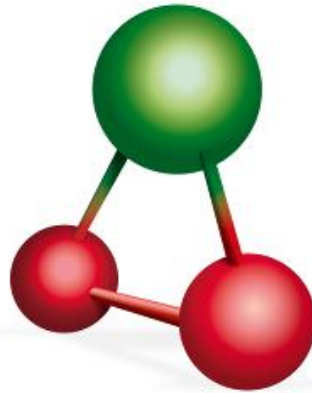
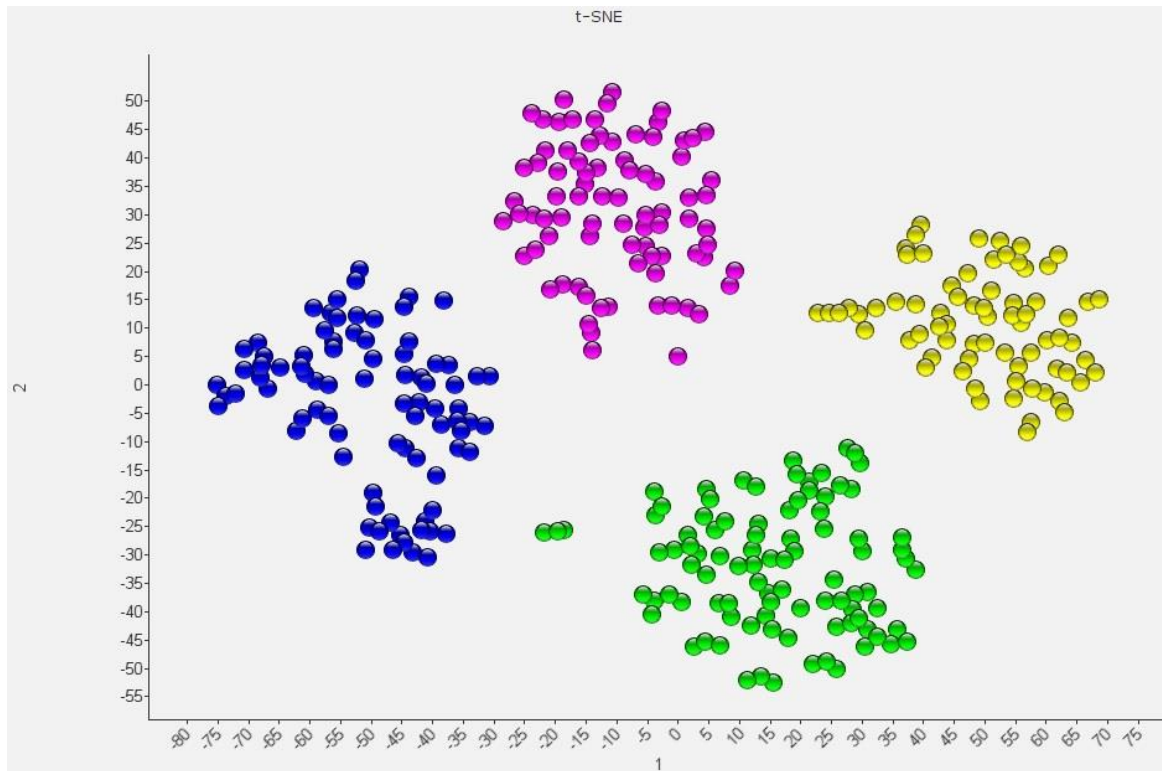


シングルセルRNA-Seqのための 情報解析

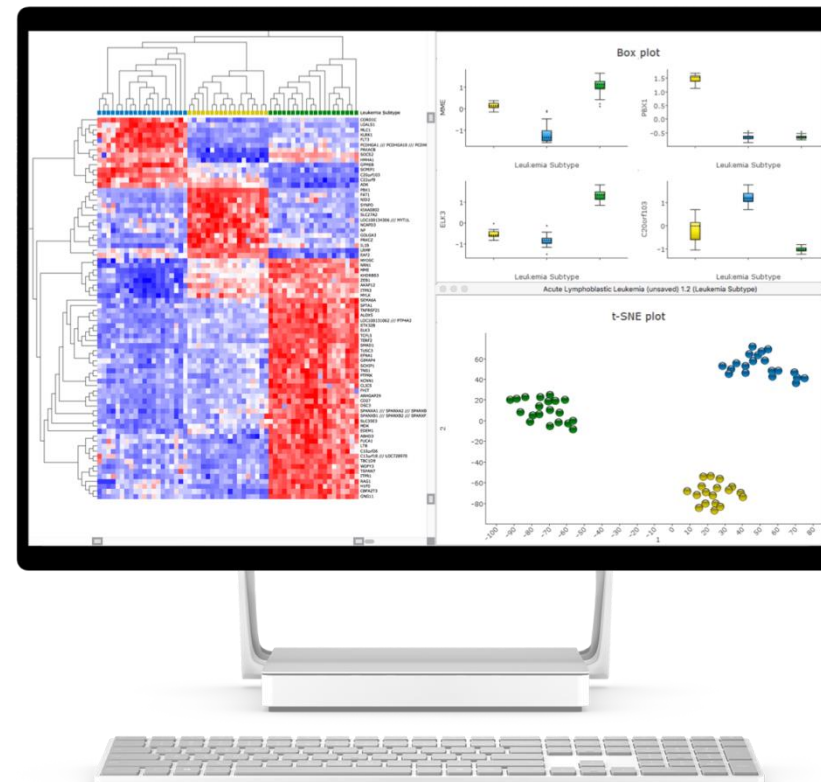


フィルジェン株式会社 バイオサイエンス部
(biosupport@filgen.jp)

- シングルセルRNA-Seqのデータ解析では、通常のRNA-Seqデータの解析手法に加え、データセット内の各細胞の遺伝子発現プロファイルの違いを俯瞰できるような、強力な情報解析アルゴリズムと、データのビジュアライズ機能を利用する必要がある
- Qlucore Omics Explorerに搭載されている各種解析アルゴリズムと、高度なグラフィックスおよび直感的なインターフェースを用いることで、シングルセルRNA-Seqのデータ解析を効率的に行うことができる



- すべてのデータを3Dプレゼンテーションとしてフルリアルタイムで操作が可能
- マウスクリックでフィルターやパラメーターの設定条件を簡単に変更し、自動でグラフに反映
- マイクロアレイデータ、RNA-Seqデータ（BAMファイル）のノーマライズが可能
- 階層クラスタリングとヒートマップ、主成分解析（PCA）、ボルケーノプロットやベン図、さらにシングルセルRNA-Seq解析用にt-SNEプロットをサポート
- 生物学的解釈を得るために、GSEA (Gene Set Enrichment Analysis) を利用可能
- サンプルの分類モデルの構築と、新サンプルへの適用を行うためのツールも搭載



対応アプリケーションエリア



Gene expression and miRNA
(Microarrays and RNA-Seq)



