

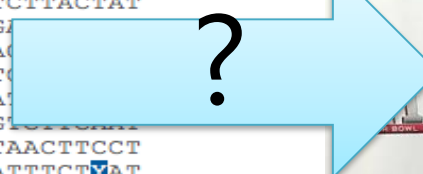
# Genetic Variation and Transcriptional Regulation

**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>ACATTAAAAAAAATTTTACTT  
TAAGTTCCAGGATACATGTGCAGGATGTGCAGGTTTGTACATAGGTAAATGTTTTATTT  
TAAATTTAATTTAACACTTTTTAT<sup>Y</sup>TTTAAGTCATACAACCTCTCATAGCCAGTAGTTAAT  
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  
CTTGAGATATTAAGCATTTAAAAATTGTTA<sup>R</sup>AGGTTTAAATTTGTTAGTAACTTGTATTGCA  
TAGAAAAATGTGCCAAATGTGTCAGTAAATAAAAAAACTTTTTT<sup>T</sup>AAAAATAAAAAATTTACAGA  
<sup>A</sup>RAATTATGAC<sup>R</sup>ATACTACAAAGAGGTTCTGTACAACCCCCCTCCCAGTTTCTCTTACTAT  
TAACATCTTAAATTAGTATGTTACATTTGTGCACAATTAGTGAACCAATATTG  
GTAC<sup>K</sup>AACTAAAGTCAGTGTTCTTTTTACTGGAGAAATGGTGTAGAAAATAA  
CACTGTGGTATGGTGGTTGCTTCTCCACCACCTACTATCTGCTTACCTAATTTCT<sup>Y</sup>AT  
TTCCAGTTTA<sup>K</sup>GTATACAGTGGCTTCCAGAAATTATTACATATA<sup>R</sup>CCCTGTGGGATACAAC  
TTGTCAACTAGAGTGGTGCTTATGTAAGTTCTTCTATCTTTAGTTTTACTGACTCTACTC  
ATTTTCAAAGTTGCTTAGTCCAGAACATTTTCACTCATACTCCTCTAGTGAAGTTGTTTC  
ATA<sup>Y</sup>GTTAGTAACACAGATTCTTTTT<sup>Y</sup>TGCAGCTGCAATCCATTTAGGGTTCCCTCTCT  
CTCCAATCTCCTAAATATTATTTTTTAAATTCATATACATCAAGGTTTATTTCTTGTGC  
TGTAAR<sup>G</sup>TTCTATAGGTTTTGACAAATACAAAGTGTC<sup>R</sup>TGTACCCATCATTACAATGTCA  
TAC<sup>R</sup>GAAATCGTTTTCACTGCT<sup>Y</sup>CTAAAAATATCCCTTGTCCCTTTGCCTATTCAACC<sup>CT</sup>TCCC  
CTCCTTTCC<sup>CA</sup>AACTCCTGGCAACCACTGATCTGTTTATCGTGGAGCTGTGTCTCTTCCA  
GAATGTCATATAAATTGAATCATACAATATGTAGACTTTTCAACCCTGGCTTATTTTGTAG  
CAATATGCATTTAACATTTCATCCATGTCCTTATGTGGCTTGTAGTTTCACTTTTTTACT  
GCTGGGTAGTATT<sup>CTR</sup>TCATAGAAATGT<sup>R</sup>CCACA<sup>R</sup>TTTGTATTATCCATT<sup>Y</sup>GCTGATTGAA  
GTATATCAATATACTTGGAACTGACTGCTAGATAG<sup>Y</sup>ATAGTAAGACTATATTTA<sup>K</sup>CTT  
TGCAAGAAACTGCCAACTGTATTTTAAAGTGGCTGTACCATTGTGCCACCAGCAACTCC  
TGCCAGTGATCCAGTATTGTGAGTTTTTGGATTTTA<sup>K</sup>CCATTCTAAAAGGTGAGTGATG  
GTATCTCATTGTGCTTTTAAATTTGTAATACTCTAATGACAAATGATGGTGGATTCTTTTT  
CATATG<sup>Y</sup>TTGTTTTCCCATTTGTATATCTCTTTAGTATGTGTCTGTTCCGGATGTTTTGTCT  
TACTTTTTTAAACTGGGTGATTGTTTTCTTTTTCTTTTTCTTTTTTTCTTTTGTAGA<sup>Y</sup>G  
GAGTCT<sup>Y</sup>GCTCTTTAGCCAGGCTGGAGTGCAGTGGCGCCATCTCGGCTCACTGCAAGC<sup>Y</sup>C  
TGCCCTCCGGGTTCAAGTGATTTTCGTACCTCAGCCTCCCGAGTAGCTGGGACTACAGGC  
GCCCCGCCACCACACCTGGCTAATTTTTTTGTATTTTTGGTGGAGACGAGGTTTCACCATG  
TCGGTCAGGCTGGTCTTAA<sup>W</sup>CTCCTGACCATAGATGATCTGCCTGTCTTGGCCTCCCAAA



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

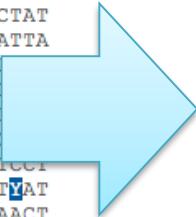
# Topics

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

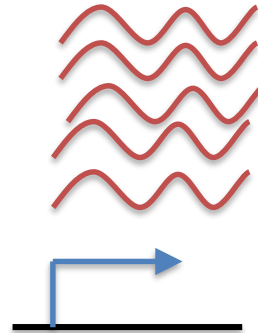


# Molecular traits as an intermediate phenotype

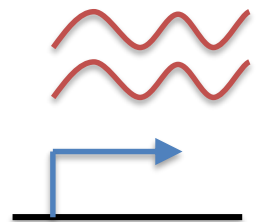
AAATAGAACTTTTGAGGTTGAATCTCTTTTAATGTAATG  
 TTACACTATACAATAGGAGTAGAAATTTGTCACCACTC  
 TTCTTTTTTTATTTTACATTAAAAAAATTTTACTT  
 GGATGTGCAGGTTTGTACATAGGTAATGTTTTATTT  
 ATTTTAAAGTCATACAACCTCTCATAGCCAGTAGTTAAT  
 GTTGATGAATTGCATCCTGTTAATAATTGCTACAGATT  
 TGATGRTCTGGGTTGGCATAAGTACATGAAGATTTAC  
 GATGAAGAAATTTAGTGTTTTTTTTTAATTTTAAGAA  
 TTTATTTCCCACTGAAAAATAAATCCACCGGGYATAA  
 AGTAACCCAACTTGAAGCTAGTTTTTCAGACTTAGGCA  
 CTCATGGTCACCCCTTGCAAGAGAAATATCTAATTGAAA  
 TTGTTAGTGCTAAAAATGACATAAAGGGATCTCACTGGG  
 TTGTTARAGGTTTAAATTGTTAGTAACTTGTTATTGCA  
 TAAATAAAAAAACTTTTTTAAAAATAAAAAATTACAGA  
 GAGGTTCTGTACAACCCCTCCAGTTTCTCTTACTAT  
 ACATTTGTCACAATTAGTGAACCAATATTGATACATTA  
 CTTTTACTGGAGAATGGTGTTAGAACTAAGGTC  
 TTGAGATGTGTATTATTTTAGGTTCTTCTCAGC  
 TATATTAACCTATGTGTACACATACATCTATGAT  
 TTATTAAGCTAAATATGAGTTTCATATGGTGTCTT  
 TATTCTAGCCTCTTCTCTTGTCTTATCTGTAACCT  
 TCTTCCACCACCTACTATCTGCTTACCTAATTTCTYAT  
 CTTCAGAATTATTACATATAACCTGTGGGATACAAC  
 TGTAAAGTTCTTCTATCTTTAGTTTACTGACTCTACTC  
 GAACATTTCACTCATACTCCTCCTAGTGAAGTTGTTTT  
 TTTTGTGCAGTCTGCATTCCATTTTAGGGTTCCCTCCT  
 TTTTAAATTCATATACATCAAGGTTTATTCTTTGTGC  
 CAAATACAAAGTGTCTGTATCCCATCATTACAATGTCA  
 AAAAATATCCCTTGTCTTTGGCTATTCAACCCTTCC  
 ACCACTGATCTGTTTATCGTGGAGCTGTGTCTCTTCCA  
 ACAATATGTAGACTTTTACCCTGGCTTATTTTGTAG  
 CATGTCCTTATGTGGCTTGTAGTTTCACTTTTACT  
 AAATGTGCCACARTTTGTATTCCATTYGCTGATTGAA  
 ATGACTGCTAGATAGYATAGTAAGACTATATTTAKCTT  
 TTTTAAAGTGGCTGTACCATTGTGCCACCAGCAACTCC  
 GTTTTTTGGATTTTAKCATTTCTAAAAGGTGAGTGATG  
 TGTAACTCTAATGACAAATGATGGTGGATTCTTTT  
 ATATCTTCTTTAGTATGTGTCTGTTCCGATGTTTGTCT  
 TTGTTTTCTTTTTCTTTTTCTTTTTCTTTTGAGAYG  
 TGGAGTGCAGTGGCGCCATCTCGGCTCACTGCAAGCYC  
 TTCGTACCTCAGCCTCCCGAGTAGCTGGGACTACAGGC  
 TTTTTTGTATTTTGGTGGAGACGAGGTTTACCATG  
 CCTGACCATAGATGATCTGCCTGTCTTGGCCTCCCAAA



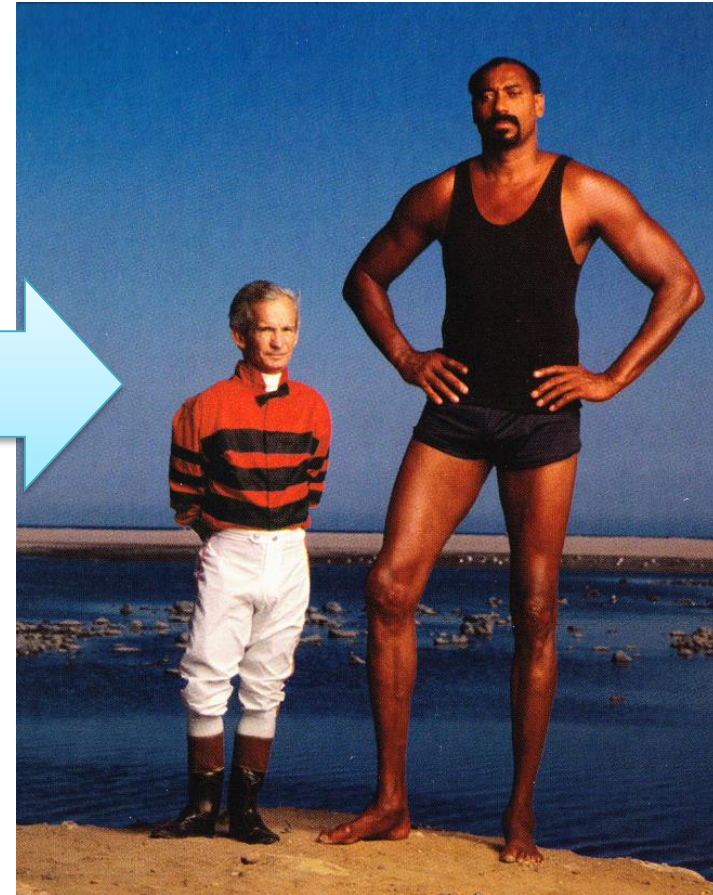
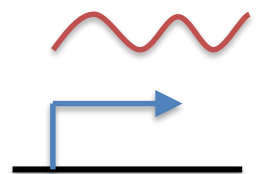
GG



GA



AA



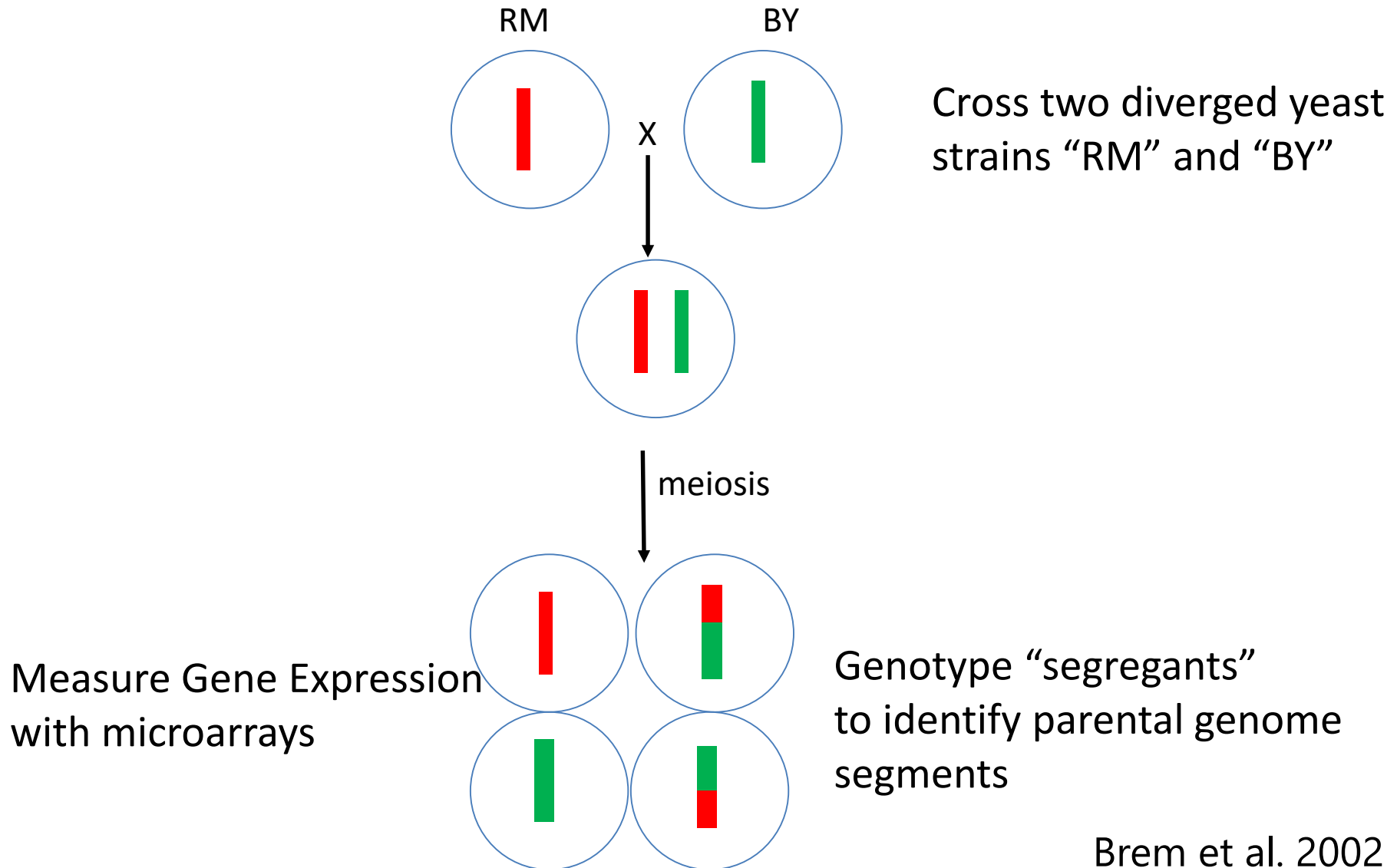
# Why use molecular traits?

- Close to underlying genetics
  - affected by small number of polymorphisms
  - require smaller sample sizes
- Can measure 1000s of traits in single experiment (e.g. RNA-seq)
- Reveal molecular basis of organismal traits
  - implicate specific cell types

# Discussion

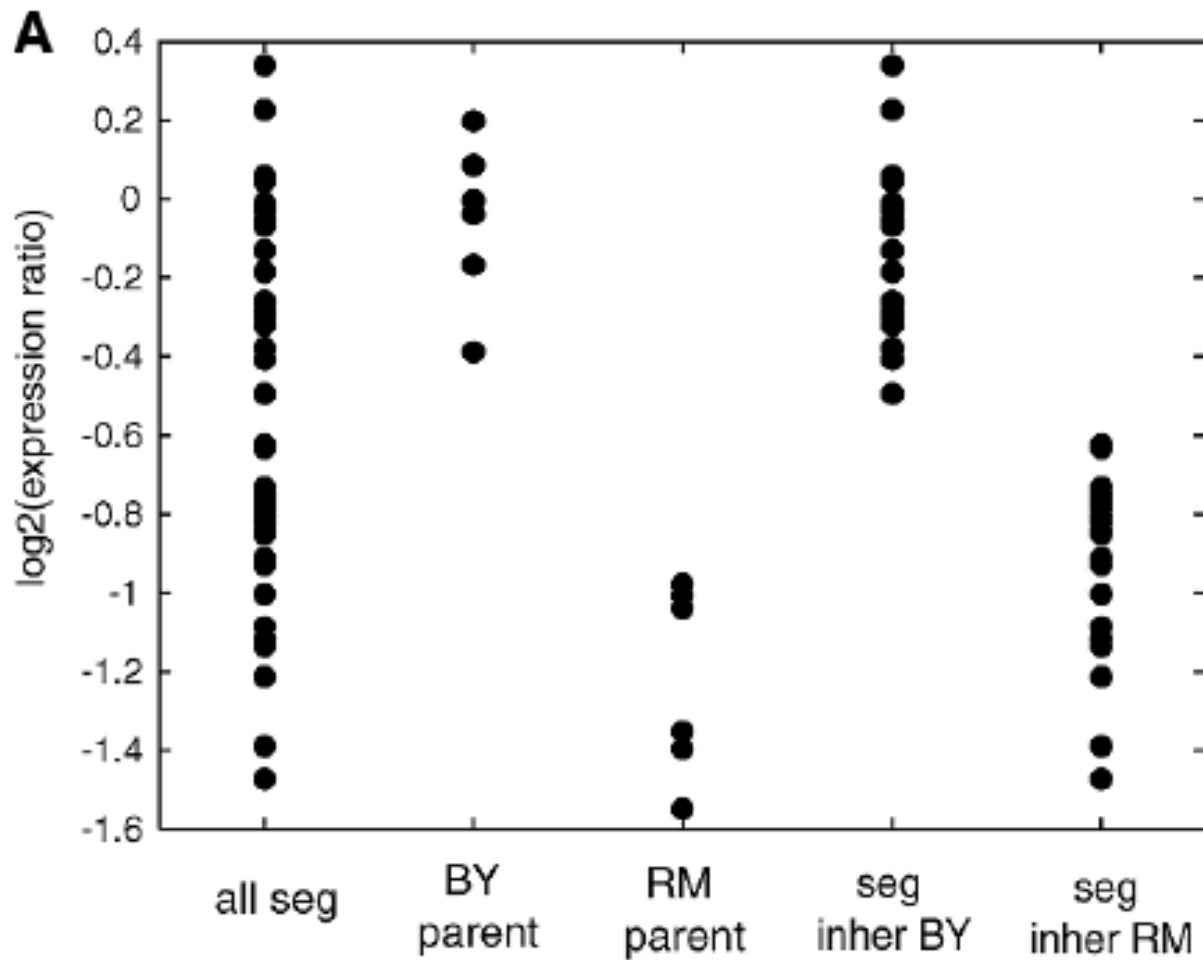
- What else could be used as molecular or cellular traits?
- How would you measure these?

# The first gene expression QTL mapping study

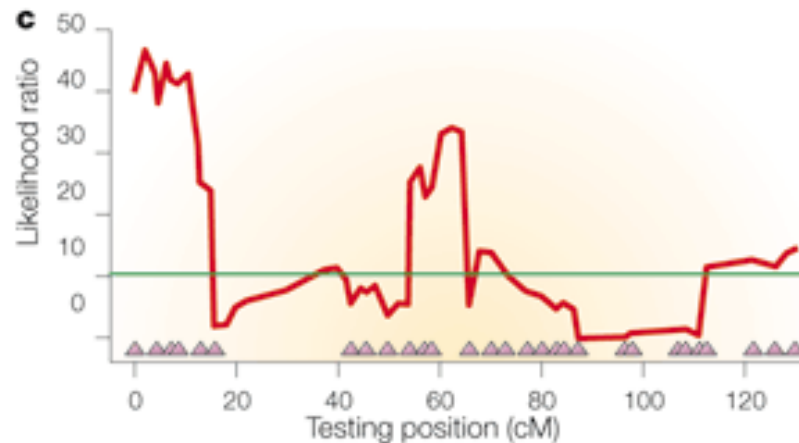
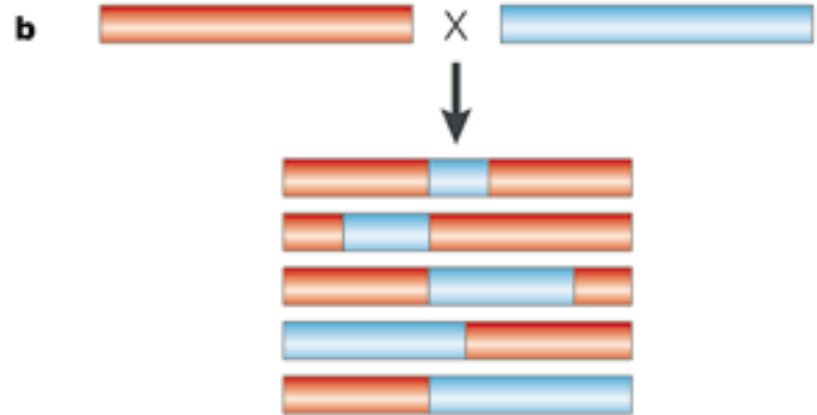
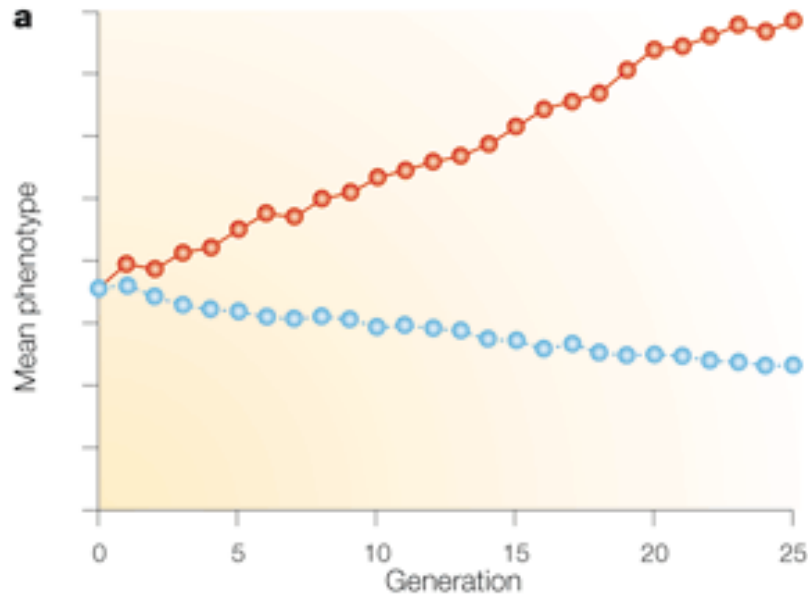




