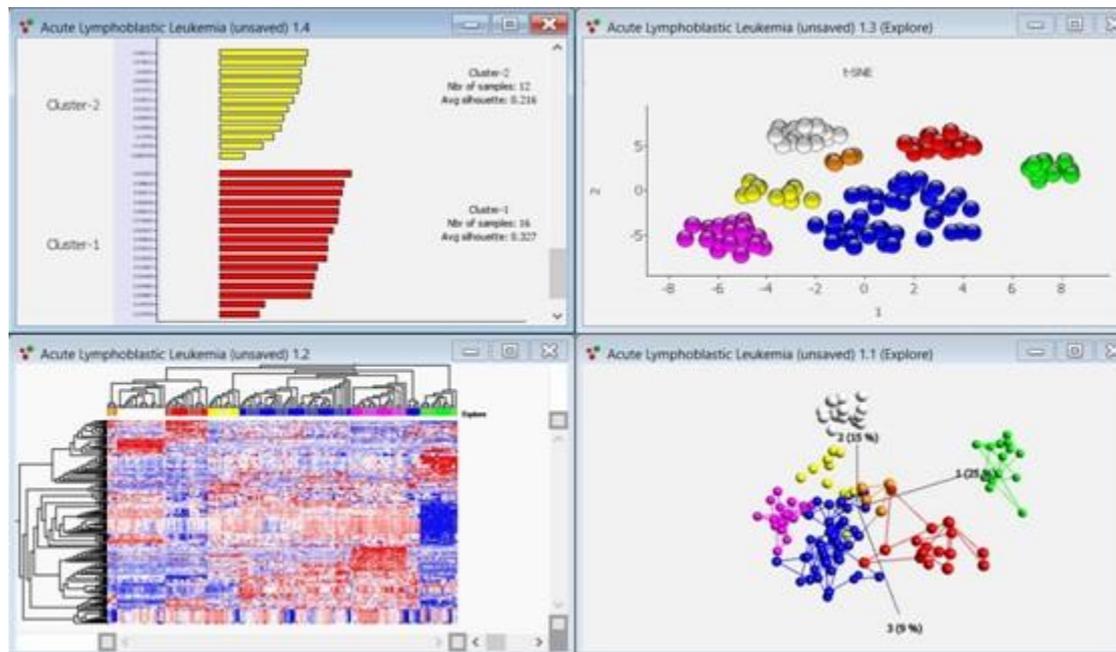


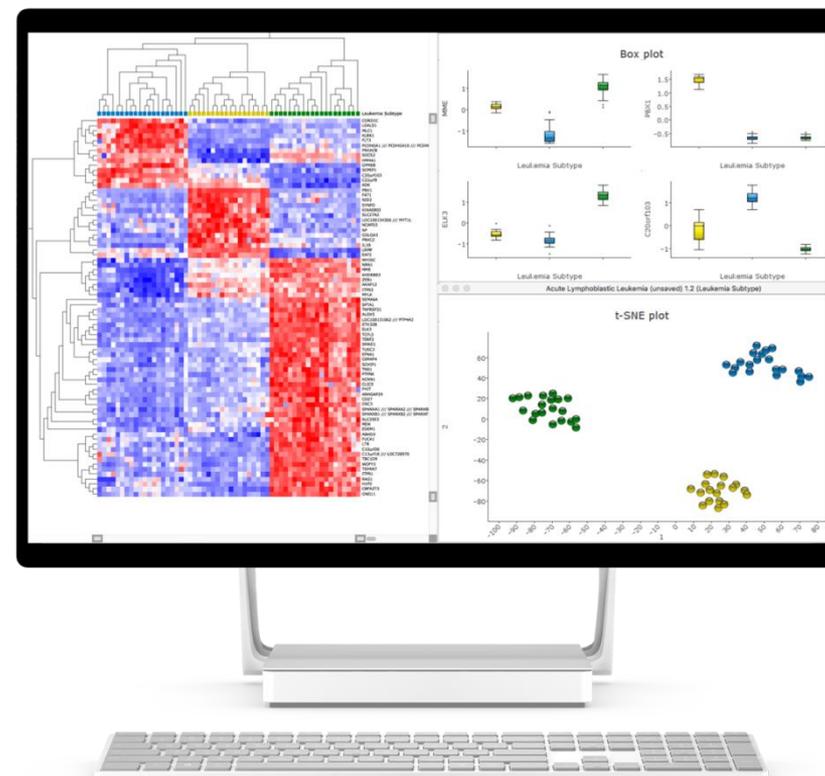
公共データベースを用いたオミックス解析

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

- 公共データベースに登録されている様々なサンプルの実験データは、自分のコンピュータにデータをダウンロードし、自由にデータ解析を行えるものがある
- Qlucore Omics Explorerには、GEOとTCGAの2種類のデータベースより、登録サンプルの遺伝子発現データをダウンロードする機能が搭載されており、自由にダウンロードすることが可能
- ダウンロード後は、ソフトウェア上で統計解析や、様々な形式でグラフ表示することで、データの探索を行う



- 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
- Affymetrix, Agilentマイクロアレイデータ、RNA-Seqデータ (BAMファイル & カウントデータファイル) のノーマライゼーションが可能
- 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



