Epigenetics & CVD Risk Prediction

Myriam Fornage, PhD

University of Texas Health Science Center at Houston
The Epigenome in Health and Disease

- **Epigenome**: Set of stable alterations to the DNA and histone proteins that alter gene expression without change in the DNA sequence.

- The epigenome as a link between the genome, the environment, and phenotypes of health & disease:
  - May mediate the long-term impact of environmental exposures on disease risk.
DNA Methylation

• is the most studied epigenetic mark
  • covalent binding of a methyl group to the 5’ carbon of cytosines occurring mainly at CpG dinucleotide sequences
  • ~30 millions CpG across the human genome and 70% of them are methylated

• plays a critical role in the regulation of gene expression
  • modulates expression of genetic information by modifying DNA accessibility to the transcriptional machinery

• is dynamic, tissue- or cell-specific, and can be influenced by, both, genes and the environment

• can be measured reliably, quantitatively, in a cost-effective manner via DNAm array

Pre-requisites for Risk Score Application

Discovery
Validation
Application
Epigenome-Wide DNA Methylation Studies (EWAS)

• **Goal**: The *integration* of DNA methylation data into our population-based research with the goal of *discovering relationships* between variation in DNA methylation with environmental exposures, genetic variation, and disease risk and disease-related traits

• Genome-wide association studies of DNAm and environmental exposures
  • DNA methylation signatures of cigarette smoking, alcohol intake, dietary vitamins intake, air pollution, dietary patterns

• Genome-wide association studies of DNAm and disease and disease-related traits
  • EWAS of blood pressure, circulating markers of inflammation, depressive symptoms, cognitive function, brain MRI traits

• GWAS of DNAm levels: Mapping of *cis* and *trans* meQTL
EWAS vs. GWAS

• Genetic factors are fixed throughout the lifetime
  • No assumption about temporality of effects
  • No issue with time of sample collection
• Genetic factors can be assumed to be randomly assigned with respect to traits
  • Population stratification is identifiable and can be corrected
• Pattern of correlation (LD) well defined in genetic data

• DNA methylation is a dynamic process
  • Collection timing matter: Optimal timing of the measurement relative to outcome of interest?
  • Issues of reverse causation need to be carefully assessed
• Confounding is often present
  • Cellular heterogeneity
  • Measured and unmeasured environmental factors
• Inter-correlation of CpGs not well-defined or exploited
• DNAm is the dependent variable in EWAS studies
Study Design and Methodologies: Blood Pressure EWAS

Two-stage EWAS

\[ \beta \sim BP + age + sex + smoking + BMI + blood cells + PC + technical covariates + family structure \]

Discovery Meta-Analysis
N=9,828
ARIC, CHS, FHS, GOLDN, LBC1936, NAS, RS-III, TwinsUK

Replication Meta-Analysis
N=7,182
Amish, ARIC, MESA, RS-III, SYS, WHI-BAA23, WHI-EMPC

Functional annotation
Gene Set Enrichment Analysis
\( e \text{FORGE} \)
percent variance explained
heritability
methylation QTLs

Gene expression
\( \pm 1\text{Mb} \)
FHS and RS
N=2,946
D\( NA_m \)
BP

FHS
N=3,679

Causal relationships
Bidirectional Mendelian randomization
Inverse variance weighted with tests for pleiotropy
ARIC, FHS, RS, and WHI-EMPC EAs
N=4,513

DNA\( m \) \( \rightarrow \) BP
BP \( \rightarrow \) DNA\( m \)

Two-step Mendelian randomization
Step One:
Assess meQTL associations in GTEx whole blood eQTL data

Gene expression
DNA\( m \) \( \rightarrow \) BP

Step Two:
Assess top GTEx whole blood eQTL association in ICBP
**EWAS of Blood Pressure – CHARGE Consortium**