# A yeast eQTL

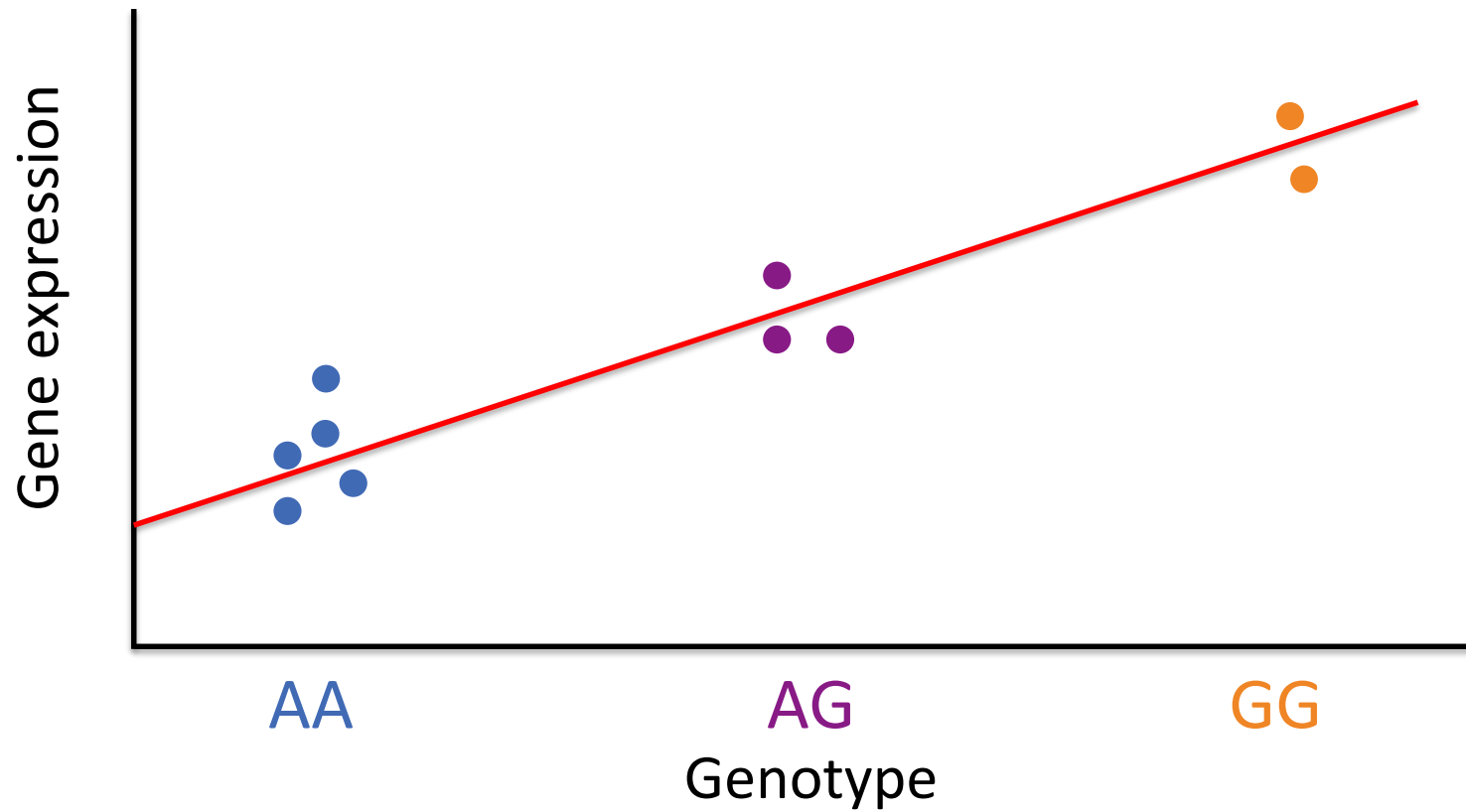


# Quantitative Trait Locus Mapping with recombinant inbred lines

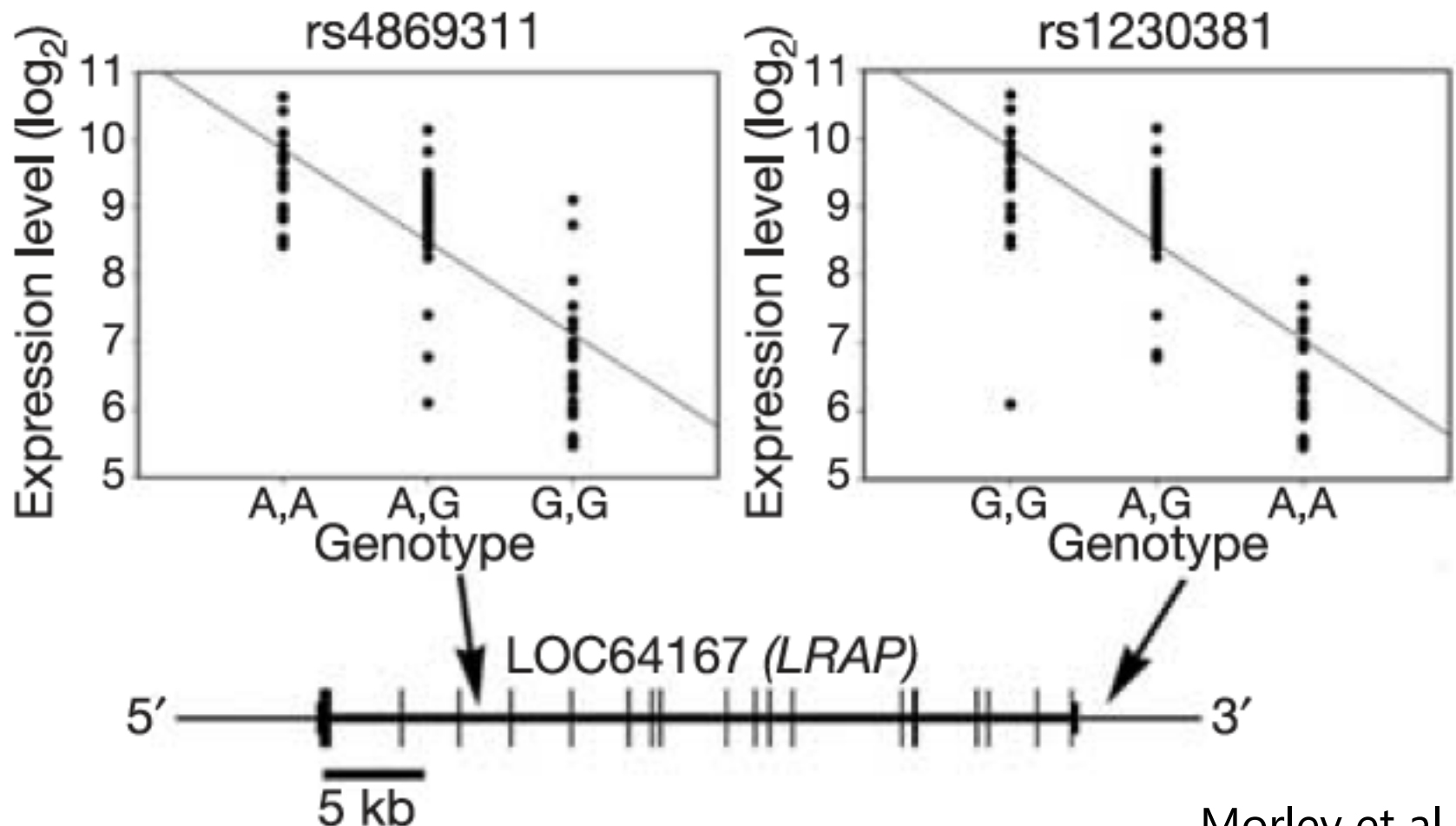


Lango-Allen et al.  
2010

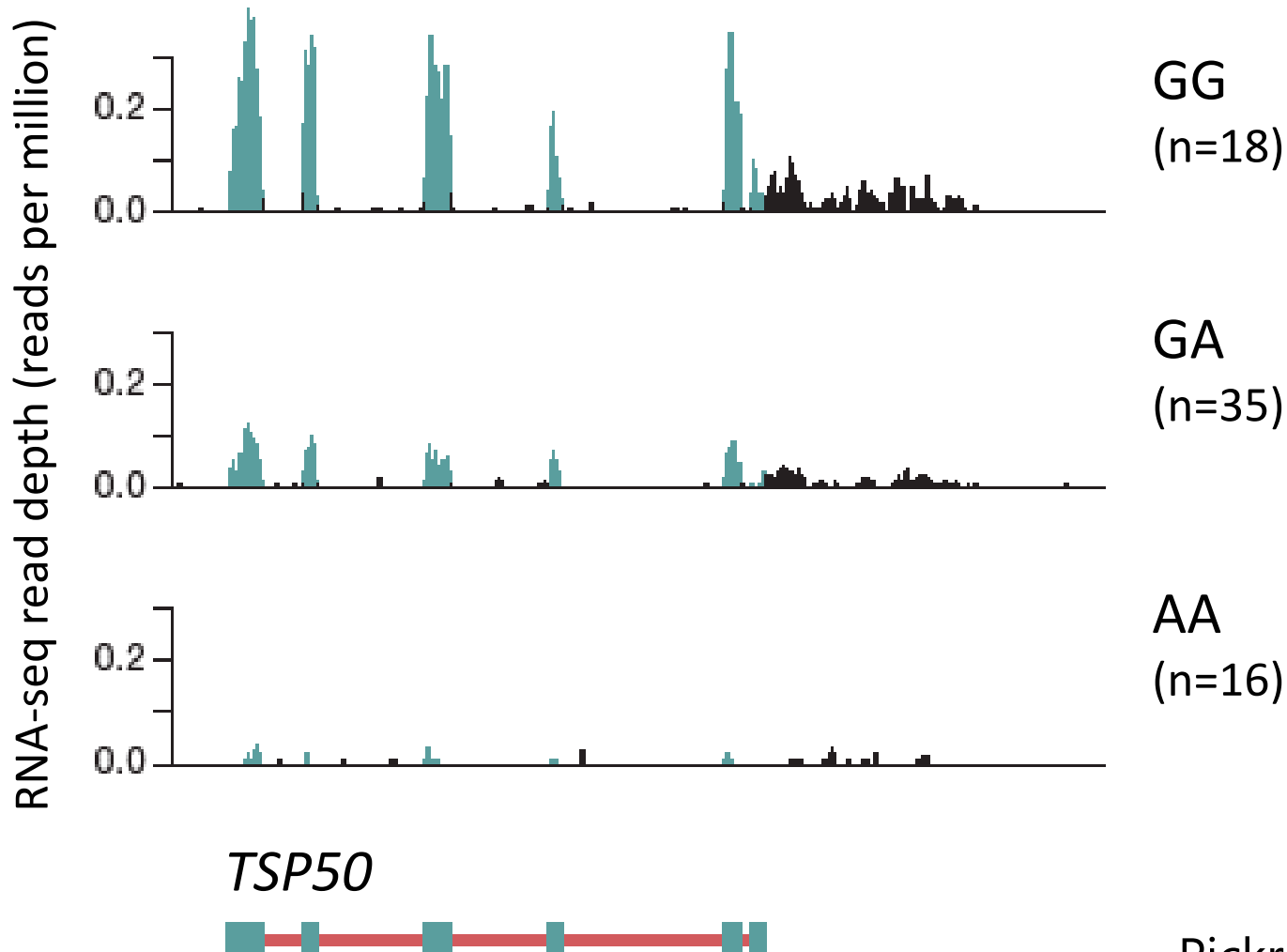
# eQTL mapping in humans



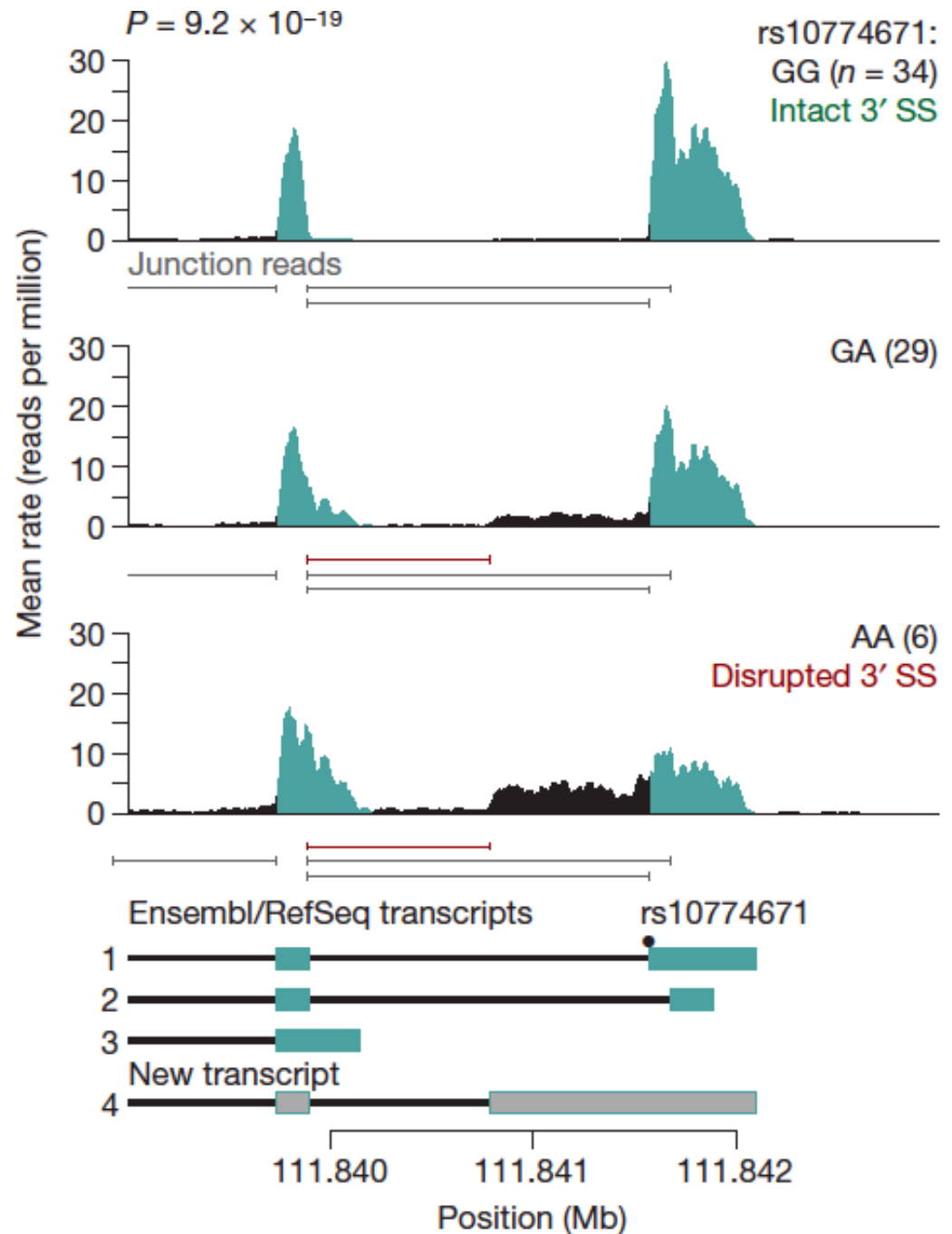
# Pedigree-based mapping of eQTLs in human B cell lines



