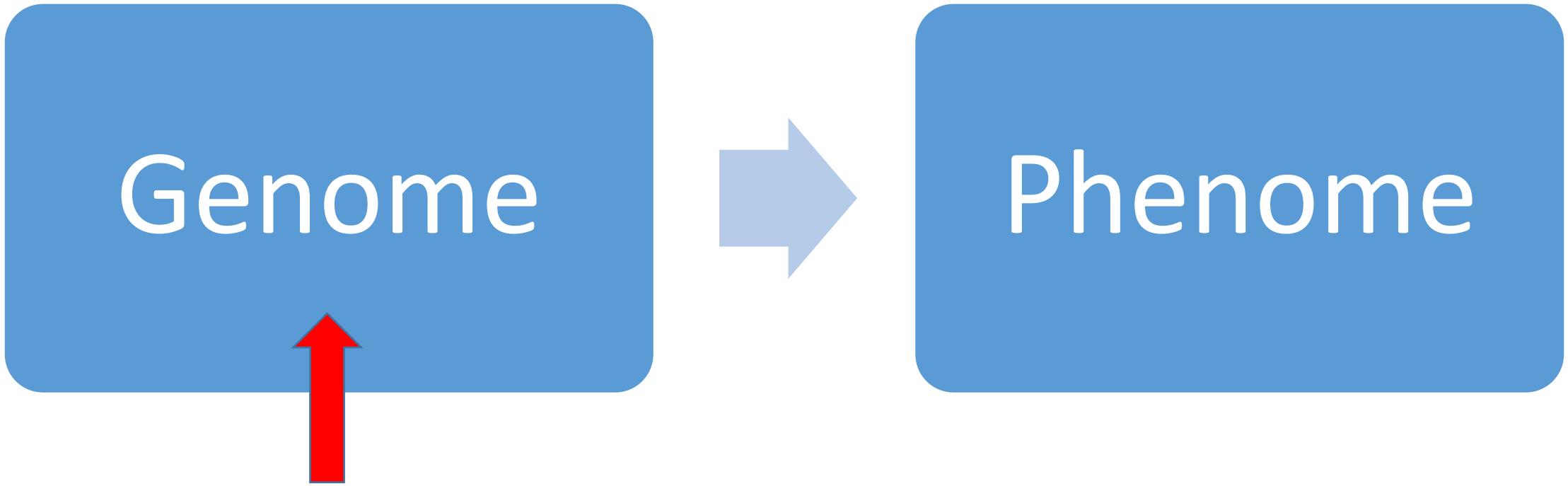
Functional Annotation in Genomics

Christian Maltecca

Motivations



Introduction to Functional Annotation

- **What is Functional Annotation?**
  - Functional annotation involves assigning biological information to genomic features, such as genes, SNPs, or other genomic regions.
  - Helps in understanding the biological role of genetic variants and their impact on phenotypic traits.
- **Why is Functional Annotation Important?**
  - Provides insights into the molecular mechanisms underlying complex traits and diseases.
  - Facilitates the interpretation of results from genome-wide association studies (GWAS) and other genomic analyses.
  - Aids in identifying potential therapeutic targets and biomarkers for disease.

Functional annotations

- External information about a variant in addition to its genotypes
  - https://useast.ensembl.org/info/genome/index.html

- Location
  - Whether SNP is within a specific genomic element
  - There are many resources to define genomic elements.
- Quantification
  - Pathogenicity scores
  - https://useast.ensembl.org/info/genome/variation/prediction/protein_function.html

Types of Functional Annotations

- **Gene-Centric Annotations:**
  - **Gene Ontology (GO):** Provides a controlled vocabulary to describe gene products in terms of their associated biological processes, cellular components, and molecular functions.
  - **Pathway Annotations:** Describes genes in the context of biochemical pathways, such as KEGG or Reactome pathways.
- **Regulatory Annotations:**
  - **Promoters and Enhancers:** Regions that regulate gene expression; identified through methods like chromatin immunoprecipitation (ChIP-seq).
  - **Transcription Factor Binding Sites (TFBS):** Locations where transcription factors bind to regulate gene expression.
- **Epigenomic Annotations:**
  - **Histone Modifications:** Modifications to histone proteins that affect chromatin structure and gene expression (e.g., H3K27ac marks active enhancers).
  - **DNA Methylation:** Addition of methyl groups to DNA, affecting gene expression without changing the DNA sequence.

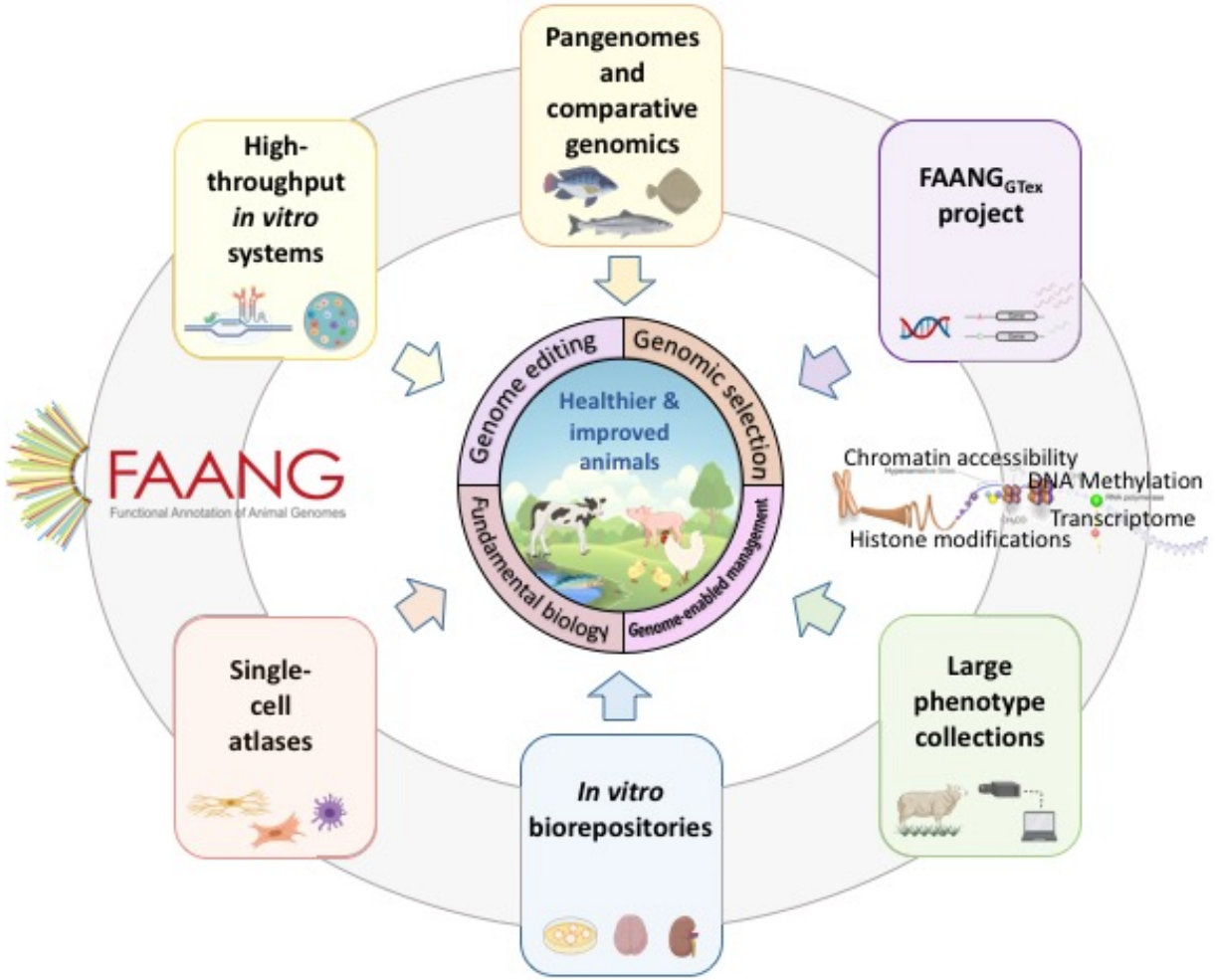
Methods for Functional Annotation

- **Experimental Methods:**
  - **ChIP-seq (Chromatin Immunoprecipitation Sequencing):** Used to identify DNA-binding sites of proteins, such as transcription factors or histones.
  - **RNA-seq (RNA Sequencing):** Provides information on gene expression levels and splicing variants.
- **Computational Methods:**
  - **Sequence Homology-Based Annotation:** Uses similarity to known sequences (e.g., BLAST) to infer function.
  - **Machine Learning Approaches:** Utilizes algorithms to predict functional elements based on genomic features (e.g., DeepSEA).
- **Integrated Approaches:**
  - **ENCODE Project:** Provides a comprehensive map of functional elements in the human genome, integrating multiple experimental data types.
  - **Roadmap Epigenomics Project:** Focuses on characterizing the epigenomic landscape across different cell types and tissues.

