Pangenome mapping enhances genotype-phenotype associations in BXD mouse family

Flavia Villani
Ph.D candidate
UTHSC

CTC-RG2024
COMPLEX TRAIT COMMUNITY

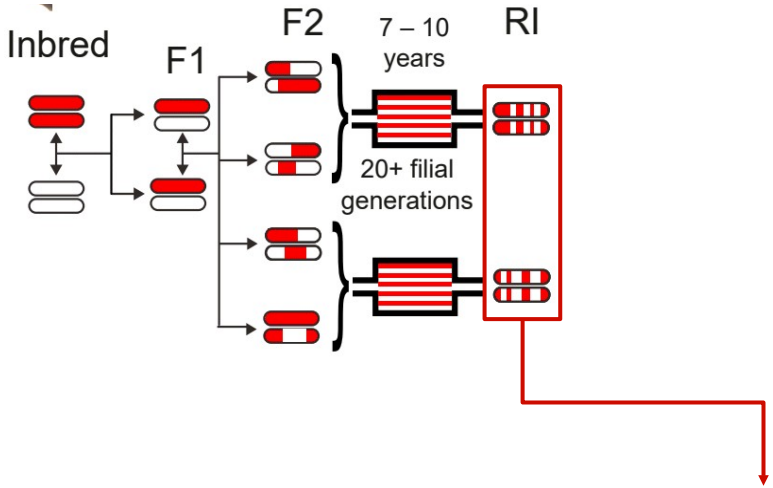
Medical College of Wisconsin, Milwaukee, USA

eQTL mapping in the BXD mouse family

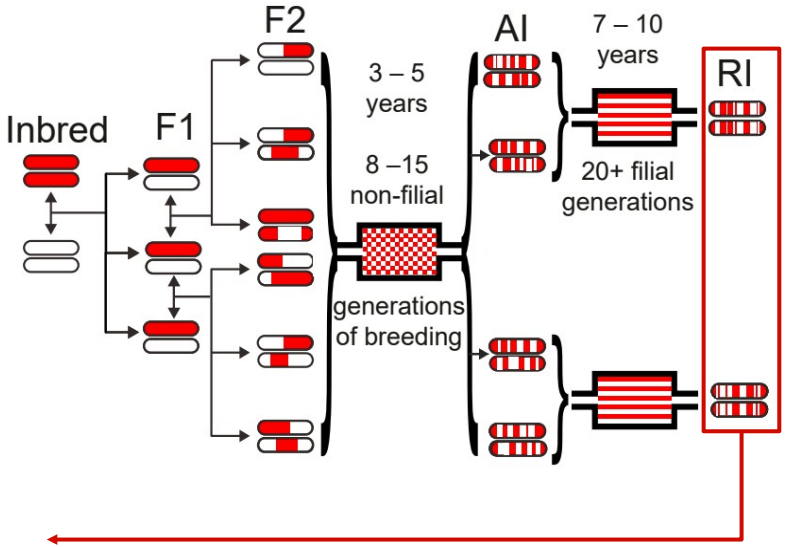
The extended BXD mouse family

C57BL/6J (mother) → **B** X **D** ← DBA/2J (father)

RECOMBINANT INBRED



ADVANCED INTERCROSS



Whole Genome Sequencing
10X Linked reads

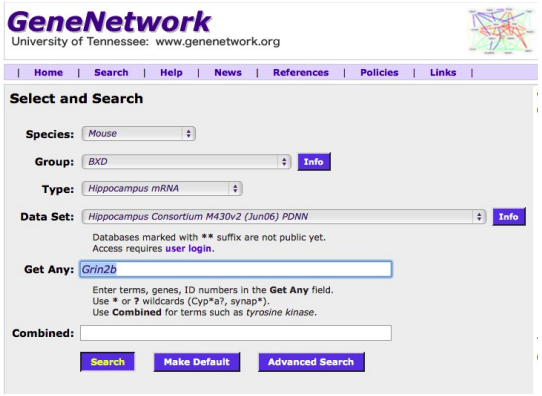
UTHSC

Mulligan, Megan K., et al. *Systems Genetics*. (2017)

Ashbrook, D. G., et al. Cell Systems (2021)

Expression QTL mapping in BXD mice

- 155 BXDs strain
- 21M mapped quantitative traits



Remap eQTL using genetic markers derived from the pangenome graph

UTHSC

Pangenomics

A pangenome is just a set of sequences…

>ATCTCTCGTAACACTGTCAGTCAGTCA
>ACTCTCATTGTGTCCCTGTACTGTCTG
>CGTGCTGCTGATCGGGGTGCTGGCTAG
>CGTGCTGATTGTCGTGGGCTGTAGGTG

…but pangenomics is a paradigm shift in genomics

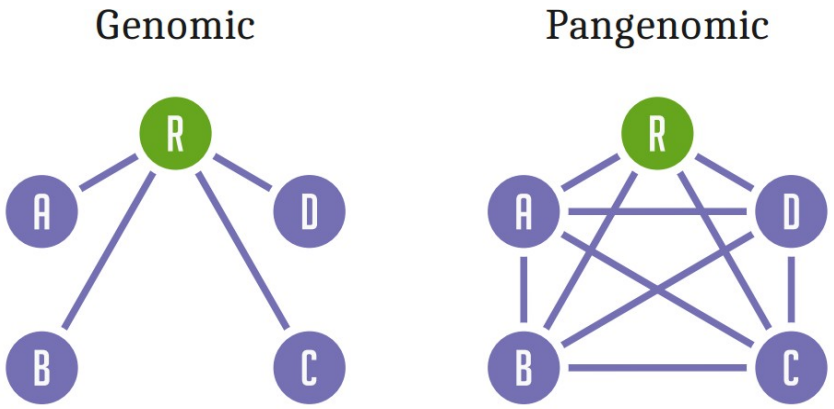


Figure from Eizenga et al., 2020.