# eQTL mapping with RNA-seq

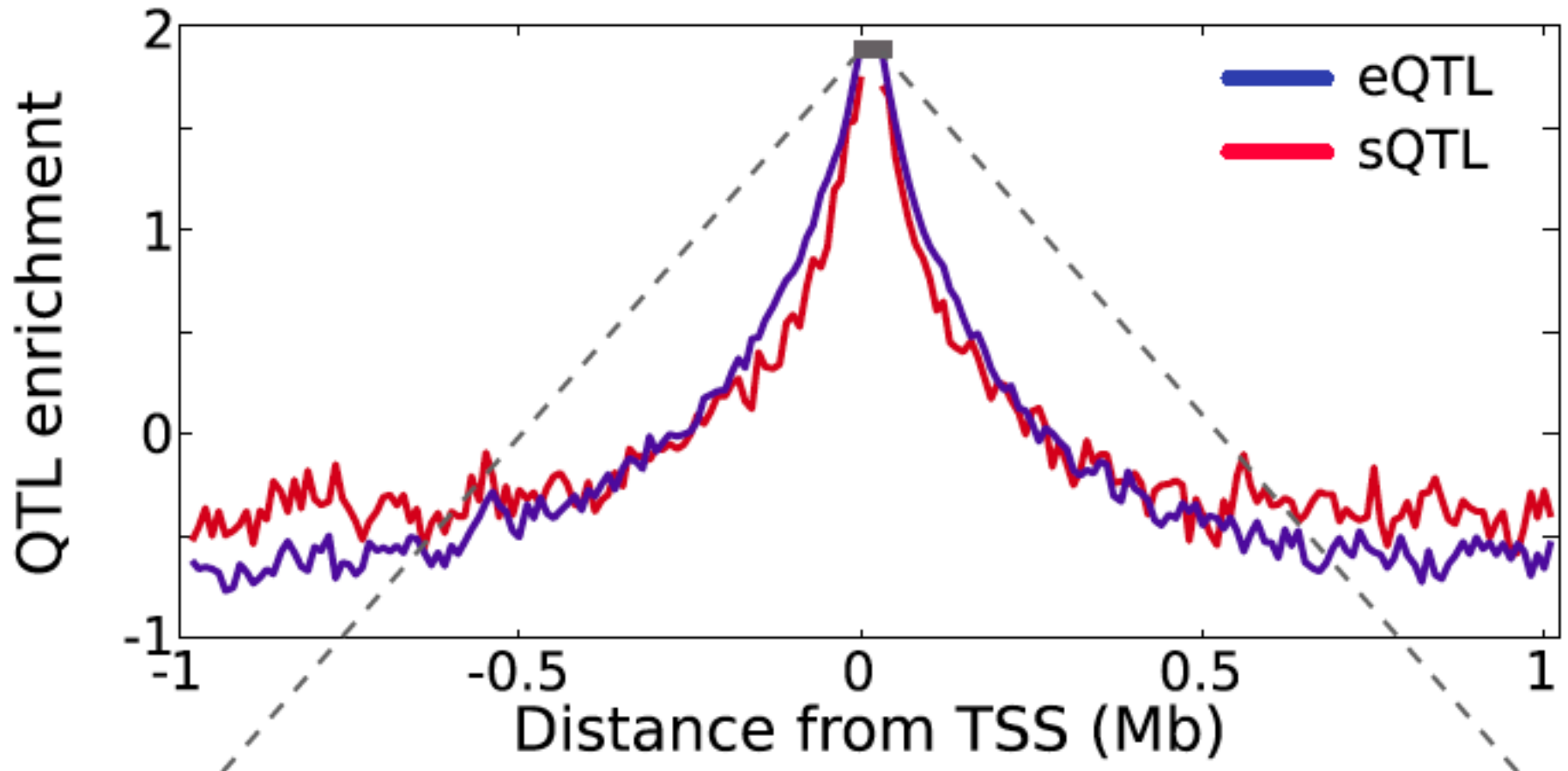


# Splicing QTLs (sQTLs)

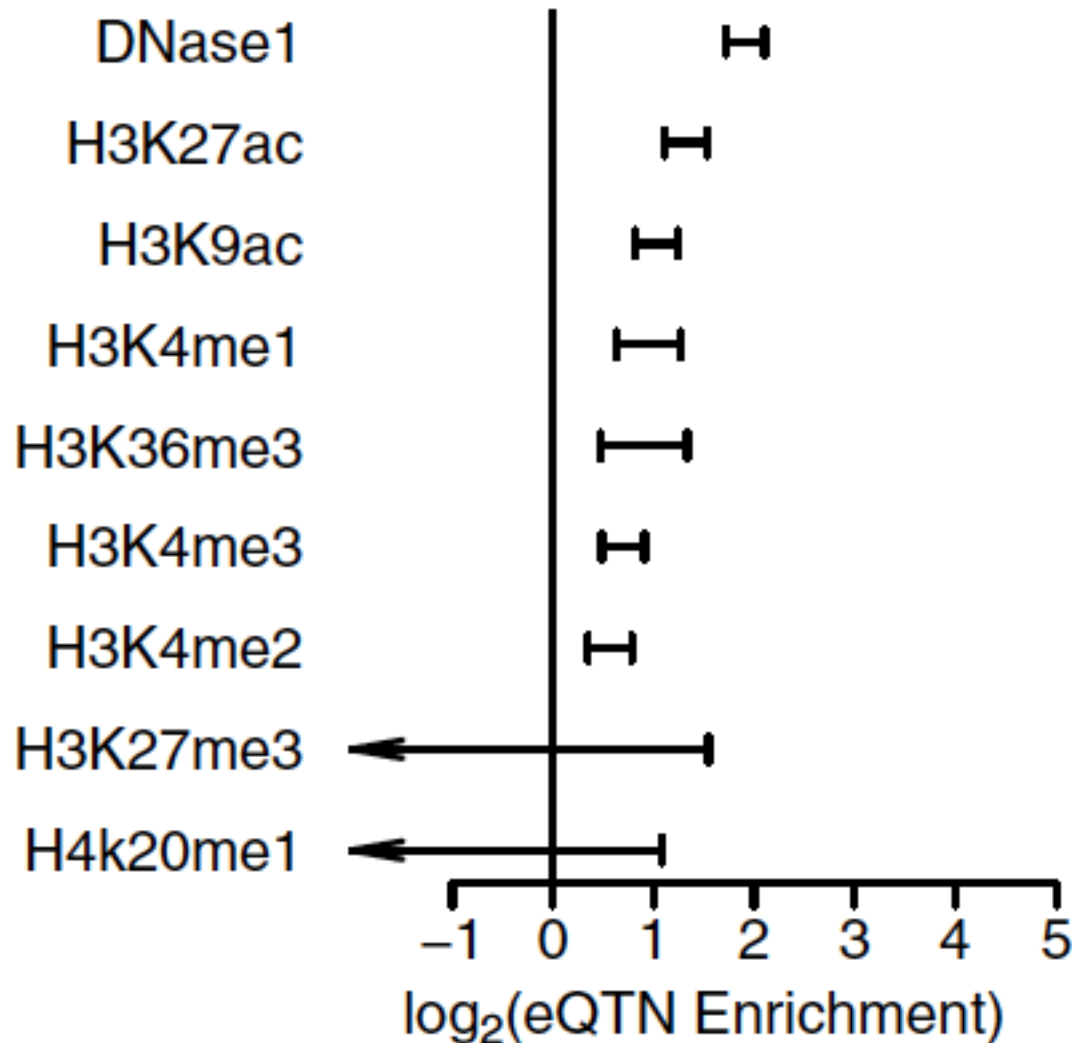




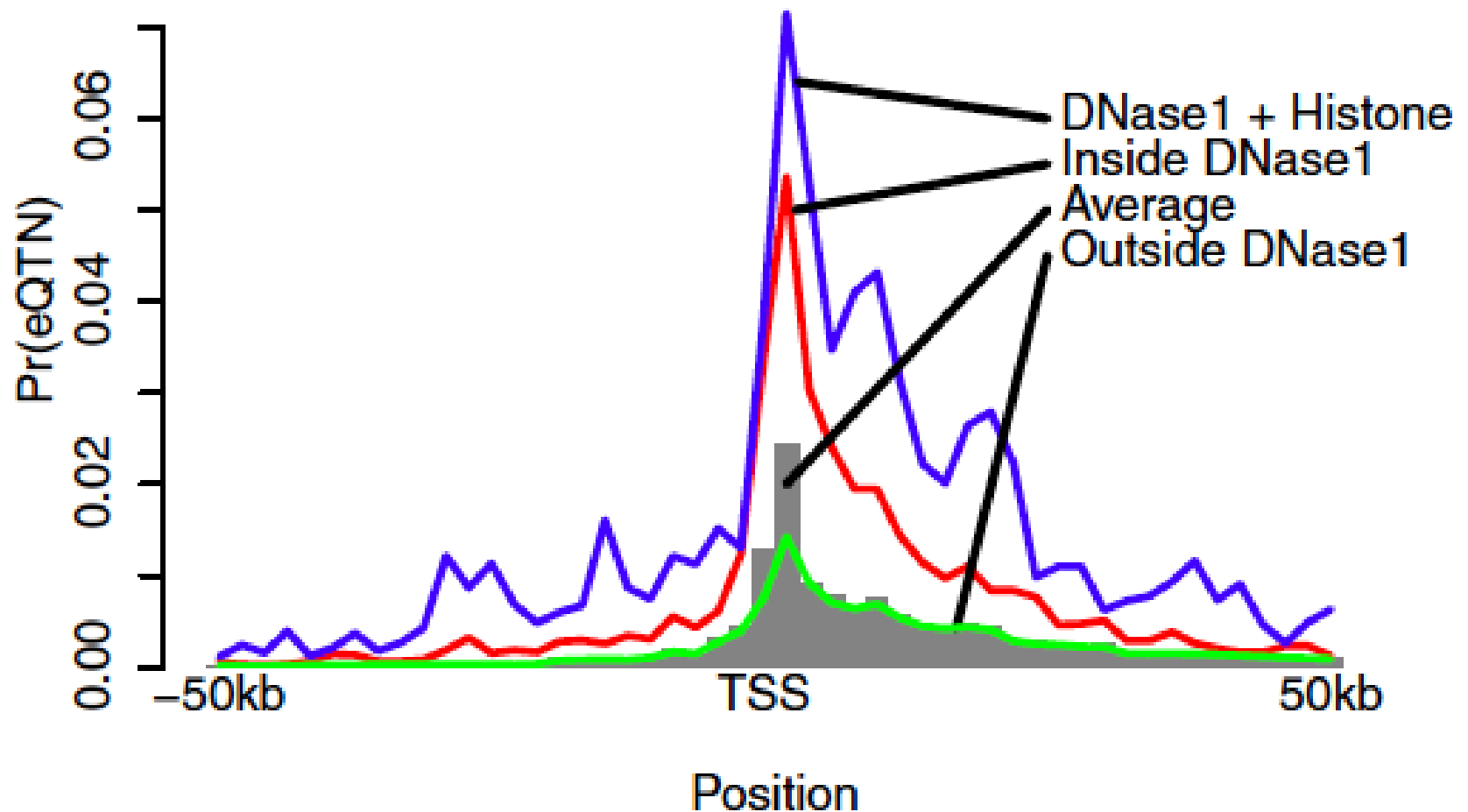
# eQTLs are enriched near transcription start sites



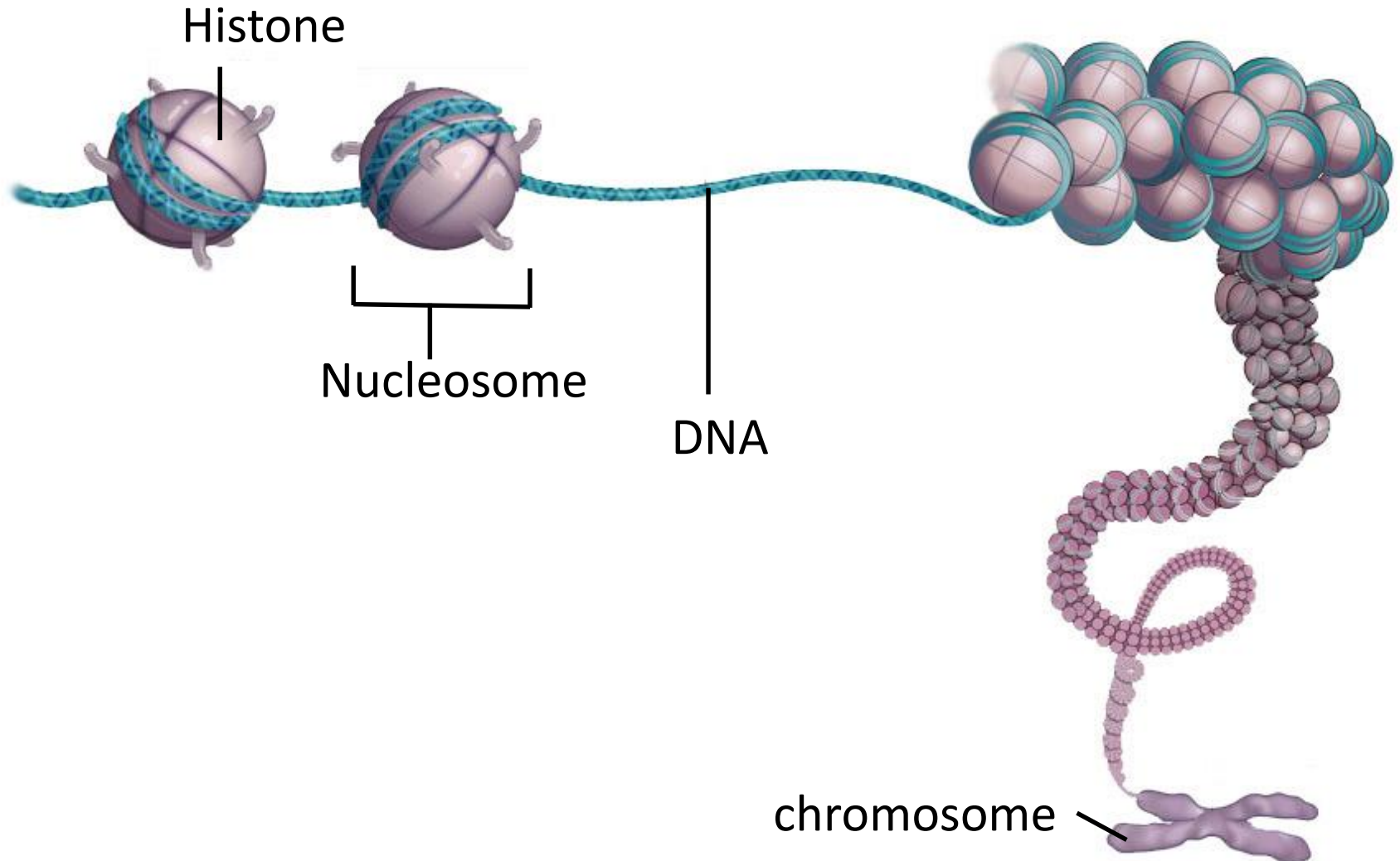
# eQTLs are enriched in regions with open/active chromatin



# Distance from TSS and DNase sensitivity are predictive of eQTLs

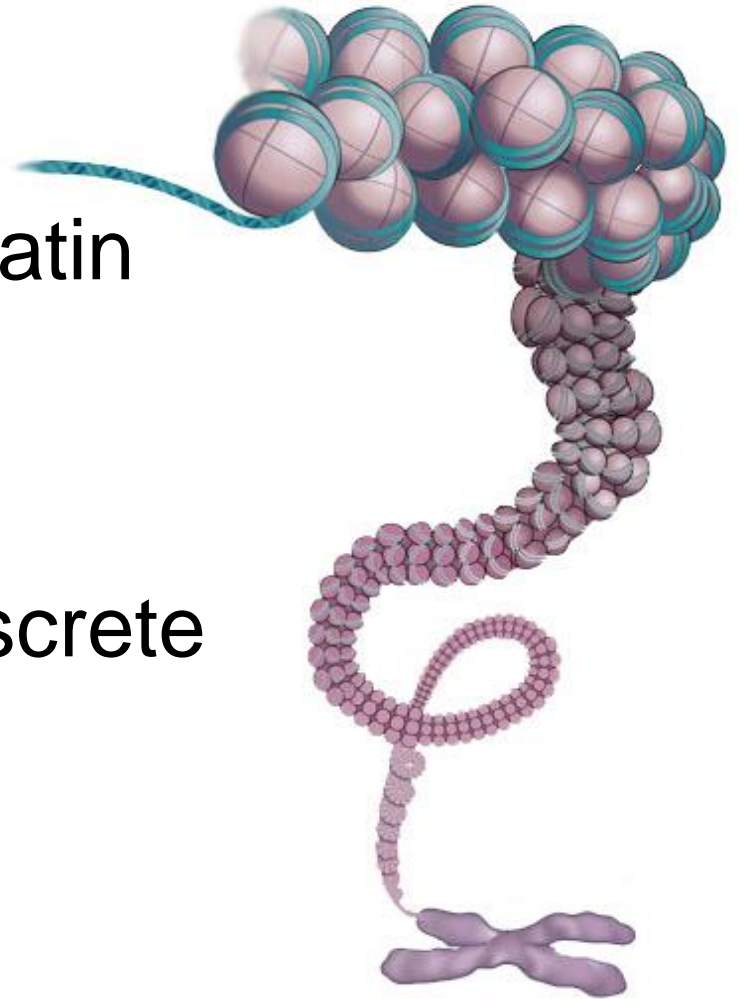


# Chromatin as a molecular trait

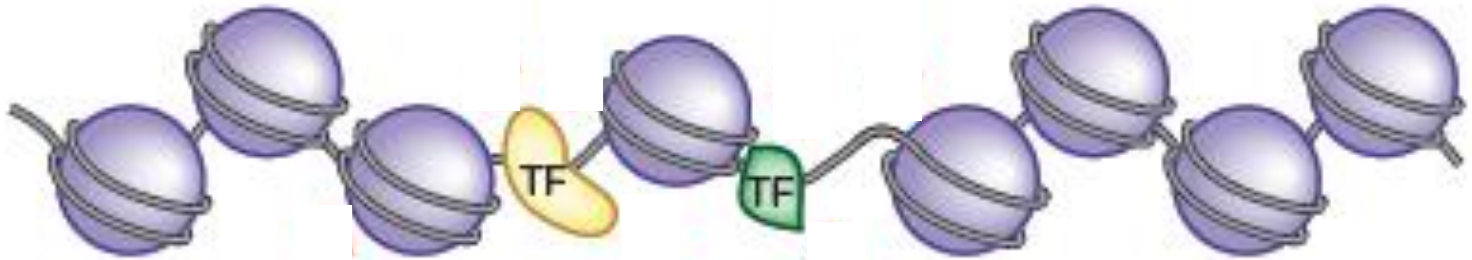


# Discussion

- What aspects of chromatin could be treated as a molecular trait?
- Do these traits have discrete or continuous values?

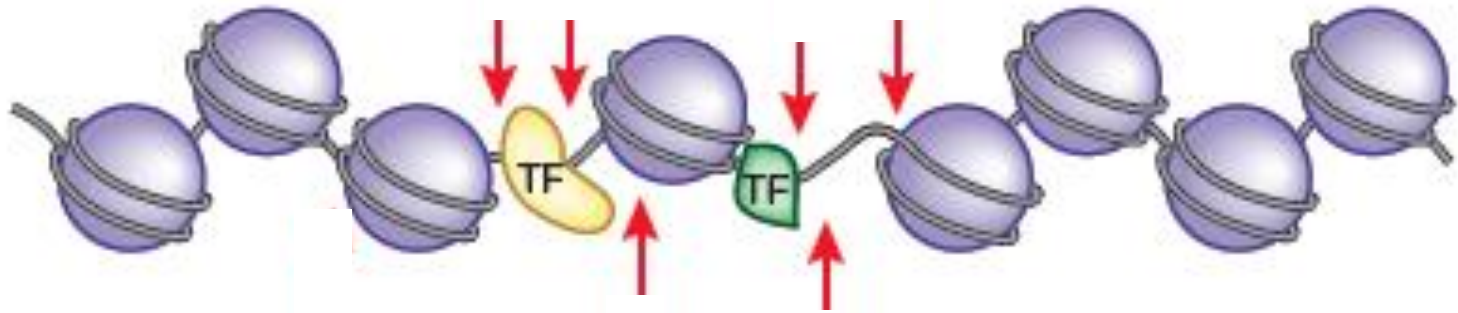


# Nucleosomes are depleted in regulatory regions



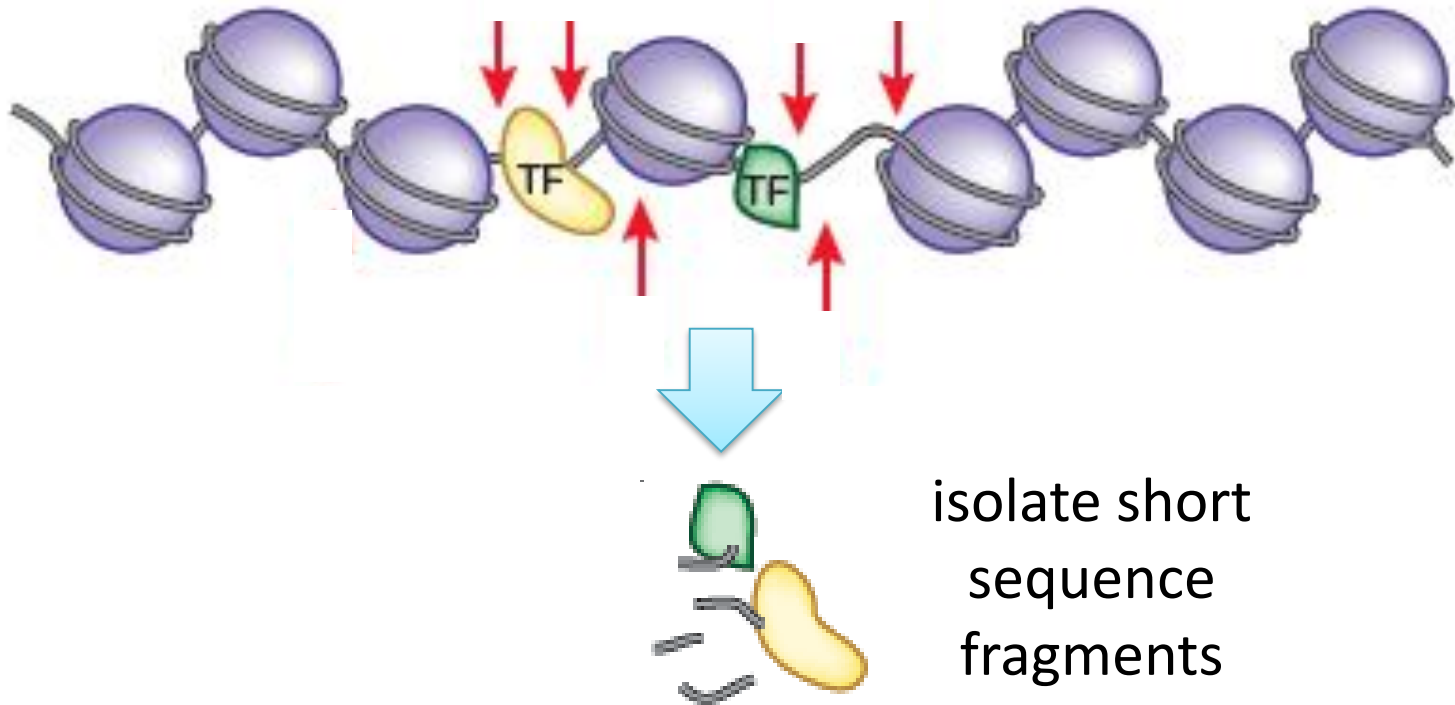


# Measuring chromatin accessibility with DNase-seq

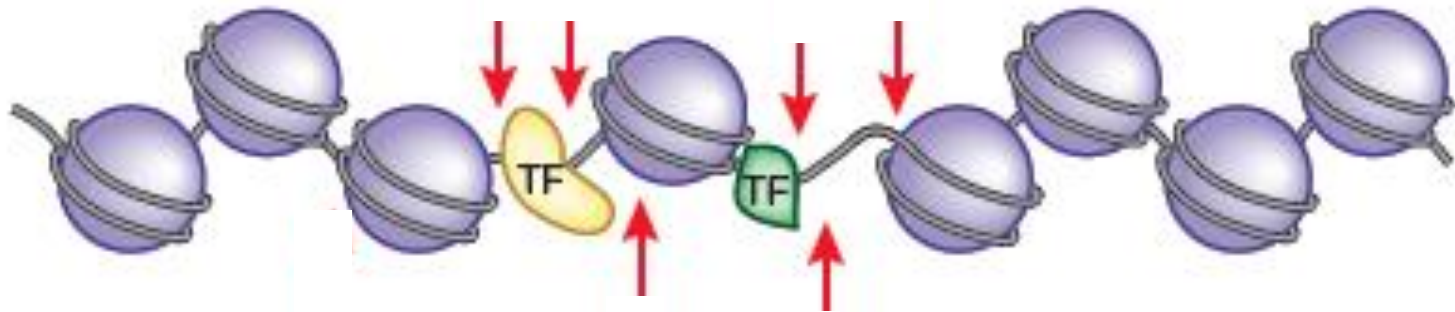


digest chromatin with Deoxyribonuclease 1 (DNase1)

# Measuring chromatin accessibility with DNase-seq



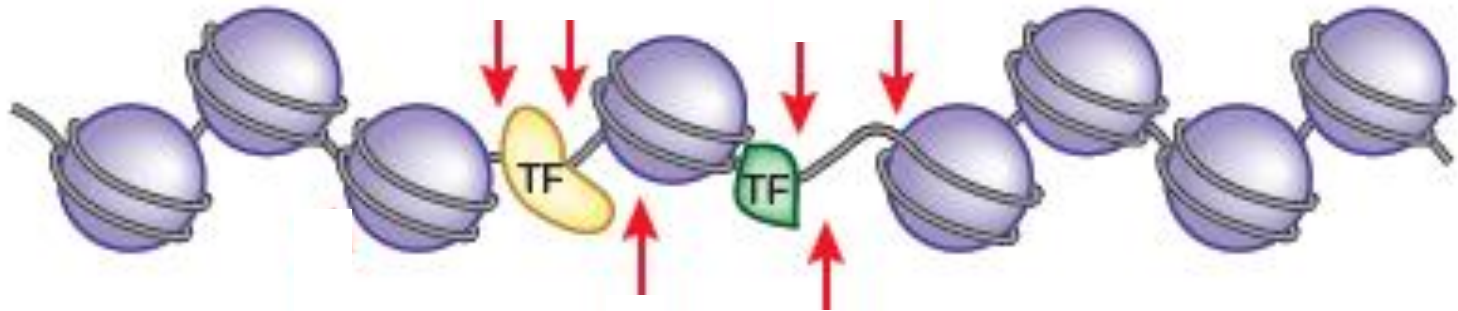
# Measuring chromatin accessibility with DNase-seq



sequence  
DNA fragments

TTGTTAAGGTTTATGTTTCGTTTAATGGTG  
TTTCTTACGACTGTACGATCAAAACGGGG  
AGGCGGCAAGCAGGTGCAGCGTTTTTATA  
GGGCTACAACACGTTGGTGCACCCAACAC

# Measuring chromatin accessibility with DNase-seq



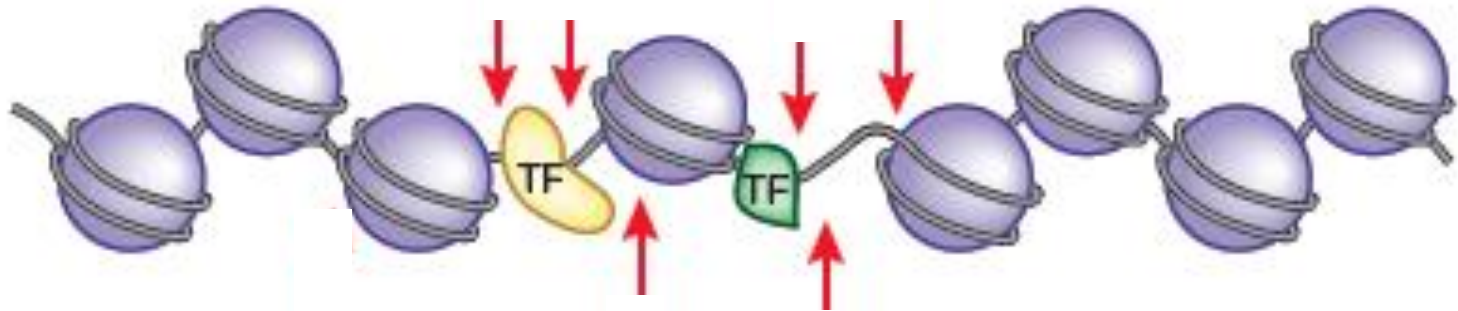
map to genome

```
          GTTCGTTTAATGGTGGCCGGAGGG
          GTTTATGTTTCGTTTAATGGTGCCG
          GTTTATGTTTCGTTTAATGGTGCCG
          AAGGTTTATGTTTCGTTTAATGGTG
          TTGTTAAGGTTTATGTTTCGTTTAA
...ACCATTGTTAAGGTTTATGTTTCGTTTAATGGTGCCGGGAGGGATTTA...
```

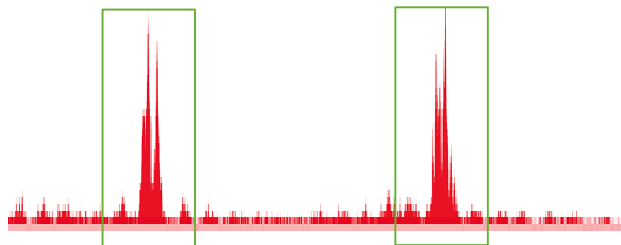


```
TTGTTAAGGTTTATGTTTCGTTTAATGGTG
TTTCTTACGACTGTACGATCAAAACGGGG
AGGCGGCAAGCAGGTGCAGCGTTTTTATA
GGGCTACAACACGTTGGTGCACCCAACAC
```