Tools and Databases for Functional Annotation

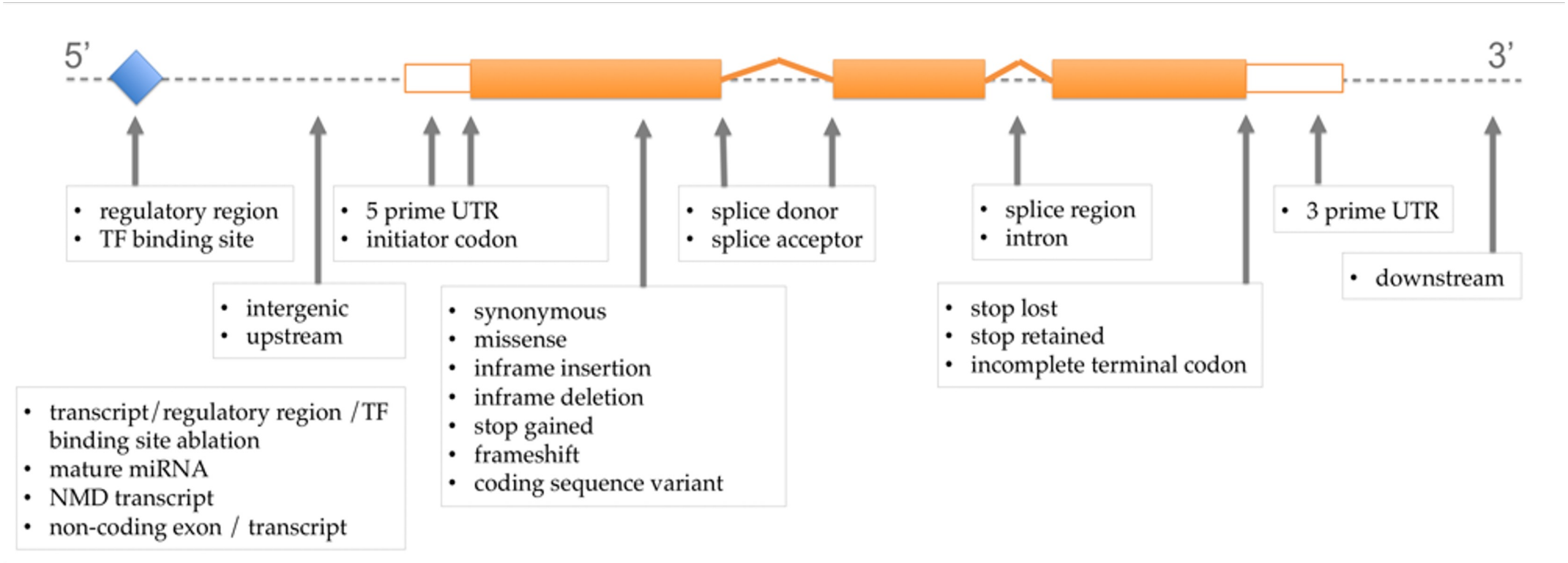
- **UCSC Genome Browser:**
  - Offers a wide range of annotations, including genes, regulatory elements, and epigenomic data.
- **Ensembl:**
  - Provides comprehensive genome annotation, including genes, variants, regulatory regions, and comparative genomics data.
- **GREAT (Genomic Regions Enrichment of Annotations Tool):**
  - Associates genomic regions with biological functions by leveraging annotations from GO, pathways, and other sources.
- **RegulomeDB:**
  - Integrates various types of functional annotation data to score regulatory potential of non-coding variants.
- **dbSNP and ClinVar:**
  - Provide annotations on known genetic variants, including their functional consequences and clinical significance.

Functional Annotation of Animal Genomes

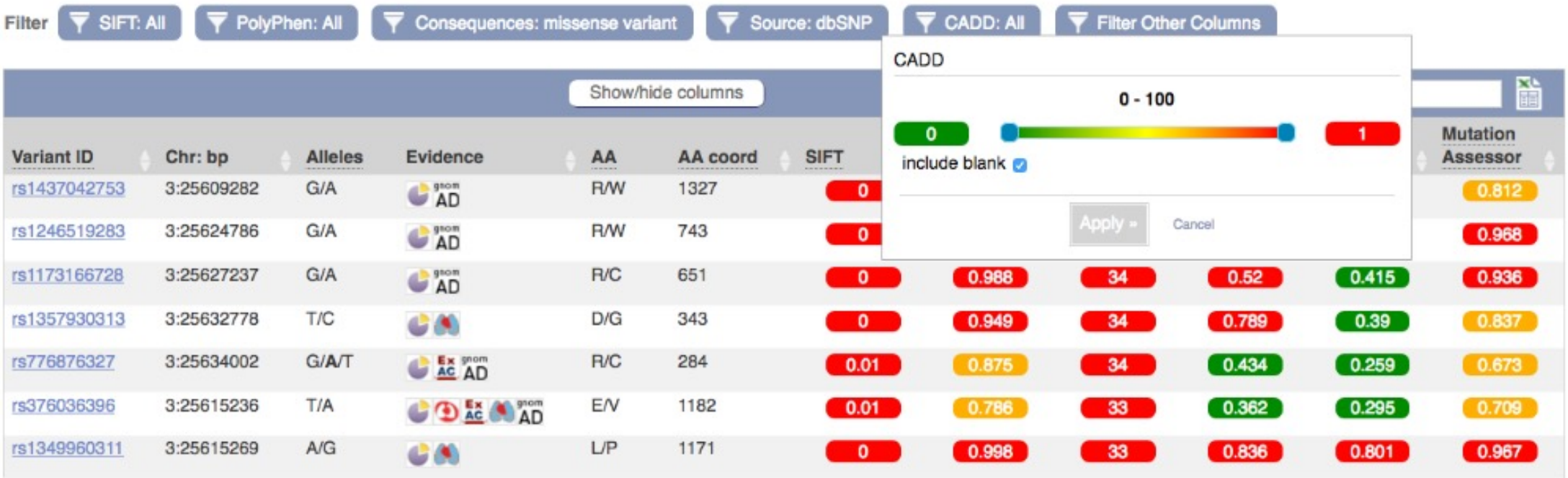


https://www.animalgenome.org/community/FAANG/

Location-based



Quantification-based



Categorical annotations

- https://useast.ensembl.org/info/genome/variation/prediction/predicted_data.html

