Genetic and epigenetic effects on allergy related diseases and traits

Liming Liang, PhD

Associate Professor of Statistical Genetics
Department of Epidemiology
Department of Biostatistics
Harvard T.H. Chan School of Public Health

Institute of Medicine
Committee on Food Allergies: Global Burden, Causes, Treatment, Prevention, and Public Policy
Workshop
August 31-September 1, 2015
Washington, DC
Genetic components in food allergy (FA) or allergy diseases and traits

- Few genetic variants robustly associated to FA
  - LOF mutation in Filaggrin gene (FLG) for peanut allergy
- First GWAS on FA outcomes identified HLA-DR -DQ regions at 6q21.32 related to peanut allergy (PA)
  - Hong et al. 2015 *Nature Communication*
- Asthma, related traits & treatment response
  - 48 GWAS studies, 141 reported genes*
  - Effect overlapping asthma, atopic dermatitis and psoriasis (Weidinger et al *Hum Mol Genet*, 2013)
- Atopic dermatitis
  - 5 GWAS studies, 38 reported genes*
- IgE
  - 7 GWAS studies, 40 reported genes*
- Self-reported allergy
  - 1 GWAS studies, 21 reported genes*

* [www.genome.gov/gwastudies](http://www.genome.gov/gwastudies) as of 2015.8.28
### Table 1 | The top loci associated with FA and the three most common types of FA in 2,197 discovery samples of European ancestry.

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Position</th>
<th>Nearest gene</th>
<th>Allele</th>
<th>MAF</th>
<th>P for SNP-phenotype association</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>FA</td>
<td>PA</td>
<td>Egg allergy</td>
<td>Milk allergy</td>
</tr>
<tr>
<td>rs12121623</td>
<td>1</td>
<td>54931396</td>
<td>SSBP3</td>
<td>ACOT11</td>
<td>G/T</td>
<td>0.13</td>
</tr>
<tr>
<td>rs1318710</td>
<td>4</td>
<td>101436201</td>
<td>EMCN</td>
<td></td>
<td>A/G</td>
<td>0.10</td>
</tr>
<tr>
<td>rs777717</td>
<td>2</td>
<td>195954734</td>
<td>LOC645314</td>
<td>SLC39A10</td>
<td>C/T</td>
<td>0.08</td>
</tr>
<tr>
<td>rs10994607</td>
<td>10</td>
<td>62760742</td>
<td>RHOBTB1</td>
<td></td>
<td>C/T</td>
<td>0.04</td>
</tr>
<tr>
<td>rs6942407</td>
<td>7</td>
<td>86861313</td>
<td>LOC100289677</td>
<td>TP53TG1</td>
<td>G/A</td>
<td>0.21</td>
</tr>
<tr>
<td>rs9275596</td>
<td>6</td>
<td>32681631</td>
<td>HLA-DQB1</td>
<td>HLA-DQA2</td>
<td>T/C</td>
<td>0.35</td>
</tr>
<tr>
<td>rs7192</td>
<td>6</td>
<td>32411646</td>
<td>HLA-DRA</td>
<td></td>
<td>G/T</td>
<td>0.39</td>
</tr>
<tr>
<td>rs862942</td>
<td>14</td>
<td>26492223</td>
<td>STXB6</td>
<td>NOVA1</td>
<td>G/T</td>
<td>0.07</td>
</tr>
<tr>
<td>rs4584173</td>
<td>8</td>
<td>135336557</td>
<td>LOC100129104</td>
<td>ZFAT</td>
<td>T/C</td>
<td>0.40</td>
</tr>
<tr>
<td>rs10878354</td>
<td>12</td>
<td>66384885</td>
<td>HMGA2</td>
<td>LLPH</td>
<td>G/A</td>
<td>0.23</td>
</tr>
<tr>
<td>rs7717393</td>
<td>5</td>
<td>155753914</td>
<td>SGCD</td>
<td></td>
<td>C/G</td>
<td>0.07</td>
</tr>
<tr>
<td>rs5961136</td>
<td>23</td>
<td>54802520</td>
<td>ITIH5L</td>
<td></td>
<td>T/G</td>
<td>0.39</td>
</tr>
<tr>
<td>rs250585</td>
<td>16</td>
<td>23401076</td>
<td>COG7</td>
<td></td>
<td>G/A</td>
<td>0.17</td>
</tr>
<tr>
<td>rs16823014</td>
<td>2</td>
<td>169817713</td>
<td>ABCB11</td>
<td></td>
<td>G/A</td>
<td>0.05</td>
</tr>
<tr>
<td>rs6498482</td>
<td>16</td>
<td>13978779</td>
<td>LOC729993</td>
<td>ERCC4</td>
<td>T/C</td>
<td>0.39</td>
</tr>
<tr>
<td>rs9898058</td>
<td>17</td>
<td>47818821</td>
<td>FAM117A</td>
<td></td>
<td>C/T</td>
<td>0.15</td>
</tr>
<tr>
<td>rs17032597</td>
<td>2</td>
<td>67055115</td>
<td>LOC10028929</td>
<td>ETAA1</td>
<td>C/A</td>
<td>0.29</td>
</tr>
<tr>
<td>rs78405116</td>
<td>11</td>
<td>1892562</td>
<td>LSU1</td>
<td></td>
<td>G/A</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Interpret functional consequence by eQTL mapping and other functional data

