Lecture 22

Personal genomics, disease epigenomics, systems approaches to disease

Predictive Medicine
Molecular Epidemiology
Mendelian Randomization

Module 5: Personalized and Systems Medicine

- L23: The three-dimensional genome.
- L24: Pharmacogenomics.
- L26/27: Term project presentations!

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
   - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models

Epidemiology

The study of the patterns, causes, and effects of health and disease conditions in defined populations
Epidemiology: Definitions and terms

- **Morbidity** level: how sick an individual is
- **Incidence**: # of new cases / # people / time period
- **Prevalence**: Total # of cases in population
- **Attributable risk**: rate in exposed vs. not exposed
- **Population burden**: yrs of potential life lost (YPLL), quality-/disability-adjusted life year (QALY/DALY)
- **Syndrome**: Co-occurring signs (observed), symptoms (reported), and other phenomena; (often hard to establish causality / risk factors)
- **Prevention challenge**: Determine disease, cause, understand whether, when, and how to intervene

Determining disease causes: study design

- **Principles of experimental design**
  - **Control**: comparison to baseline, placebo effect
  - **Randomization**: Difficult to achieve, ensure mixing
  - **Replication**: control variability in initial sample
  - **Grouping**: understand variation between subgroups
  - **Orthogonality**: all combinations of factors/treatments
  - **Combinatorics**: factorial design $n \times n \times n \times \ldots \times n$ table

- **Challenge of human subjects**
  - Legal and ethical constraints, Review boards
  - Randomization by instrumental variables
  - Clinical trials: blind (patient), double-blind (doctor too)

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
   - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models

Genetic Epidemiology

Genetic factors contributing to disease

Genome-wide association studies (GWAS)

- Identify regions that co-vary with the disease
- Risk allele G more frequent in patients, A in controls
- But: large regions co-inherited $\Rightarrow$ find causal variant
- Genetics does not specify cell type or process

All disease-associated genotypes from GWAS

- 1000s of studies, each with 1000s of individuals
  - Increasing power, meta-analyses reveal additional loci
  - More loci expected, only fraction of heritability explained
Complex disease: strong non-coding component

Monogenic / Mendelian Disease

Polygenic / Complex Disease

Human Genetic Mutation Database
April 2010 release

Catalog of GWAS studies
Hindorff et al. PNAS 2009

More loci on the way: GWAS growth continues

• When to design custom chip: continuously update
• http://www.genome.gov/admin/gwascatalog.txt

Decreasing cost of whole-genome sequencing

• Simply genotype all known variants at >0.1% freq
• Or: sequence complete diploid genome of everyone

Genetic epidemiology: What to test

• Family risk alleles, inherited with common trait
  – Specific genes, specific variants, family history
• Monogenic, actionable, protein-coding mutations
  – Most understood, highest impact, easiest to interpret
• All coding SNPs with known disease association
  – What if not druggable / treatable? Want/need know?
• All coding/non-coding associations from GWAS
  – Thousands of significant associations (1350 on 6/2012)
• All common SNPs, regardless of association
  – HapMap and 1000 Genomes capture common variants
• Genome: all SNPs, CNVs, rare/private mutations

Predictive medicine: When to screen

• Diagnostic testing: after symptoms, confirm a hypothesis, distinguish between possibilities
• Predictive risk: before symptoms even manifest
• Newborn: heel pick, store, for early treatment
• Pre-natal testing: ultrasounds, maternal serum vs. needles, probes, chorionic villus sampling
• Pre-conception testing: common/rare disorders
• Carrier testing: specific mutation in family history
• Genetics vs. biomarkers: cause vs. consequence?

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   – Genetic basis: GWAS and screening
     – Interpreting GWAS with functional genomics
     – Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   – meQTLs: Genotype-Epigenome association (cis-/trans-)
   – EWAS: Epigenome-Disease association
4. Resolving Causality
   – Statistical: Mendelian Randomization
   – Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   – Beyond single-gene models
   – Pathway-based, genome-wide models
Interpreting disease associations
Functional genomics of GWAS

Interpreting disease-association signals

(1) Interpret variants using Epigenomics
- Chromatin states: Enhancers, promoters, motifs
- Enrichment in individual loci, across 1000s of SNPs in T1D

(2) Epigenome changes in disease
- Intermediate molecular phenotypes associated with disease
- Variation in brain methylomes of Alzheimer’s patients

