

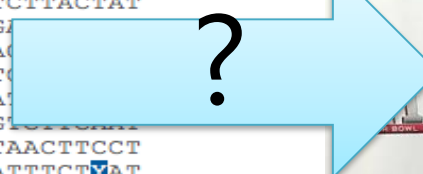
# Genetic Variation and Transcriptional Regulation

**BIOM262 - Quantitative Methods in  
Genetics and Genomics**

Graham McVicker – Salk Institute

# Which genetic variants affect human traits?

ATCCTTTCTAGAGAAAAATAGAAAAATAGAACTTTGAGGTTGAATCTCTTTTAAATGTAATG  
TTTTCTCGAATCCAAGTGTTTTTACACTATACAATAGGAGTAGAAAAATTTGTCACCACCT  
TGTGGCCAAA<sup>Y</sup>TCACCTTTTTCTTTCTTTTTTTATTTT<sup>Y</sup>ACATTAAAAAATTTTACTT  
TAAGTTCCAGGATACATGTGCAGGATGTGCAGGTTTGTACATAGGTAAATGTTTTATTT  
TAAATTTAATTTAACACTTTTTAT<sup>Y</sup>TTTAAAGTCATACAACCTCTCATAGCCAGTAGTTAAT  
ATTACCTTG<sup>M</sup>AAGTTTGGTATGGTTGATGAATTGCATCCTGTTAATAATTGCTACAGATT  
TTTGAATAAATTGCAGACCAGTTTGATG<sup>R</sup>TCCTGGGTGGGCATAAGTACATGAAGATTTAC  
TTTTCTCTGTGAGCTTTCTTGGGATGAAGAAATTTAGT<sup>G</sup>TTTTTTTTTAAATTTTAAAGAA  
AT<sup>W</sup>TTTATTATTTTTTACATGATTTATTTCCCACTGAAAAATAAATCCCACCGGG<sup>Y</sup>ATAA  
AGTGT<sup>A</sup>TTTTTTTTAAGTCACAGAGTAACCCAACCTGAAGCTAGTTTTTCAGACCTTAGGCA  
GTTTCATGCTGTAAGCCCGAGATCTCATGGTCACCCTTGCAAGAGAAATATCTAATTGAAA  
AAAAATATGAAGAGTATTAATTTTG<sup>W</sup>TAGTGCTAAAATGACATAAAGGGATCTCACTGGG  
CTTGAGATATTAAGCATTTAAATTTGTTA<sup>R</sup>AGGTTTAAATTTGTTAGTAACTTGTATTGCA  
TAGAAAAATGTGCCAAATGTGTCAGTAAATAAAAAAACTTTTTT<sup>T</sup>AAAAATAAAAAATTTACAGA  
<sup>A</sup>RAATTATGAC<sup>R</sup>ATACTACAAAGAGGTTCTGTACAACCCCCCTCCCAGTTTCTCTTACTAT  
TAACATCTTAAATTAGTATGTTACATTTGTGACAATTAGTGAACCAATATTG  
GTAC<sup>K</sup>AACATAAGTCAGTGTTCTTTTTACTGGAGAAATGGTGTAGAAACTAA  
CACTGTGGTATGGTGGTTGCTTCTCCACCACCTACTATCTGCTTACCTAATTTCT<sup>Y</sup>AT  
AGAGCAAAGAAATATATGTGTGTATATTAACCTATGTGTACACATACATCTA  
TCT<sup>R</sup>ATATGTAACAT<sup>S</sup>TGTATCTTTATTAAGCTAAATATGAGTTCATATGGTG  
TCTAATCAATTACTGTATAGATTATTTCTAGCCTCTTCTCTTGCTTATCTGTAACCTCCT  
ATTTCAAACCGTGAAAAATCTGTCTCCACCACCTACTATCTGCTTACCTAATTTCT<sup>Y</sup>AT  
TTCCAGTTTA<sup>K</sup>GTATACAGTGGCTTCCAGAAATTATTACATATA<sup>R</sup>CCCTGTGGGATACAAC  
TTGTCAACTAGAGTGGTGCTTATGTAAGTTCTTCTATCTTTAGTTTTACTGACTCTACTC  
ATTTTCAAAGTTGCTTAGTCCAGAACATTTTCACTCATACTCCTCTAGTGAAGTTGTTTC  
ATA<sup>Y</sup>GTTAGTAACACAGATTCTTTTT<sup>Y</sup>TGCAGCTGCAATCCATTTAGGGTTCCCTCCT  
CTCCAATCTCCTAAATATTATTTTTTAAATTCATATACATCAAGGTTTATTTCTTGTGC  
TGTAAR<sup>G</sup>TTCTATAGGTTTTGACAAATACAAAGTGTC<sup>R</sup>TGTACCCATCATTACAATGTCA  
TAC<sup>R</sup>GAAATCGTTTTCACTGC<sup>Y</sup>CTAAAAATATCCCTTGTCCTTTGCCTATTCAACC<sup>CT</sup>TCCC  
CTCCTTTCC<sup>CA</sup>AACTCCTGGCAACCACTGATCTGTTATCGTGGAGCTGTGTCTCTTCCA  
GAATGCATATAAATTGAATCATACAATATGTAGACTTTTCAACCCTGGCTTATTTTGTAG  
CAATATGCATTTAACATTTCATCCATGTCCTTATGTGGCTTGTAGTTTCACTTTTTTACT  
GCTGGGTAGTATT<sup>CTR</sup>TCATAGAAATGT<sup>R</sup>CCACA<sup>R</sup>TTTGTTTATCCATT<sup>Y</sup>GCTGATTGAA  
GTATATCAATATACTTGGAACTGACTGCTAGATAG<sup>Y</sup>ATAGTAAGACTATATTTA<sup>K</sup>CTT  
TGCAAGAAACTGCCAACTGTATTTTTAAAGTGGCTGTACCATTGTGCCACCAGCAACTCC  
TGCCAGTGATCCAGTATTGTGAGTTTTTTGGATTTTA<sup>K</sup>CCATTCTAAAAGGTGAGTGATG  
GTATCTCATTGTGCTTTTAAATTTGTAATACTCTAATGACAAATGATGGTGGATTCTTTTT  
CATATG<sup>Y</sup>TTGTTTTCCCATTTGTATATCTTTTAGTATGTGTCTGTTCCGGATGTTTTGTCT  
TACTTTTTTAAACTGGGTGATTGTTTTCTTTTTCTTTTTCTTTTTTTTCTTTTGGAG<sup>Y</sup>G  
GAGTCT<sup>Y</sup>GCTCTTTAGCCAGGCTGGAGTGCAGTGGCGCCATCTCGGCTCACTGCAAGC<sup>Y</sup>C  
TGCCCTCCGGGTTCAAGTGATTTTCGTACCTCAGCCTCCCGAGTAGCTGGGACTACAGGC  
GCCCCGCCACCACACCTGGCTAATTTTTTTGTATTTTTGGTGGAGACGAGGTTTCACCATG  
TCGGTCAGGCTGGTCTTAA<sup>W</sup>CTCCTGACCATAGATGATCTGCCTGTCTTGGCCTCCCAA



