

NS GWAS top SNP might not be real!

Alejandro Ochoa

StatGen, Biostatistics & Bioinformatics — Duke University

2022-08-12 — NS U01 working group

Overview

Background

- ▶ Top GWAS SNP is absurdly significant in NS data
- ▶ But disappears after merging with 1000 Genomes!

Results

- ▶ SNP is truly missing in both 1000 Genomes and gnomAD!
- ▶ Falls in a variable repeat region
 - ▶ Hard to make variant calls there
 - ▶ Location could be mismapped
- ▶ Falls on pseudogene (if location is correct), near IgE receptor genes

Top NS GWAS SNPs (array data only)

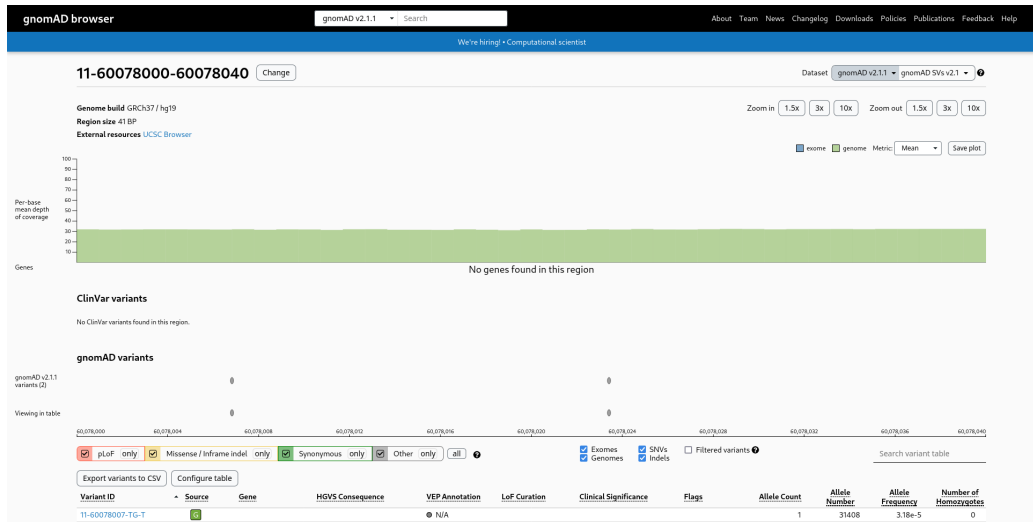
rank	chr	ID	pos	alt	ref	freq	p
1	11	JHU_11.59845492	60078020	A	C	0.0557	6.88e-51
2	8	8:30694465-AG	30836949	G	A	0.0446	2.37e-26
3	6	JHU_6.32629270	32661494	A	C	0.279	1.61e-22
4	10	rs2165031	49727796	T	A	0.166	1.54e-19
5	6	JHU_6.32631342	32663566	G	T	0.282	3.22e-19
6	6	JHU_6.32634466	32666690	T	C	0.282	4.12e-19
7	6	6:32634318-CA	32666541	A	C	0.282	5.25e-19
8	6	JHU_6.32629649	32661873	C	T	0.281	5.40e-19
9	6	JHU_6.32626345	32658569	A	G	0.183	1.69e-17
10	6	JHU_6.32603935	32636159	G	A	0.413	5.93e-17

Top SNP in more detail

Genotype	Controls	Cases
C/A	72	140
C/C	949	742
Missing	28	50

- ▶ A is rare allele (5.6% freq), A/A not observed.
- ▶ One copy of A significantly increases risk of NS
- ▶ Relatively high missingness of 3.9% (passes QC)

SNP missing in gnomAD



SNP missing in 1000 Genomes

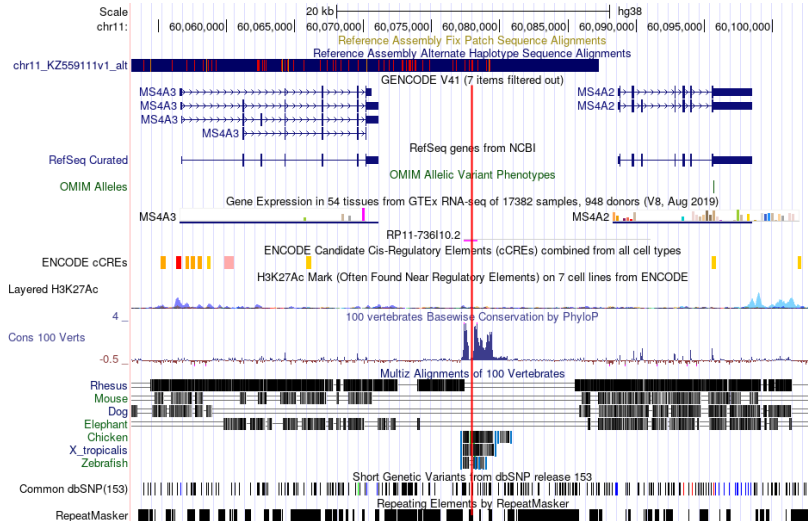
CAA repeat region is highly variable (all PASS):

pos	ref	alts
60078004	TCAACAACAACAACAA	TCAA
"	"	TCAACAA
"	"	TCAACAACAA
"	"	TCAACAACAACAA
"	"	TCAACAACAACAACAACAA
"	"	T
60078016	AC	A,*
60078018	AAC	A,*

Genome browser



Genome browser



Nearby genes

Perhaps fishing, given dubious evidence of SNP existence or accuracy of location:

- ▶ RP11-736I10.2: Expressed pseudogene, no other info.
- ▶ MS4A2: IgE Receptor Beta Subunit. Diseases associated include: Ige Responsiveness, Atopic and Allergic Asthma.
- ▶ MS4A3: CD20 Antigen-Like Protein IgE Receptor Beta Subunit

Next steps

- ▶ Get array probe info, try to find alternative mapping locations
- ▶ Can genotypes be validated experimentally?
 - ▶ On a subset of samples, by Sanger sequencing?