

SiGN-BN 実習テキスト

土井 淳

atsushi_doi@cell-innovator.com

株式会社セルイノベーター

研究開発部

福岡市東区箱崎6-10-1

九州大学 産学連携棟I アントレプレナーシップ・センター 2階

<http://www.cell-innovator.com>

cell innovator

使用するプログラムの紹介 (1)

• SiGN-BN:

- ベイジアンネットワークを用いて、**遺伝子ネットワーク**を推定するプログラム。(v1.4.0)
- <http://sign.hgc.jp/signbn/index.html>

*FX10コンパイラのバグのため、v1.3.0をお使いのかたは、v1.4.0に更新を。

• CSML (Cell Systems Markup Language):

- **遺伝子ネットワーク**やパスウェイモデルを記述する言語。XML形式。
- <http://www.csml.org>



使用するプログラムの紹介 (2)

- **CIO (Cell Illustrator Online):**

- 遺伝子ネットワークやパスウェイモデルを表示、編集、解析するためのソフトウェア。
- <https://cionline.hgc.jp/cionlineserver/apps/usersman/main>



- **XiP (eXtensible Integrative Pipeline):**

- バイオインフォマティクスの処理を自動化するプラットフォーム。GUIを用いて、遺伝子ネットワークやNGSの解析を行える。
- <http://nagasakiilab.csml.org/ja/xip>



遺伝子ネットワークの3つの入手方法

A. すでに推定済みの遺伝子ネットワークを閲覧する。

- Gene Network Database (GNDB); <http://gndb.cell-innovator.com> より、CSMLファイルを取得。無料。
- 公開データ + 公開された遺伝子リスト。

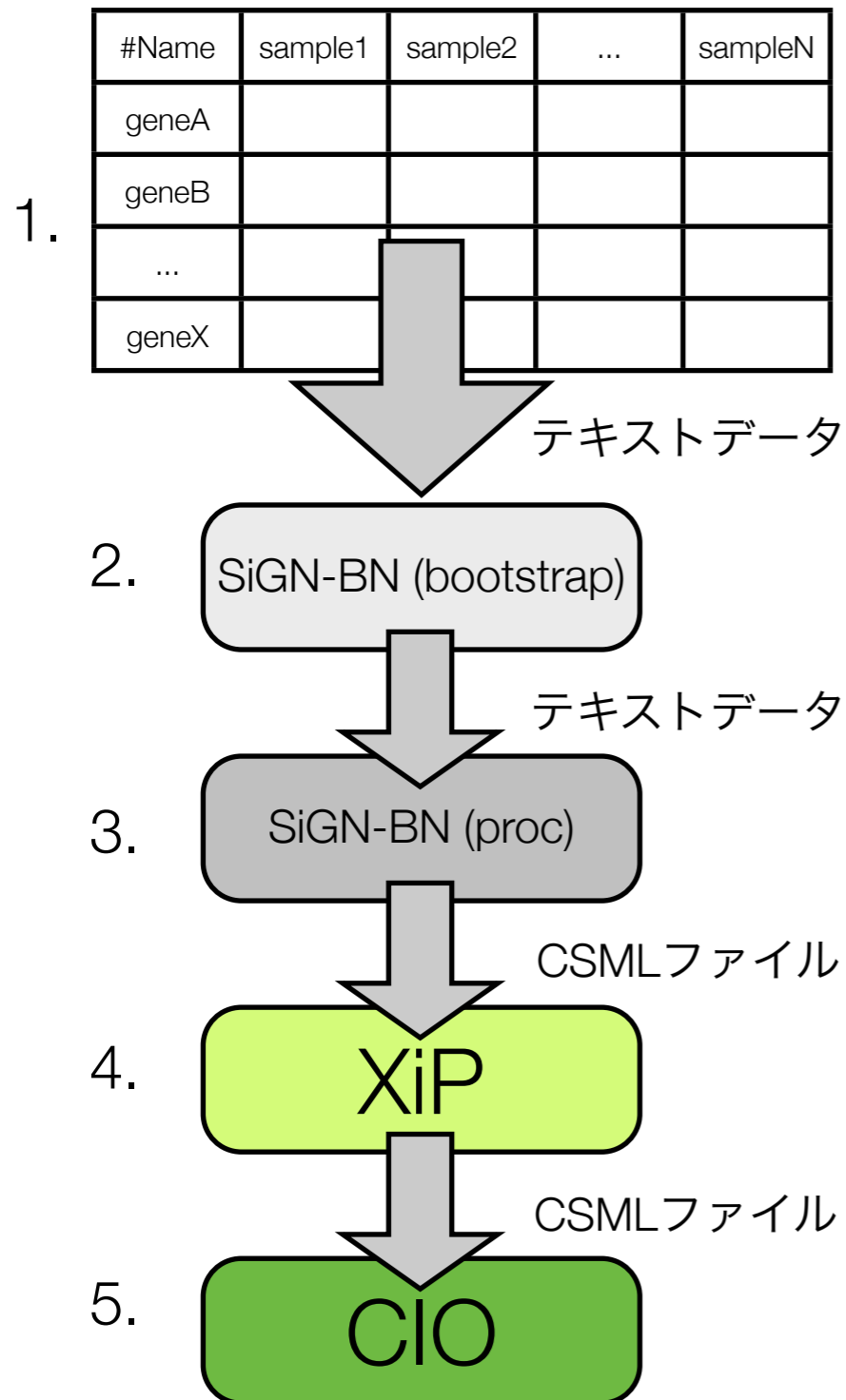
B. GEOから公開データを取得して、独自の遺伝子リストで推定。

- ローコスト。
- 公開データ + 独自の遺伝子リスト。

C. 独自のマイクロアレイデータ、独自の遺伝子リストで推定。

- ハイコスト。
- 独自データ + 独自の遺伝子リスト。

遺伝子ネットワーク解析の流れ



1. 入力データの準備。
2. **SiGN-BN** のブートストラップを実行。
3. **SiGN-BN** の proc コマンドにより遺伝子ネットワークのファイルを生成。
4. **XiP** による遺伝子ネットワークのレイアウト。
5. **CIO Player** による遺伝子ネットワークの表示。

2., 3. の処理は、リモート (SCLS) での作業。

4., 5. の処理は、ローカル (お手持ちのPC) での作業。

cell innovator

1.1 入力データの準備（ファイルの整形）

← 数十から数百サンプル →

sample1 sample2 ... sampleN

数百から1000遺伝子

\$version	1			
geneA	100	200	...	400
geneB	300	300	...	1000
...
geneX	1000	5000	...	20000

- マイクロアレイデータを Expression Data Format (EDF) の形式で準備。（エクセルなどを利用） <http://sign.hgc.jp/formats.html>
- 1行目に \$version を記入。ここでは、「1」。
- サンプル名の行は必要ない。
- **タブ区切りのテキスト**形式で保存。

1.2 入力データの Tips

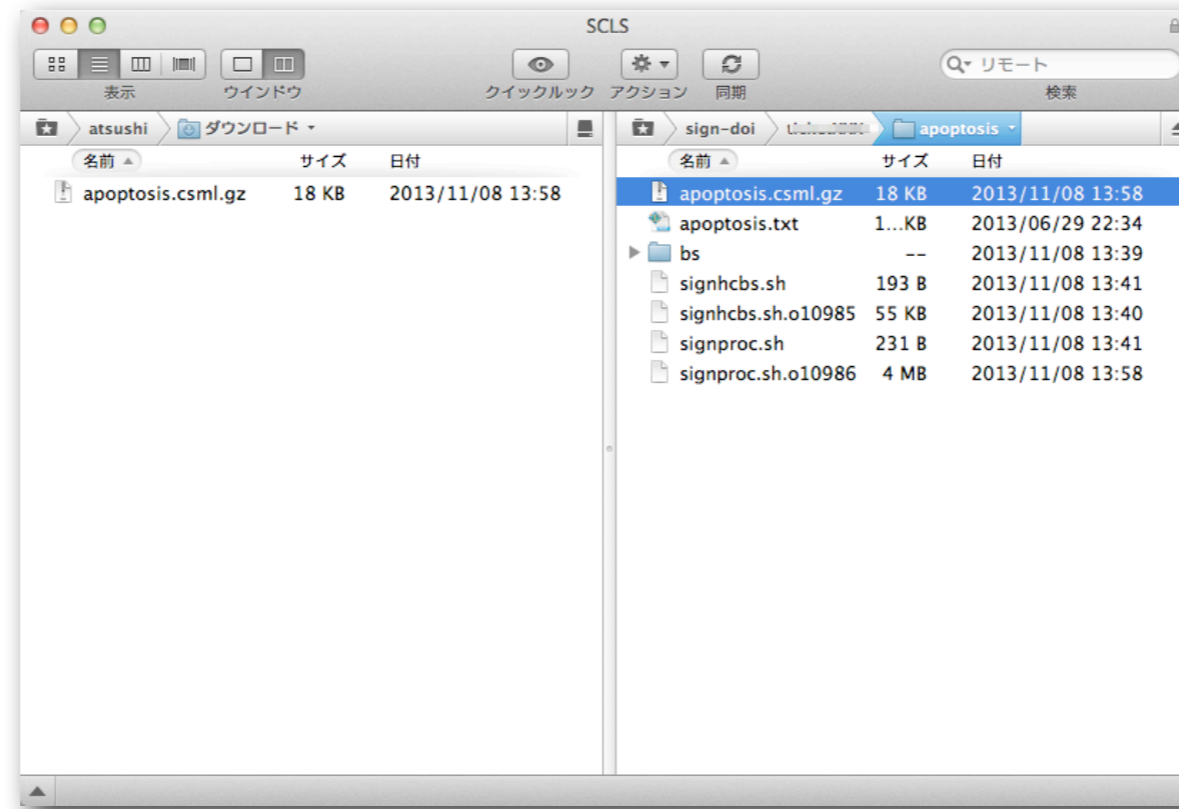
- 遺伝子数が増加すると、計算時間とメモリの消費量が増大。
 - さすがに全遺伝子での遺伝子ネットワーク推定は、まだ、できない。
 - 遺伝子数は、**1000**個くらいをおすすめ。
- 遺伝子ネットワークの推定は、数十サンプルから行えるが、サンプル数が少ないと信頼性は低下する。
 - **100**サンプル程度は欲しい。
 - 50サンプルくらいから使用した経験あり。
- 遺伝子名の重複は不可能。ProbeID を使用するか、同一遺伝子をターゲットしたプローブは、いずれか1個を選ぶ（平均値や中央値の高いものなど）。

