Systems Approach to Complex Traits Prediction and Dissection

Hae Kyung Im, PhD

Department of Health Studies

THE UNIVERSITY OF CHICAGO

January 29, 2014

Overview

- What we learned with GWAS studies
- What we are learning with NGS studies
- Limitations and challenges
- Need to shift from single variant to systems approach
- Prediction of complex traits
- Biological dissection of complex traits

Overview

- What we learned with GWAS studies
- What we are learning with NGS studies
- Limitations and challenges
- Need to shift from single variant to systems approach
- Prediction of complex traits
- Biological dissection of complex traits

Overview

- What we learned with GWAS studies
- What we are learning with NGS studies
- Limitations and challenges
- Need to shift from single variant to systems approach
- Prediction of complex traits
- Biological dissection of complex traits

Overview

- What we learned with GWAS studies
- What we are learning with NGS studies
- Limitations and challenges
- Need to shift from single variant to systems approach
- Prediction of complex traits
- Biological dissection of complex traits

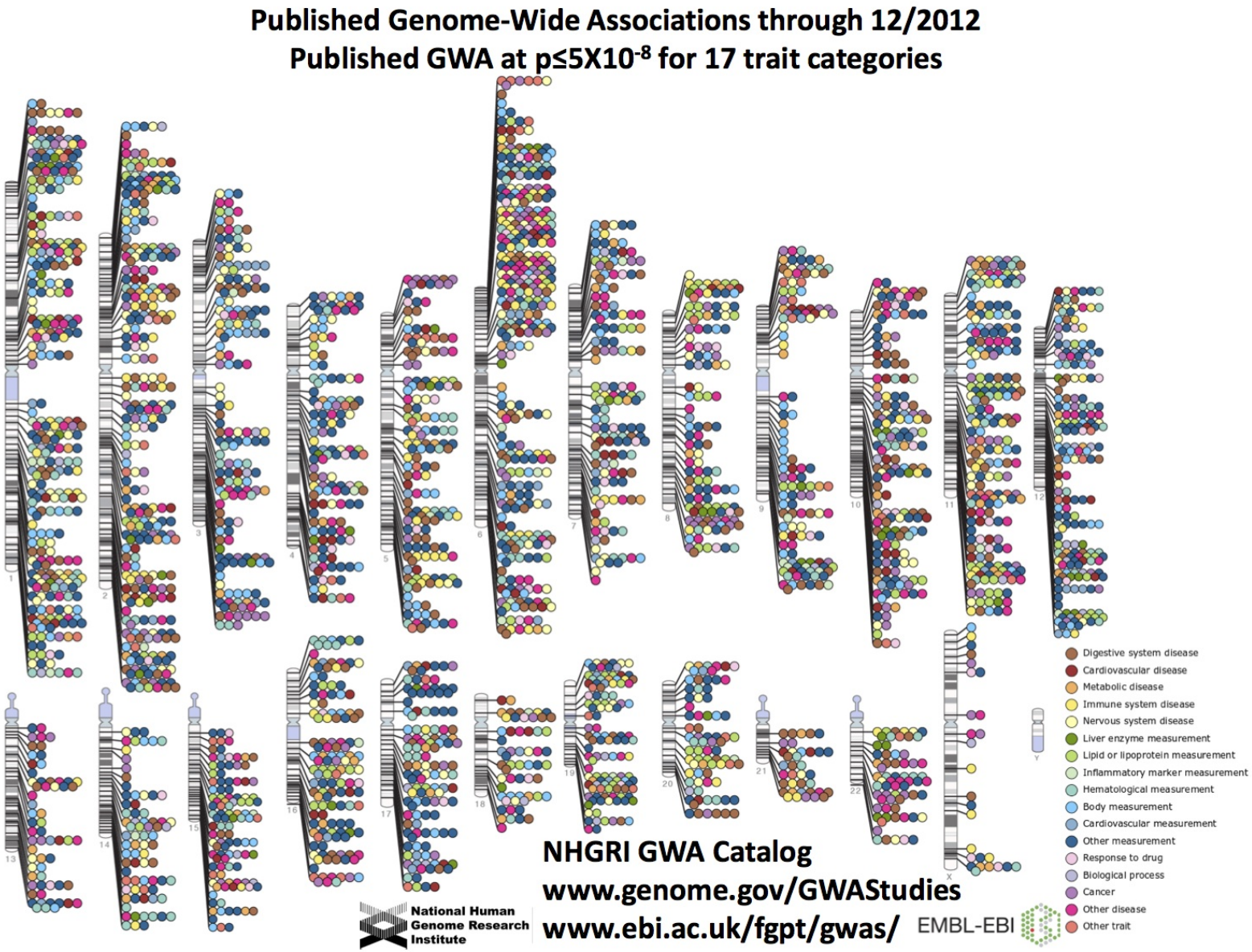
Overview

- What we learned with GWAS studies
- What we are learning with NGS studies
- Limitations and challenges
- Need to shift from single variant to systems approach
- Prediction of complex traits
- Biological dissection of complex traits

Overview

- What we learned with GWAS studies
- What we are learning with NGS studies
- Limitations and challenges
- Need to shift from single variant to systems approach
- Prediction of complex traits
- Biological dissection of complex traits

NHGRI catalog includes 11714 SNPs as of 1/2013



Only Small Portion of Variability Explained by GWAS Hits

- Genome-wide significant SNPs explain only small portion of trait variability
- Biology underlying these hits are not well understood
- Recent statistical methods compute variability explained by all common variants on a genotyping array (GCTA from Visscher Lab)
  "Chip heritability"
- For many traits, the full set of genotyped variants explain a substantial portion of the variability

Only Small Portion of Variability Explained by GWAS Hits

- Genome-wide significant SNPs explain only small portion of trait variability
- Biology underlying these hits are not well understood
- Recent statistical methods compute variability explained by all common variants on a genotyping array (GCTA from Visscher Lab)
  "Chip heritability"
- For many traits, the full set of genotyped variants explain a substantial portion of the variability

Only Small Portion of Variability Explained by GWAS Hits

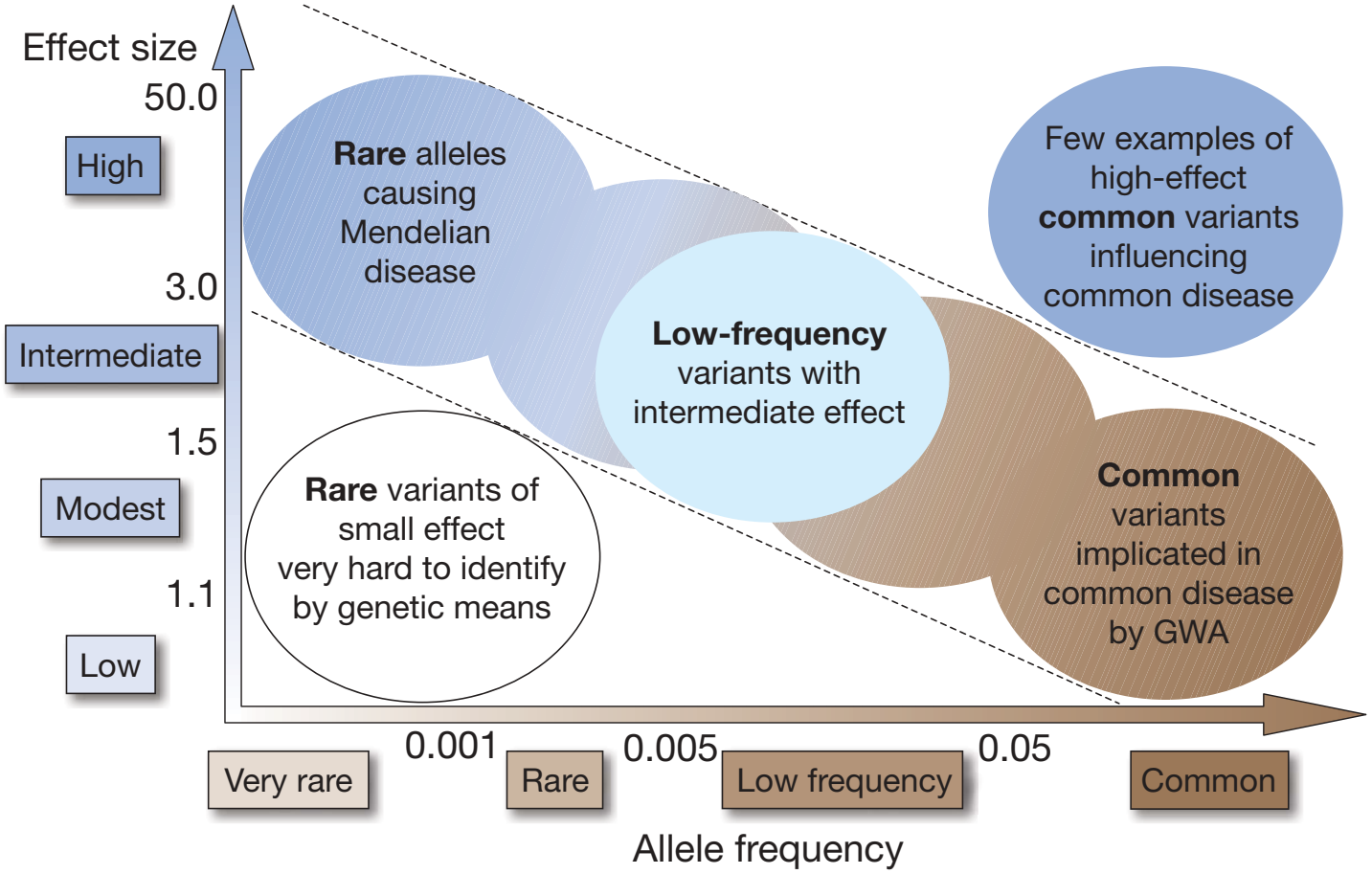
- Genome-wide significant SNPs explain only small portion of trait variability
- Biology underlying these hits are not well understood
- Recent statistical methods compute variability explained by all common variants on a genotyping array (GCTA from Visscher Lab)

  "Chip heritability"
- For many traits, the full set of genotyped variants explain a substantial portion of the variability

Only Small Portion of Variability Explained by GWAS Hits

- Genome-wide significant SNPs explain only small portion of trait variability
- Biology underlying these hits are not well understood
- Recent statistical methods compute variability explained by all common variants on a genotyping array (GCTA from Visscher Lab)
  "Chip heritability"
- For many traits, the full set of genotyped variants explain a substantial portion of the variability

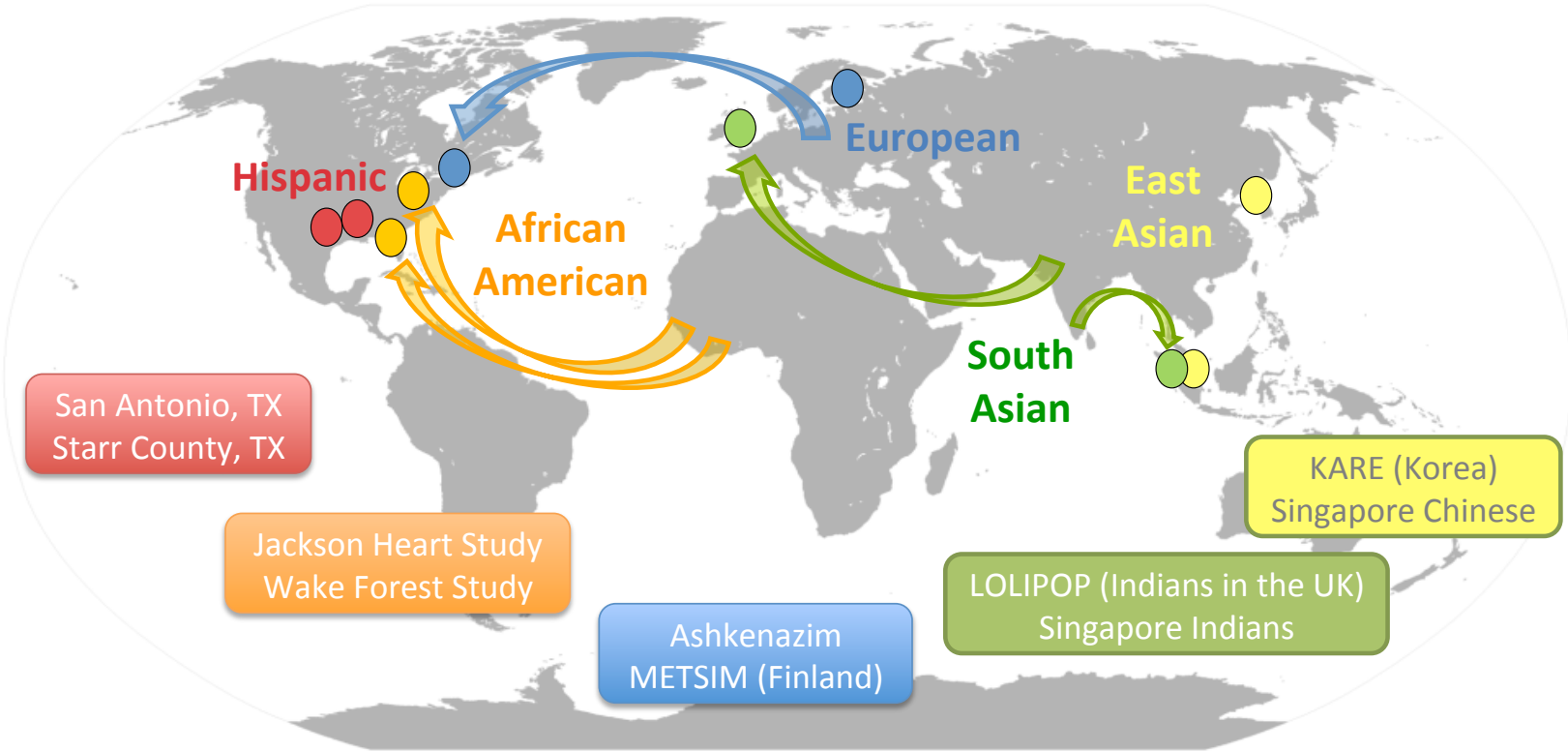
Hypothesis: Goldilock Variants Will Be Found with NGS



