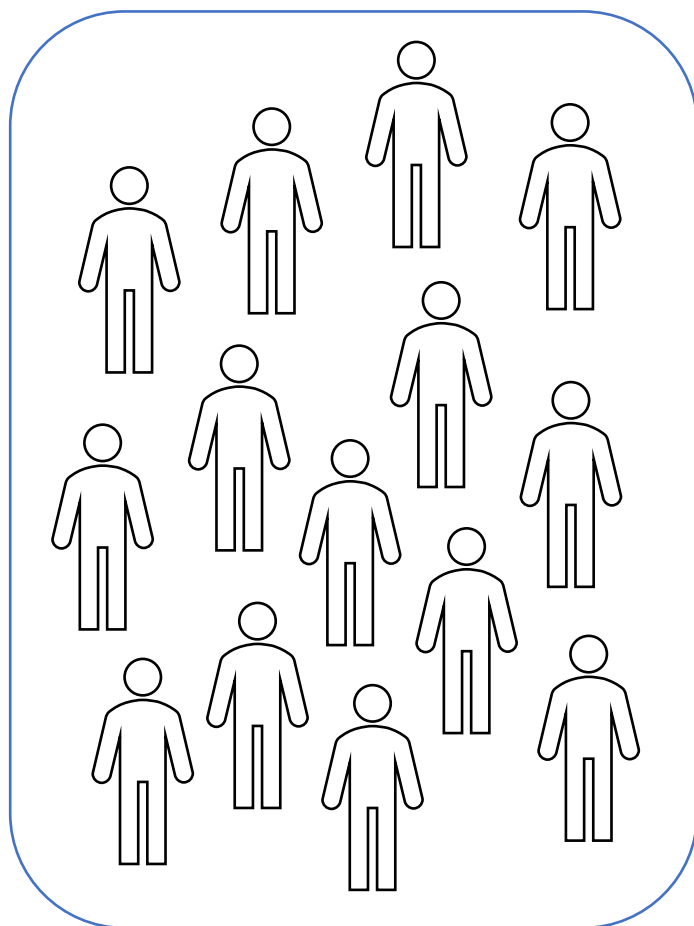
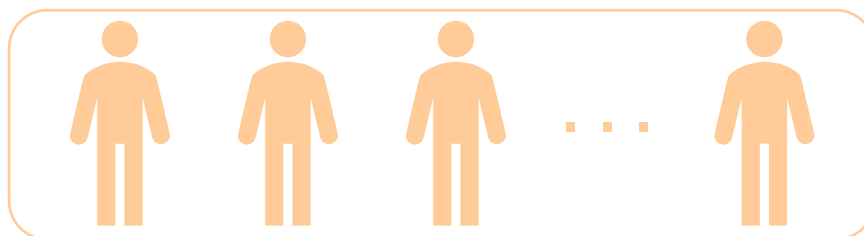
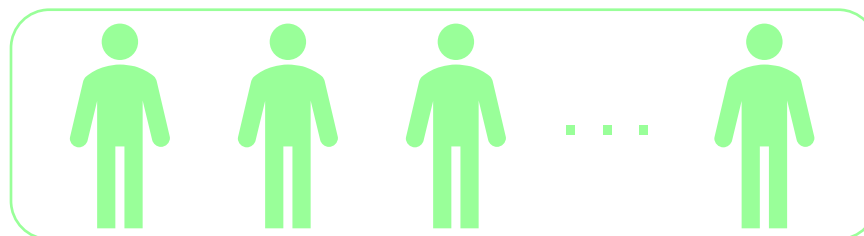
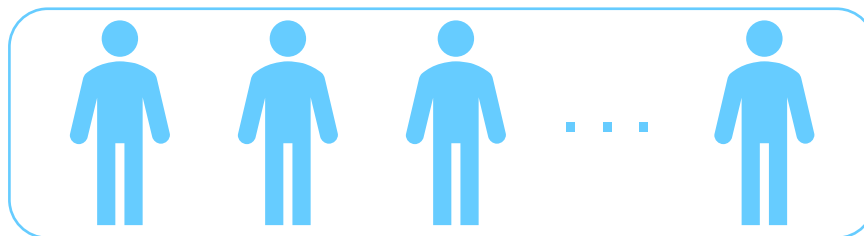
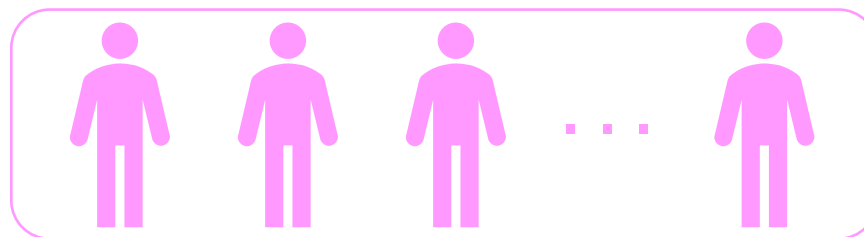


オミックスデータに基づいたサンプル分類と可視化

フィルジエン株式会社 バイオインフォマティクス部
(biosupport@filgen.jp)



サンプル集団



サンプルグループの分類

マーカー遺伝子の
探索など

Visualize and Explore

- データのクオリティチェック
- データ構造の可視化
- 新たな仮説立て

Statistical Analysis

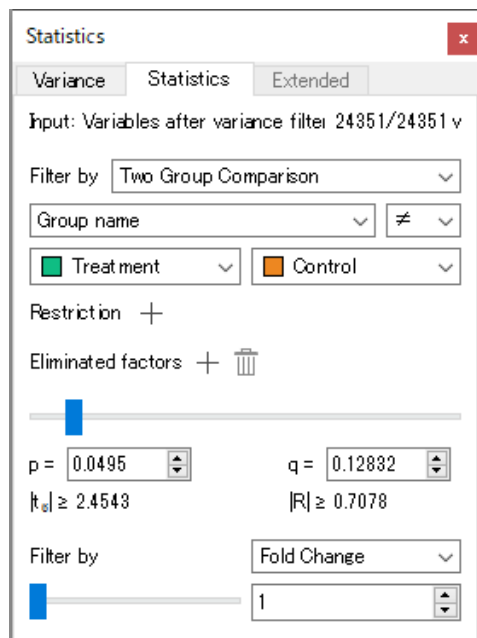
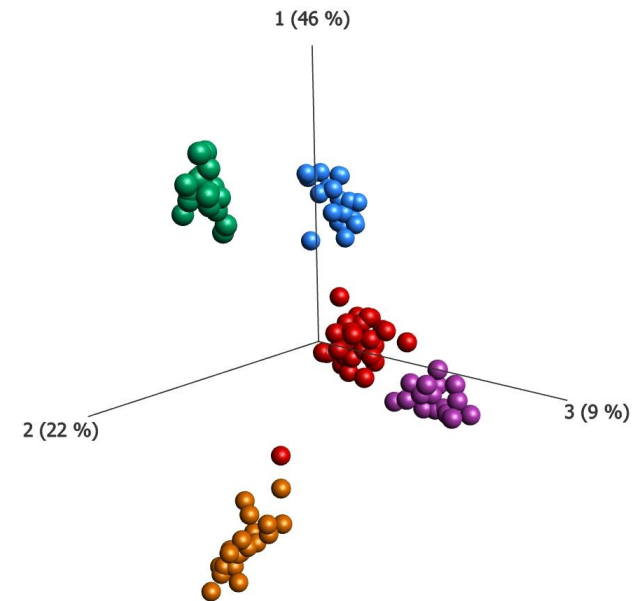
- t-test, ANOVA, 回帰分析
- Open API による統計メソッドのインテグレート
- 各種プロットやデータの作成と出力