1.3 SiGN-BN の準備

```
sign-doi@scls2:~/HCBS  sign-doi@scls2:~/apoptosis  apoptosis — sign-doi@scls2:~/apoptosis — ssh — 80x20
[sign-doi@scls2 ~]$ mkdir apoptosis
[sign-doi@scls2 ~]$ cd apoptosis
[sign-doi@scls2 apoptosis]$ cp /home/sign-doi/tmp/*.sh ./
[sign-doi@scls2 apoptosis]$ mkdir bs
[sign-doi@scls2 apoptosis]$ ls
bs/  signhcbs.sh*  signproc.sh*
[sign-doi@scls2 apoptosis]$
```

- | | |
|-------------------------|--|
| (1) 作業用ディレクトリを作成。 | <code>mkdir apoptosis</code> |
| (2) 作業用ディレクトリに移動。 | <code>cd apoptosis</code> |
| (3) 2つのファイルをコピーする。 | <code>cp /home/sign-doi/tmp/*.sh ./</code> |
| (4) ブートストラップ用ディレクトリを作成。 | <code>mkdir bs</code> |

1.4 データを転送する



- scp [オプション] [コピー元] [コピー先]
- scp -Cp apoptosis.txt sign-doi@hpci-scls.riken.jp:~/

* 実習用のデータをコピー

```
cp /home/sign-doi/tmp/apoptosis.txt ./
```

cell innovator

2.1 SiGN-BN (bootstrap)

```
atsushi — sign-doi@scls:~/apoptosis — ssh — 80x20
[sign-doi@scls apoptosis]$ cat signhcbs.sh cat でシェルスクリプトの中身を確認
#!/bin/sh
#PJM -L "rscgrp=small"
#PJM -L "node=12" <-- node で、使用するノード数を指定。
#PJM --mpi "proc=192" <-- proc で、ブートストラップの回数を指定。通常は1000回以上が望ましい。
#PJM -L "elapse=10:00"
#PJM -j

export OMP_NUM_THREADS=1

mpiexec /home/islim/sign-bn/signbn -o bs/result apoptosis.txt <-- 入力ファイルを指定。
[sign-doi@scls apoptosis]$ pjsub signhcbs.sh
[INFO] PJM 0000 pjsub Job 11410 submitted.
[sign-doi@scls apoptosis]$
```

- SiGN-BN のブートストラップを実行する。
- `pjsub signhcbs.sh`

2.2 ジョブの確認

```
atsushi — sign-doi@scls:~/apoptosis — ssh — 80x20
[sign-doi@scls apoptosis]$ pjstat

  ACCEPT  QUEUED  STGIN  READY  RUNING  RUNOUT  STGOUT  HOLD  ERROR  TOTAL
    0      1      0      0      0      0      0      0      0      1
s    0      1      0      0      0      0      0      0      0      1

JOB_ID    JOB_NAME  MD ST  USER      START_DATE      ELAPSE_LIM  NODE_REQUIRE
11410     signhcs.s NM QUE sign-doi (11/19 18:30)  0000:10:00  12
[sign-doi@scls apoptosis]$
```

- `pjstat` で、ジョブの状況を確認できる。
- 終了したら、`cat bs/result.log` で結果を確認。

3.1 SiGN-BN (proc)

```
sign-doi@scls2:~/apoptosis
[sign-doi@scls2 apoptosis]$ cat signproc.sh
#!/bin/sh
#PJM -L "rscgrp=small"
#PJM -L "node=12"
#PJM --mpi "proc=192"
#PJM -L "elapse=3:00:00"
#PJM -j

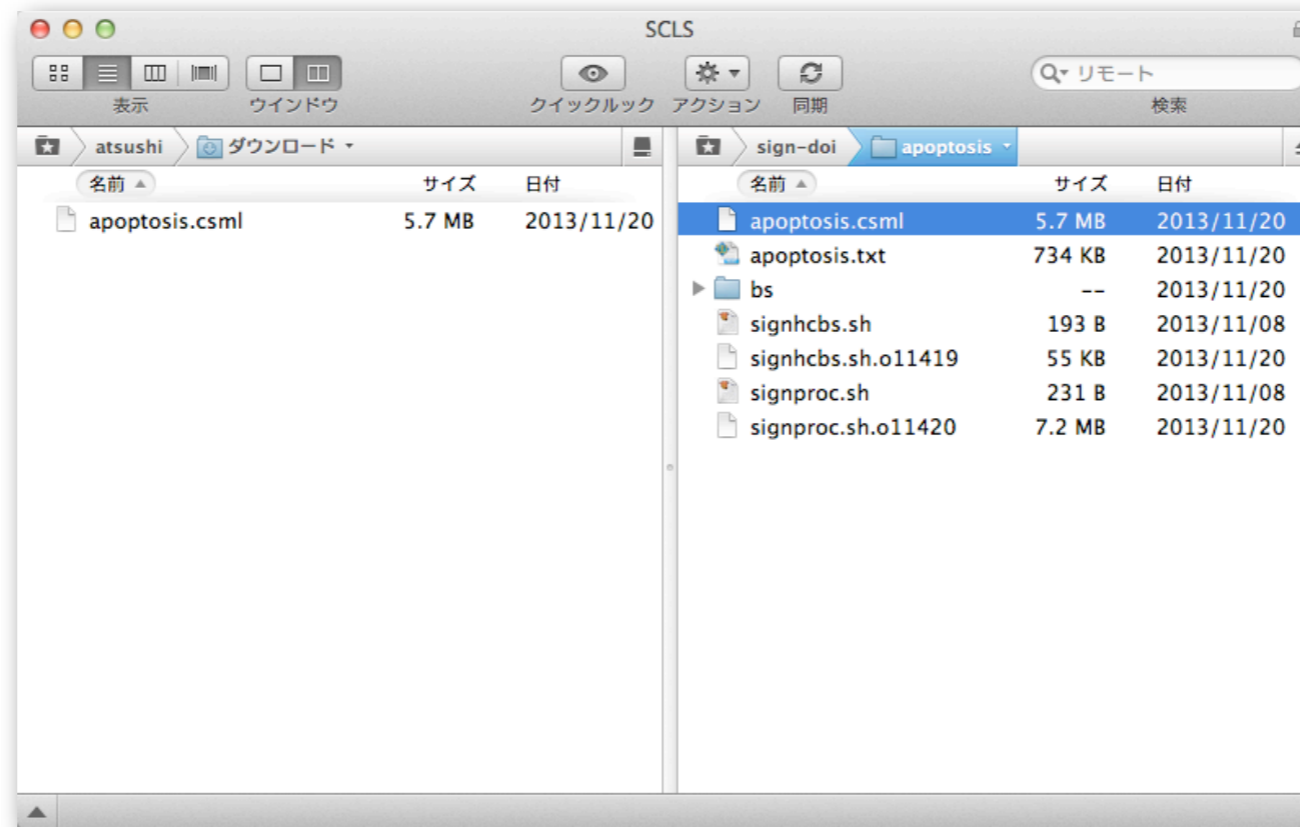
export OMP_NUM_THREADS=1

mpiexec /home/islim/sign-bn/signproc --bs prefix=bs/result --output type=csml,fi
le=apoptosis.csml
[sign-doi@scls2 apoptosis]$ pjsub signproc.sh
[INFO] PJM 0000 pjsub Job 12268 submitted.
[sign-doi@scls2 apoptosis]$
```

--output type オプションで CSML 形式を指定。その他のフォーマットもあり。

- SiGN-BN の proc を実行して、ブートストラップの結果から、遺伝子ネットワークを **csml** の形式で出力する。
- **pjsub signproc.sh**

3.2 結果のファイルを取得する



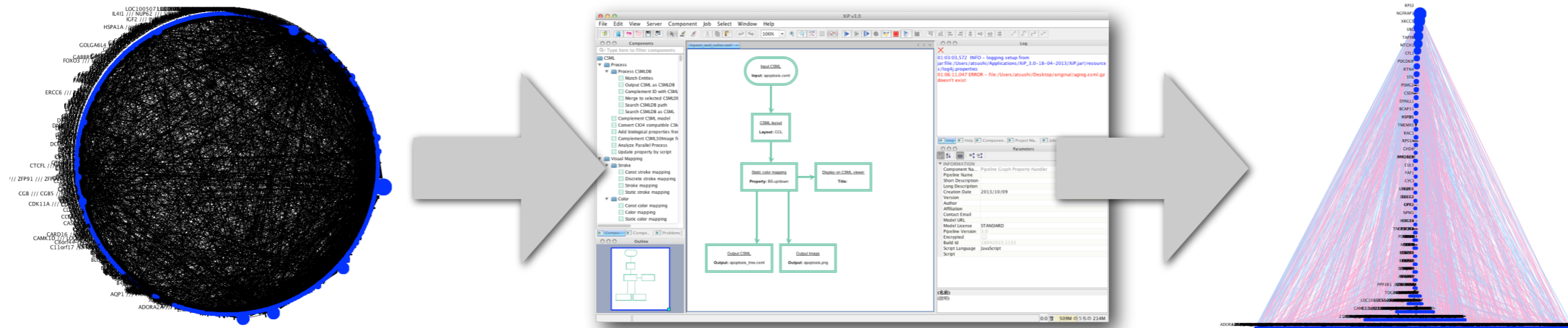
- `scp -Cp sign-doi@hpci-scls.riken.jp:~/apoptosis/apoptosis.csml ./`

* 実習用のCSMLファイルは、下記よりダウンロード可能。

<https://www.dropbox.com/s/ipof4d0t5hs0iv2/apoptosis.csml>

cell innovator

4. XiP による遺伝子ネットワークのレイアウト



- 遺伝子ネットワークは、標準状態では、円形に表示される。
- このままでは、見つらいため、グラフ理論のアルゴリズムを用いて、レイアウトを行う。
- レイアウトには、XiP というソフトウェアを利用できる。

4.1 XiP をダウンロード

The screenshot shows the main page of the XiP website. The title is "eXtensible Integrative Pipeline". Below the title is a large blue "XiP" logo. The page is in Japanese. On the left side, there is a navigation menu with links like "メインページ", "コミュニティ・ポータル", "最近の出来事", "最近の更新", "おまかせ表示", and "ヘルプ". Below the menu is a search box and a "ツールボックス" section with links like "リンク元", "関連ページの更新状況", "特別ページ", "印刷用バージョン", and "この版への固定リンク". At the bottom of the left sidebar, there is a language selection menu with "English" selected. The main content area is titled "アプリケーションの起動" (Application Startup). It features a large blue "Start" button and a "Launcher" button. Below these is a table with columns: Version, Release Date, Release Notes, Online, Online(large), and Offline. The table lists three versions: 1.0 (stable), 2.0 (stable), and 3.0 (stable), plus a "latest" row. A red arrow points to the "launcher" link in the "Online" column of the 3.0 (stable) row. To the right of the table is a "目次" (Table of Contents) section with a list of 14 items. Below the table are buttons for "Manual", "Quick Guide", and "Advanced Usage". At the bottom of the page, there is a "Video (How to start)" section with links for "for desktop" and "for laptop".