# Measuring chromatin accessibility with DNase-seq

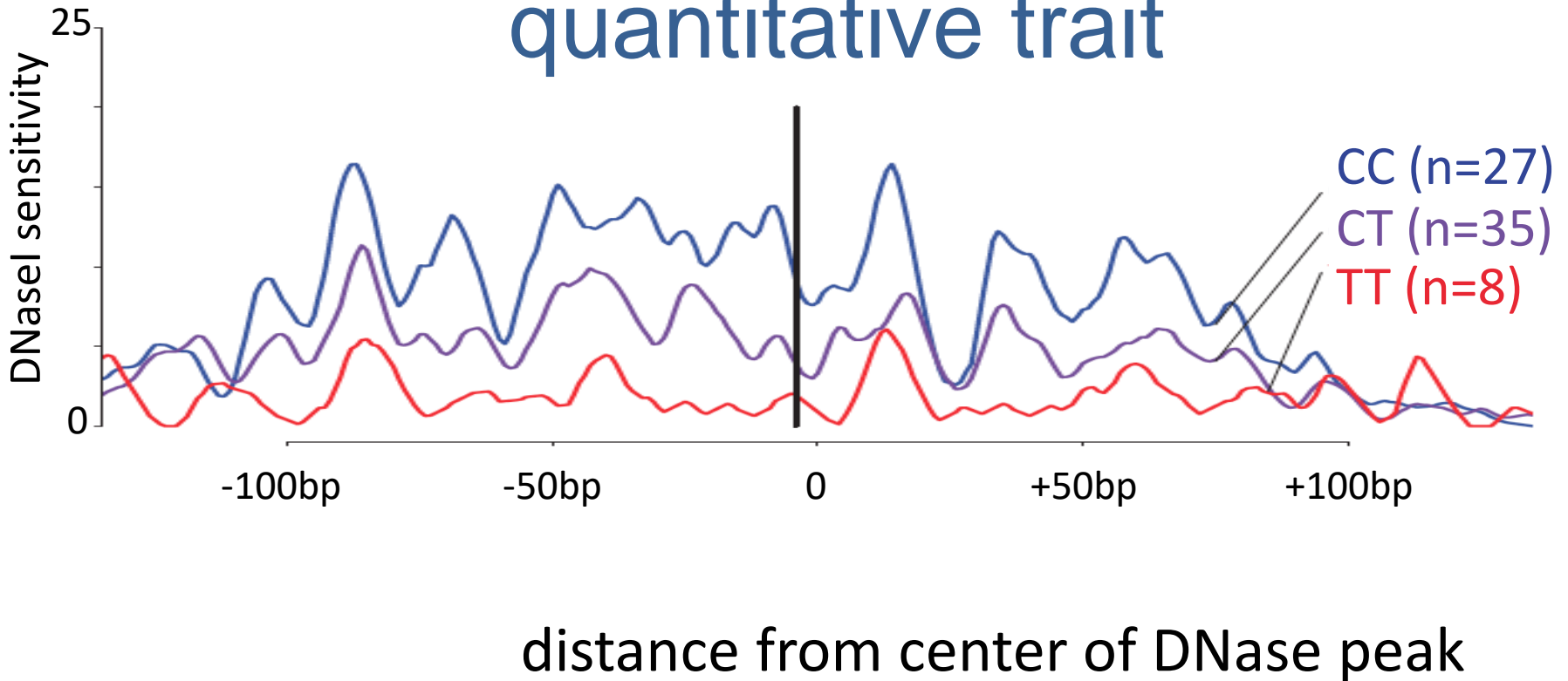


DNase hypersensitive sites



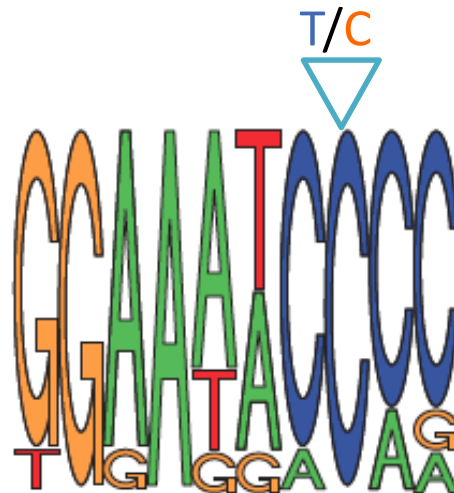
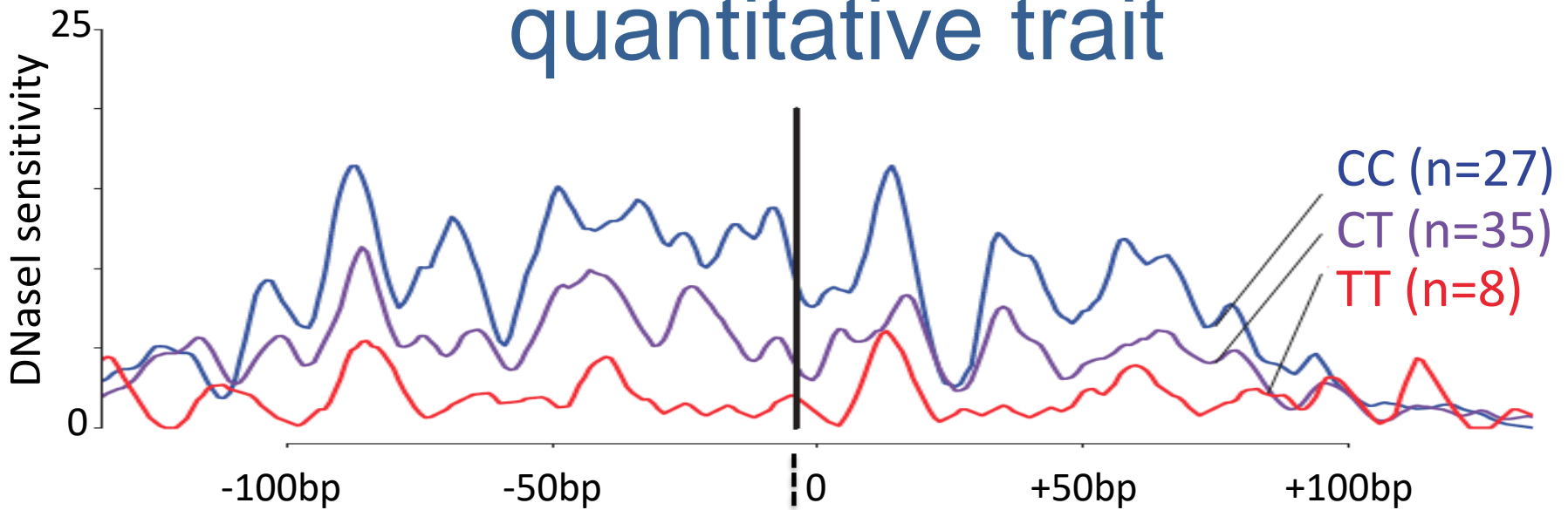
TTGTTAAGGTTTATGTTTCGTTTAATGGTG  
TTTCTTACGACTGTACGATCAAACGGGG  
AGGCGGCAAGCAGGTGCAGCGTTTTTATA  
GGGCTACAACACGTTGGTGCACCCAACAC

# DNase sensitivity as a quantitative trait



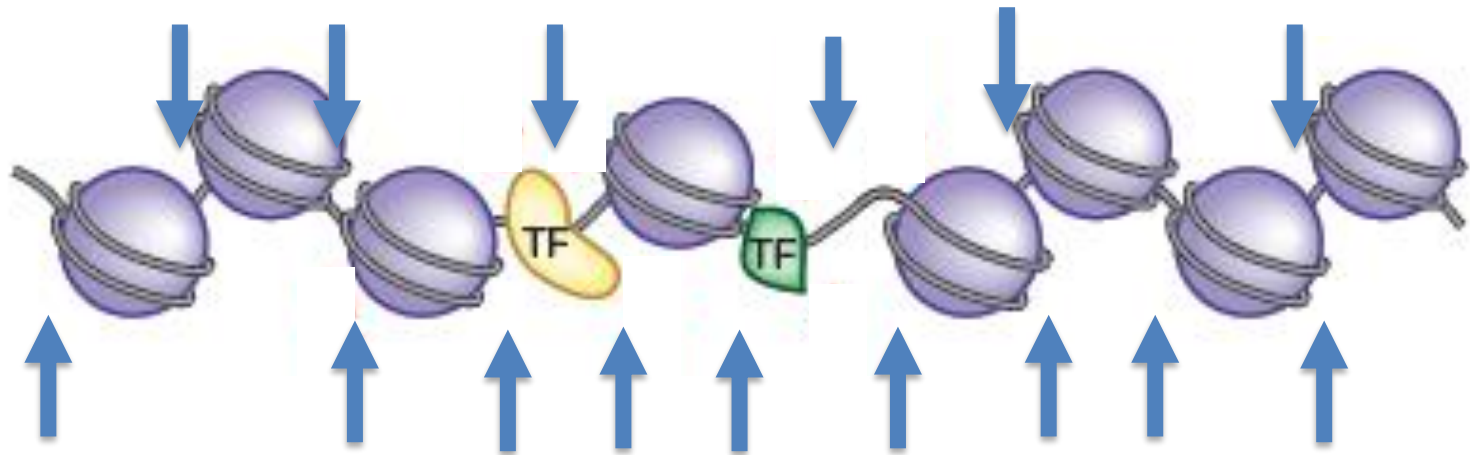


# DNase sensitivity as a quantitative trait



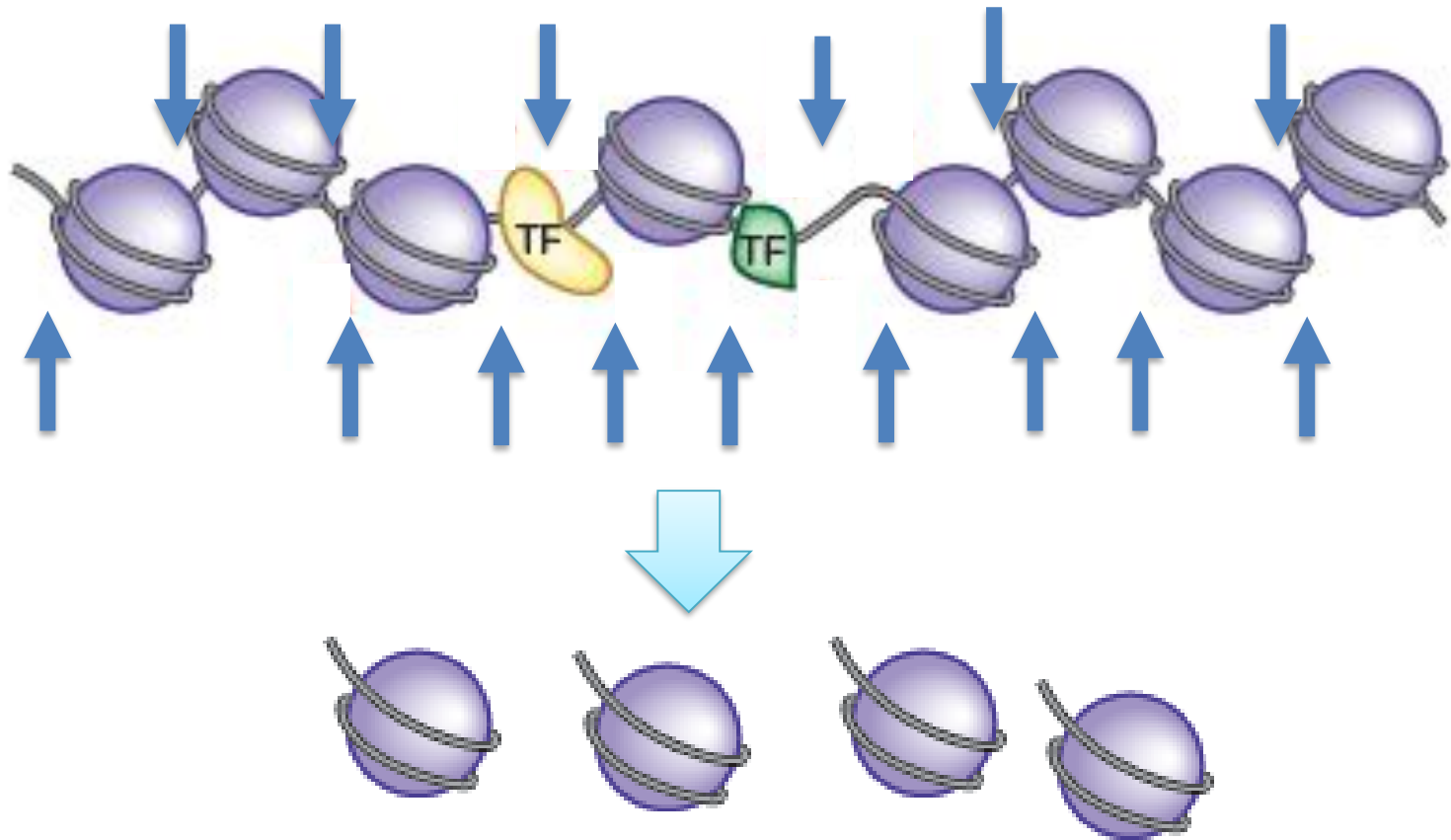
NF-κB motif

# Determining nucleosome positions with MNase-seq



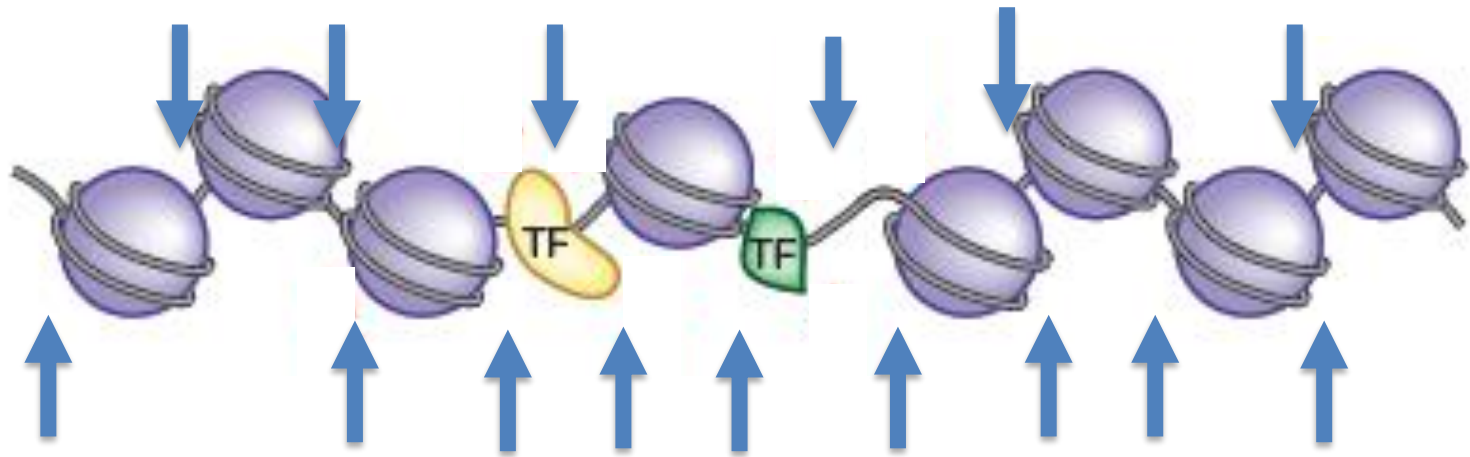
digest chromatin with micrococcal nuclease (MNase)

# Determining nucleosome positions with MNase-seq

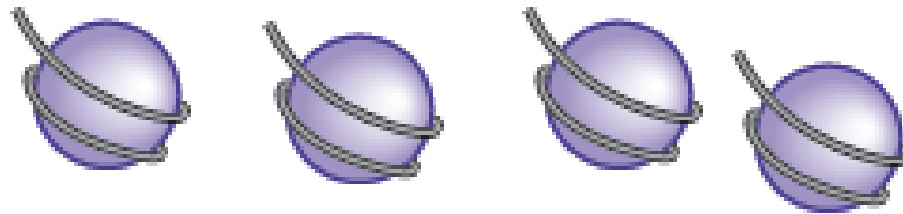


isolate nucleosome-sized fragments and sequence ends

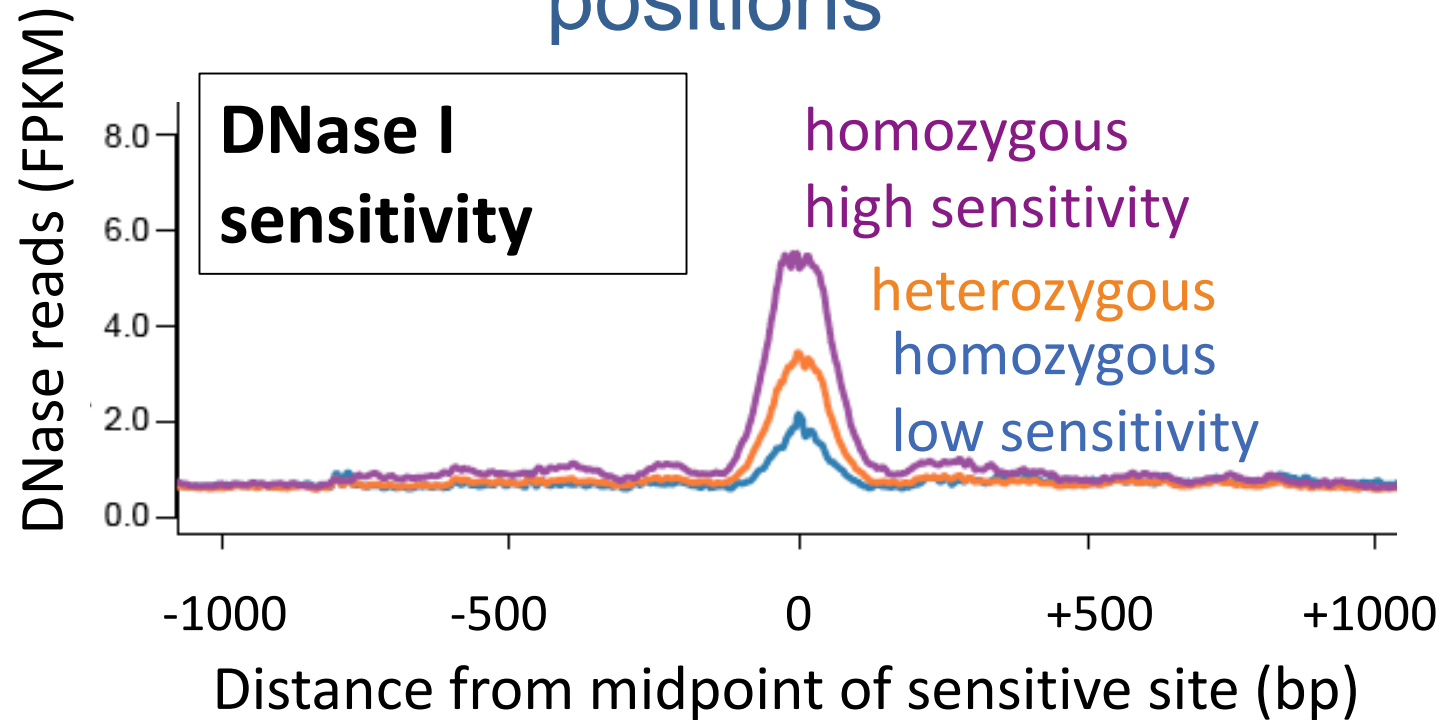
# Determining nucleosome positions with MNase-seq



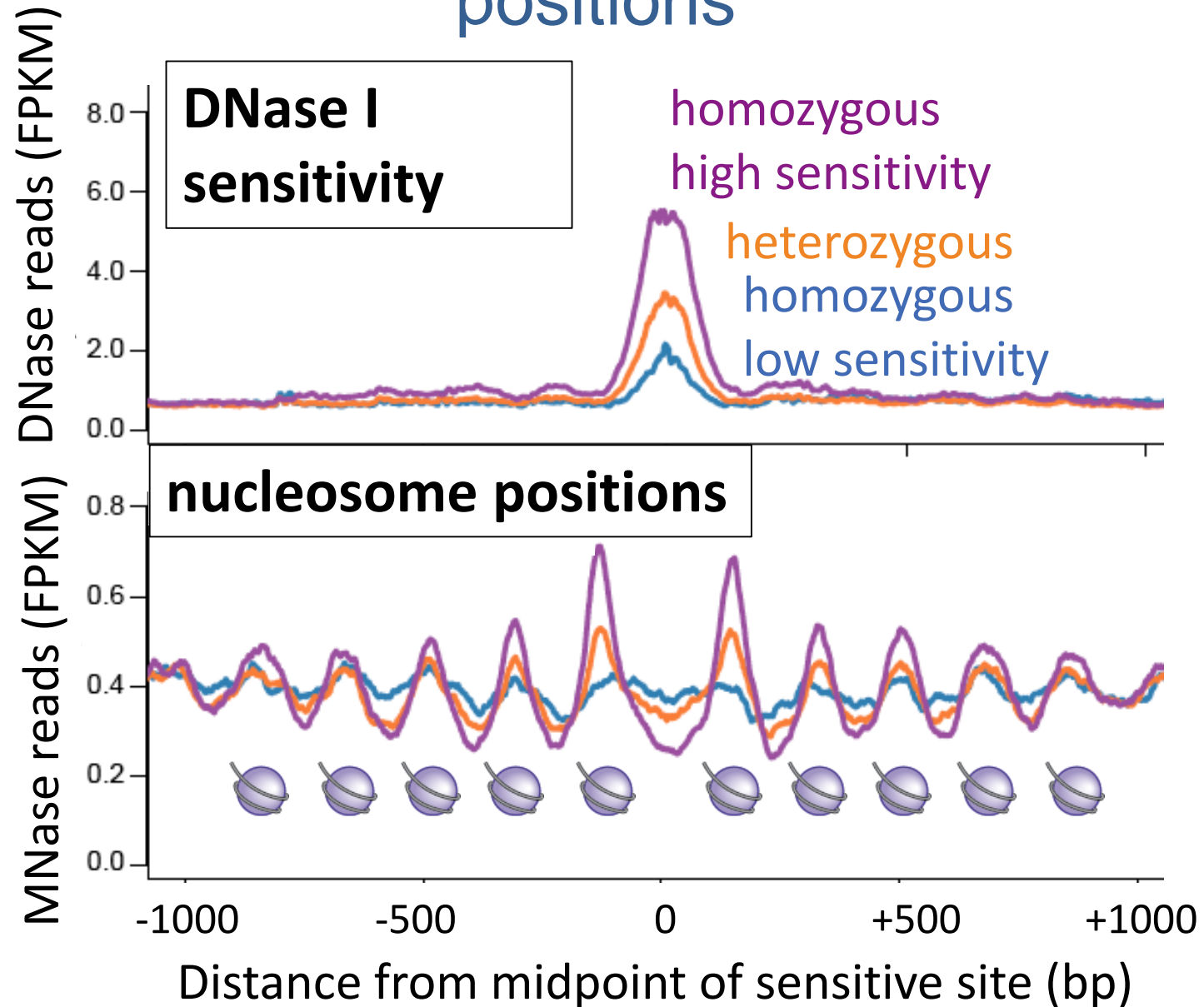
peaks are  
consistently-  
positioned  
nucleosomes



