

# Selection Pressures Affecting Age at Onset of Type 2 Diabetes

Prof. Colin NA Palmer  
Dr. Radha Venkatesan

CHARVI NANGIA



*"This research was funded by the National Institute for Health Research (NIHR) (INSPIRED 16/136/102) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care"*

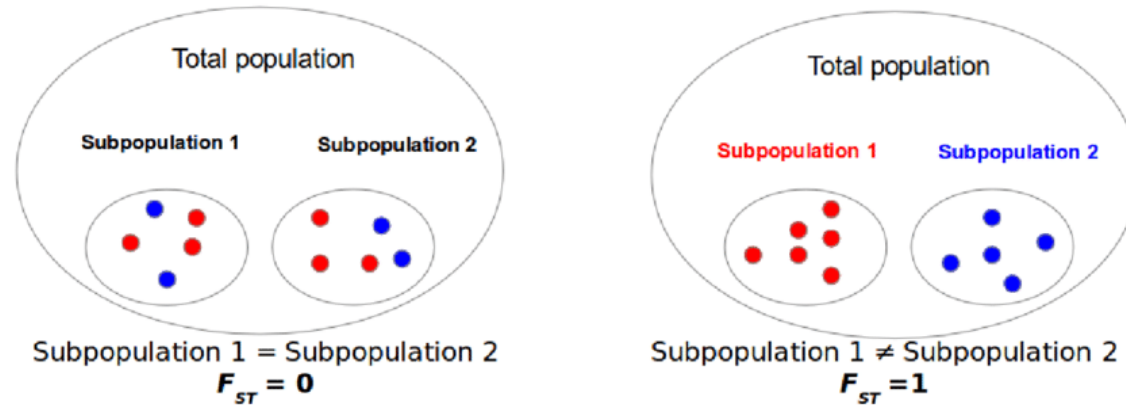
# Introduction

- There is compelling evidence that Type 2 Diabetes is strongly influenced by genetic factors.
- Most studies have been done on European populations.
- Since genetic architecture is different between populations, it is important to study the differences.
- Diabetes in India occurs at least two decades earlier in individuals as compared to a UK population.
- One way of identifying the differences in genetic architecture affecting complex traits is by studying the selection pressures affecting genes.

# Selection Pressures

- Various factors have shaped evolution across all species.
- Genetic drift, mutation, migration and selection.
- Selection – positive or negative, depending on whether it is beneficial or detrimental to the population.
- $F_{st}$  (F statistics), relies upon the differences in allele frequencies between the populations being studied.

# Fixation Index (Fst)



- A certain locus under selection pressure changes its frequency over generations.
- High values of Fst indicate greater genetic differentiation and low values indicate greater genetic similarity.
- Values range from 0 to 1.

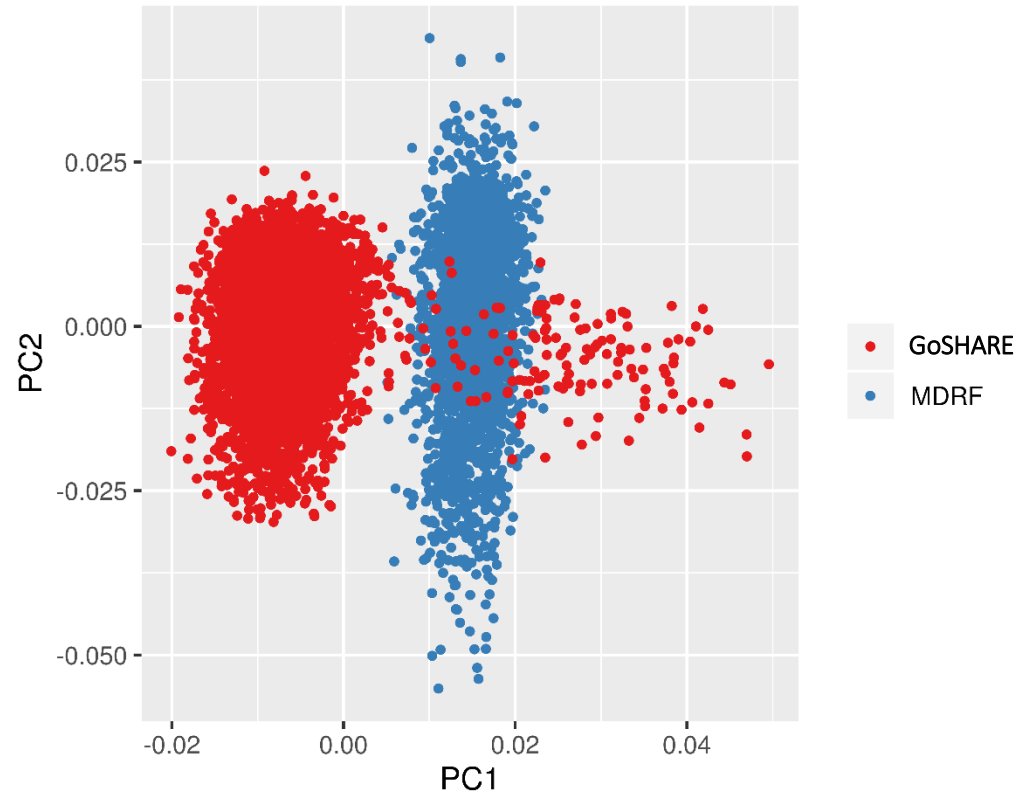
# STUDY POPULATION

- South Indian diabetes cohort from the Madras Diabetes Research Foundation (MDRF), n=6,056.
- Scottish diabetes cohort from the Scottish Health Research Registry (GoSHARE) study, n=6,681.

# METHODOLOGY

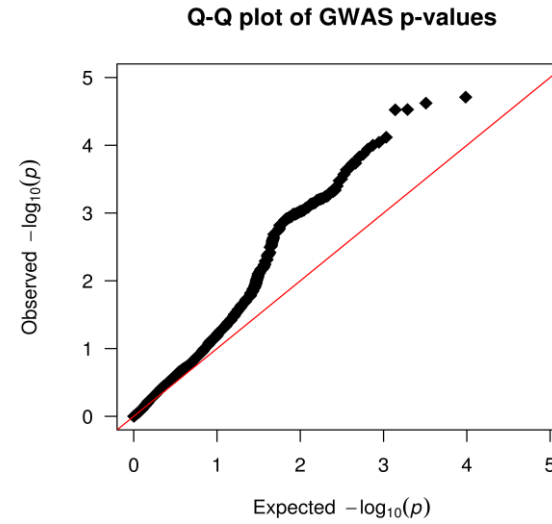
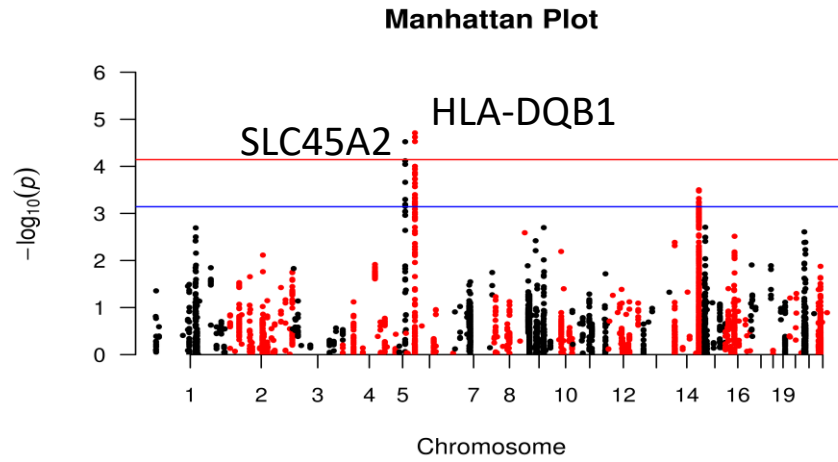
- Whole Genome FST was done on 12,737 individuals from MDRF and GoSHARE.
- ~5000 high FST SNPs which had  $FST > 0.30$ .
- Association with Age at Onset Phenotype.
- Tests for Selective Sweep – Tajima's D and Nucleotide Diversity.
- Locus Zoom
- Conditional Analysis
- Fine Mapping.