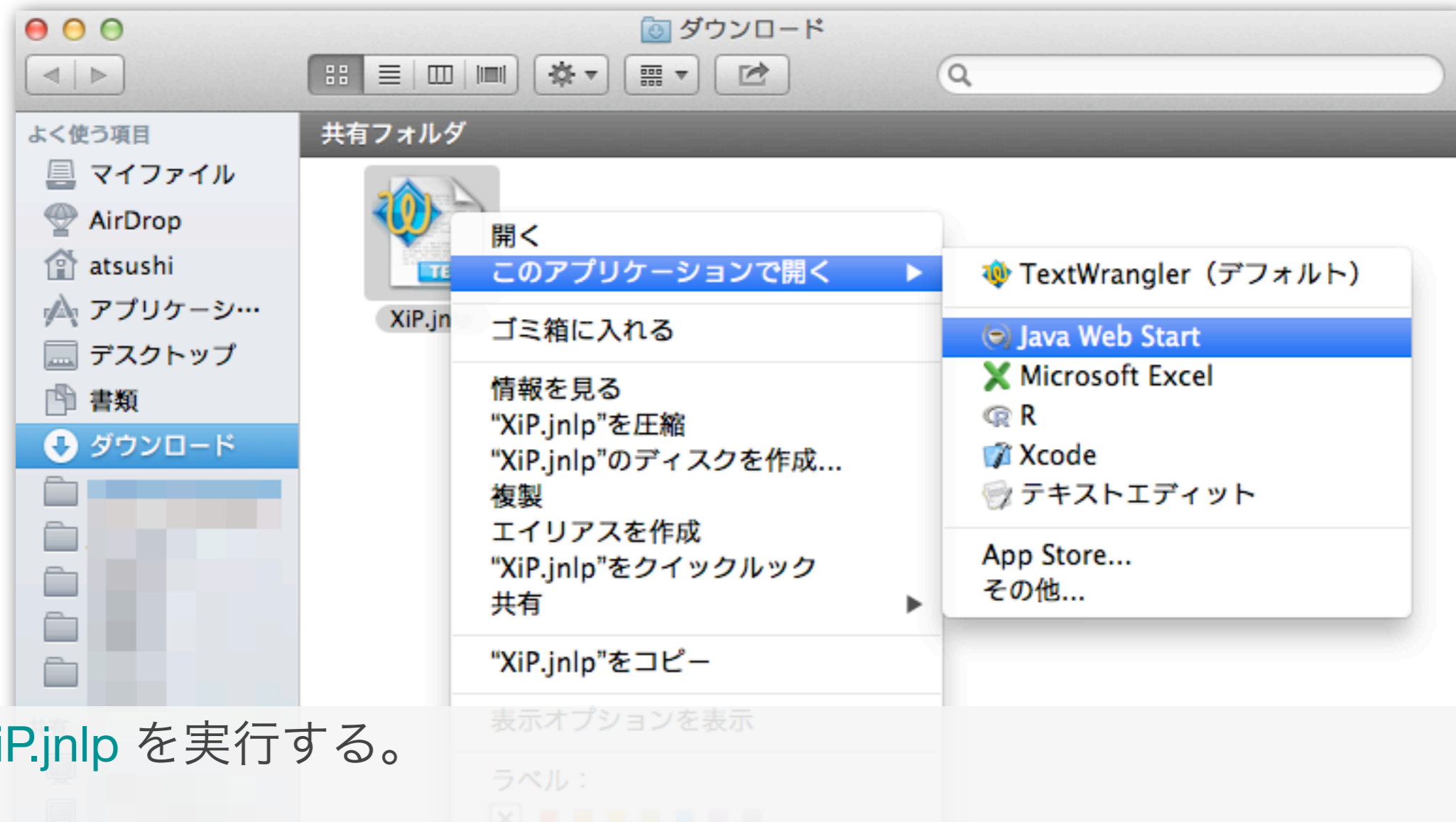
Version	Release Date	Release Notes	Online	Online(large)	Offline
1.0 (stable)	20-04-2012	release notes	launch	launch	download
2.0 (stable)	04-09-2012	release notes	launch	launch	download
3.0 (stable)	18-04-2013	release notes	launch	launch	download
latest	26-11-2013	release notes	launch	launch	download

目次 [非表示]

- 1 アプリケーションの起動
 - 1.1 Online 版
 - 1.2 Offline 版
 - 1.3 Launcher
 - 1.4 アプリケーションサイト
- 2 基本的なアプリケーションの使い方
 - 2.1 フローの作成
 - 2.2 フローの実行
- 3 トレーニングのダウンロード
- 4 パイプラインのサンプル
- 5 講習会
- 6 メニューバーについて
- 7 ウィンドウについて
- 8 ルートコンポーネントカテゴリー一覧
- 9 その他の有用な情報について
- 10 プラグインの利用
- 11 コンポーネントの開発について
- 12 HGC環境での利用について
- 13 開発 Note
- 14 文庫

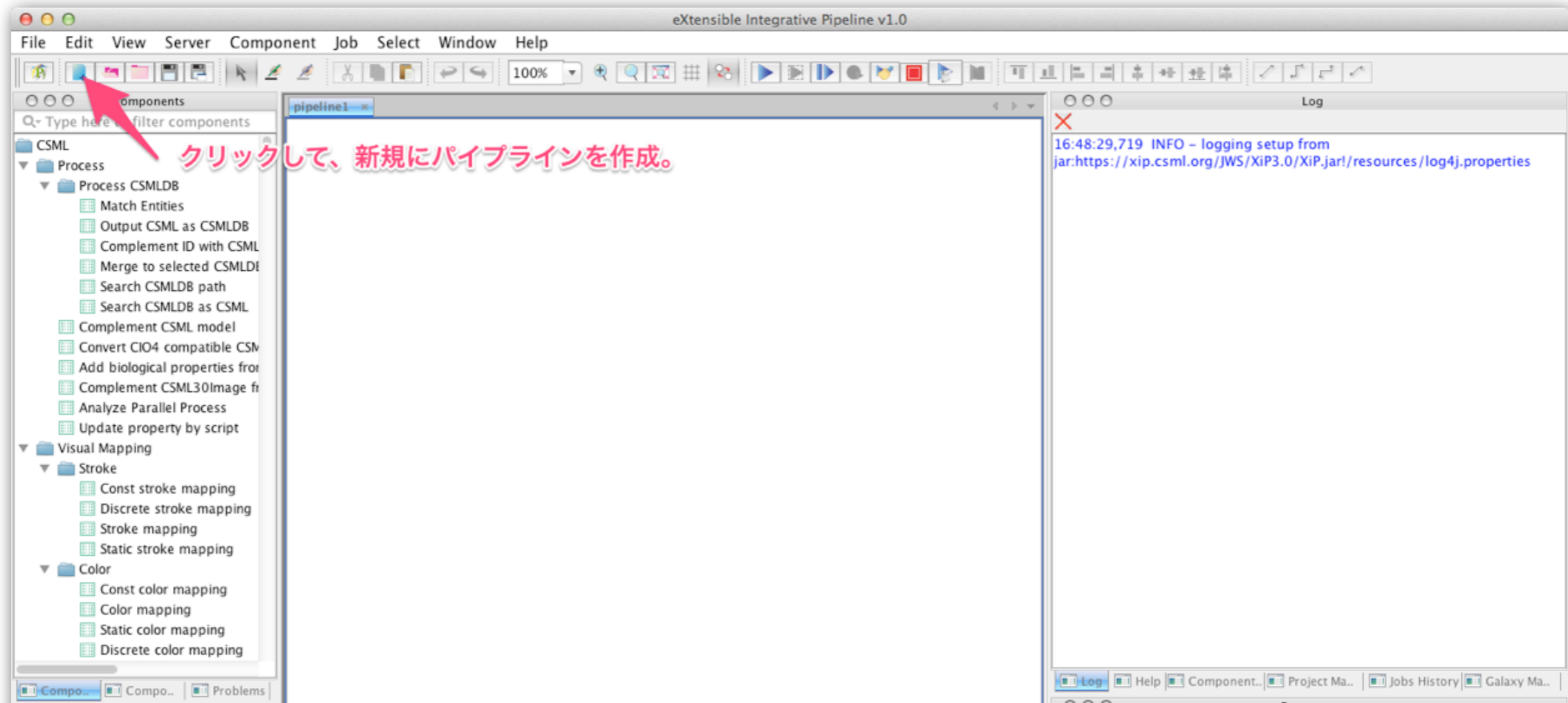
- ブラウザで <http://xip.hgc.jp> を開く。
- Online 版の **launcher** をダウンロードする。
- Offline 版も利用可能。

