

# Selection Pressures on the ACE2 Gene

Prof. Colin NA Palmer  
Dr. Radha Venkatesan

CHARVI NANGIA

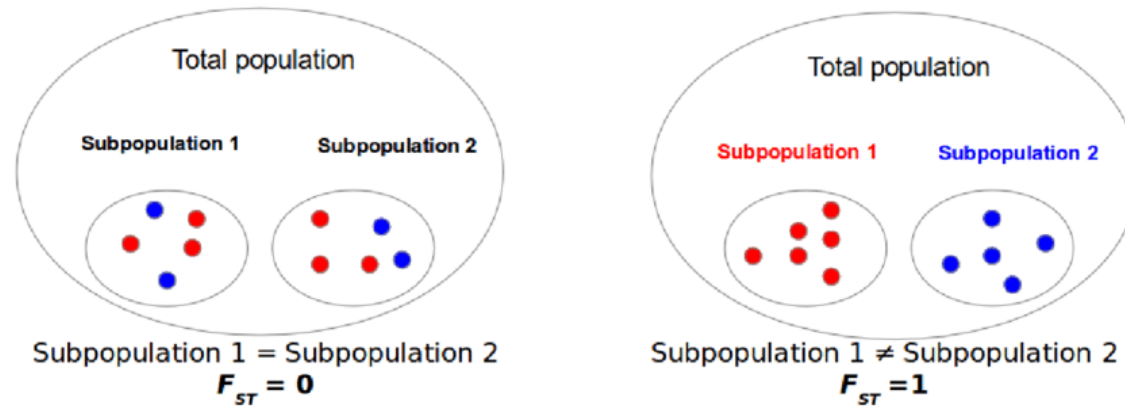


*"This research was funded by the National Institute for Health Research (NIHR) (INSPIRED 16/136/102) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of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

# 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.
- Positive selection –  $F_{st}$  (F statistics), which 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

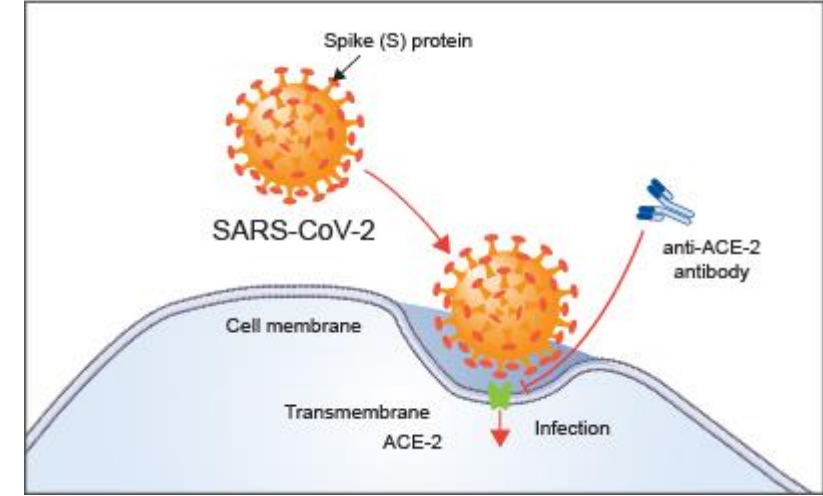
- South Indian diabetes cohort from the Madras Diabetes Research Foundation (MDRF), n=6,056.
- Scottish diabetes cohort from the Scottish Health Research Registry (GoSHARE) study, n=6,681.

# BACKGROUND

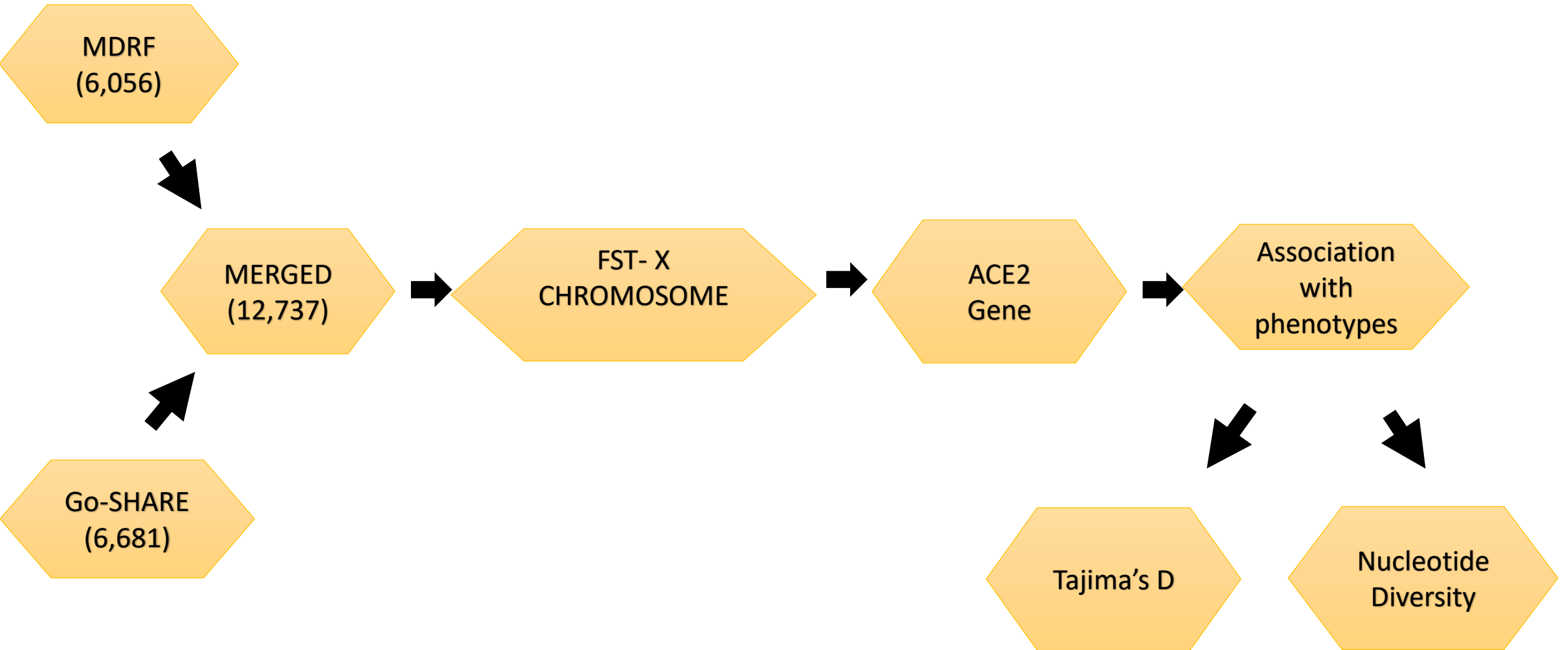


# X CHROMOSOME

- To identify gender-specific footprints of selection.
- Higher regions of selection as compared to autosomes.
- Males have one X chromosome, whereas females have two.
- ACE2 Gene – SARS-CoV2 spike protein binds to its receptor on the X chromosome.



