SUBIO PLATFORM & PLUG-INS ユーザーガイド

Subio Platform は高度にイン タラクティブな視覚化ツール を備えたオミクスデータ専用 のブラウザーで、どなたでも無 料で使えます。

Basic および Advanced のプラ グインは、あなたのアイデアを データで検証するのをお手伝 いします。

SUBIO INC.

2019/09/24

Overview7
Workflow8
インストールについて9
Subio Platform for 64-bit Windows10
Subio Platform for Mac10
バージョンアップ11
バックアップ11
Menu について
Platform12
Organism12
Genome
View
Plug-in
Help
Data Browsing
Data Manager15
Import/Export Series 😻 🏠
Region List Panel
Series Panel16
Analysis Browser
Main Graph18
2

Line Graph 🖗	
Scatter Plot (Measurements) View 📓	
Tree View 👜	
Pathway View 🂐21	
Genome View XX	
Scatter Plot (Samples) View 📓	
Venn Diagram 🌑	
Setup Series タブ26	
Setup DataSet タブ	
Sample Info.タブ26	
Datasheet タブ27	
Annotations タブ27	
Genome tab & Regions タブ28	
Chromosome タブ	
Preparation	
The Data Structure	
Platform	
Sample	
Series	
Organism	
Genome	
3	

Data Manager
Platform list
Edit Platform 🖉
Select Symbol Column 🙎
Merge Platform 📔
Assign To Other Organism 👯
Import Samples 4
Importing RNA-Seq or ChIP-Seq data
Multiple Samples in One File
Format について
1 色法(Single-Channel)のデータフォーマット
2 色法 (Dual-Channel) のデータフォーマット
Number Separators
Manage Format
Import Options
Filtering by Signal
Filtering by Location
Merging/Counting
After you import samples
Create A Series
Setup Series タブ40
4

Edit Parameters ᢞ	
Edit Flag Columns 🏲	
Normalization	
Setup DataSet タブ	
Sample Info タブ46	
The Plug-in License	
Serial Number and Activation	
Basic Plug-in Tools	
Filter	
Find Similar Pattern	
Compare To Control	
Compare 2 Groups	
Compare One to All	
Compare All Pairs	
Compare Multiple Groups	
Tree Clustering	
主成分分析 (PCA)	
Advanced Plug-in Tools	
Enrichment Analysis	
Pathway Edit Tool	
Genes Tied in Parameter59	
5	

	Genomic Location Filter	. 59
	Measurement to Region	. 61
	Summarize	. 61
	Create Intervals	. 62
	Region Score Filter	. 63
	Get Sequence	. 63
	Find Regions from Seq	. 64
	Find miRNA Targets	. 65
	Find Correlated Regions	. 68
	Scatter Plot of Regions	. 69
	Annotate Measurements	. 69
1	Need a help?	70
	Free Online Technical Support & Training	. 70

OVERVIEW



Step 0-2: データ解析というとこの部分だけを想像する人が多いようで すが、そうではありません。この作業はわたしたちや「専門家」に依頼 することができますが、それで研究の答えが得られるわけではありませ ん。GO 解析やパスウェイ解析、ネットワーク解析などというと、なに か生物学的な意味を引き出してくれそうに思いますが、実際には可能性 のあるモデルを提示するだけで、本当に細胞の中で起こっていると示す ものではありません。むしろ無数の間違ったモデルと、実際には起こっ ているのに提示されないことが混在しています。

Step 3-5: データ解析の心臓部です。最も難しいところですが、解析結 果から生物学的洞察を引き出せるのは、あなたしかいません。細胞の中 で実際に何が起こっているのかディスカッションして、浮かび上がった 仮説を検証できるのは実験しかありません。バイオインフォマティクス は、仮説と実験のアイデアを提供してくれるにすぎません。

Subio Platform とプラグインは、バイオインフォマティシャンから解 析のタスクを引き継いだバイオロジストが、誰でも自分でデータを探索 して洞察を引き出すことができるようにつくられています。

WORKFLOW



https://www.subioplatform.com/ja/analysis_guide 8

"Standard Service"の列にチェックがついている項目は、誰かに任せる ことができます。それ以外のところは、データの解釈に関するもので、 あなた自身でやる必要があります。

そこで、まずは人に任せられない項目に集中することをおすすめします。 その操作が自在にできるまで慣れてきたら、それ以外のところの操作を 少しずつ学んでいくとよいでしょう。このような理由から、このユーザ ーガイドでは、まず"Data Browsing"を学びます。

まずはマニュアルを読みながらできるだけ自分でやってみようという 人が多いのですが、それは多くの時間を無駄にします。まずは無料のオ ンラインサポートをご利用ください。ある程度データの見方を理解して からこのマニュアルを補助的に使うほうが、ずっと効率よく学べます。

インストールについて

Subio Platform のインストーラーをダウンロードします。

64-bit Windows 版と Mac OS 版の 2 種類がありますので、適切なイン ストーラーをダウンロードしてください。

Subio Platform を起動できたからといって、どんなデータでも解析で きるわけではありません。ロードできるデータの大きさは、物理的に搭 載しているメモリー容量に依存します。大きなデータセットを扱うには、 大きなメモリーが必要です。

関連項目

- ✓ Subio Platform を起動できません。どうしたらいいでしょうか?
- ✓ どのくらいのメモリーが必要ですか?
- ✓ どれくらいのサンプル同時に解析することができますか?

9

- ✓ <u>"Out of Memory" というエラーが出ました。どうすればいいです</u> か?
- ✓ Subio Platform の動作がとても遅いです。なにか解決策はありま すか?

SUBIO PLATFORM FOR 64-BIT WINDOWS

インストール

ダウンロードした ZIP ファイルを解凍するだけです。デスクトップや ドキュメントなどのユーザーディレクトリに置くことをおすすめしま す。ProgramFiles などのシステムディレクトリには置かないでくださ い。インストールディレクトリのパスに2バイト文字が含まれていると うまく動作しないことがあります。もし正常に動作しない場合は、日本 語名のフォルダー以外の場所に置いて試してみてください。

Subio Platform フォルダー自体は、同じコンピューターに複数持つこ とができます。この場合、それぞれのインスタンスは独立して動作しま す。研究者やプロジェクトごとにデータを分けて管理・解析することも 可能です。後述のプラグインは、同一のコンピューター上であれば、す べてのインスタンスで有効化できます。

SUBIO PLATFORM FOR MAC

インストール

subioplatform.dmg をダブルクリックしてマウントしてください。その 中の Subio Platform Installer.pkg をダブルクリックしてインストール すると、Applications ディレクトリに"Subio"フォルダーができます。

WINDOWS版とMAC版の操作上の違いについて

操作は Windows 版とほぼ同じですが、次の点が異なります。

10

- ✓ Control キーと書いているところは、Command (Cmd)キー
- ✓ 右クリックと書いているところは、Control キーを押しながらク リック

バージョンアップ

Subio Platform と Plug-in は常に更新されており、どなたでも無料で最 新版にバージョンアップすることができます。テクニカルサポートは最 新版のみ対応します。新しいバージョンがあるかどうか確認するには、 Help メニューから"Check for Updates..."を選択してください。

LEARN OPERATIONS

✓ <u>Subio Platform を更新する</u>

ネットワーク環境により自動更新ができない場合は、手動で下記のリン クからアップデータをダウンロードして実行してください。

 $\underline{https://www.subioplatform.com/info~general/latest~release}$

バックアップ

Subio Platform は無料で配布しており、データの保全を保証している ものではありません。くわしくは End User License Agreement をお読 みください。データのバックアップはユーザーの責任で行ってください。 完全なバックアップを取るには、インストールディレクトリをまるごと コピーすることです。

SSA ファイルをや SOA ファイルを出力しておくのでも有効です。

MENU について

PLATFORM

TRASH CAN

使っていくうちに Data Manager や Series Panel にデータが蓄積され ていきますが、いらなくなったものはゴミ箱に移動することができます。 間違って捨ててしまっても安全に元に戻せますので、安心してお使いく ださい。

IMPORT ARCHIVE

SSA (Subio Series Archive)ファイルや SOA (Subio Organism Archive) ファイルを Subio Platform にインポートすると、出力した状態を復元 することができます。ここからファイルを選択してください。

PREFERENCES

Main Graph の描画色を変更することができます。

Mode を"Quality"から"Speed"へ変更すると、描画は荒くなりますが処 理が速くなります。

Datasheet Export Mode を "**Integrated**"から"**Separated**"へ変更すると、 データを出力して R で解析するのがより簡単になります。

ORGANISM

Organism の切り替えを行います。

Genome

ゲノムをロードしたり、切り替えたり、アンロードしたりします。

12

VIEW

Main Graph に描画する View や、下段のタブ、そして Setup Series タ ブ内のヒストグラムの表示・非表示を選択することができます。デフォ ルトではすべて表示する設定になっていますが、特に大きなデータセッ トを扱う際は非表示にすることでソフトの応答性を向上できます。

LEARN OPERATIONS

✓ <u>View やタブの表示・非表示を切り替える</u>

Plug-in

有効なシリアルナンバーを使ってプラグインを activate します。また、 ご利用中のプラグインの有効期限もここで確かめることができます。

プラグインを使ったことがない方は、ぜひ5日間のフリートライアルを お試しください。

✓ <u>5 Days Free Trial</u>

Help

ONLINE SUPPORT & TRAINING

プラグイン購入の有無にかかわらず、すべての Subio Platform ユーザ ーは、無料のオンラインサポートトレーニングを受けられます。

MOVIE TUTORIAL

ソフトウェアの操作を学ぶには、マニュアルを読むよりムービーを見る ほうが適しています。一方マニュアルは、用語やコンセプト、ムービー にないショートカットなどを学ぶのに適しています。

TROUBLESHOOTING

Subio のウェブサイトの Instant Help にはたくさんの情報が蓄積され ており、ここから直接この知識ベースを検索することができます。

EXPORT TECH REPORT

問題解決のために"tech report"を送ってくださいとお願いすることが あるかもしれません。そのときは、ここからファイルを出力してメール でお送りください。

CHECK FOR UPDATES

Subio Platform の最新版がないかチェックすることができます。最新版があればバージョンアップしてください。無料です。

ABOUT SUBIO PLATFORM

Subio Platform を論文等に記載するときは、ここでバージョン番号を 確認することができます。株式会社 Subio については、Subio Inc., Amami city, Kagoshima, Japan です。