Alzheimer's Disease

Ankylosing Spondyliti

Multiple Sclerosis

Leukemia

Psoriasis

Breast Cancer

Schizophrenia

Rheumatoid Arthritis

Coronary Heart Disease

Celiac Disease

Autism

Crohn's Disease

Parkinson's Disease

Type I Diabetes

Systemic Lupus Erythematosus

ATCCTTTTCCTAGAGAAAATAGAAAATAGAACTTTGAGGTTGAATCTCTTTTAATGTAATG  
TTTTTCTCGAATCCAAGTGTTTTTACACTATACAATAGGAGTAGAAATTTGTCACCACTC  
TGTGGCCAAAYTCACTTTTTTCTTTCTTTTTTTTATTTTYACATTAAAAAAAATTTTACTT  
TAAGTTCCAGGATACATGTGCAGGATGTGCAGGTTTGTACATAGGTAAATGTTTTATTT  
TAAATTTAATTTAACACTTTTTATYTTTAAGTCATACAACCTCTCATAGCCAGTAGTTAAT  
ATTACCTTGMAAGTTTGGTATGGTTGATGAATTGCATCCTGTTAATAATTGCTACAGATT  
TTTGAATAATTGCAGACCAGTTTGATGRTCCTGGGTTGGCATAAGTACATGAAGATTTAC  
TTTTTCCTGTGAGCTTTCTTGGGATGAAGAAATTTAGTGTTTTTTTTTTAATTTTTTAAGAA  
ATWTTTTATTATTTTTTTTACATGATTTATTTCCCACTGAAAAATAAATCCCACCGGGYATAA  
AGTGTATTTTTTTTAAGTCACAGAGTAACCCAACCTTGAAGCTAGTTTTTTCAGACTTAGGCA  
GTTTCATGCTGTAAGCCCGAGATCTCATGGTCACCCTTGCAAGAGAAATATCTAATTGAAA  
AAAAATATGAAGAGTATTAATTTTGWTAGTGCTAAAATGACATAAAGGGATCTCACTGGG  
CTTGAGATATTAAGTATTAAAATTGTTARAGGTTTAAATTGTTAGTAACCTTGTTATTGCA  
TAGAAAATGTGCCAAATGTCAGTAAATAAAAAAACTTTTTTTTATAAAATAAAAATTTACAGA  
ARAATTATGACRATACTACAAAGAGGTTCTGTACAACCCCTCCCAGTTTCTCTTACTAT  
TAACATCTTAAATTAGTATGTTACATTTGTCACAATTAGTGAACCAATATTGATACATTA  
GTACKAACTAAAGTCAGTGTTTCCTTTTTACTGGAGAATGGTGTTAGAACTAAGGTCTGGG  
CACTGTGGTATGGTGGTTGCTATTGAGATGTYGTTATTTTTTAGGTTCTTTCTCAGCTGAC  
AGAGCAAAGAAATATATGTGTGTATATTAACCTATGTGTACACATACATCTATGATTATT  
TCRATATGTAACATSSTGTATCTTTATTAAGCTAAATATGAGTTCATATGGTGTCTTCAAT  
TCTAATCAATTACTGTATAGATTATTCTAGCCTCTTCCTCTTGCTTATCTGTAACCTTCCT  
ATTTCAAACCGTGAAAAATCTGTCTTCCACCACCTACTATCTGCTTACCTAATTTCTYAT  
TTCCAGTTTAKGTATACAGTGGCTTCAGAATTATTACATATARCCCTGTGGGATACAACCT  
TTGTCAACTAGAGTGGTGCTTATGTAAGTTCTTCTATCTTTAGTTTTTACTGACTCTACTC

# Topics

- Genome-wide association studies
- Molecular quantitative trait loci (QTLs)
- Gene expression QTLs
- Chromatin QTLs
- DNA methylation QTLs
- Intersection of molecular QTLs and GWAS

# Testing for genetic association



# Case-Control Association Study

Cases



Controls



AA

AG

GG

Genotype

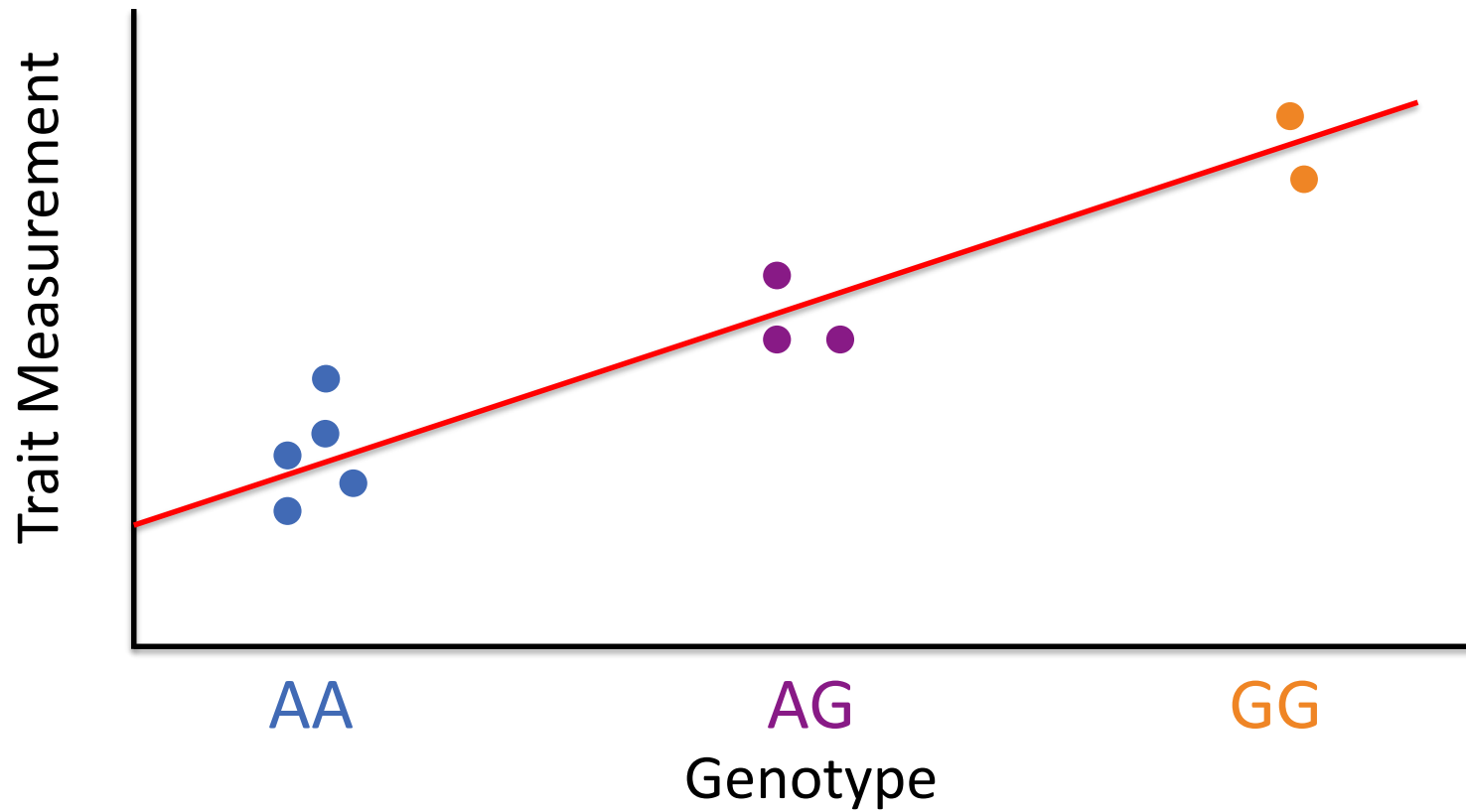
# Case-Control Association Study

	# A alleles	# B alleles
Cases	750	250
Controls	800	200

$\chi^2$  p-value = 0.009



# Quantitative Trait Association Study



# Discussion

- What are examples of human traits?
  - Quantitative Traits?
  - Binary (Case/Control) Traits?