# METHODOLOGY



# FST

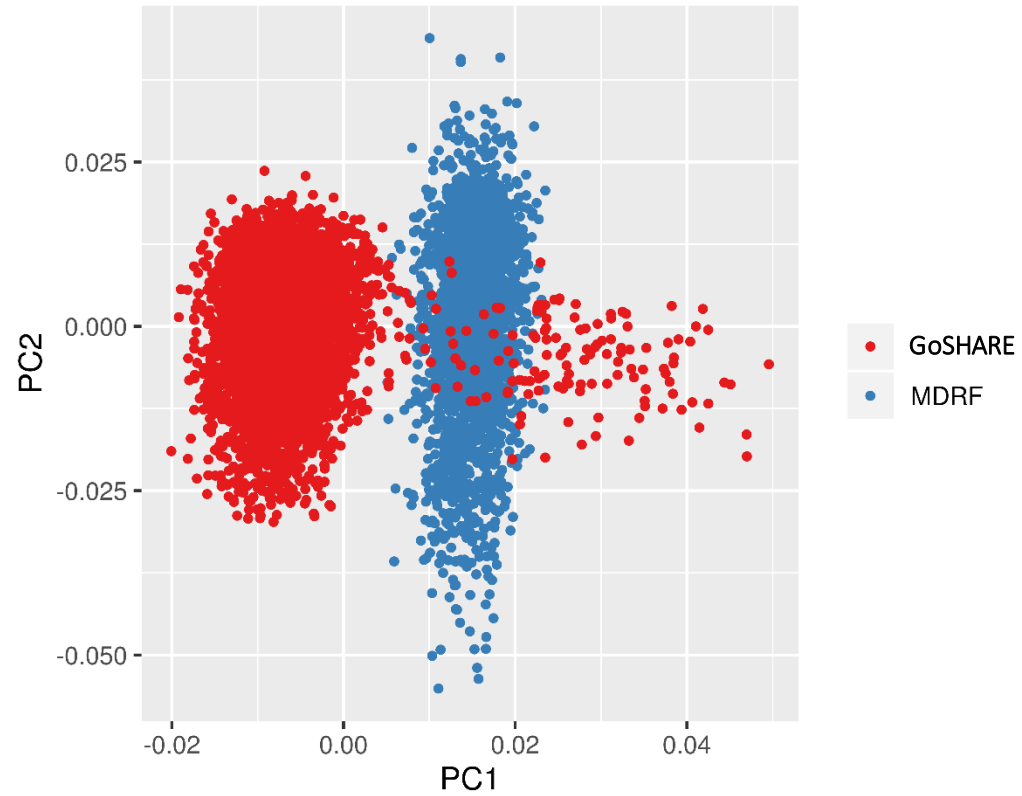
- Pairwise calculation of FST was done on the whole X chromosome.
- Fst values for 1,273,927 SNPs were obtained.
- 269 SNPs present on the ACE2 gene.
- 28 high Fst SNPs (  $Fst > 0.25$ ) were selected for further analysis.



# High Fst SNPs in the ACE2 Gene

ID	FREQ-MDRF	FREQ-GoSHARE
X:15582609	0.646	0.348
X:15582966	0.647	0.351
X:15583904	0.388	0.824
X:15586448	0.636	0.362
X:15586964	0.703	0.369
X:15588271	0.625	0.372
X:15590807	0.624	0.376
X:15597043	0.64	0.393
X:15597835	0.642	0.395
X:15598024	0.643	0.395
X:15600215	0.643	0.395
X:15600691	0.642	0.395
X:15600744	0.642	0.395
X:15601274	0.642	0.395
X:15601719	0.642	0.395
X:15603064	0.41	0.805
X:15604865	0.641	0.396
X:15607650	0.641	0.396
X:15608386	0.641	0.396
X:15608499	0.641	0.396
X:15608952	0.643	0.395
X:15610348	0.41	0.810
X:15614145	0.662	0.391
X:15615453	0.662	0.391
X:15616796	0.662	0.391
X:15617736	0.632	0.519
X:15618061	0.456	0.809
X:15618063	0.634	0.522

# PCA



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

# Phenotypes

## Age at onset

## Anthropometry :

- Height
- Weight
- BMI
- Hip Circumference
- Waist Circumference
- Waist-hip ratio