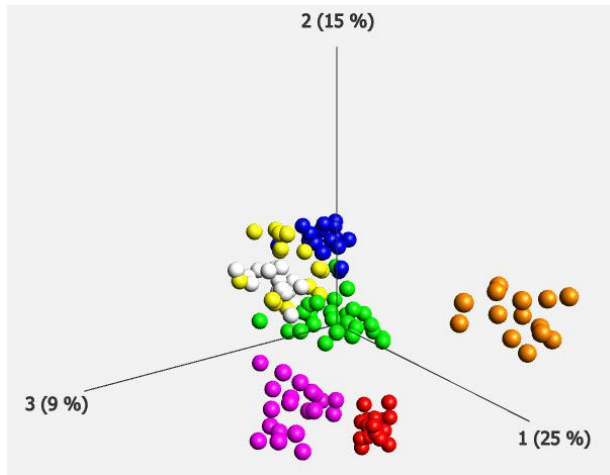
Achieve Biological Insight

- アノテーションの探索
- GO Browser
- GSEA – Gene Set Enrichment Analysis

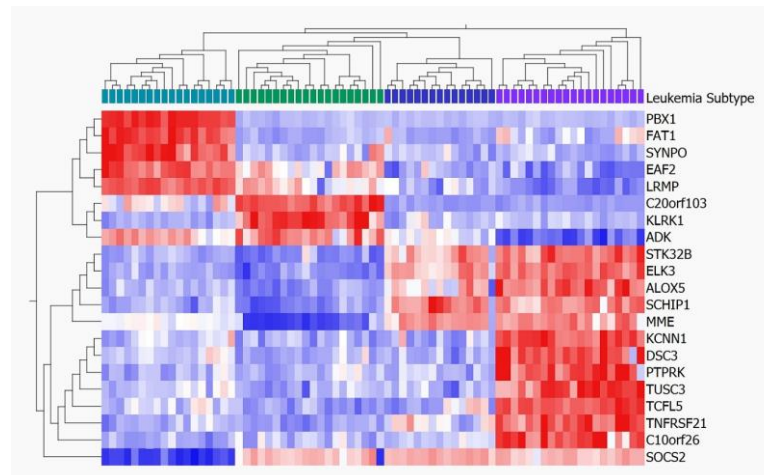
Machine learning

- データ分類モデルの構築
- kNN, SVM, RT, XG Boost
- 構築したモデルの新しいデータへの適用

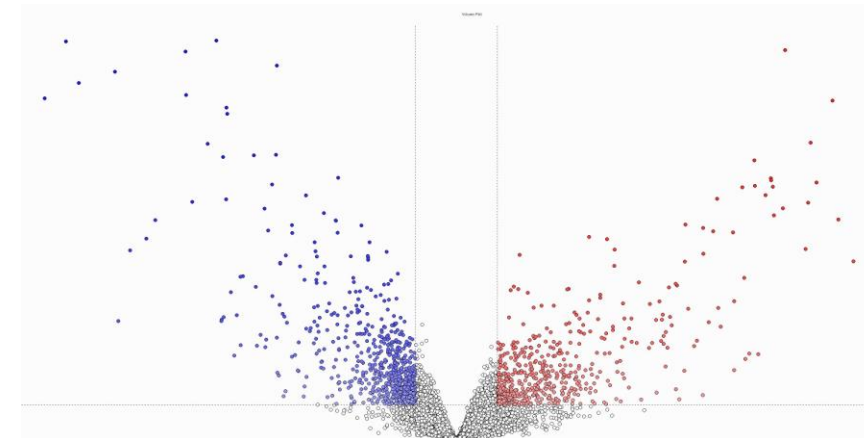




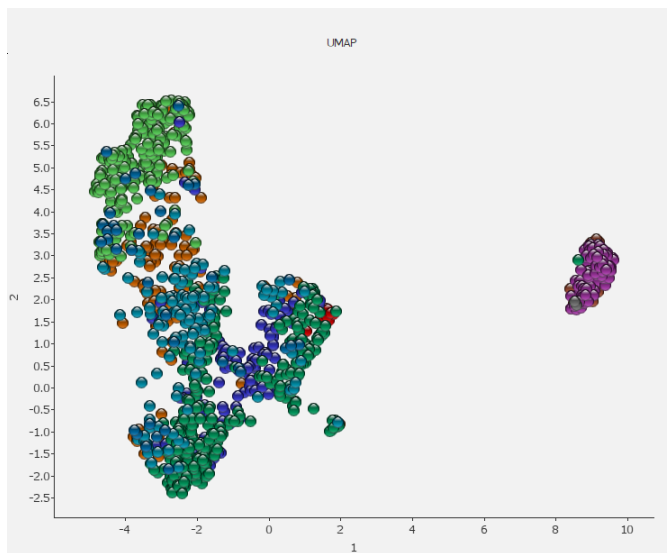
✓ 2D/3D PCA



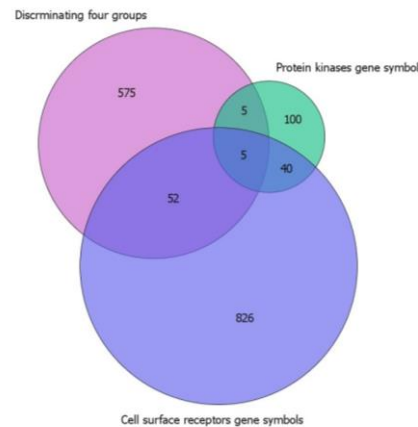
✓ ヒートマップ・階層型クラスタリング



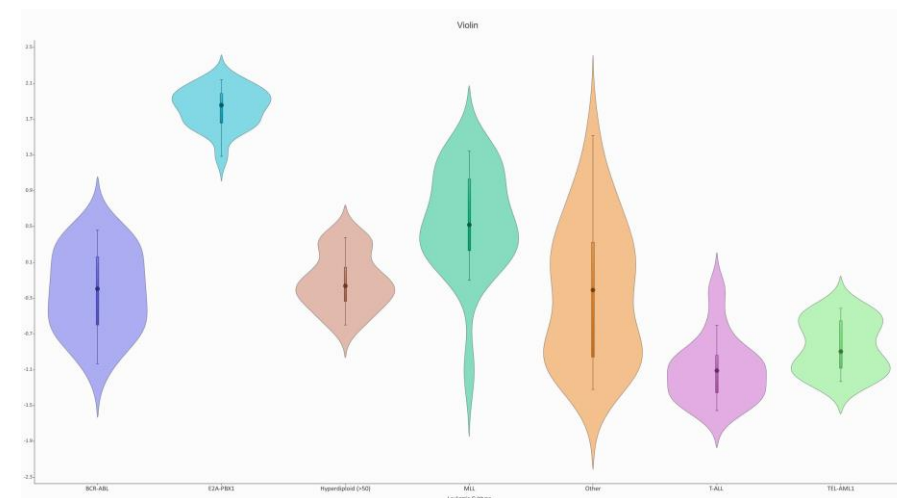
✓ ボルケーノプロット



✓ t-SNE・UMAP



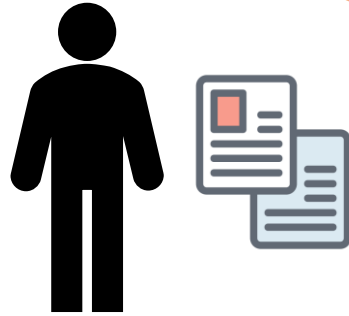
✓ ベン図



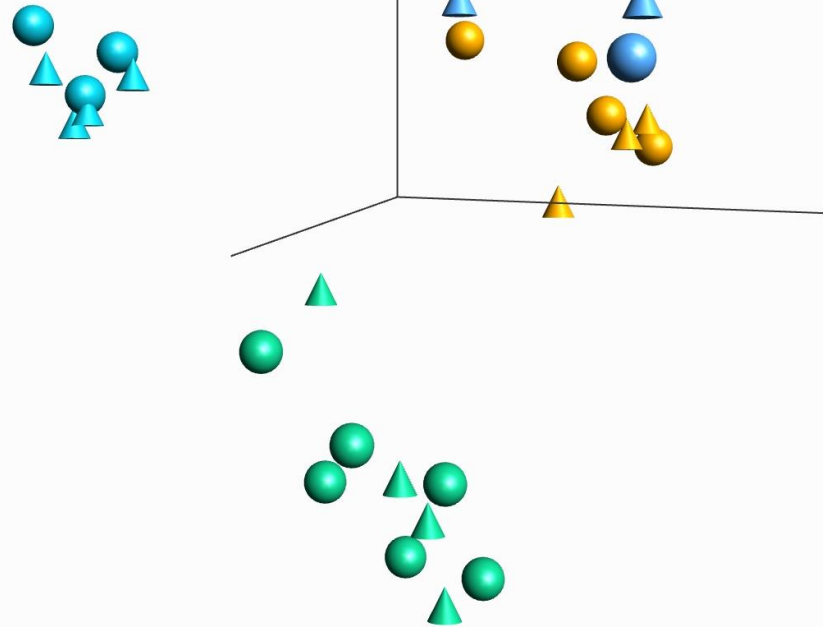
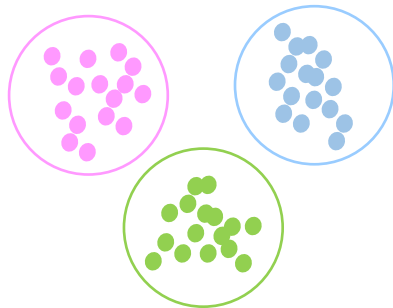
✓ バイオリンプロット