Genetic variant

Disease

Functional data

Thousands of samples

Hundreds of samples

Expression
Methylation
Acetylation
Proteomic
Metabolomic
Histone mark
DNase
Chromatin interaction

......

Methylation might mediate genetic effect on PA

Hong et al. 2015 Nat Comm
eQTL map helps functional interpretation of asthma GWAS

ORMDL3, member of a novel class of genes that encode transmembrane proteins anchored in the endoplasmic reticulum (ER).

Expressed in many tissues, particularly liver and peripheral blood lymphocytes

eQTLs from LCL can be used to interpret functions of many disease GWAS findings

- **Childhood asthma**, Nature 2007
- **Human height**, Nature 2010
- **Body mass index**, Nature Genetics 2010
- **Waist-hip ratio**, Nature Genetics 2010
- **Osteoporosis-related traits**, PLoS Genetics 2010
- **Graves’ disease**, Nature Genetics 2011
- **Concentrations of liver enzymes in plasma**, Nature Genetics 2011
- **Pancreatic cancer**, Nature Genetics 2012
- **Basal cell skin carcinoma**, Human Genetics 2012
- **Type-2 diabetes**, Diabetes 2012
- **Circulating resistin levels**, Human Molecular Genetics 2012
- **Esophageal squamous cell carcinoma**, Nature Genetics 2012
- **Human red blood cell**, Nature 2012
- **Extreme levels of anthropometric traits**, Nature Genetics 2013
PA genetic variants might regulate mRNA expression in whole blood

<table>
<thead>
<tr>
<th>MRC_Asthma mRNA Expression (Affymetrix)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe names</td>
<td>Gene names</td>
</tr>
<tr>
<td>209823_x_at</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>212998_x_at</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>208306_x_at</td>
<td>HLA-DRB1</td>
</tr>
<tr>
<td>209823_x_at</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>212998_x_at</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>208306_x_at</td>
<td>HLA-DRB1</td>
</tr>
<tr>
<td>211656_x_at</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>212999_x_at</td>
<td>HLA-DQB1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MRC_Eczema mRNA Expression (Illumina)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe names</td>
<td>Gene names</td>
</tr>
<tr>
<td>GI_4504410-S</td>
<td>HLA-DRB1</td>
</tr>
<tr>
<td>GI_24797068-S</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>GI_4504410-S</td>
<td>HLA-DRB1</td>
</tr>
<tr>
<td>GI_24797068-S</td>
<td>HLA-DQB1</td>
</tr>
</tbody>
</table>

[http://eQTL.rc.fas.harvard.edu](http://eQTL.rc.fas.harvard.edu)
Epigenetic component in allergy related traits
Immunoglobulin E (IgE) as a central mediator of allergic (atopic) inflammation

- Therapies directed against IgE benefit hay fever and allergic asthma

- Genetic association studies have not yet identified novel therapeutic targets or pathways underlying IgE regulation