# PCA

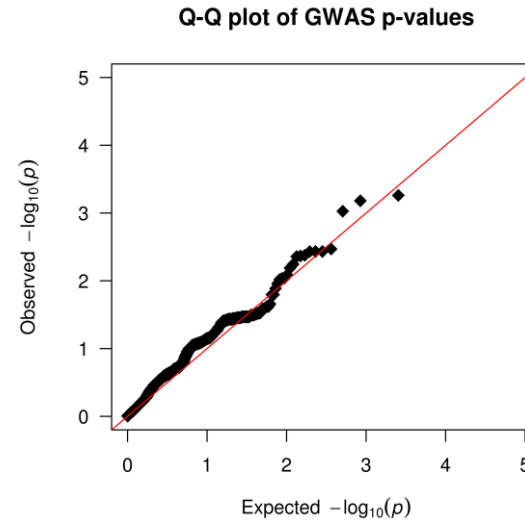
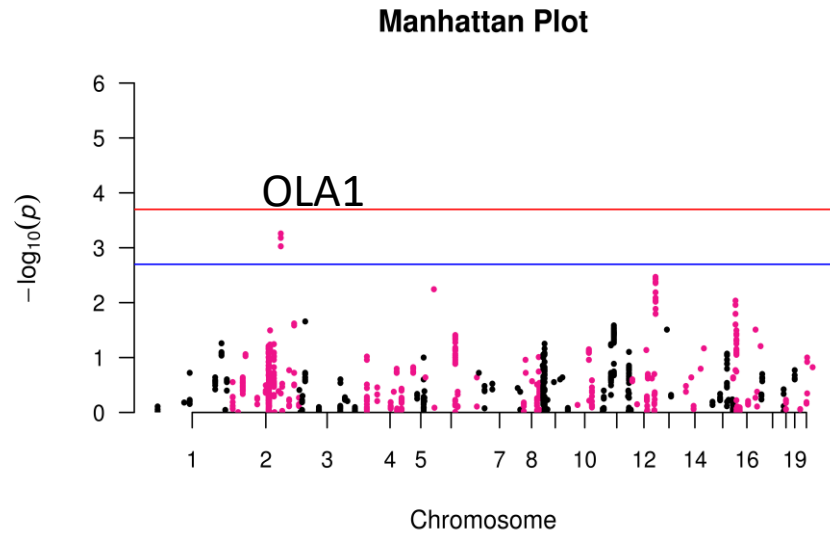


173 individuals of Indian Ancestry were found within the GoSHARE cohort who were removed from further analysis.

# AGEONSET



**GODARTS-  
GOSHARE**



**MDRF**

\*Adjusted for Sex



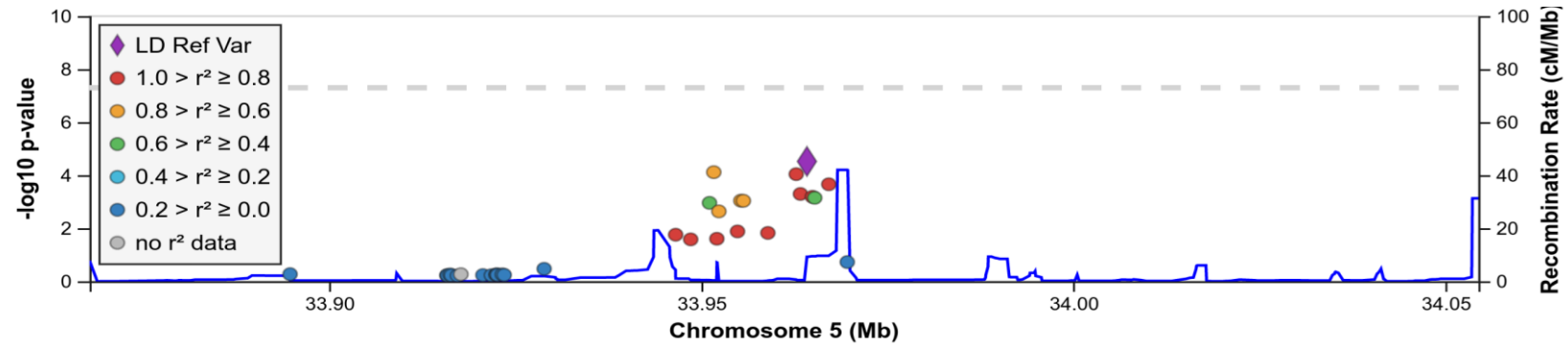
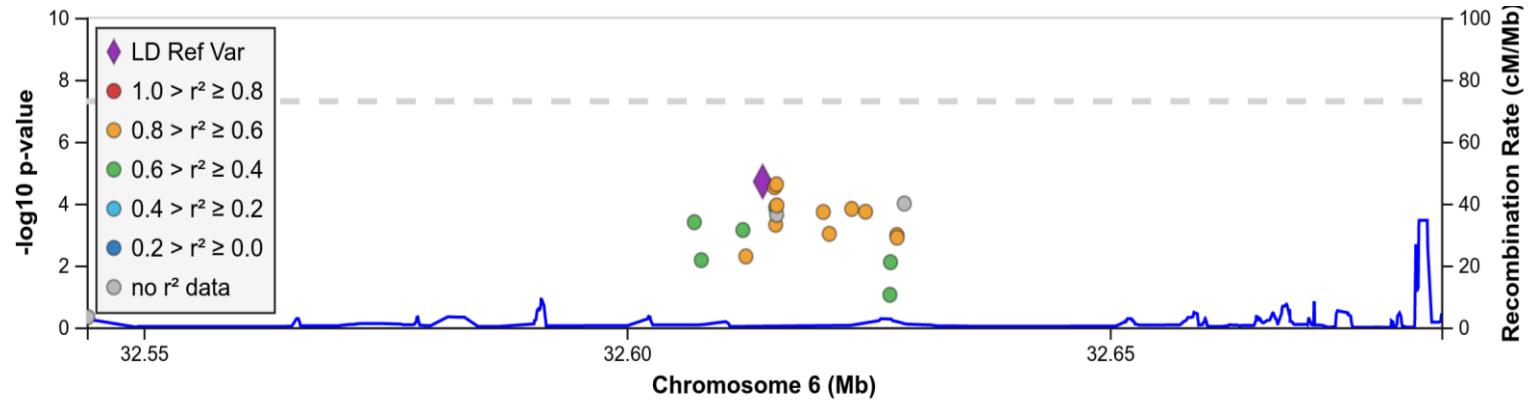
## GODARTS- GOSHARE

Chr_Pos	rsid	Ref	Alt	MAF_goshareFF	Freq_mdrfFF	beta	se	p.value	hweP_goshareFF	n_studies	effects	Location	Gene
6:32614080	rs9273242	T	C	0.385	0.82	0.634	0.148	1.95E-05	0.7962	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32615510	rs17843577	C	T	0.394	0.844	0.627	0.148	2.39E-05	0.7009	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32615312	rs17612548	A	G	0.395	0.14	0.621	0.149	2.97E-05	0.5908	4	----	intergenic	HLA-DQA1;HLA-DQB1
5:33951693	rs16891982	C	G	0.035	0.634	2.436	0.615	7.62E-05	1.44E-46	3	+++?	exonic	SLC45A2
6:32628736	rs9273532	C	T	0.39	0.14	0.576	0.148	9.99E-05	0.9385	4	----	intronic	HLA-DQB1
6:32615551	rs17843580	A	G	0.387	0.85	0.573	0.148	1.14E-04	0.757	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32615421	rs7745002	A	G	0.463	0.903	0.571	0.15	1.36E-04	0.2095	4	----	intronic	HLA-DQA1
6:32623300	rs9273330	C	T	0.404	0.853	0.572	0.151	1.49E-04	0.7606	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32624731	rs74824383	A	G	0.44	0.882	0.576	0.154	1.84E-04	0.5513	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32620359	rs17612802	C	T	0.39	0.842	0.555	0.148	1.87E-04	0.7188	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32615527	rs17843579	T	G	0.39	0.854	0.55	0.149	2.30E-04	0.9385	4	----	intergenic	HLA-DQA1;HLA-DQB1
6:32501122	rs76817401	G	T	0.074	0.516	0.724	0.199	2.70E-04	0.3712	4	----	intergenic	HLA-DRB5;HLA-DRB6

## MDRF

Chr_Pos	rsid	Ref	Alt	MAF_goshareFF	Freq_mdrfFF	BETA	SE	P_value	hweP_mdrfFF	Location	Gene
2:175090397	rs67654651	G	C	0.054	0.418	-0.723	0.209	5.50E-04	1.85E-01	intronic	OLA1
2:175083689	rs67910840	C	G	0.055	0.415	-0.717	0.211	6.60E-04	7.56E-01	intronic	OLA1
2:175093229	rs66463720	G	A	0.054	0.416	-0.692	0.209	9.40E-04	1.32E-01	intronic	OLA1