Teri Manolio et al, 2009, Nature

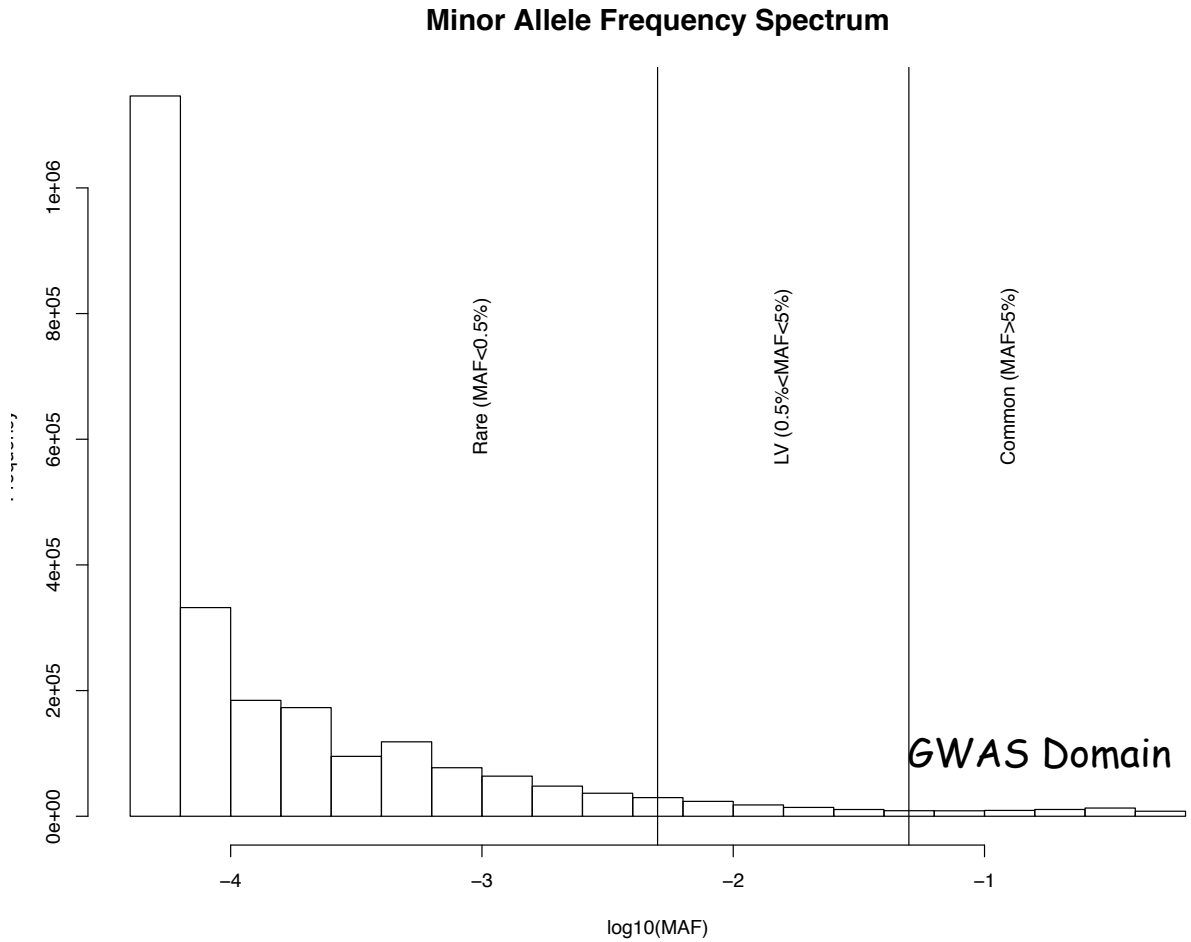
T2D-GENES Consortium Sequenced 10K Whole Exomes

Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples



1000 type 2 diabetes cases - 1000 controls from each of 5 ancestries

Most Variants Are Rare



Im on behalf of the T2D-GENES Consortium, American Diabetes Association Meeting 2013

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci
- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci
- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci
- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci
- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci

- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci

- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci

- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings

  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

What We Learned with 10K Whole Exomes

- T2D: No novel findings
- HbA1c: Known G6PD deficiency mutations
- Fasting glucose and fasting insulin: no novel findings
- Systolic/diastolic blood pressure: no novel findings
- LDL-C and HDL-C: new variants in known loci

- Merger with GoT2D Consortium
  - Adding 3000 European exome sequences did not bring in new findings
  - Exome chip studies with over 30K Europeans are starting to yield some exome-wide significant results

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
    - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

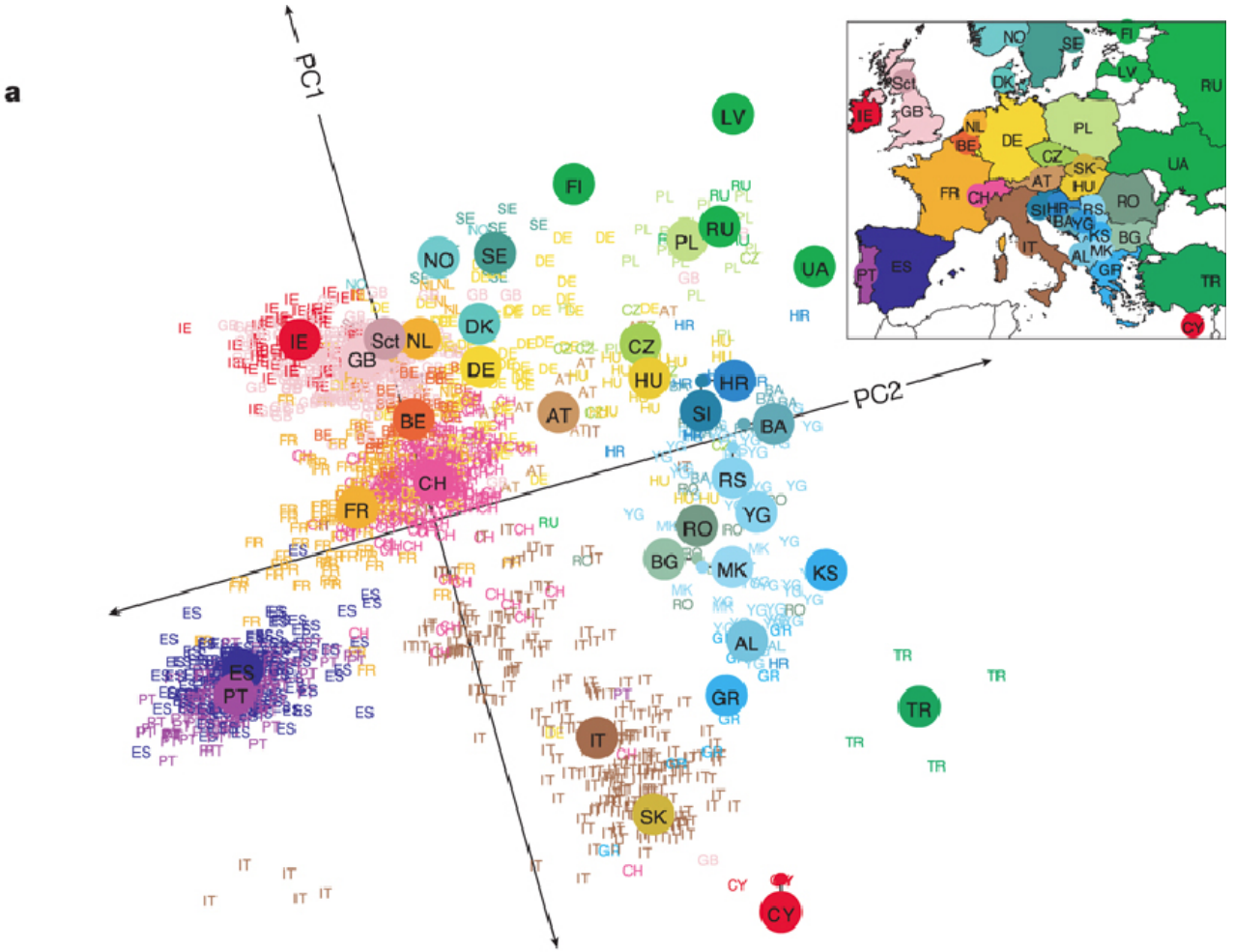
Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Conclusions so far

- Large number of loci found using GWAS/NGS designs
  - Genome-wide significant hits do not explain much of the variability
  - Underlying biology is not well understood
- Few Goldilock variants with current sample sizes (n$\sim$10K+)
- Cumulative effect of all common variants explain substantial portion of variability for many traits
- To translate to clinical practice we need
  - better predictive models
  - better ways to probe the underlying biology
  - better ways to accumulate rare variant effects

Accumulation of Small Effects Can Reveal Hidden Patterns

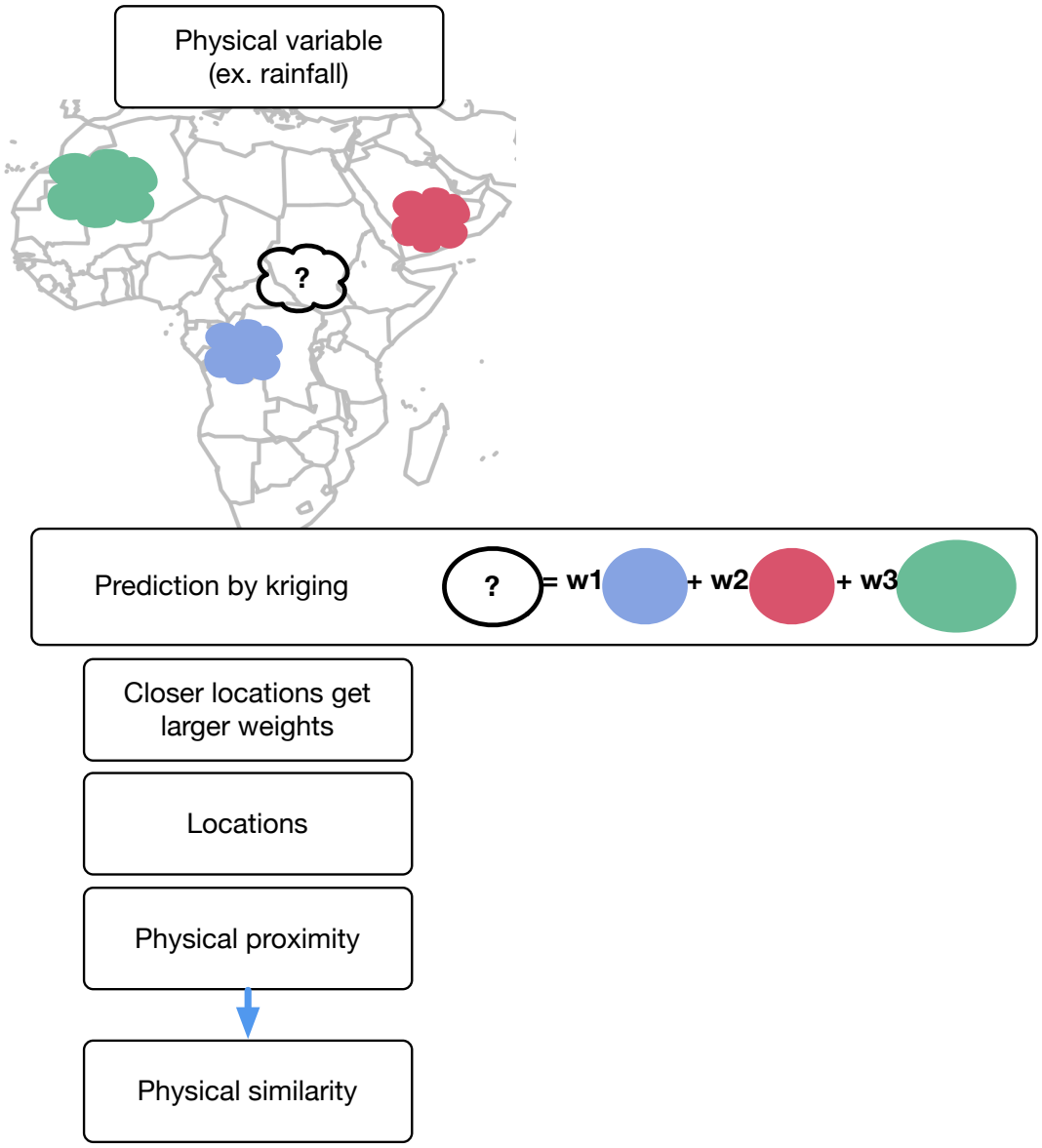


Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD: Genes mirror geography within Europe. Nature 2008, 456:98101.