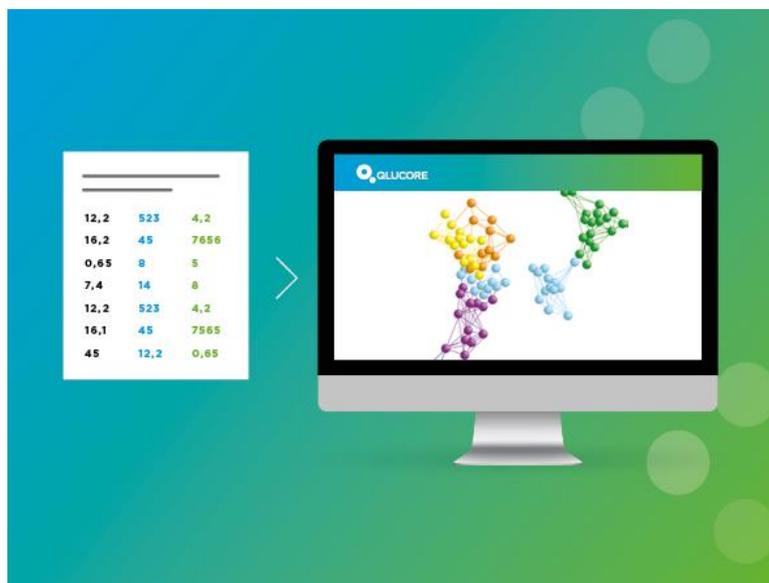
Metabolomics



Other

対応データファイル

- RNA-Seq
→ .bamファイル .txtファイル (カウントデータファイル)
- Affymetrix GeneChip WT, 3' and Clariom S 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



- NCBIが維持・管理を行っている、おもにDNAマイクロアレイや次世代シーケンサーで測定した遺伝子発現データのデータベース
- 様々な実験条件・生物種のデータが登録されており、生データや正規化済みのデータなどを自由にダウンロードすることが可能
- 実験条件などのサンプルごとの詳細情報であるサンプルアノテーション、および遺伝子名などの遺伝子アノテーションも同時にダウンロードされ、解析に用いることが可能

The screenshot shows the Gene Expression Omnibus (GEO) website. At the top, there is a navigation bar with links for 'GEO Home', 'Documentation', 'Query & Browse', and 'Email GEO'. The main heading is 'Gene Expression Omnibus' with the GEO logo. Below the heading, a brief description states: 'GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.' A search bar is located on the right with the text 'Keyword or GEO Accession' and a 'Search' button. The page is divided into three main sections: 'Getting Started', 'Tools', and 'Browse Content'. 'Getting Started' includes links for Overview, FAQ, About GEO DataSets, About GEO Profiles, About GEO2R Analysis, How to Construct a Query, and How to Download Data. 'Tools' includes links for Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, Search GEO Documentation, Analyze a Study with GEO2R, Studies with Genome Data Viewer Tracks, Programmatic Access, and FTP Site. 'Browse Content' includes a Repository Browser and a table showing the number of DataSets (4348), Series (139055), Platforms (21569), and Samples (4011310). At the bottom, there is a section for 'Information for Submitters' with links for Login to Submit, Submission Guidelines, Update Guidelines, MIAME Standards, Citing and Linking to GEO, Guidelines for Reviewers, and GEO Publications.

GEO Download

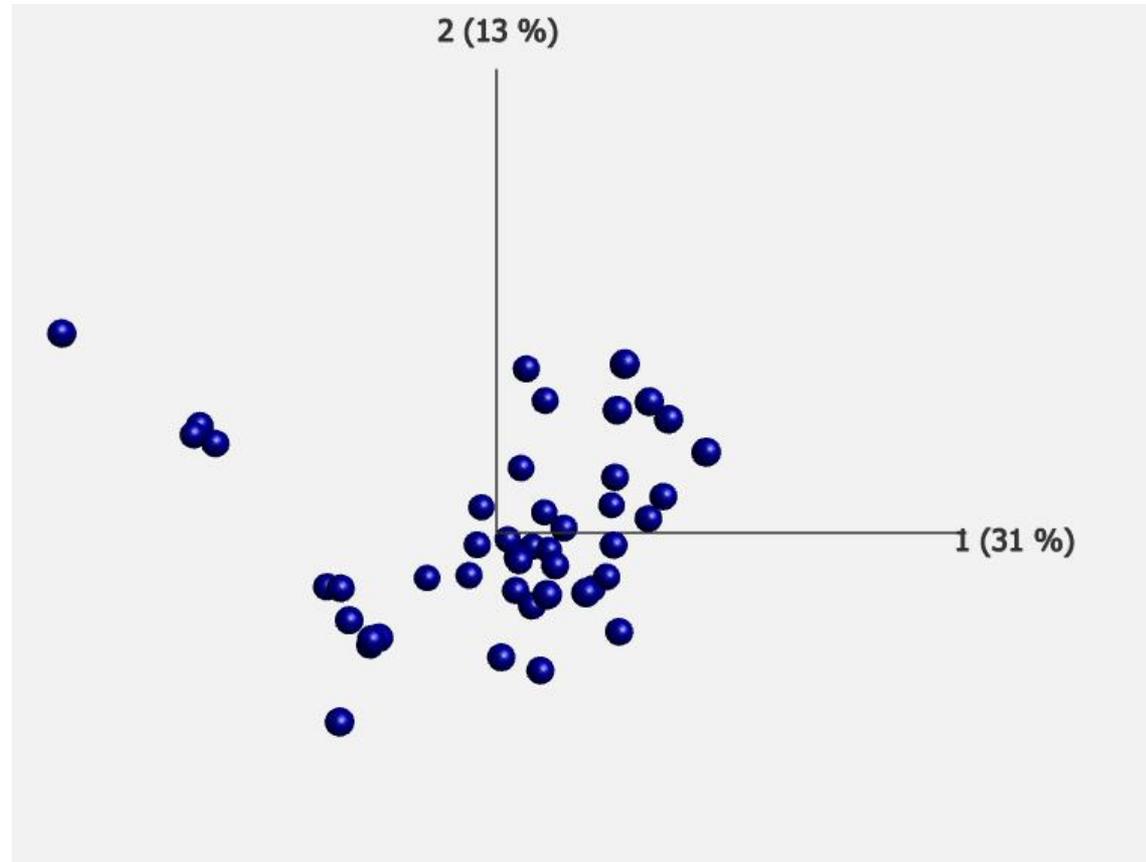
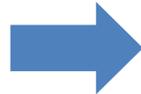
There are two types of GEO data sets:
GDS (GEO Data Set) - curated by GEO staff
GSE (GEO Series) - not curated by GEO staff