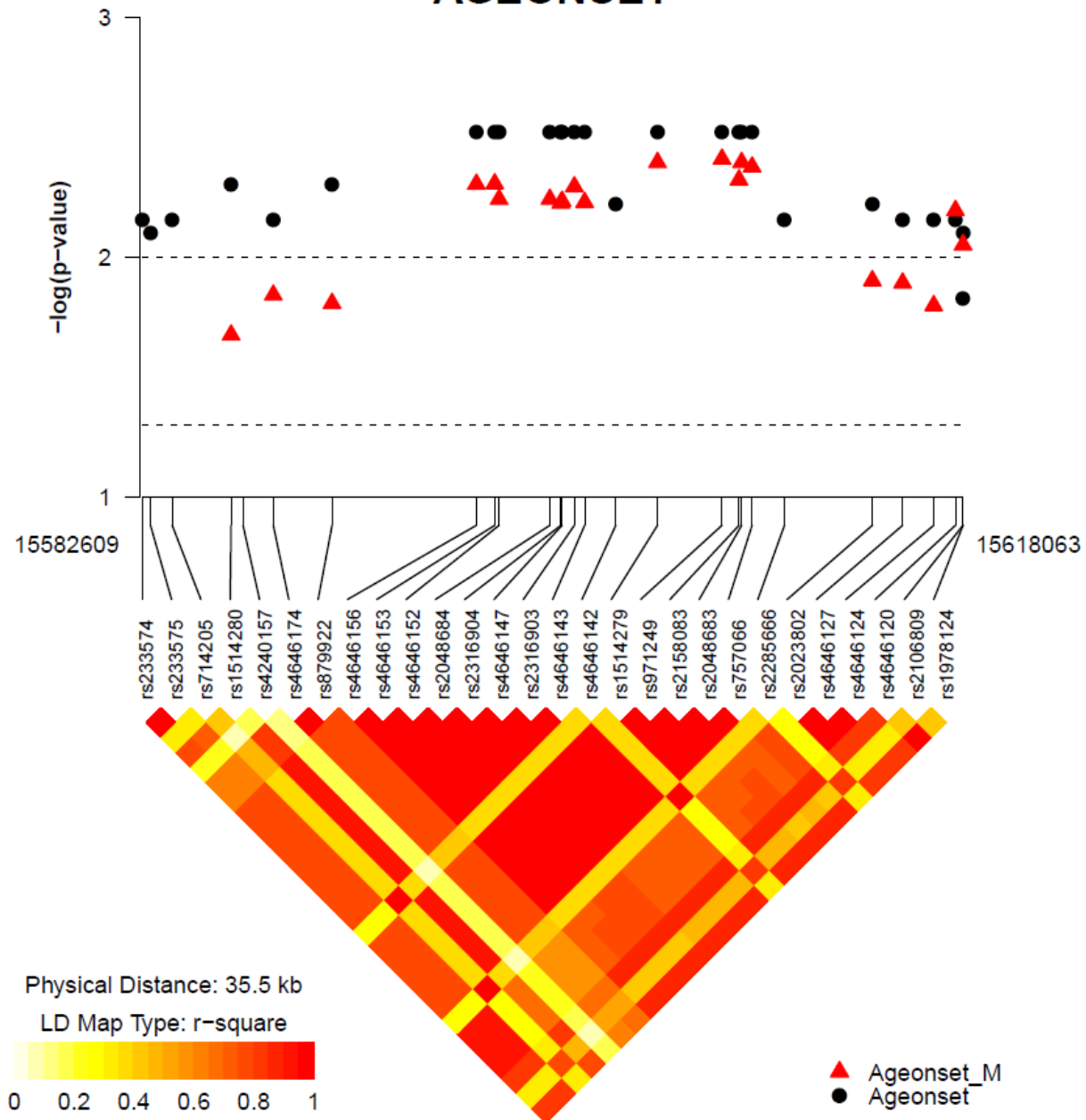
## Blood Pressure

- Systolic BP
- Diastolic BP

## Lipids

- Cholesterol
- Triglycerides
- HDL
- LDL

# AGEONSET

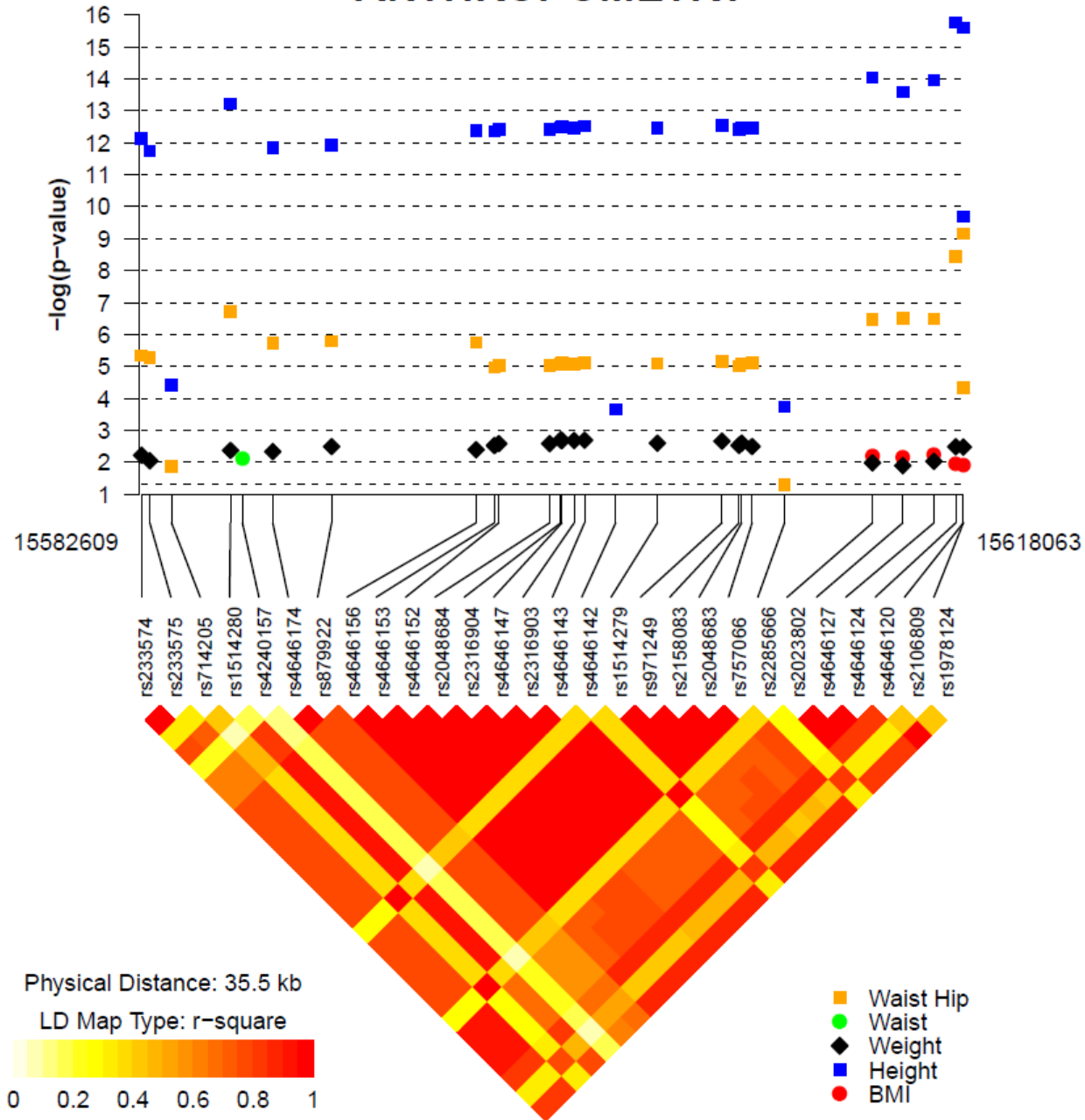


# Ageonset

SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15582609	rs233574	C	T	1.17	7.21E-03	0.761	-0.52	0.59	0.426	1.12E-02	5.55E-02
X:15582966	rs233575	A	G	1.17	7.85E-03	0.764	-0.58	0.55	0.427	9.56E-03	5.57E-02
X:15583904	rs714205	G	C	1.25	7.42E-03	0.896	0.65	0.39	0.59	6.00E-01	7.68E-02
X:15586448	rs1514280	G	A	1.17	5.18E-03	0.753	-0.54	0.54	0.454	4.53E-03	2.11E-02
X:15586964	rs4240157	C	T	-0.32	1.01E-01	0.327	-0.03	0.94	0.266	5.74E-01	5.30E-01
X:15588271	rs4646174	G	C	1.11	7.20E-03	0.754	-0.66	0.43	0.484	2.16E-03	1.44E-02
X:15590807	rs879922	G	C	1.15	5.36E-03	0.755	-0.79	0.35	0.484	8.27E-04	1.56E-02
X:15597043	rs4646156	T	A	1.28	3.29E-03	0.779	-0.98	0.3	0.472	5.57E-04	4.95E-03
X:15597835	rs4646153	T	C	1.27	3.19E-03	0.769	-0.94	0.32	0.472	8.00E-04	4.93E-03
X:15598024	rs4646152	G	A	1.27	3.19E-03	0.769	-0.84	0.37	0.472	1.36E-03	5.73E-03
X:15600215	rs2048684	C	A	1.27	3.19E-03	0.769	-0.84	0.37	0.472	1.36E-03	5.73E-03
X:15600691	rs2316904	T	C	1.27	3.37E-03	0.774	-0.82	0.38	0.473	1.49E-03	5.95E-03
X:15600744	rs4646147	A	T	1.27	3.37E-03	0.774	-0.84	0.37	0.472	1.34E-03	5.84E-03
X:15601274	rs2316903	T	G	1.27	3.38E-03	0.773	-0.95	0.31	0.473	7.28E-04	5.07E-03
X:15601719	rs4646143	C	T	1.27	3.38E-03	0.773	-0.84	0.37	0.472	1.36E-03	5.90E-03
X:15603064	rs4646142	C	G	1.27	6.20E-03	0.887	0.02	0.98	0.618	2.92E-01	2.07E-01
X:15604865	rs1514279	A	G	1.28	3.20E-03	0.773	-1.01	0.28	0.47	4.42E-04	4.02E-03
X:15607650	rs971249	C	T	1.28	3.20E-03	0.773	-1.04	0.27	0.47	3.86E-04	3.89E-03
X:15608386	rs2158083	T	C	1.28	3.02E-03	0.769	-0.91	0.34	0.471	8.73E-04	4.75E-03
X:15608499	rs2048683	G	T	1.28	3.20E-03	0.773	-1.01	0.28	0.47	4.42E-04	4.02E-03
X:15608952	rs757066	T	C	1.28	3.06E-03	0.769	-1.05	0.27	0.474	5.03E-04	4.20E-03
X:15610348	rs2285666	T	C	1.25	6.86E-03	0.883	0.12	0.88	0.612	3.37E-01	1.87E-01
X:15614145	rs2023802	A	G	1.04	6.27E-03	0.648	-0.49	0.52	0.363	3.16E-03	1.25E-02
X:15615453	rs4646127	G	A	1.02	6.91E-03	0.647	-0.57	0.45	0.364	1.91E-03	1.28E-02
X:15616796	rs4646124	C	T	1.03	6.66E-03	0.648	-0.6	0.43	0.364	1.66E-03	1.59E-02
X:15617736	rs4646120	A	G	1.02	6.80E-03	0.66	-0.32	0.65	0.366	3.91E-03	6.38E-03
X:15618061	rs2106809	G	A	0.97	1.52E-02	0.761	-0.2	0.75	0.45	2.07E-03	1.82E-01
X:15618063	rs1978124	C	T	1	7.68E-03	0.658	-0.36	0.6	0.36	2.77E-03	8.89E-03

