

Predict Functions for Noncoding Variants

Lecture 5

Outline

- Why Studying Noncoding Variants
- DEEPSEA (Zhou J. & Troyanskaya O., Nature Methods, 2015)
- LINSIGHT (Huang Y. et al. Nat. Genet. 2017)
- Integrate with GWAS Data

Part of the slides are adapted from the slides by Anthony Gitter,
Mark Craven, Colin Dewey and Daifeng Wang@wisc

Why Studying Non-coding Variants?

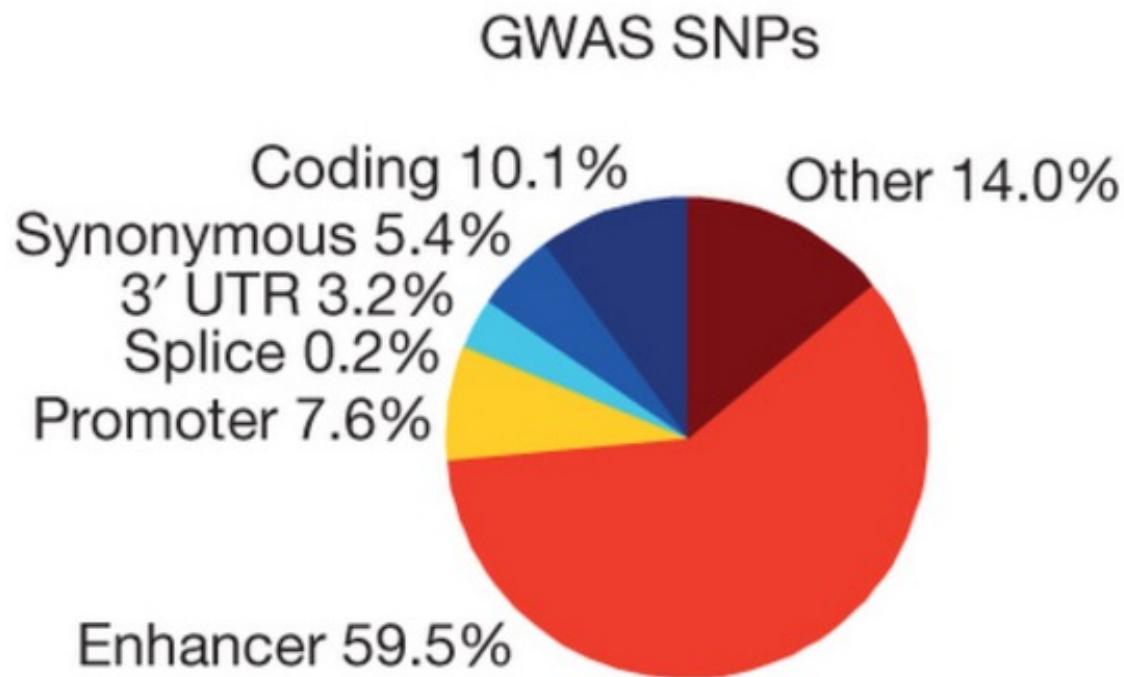
- **Noncoding variants:** germline and somatic variants located in non-coding regions, which can play an important role in complex human diseases.
- Ways noncoding variants can be functional:
 - Disrupt DNA sequence motifs
 - Promoters, enhancers
 - Disrupt miRNA binding
 - Mutations in introns affect splicing
 - Indirect effects from the above changes
- Functional effects of non-coding variants can be studied by experiments and computational prediction.

Success of GWAS



As of 2023-05-20, the GWAS Catalog contains >6K publications with >519K top associations

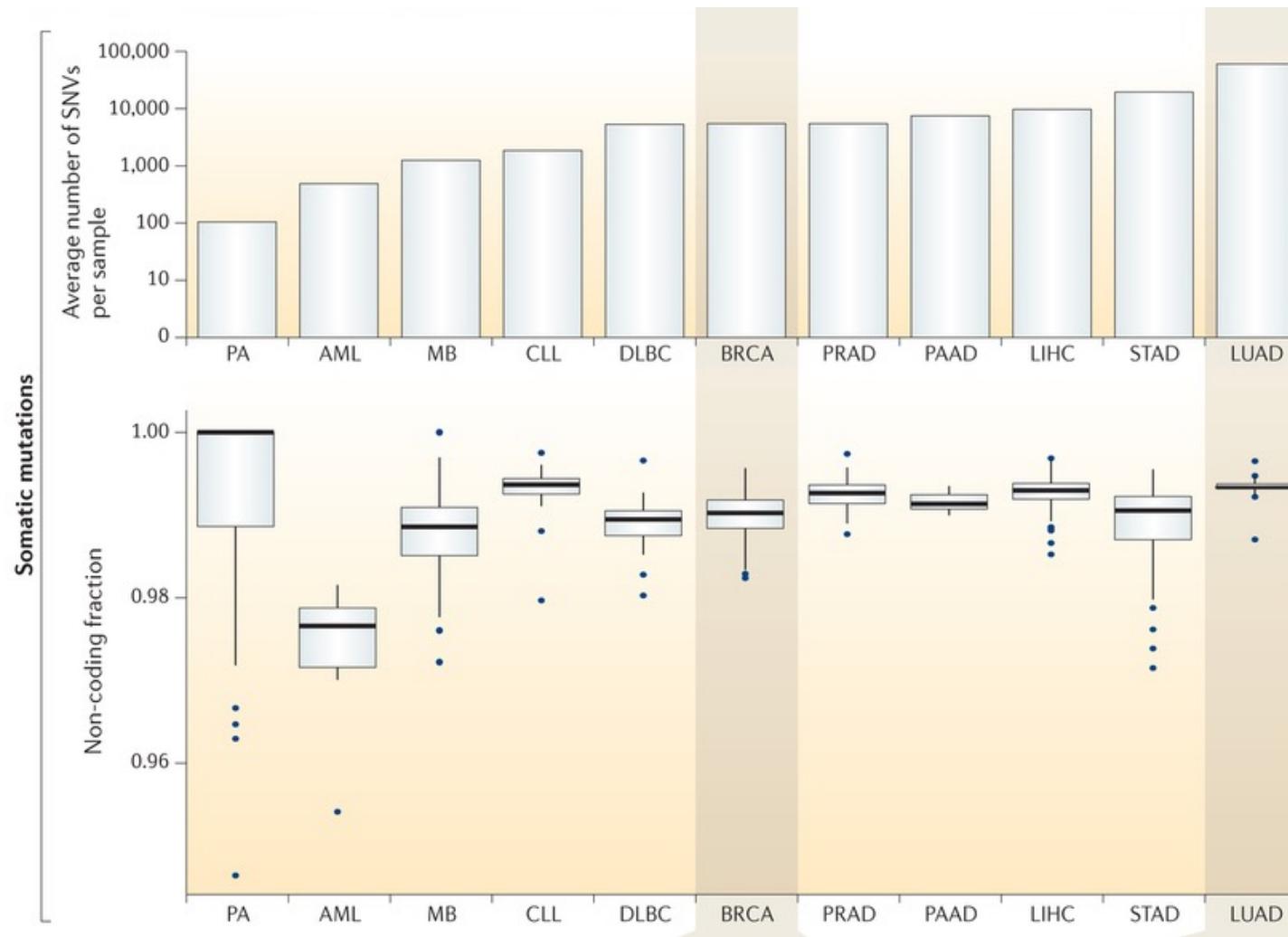
Annotations of GWAS Signals of 21 Autoimmune Diseases



90% prioritized “causal” SNPs
are noncoding

Farh K.K. et al. Nature 2015

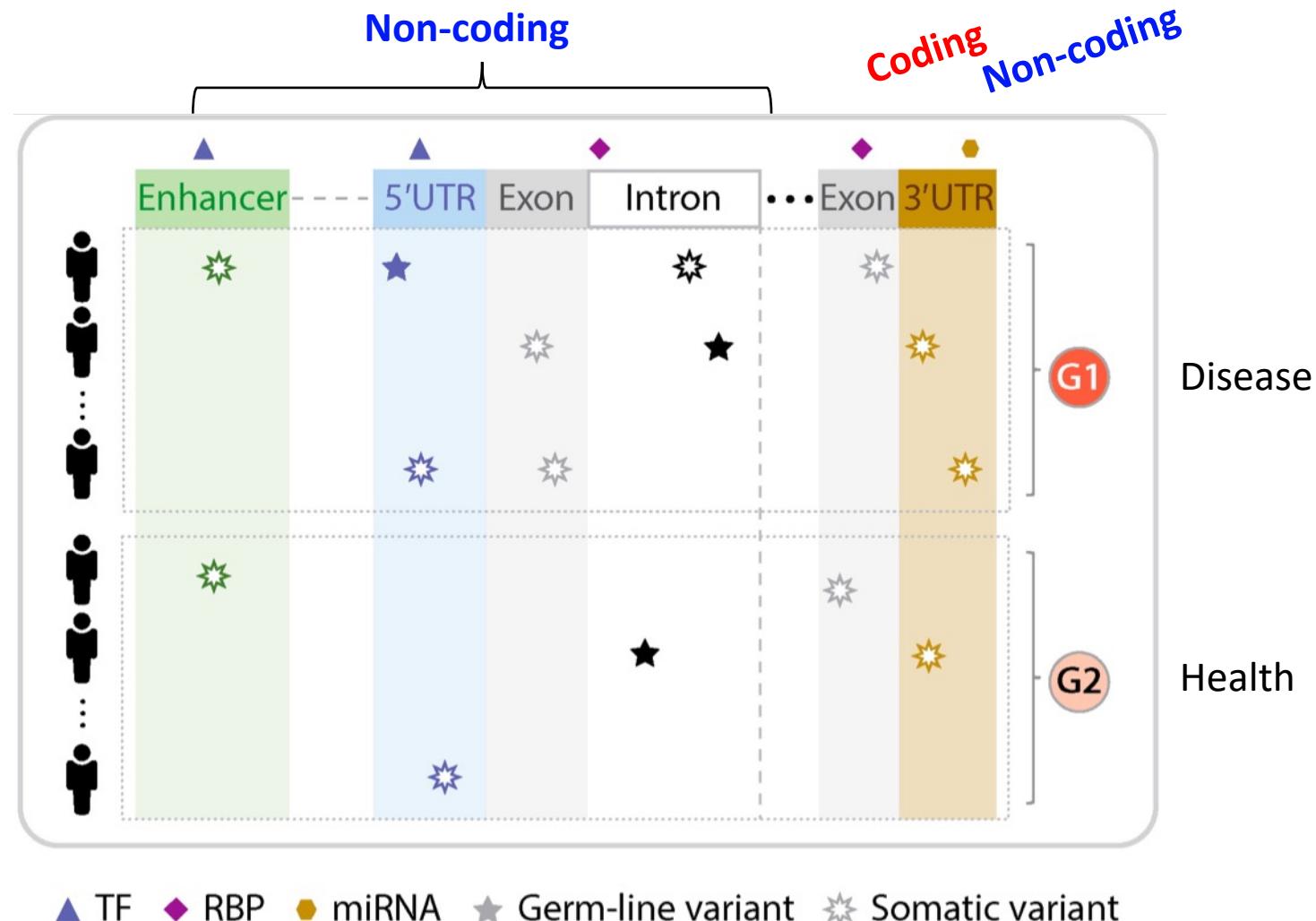
Almost all single nucleotide variants in cancer are noncoding



Khurana E. et al.
Nature Reviews Genetics 2016

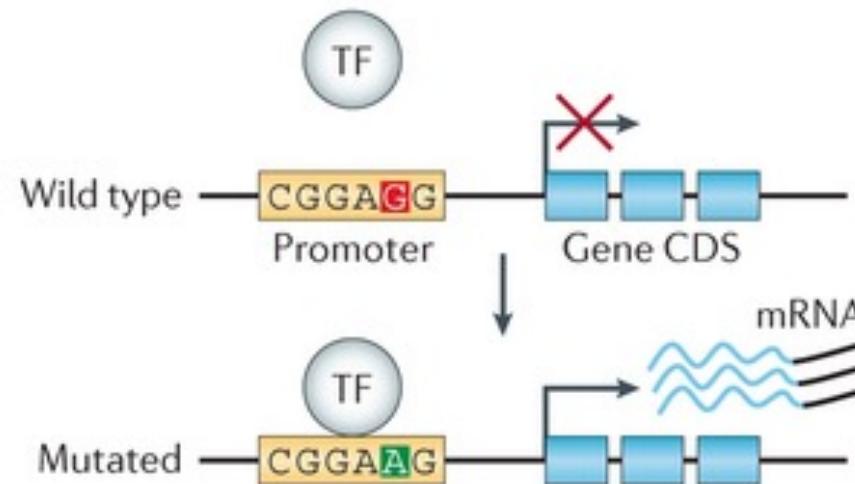
However, very few of these are driver mutations

How to link non-coding disease SNPs to genes?