Proteomics



NGS



DNA Methylation



Metabolomics



Other

.txt.gz



.chp



.txt



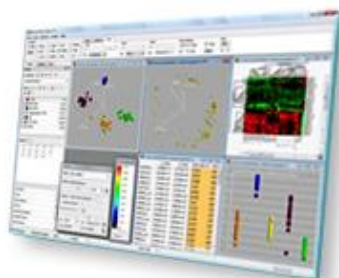
.gedata



.cel



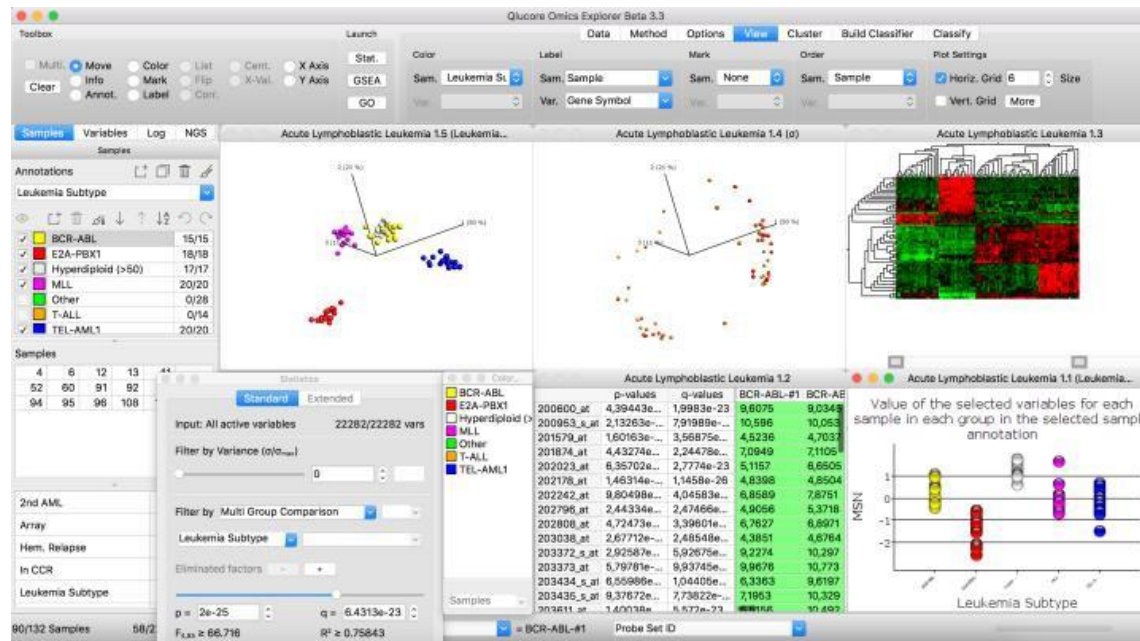
.soft.gz

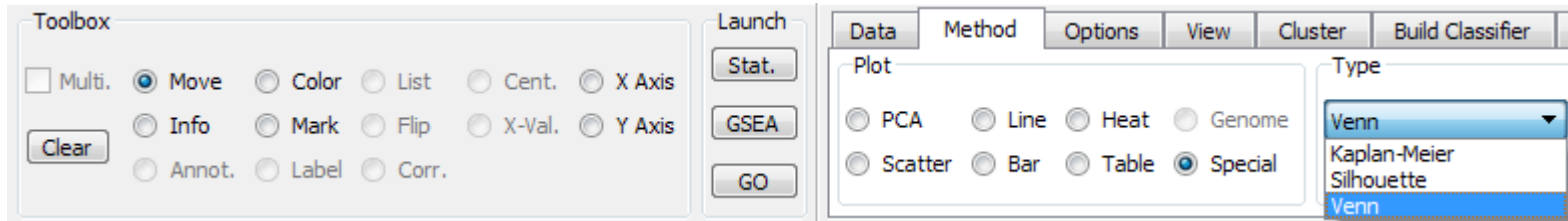


対応データファイル

- RNA-Seq
→ .bamファイル
- Affymetrix GeneChip WT, 3' Array
→ .celファイル .chpファイル
- Affymetrix GeneChip miRNA Array
→ .txtファイル
- Agilent Gene Expression Array
→ .txtファイル (Feature Extractionソフトウェア出力ファイル)
- GEO Data Set
→ .softファイル
- その他カスタムフォーマットファイル
→ .txtファイル

- Visualize:** 3D PCAプロットやヒートマップ、ボルケーノプロットやベン図の他、シングルセルRNA-Seq解析用にt-SNEプロット、生存率評価に Kaplan-Meier 曲線などをサポートし、複数のプロットグラフをインタラクティブに操作することが可能。
- Analyze:** 2グループあるいは複数グループ間の比較用統計モデルを搭載し、その他RのOpen APIをインテグレートしてソフトウェア上で使用することが可能。また Fold Change や各種クラスタリング、GSEA (Gene Set Enrichment Analysis) による生物学的解釈の評価や、サンプルの分類モデルの構築に Support Vector Machines (SVM) や Random Trees (RT) も搭載。
- Explore:** 各種マイクロアレイの他、RNA-Seqデータにおける Aligned BAMファイルのインポートからノーマライズ、GEO登録データの自動ダウンロードなどを利用可能。
- Share:** データプロット図のイメージや実データのリストなどのファイル出力、さらに操作の記録のログや動画データによる保存にも対応。

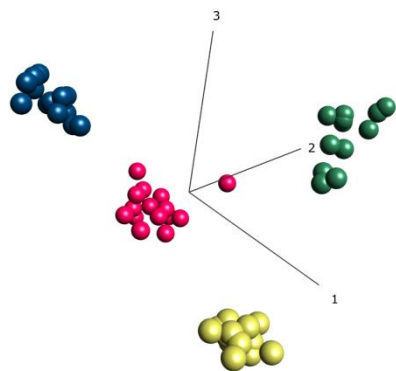




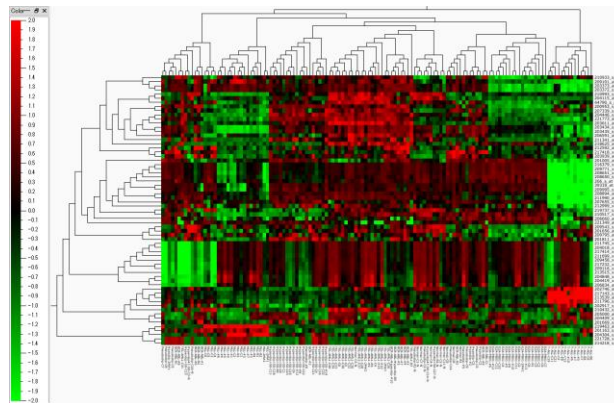
プロットの種類

- 「Method」タブの「Plot」と「Type」を選択することによって、データの表示形式を切り替える
- プロットをサンプルに対して行うか、変数（遺伝子など）に対して行うかの選択や、サンプル間の正規化の有無なども選択可能
- ユーザー定義のアノテーション情報に基づき、プロットの着色やラベル表示が可能
- データのフィルタリングを実行すると、フィルタリング結果がプロットにリアルタイムで反映される

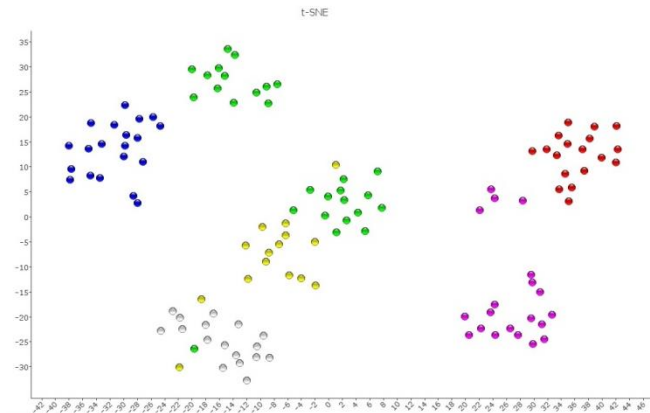
- **PCA**
- **Line**
- **Heatmap**
- **Genome**
- **Scatter**
 - Scatter
 - Volcano
 - 2D t-SNE
- **Bar**
 - Bar
 - Histogram
 - Box
- **Table**
- **Special**
 - Kaplan-Meier
 - Silhouette
 - Venn