\* Sex-stratified, sex adjusted

# ANTHROPOMETRY



# Height

SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15582609	rs233574	C	T	-0.27	0.29	0.45	0.14	0.79	0.24	0.54	7.20E-13
X:15582966	rs233575	A	G	-0.26	0.31	0.45	0.15	0.78	0.24	0.55	1.80E-12
X:15583904	rs714205	G	C	-0.27	0.34	0.55	0.38	0.36	0.34	0.28	3.90E-05
X:15586448	rs1514280	G	A	-0.3	0.21	0.45	0.03	0.96	0.27	0.6	6.10E-14
X:15586964	rs4240157	C	T	-0.01	0.96	0.2	0.17	0.38	0.15	0.51	2.00E-01
X:15588271	rs4646174	G	C	-0.26	0.29	0.46	-0.14	0.76	0.28	0.78	1.50E-12
X:15590807	rs879922	G	C	-0.26	0.29	0.46	-0.15	0.76	0.28	0.79	1.20E-12
X:15597043	rs4646156	T	A	-0.32	0.22	0.47	-0.16	0.76	0.28	0.74	4.20E-13
X:15597835	rs4646153	T	C	-0.35	0.18	0.46	-0.24	0.65	0.28	0.82	4.40E-13
X:15598024	rs4646152	G	A	-0.35	0.18	0.46	-0.16	0.76	0.28	0.7	3.80E-13
X:15600215	rs2048684	C	A	-0.35	0.18	0.46	-0.16	0.76	0.28	0.7	3.80E-13
X:15600691	rs2316904	T	C	-0.33	0.2	0.46	-0.17	0.74	0.28	0.73	3.30E-13
X:15600744	rs4646147	A	T	-0.33	0.2	0.46	-0.16	0.76	0.28	0.72	3.20E-13
X:15601274	rs2316903	T	G	-0.33	0.2	0.46	-0.23	0.66	0.28	0.83	3.40E-13
X:15601719	rs4646143	C	T	-0.33	0.2	0.46	-0.16	0.76	0.28	0.72	3.00E-13
X:15603064	rs4646142	C	G	-0.3	0.28	0.55	0.34	0.43	0.35	0.29	2.20E-04
X:15604865	rs1514279	A	G	-0.33	0.2	0.46	-0.27	0.61	0.28	0.89	3.50E-13
X:15607650	rs971249	C	T	-0.33	0.2	0.46	-0.28	0.59	0.27	0.91	2.90E-13
X:15608386	rs2158083	T	C	-0.35	0.18	0.46	-0.22	0.67	0.28	0.8	3.90E-13
X:15608499	rs2048683	G	T	-0.33	0.2	0.46	-0.27	0.61	0.28	0.89	3.50E-13
X:15608952	rs757066	T	C	-0.35	0.17	0.46	-0.28	0.59	0.28	0.9	3.50E-13
X:15610348	rs2285666	T	C	-0.27	0.33	0.54	0.47	0.26	0.35	0.21	1.80E-04
X:15614145	rs2023802	A	G	-0.16	0.47	0.39	0.03	0.94	0.2	0.72	9.10E-15
X:15615453	rs4646127	G	A	-0.14	0.54	0.39	-0.02	0.96	0.2	0.74	2.60E-14
X:15616796	rs4646124	C	T	-0.17	0.46	0.39	-0.01	0.98	0.2	0.68	1.10E-14
X:15617736	rs4646120	A	G	-0.09	0.67	0.4	-0.29	0.45	0.21	0.57	1.70E-16
X:15618061	rs2106809	G	A	-0.21	0.37	0.47	-0.09	0.79	0.26	0.71	2.10E-10
X:15618063	rs1978124	C	T	-0.05	0.82	0.4	-0.27	0.48	0.2	0.54	2.50E-16

\* Sex-stratified, age & sex adjusted

# Weight

SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15582609	rs233574	C	T	-0.494	0.311	0.852	0.9	0.39	0.467	0.202	5.99E-03
X:15582966	rs233575	A	G	-0.452	0.35	0.843	0.965	0.36	0.467	0.197	8.94E-03
X:15586448	rs1514280	G	A	-0.372	0.426	0.847	-0.135	0.89	0.505	0.835	4.25E-03
X:15588271	rs4646174	G	C	-0.365	0.428	0.863	-1.083	0.271	0.561	0.53	4.60E-03
X:15590807	rs879922	G	C	-0.355	0.444	0.866	-1.289	0.192	0.562	0.416	3.21E-03
X:15597043	rs4646156	T	A	-0.221	0.655	0.889	-1.178	0.268	0.532	0.438	4.00E-03
X:15597835	rs4646153	T	C	-0.262	0.593	0.877	-1.501	0.157	0.53	0.312	2.98E-03
X:15598024	rs4646152	G	A	-0.262	0.593	0.877	-1.515	0.155	0.53	0.308	2.60E-03
X:15600215	rs2048684	C	A	-0.262	0.593	0.877	-1.515	0.155	0.53	0.308	2.60E-03
X:15600691	rs2316904	T	C	-0.305	0.535	0.883	-1.53	0.15	0.531	0.32	1.98E-03
X:15600744	rs4646147	A	T	-0.305	0.535	0.883	-1.515	0.155	0.53	0.326	2.08E-03
X:15601274	rs2316903	T	G	-0.306	0.533	0.882	-1.53	0.149	0.53	0.319	2.08E-03
X:15601719	rs4646143	C	T	-0.306	0.533	0.882	-1.515	0.155	0.53	0.326	2.02E-03
X:15604865	rs1514279	A	G	-0.307	0.532	0.882	-1.464	0.165	0.528	0.344	2.50E-03
X:15607650	rs971249	C	T	-0.307	0.532	0.882	-1.502	0.155	0.527	0.329	2.18E-03
X:15608386	rs2158083	T	C	-0.263	0.59	0.876	-1.484	0.16	0.528	0.318	2.96E-03
X:15608499	rs2048683	G	T	-0.307	0.532	0.882	-1.464	0.165	0.528	0.344	2.50E-03
X:15608952	rs757066	T	C	-0.269	0.581	0.876	-1.406	0.19	0.532	0.359	3.20E-03
X:15614145	rs2023802	A	G	0.061	0.885	0.735	0.082	0.924	0.41	0.984	1.03E-02
X:15615453	rs4646127	G	A	0.091	0.83	0.735	0.07	0.934	0.411	0.984	1.26E-02
X:15616796	rs4646124	C	T	0.048	0.91	0.735	0.108	0.9	0.411	0.953	9.23E-03
X:15617736	rs4646120	A	G	0.043	0.918	0.744	-0.747	0.345	0.412	0.349	3.24E-03
X:15618063	rs1978124	C	T	0.058	0.889	0.739	-0.681	0.385	0.41	0.376	3.31E-03

\* Sex-stratified, age & sex adjusted



# Waist-Hip Ratio

SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15582609	rs233574	C	T	-0.003	0.236	0.005	-0.003	0.638	0.003	0.99	4.60E-06
X:15582966	rs233575	A	G	-0.003	0.208	0.004	-0.002	0.739	0.003	0.885	5.25E-06
X:15583904	rs714205	G	C	-0.002	0.56	0.005	0.006	0.232	0.004	0.21	1.35E-02
X:15586448	rs1514280	G	A	-0.003	0.148	0.005	-0.001	0.813	0.003	0.724	1.94E-07
X:15588271	rs4646174	G	C	-0.003	0.291	0.005	0	0.979	0.003	0.673	1.90E-06
X:15590807	rs879922	G	C	-0.002	0.33	0.005	-0.001	0.803	0.003	0.866	1.61E-06
X:15597043	rs4646156	T	A	-0.003	0.175	0.005	-0.001	0.863	0.003	0.705	1.77E-06
X:15597835	rs4646153	T	C	-0.002	0.335	0.005	-0.001	0.915	0.003	0.777	1.08E-05
X:15598024	rs4646152	G	A	-0.002	0.335	0.005	-0.001	0.86	0.003	0.83	9.45E-06
X:15600215	rs2048684	C	A	-0.002	0.335	0.005	-0.001	0.86	0.003	0.83	9.45E-06
X:15600691	rs2316904	T	C	-0.002	0.323	0.005	-0.001	0.865	0.003	0.817	8.19E-06
X:15600744	rs4646147	A	T	-0.002	0.323	0.005	-0.001	0.86	0.003	0.822	7.71E-06
X:15601274	rs2316903	T	G	-0.002	0.327	0.005	-0.001	0.914	0.003	0.772	8.57E-06
X:15601719	rs4646143	C	T	-0.002	0.327	0.005	-0.001	0.86	0.003	0.825	7.55E-06
X:15604865	rs1514279	A	G	-0.003	0.32	0.005	-0.001	0.903	0.003	0.778	7.96E-06
X:15607650	rs971249	C	T	-0.003	0.32	0.005	-0.001	0.894	0.003	0.786	7.10E-06
X:15608386	rs2158083	T	C	-0.002	0.332	0.005	-0.001	0.908	0.003	0.782	9.78E-06
X:15608499	rs2048683	G	T	-0.003	0.32	0.005	-0.001	0.903	0.003	0.778	7.96E-06
X:15608952	rs757066	T	C	-0.003	0.301	0.005	0	0.969	0.003	0.688	7.87E-06
X:15614145	rs2023802	A	G	-0.002	0.436	0.004	-0.004	0.419	0.002	0.649	3.37E-07
X:15615453	rs4646127	G	A	-0.002	0.388	0.004	-0.004	0.415	0.002	0.672	3.03E-07
X:15616796	rs4646124	C	T	-0.002	0.401	0.004	-0.004	0.472	0.002	0.732	3.19E-07
X:15617736	rs4646120	A	G	-0.003	0.217	0.004	-0.002	0.634	0.002	0.917	3.55E-09
X:15618061	rs2106809	G	A	0	0.828	0.005	0.001	0.789	0.003	0.749	4.70E-05
X:15618063	rs1978124	C	T	-0.003	0.156	0.004	-0.002	0.698	0.002	0.781	6.93E-10