アノテーションや分類結果の可視化

既知のサンプルアノテーション



グラフプロットから手動で分類



クラスタリングによる自動分類

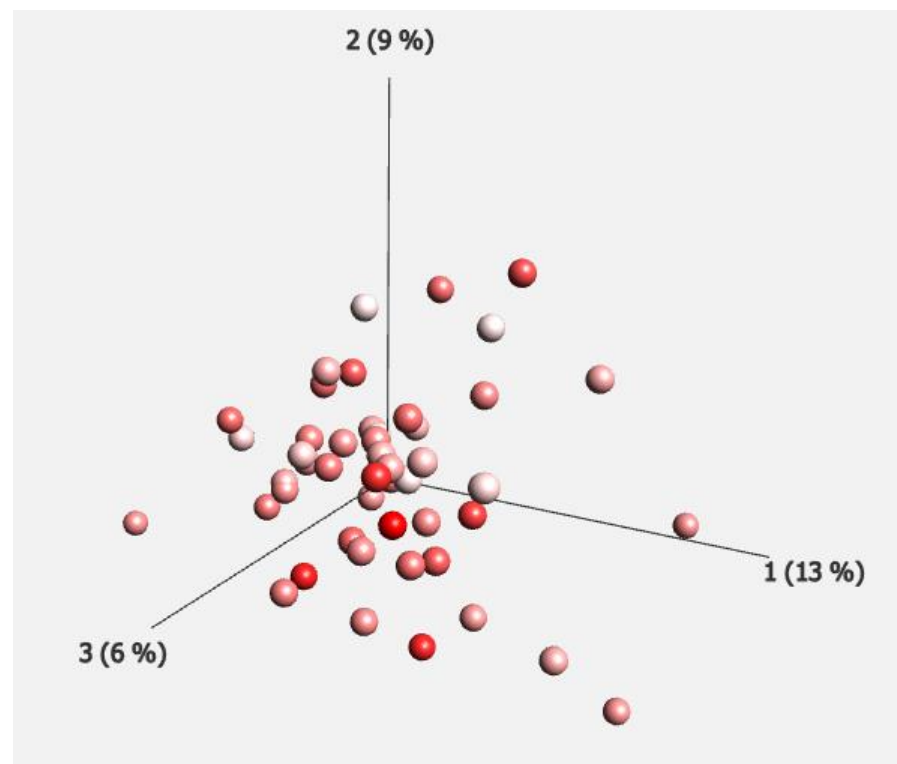
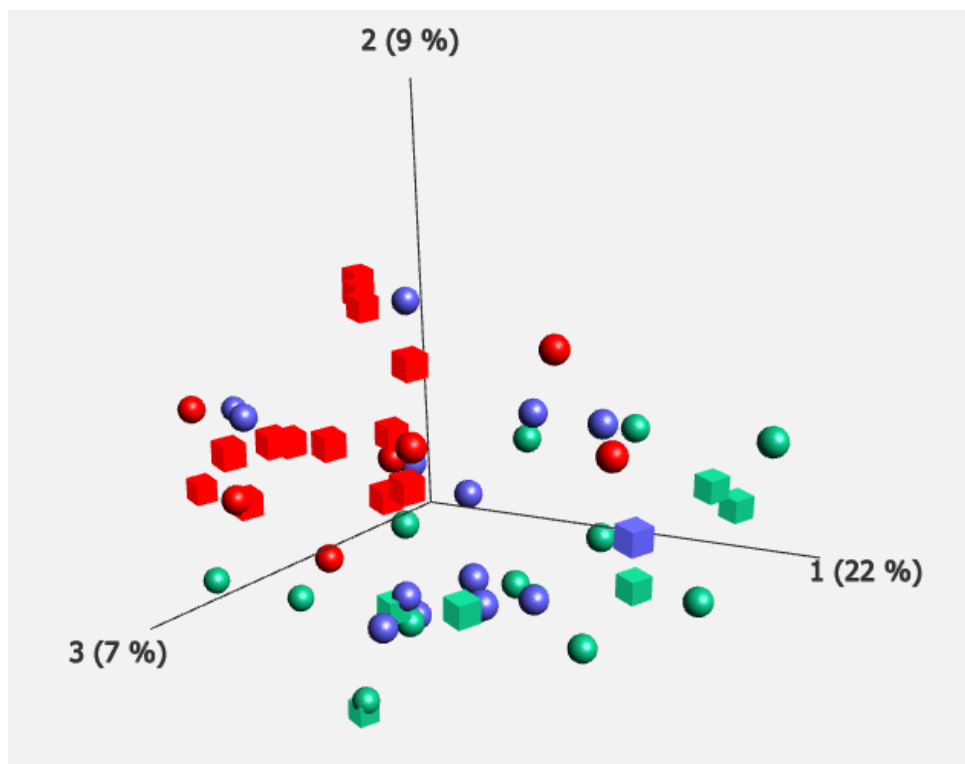


ユーザー入力のサンプルアノテーション

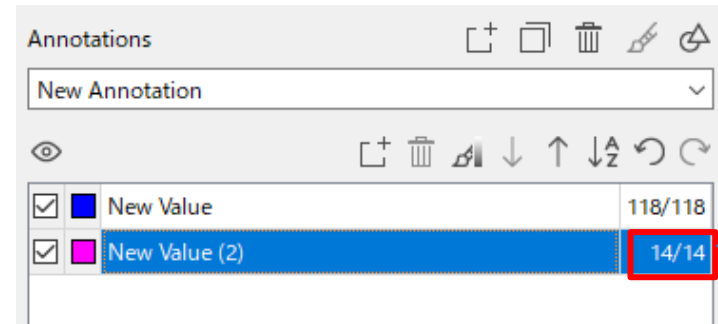
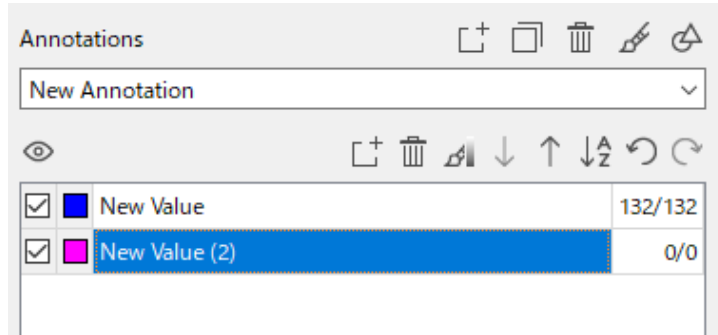
| 1 | Sample Name | Cancer Type | Gender | Age |
|----|-------------|-------------|--------|-----|
| 2 | Sample 01 | TEL-AML1 | Male | 51 |
| 3 | Sample 02 | TEL-AML1 | Male | 62 |
| 4 | Sample 03 | TEL-AML1 | Female | 39 |
| 5 | Sample 04 | BCR-ABL | Female | 36 |
| 6 | Sample 05 | MLL | Male | 49 |
| | | BCR-ABL | Female | 73 |
| 8 | Sample 07 | TEL-AML1 | Female | 54 |
| 9 | Sample 08 | BCR-ABL | Male | 62 |
| 10 | Sample 09 | BCR-ABL | Male | 38 |
| 11 | Sample 10 | MLL | Female | 27 |

ノードを色と形状に分けてプロット

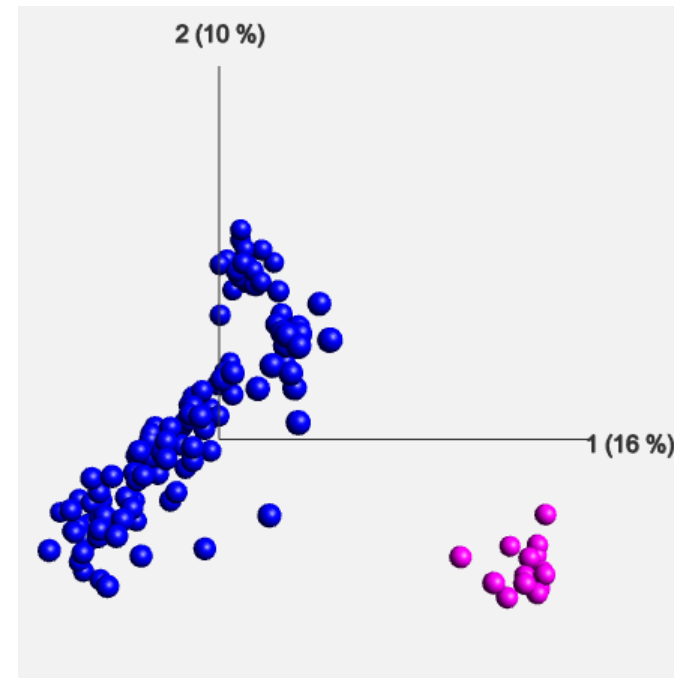
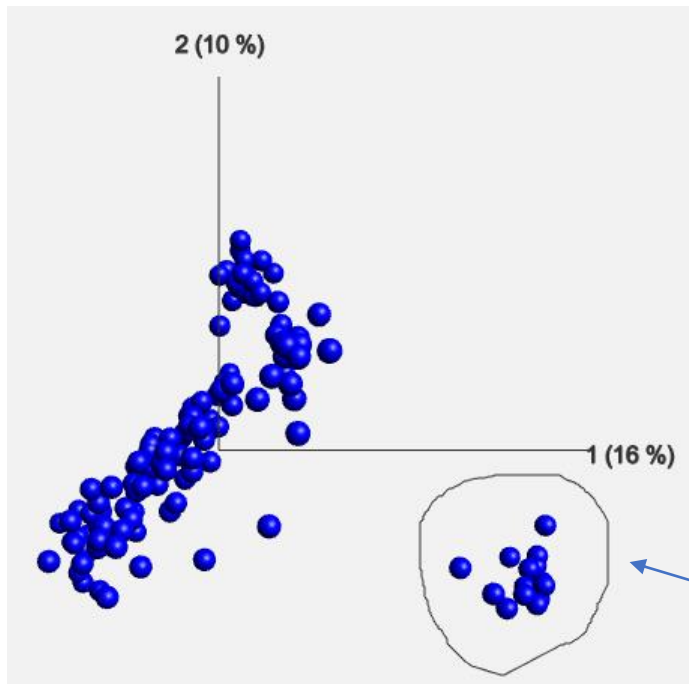
色のグラジエントをつけてプロット