**Discovery sample:** 9,828 middle-aged to older adults (EA, N = 6650; AA, N = 3178) from 9 cohorts

**Replication Sample:** 7,182 middle-aged to older adults (EA, N = 4695; AA, N = 1458; HIS, N = 1029) from 7 cohorts

<table>
<thead>
<tr>
<th>Meta-Analysis</th>
<th>tests, n</th>
<th>p-value threshold</th>
<th>SBP probes, n</th>
<th>DBP probes, n</th>
<th>total probes, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>&gt; 450,000</td>
<td>1E-7</td>
<td>25</td>
<td>9</td>
<td>31</td>
</tr>
<tr>
<td>Replication</td>
<td>31</td>
<td>0.0016</td>
<td>9</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Overall</td>
<td>&gt; 450,000</td>
<td>1E-7</td>
<td>102</td>
<td>56</td>
<td>126</td>
</tr>
</tbody>
</table>
EWAS of BP: Lessons Learned

• DNA methylation explains more of BP variance than genetic loci
  • DNAm score based on 13 replicated CpGs explained ~1.5% - 2% variance in BP
  • Genetic risk score based on known BP SNPs (N=261) explained between 0.003% and 0.1%

• Similar findings are observed for other traits

McCartney at al. 2018; PMID: 30257690
EWAS of BP: Lessons Learned

• Many identified BP-associated CpGs are heritable
  • replicated probes average $h^2 = 30$-60%; epigenome-wide average $h^2 = 12$

• meQTLs could be identified in 10 of the 13 BP-associated CpGs
  • 9 of 13 CpGs showed substantial evidence for meQTLs in EA and AA ancestries, with evidence for weak meQTLs at one additional CpG site in each ancestry
  • Seven of the 10 meQTLs showed nominal association with BP

meQTL mapping in in 4,036 EAs and 2,595 AAs and confirmed in an independent dataset (ARIES)

P-value of association of SNPs with DNAm relative to the CpG location (±25 kb)
EWAS of BP: Lessons Learned

• DNAm influences BP but also BP influences DNAm levels
  • Evidence through bidirectional Mendelian randomization
  • Instrumental variables:
    • meQTL
    • BP-associated SNPs

Instrumental Variables:
3-10 cis-meQTLs ($r^2 < 0.2$)

DNAm $\rightarrow$ BP

Forward Causality
- cg08035323

Reverse Causality
- cg00533891
- cg00574958
- cg02711608
- cg22304262
EWAS of BP: Lessons Learned

- Integration of other omics (gene expression) improves interpretability of EWAS findings

**YWHAQ Gene Expression**

- Negative association: $P = 0.04$
- Positive association: $P = 0.02$

Blood DNAm, blood gene expression, and BP measured in the same sample

**cg08035323**

(intergenic)
Assessing Functional Causality: Two-Step Mendelian Randomization

Instrumental Variables: Whole blood eQTL from GTEx

Estimates of SNP effects on gene expression from GTEx

Gene Expression

DNA \_m\ 

BP

Estimates of SNP effects on BP from ICBP GWAS

Step One

increased expression of TSPAN2

CpG-GE Q value = 8.6 \times 10^{-14}
Step One p value = 0.0074
GE-DBP Q value = 1.3 \times 10^{-16}
Step Two p value = 0.0003

decreased DNAm at cg23999170
increased diastolic BP

5 mmHg increase in diastolic BP per 0.1% decrease in DNA methylation
CpG-BP p value = 1.9 \times 10^{-13}
Application of DNAm to (Risk) Prediction

• How well does DNAm predict cardiometabolic traits?
  • DNAm scores generated in the GS cohort (N=5087) and validated in LBC1936 cohort (N=895)
  • Near perfect discriminatory power for current smokers
  • Moderate discrimination of obesity, heavy drinking, and high HDL
  • Poor discrimination of high(college) education and high LDL

ROC analysis for DNAm predictors of smoking, alcohol, education, BMI, and lipid traits in the LBC1936 cohort

McCartney at al. 2018; PMID: 30257690
Association of DNAm risk scores, polygenic risk scores, and phenotypes with mortality

