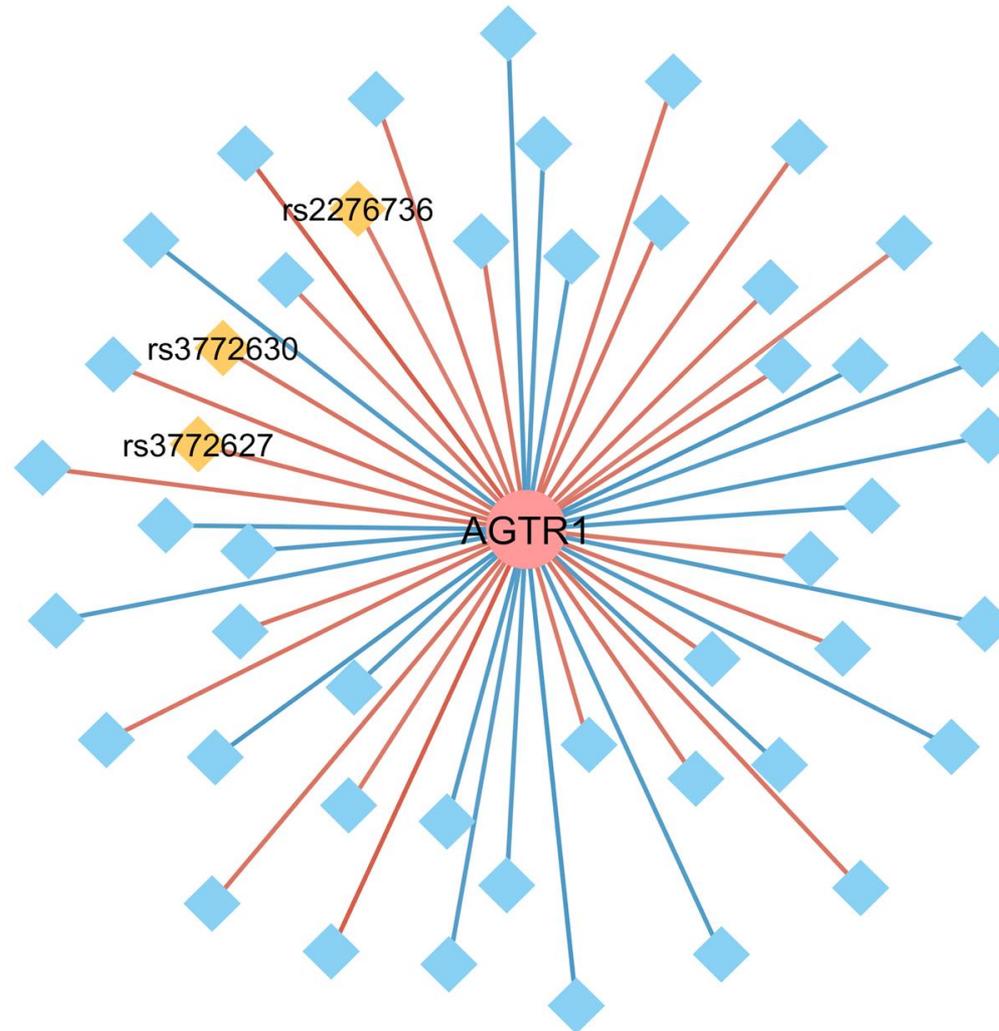
NAFLD gene set



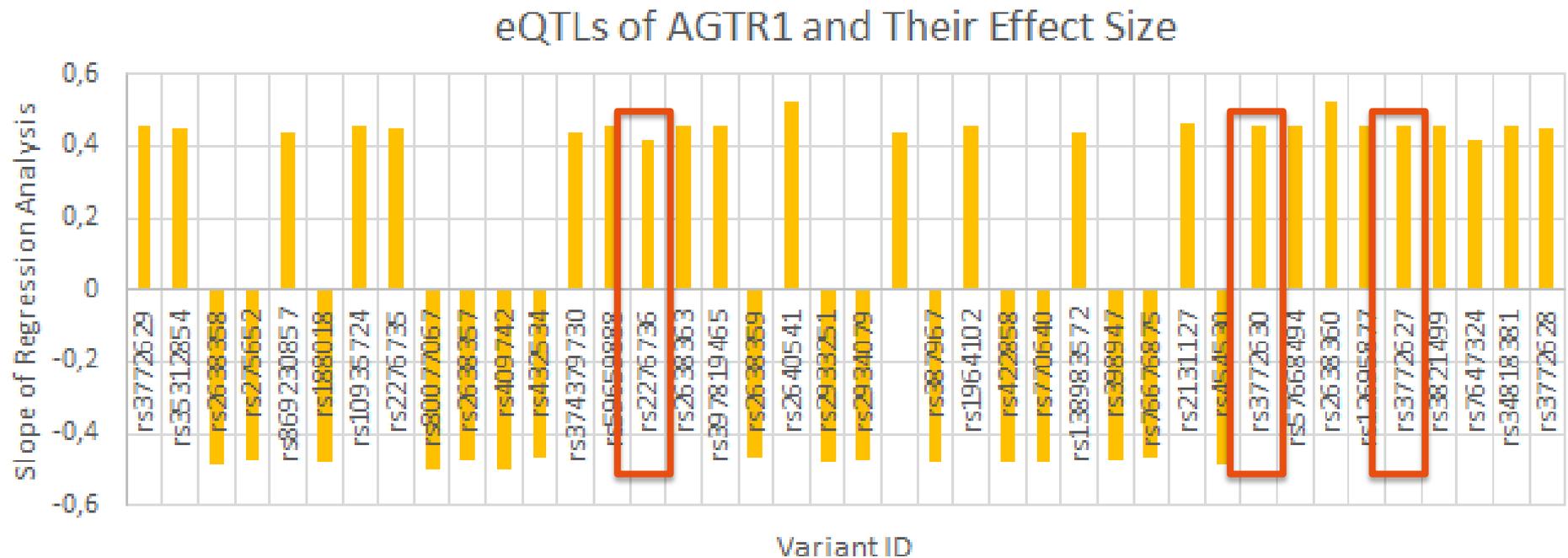
- Combined data from WikiPathways and DisGENet provided 460 genes
- 43 were found to be regulated by liver eQTLs
- Functionally related to energy homeostasis
- The genes did not share eQTLs with each other

