Genetics of NAPS2

Brandon Michael Blobner PhD

For the North American Pancreatic Study Group

Purpose

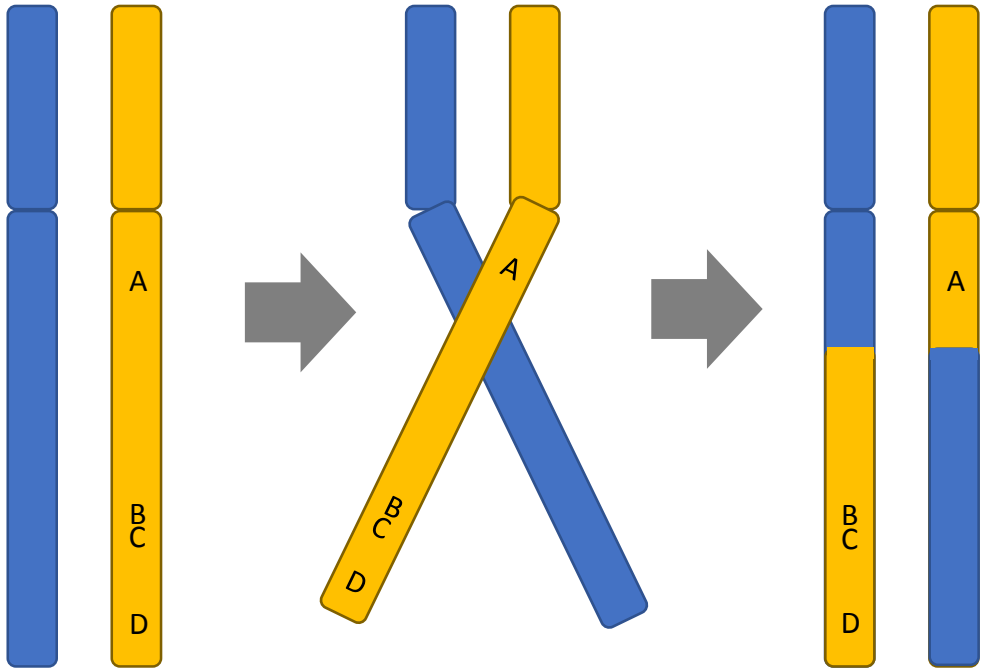
- Review the fundamentals of the Genome-wide Association Study
- Review previous NAPS2 analysis and findings
- Present new NAPS2 analyses and findings
- Provide context for the discussion on future pancreatitis analyses

Definitions

- GWAS – Genome-wide association study
- SNP – Single nucleotide polymorphism
- Locus – a position on a chromosome
- PCA – Principal components of ancestry
- GRM – Genetic relatedness matrix

Linkage Disequilibrium (LD)

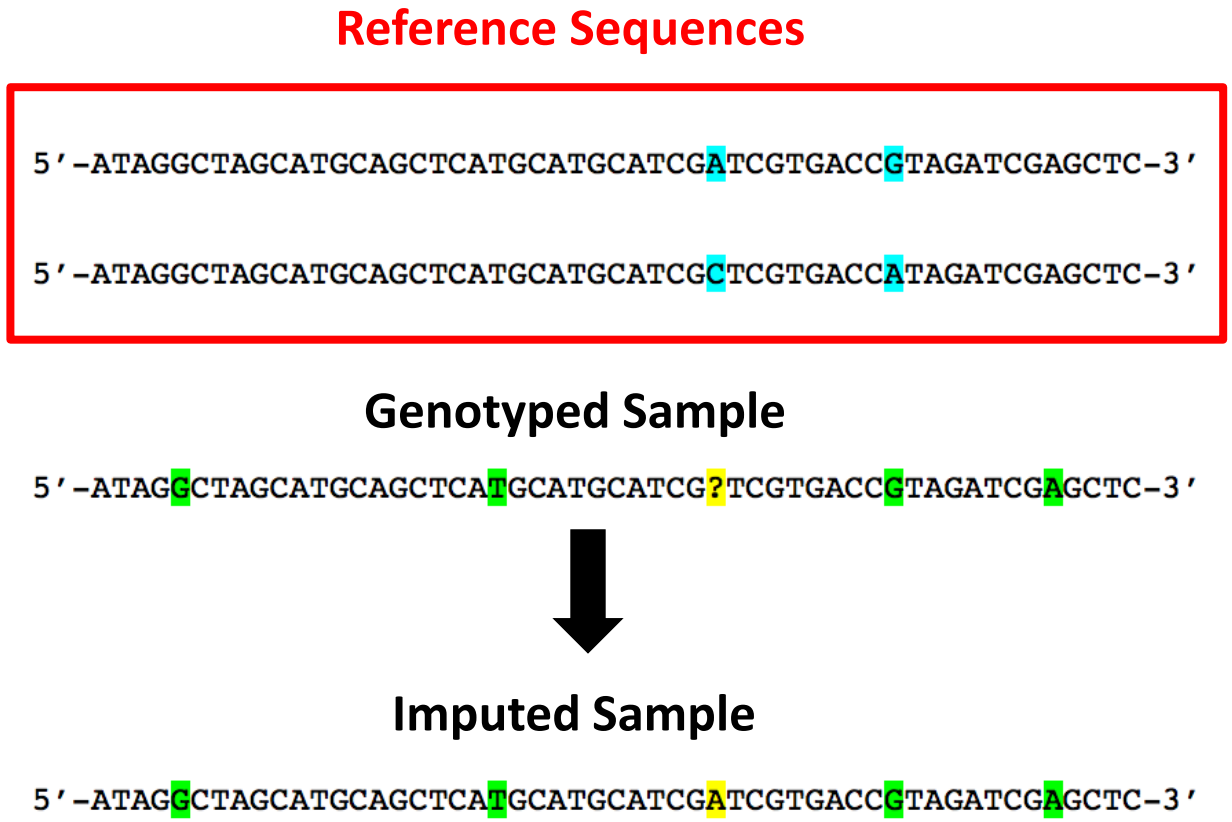
- SNPs that are in close proximity to one another on a chromosome (<100 kb) tend to be inherited together due to the decreased likelihood of a recombination event occurring between them during meiosis
- SNPs that are inherited together at a rate greater than what is expected under HWE are in LD



Over many, many recombination events SNPs B and C will be separated fewer times than SNPs A and B or SNPs C and D.