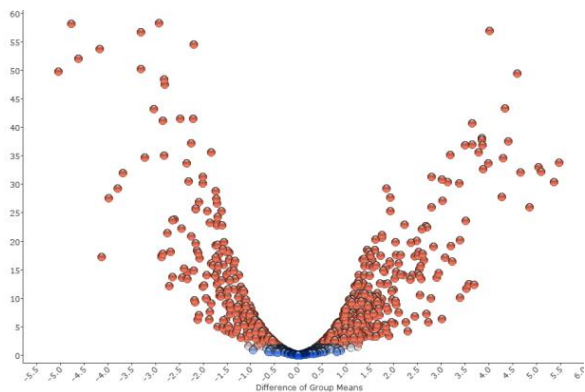
PCA



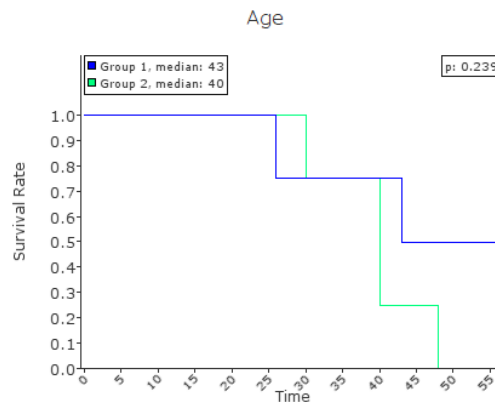
Heatmap



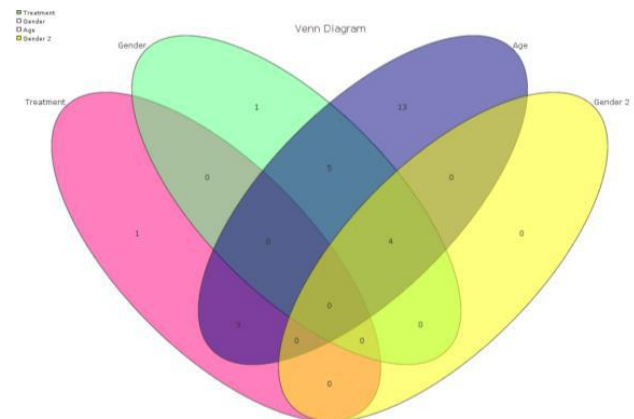
t-SNE



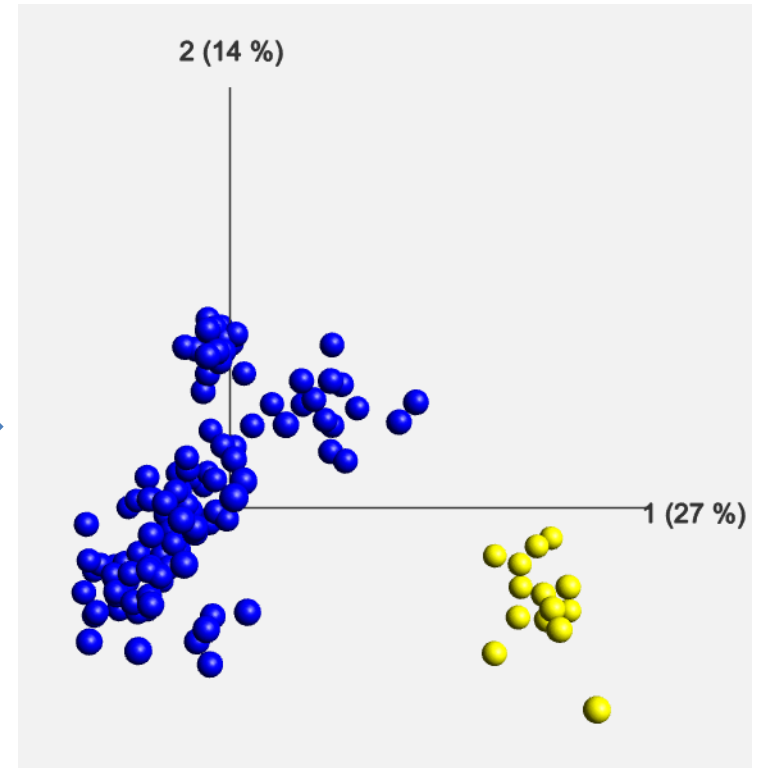
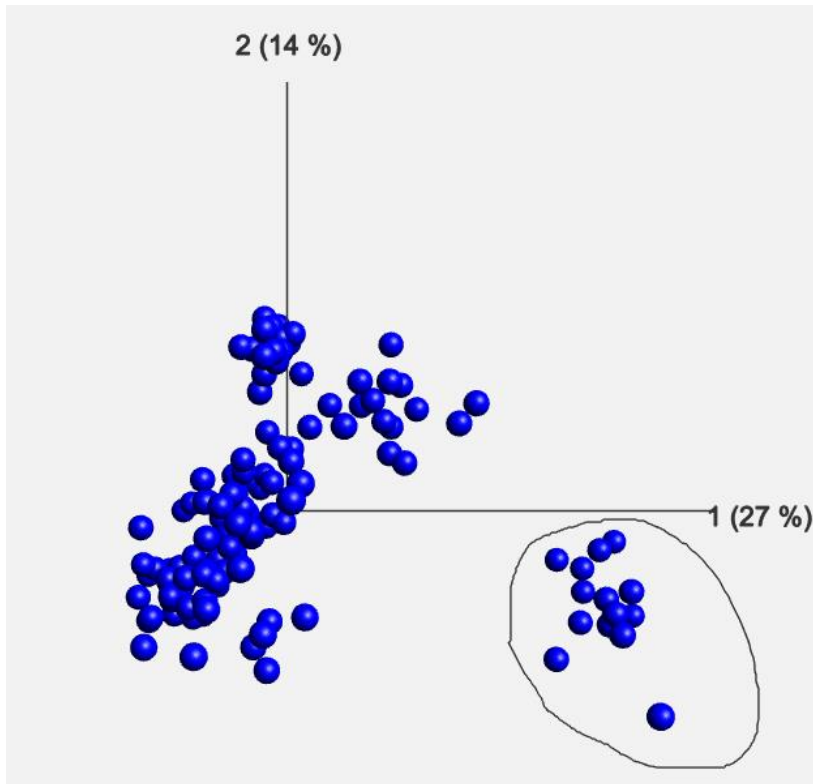
Volcano



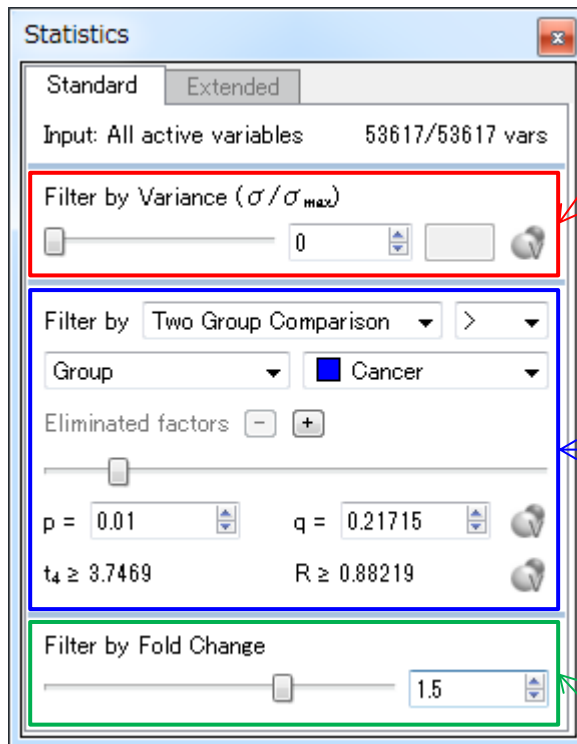
Kaplan-Meier



Venn



- プロット上のマウスカーソル操作で、サンプルのグループ分類などのアノテーションを手動で追加することができる



サンプル間のデータのばらつきによるフィルタリング

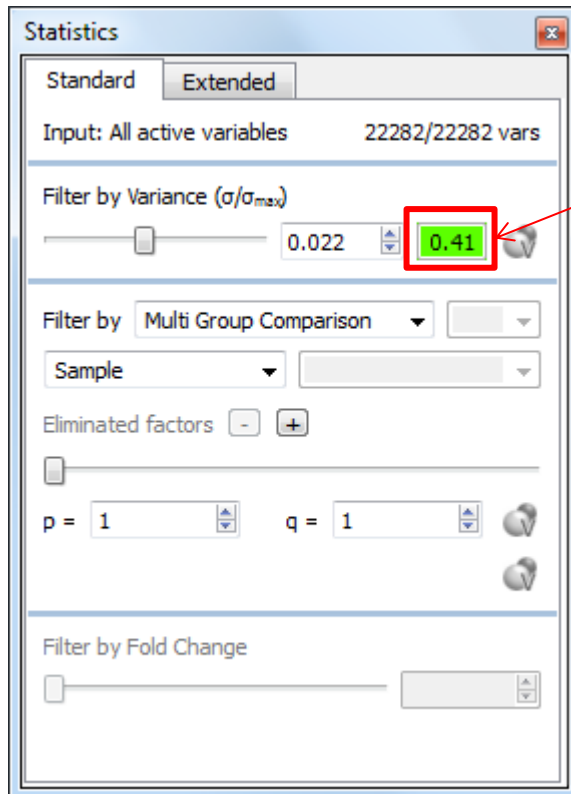
サンプルグループ間の統計検定によるフィルタリング

使用可能な統計手法

- Two Group Comparison (t-test)
- Multi Group Comparison (F-test)
- Linear Regression
- Quadratic Regression
- Rank Regression
- False discovery rates (FDR)

サンプルグループ間のFold Changeによるフィルタリング

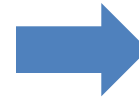
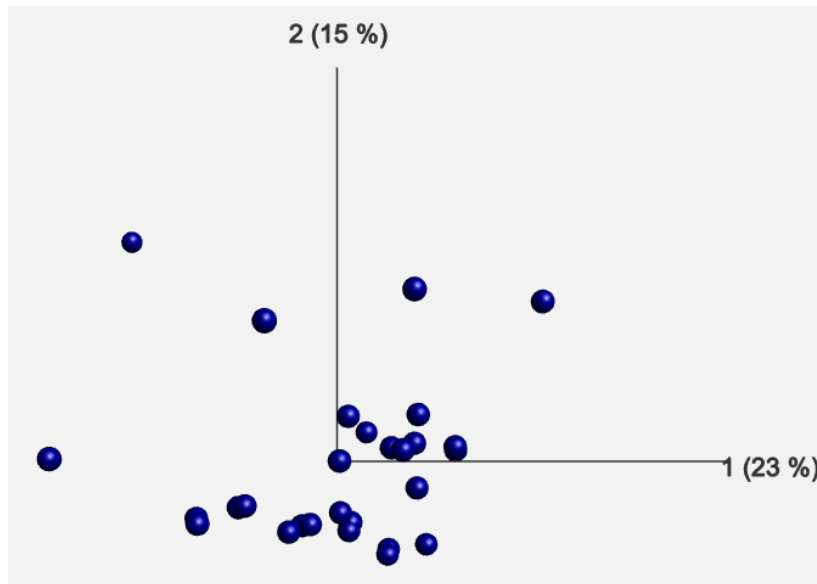
- Statistics Dockに任意の閾値を入力、あるいはスライダーを移動させ、データのフィルタリングを行う
- フィルタリングの条件を変更することにより、データプロットにもリアルタイムで反映される