LD: both a blessing & a curse

Observation: LD blocks in which there is no evidence for historical recombination

Mechanistic predictions for top disease-associated SNPs

- Lupus erythromatosus in GM lymphoblastoid
- Erythrocyte phenotypes in K562 leukemia cell line

- Disrupt activator Ets-1 motif
  - Loss of GM-specific activation
  - Loss of enhancer function
  - Loss of HLA-DRB1 expression

- Creation of repressor Gfi1 motif
  - Gain K562-specific repression
  - Loss of enhancer function
  - Loss of CCDC162 expression

Detect SNPs that disrupt conserved regulatory motifs

- Functionally-associated SNPs enriched in states, constraint
- Prioritize candidates, increase resolution, disrupted motifs
Allele-specific chromatin marks: cis-vs-trans effects

- Maternal and paternal GM12878 genomes sequenced
- Map reads to phased genome, handle SNPs indels
- Correlate activity changes with sequence differences

HaploReg: systematic ENCODE mining of variants (compbio.mit.edu/HaploReg)

- Start with any list of SNPs or select a GWA study
  - Mine publicly available ENCODE data for significant hits
  - Hundreds of assays, dozens of cells, conservation, motifs
  - Report significant overlaps and link to info/browser

Interpreting GWAS with functional genomics

1. Big questions by integrating ENCODE and GWAS data:
   - Predict relevant cell types
   - Fine-map individual regions
   - Predict common pathways
   - Understand disease architecture.
2. What reference set(s) of disease association results should be used in our tests?
   - NHGRI catalog
   - Wellcome Trust case-control studies
   - Rare variants associated with cancer, specific diseases
3. What reference set of SNPs should be used in our association studies?
   - Lead SNPs
   - 1000GP imputed SNPs or LD partners (Imputation requires availability of original study data, LD partners will give many hits for one lead SNP – how to prioritize?)
   - Low-frequency and rare variants
4. What reference regulatory annotations should be used
   - TF binding, Chromatin marks, Chromatin states, DNase, DGF
   - Regulatory motif annotations, Predicted causal regulatory motifs
5. What statistical tests and permutations should be used
   - Enrichment tests (Sampling, closed-form, empirical P-values)
   - Permute SNPs: Match MAF, LD length, # SNPs, distance to TSS, all GWAS SNPs
   - Permute regulatory annotations: Peak-shifting, Permutation, other cell types
6. Disseminate results: Haploreg, RegulomeDB, Outreach, ASHG

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
   - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models

Interpreting disease-association signals

1. Interpret variants using Epigenomics
   - Chromatin states: Enhancers, promoters, motifs
   - Enrichment in individual loci, across 1000s of SNPs in T1D

   (1) CATGACTG
   (2) CATGCTG

   Genotype
   GWAS
   Disease
   mQTLs
   Epigenome
   MWAS

   (2) Epigenome changes in disease
   - Intermediate molecular phenotypes associated with disease
   - Variation in brain methylomes of Alzheimer’s patients

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
   - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models
Molecular Epidemiology

Molecular Biomarkers of disease state:
Gene expression, DNA methylation, chromatin in specific cell types

Data Matrices – An example scenario

Genetic and epigenetic data in 750 Alzheimer’s patients/controls

EWAS: Capturing variability in the Epigenome attributable to disease

Excluding discovered and known covariates

Genotype→Methylation

Discovering mQTLs
Methylation Quantitative Trait Loci
**cis-meQTLs**

Use linear models to identify cis-meQTLs within some genomic window.

For methyl mark $m_i$ and SNP $g_j$:

$$m_i = \beta_0 + \beta_1(g_j) + \epsilon$$

- Given several predictors: is additional predictor increasing accuracy more than complexity introduced?
- Likelihood ratio testing paradigm: predict methylation with and without genotype (only works for nested models)
- Null hypothesis $H_0$: $\beta_1 = 0$: Additional model complexity doesn't explain a significant portion of variation in response

**Test using F statistic:**

- $p$ is the number of parameters in $LM_1$
- $q$ is the number of parameters in $LM_2$
- $n$ is the sample size
- RSS: Residual sum of squares
- $\beta$: parameters to learn. $\epsilon$: residual error term.