Genotyping and Imputation

- Genotyping
  - An individual's SNPs are directly measured
  - A single whole-genome genotyping array can genotype up to ~2.38 million SNPs
- Imputation
  - An individual's non-genotyped SNPs are inferred based on genotyped SNPs in that haplotype
  - Is used to expand the number of SNPs included in a GWAS by millions of SNPs without the need for additional genotyping arrays



GWAS

- 100s of thousands or millions of association tests for 100s of thousands or millions of SNPs distributed across the genome with a phenotype
  - Each SNP is analyzed independently of all other SNPs
  - Penalty for multiple testing / risk of false discovery
- Hypothesis generating analysis that identifies SNPs that are *associated* with a phenotype, but may not contribute to the phenotype. CORRELATION ≠ CAUSATION
  - Identifies regions of the genome that warrant further study to determine causality
- Results are dependent on the population studied and the definition of “case” and “control”.

An Analogy

Image from Google Earth

- You are looking to buy a house somewhere in the United States
- Your criteria for the area that you want to buy in are:
  - Good local schools
  - Walkable area
  - Close to a major city
- So you search for areas matching your criteria

An Analogy

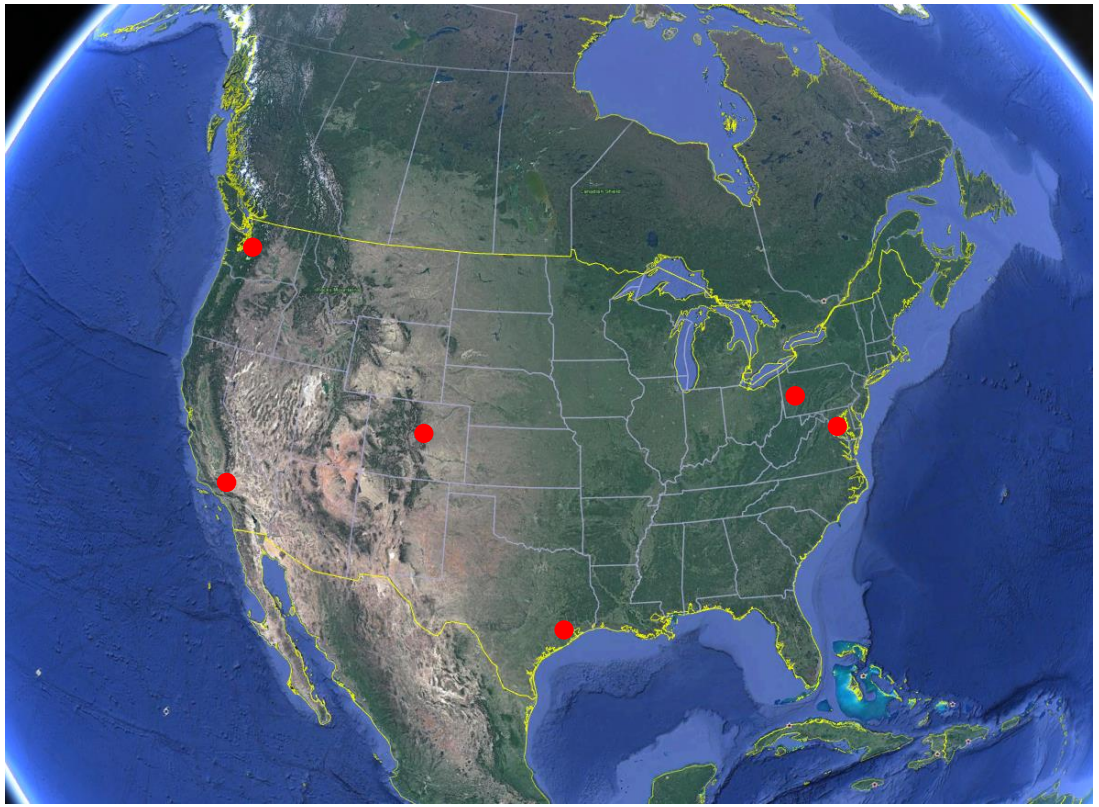
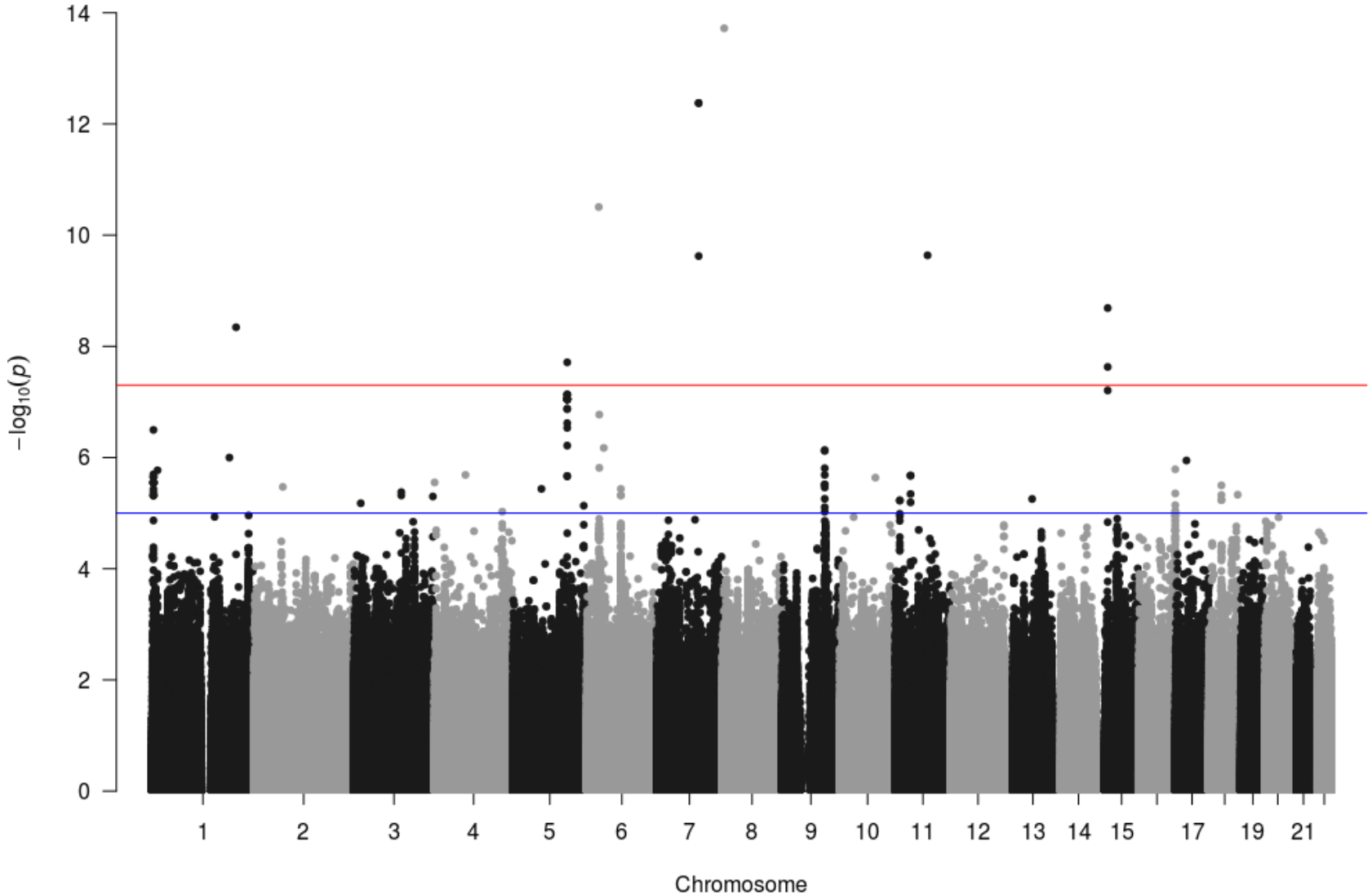