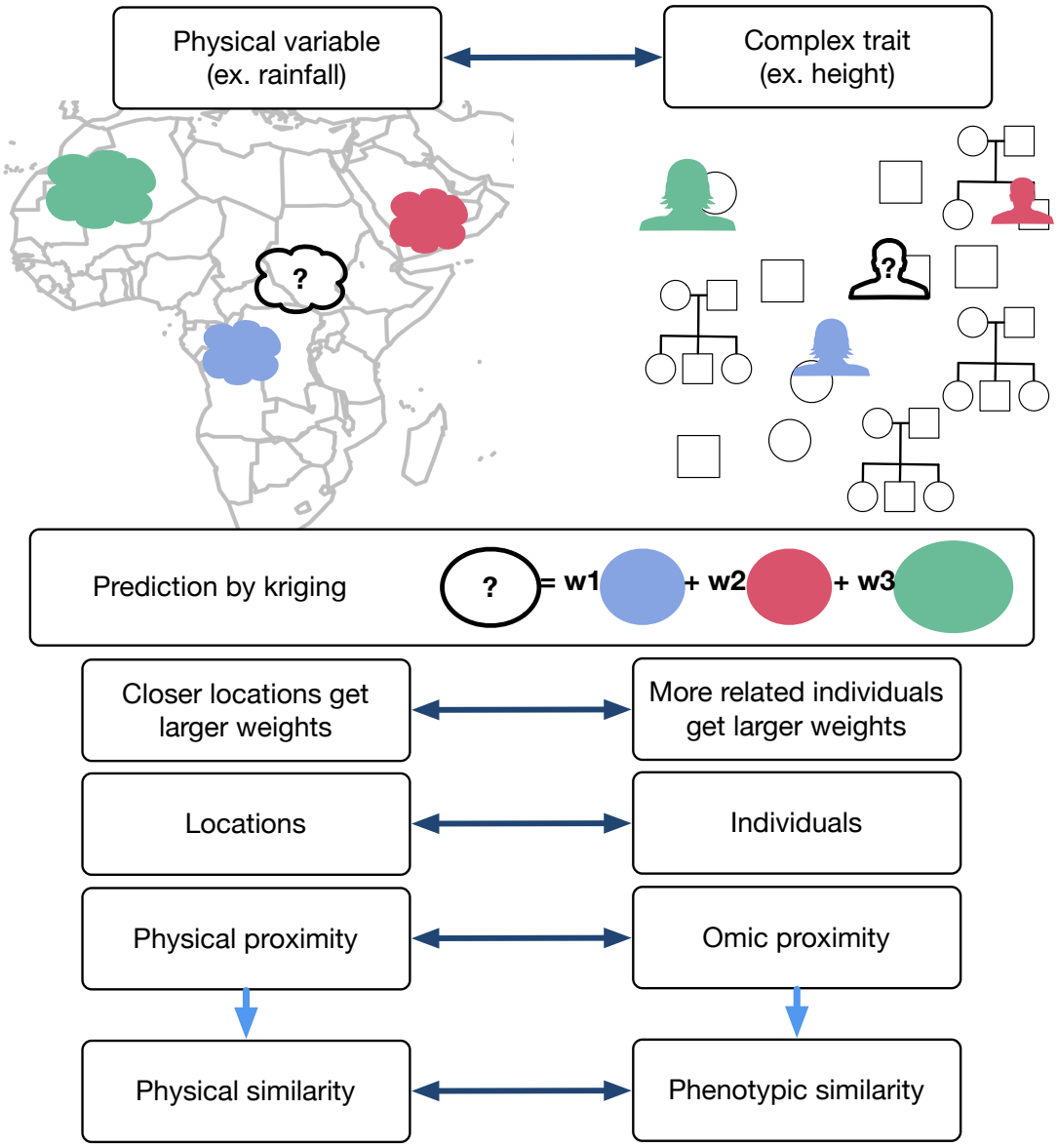
Prediction - OmicKriging

Systems approach to prediction that translates and integrates similarity in genomic and other omic data.

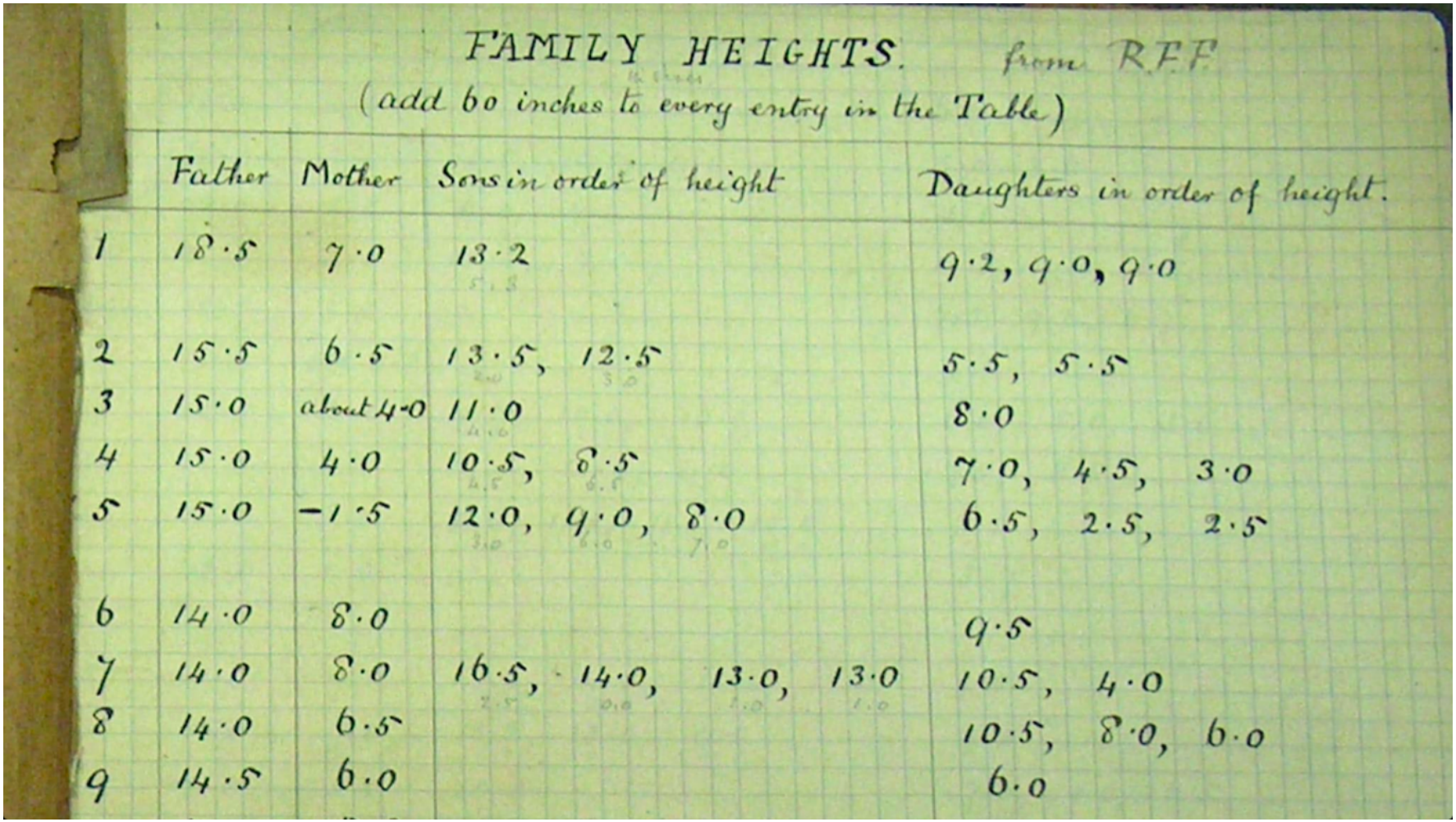
What is Kriging?



Kriging in Complex Trait Prediction



Galton's Height Data (1885)



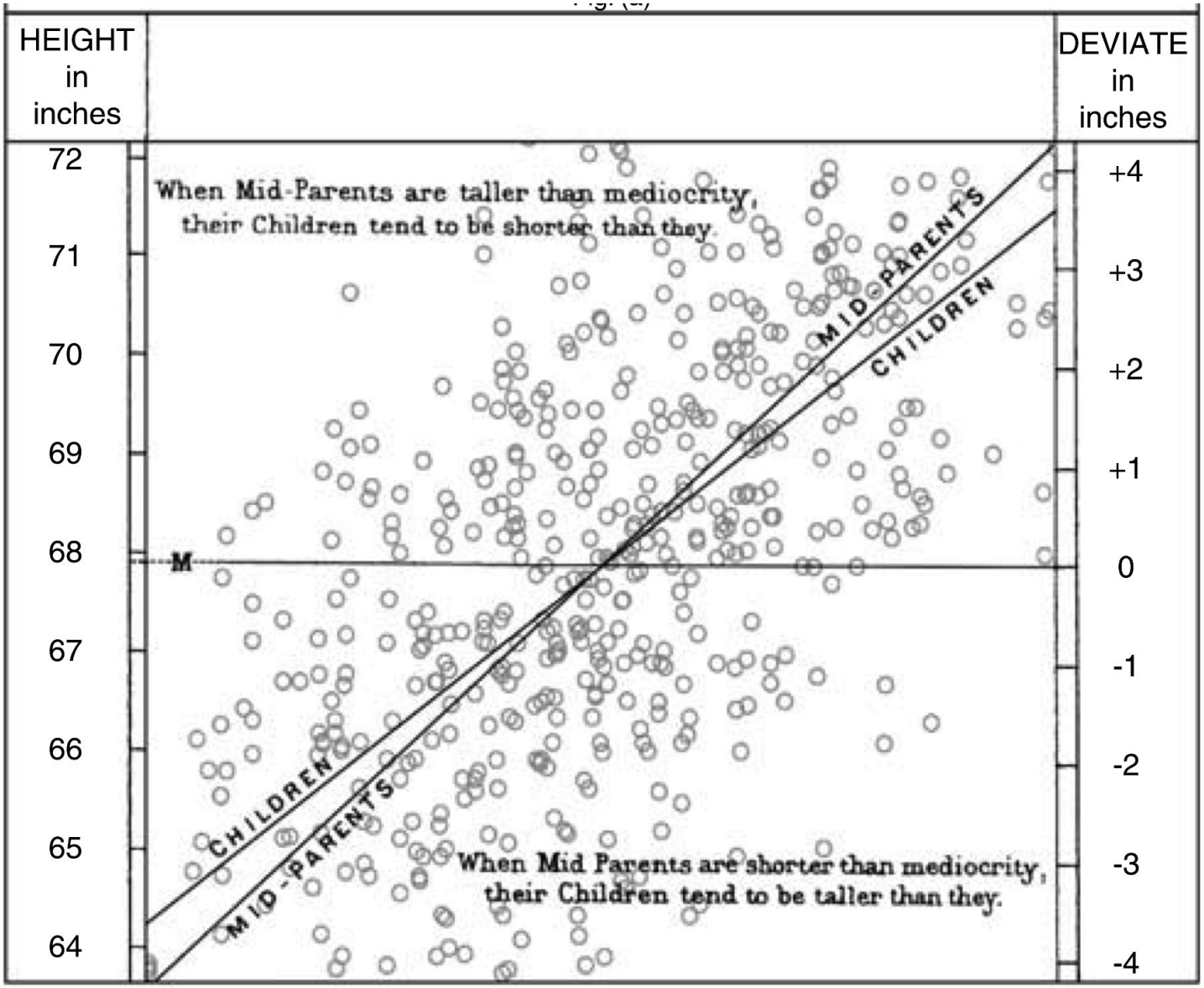
FAMILY HEIGHTS. from R.F.F.