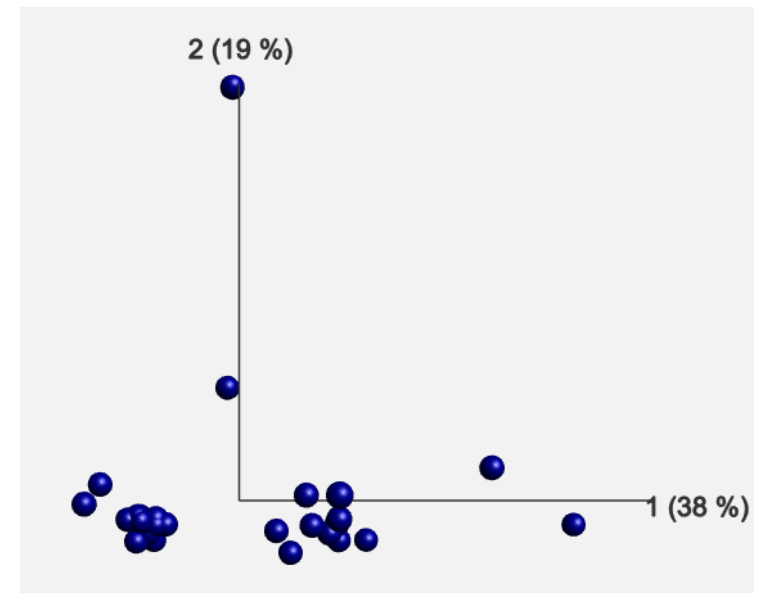
Projection Score

- PCA Plotにおいて、Filter by Varianceでフィルタリングを実施する際に計算されるスコア
- フィルタリングによりデータの次元を減らしていった際の、データの情報を数値化したものであり、この値が最大になるように調整することで、フィルタリングの閾値を決定できる
- スコアが高い場合は、緑色で表示される

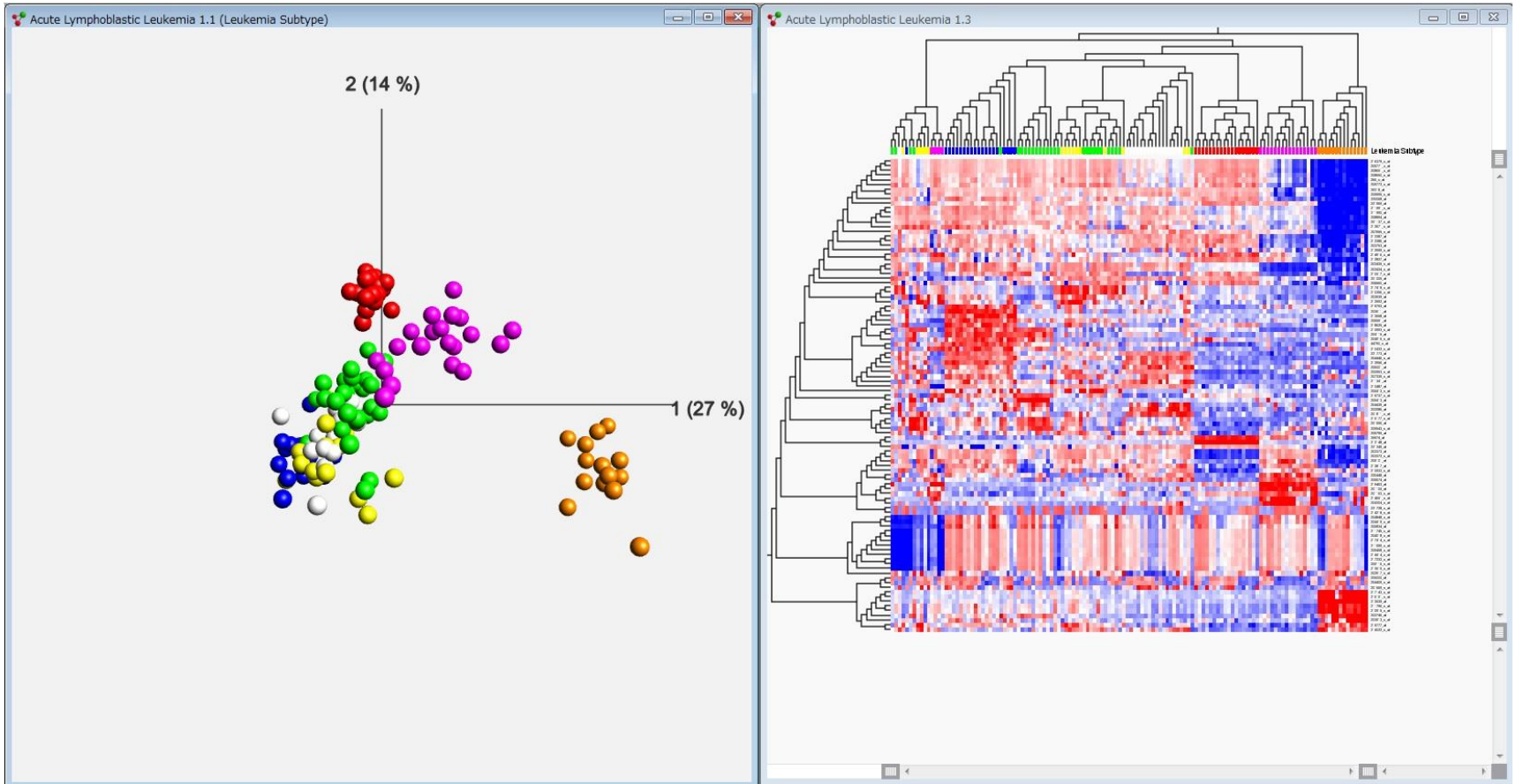
Filter by Variance ($\sigma/\sigma_{\text{max}}$)



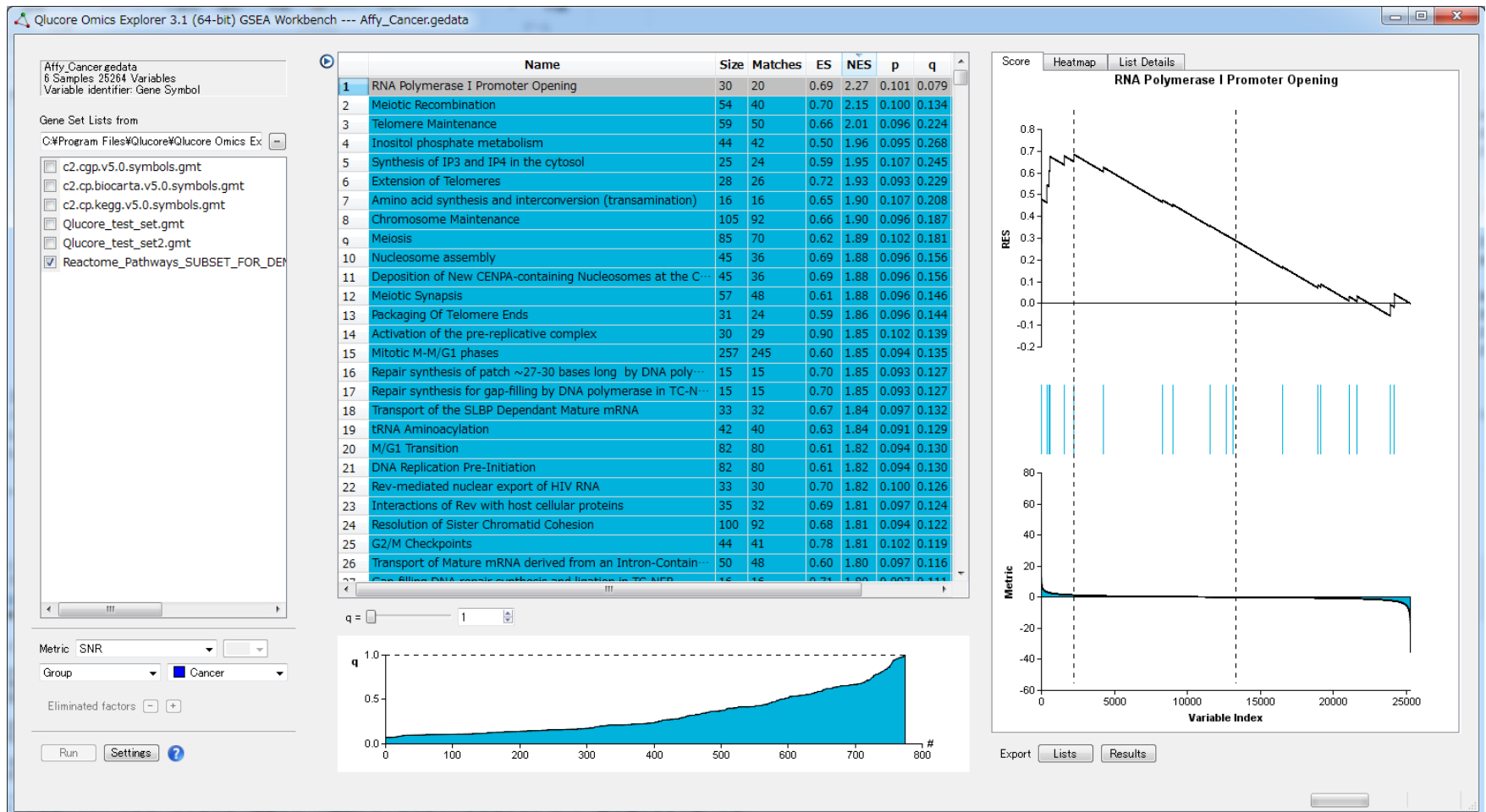
Filter by Variance ($\sigma/\sigma_{\text{max}}$)



