Using integrative genomics to assess therapeutic targets and GxE interactions

Simin Liu, MD, ScD

Professor of Epidemiology and Medicine
Director, Center for Global Cardiometabolic Health
Brown University
Overview

- **Conceptual framework**
  Different paradigms to study interactions

- **Some proof-of-concepts studies**
  1. GWAS, Mendelian randomization, pathways/networks analysis
     Identification of sex-steroids/SHBG as important drivers for diabetes and novel key regulators for CVD and diabetes *(Bottom-up)*
  2. Mg-related ion channels genes x Mg intake in diabetes risk *(Bottom-up)* and Mg short-term intervention/perturbation to gain metabolic insights *(Top-down)*
  3. GxE → Biomarkers → Outcomes of interest via mediation analysis *(Birth weight → mediating biomarkers → T2D risk. Song et al. 2015)*

- **Molecular epidemiology**
  Pharmacogoneomics and G x E research
Drug development: challenges and opportunities

Early Discovery
- Phase/Validation

Preclinical
- Target identification
- Molecule development
- In vitro & in vivo experiments
- Human observational studies

Clinical
- Phase I
- Phase II
- Phase III (Approval)
- Phase IV

Post Marketing (Phase IV)

Mendelian Randomization

15 years

$1 billion
RCT

Unknown Confounders

Received recombinant SHBG treatment

Compliance

Randomization (of recombinant SHBG treatment versus placebo)

Diseases Such as T2D

Intermediate phenotype (Plasma SHBG levels)

Randomization (of alleles)

Penetrance, functionality

T2D

MR

Unknown Confounders

Randomized controlled trial (RCT)

Mendelian Randomization (MR)

Ding et al. 2007
Three major criteria when evaluating the suitability of genotype as randomization instrument

1. G is independent of U
2. G is robustly associated with X
3. G is independent of Y given X and U (i.e. G affects Y only through X)
Linearity without interactions

- We assume general linear models for the dependencies among the variables $Y$, $X$, $G$ and $U$:

  $X = \alpha_0 + \alpha_1 G + \alpha_2 U + \varepsilon_1$

  $Y = \beta_0 + \beta_1 X + \beta_2 U + \varepsilon_2$

  and $E(\varepsilon_1) = 0$, $E(\varepsilon_2) = 0$, $E(U) = 0$

  $\beta_1$ is the parameter of interest
Two-stage-least-squares

- First obtain $\hat{\alpha}_0$ and $\hat{\alpha}_1$ by linear regression of X on G alone.
- Then find the predicted values of X by
  $$\hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 g \approx \alpha_0 + \alpha_1 g$$
- Obtain the consistent estimate $\hat{\beta}_1$ by linear regression of Y on $\hat{X}$.
  $$Y = \beta_0 + \beta_1 x + \beta_2 u + \varepsilon$$
  $$= \beta_0 + \beta_1 (\alpha_0 + \alpha_1 g + \alpha_2 u + \tau) + \beta_2 u + \varepsilon$$
  $$\approx \beta_0 + \beta_1 (\hat{x} + \alpha_2 u + \tau) + \beta_2 u + \varepsilon$$
Incorporating new strategy: at both the beginning and the end of the drug development process

Randomized Controlled Trial (RCT Phase III)
- Selection Source Population Samples
- Randomization
- Target
- LDL & HMGCoAR
- HMGCoA Reductase-inhibitor
- LDL-C
- CHD
- Time
- Placebo
- LDL-C
- CHD
- Off Target

Genetic Evaluation via Mendelian Randomization
- Source Population
- Randomization allocation of alleles based on germ line mutation (NS12916)
- LDL/HMGCoAR
- HMGCaR/aa
- HMGCaR/AA
- LDL-C
- CHD
- LDC-C

- LDL-C reduced by ~ 0.07mmol/L
- CHD risk decreased 6%
  (OR=0.94, 0.90-0.98)
Confirming the targeting effect of lower LDL-C on CHD risk [modified from Ference et al.]
Distinguishing on- from off- target effect of first-in-class compounds

Source Population

Randomization

Torcetrapib

Placebo

Bp/off target

CETP

HDL, LDL & TG

-CHD

No CETP

-LDL Unchanged

Statin

LDL

T2D

CETP variant (Taq B) analogous to CETP inhibition, but no effect on BP and CHD (Liu et al Atherosclerosis 2003)
MR meta-analysis refuted the role of HDL as therapeutic targets for MI

Instrumental variable analysis estimate of the association of genetically raised HDL cholesterol and risk of myocardial infarction using LIPG Asn396Ser as an instrument