DATA BROWSING

DATA MANAGER

Data Manager タブは画面の最上部で開きます。

上段パネルには Platform List があります。任意の Platform を選択す ると、下段パネルの Sample List や Series List では、その Platform に 関連付けられている Sample や Series が表示されます。Sample や Series はキーワードで素早く検索することができます。

Series List にある Load a Series (ボタンを押すと、そのデータが Analysis Browser で視覚化されます。Subio Platform は一度に一つの Series しか開くことができません。画面左側にある Series Panel の右 上部には Unload ()ボタンと Lock ()または Unlock ()がありま す。ダブルクリックしてロックすると、Series に誤って変更してしまう ことを防止できます。

Data Manager で **remove**(**o**)する際は、まず Series を、次に Sample を、そして最後に Platform を Remove してください。関連するオブジ ェクトがあるときは remove することができないよう安全な設計になっ ているためです。Remove したデータは **Trash Can** の中にありますの で、あやまって remove してしまっても安全に復元することができます。

Import/Export Series 🔮 🁚

プラグインを使った解析結果も含めて、Series に関するすべてのデータ を SSA ファイルにして出力することができます。Series List のツール バーにある Export Series (个)ボタンを押してください。 SSA ファイルをインポートする場合は、Series List にドラッグ&ドロ ップするか、Series List のツールバーにある import series (●)ボタン を押すか、Platform メニューから "Import Archive..."を選択します。

SSA ファイルはバックアップだけでなく、他の共同研究者とデータを 共有するのにも有用です。PC や Mac を超えて Subio Platform 上でデ ータを復元できるので、チーム内での共同作業やディスカッションを円 滑にします。ラボの責任者もレポートの代わりに生データを自分でも解 析できる状態で受け取ることができます。

✓ オミクスデータを Subio Series Archives (SSA) ファイルでかん たん共有

REGION LIST PANEL

Region Lists は、遺伝子に限らず、たとえば遺伝子の上流域や CpG islands、転写因子の結合領域、CNV 領域、SNP などゲノム上の位置に よって定義されるものの集合です。 Retgions タブや Genome タブにあ る Save as Region List (こ)ボタンや、Advanced Plug-in の解析ツール などで作られます。または、GFF や BED ファイルを Import (④) し て作ることもできます。Export (④) した場合は BED ファイルが生成 され、他のゲノムブラウザーに取り込むことができます。

SERIES PANEL

Series は Sample を集めて解析の単位にしたもので、Subio Platform で は Series の単位でロードして視覚化したり解析したりします。NCBI の Gene Expression Omnibus (GEO)のデータ構造に似た概念です。

http://www.ncbi.nlm.nih.gov/projects/geo/info/overview.html

Series Panel は画面の左側にあって、様々なデータオブジェクトを管理 しています。Add Folder ()ボタンでフォルダーを追加でき、ドラッ グ&ドロップで移動したり入れ子にしたりできますので、データを整理 するのに使ってください。オブジェクトやフォルダーを長押しすると名 称変更モードになります。編集したら Enter キーで確定してください。

ツールバーにある **import** (◆)や **export** (↑)ボタンで、Subio Platform のデータをテキストファイルにしたり、タブ区切りテキスト形式のファ イルを取り込んでオブジェクトを作ったりできます。

- ✓ <u>Measurement List を他の Series 移すには?</u>
- ✓ パスウェイのデータを他の Series に移すには?

MEASUREMENT LIST 🗾 🔏

Measurement は一つの測定値に関連付けられる ID に相当します。そ して Measurement List はその集合です。R マークついている Measurement List は、フォールド値やP 値などの数値を持っているこ とを表しています。

PROFILE 🤐

主成分分析 (PCA) の結果は **Profile** のセットとして保存され、このオ ブジェクトをもとに Scatter Plot (Samples) View (上)は描画されます。 Profile を縦軸や横軸にドラッグ&ドロップでセットすることにより、 その Profile による Sample Group のスコアが視覚化されます。

DATASET 💹 AND SAMPLE GROUP 🐠

DataSet は **Sample Group** を束ねるもので、ある意味 GEO の GDS に 相当するものといえます。

TREE 🔩

Tree は階層クラスタリングの結果として生成され、Tree View (44)で 描画されます。Tree オブジェクトは DataSet と関連付けられており、

17

作成時に使われた DataSet の配下に保存されます。Series Panel で別の DataSet を選択すると別の Tree が表れるのはそのためです。

Ратнуау 📲

Pathway は背景のイメージと、そのイメージの上に配置される Measurementのセットで、Pathway View (**谷**)で描画されます。

ANALYSIS BROWSER

ロードされた Series を閲覧して解析するのが Analysis Browser です。 上段パネルには Main Graph が描画され、下段パネルにはタブが並ん でいます。上段と下段の間にある水平のバーはドラッグして上下に動か すことができます。見やすいようにお好きなサイズに調整してください。

MAIN GRAPH

Main Graph では 6 つの種類の View を描画することができます。 View メニュー、または Main Graph のすぐ上にあるツールバーで切り替える ことができます。 すべての View は Series Panel で選択されているオブ ジェクトを描画します。

Maing Graph で描画されているものは Save Graph Image (
の)ボタン
で PNG 形式の画像ファイルとして出力できます。

Line Graph 🏁

Line Graph View はもっとも頻繁に使う View で、値の増減(変化)を 見るのに適した View です。一本の線は一つの Measurement に相当し ます。横軸上のデータポイントは選択された DataSet に含まれる Sample Group に相当します。 左上のプルダウンメニューから、Processed Signal を描画するか、Ch1 (or Ch2) Raw signal や Ch1 (or Ch2) Reserved signal を描画するか選 択することができます。縦軸のラベルをダブルクリックすると軸の表示 設定(最大値と最小値、そして目盛り)ができます。

Line Graph View の上でドラッグすると、その範囲(ピンク色の長方 形)を通る Measurement が選択状態になります。選択された Measurement は黒い線で表され、Annotations タブや Datasheet タブ では黄土色で表されます。さらに、Annotation や Datasheet テーブル の上で任意の一行をクリックすると Active 状態になり赤褐色で強調表 示されます。 Active 状態になれるのは一度に一つの Measurement だ けです。

選択した Measurement のセットは、上のツールバーにある Save a Measurement List (デオタンで Measurement List として保存できます。ドラッグによる選択と Measurement List の保存を何度か繰り返すだけで、おもしろいと思う発現パターンの遺伝子群を簡単に抽出することができます。

おおくの人が統計検定によって生物学的に重要な遺伝子を抽出できる と思っているようですが、統計学的有意と生物学的有意はまったく別物 ですからそのようなことは不可能です。そこで、ただ P 値を見るだけで なく、Line Graph で Processed Signal (対数比・変化率)と Ch1RawSignal (インテンシティ・絶対量)を切り替えながら見ることを おすすめします。発現量が高い遺伝子群は、多くが構造タンパクとエネ ルギー代謝関連の酵素です。転写因子やシグナル伝達経路の要素の多く はそう多く発現していません。変化率にだけではなく、絶対量 は生物 学的文脈では非常に重要なはずです。

横軸の下にあるスライダーは任意の Sample Group に動かすことがで きます。そうすることで Line Graph で描画される線の色が変わること に注目してください。スライダーによって支持される Sample Group に

19

おける発現比や発現量によって線の色が決まります。線の色を大まかに 見るだけで、どの Sample Group どうしが似ているか、あるいは違って いるかをほぼ見分けることができます。

LEARN OPERATIONS

✓ <u>Line Graph View</u>の使い方

SCATTER PLOT (MEASUREMENTS) VIEW

散布図上の一つ点は一つの Measurement に相当します。散布図の上で ドラッグすると、その範囲(ピンク色の長方形)にある Measurement が選択状態になります。この View は主に二つの使い方があります。一 つ目は、同じ Sample Group における Ch1RawSignal (量) と ProcessedSinal (比)の関係を視覚化することで、もう一つは、こちら のほうが馴染みのある方が多いと思いますが、二つの Sample Group 間 で Ch1RawSigna がどのくらい似ているかを見ることです。縦軸と横軸 には任意の Sample Group を Series Panel からドラッグ&ドロップで セットすることができます。

散布図の右のところで、縦軸と横軸にそれぞれ何の値を表示するかを Processed Signal、Ch1 (or Ch2) Raw Signal、Ch1 (or Ch2) Reserved Signal)の中から選択できます。

LEARN OPERATIONS

✓ <u>Scatter Plot (Measurement) の使い方</u>

Tree View 斗

Series Panel で Tree オブジェクトを選択すると、Main Graph は自動 的に Tree View に切り替わります。

20

ヒートマップの左には Measurement Tree があり、上には Sample Tree があります。Sample Tree は、右下にある"Show Sample Tree"オプシ ョンにチェックを外すことで隠すことができ、このときは Sample Group の並び順は Line Graph View と同じになります。

Data Manager の Platform List のツールバーで、"Select Symbol column"をセットしていればヒートマップの右側に遺伝子名が表示されます。ただし、表示するのに十分なスペースがあるときだけです。

Measurement Tree のノードをクリックすると、そのノードの配下にあ る Measurement が選択状態になりますので、上のツールバーの Save as Measurement List (デンボタンで Measurement List として保存す ることができます。 PDF や画像のツリーは動きませんが、 Subio Platform でデータをシェアしていれば、ユーザーが任意のクラスター を選択して遺伝子を抽出することができるのです。

一方、Sample Tree のノードをクリックすると、その配下にある Sample が Sample Info タブや Scatter Plot (Samples) View でも選択状態にな ります。Edit Parameters 画面では選択された Sample を、**Import** Status of Selection を使って"1"とマークすることができますので、発 現プロファイルから見えてきたサブグループをまとめることが簡単に できます。

最後にヒートマップは、Series Panel からテキストファイルとして export (♠)することができるので、エクセルで描画することもできます。

LEARN OPERATIONS

- ✓ <u>Tree View</u> の使い方
- ✓ Export the heatmap table of a tree

Pathway View 🦓

Series panel で Pathway オブジェクトを選択すると、Main Graph は 自動的に Pathway View に切り替わります。多くの方がデータ解析は 関連のありそうなパスウェイを抽出してその絵を取って終わりと思っ ているようですが、実際にはパスウェイ上で測定された発現パターンを 解釈するのは非常に困難です。従って PDF や画像のように静的なフォ ーマットでは不十分で、一つ一つのパスウェイ上のエレメントを丁寧に 追っていけるインタラクティブなビューワーが必要なのです。