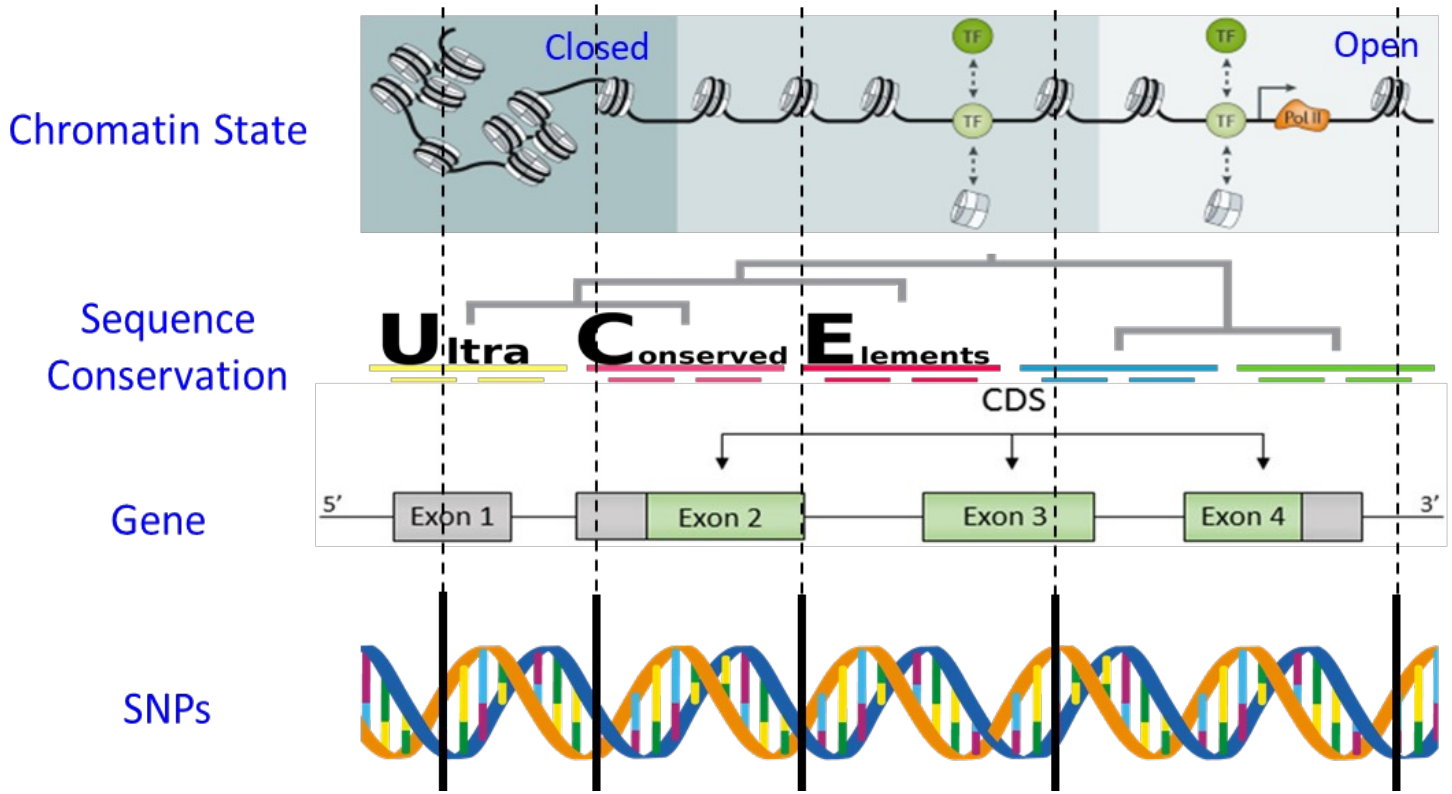
*	SO term	SO description	SO accession	Display term	IMPACT
	stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
	frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
	stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
	start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
	inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
	inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
	missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE
	protein_altering_variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	SO:0001818	Protein altering variant	MODERATE
	splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice region variant	LOW
	splice_donor_region_variant	A sequence variant that falls in the region between the 3rd and 6th base after splice junction (5' end of intron)	SO:0002170	Splice donor region variant	LOW

- Often binary
  - eQTL: whether a SNP is an identified eQTL. Yes or no.
  - CDS: whether a SNP is within CDS. Yes or no.

Continuous annotations

- Less commonly used than categorical annotations.
- Can be categorized.
- Minor allele frequency
  - <0.01, 0.01-0.05, 0.05-0.10, etc
- Conservation score
  - Constrained element
  - https://useast.ensembl.org/info/genome/compara/conservation_and_constrained.html

Link functional annotations to SNPs



Yes or no: whether a SNP is within a functional annotation category

Theory

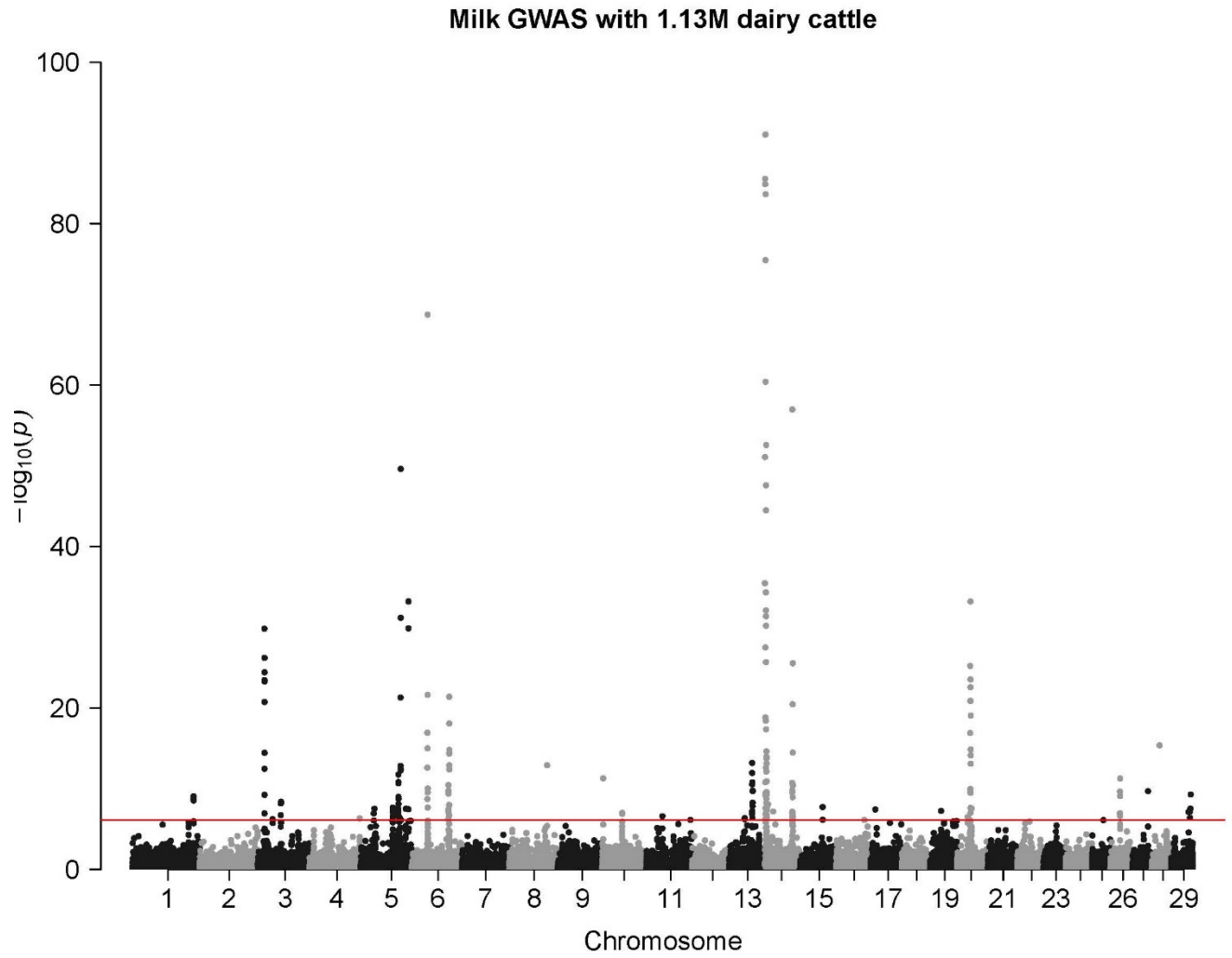
Genetic Architecture of Complex Traits

- Limit of GWAS

H_0: SNP effect size drawn from a normal distribution $N(0, \sigma_g^2/M)$.