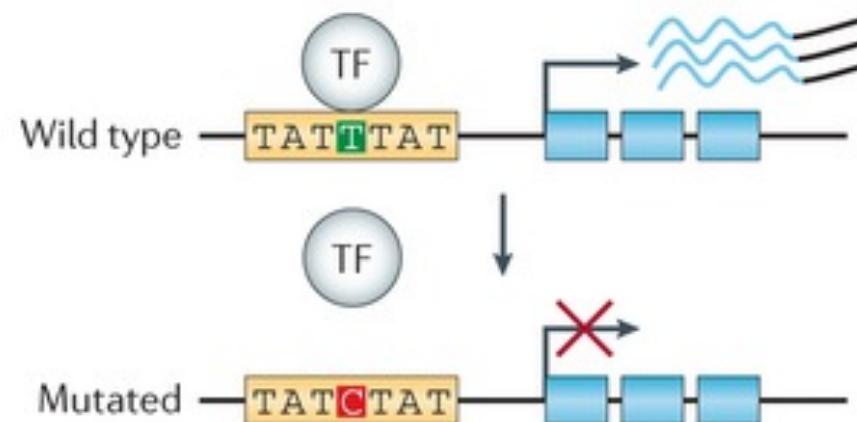
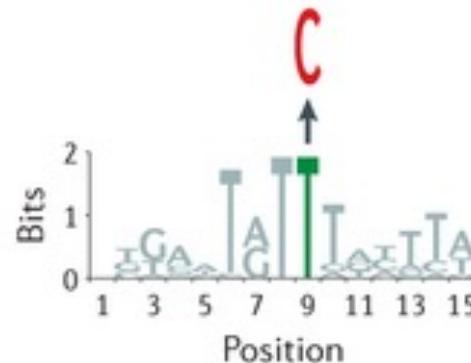


Variants Altering Motifs

Ba Gain of motif

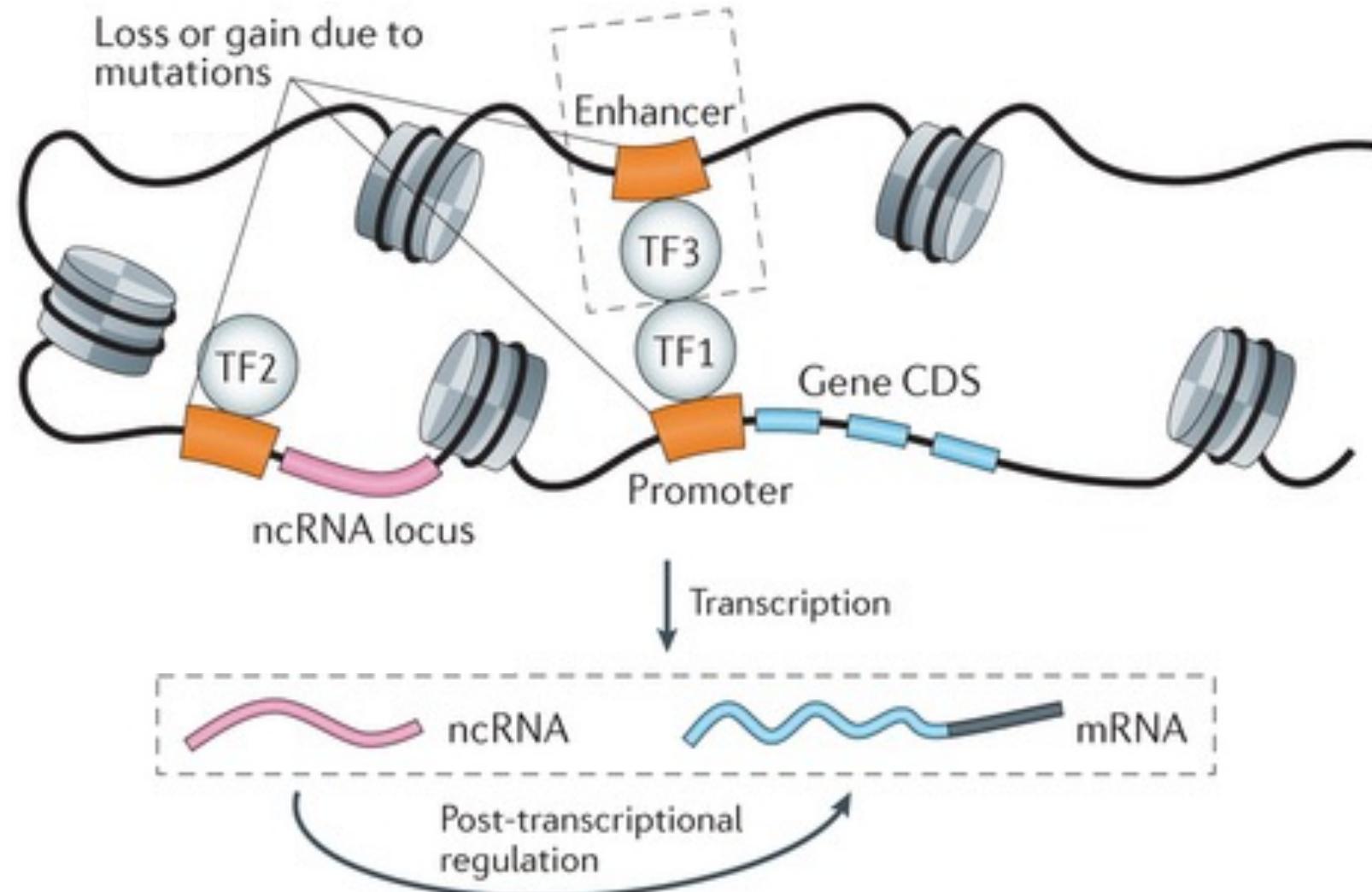


Bb Loss of motif



Khurana *Nature Reviews Genetics* 2016

Variants Affect Proximal and Distal Regulators

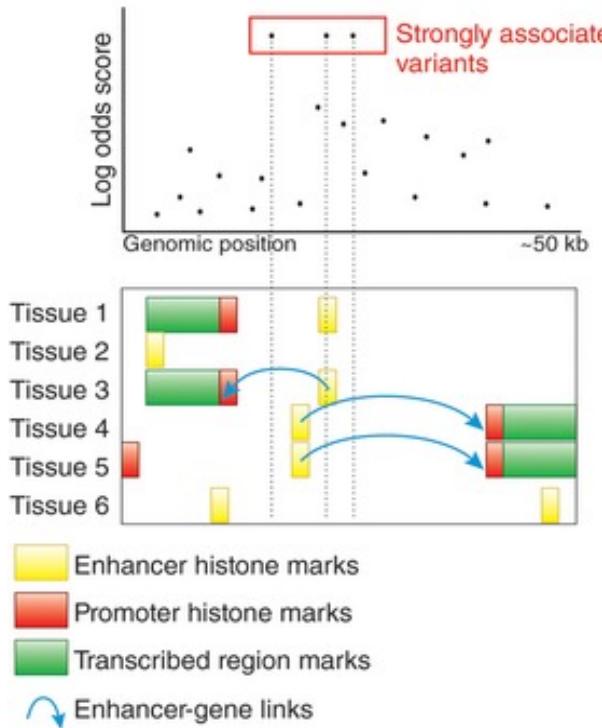


Khurana *Nature Reviews Genetics*
2016

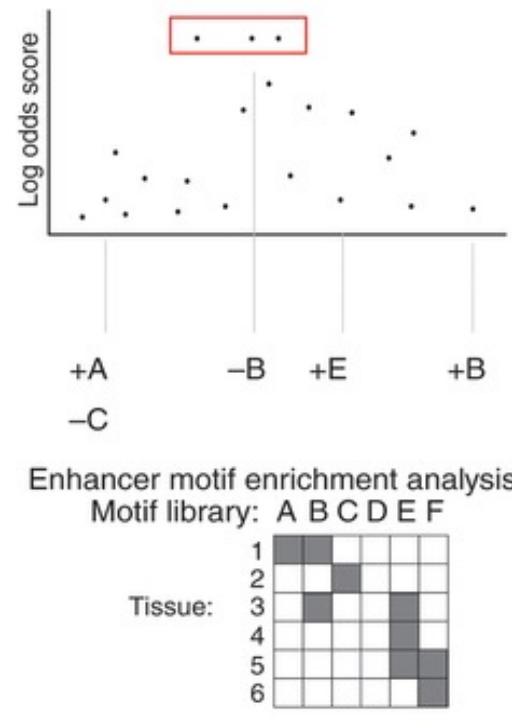
Evidence Used to Prioritize Noncoding Variants

Interpreting GWAS signals using functional and comparative genomics datasets

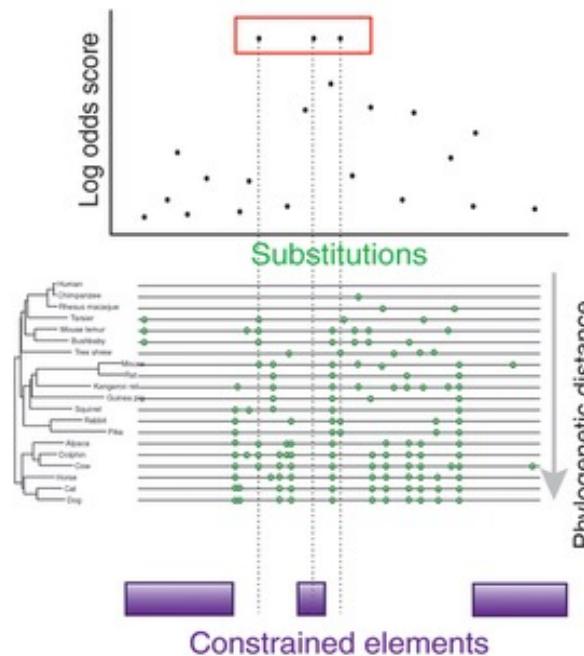
a Dissect associated haplotype using functional genomics