Enter accession number ("GDS" or "GSE" + any number of digits)

GSE2125

OK Cancel Browse GDS Browse GSE

Warning: GSE data sets do not always contain normalized data.



- GEOアクセッション番号を入力するだけで、レコードに含まれる各種データが自動でダウンロードされ、プロットグラフが表示される

- NCIが運営しているがんゲノム解析用のプラットフォーム
- がんの種類ごとに多数のサンプルデータが登録されており、遺伝子発現データのほか、遺伝子変異データやメチレーションデータなどもダウンロードが可能
- Qlucore Omics Explorerでは、25種類のコホートより、RNA-Seqデータをダウンロードすることが可能

The Cancer Genome Atlas Program

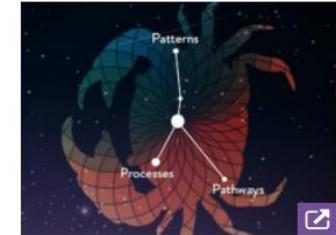
The Cancer Genome Atlas (TCGA), a landmark cancer genomics program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. This joint effort between the National Cancer Institute and the National Human Genome Research Institute began in 2006, bringing together researchers from diverse disciplines and multiple institutions.

Over the next dozen years, TCGA generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data. The data, which has already led to improvements in our ability to diagnose, treat, and prevent cancer, will remain publicly available for anyone in the research community to use.



TCGA Outcomes & Impact

TCGA has changed our understanding of cancer, how research is conducted, how the disease is treated in the clinic, and more.



TCGA's PanCancer Atlas

A collection of cross-cancer analyses delving into overarching themes on cancer, including cell-of-origin patterns, oncogenic processes and signaling pathways. Published in 2018 at the program's close.

Template Browser

The following templates are available in your template folders. To interact with a template, select it below.

10X Genomics Assistant	Load single cell data from 10X Genomics
FTP download, load, and preprocess	Download RNA abundance data from ENCODE
Heatmap sample clustering	Plots a heatmap and orders samples by pairwise
Multigroup anova	Performs an anova on a grouping annotation a
t-test	Performs a t-test and displays a heatmap and
t-test & fold-change	Filter by fold-change and p-value before displ
t-test (paired)	Performs a paired t-test and displays a heatma
TCGA RSEM	Download RNA expression data from TCGA.
Three plots	Displays a sample PCA plot, a heat map and a

This particular template will download TCGA mRNA gene expression data from GDAC (Broad Institute) and downloading constitutes agreement to the TCGA data usage policy. Only datasets with mRNA expression levels (abundances) processed are available through this template, i.e. 27 of 38 cohorts.

This is an advanced template designed to both provide easy access to data but also provide an example for developers/bioinformaticians to write custom scripts for processing on-line information and loading it into Glucose Omics Explorer.

 The following TCGA cohorts are available for download:
 Adrenocortical carcinoma (ACC)
 Bladder Urothelial Carcinoma (BLCA)
 Brain Lower Grade Glioma (LGG)
 Breast invasive carcinoma (BRCA)
 Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC)
 Cholangiocarcinoma (CHOL)
 Esophageal carcinoma (ESCA)
 Head and Neck squamous cell carcinoma (HNSC)
 Kidney Chromophobe (KIPAN)



This is a template for download and processing of TCGA data. Using it constitutes agreement to the TCGA data usage policy.

--- Input fields ---
 `Cohort` : Input only one cohort in this field and it must be the abbreviated name of that cohort.
 See below for a list of cohorts.
 `Normalize` : Write "yes" to download normalized gene expression data. Write "no" to download raw counts. If you intend to use the limma methodology we recommend using the raw data.

--- User notes ---
 It may take a few minutes to download and process the data, please be patient.
 Choose a cohort and click OK to execute when you are ready.
 We suggest saving the dataset to disk after the cohort has downloaded to save time in the future.

--- Available cohorts ---
 ACC,BLCA,LGG,BRCA,CESC,CHOL,ESCA,HNSC,KIPAN,LAML,LIHC,LUAD,DLBC,MESO,PAAD,PCPG,PRAD,SARC,SKCM,STAD,TGCT,THYM,THCA,UCS,UVM