# LOCUS ZOOM



# Conditional Analysis

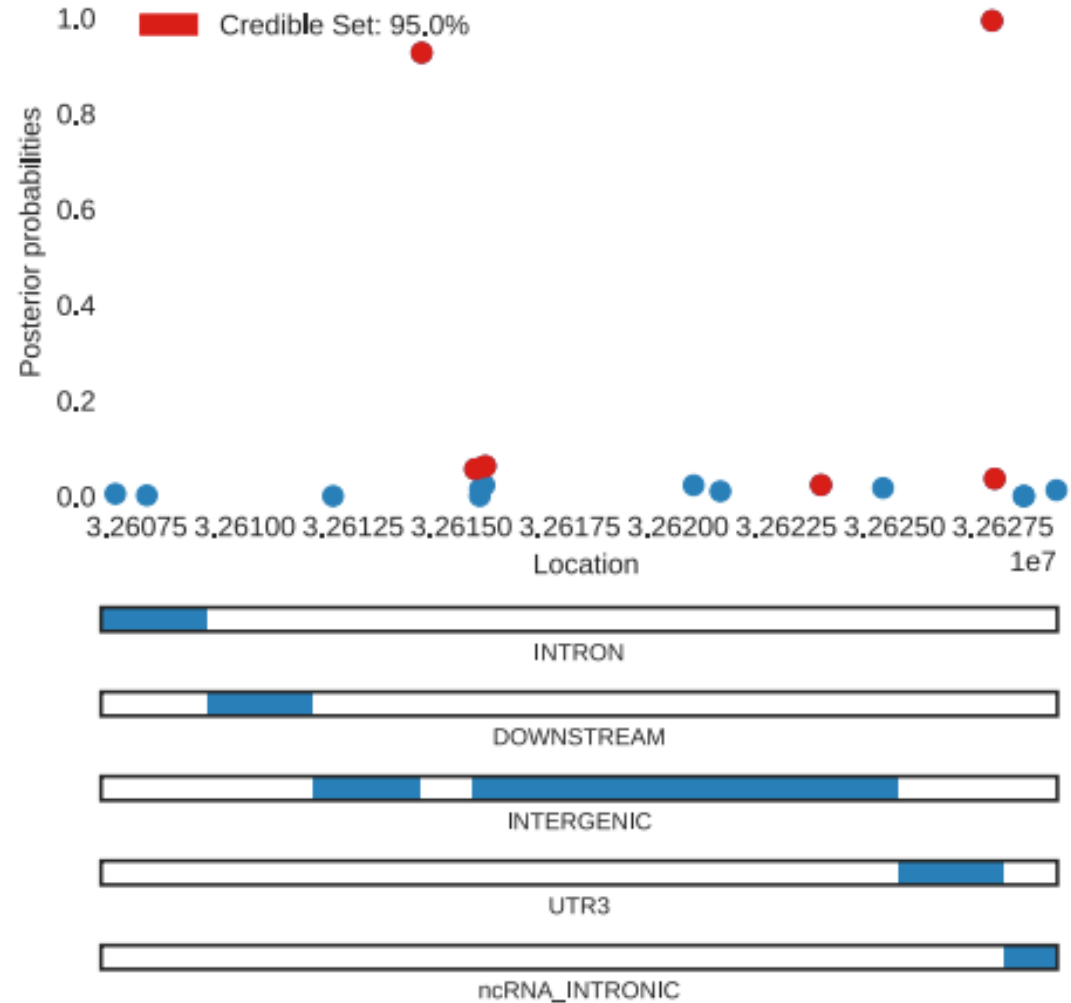
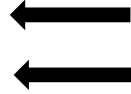
- Conditioned on the top SNP - 6:32614080 (HLA-DQA1) which gave 5:33964210 (SLC45A2) as the next hit.
- Conditioned on both 6:32614080 and 5:33964210 didn't give any other significant hits.

# Fine mapping

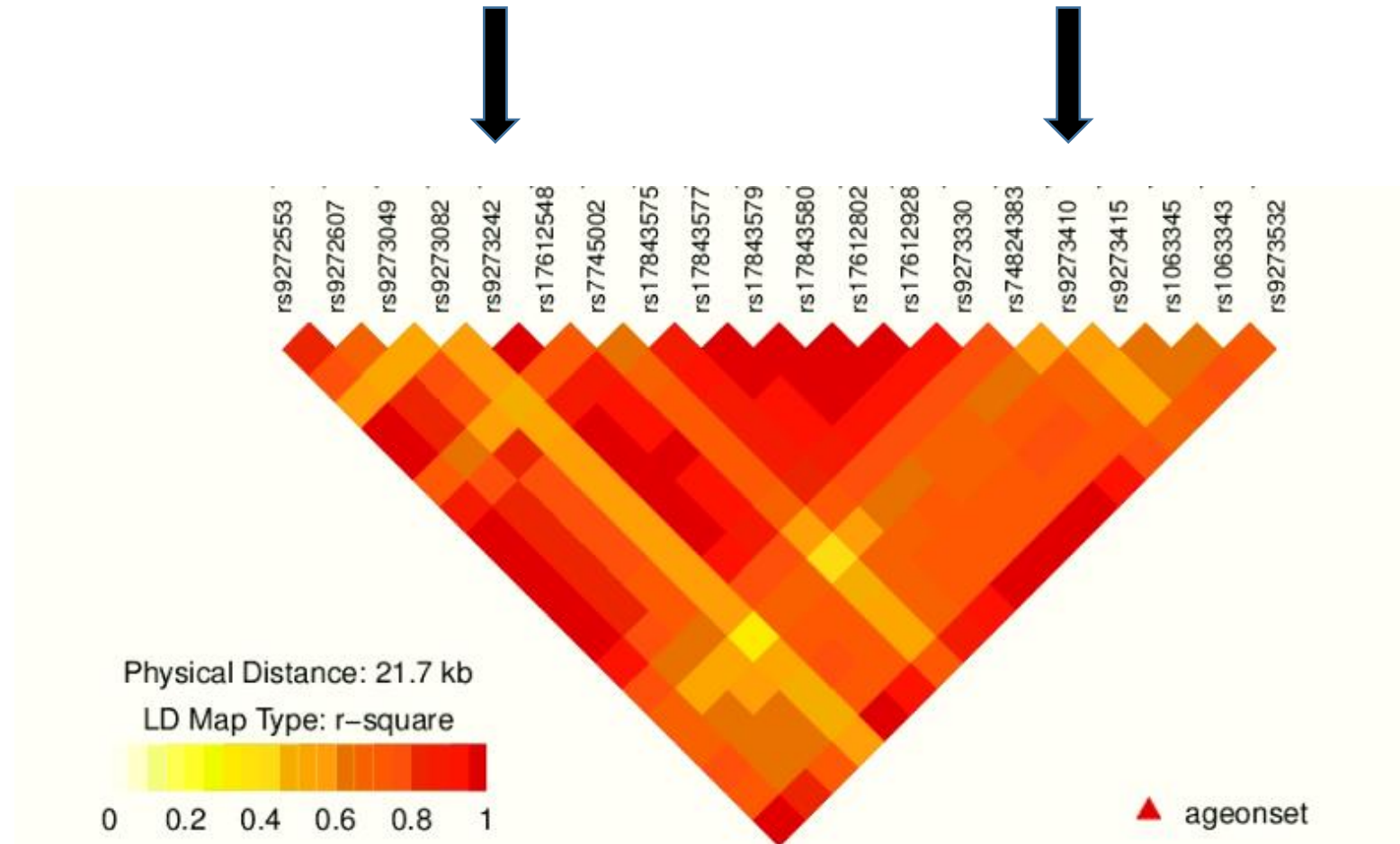
- Fine mapping was done using the software PAINTOR\_v3.
- The analysis includes -
  - Genetic association data from summary statistics
  - LD structure
  - Functional Annotation
- SNPs which were within 100kb of the top hit ( 50 kb upstream and downstream) were considered for analysis.