<https://pdfs.semanticscholar.org/4039/4cb4472673f54decbe10446339754af939ab.pdf>

eQTLs as Susceptibility Genes: AGTR1



AGTR1 eQTLs



Conclusion

RQ: What are the similarities and differences in eQTL regulation between metabolic tissues?

- eQTLs are located increasingly close to eGenes
- Proximity of eQTLs influences effect size
- Gene expression level does not impact number of eQTLs
- Densely regulated regions are 6p21 and 17q21 as well as 3q21 or 22q11 depending on tissue
 - Functionally harbor genes of immune system, cell cycle and metabolism

Conclusion

RQ: Can eQTL analysis contribute to address missing heritability in multi-factorial diseases?

- 43 of 460 NAFLD susceptibility genes are regulated by eQTLs in liver
- 3 disease-variant associations are eQTLs of Angiotensin Receptor 1 -gene
- AGTR1 has 23 eQTLs as candidate susceptibility genes

Limitations

- Small sample size in liver tissue (n=153)
- Ethnicity is not controlled for
- Donor data is from older people
 - Less eQTLs in older people found

Future directions

- Investigate eQTLs behind common metabolic processes in different metabolic tissues
- Investigate eQTLs of susceptibility genes of other diseases to uncover missing heritability
- Study power can be increased by joining eQTL data from other data in “mega-analysis”
- Liver eQTLs hold a strong foundation to pharmacogenomic research as a primary site of xenobiotic metabolism

Thank you
BiGCaT
Dr. Martina Kutmon
Dr. Lars Eijssen
Dr. Nuno Nunes

Questions?



Neo4J

- Data structuring and storage
- Property graph structure
- Data can be queried using Cypher Query Language
- Return graphs, tables, statistics

Example query: get subnetwork of all genes with >5 eQTL variants

```
MATCH (g:Gene)-[g1:GTEX_ADIPOSE]-(v:Variant)
WITH v, count(g) AS edges, collect (g) AS genes
WHERE edges=13
RETURN v, genes, edges
```



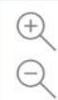
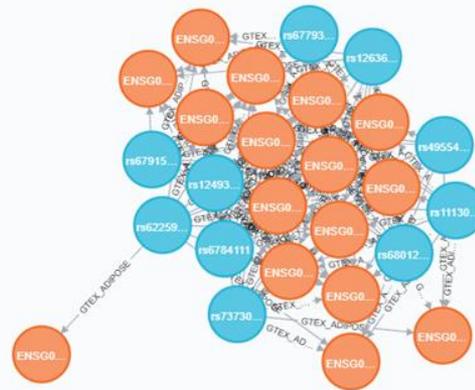
Neo4J Example Query

```
$ MATCH (g:Gene)-[g1:GTEX_ADIPOSE]-(v:Variant) WHERE v.adipose=TRUE WITH v, count(g) as edges, collect (g) AS genes WHERE edges=1...
```



*(40) Variant(11) Gene(29)

*(143) GTEX_ADIPOSE(143)