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Observational epidemiology</th>
<th>Genetically raised</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI) per 0·03 mmol/L (1 mg/dL) increase in plasma HDL cholesterol</td>
<td>Odds ratio (95% CI) per 0·03 mmol/L (1 mg/dL) increase in plasma HDL cholesterol</td>
</tr>
<tr>
<td>Atherosclerosis Risk in Communities Study</td>
<td>0·97 (0·96–0·98)</td>
<td>0·96 (0·86–1·07)</td>
</tr>
<tr>
<td>Copenhagen City Heart Study</td>
<td>0·98 (0·98–0·99)</td>
<td>1·09 (0·95–1·26)</td>
</tr>
<tr>
<td>Malmo Diet and Cancer Study, Cardiovascular Cohort</td>
<td>0·97 (0·96–0·98)</td>
<td>0·82 (0·66–1·01)</td>
</tr>
<tr>
<td>Framingham Heart Study</td>
<td>0·96 (0·94–0·98)</td>
<td>1·17 (1·00–1·37)</td>
</tr>
<tr>
<td>Health Professionals Follow-up Study</td>
<td>.</td>
<td>1·84 (0·39–8·62)</td>
</tr>
<tr>
<td>Danish Diet, Cancer, and Health Study</td>
<td>.</td>
<td>1·05 (0·79–1·41)</td>
</tr>
<tr>
<td><strong>Meta-analysis of cohort studies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 0·03 mmol/L (1 mg/dL) increase in plasma HDL cholesterol</td>
<td>0·98 (0·97–0·98)</td>
<td>1·02 (0·95–1·09)</td>
</tr>
<tr>
<td>Per 0·39 mmol/L (15 mg/dL) increase in plasma HDL cholesterol</td>
<td>0·70 (0·66–0·74)</td>
<td>1·28 (0·46–3·61)</td>
</tr>
</tbody>
</table>

Different paradigms of interactions (epistasis, epigenetic, etc.)

Mechanistic/Biological

Stochastic/probabilistic

Molecules to Man
Modified from Fajans et al, NEJM 1971

Molecules to Populations
Ding et al. NEJM 2009
Integrative genomics approach in identifying therapeutic targets

Pathway and Network Approach on GWAS data

+ Other omic data

Exome Sequencing

miRNA

Methylation

Metabolomics

GWAS results (CVD, T2D, combined CVD+T2D by ethnicities, i.e. AA, HA, and CA) (p-values/allele scores)

- MAGENTA: used to identify statistically significant SNPs associated with disease risk

- GSA-SNP: used to validate significant SNPs in GWAS

- PharmGKB: database of drug-gene interactions

- Reactome: database of biological pathways

- KEGG: database of biological pathways

Key Driver Analysis: integrates the genes from top-ranked pathways with evidence from large-scale genetic datasets from eight different tissue types to identify potential key driver (KD) genes

In-Silico Validation of Top KDS: searched for multiple mouse datasets using established pathways

Public mouse phenotype databases comparing phenotypic changes in genetically modified models with individual genes perturbed

Genes identified by CVD and T2D phenotypes using the hybrid mouse diversity panel (HMDP)
Pathways shared between CVD and T2D across all three ethnicities

**HCM**: hypertrophic cardiomyopathy
**DC**: dilated cardiomyopathy
**ARVC**: arrhythmogenic right ventricular cardiomyopathy
**Ca+**: calcium signaling pathway
**Axon**: axon guidance
**CAMs**: cell adhesion molecules
**FA**: focal adhesion
**ECM**: ECM-receptor interaction

Androgen-estrogen metabolism pathways

- **Cholesterol**
- **Δ4.5 isomerase**
- **Progesterone**
  - CYP 17*
  - CYP 17**
  - HSD3 β^2
- **17-OH-Progesterone**
  - CYP 17*
  - CYP 17**
- **Androstenedione**
  - HSD3 β^2
  - CYP 19
- **Testosterone**
  - HSD 17 β^1
  - CYP 19
- **5α-Dihydrotestosterone**
  - SRD5A^2
  - AR
  - HSD3 β^2
  - 3α, 3β-androestanediol
  - excretion
- **Estradiol**
  - HSD 17 β^1
### SNPs in sex-hormones pathway

<table>
<thead>
<tr>
<th>(i) Synthesis &amp; Transport</th>
<th>(ii) Estrogen</th>
<th>(iii) Androgen</th>
<th>(iv) Progesterone</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP11A</td>
<td>CYP1A1</td>
<td>SRD5A1/2</td>
<td>SRD5A1/2</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>CYP1B1</td>
<td>CYP3A4/5</td>
<td>AKR1C1-4/D1</td>
</tr>
<tr>
<td>CYP19A1</td>
<td>COMT</td>
<td>UGT2B15/17</td>
<td>HSD3B2</td>
</tr>
<tr>
<td>HSD3B1/B2</td>
<td>SULT1A1/1E1/2A1</td>
<td>HSD3A</td>
<td>PGR</td>
</tr>
<tr>
<td>HSD17B1/B2/B3</td>
<td>UGT1A1</td>
<td>HSD3B1/B2</td>
<td></td>
</tr>
<tr>
<td>SHBG</td>
<td>ERα</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ERβ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