# Locus 1 - HLA

Zscore	Posterior_Prob	chr_pos	p.value
-1.704581	0.993884	6:32627250	8.83E-02
-4.272847	0.926952	6:32614080	1.95E-05
-3.861015	0.062732	6:32615551	1.14E-04
-4.227756	0.06068	6:32615510	2.39E-05
-4.178486	0.0562	6:32615312	2.97E-05
-2.66083	0.036172	6:32627310	7.81E-03
-3.795609	0.023072	6:32623300	1.49E-04
-3.737664	0.022564	6:32620359	1.87E-04
-3.685656	0.022356	6:32615527	2.30E-04
-3.742726	0.016792	6:32624731	1.84E-04
-3.49102	0.014812	6:32615427	4.84E-04
-3.893139	0.012308	6:32628736	9.99E-05
-3.306831	0.009988	6:32620976	9.49E-04
-3.54191	0.004564	6:32607008	4.00E-04
-2.711439	0.001388	6:32607735	6.72E-03
-3.817205	0	6:32615421	1.36E-04
-3.385392	0	6:32612034	7.15E-04
-3.28178	0	6:32627976	1.04E-03
-3.226382	0	6:32627992	1.26E-03
-2.802334	0	6:32612339	5.09E-03



# HLA – LD Structure of Causal SNPs



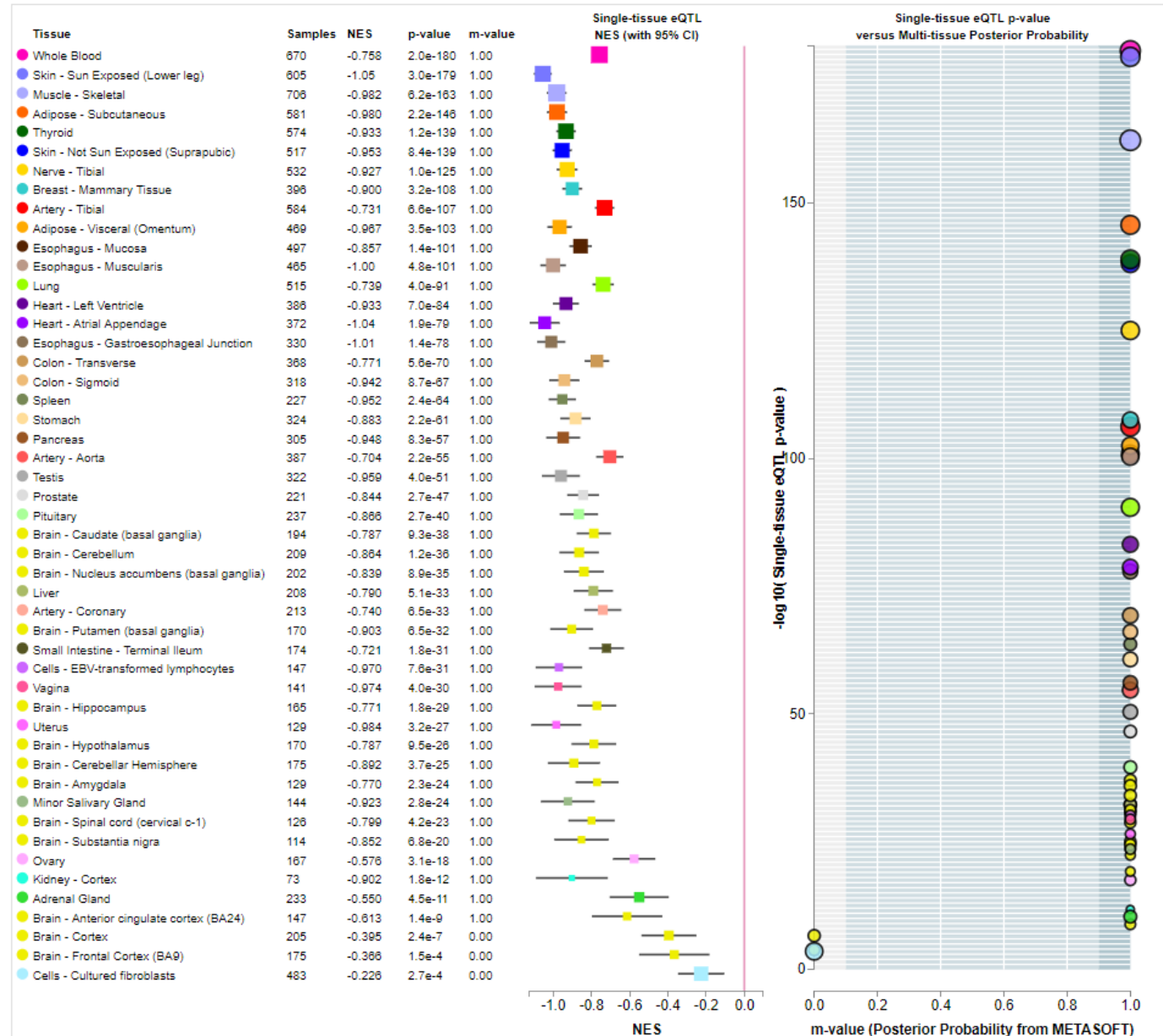
# Individual tests of Association

SNP	Ref	Alt	MAF	beta	se	p-value	n_studies
rs9273242	T	C	0.385	0.058	0.014	8.02E-05	4

SNP	Ref	Alt	MAF	beta	se	p-value	n_studies
rs9273410	C	A	0.425	0.018	0.015	0.234	4

SNP	Ref	Alt	MAF	beta	se	p-value	n_studies
rs9273242	T	C	0.385	0.058	0.014	8.02E-05	4
rs9273410	C	A	0.425	0.018	0.015	0.235	4