Haplotype resolved complete representation of genomes

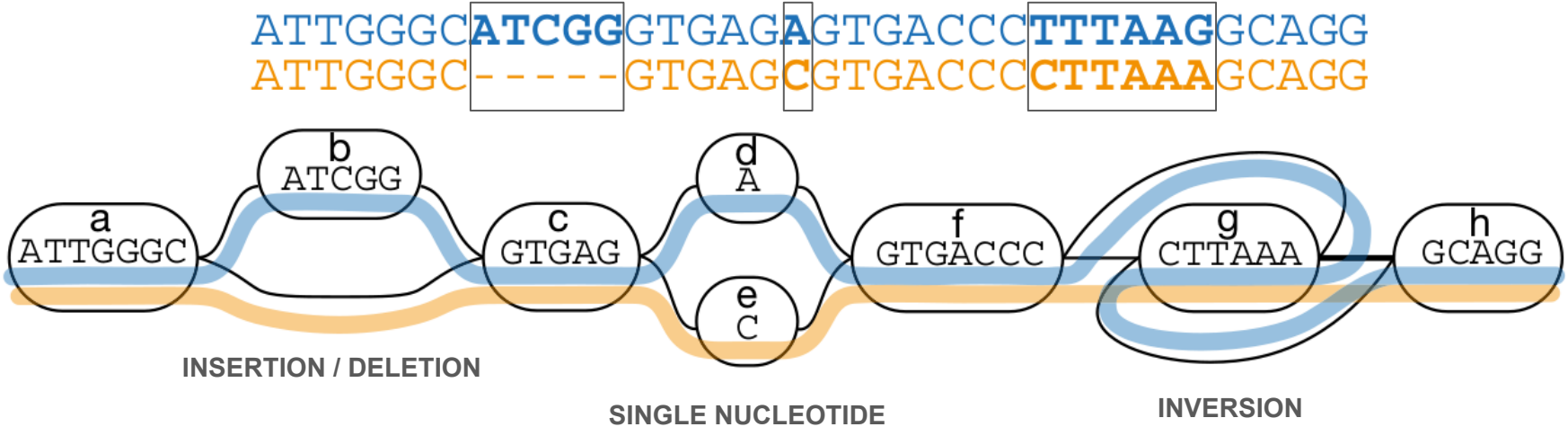
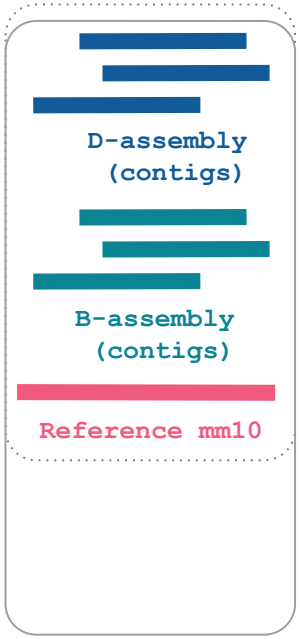


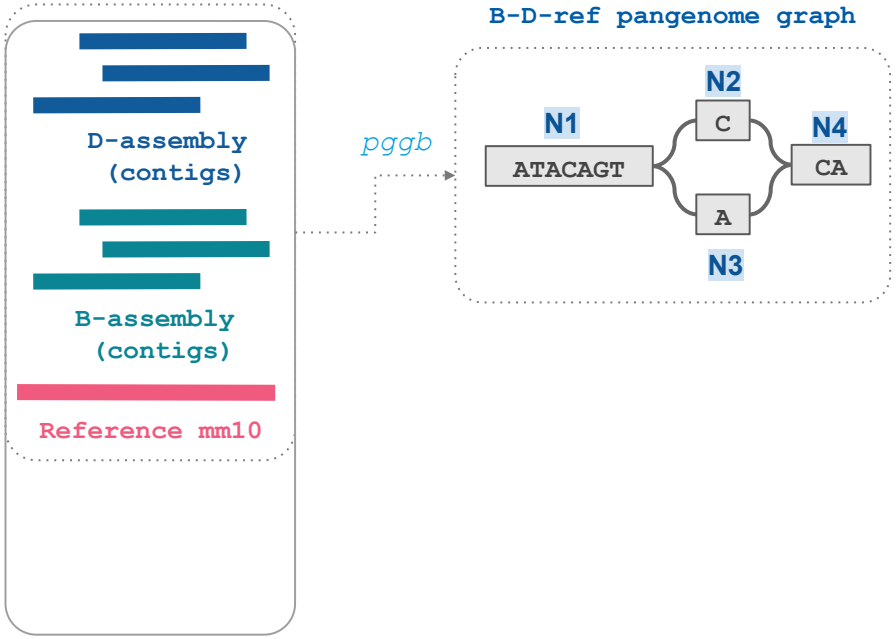
Figure from Liao *et al*, 2023

Mapping short-read *vs* a pangenome

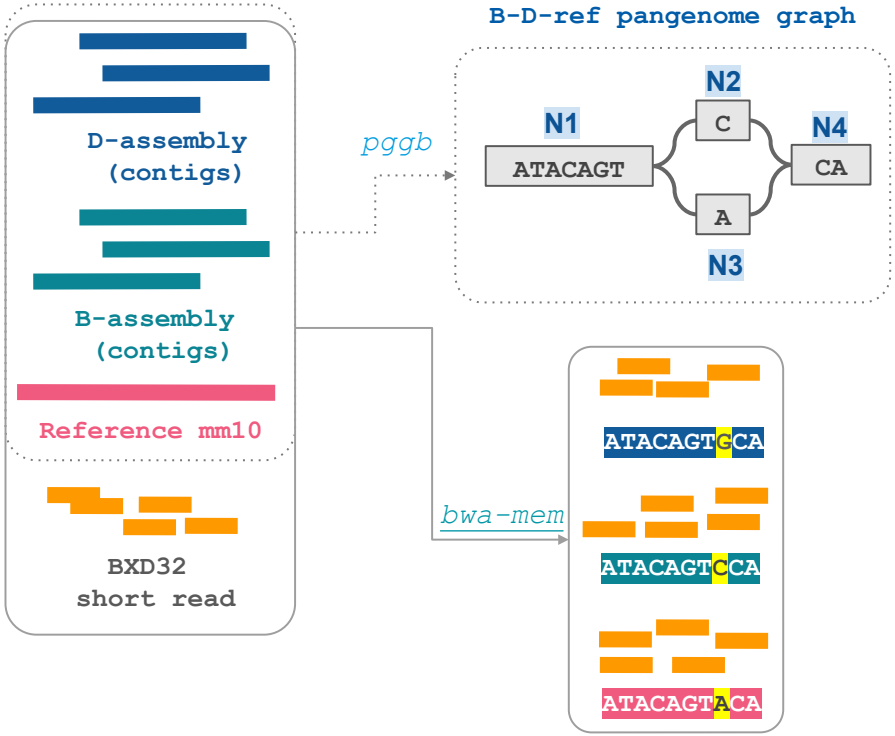
Mapping short-read against a pangenome graph



