Identifying genetic markers associated with Chronic Fatigue Syndrome (CFS)

Aliaksei Shaukunou
Mateusz Zacharecki
Julia Girtler

Agenda

01

Descriptive analysis

02

Allele distributions of each biomarker for HC and comparison against distributions from the 1000HGP

03

Information for each genetic marker from the 1000HGP

04

Genotype distributions of each biomarker for both HC and CFS and Hardy-Weinberg Equilibrium

05

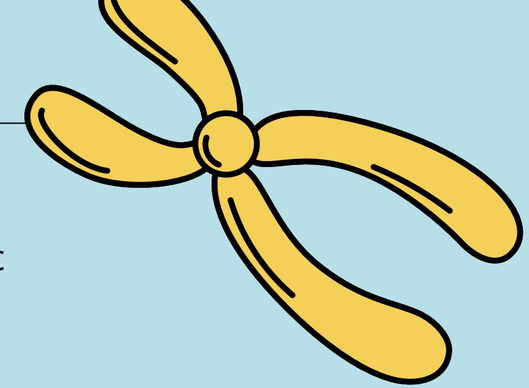
Comparison of the genotype distribution of CFS and HC for each genetic marker

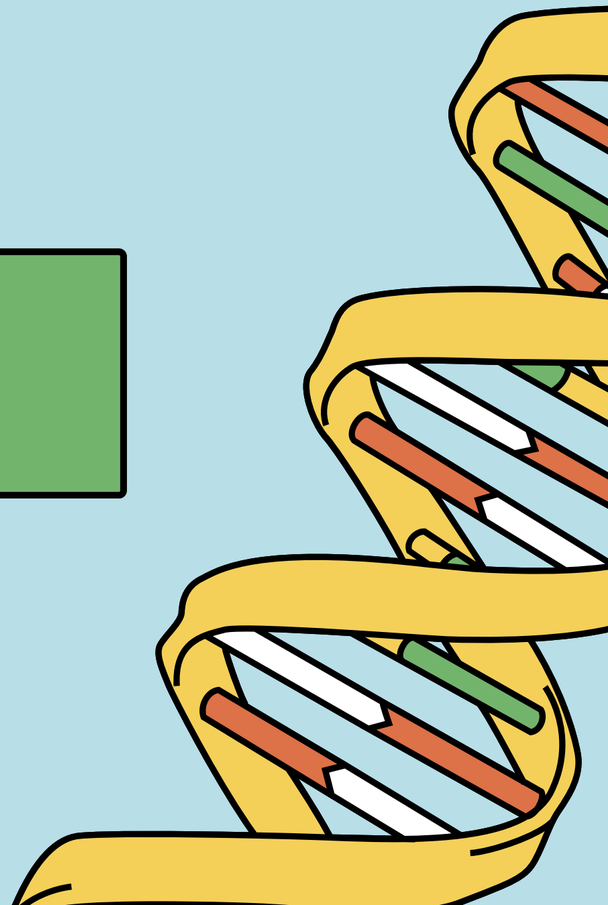
06

Best additive models

07

Can these genetic data be used as a diagnostic tool for CFS?





SECTION 1

Descriptive analysis

Data description

General

Number of patients: **437**
Total NAs: **103**

Missing values

rs2476601: **19**
rs3087243: **24**
rs3807306: **19**
rs1800629: **20**
rs1799724: **21**