M: effective number of independent markers

- Proportion of genetic variance explained by significant SNPs ≈ 18.3%



Genetic Architecture of Complex Traits (*cont.*)

- Omnigenic model
  - A small number of "core" genes
    - SNP effect size beyond polygenic effect $N(0, \sigma_g^2/M)$
  - A lot more "peripheral" genes
    - Contributing much more of heritability
    - SNP effect size NOT beyond polygenic effect $N(0, \sigma_g^2/M)$
- Core genes are easy to deal with in genomic predictions.
  - Bayesian mixture models
    - BayesB, BayesR, BSLMM
  - Fixing big QTL effects
    - LDAK
- Difficult to model peripheral genes better than GBLUP

Modeling Functional Annotations

	Annot1	Annot2	Annot3	Annot4
SNP1	1	1	0	0
SNP2	0	1	1	0
SNP3	1	1	1	1

W

$$\sigma^2_{\text{SNP1}} = \tau_1 + \tau_2$$
$$\sigma^2_{\text{SNP2}} = \tau_2 + \tau_3$$
$$\sigma^2_{\text{SNP3}} = \tau_1 + \tau_2 + \tau_3 + \tau_4$$

=Wτ

$\text{Var}(\alpha_j) = \sum_{k:j\in S_k} \tau_k$, where τ_k denotes the per-variant contribution of category k to $\text{Var}(\alpha_j)$.

Linear Mixed Model with Functional Annotations

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{D})$$

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R}\sigma_e^2)$$

$$\mathbf{D} = \begin{bmatrix} \sigma^2_{\alpha_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{\alpha_M} \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2_{\alpha_1} & \sigma^2_{\alpha_2} & \cdots & \sigma^2_{\alpha_M} \end{bmatrix}^{\mathrm{T}} = \mathbf{W}\boldsymbol{\tau}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g} + \mathbf{e}$$

$$\mathbf{g} \sim N(\mathbf{0}, \sum_{k=1}^{K} \mathbf{G}_k M_k \tau_k)$$

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R}\sigma_e^2)$$

where $\mathbf{g} = \mathbf{Z}\boldsymbol{\alpha}$, $\mathbf{G}_k = \mathbf{Z}\mathbf{W}^{(k)}\mathbf{Z}^{\mathrm{T}}/M_k$

$\mathbf{W}^{(k)}$ is a diagonal matrix whose diagonal elements are **W**'s k^{th} column.

Integrative analysis becomes VC estimation.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g} + \mathbf{e}$$

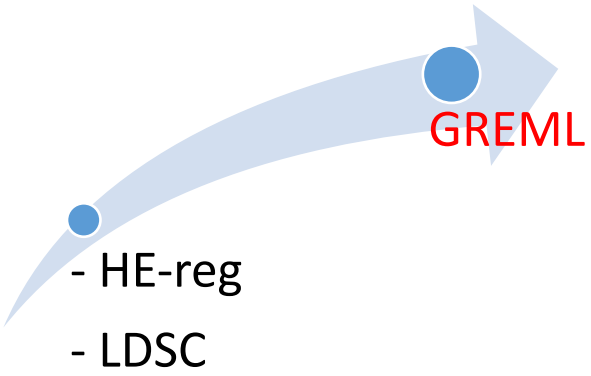
$$\mathbf{g} \sim N(\mathbf{0}, \sum_{k=1}^{K} \mathbf{G}_k M_k \tau_k)$$

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R}\sigma_e^2)$$

- We may need to model many functional annotations, so *K* is large.
- VC estimates (τ_k) may be negative.

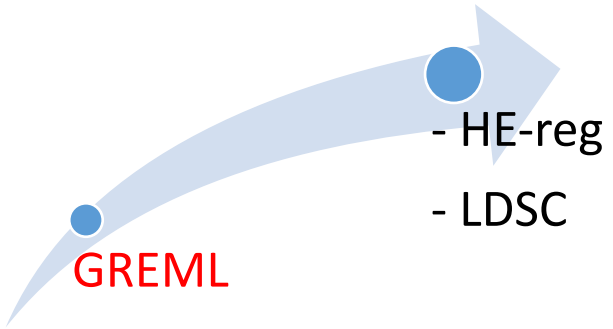
Existing VC estimation methods

Statistical efficiency



- HE or LDSC is not statistically efficient.

Stability and scalability



- GREML may fail to converge for many VCs.
- GREML may be slow for large samples.

Linkage Disequilibrium Score Regression

Introduction to Linkage Disequilibrium (LD) Score Regression

- **What is Linkage Disequilibrium (LD)?**
  - Linkage Disequilibrium (LD) refers to the non-random association of alleles at different loci. It indicates that certain combinations of alleles occur together more frequently than would be expected by chance.
  - Influenced by factors such as genetic drift, selection, mutation, recombination, and population structure.
- **What is LD Score Regression?**
  - A statistical method used to partition heritability and identify confounding biases in genome-wide association studies (GWAS).
  - Utilizes LD patterns to differentiate true polygenic signals from confounding biases in GWAS data.

Understanding LD Score

- **LD Score Definition:**
  - The LD score of a single nucleotide polymorphism (SNP) is the sum of squared correlations (r^2) between the SNP and all other SNPs within a certain genomic window.
  - Quantifies the amount of genetic variation captured by a SNP.

$$\text{LD Score } (l_j) = \sum_{i=1}^{M} r_{ij}^2$$

where (r_{ij}^2) is the squared correlation between SNP (j) and SNP (i), and (M) is the number of SNPs within the specified window.

- **Why is LD Score Important?**
  - SNPs with high LD scores tag more genetic variation, aiding in differentiating between true signal and noise in association studies.

How LD Score Regression Works

- **Core Idea:**
  - In a polygenic trait, SNPs with higher LD scores should, on average, have higher GWAS test statistics (chi-squared statistics) because they tag more causal variants.
- **Key Components:**
  - **Dependent Variable:** GWAS test statistics (chi-squared values) for each SNP.
  - **Independent Variable:** LD scores for each SNP.
- **Regression Model:**

$$E\left[\chi_j^2\right] = 1 + \frac{N \cdot h^2 \cdot l_j}{M} + \alpha$$

where:

- $\left(E\left[\chi_j^2\right]\right)$ is the expected chi-squared statistic for SNP (j),
- (N) is the sample size,
- (h^2) is the heritability explained by the SNPs,
- $\left(l_j\right)$ is the LD score of SNP (j),
- (M) is the number of SNPs, and
- (α) is an intercept term representing confounding bias.

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, \qquad \boldsymbol{\alpha} \sim N(0, \mathbf{I}\sigma_\alpha^2) \qquad \mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$$

$$\frac{\mathbf{Z}^{\mathrm{T}}\mathbf{y}}{\sqrt{N\sigma_p^2}} \sim N\left(0, \frac{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\sigma_\alpha^2}{N\sigma_p^2} + \frac{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\sigma_e^2}{N\sigma_p^2}\right),$$

$$\mathbf{s} \sim N\Big(0, N\mathbf{C}^2h^2/M + \mathbf{C}(1-h^2)\Big),$$

$$\mathbf{s} \sim N\Big(0, \mathrm{diag}\Big(N\mathbf{C}^2h^2/M + \mathbf{C}(1-h^2)\Big)\Big);$$

$$s_j \sim N\Big(0, N\ell_j h^2/M + (1-h^2)\Big),$$

Applications of LD Score Regression

- **Estimating Heritability:**
  - Used to estimate the heritability of complex traits by partitioning genetic variance from GWAS summary statistics.
- **Detecting Confounding Bias:**
  - Intercept (α) in the regression can identify confounding biases such as population stratification or cryptic relatedness.
- **Improving GWAS Results:**
  - Differentiates true polygenic signals from spurious associations caused by confounding factors.
- **Understanding Genetic Architecture:**
  - Provides insights into the genetic architecture of complex traits by assessing the genetic variance explained by SNPs across the genome.