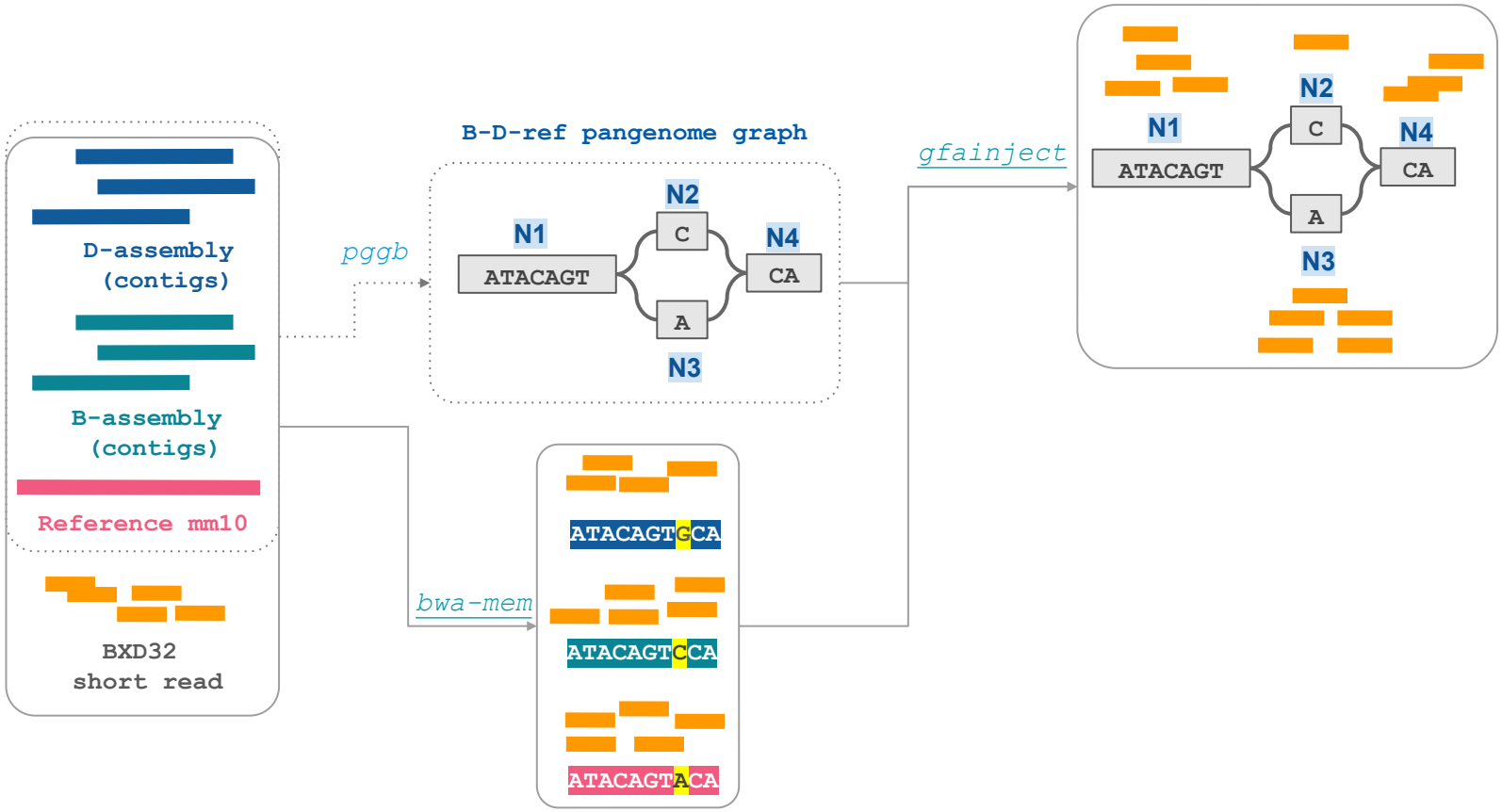
Mapping short-read against a pangenome graph



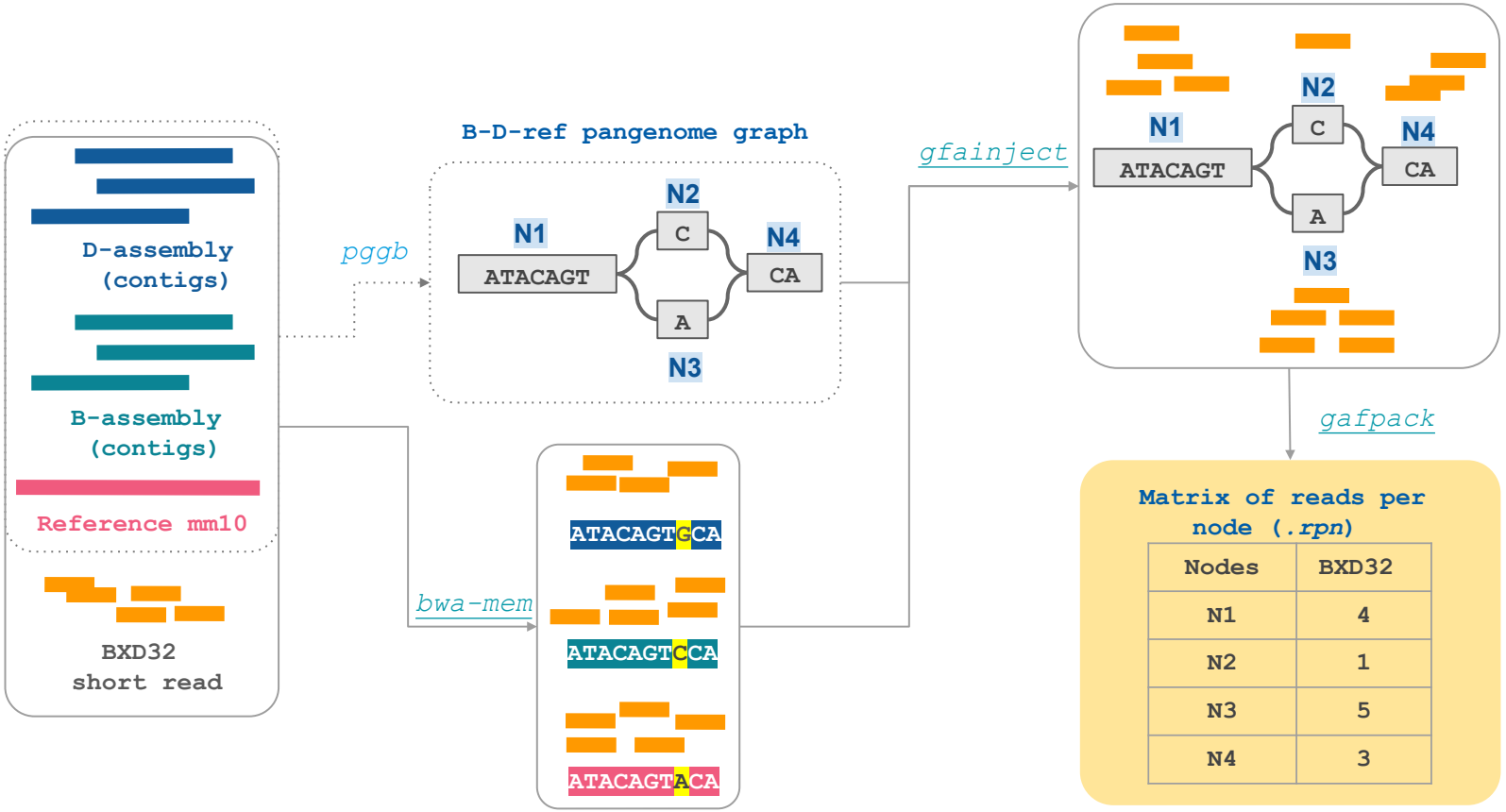
Mapping short-read against a pangenome graph



Mapping short-read against a pangenome graph



Mapping short-read against a pangenome graph



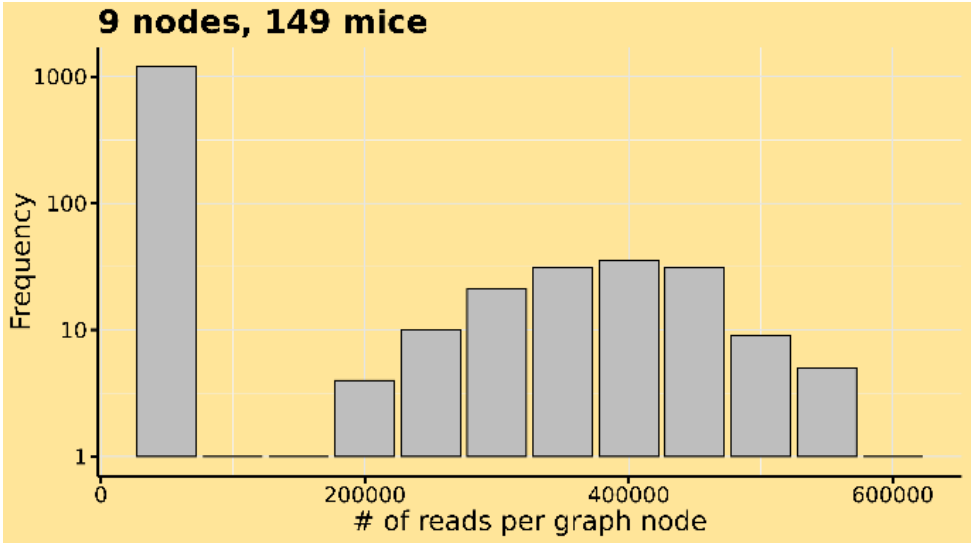
Matrix of reads per node (*.rpn*)

