

GWAS Analysis in Julia

Kayla McCue

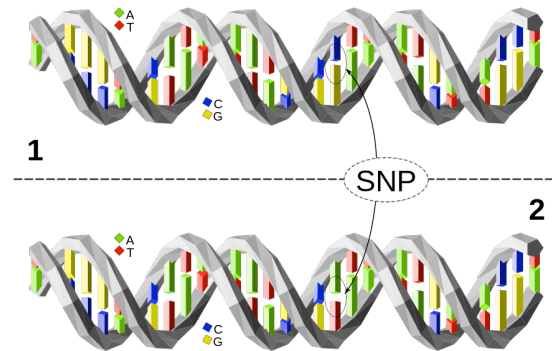
Genome Wide Association Study (GWAS)

- Want to be able to associate phenotypes with SNPs across the whole genome
- This is done using a SNP array



GWAS Vocabulary

- Single Nucleotide Polymorphism (SNP)



- Allele Frequency
- Phenotype

Analysis in Julia

- IHT Package
 - <https://github.com/klkeys/IHT.jl>
- PLINK Package
 - <https://github.com/klkeys/PLINK.jl>
- Iterative Hard Thresholding
 - ℓ_0 -penalized multivariate regression

$$\frac{1}{2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_0$$

- NP-hard, so use the heuristic projected gradient update

$$\boldsymbol{\beta}^{m+1} = P_{S_k}[\boldsymbol{\beta}^m - \mu \nabla(\boldsymbol{\beta}^m)]$$

Analysis in Julia

- Simulating data and testing IHT compared with Lasso and MCP gives favorable results
- Provides reasonable results on real GWAS data
- Julia implementation reduces run time compared to original analysis

Analysis of Multiple Simulated Datasets

	Simulation Method 1		Simulation Method 2		Simulation Method 3	
	Precision	Recall	Precision	Recall	Precision	Recall
IHT	.50	.70	.58	.70	1.00	.70
Lasso	1.00	1.00	.09	1.00	.17	.90