31 Genes (1,388 SNPs)
Sex hormone genes with significant SNPs in African-American women

<table>
<thead>
<tr>
<th>(i) Synthesis &amp; Transport</th>
<th>(ii) Estrogen</th>
<th>(iii) Androgen</th>
<th>(iv) Progesterone</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP11A</td>
<td>CYP1A1</td>
<td>SRD5A1/2</td>
<td>SRD5A1/2</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>CYP1B1</td>
<td>CYP3A4/5</td>
<td>AKR1C1-4/D1</td>
</tr>
<tr>
<td>CYP19A1</td>
<td>COMT</td>
<td>UGT2B15/17</td>
<td>HSD3B2</td>
</tr>
<tr>
<td>HSD3B1/B2</td>
<td>SULT1A1/1E1/2A1</td>
<td>HSD3A</td>
<td>PGR</td>
</tr>
<tr>
<td>HSD17B1/B2/B3</td>
<td>UGT1A1</td>
<td>HSD3B1/B2</td>
<td></td>
</tr>
<tr>
<td>SHBG</td>
<td>ERα</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ERβ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

: FDR q-values <0.05
Sex hormone genes with significant SNP-enrichment score in African-American women

<table>
<thead>
<tr>
<th>(i) Synthesis &amp; Transport</th>
<th>(ii) Estrogen</th>
<th>(iii) Androgen</th>
<th>(iv) Progesterone</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP11A</td>
<td>CYP1A1</td>
<td>SRD5A1/2</td>
<td>SRD5A1/2</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>CYP1B1</td>
<td>CYP3A4/5</td>
<td>AKR1C1-4/D1</td>
</tr>
<tr>
<td>CYP19A1</td>
<td>COMT</td>
<td>UGT2B15/17</td>
<td>HSD3B2</td>
</tr>
<tr>
<td>HSD3B1/B2</td>
<td>SULT1A1/1E1/2A1</td>
<td>HSD3A</td>
<td>PGR</td>
</tr>
<tr>
<td>HSD17B1/B2/B3</td>
<td>UGT1A1</td>
<td>HSD3B1/B2</td>
<td></td>
</tr>
<tr>
<td>SHBG</td>
<td>ERα</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ERβ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

: P-values <0.05
# Sex hormone genes with significant SNPs in Hispanic women

<table>
<thead>
<tr>
<th>(i) Synthesis &amp; Transport</th>
<th>(ii) Estrogen</th>
<th>(iii) Androgen</th>
<th>(iv) Progesterone</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP11A</td>
<td>CYP1A1</td>
<td>SRD5A1/2</td>
<td>SRD5A1/2</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>CYP1B1</td>
<td>CYP3A4/5</td>
<td>AKR1C1-4/D1</td>
</tr>
<tr>
<td>CYP19A1</td>
<td>COMT</td>
<td>UGT2B15/17</td>
<td>HSD3B2</td>
</tr>
<tr>
<td>HSD3B1/B2</td>
<td>SULT1/2/3</td>
<td>HSD3A</td>
<td>PGR</td>
</tr>
<tr>
<td>HSD17B1/B2/B3</td>
<td>UGT1A1</td>
<td>HSD3B1/B2</td>
<td></td>
</tr>
<tr>
<td>SHBG</td>
<td>ERα</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ERβ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

: FDR $q$-values <0.05
Sex hormone genes with significant SNP-enrichment score in Hispanic women

<table>
<thead>
<tr>
<th>(i) Synthesis &amp; Transport</th>
<th>(ii) Estrogen</th>
<th>(iii) Androgen</th>
<th>(iv) Progesterone</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP11A</td>
<td>CYP1A1</td>
<td>SRD5A1/2</td>
<td>SRD5A1/2</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>CYP1B1</td>
<td>CYP3A4/5</td>
<td>AKR1C1-4/D1</td>
</tr>
<tr>
<td>CYP19A1</td>
<td>COMT</td>
<td>UGT2B15/17</td>
<td>HSD3B2</td>
</tr>
<tr>
<td>HSD3B1/B2</td>
<td>SULT 1A1/1E1/2A1</td>
<td>HSD3A</td>
<td>PGR</td>
</tr>
<tr>
<td>HSD17B1/B2/B3</td>
<td>UGT1A1</td>
<td>HSD3B1/B2</td>
<td></td>
</tr>
<tr>
<td>SHBG</td>
<td>ERα</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ERβ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