他の View と同様に、パスウェイ上をドラッグすることで、描かれたピ ンクの長方形の中にある Measurement を選択できます。多くの場合、 ーつに見える箇所に複数の Measurement が重なっています。 Annotation テーブルで選択された Measurement から一つずつ Active にすることで、それぞれの Measurement の発現パターンを見ることが できます。

Space キーを押したままドラッグすると、Pathway View に表示してい る領域を動かことができます。

左上にあるオプションパネルは [●] や [●]ボタンで開いたり閉じたりでき ます。オプションパネルのスケールバーを使ってズームインしたりズー ムアウトしたりできます。緑色の長方形は Pathway View で表示して いる領域を表しており、これをドラッグして動かすことでも表示領域を 動かすことができます。グラフモードは "Color Bar"と"Bar Graph"か ら選択できます。

LEARN OPERATIONS

✓ <u>Pathway View</u>の使い方

GENOME VIEW 🐰

Genome View は Genome データがロードされているときのみ有効にな ります。Genome View がアクティブでないときは、まず Genome メニ ューから Genome を選択するか、作成してください。

統計検定で得られる統計学的有意差は生物学的有意差とは全く別物で す。P値のリストを眺めるよりも生物学的な見方をしたほうが発見につ ながるかもしれません。その一つが Genome View です。もし発現が上 昇あるいは減退する遺伝子が染色体上の狭い箇所に固まって存在する ならば、それらは転写因子による制御というよりも、エピジェネティク スや染色体の構造変化による制御を受けている可能性が考えられます。

Genome View は、Measurement や Region をゲノムの位置情報に従っ て描画します。大きく3つの部分に分かれます。最初のピンク色のトラ ックは、ロードされている Genome のエレメントを表しています。次の 緑色のトラックはそれぞれ Region List 相当します。Region List Panel から任意の Region List をドラッグ&ドロップで追加してください。ト ラックを除く場合はドラッグアウトします。そして残りは Series Panel で選択されている DataSet の Sample Group および Measurement List に相当します。

トラックの左側のラベルをダブルクリックすると、描画オプションを設 定する画面が開きます。

下段の Chromosome タブには、仮想染色体が描画されています。ピン ク色の長方形で囲まれた部分が、Genome View で表示されている領域 です。Chromosome タブ内の仮想染色体上で別の個所を囲むことで、描 画する領域が変わります。

右上の角には Genome View Navigation のコントロールがあります。 ここで Genome View で表示する領域のズームイン・ズームアウト・左 右のパンニングができます。後述のショートカットでも操作可能です。.

Zoom-in

Move-left Zoom-out

ショートカット

矢印キー(右・左)	描画領域を左右へパンニング
Ctr キー+ 矢印キー	前後の Region エレメントヘジャンプ
(右・左)	
矢印キー(上・下)	上下スクロール.
Ctr キー+矢印キー	ズームイン・ズームアウト
(上・下)	
マウスホイール	上下スクロール.
Ctr キーを押しながら	ズームイン・ズームアウト。ただし、カー
マウスホイール	ソルのある位置が中心にくる。
Space キーを押しながら	マウスと動かしただけ左右にパンニング
ドラッグ	
	1

LEARN OPERATIONS

- ✓ Genome View の使い方
- ✓ Genome データを作成する

Scatter Plot (Samples) View 💆

この View は主成分分析 (PCA) の結果を閲覧するのに用います。Series Panel から **Profile** オブジェクトを縦軸あるいは横軸にドラッグ&ドロ ップでセットすると、それにもとづいて Sample Group がスコア化され て散布図上に点として現れます。大雑把にいうと、近くにある Sample Group どうしは発現プロファイルが似ていることを表します。散布図上 をドラッグすると、ピンク色の長方形の中にある Sample が選択状態に 24

なります。選択状態になった Sample は、Sample Info タブや Tree View でも強調表示されます。Edi Parameters 画面では、選択状態の Sample を Import Status of Selection を使って"1"とマークすることができるの で、発現プロファイルが近いサブタイプを簡単に見つけてまとめること ができます。

Line Graph における Sample Group の順序は、Arrow (**イ**)ボタンに よって表現することができます。

散布図の右にあるカラーバーをダブルクリックすると、Parameter を選 択するプルダウンメニューが表れます。ここで選択した Parameter に よって散布図上の Sample Group は色分けすることができます。

LEARN OPERATIONS

✓ <u>Scatter Plot (Samples) View</u>の使い方

Venn Diagram 🏾 🍨

このツールでは二つのタブが使えます。

VENN DIAGRAM TAB

Series Panel から任意の Measurement List をドラッグ&ドロップで 3つのうちいずれかかの輪にセットします。輪から Measurement List を外すにはドラッグアウトします。複数の Measurement List をセット したら、重なり合った任意の部分をクリックすると黒くなります。もう 一度クリックすると解除されます。黒くなった状態で **Preview** ボタン や **Save List** ボタンを押すとその部分に相当する Measurement を選択 したり、Measurement List として保存したりできます。

OVERLAPS TAB

4つ以上の Measurement List をかけ合わせたい場合は、**# Overlaps** 使 ってください。

25

ここには Measurement List を好きなだけ Drag-and-drop でセットす ることができます。最後にオーバーラップ数の最小値(min) と最大値 (max) をセットして Measurement List を抽出します。すべての Measurement List の和集合がほしい時は、min に"1"、 max にセッ トした Measurement List の数を淹れてください。逆にすべての関集合 がほしい時は、min にも max にもセットした Measurement List の数 を淹れてください。Max と min の両方に 0 と入れた場合は、セットし た Measurement List のどれにも含まれない Measurement List を作 ることができます。

LEARN OPERATIONS

✓ <u>ベン図ツールの使い方</u>

SETUP SERIES タブ

この Series にどのような正規化、データの前処理が施されているかを 見ることができます。また、Ch1RawSignal や ProcessedSignal の分布 をヒストグラムで見ることができます。

Setup DataSet タブ

DataSet や Sample Group はここで定義することができます。チェックした Parameter で同じパラメータ値を持つ Sample はグループにまとめられて、それらの平均値が Sample Group の値となります。

SAMPLE INFO.タブ

Images & Parameters セクションでは Sample Image とパラメータ値 が表示されています。Notes & Files セクションでは、自由にメモを残 すことができます。また、File というところにドラッグ&ドロップで添 付ファイルをつけることができます。添付ファイルを消すには、ドラッ グアウトします。

DATASHEET タブ

Datasheet テーブルには Processed Signal や Ch1 (or 2) Raw signal、 Ch1 (or 2) Reserved signal、flag 値、Gene Symbol が表示できます。 操作は次の Annotations テーブルとほぼ同じです。

Annotations タブ

Annotations テーブルは、現在ロードされている Platform のアノテー ション情報が表示されます。一行は一つの Measurement に相当し、行 数は現在選択されている Measurement List の数と同じです。 Measurement List に含まれる Measurement の数や選択状態になって いる Measurement の数も表示されます。

Main Graph で選択した Measurement は、Annotations テーブルでも 黄土色で強調表示されています。さらに左端の列のトグルボタンで選択 状態を追加したり除いたりすることもできます。表示されているテーブ ル全体は Save a Text File (局) や Copy Table (一)で出力して他のソ フトウェアで利用することもできます。

Annotations テーブルでクリックすると、その行(Measurement)が Active 状態になり、赤黒色で強調表示され、Main Graph でも同じよう に強調表示されます。同じ行をもう一度クリックすると解除できます。

27

Annotations テーブルの上にあるテキスト入力欄では、キーワードで Search ())することができます。Search で見つかった Measurement は選択状態になります。画面右端にある Select Columns 引出しタブ(緑 色)をクリックすると、テーブルに表示する列を選択できます。表示し たい列のタイトルにチェックを入れて、表示したくないものはチェック を外してください。複数のタイトルを選択して Check Selected ())や Clear Selected ())ボタンを使うと一気に操作できて便利です。並び順 はドラッグ&ドロップで変えることができます。

WEB SERACH

Annotaions テーブルの上で右クリックすると、カーソルがさしている 遺伝子についてウェブデータベースで検索するリンクが表れます。この ウェブリンクは自由に編集することができます。

✓ ウェブリンクを定義する

GENOME TAB & REGIONS タブ

操作は Annotations タブとほとんど同じですが、表示している内容が異 なります。Regions タブは選択中の Region List に含まれる Region の 一覧を表示し、Genome タブは現在ロード中の Genome に含まれるエ ントリーの一覧を.表示します。

Chromosome タブ

Chromosome タブは Genome データをロードしているときのみ表示され、その Genome データから作られる仮想染色体を描画します。

画面右側の Select Chroms 引出しタブ(緑色)をクリックすると、描画 する染色体の選択と、その順序の設定ができます。

仮想染色体上のピンク色の長方形は、Genome View で描画している領 域を表しています。仮想染色体上の任意の箇所をドラッグして長方形を 描き直すと、Genome View もそれに合わせて描き直されます。

仮想染色体の右側に三角形が表示されていたら、選択中の Genome エ ントリー (--)、Region (--)、または Measurement (--)の位置を表してい ます。

PREPARATION

ここから先のことは、データをインポートしたり、プラグインを使って 解析作業を行ったりする人にだけ必要です。もしこれらの作業を他の人 に任せるのであれば、これより前の章で記述されている Viewer を使っ てデータを解釈することに集中してください。

THE DATA STRUCTURE

PLATFORM

Platform とは、Measurement ID とアノテーションのテーブルです。 すべての Sample と Series は必ず一つの Platform と関連付けられてい ます。同じマイクロアレイを使った実験データは、一つの Platform の に関連付けられた Sample 群として保管してください。新しいデータが 出るたびに Platform を作り直す必要はありません。

$\mathsf{S}_{\mathsf{AMPLE}}$

Sample とは、ひとつのマイクロアレイ、あるいは別の実験の一回分の 結果に相当します。実験データファイルをインポートすると、一回分の 実験データに対して Sample を一つずつ作ります。

Series

Series とは、複数の Sample を解析のためにひとまとまりにしたもので す。Subio Platform は一度にひとつの Series だけロードして、Analysis Browser で描画することができます



ORGANISM

Organism とは、複数の Platform およびそれらに関連付けられた Sample と Series をまとめて、さらに Genome と Region List 群も含 めてまとめたデータの塊です。生物学的な意味と同じようにも使えます し、自分のデータを他人のデータと分けて整理するというような目的で も使うことができます。

Organism メニューから任意の Organism を選択することで Organism が切り替わります。.