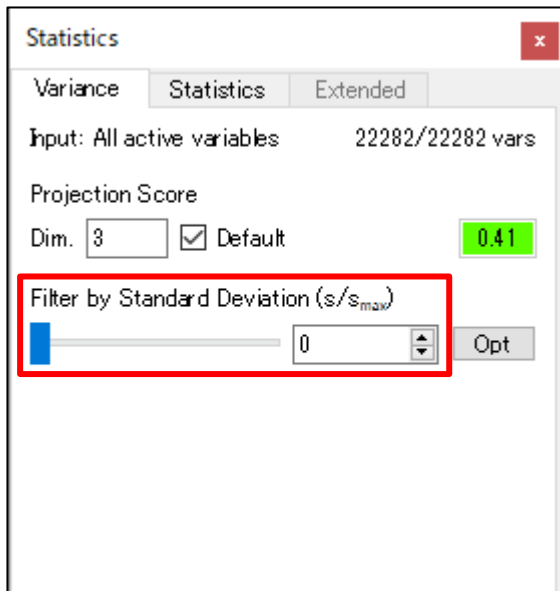


各グラフ上より任意のクラスターなどを手動でアノテーション



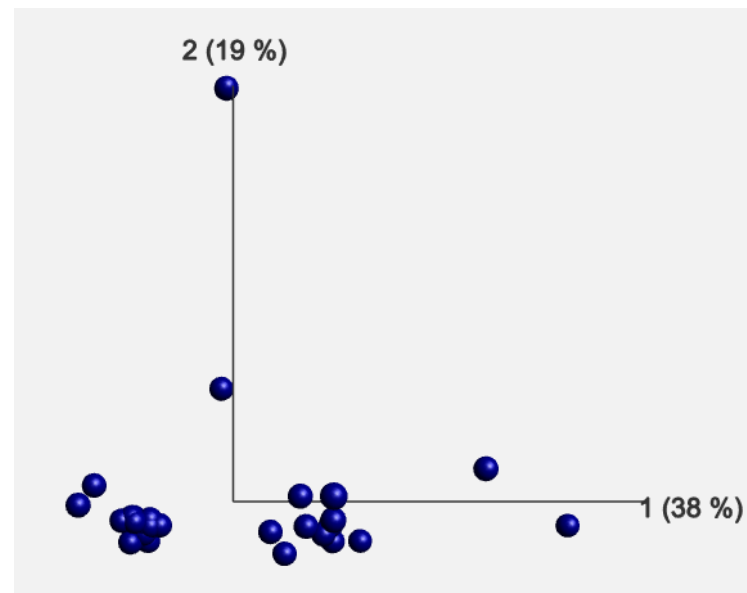
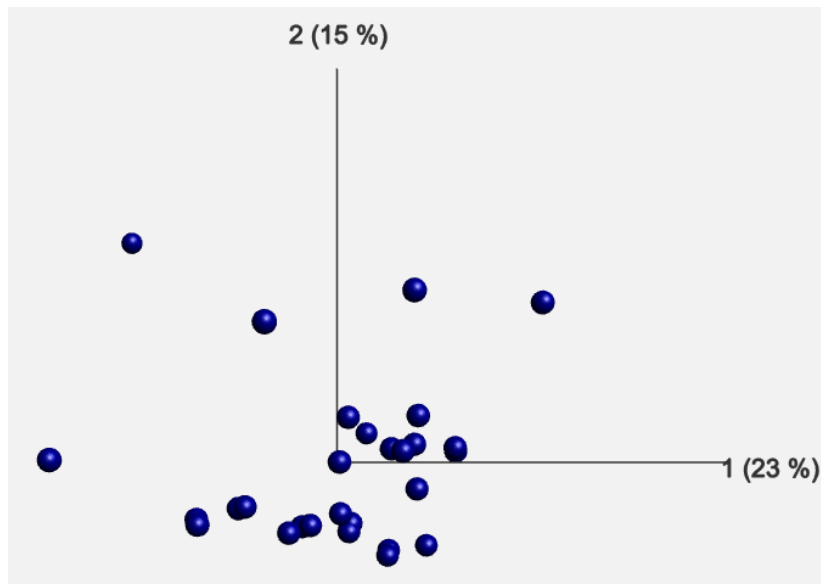
選択領域内のサンプルが
アノテーション分類される





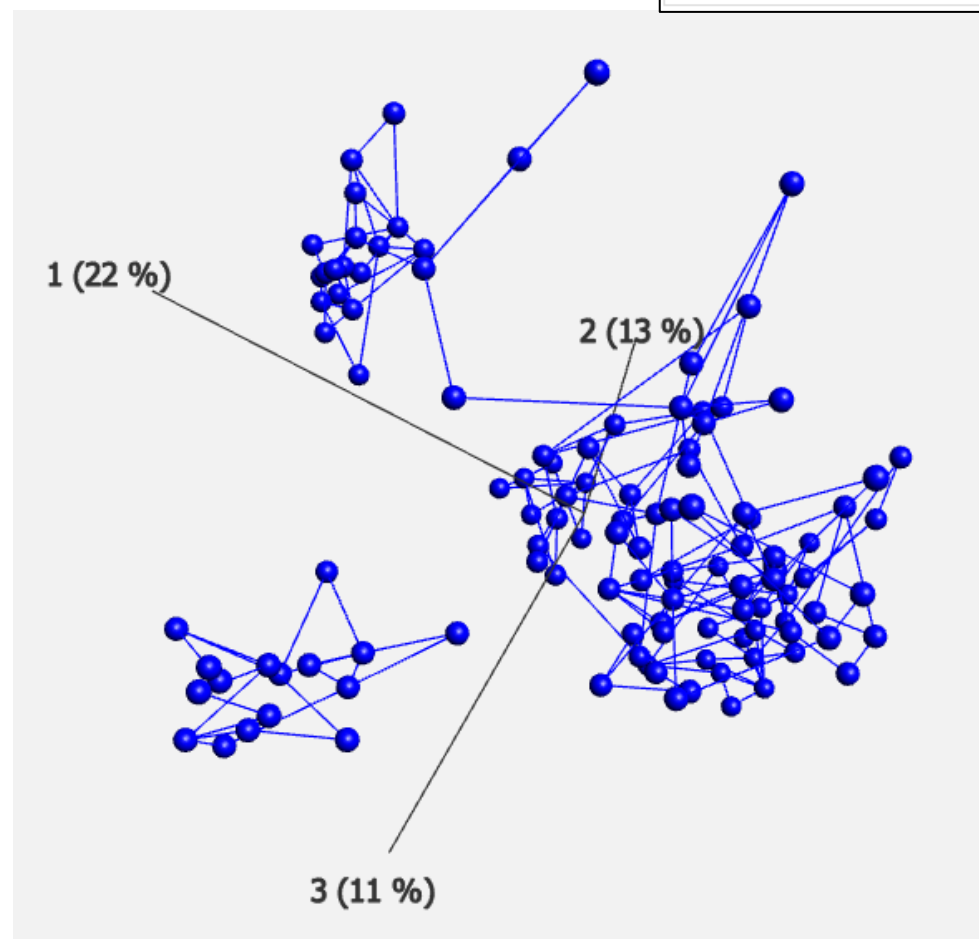
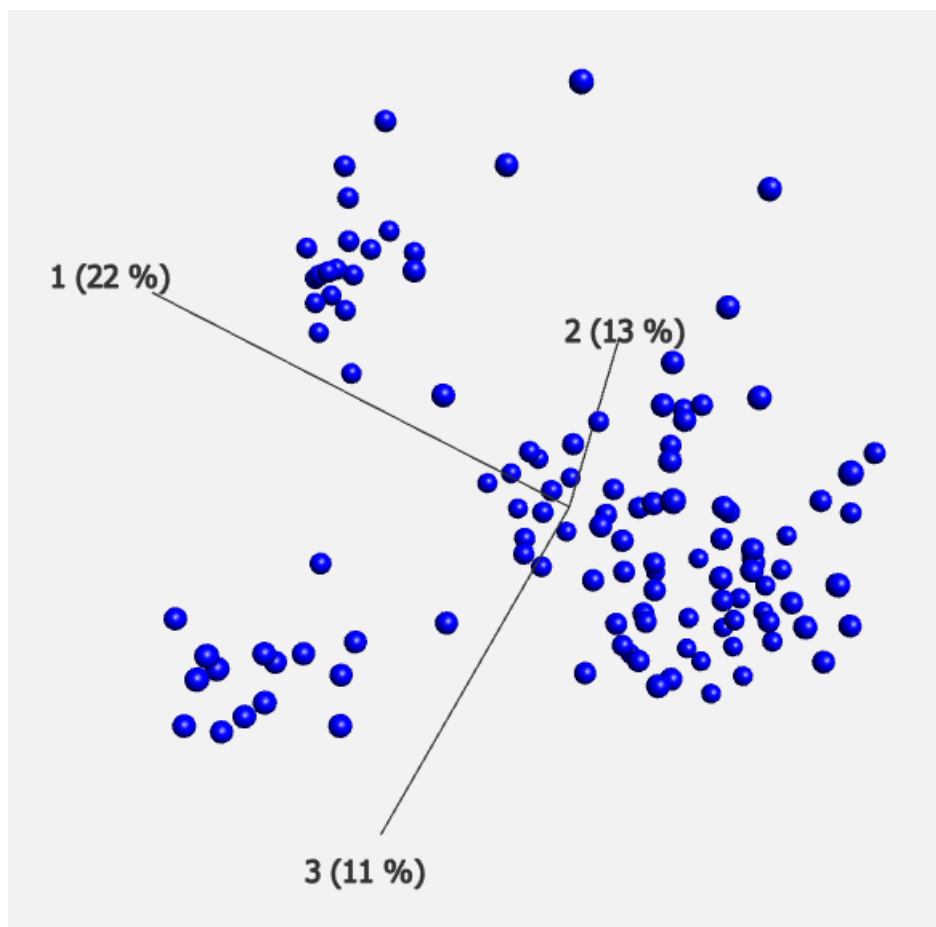
■ 分散フィルター