# dsQTLs are associated with nucleosome positions

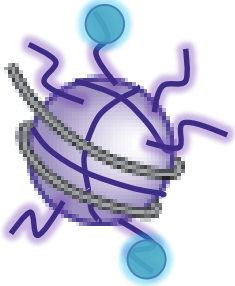


# dsQTLs are associated with nucleosome positions

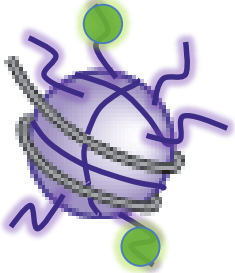




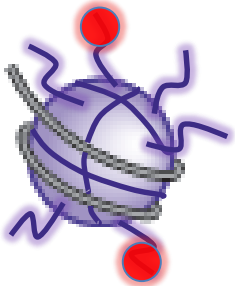
# Histone modifications



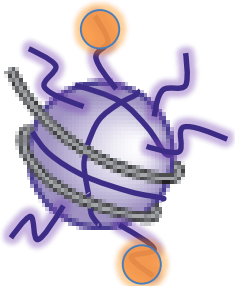
**H3K4me1**: active/open chromatin  
outside of promoters



**H3K4me3**: active promoters

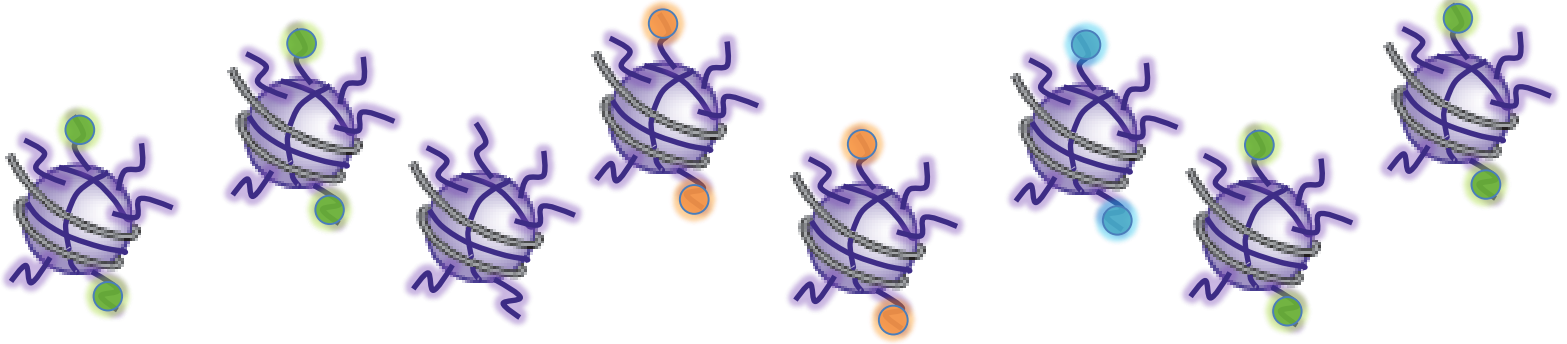


**H3K27ac**: active promoters & enhancers

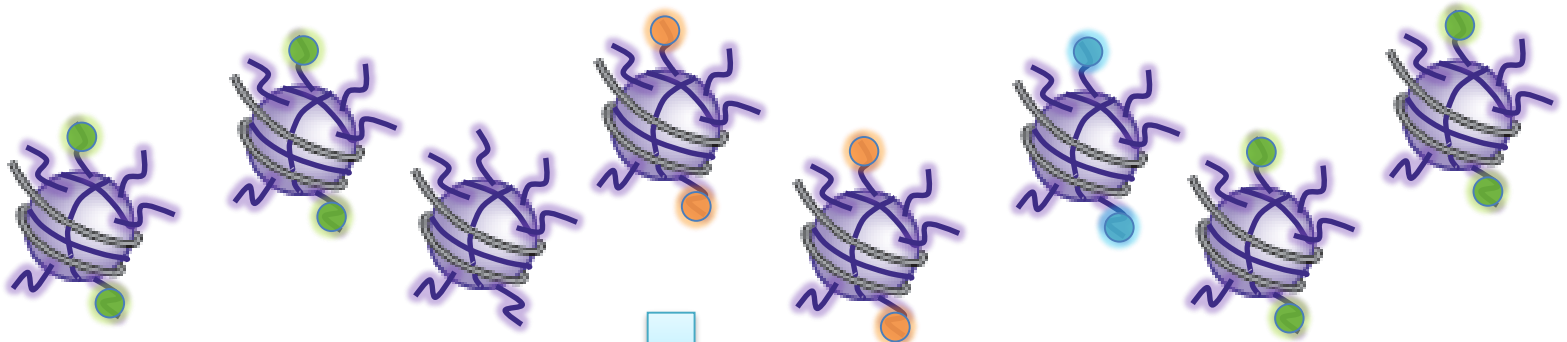


**H3K27me3**: silenced genes

# Measuring histone marks with ChIP-seq

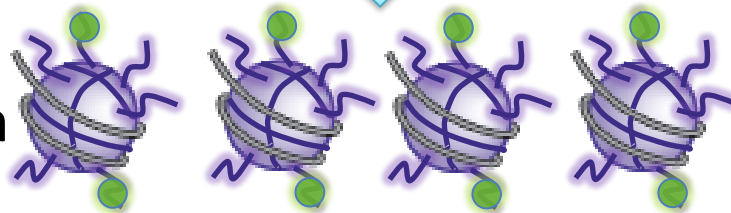


# Measuring histone marks with ChIP-seq



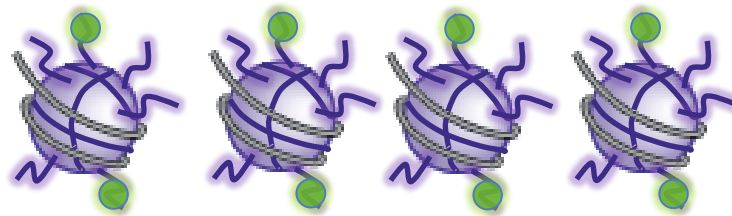
Chromatin

Immunoprecipitation  
(ChIP)



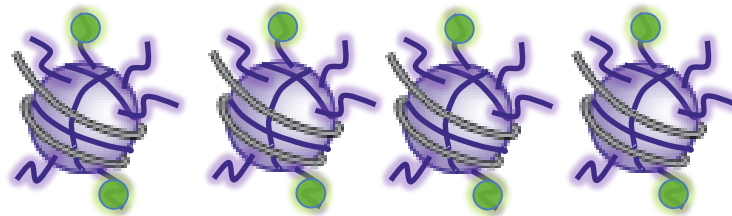
# Measuring histone marks with ChIP-seq

High-throughput  
DNA sequencing



TTGTTAAGGTTTATGTTTCGTTTAATGGTG  
TTTCTTACGACTGTACGATCAAAACGGGG  
AGGCGGCAAGCAGGTGCAGCGTTTTTATA  
GGGCTACAACACGTTGGTGCACCCAACAC

# Measuring histone marks with ChIP-seq



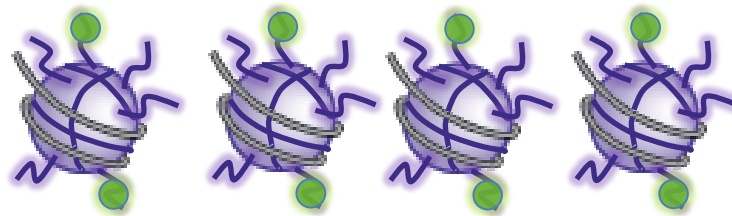
Map reads to  
genome

GTTCGTTTAATGGTGGCCGGAGGG  
GTTTATGTTTCGTTTAATGGTGCCG  
GTTTATGTTTCGTTTAATGGTGCCG  
AAGGTTTATGTTTCGTTTAATGGTG  
TTGTTAAGGTTTATGTTTCGTTTAA  
...ACCATTGTTAAGGTTTATGTTTCGTTTAATGGTGCCGGGAGGGATTTA...



TTGTTAAGGTTTATGTTTCGTTTAATGGTG  
TTTCTTACGACTGTACGATCAAACGGGG  
AGGCGGCAAGCAGGTGCAGCGTTTTTATA  
GGGCTACAACACGTTGGTGCACCCAACAC

# Measuring histone marks with ChIP-seq

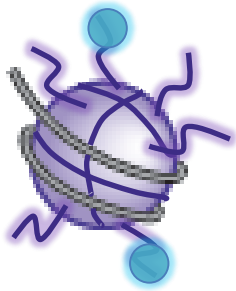


Identify ChIP-seq peaks

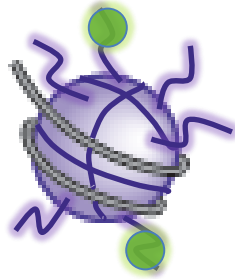


TTGTTAAGGTTTATGTTTCGTTTAATGGTG  
TTTCTTACGACTGTACGATCAAACGGGG  
AGGCGGCAAGCAGGTGCAGCGTTTTTATA  
GGGCTACAACACGTTGGTGCACCCAACAC

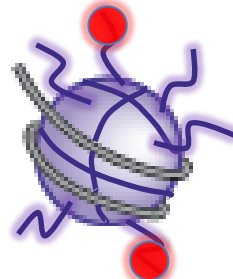
# ChIP-seq data from 10 individuals



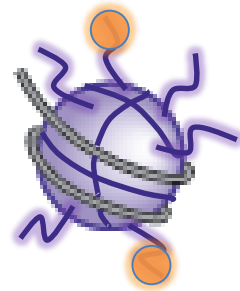
H3K4me1



H3K4me3



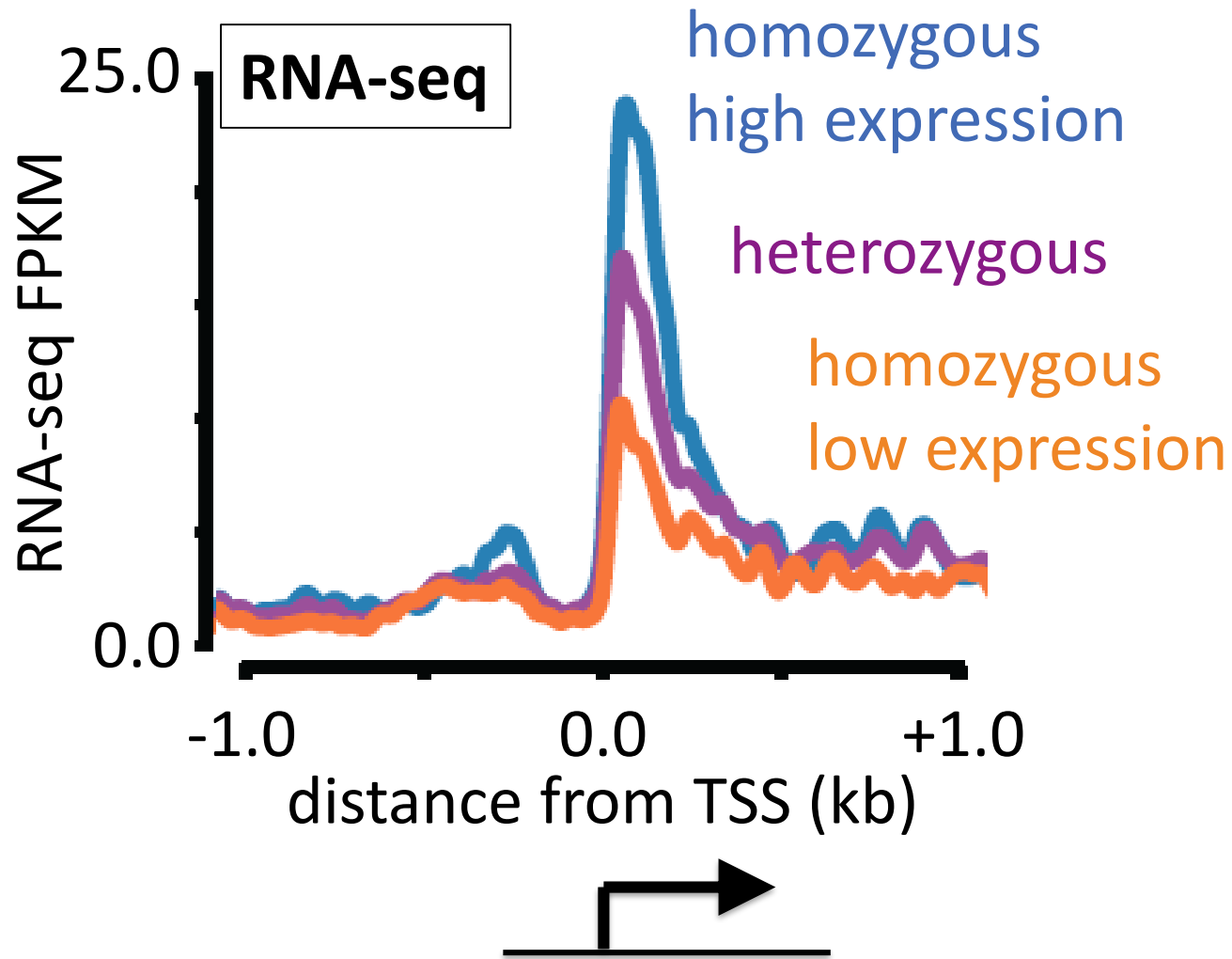
H3K27ac



H3K27me3

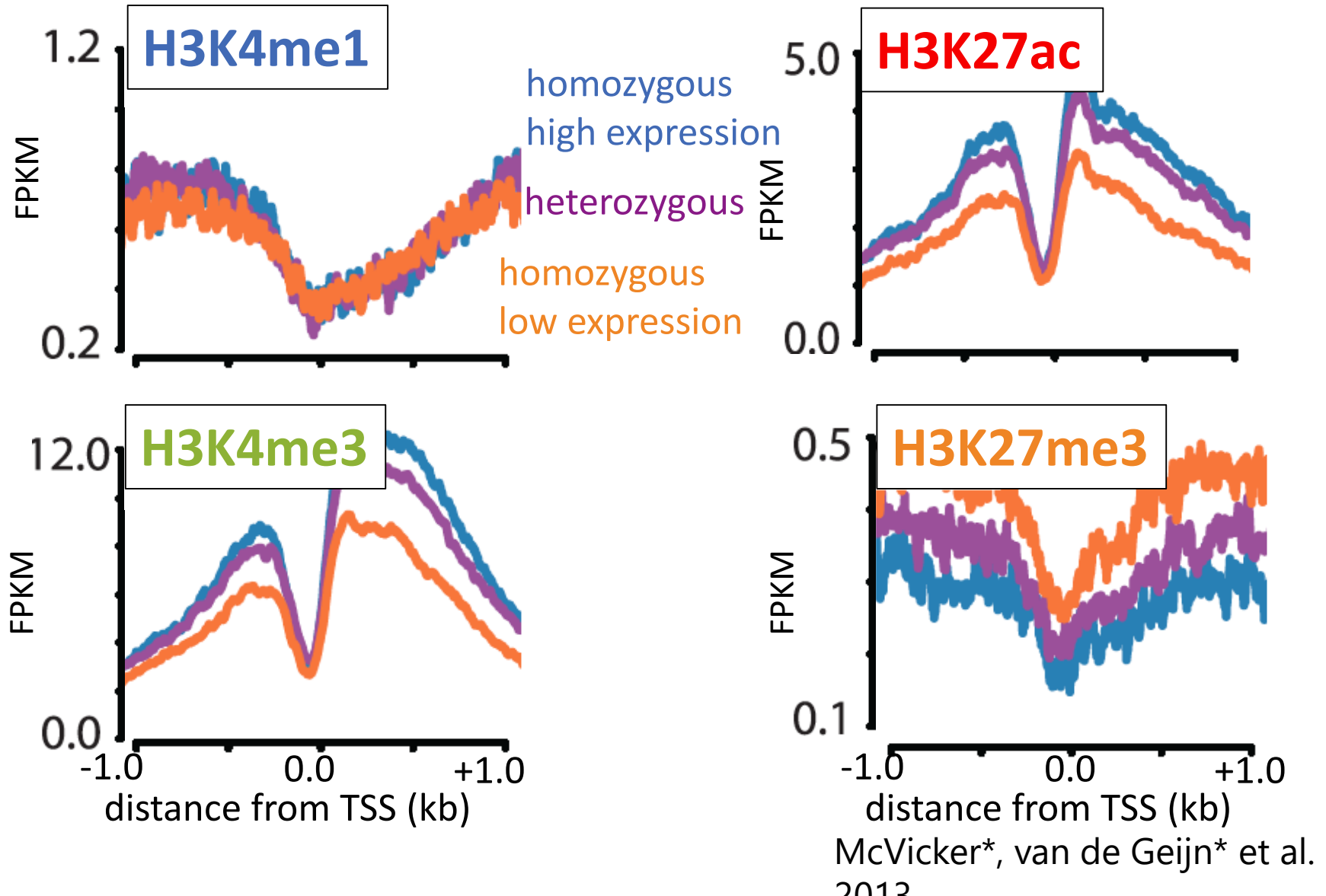


# Are eQTLs also associated with histone modifications?

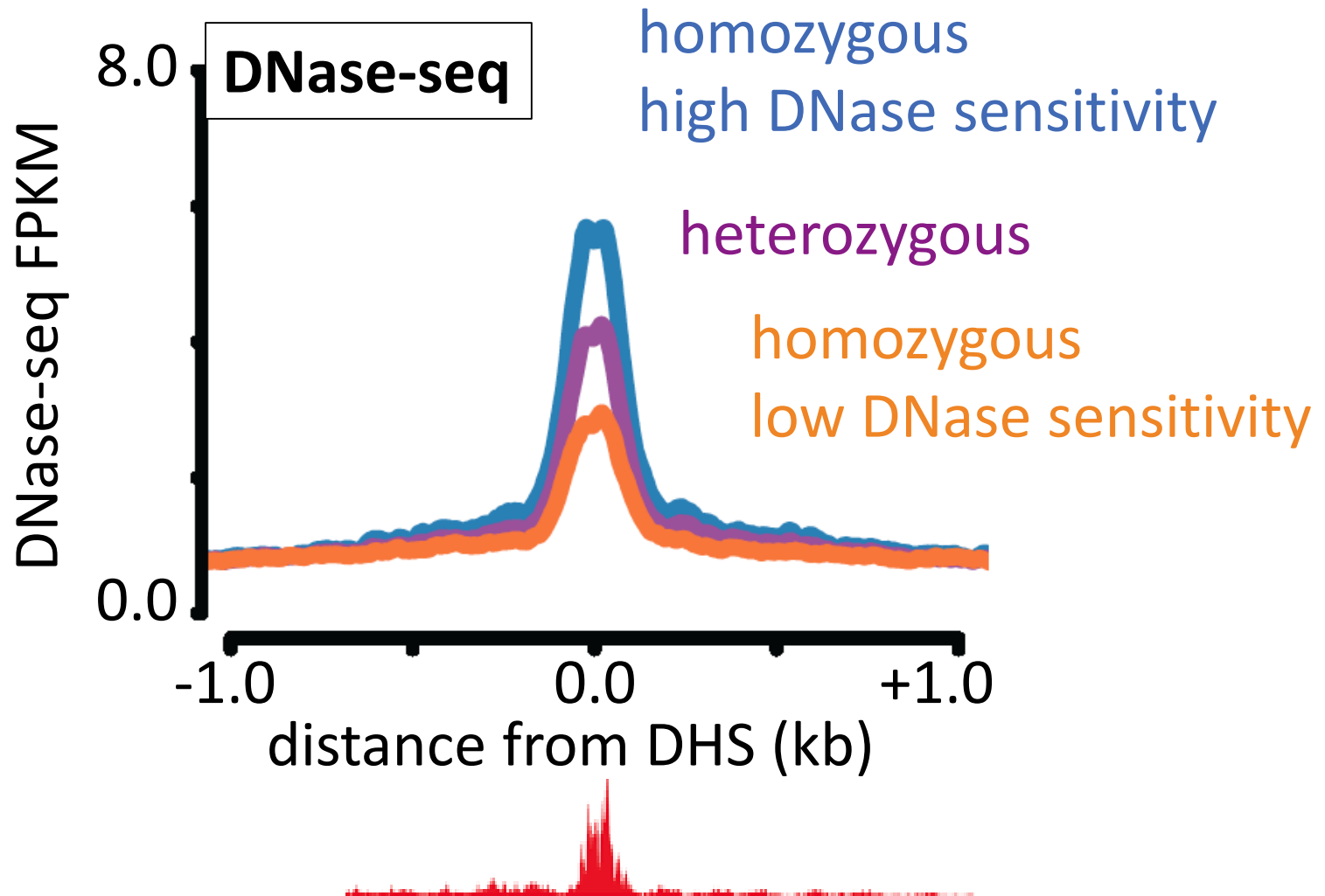




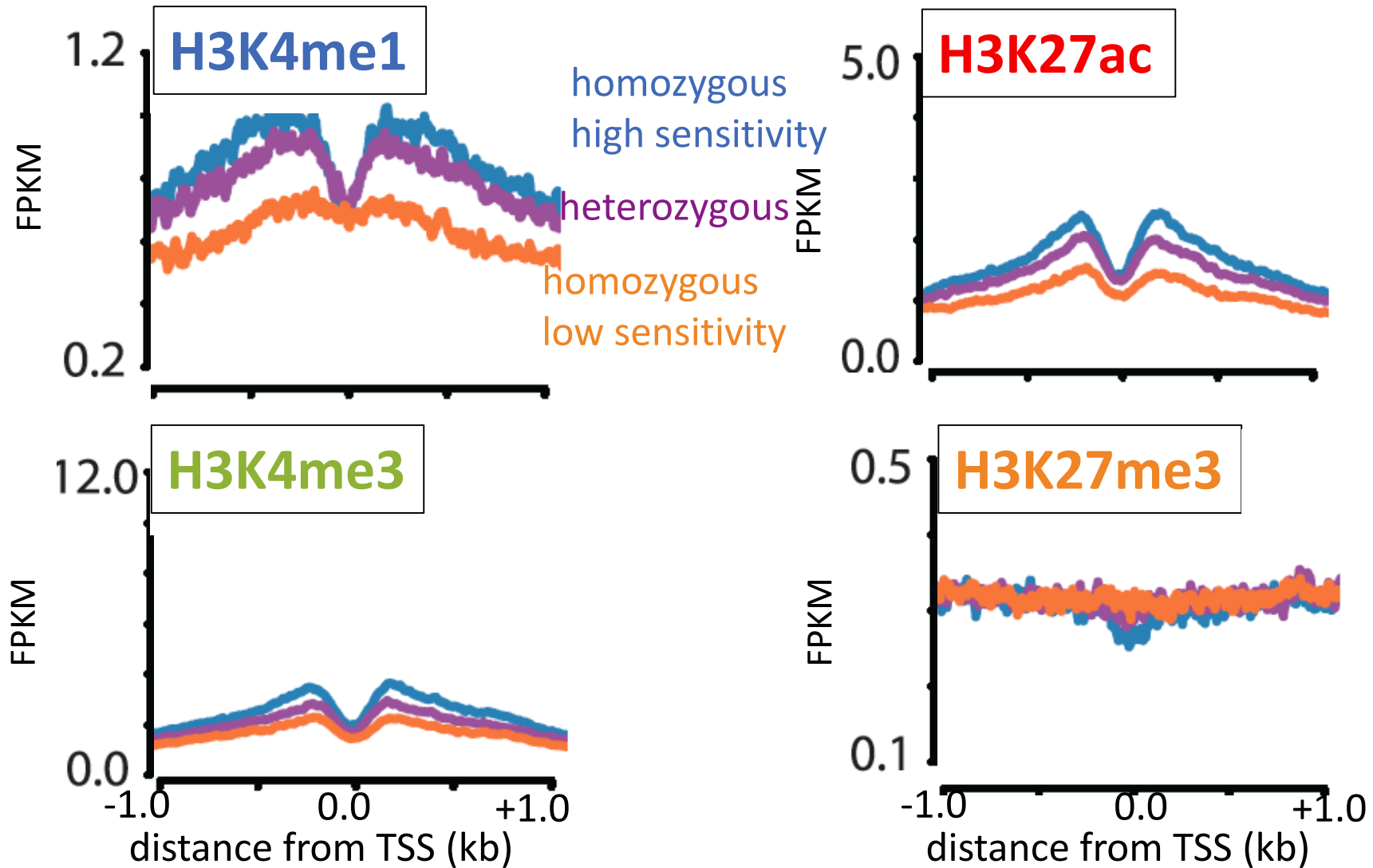
# Are eQTLs also associated with histone modifications?