Pros and Cons of LD Score Regression

- **Advantages:**
  - **Robust to Confounding:** Effectively identifies and corrects for population stratification and other confounders in GWAS data.
  - **Simple and Scalable:** Uses summary statistics and is computationally efficient for large-scale GWAS.
  - **Broad Applicability:** Applicable to both quantitative traits and case-control studies.
- **Limitations:**
  - **Requires Large Sample Sizes:** Accurate estimates of heritability and bias detection need large GWAS sample sizes.
  - **Assumptions:** Assumes GWAS results are polygenic and LD patterns are well-captured, which may not always hold.
  - **Sensitive to LD Score Calculation:** Accuracy depends on the quality of LD score calculations, which can vary across populations and reference panels.

References

1. Bulik-Sullivan, B., Loh, P.-R., Finucane, H. K., et al. (2015). *LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies*. Nature Genetics, 47, 291–295.
2. Finucane, H. K., Bulik-Sullivan, B., Gusev, A., et al. (2015). *Partitioning Heritability by Functional Annotation Using Genome-Wide Association Summary Statistics*. Nature Genetics, 47, 1228–1235.
3. Lee, J. J., Wedow, R., Okbay, A., et al. (2018). *Gene Discovery and Polygenic Prediction from a Genome-wide Association Study of Educational Attainment in 1.1 Million Individuals*. Nature Genetics, 50, 1112–1121.

Partitioning Heritability using GREML

Introduction to GREML

- **What is GREML?**
  - Genomic-Relatedness-Based Restricted Maximum Likelihood (GREML) is a statistical method used to estimate the proportion of phenotypic variance that is attributable to genetic variance (heritability).
  - GREML utilizes information from genomic data to partition heritability into components explained by different sets of genetic variants, often grouped by their minor allele frequency (MAF), functional annotation, or chromosomal location.
- **Applications of GREML:**
  - Understanding the genetic architecture of complex traits.
  - Estimating heritability attributable to specific genetic components.
  - Identifying the contribution of different genomic regions or functional annotations to the overall genetic variance of a trait.

The Basics of Heritability and GREML

- **Heritability h^2**:
  - The proportion of the total phenotypic variance of a trait that can be attributed to genetic variance.
  - Broad-sense heritability (H^2) includes all genetic variance components, while narrow-sense heritability (h^2) focuses only on additive genetic variance.
- **GREML Framework:**
  - Uses a linear mixed model to partition phenotypic variance into genetic and environmental components.
  - The model estimates the genetic variance explained by SNPs captured in a genomic relationship matrix (GRM).

$$y = X\beta + g + e$$

where:

- (y) is the vector of phenotypic values,
- (X) is the matrix of fixed effects,
- (β) is the vector of fixed effect coefficients,
- (g) is the vector of genetic effects (random effects),
- (e) is the vector of residual (environmental) effects.

Constructing the Genomic Relationship Matrix (GRM)

- **Genomic Relationship Matrix (GRM):**
  - Represents the genetic similarity between all pairs of individuals based on their SNP genotypes.
  - The GRM is constructed using genotype data to estimate the proportion of the genome shared identical-by-state (IBS) between individuals.

$$\text{GRM}_{ij} = \frac{1}{M}\sum_{k=1}^{M}\frac{(x_{ik} - 2p_k)(x_{jk} - 2p_k)}{2p_k(1-p_k)}$$

where:

- (M) is the number of SNPs,
- (x_{ik}) and (x_{jk}) are the genotypes for individuals i and j at SNP k,
- (p_k) is the allele frequency of SNP k.

Partitioning Heritability with GREML

- **Objective:**
  - Partition the heritability of a complex trait into components attributable to different sets of SNPs (e.g., by MAF, functional category, or genomic region).
- **GREML Model for Partitioning:**

$$V(y) = \sigma_g^2 K + \sigma_e^2 I$$

where:

- $V(y)$ is the phenotypic variance-covariance matrix,
- σ_g^2 is the additive genetic variance explained by SNPs,
- K is the GRM, capturing genetic relationships,
- σ_e^2 is the residual variance,
- I is the identity matrix.

Steps in GREML Analysis

1. **Prepare Data:**
   - Obtain SNP genotype data and phenotype data.
   - Quality control (QC) steps to filter SNPs and individuals (e.g., removing SNPs with low MAF, high missingness, or poor quality).
2. **Construct the GRM:**
   - Calculate the GRM using quality-controlled SNPs.
3. **Fit the GREML Model:**
   - Use software like GCTA, BOLT-LMM, or LDAK to fit the linear mixed model and estimate variance components.
4. **Partition Heritability:**
   - Decompose total genetic variance into components explained by different sets of SNPs (e.g., different MAF bins or functional categories).

Interpretation of GREML Results

- **Total Heritability Estimate:**
  - The overall proportion of phenotypic variance explained by all SNPs in the model.
- **Partitioned Heritability Estimates:**
  - Heritability explained by specific subsets of SNPs, allowing for interpretation of the contribution of various genetic factors.
- **Implications for Genetic Architecture:**
  - Helps in understanding the distribution of genetic effects across the genome.
  - Provides insights into whether certain genomic regions or functional annotations contribute disproportionately to the trait’s heritability.

Advantages and Limitations of GREML

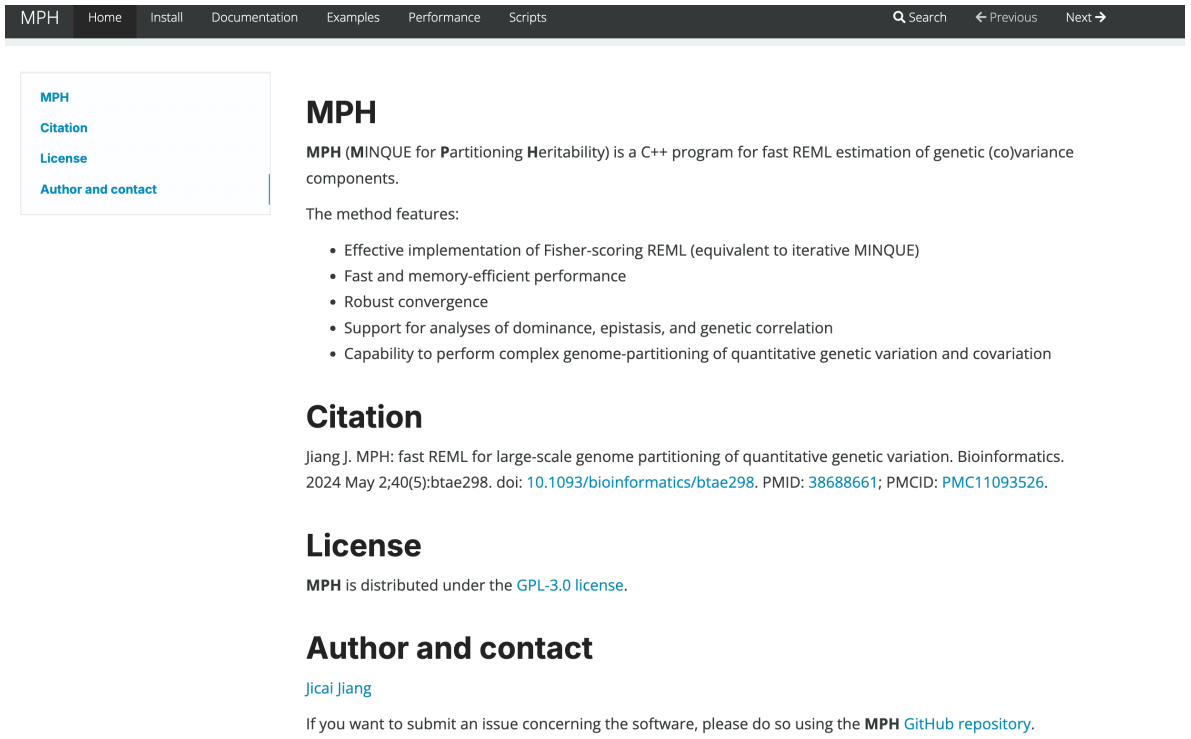
- **Advantages:**
  - Allows estimation of SNP-based heritability using only genotype data.
  - Can partition heritability based on functional annotations or other criteria.
  - Robust to population stratification if properly accounted for in the GRM.
- **Limitations:**
  - Requires large sample sizes for accurate heritability estimates.
  - Sensitive to the quality and representativeness of the reference population used to calculate the GRM.
  - Assumes a polygenic model, where many variants of small effect contribute to the trait.