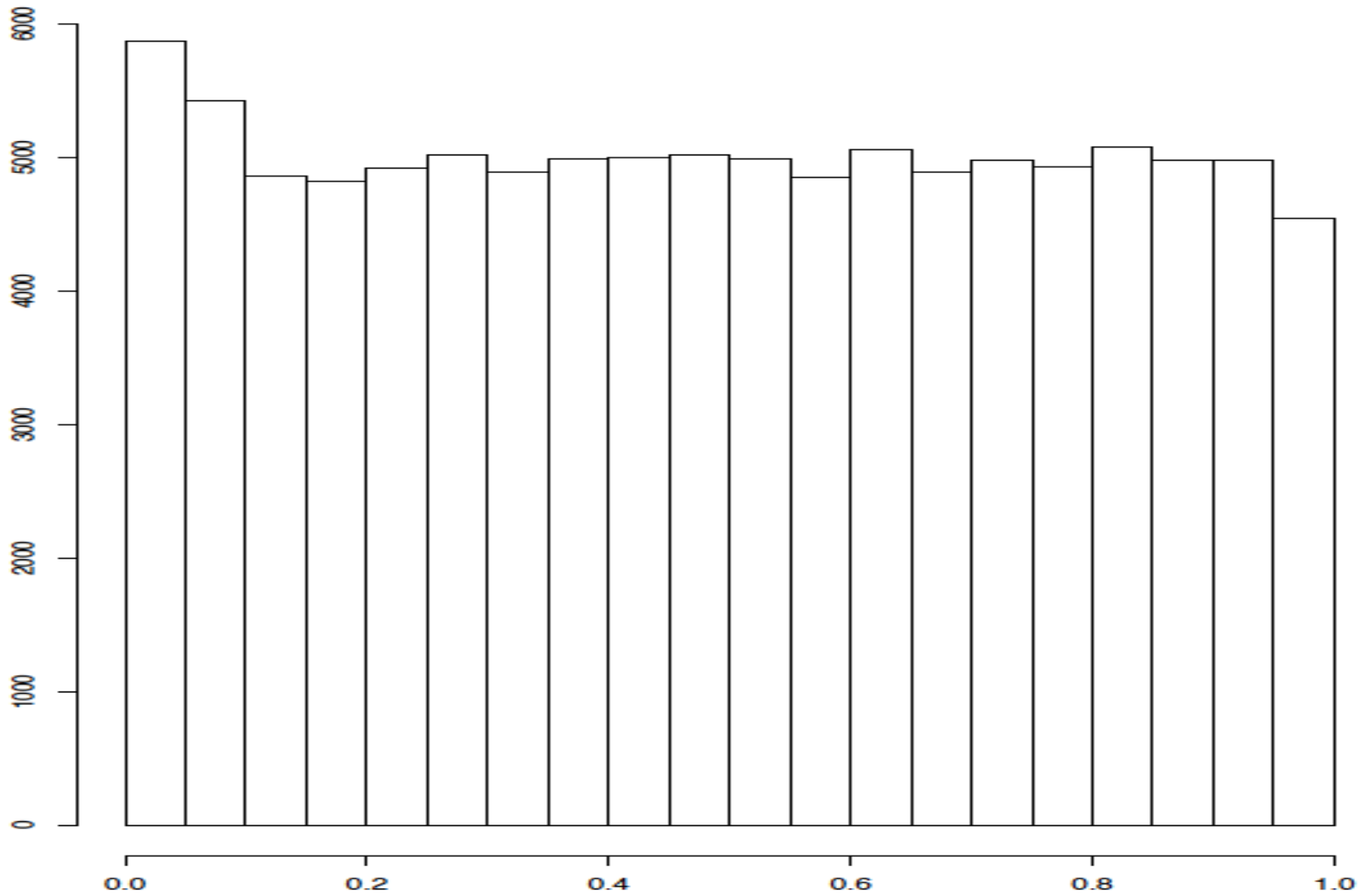
# In class exercise

- GWAS study results from Rheumatoid Arthritis (Okada et al. 2014)
- Meta-analysis of European and Asian individuals
- 29,880 RA cases and 73,758 controls
- ~10 million SNPs tested
- **Datafile:** RA\_GWAS.txt  
(reduced to random set of 100k SNPs)

# Exercise

- Read RA\_GWAS.txt table into R
- Make a histogram of the p-values

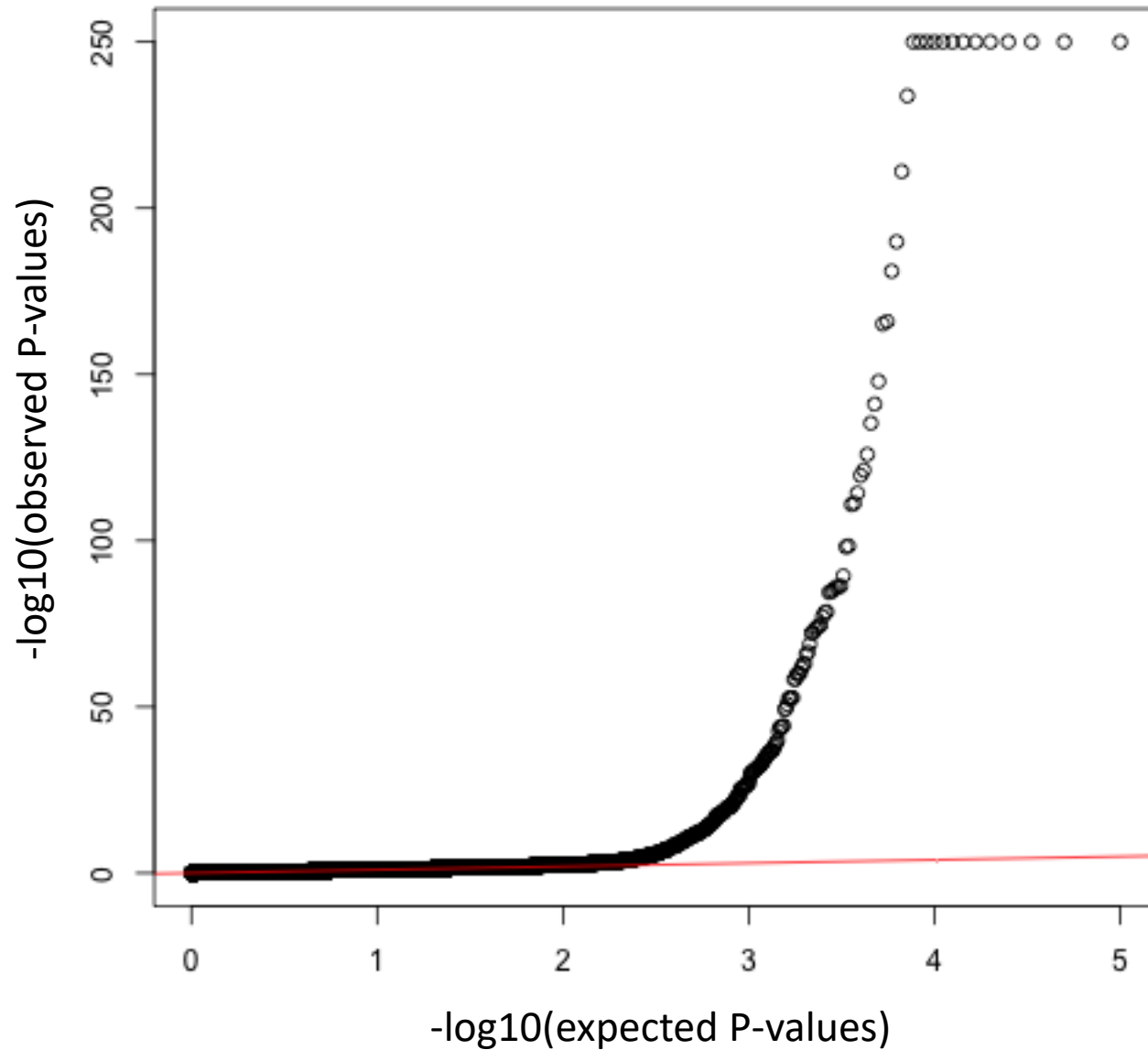
# P-value distribution



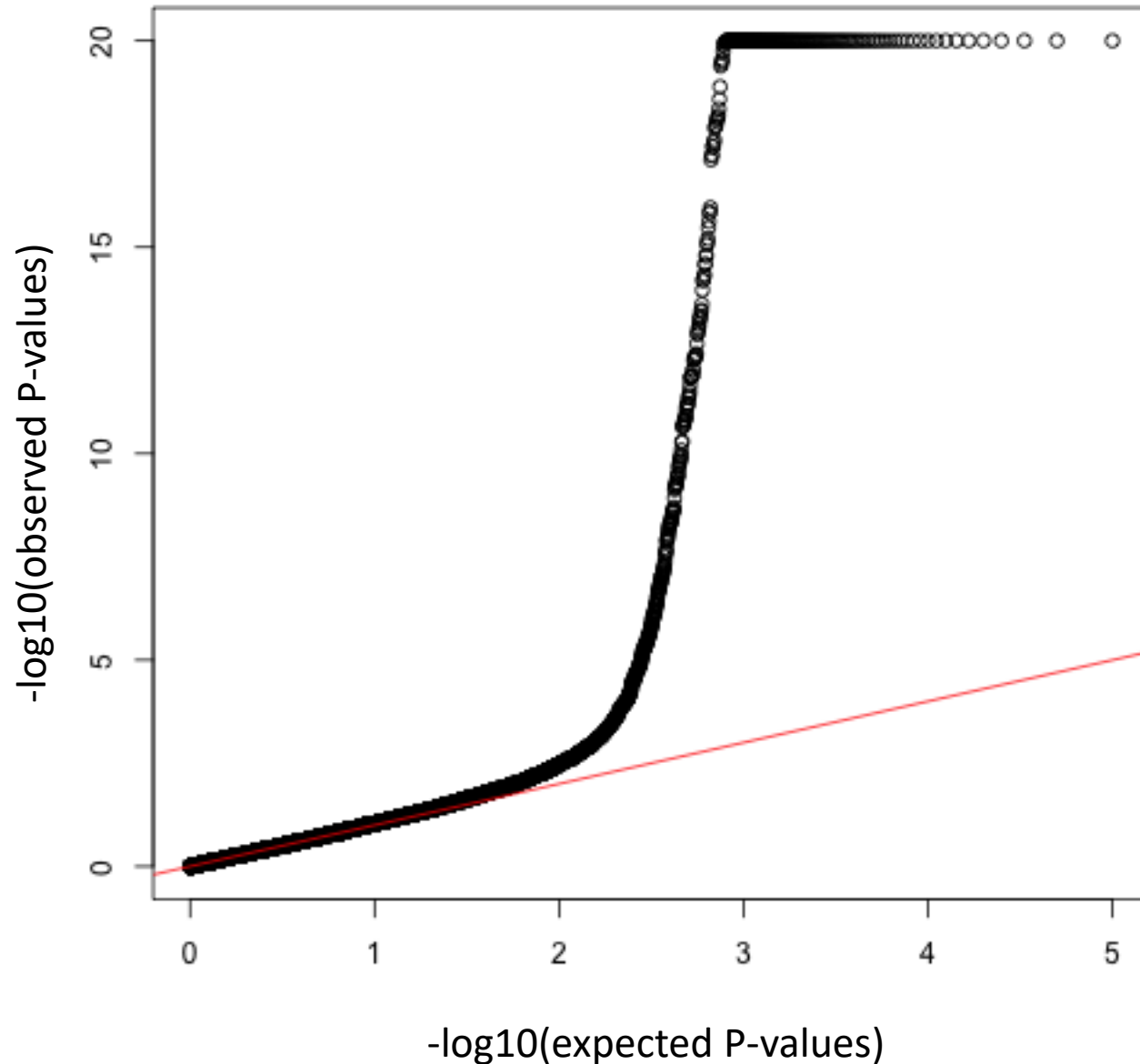
# Exercise

- Make a quantile-quantile plot of  $-\log_{10}$  expected vs. observed p-values
- Expected p-values are uniformly distributed