- サンプル間の発現変動のばらつきが小さい遺伝子をフィルタリングすることによって、特徴的なクラスターを抽出しやすくなる
- フィルタリングのスライダーバーを調整すると、PCAプロットなどのグラフがリアルタイムで変化するため、適切なフィルタリングの閾値を見つけやすい



■ ネットワーク

- PCAプロットにおいて、遺伝子発現プロファイルが似ているサンプル同士を線で繋げて表示することで、クラスターが識別しやすくなる
- 上位何サンプルまで繋げるかや、サンプル間の距離を指定することも可能



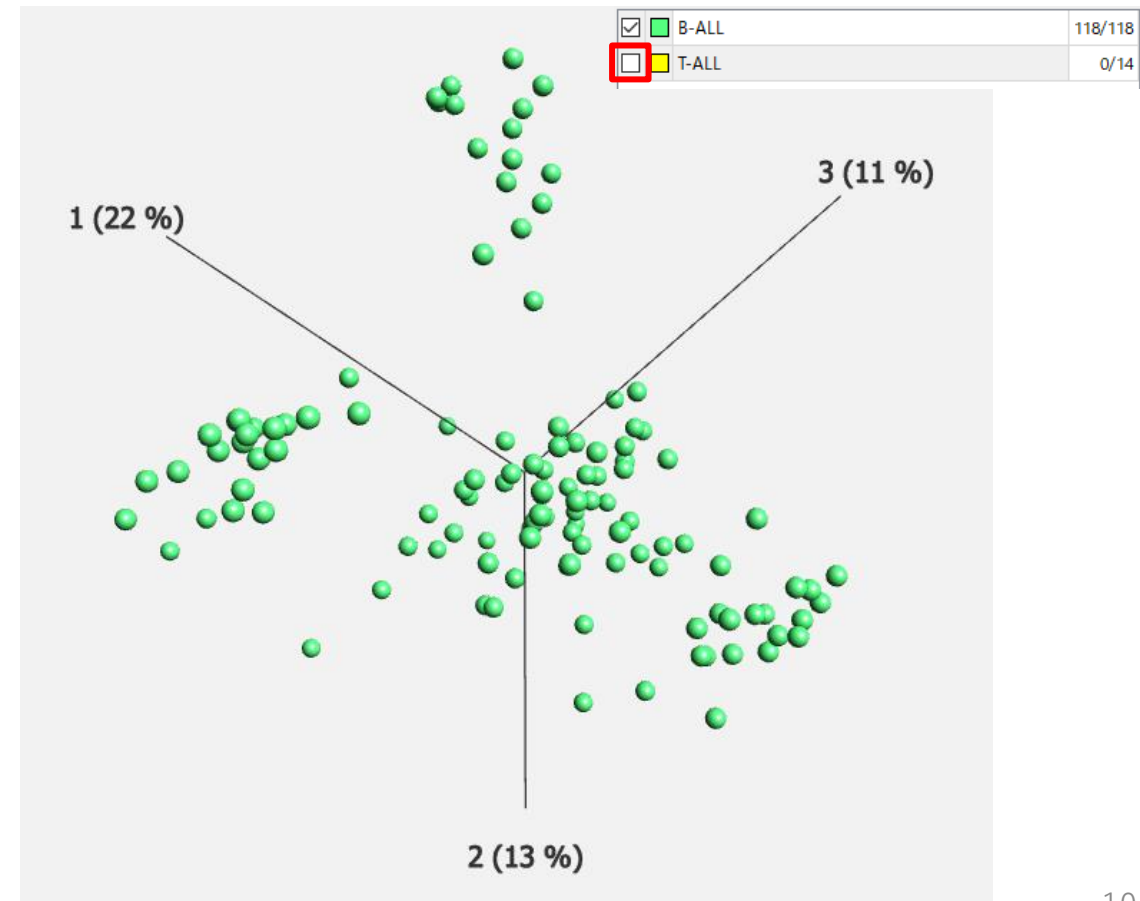
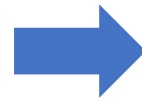
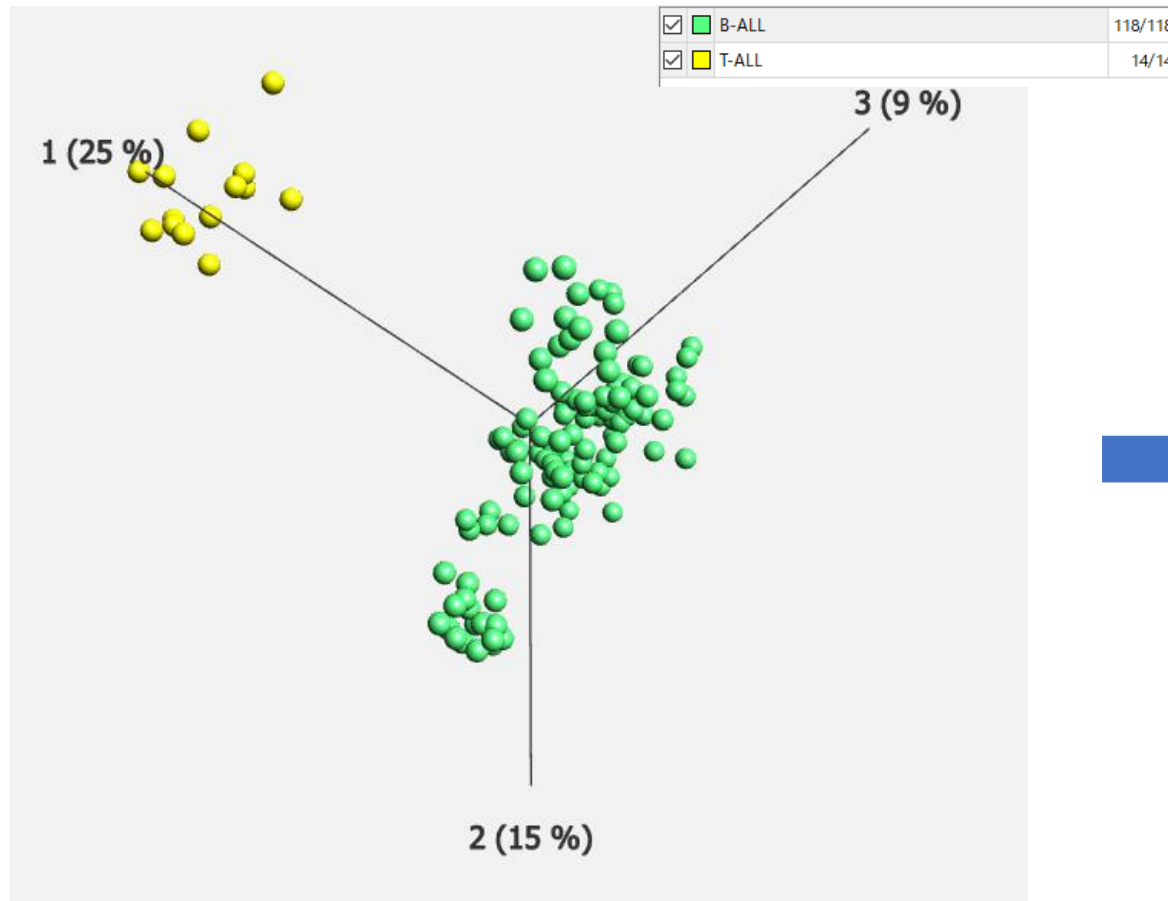
Network

