|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |
| TranscriptA | 0 | 16 | 1 | 0 | 1228 | 746 | 166 | 574 | 5207 | 25693 | 1018 | 2622 | 364608 | 58221 | 210304 | 239869 |
| TranscriptB | 36 | 328 | 190 | 178 | 36 | 4 | 0 | 0 | 2581 | 7068 | 298 | 17049 | 23 | 2542 | 799 | 931 |

Table 1. Count level measurement of two transcripts (genes) in four groups

|  |  |  |  |
| --- | --- | --- | --- |
|  | EPIG | EPIG-seq | ORIGEN |
| Data type | Continuous | Count level | Count level |
| Distribution assumption | Gaussian | Poisson | NonParametric |
| Correlation measurement | Person's | CYs | Order statistics |
| Measurement or spreading | Variance | Dispersion | Residue |
| Magnitude measurement | LogRatio | Wilcoxon test | Difference Between means |
| Significant test | SignalToNoise | Variance-to-mean ratio (VMR) | mdFDR means |
| Deliverable results | Co-expression pattern | Co-expression pattern | Co-expression pattern |

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 | 1.5 | 1.5 | 1.5 |

Table 2. Simulated RNA-seq data with 5 meaningful patterns, with their respective weight for a given pattern in a given group.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Extracted |  |  |  |
| Seeded | Group\_A | Group\_B | Group\_C | Group\_D | Group\_E | Sensitivity (%) |
| Group\_A | 132 | 11 | 7 | 0 | 0 | 88 |
| Group\_B | 0 | 169 | 0 | 0 | 0 | 100 |
| Group\_C | 26 | 5 | 135 | 0 | 0 | 81.33 |
| Group\_D | 0 | 0 | 0 | 166 | 15 | 91.71 |
| Group\_E | 23 | 0 | 9 | 43 | 111 | 59.68 |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample #** | **GS** | **MS** | **# of Patterns** | **# of Genes** |
| 1 | 0.31 | 0.54 | 6 | 192 |
| 2 | 0.37 | 0.51 | 4 | 169 |
| 3 | 0.21 | 0.52 | 6 | 344 |
| 4 | 0.41 | 0.59 | 4 | 197 |

Table 4. Clustering assessment on sampled BRCC data, GS: General Silhouette, MS: Maximum Silhouette

|  |  |  |  |
| --- | --- | --- | --- |
| Pattern # | # of Genes | Breast Cancer Related Biology | p-value |
| 1 | 29 | GO:0042803 - Protein homodimerization activity (S100A16, CENPF, APOE, PLOD1, TOP2A) | 2.20E-03 |
| 2 | 114 | Extra cellular matrix-receptor interaction | 2.8E-04 |
| 3 | 46 | Peroxisome proliferator-activated receptor signaling pathway | 6.0E-02 |
| 4 | 8 | CD59, ITGB1 and 5 ribosomal protein genes | **-** |

Table5A Pathway analysis on four significant patterns extracted from BRCC sampled data

|  |  |  |  |
| --- | --- | --- | --- |
| Pattern # | # of Genes | MOA Related Biology | p-value |
| 1 | 9 | GO:0006631 - Fatty acid metabolic process | 3.8E-06 |
| 2 | 10 | GO:0055114 - Oxidation reduction process | 2.3E-02 |
| 3 | 10 | GO:0042592 - Homeostatic process | 6.0E-02 |
| 4 | 4 | **-** | **-** |

Table5B Pathway analysis on four significant patterns extracted from SEQC MOA data