Nodes	BXD32
N1	4
N2	1
N3	5
N4	3

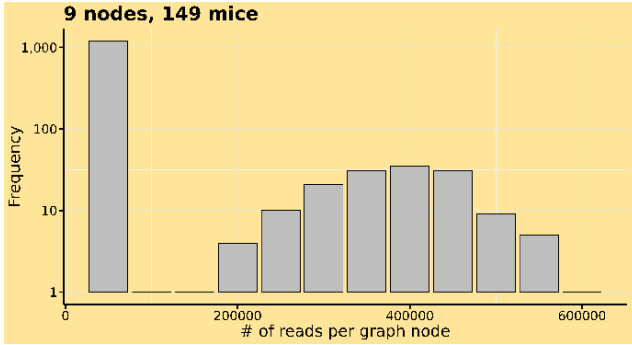
Genotyping from pangenome

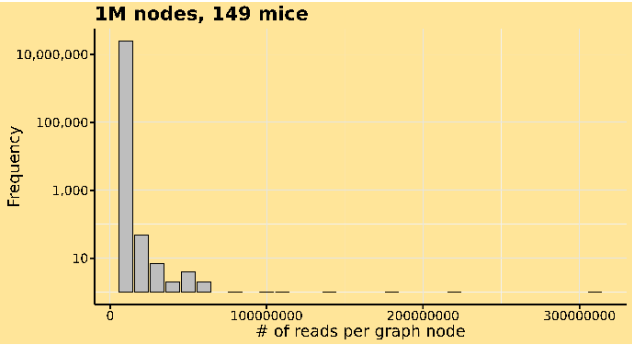
Matrix of reads per node

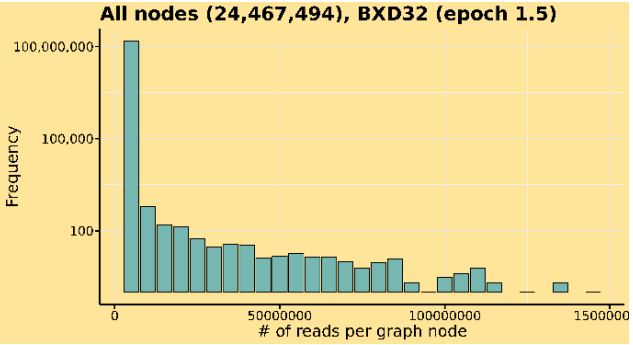
Nodes	BXD28	BXD22	BXD34	BXD40	BXD44	…
N1	348940	439853	384809	342335	232189	…
N2	2940	5277	6911	5884	4951	…
N3	125	218	377	392	211	…
N4	3	8	5	8	6	…
N5	0	3	7	5	2	…
N6	0	24	9	48	58	…
…	…	…	…	…	…	…