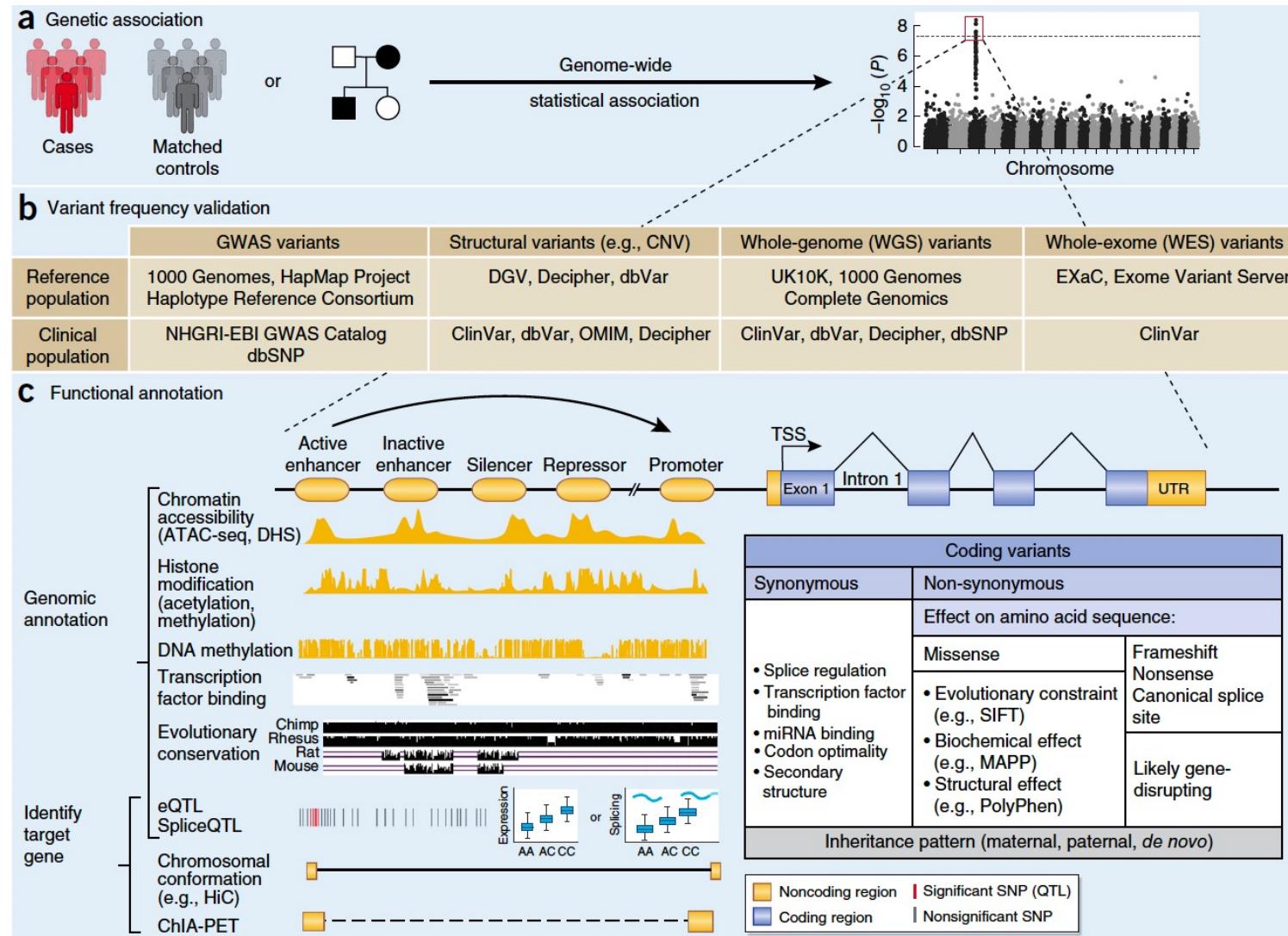
b Dissect associated haplotype using regulatory genomics



c Dissect associated haplotype using comparative genomics



Functional Genomics from Variants to Genes



Disease-associated genomic variants



How do variants function?

Visualizing Evidence



<https://regulomedb.org/regulome-search/>

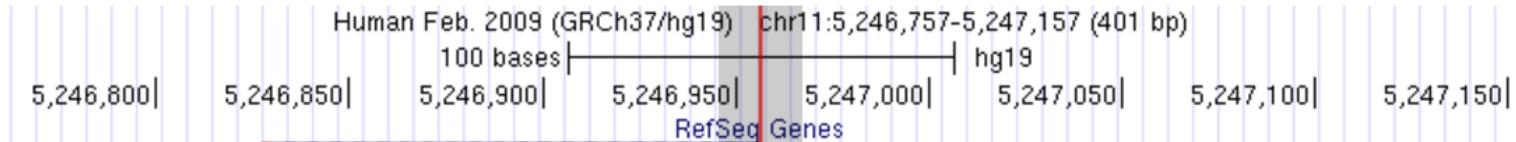
Data supporting chr11:5246957 (rs33914668)

Summary of evidence

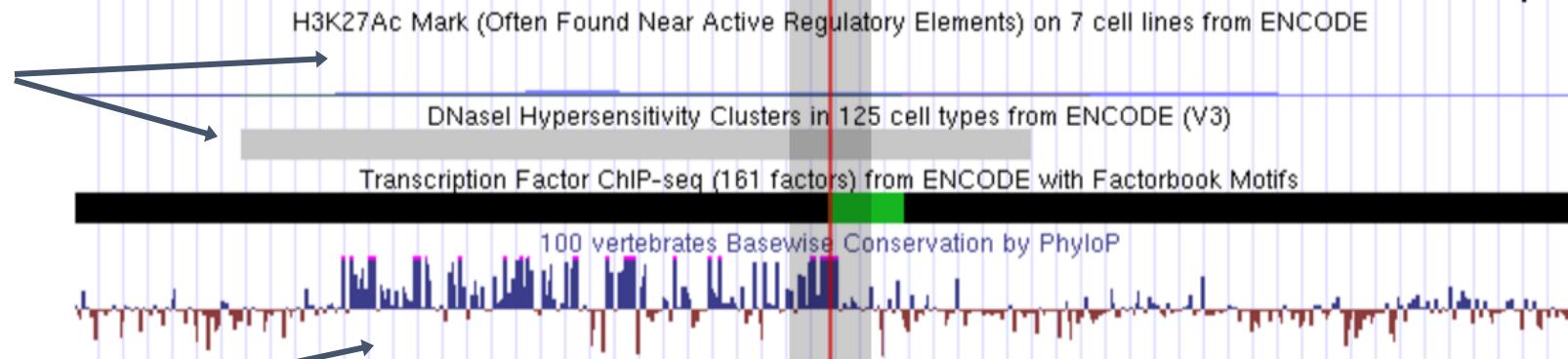
Score: 2a

Likely to affect binding

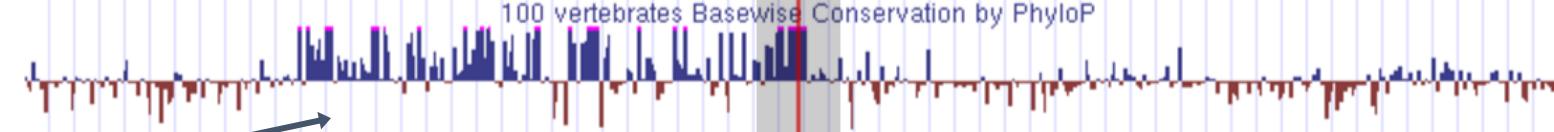
Genes



Epigenetic annotations



Conservation



Affected motifs



Boyle Genome Research 2012

Prioritizing GWAS Signals with Epigenetics

Summary

- Disrupted regulatory elements one of the best understood effects of noncoding SNPs
- Make use of extensive epigenetic datasets
- Similar strategies have actually worked
 - rs1421085 in *FTO* region and obesity
 - Claussnitzer *New England Journal of Medicine* 2015
- Epigenetic data at a genomic position is often in the presence of the reference allele
 - Don't have measurements for the SNP allele

DeepSEA: Deep Learning-Based Sequence Analyzer

- Given:
 - A sequence variant and surrounding sequence context
 - Large-scale chromatin-profiling data
- Do:
 - Predict TF binding, DNase hypersensitivity, and histone modifications in multiple cell and tissue types
 - Predict variant functionality with single-nucleotide sensitivity

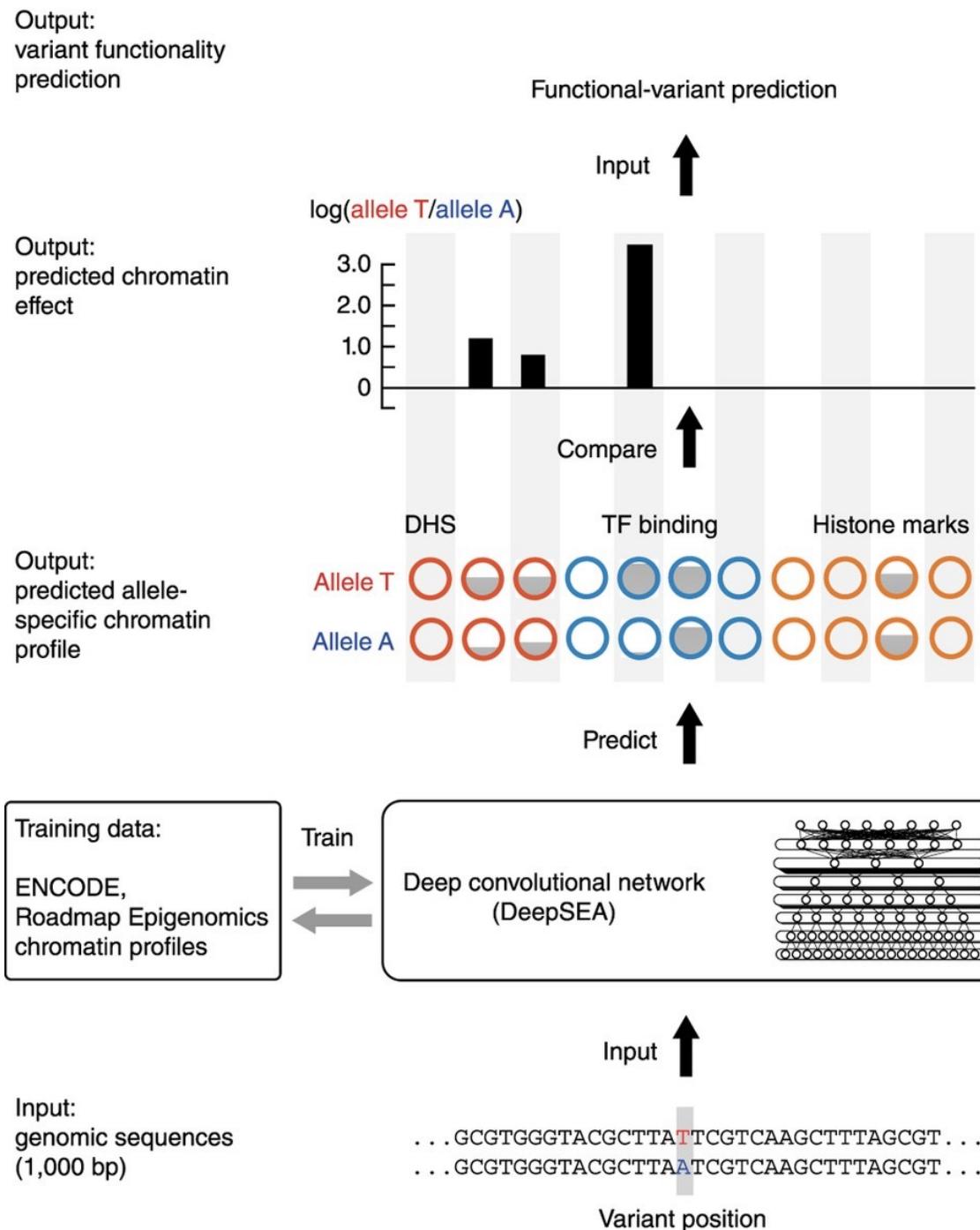
Zhou and Troyanskaya *Nature Methods* 2015

DeepSEA: Deep Learning-Based Sequence Analyzer

- First directly learn regulatory sequence code from genomic sequence by learning to simultaneously predict large-scale chromatin-profiling data, including **TF binding, DNase I sensitivity, and histone-mark profiles**
- Three major features
 - Integrating sequence information from a wide sequence context (e.g., 1kbp): *sequence surrounding the variant position determines the regulatory properties of the variant*
 - Learning sequence code at multiple spatial scales with a hierarchical architecture: *allows scaling to such long sequence input and learning sequence dependencies at multiple scales*
 - Multitask joint learning of diverse chromatin factors sharing predictive features: *computational efficient, allows predictive strength to be shared across a wide range of chromatin feature profiles for TF binding, DHSs and histone marks*