EDIT ORGANISM

Organism メニューから"Edit..."を選ぶと、新規 Organism を作成した り、既存の Organism を削除したり名称変更したりすることができま す。SOA(Subio Organism Archive)ファイルを生成したり、SOA ファ イルからデータを再構築したりするのもここです。SOA ファイルは、 データをバックアップしたり、他のコンピューターにデータを移したり、 共同研究者とデータを共有したりするのに便利です。

Genome

Genome とは、位置情報によって定義される genomic element のセッ トで、Genome View や Chromosome タブを描画するのに必要です。こ れらの機能を使わない場合は、この機能について無視しても結構です。 Genome メニューから既存の Genome をロードしたり、切り替えたり、 アンロードしたりすることができます。

新しい Genome は BED ファイルまたは GFF ファイルから作ることが できます。UCSC Table Browser からダウンロードした RefSeq Genes の BED ファイルから作るのが簡単でしょう。

<u>http://genome.ucsc.edu/cgi-bin/hgTables</u>

LEARN OPERATIONS

✓ <u>Creatig A Genome</u>

DATA MANAGER

PLATFORM LIST

このリストでは、一つの Platform のみ選択できます。Platform を選択 すると、その Platform に関連付けられた Sample と Series が下段のテ ーブルで表示されます。リストの右にある Column Headers セクショ

ンは、選択した Platform が持つアノテーションのタイトルを表示して います。

New Platform

新しい Platform を作成するには、Measurement ID アノテーションの テーブルをタブ区切りテキスト形式のファイルで準備します。市販のカ タログマイクロアレイを使っている場合は、メーカーのサイトからダウ ンロードできます。ヘッダー行は必須ですが、1行目でなくてもかまい ません。

<u>GEO</u>サイトからダウンロードした GSE の SOFT formatted family file からも Platform を作れます。

LEARN OPERATIONS

✓ <u>Creating A Platform</u>

Edit Platform 🧷

Platform List のツールバーから Edit Platform(▲)ボタンを押すと、ア ノテーションテーブルを更新したり編集したりできます。編集画面の右 端にある Select Columns 引出しタブ(緑色)をクリックして開くと、 アノテーションテーブルのタイトルが表示されます。チェックするとテ ーブルに表示されて、チェックを外すと非表示になります。引出しタブ の中でタイトルをダブルクリックすると、タイトルを変更でき、またド ラッグ&ドロップで順序を変更できます。

列に対する操作は、テーブル中で選択しているセルを含む列全体が対象 となります。同様に行に対する操作は、選択中のセルを含む行全体が対 象となります。

Split Column 🏂

編集画面のツールバーにある Split Columns (♥)ボタンは、その隣にあ るテキストボックスに入力した文字列のところで値を区切って、別々の 列に分割してくれます。区切り文字は一文字でなくてもかまいません。 たとえばアフィメトリクス社の提供するアノテーションテーブルの区 切り文字には "///"が使われており、そのような場合はスラッシュを三つ 入力して Split Columns ボタンを押してください。

Lоок Up 🔳

編集画面のツールバーにある Look Up (●)ボタンを押すと、外部のア ノテーションテーブル(タブ区切りテキスト形式のファイルまたは SOFT formatted family file)から情報を取り込むことができます。 Platform と外部テーブルが共通して持っている ID であれば、なんでも 照合してマッチした Measurement に対して情報を取り込みます。 Platform と外部テーブルが同じタイトルを持っている列については、 "replace entirely"(列全体を上書きする)または"fill blanks only"(空 白のセルのみ値を埋める)のいずれかを選択できます。上書きしたくな い場合は、外部テーブルのタイトルを変更してやりから直してください。

Icon	Tool tip	Shortcut
1	Undo	Ctr + z
^	Redo	Ctr + y
	Select All	Ctr + a
5	Сору	Ctr + c
se.	Cut	Ctr + x
5	Paste	Ctr + v
3	Clear	Delete
Ċ)	Repeat	Ctr + d
*	Split Column	-
	Look Up	-

ボタンとショートカット

2	Find	Ctr + f	
" "	Add Column	-	
ň	Delete Column	-	
*	Add row		
×	Delete row		
-	Move by Cell	Arrow keys	
-	Move and Select	Shift + Arrow keys	
-	Move by Block	Ctr + Arrow keys	

LEARN OPERATIONS

✓ Platform の遺伝子アノテーションを更新する

Select Symbol Column 🖉

Platform を作成したら、忘れずにこの設定を行ってください。ここで設 定した列の値が、Tree View や Datasheet タブで表示されますので、デ ータがぐっとわかりやすくなります。また、Advanced Plug-in に含ま れる"Find miRNA Targets"ツールは、ここで miRNA の名前と Gene Symbol の列が設定されていなければ正常に動作しません。

Merge Platform 📕

同じマイクロアレイの Paltform が複数できてしまったときなどに、こ れらを一つにまとめて整理することができます。特に、SSA ファイルを インポートすると、そのたびに新しい Platform が作られますので、い らなくなったデータを削除したり、Platform を統合して整理したりす る作業を必要に応じて行う必要があります。 PARENT となる Platform を一つ選んで、その Platform と統合する CHILD の Platform を複数選択して実行すると、すべての Sample と Series が Parent Platform に移行して Child Platform はなくなります。

統合する前に、Merged details をよく確認してください。Parent と Child が全く同じであれば、Parent に追加する列も行も0になります。

ー度統合してしまうと元には戻せませんので、この作業を行う前にデー タのバックアップを取ることをおすすめします。

Assign To Other Organism 👭

選択した Platform とそれに関連付けられている Sample と Series をま とめて別の Organism に移行することができます。

Import Samples 👒

Subio Platform は**タブ区切りテキスト形式**の数値テーブルであれば、 どんなデータでも取り込んで解析することができます。Affymetrix 社 の GeneChip、Agilent 社のマイクロアレイ、Illumina 社の BeadArray や次世代シーケンスから出てきた FPKM・BED・GFF ファイルなどで す。また、Affymetrix 社の BAR ファイル、次世代シーケンスから作ら れる BAM/SAM ファイルもインポートできます。GEO データベースの データを取り込みたい場合は、GSE レコードの SOFT formatted family file をインポートします。

通常は、特定の Platform に実験データを取り込んで Sample を作成し ますが、実験データを取り込むと同時に Platform を作成することも可 能です。 Platform List のツールバーにある、**Import Samples**(¹⁴⁴)ボタンを押す か、Sample List にファイルを**ドラッグ&ドロップ**して **Import Samples Wizard** を開始します。

IMPORTING RNA-SEQ OR CHIP-SEQ DATA

遺伝子レベルで数値化された FPKM のテーブルなどは、マイクロアレ イデータと同じように扱うことができます。ゲノムを一定の幅で区切っ たビンごとにカウントされた BED や GFF フォーマットのファイルも 同様です。しかし、マップされたタグの状態であれば、BAM または SAM フォーマットになっていれば Subio Platform に取り込むときにビンご とのカウントに変換することができます。ただし、この処理には非常に 大きなメモリーが必要になりますので、32GB または 64GB のメモリー を搭載したマシンをお使いください。それよりも小さなメモリー容量の マシンでしたら、テキストファイルにするところまでは Galaxy などの ウェブツールで処理することをおすすめします。そのようなデータ処理 を含めて弊社の解析サービスにご依頼いただくことも可能です。

MULTIPLE SAMPLES IN ONE FILE

複数の実験データが一つにまとまっているようなファイルをインポー トする場合は、**Multiple Samples in One File** オプションを使うと一つ 一つの Sample に分割してインポートすることができます。

LEARN OPERATIONS

✓ エクセルのデータを Subio Platform にインポートする

FORMATについて

実験データの生ファイルを見ると、たくさんの列があって複雑に見える かもしれませんが、Subio Platform が必要なのはこの中のたった二つ、 Measurement ID と Ch1 Raw Signal に相当する列だけです。2 色法の データならこれに加えて Ch2 Raw Signal の列を指定する必要がありま す。Ch1 Reserved Signal と Ch2 Reserved Signal は解析には必要ない ですが、データの質に関する情報を視覚化するために取り込みたい場合 に指定すればいいです。Format とは、実験データファイルのどこにタ イトル行があって、そしてどこに上記に相当する列があるかを指定する ものです。

フラグ列はここでは指定せず、Setup Series タブで指定します。これに より、いつでもフラグ列を変更できるようになっています。

Format の設定を保存しておけば、次に同じ形式のデータを取り込むと きに呼び出せるようになります。

1 色法 (SINGLE-CHANNEL) のデータフォーマット

Measurement ID と Ch1 Raw Signal に相当する列を指定してくださ い。Ch1 Raw Signal には通常シグナル強度に相当する値を持つ列を指 定します。ノーマライズ済みの対数比の値しかない場合は、これを指定 するのでも構いませんが、信頼性に関する情報は極端に少なくなります。

2 色法 (DUAL-CHANNEL) のデータフォーマット

Measurement ID と Ch1 Raw Signal に加えて、Ch2 Raw Signal の列 指定も必須となります。Ch1 Raw Signal には実験したほうのサンプル のシグナル強度を指定し、Ch2 Raw Signal にはコントロールサンプル のシグナル強度を指定します。Subio Platform では、必ず Ch1/Ch2 の 比が取られて Processed Signal になります。Dye-swap をした実験では green/red と red/green の二つのフォーマットを作って、二回に分けて

実験データをインポートして、最終的に test sample (ch1) / reference sample (Ch2)の比になるようにしてください。

NUMBER SEPARATORS

アメリカ形式とヨーロッパ形式の数値フォーマット(桁区切りと小数点) を自動的に認識するようになっていますが、もし間違っている場合は手 動で修正してください。

MANAGE FORMAT

保存済みの Format を見ることができます。

IMPORT OPTIONS

FILTERING BY SIGNAL

実験データをインポートするときに、データが大きすぎて解析不能にな るのを避けるため、閾値よりも大きい、小さい、あるいはその両方の値 だけを抽出してインポートすることができます。

FILTERING BY LOCATION

特定の領域、たとえばすべての遺伝子と重複する領域やある遺伝子群の 上流域のみなどに絞り込んで、その範囲の Measurement のみをインポ ートすることができます。

Merging/Counting

個別のMeasurementをインポートする代わりに、近傍のMeasurement をまとめたり (Intervals)、ゲノムを既定のサイズに分割したビンごと にまとめたりすることで、Measurementの数を減らすことも可能です。 38