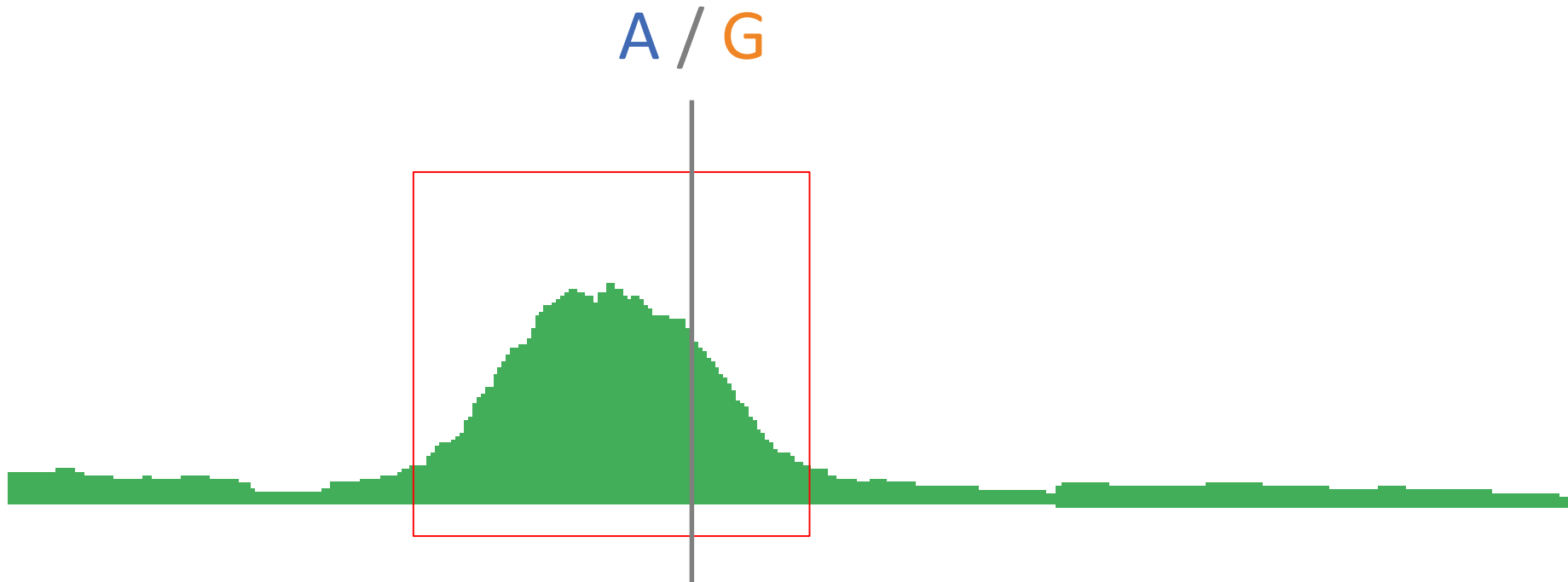
# Are dsQTLs also associated with histone modifications?



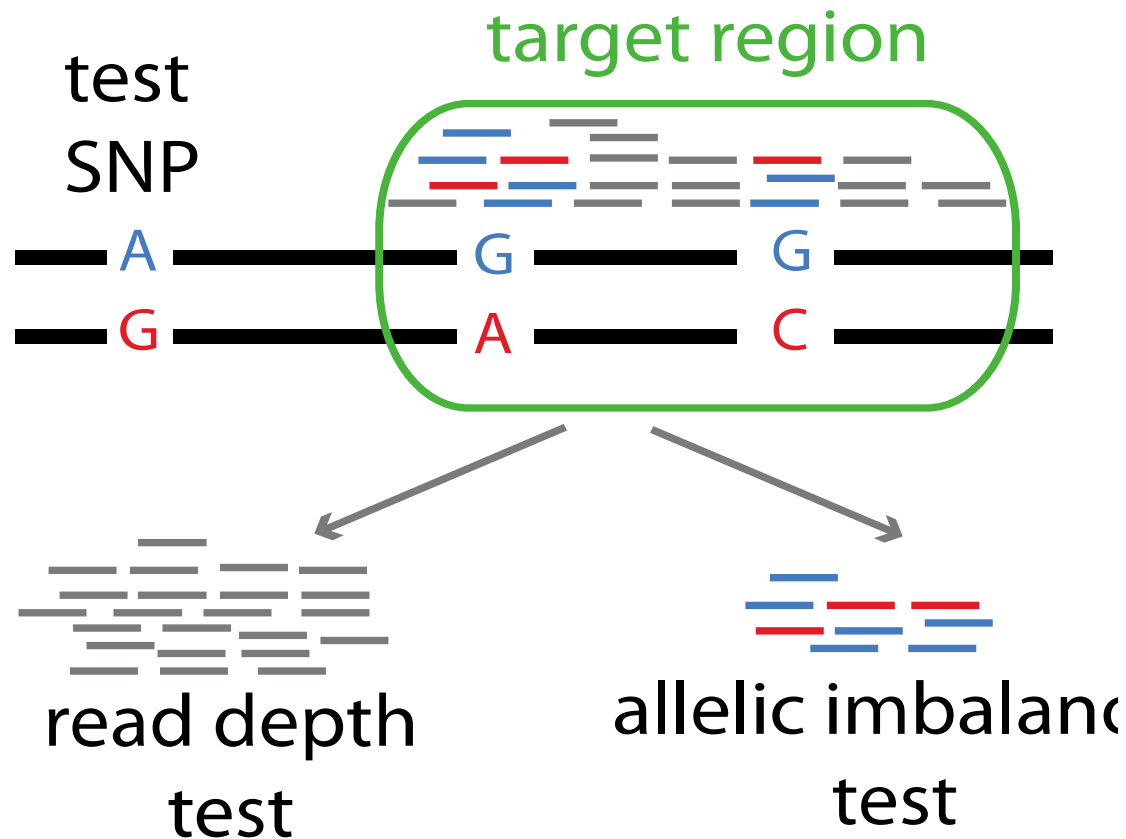
# Are dsQTLs also associated with histone modifications?



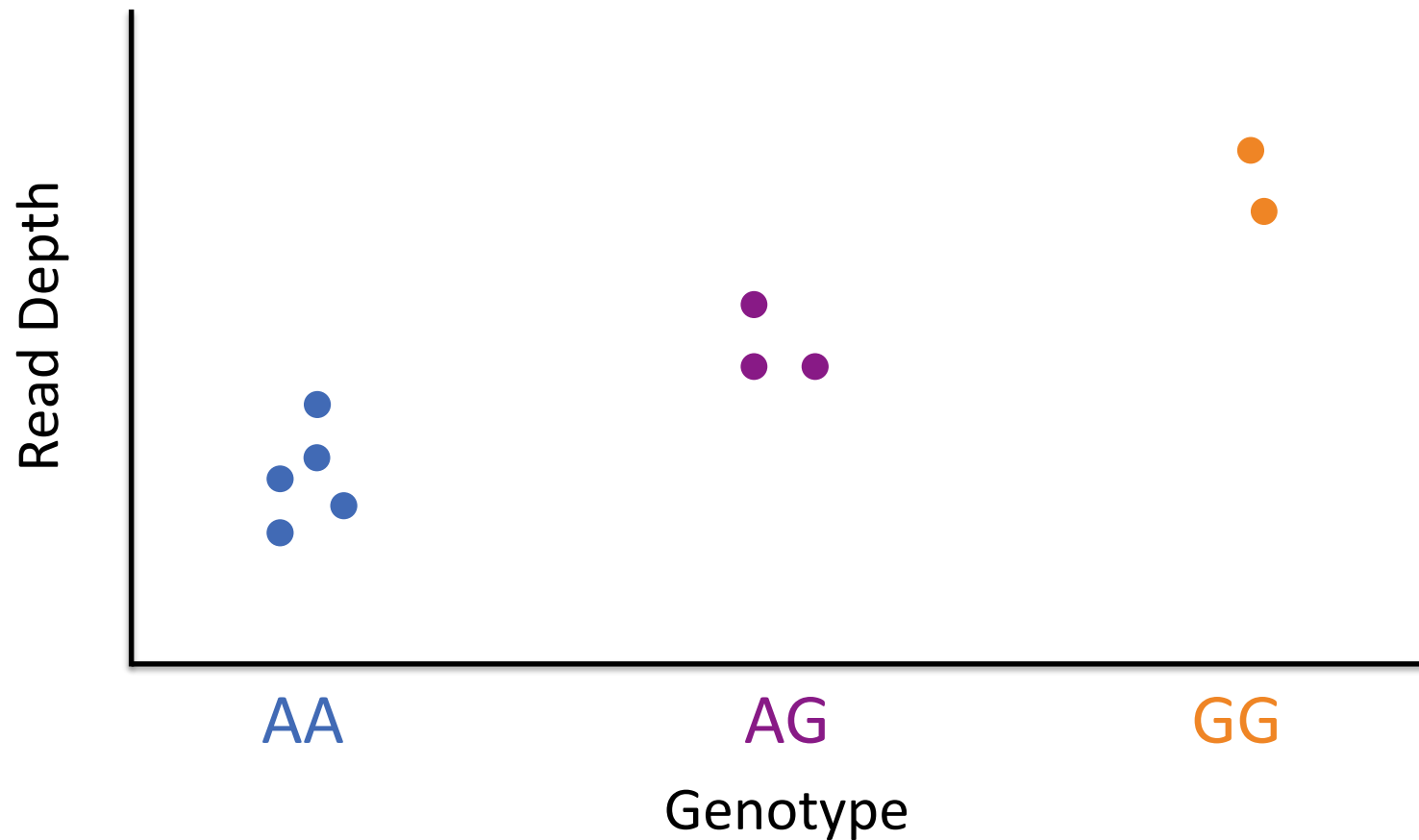
# Can we identify Histone Mark QTLs?



# Combined Haplotype Test



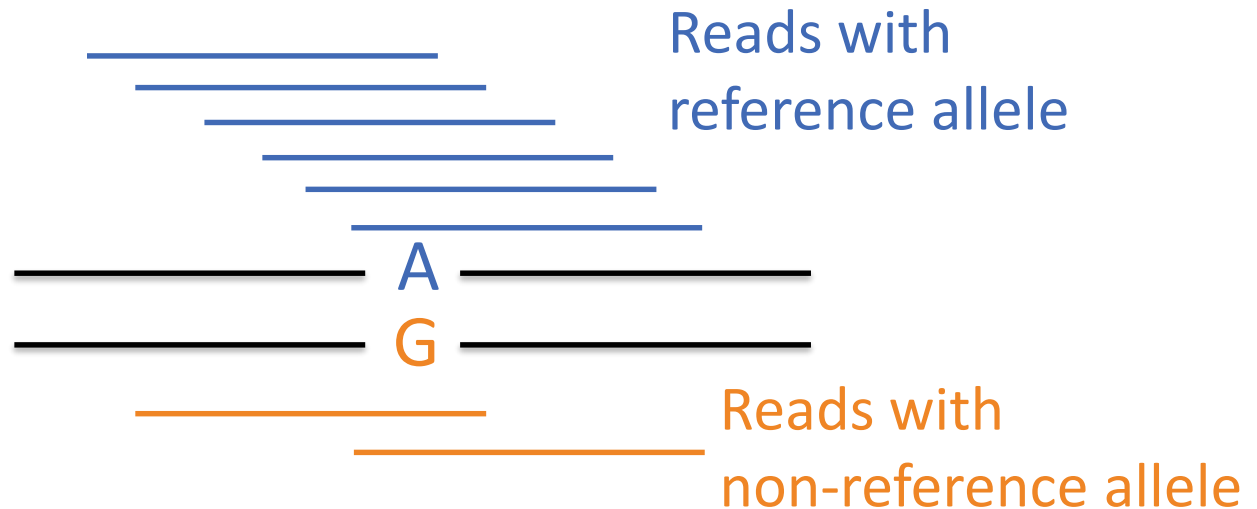
# Read Depth Association Test



Model read counts with Poisson distribution:

$$\lambda = \begin{matrix} 2\alpha & \alpha+\beta & 2\beta \end{matrix}$$

# Allelic Imbalance Test



Model reference proportion with binomial distribution:

$$p = \frac{\alpha}{\alpha + \beta}$$

# Haplotype Imbalance Test



- Phase SNPs
- Test allelic imbalance across entire haplotype



# Combined Haplotype Test

Read Depth Association Test

$$\lambda = \begin{cases} 2\alpha & \text{if homozygous AA} \\ \alpha + \beta & \text{if heterozygous AB} \\ 2\beta & \text{if homozygous BB} \end{cases}$$