Group

CFS (Chronic Fatigue Syndrome)	HC (Healthy Control)
227	210

Gender

male	female
253	184

Biomarkers' distribution

rs2476601

AA	AG	GG	*NA*
2	76	340	19

rs3087243

AA	AG	GG	*NA*
63	202	148	24

rs3807306

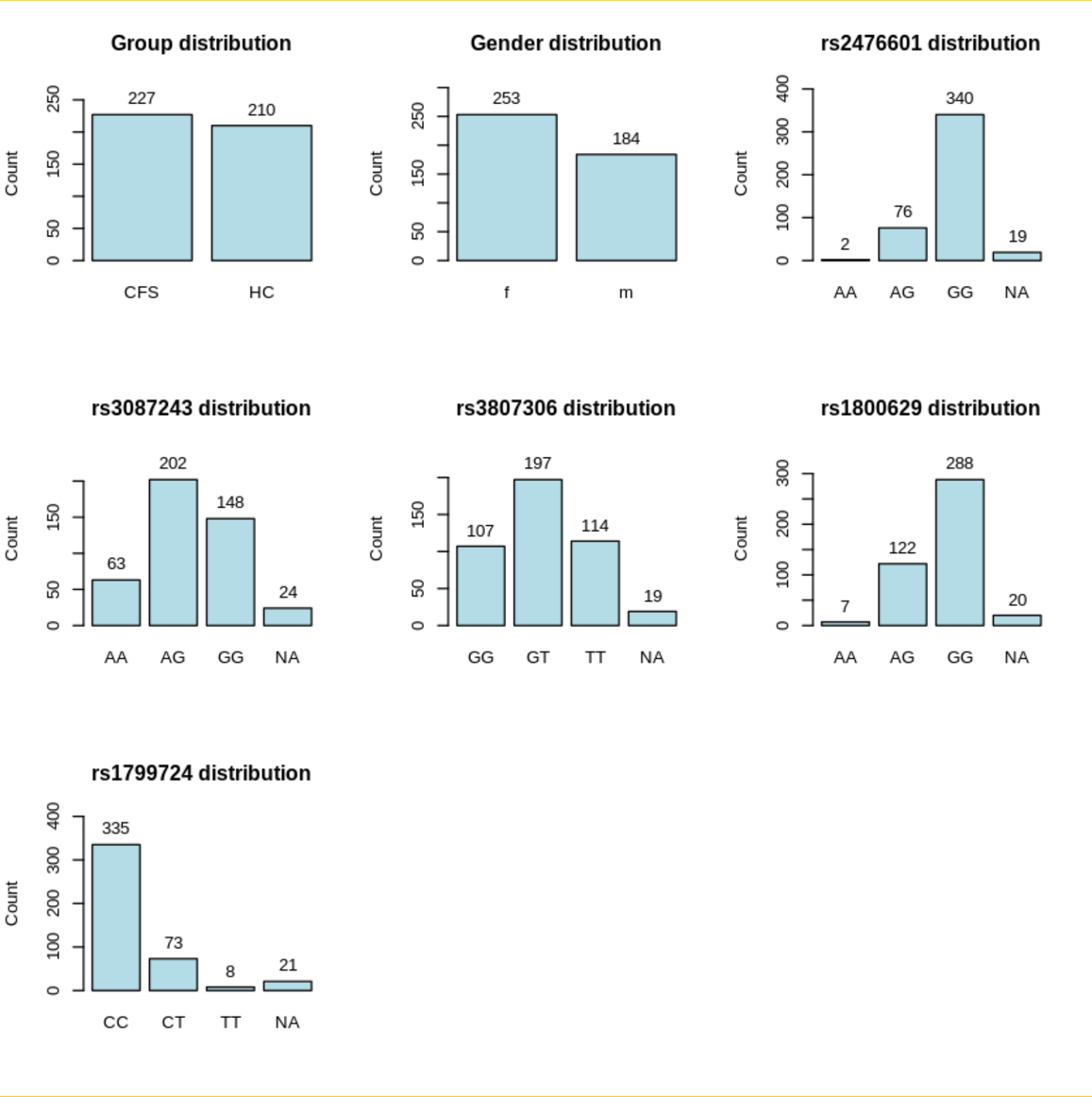
AA	AG	GG	*NA*
107	197	114	19

rs1800629

AA	AG	GG	*NA*
7	122	288	20

rs1799724

AA	AG	GG	*NA*
335	73	8	21



Distribution barplots

SECTION 2

Information for each genetic marker from the 1000HGP

rs2476601

Chromosome: **1**

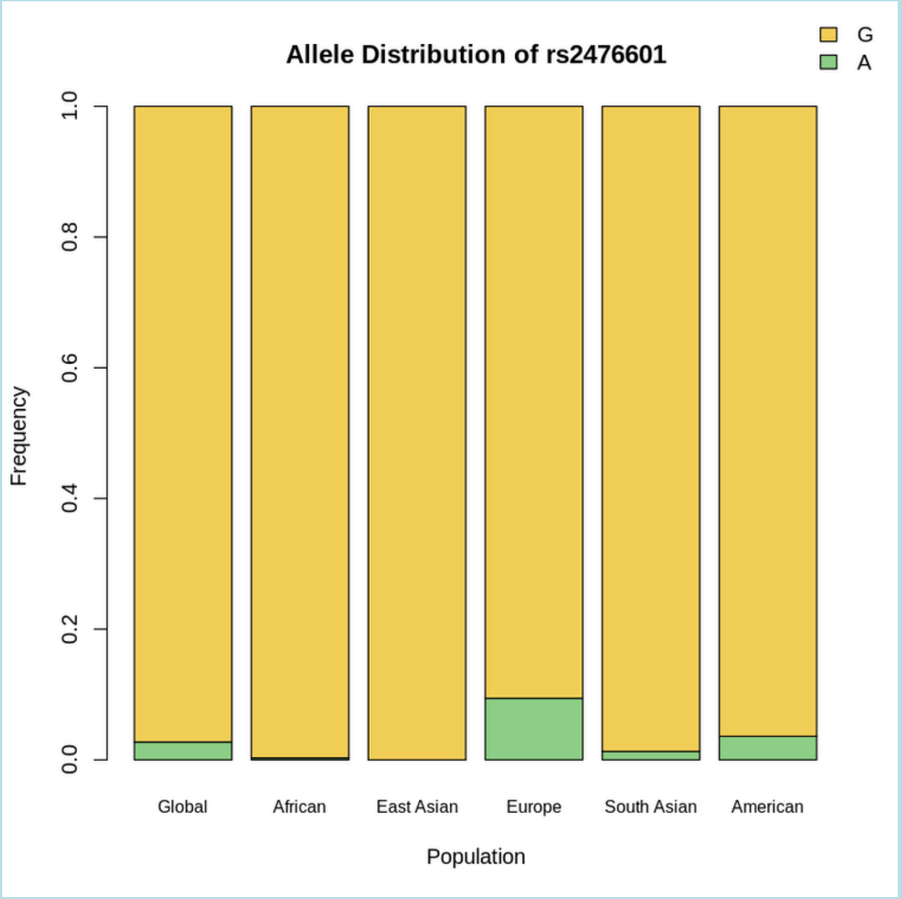
Location: **113834946**

(GRCh38.p14)