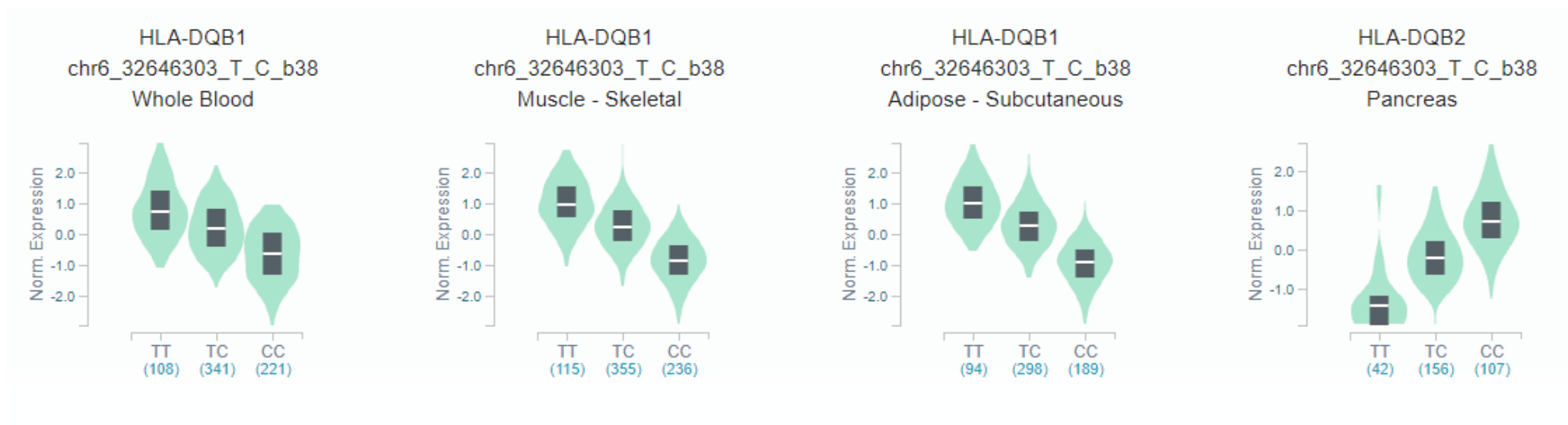
# Multi-Tissue eQTL rs9273242



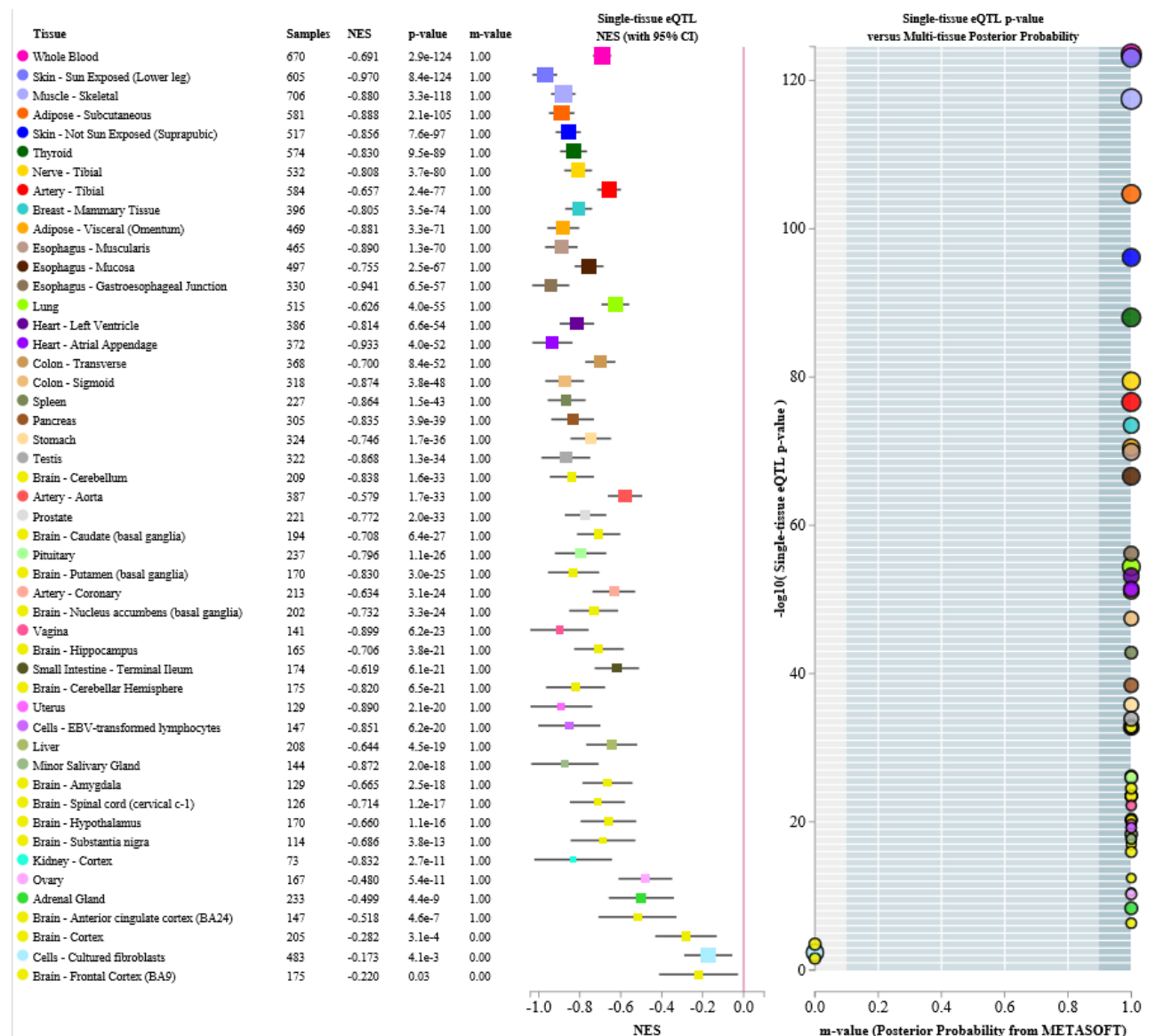


# Single-Tissue eQTL rs9273242

Minor Allele - T



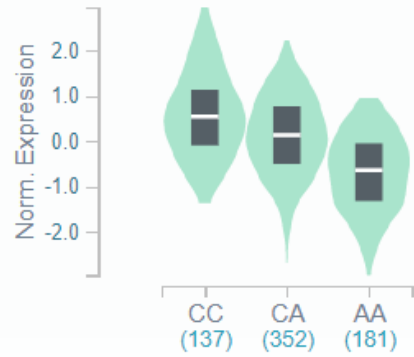
# Multi-tissue eQTL – rs9273410



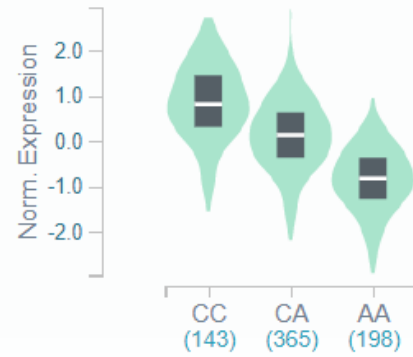
# Single tissue eQTL – rs9273410

Minor Allele - C

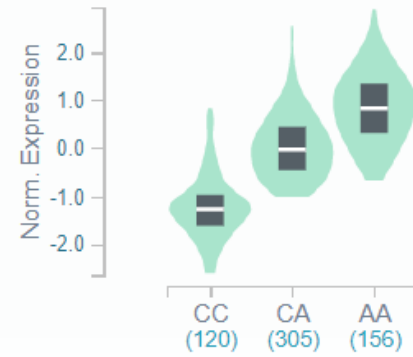
HLA-DQB1  
chr6\_32659473\_C\_A\_b38  
Whole Blood



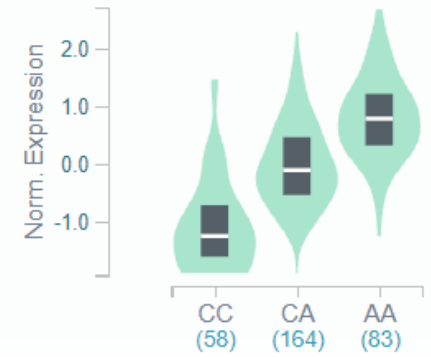
HLA-DQB1  
chr6\_32659473\_C\_A\_b38  
Muscle - Skeletal