2 Nearest Dir. Conn.

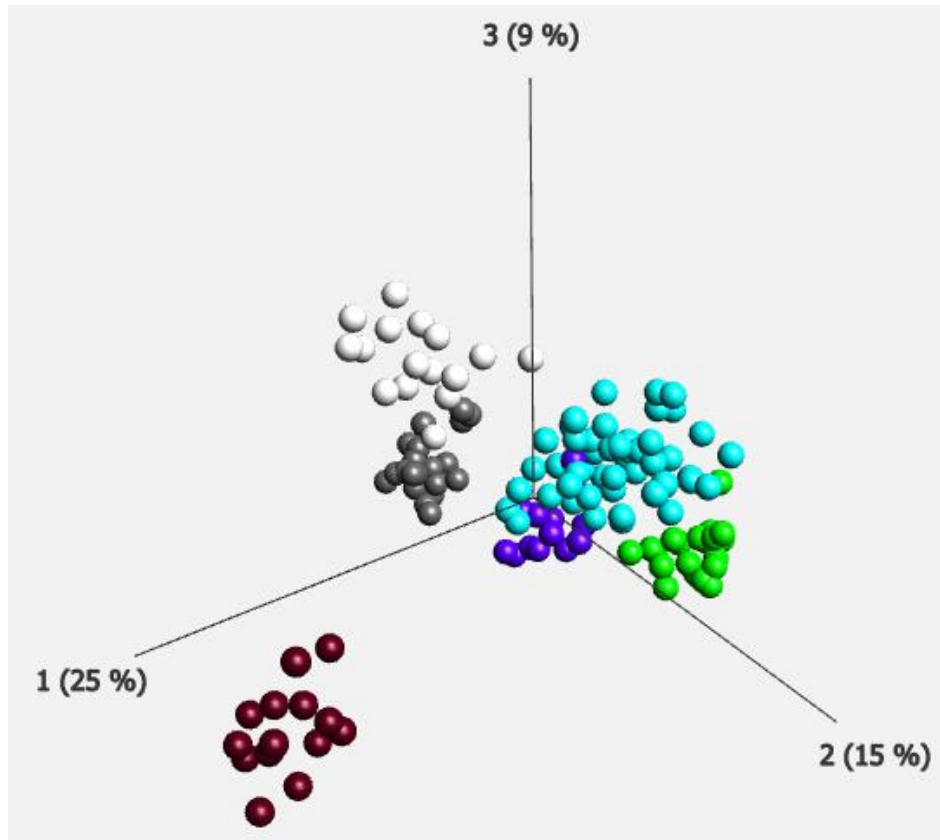
Dist. 0

■ アノテーション表示の切り替え

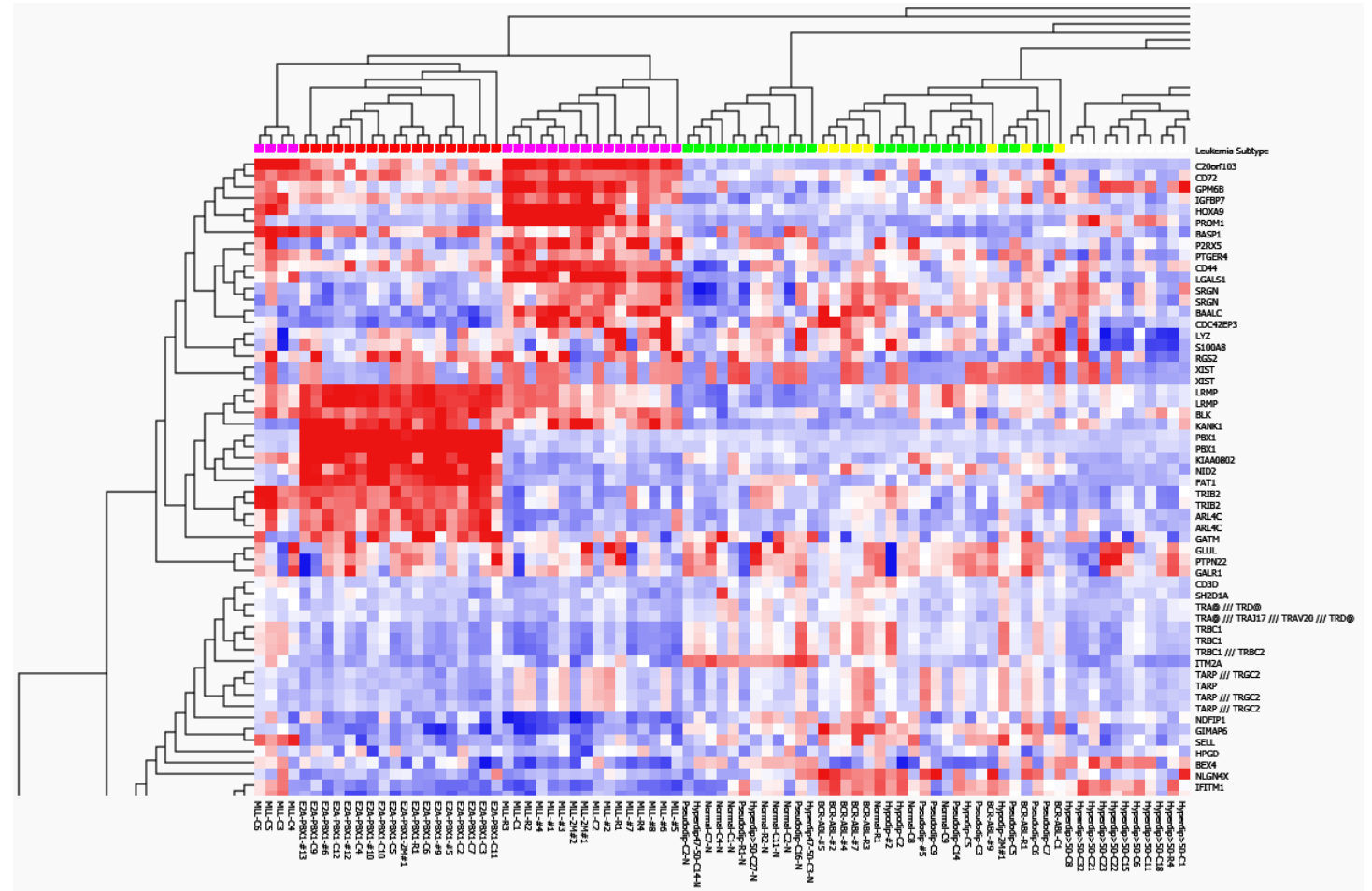
- アノテーションの表示／非表示をワンクリックで切り替えが可能で、これによりデータセットから任意のサンプルクラスターなどを削除し、プロットの再計算が可能
- 外れ値をもつサンプルなどを削除することで、グラフがクリアになる