+

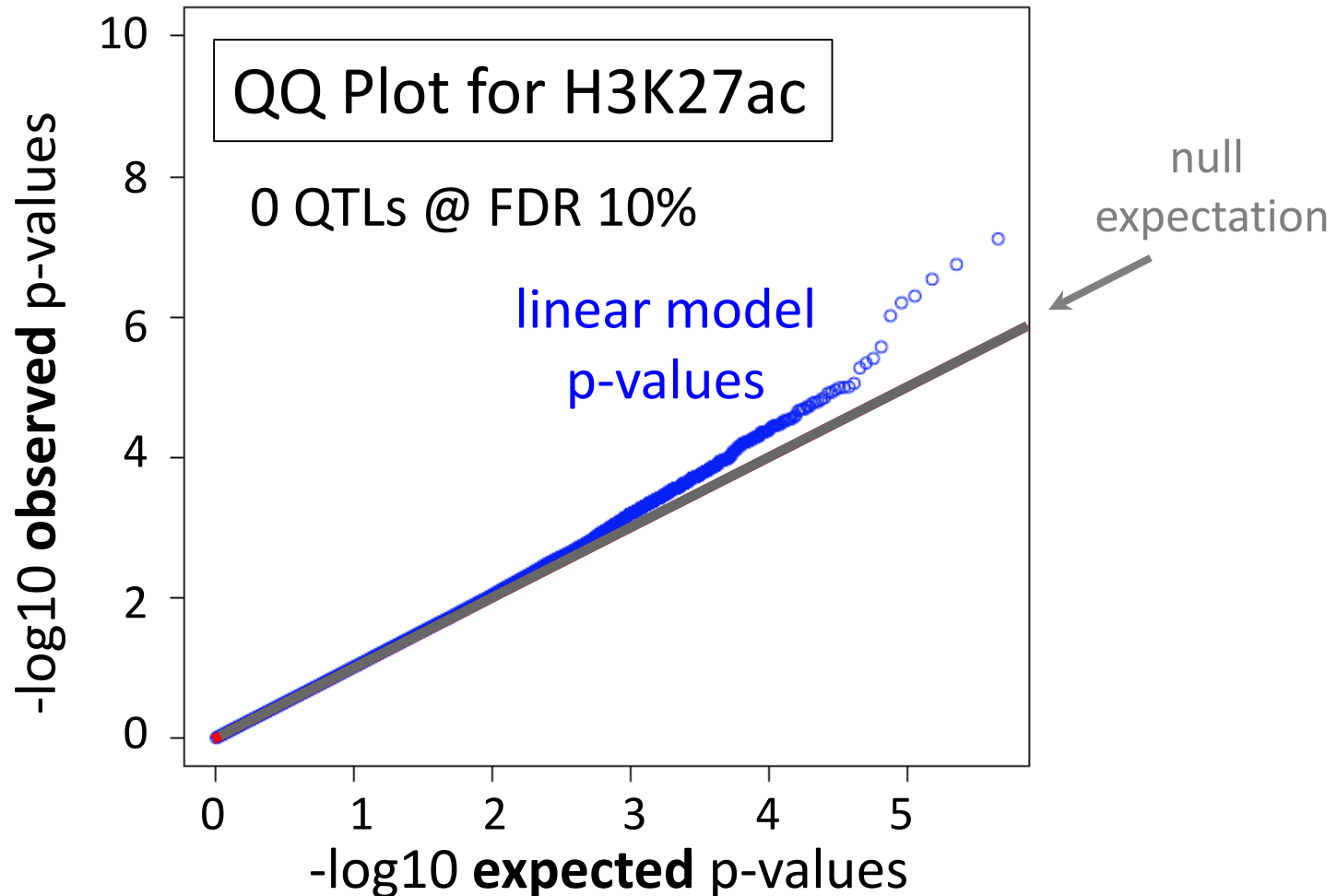
Haplotype Imbalance Test

$$p = \frac{\alpha}{\alpha + \beta}$$

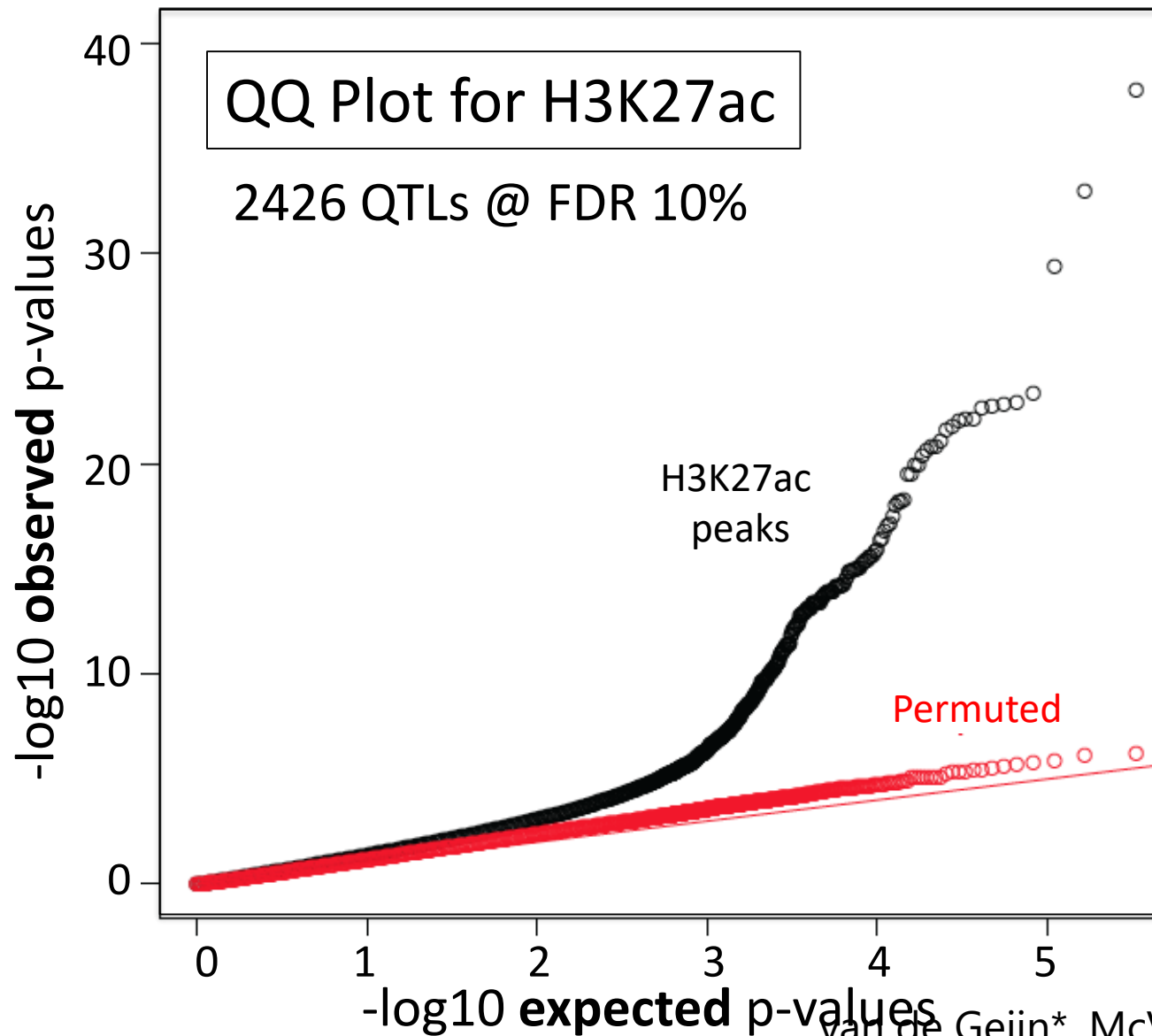
null hypothesis,  $H_0: \alpha = \beta$

alternative hypothesis,  $H_1: \alpha \neq \beta$

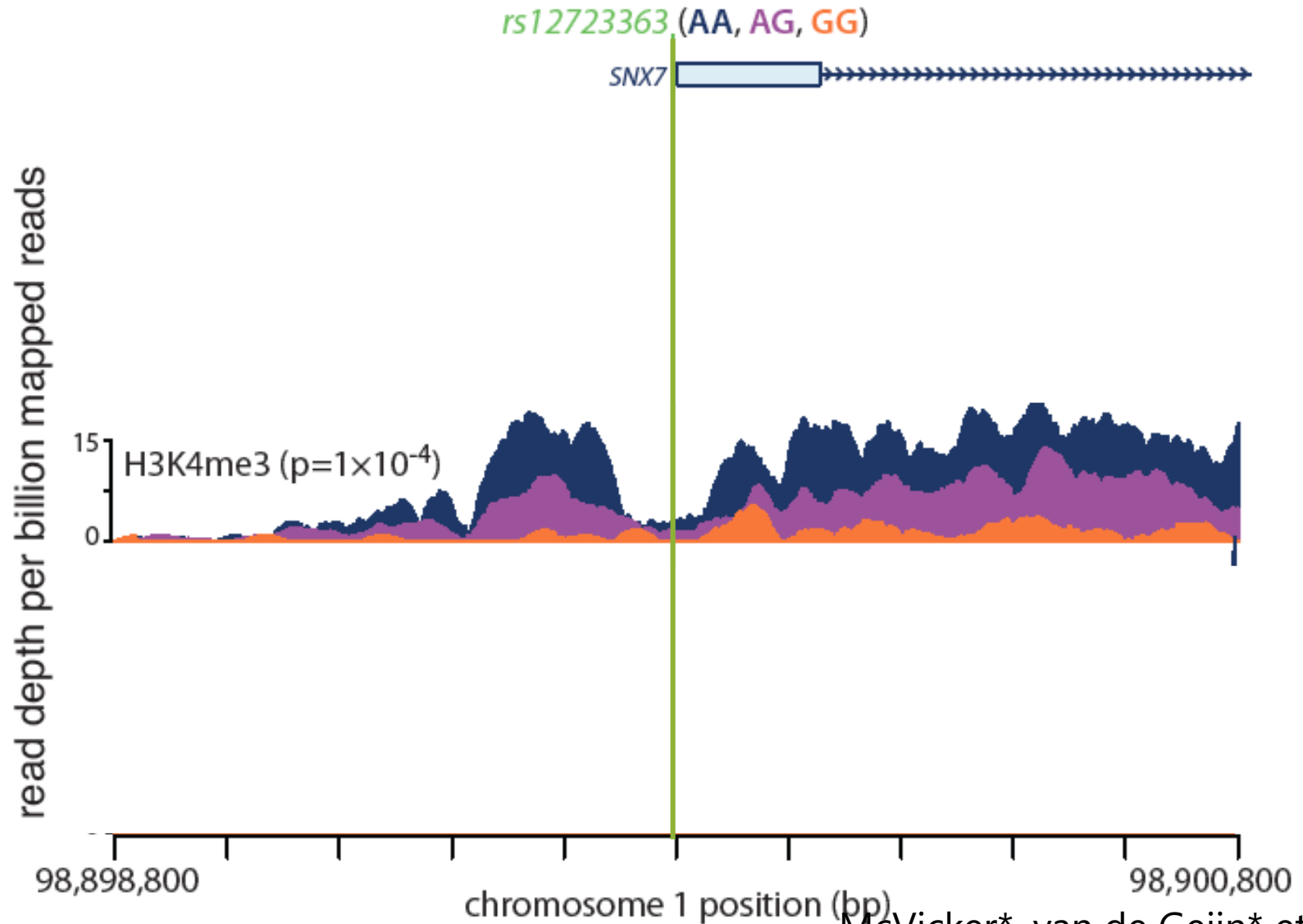
# Standard mapping of H3K27ac QTLs with 10 Individuals