(add 60 inches to every entry in the Table)

	Father	Mother	Sons in order of height	Daughters in order of height.
1	18·5	7·0	13·2	9·2, 9·0, 9·0
2	15·5	6·5	13·5, 12·5	5·5, 5·5
3	15·0	about 4·0	11·0	8·0
4	15·0	4·0	10·5, 8·5	7·0, 4·5, 3·0
5	15·0	−1·5	12·0, 9·0, 8·0	6·5, 2·5, 2·5
6	14·0	8·0		9·5
7	14·0	8·0	16·5, 14·0, 13·0, 13·0	10·5, 4·0
8	14·0	6·5		10·5, 8·0, 6·0
9	14·5	6·0		6·0

Hanley JA: Transmuting Women into Men. The American Statistician 2004, 58:237243.

Galton Was Kriging with Kinship Matrix (1885)



Kriging = BLUP (Best Linear Unbiased Prediction)

- Galton (1885): parent to offspring
- Fisher (1918) and Wright (1921): pedigree
- Formalized by Henderson (1950,1975) and Goldberger (1962)
- Kriging can be interpreted as the posterior mean of the genetic component given observations ($Y = G + \epsilon$)

- Kriging translates genetic similarity into phenotypic prediction

Kriging = BLUP (Best Linear Unbiased Prediction)

- Galton (1885): parent to offspring
- Fisher (1918) and Wright (1921): pedigree
- Formalized by Henderson (1950,1975) and Goldberger (1962)
- Kriging can be interpreted as the posterior mean of the genetic component given observations ($Y = G + \epsilon$)

- Kriging translates genetic similarity into phenotypic prediction

Kriging = BLUP (Best Linear Unbiased Prediction)

- Galton (1885): parent to offspring
- Fisher (1918) and Wright (1921): pedigree
- Formalized by Henderson (1950,1975) and Goldberger (1962)
- Kriging can be interpreted as the posterior mean of the genetic component given observations ($Y = G + \epsilon$)

- Kriging translates genetic similarity into phenotypic prediction

Kriging = BLUP (Best Linear Unbiased Prediction)

- Galton (1885): parent to offspring
- Fisher (1918) and Wright (1921): pedigree
- Formalized by Henderson (1950,1975) and Goldberger (1962)
- Kriging can be interpreted as the posterior mean of the genetic component given observations ($Y = G + \epsilon$)

- Kriging translates genetic similarity into phenotypic prediction

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

OmicKriging Integrates Heterogeneous Data for Prediction

Extends Kriging framework to integrate heterogeneous data sources by combining similarity matrices

$$\Sigma = \theta_1 \Sigma_{\text{Source1}} + \theta_2 \Sigma_{\text{Source2}} + \cdots + (1 - \theta_1 - \theta_2 - \cdots)\mathbb{I}$$

- Multiple omics data
- Prior GWAS results by giving more weight to markers with larger effect sizes
- eQTL information by giving more weight to regions with regulatory evidence
- Even geographic similarity

- θ's selected to maximize prediction performance

Poly-Omic Model

$$Y_i = a + G_i + T_i + O_i + \epsilon_i$$

$$G_i = \sum_{l=1}^{M} \beta_l^G X_{il}^G \qquad \text{genetic component}$$

$$T_i = \sum_{l=1}^{L} \beta_l^T X_{il}^T \qquad \text{transcriptomic component}$$

$$O_i = \sum_{l=1}^{L'} \beta_l^O X_{il}^O \qquad \text{other omic component}$$

$$(\beta_G, \beta_T, \beta_O)' \sim \mathrm{N}(0, \boldsymbol{\Sigma}_\beta)$$

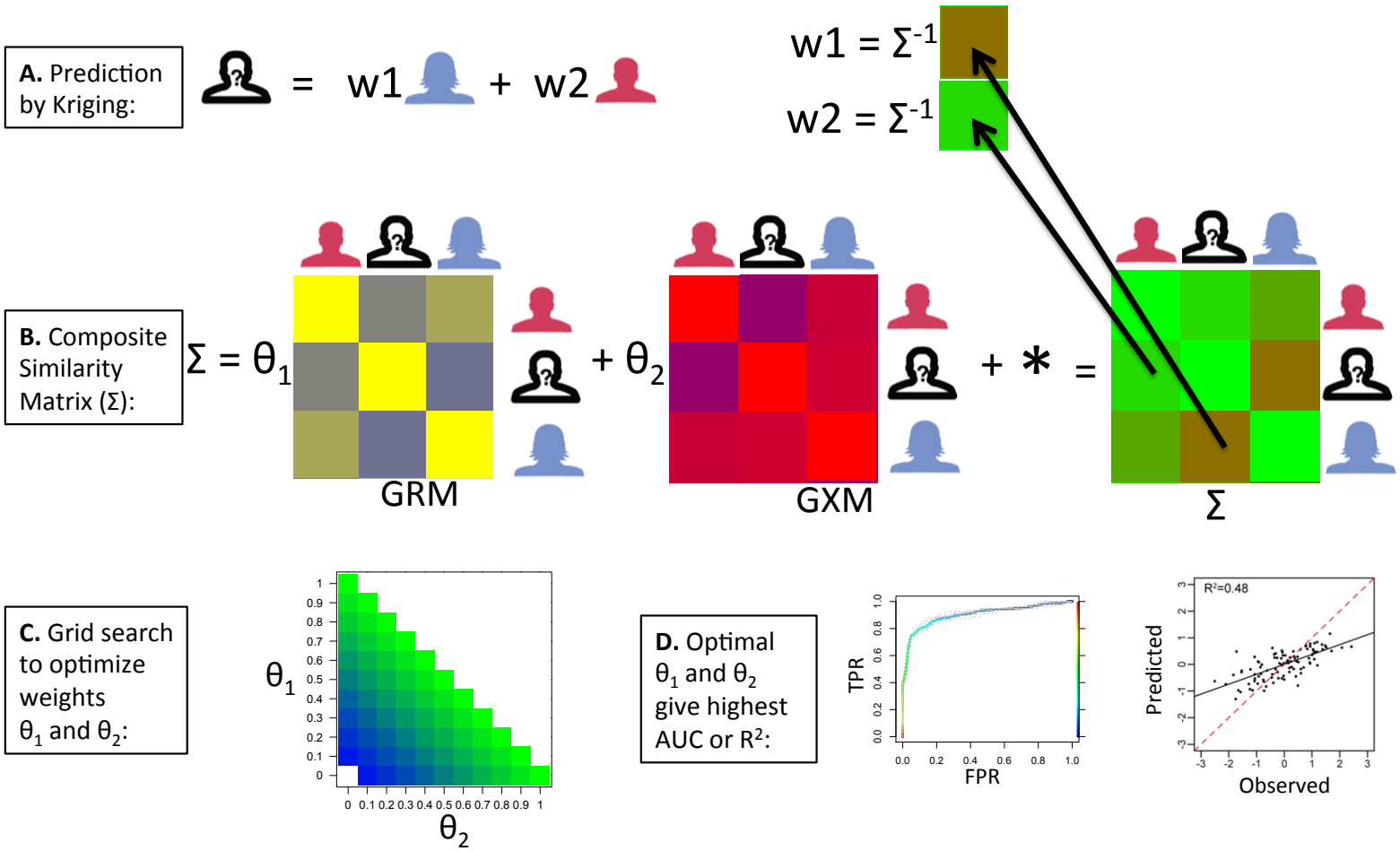
Optimal Similarity Matrix

$$Y_i = a + G_i + T_i + O_i + \epsilon_i$$

Assuming independence of β's

$$\begin{aligned}\Sigma_{i,j} &= \theta_G \sum_{l=1}^{M} X_{il}^G X_{jl}^G + \theta_T \sum_{l=1}^{L} X_{il}^T X_{jl}^T + \theta_O \sum_{l=1}^{L'} X_{il}^O X_{jl}^O + \theta_\epsilon \delta_{ij} \\ &= \theta_G \Sigma_G + \theta_T \Sigma_T + \theta_O \Sigma_O + \theta_\epsilon \delta_{ij}\end{aligned}$$

Integration of Multiple Sources



Predicting Disease Risk using OmicKriging

- Wellcome Trust Case Control Consortium
  - BD: bipolar disorder
  - CAD: coronary artery disease
  - HT: hypertension
  - T1D: type 1 diabetes
  - T2D: type 2 diabetes
  - CD: Crohn's disease
  - RA: rheumatoid arthritis
- Two approaches
  - Whole genome prediction no prior information ($\sim$400K snps)
  - Whole genome prediction with more weight to SNPs implicated in previous studies

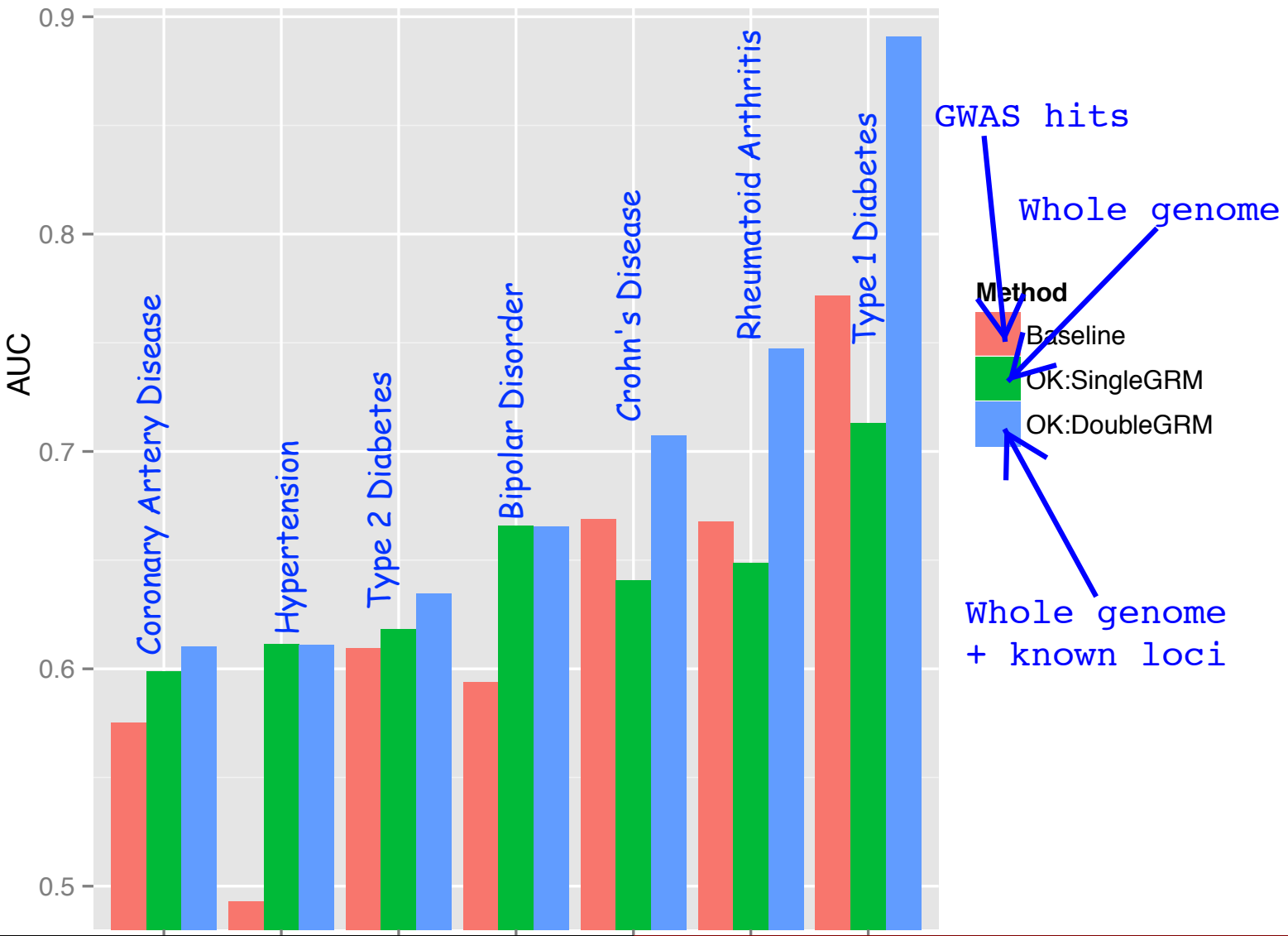
Predicting Disease Risk using OmicKriging