クラスタリングによりサンプルを自動で分類

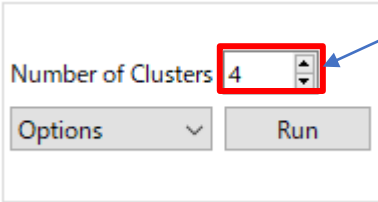


✓ K-meansクラスタリング

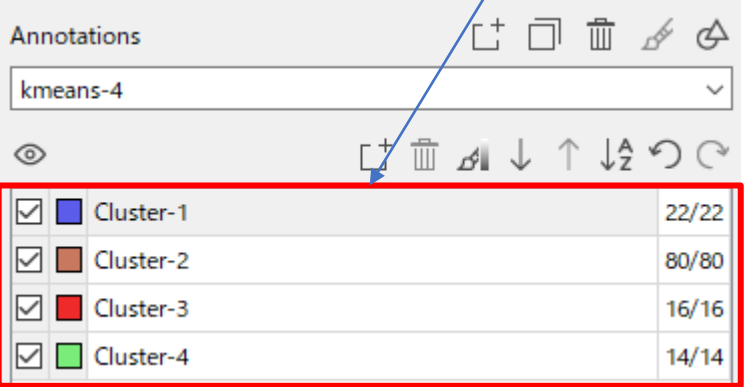


✓ 階層型クラスタリング

ユーザー指定のクラスター数

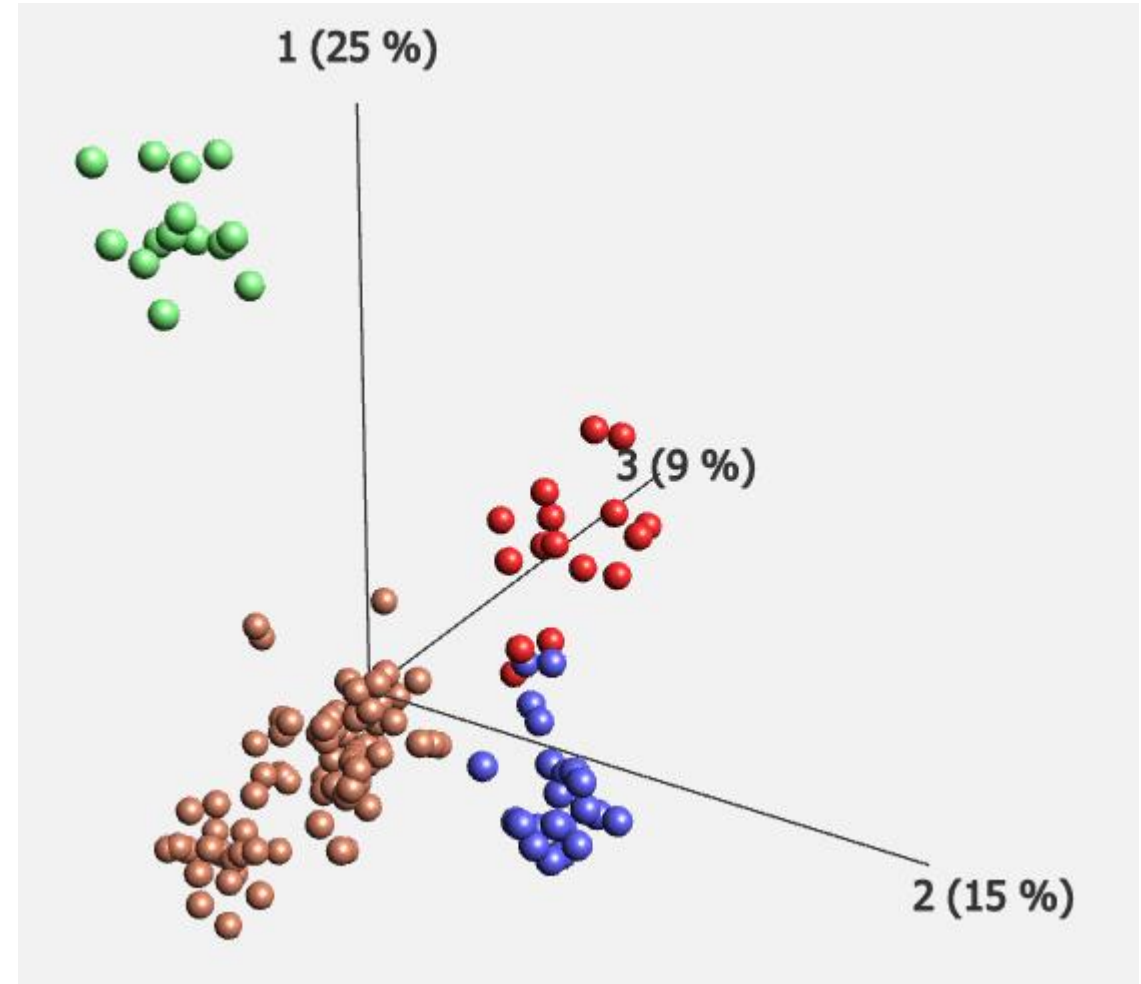


自動生成されたサンプルアノテーション



| Cluster | Count |
|-----------|-------|
| Cluster-1 | 22/22 |
| Cluster-2 | 80/80 |
| Cluster-3 | 16/16 |
| Cluster-4 | 14/14 |

グラフにプロット



■ K-meansクラスタリング

- データセットをユーザー指定のクラスター数に自動で分類
- クラスタリング結果はサンプルアノテーションとして保存され、各グラフに反映が可能

Data Method Options View Cluster

Plot

PCA Line Heat Genome

Scatter Bar Table Special

ヒートマップ表示の選択

+

Mark Order Size

None H. clustering Auto 10

None H. clustering Auto 10

Order Samples According To

The annotation

Order numerically

Hierarchical clustering

The value of variables for which equals

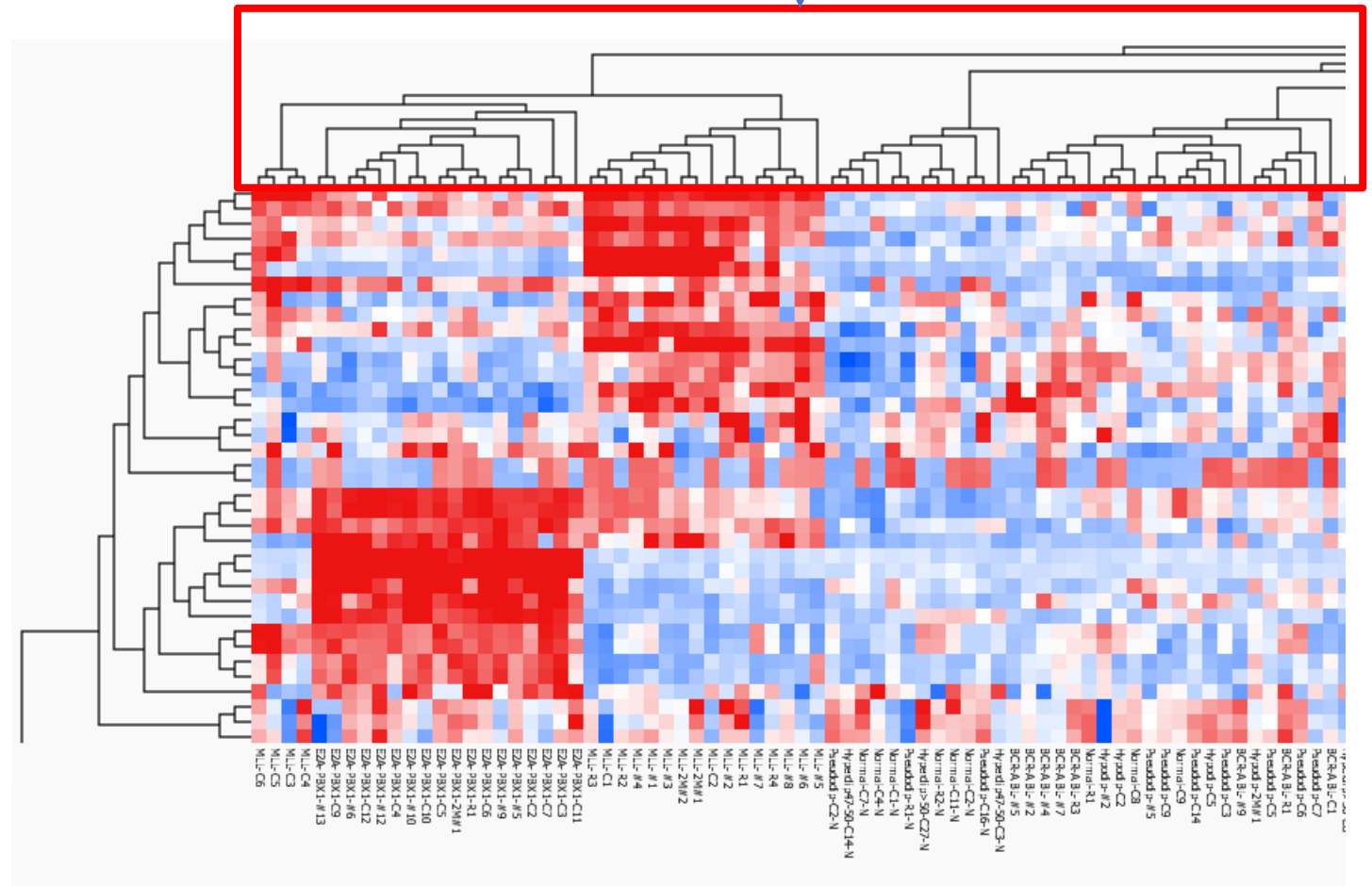
PCA component

Direction

Decreasing Increasing

サンプル表示の順番に Hierarchical clustering を選択

サンプルの分類結果



■ 階層型クラスタリング

- データセットを階層構造をもったクラスターに自動で分類
- サンプル・遺伝子の両方に対しての分類とヒートマップによる可視化が可能

分類結果のサンプルアノテーションを利用し、
発現変動遺伝子の抽出

Statistics ✕

Variance Statistics Extended

Input: Active vars (variance filtered) 207/22282

Filter by Two Group (t-test)

Leukemia Subtype ≠

BCR-ABL [All]