- Projection Scoreが高いフィルタリング閾値を採用することによって、閾値の選択の根拠を客観的に示すことができる



- 複数の種類のプロットを表示させ、データのフィルタリング条件などを連動させることができる



- 特定の生理作用（疾患、薬物刺激など）や遺伝子機能カテゴリー（Gene Ontology, パスウェイなど）の条件で発現する遺伝子のグループ（遺伝子セット）情報を用いて、発現変動遺伝子セットに、それらグループの遺伝子がどれだけ多く含まれているかを検定する

Cluster: K-means Clusteringなどのアルゴリズムを使い、データを指定数のクラスターに分類

Data	Method	Options	View	Cluster	Build Classifier	Classify
Number of Clusters <input type="text" value="4"/>						
Options <input type="button" value="Run"/>						

Build Classifier: サンプルデータと既知のグループ情報をトレーニングセットとして用いて、サンプルデータの分類を行うためのモデルを作成

Data	Method	Options	View	Cluster	Build Classifier	Classify
Method	<input type="text" value="KNN"/>	Ranking	<input type="text" value="Multi Group Comparison"/>	Validation	<input type="text" value="no validation set"/>	<input type="button" value="Settings"/>
Key	<input type="text" value="Gene"/>	Filter Value	<input type="text"/>	Validation Key	<input type="text"/>	<input type="button" value="Build"/>

Classifier: Build Classifierで作成した分類モデルを、グループ情報未知サンプルデータに適用

Data	Method	Options	View	Cluster	Build Classifier	Classify
Classifier						
<input type="button" value="Load"/> <input type="text"/>						
<input type="button" value="Unload"/> <input type="button" value="Show Report"/>						
Normalization Mode						
<input type="radio"/> As Dataset						
<input type="radio"/> As Classifier						
<input type="button" value="Apply"/>						

ソフトウェアデモンストレーション

- 合計329サンプルのhESC細胞のシングルセルRNA-Seqデータ
- 細胞の分化過程の観察のため、4つのタイムポイントごとにサンプリングを行っている

NCBI GEO Accession Display

Scope: Self Format: HTML Amount: Quick GEO accession: GSE109979 GO

Series GSE109979 Query DataSets for GSE109979

Status	Public on Feb 15, 2018
Title	Single-Cell RNA Sequencing Reveals Metallothionein Heterogeneity during hESC Differentiation to Definitive Endoderm [scRNA-Seq]
Organism	Homo sapiens
Experiment type	Expression profiling by high throughput sequencing
Summary	Differentiation of human pluripotent stem cells toward definitive endoderm (DE) is the critical first step for generating cells comprising organs such as the gut, liver, pancreas and lung. This in-vitro differentiation process generates a heterogeneous population with a proportion of cells failing to differentiate properly and maintaining expression of pluripotency factors such as Oct4. RNA-sequencing of single cells collected at four time points during a 4-day DE differentiation identified high expression of metallothionein genes in the residual Oct4-positive cells that failed to differentiate to DE. Using X-ray fluorescence microscopy and multi-isotope mass spectrometry, we discovered that high intracellular zinc level corresponds with persistent Oct4 expression and failure to differentiate. We further show that differentiation-arrested phenotype is inversely correlated with zinc concentration in the differentiation media. This study improves our understanding of in-vitro DE differentiation and provides actionable options to improve DE differentiation efficiency.
Overall design	RNA-sequencing of 329 single cells collected at four time points during a 4-day DE differentiation to identify mechanisms leading to cellular heterogeneity during differentiation
Contributor(s)	Lu J , Lerou P , Li H
Citation(s)	Lu J, Baccei A, Lummertz da Rocha E, Guillermier C et al. Single-cell RNA sequencing reveals metallothionein heterogeneity during hESC differentiation to definitive endoderm. <i>Stem Cell Res</i> 2018 Apr;28:48-55. PMID: 29427839

