Epigenomic Toolset

Mapping the Ontogenetic Pathways of Cellular Differentiation from the Human Epigenome Atlas Data Using the *Epigenomic Toolset* Within the Genboree Workbench

Aleks Milosavljevic
Baylor College of Medicine

Epigenomics: A Roadmap to the Living Genome
Boston, October 22\textsuperscript{nd}, 2013
Acknowledgements

Epigenomics Data Analysis and Coordination Center (EDACC)

Cristian Coarfa
Alan Harris
Viren Amin
Vitor Onuchic
Andrew Jackson
Sriram Raghuraman
Sameer Paithankar
Matt Roth
Kevin Riehle
Tim Charnecki

Joe Costello
Cydney Nielsen
Martin Hirst

Govind Ramamoorthy
Eleonora Laritsky
Robert Waterland

Steven Jones

Toutai Mituyama

AM 10/21/2013
The scope of the Epigenomic Toolset

Phylogeny

Reconstruct ontogeny
Ontogenetic pathways of cellular differentiation

Interpret epigenomic perturbations
Disease-associated
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. Inform experiment design by identifying informative marks

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
**Epigenomic Data Slice**

Regions Of Interest (ROIs)
(enhancers, promoters, ChIP-seq peaks, etc.)

<table>
<thead>
<tr>
<th></th>
<th>ROI 1</th>
<th>ROI 2</th>
<th>ROI 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell type 1</td>
<td>0.8</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Cell type 2</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Cell type 3</td>
<td>0.6</td>
<td>0.8</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Epigenomic marks (e.g., histone marks)

AM 10/21/2013
Heatmaps and Dendograms

<table>
<thead>
<tr>
<th></th>
<th>ROI 1</th>
<th>ROI 2</th>
<th>ROI 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell type 1</td>
<td>0.8</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Cell type 2</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Cell type 3</td>
<td>0.6</td>
<td>0.8</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Epigenomic Data Slice

Self-correlation

Dendogram

Heatmap

AM 10/21/2013
Dendogram of 100 epigenomes constructed from ~0.5M enhancers and H3K4me1

- Immune
- Endoderm
- Cardiac
- Ectodermal Epithelial
- Gastrointestinal
- Brain
- Fetal Organs
- Mesenchymal Stem Cells
ROIs with lineage-specific active epigenomic marks

- **lincRNA TSS**
  - 52% 1 lineage
  - 13% >1 lineages

- **Enhancers**
  - 60% 1 lineage
  - 20% >1 lineages

- **Promoters**
  - 41% 1 lineage
  - 8% >1 lineages

---

**Gene Counts**:

- Adult Brain: 41441
- Cardiac: 18675
- Epidermal: 20097
- Endoderm: 12309
- Fetal Muscle: 6467
- Fetal Organ: 995
- GastroInt: 13531
- HSCs: 21931
- Mesenchymal: 51164
- Myeloid: 13993
- NeuroProg: 19849
- Placenta: 7757
- Smooth Muscle: 16427
- Stem Cell: 87597
- T-cells: 59808
- **Total**: 392041
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. **Inform experiment design by identifying informative marks**

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
**Hypothesis:** a branch shared by many trees (confirmed by many mark-ROI combinations) corresponds to an ontogenetic pathway.
Informative epigenomic marks and mark-ROI combinations

ROIs

Confirmation frequency score

Marks
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. Inform experiment design by identifying informative marks

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
Mapping epigenomic changes using Spark tool

Enhancers

Enrichment

CD34

CD14

ENCODE TFBS

14

SUZ12
ZNF274
BRF1
HSF1
ZBTB7A
EGR1

TAL1
POU5F1
NANOG
IRF4

1
2
3
4
5
6
7
8
9
10

H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3

H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3
CG content

AM 10/21/2013
Pathway enrichment using H3K4me1 over enhancers and GREAT tool

<table>
<thead>
<tr>
<th>Pathway</th>
<th>-log10(Binomial p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>immune system process</td>
<td>256.86</td>
</tr>
<tr>
<td>defense response</td>
<td>255.45</td>
</tr>
<tr>
<td>immune response</td>
<td>232.62</td>
</tr>
<tr>
<td>inflammatory response</td>
<td>151.45</td>
</tr>
<tr>
<td>innate immune response</td>
<td>134.07</td>
</tr>
<tr>
<td>epidermis development</td>
<td>104.01</td>
</tr>
<tr>
<td>keratinocyte differentiation</td>
<td>72.65</td>
</tr>
<tr>
<td>epidermal cell differentiation</td>
<td>65.18</td>
</tr>
<tr>
<td>hair cycle</td>
<td>46.58</td>
</tr>
<tr>
<td>hair follicle development</td>
<td>46.27</td>
</tr>
<tr>
<td>muscle system process</td>
<td>143.43</td>
</tr>
<tr>
<td>muscle contraction</td>
<td>124.42</td>
</tr>
<tr>
<td>regulation of metal ion transport</td>
<td>73.78</td>
</tr>
<tr>
<td>regulation of calcium ion transport</td>
<td>63.33</td>
</tr>
<tr>
<td>regulation of heart contraction</td>
<td>52.48</td>
</tr>
<tr>
<td>synapse organization</td>
<td>60.95</td>
</tr>
<tr>
<td>glial cell differentiation</td>
<td>53.60</td>
</tr>
<tr>
<td>cell recognition</td>
<td>50.87</td>
</tr>
<tr>
<td>oligodendrocyte differentiation</td>
<td>46.31</td>
</tr>
<tr>
<td>regulation of stem cell differentiation</td>
<td>44.81</td>
</tr>
</tbody>
</table>
GWAS regions enriched for enhancers with immune-lineage specific H3K4me1 mark