- Top genetic loci identified from large scale GWAS explain ~1-2% of IgE variation
<table>
<thead>
<tr>
<th></th>
<th>MRCA</th>
<th>Swansea</th>
<th>SLSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>355</td>
<td>149</td>
<td>160</td>
</tr>
<tr>
<td>Family/unrelated</td>
<td>Family</td>
<td>Unrelated</td>
<td>Family</td>
</tr>
<tr>
<td>Age (Mean, range)</td>
<td>28, 2-61</td>
<td>21, 18-30</td>
<td>29, 5-79</td>
</tr>
<tr>
<td>% Female</td>
<td>172 (48.5%)</td>
<td>72 (48.3%)</td>
<td>80 (50.0%)</td>
</tr>
<tr>
<td>N (%) Asthmatic</td>
<td>175 (49.3%)</td>
<td>34 (22.8%)</td>
<td>69 (43.1%)</td>
</tr>
<tr>
<td>N (%) Smokers</td>
<td>45 (12.7%)</td>
<td>33 (22.1%)</td>
<td>28 (17.5%)</td>
</tr>
<tr>
<td>Eosinophil count</td>
<td>0.41 +/- 0.38</td>
<td>0.25 +/- 0.21</td>
<td>0.24 +/- 0.21</td>
</tr>
<tr>
<td>Geometric Mean Serum IgE (Range) IU/L</td>
<td>320, 1-4999</td>
<td>663, 0-18800</td>
<td>412, 2-7653</td>
</tr>
<tr>
<td>Methylation platform</td>
<td>Illumina 27K</td>
<td>Illumina 450k</td>
<td>Illumina 450K</td>
</tr>
<tr>
<td>Genotype</td>
<td>Illumina 100K + 300K, Imputation</td>
<td>Illumina 600k</td>
<td>N/A</td>
</tr>
<tr>
<td>Gene expression</td>
<td>Affy U133+2, exon, RNA-seq from LCL</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* MRCA: British MRC panel for asthma
* Swansea: Student of Swansea university, UK
* SLSJ: Familial asthma collection in the Saguenay–Lac-Saint-Jean region

Cell type specific methylation pattern is a challenge for epigenome-wide association study.

http://en.wikipedia.org/

Houseman et al., 2012
Novel methodology needed for high dimensional genomic data mediating genetic effect on phenotypic variation

- Expression
- Methylation
- Metabolomics

- Individual Loci
- Networks

- Clinical Phenotypes
- Cells and Tissues

- Variance Components
  - Environment
  - Polymorphism
  - Inheritance
Epigenome-wide association scan with IgE

CpG sites associated with IgE (stratified by EOS counts)

Blue = Subjects with eosinophil (EOS) counts > median

Is eosinophil (EOS) a confounder or source of true association signal?

Mediation Analysis* as a descriptive presentation of EOS in between IL4 methylation and IgE association

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>p_value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct effect</td>
<td>-17.6</td>
<td>.000001250</td>
<td>(-24.73, -10.49)</td>
</tr>
<tr>
<td>Indirect effect via EOS</td>
<td>-13.1</td>
<td>.000000005</td>
<td>(-17.45, -8.71)</td>
</tr>
<tr>
<td>Total Effect</td>
<td>-30.7</td>
<td>5.5274E-21</td>
<td>(-37.09, -24.29)</td>
</tr>
</tbody>
</table>

Mendelian Randomization for causal effect of IL4 methylation on IgE gave pvalue 0.00068

* Linda Valeri & Tyler VanderWeele, Psychological Methods 2013
Methylation-IgE association within eosinophil

Liang et al, Nature 2015
Top IgE related genes were activated in EOS

Liang et al., Nature 2015
What we have learned ...

- 36 CpG loci reported, replicated and confirmed for association between IgE and methylation in EOS
  - Top 3 loci explain 13% IgE variation
  - Familial variance (H^2) due to genetics was 40% and methylation was 36%
  - Related genes encode eosinophil products, related to phospholipid inflammatory mediators, specific transcription factors, and mitochondrial proteins

- Top genes are potential drug targets for allergic diseases

- **Methylation profile** at these top loci may be useful to classify patients with EOS IgE mediated allergic inflammation and personalize their therapy

Integrated omics studies
-- current research opportunity and challenge

Dietary intake/treatment

GWAS

Gene expression

Metabolomics

Methylation
Omic-epidemiology would be a useful systematic approach to investigate FA
Future research for FA