- Wellcome Trust Case Control Consortium
  - BD: bipolar disorder
  - CAD: coronary artery disease
  - HT: hypertension
  - T1D: type 1 diabetes
  - T2D: type 2 diabetes
  - CD: Crohn's disease
  - RA: rheumatoid arthritis
- Two approaches
  - Whole genome prediction no prior information ($\sim$400K snps)
  - Whole genome prediction with more weight to SNPs implicated in previous studies

Predicting Disease Risk using OmicKriging

- Wellcome Trust Case Control Consortium
  - BD: bipolar disorder
  - CAD: coronary artery disease
  - HT: hypertension
  - T1D: type 1 diabetes
  - T2D: type 2 diabetes
  - CD: Crohn's disease
  - RA: rheumatoid arthritis
- Two approaches
  - Whole genome prediction no prior information ($\sim$400K snps)
  - Whole genome prediction with more weight to SNPs implicated in previous studies

Whole Genome with More Weight to Known Loci Outperforms GWAS Hits Only Predictions



Several Large Effect Variants Affect T1D, CD, RA

Many Variants of Modest Effects in BD, HT, CAD, T2D

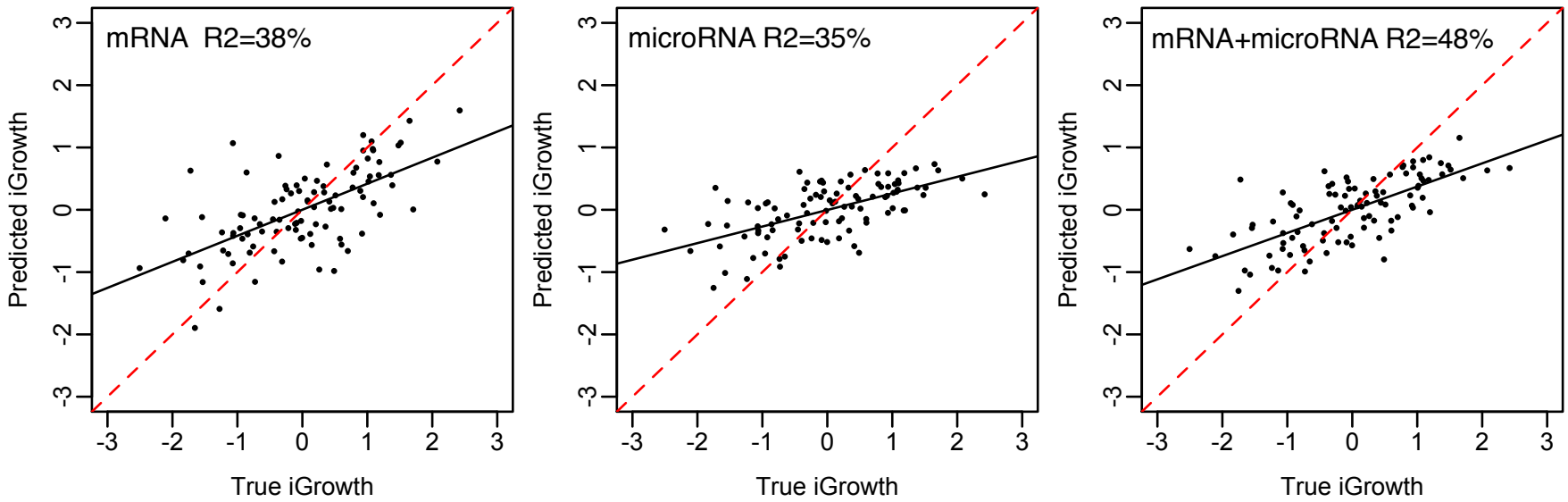
OmicKriging Using Known Loci Outperforms Other Models

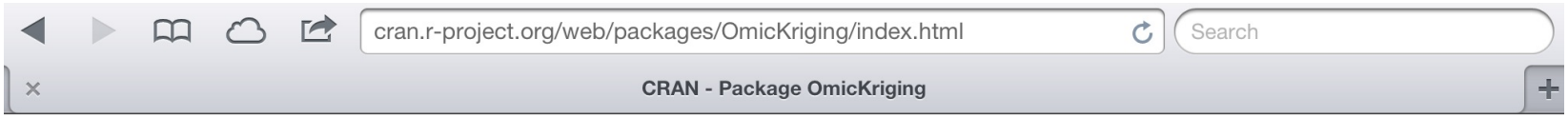
- OmicKriging using priors performs best
- Genetic architecture determines gain in performance
  - Autoimmune diseases (type 1 diabetes, Crohn's disease, rheumatoid arthritis) have some large effects + polygenic background
    Whole Genome no prior $<$ GWAS hits $<$ Whole genome + priors
  - Bipolar disorder, hypertension, type 2 diabetes, and coronary artery disease has more polygenic architecture
    GWAS hits $<$ Whole Genome no prior $<$ Whole genome + priors

Integrating Multiple Omics for Cell Growth Prediction

- Intrinsic cellular growth phenotype (Im et al 2012 PLoS Genetics)
  - 99 HapMap cell lines
  - CEU and YRI
  - Genotype, mRNA, microRNA
- Genes associated with iGrowth are prognostic of survival in cancer patients

Integrating Whole Genome mRNA and microRNA Improves Prediction





OmicKriging: OmicKriging for Phenotypic Prediction

This package provides functions to generate correlation matrices from SNP, gene expression, methylation or other 'omics' datasets and to use these matrices to predict the phenotype of an individual by using the phenotypes of the remaining individuals through kriging. Kriging is a geostatistical method for optimal prediction or best unbiased linear prediction. It consists of predicting the value of a variable at an unobserved location as a weighted sum of the variable at observed locations. Intuitively, it works as a reverse linear regression: instead of computing correlation (univariate regression coefficients are simply scaled correlation) between a dependent variable Y and independent variables X, it uses known correlation between X and Y to predict Y.

Version:	1.0
Depends:	R (≥ 2.10.0)
Published:	2013-03-09
Author:	Hae Kyung Im, Heather E. Wheeler
Maintainer:	Hae Kyung Im <haky at uchicago.edu>
License:	GPL (> 3)
NeedsCompilation:	no
CRAN checks:	OmicKriging results

Downloads:

Package source:	OmicKriging_1.0.tar.gz
MacOS X binary:	OmicKriging_1.0.tgz
Windows binary:	OmicKriging_1.0.zip
Reference manual:	OmicKriging.pdf

Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

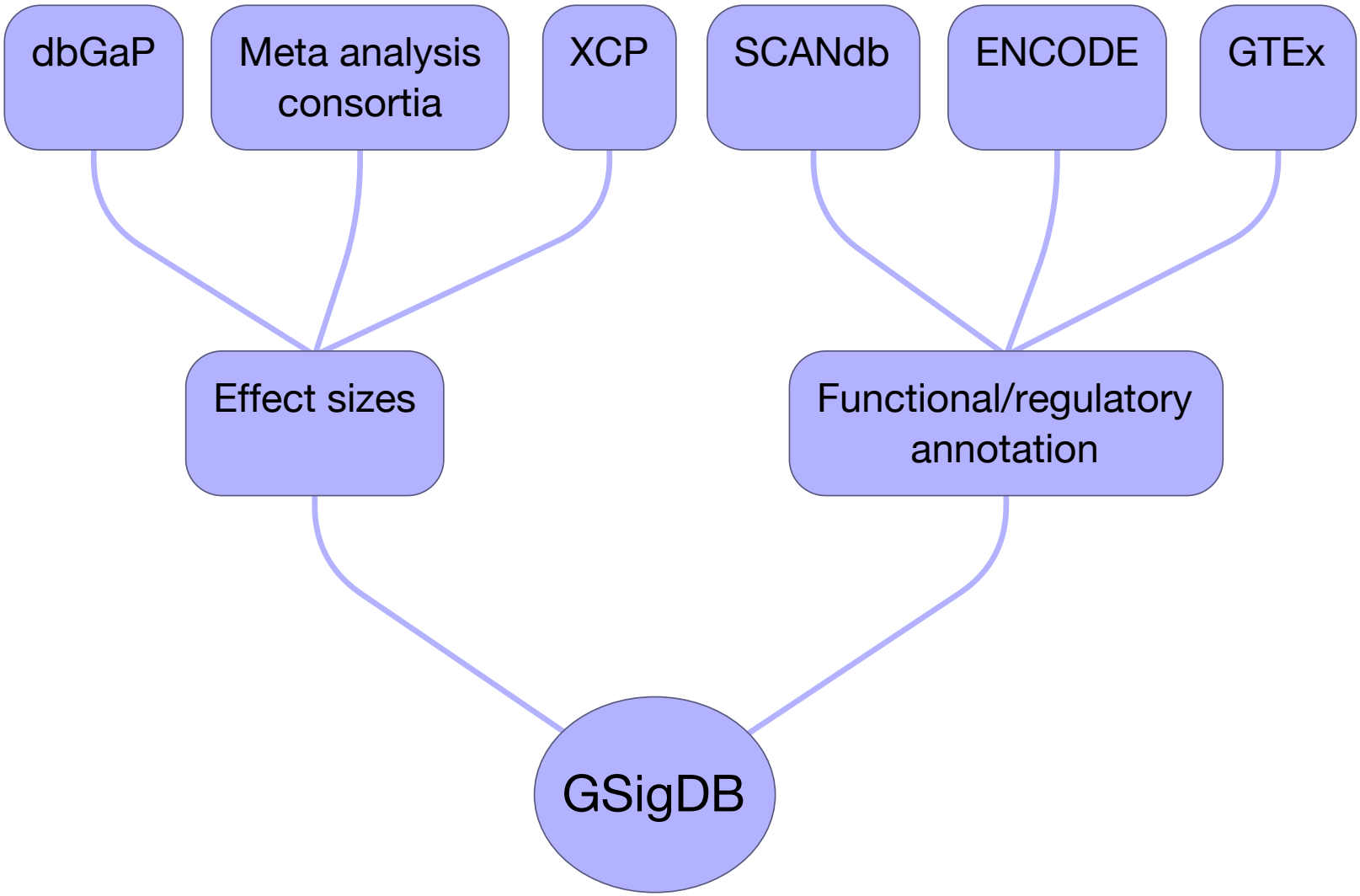
Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

Summary Prediction Method

- Proposed a systems approach to complex trait prediction that leverages and integrates multiple omic similarity
  - Allow easy integration of heterogenous data and prior functional evidence
  - Publicly available software for OmicKriging
- Clinically relevant prediction can be achieved even if we do not know the individual variant's contribution
  - Recent power calculations show that it may take sample sizes of 1 Million or more (Chatterjee et al NG 2013)
  - Advances in high throughput assays and electronic medical record systems
  - Methods and infrastructure that can process these deluge of data and keep us at the forefront of genomic medicine

GSigDB: Database of Predictive Models



Systems Dissection of Complex Traits

Explore biology and mechanisms underlying complex traits using whole genome prediction of biomarkers as probes

Important Biological Questions We Will Be Able to Answer

- Do inflammation markers affect risk of pancreatic cancer?
- Do genetic risk factors increasing risk of developing cancer affect the response to chemotherapy?
- Does intrinsic cellular growth rate have an effect on cancer risk?
- Does inflammation have a role in T1D, T2D, Crohns disease, or ulcerative colitis?
- Does genetic liability to hypertension predispose to drug induced hypertension toxicity?

Important Biological Questions We Will Be Able to Answer

- Do inflammation markers affect risk of pancreatic cancer?
- Do genetic risk factors increasing risk of developing cancer affect the response to chemotherapy?
- Does intrinsic cellular growth rate have an effect on cancer risk?
- Does inflammation have a role in T1D, T2D, Crohns disease, or ulcerative colitis?
- Does genetic liability to hypertension predispose to drug induced hypertension toxicity?

Important Biological Questions We Will Be Able to Answer

- Do inflammation markers affect risk of pancreatic cancer?
- Do genetic risk factors increasing risk of developing cancer affect the response to chemotherapy?
- Does intrinsic cellular growth rate have an effect on cancer risk?
- Does inflammation have a role in T1D, T2D, Crohns disease, or ulcerative colitis?
- Does genetic liability to hypertension predispose to drug induced hypertension toxicity?