\* Sex-stratified, age & sex adjusted

# BMI

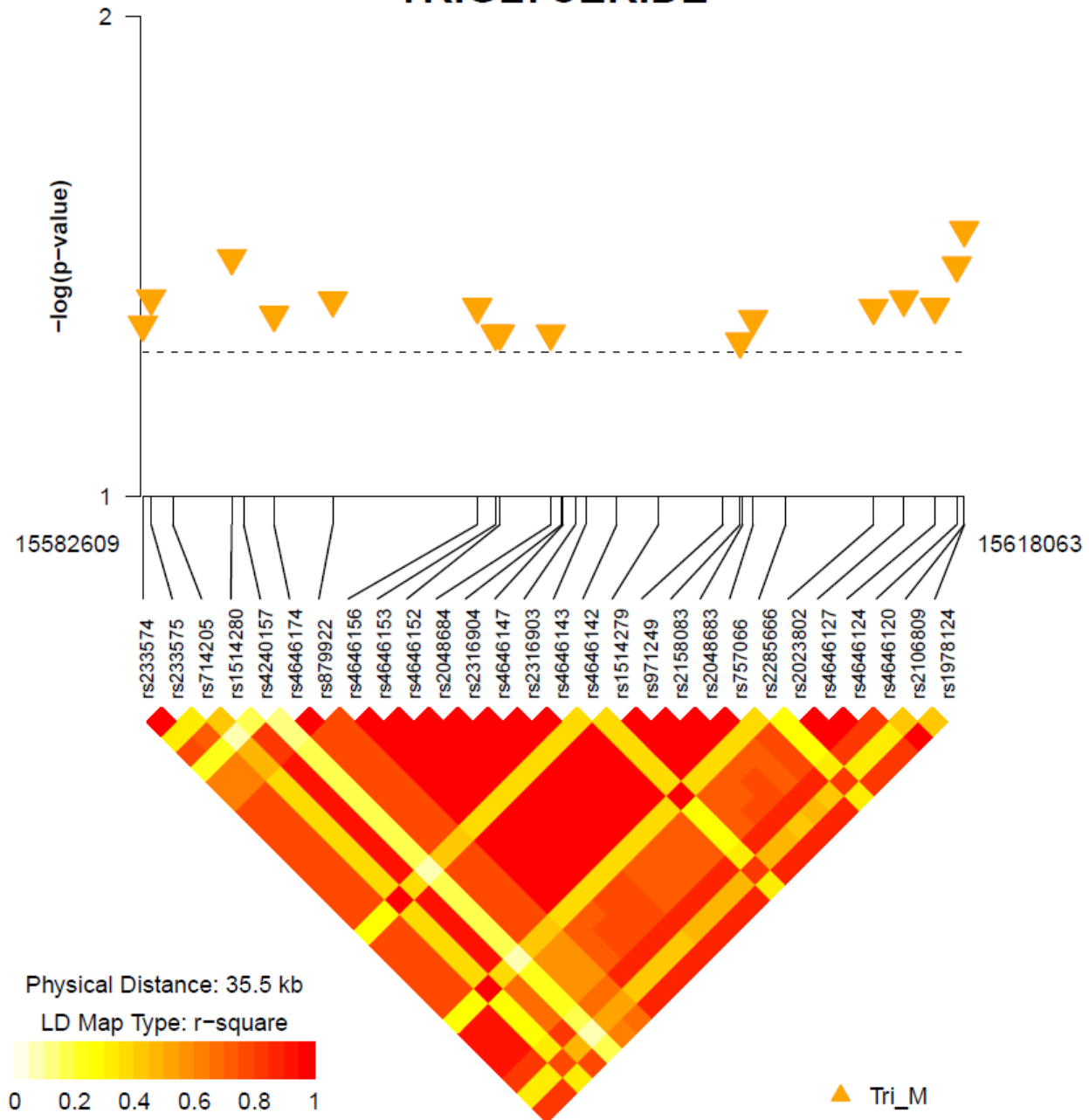
SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15614145	rs2023802	A	G	0.083	0.53	0.226	0.011	0.974	0.16	0.853	6.74E-03
X:15615453	rs4646127	G	A	0.085	0.52	0.225	0.022	0.947	0.16	0.871	7.12E-03
X:15616796	rs4646124	C	T	0.079	0.547	0.226	0.035	0.917	0.16	0.909	5.64E-03
X:15617736	rs4646120	A	G	0.05	0.702	0.229	-0.226	0.47	0.161	0.38	1.17E-02
X:15618063	rs1978124	C	T	0.042	0.748	0.227	-0.211	0.497	0.16	0.418	1.20E-02

# Waist Circumference

SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15586964	rs4240157	C	T	-0.175	0.338	0.307	-0.7953	0.01957	0.2558	0.1013	8.12E-03

\* Sex-stratified, age & sex adjusted

# TRIGLYCERIDE

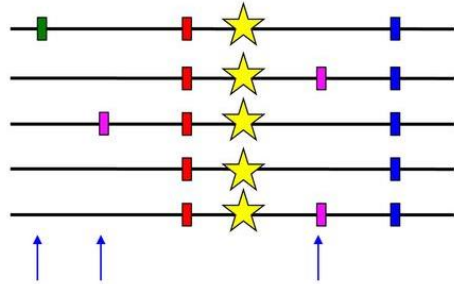


# Triglyceride

SNP	rsid	A1	A2	BETA_M	PVAL_M	SE_M	BETA_F	PVAL_F	SE_F	P_DIFF_BETA	PVAL
X:15582609	rs233574	C	T	-10.54	4.39E-02	9.03	-0.39	0.97	4.27	0.36	0.073
X:15582966	rs233575	A	G	-10.91	3.87E-02	9.09	-1.08	0.91	4.27	0.38	0.058
X:15583904	rs714205	G	C	-10.12	8.70E-02	11.32	-5.38	0.5	6.2	0.63	0.361
X:15586448	rs1514280	G	A	-11.21	3.19E-02	9.36	3.3	0.72	4.66	0.18	0.111
X:15588271	rs4646174	G	C	-10.34	4.19E-02	9.33	2.96	0.74	4.98	0.2	0.161
X:15590807	rs879922	G	C	-10.5	3.91E-02	9.34	1.5	0.87	4.97	0.25	0.126
X:15597043	rs4646156	T	A	-11.06	4.03E-02	9.63	4.88	0.63	4.9	0.17	0.134
X:15597835	rs4646153	T	C	-10.64	4.58E-02	9.48	9.75	0.34	4.9	0.08	0.177
X:15598024	rs4646152	G	A	-10.64	4.58E-02	9.48	9.03	0.38	4.9	0.09	0.165
X:15600215	rs2048684	C	A	-10.64	4.58E-02	9.48	9.03	0.38	4.9	0.09	0.165
X:15608386	rs2158083	T	C	-10.55	4.76E-02	9.47	9.64	0.34	4.9	0.08	0.178
X:15608952	rs757066	T	C	-10.83	4.27E-02	9.5	9.39	0.36	4.91	0.08	0.17
X:15614145	rs2023802	A	G	-9.505	4.05E-02	7.96	8.09	0.33	3.82	0.06	0.174
X:15615453	rs4646127	G	A	-9.554	3.89E-02	7.94	7.74	0.35	3.83	0.07	0.174
X:15616796	rs4646124	C	T	-9.505	4.03E-02	7.96	8.41	0.31	3.83	0.06	0.183
X:15617736	rs4646120	A	G	-9.915	3.30E-02	8.24	7.76	0.3	3.77	0.05	0.202
X:15618063	rs1978124	C	T	-10.18	2.79E-02	8.16	9.9	0.18	3.73	0.02	0.234