- More genetic/epigenetic genome-wide association studies for FA outcomes are needed
- Prenatal or early life epigenetic profile related to FA using prospective birth cohorts
  - Profile/Network approach in addition to individual gene approach
- Patient stratification based on genetic and epigenetic variants
  - Risk, subtype and degree of FA
  - Treatment efficacy
- Cell type specific epigenetic/expression effect on FA
- Measure or predict methylation in target tissue related to FA
  - Ma et al Nucleic Acids Research 2014
Acknowledgements

- Harvard SPH
  - Murray Mittleman
  - Andrea Baccarelli
  - Joel Schwartz
  - Elissa Wilker
  - Oliver Hofmann
  - Baoshan Ma
  - Ga Liao

- Imperial College
  - William Cookson
  - Miriam Moffatt
  - Saffron Willis-Owen
  - Kenny Wong
  - Aristeia Binia

- McGill University and Génome Québec Innovation Centre
  - Mark Lathrop
  - Tomi Pastinen
  - Elin Grundberg
  - Stephan Busche

- Swansea University
  - Julian M. Hopkin
  - Gwennyth Davies

- Université du Québec à Chicoutimi
  - Catherine Laprise

- University of Colorado
  - David Schwartz
  - Ivana Yang

- Ontario Institute for Cancer Research
  - Thomas Hudson

- Jewish General Hospital and Lady Davis Research Institute
  - Marie Hudson

- SciLifeLab, Uppsala University
  - Lars Rönnblom
Gene expression quantitative trait loci (eQTL) mapping

a Cis (local)

b Trans (distal)

Cheung et al. Nat Review Genet 2009
Heritability and shared environmental components
Variance component model for environmental and genetic/epigenetic effect

- $\text{Methy\_parent} = \text{Sex} + \text{Age} + \text{Ep} + \text{G} + \text{R}$
- $\text{Methy\_children} = \text{Sex} + \text{Age} + \text{Ec} + \text{G} + \text{R}$
  - $(\text{Ep}, \text{Ec})^T \sim \text{MVN}(0, \Sigma)$, $V(\text{Ep}) = \text{Vep}$, $V(\text{Ec}) = \text{Vec}$ and $\text{Cov}(\text{Ep},\text{Ec}) = \text{Cpc}$

- Heritability in children = $\frac{\text{VG}}{\text{Vec} + \text{VG} + \text{VR}}$

- Genetic/epigenetic component in Parent-Child correlation = $0.5*\text{VG} / (\text{Cpc} + 0.5*\text{VG})$. 
Genetic contribution after adjusting by Sex and Age
Genetic loci affecting methylation
Methylation QTL mapping

• To find genetic variants associated to methylation level

• Take methylation level at each gene as a quantitative trait

• Test for association with each SNPs on the genome
  – 27,000 x 400,000 association tests using genotyping array
  – 27,000 x 8,000,000 association tests using imputed 1000Genomes SNPs
Association of meQTLs to SNPs (meSNPs)

- **2079** genome-wide significant meSNPs in *cis* (within 1Mb of the meQTL) and **522** *trans* meSNPs (44 >1Mb and 478 on different chromosomes)

- The peak meSNP explained an average of 22.3% total residual variance (range 5.6%-80.6%)
Shared genetic regulators for mRNA expression and DNA methylation

• Gene expression in LCL available from 400 samples including those with methylation data

• 666 peak meSNPs (out of 2618 genome-wide significant) were in linkage disequilibrium ($R^2 > 0.5$) with a significant peak eSNP (327 in LD with $R^2 > 0.8$), suggested that up to 25% of the meSNPs were also affecting transcription
Methylation loci affecting clinical phenotypes
Difference between GWAS and EWAS

• Epigenetic markers
  – DNA methylation marker is most stable and amenable to high throughput analysis
  – **Infinium HumanMethylation450 BeadChip**
  – Beta value and M value

• Hard to establish causal effects
  – Disease may cause methylation changes
  – Longitudinal studies/statistical models to establish causal effects

• DNA methylation varies between tissues/cell types
  – Most EWAS use tissues
  – Cell composition changes in response to disease or exposure
  – Cell mixture is a potential confounder
Network association analysis
Co-methylation network

• Correlation of methylation may indicate genes share similar pathway
• Analyze pathway genes together may increase power to detect association
• Construction of networks
  – Based on marginal correlation of two genes
    • could mean direct or indirect regulation
  – Based on conditional correlation of two genes
    • more likely to be direct regulation
Clustering of genes into network modules
Identify network associated to complex traits

