Objective (i.e., why you woke up this early)

- To review (all) the steps for designing and carrying out an epigenetic study
  - Focus on environmental exposure
  - Applications to epidemiology
Step by step

Step 1
Intro to epigenetics

Step 2
Epigenetics & Environment

Step 3
Study Design

Step 4
Starting a Study

Step 5
Lab Analysis

Step 6
Data Analysis

Step 7
Wrap up
Step by step

Step 1
Intro to epigenetics
A Symphonic Example

DNA

Phenotype

Epigenetics
Epigenetics & Music Use the Same Markings
Epigenetics & Music Use the Same Markings

- pencil markings (can be erased)
- markings in ink (permanent)
Epigenetic markings

DNA methylation
Methyl marks added to certain DNA bases repress gene transcription

Histone modifications
A combination of different molecules can attach to the ‘tails’ of proteins called histones. These alter the activity of the DNA wrapped around them
Why Epigenetics in Environmental Health?

• The epigenome is environmentally sensitive
  – May provide records of past exposures and help to reconstruct risk-factor experience

• The epigenome harbors profiles potentially useful to identify at-risk individuals
  – May help to predict future risks of disease

• Epigenomic investigations might bring further mechanistic understanding
  – Challenges in human studies need to be considered
Controlled air particle exposure

N=11 participants (5M/6F)

Age: 27.2 yrs [19-48]

Target concentrations:
- Fine CAPs: 250 µg/m³
- Coarse CAPs: 200 µg/m³

Bellavia et al. JAHA 2013
Proportions of fine CAPs effects mediated by methylation

Effects mediated by **Alu hypomethylation** on Systolic and Diastolic BP

- Exposure to Fine CAPs
  - Alu Hypomethylation
    - Systolic Blood Pressure
      - Proportion of Mediation: 10%
    - Diastolic Blood Pressure
      - Proportion of Mediation: 32%

Effects mediated by **TLR4 hypomethylation** on Systolic and Diastolic BP

- Exposure to Fine CAPs
  - TLR4 Hypomethylation
    - Systolic Blood Pressure
      - Proportion of Mediation: 22%
    - Diastolic Blood Pressure
      - Proportion of Mediation: 44%

Bellavia et al. JAHA 2013
Joint effect of employment length and metal exposure on H3K4 dimethylation

Cantone et al. EHP 2011
Step by step

Step 1
Intro to epigenetics

Step 2
Epigenetics & Environment

Step 3
Study Design
Objectives of studies in epigenetic epidemiology

Michels K, Epigenetic Epidemiology Springer 2012
Chapter 3, Study Design
THE ISSUE OF REVERSE CAUSATION

Data from the COPD EWAS Study

Work by Dawn Demeo, et al.
The case of SERPINA1 and COPD

Manhattan Plot of Methylation by COPD Status (ICGN)
(Model I)

-log10 P value

Qiu et al. AJRCCM 2012
SERPINA1 methylation lower in COPD cases (?)

Pyrosequencing Validation for cg02181506

n=943, p-value (GLMM Model I)=4.3e-17

Qiu et al. AJRCCM 2012
The ideal world

- Epigenetic Inheritance Systems
- Stochastic Events
- Germ-line Genetic Variation
- Environment

Epigenome → Intermediate Phenotypes / Biomarkers → Disease

Relton and Davey-Smith Int J Epidemiol 2012
The real world

- Epigenetic Inheritance Systems
- Stochastic Events
- Germ-line Genetic Variation
- Environment

Epigenome → Intermediate Phenotypes / Biomarkers → Disease

Relton and Davey-Smith Int J Epidemiol 2012
The real world

- Epigenetic Inheritance Systems
- Stochastic Events
- Germ-line Genetic Variation
- Environment

Epigenome → Intermediate Phenotypes / Biomarkers → Disease

Relton and Davey-Smith Int J Epidemiol 2012
Reverse Causation

- In reverse causation:
  - Cause and effect are reversed
- COPD Study:
  - Study aim: determining whether methylation → COPD
  - We need to exclude that COPD → methylation
  - Need to determine temporal sequence of events
- Study design
  - Cross-sectional and case-control studies are susceptible to reverse causation
  - Longitudinal studies should be preferred in epigenetic epidemiology
Step by step

Step 1: Intro to epigenetics
Step 2: Epigenetics & Environment
Step 3: Study Design
Step 4: Starting a Study
Biospecimens

• DNA methylation
  – Any DNA sample of reasonable quality would do

• Histone Modifications
  – High amounts of sample needed
  – DNA cannot be used (histones are proteins)

• Epigenetics marks are tissue specific
  – Consider function of the gene in the type of samples available
HOW MANY EPIGENOMES?

