The Role of Genomic Prediction in Precision Medicine

Hae Kyung Im, PhD

March 6, 2015
Overview

- Precision Medicine Initiative
- Monogenic vs. polygenic traits
- Review of prediction methods
- Poly-Omic integration: OmicKriging
- Role of regulatory variants in complex traits
- PrediXcan
- Prediction of gene expression traits
Precision Medicine

- Obama: Precision Medicine Initiative $215M for 2016 Budget
- Instead of “one-size fits-all-approach”
- “Right treatment, at the right time to the right person”
- Innovative approach to disease prevention and treatment based on individual differences in genes, environments, and lifestyles

http://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative
Precision Medicine Implementation

**Prediction**
- Disease
  - risk stratification
  - intervention strategies
- Adverse events
- Efficacy of treatment

**Dissection**
- Etiology of complex traits
- Mechanism by which genetic variation drives phenotypic variation
- Druggable targets
In year 2000, president Clinton announced the completion of the first draft of the human genome, which would "revolutionize the diagnosis, prevention, and treatment of most, if not all, human diseases.

Francis Collins predicted that diagnosis of genetic diseases would be accomplished by 2010 and that treatments would start to roll out perhaps by 2015.

Why are we not there yet?
The Promise of the Human Genome Sequencing Project

- In year 2000, president Clinton announced the completion of the first draft of the human genome, which would "revolutionize the diagnosis, prevention, and treatment of most, if not all, human diseases.

- Francis Collins predicted that diagnosis of genetic diseases would be accomplished by 2010 and that treatments would start to roll out perhaps by 2015.

- Why are we not there yet?
In year 2000, president Clinton announced the completion of the first draft of the human genome, which would revolutionize the diagnosis, prevention, and treatment of human diseases.

Francis Collins predicted that diagnosis of genetic diseases would be accomplished by 2010 and that treatments would start to roll out perhaps by 2015.

Why are we not there yet?

Genetic architecture is much more complex than anticipated.
Monogenic vs. Polygenic Architecture
Genetic Architecture of Complex Traits
Genetic Architecture of Complex Traits
Single Variants Not Relevant for Highly Polygenic Traits
Genes mirror geography within Europe

John Novembre\textsuperscript{1,2}, Toby Johnson\textsuperscript{4,5,6}, Katarzyna Bryc\textsuperscript{7}, Zoltán Kutalik\textsuperscript{4,6}, Adam R. Boyko\textsuperscript{7}, Adam Auton\textsuperscript{7}, Amit Indap\textsuperscript{7}, Karen S. King\textsuperscript{8}, Sven Bergmann\textsuperscript{4,6}, Matthew R. Nelson\textsuperscript{8}, Matthew Stephens\textsuperscript{2,3} & Carlos D. Bustamante\textsuperscript{7}
Whole Genome Prediction Methods
Additive Genetic Model

\[ Y = \sum_{k=1}^{M} \beta_k X_k + \epsilon \]

Univariate Regression

GWAS

Penalized regression

Ridge

LASSO

Elastic Net

\[ \| Y - X_k \beta_k \|_2 \]

\[ \| Y - \sum_k X_k \beta_k \|_2 + \lambda_1 \| \beta \|_1 + \lambda_2 \| \beta_2 \|_2 \]
Common polygenic variation contributes to risk of schizophrenia and bipolar disorder

The International Schizophrenia Consortium*

\[ Y = \sum_{k=1}^{M} \hat{\beta}_k^{GWAS} X_k \]
Best Linear Unbiased Prediction (BLUP)/Ridge

REPORT

GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,1,* S. Hong Lee,1 Michael E. Goddard,2,3 and Peter M. Visscher1

Penalized regression

\[
Y = \sum_{k=1}^{M} \hat{\beta}_k \text{Ridge} X_k
\]

\[
\|Y - \sum_k X_k \beta_k\|_2^2 + \lambda_2 \|eta_2\|_2
\]
Regularization and variable selection via the elastic net