• Test the overall effect from a network module
  – Use module eigen-gene (e.g. first PC) to represent a network module

• Test for enrichment of associated loci in a network module
  – Could be more powerful when many loci have small to modest effect size
<table>
<thead>
<tr>
<th>MEdarkorange</th>
<th>MEorange</th>
<th>MEtan</th>
<th>MEwhite</th>
<th>MEpurple</th>
<th>MElightgreen</th>
<th>MEyellow</th>
<th>MEblack</th>
<th>METurquoise</th>
<th>MEcyan</th>
<th>MEDarkgrey</th>
<th>MESTeelBlue</th>
<th>MECyanlight</th>
<th>MEmidnightblue</th>
<th>MERed</th>
<th>MERoyalblue</th>
<th>MESkyblue</th>
<th>MEPink</th>
<th>MEblue</th>
<th>MElightblue</th>
<th>MEGreenyellow</th>
<th>MEdarkgreen</th>
<th>MGreen</th>
<th>MEMagenta</th>
<th>MElightyellow</th>
<th>MEdarkturquoise</th>
<th>MEdarkred</th>
<th>MEbrown</th>
<th>MESalmon</th>
<th>MEgrey60</th>
<th>MEsaddlebrown</th>
<th>MEgrey</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.106(1)</td>
<td>0.057(1)</td>
<td>-0.045(2)</td>
<td>0.072(1)</td>
<td>0.087(1)</td>
<td>-0.01(1)</td>
<td>-0.038(2)</td>
<td>0.18(1)</td>
<td>-0.006(1)</td>
<td>0.021(1)</td>
<td>0.042(1)</td>
<td>0.01(1)</td>
<td>0.031(1)</td>
<td>0.026(2)</td>
<td>0.012(1)</td>
<td>0.051(1)</td>
<td>0.021(1)</td>
<td>0.032(1)</td>
<td>0.006(1)</td>
<td>0.021(1)</td>
<td>0.031(1)</td>
<td>0.028(1)</td>
<td>0.031(1)</td>
<td>0.021(1)</td>
<td>0.033(1)</td>
<td>0.031(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.034(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>-0.004(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.63(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.21(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td>0.021(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td>0.001(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Module–trait relationships: Effect size (P-value), colored by $-\log_{10}(P-value)$

- **Sex**: 0.28 | **Age**: 0.28 | **LNIGE**: 0.28 | **DDAST**: 0.28 | **Par**: 0.28 | **EOS**: 0.28 | **PLA**: 0.28 | **HC**: 0.28 | **RC**: 0.28 | **MCV**: 0.28 | **MCHGBP**: 0.28 | **NEU**: 0.28 | **LYM**: 0.28 | **MON**: 0.28 | **BAS**: 0.28 | **Season**: 0.28 | **AgeSex**: 0.28 | **SexTotal**: 0.28 | **AgeTotal**: 0.28 | **Sex**: 0.28 | **Age**: 0.28 | **LNIGE**: 0.28 | **DDAST**: 0.28 | **Par**: 0.28 | **EOS**: 0.28 | **PLA**: 0.28 | **HC**: 0.28 | **RC**: 0.28 | **MCV**: 0.28 | **MCHGBP**: 0.28 | **NEU**: 0.28 | **LYM**: 0.28 | **MON**: 0.28 | **BAS**: 0.28 | **Season**: 0.28 | **AgeSex**: 0.28 | **SexTotal**: 0.28 | **AgeTotal**: 0.28 | **Sex**: 0.28 | **Age**: 0.28 | **LNIGE**: 0.28 | **DDAST**: 0.28 | **Par**: 0.28 | **EOS**: 0.28 | **PLA**: 0.28 | **HC**: 0.28 | **RC**: 0.28 | **MCV**: 0.28 | **MCHGBP**: 0.28 | **NEU**: 0.28 | **LYM**: 0.28 | **MON**: 0.28 | **BAS**: 0.28 | **Season**: 0.28 | **AgeSex**: 0.28 | **SexTotal**: 0.28 | **AgeTotal**: 0.28
Identify important hub genes for network modules
Network analysis is also useful for metabolite data

Wang 2011
Metabolite network modules identified based on NHS and HPFS samples

Liu et al. (unpublished data)
Metabolite network models associated with baseline BMI

