Array Based Genome Wide DNA Methylation and Genotyping Data Analysis

Jenny Wu, Ph.D.
April 4, 2023
Illumina Infinium Arrays at GRTH

- iScan System

- Bead based microarray technology for methylation and genotyping

- Data Analysis support at GRT hub

- Methylation data analysis workshop on 4/25
Infinium Methylation Assay

• Two types of beads to detect methylation

• Target site matches with the probe, enabling single-base extension and detection. Single-base mismatch inhibits extension
Illumina Methylation Arrays

- Infinium Human MethylationEPIC Array
  - V2 has 935k methylation sites with coverage of RefSeq genes, CpG islands, enhancer regions, open chromatin sites etc.
  - Compatible with FFPE samples
  - 8 Samples per BeadChip
- Infinium Mouse Methylation Array
  - >285k methylation sites per sample at single-nucleotide resolution
  - 24 samples per BeadChip
Infinium Methylation Assay Workflow

Bisulfite Conversion → Whole Genome amplification and enzyme fragmentation → Array Processing Scanning

**Template:** (Double Stranded)

A: 5'GACGGTCCAGTCACAGCTGCT-3'

B: 3'-TTCGCAACGGTCCAGTCACAGCTGCT-5'

**Bisulfite Converted:** (Single Stranded)

A: 5'GATCGTTTTAACGGTACGTTTACGTTTACGTTT-3'

B: 3'-TTCCGCAACGGTTTACGTTTACGTTTACGTTT-5'

---

Genomics Research and Technology Hub
Methylation Array Work Flow

**Data Generation**
- Lab Preparation
- Array Processing
- idat

**QC and Pre Processing**
- Quality Control and Normalization
- Bisulfite conversion
- Beta value density
- SWAN
- GenomeStudio, Champ, missMethyl

**Methylation Calling**
- M values and Beta values for each site
- GenomeStudio, minfi, Champ, SeSame

**Differential Methylation**
- Identify differentially methylated sites and regions
- Limma, bumphunter, DMRFinder

Common Software Tools

Genomics Research and Technology Hub
Illumina GenomeStudio

- Basic QC, visualization and analysis and generate reports.
- Does not container control probes
- No advanced analysis

https://support.illumina.com/array/array_software/genomestudio/downloads.html
R Based Tools for Methylation

- End to End solution for methylation Analysis
  - Preprocessing, QC assessment, normalization, methylation calling, plotting functionality
  - DMC and DMR analysis

- **Minfi, Champ, SeSaMe**

...
Methylation Array Pricing

• Reagent Expenses:
  • Human: Infinium Methylation EPIC v2.0 Kit – 8 samples per chip
  • Mouse: Infinium Mouse Methylation BeadChip kit - 24 samples per chip

• Labor Expense: $425 per chip

<table>
<thead>
<tr>
<th># samples</th>
<th>Human Kit ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2,392</td>
</tr>
<tr>
<td>16</td>
<td>4,576</td>
</tr>
<tr>
<td>32</td>
<td>8,320</td>
</tr>
<tr>
<td>96</td>
<td>24,960</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># samples</th>
<th>Mouse Kit ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>5,366</td>
</tr>
<tr>
<td>48</td>
<td>9,734</td>
</tr>
<tr>
<td>96</td>
<td>19,468</td>
</tr>
</tbody>
</table>
Illumina Global Diversity Array (GDA)

- ~1.9M Variants built for All of Us and optimized for precision medicine
- Population, precision, and clinical research
- Built for Multiple Ethnicities: High coverage of admixed U.S. population
- Capture low frequency variants

Genomics Research and Technology Hub
Genotyping Array Work Flow

Data Generation
- Lab Preparation
- Array Processing
  - Beeline
  - GTC2VCF

QC and Pre Processing
- Quality Control:
  - MAF filtering
  - Relatedness
  - Population-Stratification
- Plink, plinkQC

Ancestry Estimation
- Identify Ref Panel
  - (1000 genome, HapMap3)
  - Ancestry Estimation
- STRUCTURE, Admixture

Polygenic Risk Score
- Identify base GWAS
  - (GWAS catalog)
  - Imputation
  - PRS Prediction
  - Validation Prediction
- Beagle, MaCH
  - PRSice2, LDpred

Medical Applications
- Patient Stratification
  - Sub-phenotyping
- Gene-gene, gene-environment interaction
  -......

Common Software Tools

Genomics Research and Technology Hub
Overall Workflow for Users

• IDAT files -> GenomeStudio = Final report and sample call rates.
• IDAT files -> Beeline = GTC files -> Illumina GTCToVCF = VCF files
• Final reports, VCF files can be analyzed using various program and packages.