Schematic Overview of DeepSEA

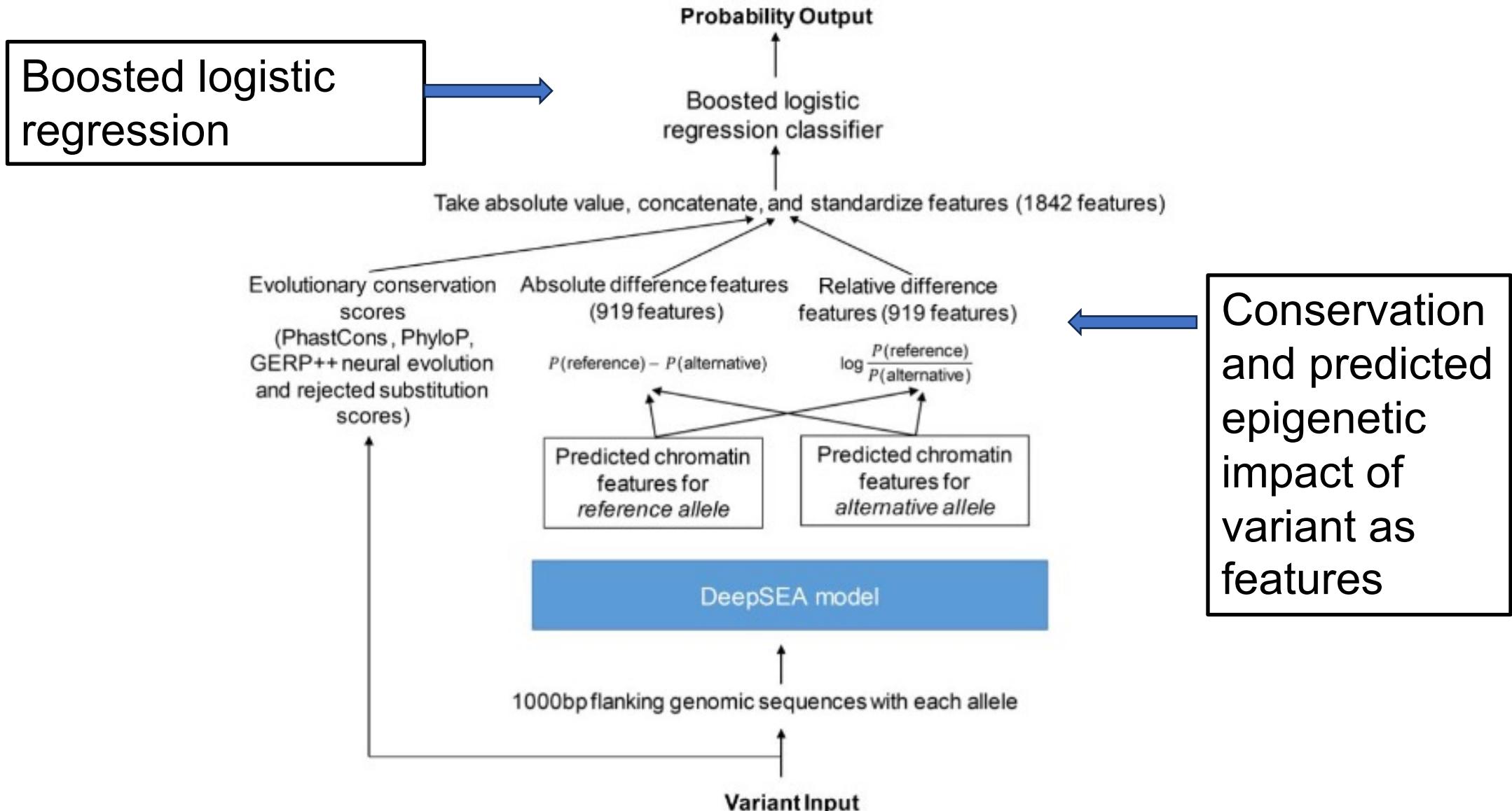
For model training: Genome-wide chromatin profiles from ENCODE and Roadmap Epigenomics projects, including 690 TF binding profiles for 160 different TFs, 125 DHS profiles and 104 histone-mark profiles



Predicting Functional Variants

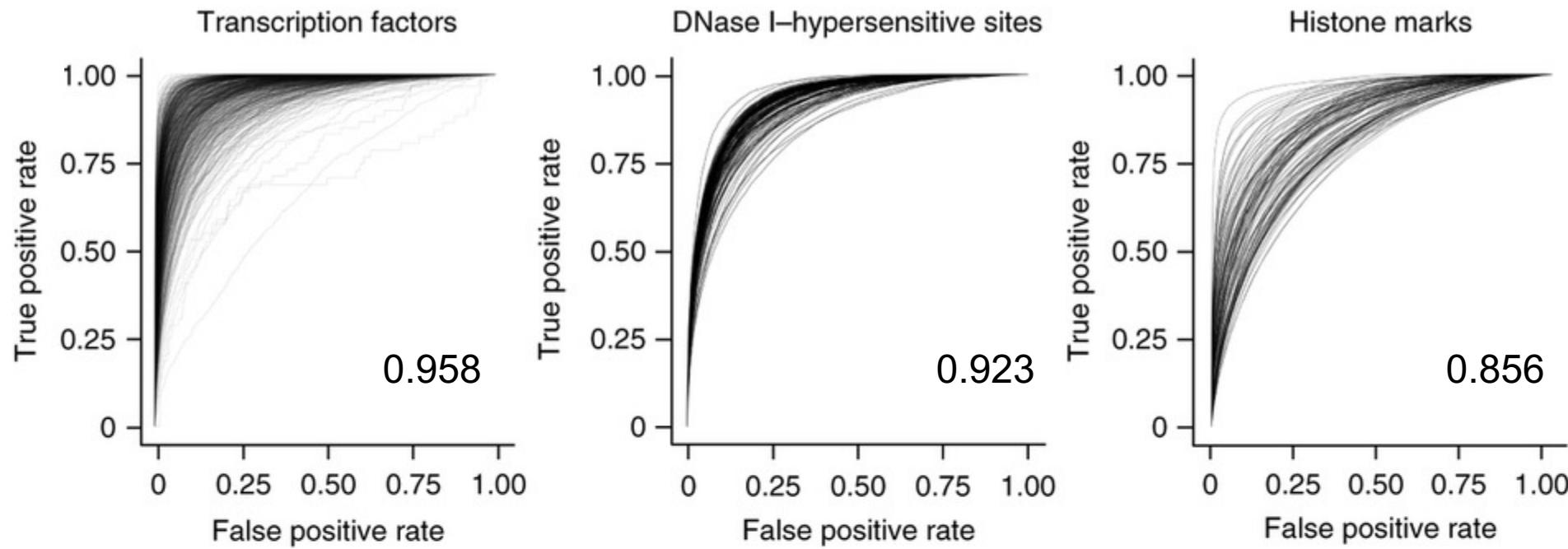
- Can predict epigenetic signal for any novel variant (SNP, insertion, deletion)
- Define novel features to classify variant functionality
 - Difference in probability of signal for reference and alternative allele
- Train on SNPs annotated as regulatory variants in GWAS and eQTL databases

Predicting Functional Variants



Predicting Epigenetic Annotations

- Compute median AUROC for three types of classes



Zhou and Troyanskaya *Nature Methods* 2015

DeepSEA Summary

- Ability to predict how unseen variants affect regulatory elements
- Accounts for sequence context of motif
- Parameter sharing with convolutional layers
- Multitask learning to improve hidden layer representations

- Does not extend to new types of cells and tissues
- AUROC is misleading for evaluating genome-wide epigenetic predictions

Predicting New TF-cell Type Pairs

Training data

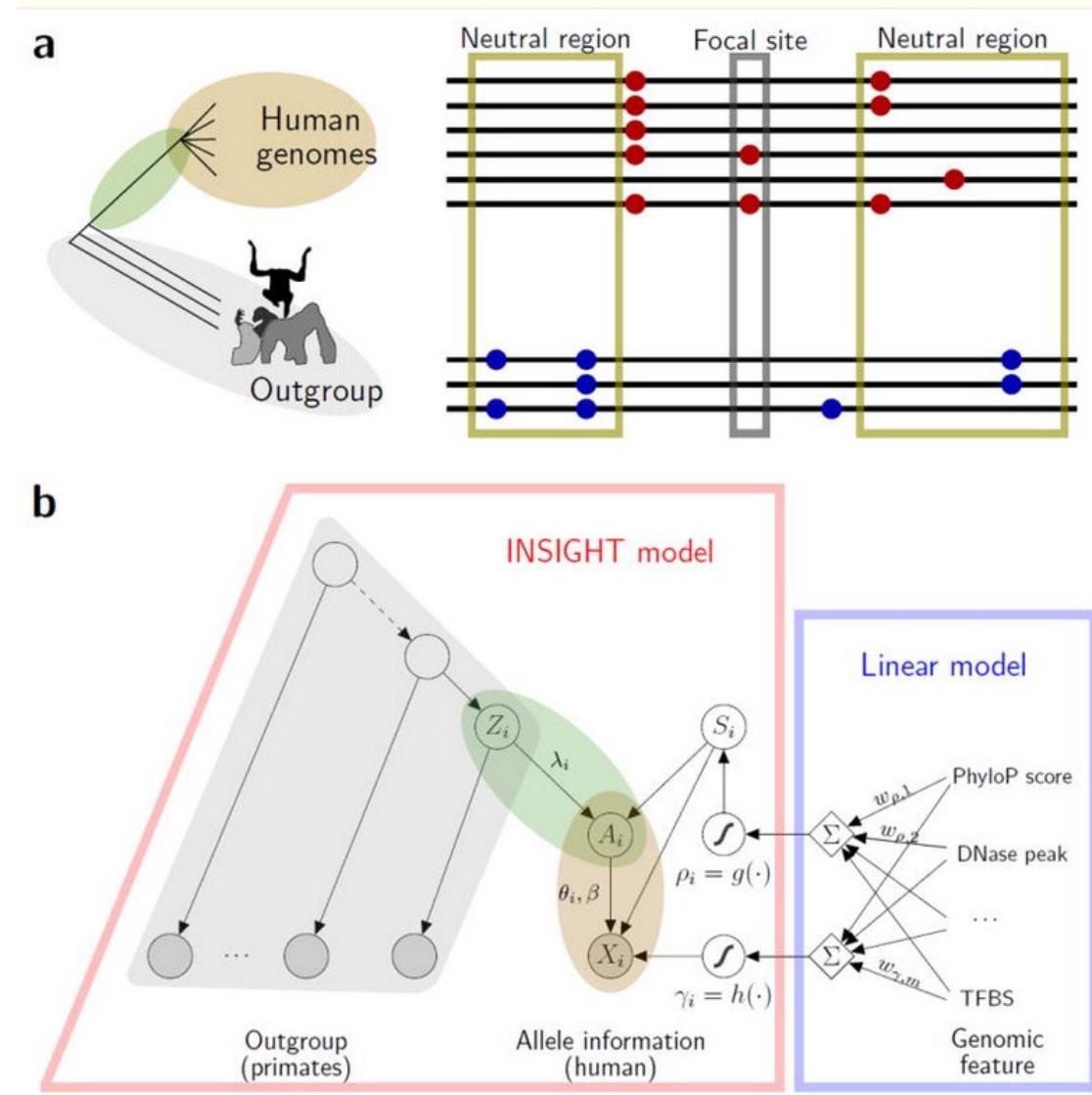
TF	Cell type
A	1
A	2
A	3
B	3
B	4
C	1
C	4

- DeepSEA cannot predict pairs not present in training data
 - Can predict TF A in cell type 1
 - Not TF A in cell type 4
- New methods can
 - [TFImpute](#)
 - [FactorNet](#)
 - [Virtual ChIP-seq](#)

LINSIGHT: Linear INSIGHT (Huang Y. et al. Nat. Genet. 2017)

- Combines a generalized linear model for functional genomic data with a probabilistic model of molecular evolution
- Based on existing **INSIGHT/fitCons** framework but has vastly improved speed, scalability, genomic resolution, and prediction power
- The main idea behind **LINSIGHT** is to bypass the clustering step of fitCons and instead couple the probabilistic INSIGHT model directly to a generalized linear model for genomic features.