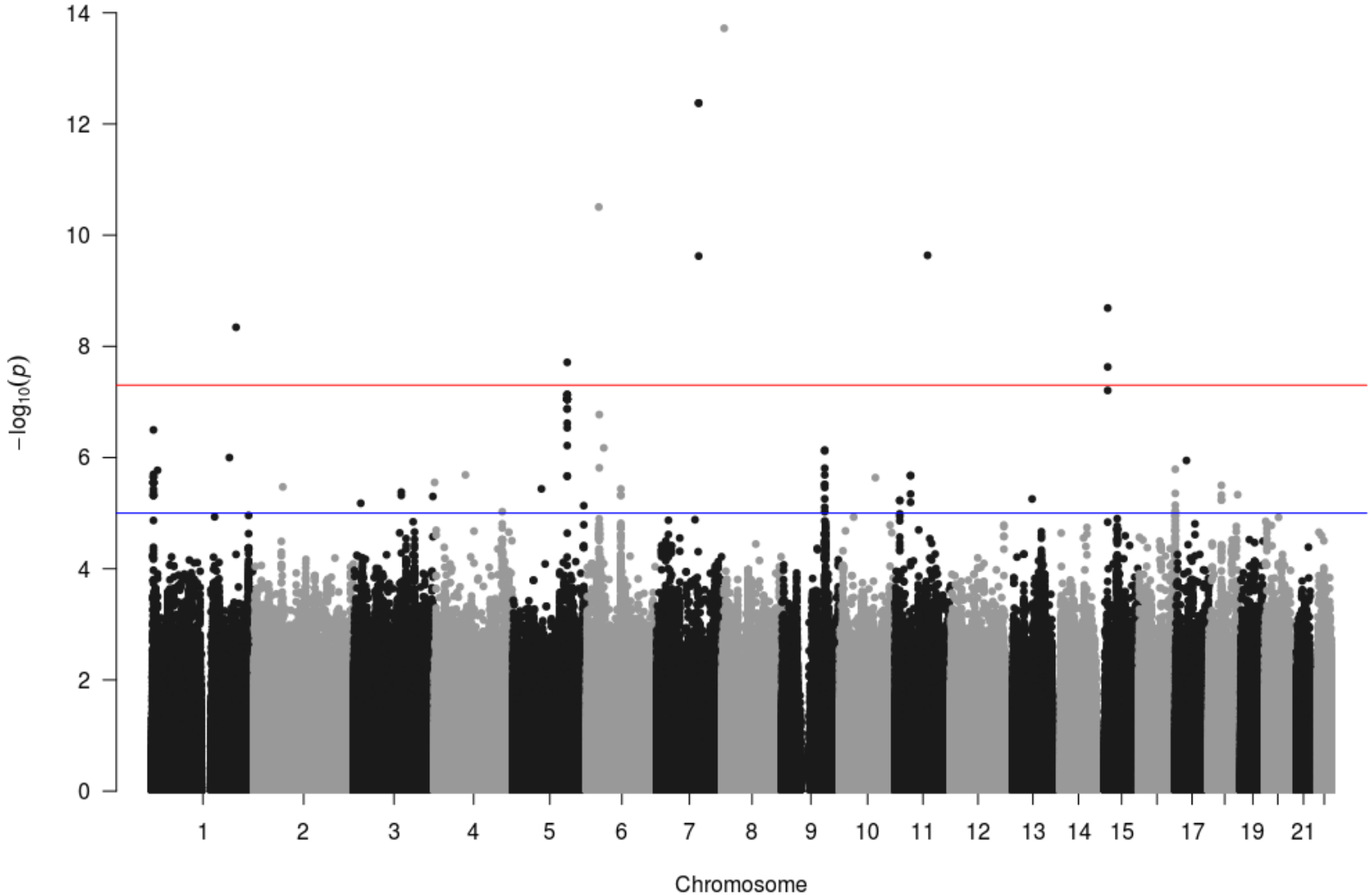
Image from Google Earth

- You are looking to buy a house somewhere in the United States
- Your criteria for the area that you want to buy in are:
  - Good local schools
  - Walkable area
  - Close to a major city