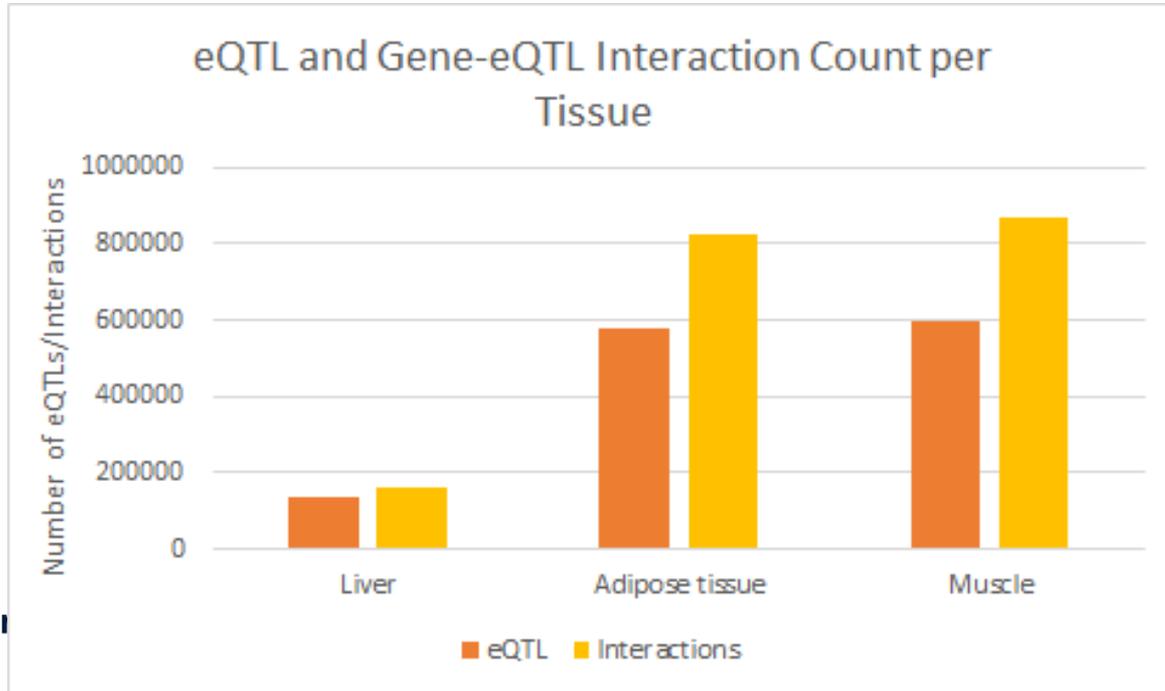
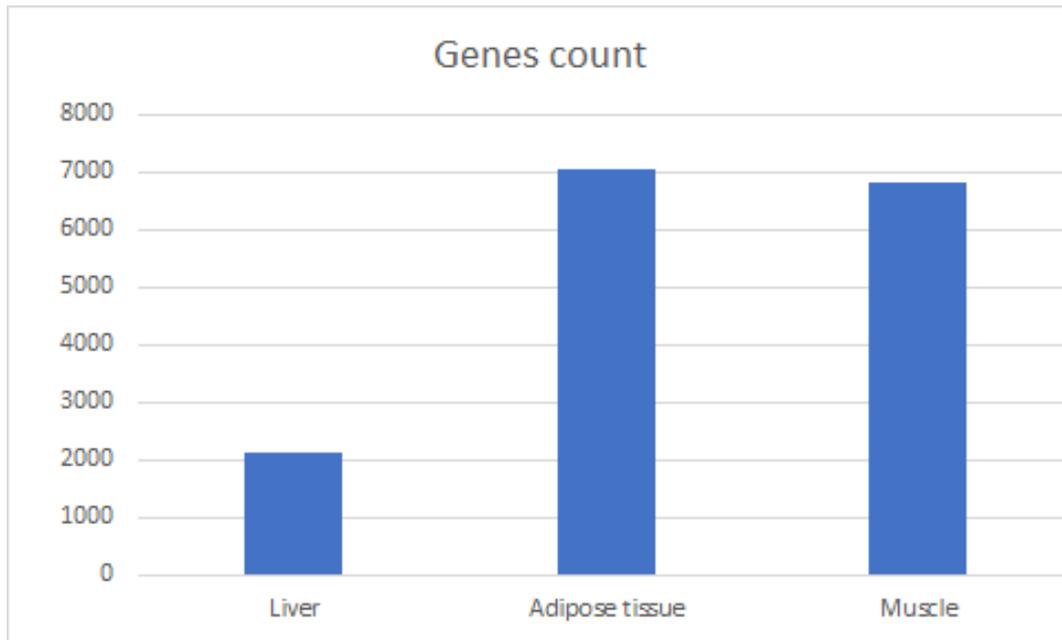


Displaying 40 nodes, 143 relationships.

Automation in R

- Statistics and graphs (some graphs were done in excel)
- Covert Neo4J queries into Cytoscape

```
77 query<- "MATCH (v:Variant)
78 WITH v, size((v)-[:GTEX_MUSCLE]->(:Gene)) as degree
79 WHERE degree=13
80 MATCH (v)-[r:GTEX_MUSCLE]->(g:Gene)
81 RETURN g,r,v"
82 G <- query %>% call_neo4j(con, type = "graph")
83 igraph <- getIGraph(G)
84 createNetworkFromIgraph(igraph)
```