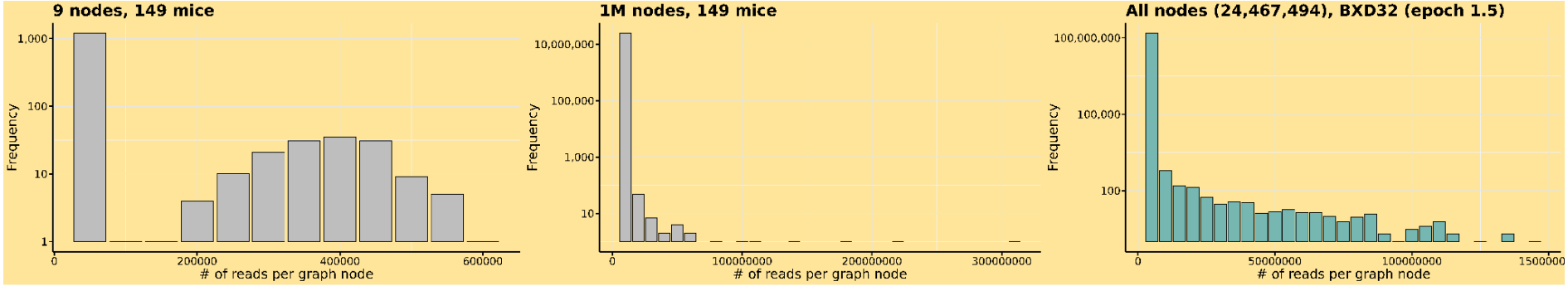
Matrix of reads per node

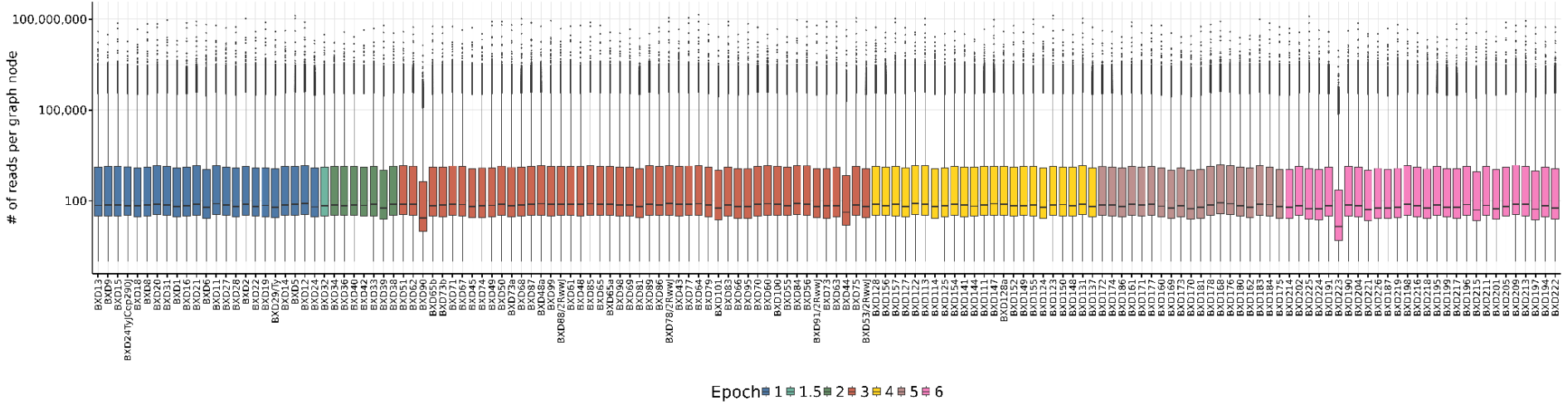






Matrix of reads per node





Application to eQTL mapping

Standard approach

Genotype matrix

rsID	A1	A2	Ind1	Ind2
rs3144	A	T	1	0

Phenotype matrix

ID	Value
Ind1	1.2
Ind2	2.3

Marker map

rsID	Pos	Chr
rs3144	1200	1

GEMMA: Genome-wide Efficient Mixed Model Association

Standard approach

Genotype matrix

rsID	A1	A2	Ind1	Ind2
rs3144	A	T	1	0

Phenotype matrix

ID	Value
Ind1	1.2
Ind2	2.3

Marker map

rsID	Pos	Chr
rs3144	1200	1

Pangenomic approach

Matrix of mapped reads

Node	A1	A2	Ind1	Ind2
N1	A	A	1	2

Phenotype matrix

ID	Value
Ind1	1.2
Ind2	2.3

Marker map

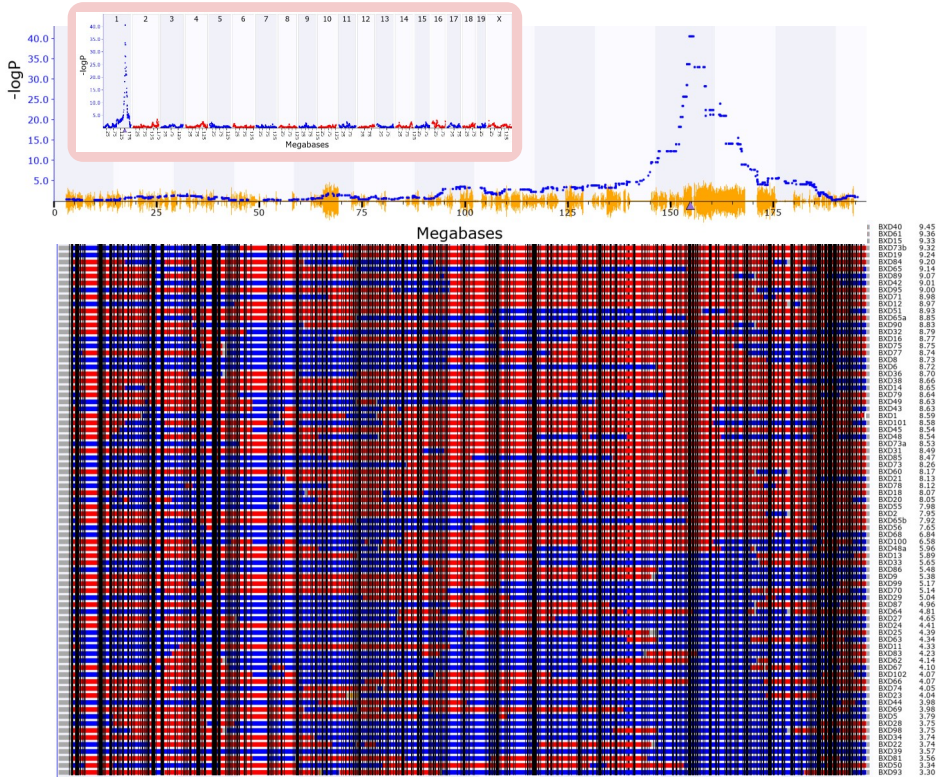
Node	Pos	Chr
N1	10	1

GEMMA: Genome-wide Efficient Mixed Model Association

eQTL mapping

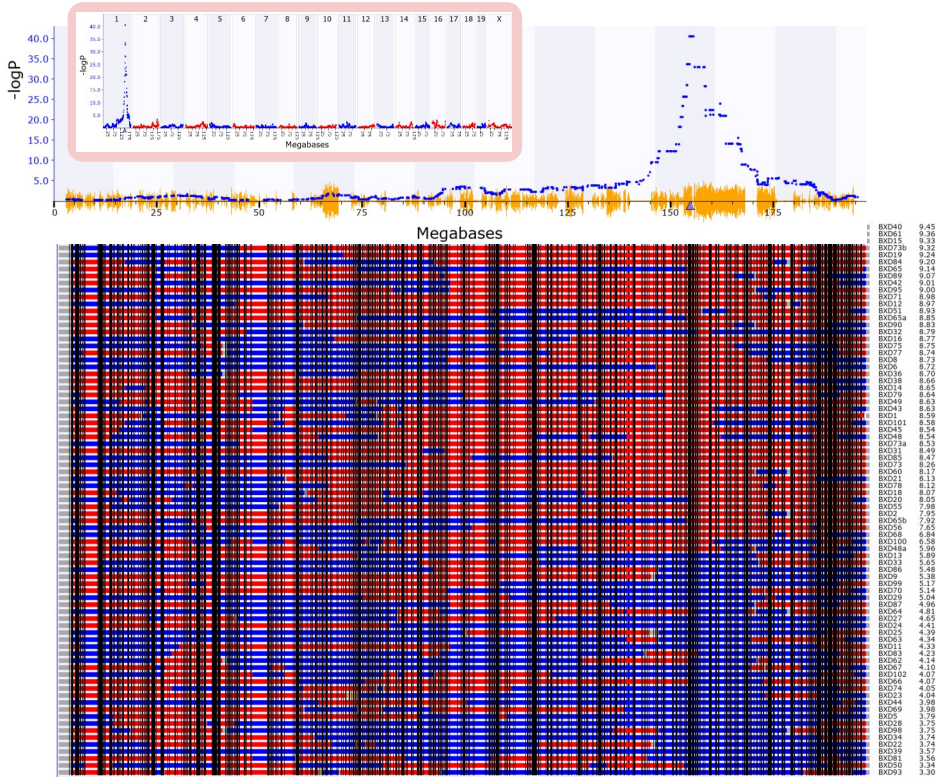
UTHSC

A very strong signal as a positive control

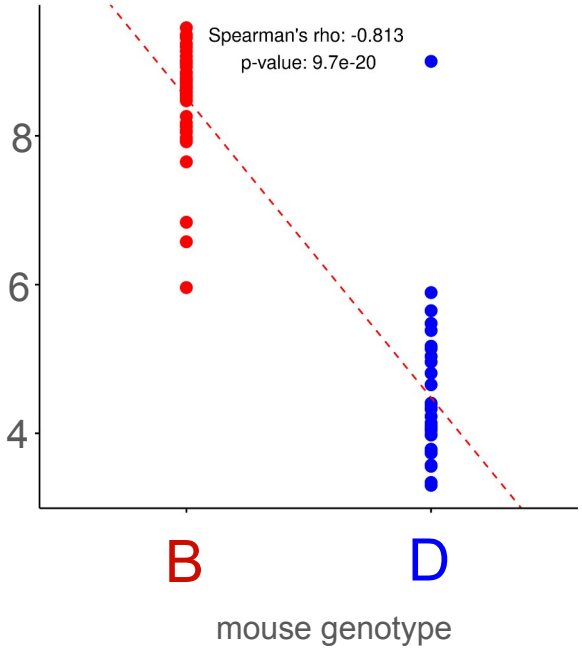


https://genenetwork.org/show_trait?trait_id=10358973&dataset=UTHSC_SPL_RMAEx_1210