LINSIGHT: Linear INSIGHT (Huang Y. et al. Nat. Genet. 2017)



LINSIGHT: Linear INSIGHT (Huang Y. et al. Nat. Genet. 2017)

- fitCons is an evolution-based method, explicitly characterizing the influence of natural selection at each genomic site of interest using a full probabilistic evolutionary model and patterns of genetic variation within and between species.
- FitCons makes a distinction between functional genomic and comparative genomic data, first defining several hundred clusters of genomic positions with distinct functional genomic “signatures,” and then estimating the fraction of nucleotides under natural selection within each cluster from polymorphism and divergence data.
- These estimates are obtained using the INSIGHT evolutionary model, which are interpreted as the probabilities that mutations in each cluster of genomic sites will have fitness consequences ([fitCons scores](#)).
- [LINSIGHT scores](#) are available as a track on the Cold Spring Harbor Laboratory mirror of the UCSC Genome Browser (hg19 assembly).

Fig. 2. Summary of LINSIGHT scores across the noncoding human genome (3.001 billion nucleotide sites).

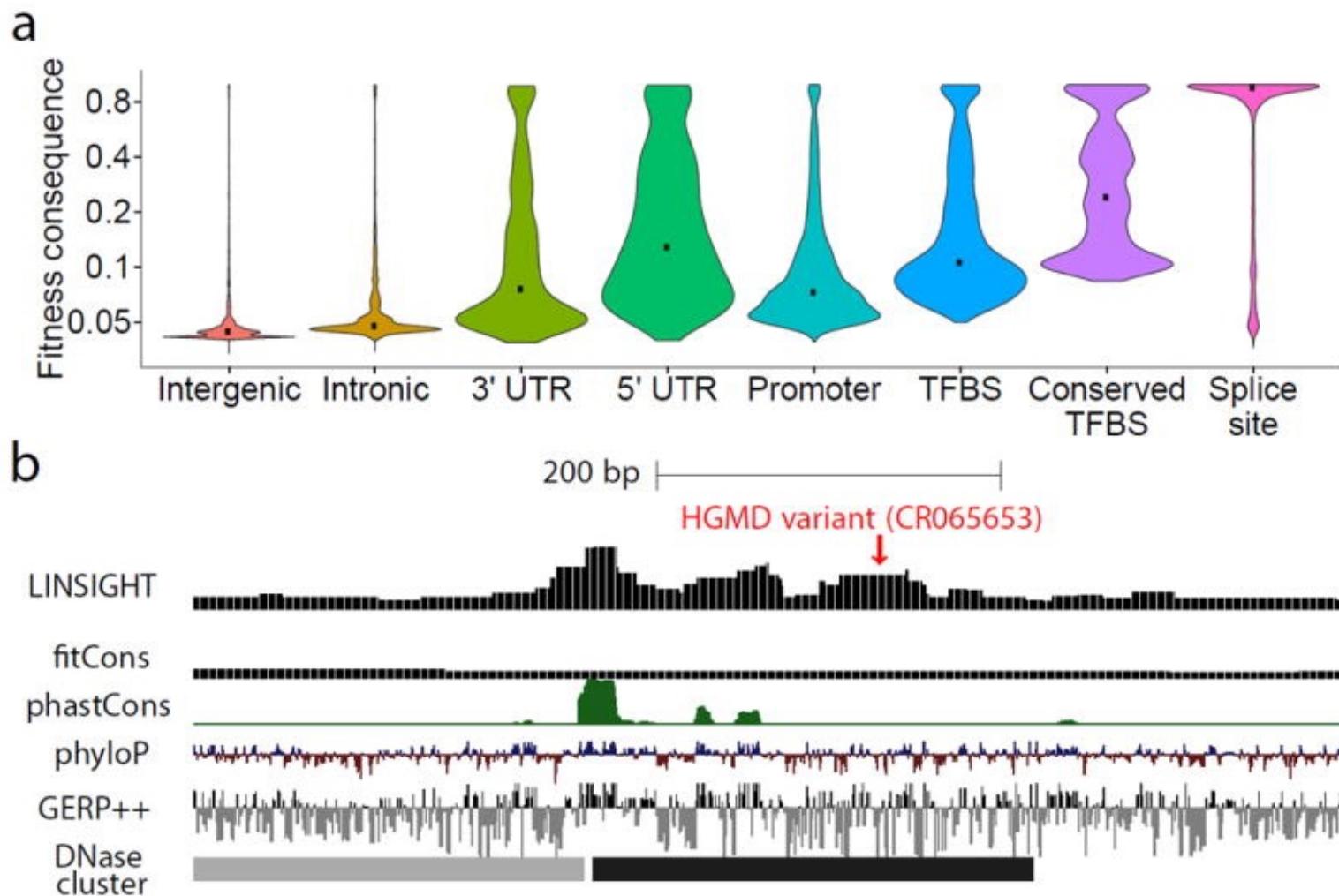
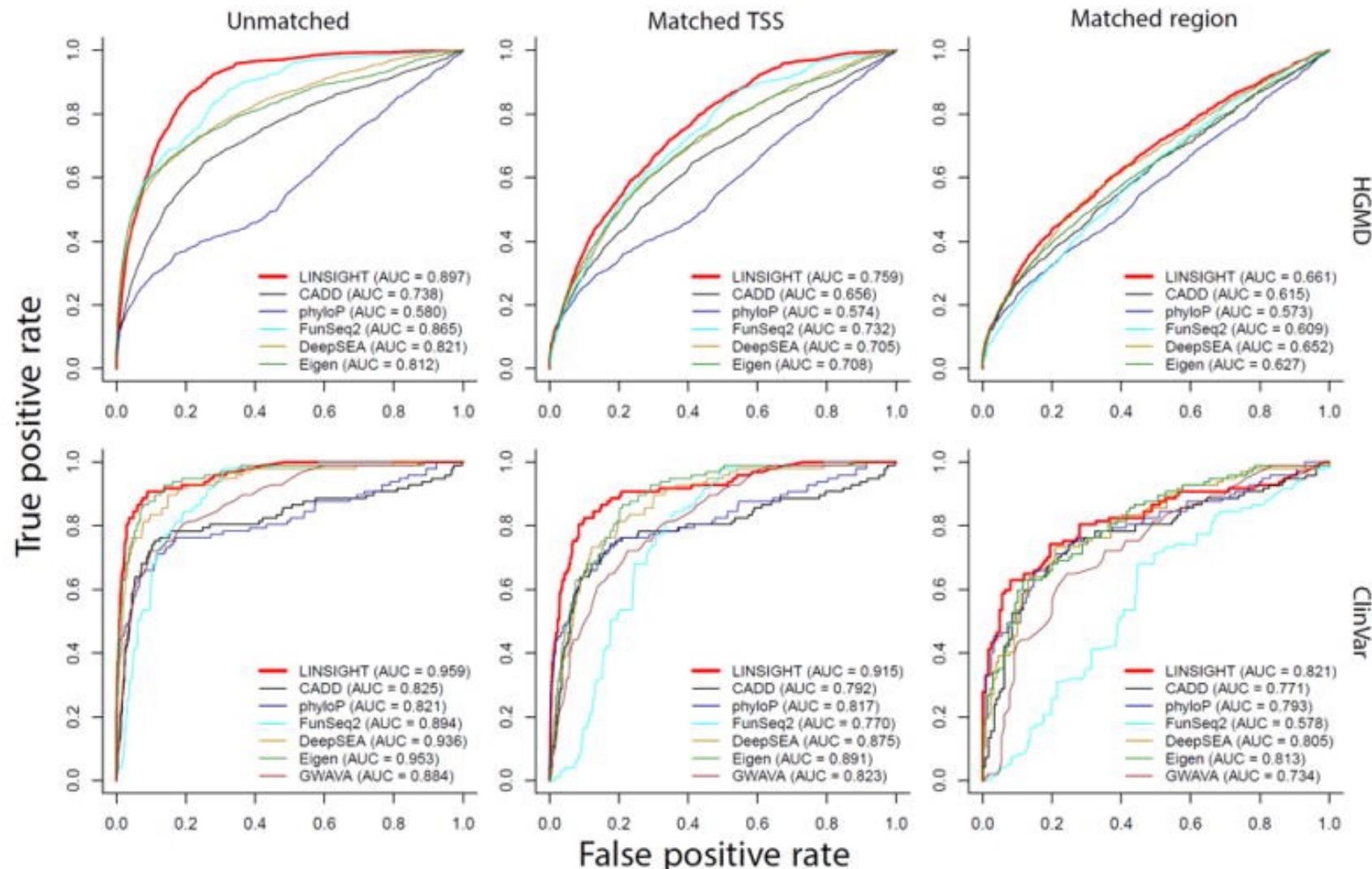


Fig. 3. Prediction power of various computational methods for distinguishing disease-associated noncoding variants from variants not likely to have phenotypic effects.



How to integrate such predicted functional annotations of non-coding variants with GWAS data?

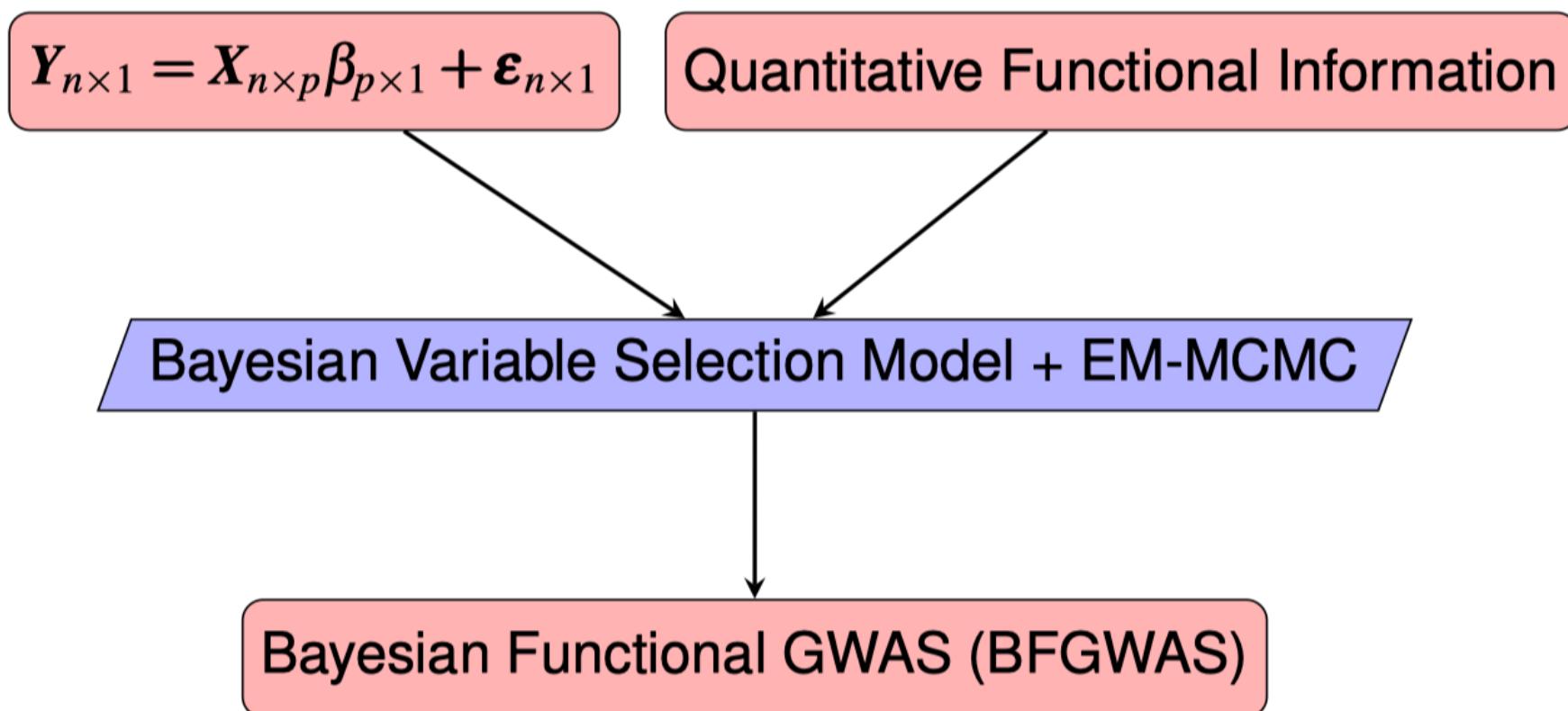
- Interpretate GWAS results?
- Link GWAS signals to function genes?
- Prioritize potential “causal” variants?