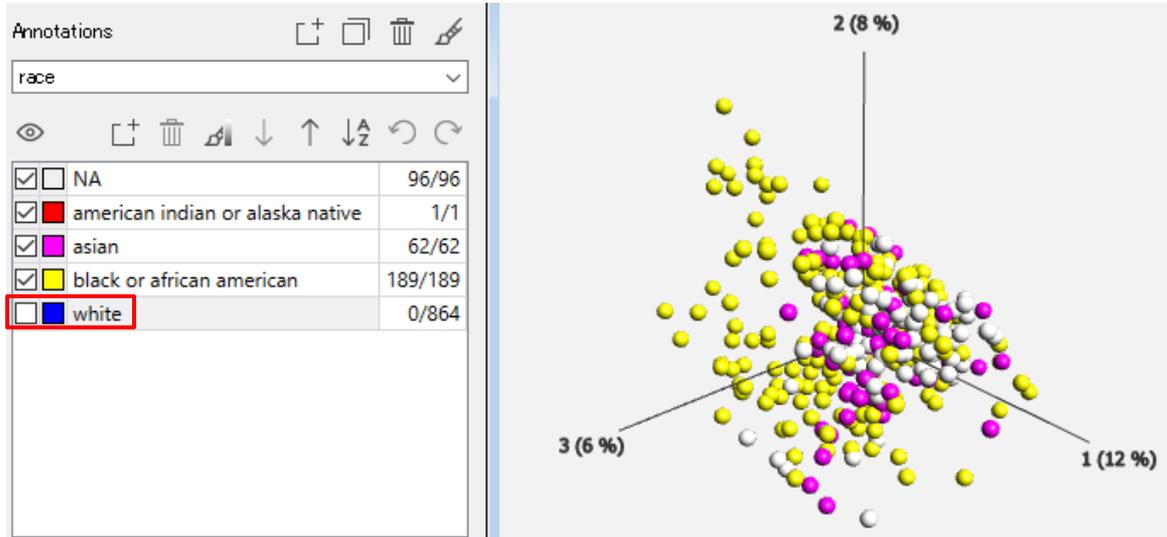
--- Example ---
 The ESCA cohort is fast to download and shows an immediate difference between histology types.
 Type ESCA into the cohort field and click OK.
 This template will automatically color code different histology.

Cohort

Normalize

Execute Close

- コホート名と、取得するデータの正規化の有無を指定



- サンプルアノテーションも同時にダウンロードされるため、アノテーションに基づいたデータの表示や絞り込みなども可能

コホート一覧

- ✓ Adrenocortical carcinoma (ACC)
- ✓ Bladder Urothelial Carcinoma (BLCA)
- ✓ Brain Lower Grade Glioma (LGG)
- ✓ Breast invasive carcinoma (BRCA)
- ✓ Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC)
- ✓ Cholangiocarcinoma (CHOL)
- ✓ Esophageal carcinoma (ESCA)
- ✓ Head and Neck squamous cell carcinoma (HNSC)
- ✓ Kidney Chromophobe (KICH)
- ✓ Acute Myeloid Leukemia (LAML)
- ✓ Liver hepatocellular carcinoma (LIHC)
- ✓ Lung adenocarcinoma (LUAD)
- ✓ Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC)
- ✓ Mesothelioma (MESO)
- ✓ Pancreatic adenocarcinoma (PAAD)
- ✓ Pheochromocytoma and Paraganglioma (PCPG)
- ✓ Prostate adenocarcinoma (PRAD)
- ✓ Sarcoma (SARC)
- ✓ Skin Cutaneous Melanoma (SKCM)
- ✓ Stomach adenocarcinoma (STAD)
- ✓ Testicular Germ Cell Tumors (TGCT)
- ✓ Thymoma (THYM)
- ✓ Thyroid carcinoma (THCA)
- ✓ Uterine Carcinosarcoma (UCS)
- ✓ Uveal Melanoma (UVM)

① 発現変動遺伝子の探索

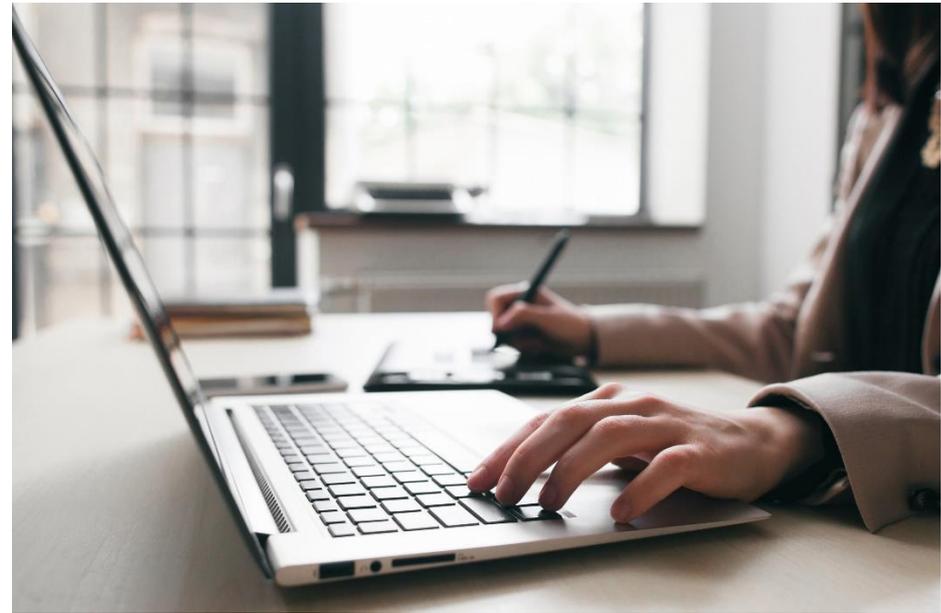
- ✓ 発現量有意差解析
- ✓ ボルケーノプロット
- ✓ ヒートマップ

② 興味のある遺伝子の発現量の確認

- ✓ 発現スケールによる色付け
- ✓ ボックスプロット

③ データセット間での遺伝子リストの比較

- ✓ ベン図



- 発現変動遺伝子の探索には、統計解析用フィルター機能を使用
- サンプルのグループ分けには、遺伝子発現データとともにダウンロードされるサンプルアノテーションをそのまま利用可能
- サンプルアノテーションの種類に応じて、検定手法を選択
 - ✓ 2群間・3群間以上の比較解析
 - ✓ 回帰解析
- 発現変動遺伝子のリストは、テーブルビューまたはIDリストより確認し、ボルケーノプロットやヒートマップで表示も可能