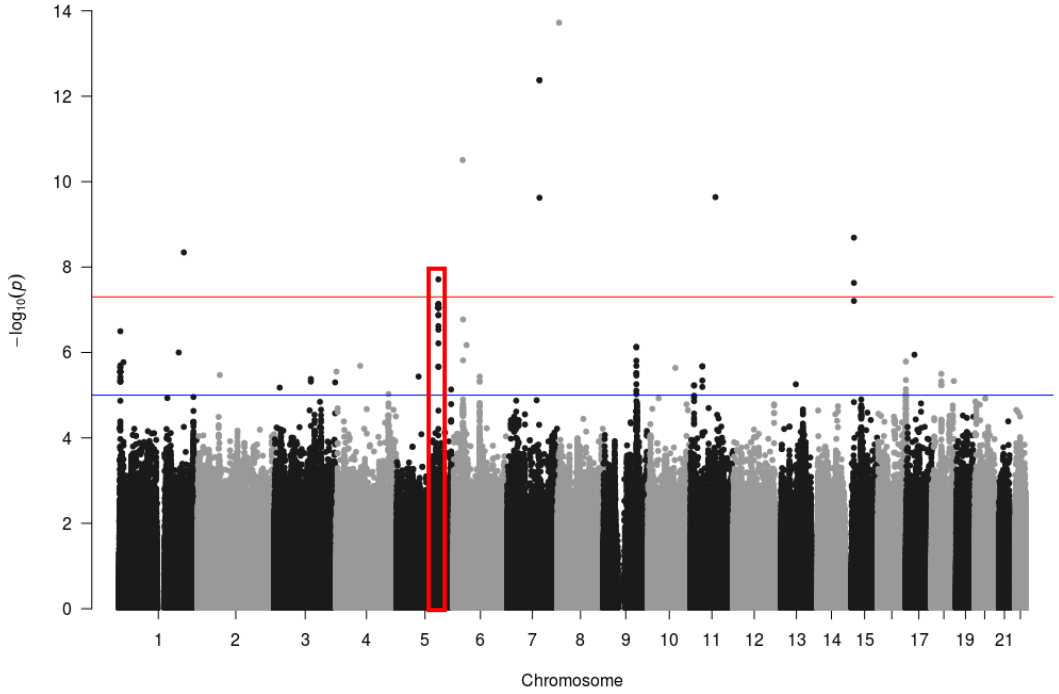

14
12
10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
Chromosome

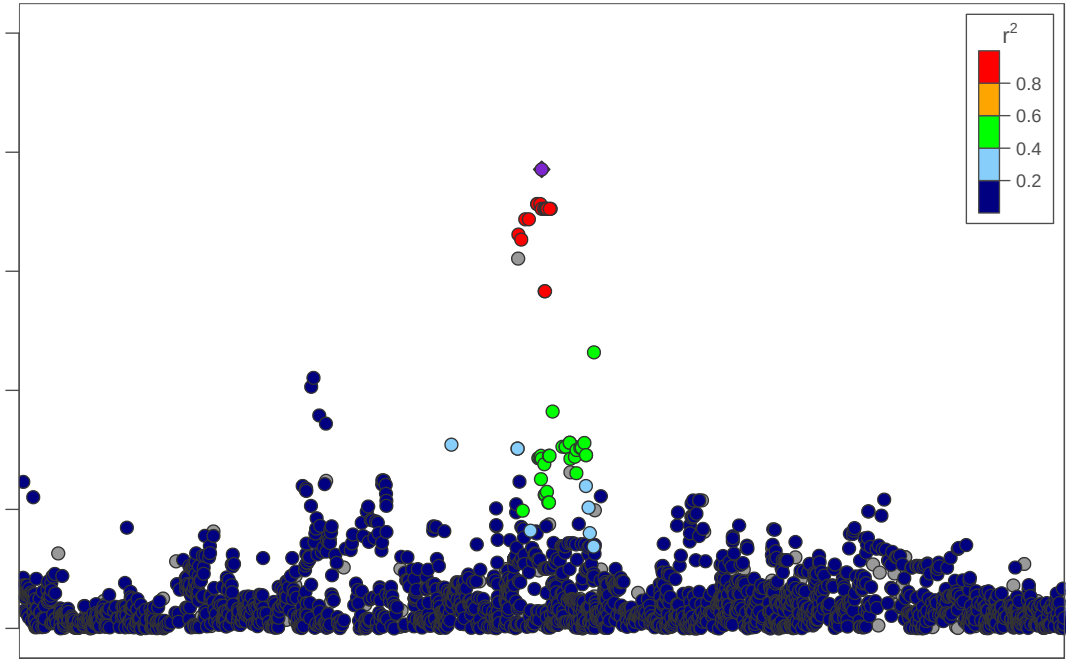
Your search (GWAS) gives you a neighborhood (locus), but not a specific house (SNP)