# P-value QQPlot

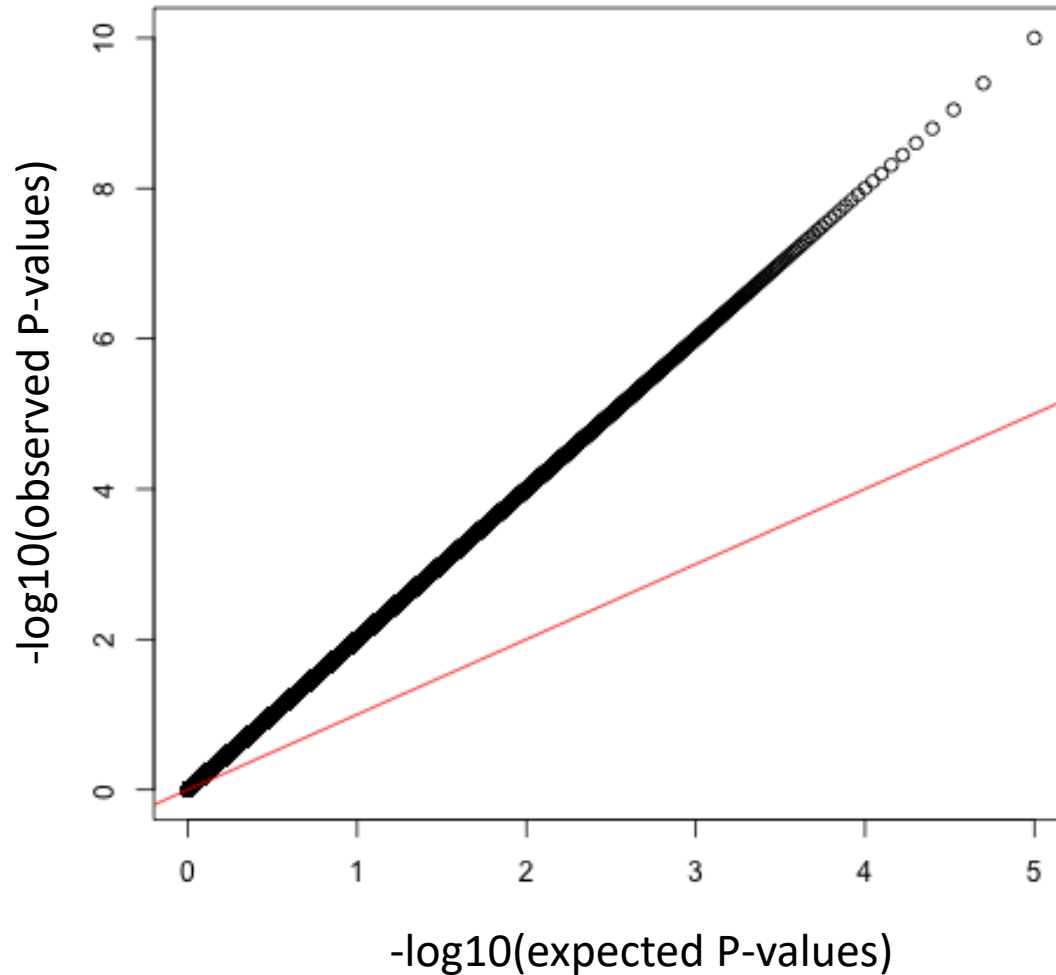


# truncated $-\log_{10}$ P-value QQPlot





# example of deflated p-values



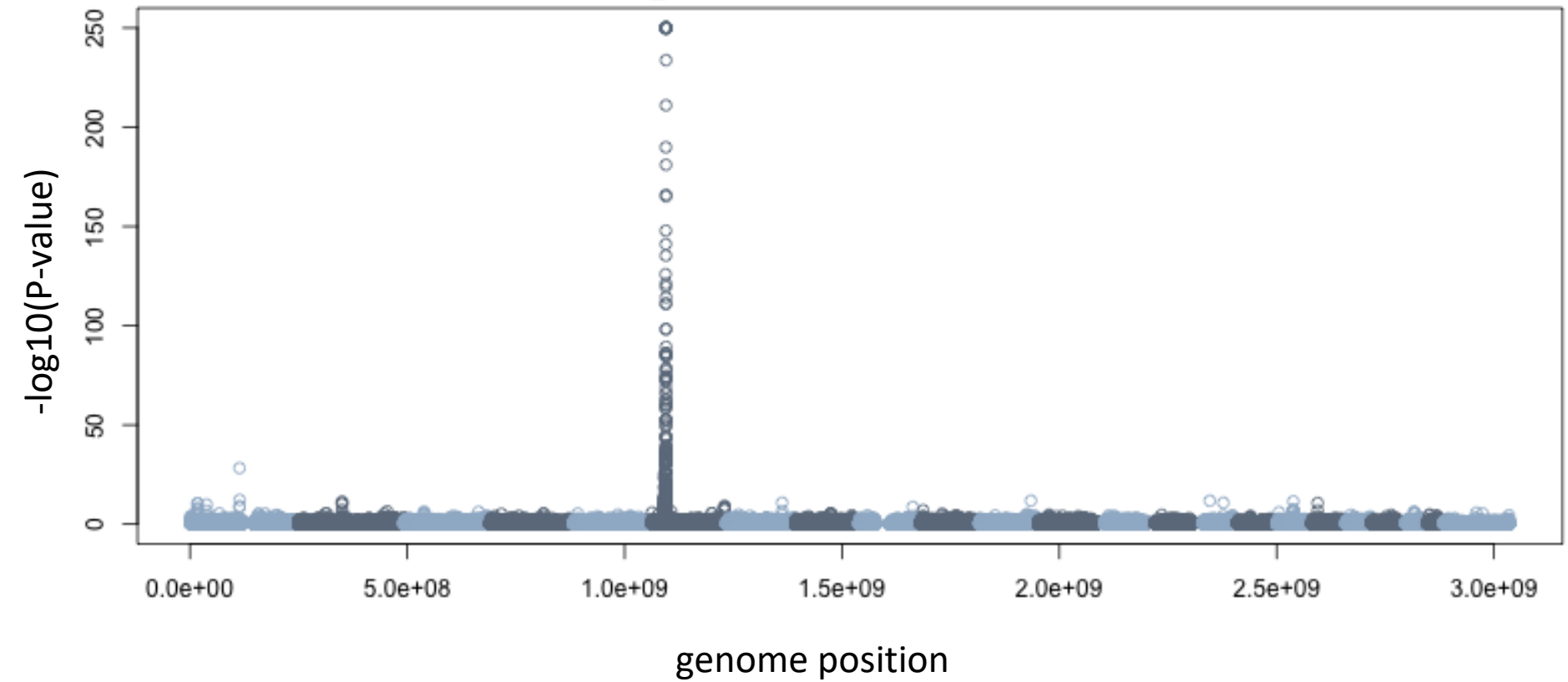
# Discussion

- What can cause "deflated" p-values?
- Inflated p-values?
- What other information can QQPlot of  $-\log_{10} P$ -values provide?