<u>Alleles:</u>

Referential Allele: **A**

Alternative Allele: **G**



rs3087243

Chromosome: 2

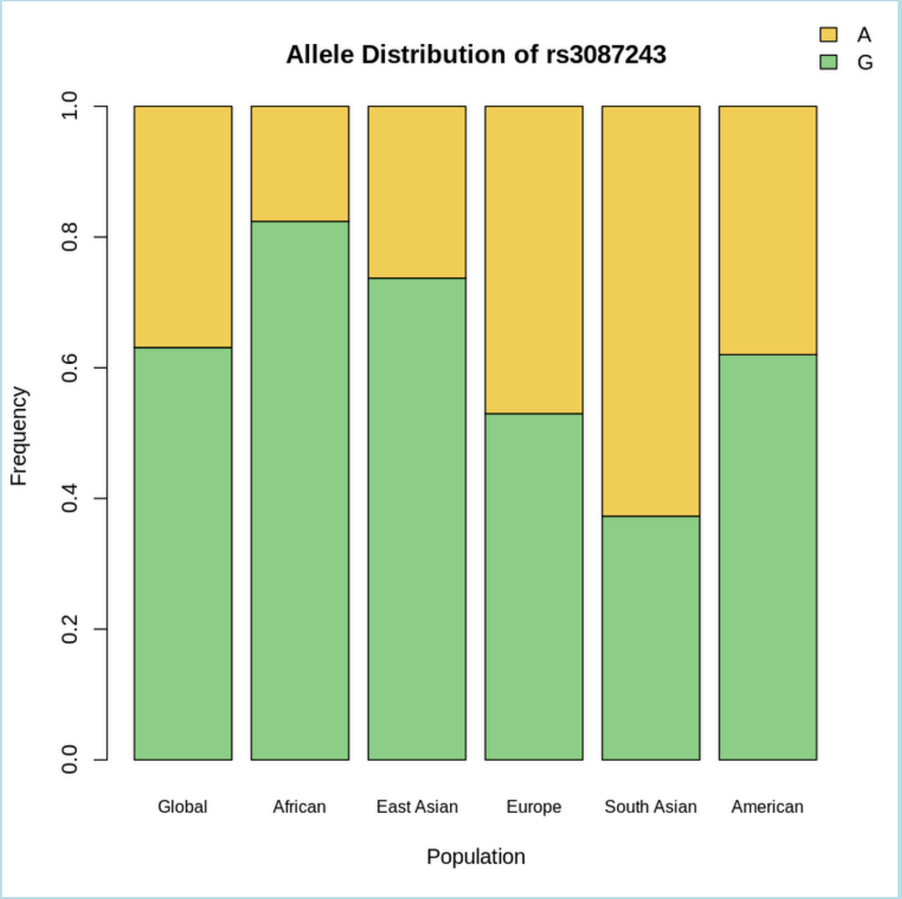
Location: **203874196**

(GRCh38.p14)

<u>Alleles:</u>

Referential Allele: **G**

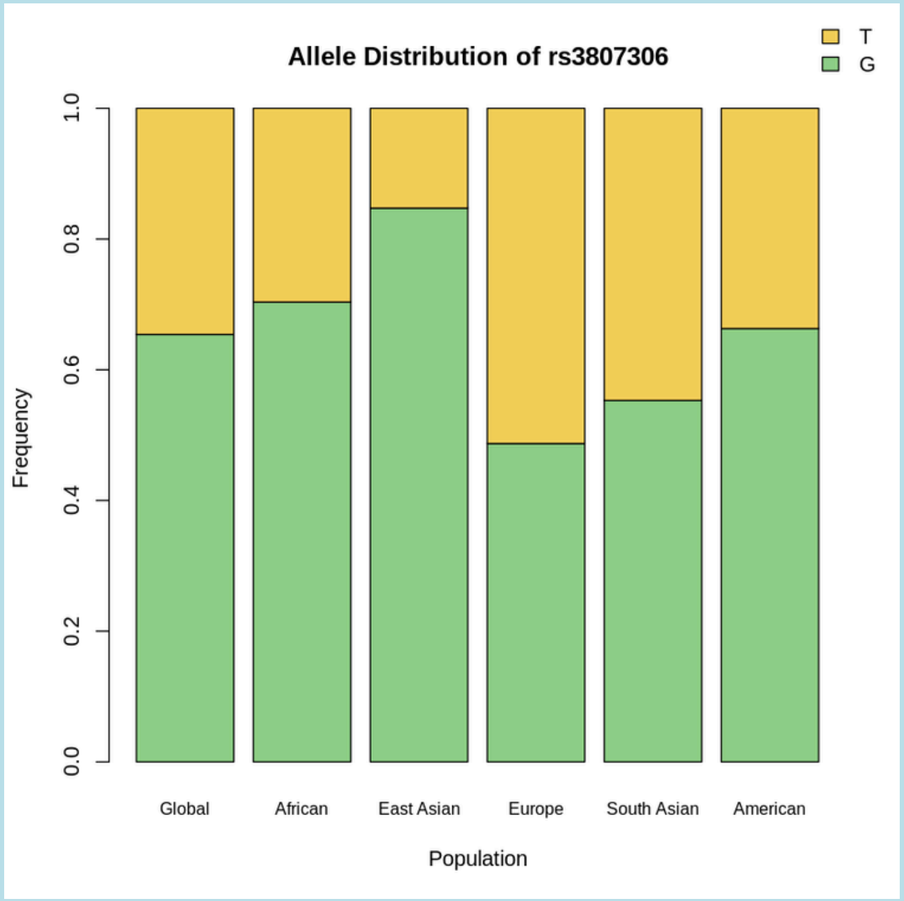
Alternative Allele: **A**



rs3807306

Chromosome: **7**
Location: **128940626**
(GRCh38.p14)