Hui Zou and Trevor Hastie
Stanford University, USA

\[
Y = \sum_{k=1}^{M} \hat{\beta}_k^{E-N} X_k
\]

\[
\| Y - \sum_k X_k \beta_k \|_2 + \lambda_1 \| \beta \|_1 + \lambda_2 \| \beta_2 \|_2
\]
Polygenic Modeling with Bayesian Sparse Linear Mixed Models

Xiang Zhou¹*, Peter Carbonetto¹, Matthew Stephens¹,²*

\[ Y = \sum_{k=1}^{M} \beta_k^L X_k + \sum_{k=1}^{M} \beta_k^S X_k + \epsilon \]

\[ \beta_k^L \sim N(0, \sigma^2_L) \]

\[ \beta_k^S \sim N(0, \sigma^2_S) \]

MultiBLUP: improved SNP-based prediction for complex traits

Doug Speed and David J Balding

*Genome Res.* published online June 24, 2014
Access the most recent version at doi:10.1101/gr.169375.113
Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies

Nilanjan Chatterjee¹, Bill Wheeler², Joshua Sampson¹, Patricia Hartge¹, Stephen J Chanock¹ & Ju-Hyun Park¹,³

---

**Figure 1** Predictive Performance

- **c** T1D without MHC region
- **d** T2D
- **e** Prostate cancer

- **AUC**
- **PCC**

---

**Millions of Samples Needed to Achieve Maximum Predictive Performance**

Nature Genetics 2013
OmicKriging: Integration of Multiple Omics Data
What is Kriging?

Prediction by kriging

\[ \text{?} = w_1 + w_2 + w_3 \]

Closer locations get larger weights

Locations

Physical proximity

Physical similarity

Physical variable (ex. rainfall)

Complex trait (ex. height)

What is Kriging?
What is Kriging?

Physical variable (ex. rainfall) → Complex trait (ex. height)

Prediction by kriging

\[ ? = w_1 + w_2 + w_3 \]

- Closer locations get larger weights
- More related individuals get larger weights
- Locations
- Individuals
- Physical proximity
- Omic proximity
- Physical similarity
- Phenotypic similarity
Kriging

Predicted $Y$ is the weighted average of the observations

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

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

Without covariates

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

$\rho$ the correlation between the new value and the observed values and

$\Sigma$ the correlation matrix of the observations.
<table>
<thead>
<tr>
<th>Father</th>
<th>Mother</th>
<th>Sons in order of height</th>
<th>Daughters in order of height</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.5</td>
<td>7.0</td>
<td>13.2</td>
</tr>
<tr>
<td>2</td>
<td>15.5</td>
<td>6.5</td>
<td>13.5, 12.5</td>
</tr>
<tr>
<td>3</td>
<td>15.0</td>
<td>about 4.0</td>
<td>11.0</td>
</tr>
<tr>
<td>4</td>
<td>15.0</td>
<td>4.0</td>
<td>10.5, 8.5</td>
</tr>
<tr>
<td>5</td>
<td>15.0</td>
<td>-1.5</td>
<td>12.0, 9.0, 8.0</td>
</tr>
<tr>
<td>6</td>
<td>14.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>14.0</td>
<td>8.0</td>
<td>16.5, 14.0, 13.0, 13.0</td>
</tr>
<tr>
<td>8</td>
<td>14.0</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>14.5</td>
<td>6.0</td>
<td></td>
</tr>
</tbody>
</table>

Galton Was Kriging with Kinship Matrix (1885)

<table>
<thead>
<tr>
<th>HEIGHT in inches</th>
<th>DEVIATE in inches</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>+4</td>
</tr>
<tr>
<td>71</td>
<td>+3</td>
</tr>
<tr>
<td>70</td>
<td>+2</td>
</tr>
<tr>
<td>69</td>
<td>+1</td>
</tr>
<tr>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>-1</td>
</tr>
<tr>
<td>66</td>
<td>-2</td>
</tr>
<tr>
<td>65</td>
<td>-3</td>
</tr>
<tr>
<td>64</td>
<td>-4</td>
</tr>
</tbody>
</table>

When Mid-Parents are taller than mediocrity, their Children tend to be shorter than they.