# Exercise

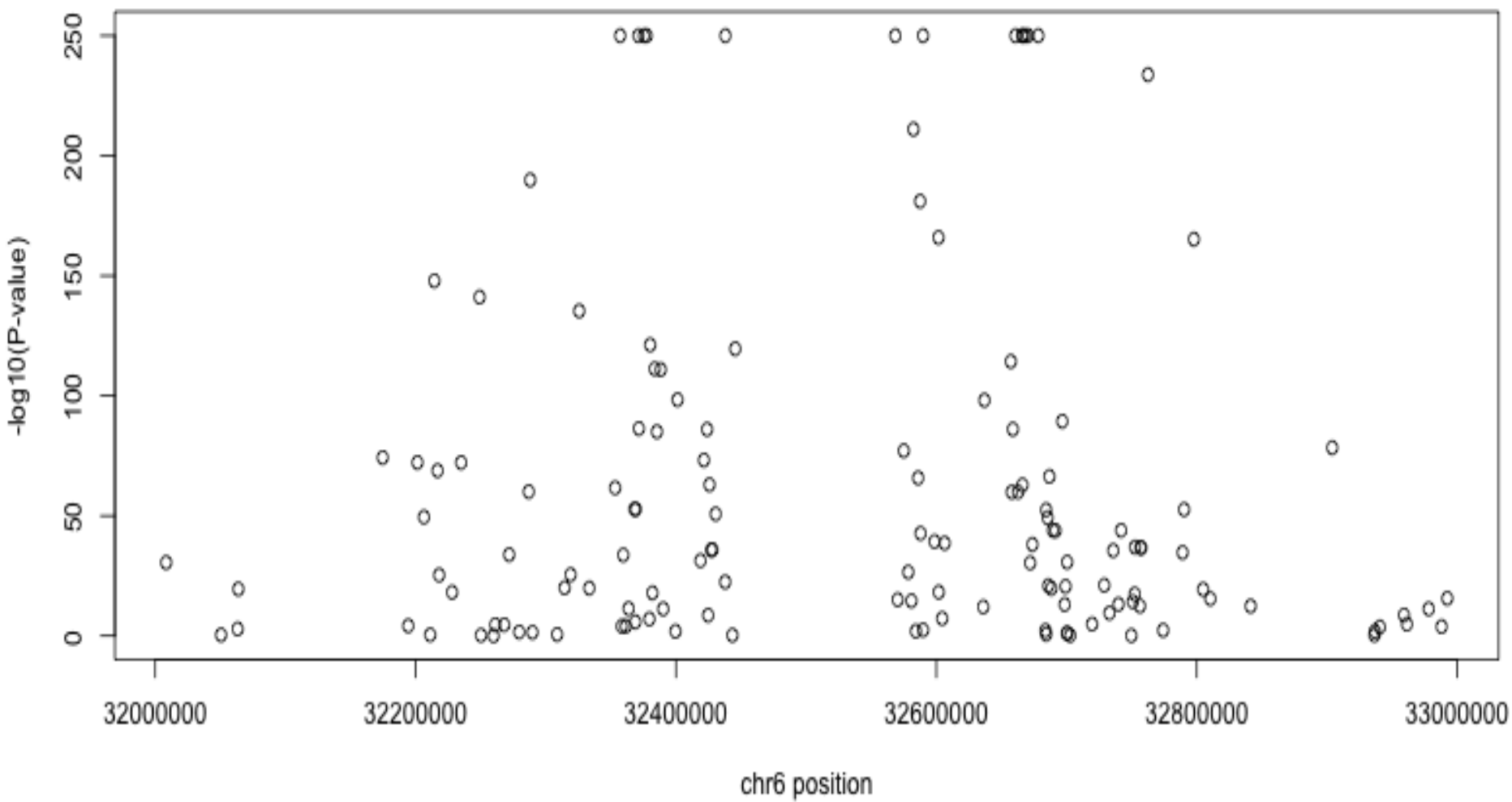
- Make a “manhattan” plot with SNP genome position on X-axis,  $-\log_{10}$  p-values on Y-axis
- Try to make neighboring chromosomes different colors

What region is this signal from?



# Exercise

- What SNP is giving the lowest p-value?
- What gene(s) are in this region (look up in genome browser)



# MHC Class 2

chr6 (p21.33-p21.32) p22.3 q12q13 15 6q21 q27

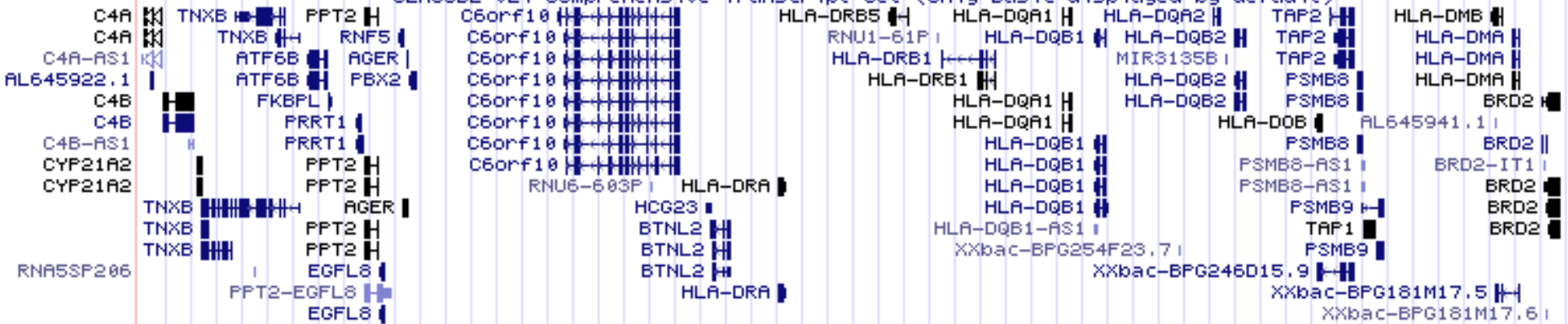
Scale  
chr6:

500 kb

32,500,000

hg38

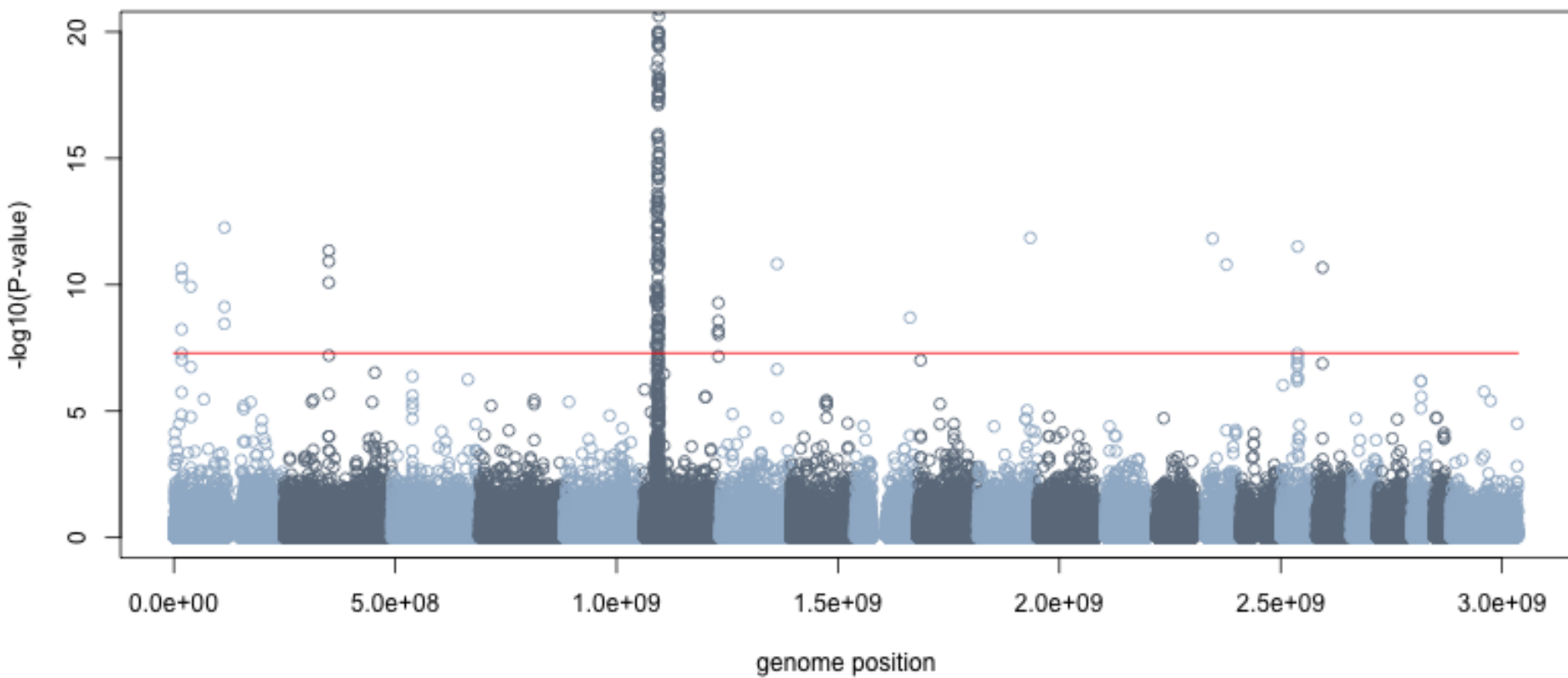
GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)



# Exercise

- What about outside of the MHC region?
- remake manhattan plot, but threshold low p-values to  $1e-20$
- Draw a line indicating genome-wide significance at  $p=5e-8$
- Roughly how many significant hits are there?