手順1： サンプルデータのインポート

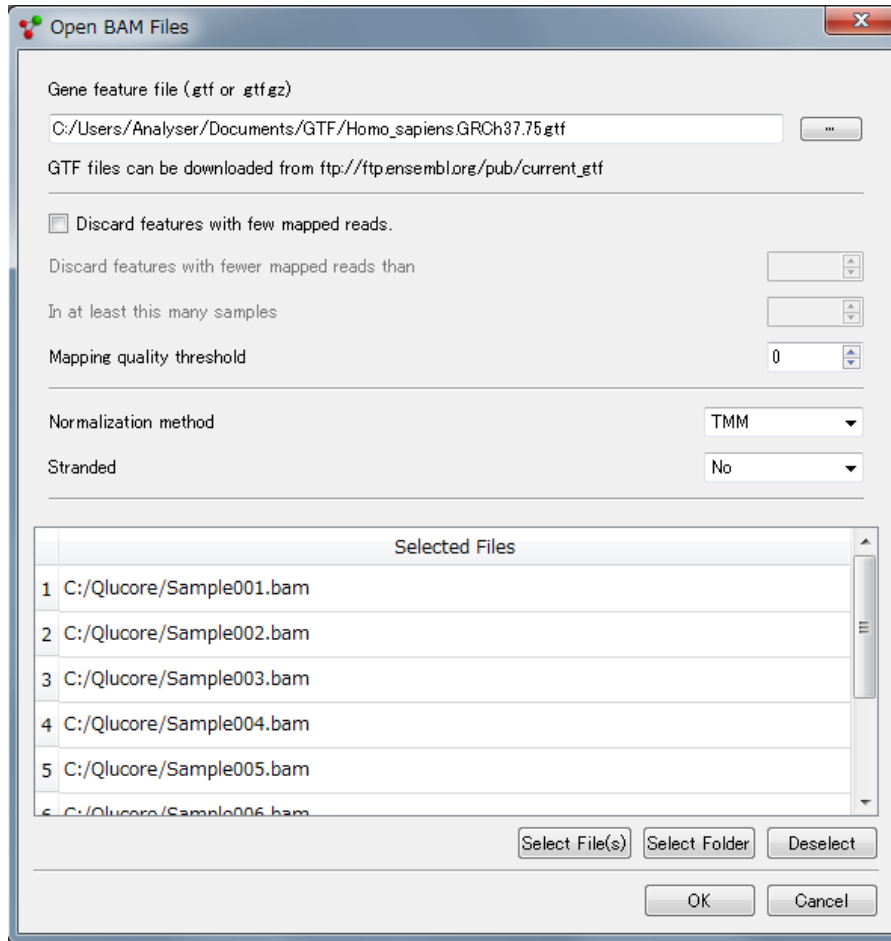
- RNA-Seq BAMファイル、または数値化データファイルをインポート

手順2： 遺伝子の抽出

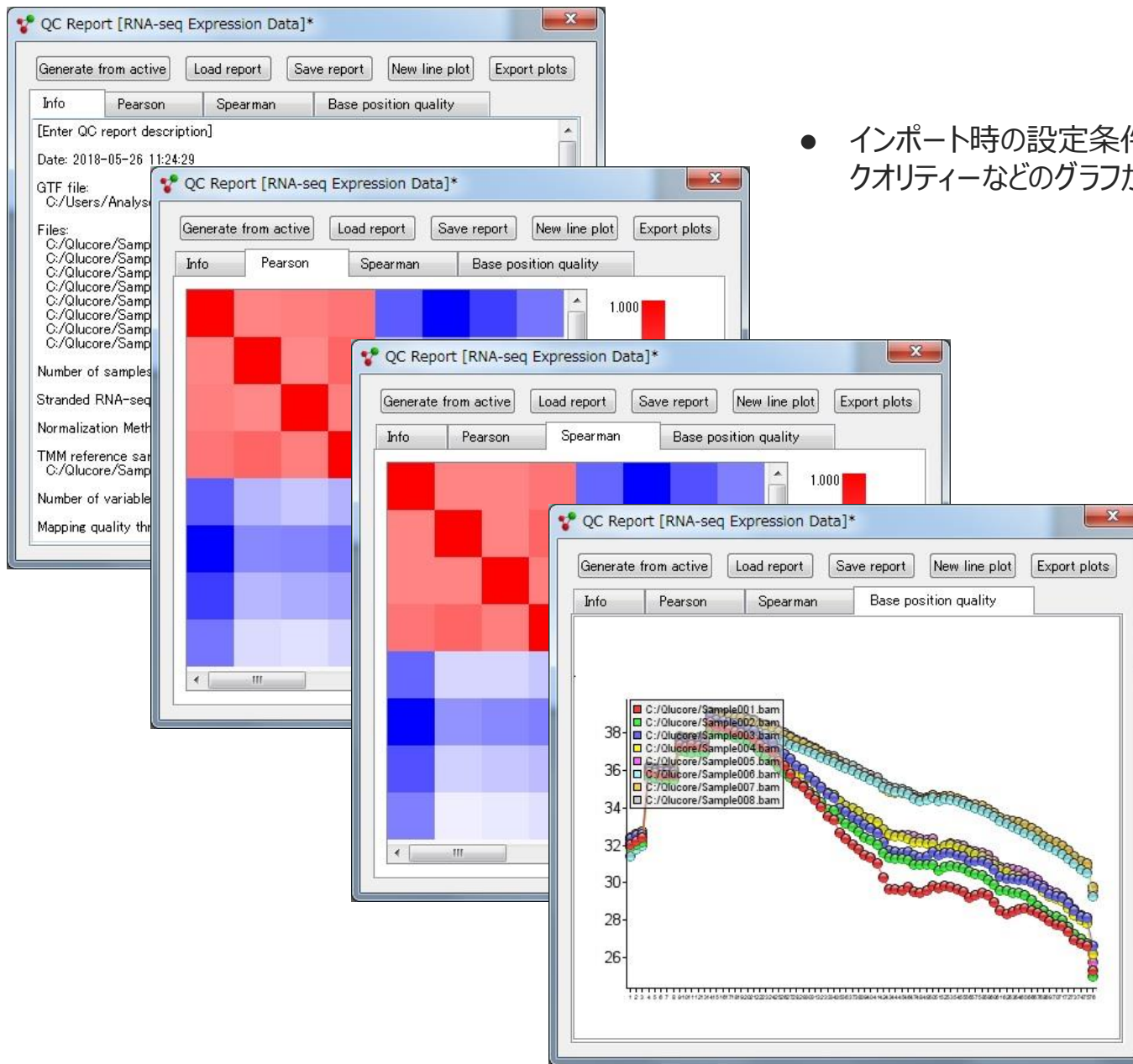
- Projection Scoreなどを用いて、解析に用いる遺伝子を抽出

手順3： 各種プロットグラフの作成

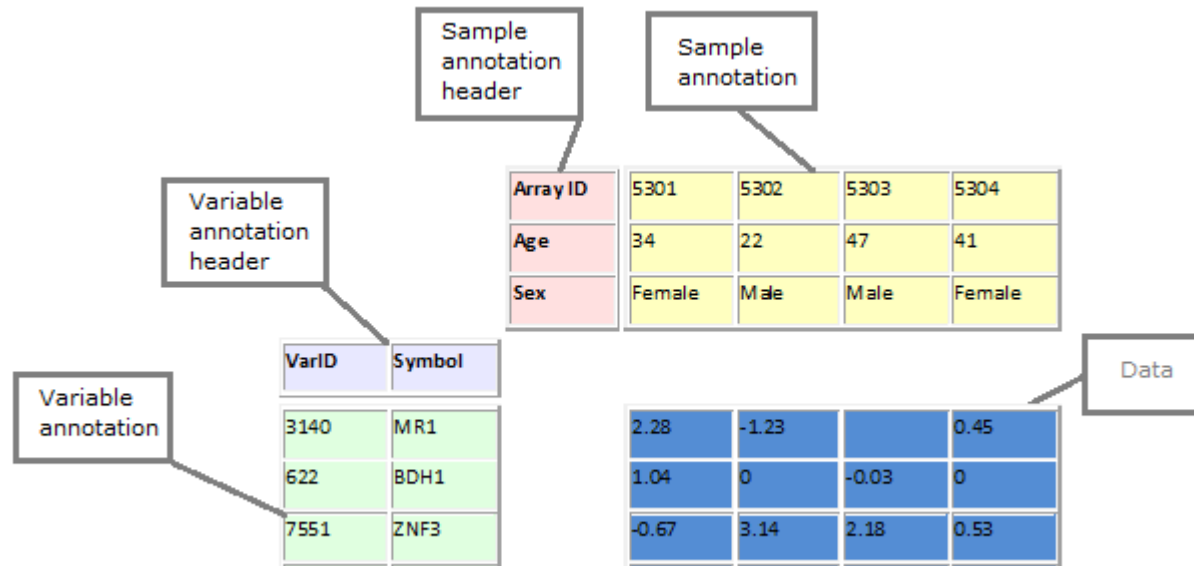
- PCA、t-SNE、階層クラスタリングなどのグラフ作成



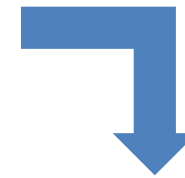
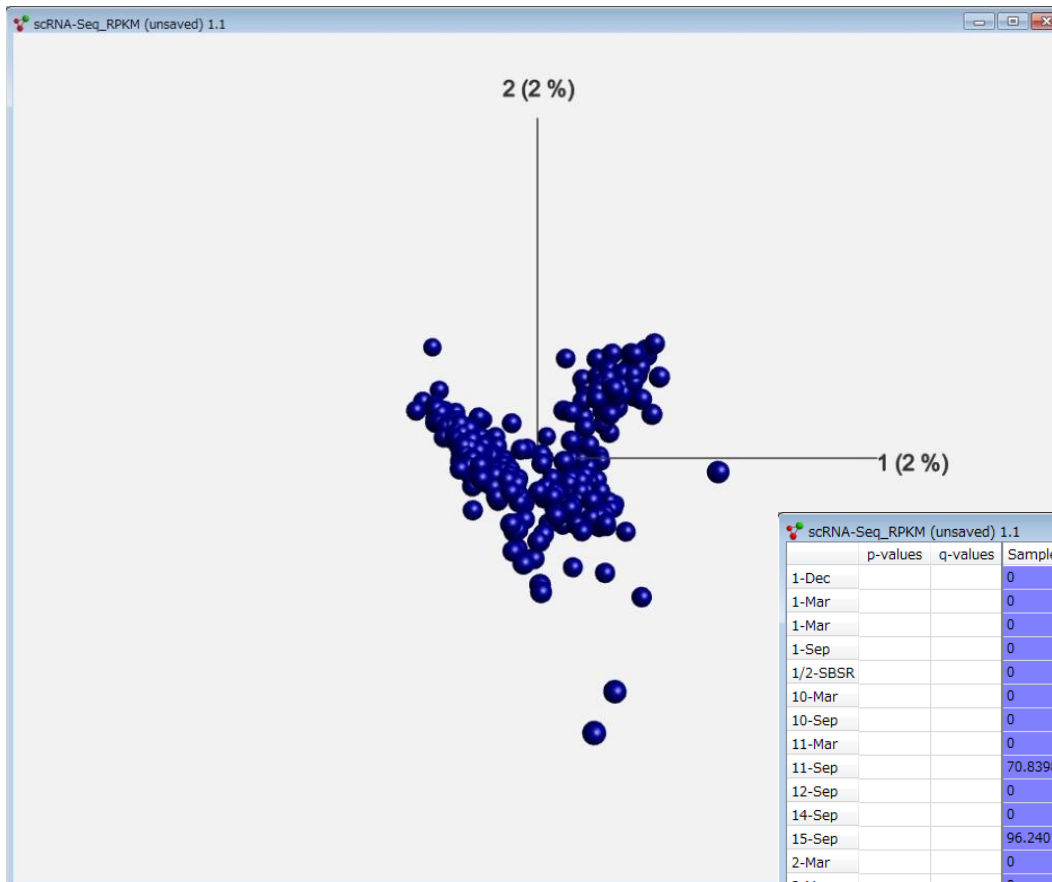
- BAMファイルのインポートメニューでは、サンプルごとのBAMファイルに加え、遺伝子アノテーションデータのGTFファイルも指定する必要がある
- ノーマライゼーション手法としては、TMM、FPKM、TPMを選択可能
- マッピング時の各種クオリティー情報によるフィルタリングや、ストランドの設定を行うことも可能



- インポート時の設定条件や、サンプル間の相関係数、クオリティーなどのグラフが出力される

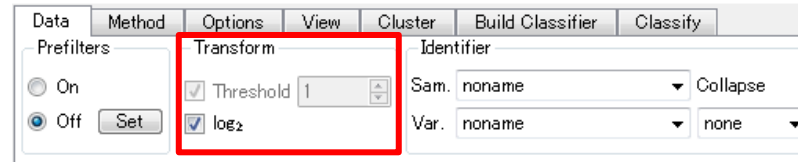


- すでに数値化されているデータをインポートする場合は、タブ区切りまたはカンマ区切りのテキストファイルを使用する
- インポート時にサンプル名や遺伝子名などのフィールドを自由に指定することができるため、様々なデータフォーマットのファイルに対応することができる