When Mid Parents are shorter than mediocrity, their Children tend to be taller than they.
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)
- G-BLUP: genetic relatedness estimated using genotype
- BLUP/Kriging can be interpreted as the posterior mean of the genetic component given observations ($Y = G +$ error)
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)
- G-BLUP: genetic relatedness estimated using genotype
- BLUP/Kriging can be interpreted as the posterior mean of the genetic component given observations \((Y = G + \text{error})\)

BLUP/Kriging translates genetic similarity into phenotypic prediction
Polyomic Model

\[ Y_i = a + G_i + T_i + O_i + \epsilon_i \]

\[ G_i = \sum_{l=1}^{M} \beta_i^G X_{il}^G \]  
\hspace{1cm} \text{genetic component}

\[ T_i = \sum_{l=1}^{L} \beta_i^T X_{il}^T \]  
\hspace{1cm} \text{transcriptomic component}

\[ O_i = \sum_{l=1}^{L'} \beta_i^O X_{il}^O \]  
\hspace{1cm} \text{other omic component}

\[ (\beta_G, \beta_T, \beta_O)' \sim N(0, \Sigma_\beta) \]
Optimal Similarity Matrix

\[ Y_i = a + G_i + T_i + O_i + \epsilon_i \]

Assuming independence of \( \beta \)'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

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

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

**mRNA**
- mRNA $R^2=38\%$

**microRNA**
- microRNA $R^2=35\%$

**mRNA + microRNA**
- mRNA+microRNA $R^2=48\%$
Application to Wellcome Trust Case Control Consortium

- WTCCC
- 7 disease cases and 2 control sets:
  - Coronary Artery Disease (2000)
  - Hypertension (2000)
  - Type 2 Diabetes (2000)
  - Bipolar Disorder (2000)
  - Crohn’s Disease (2000)
  - Rheumatoid Arthritis (2000)
  - Type 1 Diabetes (2000)
  - 1958 Birth Cohort (1500)
  - UK National Blood Services (1500)
GWAS hits vs. Whole Genome Prediction (OmicKriging)

- **Coronary Artery Disease**
- **Hypertension**
- **Type 2 Diabetes**
- **Bipolar Disorder**
- **Crohn's Disease**
- **Rheumatoid Arthritis**
- **Type 1 Diabetes**

**Methods**:
- **Baseline**
- **OK:SingleGRM**
- **OK:DoubleGRM**
- **OmicKriging: WTCCC 7 Diseases**

**AUC** values for each disease category using different methods.
OmicKriging: OmicKriging for Phenotypic Prediction

This package provides functions to generate a correlation matrix from a genetic dataset and to use this matrix 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: 1.3
Depends: R (≥ 2.15.1), doParallel
Imports: ROCR, irbA
Published: 2014-06-18
Author: Hae Kyung Im, Heather E. Wheeler, Keston Aquino Michaels, Vassily Trubetskoy
Maintainer: Hae Kyung Im <haky at uchicago.edu>
License: GPL (≥ 3)
NeedsCompilation: no
Materials: README
CRAN checks: OmicKriging results

Downloads:
Reference manual: OmicKriging.pdf
Vignettes: Application Tutorial: OmicKriging
Package source: OmicKriging_1.3.tar.gz
Windows binaries: r-devel: OmicKriging_1.3.zip, r-release: OmicKriging_1.3.zip, r-oldrel: OmicKriging_1.3.zip
OS X Snow Leopard binaries: r-release: OmicKriging_1.3.tgz, r-oldrel: OmicKriging_1.3.tgz
OS X Mavericks binaries: r-release: OmicKriging_1.3.tgz
Old sources: OmicKriging archive
Poly-Omic Prediction of Complex Traits: OmicKriging

Heather E. Wheeler,1 Keston Aquino-Michaels,2 Eric R. Gamazon,2 Vassily V. Trubetskoy,2 M. Eileen Dolan,1 R. Stephanie Huang,1 Nancy J. Cox,2 and Hae Kyung Im3*

1Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, Illinois, United States of America; 2Section Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, United States of America; 3Department of Health Studies, L Chicago, Chicago, Illinois, United States of America

Received 26 November 2013; Revised 11 March 2014; accepted revised manuscript 12 March 2014.
Published online 2 May 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21808

ABSTRACT: High-confidence prediction of complex traits such as disease risk or drug response is an ultimate goal of personalized medicine. Although genome-wide association studies have discovered thousands of well-replicated polymor
- OmicKriging is a systems approach to complex trait prediction that leverages and integrates multiple omic data
- We can attain relevant prediction even if we do not know the individual variant’s contribution
- Important tool for integrating the vast amounts of data to be generated with the precision medicine initiative
Role of Regulatory Variation in Complex Traits
- Most trait-associated SNPs are not coding
- Mechanism via regulation of gene expression levels
Altered Protein Levels Influences Disease Risk

Albert & Kruglyak 2015 NGReviews
Regulatory variants explain much more heritability than coding variants across 11 common diseases

Alexander Gusev, S Hong Lee, Benjamin M Neale, et al.

Figures

Figure 1. Functional partitioning of SNP-heritability across eleven traits. For each trait, the liability explained by SNPs in six functional categories was jointly estimated, with the meta-analysed average shown in filled bars. The null expectation, equal to the percent of SNPs in each category, is shown by dashed, unfilled bars, with p-value reporting the difference from this expectation. Fold-enrichment relative to the null expectation shown in parenthesis below each category label. Left panel shows results from analyses of genotyped SNPs only, right panel shows analysis of genotyped and 1,000 Genomes imputed SNPs. Error bars define 95% confidence interval.

- Coding: 4.1x
- UTR: 3.5x
- Promoter: 2.2x
- DHS: 1.6x
- Intron: 0.8x
- Intergenic: 0.6x

Mean observed Expected (% SNPs)

- Coding: 13.8x
- UTR: 8.4x
- Promoter: 2.8x
- DHS: 5.1x
- Intron: 0.1x
- Intergenic: −0.1x

DHS: DNAse hypersensitivity sites, control accessibility of the region thus levels of transcription

<1e−20

5.5e−12

4.7e−04

4.3e−03

1.2e−01
PrediXcan

Nature Genetics - under revision
Motivation for PrediXcan

- Lack of mechanistic understanding of most GWAS discoveries
- Large proportion of variation explained by regulatory variants
- We propose PrediXcan that tests the proposed mechanism
Genetic Control of Disease Through Gene Regulation

G → t(G) → t(G) + e → Disease Status

Genetic Variation

Genetically Determined Expression

Disease Liability

Affected

Unaffected
Mechanisms Tested by PrediXcan

- **PrediXcan**
  - **GReX** (Genetically regulated expression)
  - Other factors
  - Trait-altered component

Gene Expression Decomposition
### Genetic Variation

<table>
<thead>
<tr>
<th>n individuals</th>
<th>M SNPs</th>
<th>id</th>
<th>id2</th>
<th>id3</th>
<th>\ldots</th>
<th>idn</th>
</tr>
</thead>
<tbody>
<tr>
<td>id1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>id2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>id3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>idn</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Observed Transcriptome

<table>
<thead>
<tr>
<th>m genes</th>
<th>Tissue-1</th>
<th>Tissue-2</th>
<th>Tissue-p</th>
</tr>
</thead>
<tbody>
<tr>
<td>gm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>id1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>id2</td>
<td>2.2</td>
<td>1.7</td>
<td>1.2</td>
</tr>
<tr>
<td>id3</td>
<td>1.3</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>idn</td>
<td>1.2</td>
<td>2.2</td>
<td>3.1</td>
</tr>
</tbody>
</table>

PrediXcan uses Reference Transcriptome
PredictDB: Public Database of Weights for GReX

Additive model of gene expression trait trained in reference transcriptome datasets

\[ T = \sum_{k} w_k X_k + \epsilon \]