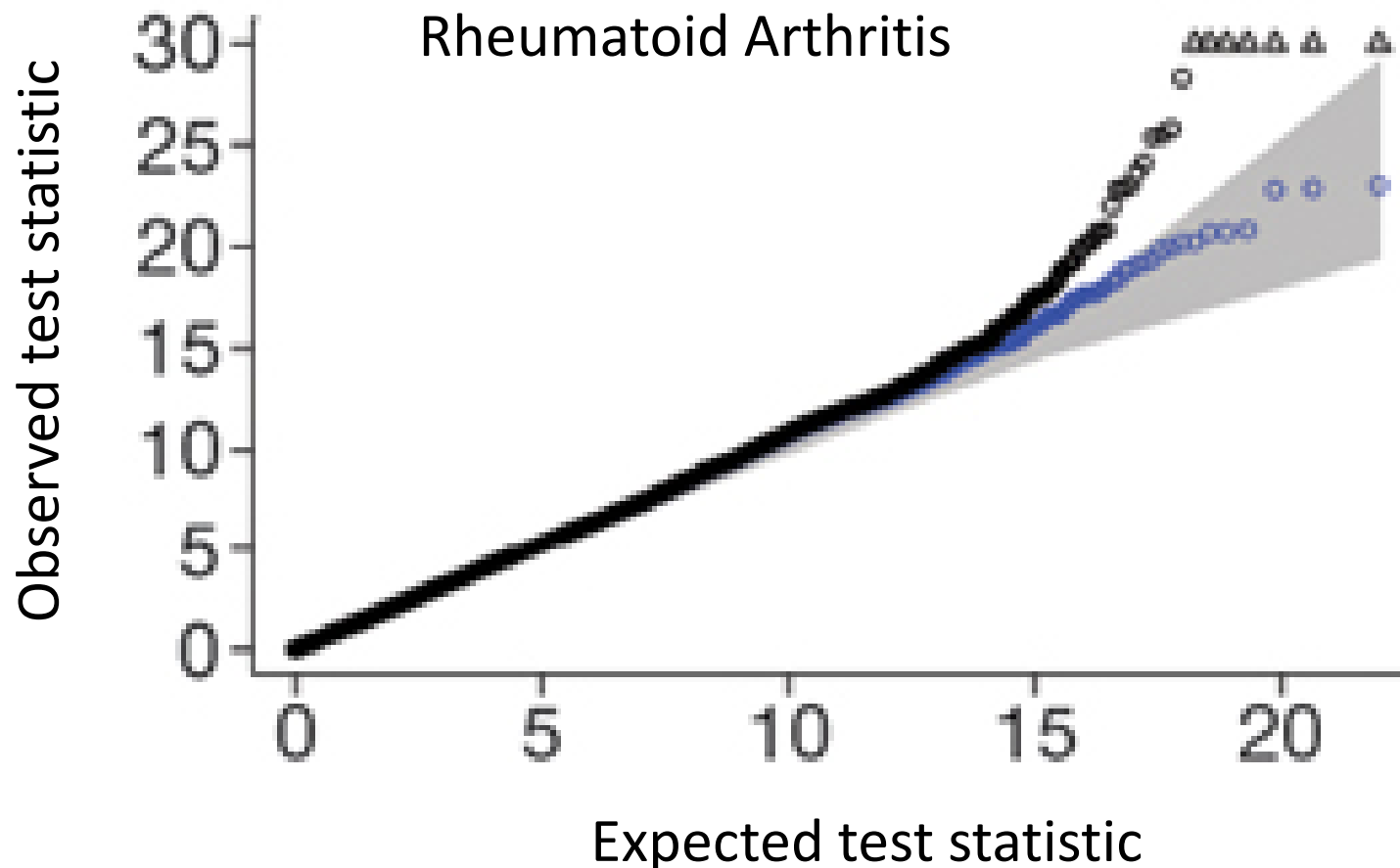


# Discussion

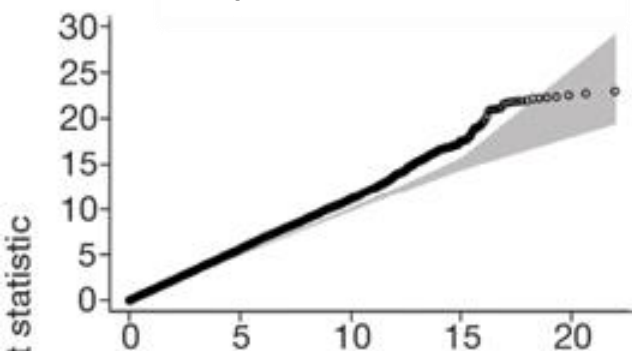
- Why use  $5e-8$  as a threshold for genome-wide significance?
- Why not simply correct for the number of SNPs tested?

# Wellcome Trust Case Control Study

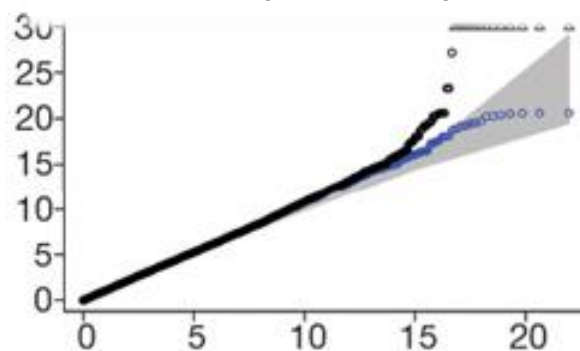
- 14,000 cases for 7 diseases
- 3,000 shared controls