4.2 XiP の起動



- XiP.jnlp を実行する。
- Mac の場合は、右クリックメニューから、“Java Web Start” を選択。

4.3 新規パイプラインの作成



- アプリケーションのダウンロード後に、XiPが起動する。
- 左上のアイコンをクリックして、新規パイプラインを作成。

4.4 input コンポーネントを配置

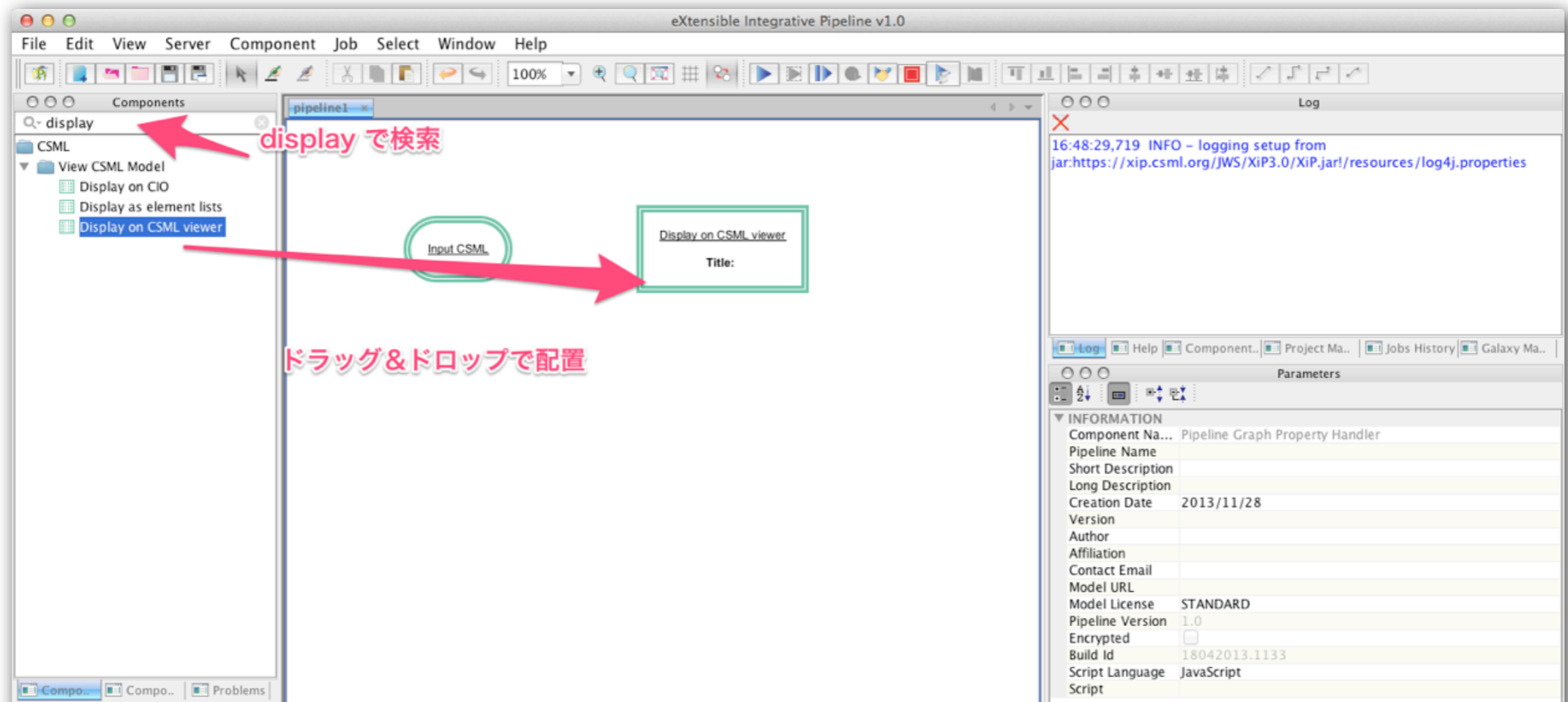
input で検索

ドラッグ&ドロップでコンポーネントを配置。

- input と入力して、コンポーネントを検索。
- **Input CSML** コンポーネントを、ドラッグ&ドロップで配置。

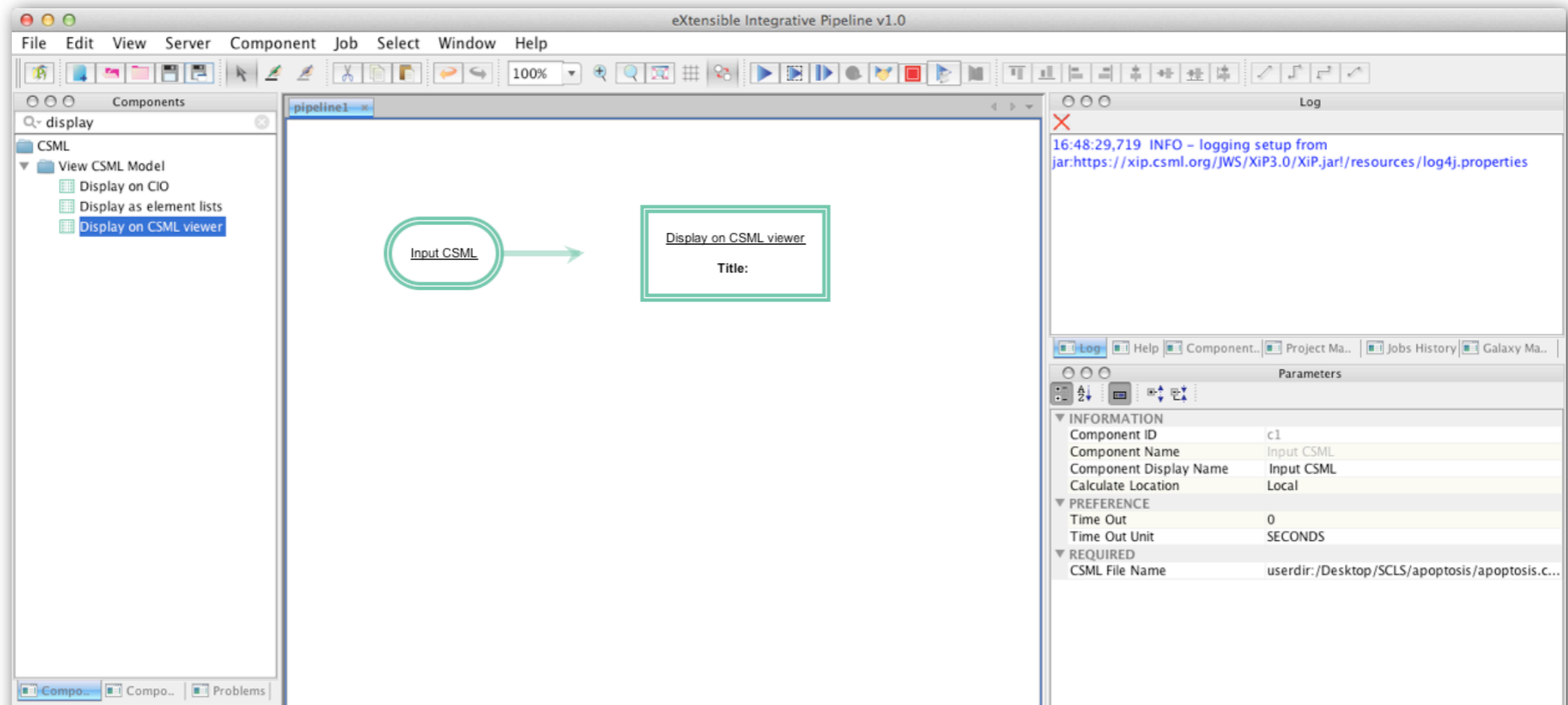
INFORMATION	
Component Na...	Pipeline Graph Property Handler
Pipeline Name	
Short Description	
Long Description	
Creation Date	2013/11/28
Version	
Author	
Affiliation	
Contact Email	
Model URL	
Model License	STANDARD
Pipeline Version	1.0
Encrypted	<input type="checkbox"/>
Build Id	18042013.1133
Script Language	JavaScript
Script	

4.5 display コンポーネントを配置



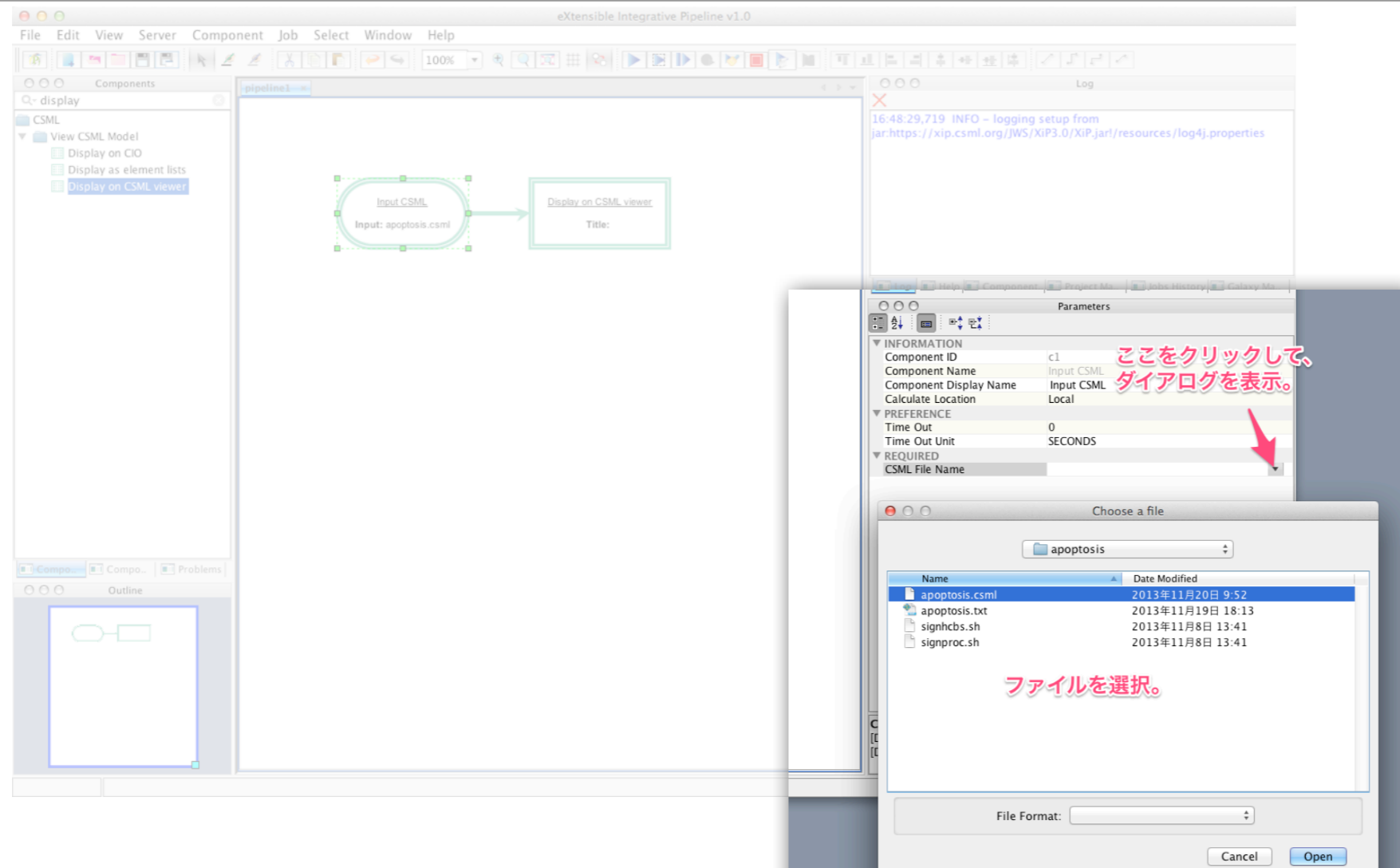
- display と入力して、コンポーネントを検索。
- display on CSML viewer コンポーネントを、ドラッグ&ドロップで配置。