References

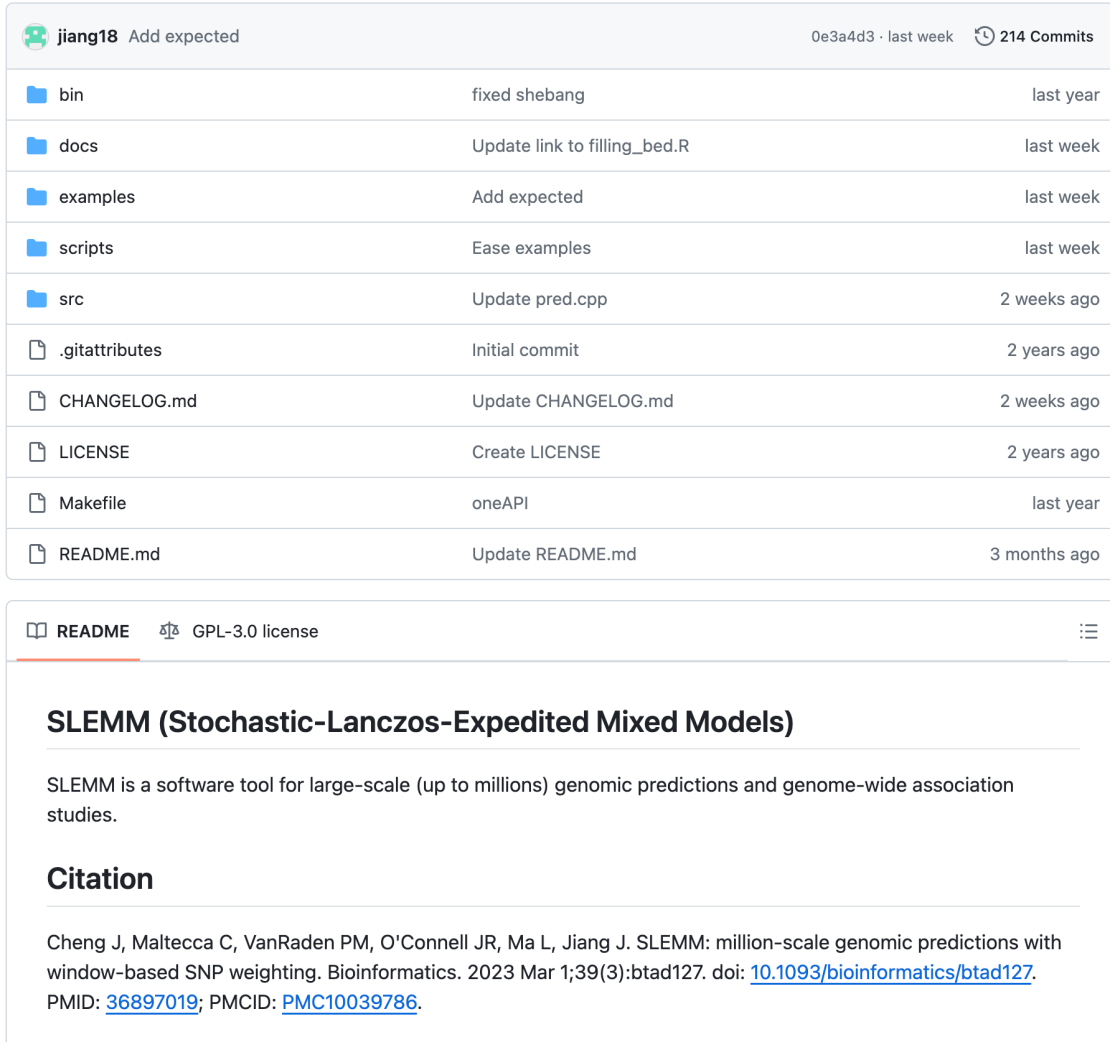
1. Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics*, 88(1), 76-82.
2. Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. *American Journal of Human Genetics*, 88(3), 294-305.
3. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and Pitfalls in the Application of Mixed-Model Association Methods. *Nature Genetics*, 46(2), 100-106.

