Recurrent single-molecule epistates define tumor methylome differences

Huy Q. Dinh
Ben Berman lab @ USC Epigenome Center
University of Southern California
Tumor whole genome DNA methylation profiling by bisulfite sequencing

- ~ 10K samples profiled by Illumina Infinium 27K & 450K microarray platforms within TCGA projects.

- Whole Genome Bisulfite Sequencing (WGBS)
  - Methylation profile of 28M CpGs with more sequence variation information.
  - WGBS on 47 TCGA patient samples completed
  - >15x sequence coverage / sample

- Identification of recurrent Differentially Methylated Regions (DMRs) across multiple tumors based on single-molecule analysis of WGBS.

<table>
<thead>
<tr>
<th>Sample</th>
<th>#</th>
<th>Sample</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUAD</td>
<td>6</td>
<td>GBM</td>
<td>6</td>
</tr>
<tr>
<td>LUSC</td>
<td>5</td>
<td>BRCA</td>
<td>6</td>
</tr>
<tr>
<td>BLCA</td>
<td>7</td>
<td>STAD</td>
<td>5</td>
</tr>
<tr>
<td>UCEC</td>
<td>6</td>
<td>COAD</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>READ</td>
<td>3</td>
</tr>
</tbody>
</table>
Tumor heterogeneity in WGBS

WGBS reads

- methylated
- unmethylated

Average Methylation in 8-CpG window

Epipolymorphism (Landan et al, NatGen 2012)

- none
- high
- Low/intermediate

Epistates

Learning epistate mixture based on Expectation Maximization (EM)
Extended from allelic methylation model (Fang et al, PNAS 2012)
Estimating within-sample epistate frequencies

<table>
<thead>
<tr>
<th>Prob(α)</th>
<th>Prob(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.999</td>
<td>6e-10</td>
</tr>
<tr>
<td>0.999</td>
<td>5e-10</td>
</tr>
<tr>
<td>0.999</td>
<td>9e-09</td>
</tr>
<tr>
<td>0.999</td>
<td>9e-09</td>
</tr>
<tr>
<td>0.999</td>
<td>9e-09</td>
</tr>
<tr>
<td>0.999</td>
<td>9e-09</td>
</tr>
</tbody>
</table>

Averaged Expectation of epistate β frequency

Normal colon

- methylated
- unmethylated

Tumor colon

- methylated
- unmethylated

Pooling all reads

0 0 0.6 0 0 0.6 0 0
1 1 0.6 1 1 0.6 1 1

E(β reads) = \sum_{r \in R} 1 \times P(r | \beta)
E(α reads) = 1 - E(β reads)
Using epistates to identify DMRs

Colon
- Normal
- Tumor 1
- Tumor 2

Lung
- Normal
- Tumor 1
- Tumor 2

Tumor-specific epistate

Colon-specific epistate
Genome-wide scanning for epistates

Pool reads across all samples

Normal tissue reads

Tumor tissue reads

Read-length window scanning for whole genome (~150-200 bp)

(1) Good fit to two epistate model?

(2) Determining tumor-specific and tissue-specific loci
Estimating tumor purity based on epistate frequencies
Epistate estimate tumor purity highly correlated with ABSOLUTE estimation.

(Carter et al, NatGen 2012; Zack et al, NatGen 2013)
Epistates can be detected at low frequency

Methylated epistate found in only COAD tumors

COAD1: 40.34%  COAD2: 13.82%  COAD3: 15.79%
Epistate DMR example: Epigenetically silenced distal element (SMAD3)

SMAD3 mediate multiple signaling pathways including Wnt signaling and cell cycle.
Summary

- We developed **single-molecule epistate method** for analyzing multiple WGBS samples/tumors.

- **Epistate applications:**
  - Methylation-based tumor purity estimation.
  - Differentially Methylated Regions finding reveals epigenetically silenced (regulatory) regions at (low) epistate frequencies.
Acknowledgements

• Berman Lab
  – Yaping Liu
  – Lijing Yao
  – Ben Berman

• JHU
  – Steve Baylin

• USC Epigenome Center
  – Peter Laird
  – Hui Shen
  – Toshi Hinoue
  – Dan Weisenberger
  – Charlie Nicolet
  – Zack Ramjan
  – Kim Siegmund
  – Natalia Shatokhina
  – Selene Tyndale
  – Helen Truong

The Cancer Genome Atlas

NCI grant #1U24CA143882