\* Sex-stratified, age & sex adjusted

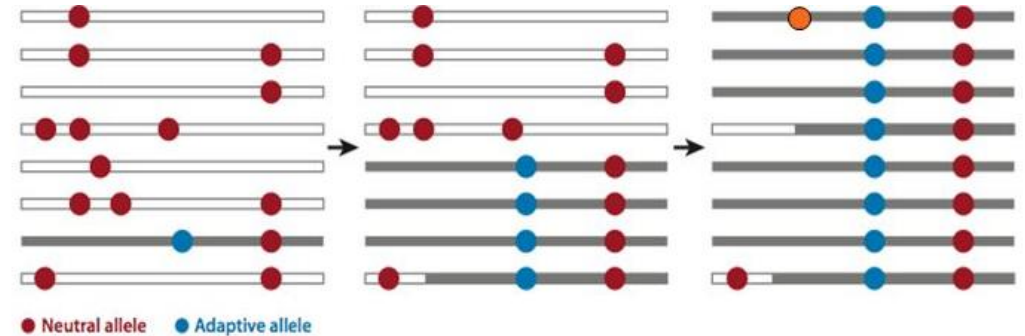
# Tajima's D and Nucleotide Diversity



$S = 3$

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

avg. # difs between each pair of sequences



$\pi$

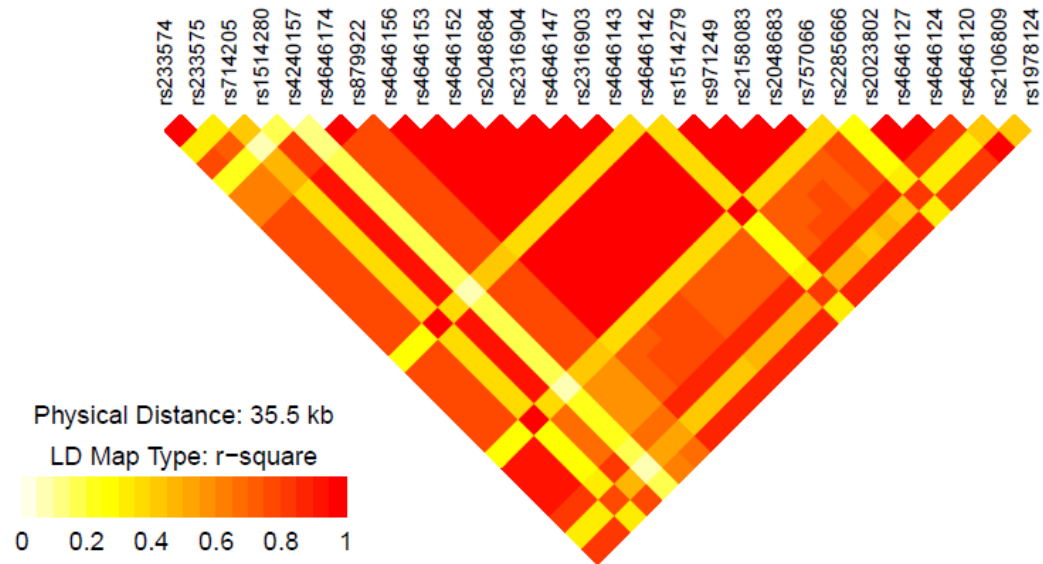
Observed – Expected =  $\pi - \Phi$

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

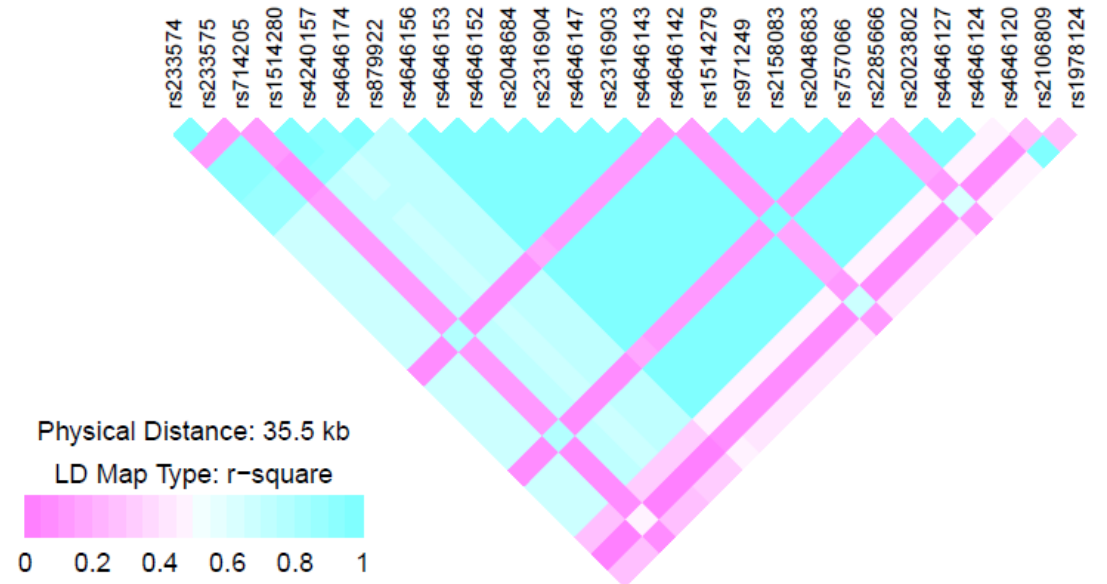
- Nucleotide Diversity  $> 0$  = No sweep
- Nucleotide Diversity = 0 = sweep

# LD STRUCTURE

## MDRF

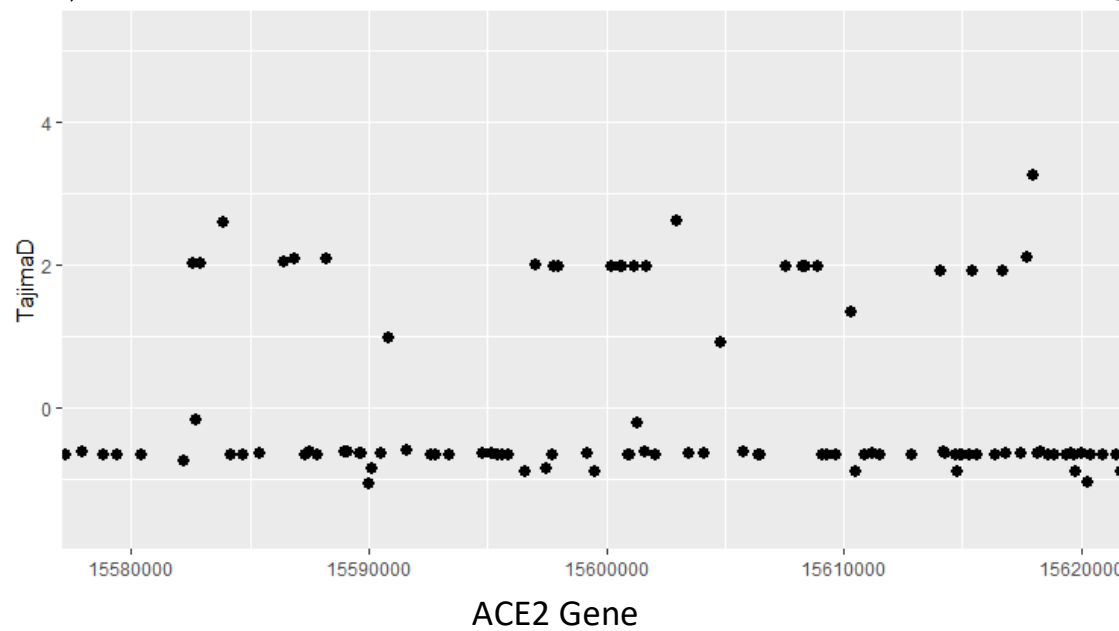
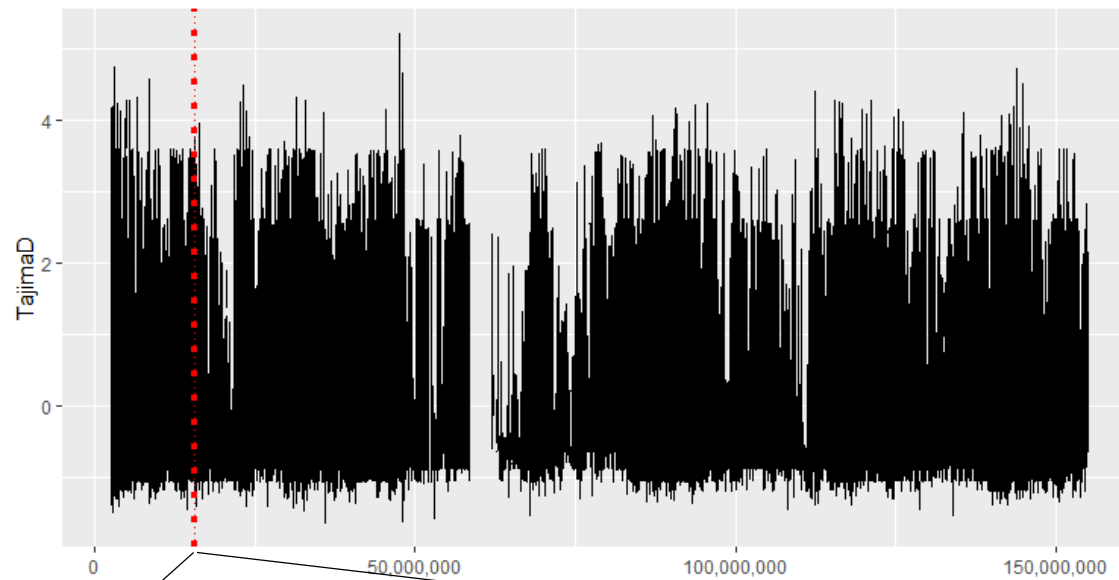