Restriction +

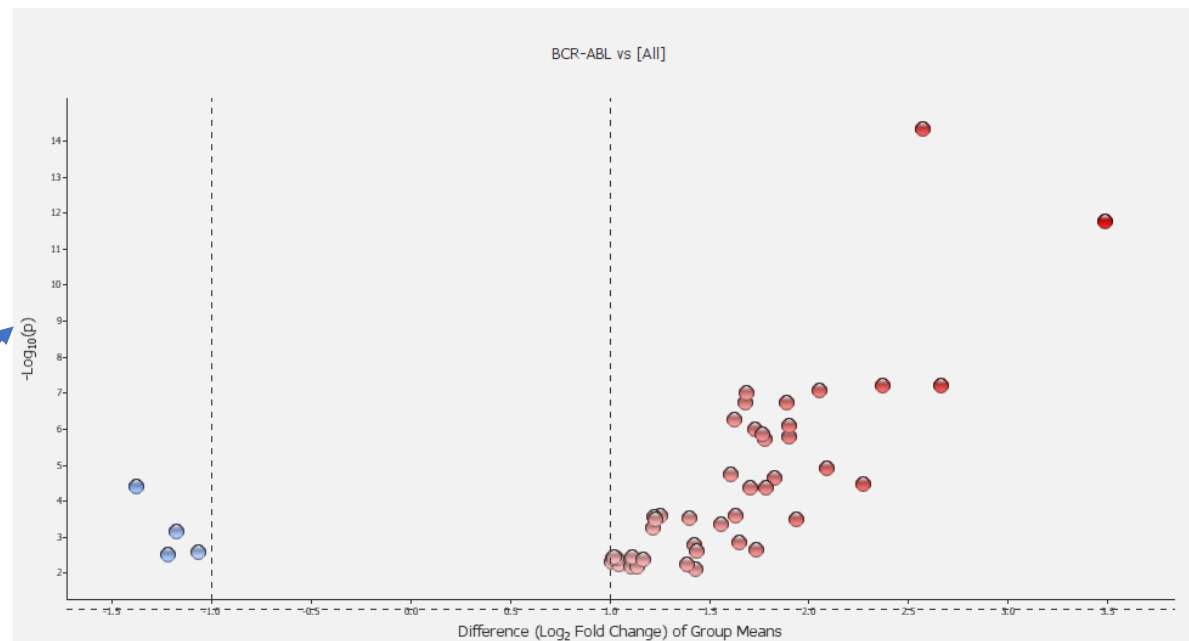
Eliminated factors + ✕

p = 0.01 q = 0.033422

$|t_{130}| \geq 2.6142$ $|R| \geq 0.22348$

Filter by Fold Change

2

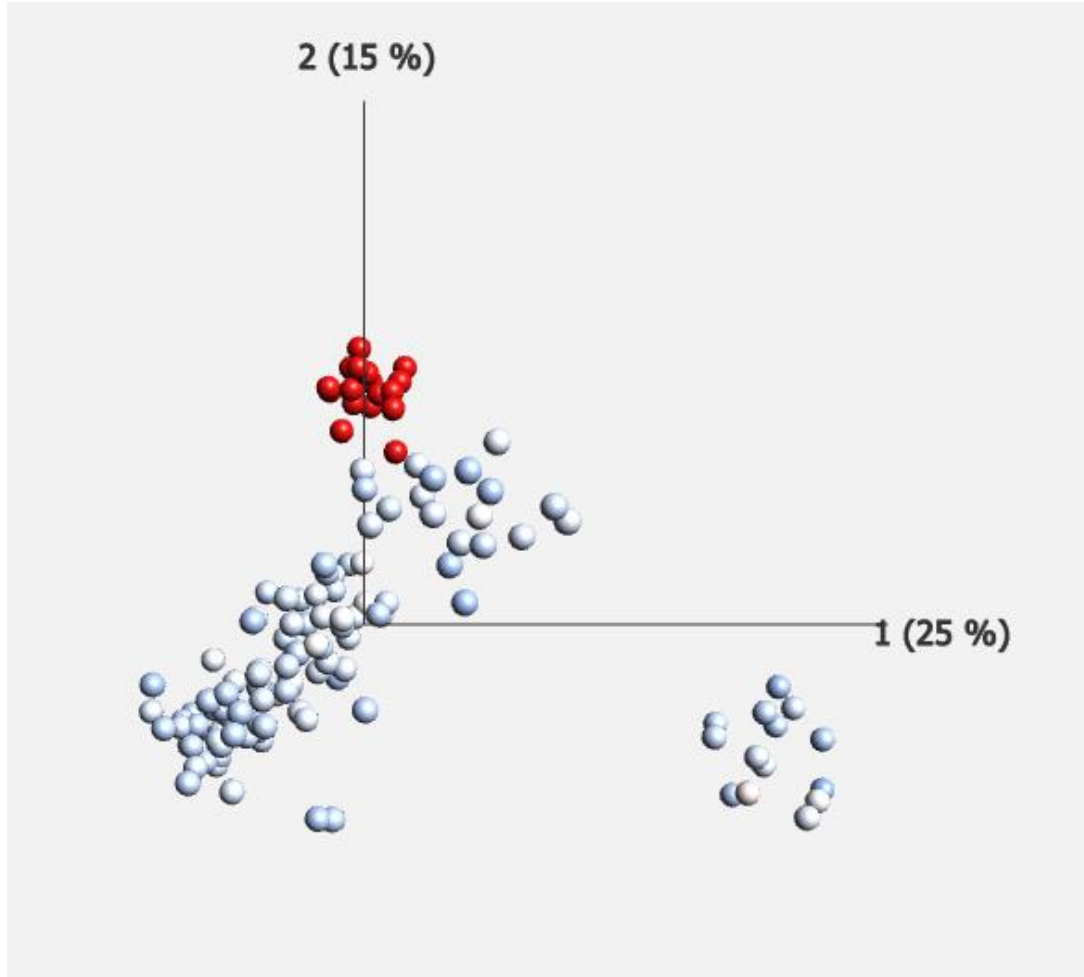


✓ ボルケーノプロット

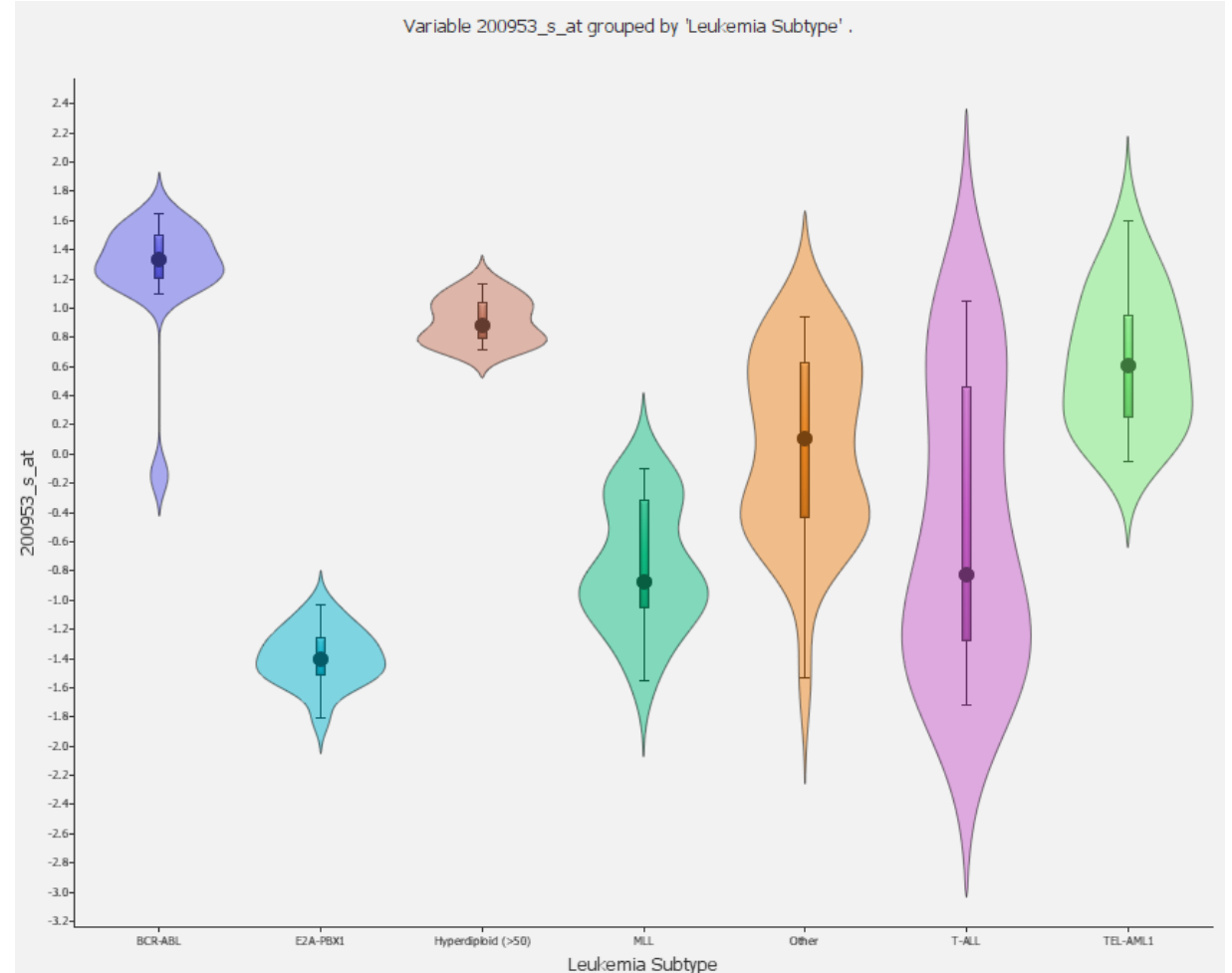
| IDs | Probe Set ID | Gene Symbol | p-value | q-value | Fold change |
|-----|--------------|-------------|-------------|-------------|-------------|
| 1 | 200953_s_at | CCND2 | 6.3055E-08 | 3.2631E-06 | 6.33025 |
| 2 | 201028_s_at | CD99 | 1.79425E-07 | 4.87219E-06 | 3.70553 |
| 3 | 201029_s_at | CD99 | 1.02775E-06 | 1.93404E-05 | 3.31656 |
| 4 | 201601_x_at | IFITM1 | 1.73129E-05 | 0.000223985 | 3.04406 |
| 5 | 201656_at | ITGA6 | 3.24838E-05 | 0.000373564 | 4.82861 |
| 6 | 201811_x_at | SH3BP5 | 0.000249063 | 0.00226764 | 3.09116 |
| 7 | 201858_s_at | SRGN | 0.00025196 | 0.00226764 | 2.38158 |
| 8 | 201859_at | SRGN | 0.00052879 | 0.00377446 | 2.32264 |

✓ 遺伝子リスト

選択した遺伝子の発現量を様々なグラフで可視化



✓ 遺伝子発現量のグラジエントをつけたプロット



✓ バイオリンプロット

お問い合わせ先：フィルジエン株式会社

TEL: 052-624-4388 (9:00～18 : 00)

FAX: 052-624-4389

E-mail: biosupport@filgen.jp