<table>
<thead>
<tr>
<th>Metabolite Modules</th>
<th>Effect Size</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEturquoise</td>
<td>1.275262056</td>
<td>7.44E-16</td>
</tr>
<tr>
<td>MEMagenta</td>
<td>1.242871325</td>
<td>2.51E-12</td>
</tr>
<tr>
<td>MEyellow</td>
<td>1.121665251</td>
<td>1.75E-05</td>
</tr>
<tr>
<td>MEgreen</td>
<td>0.926034794</td>
<td>0.000147761</td>
</tr>
<tr>
<td>MEbrown</td>
<td>0.511169155</td>
<td>0.002076766</td>
</tr>
<tr>
<td>MERed</td>
<td>0.372516746</td>
<td>0.021246309</td>
</tr>
<tr>
<td>MEblue</td>
<td>-0.369653664</td>
<td>0.061785336</td>
</tr>
<tr>
<td>MEgrey</td>
<td>0.192287388</td>
<td>0.231193246</td>
</tr>
<tr>
<td>MEblack</td>
<td>0.18776424</td>
<td>0.265299556</td>
</tr>
<tr>
<td>MEpink</td>
<td>-0.158870904</td>
<td>0.308250161</td>
</tr>
</tbody>
</table>

Liu et al. (unpublished data)
Epigenetic profiling
Methylation is associated with age and may reflect aging process.
Tumor tissues show a faster methylation aging rate
Δage = Methylation Age – chronological Age is associated all-cause mortality
Δage = Methylation Age – chronological Age
is a heritable trait
Cross tissue methylation prediction and calibration
Why predict methylation across tissues

- DNA methylation were correlated across tissues (Byun et al. Hum Mol Genet 2009, Caliskan et al. Hum Mol Genet 2011)

- Target tissues are hard to obtained for large scale epidemiological studies

- Can we use one tissue (e.g. blood) as surrogate to study the methylation in other tissues (e.g. atrial tissue)?
Blood, Artery and Atrium

• A pilot study for atrial fibrillation
  – Patients undergo coronary artery bypass graft at Beth Israel Deaconess Medical Center
  – We collected 3 tissues on the same individual
    • Atrium (n=17)
    • Internal mammary artery (n=16)
    • Peripheral leukocytes (n=20)
  – After QC, 14 samples have all three tissues
  – Illumina 450K methylation array
    • Normalized using the Tost pipeline

Ma et al Nucleic Acids Research 2014
Blood and Atrium for 14 subjects

Red: Atrium vs. Blood
Blue: Atrium vs. Predicted Atrium (linear model)

Ma et al Nucleic Acids Research 2014
Prediction with high accuracy

Range of true values

**Red**: Artery vs. Blood (R²=0.817)
**Green**: Linear model (R²=0.981)
**Purple**: SVM (R²=0.985)

Ma et al. Nucleic Acids Research 2014
Clustering using atrium methylation

Ma et al. Nucleic Acids Research 2014
Clustering using blood methylation

Ma et al. Nucleic Acids Research 2014
Clustering using blood methylation after calibration

Ma et al. Nucleic Acids Research 2014
What we have learned so far...

- Methylation can be an important factor underlying complex traits

- Whole genome association scan of methylation loci is feasible and powerful
  - Microarray based measurement needs careful preprocessing and modeling
  - Confounding due to cell heterogeneity
    - Adjusting cell proportions/counts
    - PCA and SVA
  - Multiple testing (FDR is preferred than Bonferroni due to correlation between CpGs)

- Interesting associations have been identified
  - DNA variant -> Methylation
  - Methylation <-> Complex traits / diseases
  - DNA variant -> Methylation -> Disease (mediation effect)
  - Individual loci and network module

- High throughput technologies are more and more affordable
  - E.g. Illumina 450K array is one of current popular choices
  - Whole genome sequencing of bisulfite converted DNA could be the next
Future methodology directions

• Integrative genomics from different sources
  – Multi-type functional annotation for association analysis
  – Exposure -> High dimensional mediators -> Outcome
  – Exposure -> High dimensional confounders -> Outcome

• Risk prediction
  – Genetic risk
  – Risk based on Methylation/Expression/Metabolites/other omics

• Personalized treatment
  – Gene x Treatment interaction

• Drug target discovery
  – Use GWAS/EWAS to narrow down to targets with high probability to pass clinical trial