4.6 コンポーネントを接続



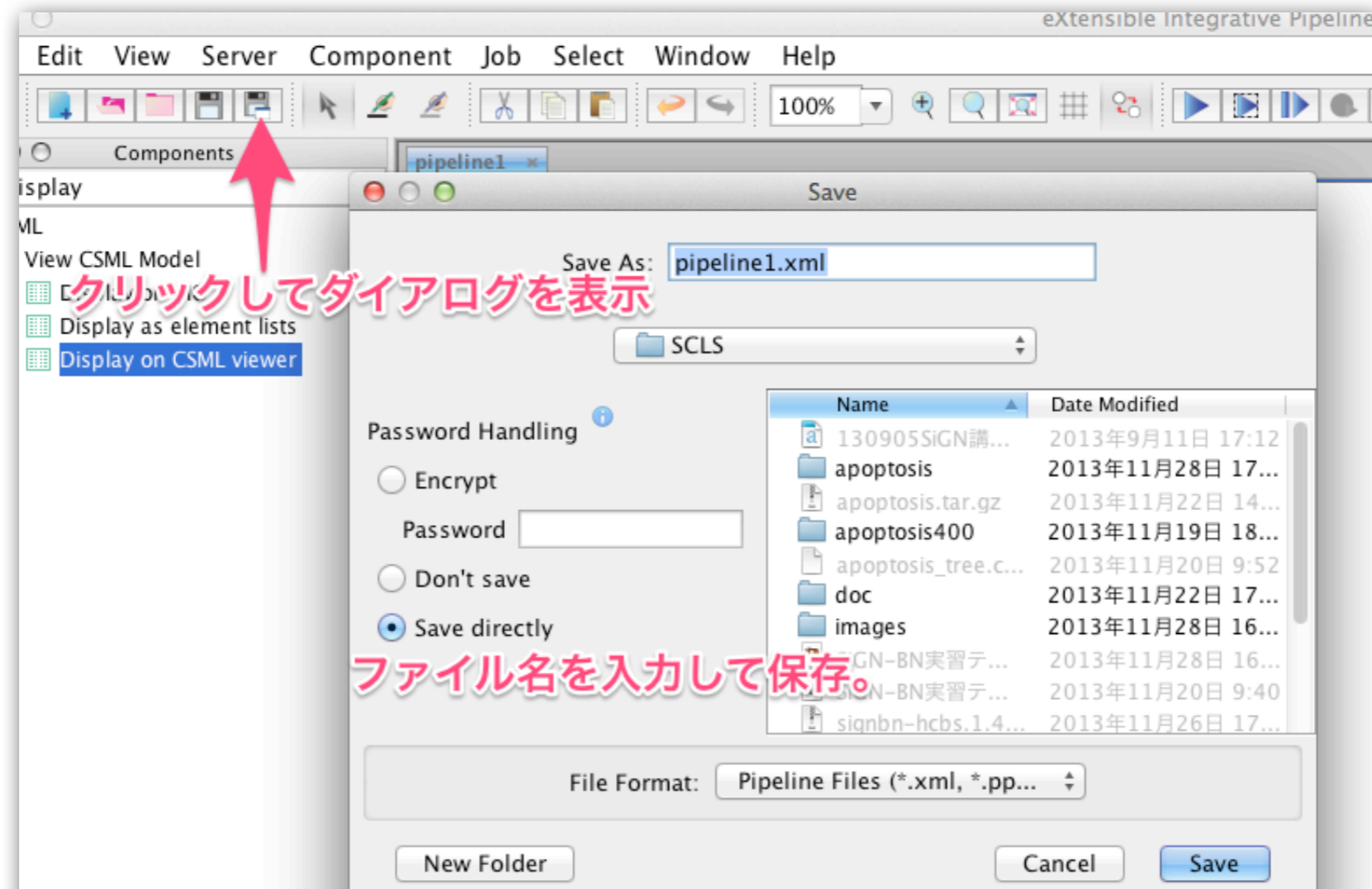
- 各コンポーネントの中心から、コネクタをドラッグで引き出せる。
- **input CSML** コンポーネントから、コネクタを **Display on CSML viewer** コンポーネントに接続する。

4.7 Input CSML コンポーネントに入力ファイル名を設定



- **Input CSML** コンポーネントのパラメーターウィンドウの、**CSML File Name** の部分に、読み込ませる CSML ファイルを設定する。

4.8 作成したパイプラインを保存



- Save as のアイコンをクリックして、ダイアログを表示する。
- ファイル名を入力して、作成したパイプラインを保存する。

4.9 パイプラインの実行

Input CSML
Input: apoptosis.csml

Display on CSML viewer
Title:

クリックして実行

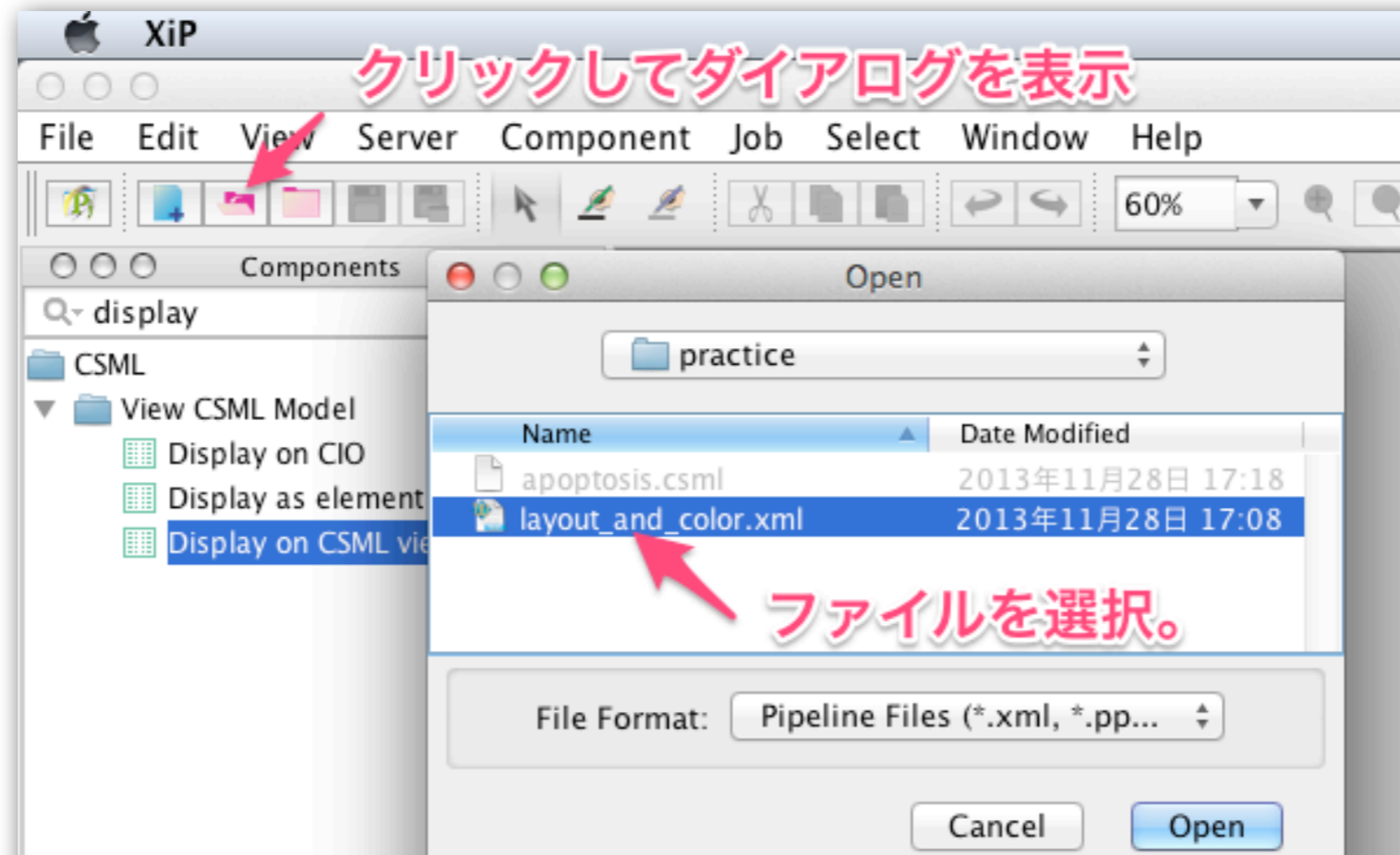
Plot Frame 1
VIEW

遺伝子ネットワークが表示される。

Create Snapshot

- 再生ボタンをクリックすると、パイプラインが実行される。
- ここでは、**遺伝子ネットワーク**の**CSML** ファイルが読み込まれ、CSML viewer にネットワーク図が表示される。（レイアウトなしの場合、円形に表示される。）