HLA-DQB2  
chr6\_32659473\_C\_A\_b38  
Adipose - Subcutaneous

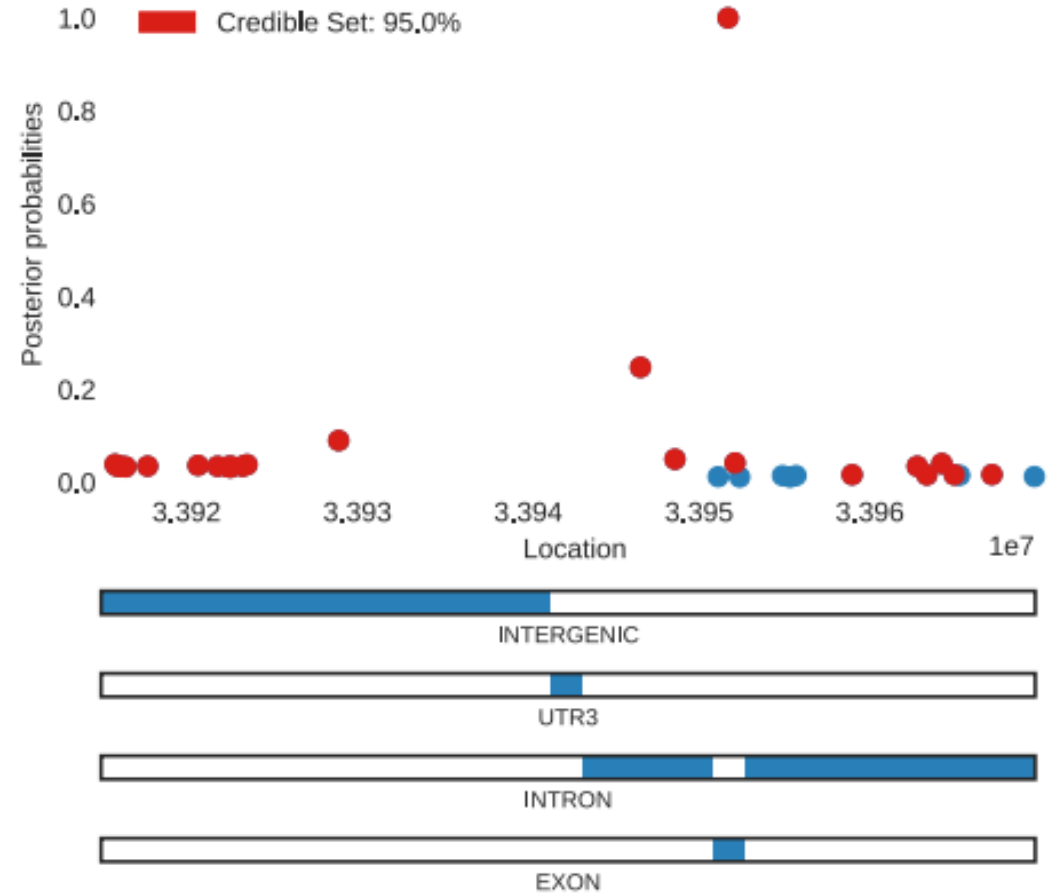


HLA-DQB2  
chr6\_32659473\_C\_A\_b38  
Pancreas



# Locus 2 –SLC45A2

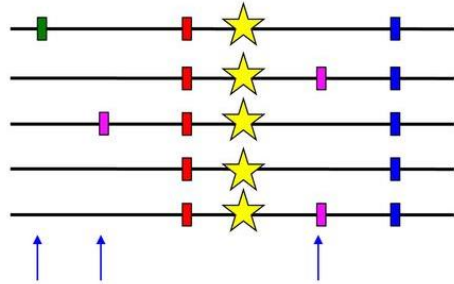
Zscore	Posterior_Prob	chr_pos	p.value
3.9584	0.99998	5:33951693	7.62E-05
2.378766	0.247496	5:33946571	1.74E-02
-0.962388	0.089668	5:33928902	3.36E-01
2.225757	0.049476	5:33948589	2.61E-02
2.2462	0.041336	5:33952106	2.47E-02
4.175878	0.040108	5:33964210	3.00E-05
0.539553	0.038604	5:33915813	5.90E-01
0.567438	0.037804	5:33923539	5.70E-01
0.555976	0.036132	5:33920667	5.78E-01
0.615744	0.034556	5:33917717	5.38E-01
0.568234	0.034408	5:33922611	5.70E-01
0.538566	0.034312	5:33915935	5.90E-01
3.917014	0.034288	5:33962770	9.05E-05
0.545314	0.034264	5:33921810	5.86E-01
0.604226	0.034252	5:33923272	5.46E-01
0.598419	0.033544	5:33916269	5.50E-01
0.582141	0.033536	5:33922410	5.60E-01
0.541108	0.033096	5:33916434	5.88E-01
0.612774	0.031968	5:33922558	5.40E-01
0.519598	0.031108	5:33917250	6.03E-01
3.69885	0.016364	5:33967145	2.18E-04
2.435579	0.016104	5:33958959	1.49E-02
3.413294	0.015892	5:33964938	6.46E-04
3.479837	0.015596	5:33963333	5.05E-04
3.388511	0.01486	5:33965257	7.07E-04
3.317006	0.01426	5:33955673	9.15E-04
2.48264	0.014212	5:33954880	1.31E-02
1.320506	0.012188	5:33969628	1.87E-01
3.266174	0.011824	5:33951116	1.10E-03
3.050567	0.010936	5:33952378	2.29E-03
3.319983	0.010892	5:33955326	9.05E-04



# SLC45A2- rs16891982

- Shows the highest probability of causality.
- Located in the exonic region of the gene.
- Missense variant - Leu374Phe
- Has been reported in more than 70 publications.
- Reported for skin, hair and eye pigmentation.
- Has also been reported for skin melanomas.

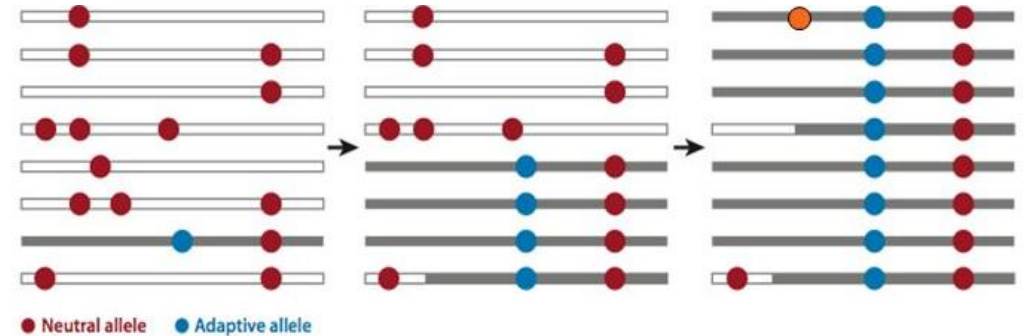
# Tajima's D and Nucleotide Diversity



$S = 3$

$$\pi = \frac{(2 + 2 + 1 + 2) + (2 + 1 + 0) + (1 + 2) + (1)}{10 \text{ pairwise comparisons}} = 1.4$$

avg. # difs between each pair of sequences



$\pi$

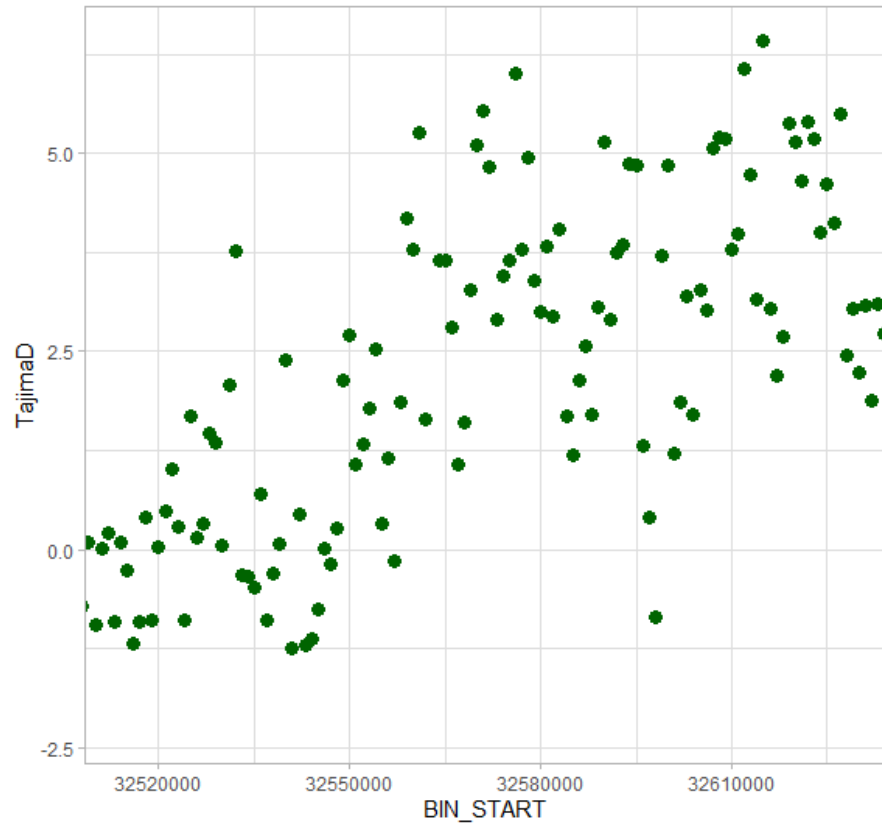
Observed – Expected =  $\pi - \Phi$

- Tajima's  $D > 1$  = No sweep
- Tajima's  $D = 1$  = No selection
- Tajima's  $D < 1$  = sweep

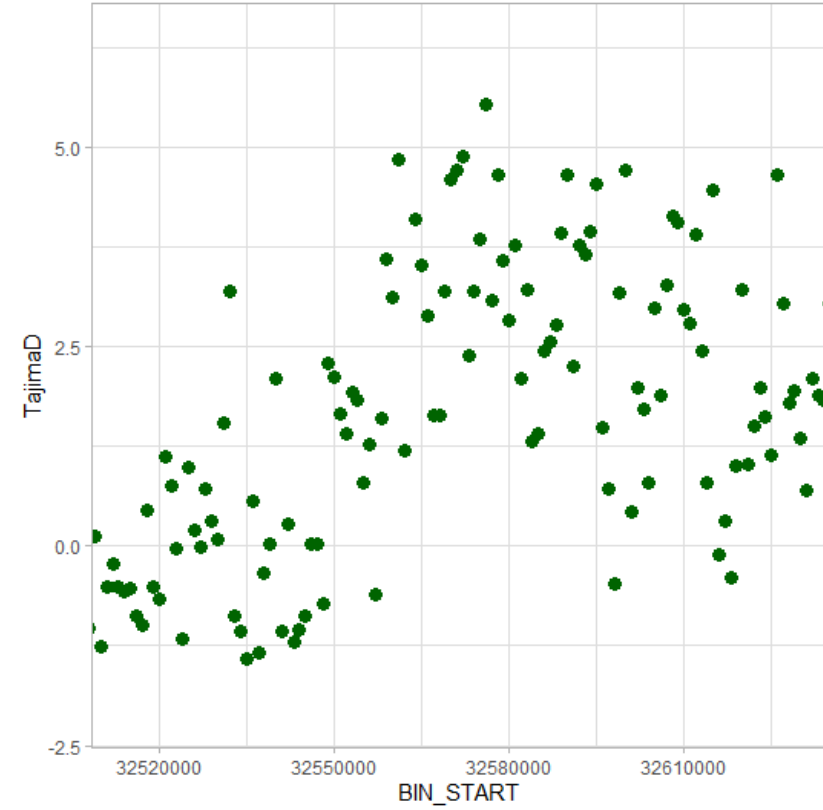
- Nucleotide Diversity  $> 0$  = No sweep
- Nucleotide Diversity = 0 = sweep