BFGWAS_QUANT: Bayesian Functional GWAS Method Accounting for Multivariate Quantitative Functional Annotations

Motivations

- Account for linkage disequilibrium (LD, correlation among genetic variants) to fine-map independent association
- Account for multivariate quantitative functional annotations to prioritize potential causal variants
- Use publicly available summary-level GWAS data of large sample sizes
- Understand biological mechanisms underlying genetic associations

BFGWAS Diagram



J. Yang et. al. AJHG 2017.

Bayesian Hierarchical Model Framework

Multivariable linear regression model with **Standardized** phenotype and genotype vectors:

$$Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1}, \quad \epsilon \sim MN(0, I). \quad (1)$$

Prior:

- $\beta_i \sim \pi_i N(0, \frac{1}{n} \tau_\beta^{-1}) + (1 - \pi_i) \delta_0(\beta_i); i = 1, \dots, p$
- With augmented quantitative functional annotation data vector $A_i = (1, A_{i,1}, \dots, A_{i,J})$ for variant $i = 1, \dots, p$,

$$\text{logit}(\pi_i) = A_i' \alpha; \quad \pi_i = \frac{e^{A_i' \alpha}}{1 + e^{A_i' \alpha}}; \quad \alpha = (\alpha_0, \alpha_1, \dots, \alpha_J)$$

- Introduce a latent indicator vector $\gamma_{p \times 1}$, equivalently

$$\gamma_i \sim \text{Bernoulli}(\pi_i), \quad \beta_{-\gamma} \sim \delta_0(\cdot), \quad \beta_\gamma \sim MVN_{|\gamma|}(0, \frac{1}{n} \tau_\beta^{-1} I_\gamma)$$

Shared Parameters by Genome-wide Variants

- Fix $\tau_\beta \in (0, 1]$: Inverse of effect size variance in the multivariable regression model.
 - $\tau_\beta = 1$ assumes same prior effect size variance as the marginal effect sizes.
 - $\tau_\beta < 1$ assumes larger magnitude for effect sizes in the multivariable model than the marginal ones.
- $\alpha_j \sim N(0, 1), j = 1, \dots, J$: Enrichment parameters
- Fix $\alpha_0 \in (-13.8, -9)$ to induce a sparse model.
 - $\alpha_0 = -13.8$ assumes prior causal probability 10^{-6} when $\alpha_j = 0, j = 1, \dots, J$.

Parameters of Interest

- Enrichment parameters:
 - $(\alpha_1, \dots, \alpha_J)$: for J annotations
- Variant-specific parameters (association evidence):
 - β_i : Effect-size
 - $\hat{\pi}_i = E[\gamma_i]$: Causal Posterior Probability (CPP)
- Region-level (Association evidence):
 - $\text{Regional CPP} = E[\max(\gamma_{i_1}, \dots, \gamma_{i_k})]$: Regional probability of having at least one causal variant
 - $\text{Sum CPP} = \sum_{i=1}^p \hat{\pi}_i I(\pi_i > 0.01)$: Expected number of causal SNPs

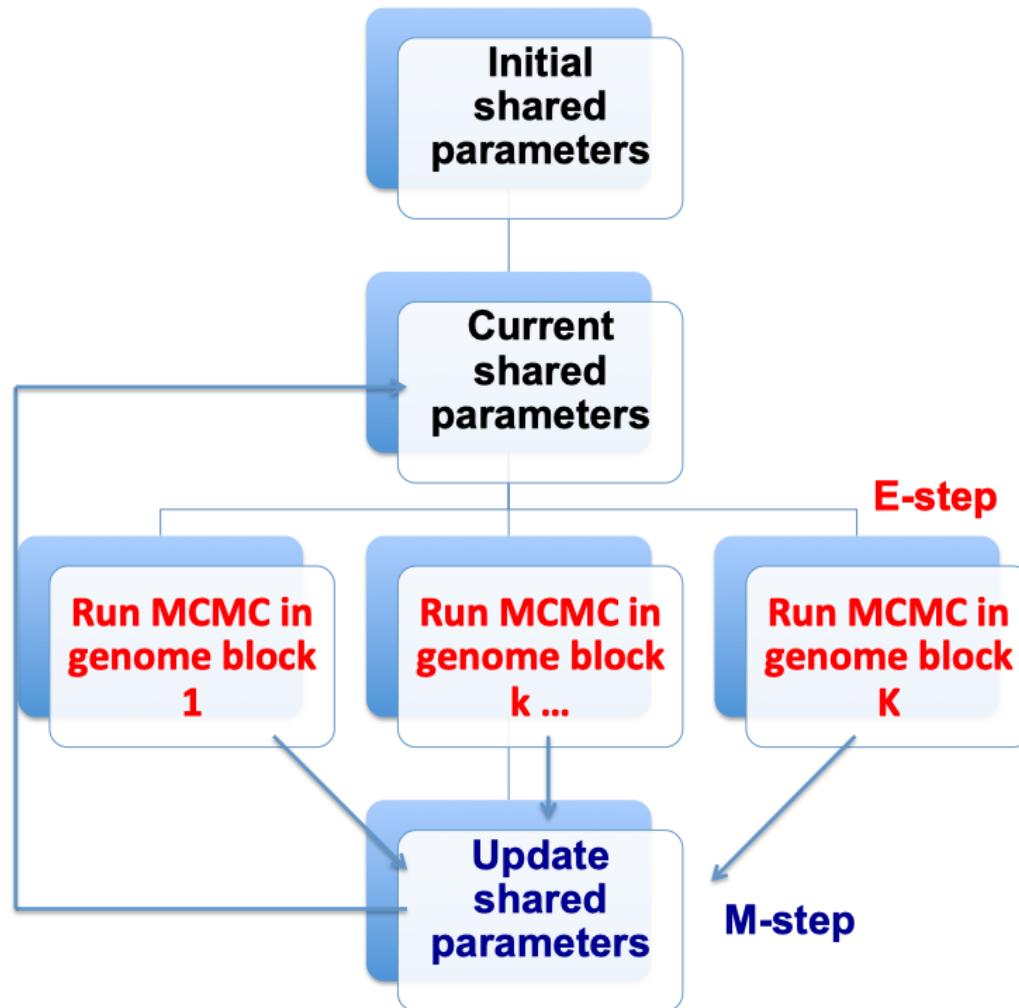
Bayesian Inference

- Joint posterior distribution

$$P(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi}(\boldsymbol{\alpha}) | Y, X, A) \propto (2)$$
$$P(Y|X, \boldsymbol{\beta}, \boldsymbol{\gamma}) P(\boldsymbol{\beta} | \boldsymbol{\gamma}, \tau_\beta) P(\boldsymbol{\gamma} | \boldsymbol{\pi}(\boldsymbol{\alpha}), A) P(\boldsymbol{\pi}(\boldsymbol{\alpha}))$$

- Product of Likelihood and Prior
- MCMC by Metropolis-Hastings algorithm with a proposal strategy for $\boldsymbol{\gamma}$
- Challenges of Standard MCMC: Memory Usage and Convergence Rate for $\sim 10M$ genome-wide variants

EM-MCMC Algorithm



Enable
Genome-wide
Analysis

Improve MCMC
Convergence
Rate

MCMC Algorithm

Given shared parameters $(\alpha_0, \alpha_1, \dots, \alpha_J), \tau_\beta$:

- Propose a new indicator vector γ
- Posterior conditional distribution for $\beta_{|\gamma|}$:

$$P(\beta_{|\gamma|} | Y, X, \gamma, \tau_\beta) \sim MVN_{|\gamma|}(\mu_{\beta_{|\gamma|}}, \Sigma_{\beta_{|\gamma|}});$$

$$\mu_{\beta_{|\gamma|}} = \Sigma_{\beta_{|\gamma|}} X^T Y, \Sigma_{\beta_{|\gamma|}} = \frac{1}{n} (R + \tau_\beta I_{m \times m})^{-1}, R = \frac{1}{n} X^T X$$

- Conditional posterior likelihood:

$$P(\gamma | Y, X, \pi(\alpha), \tau_\beta) \propto$$

$$\sqrt{|\Sigma_{\beta_{|\gamma|}}|} \cdot (n\tau_\beta)^{\frac{m}{2}} \cdot \exp \left\{ -\frac{n}{2} + \frac{1}{2} (X^T Y)^T \Sigma_{\beta_{|\gamma|}} (X^T Y) \right\} \cdot \prod_{i=1}^P P(\gamma_i | \pi(\alpha_i))$$

MCMC Algorithm

- Apply Metropolis-Hastings algorithm
- If accepted, update effect-size estimates:

$$\hat{\boldsymbol{\beta}}_{|\gamma|} = \boldsymbol{\mu}_{\boldsymbol{\beta}_{|\gamma|}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{|\gamma|}} \mathbf{X}^T \mathbf{Y}$$

- Reference LD and GWAS Summary statistics can be used to derive values for $(\mathbf{R}, \mathbf{X}^T \mathbf{Y})$ in the MCMC algorithm:
 - \mathbf{R} : Reference LD correlation matrix of the same ancestry
 - $\mathbf{X}_i^T \mathbf{Y} = \sqrt{n} Z_{score_i}$: Z_{score_i} is the single variant Z-score test statistic using standardized Y and X_i for SNP i

Update Enrichment Parameters by Maximum A Posteriori (MAP)

- The expected log-posterior-likelihood function of α :

$$l(\alpha) = E_{\gamma}[\ln(P(\alpha|\gamma, \mathbf{A}))] \propto \sum_{i=1}^p \left[\hat{\gamma}_i \ln \left(e^{A'_i \alpha} \right) - \ln \left(1 + e^{A'_i \alpha} \right) \right] - \frac{\alpha' \alpha}{2}$$

- Enrichment parameters α are estimated by using the following gradient and hessian functions:

$$\frac{dl(\alpha)}{d\alpha} = \sum_{i=1}^p \left[\hat{\gamma}_i A'_i - \left(1 + e^{-A'_i \alpha} \right)^{-1} A'_i \right] - \alpha'$$

$$\frac{d^2 l(\alpha)}{d\alpha d\alpha'} = - \sum_{i=1}^p \left[\frac{e^{-A'_i \alpha}}{(1 + e^{-A'_i \alpha})^2} (A_i A'_i) \right] - I$$

Parameters of Interest

- Significant causal SNPs with $CPP > 0.1068$ (equivalent to p-value $< 5 \times 10^{-8}$), effect size estimates $\hat{\beta}_i$ and posterior causal probability $\hat{\pi}_i$
- Estimates of enrichment parameters $(\alpha_1, \dots, \alpha_J)$
- Sum of CPP estimates for variants with $\hat{\pi}_i > 0.01$ estimates the number of expected GWAS signals

eQTL based Annotations

Derived from frontal cortex brain tissues and Microglia cells:

- **Allcis-eQTL**: Binary annotation denoting if a SNP is a significant cis-eQTL
- **95%CredibleSet**: Binary annotation denoting if a SNP is in within a fine-mapped 95% credible set of cis-eQTL by CAVIAR
- **MaxCPP**: Maximum cis-CPP per SNP across all genes
- **BGW_MaxCPP**: Maximum CPP (cis- or trans-) per SNP across all genes derived by our BGW-TWAS method
- **Microglia-eQTL** : Binary annotation denoting if a SNP is a significant cis-eQTL of Microglia cell type

Histone Modifications based Annotations

Derived from the epigenomics data in the brain mid frontal gyrus region from the ROADMAP Epigenomics database:

- H3K4me1 (primed enhancers)
- H3K4me3 (promoters)
- H3K36me3 (gene bodies)
- H3K27me3 (polycomb regression)
- H3K9me3 (heterochromatin)

All binary annotations denoting if the SNP is located in the peak regions of the above histone modifications.