Images from Google Earth


14
12
10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
Chromosome

Locus Zoom Plot





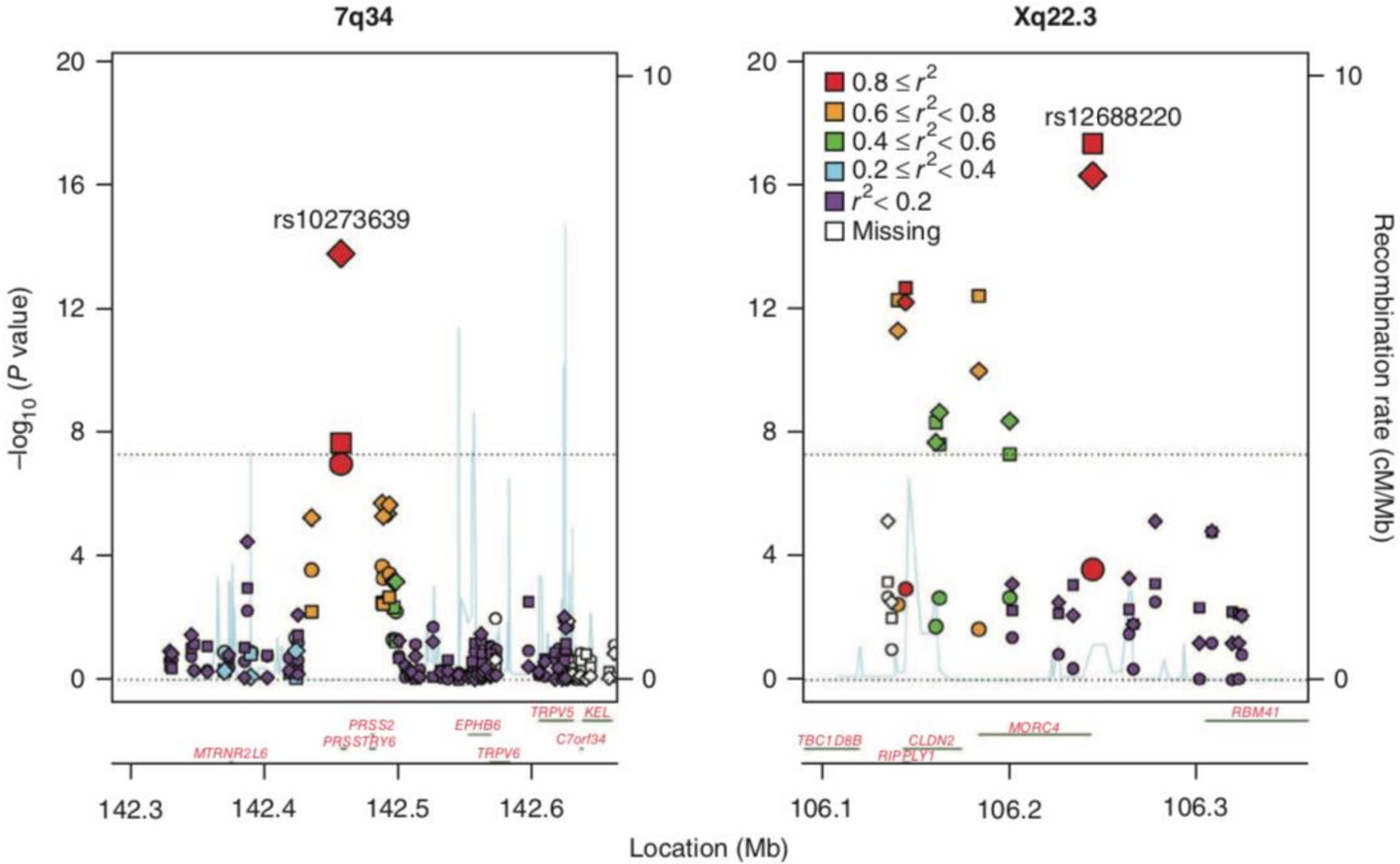
NAPS2

- North American Pancreatitis Study 2
- Designed to advance knowledge of genetic, environmental, and metabolic factors that contribute to pancreatitis
- Recurrent acute pancreatitis
  - Two or more episodes of acute pancreatitis
  - No imaging evidence of CP
- Chronic Pancreatitis
  - Imaging evidence of CP by CT or ERCP
  - Histology evidence of CP by MRCP or EUS