<u>Alleles:</u>
Referential Allele: **G**
Alternative Allele: **T**



rs1800629

Chromosome: **6**

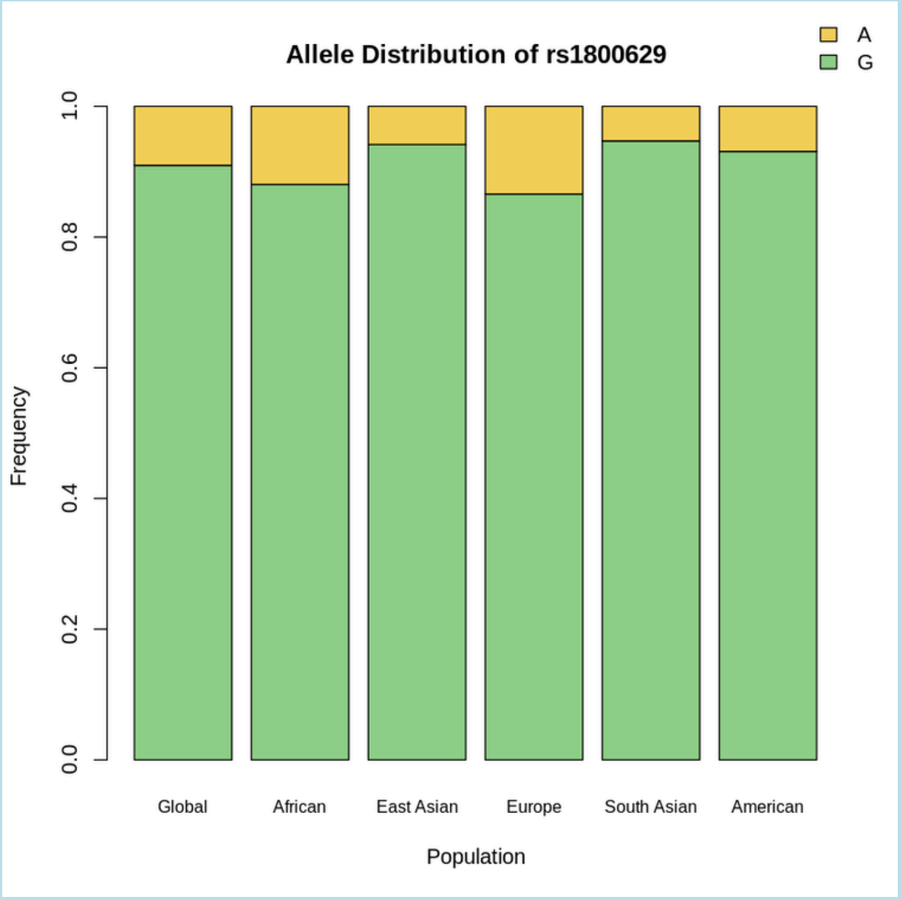
Location: **31575254**

(GRCh38.p14)

<u>Alleles:</u>

Referential Allele: **G**

Alternative Allele: **A**



rs1799724

Chromosome: **6**

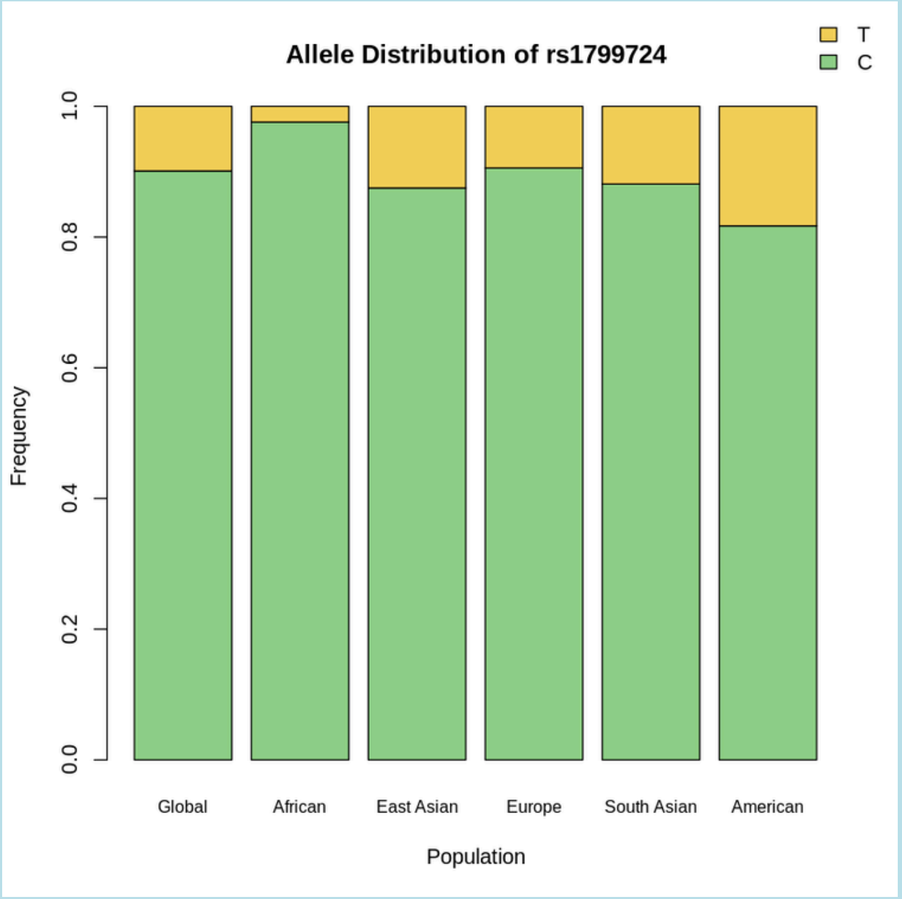
Location: **31574705**

(GRCh38.p14)