Cytoscape

- Data visualization



Datasets of interest

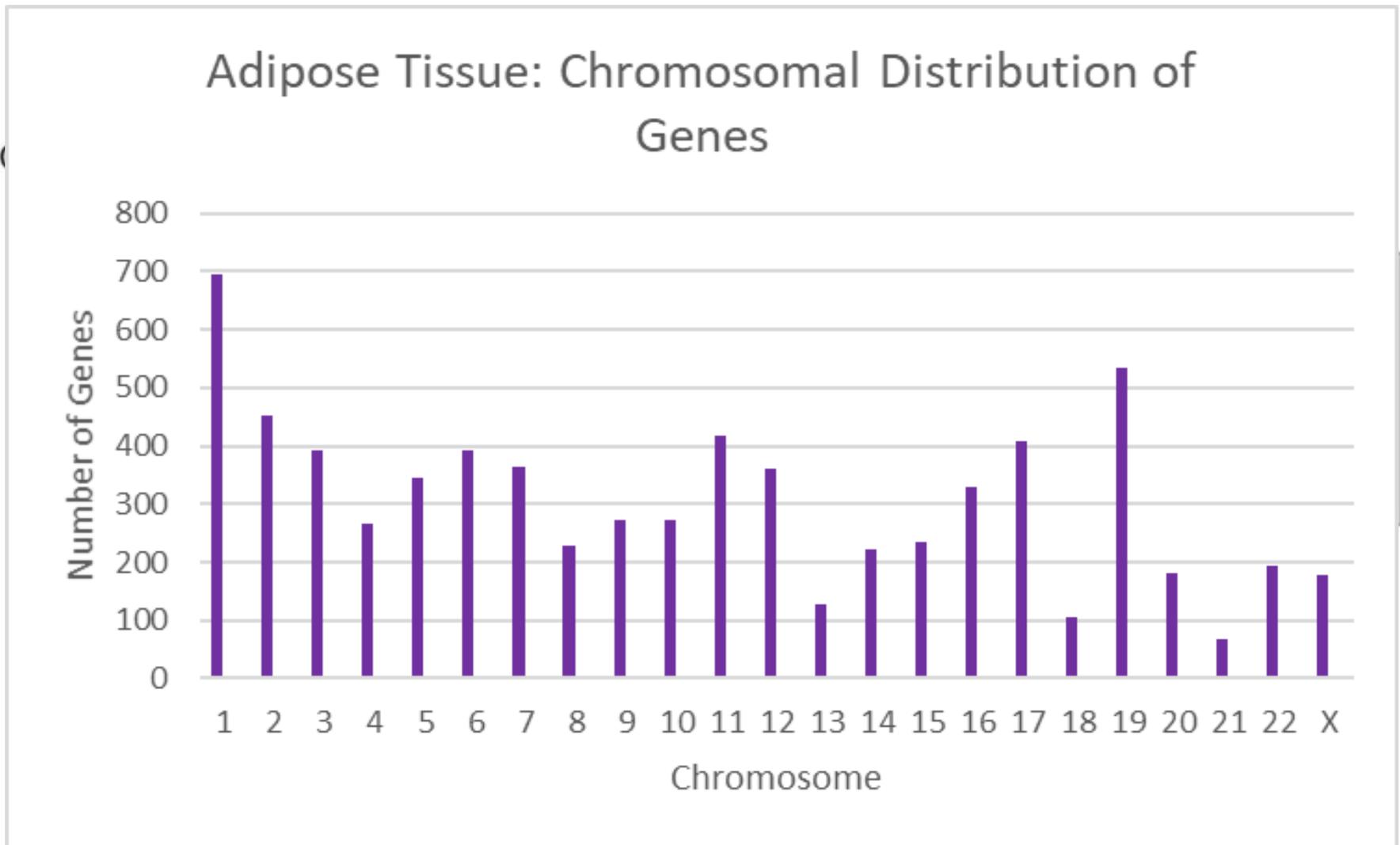
- Downloaded from Disgenet and Wikipathways
- Queried through Neo4J visualized in cytoscape