## GoSHARE



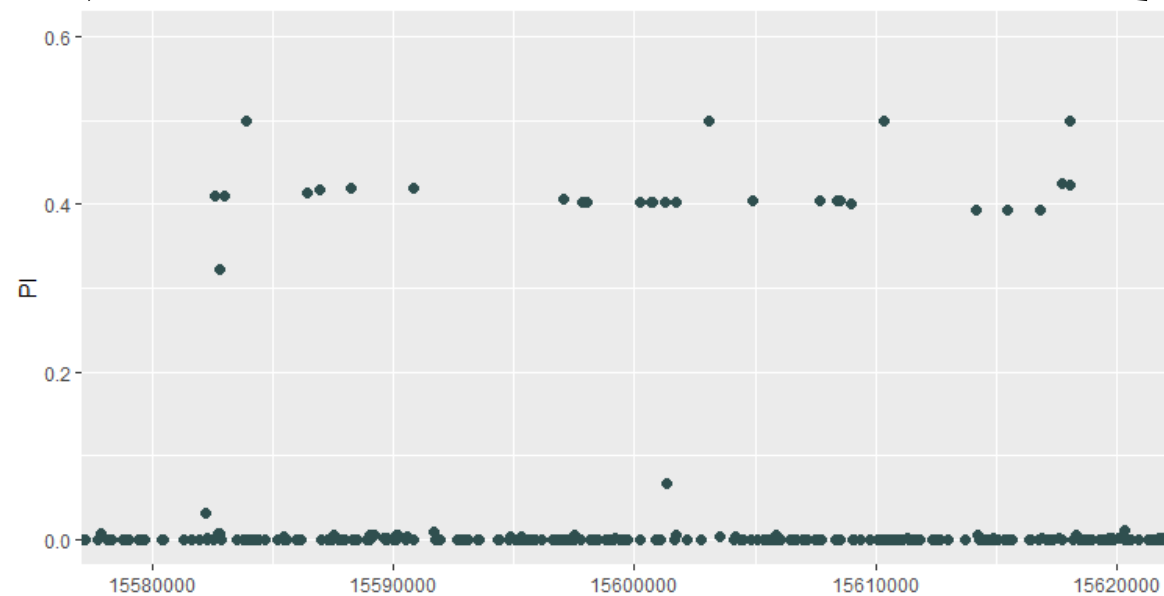
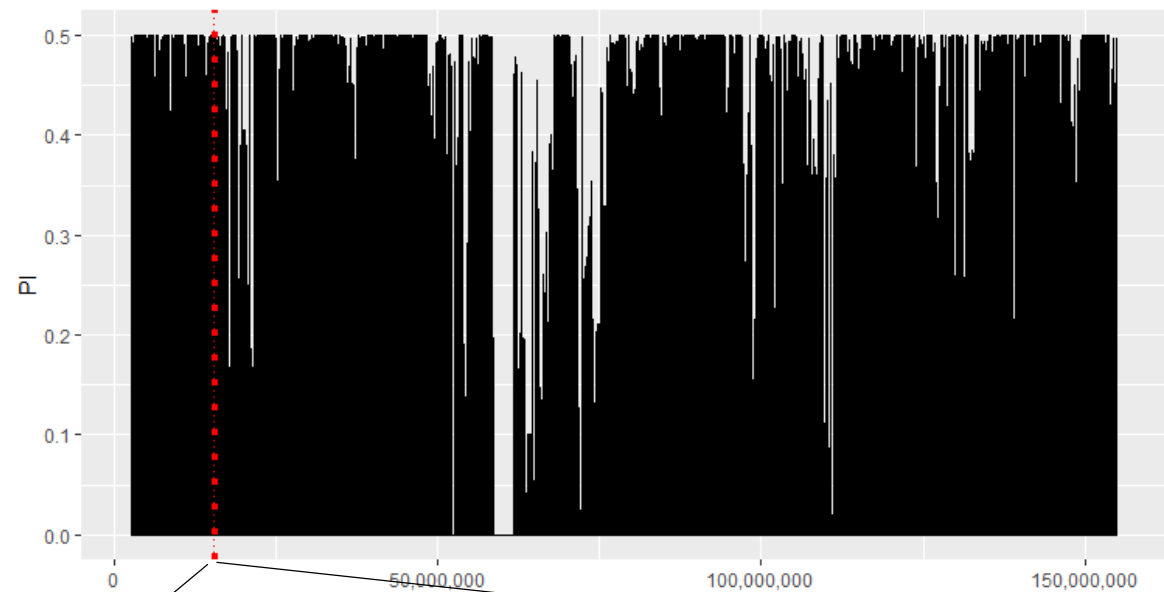
# Tajima's D MDRF

## X CHROMOSOME



# Nucleotide Diversity MDRF

## X CHROMOSOME

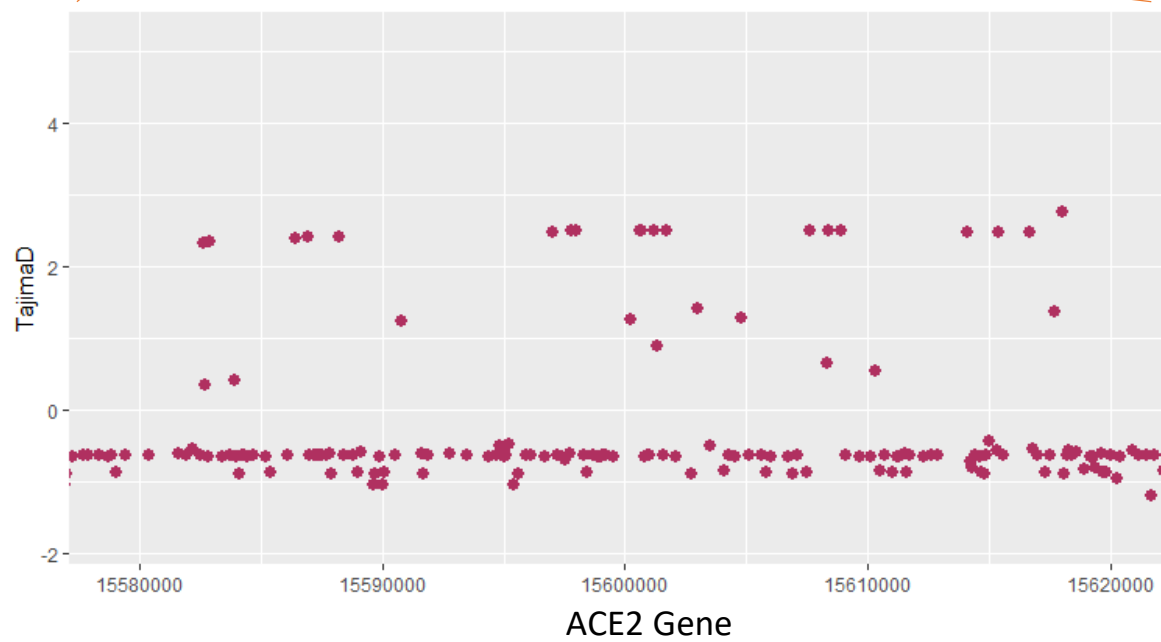
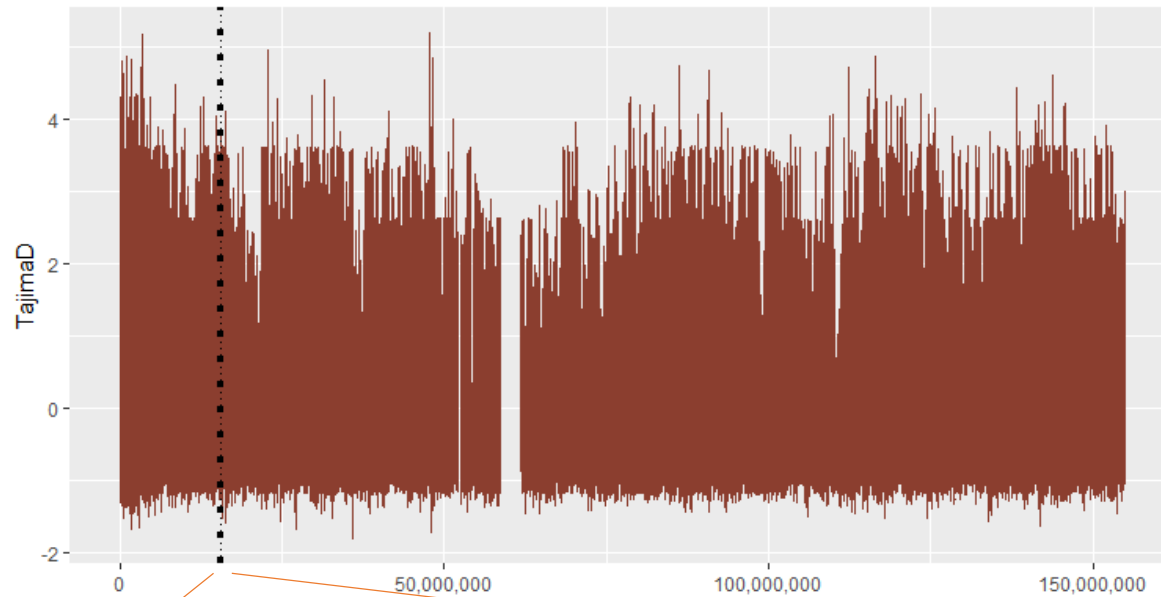