Weights stored in PredictDB
**PrediXcan Imputes Transcriptome & Tests Assoc.**

### Genetic Variation

<table>
<thead>
<tr>
<th>n' individuals</th>
<th>M SNPs</th>
<th>&quot;Imputed&quot; Transcriptome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>id</strong></td>
<td><strong>g1</strong></td>
<td><strong>g2</strong></td>
</tr>
<tr>
<td>id1</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>id2</td>
<td>1</td>
<td>2.3</td>
</tr>
<tr>
<td>id3</td>
<td>2</td>
<td>3.3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>idn'</td>
<td>1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Weights stored in PredictDB

Reference transcriptome

Association Test

**GReX**

PrediXcan on GWAS Data

Translating Genomics
PrediXcan: Mechanism-driven Gene-Based Test

- Directly tests the molecular mechanism through which genetic variants affect phenotype
- Genes more attractive than genetic variants
  - A lot is known about their function
  - Follow up experiments can be easily devised
  - Reduced multiple testing burden
- Direction of effects
  - Positive effects: down regulation is therapeutic option
  - Negative effects: more likely to harbor deleterious rare variants
- No reverse causality issues
- Can be systematically applied to existing GWAS studies
- Tissue-specificity can be inferred
Reference Transcriptome Data

- GTEx - Genotype of Tissue Expression
  - Large scale Common Fund project
  - 900 organ donors
  - 45 tissues
  - RNAseq, whole exome seq, whole genome seq
- GEUVADIS
  - RNAseq 462 individuals from the 1000 Genomes Project
- Cerebellum expression (Array GSE35974)
- Framingham, n>5000m, whole blood
- Depression Genes & Networks, n>900, whole blood
Good Prediction Performance

**Prediction R^2**

Training with GTEx
Testing in 1K Genomes

Sahar Mozaffari

**Replication R^2**

Replicate RNAseq
Pickrell et al 2010 vs. 1K Genomes 2013
Examples of Well Predicted Genes

Sahar Mozaffari

Training with GTEx
Testing in 1K Genomes
Genes Associated with Rheumatoid Arthritis

- RSBN1: Type 1 diabetes gene, an immune disease
- HLA genes
- PSME1: involved in processing of class I MHC peptides
PrediXcan Results for Crohn’s Disease and Hypertension

Crohn’s Disease

IRGM is a known Crohn’s gene

Hypertension

Whole blood may not be relevant tissue
PrediXcan Outperforms VEGAS

Eric Gamazon
Enrichment of Known Crohn’s Genes Among Findings

100 qqplot with random samples of 205 genes

Above this line implies enrichment
No Enrichment Among Hypertension Findings

100 qqplots with random samples of 133 genes

Above this line would imply enrichment

Whole blood may not be relevant tissue
PrediXcan: a Gene Discovery Approach

- PrediXcan is a powerful gene based association test
- It directly tests the molecular mechanism through which genetic variants affect phenotype
- Reduced multiple testing burden compared to single variant approach
- Unlike other gene based tests, it provides direction of effects
- Advantages relative to gene expression studies
  - Applicable to any GWAS datasets
gene expression levels are predicted from genotype data
  - No reverse causality
disease status does not affect germline DNA
  - Multiple Tissues can be evaluated
tissue expressions are only needed to build prediction models
Prediction of Gene Expression Traits
Genetic Architecture to Improve Prediction

- Local and distant regulation (heritability)

- Sparsity/Polygenicity

- This information guides us to improve prediction, i.e. estimates of GReX
Local/Distant Heritability Estimation

- Gene expression trait model

\[ Y = \sum_{\text{local}} \beta_k^{\text{local}} X_k + \sum_{\text{distant}} \beta_k^{\text{distant}} X_k + \epsilon \]

- REML to estimation of local and distant contributions jointly
- Covariance of local component: GRM using SNPs nearby
- Covariance of distant component: GRM using distant SNPs
- We use GCTA as REML calculator

Total Heritability = Local H2 + Distant H2
Whole Blood Expression Data: DGN

- Whole blood from Depression Genes and Networks study
- n = 922
- RNA-seq
Local Heritability Can Be Well Estimated