Jicai Jiang NC State University

https://jiang18.github.io/mph/



https://github.com/jiang18/slemm

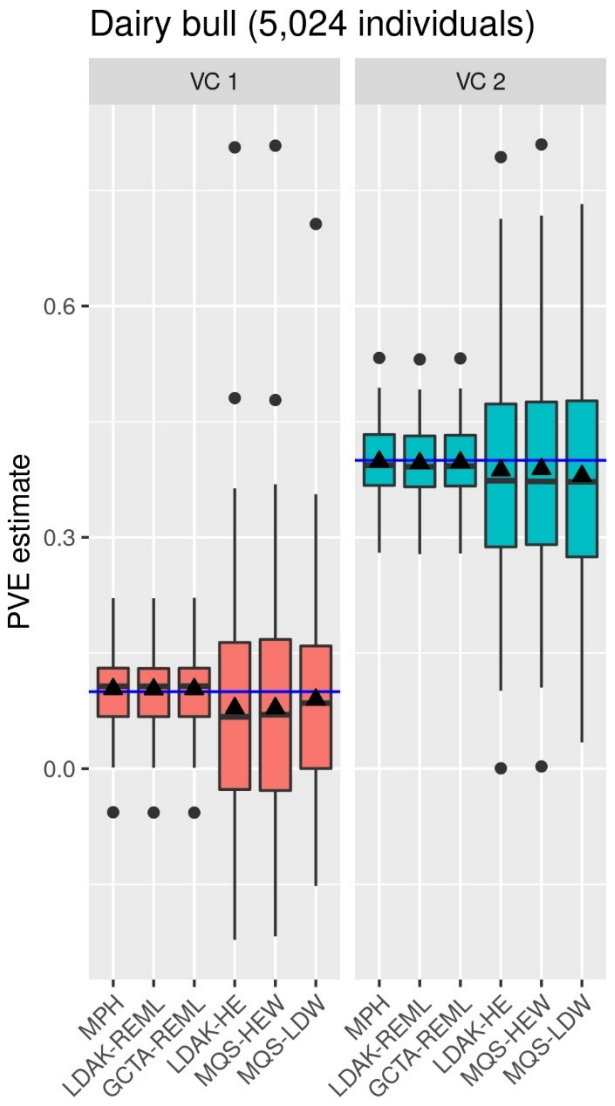


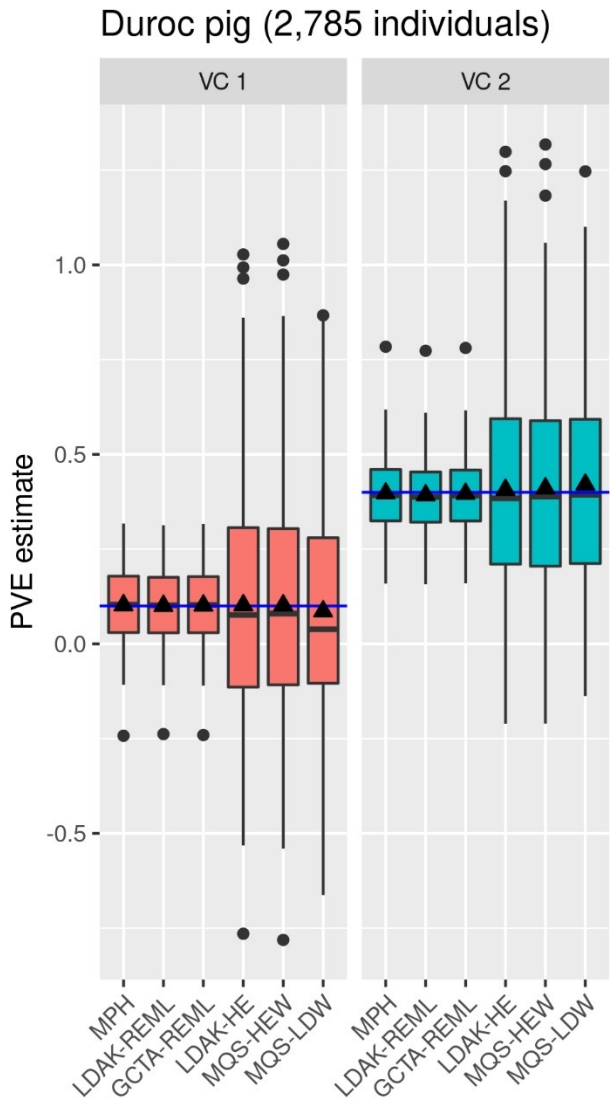
LDSC and stratified LDSC

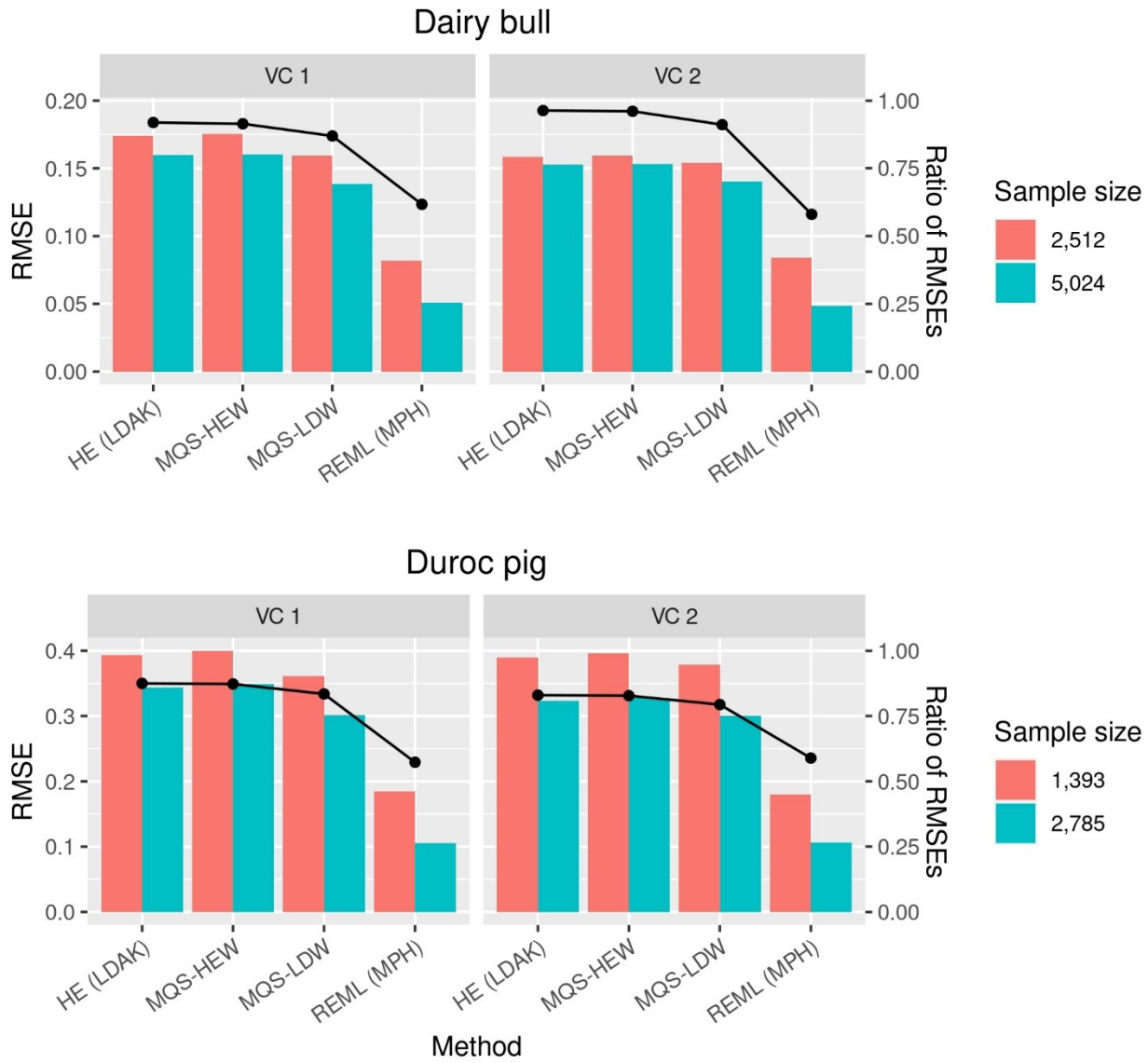
$$t_j \sim \mathrm{N}\left(0, \frac{n}{m}\tilde{l}_j h^2 + c\right) \longrightarrow t_j \sim \mathrm{N}\left(0, \frac{n}{m}\sum_{k=1}^{K}\widetilde{l_{jk}}\tau_k + c\right)$$

c is estimated in LDSC, which can measure the genomic inflation of GWAS statistics.

Parameter estimation is based on iteratively reweighted least squares.







MQS-LDW (like LDSC) lacks statistical power.

MQS-LDW responds poorly to increasing sample size.

Solution

- We need to use REML for the integration of functional annotations.

- Better REML implementation
  - Much more computationally efficient
  - More robust

  - MPH (https://jiang18.github.io/mph/)

Significance Tests and Predictions

- Does a function annotation contribute to heritability?
  - Wald tests on τ_k
  - LRT tests
- Functional-annotation-informed variance components
  - Genetic variance estimate for each marker
- Genomic predictions
  - BLUPs enhanced by the integration of functional annotations

Functional enrichment

	Annot1	Annot2	Annot3	Annot4
SNP1	1	1	0	0
SNP2	0	1	1	0
SNP3	1	1	1	1

W

$$\sigma^2_{\text{SNP1}} = \tau_1 + \tau_2$$
$$\sigma^2_{\text{SNP2}} = \tau_2 + \tau_3$$
$$\sigma^2_{\text{SNP3}} = \tau_1 + \tau_2 + \tau_3 + \tau_4$$

=Wτ

$\text{Var}(\alpha_j) = \sum_{k:j \in S_k} \tau_k$, where τ_k denotes the per-variant contribution of category k to $\text{Var}(\alpha_j)$.

More generally, $\text{Var}(\alpha_j) = \sum_k W_{jk}\tau_k$ for a binary (W_{jk}=0/1) or continuous functional annotation.

Functional enrichment (*cont.*)

$$\begin{bmatrix} \sum\text{var}(\alpha_1) \\ \sum\text{var}(\alpha_2) \\ \vdots \\ \sum\text{var}(\alpha_M) \end{bmatrix} = \mathbf{W}\hat{\boldsymbol{\tau}}$$

Genetic variance explained by SNPs in annotation category k

$$v_k = \sum_{j:j\in S_k} \mathrm{Var}(\alpha_j)$$

Per-SNP enrichment in annotation category k

$$\rho_k = \left(\sum_{j:j\in S_k} \mathrm{Var}(\alpha_j)/M_k\right) \Big/ \left(\sum_j \mathrm{Var}(\alpha_j)/M\right)$$

Functional enrichment (*cont.*)

Let **Q** be a variant-to-annotation incidence matrix.

