# Figure 1

## GWAS Summary statistics



## LD information



| SNP | rs1 | rs2 | rs3 | rs4 | rs5 | rs6 | rs7 |
|-----|-----|-----|-----|-----|-----|-----|-----|
| -log10P | 10.8 | 12.5 | 9.1 | 9.1 | 5.3 | 6.2 | 3.7 |
| DHS | | | | | | | |
| TFBS | | | | | | | |
| LD | | | | | | | |

## Regulatory/functional data
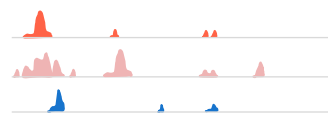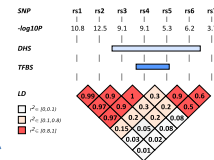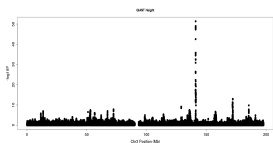


## 1. Greedy *LD pruning* ($r^2 > 0.01$)
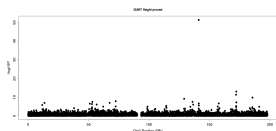


## 2. *LD tagging* ($r^2 > 0.8$) *annotation overlap*

### A annotations



V variants
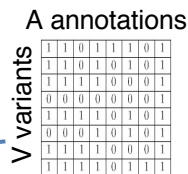
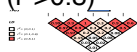## 3. Generalized linear model with *feature correction*

$$\text{logit } E(y) = 1\alpha + X_{TSS}\beta_{TSS} + X_{TAGS}\beta_{TAGS} + X_{Ai}\beta_{Aj}$$
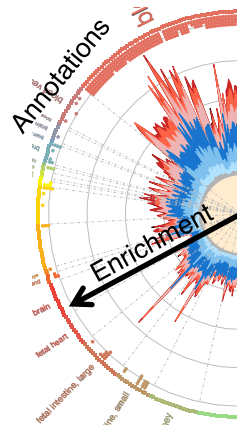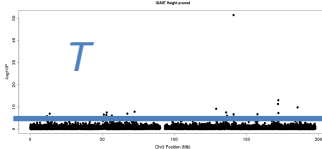
TSS distance



TSS

LD proxies ($r^2 > 0.8$)

Annotation j

$y_{SNPi} \sim$ Bernoulli

$$y_{SNPi} = \begin{cases} 1 \text{ if P-value} < T \\ 0 \text{ otherwise} \end{cases}$$

$T$





## 4. Model selection for multiple annotations

- Sort annotations in order of significance
- Iteratively add another annotation to the model if it improves the model
  $$\text{logit } E(y) = 1\alpha + X_{TSS}\beta_{TSS} + X_{TAGS}\beta_{TAGS} + X_{A1}\beta_{A1} + \ldots + X_{Aj}\beta_{Aj}$$
- Report final model and tree of retained/discarded annotations



Annotations: A1, A2, A3, A4, A5, A6

Enrichment