Local (joint) Heritability

h2
0.00
0.25
0.50
0.75
1.00

1:nrow(data)
0
5000
10000

@hakyim
Distant Heritability Not Reliable
- Only local component can be assessed
- LASSO performs slightly better than E-N 0.50 in cross validated R2
Performance vs sparsity

Elastic Net DGN-WB chr22 (341 genes)
E-N & LASSO Outperform Polygenic Score
Whole Blood DGN (n=922) + 38 GTEx Tissue Models

<table>
<thead>
<tr>
<th>Adipose-Subcutaneous 0.5.db</th>
<th>AdrenalGland_0.5.db</th>
<th>Artery-Aorta_0.5.db</th>
<th>Artery-Coronary_0.5.db</th>
<th>Brain-Anterior...4)_0.5.db</th>
<th>Brain-Caudate...)_0.5.db</th>
<th>Brain-Cerebell...re_0.5.db</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain-Cerebellum_0.5.db</td>
<td>Brain-Cortex_0.5.db</td>
<td>Brain-Frontal...0.5.db</td>
<td>Brain-Hippocampus...0.5.db</td>
<td>Brain-Hypothalamus...0.5.db</td>
<td>Brain-Nucleus...a)_0.5.db</td>
<td>Brain-Putamen...)_0.5.db</td>
</tr>
<tr>
<td>Breast-Mammary...0.5.db</td>
<td>Cells-EBV...0.5.db</td>
<td>Cells-Transverse...0.5.db</td>
<td>Colon-Sigmoid_0.5.db</td>
<td>Colon-Transverse_0.5.db</td>
<td>DGN-WB_0.5.db</td>
<td>Esophagus-Gastroesop..._0.5.db</td>
</tr>
<tr>
<td>Esophagus-Mucosa_0.5.db</td>
<td>Esophagus-Muscularis_0.5.db</td>
<td>Heart-Atrio...0.5.db</td>
<td>Heart-LeftVentricle_0.5.db</td>
<td>Liver_0.5.db</td>
<td>Lung_0.5.db</td>
<td>Nerve-Tibial_0.5.db</td>
</tr>
<tr>
<td>Ovary_0.5.db</td>
<td>Pancreas_0.5.db</td>
<td>Pituitary_0.5.db</td>
<td>Skin-NotSunExposed...0.5.db</td>
<td>Skin-SunExposed...g)_0.5.db</td>
<td>SmallIntestine...m_0.5.db</td>
<td>Spleen_0.5.db</td>
</tr>
<tr>
<td>Stomach_0.5.db</td>
<td>Testis_0.5.db</td>
<td>Thyroid_0.5.db</td>
<td>WholeBlood_0.5.db</td>
<td>WholeBlood_0.5.db</td>
<td>WholeBlood_0.5.db</td>
<td>WholeBlood_0.5.db</td>
</tr>
</tbody>
</table>
Challenges in Pharmacogenomic Predictions
# Pharmacogenomic Findings

<table>
<thead>
<tr>
<th>Evidence Level</th>
<th>Counts</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>1b</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>2a</td>
<td>96</td>
<td>6</td>
</tr>
<tr>
<td>2b</td>
<td>74</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1175</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>145</td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1547</td>
<td>100</td>
</tr>
</tbody>
</table>

Only Level 1a findings have clinical guidelines

https://www.pharmgkb.org/
Challenges of Pharmacogenomic Studies

- Smaller sample size
- Even more important to integrate prior data
- Integrate other functional data
- Heritability estimates are harder
  - Limited family data
  - Usually samples greater than 1K are needed for GCTA
Bevacizumab Induced Hypertension

- Bevacizumab is a humanized monoclonal antibody that inhibits VEGF induced angiogenesis
- Hypertension is a common adverse event to bevacizumab treatment
- The incidence of hypertension with bevacizumab is 20-30%, while grade 3 or greater hypertension occurs in only 10-15% of patients.

