The SEG identification framework extracts a set of stability features from a Gamma-Gaussian mixture model (Ghazanfar et al., 2016) and also other gene characteristics, including:

1. Mixing proportion
2. Standard deviation
3. Zero proportion
4. F-statistics

and derives a stability index for each gene on the single-cell level. The higher stability index indicates the gene is more likely to be a stably expressed gene.

**Stably expressed genes (SEGs) identification and evaluation framework**

**Mixture model**

- **SEGs**
- **Gaussian**
- **Gene expression across cells**

**Additional scRNA-seq characteristics**

- **Evaluation metrics**
  - **ARI**
  - **Purity**
  - **FM**
  - **Jaccard**

**Mean-variance plot**

In comparison to the housekeeping genes (HKG) defined previously using bulk transcriptomes, SEGs identified on the single-cell level have significantly smaller expression variances across individual cells.

**PCA**

The PCA plots generated from using SEGs show much less separation with respect to the developmental stages.

**Evaluation metrics**

- **Lower evaluation metrics**
- **Lower concordance with cell types**
- **More stably expressed across cell types**

**Characterisation of stably expressed genes**

**Early human development (Petropoulos et al., 2016)**

Human SEGs

**Mouse SEGs**

**Early mouse development (Deng et al., 2016)**

**Benchmark results using independent scRNA-seq datasets**

<table>
<thead>
<tr>
<th>ID</th>
<th>Publication</th>
<th>Description</th>
<th>Organism</th>
<th>f cell</th>
<th>f class</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS685270</td>
<td>Villain et al. (2016)</td>
<td>Peripheral blood mononuclear cells</td>
<td>Human</td>
<td>1104</td>
<td>5</td>
</tr>
<tr>
<td>GS67548</td>
<td>Chiu et al. (2016)</td>
<td>Pluripotent stem cells and endoderm progenitors</td>
<td>Human</td>
<td>1018</td>
<td>2</td>
</tr>
<tr>
<td>GS70295</td>
<td>Trosh et al. (2016)</td>
<td>Multicellular metastatic melanoma</td>
<td>Human</td>
<td>4645</td>
<td>7</td>
</tr>
<tr>
<td>GS660357</td>
<td>Damanskis et al. (2015)</td>
<td>Adult and fetal brain</td>
<td>Human</td>
<td>466</td>
<td>8</td>
</tr>
<tr>
<td>GS683911</td>
<td>Zelis et al. (2015)</td>
<td>Mouse cortex and hippocampus</td>
<td>Mouse</td>
<td>3005</td>
<td>7</td>
</tr>
<tr>
<td>GS655253</td>
<td>Treutlein et al. (2014)</td>
<td>Developmental lung epithelial cells</td>
<td>Mouse</td>
<td>198</td>
<td>4</td>
</tr>
<tr>
<td>GS661521</td>
<td>Scialdone et al. (2014)</td>
<td>Mesoderm diversification</td>
<td>Mouse</td>
<td>1205</td>
<td>4</td>
</tr>
<tr>
<td>GS681333</td>
<td>Baró et al. (2016)</td>
<td>Pancreatic inter- and intra-cells</td>
<td>Mouse</td>
<td>822</td>
<td>13</td>
</tr>
</tbody>
</table>

**Organism**: Human, Mouse

<table>
<thead>
<tr>
<th>SEGs</th>
<th>aHK microarray</th>
<th>bHK microarray</th>
<th>cHKG microarray</th>
<th>dHKG microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARI</td>
<td>55±8</td>
<td>42±3</td>
<td>38±4</td>
<td>48±5</td>
</tr>
<tr>
<td>FM</td>
<td>56±1</td>
<td>51±3</td>
<td>52±3</td>
<td>51±3</td>
</tr>
<tr>
<td>FM</td>
<td>49±1</td>
<td>39±1</td>
<td>53±1</td>
<td>48±1</td>
</tr>
</tbody>
</table>
| K-means clustering outputs using SEGs derived from scRNA-seq data showed the lowest concordance to their pre-defined class labels.**

**Conclusion**


Reference:
Contact: yingxin.lin@sydney.edu.au pengyi.yang@sydney.edu.au jian.yang@sydney.edu.au