<u>Alleles:</u>

Referential Allele: **C**

Alternative Allele: **T**



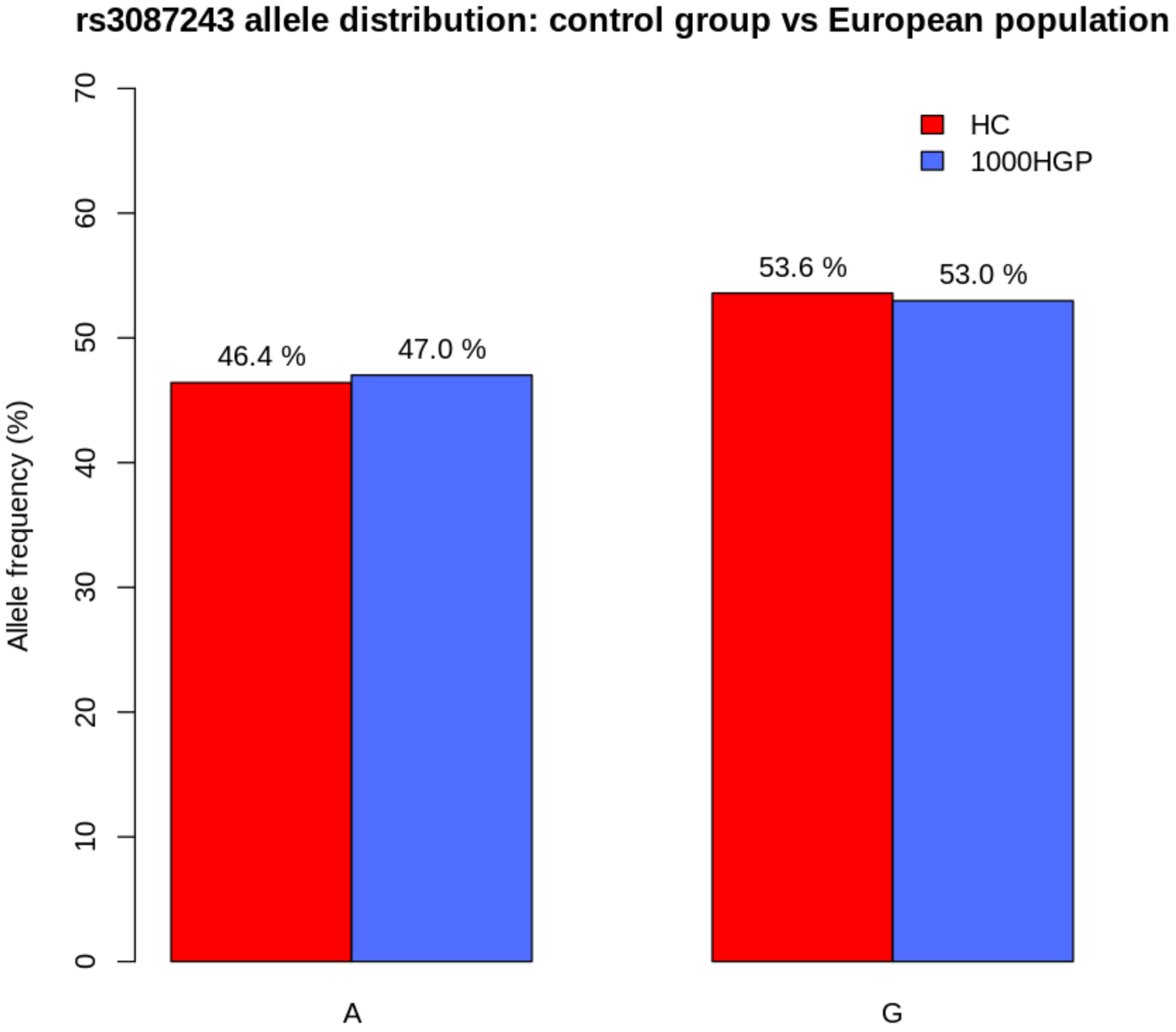
SECTION 3

Comparison of allele distributions in the control group against those of 1000 Genomes populations

Comparison of allele distributions in the control group against those of 1000 Genomes populations

SNP ID / Population	rs2476601	rs3087243	rs3807306	rs1800629	rs1799724
African	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.002037	< 2.2e-16
East Asian	< 2.2e-16	< 2.2e-16	< 2.2e-16	1.358e-15	0.9399
European	0.08891	0.8393	0.6907	0.03347	0.07243
South Asian	1.838e-12	8.033e-11	0.02759	< 2.2e-16	0.8176
American	0.001189	0.0008194	9.968e-12	3.446e-12	0.001214