4.10 既存のパイプラインの読み込み

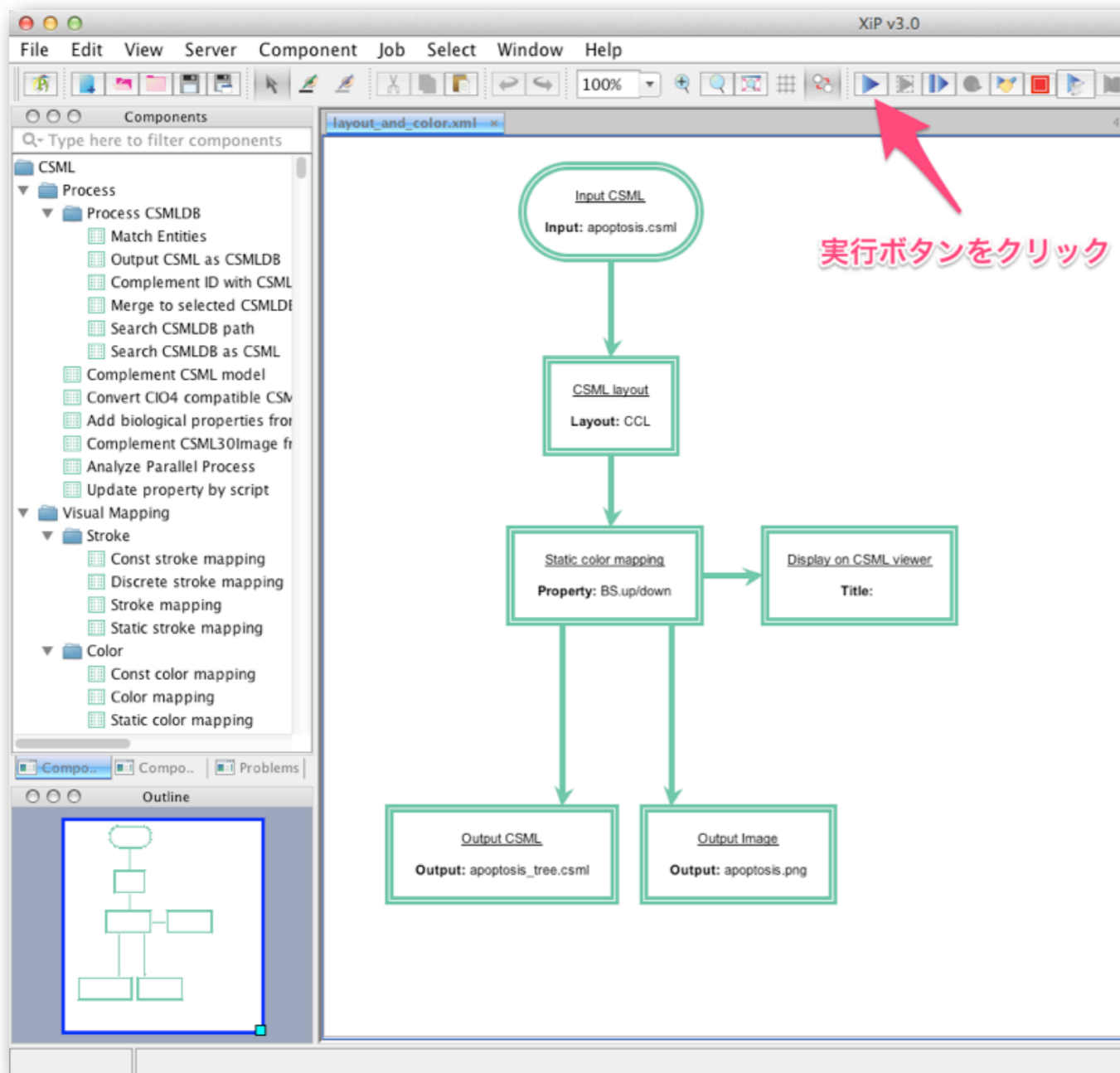


- Open のアイコンをクリックして、ダイアログを表示する。
- ファイル名を選択して、既存のパイプラインを読み込む。

*実習用のパイプラインは、下記よりダウンロード可能。

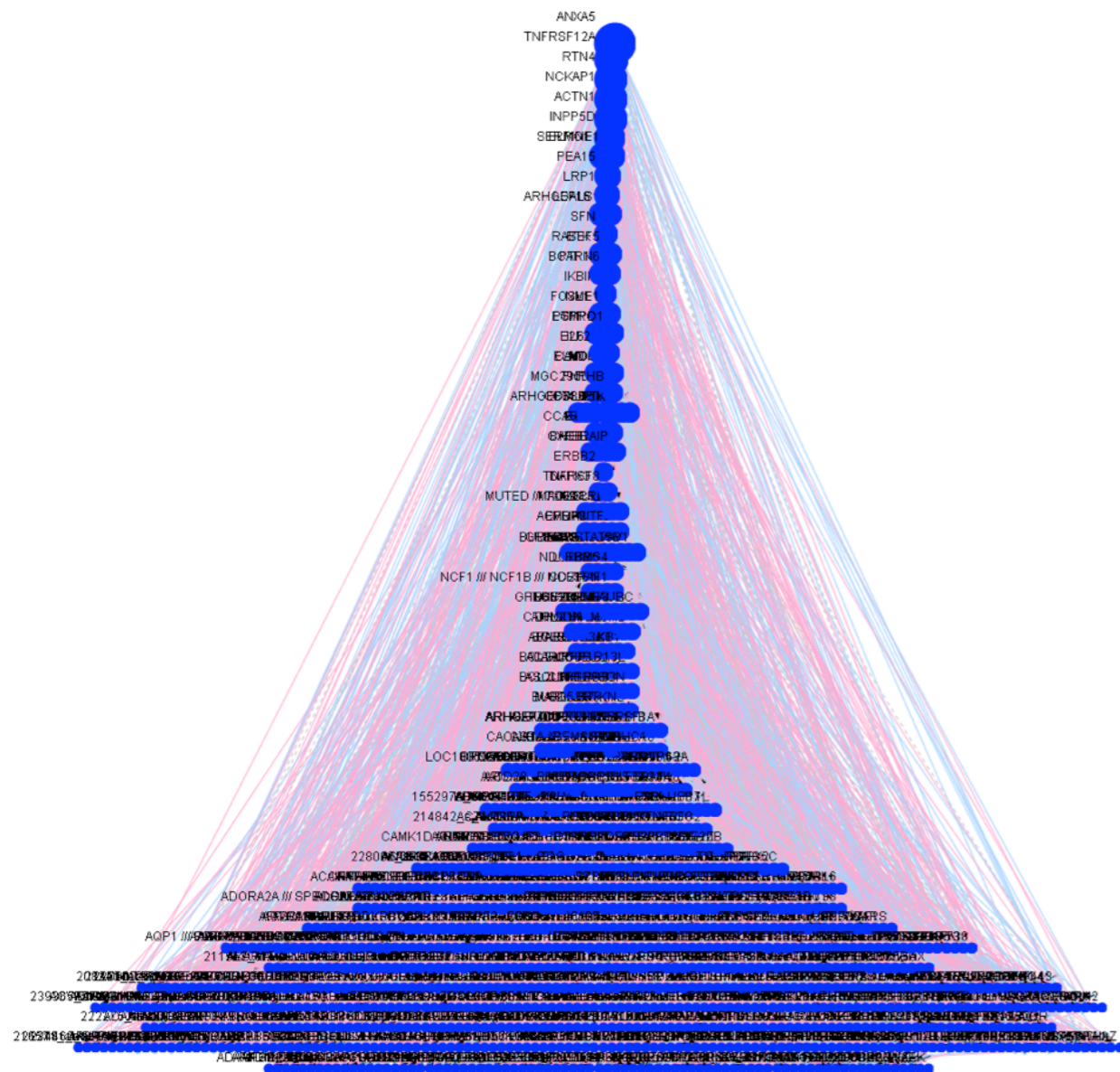
https://www.dropbox.com/s/gdqymcrbnja6d9w/layout_and_color.xml

4.11 既存のパイプラインの実行



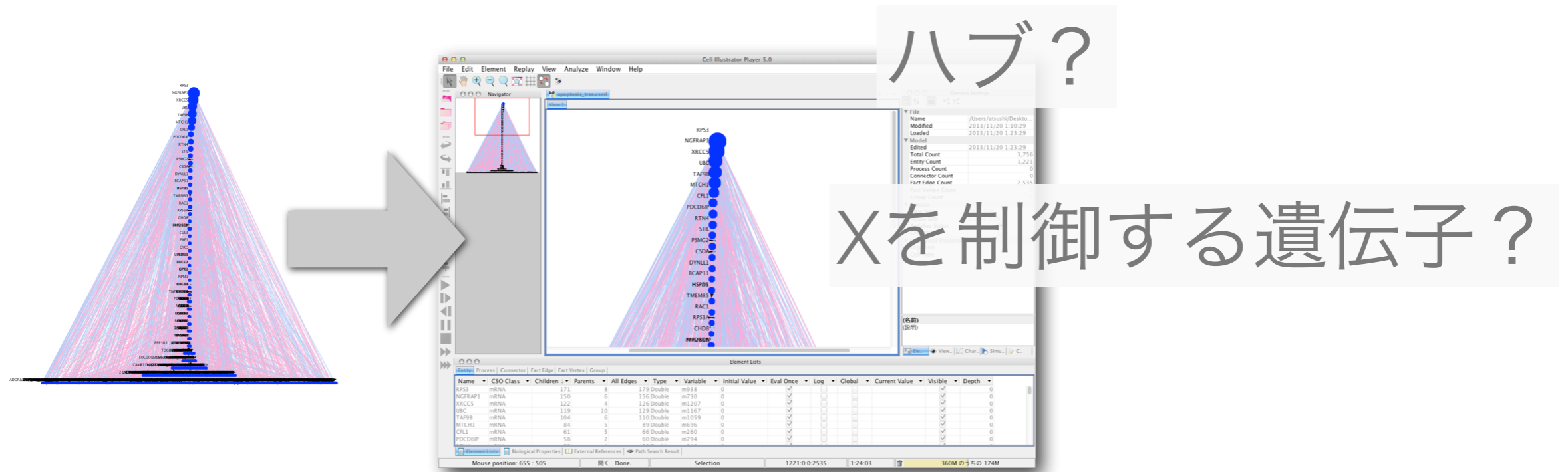
- 必要に応じて、各コンポーネントのパラメーター（入力ファイル名など）を変更。
- 実行ボタンをクリックして、既存のパイプラインを実行。
- この例では、CCLアルゴリズムによって、ツリー上にレイアウトされた遺伝子ネットワークの CSML ファイル ([apoptosis.tree.csml](#)) と画像ファイル ([apoptosis.png](#)) が出力される。

4.12 ツリー状にレイアウトされた遺伝子ネットワーク



- CCL アルゴリズムによりツリー状にレイアウトされた遺伝子ネットワークでは、子の数が多い遺伝子ほど、上部に配置される。
- また、CSML では、子の数が多い遺伝子ほど、大きいサイズのノードで表示させている。
- エッジは、推定されたスコアに応じて、up, down, unknown の3タイプの制御関係を示す。（それぞれ、ピンク、水色、グレーで色づけ。）

5. CIO による遺伝子ネットワークの表示と解析



- 遺伝子ネットワークの表示と解析には、 [Cell Illustrator Online \(CIO\)](#) を利用。
- 閲覧だけであれば、無償の [Cell Illustrator Player](#) が利用できる。

cell innovator

5.1 Cell Illustrator Player をダウンロード

Cell Illustrator Online (CIO)

Cell Illustrator Online 5.0

Main menu

- Registration
- Cell Illustrator
- Cell Illustrator Player

Cell Illustrator Online 5.0

Java Web Start is required to run CIO. Download the latest Java Runtime Environment (JRE) from <http://java.com> and install it on your PC.

Get your login account through the Registration and try CIO.

After a successful registration you may start CIO from the web browser by pressing the Cell Illustrator button above.

If you are already registered you can check your [License Status](#).

NEW CIO 5.0 standalone package is available for download for the licensed users from the [License Status](#) page.