このオプションは ChIP-seq や Methyl-Seq などのようにサンプル毎に 微妙に一が異なるデータをインポートする際にも有用です。

LEARN OPERATIONS

✓ <u>Import Experimental Data.</u>

AFTER YOU IMPORT SAMPLES

Import Sample Wizard を完了すると、新しい Sample が Sample List に現れます。すぐに Sample Information (シ)ボタンをクリックして、 これらの Sample が何かを記述しておくことをおすすめします。

Edit Sample Information 画面の操作は、Edit Platform **《**画面とほぼ 同じです。違いは Image 列と Files 列があって、そこにはドラッグ&ド ロップでファイルを入れることです。Image 列は 3MB までの画像ファ イル (BMP, JPG, GIF, PNG) のみ受け付けます。

GEO のサンプル情報は、関連する SOFT formatted family file をイン ポートすることで簡単に埋めることができます。Look Up ツールを使っ て情報をインポートしてください。

CREATE A SERIES

解析したい Sample にチェックを入れて、Sample List ツールバーにある Create Series ()ボタンを押します。

SETUP SERIES タブ

新しし Series ができると、**Analysis Browser** が開きデータが描画され ます。まず、Setup Series タブでこの Series の基本設定をします。

Edit Parameters 🟒

Parameter とは、sample ごとにその生物学的性質や実験条件などを記述したものです。Edit Parameters *L*ボタンを押して編集します。

既に Sample Information を登録している場合や、おなじ Sample を使った解析済みの Series がある場合などは、それらの値をインポートすることができます。

一つの列が一つの意味を持つように編集します。たとえば、"Sample Name"列に"WT_NoTreatment_Replicate1"という値が入っていたとします。そうすると、"Strain"列に"WT"、"Treatment"列に"No"、
"Replicate"列に"1"というように3つの列に分割して値を入れます。

それぞれの Parameter について、そのタイプを"Categorical"または "Numerical"から選択します。このタイプは、この後の Parameter によ る色分けやソートしたときの並び順に影響してきます。

テキストファイルや SOFT formatted family ファイルから情報をイン ポートする場合は、編集ツールの Look Up ツールを使ってください。

IMPORT SAMPLE INFORMATION

40

Sample information の列がリストアップされていますので、その中か ら取り込みたいタイトルを選んでインポートしてください。

IMPORT SERIES PARAMETERS

解析済みの Series が他にある場合は、その Series から Parameter を 持ってくることができます。インポートしたい Series を選択したら、 その Series に与えられている Parameter のリストが表れますので、そ のなかから取り込みたい Parameter を選んでインポートしてください。

IMPORT STATUS OF SELECTION

Scatter Plot (Samples) View や Tree View で Sample Group を選択し ている場合は、"selection"という列を追加して、そこに選択している Sample だけに"1"という値を与えます。この機能は、特に数百におよぶ Sample をそれ以外の Sample から区別したいときにとても有用です。

Edit Flag Columns

フラグ値はインテンシティに付加されていて、その測定値の信頼度につ いて情報を与えてくれます。この情報は、統計解析の前に行うクオリテ ィコントロールの段階で使えるのはもちろんですが、場合によっては解 析そのものにも使えます。

Subio Platform では、4 つまでフラグ列として選択できます。フラグ列 として選択したら、その値が "categorical"なのか"numerical"なのかを 選択してください。この設定はエラー! 参照元が見つかりません。の Filter ツールの挙動に影響します。

フラグカラムの選択状態は、Flag Set として保存することができます。 保存しておけば、次回以降は Flag Set を選択するだけで同じ状態を簡 単に再現できるようになります。

LEAN OPERATIONS

✓ <u>マイクロアレイのフラグ値とは?</u>

NORMALIZATION

Subio Platform でいうところの Normalization とは、Ch1 Raw Signal から Processed Signal を生成する一連の処理のことです。

保存されている Normalization をプルダウンメニューから呼び出すと、 自動的にブロックがセットされます。

PRESET SCENARIOS FOR 1-COLOR DATA SETS

Profiling*: コントロールサンプルがない実験に汎用的に使えます。

Compare to Control*: コントロールサンプルが一つあるいは複数あ る場合に汎用的に使えるセットです。最後のブロックのオプション設定 で、どのサンプルがコントロールなのかを指定する必要があります。

Import Log Ratio Data*: すでに対数比になっているデータをインポ ートした場合に使ってください。元は2色法のデータでも二つのチャン ネルの値の対数比になっているデータを取り込んだ場合も同様です。

PRESET SCENARIOS FOR 2-COLOR DATA SETS

Direct Comparison*:,2 色法の実験データで、二つのチャンネルで二 つの条件を直接比較している場合に使います。

Common Reference Design*: 共通のリファレンスサンプルを特定 のチャンネルで測定して、サンプル間の比較は間接的に行う実験デザイ ンの時に使ってください。

NORMALIZATION BOARD

Normalization を編集したい場合は、Normalization Board 上の Normalize Block を編集します。

42

Normalize block は Normalization で行われる処理の単体で、画面右の Normalize Blocks 引出しタブ(緑色)をクリックすると現れる倉庫か ら取り出したり(<)、仕舞ったり()することができます。

Normalization Board に並べられたブロックはドラッグ&ドロップで順 序を入れ替えることができます。それぞれのブロックの左側には黒い三 角形のアイコンがあります。これをクリックすることで、そのブロック のオプション設定パネルを開いたり閉じたりします。

Normalization Board 上でブロックを選択すると、そのブロックの結果 の値の分布が左のヒストグラムで描画されます。前のブロックの結果と そのブロックの結果を見比べることで、そのブロックの処理が具体的に 何をやってどのような効果を持つのかを直感的に理解できるようにな っているので、どなたでも適切な判断ができます。最後のブロックをク リックしたときに表示されているヒストグラムは、Processed Signal の 分布を表します。

編集し終えたら、最後に必ず"**Do Normalize**"ボタンを押してください。 これで Processed Signal が書き換わります。逆にいうと、このボタンを 押すまでは実際に影響を与えませんので、好きなだけ Normalization の 試行錯誤をすることができます。

Normalization Board の上のプルダウンで"**Save Normalization**"を選 択すると、そのブロックの設定に名前を付けて保存できます。こうして おくと、ワンクリックでその設定を呼び出せるようになって便利です。

NORMALIZE BLOCKS

PREPROCESS

LOW SIGNAL CUTOFF

Ch1 Raw Signal (と Ch2 Raw Signal) のうち、低すぎる値を除去するか、別の値に置き換えるかすることができます。

無料のオンラインサポート&デモンストレーションを申し込む

.....

LOG TRANSFORMATION

比とは、増加側(1から∞)と減少側(1から0)を対等に扱うことがで きません。しかし対数に変換していれば、対数比にしたときに2倍と 1/2倍が0から等距離に来ますのでこの問題を回避できます。

.....

IMPORT LOG RATIO DATA

もし対数を Ch1 Raw Signal としてインポートした場合は、その値が対 数であることをこのブロックを使って Subio Platform に教えてあげる 必要があります。

TRANSFORM SIGNALS

このブロックは、データに四則演算による変更を加えます。すべての遺 伝子を対象にすることもできますし、Measurement List を指定するこ とで特定の Measurement に対してだけ処理をすることも可能です。

FILL MISSING VALUES

このブロックは、値が何もないところを指定した値で埋めます。このブ ロックは、RNA-Seq のデータを解析するときに特に重要です。RNA-Seq のデータは多くの欠損値を含むことが多いからです。

.....

STANDARDIZATION

GLOBAL NORMALIZATION

Global normalization は単純ですが、広く使われている強力な手法です。 このブロックは、中央値や平均値、あるいは任意のパーセンタイルの値 が揃うようにヒストグラムを左右に動かして調整する効果を持ちます。 これによってサンプル間の統計誤差を最小化して生物学的な変化を浮 かび上がらせることができます。

場合によっては、コントロール遺伝子として抽出した Measurement 群 だけを使って同様の処理をしたいことがあります。そのような時は、 44

Use Control Genes ボタンを押して、コントロール遺伝子だけを含む Measurement List をセットしてください。

QUANTILE NORMALIZATION

各サンプルにおいて値が小さいほうから並べられます。すべてのサンプ ルから一番小さい値を取り出して、その平均値に置き換えます。次に2 番目に小さい値を取りだして、というのを繰り返します。そうすること で、すべてのサンプルの分布が完全に一致することになります。この手 法は前出の Global Normalization よりもはるかに強力にデータの外見 を一様にすることができます。しかしこのような矯正が、生物学的に意 味があるかどうかは熟慮の必要があります。たとえば、異なる種類の細 胞、あるいは異なる発生段階の細胞から得られたデータを比較する場合 などには、このような矯正はかえって問題となるでしょう。

MAKING RATIO

CENTERING

特にコントロールサンプルがない実験の場合、このブロックを使って全体の平均値や中央値に対する比にします。たとえコントロールサンプルがなくても比にするのは、とりあえず発現強度を無視して増減という変化にだけ注目することで、複雑さの次元を一つ下げることができるからです。

このブロックは1色法のデータだけでなく、共通リファレンスを使った 2色法の実験デザインの場合も使えます。

RATIO TO CONTROL SAMPLES

このブロックはコントロールと指定したサンプルに対する比に変換します。

このブロックは1色法のデータだけでなく、共通リファレンスを使った 2色法の実験デザインの場合も使えます。

45

RATIO (CH1/CH2)

2 色法のデータを Ch1 Raw Signal / Ch2 Raw Signal に変換します。

.....

LOWESS (CH1/CH2)

LOWESS による正規化を施した Ch1 Raw Signal / Ch2 Raw Signals の比に変換します。

LEARN OPERATIONS

✓ <u>ノーマライズを編集する</u>

LEARN "WHAT ARE PROCESSED SIGNALS? WHY DO YOU TURN SIGNALS INTO LOG RATIOS?"

✓ Processed Signals とは? なぜ対数比に変換するの?