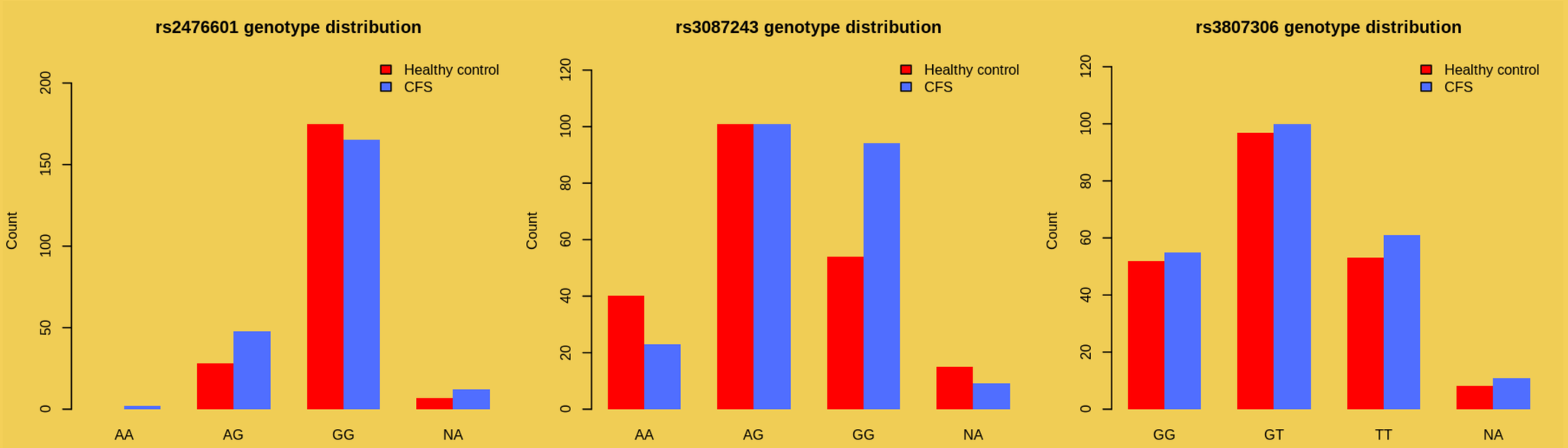
Table: p-values of exact binomial tests for equality of distributions

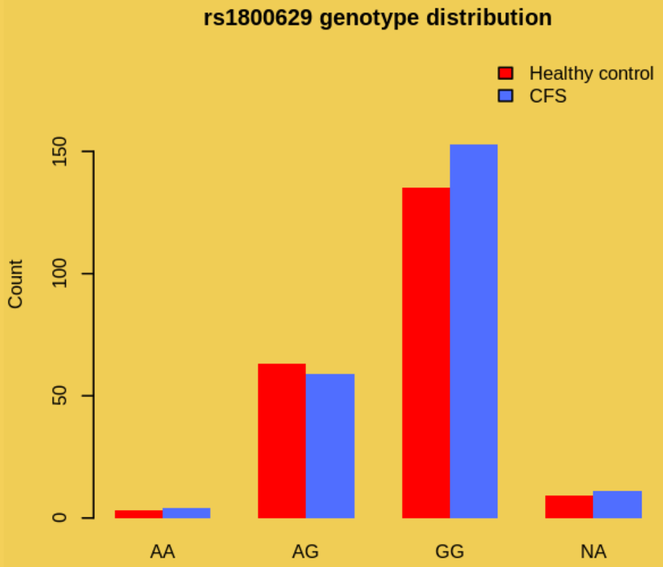

rs3087243 allele distribution: control group vs European population
HC
1000HGP
Allele frequency (%)
0
10
20
30
40
50
60
70
46.4 %
47.0 %
53.6 %
53.0 %
A
G

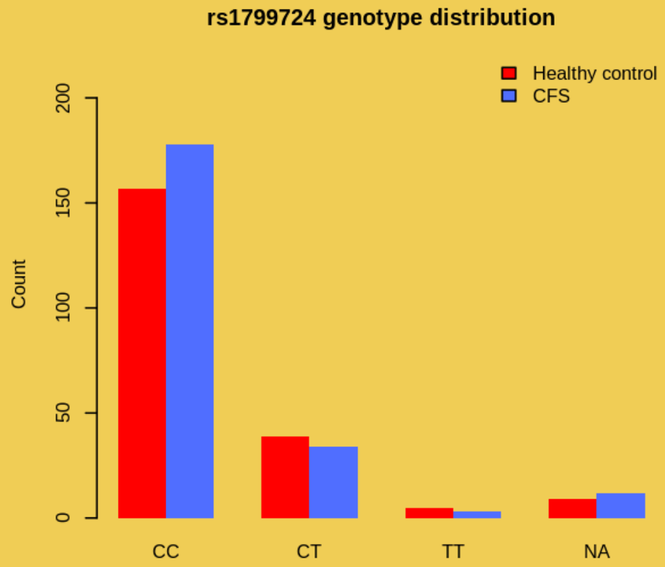
SECTION 4

Genotype distributions in control and case groups. Hardy-Weinberg Equilibrium

Genotype distribution in case and control groups







Hardy-Weinberg Equilibrium

Hardy–Weinberg equilibrium was tested in control and case groups using an exact test