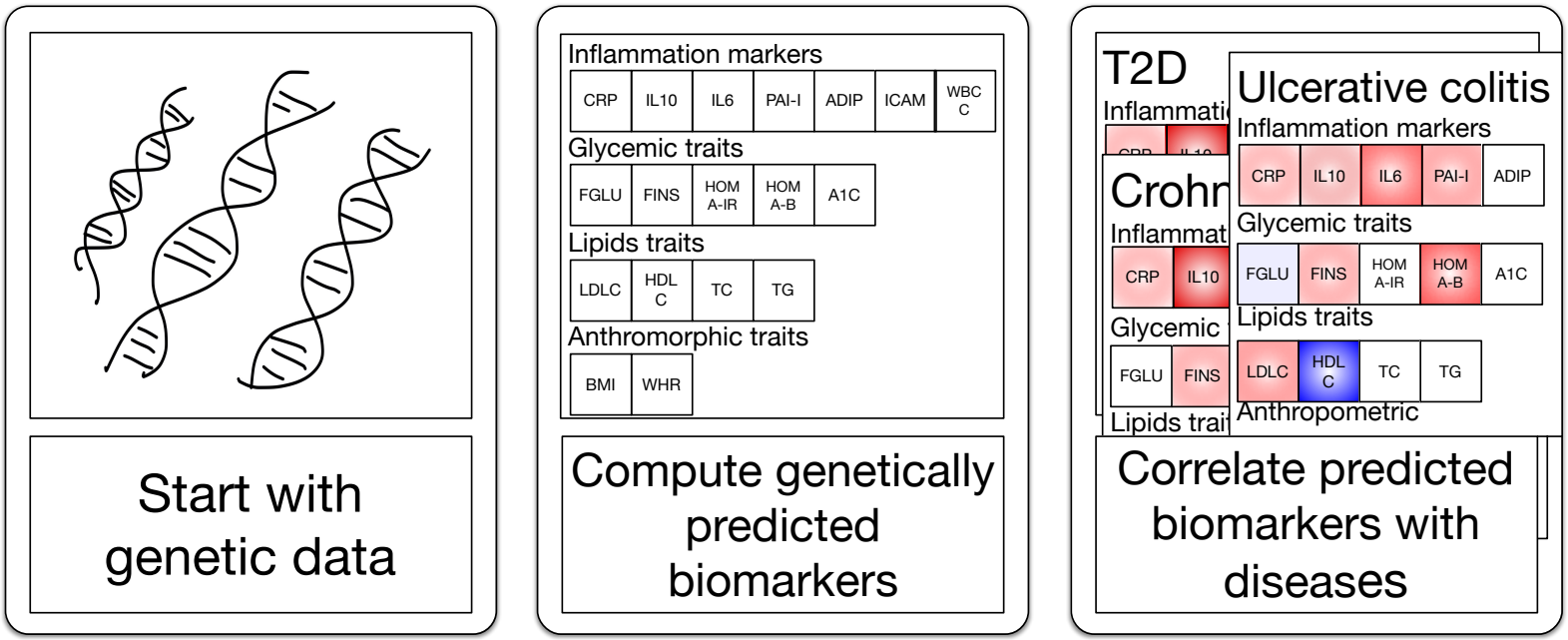
Important Biological Questions We Will Be Able to Answer

- Do inflammation markers affect risk of pancreatic cancer?
- Do genetic risk factors increasing risk of developing cancer affect the response to chemotherapy?
- Does intrinsic cellular growth rate have an effect on cancer risk?
- Does inflammation have a role in T1D, T2D, Crohns disease, or ulcerative colitis?
- Does genetic liability to hypertension predispose to drug induced hypertension toxicity?

Important Biological Questions We Will Be Able to Answer

- Do inflammation markers affect risk of pancreatic cancer?
- Do genetic risk factors increasing risk of developing cancer affect the response to chemotherapy?
- Does intrinsic cellular growth rate have an effect on cancer risk?
- Does inflammation have a role in T1D, T2D, Crohns disease, or ulcerative colitis?
- Does genetic liability to hypertension predispose to drug induced hypertension toxicity?

Dissection Approach



We Observe Nature's Experiments

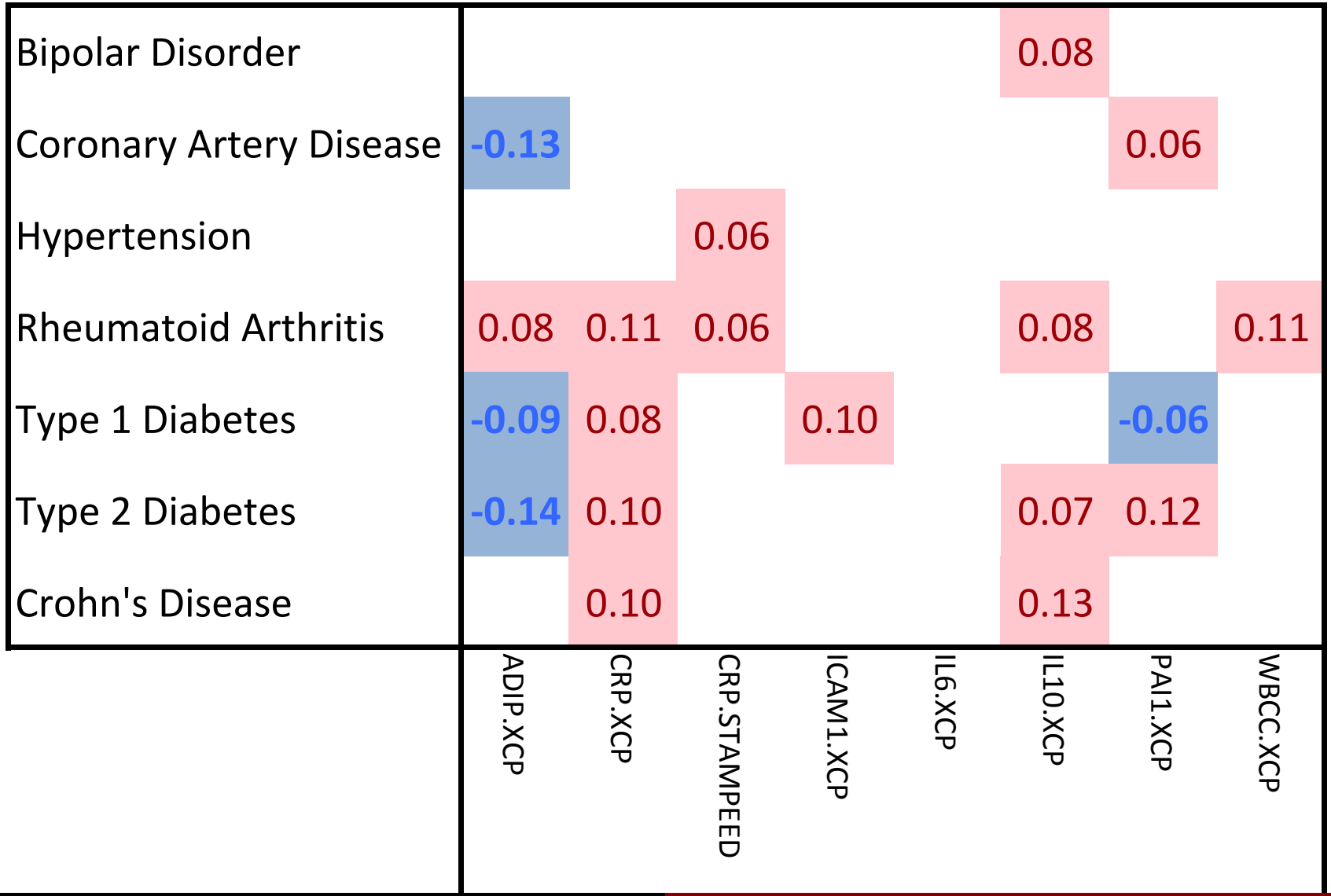
- At meiosis, individuals get "randomized" to high or low levels of a biomarker
- Individuals with high genetic score will be "exposed" to higher levels of the biomarker
- Is there a correlation between this lifetime "exposure" and disease risk?

Biomarkers Currently Available in gSigDB

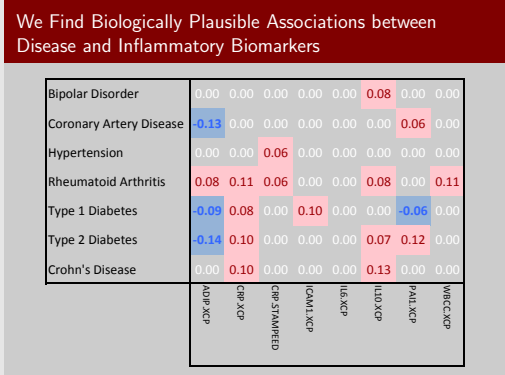
- Inflammatory markers (C reactive protein, soluble intercellular adhesion molecule 1, IL-6, IL-10, white blood cell counts, adiponectin)
- prothrombotic markers (fibrinogen, plasminogen activator inhibitor 1)
- Glycemic traits (fasting glucose, fasting insulin, insulin resistance, beta cell function)
- Lipid traits
- Blood pressure traits
- Cellular traits (proliferation rate)

- Disease risk (T1D, T2D, CD)

We Find Biologically Plausible Associations between Disease and Inflammatory Biomarkers



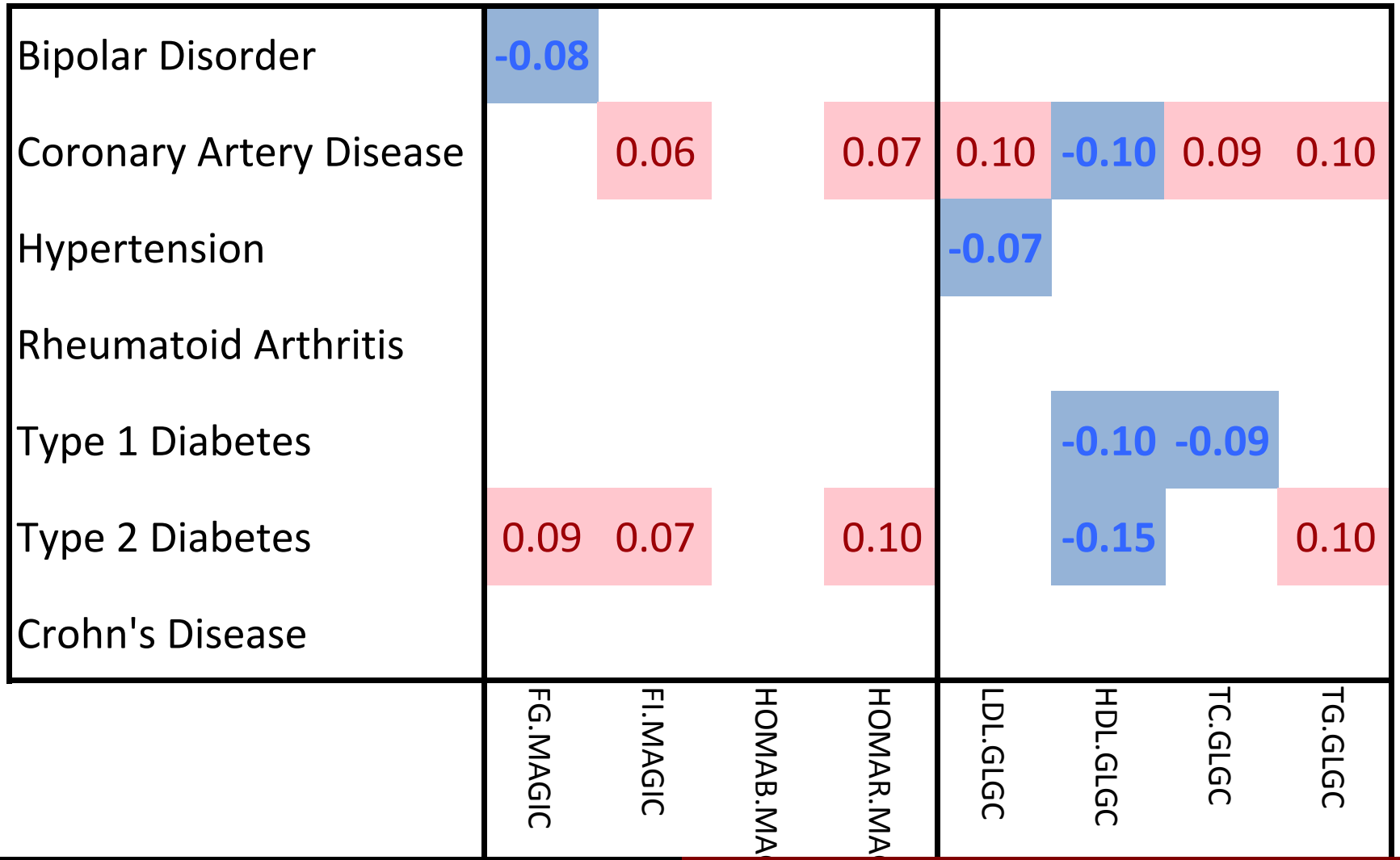
We Find Biologically Plausible Associations between Disease and Inflammatory Biomarkers

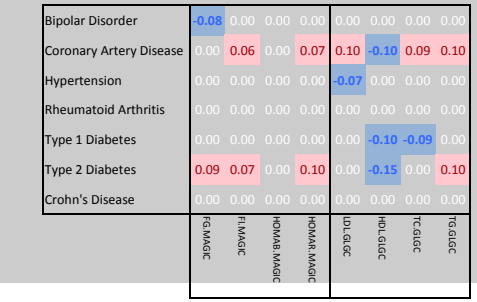


Adiponectin dysregulation and insulin resistance in type 1 diabetes. Pereira RI, Snell-Bergeon JK, Erickson C, Schauer IE, Bergman BC, Rewers M, Maahs DM. J Clin Endocrinol Metab. 2012 Apr;97(4):E642-7. doi: 10.1210/jc.2011-2542.