Statistics

Variance Statistics Extended

Input: Variables after variance filtering 42405/42405 vars

Filter by Two Group Comparison ≠

transfected with empty vector (pcDNA3)

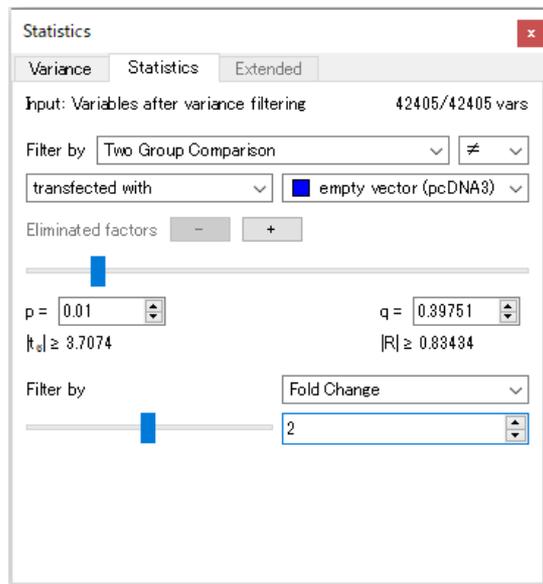
Eliminated factors - +

p = 0.005 q = 0.3682

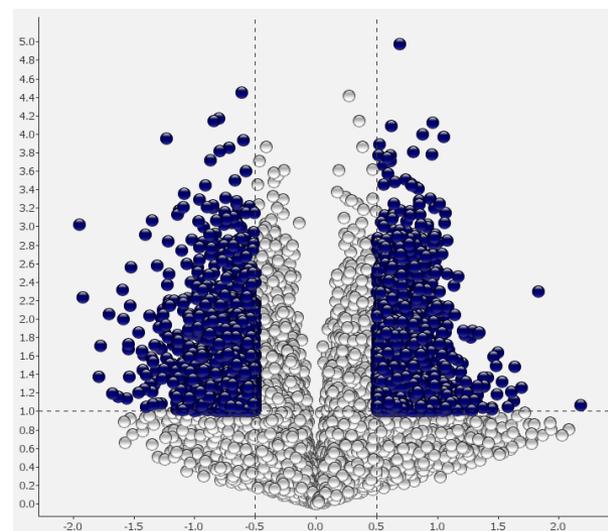
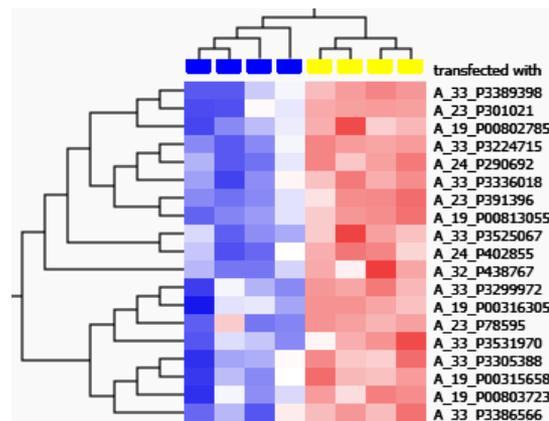
$|t| \geq 4.3168$ $|R| \geq 0.86974$

Filter by Fold Change

2



	p-values	q-values	TPC-1_con	TPC-1_con	TPC-1_con	TPC-1_con	TPC-1_emf	TPC-1_emf	TPC-1_emf	TPC-1_emf
A_19_P003	0.006009	0.3772	3.8992	4.0592	3.8501	4.3071	3.1368	3.5057	2.8421	2.4048
A_19_P003	0.009897	0.3975	4.0592	3.7464	3.9409	4.0754	3.1368	2.6827	3.1672	1.7316
A_19_P003	0.002022	0.3521	2.9627	3.4816	2.6827	3.2358	4.2561	3.8375	4.213	4.1433
A_19_P003	0.0004371	0.3392	3.9928	4.3825	4.1382	4.3071	2.9394	2.9037	3.1672	3.4654
A_19_P003	0.001414	0.3392	3.3552	3.4324	3.3463	3.7055	4.6393	4.2081	4.9822	4.3207
A_19_P008	0.000944	0.3392	3.8992	3.1368	3.8992	3.3195	1.7316	0.89138	2.1032	1.7316