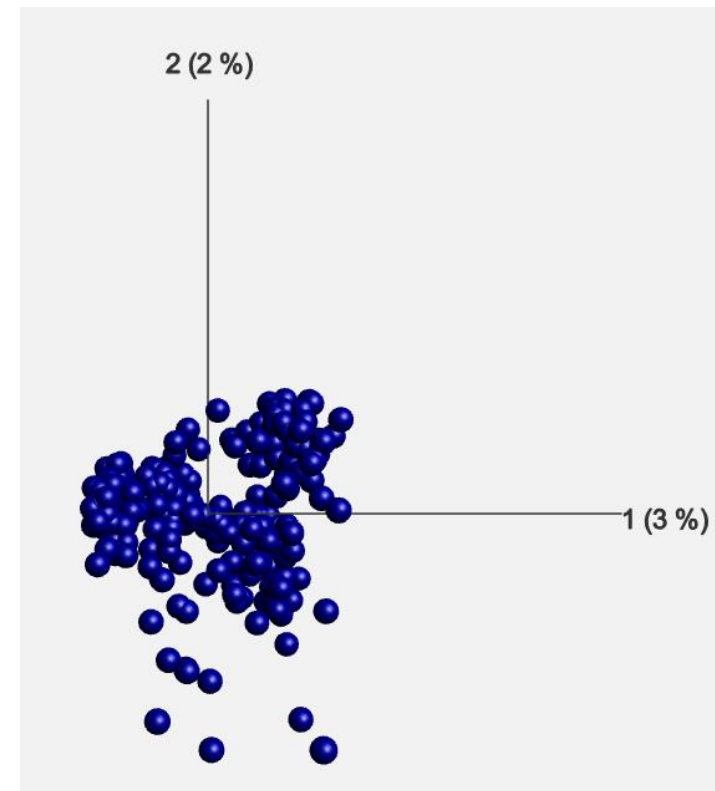


	p-values	q-values	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C
1-Dec			0	0	0	4.92575	1.48805	0	0
1-Mar			0	10.678	0	0	8.2994	0.17993	10.8816
1-Mar			0	0	0	8.5988	26.9508	0	0
1-Sep			0	0	0	0	0	0	0
1/2-SBSR			0	0	21.8371	23.7318	32.6089	0	0
10-Mar			0	0	0	0	0	0	0
10-Sep			0	14.2747	7.36993	12.0724	7.15103	11.0384	11.2512
11-Mar			0	0	0	0	0	0	0
11-Sep			70.8398	63.8894	57.2022	32.0308	39.0407	49.5489	15.3783
12-Sep			0	0	0	0	0	0	0
14-Sep			0	0	0	0	0	0	0
15-Sep			96.2401	72.1834	111.251	147.773	139.111	168.891	145.914
2-Mar			0	0	0	8.81421	0	0	0
2-Mar			0	3.07719	0	0.466544	0	0	0
2-Sep			113.972	210.395	199.352	181.716	149.248	106.687	208.858
3-Mar			0	13.5118	6.21246	0.186234	0.450086	0	17.7037
3-Sep			0	10.2881	3.86304	0	3.19855	0	2.62108
4-Mar			0	0	0	0	0	0	0
4-Sep			0	0	0.561265	0	0	0.661179	0
5-Mar			5.01165	3.32756	0.218159	0	1.89664	2.26155	0

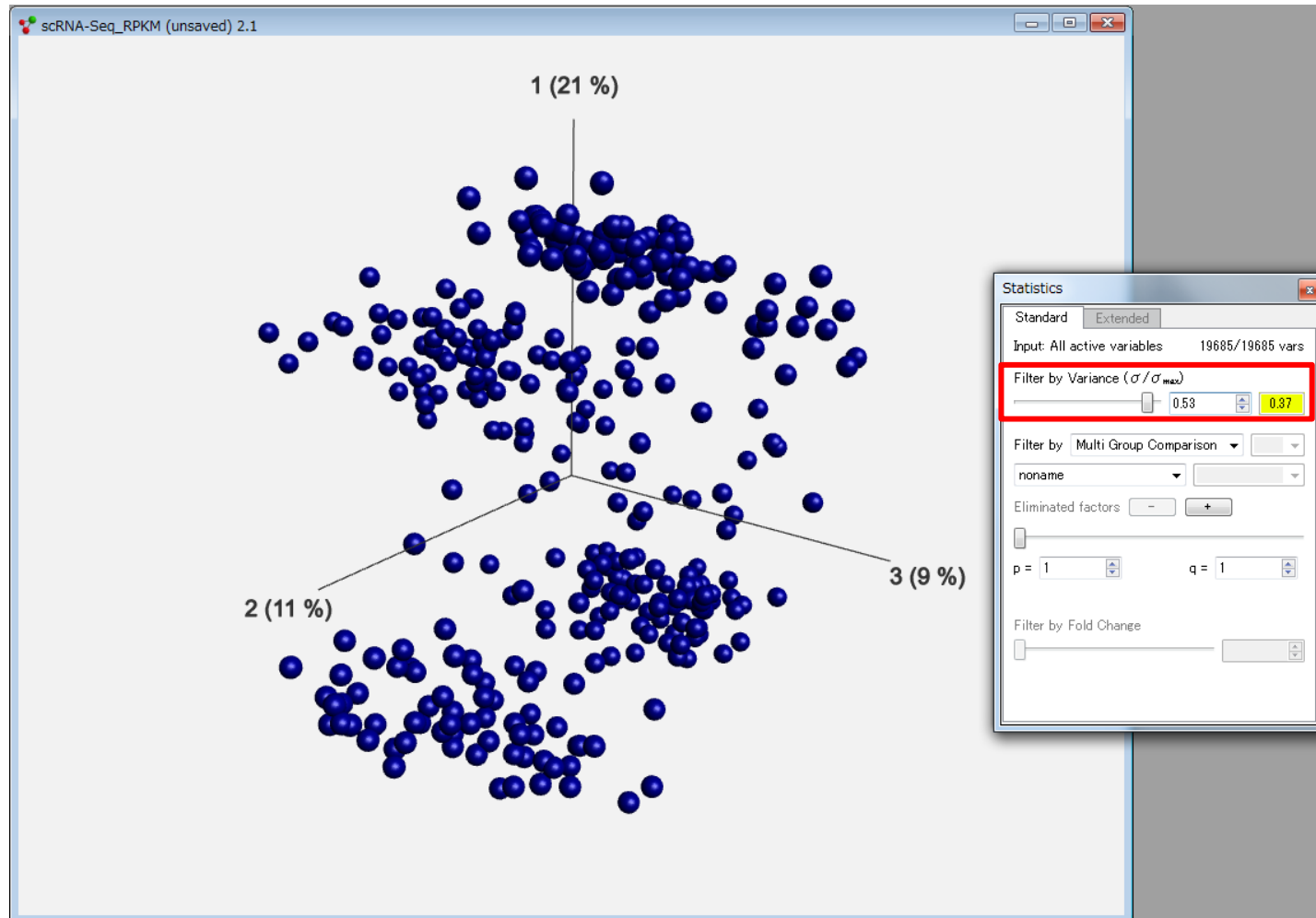
- インポートが完了すると、PCAプロットが表示される
- プロットを変更することで、数値化データをテーブルで確認することも可能