# Combined Haplotype Test Mapping of H3K27ac QTLs

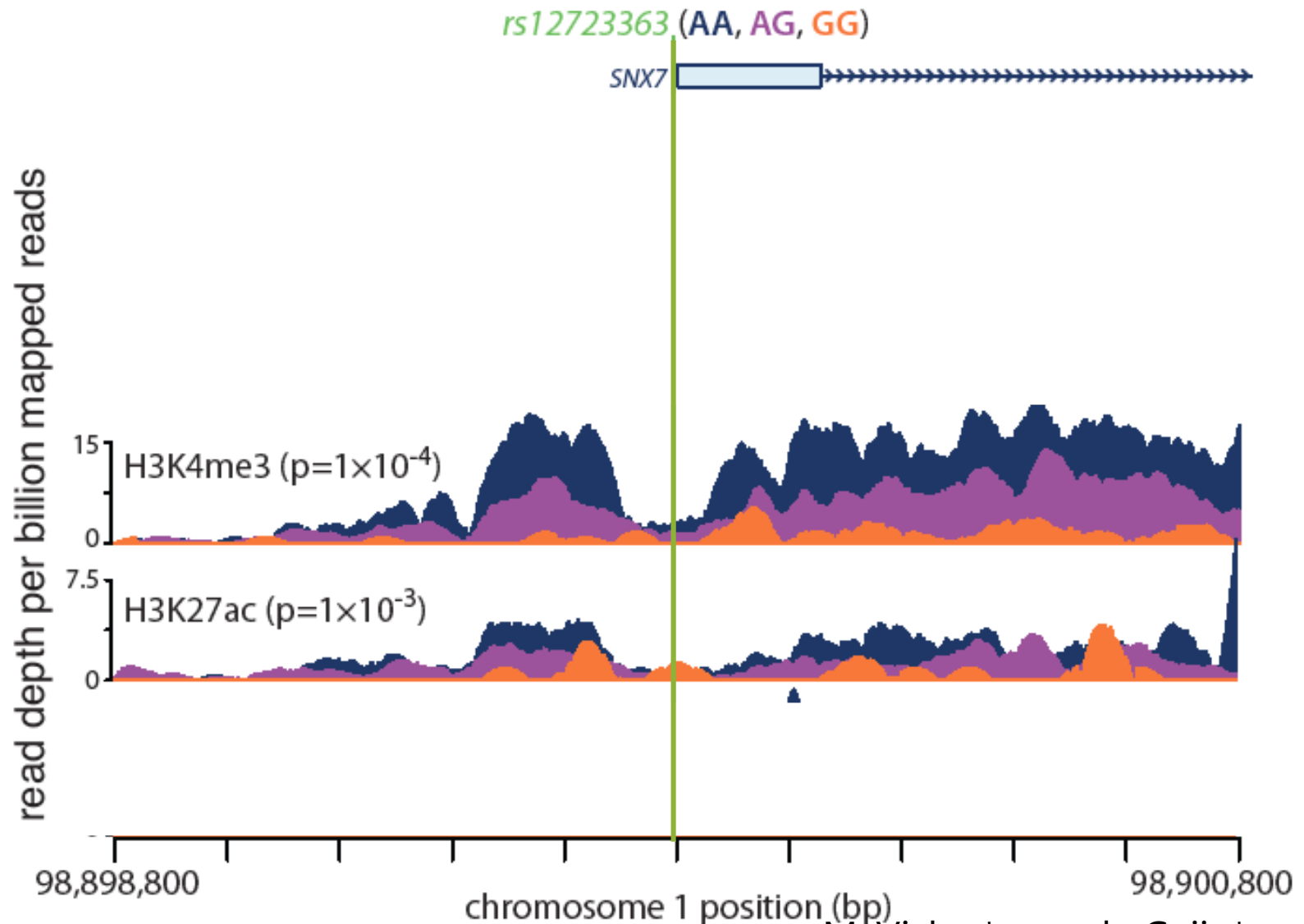


# Example of a histone mark QTL

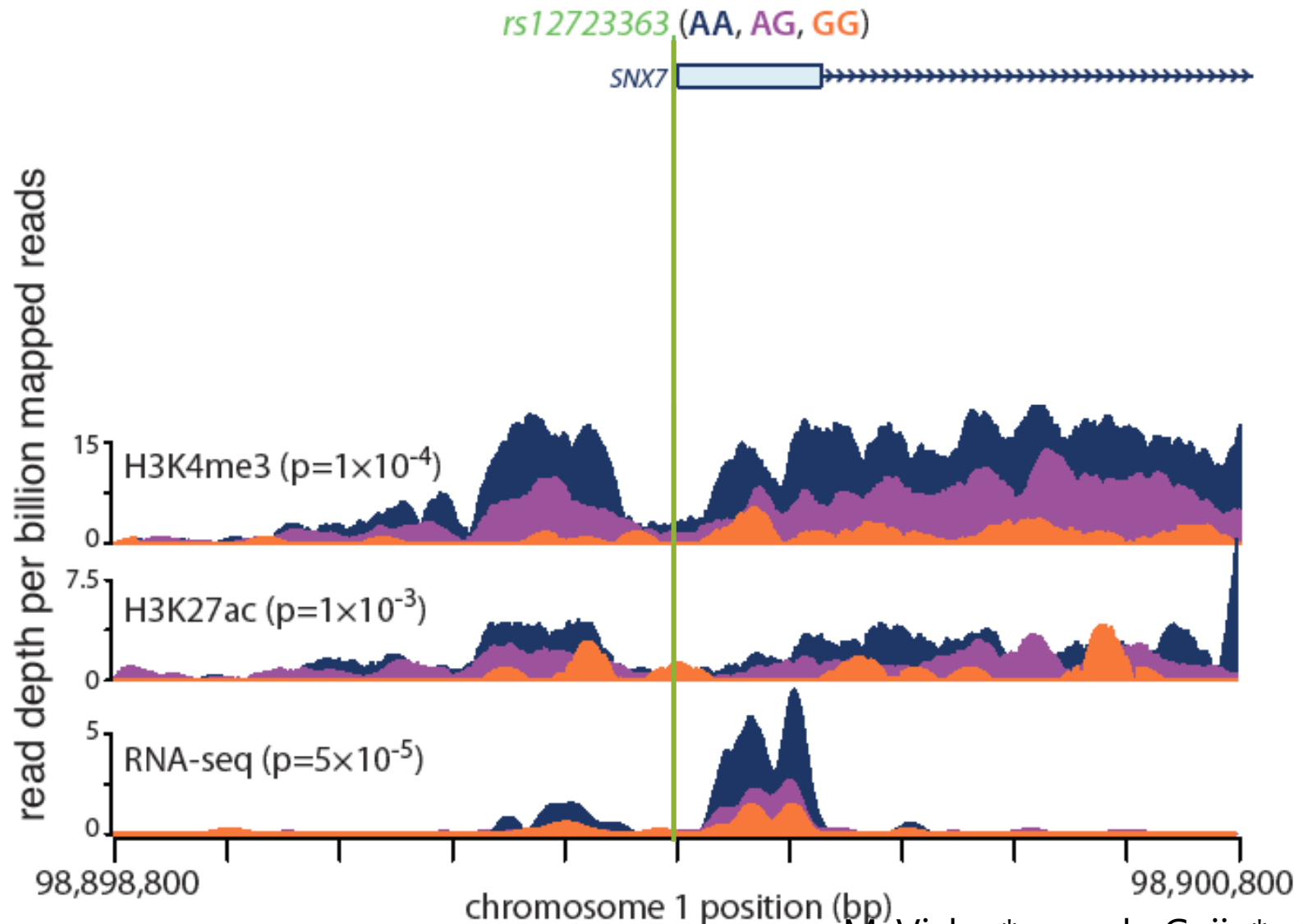


McVicker\*, van de Geijn\* et al.  
2013

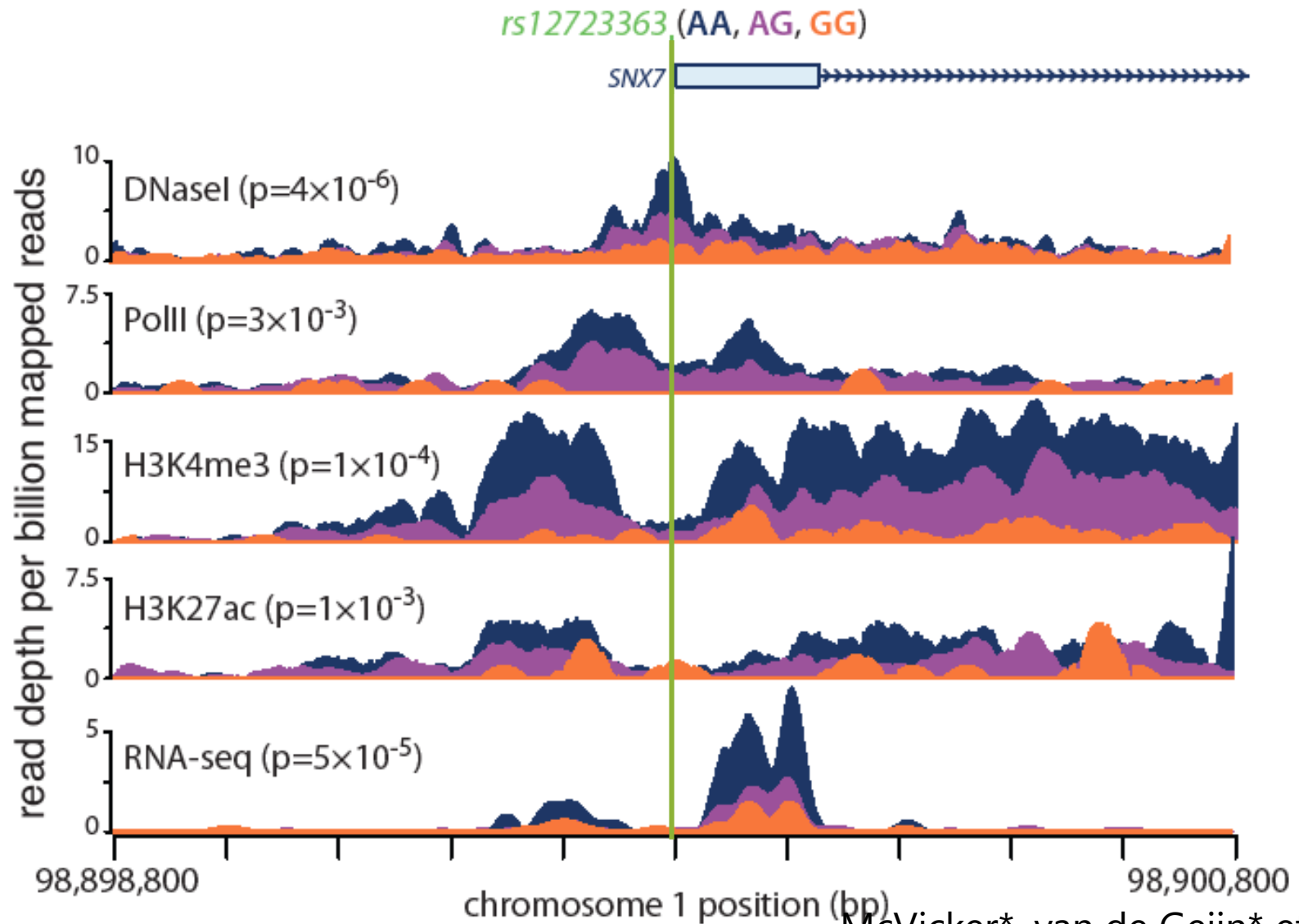
# Example of a histone mark QTL



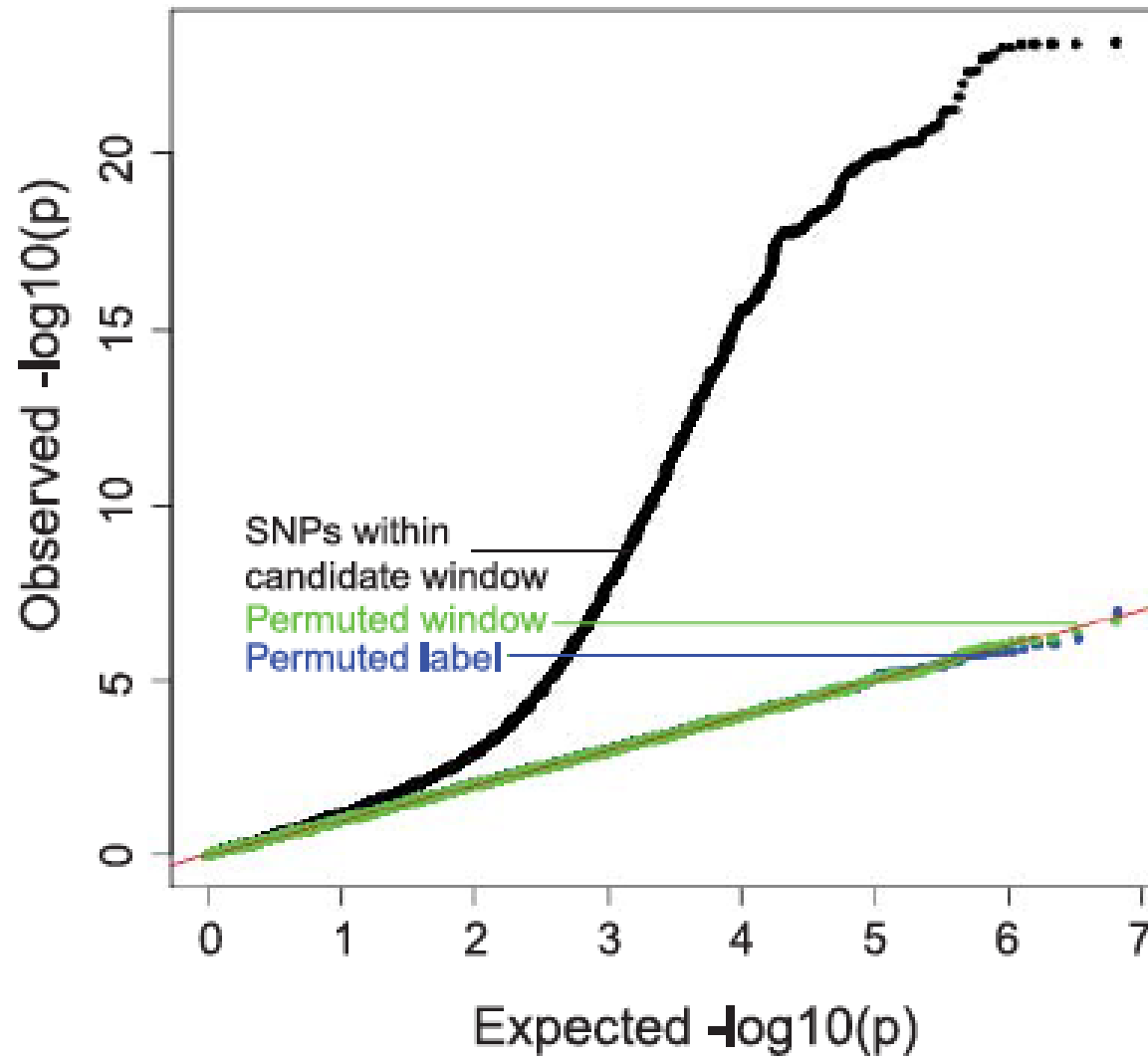
# Example of a histone mark QTL



# Example of a histone mark QTL

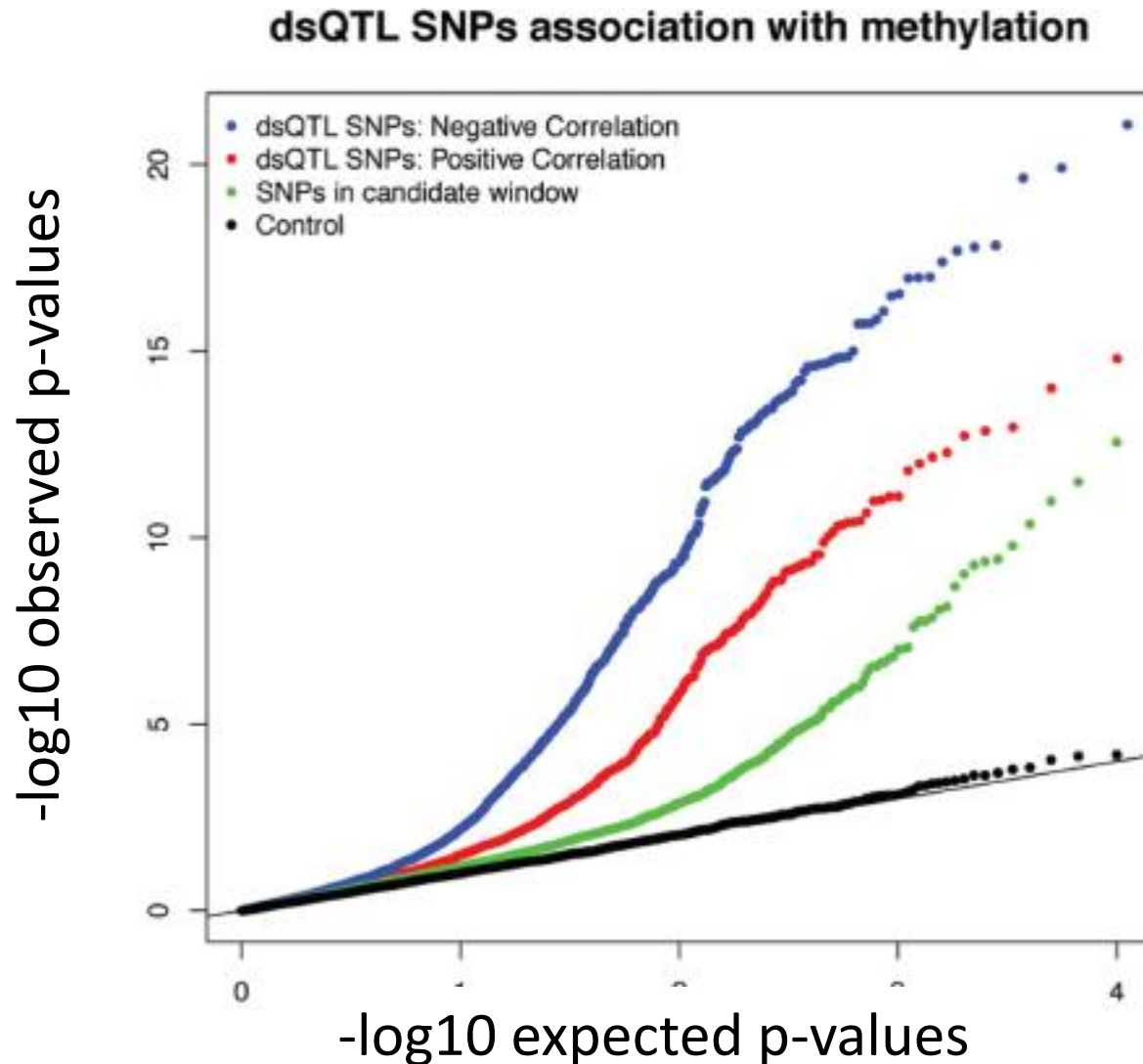


# DNA methylation QTLs

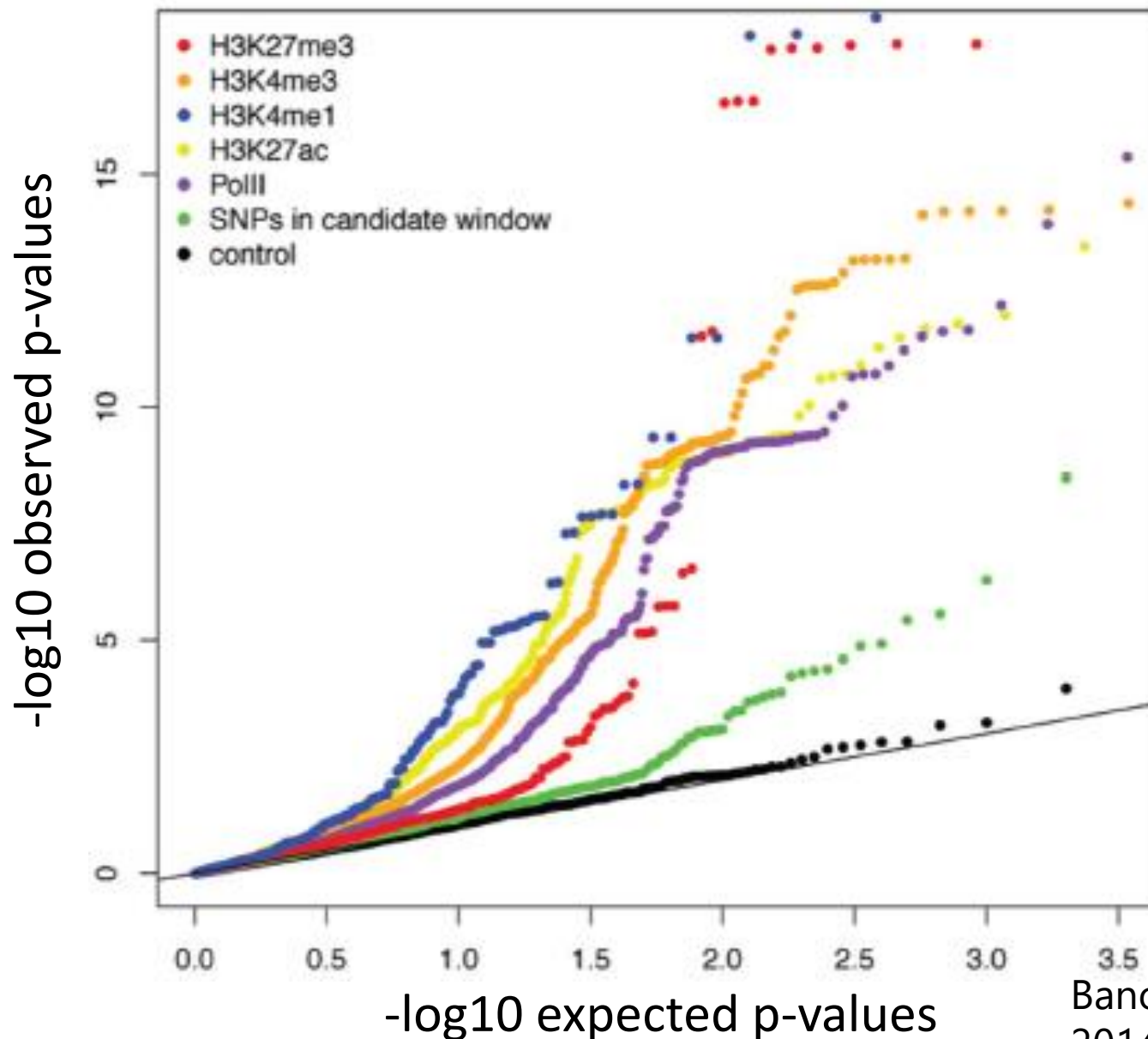


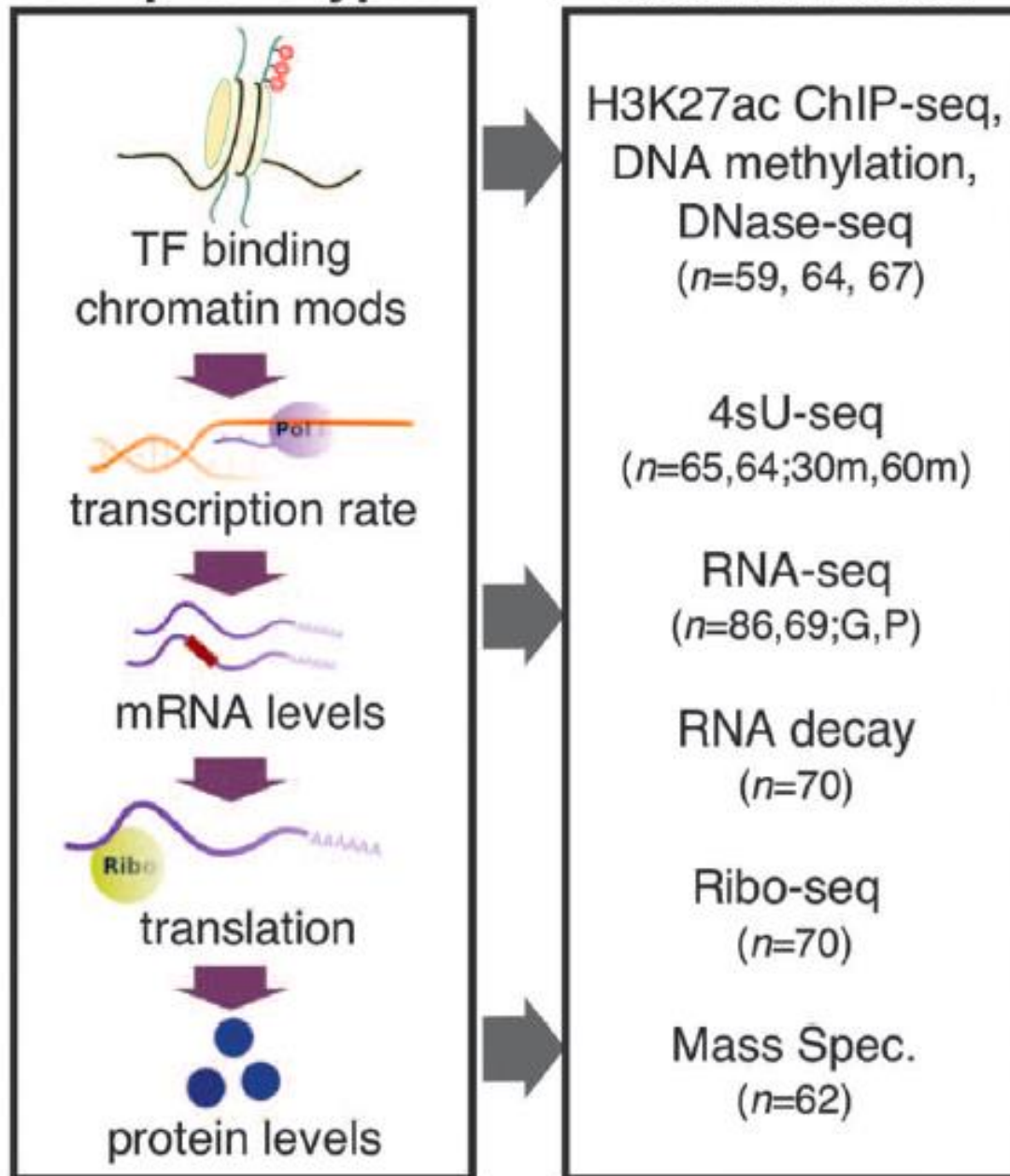


# DNA Methylation QTLs are often also DNase sensitivity QTLs

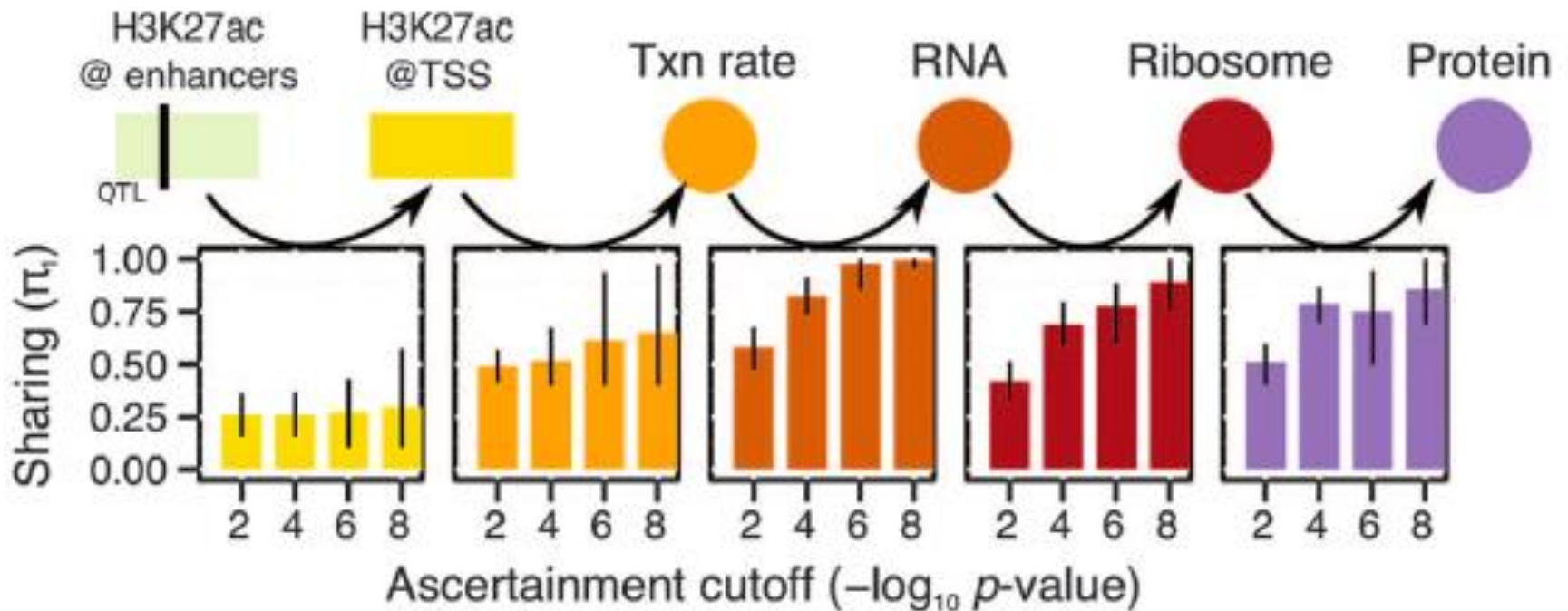


# Histone mark QTLs are often DNA methylation QTLs

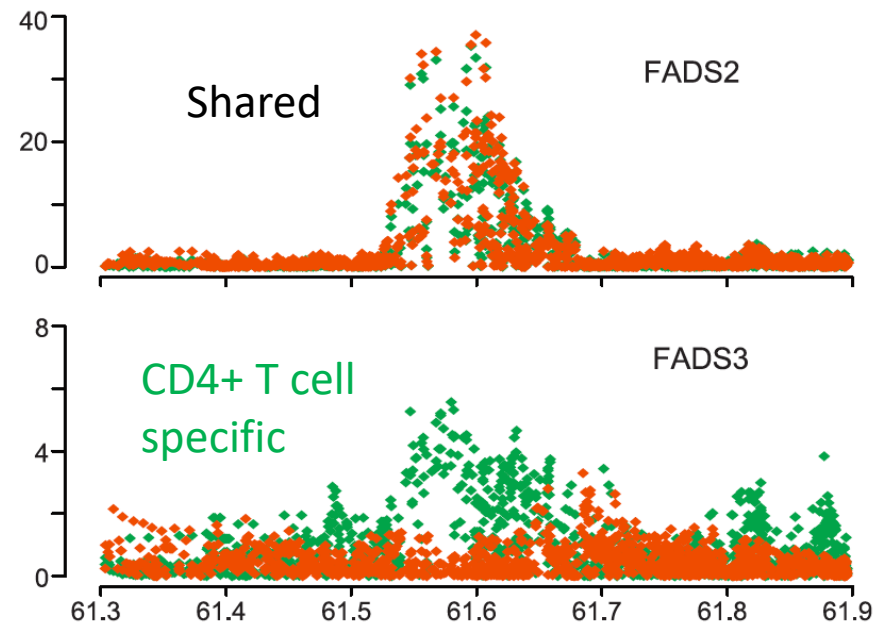
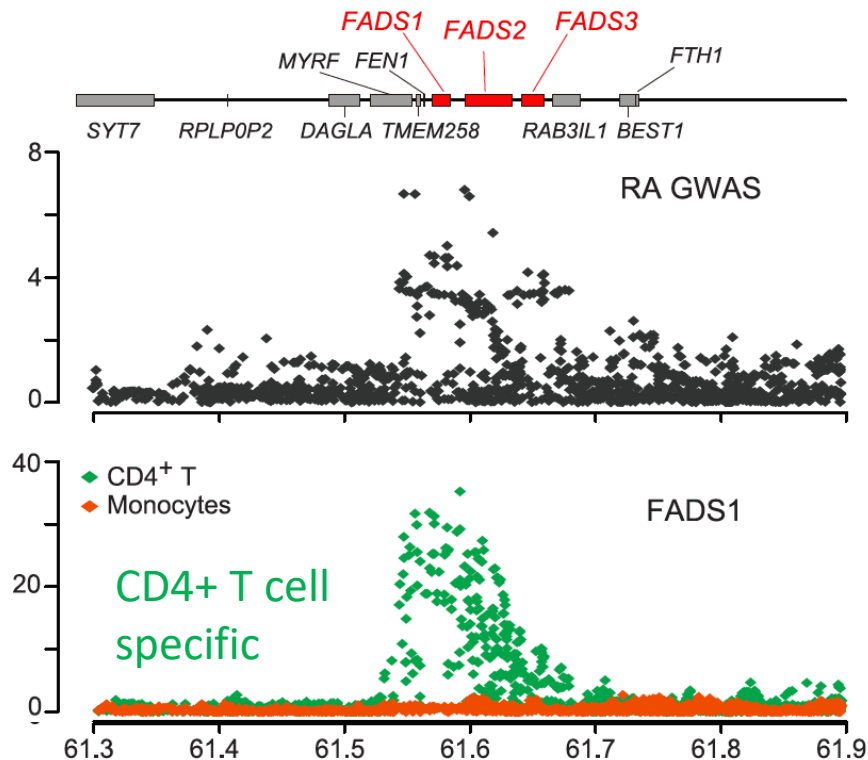


**A****Cell phenotypes****Quantification****Regulatory cascade**

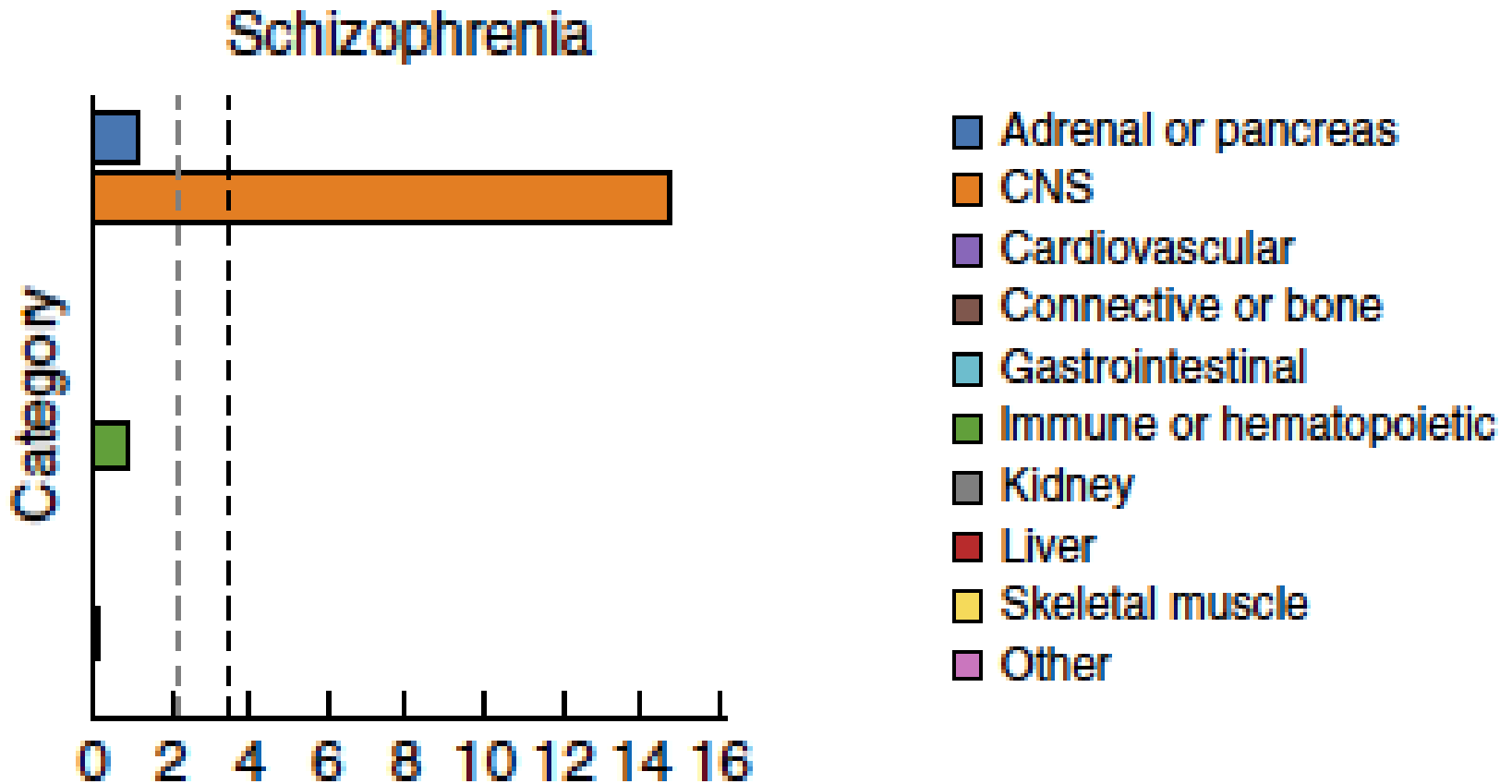
# Sharing of regulatory QTLs

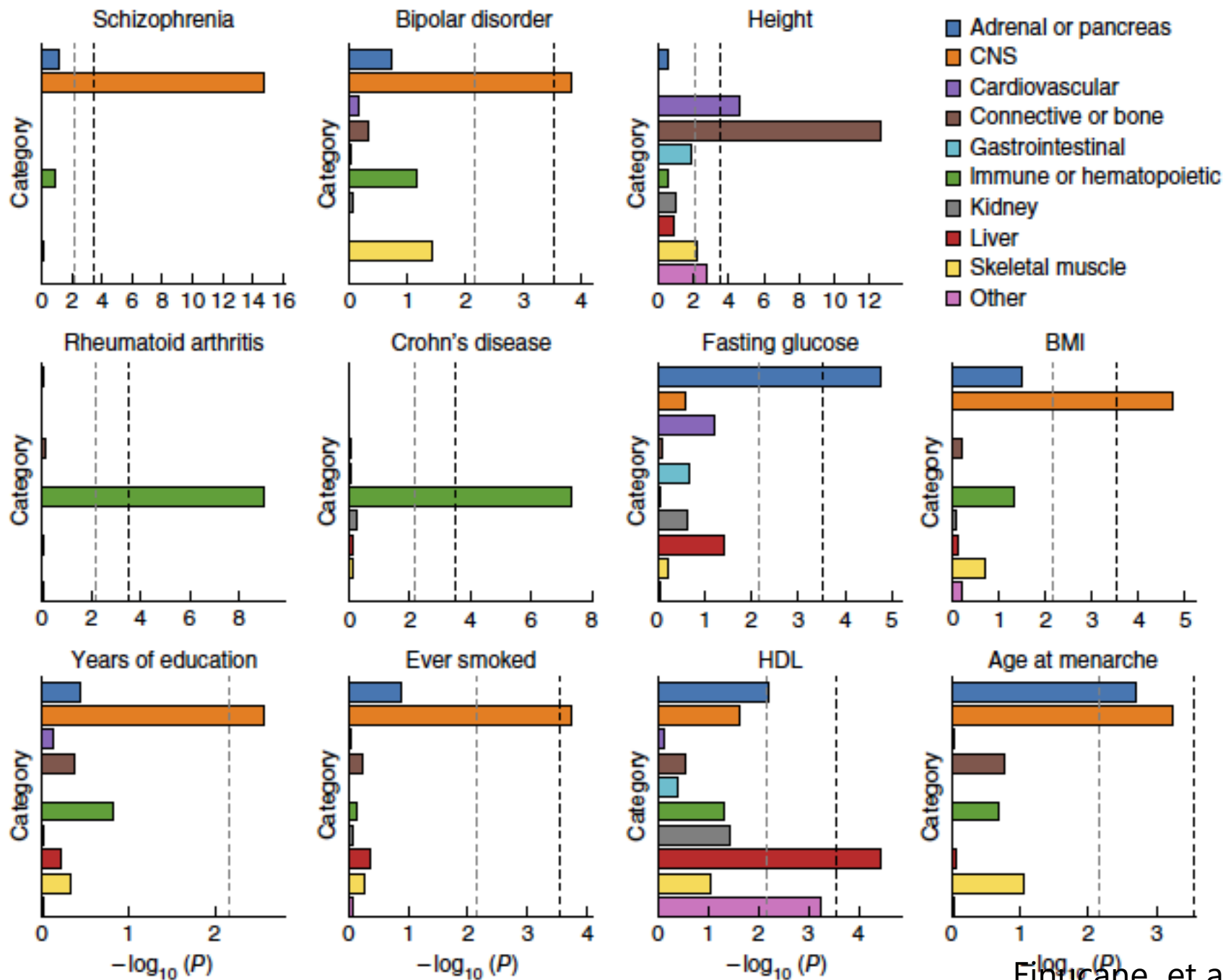


# Intersecting eQTLs with GWAS



# GWAS enrichment in cell-type specific annotations





# Summary

- Many molecular traits can be associated with genetic variants
- Molecular QTLs can reveal mechanism underlying organismal traits
- Smaller samples are needed to map molecular QTLs than organismal traits
- Genetic associations are challenging to interpret