# TajimasD - HLA

GoShare

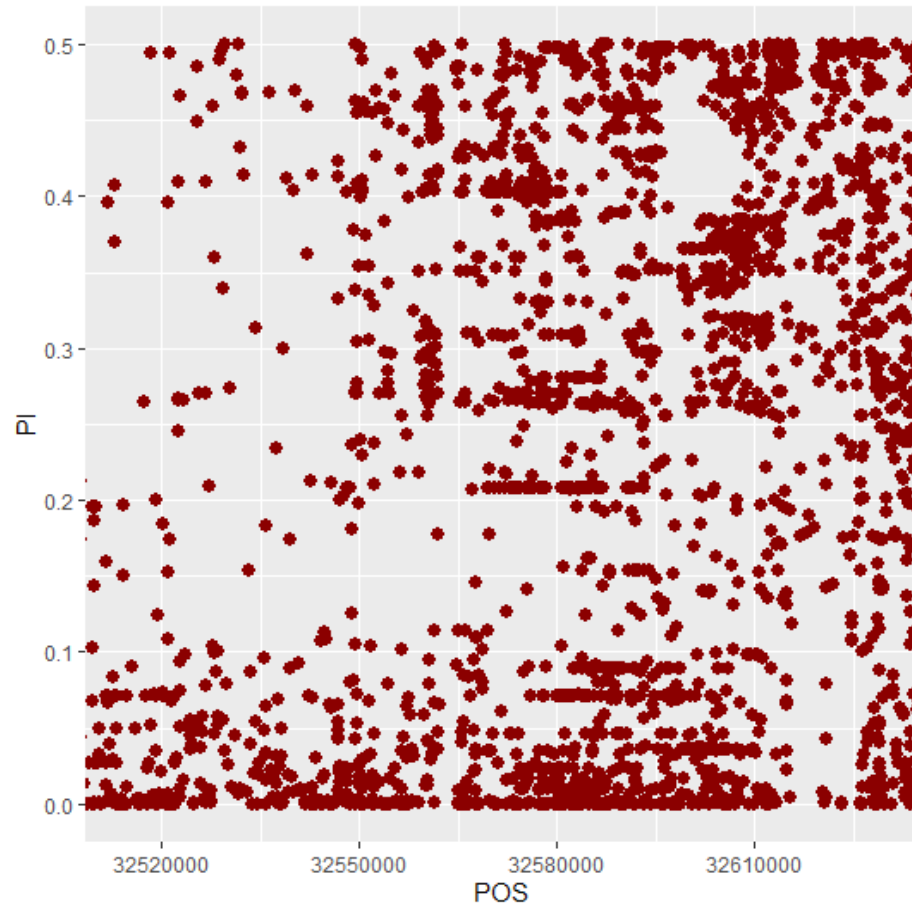


MDRF

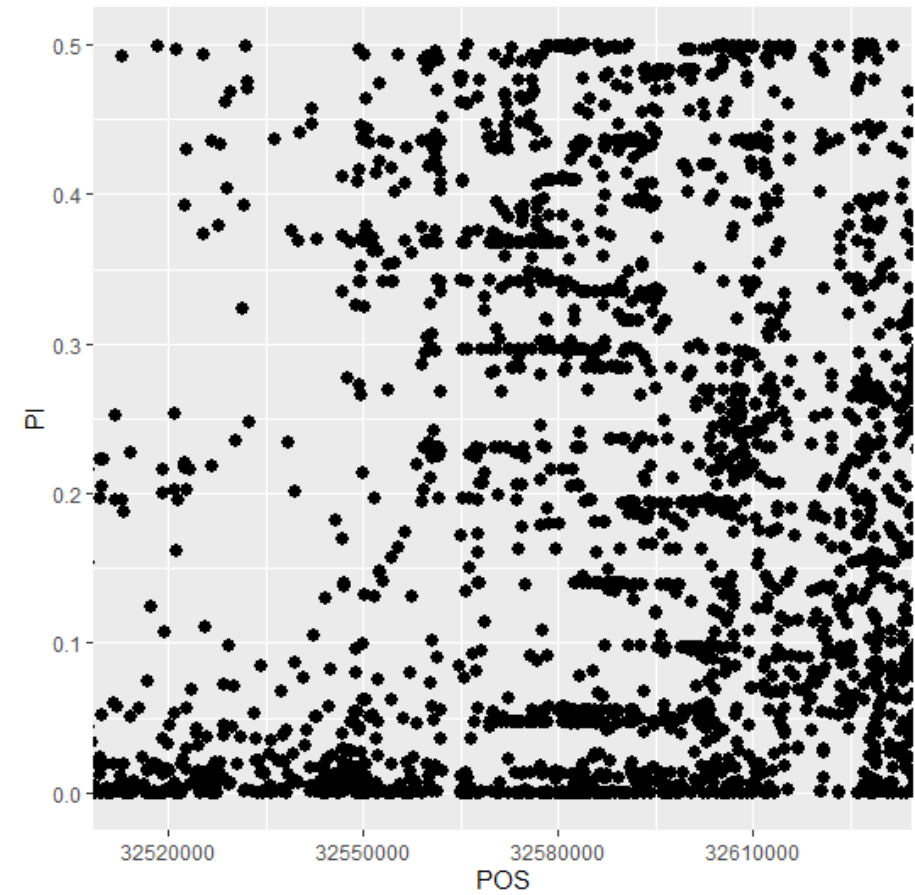


# Nucleotide Diversity - HLA

GoShare



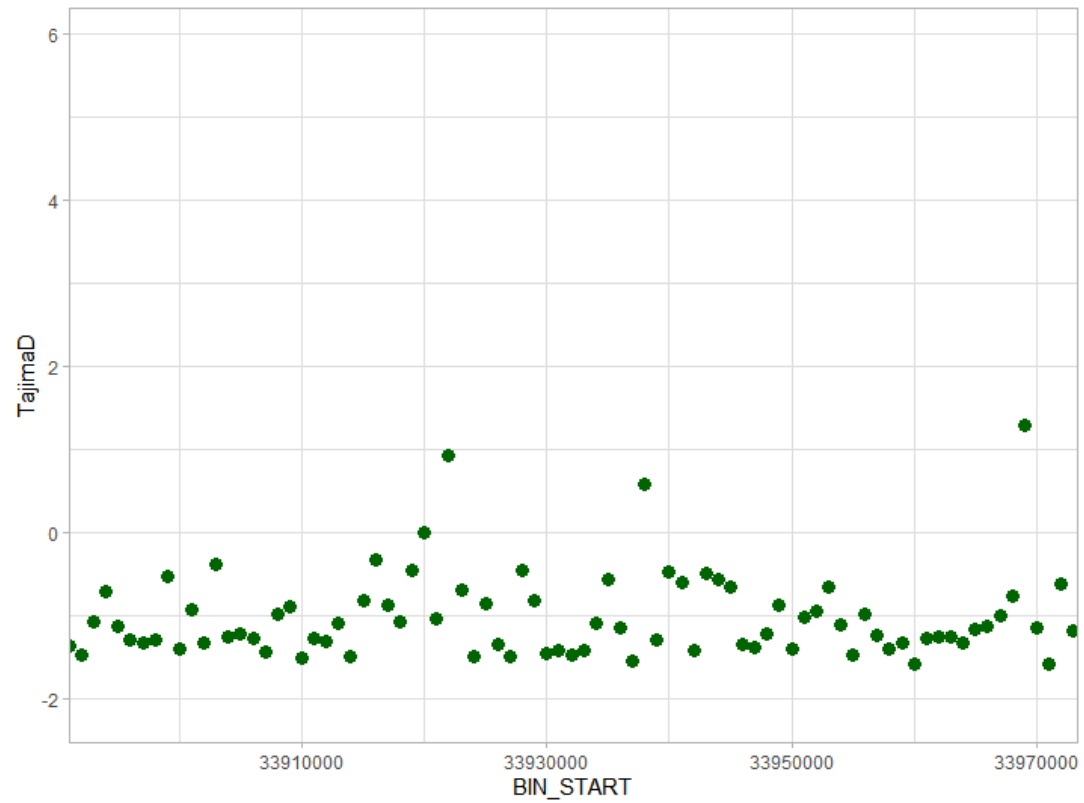
MDRF



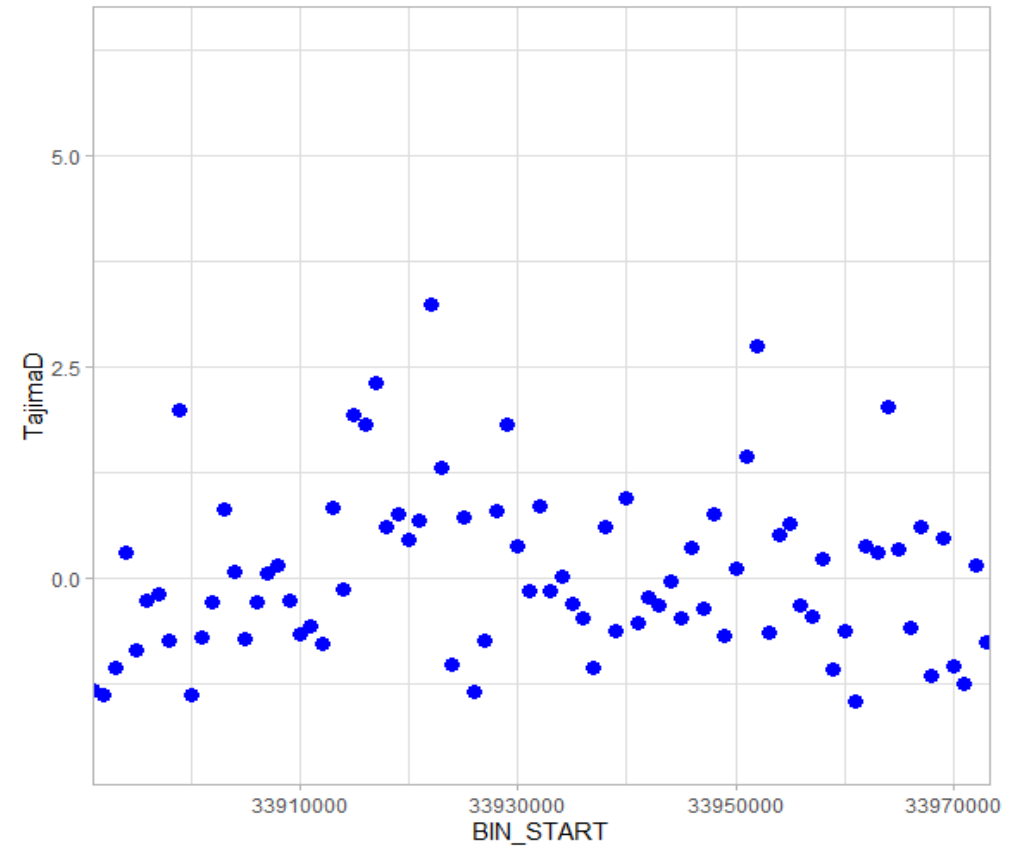


# TajimasD - SLC45A2

GoShare

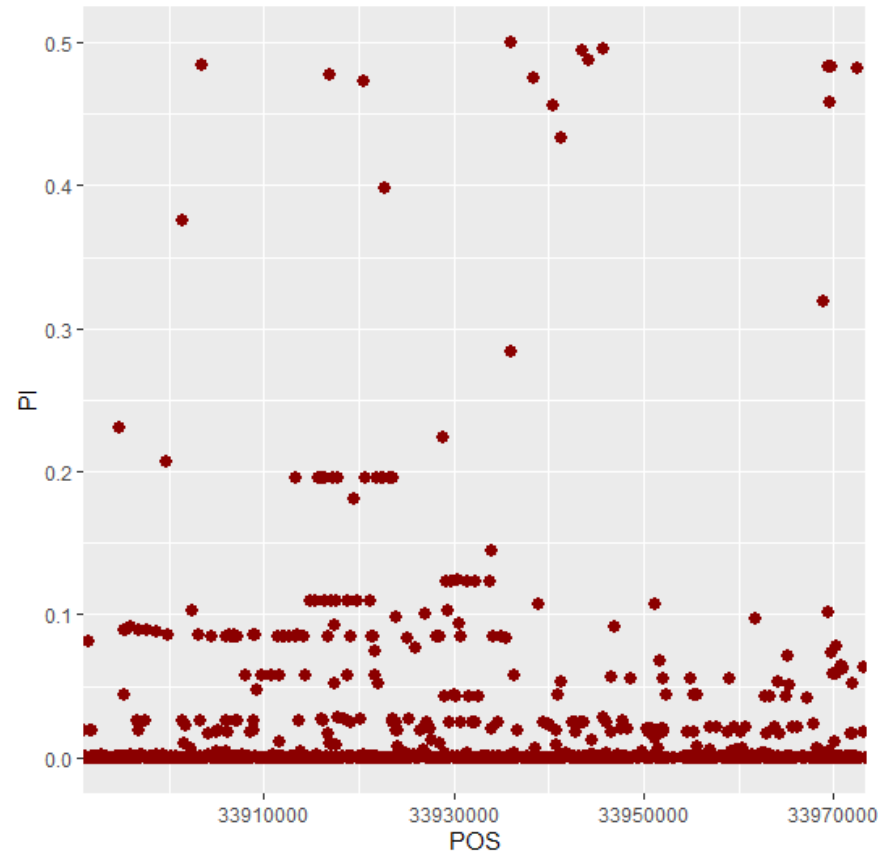


MDRF

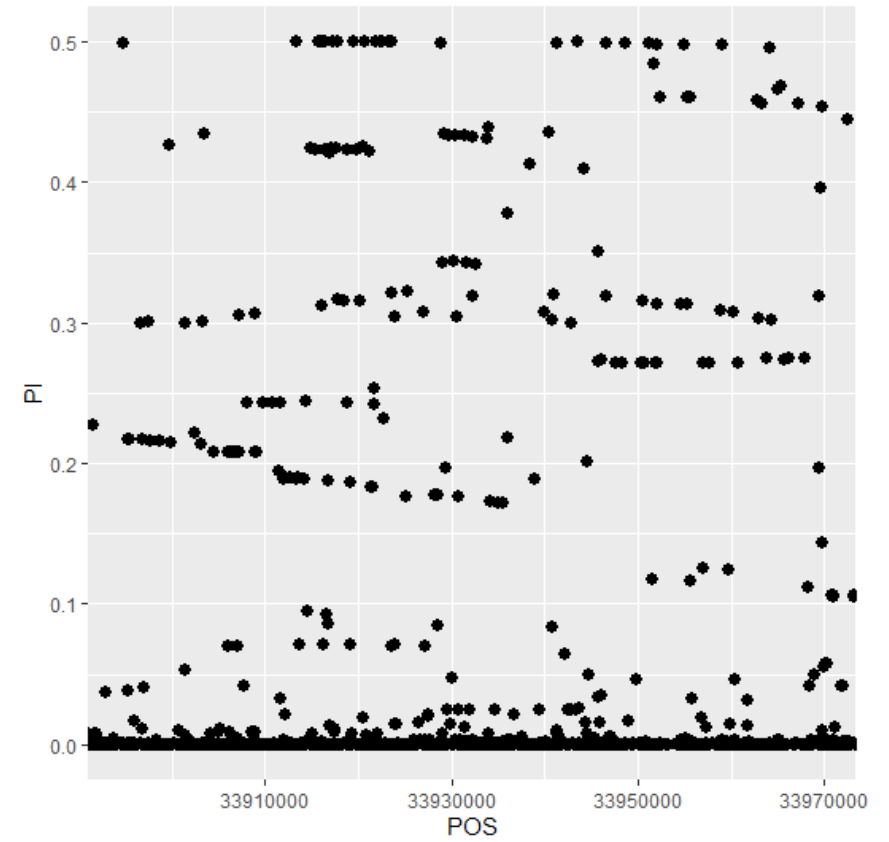


# Nucleotide Diversity - SLC45A2

GoShare



MDRF



# Discussion

- From among ~5000 FST SNPs, 2 regions – HLA and SLC45A2 were found to be associated with age at onset of Type 2 Diabetes in the Scottish population.
- Within HLA, two SNPs – rs9273242 and rs9273410 showed the highest probability of causality.
- rs9273242 has been known to be associated with Asthma, Celiac disease and hypothyroidism.
- Apart from the above traits, rs9273410 has been associated with Age at which diabetes diagnosed in UK Biobank (  $p=3.2e-43$ ).
- No selective sweep was observed in the HLA gene in either populations.

- Within SLC45A2, rs16891982 showed the highest probability of causality.
- The region had shown presence of selective sweep in the Scottish population, but not in the Indian population.
- OLA1 region in MDRF , though not highly significant , shows considerable promise.
- It reduces age onset by almost a year and is very frequent in Indians.
- It interacts with BRCA1 and is implicated in breast cancers.
- The region needs to be replicated and further analysis done to better understand its importance.

Thank you