Setup DataSet タブ

Parameter を設定したら、Setup DataSets タブで共通の Parameter 値 をもつ Sample を Sample Group にまとめて、さらにその Sample Group の並び順を整理して DataSet を定義します。

LEARN OPERATIONS

✓ <u>Series と DataSet のセットアップ</u>

SAMPLE INFO タブ

IMAGES & PARAMETERS (INFORMATION OF EACH SAMPLE)

Edit Sample Information 画面でそれぞれの Sample に付加された **Image**や**File**が、Parameter とともに表示されます。これらのイメー

46

ジやファイルはダブルクリックすると適切なアプリケーションで開く ことができます。

NOTES & FILES (INFORMATION OF THE CURRENT SERIES)

Notes 欄にはユーザーが自由にメモを残せます。ULR が書いてあれば 自動的にウェブリンクになります。Data Manager の Series List のツ ールバーでキーワード検索する対象がこの Notes ですので、できるだけ 情報を入れておいたほうが後で Series を探すのが簡単になります。

Files 欄にはドラッグ&ドロップで任意のファイルを添付することがで きます。ダブルクリックすると自動的に適切なアプリケーションで開き ます。添付したファイルを除く場合は、ドラッグアウトしてください。

Notes & Files セクションの左上にある黒い三角マークは、このセクションを開いたり閉じたりするのに使います。

THE PLUG-IN LICENSE

Plug-in は Subio Platform に有償で解析機能を追加するものです。 Serial Number を行使すると、行使したコンピューターでプラグインを 規定の期間使用できるようになります。複数のコンピューターでプラグ インを使いたい場合は、それぞれのコンピューターごとにライセンスが 必要です。

SERIAL NUMBER AND ACTIVATION

Plug-in Serial Number を購入すると、通常は email で届きます。Plugins メニューから"**Plug-in Manger**"を選択して、"Basic Plug-in"または "Advanced Plug-in"タブを開いてください。Serial Number をコピー& ペーストしたら Activate ボタンを押します。アクティベーションに成 功したら Subio Platform を起動すると Analysis Browser の右端にあ る Plug-ins 引出しタブ (ピンク色)を開くとプラグインのツールボタ ンが現れます。

もしそのコンピューターがインターネット経由でライセンスサーバー にアクセスできなかったら、エラーログを出力するダイアログ(下図) が表示されます。OK ボタンをおしてログファイルを保存し、これをメ ールで <u>support@subio.jp</u> 宛にお送りください。ライセンスファイルが 返送されて来たら、それをインストールディレクトリの中にある activationkeys フォルダーの中に保存してから、Subio Platform を起動 してください。



Windows 10 では Mac アドレスをランダム化することができるように なっていますが、プラグインをアクティベートする時、および使用する ときはこの機能を使っているとプラグインを利用できない可能性が高 いです。プラグインをご使用の方は、「ランダムなハードウェアアドレ ス」というオプションにチェックしないでください。

✓ <u>プラグインを使う場合、Windows10 の「ランダムなハードウェア</u>
 アドレス」を有効にしないでください。

BASIC PLUG-IN TOOLS

FILTER

Filter ツールは Basic Plug-in の中でももっとも頻繁に使うツールで、 主に実験データの統計解析の前に行うクオリティコントロールで使い ます。

クオリティコントロールでは、主に次のような遺伝子を除きます。

インテンシティが常に低すぎる遺伝子群

ユニバーサルレファレンスの RNA サンプルでもない限り、全ての遺 伝子が発現していることはありえません。しかし、マイクロアレイに は網羅的にプローブが設計されているので、発現していない遺伝子に も何らかの測定値が付きます。それから、測定装置の感度限界よりも 低い発現量の遺伝子は正確な値を得ることはできません。

実際の測定値は、インテンシティが低い領域において繰り返し実験の 測定値のばらつきが大きくなることが知られています。繰り返し実験 の測定値が再現性を持つ領域を「シグナル領域」、測定値が再現しな

49

い領域を「ノイズ領域」とすると、ノイズ領域には上記二つの要素が 混在していると考えられます。

大事なのはシグナル領域とノイズ領域の境界を見極めることです。論 文や教科書に書いてあるのを真似して閾値をそのまま使う人がいま すが、これは絶対にしてはいけません。シグナル領域とノイズ領域の 境界はデータごとに異なるからです。フラグ値を使ってフィルタリン グするのも、測定値の信頼度はインテンシティに依存しているので、 やっているのは同じことです。

フィルターによって、常にシグナル領域にある遺伝子を抽出すると、 残った Measurement の値はどれも信頼性が高いということになりま す。しかし、例えば Normal と Diseased という二つの状態を比べて いるときなど、Normal で発現していなかったのに Diseased では発 現している遺伝子や、その逆の動きをしている遺伝子は、常にシグナ ル領域にある遺伝子群の中には入っていません。そこで、フィルター をかける時は、常にノイズ領域にある遺伝子を除くということを考え て行ってください。

発現変動がみられない遺伝子群

網羅的解析の大前提は、「ほとんどの遺伝子は変動しておらず、ごく 一部の遺伝子だけが変動している」ということです。しかし、変動し ていない遺伝子でも繰り返し実験でまったく同じ値を得ることはあ り得ません。そこで、今度はまったく同じ値を全く変動していない遺 伝子がどれくらいばらついて観測されるかという境界を見極める必 要があります。変動していない遺伝子においても観測されうるばらつ きの中に常に納まっていたら除くということを考えて行ってくださ い。

変動しない遺伝子がノイズとなって最も悪影響を受けるのがクラス タリングです。クラスタリングの前にこのようなフィルタリングを行 っておくことがとても重要です。

50

Filter ツールは汎用的に使えるよう設計されていますので、上記以外の コンセプトでフィルタリングを行うことも、もちろん可能です。

フィルタリングによって、大事な遺伝子を解析の前に外してしまうこと をおそれる方がいますが、その必要はありません。しっかりとクオリテ ィコントロールをして、信頼性の高い統計解析を行った後、その結果と して得られた特定の発現パターンを示す遺伝子群をテンプレートにし て、次に説明される Find Similar Pattern ツールを使って、その発現パ ターンに似ている遺伝子をすべての Measurement の中から拾いなおす ことが可能だからです。

Filter ツールで表示されるヒストグラムを、上記の境界を見極めて閾値 を決めるのにとても役に立つのですが、非常に大きなデータを扱ってい るときは処理が非常に重くなります。そのような場合は、ヒストグラム の描画を省くことができます。

LEARN OPERATIONS

- ✓ <u>フィルターの使い方</u>
- ✓ <u>マイクロアレイのフラグ値とは?</u>

FIND SIMILAR PATTERN

選択中の Measurement 群の平均パターンに似ている、あるいは逆向き の発現パターンを持つ Measurement を抽出します。相関係数が1に近 いほど「似ている」ことを表し、-1に近いほど「逆向き」であること を表します。

LEARN OPERATIONS

✓ 発現パターンの似た遺伝子の抽出

COMPARE TO CONTROL

Compare to Control ツールは、Processed Signal が 0 から離れている measurement を抽出します。これは、Normalization で Ratio to Control Samples ブロックを適用していて、Processed Signal が Control Sample に対する対数比になっていることを前提として使うものです。直接比較 している 2 色法のデータの解析にも適しています。このツールを使う重 要な目的がもう一つあり、それが paired T-test の適用です。それ以外 の比較であれば、後述の Compare 2 Groups ツールを使うことを検討し てください。

セットした Sample Group が複数の Sample から成る場合は、One Sample T-test による P 値の計算が可能なので、volcano plot が描画で きます。もし一つの Sample から成る場合は、Fold 値を表すヒストグラ ムのみ描画可能です。

LEARN OPERATIONS

- ✓ <u>T 検定により発現差のある遺伝子を抽出する</u>
- ✓ Paired T-test で解析するにはどうすればいいですか?

COMPARE 2 GROUPS

Compare 2 Groups ツールは、2 群間で発現差のある遺伝子を抽出します。

セットした Sample Group の両方が複数の Sample から成る場合は、**T**-**Test** (Welch's T-test) または **T-test (equal variance)** (Student's T-test) 、**Mann-Whitney U-test** のいずれかによる P 値の計算が可能な ので、Volcano Plot を描画できます。**P-values** は **BH FDR** (Benjamini-Hochberg) に補正することもできます。

もし両方またはいずれかの Sample Group が一つの Sample から成る 場合は、Fold 値を表すヒストグラムのみ描画可能です。

- ✓ 網羅的遺伝子発現データ解析の理論と実践
- ✓ <u>T検定がなぜ理論的には使えないのか、そして、それでもなぜ実用</u>
 には使えるのか

LEARN OPERATIONS

✓ <u>T 検定により発現差のある遺伝子を抽出する</u>

COMPARE ONE TO ALL

Compare One to All ツールは、セットした Sample Group と、セット した DataSet に含まれるすべての Sample Group との2群間比較を一 括して行います。もしフラグカラムが設定してあれば、解析を行う Measurement をフラグ値で指定して実行することができます。

結果として得られる、Filter Passing Genes と Differentially Expressed Genes (DEGs)の Measurement List を個別に保存することもできます し、DEGs を出現頻度によりまとめた Measurement List を保存するこ ともできます。

後述の Compare Multiple Groups は、擬陽性を最小限に抑えることを 目的として使えますが、こちらのツールや、次の Compare All Pairs は 偽陰性を最小限に抑えるためのツールとして有用です。

LEARN OPERATIONS

✓ 発現差のある遺伝子を一括して抽出する

COMPARE ALL PAIRS

Compare All Pairs ツールは、セットした DataSet に含まれる SampleGroup から生成可能なすべてのペアについて、2群間比較を一 括して行います。

Sample Goup の数が 30 以下の場合は、すべての組み合わせの DEGs のリストを保存することができます。30 以上の場合は、出現頻度によりまとめたリストのみ保存可能です。

結果は3つのタブで表示されます。"Table"タブは、Filter Passing Genes または Differentially Expressed Genes (DEGs)の数を単純なテ ーブルで表示します。"Chart of DEGs"タブは、DEGsの数を2次元の 図で表示します。セルの色が赤いほど、DEGsの数が多いことを表しま す。"Chart of Correlation"タブは、Filter Passing Genes の ProcessedSignalの相関係数を二次元の図で表示します。セルの色が青 いほど相関の高いペアで、赤いほど逆相関になっているペアであること を表します。

LEARN OPERATIONS

✓ 発現差のある遺伝子を一括して抽出する

COMPARE MULTIPLE GROUPS

Compare Multiple Groups は、比較したいグループが3つ以上あると きに使います。統計学の教科書では、3 群以上の比較を行う場合は T-Test ではなく ANOVA を使うことが推奨されます。その理由は、たと えば5つのグループを比較する場合、2 群間の比較を行う T-test を有 意水準5%で繰り返すならば 10 回の検定が必要でそのたびに偽陽性が 5%の確率で入ってくるので、全体として偽陽性が膨れ上がることが考 えられるためです。