Keston Aquino Michaels
Bevacizumab Trials

- CALGB 90401
  - a randomized double-blinded placebo controlled phase III trial comparing docetaxel and prednisone with and without bevacizumab in men with hormone refractory prostate cancer
  - n = 664 (with genotype data after QC)
  - PI: Howard McLeod

- CALGB 80303
  - a randomized phase III trial of gemcitabine plus bevacizumab versus gemcitabine plus placebo in patients with advanced pancreatic cancer
  - n = 152 (with genotype data after QC)
  - PI: Federico Innocenti
Bevacizumab Induced Hypertension

- Is primary hypertension risk score predictive of bevacizumab induced hypertension
  - Hypertension results from Cross Consortia Pleiotropy group (n~20K)
- Can we predict drug induced hypertension?
  - 90401 training set
  - 80303 test set
Primary Hypertension Score Predicts Bev-induced HT

AUC = 0.62

Keston Aquino Michaels
Bev-Hypertension Predicted Within Study

Two steps LASSO + Random Forest

Cross Validated in Training Set

Keston Aquino Michaels

AUC = 0.71
Pharmacogenomics Beyond Single Variants

Bev-Hypertension Predicted in Independent Study

AUC = 0.68

Two steps LASSO + Random Forest

Validated in Independent Set
AUC down to 0.68 from 0.71

Keston Aquino Michaels
Bev-Hypertension Predicted in Independent Study

AUC = 0.72

Two steps LASSO + Random Forest + Primary Hypertension Score

Validated in Independent Set

Keston Aquino Michaels
Summary Pharmacogenomics

- Most single variant findings have limited clinical utility
- Whole genome approaches to prediction improves utility
- Bevacizumab induced hypertension example
  - primary hypertension results help in predicting drug induced hypertension
  - successfully predicted bevacizumab induced hypertension in independent study
  - combining primary + bevacizumab induced HT leads to improved prediction
Summary

- Shift from monogenic to polygenic paradigm
- Systems approach to genomics
  - Most single variant findings have limited clinical utility
  - Whole genome approaches to prediction improves utility
- Larger sample sizes will be needed, 1Million+
- OmicKriging: prediction method that integrates heterogeneous sources of data well suited for data from the Precision Medicine Initiative
- Large role of regulation variants in complex traits
- PrediXcan: novel gene based test that test mechanism
- Prediction of gene expression traits
Conclusion

- recognizing the complexity of the genetic architecture and mechanisms of genetic control,
- collecting deep phenotype data from large number of individuals,
- broadly sharing data and results, and
- integrating multiple sources of data
- using mechanism-driven tests

We will achieve the promise of precision medicine
Thank You!

Contributors
- Heather Wheeler
- Nancy Cox
- Eric Gamazon
- Keston Aquino Michaels
- Sahar Mozaffari
- Kaanan P. Shah
- Nicholas Knoblauch
- Vassily Trubetskoy
- GTEx Consortium

Data sources
- WTCCC
- GAINS/Bipolar Disorder
- GoKinD
- Disease Genes & Networks

Funding
- HKI was funded in part by Uchicago CTSA NCI K12CA139160
- University of Chicago Diabetes Research and Training Center: P60 DK20595, P30 DK020595
- Genotype of Tissue Expression GTEx R01 MH090937 and R01 MH101820
- P50DA037844 - Integrated GWAS of complex behavioral and gene expression traits in outbred rats
- Pharmacogenomics of Anticancer Agents PAAR UO1GM61393
- Pharmacogenomics Research Network (PGRN) Statistical Analysis Resource (P-STAR) U19 HL065962
- Conte Center grant P50MH094267
Lipid Markers AUC

Manickam et al 2011 J Clinical Lipidology

![Graph showing Lipid Markers AUC](image)

**Lipid Markers AUC**

- LDL AUC ~ 0.52
- HDL AUC = 0.60
Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS

Dan L. Nicolae¹,²,³, Eric Gamazon¹, Wei Zhang¹, Shiwei Duan¹, M. Eileen Dolan¹,², Nancy J. Cox¹

Annotation to Enhance Discovery from GWAS

Plos Genetic 2010