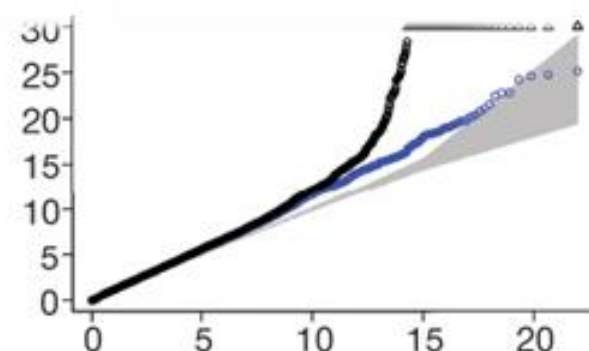
### Bipolar Disorder



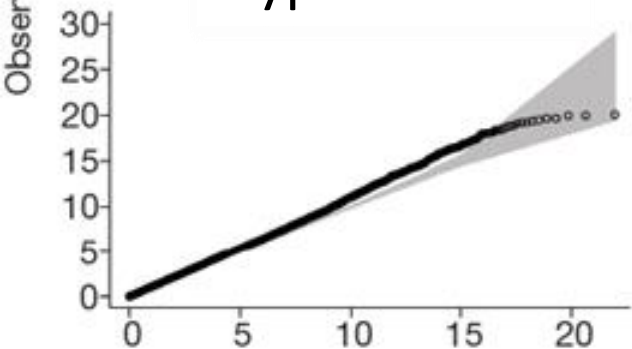
### Coronary Artery Disease



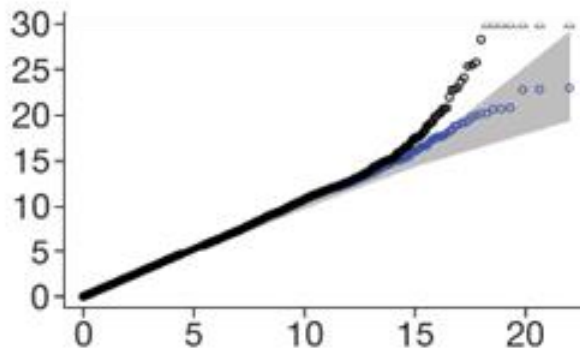
### Crohn's Disease



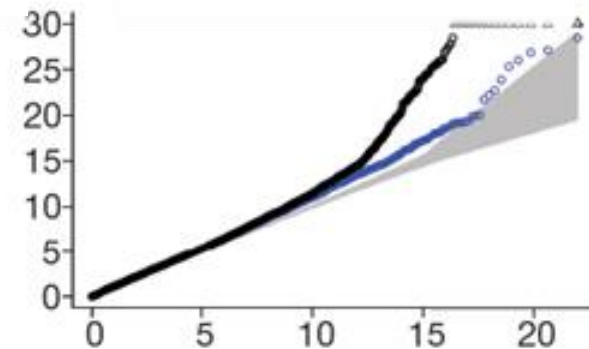
### Hypertension



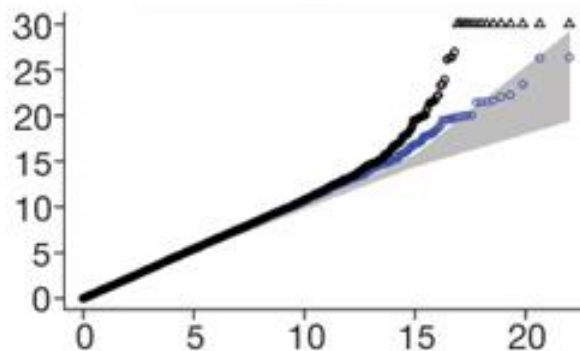
### Rheumatoid Arthritis



### Type 1 Diabetes

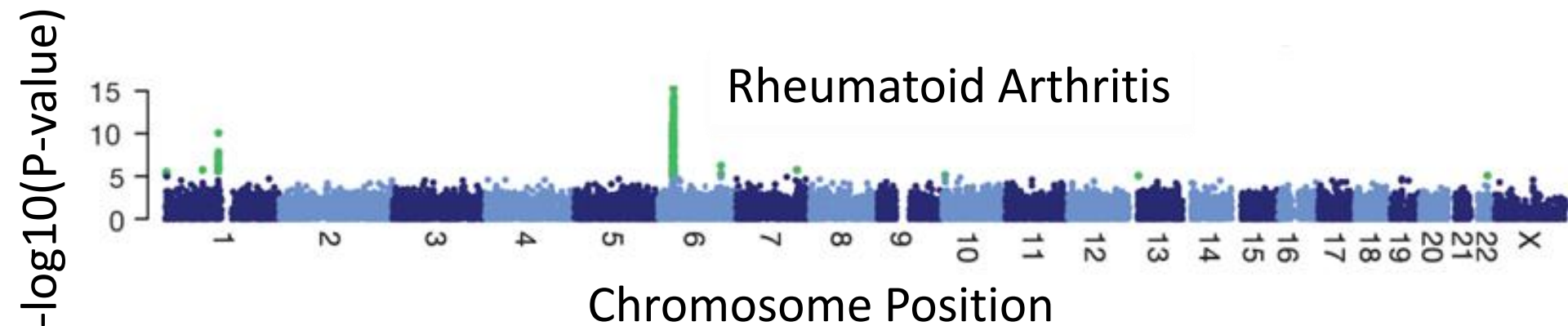


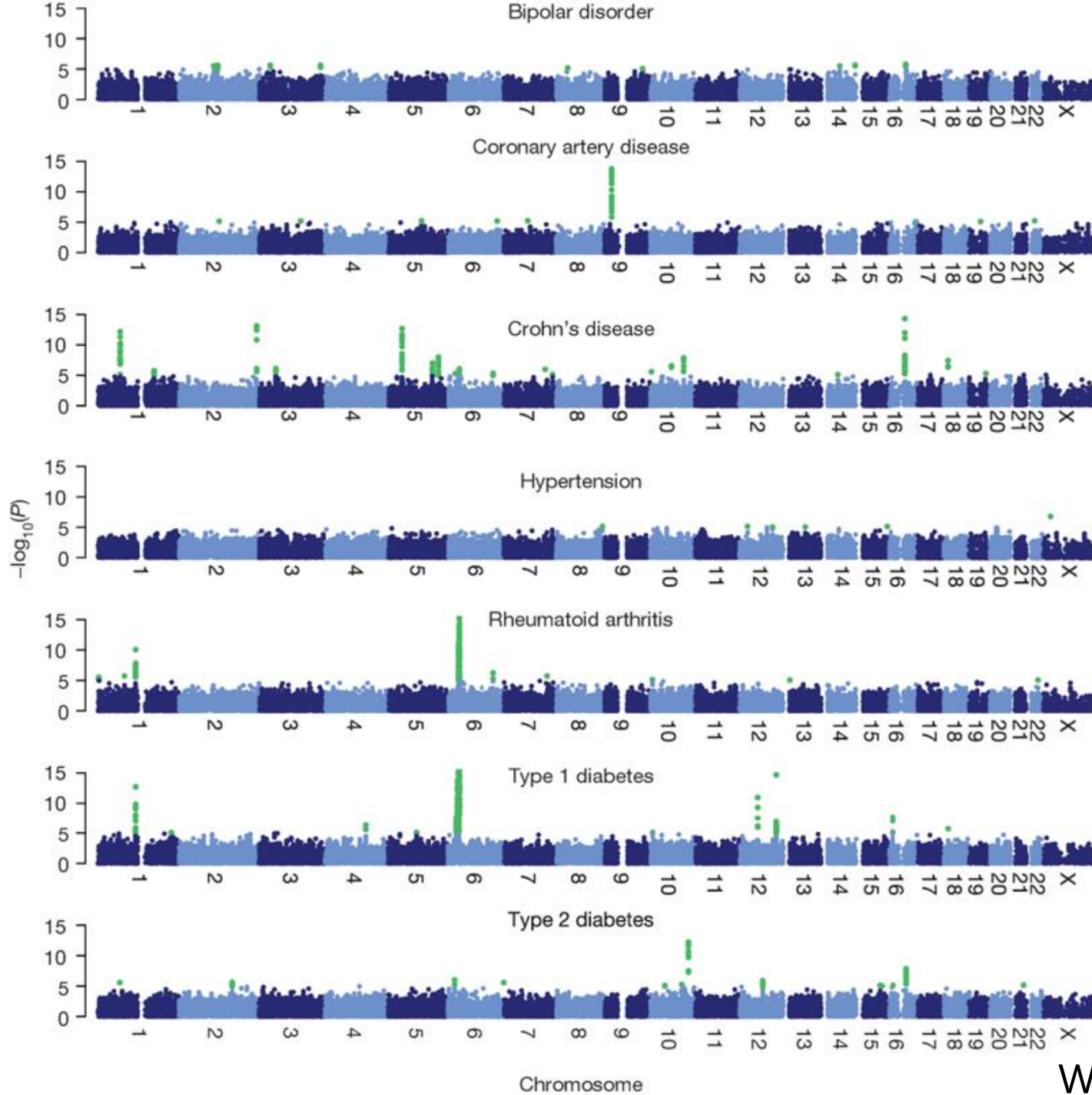
### Type 2 Diabetes



Expected chi-squared value

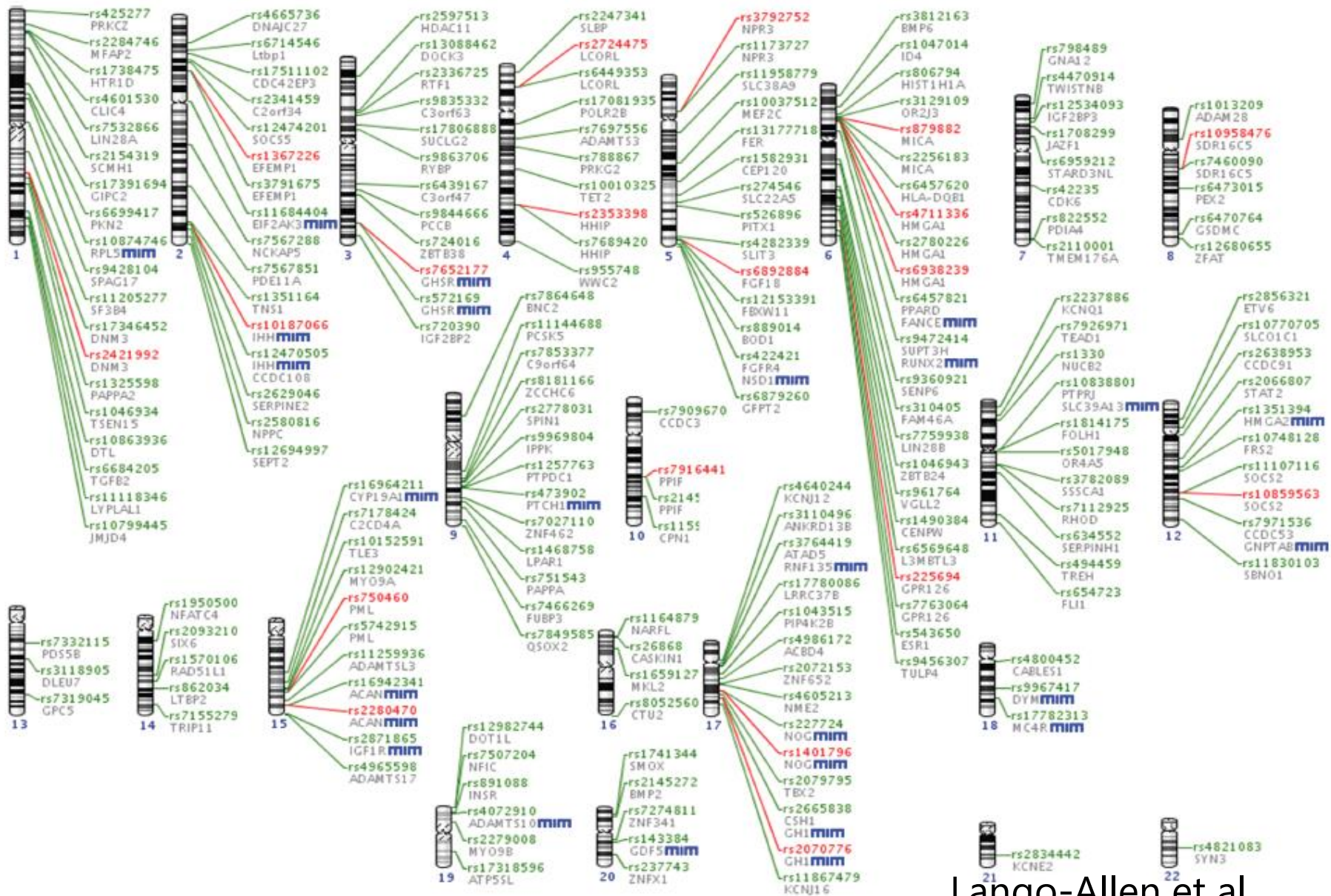
# Manhattan Plot of Association Signals





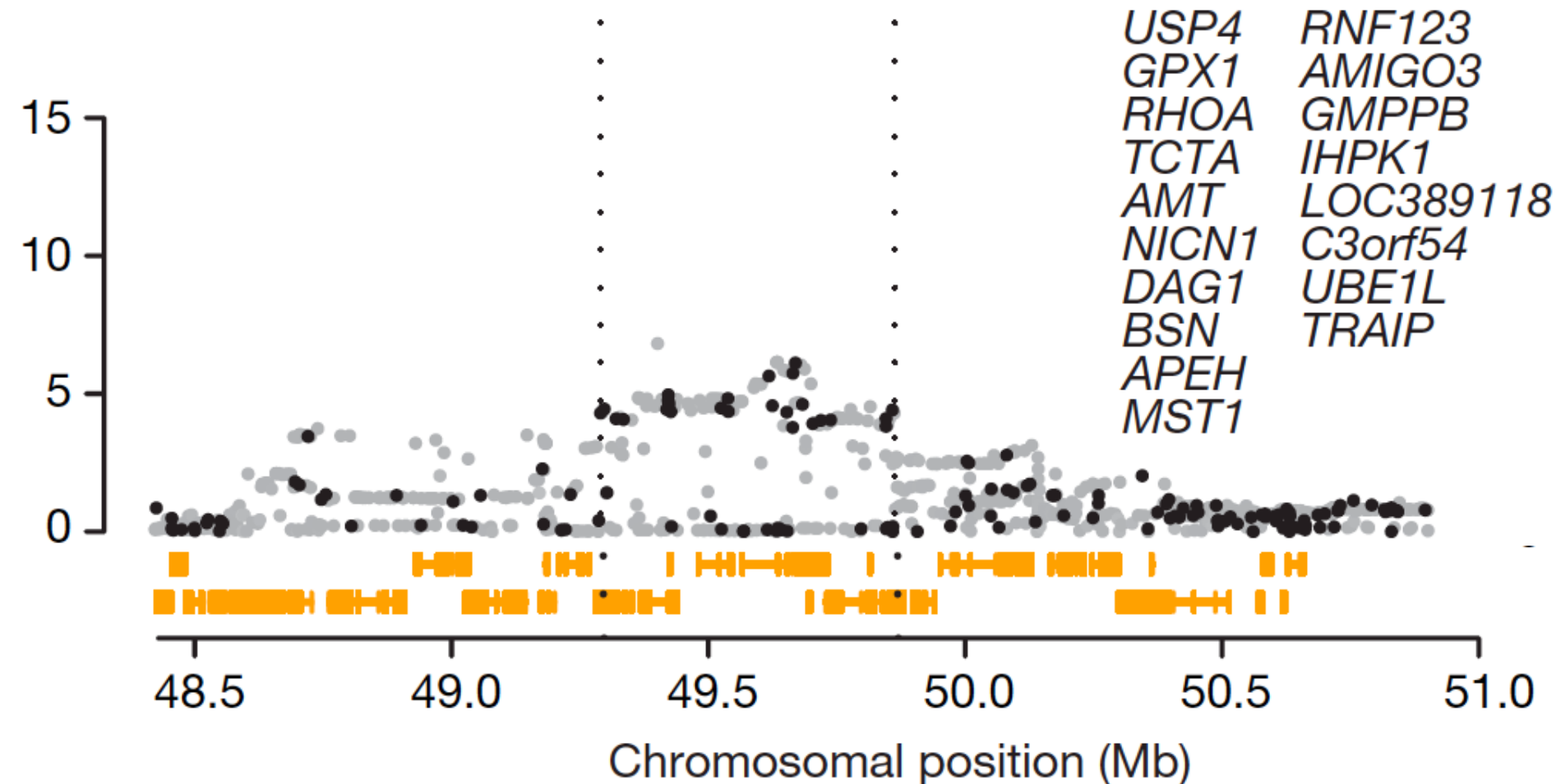


# QTLs for human height



Lango-Allen et al.  
2010

# Crohn's Disease Hit Region, Chromosome 3





# Discussion

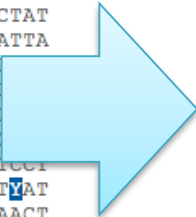
- What are advantages of GWAS over candidate gene studies?
- What are some disadvantages / limitations of GWAS?

# Difficulty in interpreting GWAS hits

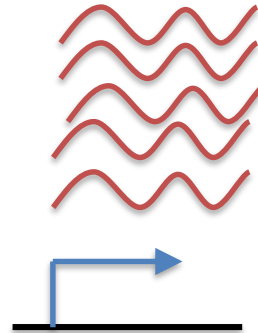
- Most associated variants (hits) are far from known genes
- Which gene(s) do they effect?
- What is their function / mechanism?
- What cell types are the variants active in?
- Which variants are causal?

# Molecular traits as an intermediate phenotype

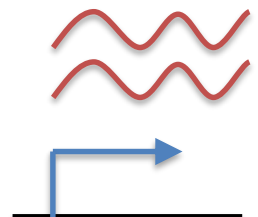
AAATAGAACTTTTGAGGTTGAATCTCTTTTAATGTAATG  
 TTACACTATACAATAGGAGTAGAAATTTGTCACCACTC  
 TTCTTTTTTTATTTTACATTAAAAAAATTTTACTT  
 GGATGTGCAGGTTTGTACATAGGTAAATGTTTTATTT  
 ATTTTAAAGTCATACAACCTCTCATAGCCAGTAGTTAAT  
 GTTGATGAATTGCATCCTGTTAATAATTGCTACAGATT  