: P-values <0.05
**SHBG SNPs, serum levels, & T2D risk**

**Mendelian Randomization:**

- **Blacks:** OR = 0.21 (95%CI = 0.14-0.30)
- **Hispanics:** OR = 0.27 (95%CI = 0.16-0.46)
- **Asians:** OR = 0.34 (95%CI = 0.22-0.53)

$p_{\text{trend}} = 0.0004$
SHBG and T2D risk (by race)
Ding et al. NEJM 2009, Chen et al. WHI, submitted

* p-for-heterogeneity = 0.82

2. Chacko et al. Magnesium supplementation, metabolic and inflammatory markers, and global genomic and proteomic profiling: a randomized, double-blind, controlled, crossover trial in overweight individuals. AJCN 2011

Research questions

1) How are genetic variations in ion channels related to Mg homeostasis and glucose metabolism associated with risk of T2D?

2) What is the effect of on T2D risk due to Mg-related genetic variants among individuals with low/high Mg intake?

3) Develop “top-down” intervention / perturbation to the system (high vs. low Mg status/different genotypes vs. phenotypes)
Effects of Mg on Gene Expression

C1q and tumor necrosis factor-related protein-9

Pro-platelet basic protein ligand

C1q and tumor necrosis factor
Effects of Mg supplementation on Insulin

Change: -2.1 mcU/mL after Mg vs. after placebo
Effects of Mg on metabolites profiling

- 2311 Da
  p = 0.014
  AUC = 0.90

- 3803 Da
  p = 0.004
  AUC = 0.90

- 4537 Da
  p = 0.0093
  AUC = 0.90

- 19.26 kDa
  p = 0.008
  AUC = 0.95

- 10.74 kDa
  p = 0.006
  AUC = 0.94

- 22.0 kDa
  p = 0.0054
  AUC = 0.95
Reconstruct a pathogenetic network linking reproduction, immunity, and metabolism in diabetes development.
Birth weight and diabetes in WHI
Song et al. Diabetologia 2015

Low Birth Weight

- Leptin and Leptin Receptor
- Sex Steroids and SHBG
- Chronic Inflammation
- Endothelial Dysfunction
- Cellular Aging
- Blood Pressure

Insulin Resistance and β-cell Dysfunction

Type 2 Diabetes

High blood pressure
Insulin resistance
Endothelial dysfunction
Low SHBG
STATISTICAL ANALYSIS – Mediation Analysis

**Natural Direct Effect**: Set \( M \) to \( M_{X=0} \), compare \( Y_{X=1} \) with \( Y_{X=0} \)

**Natural Indirect Effect**: Set \( X \) to \( X=1 \), compare \( Y_{M_{X=1}} \) with \( Y_{M_{X=0}} \)

**Models**

1) \( \text{logit} [P (Y = 1| x, m, c)] = \theta_0 + \theta_1 x + \theta_2 m + \theta_3 c \)

2) \( \mathbb{E} [M | x, c] = \beta_0 + \beta_1 x + \beta_2 c \) (*Weighted to account for case-control design)

**Natural Direct Effect**: \( OR_{\text{NDE}} = \exp (\theta_1) \)

**Natural Indirect Effect**: \( OR_{\text{NIE}} = \exp (\theta_2 \beta_1) \)

**Proportion Mediated (on RD scale)**: \( \frac{OR_{\text{NDE}} \times (OR_{\text{NIE}} - 1)}{OR_{\text{NDE}} \times OR_{\text{NIE}} - 1} \)

**95% CI**: Bootstrapping
Opportunities and challenges

- Improved power in measures of exposures and outcomes
- Identification of key drivers as indicators of interactions
- Enhanced molecular insights in pathogenesis
- Short-term perturbation possible
- Difficulties in obtaining/storing/assaying biomaterials (more errors?)
- Difficulties integrating models at various levels
- Rapid evolution of technologies/assays
- Interpretation (Does the major markers really capture the biology? And for what timeframe?)
Acknowledgements:

Particularly my former students and fellows Drs. Eric Ding, Yiqing Song, Sean Hsu, James Sul, Yan Song, Yuko You, Sara Chacko, Chun Chao, Atsushi Goto, Brian Chen, Michelle Cho, Cathy Lee, Katie Brennan, and Katie Chan.

Faculty and staff at Harvard, UCLA, and Brown (Sam Dudley, Chuck Eaton, Hank Wu, Yen Huang, Xin Luo, Geetha G, Haiyan Xu, and Isabel Zhang), as well as participants and colleagues at WHI.

Funding Supports from AHA, NIH, CDC, BWF, & CA.