- Study centers
  - Brigham and Women's Hospital
  - Dartmouth-Hitchcock
  - Indiana University
  - Medical University of South Carolina
  - Mayo Clinic Jacksonville
  - University of Michigan
  - University of Pittsburgh
  - St. Louis University
  - Aurora Healthcare
  - University of Alabama at Birmingham
  - University of Florida
  - Griffin Hospital – Yale Affiliate
  - Virginia Commonwealth University

Previous NAPS2 Analyses

- Whitcomb et al. 2012
  - 625,739 genotyped SNPs from
    - Stage 1: 676 cases and 4,507 controls
    - Stage 2: 910 cases and 4,177 controls



Whitcomb et al. 2012

Genetics of Pancreatitis

- 9 studies have reported a total of 180 SNPs from 7 genome-wide significant loci ($p<5e^{-8}$) and a number of additional suggestive significant loci ($p<1e^{-5}$) for pancreatitis
- Variants in CFTR, PRSS1/PRSS2, SPINK1, CLDN2, CTRC, etc. have been previously associated with pancreatitis
- Much of the heritability of pancreatitis remains unexplained

Genetics of Pancreatitis

- 9 studies have reported a total of 180 SNPs from 7 genome-wide significant loci ($p<5e^{-8}$) and a number of additional suggestive significant loci ($p<1e^{-5}$) for pancreatitis
- Variants in CFTR, PRSS1/PRSS2, SPINK1, CLDN2, CTRC, etc. have been previously associated with pancreatitis
- Much of the heritability of pancreatitis remains unexplained

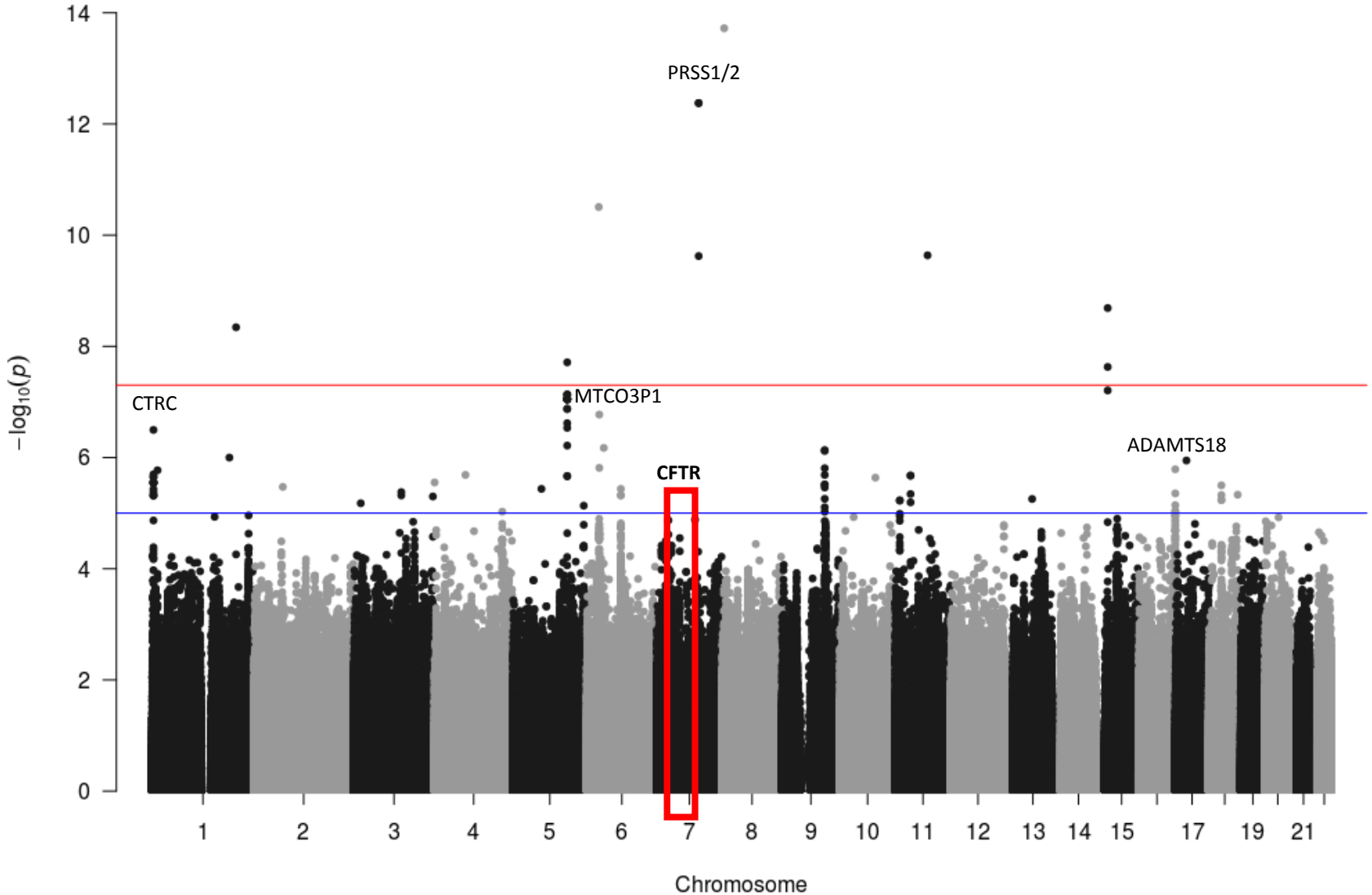
Can we identify additional pancreatitis-associated loci in NAPS2?