A very strong signal as a positive control

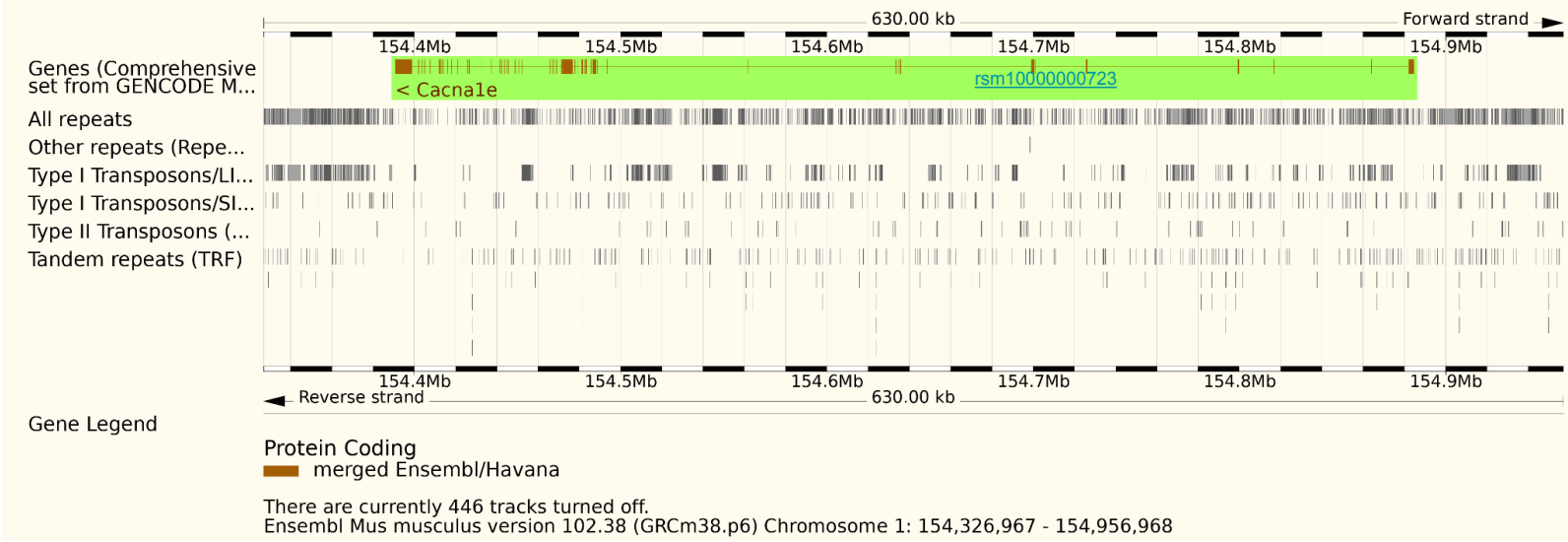


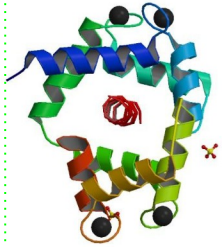
Cacna1e
gene expression
(spleen)



https://genenetwork.org/show_trait?trait_id=10358973&dataset=UTHSC_SPL_RMAEx_1210

Cacna1e gene (chr1)

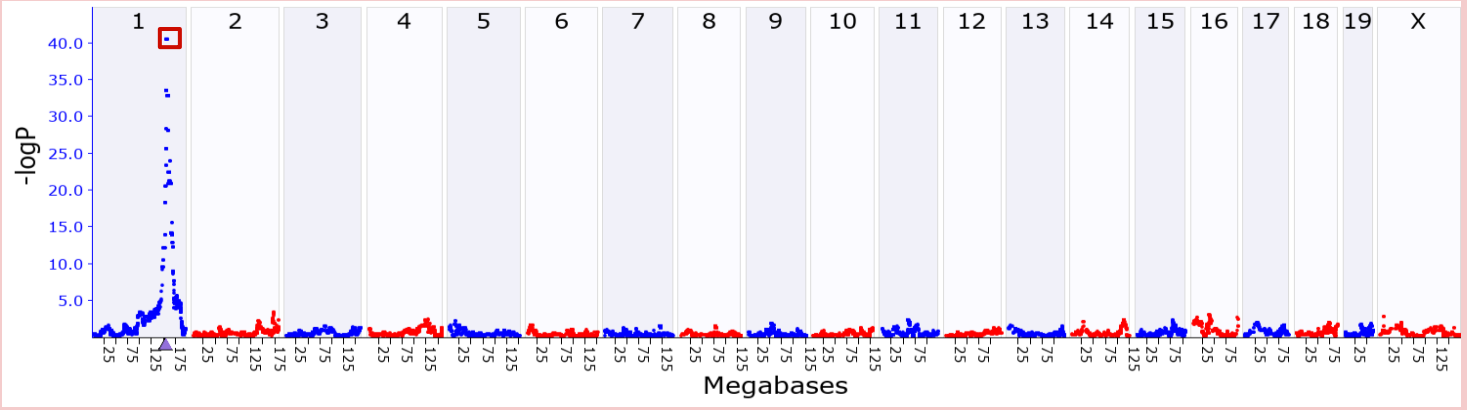




Cacna1e encodes an integral membrane protein that belongs to the calcium channel alpha-1 subunits family.

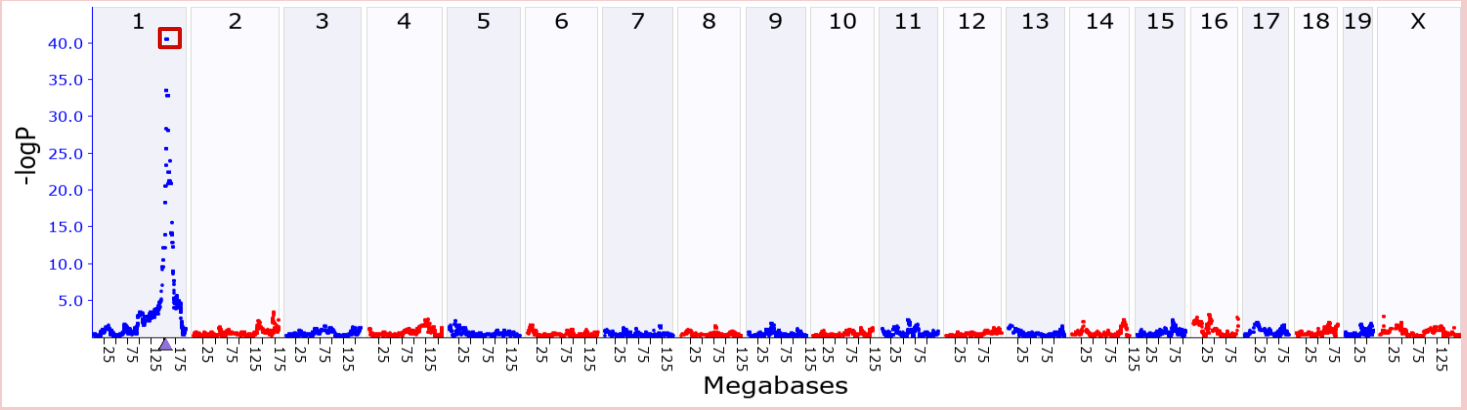
Signal replication for *Cacna1e* on chromosome 1