Under null hypothesis: $F = \frac{(RSS_{LM1} - RSS_{LM2}) / (q - p)}{RSS_{LM2} / (n - q)}$

Is distributed as F distribution with $(q-p, n-q)$ degrees of freedom

- If $F$ statistic significant: reject null: This p-value is what we report in a meQTL study
- Otherwise, no meQTL: i.e. $RSS_{LM1} - RSS_{LM2}$ too small vs. increase in model complexity

**Alternative methods of detection:**

- Permutation:
  - Correlate methylation and genotype.
  - For $i$ in 1 -> nperm:
    - Permute genotypes
    - Correlate methylation and genotype
  - Generate empirical p-value from permuted correlations
- LMM: Linear mixed models.

---

**Most epigenomic variability is genotype-driven**

Manhattan plot of 450,000 methylation probes

- Genome-wide significance at $p<3 \times 10^{-10}$
- Prune for probes disrupted by SNP
  - 140,000 CpGs associated with genotype at 1% FDR
  - 55,000 at Bonferroni-corrected P-value of $10^{-2}$

**Scaling of discovery power with individuals**

- Number of meQTLs continues to increase linearly
- Weak-effect meQTLs: median $R^2<0.1$ after 400 indiv.

---

**Importance of search parameters for QTL finding**

- Search window
- Minor allele frequency
- Correcting for co-variates

---

**The indecision tree ...**

- # components
- PCA
- BG + QN
- SVA
- Non-specific gene filtering
- Search Radius
- Minor Allele Frequency
- Normalization
- Non-specific procedures
- # samples
- 10K

Slide credit: Benjamin Raby

QN – Quantile Normalization

IQR – Inter-quartile range
**Effect of eQTL-gene distance on QTL discovery**

- Not always a monotonic increase
- Depends on the study, the specific phenotype, and the regulatory architecture of the disease

**Effect of minor allele frequency (MAF)**

- Rare variants: Fewer datapoints \(\rightarrow\) greater test statistic variance
- Drop in hits: rare variants (a) inflate multiple comparisons, (b) are underpowered, (c) corrupt FDR broadly

**Effect of principal component correction**

- Correcting for too few PCs \(\rightarrow\) noise remains
- Correcting for too many \(\rightarrow\) eliminate signal

**Goal: Personalized and Predictive Medicine**

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
     - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models

**eWAS**

Link methylation \(\rightarrow\) phenotype (\(\sim\) cis-eQTLs):
- linear models and hypothesis testing
- Predict phenotype using methylation

\[ LM1: p_i = \beta_0 + \varepsilon \]
\[ LM2: p_i = \beta_0 + \beta_1 m_i + \varepsilon \]
**eWAS**

- Link methylation to phenotype (~cis-eQTLs): linear models and hypothesis testing
- Predict phenotype using methylation

**Problem:** variance due to phenotype probably very small (unless your phenotype is cancer)

Needle in a haystack

Control for other sources of variance to make the variance due to the phenotype stand out.

If phenotype is Alzheimer’s (AD), gender incorporates more variance into your M matrix than does AD.

**LM1:** AD = β₀ + β₁(gender) + ε

**LM2:** AD = β₀ + β₁(mj) + β₂(gender) + ε

 Might have many environmental variables to control for.