しかし、実践的なオミクスデータ解析としては T-test を繰り返すほう がいいということも考えられるのです。その理由は、例えば上記の例で 1つのグループがコントロールに相当する場合、総当たりの10回では なく、コントロールとの比較の4回の検定で済むことが考えられます。 もう一つの理由は、偽陽性を低く抑えるということは、偽陰性が増える

54

ことを意味します。言い換えると、検出力が弱くなるということです。 オミクスの使用目的が「間違いのない結果を出す」ことではなく、「候 補をあげる」ことであれば偽陽性よりも偽陰性が増えることを警戒する 必要があるからです。あがってきた候補遺伝子から、細胞内で起こって いることを推測して、仮説を立てて、その仮説を実験で検証するという 全体の流れのなかでオミクスを利用するのであれば、候補をあげる段階 で偽陽性を厳しく排除することはむしろ弊害となります。

✓ 網羅的遺伝子発現データ解析の理論と実践

LEARN OPERATIONS

✓ ANOVAにより発現差のある遺伝子を抽出する

TREE CLUSTERING

階層型クラスタリングは、発現パターンの似た Measurement が集めて 分類したり、発現プロファイルの似た Sample Group を集めて分類した りするのに便利です。特に Sample Group が4つ以上の時に有効で、3 つ以下の場合は前述の発現差解析で用は足ります。

クラスタリングを実行する前に、ノイズとなる Measurement を、Filter ツールを使って除去しておくことが重要です。フィルタリングを行うも う一つの効用は、Measurement を減らすことで計算時間と消費メモリ ーを大幅に削減できることです。メモリー不足のエラーが出てクラスタ リングが実行できないときは、Measurement の数を減らしてみてくだ さい。事前にフィルタリングをしたくないという人がいますが、 Measurement を減らしてクラスタリングを実行した後、得られたクラ スターの Measurement 群をテンプレートにして Find Similar Pattern ツールを使えば、すべての Measurement の中から拾いなおすことがで きるので、心配いりません。 Similarity measure オプションでは、近似度の計算方法を選択すること ができますが、なにを選択すればいいか分からない場合は以下を参考に してください。

Pearson Correlation: いわゆる相関係数と呼ばれるものです。 Normalization で Centering ブロックを適用している場合に適してい ます。

Uncentered Correlation: Normalization で Centering を適用してい る場合でも Compare to Control を適用している場合でもどちらにも適 していますので、Pearson Correlation よりも幅広いケースで最適です。

Spearman Correlation: 外れ値の影響を受けにくいです。Sample Group が 10 以上ある場合は、こちらを第一の選択肢にしてください。

Euclidian (distance): 上記3つのいずれかで計算したクラスタリングの結果から、特定のクラスターを取り出して、これをさらに分類する時に適しています。

クラスタリングの計算が終わると、Tree オブジェクト(③)が DataSet に関連付けられて Series Panel に保存されます。このデータは Tree View (44)で閲覧します。

LEARN OPERATIONS

- ✓ <u>階層型クラスタリング</u>
- ✓ <u>Tree View</u> の使い方
- ✓ <u>T検定がなぜ理論的には使えないのか、そして、それでもなぜ実用</u>
 <u>には使えるのか</u>

主成分分析 (PCA)

多変量データはとても複雑なので、次元を落として単純化してみること で理解しやすくなります。Sample Group がたくさんある DataSet を 俯瞰して大雑把に把握するのに、主成分分析は威力を発揮します。 Scatter Plot (Samples) 上で近くにあれば発現プロファイルが近いこ とを表し、遠ければ発現プロファイルが似ていないことを表します。さ らに、0の線を越えて反対側にあれば、逆向きの発現プロファイルであ ることを表します。時系列データでは、時間経過とともに発現プロファ イルがどのように変化していくのかを理解する助けとなります。

PCA を実行すると Profile オブジェクト()が Series Panel の Profiles フォルダーの中に保存されます。Scatter Plot (Samples) View ()の 縦軸あるいは横軸に Profile をドラッグ&ドロップでセットすると、 Sample Group がその Profile により計算されるスコアをもとに配置さ れます。

Details>ボタンをクリックすると、各主成分の係数を見ることができま す。さらに、係数が大きいものと小さいものを、その主成分に貢献度の 高い Measurement 群としてプレビューしたり保存したりできます。

LEARN OPERATIONS

- ✓ <u>主成分分析 (PCA)</u>
- ✓ <u>Scatter Plot (Samples) View</u>の使い方

ADVANCED PLUG-IN TOOLS

Advanced Plug-in は、Basic Plug-in で統計解析を行った後、生物学的 解析をするときに有用です。統計解析と、その結果の生物学的解釈は別 の人が担当することも多いと思います。Basic と Advanced プラグイン は別々のコンピューターで使えます。プラグインの機能がわかれている ことで必要な人に必要な機能を割り当てることができるので、チーム全 体としてコストを下げられます。

また、Advanced Plug-in は、ゲノム上の位置と関連する解析、たとえ ば ChIP-chip、CGH アレイ、プロモーターアレイ、メチレーションア レイ、タイリングアレイ、miRNA マイクロアレイ、RNA-Seq、ChIP-Seq、Methyl-Seq などには必須の解析ツール群です。

ENRICHMENT ANALYSIS

Platform のアノテーションテーブルの列を一つ選んで、選択した Measurement List がどのようなキーワードをエンリッチしているかを 調べます。Gene Ontology (GO)の列に使えば、GO エンリッチメント解 析になるのはもちろんですが、サイトバンドやプロテインドメイン、pfam など他の情報に対しても汎用的に使えます。

LEARN OPERATIONS

✓ Enrichment Analysis Tool の使い方

PATHWAY EDIT TOOL

Pathway は背景のイメージと、そのイメージの上に配置される Measurement のセットです。Pathway Edit Tool は新しい Pathway オ ブジェクトを作ったり、既存の Pathway オブジェクトを編集したりす

るツールですが、その基本はイメージ上の任意の位置に Measurement を配置することです。

KEGG pathway converter を使うと、任意の KEGG パスウェイの URL を指定するだけで、Pathway オブジェクトを作るのに必要なイメージ ファイルと、そのイメージ上の遺伝子の位置を示したテキストファイル を生成してくれるので、Pathway Edit Tool で Pathway オブジェクト に変換する作業を半自動化してくれます。

他のパスウェイデータベースについても同様の補助ツールがほしい場 合は<u>ご連絡ください</u>。

LEARN OPERATIONS

✓ Pathway Edit Tool の使い方

GENES TIED IN PARAMETER

Genes Tied in Parameter ツールは、たとえば年齢、経過時間、濃度、 ステージなどの数値データと相関(あるいは逆相関)関係にある遺伝子 を抽出します。

LEARN OPERATIONS

✓ <u>Genes Tied in Parameter の使い方</u>

GENOMIC LOCATION FILTER

Genomic Location Filter ツールは、Measurement や Region を染色体 上の相対的な位置関係によって抽出するツールです。

Reference Track は検索の開始位置を指定するものです。一方 **Query Track** は、この中から特定の領域にあるものを抽出します。 Query Track に Measurement List をセットしていれば、結果として Measurement

List が出力されます。Region List または現在ロードされている Genome が Query Track にセットされている場合は、Region List が出力されます。

位置検索モードは、次の中から選びます。

Upstream/Downstream

Query Track の中から、Reference Track のエレメントの上流(または 下流)域にあるものを抽出します。Reference Track にセットするもの は strand の情報が必要です。

Distance に検索領域の長さを入れます。Search mode が "Whole"の場 合は、Reference Track のエレメント全体を検索の基準とします が、"Blocks"を選択した場合はブロック(通常はエクソンに相当)が検 索の基準となります。Fringe は、指定した長さを検索領域の両側に追加 します。

Coverage が"completely"の場合は、Query Track のエレメントが

検索領域に完全に入っている場合だけ抽出し、"partially"の場合は一部 でも重複していれば抽出されます。

UPSTREAM/DOWNSTREAM OF QUERY TRACK

Query Track のエレメントのうち、Reference Track のエレメントをその上流(または下流)域に持つものを抽出します。

たとえば、Reference Track に ChIP-Seq から得られた転写因子結合領 域 (strand 情報なし)をセットして、Query Track に Genome (strand 情報をもつ遺伝子群)をセットすると、上流に転写因子が結合している 遺伝子が抽出できます。

OVERLAPPING

Query Track のエレメントのうち、Reference Track のエレメントと重 複しているものを抽出します。

このモードを使うと二つの Region List の積集合または差集合に相当する List が作れます。

Merge

二つの Region List、または Region List と Measurement List の和集 合を作り、Region List として保存できます。

LEARN OPERATIONS

- ✓ Genomic Location Filter の使い方
- ✓ Genomic Location Filter を使ったケーススタディ

MEASUREMENT TO REGION

Advanced Plug-in のプラグインにはインプットとして Region List し か受け付けないものがあります。そこでこのユーティリティツールを使 って、Measurement List と Sample Group を Region List に変換しま す。DataSet をインプットにセットすると、その配下の Sample Group をまとめてセットするという意味になり、結果として Region List 群が フォルダーの中にできます。

LEARN OPERATIONS

- ✓ Measurement to Region および Region Score Filter の使い方
- ✓ <u>Measurement to Region</u>を使ったケーススタディ

SUMMARIZE

✓ 個々のエレメントのスコアはふらついていても、近傍のものをまと
 めて俯瞰してみると、増えている(減っている)といった傾向が見

61

えてくることがあります。 Summarize ツールは、近傍のエレメン トをまとめるツールの一つで、似たような働きをするツールに後述 の RNA-Seq の実験データ を、FPKM でなく BAM で見る

Create Intervals ツールがあります。

オプション設定の詳細は Genomic Location Filter を参考にしてください。

Upstream/Downstream

Query Track のエレメントのうち、Reference Track のエレメントの上 流(あるいは下流)域にあるものをまとめます。

OVERLAPPING

Query Track のエレメントのうち、Reference Track のエレメントと重 複数ものをまとめます。

Βιν

このオプションを選択すると、Reference Track は必要ありません。 Query Track のエレメントを一定の長さで区切ったゲノムビンごとに まとめます。

LEARN OPERATIONS

- ✓ Summarize および Create Intervals の使い方
- ✓ <u>RNA-Seq</u>の実験データ を、FPKM でなく BAM で見る

CREATE INTERVALS

近傍のエレメントを連結することで、エレメントの数を減らすことがで きます。となりあった二つのエレメントが Maximum gap で指定した 値よりも近ければ結合します。そして、出力するリストから Minimum