Data from the ELEMENT/PROGRESS Mexico Birth Cohort
Epigenetic markings are **Tissue Specific**.

Potentially **each tissue or cell type** has a specific methylation profile.
Environmental causes of fetal growth
Epigenomics in the ELEMENT study (R01ES020268, Wright, Baccarelli, PIs)

• Umbilical vessels and blood are critical to maternal-fetal nutrient transfer
• Ongoing case-control study nested in the ELEMENT/PROGRESS Mexico Cohort
• Fetal growth and potential environmental determinants
Top 20 methylation sites in umbilical **blood, artery, vein** associated with birth weight

**Pilot Study**

20 small birth weight vs. 20 normal birth weight children

<table>
<thead>
<tr>
<th>Rank</th>
<th>Blood SYMBOL</th>
<th>Blood p.value</th>
<th>Artery SYMBOL</th>
<th>Artery p.value</th>
<th>Vein SYMBOL</th>
<th>Vein p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KLF15</td>
<td>0.00005</td>
<td>POLR2J2</td>
<td>0.00003</td>
<td>SFT2D1</td>
<td>0.00006</td>
</tr>
<tr>
<td>2</td>
<td>EBPL</td>
<td>0.00018</td>
<td>IER5</td>
<td>0.00009</td>
<td>SLITRK4</td>
<td>0.00007</td>
</tr>
<tr>
<td>3</td>
<td>ITPKB</td>
<td>0.00022</td>
<td>LR8</td>
<td>0.00009</td>
<td>MGAT5B</td>
<td>0.00014</td>
</tr>
<tr>
<td>4</td>
<td>TOPORS</td>
<td>0.00023</td>
<td>CFHR5</td>
<td>0.00016</td>
<td>CGRRF1</td>
<td>0.00015</td>
</tr>
<tr>
<td>5</td>
<td>ZNF365</td>
<td>0.00026</td>
<td>HTATIP2</td>
<td>0.00018</td>
<td>NOX1</td>
<td>0.00017</td>
</tr>
<tr>
<td>6</td>
<td>ACOT4</td>
<td>0.00026</td>
<td>SAMD7</td>
<td>0.00028</td>
<td>IQWD1</td>
<td>0.00020</td>
</tr>
<tr>
<td>7</td>
<td>CCNJ</td>
<td>0.00030</td>
<td>ALOX15B</td>
<td>0.00035</td>
<td>NOL9</td>
<td>0.00026</td>
</tr>
<tr>
<td>8</td>
<td>DCK</td>
<td>0.00033</td>
<td>POLD2</td>
<td>0.00036</td>
<td>SFXN1</td>
<td>0.00030</td>
</tr>
<tr>
<td>9</td>
<td>DDX27</td>
<td>0.00048</td>
<td>SPIB</td>
<td>0.00045</td>
<td>RPS15A</td>
<td>0.00031</td>
</tr>
<tr>
<td>10</td>
<td>C17orf88</td>
<td>0.00052</td>
<td>MAPK9</td>
<td>0.00050</td>
<td>CAMK2D</td>
<td>0.00042</td>
</tr>
<tr>
<td>11</td>
<td>WDR37</td>
<td>0.00057</td>
<td>CER1</td>
<td>0.00055</td>
<td>PRSS27</td>
<td>0.00044</td>
</tr>
<tr>
<td>12</td>
<td>PGPEP1</td>
<td>0.00063</td>
<td>WDFY3</td>
<td>0.00057</td>
<td>GJA3</td>
<td>0.00048</td>
</tr>
<tr>
<td>13</td>
<td>RPS15A</td>
<td>0.00063</td>
<td>LETM2</td>
<td>0.00058</td>
<td>C1orf26</td>
<td>0.00050</td>
</tr>
<tr>
<td>14</td>
<td>REST</td>
<td>0.00065</td>
<td>RGSL1</td>
<td>0.00064</td>
<td>LYPLA1</td>
<td>0.00055</td>
</tr>
<tr>
<td>15</td>
<td>C19orf22</td>
<td>0.00071</td>
<td>BSDC1</td>
<td>0.00069</td>
<td>FLJ45803</td>
<td>0.00056</td>
</tr>
<tr>
<td>16</td>
<td>VPS54</td>
<td>0.00073</td>
<td>MSR1</td>
<td>0.00074</td>
<td>PAK1</td>
<td>0.00064</td>
</tr>
<tr>
<td>17</td>
<td>PDE9A</td>
<td>0.00074</td>
<td>FOXF2</td>
<td>0.00074</td>
<td>C21orf129</td>
<td>0.00081</td>
</tr>
<tr>
<td>18</td>
<td>CLDN11</td>
<td>0.00075</td>
<td>ZNF592</td>
<td>0.00080</td>
<td>SENP6</td>
<td>0.00091</td>
</tr>
<tr>
<td>19</td>
<td>SPAG1</td>
<td>0.00076</td>
<td>PTPDC1</td>
<td>0.00083</td>
<td>TNFRSF1A</td>
<td>0.00099</td>
</tr>
<tr>
<td>20</td>
<td>ARIH1</td>
<td>0.00076</td>
<td>FLJ20920</td>
<td>0.00089</td>
<td>SLC22A6</td>
<td>0.00106</td>
</tr>
</tbody>
</table>
**Blood Counts and Methylation**
(combined-analysis of 5 studies)