Standard approach

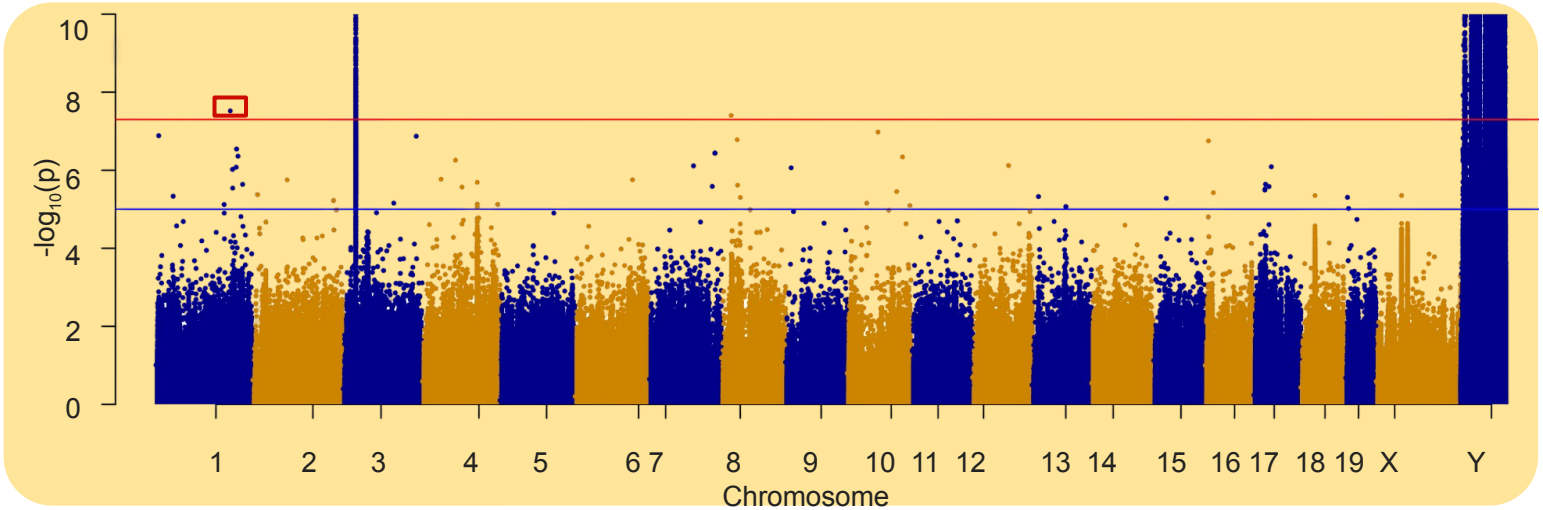


Signal replication for *Cacna1e* on chromosome 1

Standard approach



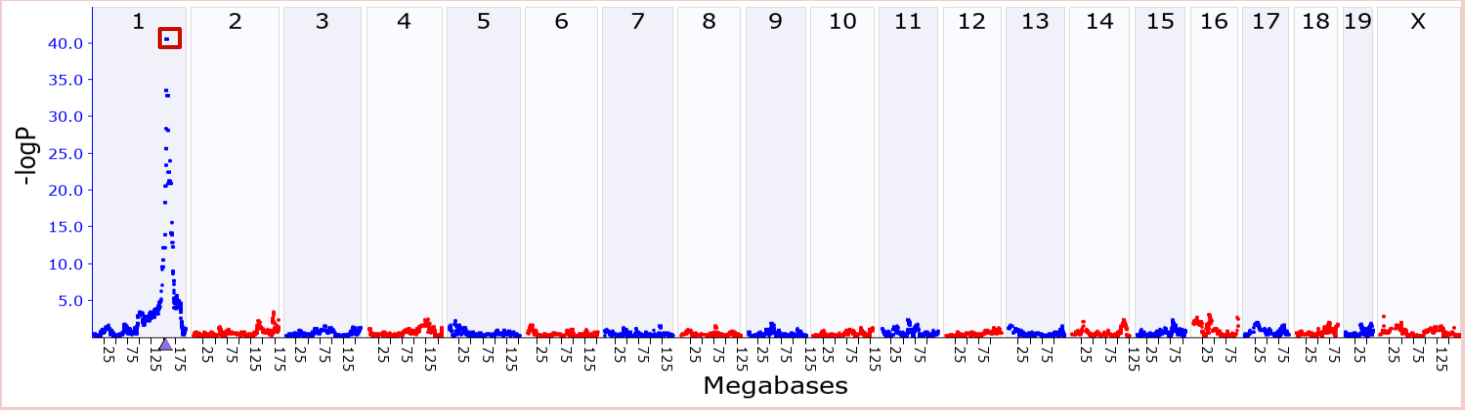
Pangenomic approach



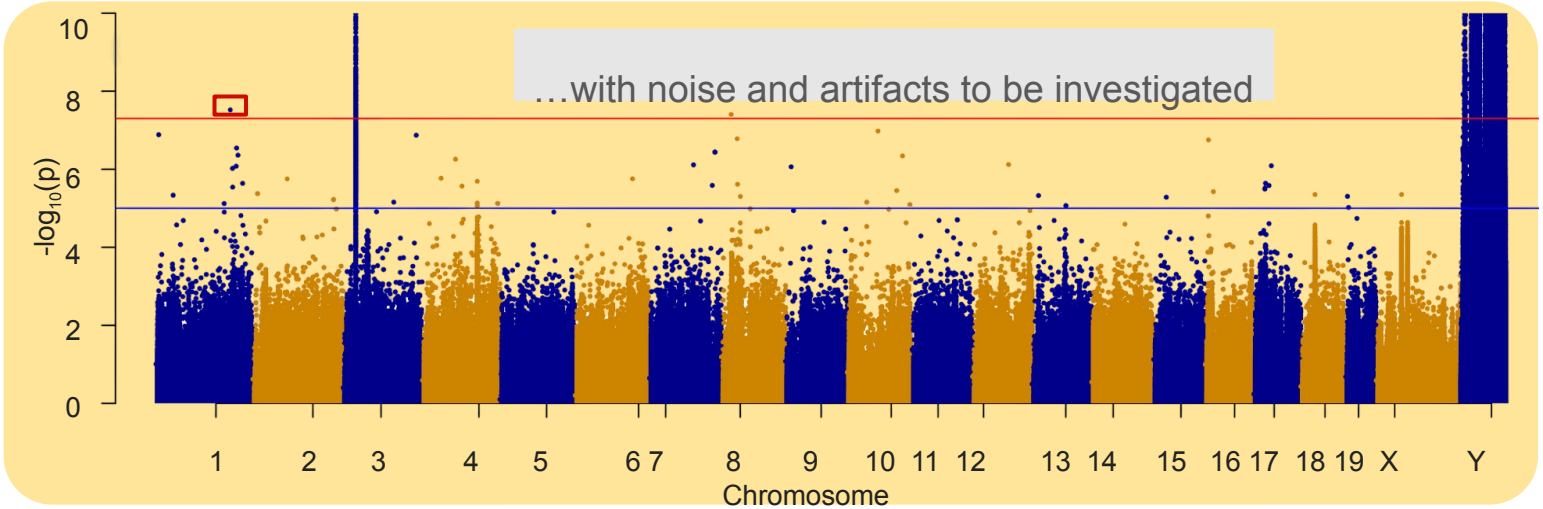
UTHSC

Signal replication for *Cacna1e* on chromosome 1

Standard approach



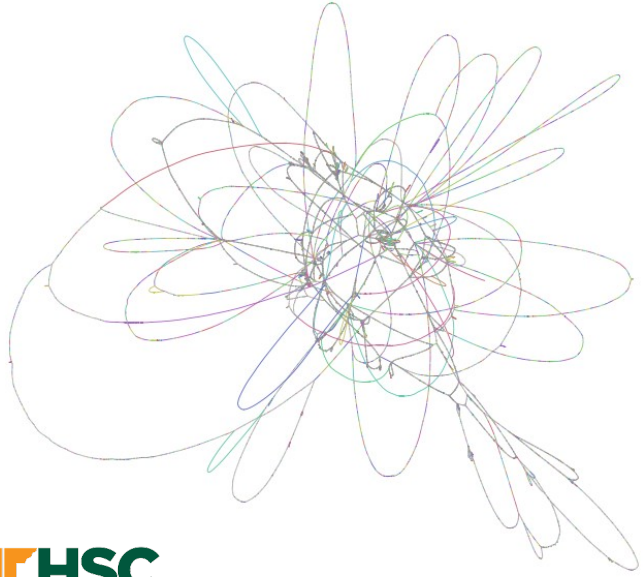
Pangenomic approach



UTHSC

Conclusions

- Pipeline for deriving genotype markers mapping short-read against pangenome graphs
- Application to eQTL mapping in BXD



Future directions

- Apply quality filters
- Manage repeats
- Extend the analyses to all traits in GN

UTHSC

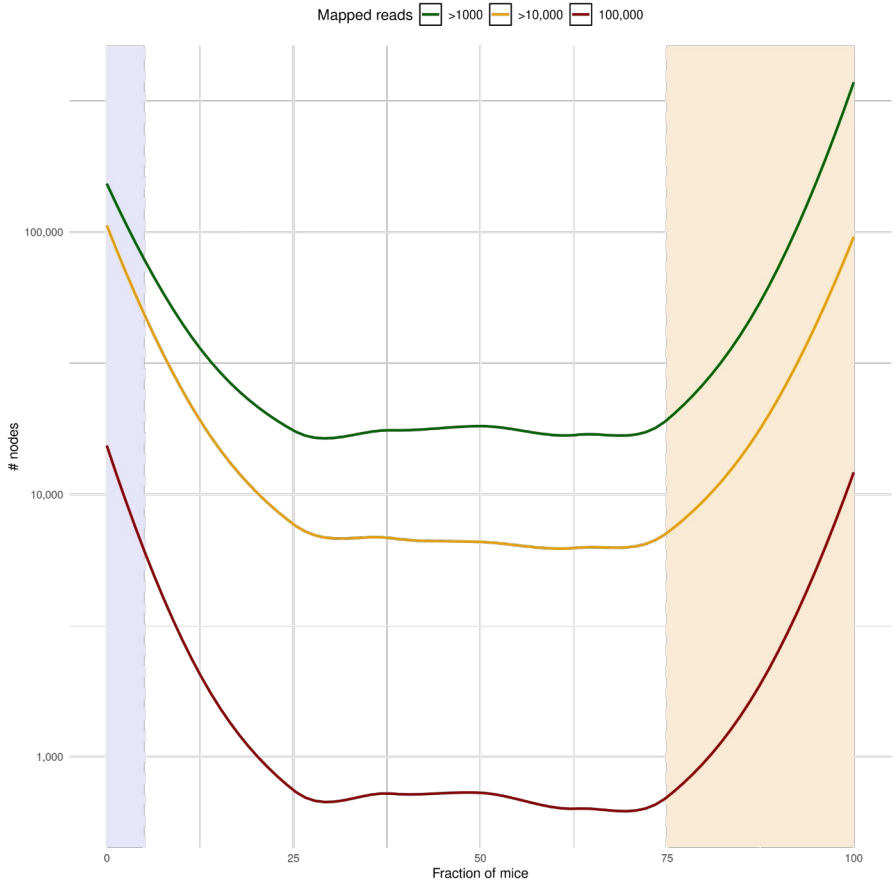
Acknowledgement