Fold enrichment

<table>
<thead>
<tr>
<th>Condition</th>
<th>Fold Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean_platelet_volume</td>
<td>4</td>
</tr>
<tr>
<td>Multiple_Myeloma</td>
<td>3</td>
</tr>
<tr>
<td>betaThalassemia</td>
<td>2</td>
</tr>
<tr>
<td>Inflammatory_bowel_disease</td>
<td>2</td>
</tr>
<tr>
<td>Asthma</td>
<td>2</td>
</tr>
<tr>
<td>Rheumatoid_Arthritis</td>
<td>2</td>
</tr>
<tr>
<td>GWAS_of_Parkinson_disease</td>
<td>2</td>
</tr>
<tr>
<td>lupus_erythematosus</td>
<td>2</td>
</tr>
<tr>
<td>leprosy</td>
<td>2</td>
</tr>
<tr>
<td>schizophrenia</td>
<td>2</td>
</tr>
<tr>
<td>multiple_sclerosis</td>
<td>2</td>
</tr>
<tr>
<td>Celiac_disease</td>
<td>2</td>
</tr>
<tr>
<td>C-reactive_protein_blood_levels</td>
<td>2</td>
</tr>
<tr>
<td>venous_thromboembolism</td>
<td>1</td>
</tr>
<tr>
<td>glioma</td>
<td>1</td>
</tr>
<tr>
<td>GWAS_of_Crohn_disease</td>
<td>1</td>
</tr>
</tbody>
</table>
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. Inform experiment design by identifying informative marks

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
Dendogram including HL-60, a *promyelocytic leukemia* cell line confirms *myeloid lineage*
Spark analysis involving closest reference epigenomes reveals lack of differentiation

Epigenomic patterns over 1025 enhancers that change epigenomic state during myeloid differentiation

H3K27me3

H3K4me1
Pathway analysis reveals up-regulation of oncogenetic pathways

Analysis of enhancers with up-regulated H3K4me1 using GREAT tool
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. Inform experiment design by identifying informative marks

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
Samples from pure cell types

Methylation Profiling

100% methylated

Site A      Site B

Methylation Profiling

100% methylated

Site A      Site B
Samples from a complex tissue
Tumor vs. normal tissue

Tumor Samples

Normal Samples

Site A

Site B
Marker CpGs for Illumina arrays

LIMMA tool within the Genboree Workbench used to select marker CpGs that show lineage-specific methylation

- Illumina 450k array CpGs
- Illumina 27k array CpGs
- Shared 450K and 27K array CpGs

Tracks with Illumina marker CpGs available at www.genboree.org (Use case 21)
Comparison against reference epigenomes

ROIs: Marker CpG loci

Epigenome Atlas

Breast normal/tumor

AM 10/21/2013
Comparison against reference epigenomes

Comparing methylation arrays against the methylomes in the Epigenome Atlas

Step 1: Array import via the Genboree Workbench:
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. Inform experiment design by identifying informative marks

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
Virtual integration across the International Human Epigenome Consortium (IHEC)

Vancouver
Steven Jones
Martin Hirst

Tokyo
Toutai Mituyama

Houston
NIH Roadmap Epigenomics

Samples (by Type)

Epigenomic Assays
Virtual integration

Analysis
Only the “data slices” transferred across sites
(10MB vs 10GB)

Visualization in UCSC Genome Browser
Applications of the Epigenomic Toolset

1. Identify cell-type specific epigenomic states

2. Inform experiment design by identifying informative marks

3. Interpret epigenomic changes during cellular differentiation

4. Detect and interpret epigenomic perturbations in disease (cancer cells) relative to epigenomes of normal cells

5. Detect shifts in cell type composition in complex tissues (tumor vs. normal) using epigenomes of constituent cells

6. Enable world-wide collaboration and integrative analysis by virtual integration of data, tools, and computing resources
ASHG workshop tomorrow

Use cases available here

www.genboree.org