tradeSeq: TRAjectory-based Differential Expression for SEQuencing data

Talk given at the retreat of the Berkeley Stem Cell Center on 04/14

Hector Roux de Bézieux
Group in Biostatistics
Sandrine Dudoit’s lab
GitHub: HectorRDB
Website: http://hectorrdb.github.io
### Single-cell RNA-Seq

**Unmixing the smoothie**

<table>
<thead>
<tr>
<th>Bulk RNA - Seq</th>
<th>Single-cell RNA - Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Smoothie" /></td>
<td><img src="image2" alt="Smoothie" /></td>
</tr>
<tr>
<td>VS</td>
<td>VS</td>
</tr>
<tr>
<td><img src="image3" alt="Fruit Bowl" /></td>
<td><img src="image4" alt="Fruit Bowl" /></td>
</tr>
</tbody>
</table>
Dimensionality reduction

- Bone-marrow stem cells from the \texttt{monocle 3 vignette}
- 2660 cells and 3004 genes
- Now 2660 cells in two dimensions
- Many methods exist (PCA, t-SNE, UMAP)
Trajectory inference

Finding developmental paths

Each cell has a pseudotime, which measure how far along it is in the developmental process
cluster-based DE is artificial

Genes are now expressed in a continuous manner (since 2014)

Differential Expression is still cluster-based, i.e. discrete.
Trajectory-based DE

We developed tradeSeq, an algorithm that leverages the continuous nature of scRNA-Seq.

- Available as an R package on Github (statOmics/tradeSeq). Soon on Bioconductor.

- Modular tool that work with any dimensionality reduction and trajectory inference method.
Statistical model

Can accommodate
- Design matrix
- Different sequencing depth
- Weights
An investigation tool

### Differential Expression Tests

<table>
<thead>
<tr>
<th>Lineages</th>
<th>Within the orange lineage</th>
<th>Between the orange and blue lineages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>associationTest</td>
<td>startVsEndTest</td>
</tr>
<tr>
<td><img src="image1" alt="Lineage" /></td>
<td>DE</td>
<td>DE</td>
</tr>
<tr>
<td><img src="image2" alt="Lineage" /></td>
<td>Not DE</td>
<td>Not DE</td>
</tr>
<tr>
<td><img src="image3" alt="Lineage" /></td>
<td>DE</td>
<td>Not DE</td>
</tr>
<tr>
<td><img src="image4" alt="Lineage" /></td>
<td>DE</td>
<td>DE</td>
</tr>
<tr>
<td><img src="image5" alt="Lineage" /></td>
<td>DE</td>
<td>DE</td>
</tr>
<tr>
<td><img src="image6" alt="Lineage" /></td>
<td>DE</td>
<td>DE</td>
</tr>
</tbody>
</table>
Outperforms existing methods
Provides unique insights

Gene Irf8 in the bone marrow dataset
Also works with multiple lineages
Take-aways

tradeSeq is

- An open-source R package to investigate differential expression along and between trajectories.

- A modular tool that work with any dimensionality reduction and trajectory inference method.

- An investigation tool that can answer various biologically relevant questions and provide unique and novel insights into complex datasets.
Thank you for listening

Any questions?