Motivation

- Regulatory sequence involves complex hierarchical patterns that are difficult for existing computational methods to model
- Deep Learning techniques show great promise in this area [1-3] but are considered uninterpretable “Black Boxes”, limiting their usefulness for making biological discoveries

Our contributions

- Novel algorithm (DeepLIFT) for explaining predictions of a given deep learning model for particular input examples
- Novel algorithm (MoDISco) for extracting recurring patterns (motif discovery) using a deep learning model

Example problem

- Goal: learn key regulatory sequences governing hematopoiesis
- Approach:
  1. Experimentally identify biochemically active regions in different cell-lines during the hematopoiesis lineage
  2. Train deep learning model to predict activity from seq.
  3. Interpret the model to learn key regulatory sequences

Model architecture overview

Output: Accessible (+1) vs not accessible (0)

"Fully connected" layers incorporate all info together

Later layers build on patterns of previous layer

Learned pattern detectors

Input: DNA sequence represented as ones and zeros

Results (MoDISco)

Superior motif discovery for Nanog

**Positive set:** 5,473 reproducible Nanog peaks in H1-E esc from ENCODE

**Negative set:** 258,987 H1 ESC Dnase-seq peaks

**Visualizing individual pattern projections:** DeepLIFT (Alipanahi et al.)

<table>
<thead>
<tr>
<th>Motif Cluster</th>
<th>H3</th>
<th>H4</th>
<th>H2A</th>
<th>H2B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corresponding MoDISco motif</th>
<th>H3</th>
<th>H4</th>
<th>H2A</th>
<th>H2B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

4 Motif clusters identified by MoDISco:

Zic3

Oct4-Sox2-Nanog

Nanog

Sox2

Results of logistic regression model trained to predict Nanog binding using the top 3 motif hits, per motif, per region

**Fusion motif from subclustering:**

**Co-binding between Zic3 and Nanog?**

**Individual examples:**

**Protein-protein interaction:**

Method: DeepLIFT (Deep Learning Important Features)

Compute behaviour under "reference"

Use difference from reference to find importance scores

- Consider $i_1 < 0$, $i_2 > 0$. Using difference-from-reference, we see $h_3 < 0$ below its reference value DeepLIFT assigns an importance of $(-0.2/h_3) - (-0.3/i_1)$.
- By considering different orders for positive and negative terms, we can also improve assignment of importance scores:

  * **AND\/*min operation:**
    - Consider $i_1 > 0$, $i_2 < 0$.
    - Standard breakdown: $y = (i_1 < 0) \cdot (i_2 < 0) \cdot (-0.2 + (-0.3)) = 0.5$.
    - Average over both orders: $y = (i_1 < 0) \cdot (i_2 < 0) \cdot (-0.2 + (-0.3)) = 0.5$.

- **OR\/*/max operation:**
  - Consider $i_1 < 0$, $i_2 < 0$.
  - Standard breakdown: $y = (i_1 < 0) \cdot (i_2 < 0) \cdot (-0.2 - 0.3) = 0.5$.
  - Average over both orders: $y = (i_1 < 0) \cdot (i_2 < 0) \cdot (-0.2 - 0.3) = 0.5$.

- **Other possible breakdowns:**
  - $y = \max(0, i_1 / 2 + i_2 / 2)$

Proof-of-concept: morphing an "8" to a 3 or a 6

Deep learning model is trained to recognize handdrawn digits from the MNIST database. Pixels are ranked by difference of importance for original class (e.g., 8) and target class (e.g., 3 or 6) by different methods. Up to 20% of pixels more important to original class than target class erased.

Method: MoDISco (Motif Discovery from Importance Scores)