クリックして、ファイルをダウンロード。

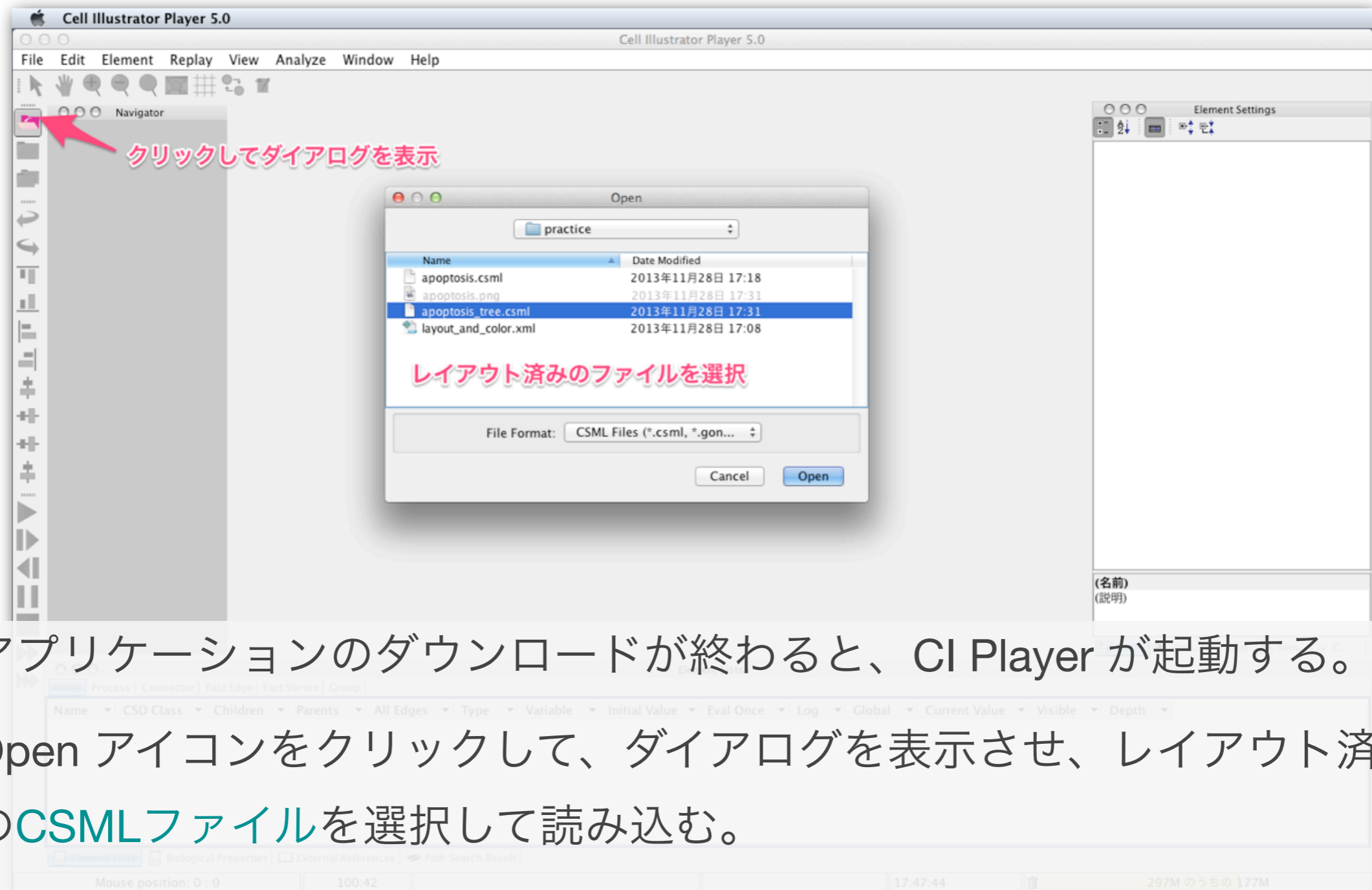
- ブラウザで <https://cionline.hgc.jp/cionlineserver/apps/usersman/main> を開く。
- XiP同様に **launcher** をダウンロードする。

5.2 Cell Illustrator Player を起動



- `launchCIOPlayer.jnlp` を実行する。
- Mac の場合は、右クリックメニューから、“Java Web Start” を選択。

5.3 CSMLファイルを読み込む



5.4 CI Player で表示した遺伝子ネットワーク

The screenshot displays the Cell Illustrator Player 5.0 interface. The main window shows a gene network diagram with nodes and edges. A callout box labeled "全体図" (Overall Diagram) points to a thumbnail in the Navigator panel. Another callout box labeled "ビュー" (View) points to the main diagram area. A third callout box labeled "各種設定情報" (Various Setting Information) points to the Element Settings panel on the right. A fourth callout box labeled "各種パラメーターのテーブル" (Table of Various Parameters) points to the Element Lists table at the bottom.

各種設定情報 (Element Settings)

File	
Name	/Users/atsushi/Desktop...
Modified	2013/11/20 1:10:29
Loaded	2013/11/20 1:23:29

Model	
Edited	2013/11/20 1:23:29
Total Count	3,756
Entity Count	1,221
Process Count	0

Canvas	
View Size	768; 530
Actual Size	1,190; 1,062
Antialias Status	<input checked="" type="checkbox"/>
Grid Status	<input type="checkbox"/>

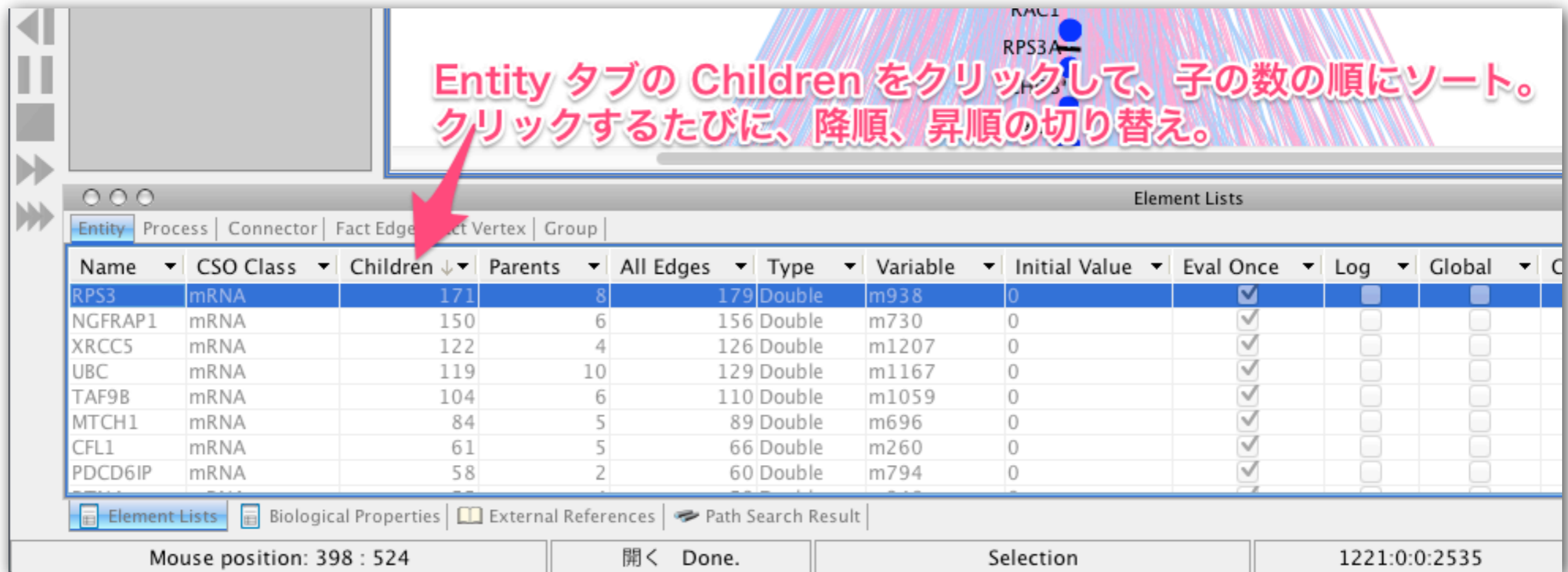
Biological Properties	
Organism	
Cell Type	
Tissue	

(名前)
(説明)

Element Lists

Name	CSO Class	Children	Parents	All Edges	Type	Variable	Initial Value	Eval Once	Log	Global	Current Value	Visible	Depth
RPS3	mRNA	171	8	179	Double	m938	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
NGFRAP1	mRNA	150	5	155	Double	m730	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
XRCC5	mRNA	122	4	126	Double	m730	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
UBC	mRNA	119	10	129	Double	m730	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
TAF9B	mRNA	104	6	110	Double	m730	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
MTCH1	mRNA	84	3	87	Double	m938	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
CFL1	mRNA	61	5	66	Double	m260	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0
PDCD6IP	mRNA	58	2	60	Double	m794	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	0

5.5 テーブルによるハブ遺伝子の選択（ソート）



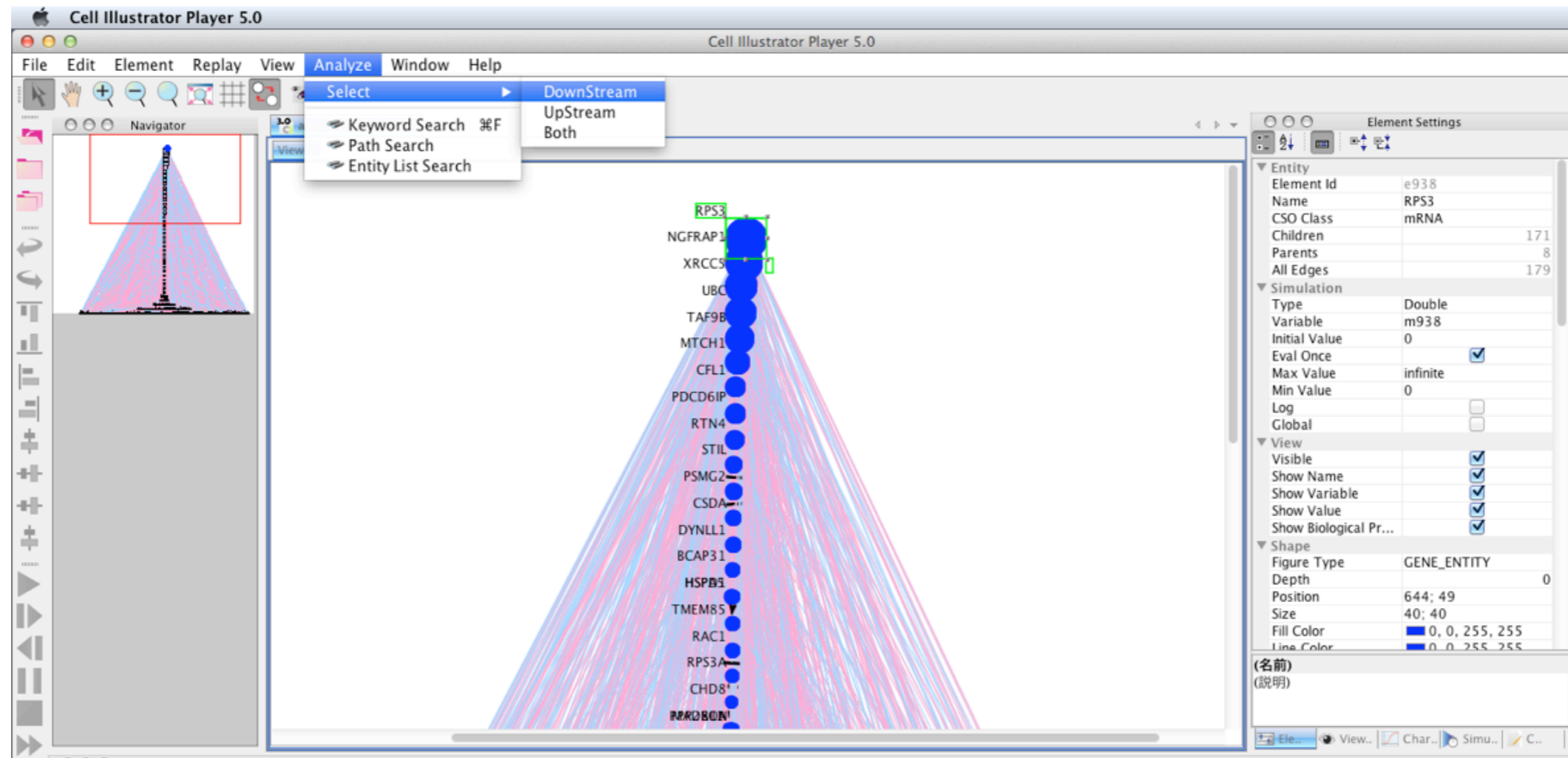
Entity タブの Children をクリックして、子の数の順にソート。
クリックするたびに、降順、昇順の切り替え。