Results

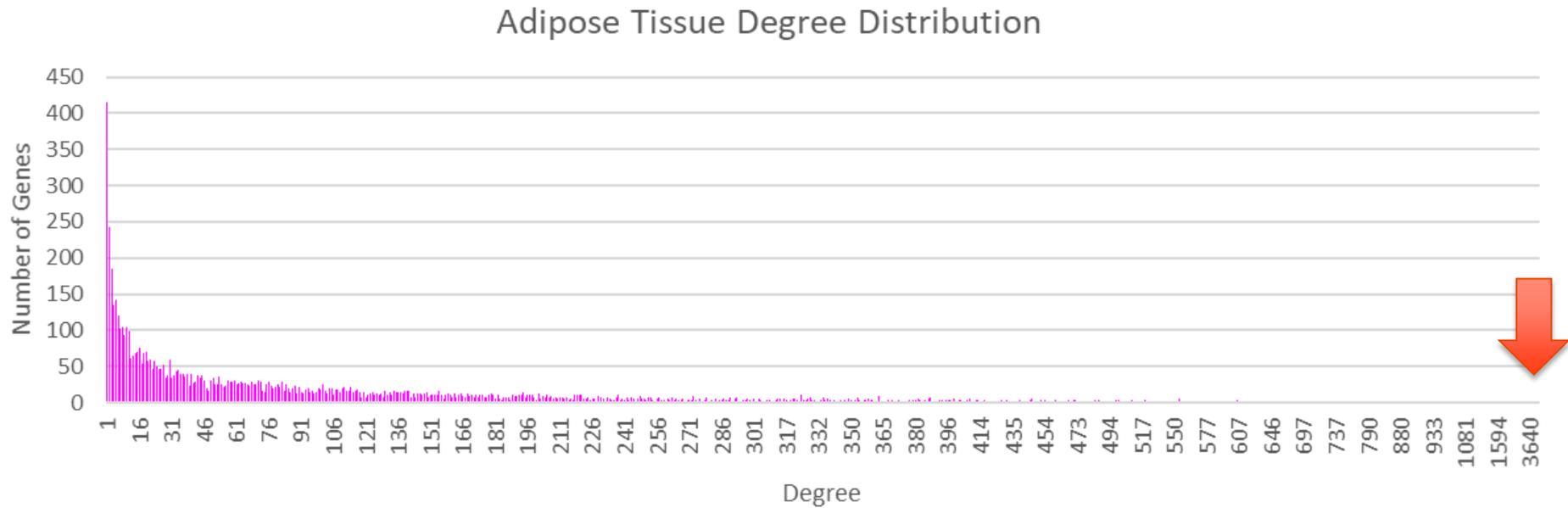
Adipose Tissue eQTL Analysis



Adipose Tissue General eQTL Profile



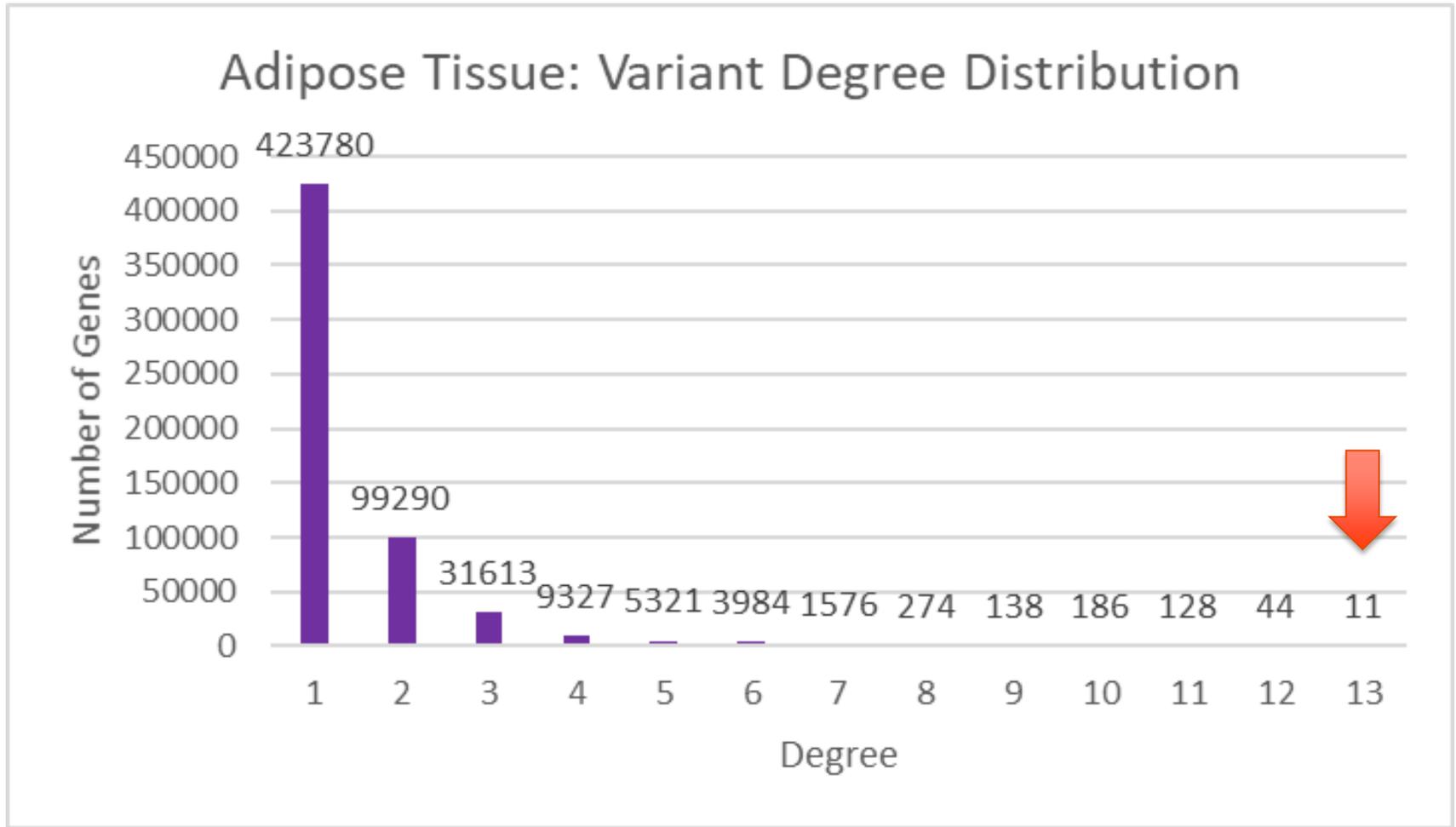
Adipose Tissue eGenes



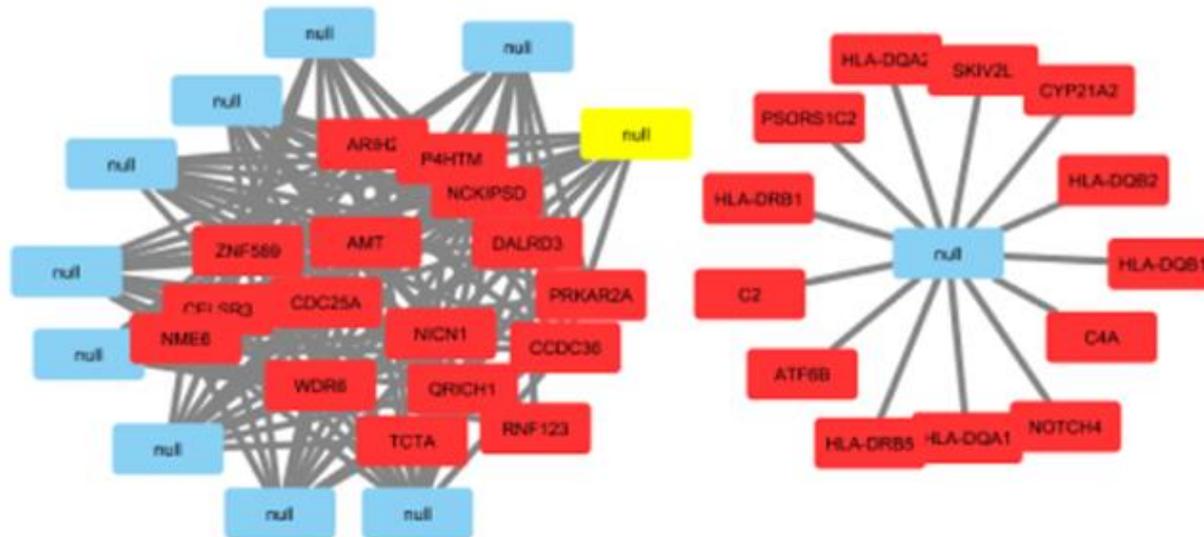
Adipose Tissue eGenes

Degree	Gene	Position
4127	HLA-DQB2	6p21.32
4090	HLA-DQA2	6p21.32
4082	HLA-DQB1	6p21.32
3647	ZFP57	6p22.1
3640	HLA-DRB5	6p21.32
3266	LRRC37A	17q21.31
3105	LRRC37A2	17q21.31
3086	KANSL1	17q21.31
2954	ARL17A	17q21.31
2934	CHR1	17q21.31