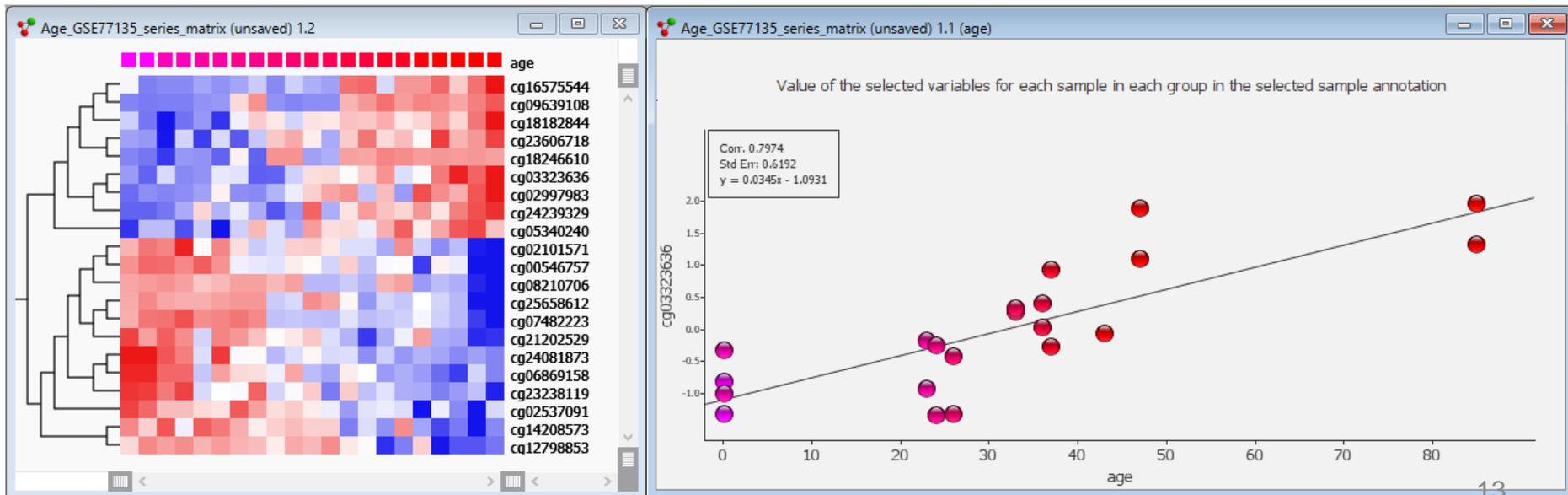


- 2グループ間有意差検定を行い、解析結果をデータテーブル、ヒートマップ、ボルケーノプロットで表示

- 連続値のサンプルアノテーションを検定に用いる場合は、回帰解析を使用する
- サンプルアノテーションによるヒートマップへのラベル付けや、スキャタープロットで各遺伝子の発現スケールをグラフ表示

age

<input checked="" type="checkbox"/>	0.1	2/2
<input checked="" type="checkbox"/>	0.2	2/2
<input checked="" type="checkbox"/>	23	2/2
<input checked="" type="checkbox"/>	24	2/2
<input checked="" type="checkbox"/>	26	2/2
<input checked="" type="checkbox"/>	33	2/2
<input checked="" type="checkbox"/>	36	2/2
<input checked="" type="checkbox"/>	37	2/2
<input checked="" type="checkbox"/>	43	1/1
<input checked="" type="checkbox"/>	47	2/2
<input checked="" type="checkbox"/>	85	2/2



- データセット内から任意の遺伝子を検索したい場合は、データのサーチ機能を使用する
- 遺伝子アノテーションも同時にダウンロードしている場合は、任意のフィールドで検索を行うことも可能
- 検索リストより、任意の遺伝子をクリックすることで、選択遺伝子を各種プロットに反映させる

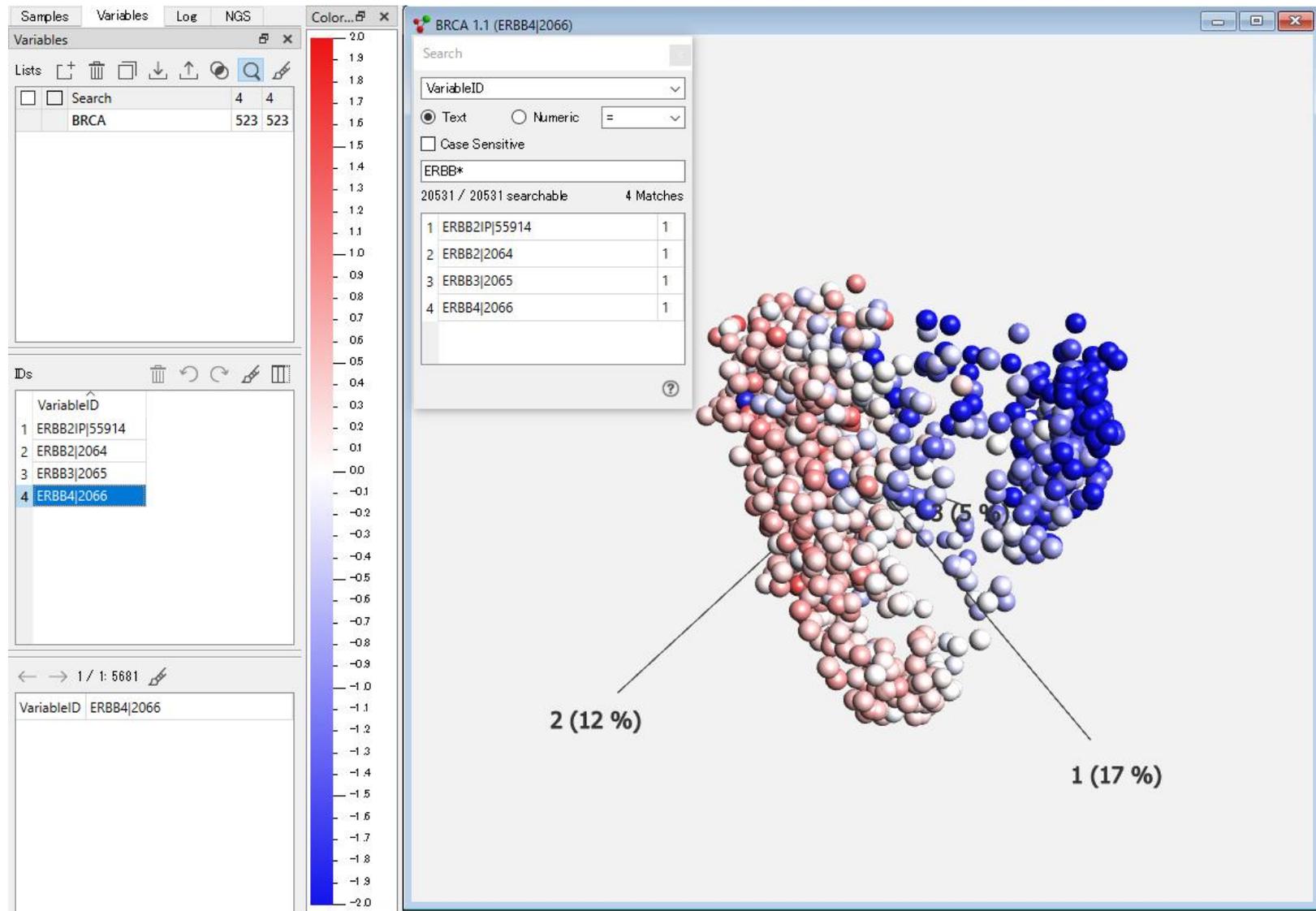
The screenshot shows a search window titled "Search" with a close button (x) in the top right corner. The search criteria are as follows:

- Search field: VariableID
- Search type: Text, Numeric
- Operator: =
- Case Sensitive:
- Search term: ERBB*

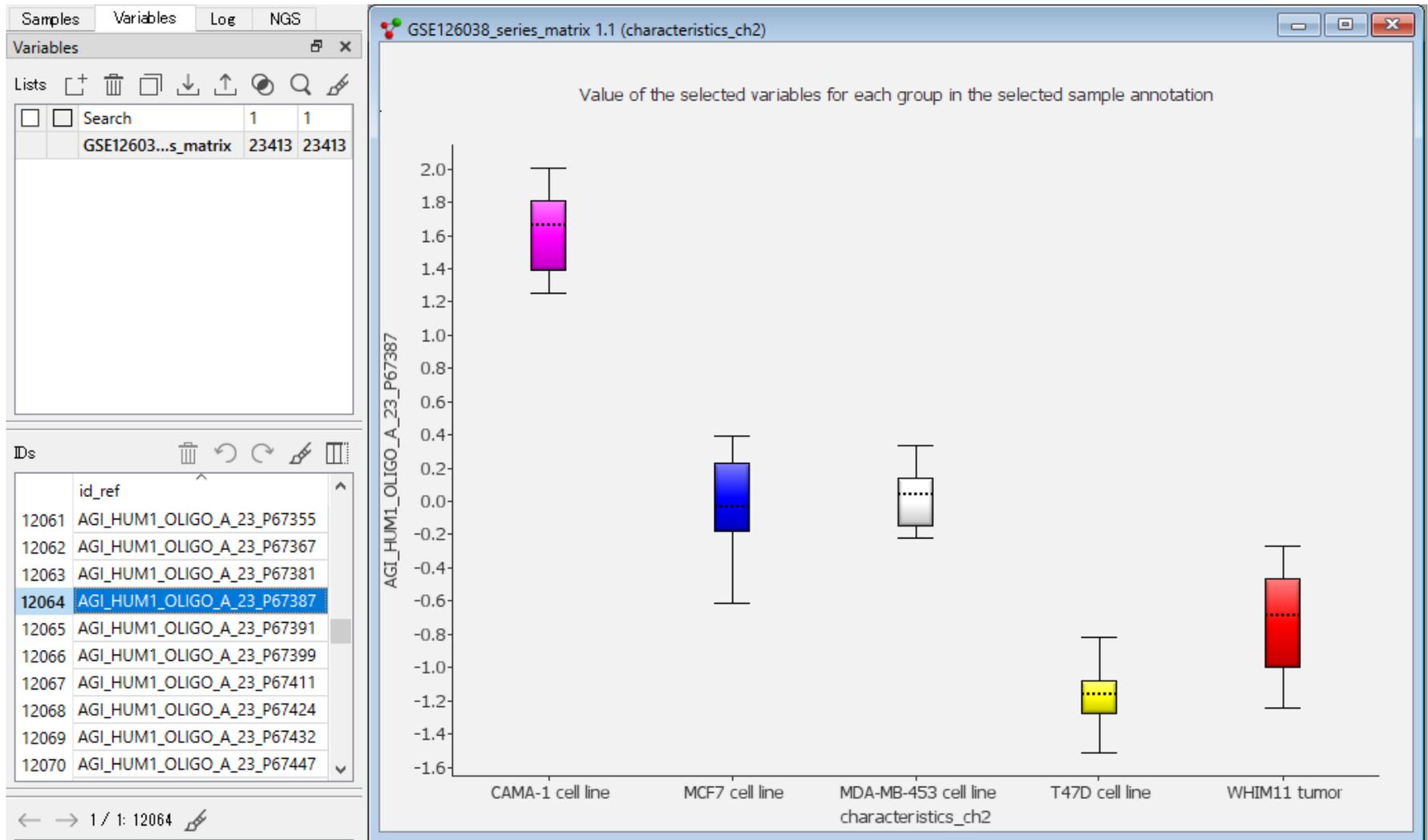
Below the search criteria, it indicates "20531 / 20531 searchable" and "4 Matches". The results are displayed in a table:

1	ERBB2IP 55914	1
2	ERBB2 2064	1
3	ERBB3 2065	1
4	ERBB4 2066	1

A help icon (?) is located at the bottom right of the search window.