<table>
<thead>
<tr>
<th></th>
<th>Alu</th>
<th>LINE-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta *</td>
<td>P-value *</td>
</tr>
<tr>
<td>White blood cells, $10^3$cell/mm$^3$</td>
<td>0.002</td>
<td>0.938</td>
</tr>
<tr>
<td>Neutrophils, %</td>
<td>0.009</td>
<td>0.226</td>
</tr>
<tr>
<td>Lymphocytes, %</td>
<td>-0.009</td>
<td>0.246</td>
</tr>
<tr>
<td>Monocytes, %</td>
<td>-0.001</td>
<td>0.981</td>
</tr>
<tr>
<td>Eosinophils, %</td>
<td>-0.014</td>
<td>0.643</td>
</tr>
<tr>
<td>Basophils, %</td>
<td>0.005</td>
<td>0.968</td>
</tr>
</tbody>
</table>

* Adjusted for age, gender and study.

---

**Zhu et al., Int J Epidemiology 2012**

- Need to account for signals from cell type differences
  - Adjust for cell type in multivariate analysis
  - Normalize methylation data for cell types before data analysis
DNA methylation arrays as surrogate measures of cell mixture distribution

Houseman et al, BMC Bioinformatics 2012
Step by step

Step 1
Intro to epigenetics

Step 2
Epigenetics & Environment

Step 3
Study Design

Step 4
Starting a Study

Step 5
Lab Analysis
DNA Methylation
  • Global methylation content
    – How much methylation in a test DNA, regardless of the position
  • Gene-specific analysis
    – How much methylation at or nearby a candidate gene
  • Genome-wide scans
    – Microarrays, Next Generation Sequencing

Histone Modifications
  • Global modification content
  • Gene-specific analysis
  • Genome-wide scans
DNA methylation & histone modification analysis

• Molecular biology: largely related to genetics
  – methods that analyze the DNA sequence
• Epigenetic marks: do not modify the underlying DNA sequence
• Workarounds:
  – Bisulfite treatment
  – Antibody-based methods
    (or alternatively methyl-binding proteins)
Bisulfite treatment

Modifies non methylated cytosines

Differentiation of methylated and non methylated cytosines

Any method that can analyze sequence

DNA methylation

Antibodies

Bind modified or methylated cytosines, modified histones

DNA enriched with the mark of interest (Ab specificity)

Any method that can quantify enrichment

Histone modifications
Bisulfite modification of DNA

- Prior to PCR, DNA is treated with sodium bisulfite
- Non-methylated C is permanently modified to U
- In PCR, U and T are equivalent
Scans of the (entire?) epigenome
Genome-wide scans

• Microarrays
  – **Illumina Infinium for DNA methylation** (bisulfite treatment)
    • 484,000 CpG sites (450K, released in 2011)
  – Nimblegen for DNA methylation or Histone Modifications (Ab-based)
    • 2 Million probes

• Next Generation sequencing
  – Various platforms
  – Bisulfite treatment (DNA methylation)
  – Ab-based (DNA methylation or Histone modifications)
Features covered in the 450k Infinium BeadChip