Adipose Tissue Variants



Adipose Tissue Variants



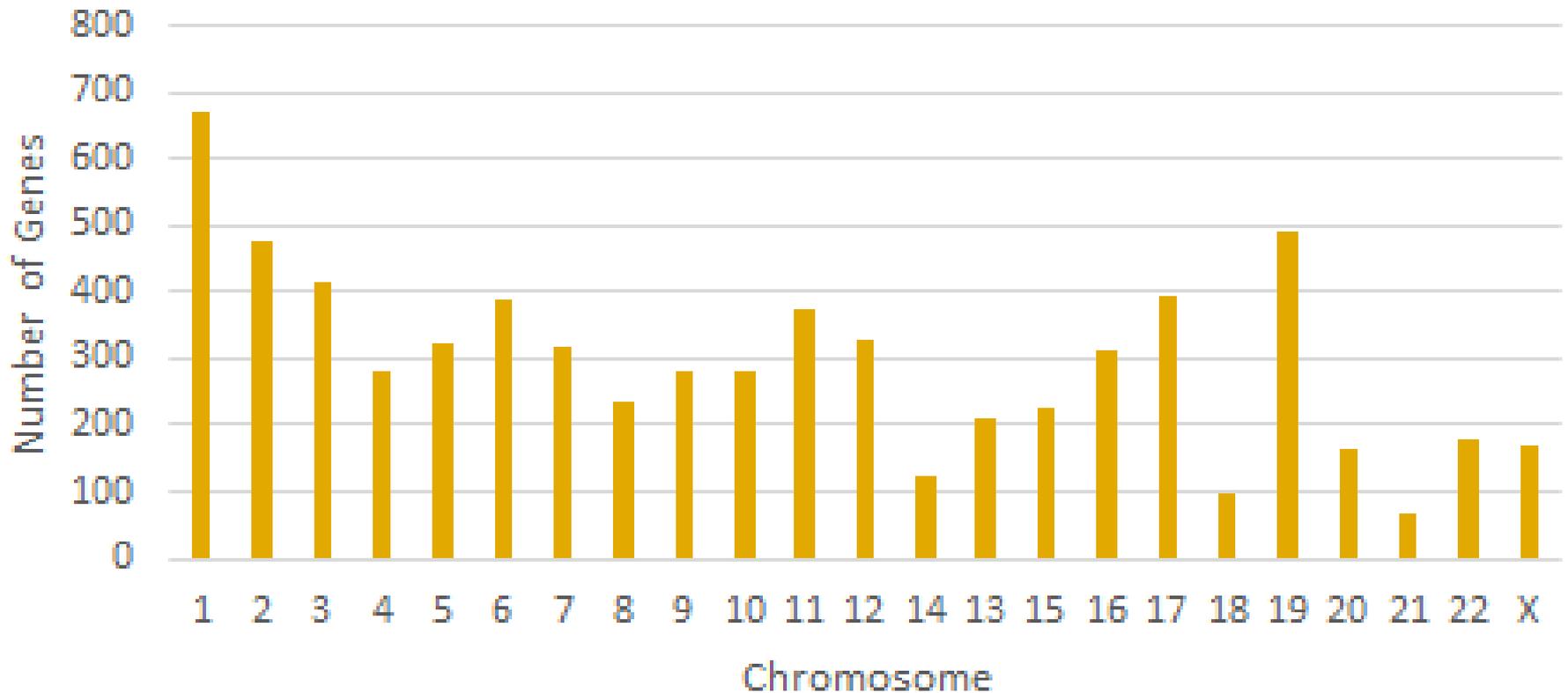
Results

Skeletal Muscle eQTL Analysis



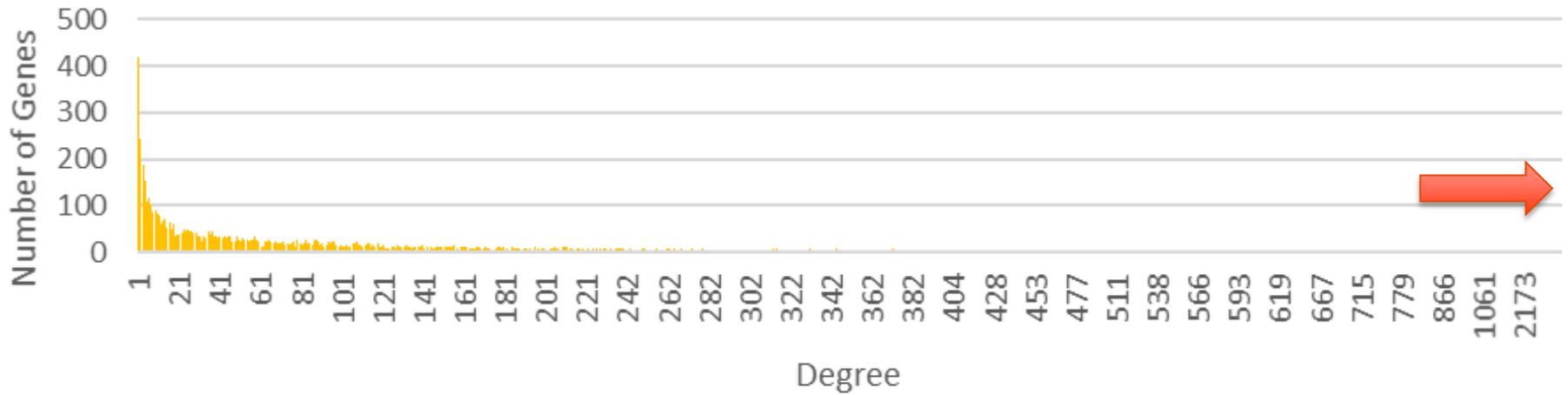
Skeletal Muscle General eQTL Profile

Skeletal Muscle: Chromosomal distribution of Genes



Skeletal Muscle eGenes

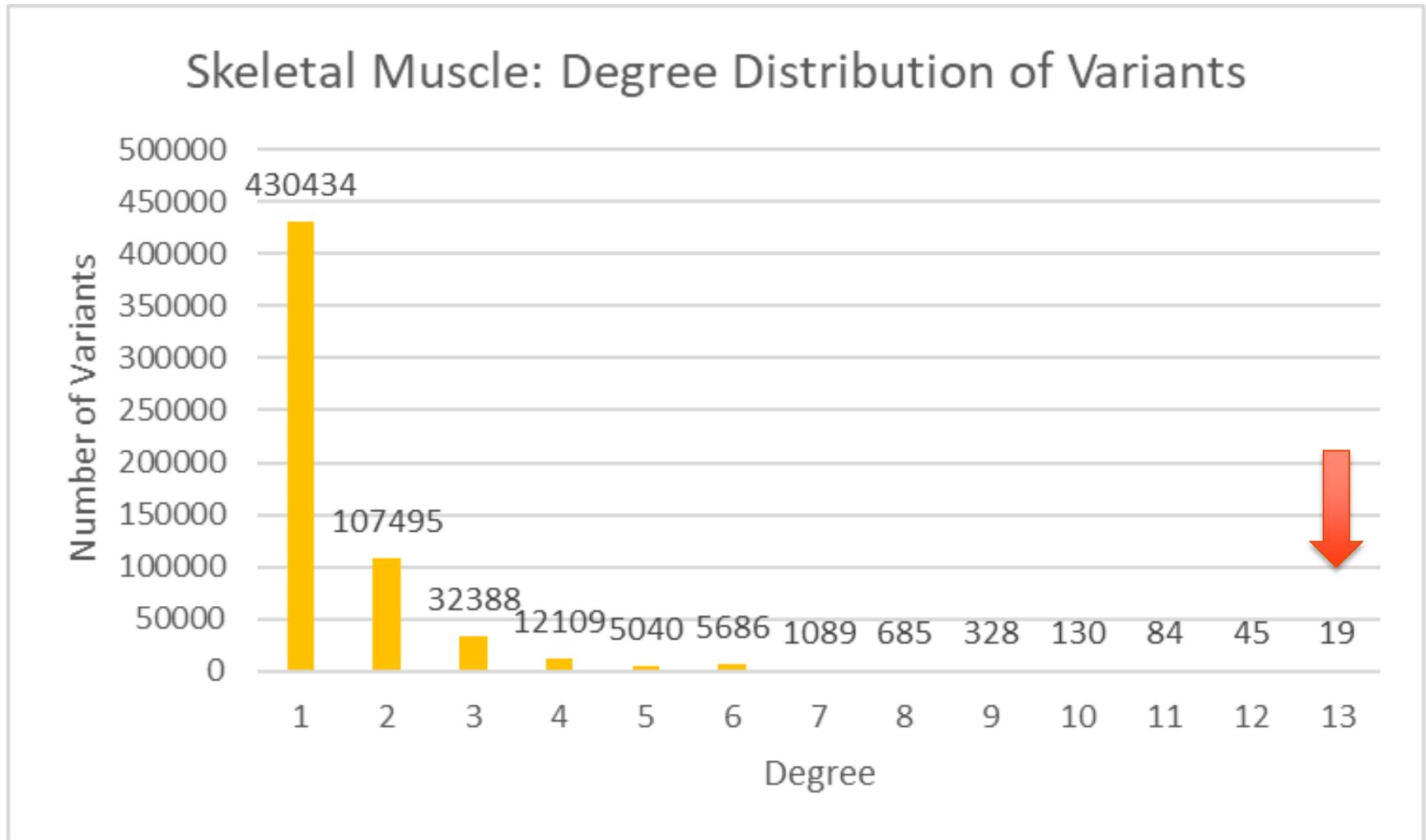
Skeletal Muscle Degree Distribution



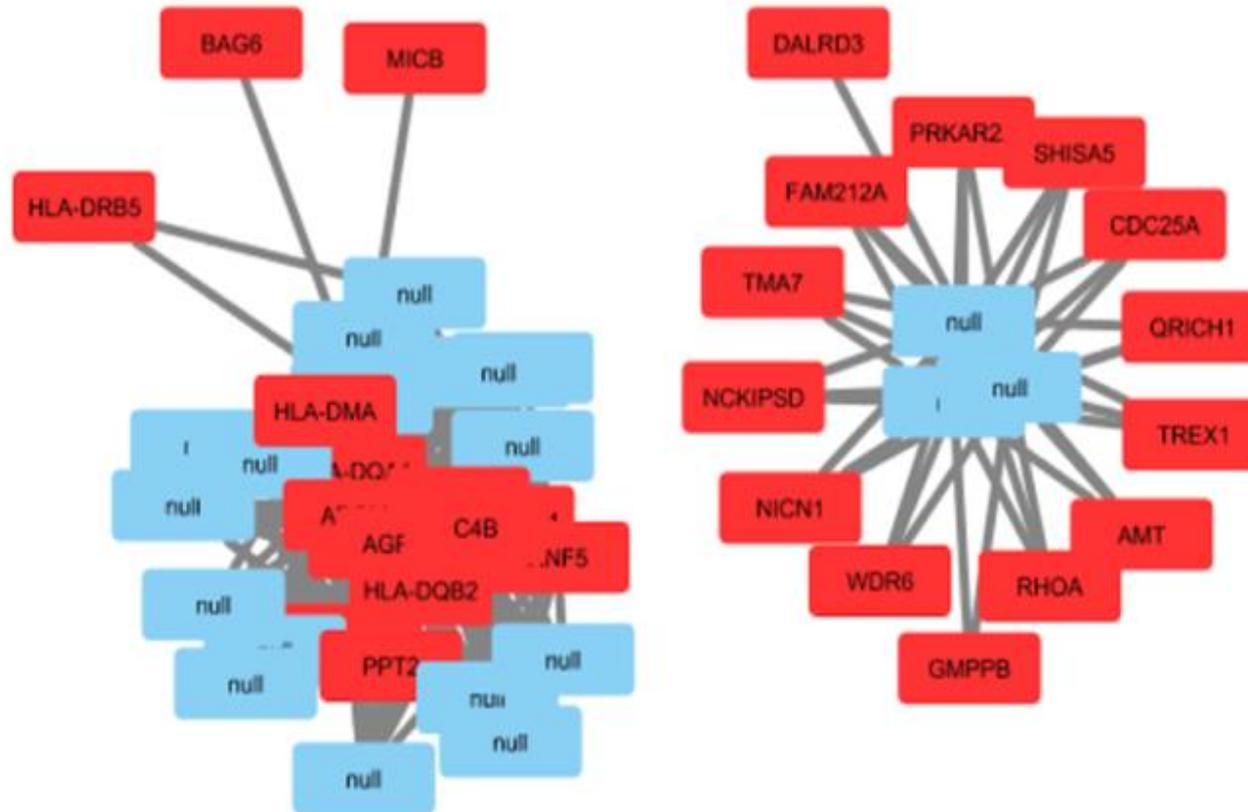
Skeletal Muscle eGenes

Degree	Gene	Position
4695	HLA-DQB1	6p21.32
4643	HLA-C	6p21.33
4620	HLA-DQB2	6p21.32
2575	HLA-DQA2	6p21.32
3764	HLA-DRB5	6p21.32
3448	HLA-DQA1	6p21.32
3234	LRRC37A	17q21.31
3139	KANSL1	17q21.31
3134	LRRC37A2	17q21.31
3077	PLEKHM1	17q21.21

Skeletal Muscle Variants



Skeletal Muscle Variants



- Genes experience co-regulation
- Genes are more closer to each other (due to length?)