Group \ SNP ID	rs2476601	rs3087243	rs3807306	rs1800629	rs1799724
Control	0.605	0.666	0.575	0.214	0.184
Case	0.747	0.651	0.278	0.794	0.403

Table: p-values of exact tests for Hardy–Weinberg equilibrium

SECTION 5

Comparison of the genotype distribution of CFS and HC for each genetic marker

rs2476601

	AA	AG	GG
HC	0	28	175
CFS	2	48	165

rs3087243

	AA	AG	GG
HC	40	101	54
CFS	23	101	94

rs3807306

	GG	GT	TT
HC	52	97	53
CFS	55	100	61

rs1800629

	AA	AG	GG
HC	3	63	135
CFS	4	59	153

rs1799724

	CC	CT	TT
HC	157	39	5
CFS	178	34	3

Genotype Distribution Analysis

SIGNIFICIANT RESULTS

rs2476601

Chi-squared test:
p-value(adj) = 0.06767
Fisher's test:
p-value(adj) = 0.03723

rs3087243

Chi-squared test:
p-value(adj)= 0.00421
Fisher's test:
p-value(adj) = 0.00404

CHI-SQUARED TEST
FISHER'S TEST
BENJAMINI-HOCHBERG CORRECTION

CONCLUSION

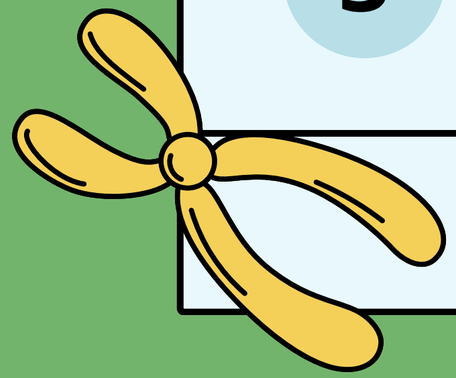
Based on the results ($p_{adj}<0.05$), we reject the null hypothesis of independence. This confirms a significant difference in genotype distribution between CFS patients and the HC.

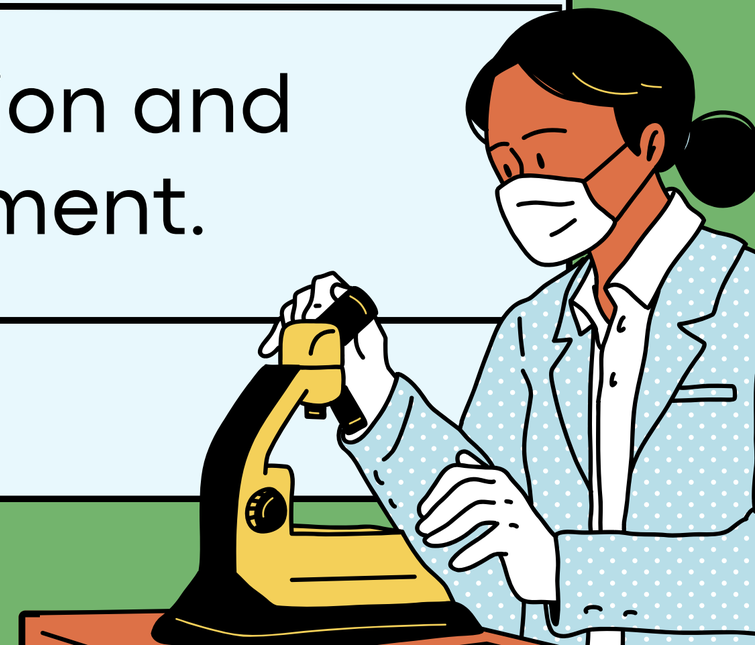
SECTION 6

Best additive models

Methodology

1. Genotype Encoding: Additive mapping (0, 1, 2) of risk alleles.
2. Binary Logistic Regression (GLM) adjusted for gender.
3. Multivariate model construction and predictive accuracy assessment.





FINAL MODEL

Group ~ rs3087243 + rs2476601 + Gender

p-values:
rs3087243 - **0.00361**
rs2476601 - **0.006930**
Gender- 0.09897

CONFUSION MATRIX AND ACCURACY

Actual \ Predicted	0	1
0	**121**	**67**
1	**87**	**119**

Final Model Accuracy:
60.91 %

CONCLUSIONS:

- Both genetic markers (rs3087243 & rs2476601) ***contribute significantly*** to the model, confirming their association with CFS.
- The accuracy demonstrates a verified biological association but limited diagnostic utility as a standalone tool.
- This specific model, gender did not show statistical significance.

SECTION 7

Can these genetic data be used as a diagnostic tool for CFS?