A small-molecule AdipoR agonist for type 2 diabetes and short life in obesity. Okada-Iwabu M, Yamauchi T, Iwabu M, Honma T, Hamagami KI, Matsuda K, Yamaguchi M, Tanabe H, Kimura-Someya T, Shirouzu M, Ogata H, Tokuyama K, Ueki K, Nagano T, Tanaka A, Yokoyama S, Kadowaki T. Nature. 2013 Oct 30. doi: 10.1038/nature12656. [Epub ahead of print]

We Find Biologically Plausible Associations between Disease and Metabolic Biomarkers





Is bipolar disorder an endocrine condition? Glucose abnormalities in bipolar disorder. Garcia-Rizo C, Kirkpatrick B, Fernandez-Egea E, Oliveira C, Meseguer A, Grande I, Undurraga J, Vieta E, Bernardo M. Acta Psychiatr Scand. 2013 Sep 12. doi: 10.1111/acps.12194.

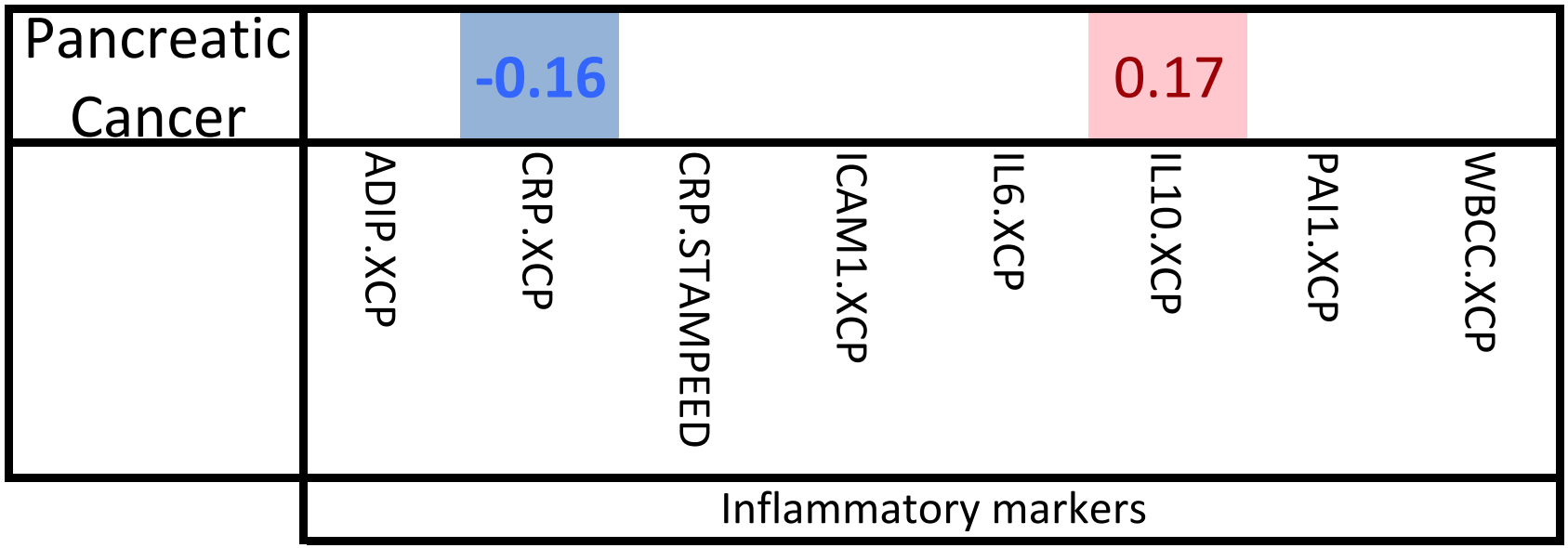
Fasting insulin concentrations and incidence of hypertension, stroke, and coronary heart disease: a meta-analysis of prospective cohort studies. Xun P, Wu Y, He Q, He K. Am J Clin Nutr. 2013 Oct

A higher fasting insulin concentration or hyperinsulinemia was significantly associated with an increased risk of hypertension and CHD but not stroke.

We Find Biologically Plausible Associations between Disease and Other Biomarkers

	BMI.GIANT	HEIGHT.GIA	WHRadjBMI	iGrowth.CEU	BRTHW.EGG	CD.IIBD	T1D.CHOP_r	T1D.CHOP_V
Bipolar Disorder				0.07		0.61	-0.08	-0.11
Coronary Artery Disease				0.07		0.57		
Hypertension						0.61		
Rheumatoid Arthritis			0.07			0.54	0.13	0.36
Type 1 Diabetes			-0.06	0.06		0.47	0.80	2.40
Type 2 Diabetes	0.16					0.58		
Crohn's Disease	0.07					2.85		0.14

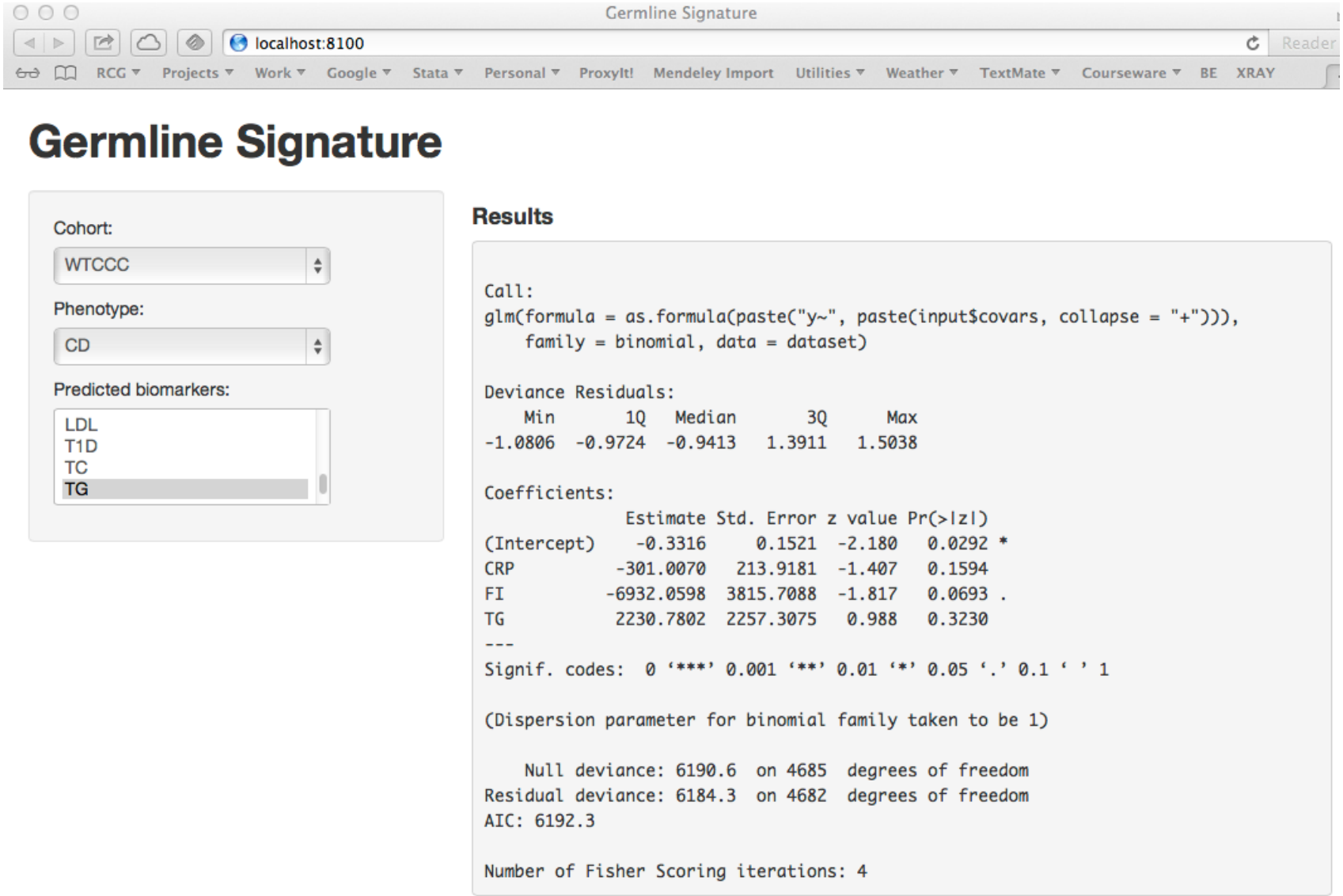
We Find Biologically Plausible Associations Between Pancreatic Cancer and Inflammatory Markers



Challenges to Address

- Detection dependent on prediction algorithm
- Improve predictive model
- Artifacts
  - overlapping samples
  - population stratification
- Interpretation
- Replication by independent
  - training set
  - testing set
- results by cohort rather than meta analysis

Web-based Discovery Engine



Summary Dissection Method

- Proposed a method to dissect biology of complex traits
  - genetic scores as proxies for lifetime exposure to given biomarker level
  - test effect of this "exposure" on disease risk
- No endophenotype data is needed, we compute it from genetic data
- Public database of predictive models for biomarkers - under construction
- Web resource to allow other investigators test their own hypotheses - under construction

Summary Dissection Method

- Proposed a method to dissect biology of complex traits
  - genetic scores as proxies for lifetime exposure to given biomarker level
  - test effect of this "exposure" on disease risk
- No endophenotype data is needed, we compute it from genetic data
- Public database of predictive models for biomarkers - under construction
- Web resource to allow other investigators test their own hypotheses - under construction

Summary Dissection Method

- Proposed a method to dissect biology of complex traits
  - genetic scores as proxies for lifetime exposure to given biomarker level
  - test effect of this "exposure" on disease risk
- No endophenotype data is needed, we compute it from genetic data
- Public database of predictive models for biomarkers - under construction
- Web resource to allow other investigators test their own hypotheses - under construction

Summary Dissection Method

- Proposed a method to dissect biology of complex traits
  - genetic scores as proxies for lifetime exposure to given biomarker level
  - test effect of this "exposure" on disease risk
- No endophenotype data is needed, we compute it from genetic data
- Public database of predictive models for biomarkers - under construction
- Web resource to allow other investigators test their own hypotheses - under construction

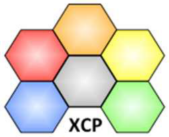
Summary Dissection Method

- Proposed a method to dissect biology of complex traits
  - genetic scores as proxies for lifetime exposure to given biomarker level
  - test effect of this "exposure" on disease risk
- No endophenotype data is needed, we compute it from genetic data
- Public database of predictive models for biomarkers - under construction
- Web resource to allow other investigators test their own hypotheses - under construction

Summary Dissection Method

- Proposed a method to dissect biology of complex traits
  - genetic scores as proxies for lifetime exposure to given biomarker level
  - test effect of this "exposure" on disease risk
- No endophenotype data is needed, we compute it from genetic data
- Public database of predictive models for biomarkers - under construction
- Web resource to allow other investigators test their own hypotheses - under construction

Summary result sources

- 

Inflammation markers: Cross Consortia Pleiotropy (XC-Pleiotropy) Consortium PMI Working Group
- **EGG results:** Data on birth weight trait has been contributed by EGG Consortium and has been downloaded from www.egg-consortium.org.
- **GIANT results:** Retrieved from `http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files`
- **T1D results:** Jonathan Bradfield, Hakon Hakonarson, and Struan Grant (personal communication)
- **MAGIC results:** Data on glycaemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org
- **Crohn's disease results:** Data downloaded from the International Inflammatory Bowel Disease Genetics Consortium website (IIBDGC) `http://www.ibdgenetics.org/downloads.html`
- **GLGC Lipids results:** Downloaded from `http://www.sph.umich.edu/csg/abecasis/public/lipids2010/`
- **CRP results data from STAMPEED:** The aggregate genomic datasets used for analysis described in this manuscript were obtained from the database of Genotypes and Phenotypes (dbGaP) found at `http://www.ncbi.nlm.nih.gov/gap` through dbGaP accession number phs000221.v1.p1.