Application Studies by BFGWAS_QUANT

Phenotype: Clinical diagnosis of AD dementia

Individual-level GWAS Data

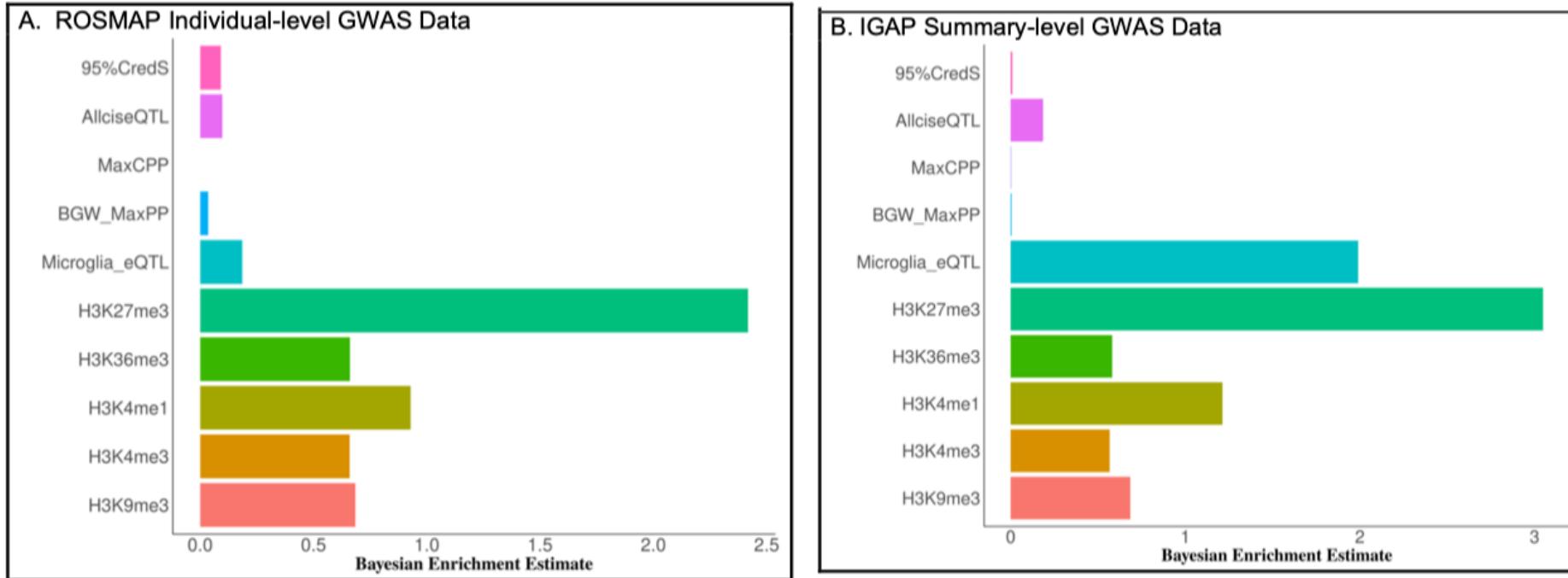
- Religious Orders Study and Rush Memory and Aging Project (ROS/MAP): 1,417 WGS samples of European ancestries
- Adjust for covariates: age, sex, smoking status, study index (ROS or MAP), and top 3 genotype PCs

IGAP Summary-level GWAS Data

- Meta-analysis summary data from 4 cohorts with ~ 54K samples.
- Reference LD was derived from the ROS/MAP WGS genotype data

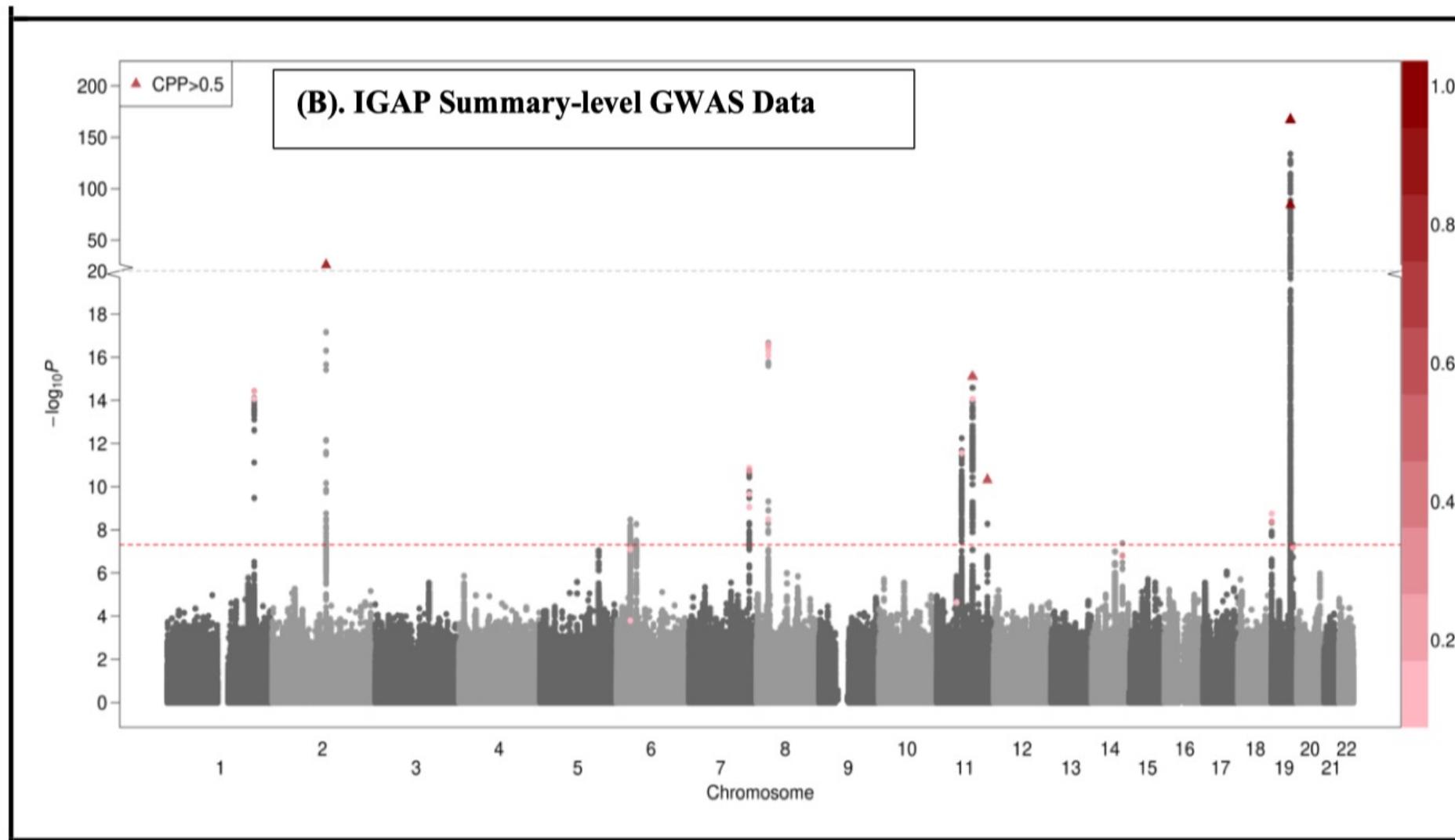
Considered 10 eQTL and histone modification based annotations

Estimates of Enrichment Parameters in Real Data



- **Microglia is a known related cell type in the brain for Alzheimer's disease.**
- **H3K27me3 is an epigenetic modification to the DNA packaging protein Histone H3, the tri-methylation of lysine 27 on histone H3 protein, which is associated with the downregulation of nearby genes via the formation of heterochromatic regions.**

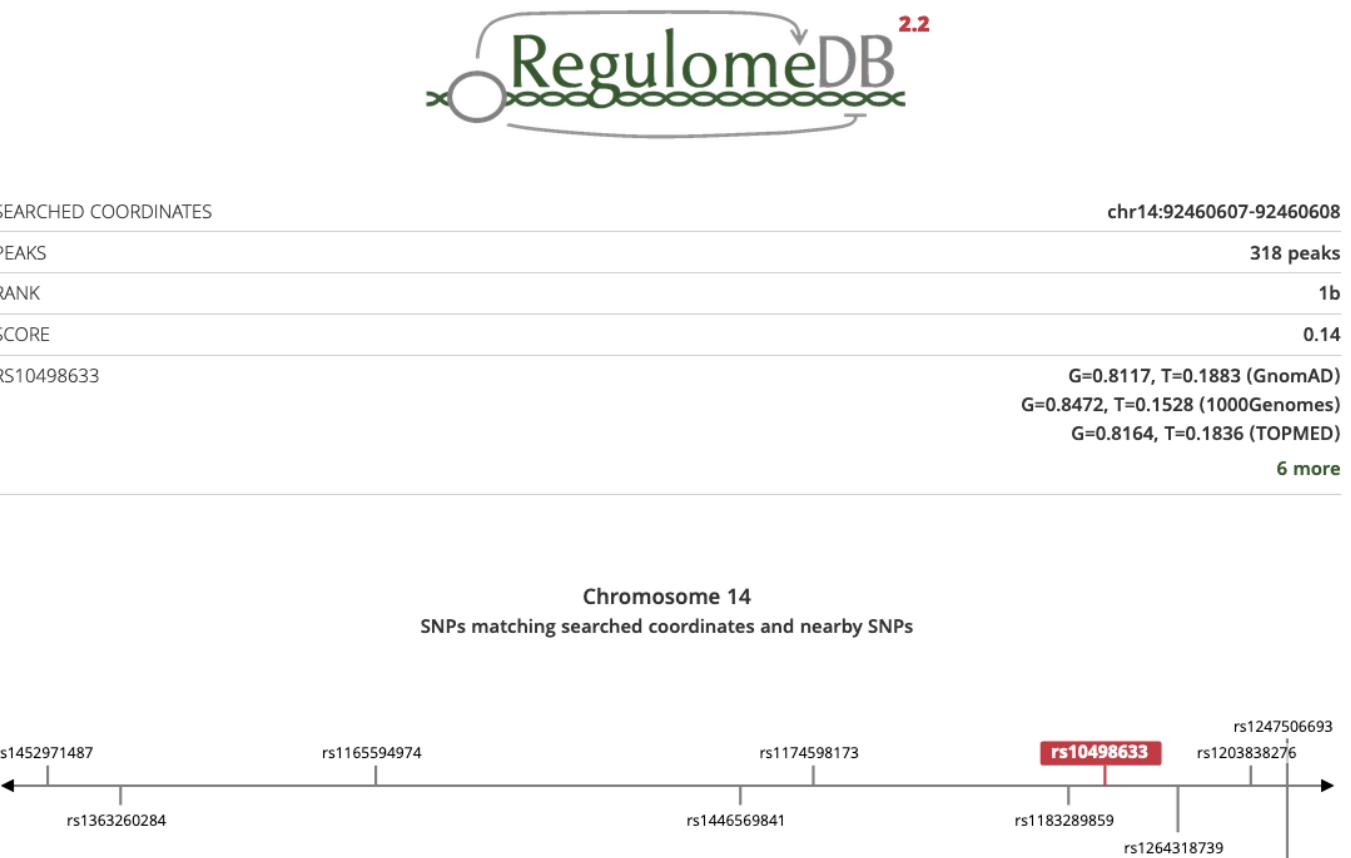
BFGWAS Results of AD Dementia



Sum_CPP=54.3

Table 2. Significant SNPs with Bayesian CPP > 0.1068 by BFGWAS_QUANT for studying AD using the IGAP summary-level GWAS data. SNPs with single variant test P-value > 5×10^{-8} were shaded in gray.