**Role of QTLs in disease phenotypes**

**Majority of AD-associated GWAS SNPs are meQTLs**

<table>
<thead>
<tr>
<th>Rank</th>
<th>rsId</th>
<th>Gene</th>
<th>Description</th>
<th>meQTL P-value</th>
<th>meQTL SNP</th>
<th>meQTL Gene</th>
<th>SNP state</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs11767557</td>
<td>EPHA1</td>
<td>Ephrene A receptor 1</td>
<td>1.12E-13</td>
<td>rs12703526</td>
<td>cg18997129</td>
<td>24_Quies3</td>
</tr>
<tr>
<td>2</td>
<td>rs1532278</td>
<td>CLU</td>
<td>Clusterin</td>
<td>6.68E-125</td>
<td>rs17057441</td>
<td>cg18814083</td>
<td>22_Quies1</td>
</tr>
<tr>
<td>3</td>
<td>rs3865444</td>
<td>CD33</td>
<td>Myeloid transmembrane receptor</td>
<td>1.13E-10</td>
<td>rs12971624</td>
<td>cg11581627</td>
<td>22_Quies1</td>
</tr>
<tr>
<td>4</td>
<td>rs561605</td>
<td>PICALM</td>
<td>Phosphatidylinositol binding clathrin assembly protein</td>
<td>7.11E-77</td>
<td>rs17817919</td>
<td>cg24166175</td>
<td>22_Quies1</td>
</tr>
<tr>
<td>5</td>
<td>rs610932</td>
<td>MS4A2</td>
<td>Immunoglobulin receptor subunit</td>
<td>1.37E-33</td>
<td>rs562028</td>
<td>cg16954525</td>
<td>22_Quies1</td>
</tr>
<tr>
<td>6</td>
<td>rs6701713</td>
<td>CR1</td>
<td>Complement Receptor 1</td>
<td>1.39E-21</td>
<td>rs3849266</td>
<td>cg19373649</td>
<td>22_Quies1</td>
</tr>
<tr>
<td>7</td>
<td>rs7561528</td>
<td>BIN1</td>
<td>Bridging Integrator Nucleocytoplasmic adaptor protein</td>
<td>3.73E-176</td>
<td>rs4663104</td>
<td>cg02887598</td>
<td>2_TssF</td>
</tr>
<tr>
<td>8</td>
<td>rs9349407</td>
<td>CD2AP</td>
<td>Actin Cytoskeleton Regulating Scaffold</td>
<td>7.12E-63</td>
<td>rs2275446</td>
<td>cg16361253</td>
<td>22_Quies1</td>
</tr>
</tbody>
</table>

**GWAS variants are enriched for eQTL**

Correlated with strength of association

Independent of allele frequency

Nicolae et al. PLoS genetics. 2010

Murphy A et al. Hum Mol Genet. 2010

Slide credit: Benjamin Raby
Can we leverage eQTL data for disease-variant identification?

- Multivariate regression model building:
  - Outcome: Disease-susceptibility variants identified by GWAS and all variants with \( r^2 > 0.8 \)
  - Variables: Distance from TSS, MAF, Chromatin state, and eQTL FDR
- Training set: random set of 2.7 million variants mapping to within 50 kb of transcript
- Test set: remaining ~2.0 million variants

**Multivariate Predictors of GWAS-tagging variants**

- **Distance from Gene (25-50 kb ref)**
  - Within gene
  - < 1kb
  - 1kb – 5kb
  - 5kb – 10kb
  - 10kb-25kb
- **Minor Allele Frequency (25-5 ref)**
  - 0.00 – 0.005
  - 0.005-0.05
  - 0.05-0.10
  - 0.10-0.25
- **eQTL FDR (> 2 ref)**
  - 0 – 0.05
  - 0.05-0.10
  - 0.10-0.20
- **Chromatin State (Heterochr ref)**
  - Repetitive/CNV
  - Insulator
  - Weak Txn
  - Weak Enhancer I
  - Elongation
  - Txn Transition
  - Weak Promoter
  - Active Promoter
  - Poised Promoter
  - Repressed
  - Weak Enhancer II
  - Repetitive CNV
  - Strong Enhancer I
  - Strong Enhancer II
  - None

**Role of enhancers vs. promoters in Alzheimer’s disease association**

- Highly distinct signatures for promoters vs. enhancers
- Enhancers hemi-methylated and highly variable
- Promoters show least variable methylation
- Enrichment for meQTLs
- SNP-associated CpGs in enhancers, not promoters
- TSS-flanking region
- Transcribed
- Repressed
- Promoter methylation less affected by genetics
- Enhancer methylation highly genotype-driven
- TSS-flanking and repressed regions also genetic
AD-associated probes in distal enhancers

- After cleaning with known and inferred covariates.
- Distal and transcribed enhancers enriched.
- Proximal regulators (promoters) depleted.

ICA covariate correction cleans up enhancer signal

Before:
Orange: Enrichment of enhancer probes for association with the real phenotype.
Grey: Enrichment of enhancer probes for a scrambled phenotype.

Empirical p=0.06
Empirical p<0.0001

After:
(After conditioning on 7 surrogate variables discovered with ICA.)

AD predictive power highest in enhancers

Top predictive features are:
- Enhancer methylation
- All methyl
- TSS, Het
- Genetics (incl. APOE)

- Causality?
- Common pathways?