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Final Summary

- Successes and limitations of GWAS/Sequencing studies
- Need to shift to cumulative effect paradigm given polygenic architecture of most complex disorders
- To speed up clinical translation, we propose systems approaches to prediction and dissection of complex disorders
- Proposed a prediction method that integrates multiple omics and prior knowledge
- Proposed a dissection method that probes the biology of complex disorders
  - Building publicly available database of predictive models of biomarkers
  - Building a web resource to allow others test their own hypotheses
- These approaches can be applied to existing data to gain new insight into the biology
- Both methods are applicable to common and rare variants

Acknowledgement

Nancy J. Cox
Keston Aquino-Michaels
Heather Wheeler
Eric R. Gamazon
R. Steph Huang
M. Eileen Dolan
Vassily Trubetskoy

CALGB 80303
Federico Innocenti
Kouros Owzar
Patrick Evans
Yusuke Nakamura

Funding Sources

GTEx
R01 MH090937 and R01 MH101820

PAAR NIH/NIGMS UO1GM61393

UCCC Support Grant
UC Cancer Center Support Grant
NCI P30 CA014599-36

UC DRTC
University of Chicago Diabetes Research and Training Center; P60 DK20595

Conte Center grant
P50MH094267

K12 Award

THE UNIVERSITY OF CHICAGO
INSTITUTE FOR TRANSLATIONAL MEDICINE
CLINICAL AND TRANSLATIONAL SCIENCE AWARD
BIOLOGICAL SCIENCES DIVISION
5841 S. Maryland Avenue, MC 6026
Chicago, IL 60637

January 10, 2014

Hae Kyung Im, PhD
University of Chicago
Department of Health Studies
5841 S Maryland Ave. Room – AMB R321B, MC 2007
Chicago, Illinois 60637

Re: **CTSA KL2/K12 SCHOLAR AWARDS**

I am pleased to inform you that you have been selected as a University of Chicago Paul Calabresi Clinical Oncology K12 Scholar. Congratulations and welcome to the program! The Institute for Translational Medicine (ITM) will provide you with:

- $94,950/year to be used for salary and fringe benefits at 75% effort
- $3,000/year in travel-related expenses, and
- $20,000 in other expenses, to be used as research and developmental support

Your award period is 08/01/13 to 07/31/15.

Diabetes Research and Training Center Pilot Grant

The University of Chicago
Diabetes Research and Training Center
Pilot and Feasibility Grant Program

5841 S. Maryland Avenue, MC 1027
Chicago, IL 60637
(773) 702-5849

January 27th, 2014

Dear Dr. Im,

We are very pleased to inform you that your Pilot and Feasibility application to the University of Chicago DRTC has been recommended for funding for the amount of $20,000 through January 31st, 2015. We realize that this is not the full amount that you requested, but we are trying to support as many meritorious grants as possible with finite funds. We hope that receipt of this award will assist you with your on-going research, and a second year of funding will be available upon demonstration of significant progress during the first year. We will work with the grants office to set up your account by February 1, 2014.

We request that all DRTC P&F grant recipients attend the Annual DRTC Diabetes Day which will be held on the University of Chicago campus May 17th, 2014. Attendance is required in order to be considered for a second year of funding. If there is a specific conflict, please let me know. We also request that the DRTC grant P30 DK020595 is acknowledged in any resulting future publications or presentations.

Thank You!

Kriging

Predicted Y is the weighted average of the observations

$$\text{Prediction}(Y_{\text{new}}) = \omega_1 Y_1 + \omega_2 Y_2 + \cdots + \omega_n Y_n$$

$$\omega_i = \text{function(all } n(n+1)/2 \text{ pairs of correlations)}$$

Without covariates

$$\boldsymbol{\omega}' = \boldsymbol{\rho}' \boldsymbol{\Sigma}^{-1}$$

$\boldsymbol{\rho}$ the correlation between the new value and the observed values and

$\boldsymbol{\Sigma}$ the correlation matrix of the observations.

- Prediction

$$(Y_{\text{new}}) = \omega_1 Y_1 + \omega_2 Y_2 + \cdots + \omega_n Y_n$$

- No covariates

$$\boldsymbol{\omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}$$

- With covariates

$$\boldsymbol{\omega} = \Sigma^{-1}(\boldsymbol{\rho} + \mathbb{Z}\mathbf{m})$$

$$\mathbf{m} = (\mathbb{Z}'\boldsymbol{\Sigma}^{-1}\mathbb{Z})^{-1}(\mathbf{z} - \mathbb{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho})$$

Optimal Similarity Matrix

- Depends on genetic architecture of trait
- Linear Mixed Model (LMM)

Yang, Lee, Goddard, and Visscher, (2011), AJHG

$$\begin{aligned} Y_i &= \mu + \sum_{k=1}^{p} \beta_k X_{i,k} + e_i \\ \beta &\sim \mathrm{N}(0, \sigma^2_\beta) \end{aligned}$$

Optimal Similarity $\longrightarrow XX' = \mathrm{GRM}$

Optimal Similarity Matrix

- Bayesian Sparse Linear Mixed Model (BSLMM)

  Zhou, Carbonetto, and Stephens (2013) Plos Genetics

$$Y_i = \mu + \sum_{k=1}^{m} \beta_{L,k} X_{i,k} + \sum_{k=1}^{p} \beta_{S,k} X_{i,k} + e_i$$

$$\beta_x \sim \mathrm{N}(0,\sigma_x^2);\ \beta_S \sim \mathrm{N}(0,\sigma_S^2);\ m << p$$

Optimal Similarity $\longrightarrow \theta_L GRM_L + \theta_S GRM_S$

Optimal Similarity Matrix

$$Y_i = a + G_i + T_i + O_i + \epsilon_i$$

Assuming independence of β's

$$\Sigma_{i,j} = \theta_G \sum_{l=1}^{M} X_{il}^G X_{jl}^G + \theta_T \sum_{l=1}^{L} X_{il}^T X_{jl}^T + \theta_O \sum_{l=1}^{L'} X_{il}^O X_{jl}^O + \theta_\epsilon \delta_{ij}$$

More generally

$$\Sigma_{i,j} = \theta_G \sum_{l=1}^{M} X_{ik}^G X_{jk}^G + \theta_T \sum_{l=1}^{L} X_{ik}^T X_{jk}^T + \theta_O \sum_{k=1}^{L'} X_{ik}^O X_{jk}^O + \theta_\epsilon \delta_{ij} + \sum_{k \neq l} \text{cov}(\beta_k, \beta_l) X_{ik} X_{jl}$$

Predicting Disease Risk using OmicKriging

- Wellcome Trust Case Control Consortium
  - BD: bipolar disorder
  - CAD: coronary artery disease
  - HT: hypertension
  - T1D: type 1 diabetes
  - T2D: type 2 diabetes
  - CD: Crohn's disease
  - RA: rheumatoid arthritis
- Two approaches
  - Whole genome prediction no prior information (∼400K snps) $\beta_k \sim N(0, \sigma^2_\beta)$
  - Whole genome prediction with more weight to known $\beta_k \sim \pi N(0, \sigma^2_a + \sigma^2_b) + (1 - \pi) N(0, \sigma^2_b)$

Predicting Disease Risk using OmicKriging

- Wellcome Trust Case Control Consortium
  - BD: bipolar disorder
  - CAD: coronary artery disease
  - HT: hypertension
  - T1D: type 1 diabetes
  - T2D: type 2 diabetes
  - CD: Crohn's disease
  - RA: rheumatoid arthritis
- Two approaches
  - Whole genome prediction no prior information (∼400K snps) $\beta_k \sim N(0, \sigma^2_\beta)$
  - Whole genome prediction with more weight to known $\beta_k \sim \pi N(0, \sigma^2_a + \sigma^2_b) + (1 - \pi) N(0, \sigma^2_b)$

Predicting Disease Risk using OmicKriging

- Wellcome Trust Case Control Consortium
  - BD: bipolar disorder
  - CAD: coronary artery disease
  - HT: hypertension
  - T1D: type 1 diabetes
  - T2D: type 2 diabetes
  - CD: Crohn's disease
  - RA: rheumatoid arthritis
- Two approaches
  - Whole genome prediction no prior information (∼400K snps) $\beta_k \sim N(0, \sigma^2_\beta)$
  - Whole genome prediction with more weight to known $\beta_k \sim \pi N(0, \sigma^2_a + \sigma^2_b) + (1 - \pi) N(0, \sigma^2_b)$

Statin Response GWAS Findings Not Replicated

Impact of common genetic variation on response to simvastatin therapy among 18 705 participants in the Heart Protection Study

Jemma C. Hopewell[1]*, Sarah Parish[1], Alison Offer[1], Emma Link[1], Robert Clarke[1], Mark Lathrop[2], Jane Armitage[1], and Rory Collins[1], on behalf of the MRC/BHF Heart Protection Study Collaborative Group

[1]Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), University of Oxford, Oxford, UK; and [2]Centre National de Génotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry, France

Received 23 April 2012; revised 3 August 2012; accepted 18 September 2012; online publish-ahead-of-print 24 October 2012

See page 949 for the editorial comment on this article (doi:10.1093/eurheartj/ehs439)

Aims Statins reduce LDL cholesterol (LDL-C) and the risk of vascular events, but it remains uncertain whether there is clinically relevant genetic variation in their efficacy. This study of 18 705 individuals aims to identify genetic variants related to the lipid response to simvastatin and assess their impact on vascular risk response.

Methods and results A genome-wide study of the LDL-C and apolipoprotein B (ApoB) response to 40 mg simvastatin daily was performed in 3895 participants in the Heart Protection Study, and the nine strongest associations were tested in 14 810 additional participants. Selected candidate genes were also tested in up to 18 705 individuals. There was 90% power to detect differences of 2.5% in LDL-C response (e.g. 42.5 vs. 40% reduction) in the genome-wide study and of 1% in the candidate gene study. None of the associations from the genome-wide study was replicated, and nor were significant associations found for 26 of 36 candidates tested. Novel lipid response associations with variants in *LPA, CELSR2/PSRC1/SORT1*, and *ABCC2* were found, as well as confirmatory evidence for published associations in *LPA, APOE*, and *SLCO1B1*. The largest and most significant effects were with *LPA* and *APOE*, but were only 2–3% per allele. Reductions in the risk of major vascular events during 5 years of statin therapy among 18 705 high-risk patients did not differ significantly across genotypes associated with the lipid response.

Conclusions Common genetic variants do not appear to alter the lipid response to statin therapy by more than a few per cent, and there were similar large reductions in vascular risk with simvastatin irrespective of genotypes associated with the lipid response to simvastatin. Consequently, their value for informing clinical decisions related to maximizing statin efficacy appears to be limited.

Keywords Pharmacogenetics • Statins • LDL-C • ApoB

Cholesterol and Pharmacogenetics Simvastatin Study

- 562 genotyped individuals (QC)
- gene expression for 480 individuals
  - LCL were established
  - Simvastatin treated and control gene expression
- Outcome: change in patient's LDL-C after simvastatin treatment

OmicKriging Outperforms GWAS Hits-only Prediction

- Used SNPs 50Kb around 45 statin response loci from the Heart Protection Study ($\sim$10K SNPs)
- Baseline model (45 SNPs) has no predictive power ($R^2 \sim 0$)
- Using top 1K SNPs improves prediction ($R^2 = 1.6\%$)
- OmicKriging $R^2 = 3.7\%$

We Assume an Additive Genetic Model for Biomarkers

$$\text{biomarker level} = \underset{\text{component}}{\text{Genetic}} + \underset{\text{component}}{\text{Environmental}}$$

$$\text{Genetic component} = \sum_{k=1}^{M} \beta_k X_k$$

X_i	# of reference alleles
β_k	true effect sizes

gSigDB: Database of predictive models