62

Size で指定した値よりも小さければ除外することができます。Region List をインプットしている場合のみ、出力される Region List にスコア が付きます。このツールを使う前に **Measurement to Region** ツールを 使って Measurement List を Region List に変換してください。

LEARN OPERATIONS

- ✓ <u>Summarize および Create Intervals</u>の使い方
- ✓ <u>Create Intervals を使ったケーススタディ</u>

REGION SCORE FILTER

Region List () の中には、"R"マークのついたもの ()があり、これ はスコアを持っていることを表しています。Region Score Filter はこの ような Region List をインプットとしてスコアでフィルタリングしてサ ブセットを抽出します。

Region Score Filter 画面に描画されるヒストグラムは閾値を決めるの に大変役立つものですが、巨大な Series を解析しているときは描画に 大変時間がかかります。そのような時はヒストグラムを非表示できます。

LEARN OPERATIONS

- ✓ <u>Measurement to Region および Region Score Filter の使い方</u>
- ✓ <u>Genome View</u>の使い方

GET SEQUENCE

Reference Track に含まれるエレメントから相対位置で指定された領域の塩基配列を取得します。

このツールを使うためには、ゲノムのシーケンスデータをあらかじめダウンロードしておく必要があります。たとえば、下記のリンクからヒト

63

ゲノムのシーケンスデータファイルをダウンロードできます。染色体ご とに分けられた.fa.gz ファイルをすべてダウンロードして、一つのフォ ルダーのなかに保存してください。Gz 圧縮は解凍して保存してもいい ですし、回答せずに保存してもいいです。

✓ <u>ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/Assembled_chromoso</u> mes/seq/

このツールを実行すると、下段に結果の塩基配列が表示されま す。"Forward"を"Reverse"に切り替えると、塩基配列が逆向きになりま す。ピンク色のついている領域は、Reference Trackのエレメントと重 複した領域であることを表しています。塩基配列はドラッグしてコピー できます。Save FASTA ボタンで FASTA ファイルを出力できます。 1,500bp よりも長い場合は画面には表示されていませんが、出力した FASTA ファイルにはちゃんと入っています。出力した FASTA ファイ ルは、ホモロジー検索や転写因子結合サイトのモチーフ検索などを行う ためのさまざまな外部ツールで利用できます。

LEARN OPERATIONS

- ✓ <u>Get Sequence</u>の使い方
- ✓ <u>Get Sequence を使ったケーススタディ</u>

FIND REGIONS FROM SEQ

Find Regions from Seq は、特定の配列を持つ領域を検索して Region List として保存します。塩基配列は ATCG の4文字のほかにも、IUPAC で定義されている W (A または T) などのコードが使えます。

✓ <u>http://en.wikipedia.org/wiki/Nucleic_acid_notation</u>

このツールを使うためには、ゲノムのシーケンスデータをあらかじめダウンロードしておく必要があります。ゲノム全体を対象に検索すること

もできますが、Reference Track を使って検索領域を狭めることで、計 算時間が短くなり、結果の解釈も簡単になるのでおすすめです。

LEARN OPERATIONS

- ✓ Find Regions from Seq の使い方
- ✓ <u>Find Regions from Seq を使ったケーススタディ</u>

FIND MIRNA TARGETS

Find miRNA Targets ツールは、miRNA とそのターゲットと考えられ る候補遺伝子のペアで、発現パターンが逆相関を示すものを抽出します。

このツールを使うには、まず miRNA-Target ペアの定義が必要です。 下の miRNA-ターゲット遺伝子データベースのリンクから、ペアの定義 ファイルをダウンロードしてください。必ずすべてのデータベースから ダウンロードしなければいけないわけではありません。Excel ファイル はタブ繰リテキスト形式のファイルに変換してください。

Database	URL	Download File		
Validated mil	Validated miRNA Target Database:			
miRecords	http://mirecords.biolead.org/	miRecords_versi		
		on3.xls		
mirTarbase	http://mirtarbase.mbc.nctu.edu.tw/	MTI.xls		
Predicted miRNA Target Database:				
TargetScan	http://www.targetscan.org/	Conserved site		
		context+ scores		
PITA	http://genie.weizmann.ac.il/	PITA Targets		
		catalog		
miRanda	http://www.microrna.org/	Good mirSVR		
		score,		
		Conserved		
		miRNA		

Settings...ボタンを押すと、miRNA-Target Database Settings 画面が 開きます。それぞれのデータベースからダウンロードしたファイルをセ ットしてください。Predicted databases の定義は、score の閾値を設定 することにより、信頼性の高いペアだけを使うこともできます。スコア が小さいほど信頼性は高いと考えられ、デフォルトでは半分のペアが使 われるるよう中央値がセットされています。最後に、これらのデータベ ースでの定義の"intersection"を使うか、"union"を使うかを選択してく ださい。

miRNA マイクロアレイの Platform と遺伝子発現マイクロアレイの Platform、そしてそれぞれのプラットフォームに対となる実験データセ ットが必要です。たとえば、まず miRNA アレイのデータを解析する Series をロードします。特定の発現パターンを示す miRNA 群を抽出し たら、**Measurement to Region** ツールを使って Region List に変換しま す。このとき、Platform の chrom, chromStart, chromEnd の列に値が なければ Region へ変換できません。ダミーでも構わないので必ず値を 埋めておいてください。また、Select Symbol Column で指定した列に は miRNA の名前が必要です。

次に、対となる遺伝子発現データの Series をロードしたら、**Find** miRNA Targets ツールを開いてください。

Associate with Region List ボタンを押して Sample Group – Region List Pair 画面を開きます。ドラッグ&ドロップで一つ一つの Sample Group と Region List を対応付けることもできますが、Set Region Lists by A Folder ボタンを使ってまとめてセットすることもできます。現在 は遺伝子発現データを開いているので、Platform Represents オプショ ンでは"gene"モードを選択して実行します。

各 miRNA とターゲット遺伝子の発現パターンの相関係数がヒストグ ラムで表示されます。ここで、相関係数が-1 付近の Measurement List

66

を保存すると、miRNAの制御を受けている可能性の高い遺伝子群となります。



Details...ボタンを押すと詳細画面に切り替わり、各ペアにおける miRNA と遺伝子発現の関係が散布図で表されます。

散布図上でドラッグして点を選択すると、その点が表している Sample が Sample Info タブで強調表示されます。下のテーブルでは miRNA-タ ーゲット遺伝子ペアを表していて、行をクリックするたびに散布図が描 きかわります。



LEARN OPERATIONS

- ✓ <u>Find miRNA Targets</u>の使い方
- ✓ <u>Find miRNA Targets を使ったケーススタディ</u>

FIND CORRELATED REGIONS

異なる Platform に属するデータを直接比較することはできません。そ こで比較したいデータはそれぞれの Series から Measurement to Region ツールを使って Region List に変換してください。

たとえば、ChIP-Seq のデータと遺伝子発現マイクロアレイのデータを 統合解析したい場合は、まず、ChIP-Seq のデータと遺伝子発現データ をそれぞれ別の Series を作って解析してください。

ChIP-Seq のデータ解析では、ノイズを除去して意味のありそうな結合 部位を抽出するでしょう。そうしたら Measurement to Region ツール を使って、抽出した領域(Measurement List)に Processed Signal を っけて Region List に変換します。一つの Region List が一つの Sample Group に対応しています。次に遺伝子発現データの Series を開きます。 この Series は先ほどの ChIP-Seq と対になっていて、同じ Sample Group を持っています。

それでは、転写因子の結合状態に相関して(あるいは逆相関して)発現 レベルが変動している遺伝子を抽出しましょう。

Find Correlated Regions ツールを開きます。フィルターをかけてノイ ズを除去した Measurement List をセットします。Associate with Region Lists ボタンを押して、Region List と Sample Group を対応付 けます。遺伝子の上流域の結合状態と遺伝子の発現レベルの相関を計算 したい場合は、Position of Regions で"Upstream"を選択します。 Distance で上流域の長さを決定して実行します。

結合サイトと遺伝子のペアでそれぞれ、結合パターンと発現パターンの 相関係数を計算して、ヒストグラムで表示されます。1に近い(相関) ところや-1に近い(逆相関)とことを抽出することができます。 Details ボタンを押すと詳細画面に切り替わります

LEARN OPERATIONS

- ✓ Find Correlated Regions の使い方
- ✓ <u>Find Correlated Regions</u>を使ったケーススタディ

SCATTER PLOT OF REGIONS

Scatter Plot Regions ツールでは、二つの Region の関係を散布図で表示します。二つの Region Lists を **Region List 1** と **Region List 2** にセットしたら、二つの Region List のエレメントどうしの相対的な位置関係を **Relative Positions** で選択します。

散布図上をドラッグすると Region が選択状態になり、Regions タブや Chromosome タブでも自動的に強調表示されます。

LEARN OPERATIONS

✓ <u>Scatter Plot of Regions</u>の使い方

ANNOTATE MEASUREMENTS

BED/GFF/BAM/SAM ファイルをインポートして Sample を作った場 合、Platform はゲノム上の位置を ID にして自動的に作られますが、遺 伝子アノテーション等を一切持っていない Platform ができます。 Annotate Measurement ツールは、現在ロードしている Genome のア ノテーションテーブルから、その重複領域、上流域、下流域などにある Measurement に情報をコピーします。

LEARN OPERATIONS

✓ Annotate Measurements の使い方

NEED A HELP?

Subio Platform およびプラグインをご使用中にわからないことや困っ たことがありましたら、**Help** メニューで情報を探すか、<u>Contact Us</u>よ りご連絡ください。

場合によっては、システムの状況を調べるために、**Tech Report** を送っ ていただくようお願いすることがあります。このようなときは、Help メ ニューから **Export Tech Report** を選択して、生成される ZIP ファイル をメールでお送りください。また、別の場合では SSA ファイルを送っ ていただくようお願いすることがあります。このようなときは、Data Manager の下段の Series タブで、該当する Series を選んで **Export Series** ボタンを押してください。SSA ファイルはメールで送るには大 きすぎるので、下記のサイトから Hightail 経由でお送りください。

✓ Subio に安全にデータを送るにはどうすればいいですか?

皆様のご協力に感謝します。

FREE ONLINE TECHNICAL SUPPORT & TRAINING

プラグイン購入の有無にかかわらず、どなたでもオンラインサポートお よびオンライントレーニングを無料で受けられます。ウェブ会議システ ムを使ってお客様の画面を見せていただきながら、操作を指示させてい ただきます。異なる場所から同時に複数の方が参加することも可能です。

We Hope You Enjoy Exploring The World of Omics Data!