 TGATGRTCTGGGTTGGCATAAGTACATGAAGATTTAC  
 GATGAAGAAATTTAGTGTTTTTTTTTAATTTTAAGAA  
 TTTATTTCCCACTGAAAAATAAATCCACCGGGYATAA  
 AGTAACCCAACTTGAAGCTAGTTTTTCAGACTTAGGCA  
 CTCATGGTCACCCCTTGCAAGAGAAATATCTAATTGAAA  
 TTGTTAGTGCTAAAAATGACATAAAGGGATCTCACTGGG  
 TTGTTARAGGTTTAAATTGTTAGTAACTTGTTATTGCA  
 TAAATAAAAAAACTTTTTTAAAAATAAAAAATTACAGA  
 GAGGTTCTGTACAACCCCTCCAGTTTCTCTTACTAT  
 ACATTTGTCACAATTAGTGAACCAATATTGATACATTA  
 CTTTTACTGGAGAATGGTGTAGAAACTAAGGTC  
 TTGAGATGTGTATTATTTTAGGTTCTTCTCAGC  
 TATATTAACCTATGTGTACACATACATCTATGAT  
 TTATTAAGCTAAATATGAGTTTCATATGGTGTCTT  
 TATTCTAGCCTCTTCTCTTGTCTTATCTGTAACCT  
 TCTTCCACCACCTACTATCTGCTTACCTAATTTCTYAT  
 CTTCAGAATTATTACATATAACCTGTGGGATACAAC  
 TGTAAAGTTCTTCTATCTTTAGTTTACTGACTCTACTC  
 GAACATTTCACTCATACTCCTCCTAGTGAAGTTGTTTT  
 TTTTGTGCAGTCTGCATTCCATTTTAGGGTTCCCTCCT  
 TTTTAAATTCATATACATCAAGGTTTATTCTTTGTGC  
 CAAATACAAAGTGTCTGTATCCCATCATTACAATGTCA  
 AAAAATATCCCTTGTCTTTGGCTATTCAACCTTCC  
 ACCACTGATCTGTTTATCGTGGAGCTGTGTCTCTTCCA  
 ACAATATGTAGACTTTTACCCTGGCTTATTTTGTTAG  
 CATGTCCTTATGTGGCTTGTAGTTTCACTTTTACT  
 AAATGTGCCACARTTTGTATTCCATTYGTGATTGAA  
 ATGACTGCTAGATAGYATAGTAAGACTATATTTAKCTT  
 TTTTAAAGTGGCTGTACCATTGTGCCACCAGCAACTCC  
 GTTTTTTGGATTTTAKCATTTCTAAAAGGTGAGTGATG  
 TGTAACTCTAATGACAAATGATGGTGGATTCTTTT  
 ATATCTTCTTTAGTATGTGTCTGTTCCGATGTTTGTCT  
 TTGTTTTCTTTTTCTTTTTCTTTTTCTTTTGAGAYG  
 TGGAGTGCAGTGGCGCCATCTCGGCTCACTGCAAGCYC  
 TTCGTACCTCAGCCTCCCGAGTAGCTGGGACTACAGGC  
 TTTTTTGTATTTTGGTGGAGACGAGGTTTACCATG  
 CCTGACCATAGATGATCTGCCTGTCTTGGCCTCCCAAA



GG



GA



AA