AD prediction reveals likely biological pathways

HEB/Tcf12: proliferating neural and progenitor cells
GATA: cell growth, blood, cell development
TLX1/NFIC: Neuronal cell fates

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
   - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models

Risk factor causality w/ instrumental variables

If X ⇔ Y are correlated, possible scenarios are:
- X → Y
- Y ⇔ X
- X ⇔ U ⇔ Y

To distinguish, need controlled random experiment

- Is risk factor X causing disease Y (or a consequence)?
  - E.g. alcohol addiction, smoking, blood cholesterol, fever, stress
  - Randomized experiment, with and without X: feasibility? ethics?
- G ⇔ randomized experiment (e.g. random Mendelian inheritance), as only some subjects have genotype
- G (i.v.) must be correlated with Y but only through X
  - i.e. if X known, G gives no additional information about Y
In silico thought experiment

Hemi-methylation associated with meQTL yields a p-value that’s 30 orders of magnitude lower for the AD phenotype.

Mendelian randomization approach

Account for variance due to genotype, how much does methylation add?

From G, include probe-specific terms for cis-meQTLs, as well as including trans-meQTLs in all comparisons.

With variability due to genotype and environmental covariates removed, the effect due to phenotype should become more prevalent.

Goal: Personalized and Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   – Genetic basis: GWAS and screening
   – Interpreting GWAS with functional genomics
   – Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   – meQTLs: Genotype-Epigenome association (cis-/trans-)
   – EWAS: Epigenome-Disease association
4. Resolving Causality
   – Statistical: Mendelian Randomization
   – Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   – Beyond single-gene models
   – Pathway-based, genome-wide models

Rank-based functional testing of weak associations

Enrichment peaks at 10,000s of SNPs down the rank list, even after LD pruning!

Weak-effect T1D hits in 50k T-cell enhancers

- Rank all SNPs based on GWAS signal strength
- Functional enrichment for cell types and states

LD-pruning (CEU r²>.2): 50k ➔ 41k independ. loci
Cell type specificity stronger for enhancers

- T/B-cells also enriched for promoters, transcribed
- Enhancer enrichment much more cell type specific

T1D/RA-enriched enhancers spread across genome

- High concentration of loci in MHC, high overlap
- Yet: many distinct regions, 1000s of distinct loci

Understanding interactions:
Linking disease variants into networks

- SP3 no direct assoc but clusters w/ many MS hits
- SP3 is indeed down-regulated in MS patients

Example 1: SP3 predicted role in MS

Example 2: MAZ predicted role in T1D

- MAZ no direct assoc, but clusters w/ many T1D hits
- MAZ indeed known regulator of insulin expression
Personal genomics tomorrow: Already 100,000s of complete genomes

- Health, disease, quantitative traits:
  - Genomics regions \(\rightarrow\) disease mechanism, drug targets
  - Protein-coding \(\rightarrow\) cracking regulatory code, variation
  - Single genes \(\rightarrow\) systems, gene interactions, pathways

- Human ancestry:
  - Resolve all of human ancestral relationships
  - Complete history of all migrations, selective events
  - Resolve common inheritance vs. trait association

- What’s missing is the computation
  - New algorithms, machine learning, dimensionality reduction
  - Individualized treatment from 1000s genes, genome
  - Understand missing heritability
  - Reveal co-evolution between genes/elements
  - Correct for modulating effects in GWAS

Challenge ahead: From research to clinic

1. Systematic medical genotyping / sequencing
   - Currently a curiosity, future: medical practice
2. Systematic medical molecular profiling
   - Functional genomics in relevant cell types
3. Systematic perturbation studies for validation
   - 1000s of regulatory predictions x 100s cell types
4. Systematic repurposing of approved drugs
   - Systems-biology view of drug response
5. Genomics of drug response in clinical trials
   - Personalized drug prescription and combinations
6. Partnerships: academia, industry, hospitals
   - Interdisciplinary training in each of the institutions

Summary: Personalized & Predictive Medicine

1. Intro to Epidemiology: basis of human disease
2. Genetic Epidemiology:
   - Genetic basis: GWAS and screening
   - Interpreting GWAS with functional genomics
   - Calculating functional enrichments for GWAS loci
3. Molecular epidemiology
   - meQTLs: Genotype-Epigenome association (cis-/trans-)
   - EWAS: Epigenome-Disease association
4. Resolving Causality
   - Statistical: Mendelian Randomization
   - Experimental: Massively Parallel Assays
5. Systems Genomics and Epigenomics of disease
   - Beyond single-gene models
   - Pathway-based, genome-wide models