- 興味のある遺伝子をデータセット内から検索し、発現スケールに基づきプロットを色付け



- ボックスプロットにより、興味のある遺伝子の発現量をサンプルアノテーション間で比較

- 統計解析用フィルターを用いて検索した発現変動遺伝子は、リスト登録を行うことで、後からデータの呼び出しが可能
- リスト登録されたデータの詳細は、IDリストより確認する

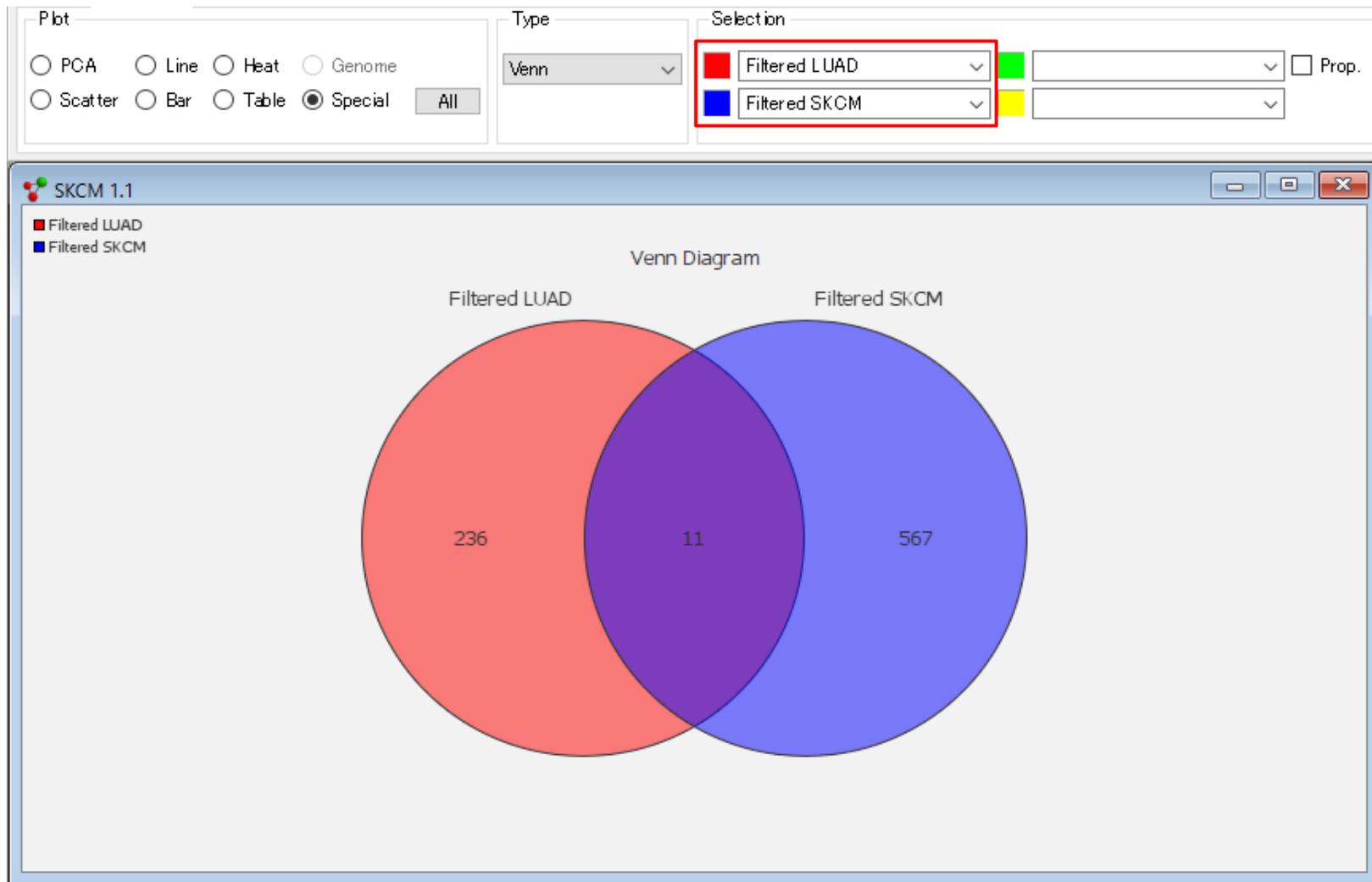
The screenshot displays the 'Variables' window in the Filgen software. It features a 'Lists' section with a table of variables and their counts, and an 'IDs' section with a list of variable IDs.

Variables

Search	Count	Count
<input type="checkbox"/> Search	0	0
<input type="checkbox"/> LUAD	20531	20531
<input type="checkbox"/> SKCM	20531	20531
<input type="checkbox"/> LUAD_P0.05	1470	1470
<input type="checkbox"/> LUAD_P0.01	358	358
<input type="checkbox"/> SKCM_P0.05	2455	2455
<input type="checkbox"/> SKCM_P0.01	915	915

IDs

VariableID
1 ? 57714
2 ABCC13 150000
3 ABCD2 225
4 ABHD11 83451
5 ABHD3 171586
6 ABRA 137735
7 ACADM 34
8 ACAP2 23527
9 ACER2 340485
10 ACTR2 10097



- 登録済みの遺伝子リストを選択して、複数実験の発現変動遺伝子の比較をベン図で表示

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