Convention

1

Removing 43 patients with missing data in order to compute AUC and ROC curve

2

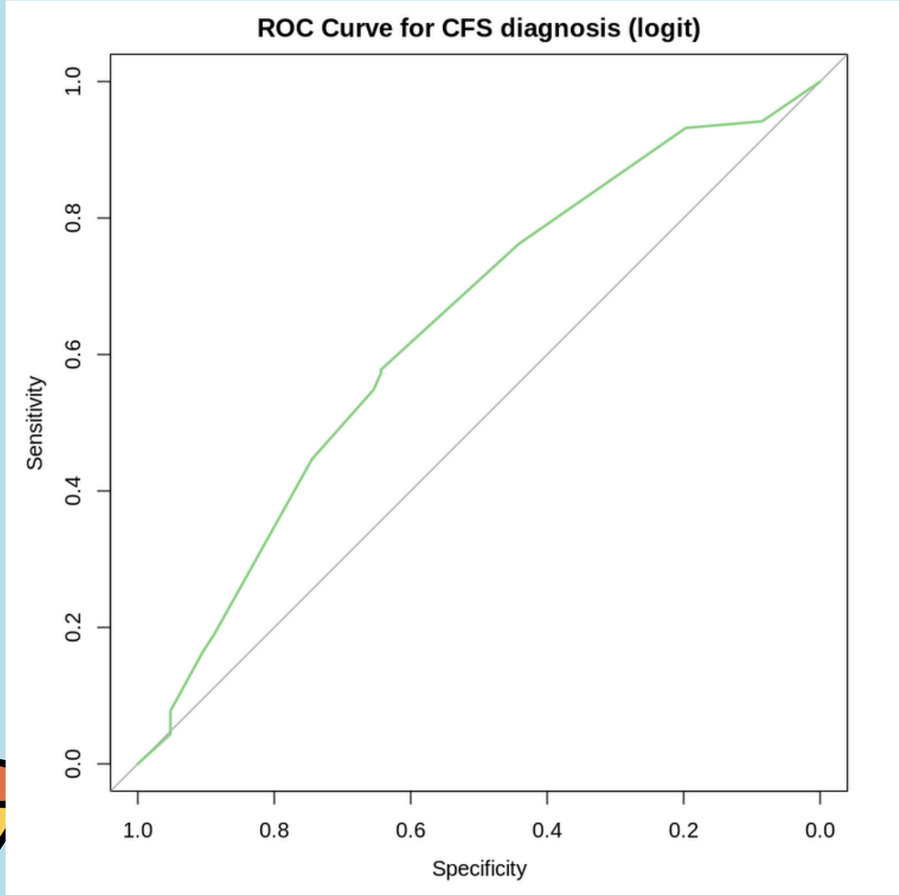
Variables chosen for the best additive model used in modeling

3

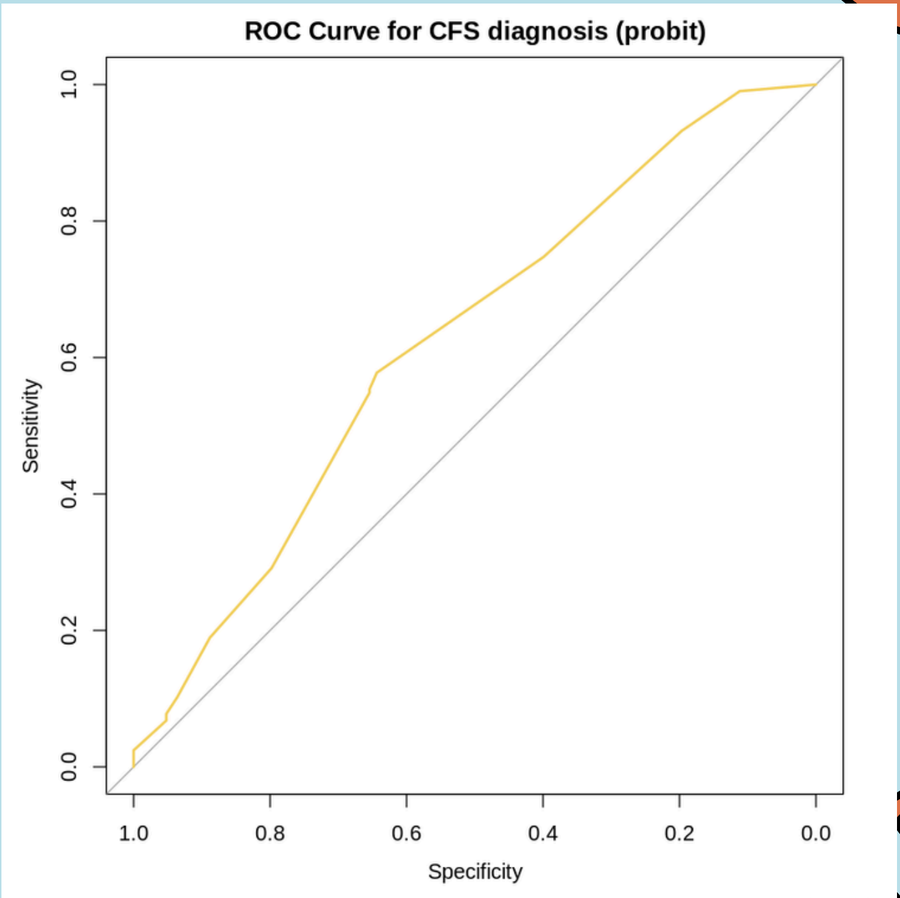
Both logistic and probit regression used as GLMs

AUC & ROC curves

AUC: 0.6363 (0.582, 0.69)



AUC: 0.6235 (0.569, 0.678)



OPTIMAL CUTPOINTS

REGRESSION	CRITERION	CUTOFF	SENSITIVITY	SPECIFICITY
Logistic	ROC01	0.5116	0.5776	0.6436
	SpEqualSe	0.5116	0.5776	0.6436
	MaxEfficiency	0.4631 0.5116	0.7621 0.5776	0.4414 0.6436
Probit	ROC01	0.5085	0.5776	0.6436
	SpEqualSe	0.5085	0.5776	0.6436
	MaxEfficiency	0.5085	0.5776	0.6436