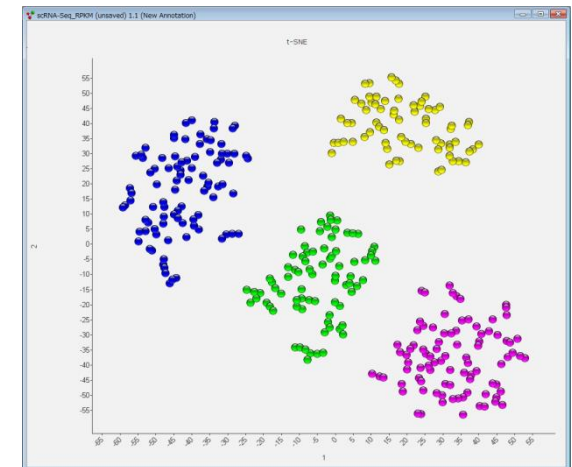
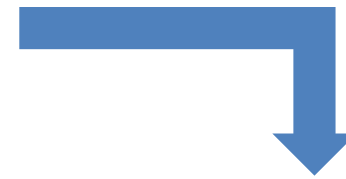
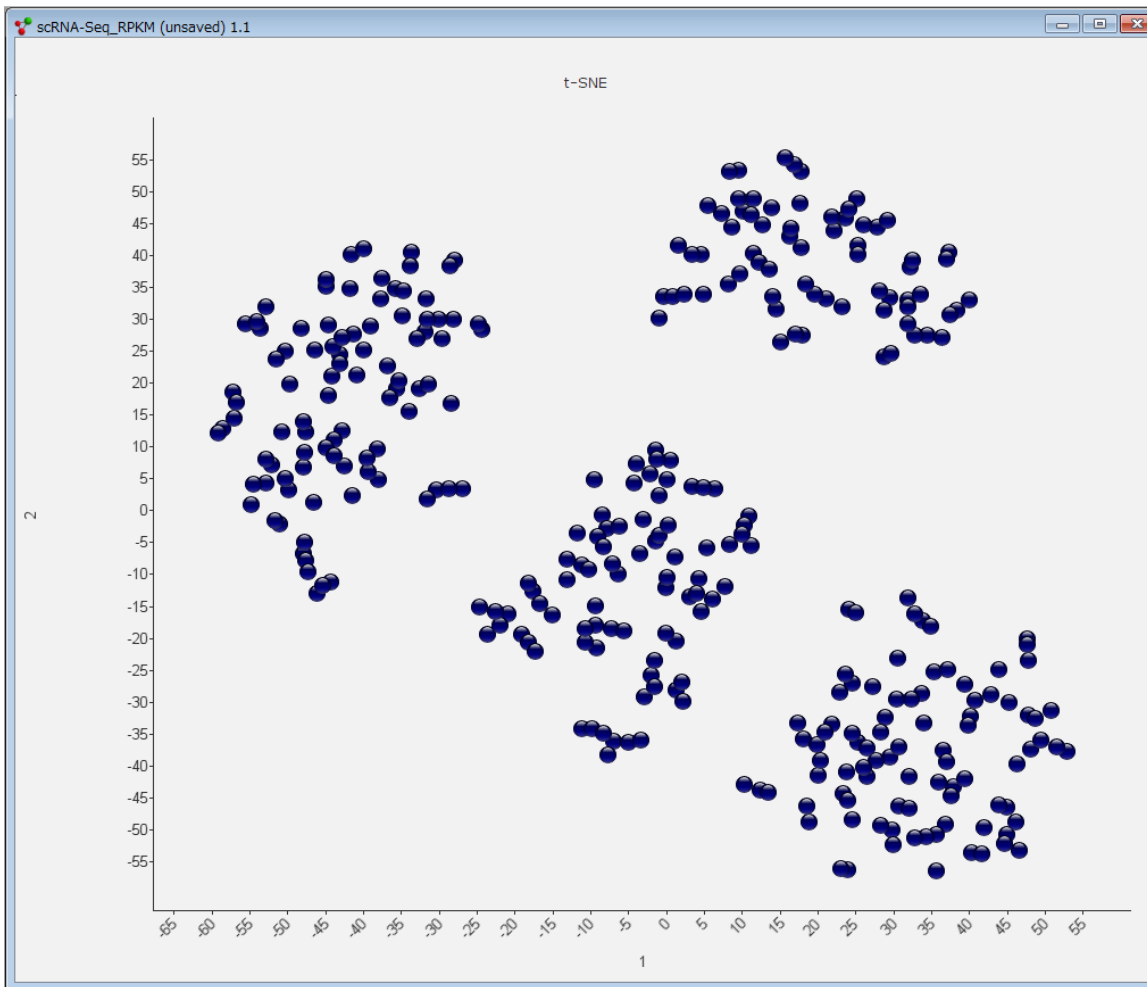
Sample0C	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C	Sample0C
0	0	0	2.30034	0.573427	0	0	0	2.39552
0	3.41657	0	0	3.05301	0	3.44382	0	2.55318
0	0	0	3.10413	4.75226	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	4.44871	4.56875	5.02719	0	0	0	0
0	0	0	0	0	0	0	0	0
0	3.83539	2.88165	3.59364	2.83815	3.46446	3.492	5.91135	2.60136
0	0	0	0	0	0	0	0	0
6.14649	5.9975	5.838	5.00139	5.28691	5.63078	3.94282	6.63331	5.91725
0	0	0	0	0	0	0	0	0
6.58857	6.17359	6.79767	7.20724	7.12009	7.39995	7.18898	8.14414	7.16083
0	0	0	3.13983	0	0	0	0	0
0	1.62162	0	0	0	0	0	0	0
6.83254	7.71696	7.63917	7.50554	7.22157	6.73724	7.70638	7.95609	6.77065
0	3.75615	2.63516	0	0	0	4.14598	0	0
0	3.3629	1.94974	0	1.67742	0	1.39016	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
2.32529	1.73447	0	0	0.923449	1.17731	0	0.392573	0



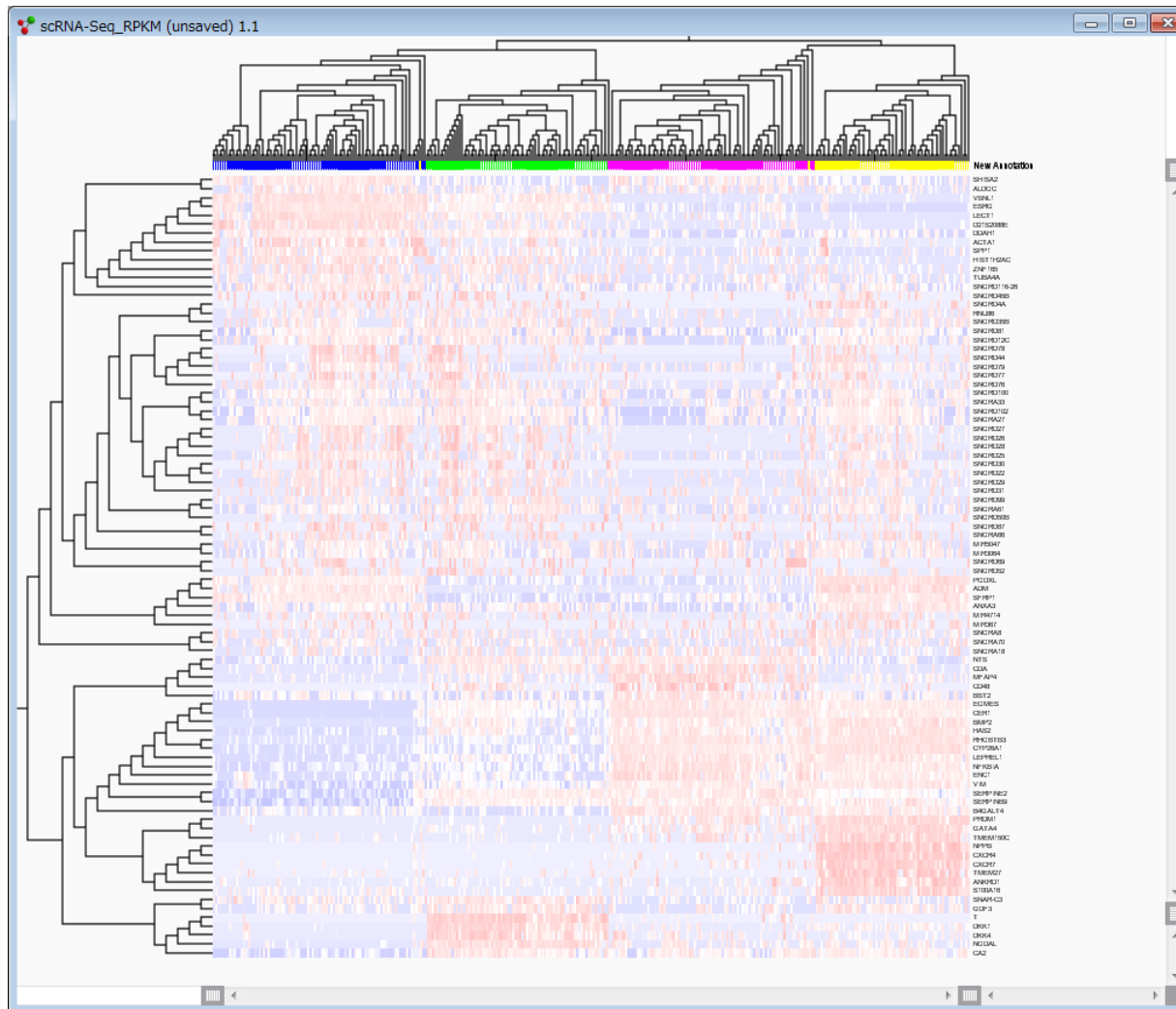
- 対数データになっていない場合は、必要に応じて対数変換を行う
- 変換を行うと、各プロットの表示も自動で変化する



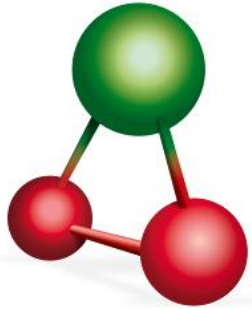
- インポートした全遺伝子セットから、分散フィルターを使用し、サンプル間の分散が大きい遺伝子のみを抽出する
- Projection Scoreが最大になるように調整することで、フィルターの閾値を決定できる



- 抽出遺伝子データを用いてt-SNEプロットを表示させた結果、サンプルが4つのグループに分類されることがわかる
- 分類がわかりやすいように、必要に応じてManual Annotationでサンプルの色付けを行う



- プロットをヒートマップに切り替え、階層クラスタリングを実施することで、各グループに特徴的な発現遺伝子のクラスター等を確認することが可能



ソフトウェアの詳細は、以下サイトをご覧ください。

弊社Webサイト：

<http://www.filgen.jp/Product/BioScience21-software/index15-qlucore.html>

メーカーサイト：

<http://www.qlucore.com/>

お問い合わせ先：フィルジェン株式会社

TEL 052-624-4388 (9:00～17:00)

FAX 052-624-4389

E-mail: biosupport@filgen.jp