<table>
<thead>
<tr>
<th>Trait</th>
<th>Predictor</th>
<th>HR</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td>Phenotypic</td>
<td>0.93</td>
<td>0.82 – 1.07</td>
<td>0.362</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>1.24</td>
<td>1.08 – 1.43</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.05</td>
<td>0.92 – 1.21</td>
<td>0.479</td>
</tr>
<tr>
<td>Smoking</td>
<td>Phenotypic</td>
<td>1.91</td>
<td>0.98 – 3.70</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>1.29</td>
<td>1.05 – 1.57</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.09</td>
<td>0.86 – 1.37</td>
<td>0.801</td>
</tr>
<tr>
<td>Education</td>
<td>Phenotypic</td>
<td>0.9</td>
<td>0.78 – 1.05</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>0.81</td>
<td>0.71 – 0.93</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>0.96</td>
<td>0.84 – 1.11</td>
<td>0.59</td>
</tr>
<tr>
<td>BMI</td>
<td>Phenotypic</td>
<td>1.14</td>
<td>0.99 – 1.32</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>1.01</td>
<td>0.87 – 1.17</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.10</td>
<td>0.95 – 1.28</td>
<td>0.184</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>Phenotypic</td>
<td>0.86</td>
<td>0.74 – 1.00</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>0.98</td>
<td>0.83 – 1.14</td>
<td>0.774</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.14</td>
<td>1.00 – 1.31</td>
<td>0.064</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>Phenotypic</td>
<td>0.92</td>
<td>0.77 – 1.09</td>
<td>0.324</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>0.92</td>
<td>0.78 – 1.08</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.08</td>
<td>0.94 – 1.25</td>
<td>0.274</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>Phenotypic</td>
<td>0.9</td>
<td>0.78 – 1.05</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>1.01</td>
<td>0.86 – 1.19</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.10</td>
<td>0.95 – 1.28</td>
<td>0.181</td>
</tr>
<tr>
<td>Waist-to-hip ratio</td>
<td>Phenotypic</td>
<td>1.24</td>
<td>1.08 – 1.42</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>0.93</td>
<td>0.82 – 1.07</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.18</td>
<td>1.03 – 1.36</td>
<td>0.016</td>
</tr>
<tr>
<td>% body fat</td>
<td>Phenotypic</td>
<td>1.08</td>
<td>0.93 – 1.23</td>
<td>0.328</td>
</tr>
<tr>
<td></td>
<td>Epigenetic</td>
<td>1.18</td>
<td>1.03 – 1.36</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Genetic</td>
<td>1.18</td>
<td>1.03 – 1.36</td>
<td>0.016</td>
</tr>
</tbody>
</table>
Application of DNAm to Age Prediction

- DNAm-based age estimators
  - Age has a strong impact on genome-wide DNAm levels
  - DNAm age estimators are based on sets of CpGs selected to best estimate chronological age

- **Age acceleration**: Deviation of the DNA methylation-predicted age from the chronological age – Index of an individual’s rate of aging

- Discrepancies between a person’s DNA methylation age and chronological age may be detrimental to health
  - Association between blood DNA methylation-derived measures of accelerated aging and all-cause mortality (Marioni et al. 2015)
Application of DNA methylation (DNAm) to Age Prediction

Horvath and Raj, 2018. PMID: 29643443
Conclusions

• EWAS identifies new genomic regions influencing complex traits not previously implicated by GWAS but care must be taken in the interpretation of epigenetic associations

• DNAm scores explain a substantial proportion of phenotypic variance and are able to predict health and lifestyle factors with some success

• Data suggest a potential application of DNAm signatures as proxies for self-(un)reported phenotypes, such as smoking

• DNAm age biomarkers of aging for identifying anti-aging interventions?
  • DNAm is dynamic and tissue-specific. The predictive abilities of DNAm may depend on the characteristics of the population/tissue in which the score was derived
DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation


- CHARGE Epigenetics Working Group