Vincenza Colonna
Erik Garrison
David Ashbrook
Rob Williams
Hao Chen
Pjotr Prins
Andrea Guarracino
Farnaz Salehi

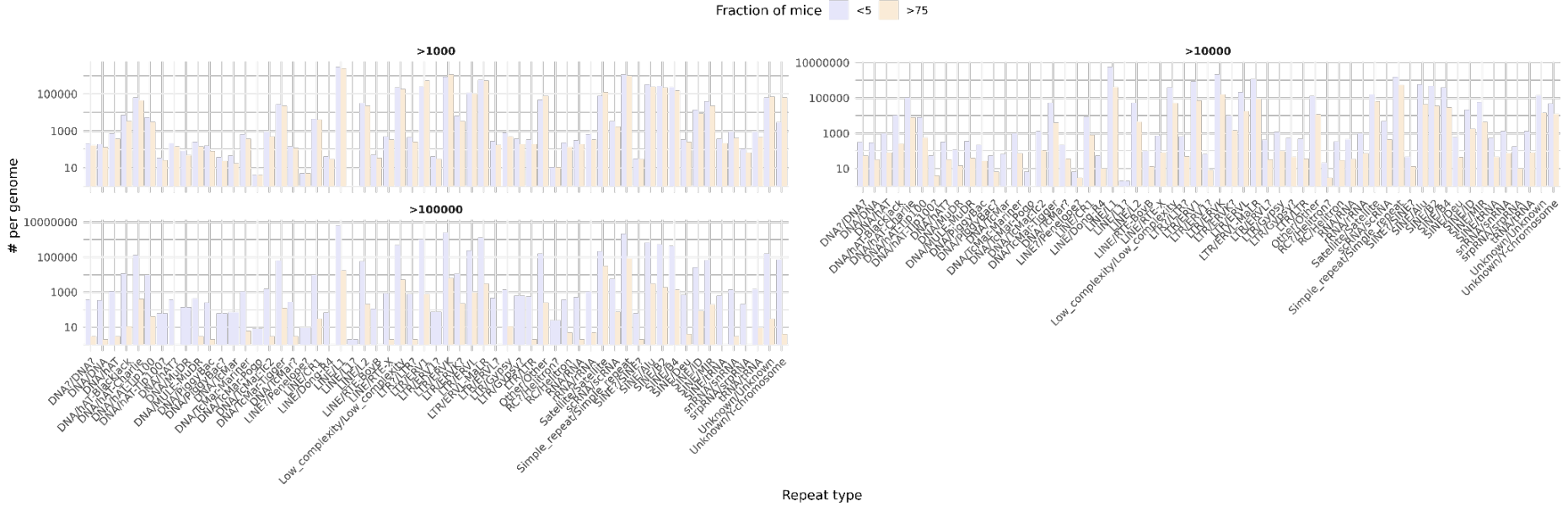
Colonna's lab
Pangenome group
and many others!!

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

THANKS!


Mapped reads
>1000
>10,000
100,000
nodes
100,000
10,000
1,000
0
25
50
75
100
Fraction of mice

Repeat overlap
1000.non-overlap
100000.non-overlap
10000.notinmm10
1000.overlap
100000.overl
10000.non-overlap
1000.notinmm10
100000.notinmm10
10000.overlap
0-5%
75-100%
Count
40000000
30000000
20000000
10000000
0
>1000
>10,000
>100,000
>1000
>10,000
>100,000
Reads mapped


Fraction of mice
<5
>75
>1000
>10000
>100000
per genome
Repeat type