ACE2 Gene



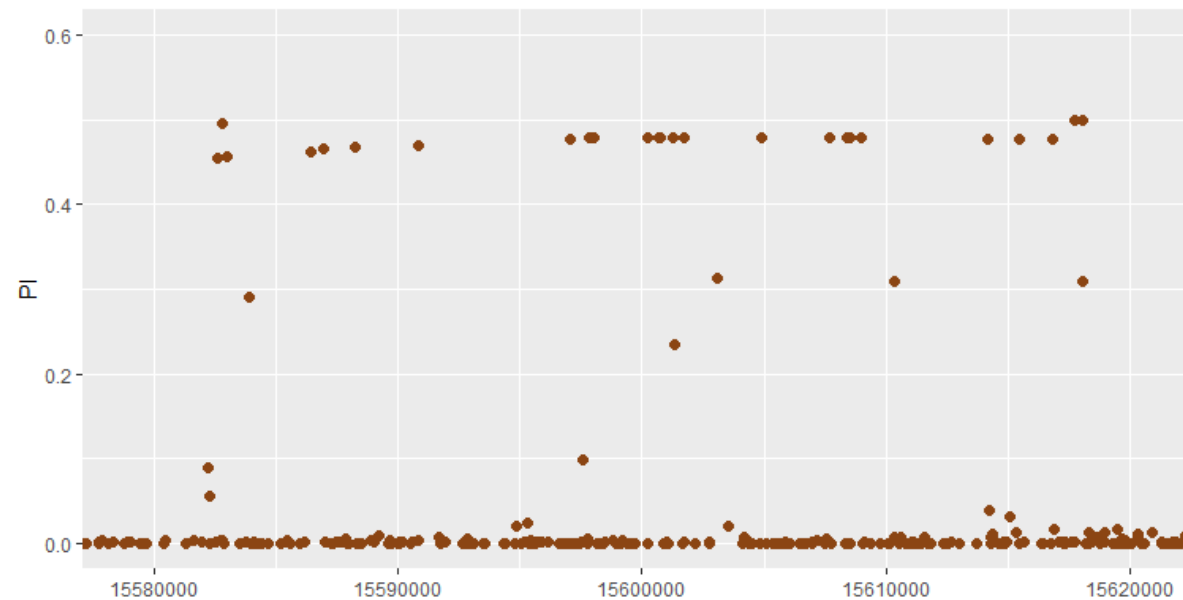
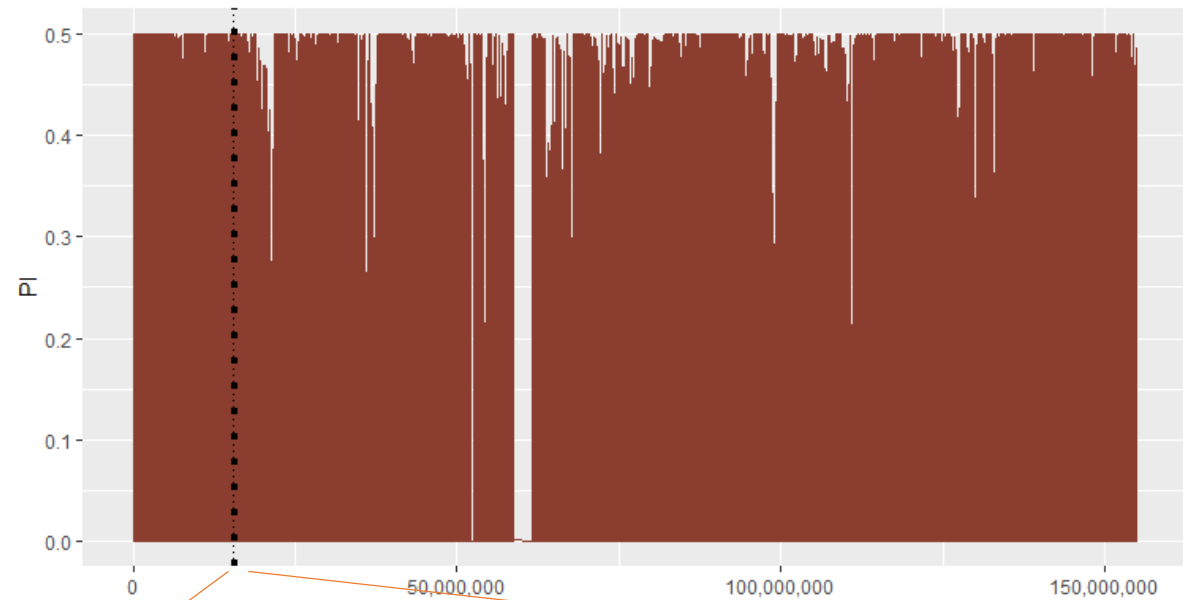
# Tajima's D GoSHARE

## X CHROMOSOME



# Nucleotide Diversity GoSHARE

## X CHROMOSOME



ACE2 Gene

# Balancing Selection

- The above results point towards the alleles showing balancing selection.
- The genetic signature of balancing selection is the presence of polymorphic sites at intermediate (balanced) frequencies in both the populations, as against what would be expected under neutrality.
- Causes could be heterozygote advantage or negative frequency dependent selection (allele's fitness is inversely proportional to its frequency).
- It can also occur when allelic sites evolve independently of one another due to higher rate of recombination.

# SUMMARY

- 28/269 alleles in the ACE2 gene showed high  $F_{st}$ .
- These alleles were associated with 7/13 phenotypes in MDRF.
- None of the alleles were associated with any phenotype in GoSHARE.
- The alleles, though all linked, showed different LD structure between the two populations.
- No evidence was found for a selective sweep, as indicated by high Tajima's  $D$  and Nucleotide diversity values.
- It could be a recent selection event.
- It comes under the category of balanced selection.

# Way Forward...

- Conditional analysis to identify causality.
- McDonald-Kreitman test (dN/dS).
- Ratio of non-synonymous nucleotide changes per non-synonymous site to synonymous changes per synonymous site.
- $dN/dS < 1$  = non-synonymous changes are deleterious.
- $dN/dS > 1$  = non-synonymous changes are advantageous and can rapidly proceed towards fixation.
- Confirms balancing selection.
- To look for selection events in the whole X Chromosome.

**Thank You!**