	Annot1	Annot2	Annot3	Annot4	Annot1,3
SNP1	1	1	0	0	1
SNP2	0	1	1	0	1
SNP3	1	1	1	1	1
	↑ $\boldsymbol{q}_1$	↑ $\boldsymbol{q}_2$	↑ $\boldsymbol{q}_3$	↑ $\boldsymbol{q}_4$	↑ $\boldsymbol{q}_5$

Genetic variance explained by SNPs in annotation category k

$$v_k = \mathbf{q}_k^{\mathrm{T}} \mathbf{W} \hat{\boldsymbol{\tau}}$$

Per-SNP enrichment in annotation category k

$$\rho_k = \frac{\mathbf{q}_k^{\mathrm{T}} \mathbf{W} \hat{\boldsymbol{\tau}} / M_k}{\mathbf{1}^{\mathrm{T}} \mathbf{W} \hat{\boldsymbol{\tau}} / M}$$

Functional enrichment (*cont.*)

Note that $\hat{\boldsymbol{\tau}}$ and $\mathrm{var}(\hat{\boldsymbol{\tau}})$ are computed by REML.
Given $\hat{\boldsymbol{\tau}}$ and $\mathrm{var}(\hat{\boldsymbol{\tau}})$, standard errors for enrichment estimates can be readily computed.

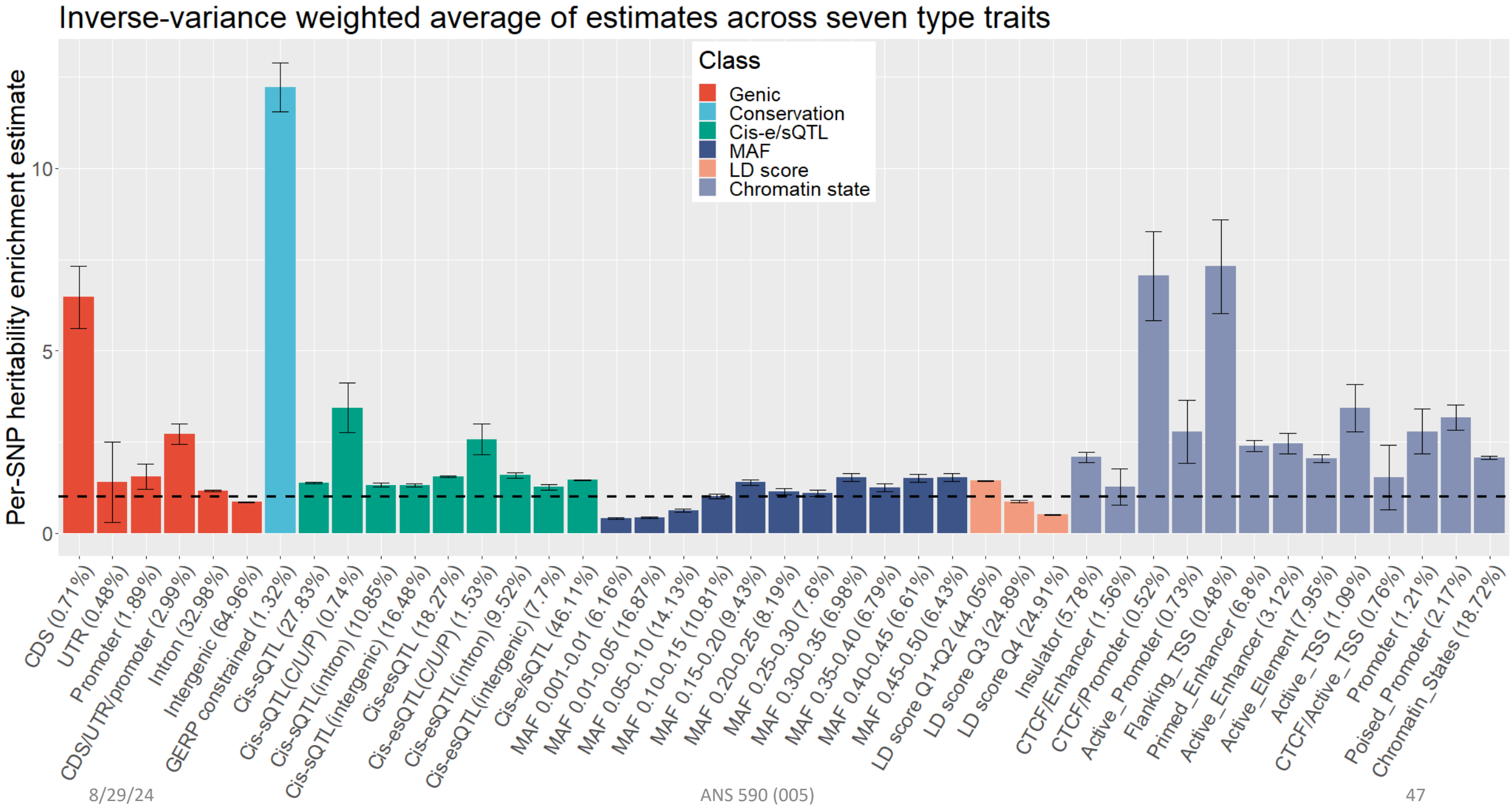
Genetic variance explained by SNPs in annotation category k

$$\mathrm{var}\left(v_k\right) = \mathrm{var}\left(\mathbf{q}_k^{\mathrm{T}}\mathbf{W}\hat{\boldsymbol{\tau}}\right) = \mathbf{q}_k^{\mathrm{T}}\mathbf{W}\,\mathrm{var}\left(\hat{\boldsymbol{\tau}}\right)\mathbf{W}^{\mathrm{T}}\mathbf{q}_k$$

Per-SNP enrichment in annotation category k

$$\mathrm{var}\left(\rho_k\right) = \mathrm{var}\left(\frac{\mathbf{q}_k^{\mathrm{T}}\mathbf{W}\hat{\boldsymbol{\tau}}/M_k}{\mathbf{1}^{\mathrm{T}}\mathbf{W}\hat{\boldsymbol{\tau}}/M}\right)$$

Delta method


Inverse-variance weighted average of estimates across seven type traits
Per-SNP heritability enrichment estimate
0
5
10
Class
Genic
Conservation
Cis-e/sQTL
MAF
LD score
Chromatin state
CDS (0.71%)
UTR (0.48%)
Promoter (1.89%)
CDS/UTR/promoter (2.99%)
Intron (32.98%)
Intergenic (64.96%)
GERP constrained (1.32%)
Cis-sQTL (27.83%)
Cis-sQTL(C/U/P) (0.74%)
Cis-sQTL(intron) (10.85%)
Cis-sQTL(intergenic) (16.48%)
Cis-esQTL (18.27%)
Cis-esQTL(C/U/P) (1.53%)
Cis-esQTL(intron) (9.52%)
Cis-esQTL(intergenic) (7.7%)
Cis-e/sQTL (46.11%)
MAF 0.001-0.01 (6.16%)
MAF 0.01-0.05 (16.87%)
MAF 0.05-0.10 (14.13%)
MAF 0.10-0.15 (10.81%)
MAF 0.15-0.20 (9.43%)
MAF 0.20-0.25 (8.19%)
MAF 0.25-0.30 (7.6%)
MAF 0.30-0.35 (6.98%)
MAF 0.35-0.40 (6.79%)
MAF 0.40-0.45 (6.61%)
MAF 0.45-0.50 (6.43%)
LD score Q1+Q2 (44.05%)
LD score Q3 (24.89%)
LD score Q4 (24.91%)
Insulator (5.78%)
CTCF/Enhancer (1.56%)
CTCF/Promoter (0.52%)
Active_Promoter (0.73%)
Flanking_TSS (0.48%)
Primed_Enhancer (6.8%)
Active_Enhancer (3.12%)
Active_Element (7.95%)
Active_TSS (1.09%)
CTCF/Active_TSS (0.76%)
Promoter (1.21%)
Poised_Promoter (2.17%)
Chromatin_States (18.72%)