Computation Helps Tell Us How Muscles Are Made

Walter L. (Larry) Ruzzo

Computer Science & Engineering, UW
Genome Sciences, UW
Fred Hutchinson Cancer Research Center
Seattle, USA

http://www.cs.washington.edu/homes/ruzzo
http://bio.cs.washington.edu
3 World-Changing Technologies

~15 Giga-bases of DNA seq:
  a few days,
  a few $1000,
  a few terabytes
  (= 5 x your genome)
The “Text Book” Story: Transcription Factors Regulate Genes

DNA

TF TSS Gene (on or off?)

Promoter

Protein “Transcription Factor”
MyoD

Transcription factor involved in myogenesis

I.e., differentiation of muscle precursors (myoblasts) into muscle cells (myotubes)

There are other players, but H. Weintraub (1991) showed that this is the key factor – artificially inserting MyoD in many cell types (not just myoblasts) drives conversion into myotubes

“The molecular essence of commitment”
“The Standard Model”

MyoD binding absent or rare in myoblasts
Adding it (in myoblasts or many other cell types) starts a cascade leading to myotubes
~1000 genes show changed activity levels between myoblasts & myotubes
Expectation: MyoD drives those changes, by binding their promoters
Rest of talk in a nutshell

That elegant model insulted by data
MyoD Experimental Design

Myoblasts -> Chromatin IP with Myod antibody

- Crosslink DNA and proteins (optional) and isolate chromatin
- Sonicate or digest chromatin
- Immunoprecipitate, reverse crosslinking, purify DNA

Myotubes

Illumina Sequencing

5-20 million reads of 35-55 bp per sample

Map to Genome

(C.f. talk (3:55, rm 403)/poster (rm 407) on reconfigurable computing)
Identification of Binding Regions

Many overlapping reads (presumably) mark binding sites

Scattered, low-coverage reads = background noise

Many reads = High sharp peaks
Myod Locus

Chromosome 7

Myotube

Myoblast

Myod regulates itself; data confirms old experiments, adds novel sites.
Computational Questions

Did we sequence deeply enough?

How do we say “this peak increased or decreased” between myotubes and myoblasts?

What DNA sequence does MyoD bind?

Are there other sequence motifs nearby (binding cooperating factors)?

No time to describe today, but, trust me, interesting computational questions. One example below...
One Analysis Question: How tall must a peak be?

Estimate Poisson null model from “islands” of height 1, 2.

How likely is height 6? 12?

theory vs actual: very unlikely by chance.
"The Standard Model"

MyoD binding absent or rare in myoblasts
Adding it (in myoblasts or many other cell types) starts a cascade leading to myotubes
~1000 genes show changed activity levels between myoblasts & myotubes

Expectation: *MyoD drives those changes, by binding their promoters*
How Many Peaks?

As opposed to the 1000 genes changed, we find MyoD bound to

- **26,000** places (at 12-read cutoff; FDR < $10^{-6}$)
- **60,000** places (at ~.01 FDR)

In myotubes and myoblasts *both*

*(Excludes X, Y, repetitive regions)*
Where are peaks?

Concentrated near starts of genes

Binds in 41% of genes

But majority are far from any gene

50% are > 10kb from nearest TSS
### What's it doing?

#### Table 2. Gene Ontology Analysis on Differentially Bound Peaks in Myoblasts versus Myotubes

<table>
<thead>
<tr>
<th>GOID</th>
<th>Term</th>
<th>P Value</th>
<th>ORᵃ</th>
<th>Countᵇ</th>
<th>Sizeᶜ</th>
<th>Ontᵈ</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0005856</td>
<td>cytoskeleton</td>
<td>2.05E-11</td>
<td>2.40</td>
<td>94</td>
<td>490</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0043292</td>
<td>contractile fiber</td>
<td>6.98E-09</td>
<td>5.85</td>
<td>22</td>
<td>58</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0035583</td>
<td>myofibril</td>
<td>1.96E-08</td>
<td>5.74</td>
<td>21</td>
<td>56</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0043292</td>
<td>contractile fiber part</td>
<td>2.58E-08</td>
<td>5.97</td>
<td>20</td>
<td>52</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>sarcomere</td>
<td>4.95E-08</td>
<td>6.04</td>
<td>19</td>
<td>49</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>cytoskeletal protein binding</td>
<td>3.69E-07</td>
<td>2.52</td>
<td>47</td>
<td>227</td>
<td>MF</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>skeletal muscle development</td>
<td>2.50E-06</td>
<td>4.13</td>
<td>20</td>
<td>65</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>actin cytoskeleton</td>
<td>4.73E-06</td>
<td>3.08</td>
<td>27</td>
<td>111</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0016071</td>
<td>actin binding</td>
<td>7.52E-06</td>
<td>2.59</td>
<td>34</td>
<td>159</td>
<td>MF</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>muscle contraction</td>
<td>1.93E-05</td>
<td>4.22</td>
<td>16</td>
<td>51</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>cytoskeletal part</td>
<td>2.23E-05</td>
<td>2.03</td>
<td>51</td>
<td>294</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>I band</td>
<td>2.27E-05</td>
<td>5.67</td>
<td>12</td>
<td>32</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>muscle system process</td>
<td>2.54E-05</td>
<td>4.11</td>
<td>16</td>
<td>52</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>actin filament-based process</td>
<td>2.89E-05</td>
<td>2.73</td>
<td>27</td>
<td>119</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>muscle development</td>
<td>5.06E-05</td>
<td>2.69</td>
<td>26</td>
<td>116</td>
<td>BP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOID</th>
<th>Term</th>
<th>P Value</th>
<th>ORᵃ</th>
<th>Countᵇ</th>
<th>Sizeᶜ</th>
<th>Ontᵈ</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0043292</td>
<td>extracellular region part</td>
<td>4.59E-09</td>
<td>3.43</td>
<td>37</td>
<td>229</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0003576</td>
<td>extracellular region</td>
<td>1.88E-08</td>
<td>2.54</td>
<td>56</td>
<td>457</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>enzyme linked receptor protein signaling path</td>
<td>7.88E-07</td>
<td>3.17</td>
<td>29</td>
<td>188</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0005857</td>
<td>extracellular space</td>
<td>1.70E-06</td>
<td>3.83</td>
<td>21</td>
<td>116</td>
<td>CC</td>
</tr>
</tbody>
</table>
What else is it doing?
Another layer of regulation!

DNA wraps “histones”

Strong MyoD peaks correlate with up-shift in histone 4 acetylation, genome-wide

But no association with histone 3 lysine 4 tri-methylation (a control expt)
Summary

DNA sequencing providing floods of data
Many fundamental biological questions being attacked
Deep need for sophisticated computational analysis
One such story:

- MyoD present/bound in both myoblasts & myotubes
- Binds most genes, not just differentially expressed ones
- Significant genome-wide binding
- Broad chromatin modifications (histone H4 acetylation)
- Differentially bound peaks are (weakly) associated with changed expression
- Motif discovery possible, including cofactors

Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming.
Cao, Yao, Sarkar, Lawrence, Sanchez, Parker, MacQuarrie, Davison, Morgan, Ruzzo, Gentleman, Tapscott. Developmental Cell. 2010 Apr 20;18(4):662-74.