Classification Settings

- Classification rule: LDA
- Total number of features after t-test: 5
- SAGE-tags range: 50K to 100K
- SAGE-tags and NGS-reads noise COV: 0.1

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature size $D$</td>
<td>20000</td>
</tr>
<tr>
<td>Training sample size $n$</td>
<td>60, 120, 180</td>
</tr>
<tr>
<td>Test sample size $n_t$</td>
<td>3000</td>
</tr>
<tr>
<td>Class 0 $(m_0, \sigma_0)$</td>
<td>(0.0, 0.6)</td>
</tr>
<tr>
<td>Class 1 $(m_1, \sigma_1)$</td>
<td>(1.0, 0.6)</td>
</tr>
<tr>
<td>Correlation $\rho$</td>
<td>0.4</td>
</tr>
<tr>
<td>Block size $l$</td>
<td>5</td>
</tr>
<tr>
<td>Global markers $D_{gm}$</td>
<td>10</td>
</tr>
<tr>
<td>Subclasses $c$</td>
<td>2</td>
</tr>
<tr>
<td>Heterogenous markers per subclass $D_{hm}$</td>
<td>50</td>
</tr>
<tr>
<td>High-variance non-markers $D_{hv}$</td>
<td>2000</td>
</tr>
<tr>
<td>Low-variance non-markers $D_{lv}$</td>
<td>17890</td>
</tr>
</tbody>
</table>

NGS-reads Settings

- **Group 1**
  - 1K to 50K
  - 250K to 300K
  - 500K to 550K
- **Group 2**
  - 5M to 5M+50K
  - 10M to 10M+50K
  - 15M to 15M+50K
- **Group 3**
  - 25M to 25M+50K
  - 32.5M to 32.5M+50K
  - 40M to 40M+50K
- **Group 4**
  - 50M to 50M+50K
  - 75M to 75M+50K
  - 100M to 100M+50K
Figure 1: Group 1: (a) With feature selection, $\sum_j X_{i,j} \in (1K, 50K)$ for NGS-reads; (b) Without feature selection, $\sum_j X_{i,j} \in (1K, 50K)$ for NGS-reads; (c) With feature selection, $\sum_j X_{i,j} \in (250K, 300K)$ for NGS-reads; (d) Without feature selection, $\sum_j X_{i,j} \in (250K, 300K)$ for NGS-reads; (e) With feature selection, $\sum_j X_{i,j} \in (500K, 550K)$ for NGS-reads; (f) Without feature selection, $\sum_j X_{i,j} \in (500K, 550K)$ for NGS-reads.
Figure 2: Group 2: (a) With feature selection, $\sum_j X_{i,j} \in (5M, 5M + 50K)$ for NGS-reads; (b) Without feature selection, $\sum_j X_{i,j} \in (5M, 5M + 50K)$ for NGS-reads; (c) With feature selection, $\sum_j X_{i,j} \in (10M, 10M + 50K)$ for NGS-reads; (d) Without feature selection, $\sum_j X_{i,j} \in (10M, 10M + 50K)$ for NGS-reads; (e) With feature selection, $\sum_j X_{i,j} \in (15M, 15M + 50K)$ for NGS-reads; (f) Without feature selection, $\sum_j X_{i,j} \in (15M, 15M + 50K)$ for NGS-reads.
Figure 3: Group 3: (a) With feature selection, $\sum_j X_{i,j} \in (25M, 25M + 50K)$ for NGS-reads; (b) Without feature selection, $\sum_j X_{i,j} \in (25M, 25M + 50K)$ for NGS-reads; (c) With feature selection, $\sum_j X_{i,j} \in (32.5M, 32.5M + 50K)$ for NGS-reads; (d) Without feature selection, $\sum_j X_{i,j} \in (32.5M, 32.5M + 50K)$ for NGS-reads; (e) With feature selection, $\sum_j X_{i,j} \in (40M, 40M + 50K)$ for NGS-reads; (f) Without feature selection, $\sum_j X_{i,j} \in (40M, 40M + 50K)$ for NGS-reads.
Figure 4: Group 4: (a) With feature selection, \( \sum_j X_{i,j} \in (50M, 50M + 50K) \) for NGS-reads; (b) Without feature selection, \( \sum_j X_{i,j} \in (50M, 50M + 50K) \) for NGS-reads; (c) With feature selection, \( \sum_j X_{i,j} \in (75M, 75M + 50K) \) for NGS-reads; (d) Without feature selection, \( \sum_j X_{i,j} \in (75M, 75M + 50K) \) for NGS-reads; (e) With feature selection, \( \sum_j X_{i,j} \in (100M, 100M + 50K) \) for NGS-reads; (f) Without feature selection, \( \sum_j X_{i,j} \in (100M, 100M + 50K) \) for NGS-reads.