|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | EPIG | EPIG-seq | SAMseq | Deseq |
| Data type | Continuous | Count level | Count level | Count level |
| Distribution assumption | Gaussian | Poisson | Poisson | Negative Binomial |
| Correlation measurement | Person's | Cys | N/A | N/A |
| Measurement or spreading | Variance | Dispersion | Dispersion | Dispersion |
| Magnitude measurement | LogRatio | Wilcoxon test | Wilcoxon test | FoldChange |
| Significant test | SignalToNoise | Variance-to-mean ratio (VMR) | NonParametrics | NegBinom test |
| Deliverable results | Co-expression pattern | Co-expression pattern | DEG | DEG |

Table 1. Comparison of methods

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pattern number | Number of genes | Group1 | Group2 | Group3 | Group4 |
| 1 | 200 | 1 | 1.5 | 2.5 | 4 |
| 2 | 200 | 1 | -1.5 | -2.5 | -4 |
| 3 | 200 | 1 | -1.5 | 4 | 2.5 |
| 4 | 200 | 1 | 4 | 2.5 | 1.5 |
| 5 | 200 | 1 | 4 | 4 | 4 |
| 6 | 19000 | 1 | 1 | 1 | 1 |

Table 2. Simulated RNA-seq data with 5 meaningful patterns

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Seeded |  |  |  |
| Extracted | Group\_A | Group\_B | Group\_C | Group\_D | Group\_E | Group\_F |
| Group\_A | 147 | 0 | 81 | 0 | 48 | 40 |
| Group\_B | 0 | 172 | 0 | 0 | 0 | 17 |
| Group\_C | 20 | 0 | 92 | 0 | 11 | 6 |
| Group\_D | 0 | 0 | 0 | 169 | 30 | 8 |
| Group\_E | 11 | 1 | 5 | 12 | 100 | 1 |

Table 3. EPIG-seq results on simulated RNA-seq data with 5 meaningful patterns (Group A – E)

|  |  |
| --- | --- |
| Pattern results compared | NMI |
| Exp1 vs. Exp2 | 0.197 |
| Exp1 vs. Exp3 | 0.464 |
| Exp1 vs. Exp4 | 0.373 |
| Exp2 vs. Exp3 | 0.461 |
| Exp2 vs. Exp4 | 0.332 |
| Exp3vs. Exp4 | 0.280 |

Table 4 Comparison of clusterings using EPIG-Seq on sampled TCGA breast cancer RNA-seq data