CHR	rsID	Gene	Function	CPP	Beta	P-value
1	rs6656401	CR1	Intron	0.119	-0.017	8.67E-15
1	rs7515905	CR1	Intron	0.206	-0.019	3.75E-15
1	rs1752684	CR1	Regulatory	0.125	-0.017	3.77E-15
1	rs679515	CR1	Intron	0.220	-0.018	3.60E-15
2	rs4663105	BIN1	Regulatory	0.631	0.050	1.26E-26
2	rs6733839	BIN1	Regulatory	0.796	0.053	1.24E-26
6	rs9270999	HLA-DRB1	Intron	0.181	0.001	8.04E-08
6	rs9273472	HLA-DRB1	Intron	0.110	0.074	1.63E-04
7	rs10808026	EPHA1	Intron	0.123	-0.020	1.36E-11
7	rs11762262	EPHA1	Intron	0.117	-0.011	2.21E-10
7	rs11763230	EPHA1	Intron	0.325	-0.020	1.86E-11
7	rs11771145	EPHA1	Intron	0.173	-0.021	8.69E-10
8	rs28834970	PTK2B	Intron	0.137	0.066	3.22E-09
8	rs2279590	CLU	Intron	0.166	0.021	4.47E-17
8	rs4236673	CLU	Intron	0.123	0.020	3.25E-17
8	rs11787077	CLU	Intron	0.247	0.022	2.94E-17
8	rs9331896	CLU	Intron	0.154	0.022	8.38E-17
8	rs2070926	CLU	Intron	0.278	0.023	2.69E-17
11	rs11039390	NUP160	Downstream	0.145	-0.004	2.31E-05
11	rs4939338	MS4A6E	Upstream	0.139	0.011	2.79E-12
11	rs7110631	PICALM	Intergenic	0.134	0.014	8.77E-15
11	rs10792832	RNU6-560P	Regulatory	0.633	0.027	7.89E-16
11	rs11218343	SORL1	Regulatory	0.643	-0.046	4.77E-11
14	rs10498633	SLC24A4	Intron	0.371	-0.059	1.55E-07
19	rs3752246	ABCA7	Missense	0.361	-0.027	4.27E-09
19	rs4147929	ABCA7	Regulatory	0.111	-0.030	1.77E-09
19	rs41289512	PVRL2	Regulatory	1.000	0.132	1.81E-167
19	rs6857	PVRL2	3' UTR	1.000	0.359	0
19	rs769449	APOE/TOMM40	Regulatory	1.000	0.292	0
19	rs56131196	APOC1	Regulatory	1.000	0.251	0
19	rs78959900	APOC1	Downstream	1.000	-0.096	8.22E-85
19	rs12459419	CD33	Missense	0.245	-0.027	6.66E-08



eQTL Data

eQTL data						
Method	▲ QTL location	▼ Biosample	▼ Target genes	▼ Dataset	▼ File	▼
eQTLs	chr14:92460607..92460608	subcutaneous adipose tissue	SLC24A4	ENCSR194AZV	ENCFF202BMV	
eQTLs	chr14:92460607..92460608	fibroblast derived cell line	RIN3	ENCSR272FLO	ENCFF705GVN	
eQTLs	chr14:92460607..92460608	venous blood	SLC24A4	ENCSR785ONK	ENCFF396TJI	

RIN3 Binds to BIN1 and CD2AP to increase APP-CTFs in early endosomes

Developing topics

Ruinan Shen✉, Chengbiao Wu

First published: 07 December 2020 | <https://doi.org/10.1002/alz.047161> |

Citations: 1

 PDF  TOOLS  SHARE

Abstract

Background

About one-third of the genes related to Alzheimer's disease (AD) identified by GWAS encode proteins that function predominantly in the endocytic pathways. Among them, the Ras and Rab Interactor 3, RIN3 is a guanine nucleotide exchange factors (GEFs) for the Rab5 small GTPase family and has been implicated to be a risk factor for both late onset AD (LOAD) and sporadic early onset AD (sEOAD). However, how RIN3 is linked to AD pathogenesis is currently undefined.

Methods

Quantitative PCR and immunoblotting were used to measure gene expression levels in mouse brain, cultured basal forebrain cholinergic neuron (BFCNs) or PC12 cells. Immunostaining was used to define subcellular localization of RIN3 and visualize endosomal changes in cultured primary BFCNs and PC12 cells. Recombination RIN3-flag-tagged protein was purified from HEK293T cells and was used to define RIN3-interactomes by mass spectrometry. Live imaging of primary neurons was used to examine axonal transport.

Results

Expression of RIN3 was upregulated in the brain tissues of young APP/PS1 mice. The increase was specific for the hippocampus and cortex since RIN3 expression level did not show an increase in either the olfactory bulb or the cerebellum. RIN3 mRNA level was also found to be elevated in embryonic day (E18) 18 in primary basal forebrain cholinergic neurons (BFCNs) cultured from APP/PS1 mouse.

Concomitant early endosome enlargement was seen in these BFCNs.

We have demonstrated that RIN3 interacted with two other AD risk factors: BIN1(bridging integrator 1) and CD2AP (CD2 associated protein). Both BIN1 and CD2AP were recruited to early endosomes through binding to the proline rich domain (PRD) of RIN3.

Overexpression of RIN3 or CD2AP promoted APP cleavage to increase APP carboxyl terminal fragments (CTFs) in PC12 cells, while upregulating RIN3 or BIN1 neuronal isoform increase phosphorylated Tau level. The effect of upregulation of RIN3 expression to increase APP CTFs was ameliorated by a dominant negative Rab5 construct (Rab5^{S34N}). Increased expression of RIN3 impaired axonal trafficking of both APP and BACE1.

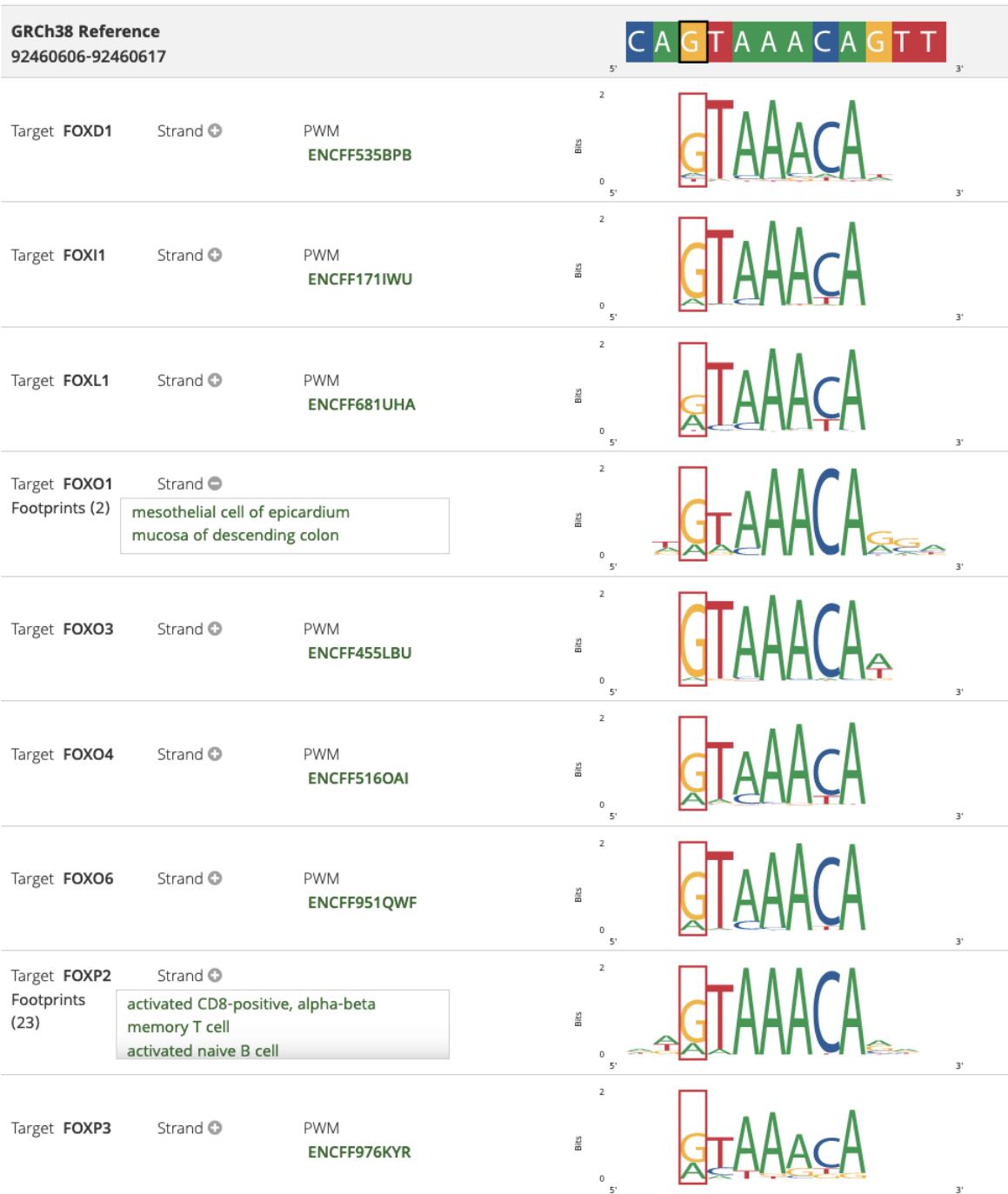
Conclusion

RIN3 is significantly unregulated and correlated with endosomal dysfunction in APP/PS1 mouse. Through interacting with BIN1 and CD2AP, increased RIN3 expression alters axonal trafficking and procession of APP.

ChIP Data

ChIP data							
Method ▾	Peak location	▼ Biosample	▼ Targets ▾	Organ	▼ Dataset	▼ File	▼ Value ▾
ChIP-seq	chr14:92460092..92460692	endothelial cell	CTCF	epithelium	ENCSR031PXV	ENCFF509KKI	31.92589
ChIP-seq	chr14:92460324..92460924	endothelial cell	CTCF	epithelium	ENCSR031PXV	ENCFF509KKI	7.87005
ChIP-seq	chr14:92460137..92460721	HCT116	CTCF	large intestine, epithelium, colon, intestine	ENCSR048RGR	ENCFF470EAN	34.66111
ChIP-seq	chr14:92460116..92460656	HCT116	CTCF	large intestine, epithelium, colon, intestine	ENCSR107GWZ	ENCFF673EMS	32.50503
ChIP-seq	chr14:92460042..92460758	CD14-positive monocyte	CTCF	bodily fluid, blood	ENCSR162KZY	ENCFF476EKR	20.86036
ChIP-seq	chr14:92460145..92460635	type B pancreatic cell	CTCF	endocrine gland, epithelium, pancreas	ENCSR458PYQ	ENCFF190FKP	13.86670
ChIP-seq	chr14:92460307..92460747	A549	SMC3	lung	ENCSR620NWG	ENCFF511SKA	7.15090
ChIP-seq	chr14:92460004..92460708	activated CD4-positive, alpha-beta T cell	CTCF	bodily fluid, blood	ENCSR806WWX	ENCFF951RPD	35.31633
ChIP-seq	chr14:92460180..92460724	HCT116	CTCF	large intestine, epithelium, colon, intestine	ENCSR954MUZ	ENCFF498FLW	16.87120

Motifs



By TFBSShape Webtool

<https://tfbsshape.usc.edu/Detail/jaspar/MA0139.1>



Web Resources

- DEEPSEA Webtool:
 - <https://hb.flatironinstitute.org/deepsea/>
- LINSIGHT (Linux C++ tool):
 - <https://github.com/CshlSiepelLab/LINSIGHT>
- BFGWAS_QUANT
 - https://github.com/yanglab-emory/BFGWAS_QUANT