Current Pancreatitis GWAS

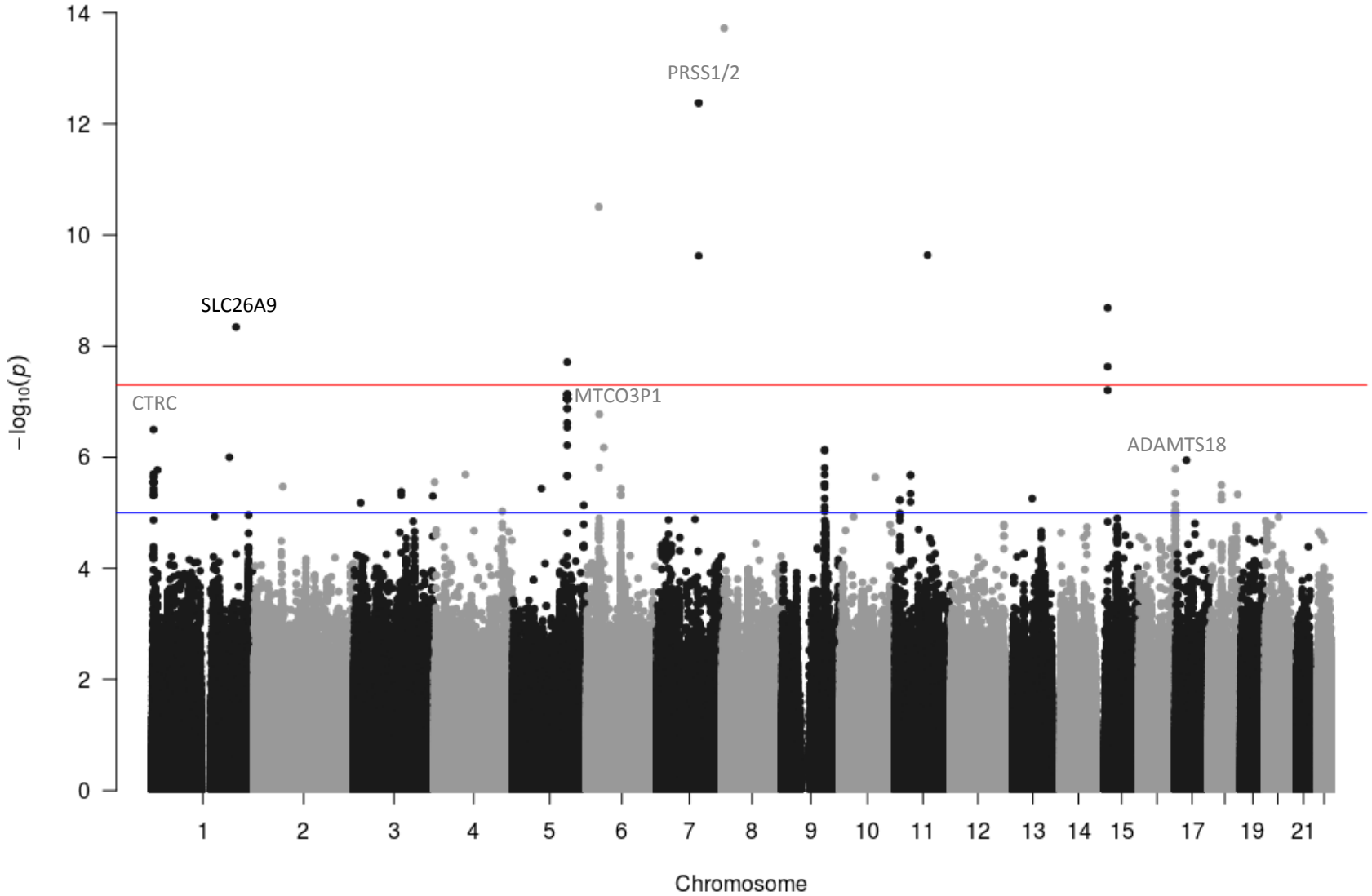
- 1492 cases of recurrent acute or chronic pancreatitis
- 869 controls
- Population
  - Mean age 49.4±15
  - 47% Male
  - 76% European ancestry
  - 57% Smokers
  - 18% “Very Heavy” alcohol consumption

- 9,838,266 SNPs
- MAF ≥ 0.01
- Covariates
  - Age
  - Sex
  - BMI
  - Alcohol consumption
  - Smoking
  - Principal Components of Ancestry
  - Genetic Relatedness Matrix

Pancreatitis vs Controls

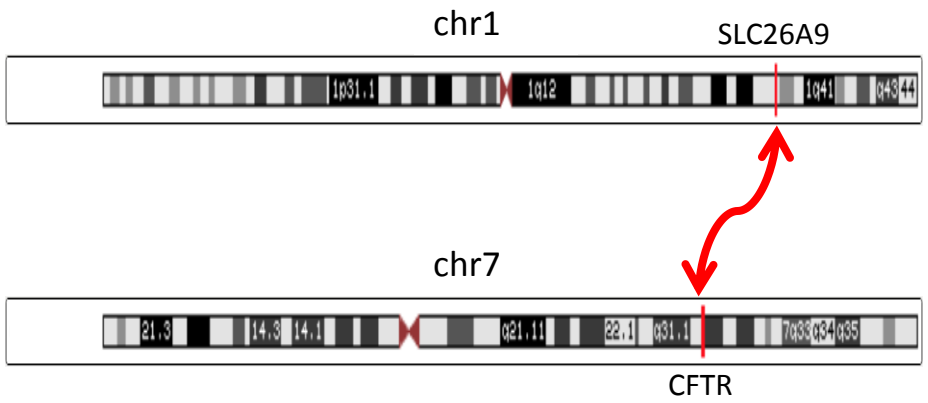


Pancreatitis vs Controls



Epistasis analysis

- Epistasis – when the genotype at one locus affects the expression of the phenotype of another locus
- 623 CFTR snps present in the NAPS2 cohort were analyzed for epistasis with the sentinel snp of the SLC26A9 locus, rs7366689 (MAF = 19.2%)
- CFTR snp rs17547853 is in epistasis with rs7366689, $p < 0.0001$