The 450K BeadChip covers a total of 77,537 CpG Islands and CpG Shores (N+S)

<table>
<thead>
<tr>
<th>Region Type</th>
<th>Regions</th>
<th>CpG sites covered on 450K BeadChip array</th>
<th>Average # of CpG sites per region</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpG Island</td>
<td>26,153</td>
<td>139,265</td>
<td>5.08</td>
</tr>
<tr>
<td>N Shore</td>
<td>25,770</td>
<td>73,508</td>
<td>2.74</td>
</tr>
<tr>
<td>S Shore</td>
<td>25,614</td>
<td>71,119</td>
<td>2.66</td>
</tr>
<tr>
<td>N Shelf</td>
<td>23,896</td>
<td>49,093</td>
<td>1.97</td>
</tr>
<tr>
<td>S Shelf</td>
<td>23,968</td>
<td>48,524</td>
<td>1.94</td>
</tr>
<tr>
<td>Remote/Unassigned</td>
<td>-</td>
<td>104,926</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>485,553</strong></td>
<td></td>
</tr>
</tbody>
</table>

The 450K BeadChip covers a total of 77,537 CpG Islands and CpG Shores (N+S)

The 450K BeadChip covers a total of 20,617 genes
(m)RRBS - (multiplexed) Reduced Representation Bisulfite Sequencing

- Gel-free, high-throughput version of RRBS
- Method utilizes cutting pattern of MspI (C^CGG) to systematically digest DNA, enrich for CpG dinucleotides
  - "reduced representation" of the genome, which covers the majority of CpG islands and promoters, and a reasonable number of exons, shores and enhances
- Advantages
  - only need 50-200ng DNA
  - can be from any species
  - cost and time (reduced sequencing, 96-well format, multiplexing)
Number of CpG covered by mRRBS

Our 28 of 42 passing samples (67%)
Other methods for deep bisulfite sequencing

• Straight up sequencing
  – Bona fide genome wide
  – Current costs around $5,000/sample
• Agilent SureSelect
  – Genome-scale bisulfite sequencing
    (about 2 million GpG sites)
  – Custom CpG sites
• Padlock
  – 4,000-400,000 CpG sites
• RainDance
  – 500-1000 amplicons
  – Proposed for replication of EWAS studies
Sources of analytical variability

• Several factors can affect results
  – DNA/sample quality
  – Plate effects
  – Batch effect
  – Row/column effect

• How to handle this
  – Best laboratory practice
  – Randomize/balance samples
  – Universal DNA/Replicates
  – Bioinformatics/Statistical analysis
Step by step

- Step 1: Intro to epigenetics
- Step 2: Epigenetics & Environment
- Step 3: Study Design
- Step 4: Starting a Study
- Step 5: Lab Analysis
- Step 6: Data Analysis
Data preprocessing
Illumina 450K - Infinium 1 vs. Infinium 2

Dedeurwaerder et al., Epigenomics 2011
mRRBS – analytic pipeline
Applying existing methods to new types of data
Illumina Infinium 450K peculiar kind of data

**Beta-value:**

\[ \beta = \frac{M}{U + M} \]

- \( \beta = 0 \): All cells are non-methylated
- \( \beta = 1 \): All cells are methylated
Illumina Infinium 450K

peculiar kind of data

Beta-value:

\[ \beta = \frac{M}{U + M} \]

- \( \beta = 0 \) : All cells are non-methylated
- \( \beta = 1 \) : All cells are methylated

M value:

\[ Mvalue = \log \left( \frac{M}{U} \right) \]

- \( M = -\infty \) : All cells are non-methylated
- \( M = +\infty \) : All cells are methylated
Illumina Infinium 450K
peculiar kind of data

Beta-value:

\[ \beta = \frac{M}{U + M} \]

\[ \beta = 0 \quad : \quad \text{All cells are non-methylated} \]
\[ \beta = 1 \quad : \quad \text{All cells are methylated} \]

• Handling β-values
  – Linear regression and hope for the best
  – Robust regression (Joubert BR et al EHP 2012)
  – Beta-regression (Seow WJ, PLOSLone 2012)
EWAS analyses (e.g. Illumina Infinium 450k)

- Type 1 error (False positives)
  - Bonferroni (p-value x 486K)
  - Benjamini-Hochberg False Discovery Rates (FDRs or q-values)
- Type 2 error (False negatives)
  - Adequate power (high sample size)
  - Meta-analyses
EWAS validation – Study design

- **Discovery only (Single study)**
  - Prone to false positive findings (negative too)

- **Split sample design**
  - Divide population in two
  - Split 1: EWAS; Split 2: candidate gene analysis
  - Overall power lower than same-size discovery only
  (Skol AD, Nat Genet 2006).