rs7366689 interacts with smoking in pancreatitis

Cases

	TT	TC	CC	
Smoker	644	160	18	822
Non Smoker	121	39	10	170
	765	199	28	992

Controls

	TT	TC	CC	
Smoker	261	70	55	386
Non Smoker	47	22	56	125
	308	92	111	511

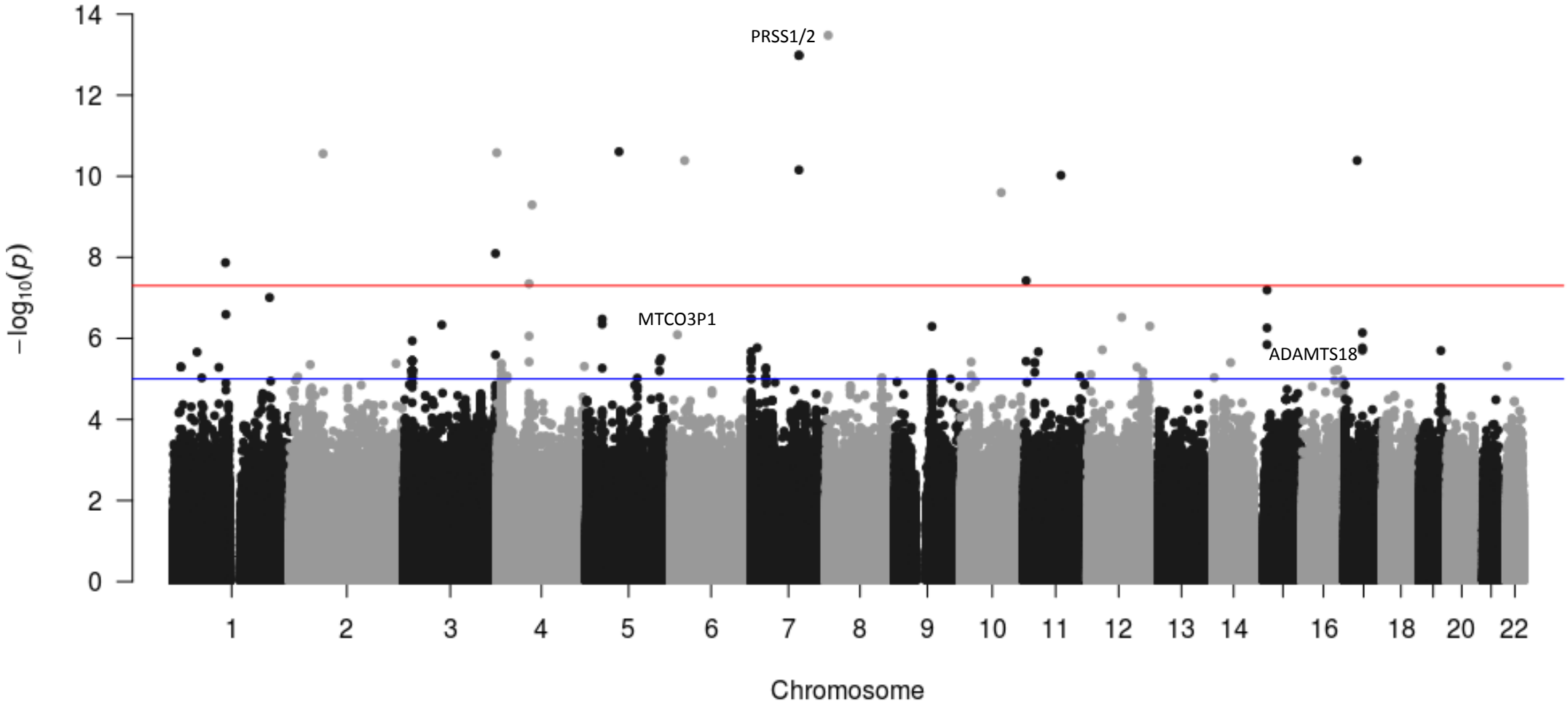
- T allele occurs more frequently in individuals with pancreatitis that smoke *p*-value = 0.01
- Formula = (pancreatitis ~ genotype*smoking)
- Genotype*smoking *p*-value = 0.001

Research Question:
Are loci associated with RAP in the NAPS2 cohort?

RAP GWAS mega analysis

- 836 cases of recurrent acute pancreatitis
- 1296 controls
- 9,679,811 SNPs
- MAF ≥ 0.01
- Covariates
  - Age
  - Sex
  - Principal Components of Ancestry
  - Genetic Relatedness Matrix

RAP vs Controls



Conclusions

- GWAS results depend on the precise phenotype being evaluated and appropriate controls
- AP, RAP and CP are complex, and analysis of risk within this broad phenotype is needed for fibrosis, acinar dysfunction, diabetes and cancer are needed.
- Early GWAS chips and populations focused on populations of European ancestry: New global SNP representation is needed.
- SNP analysis is limited. Future studies should consider:
  - Candidate gene analysis (e.g. CFTR)
  - Whole genome sequencing
  - Direct evidence of which variant results in altered cell function
  - Integration of cell dysfunction with disease detection and management.