- **Single study cross-validation**
  - Sample a study subset (e.g., 90%) multiple times

- **Discovery > Replication**
  - Two independent studies
  - Ensure validation + generalizability

- **Meta-analysis**
  - Uses estimates from multiple populations
  - Needed to achieve large sample size
  - Allows for evaluating generalizability
Step by step

Step 1: Intro to epigenetics
Step 2: Epigenetics & Environment
Step 3: Study Design
Step 4: Starting a Study
Step 5: Lab Analysis
Step 6: Data Analysis
Step 7: Wrap up
Challenges in epigenetics

• How many epigenomes in one body?
  – Tissue specificity
    • Most studies in humans are on blood DNA
    • Need to investigate tissues relevant for the exposure-disease of interest (challenging in epidemiology)

• How many epigenomes in one lifetime?
  – The epigenome changes over time
    • Reverse causation is always a potential issue
    • Need for longitudinal studies

• How many epigenomic markings in one epigenome?
  – DNA methylation, histone modifications, others
    • Which is most informative?
Opportunities in epigenetics

• How many epigenomes in one body?
  – Opportunities for screenings of multiple epigenomes:
    • Multiple tissues (e.g., ELEMENT study)
    • Multiple cell types (e.g., blood subpopulations)

• How many epigenomes in one lifetime?
  – Opportunities for lifecourse epigenetics:
    • The epigenome might record recent or past experiences
    • The epigenome might predict future risk of disease

• How many epigenomic markings in one epigenome?
  – Integrate multiple epigenomic markings
    • Coordinated and complementary mechanisms
Some readings

• Michels K. *Epigenetic Epidemiology*. Springer, 2012


Consortia & Initiatives

The NIH Roadmap Epigenomics Program
A NIH Initiative to foster epigenomic research, develop comprehensive reference epigenome maps, and generate new technologies for epigenomic analyses.
http://nihroadmap.nih.gov/epigenomics/

The Epigenome Network of Excellence
An EU-funded network of institutions and research groups
http://www.epigenome-noe.net/WWW/index.php

The Human Epigenome Projects
A public/private collaboration to catalogue Methylation Variable Positions (MVPs) in the human genome
http://www.epigenome.org/

NAME21
A German National Initiative to analyze DNA methylation Patterns of Genes on Chromosome 21
http://biochem.jacobs-university.de/name21/
Online databases

The Human Epigenome Atlas
The atlas includes human reference epigenomes and the results of their integrative and comparative analyses.
http://www.genboree.org/epigenomeatlas/index.rhtml

MethDB
A searchable database for DNA methylation and environmental epigenetic effects
http://www.methdb.de/

Human Histone Modification Database (HHMD)
A searchable database of information from experimental data to facilitate understanding of histone modifications. It incorporates 43 location-specific histone modifications in humans.
http://bioinfo.hrbmu.edu.cn/hhmd

NCBI Epigenomics
An online repository of epigenetic datasets

GenelImprint
A catalogue of imprinted genes
http://www.geneimprint.com/site/genes-by-species

Catalogue of Parent of Origin Effects
Searchable database of imprinted genes and related effects
http://igc.otago.ac.nz/home.html
Tools and Other Resources

**MethPrimer**
*Primer Design for Methylation PCR*
http://www.urogene.org/methprimer/index1.html

**MethBlast**
*A sequence similarity program that checks your primers for bisulfite converted DNA by blasting them against unmethylated and methylated genomic sequences of man, mouse and rat*
http://medgen.ugent.be/methBLAST/

**Methylator**
*Methylator attempts to predict whether CpGs in a DNA sequence are likely to be methylated or not*
http://bio.dfci.harvard.edu/Methylator/

**RMAP**
*RMAP is a tool to map reads from the next-generation sequencing technology that supports bisulfite-treated reads mapping.*
http://rulai.cshl.edu/rmap/

**Chromatin Structure & Function**
*Information on chromatin biology, histones and epigenetics*
http://www.chromatin.us/chrom.html
Thanks!