Name	CSO Class	Children	Parents	All Edges	Type	Variable	Initial Value	Eval Once	Log	Global
RPS3	mRNA	171	8	179	Double	m938	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
NGFRAP1	mRNA	150	6	156	Double	m730	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
XRCC5	mRNA	122	4	126	Double	m1207	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
UBC	mRNA	119	10	129	Double	m1167	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TAF9B	mRNA	104	6	110	Double	m1059	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
MTCH1	mRNA	84	5	89	Double	m696	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CFL1	mRNA	61	5	66	Double	m260	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PDCD6IP	mRNA	58	2	60	Double	m794	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Mouse position: 398 : 524 開く Done. Selection 1221:0:0:2535

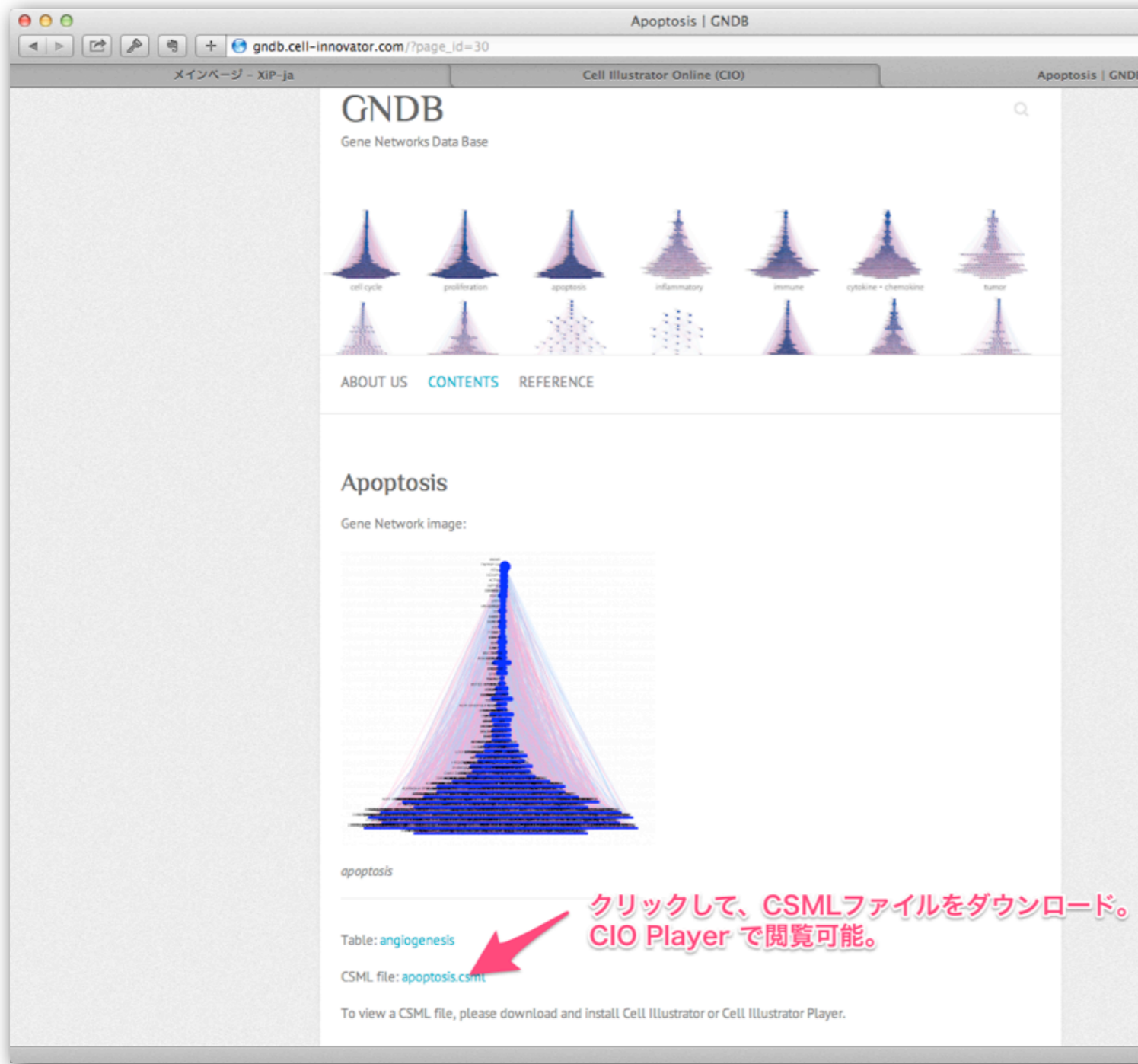
- 下部のテーブルの Entity タブのうち、Children 列をクリックして、子の順にソートできる。クリックするたびに降順、昇順を切り替えられる。
- テーブルの遺伝子名をクリックして、直接、ノードを選択できる。

5.5 下流の遺伝子の選択



- 遺伝子を選択した状態で、メニューから、**Analyze --> Select --> DownStream** をクリック。
- 下流の遺伝子が選択される。（**CIO** では、この部分だけの抽出が可能。）

5.6 既存の遺伝子ネットワークの利用



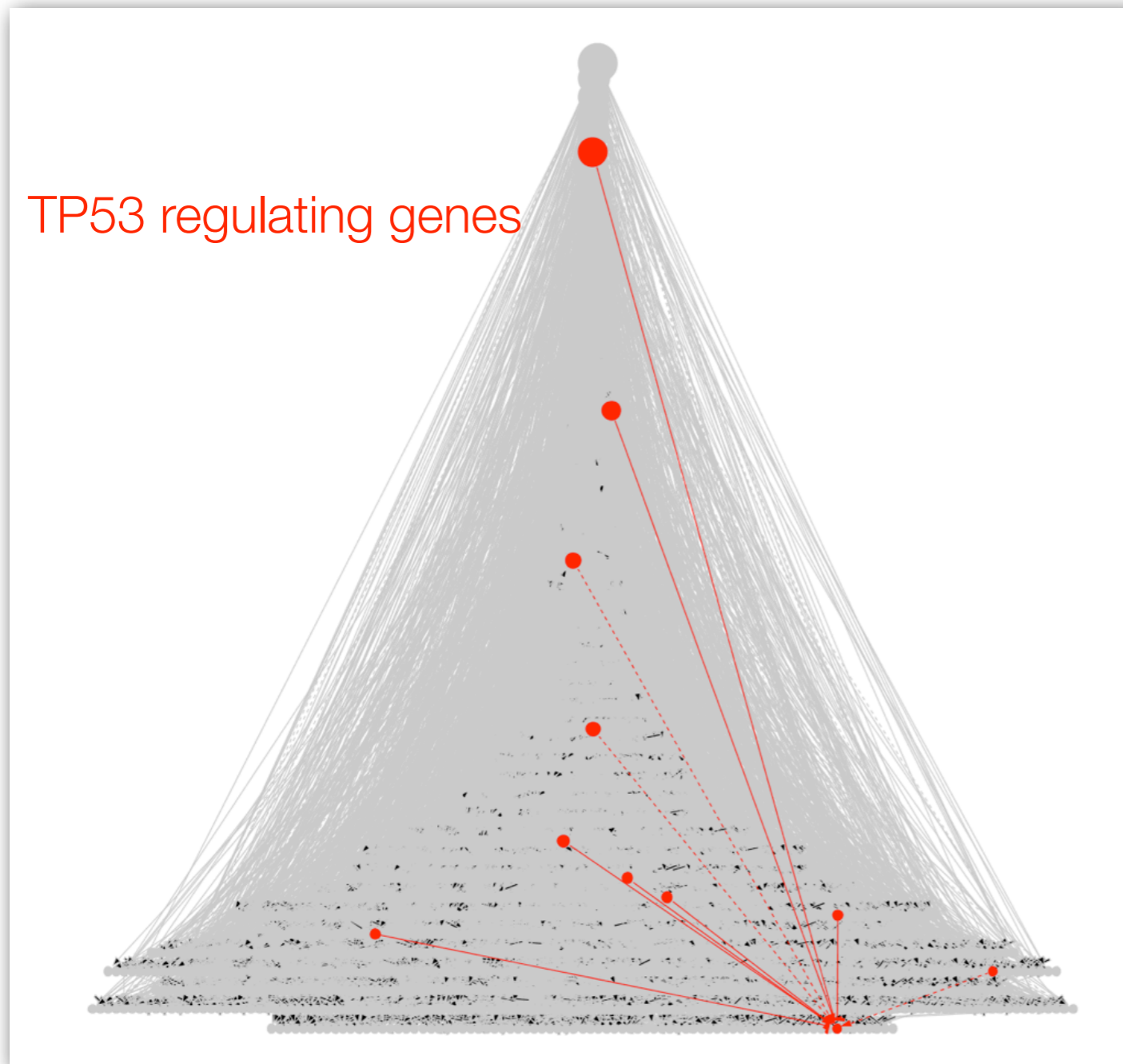
The screenshot shows the GNDDB (Gene Networks Data Base) website. The main content area displays a gene network image for 'Apoptosis'. Below the network image, there is a table with the following information:

Table: angiogenesis
CSML file: apoptosis.csml

A red arrow points to the 'CSML file: apoptosis.csml' link. Below the table, there is a note: 'To view a CSML file, please download and install Cell Illustrator or Cell Illustrator Player.'

- **GNDDB** (<http://gndb.cell-innovator.com>)などで、公開されている **CSML** ファイルを取得すれば、CI Playerで閲覧できる。
- ハブ遺伝子、上流、下流の遺伝子の探索などが可能。
- **GNDDB** は、Cancer Cell Line Encyclopedia (**CCLE**) で公開されたマイクロアレイデータを利用して推定。

5.7 既存の遺伝子ネットワークの利用



- TP53 を選択して、**Analyze** --> **Select** --> **UpStream** をクリック。
- TP53 の**遺伝子発現に影響を与えているように見える遺伝子**を選択できる。
- 左図は、CIO で、さらに、**Extract Subnet** を行ったもの。

リソース

- データ
 - CCLE; <http://www.broadinstitute.org/ccle/home>
 - GEO; <http://www.ncbi.nlm.nih.gov/geo/>
 - GNDB; <http://gndb.cell-innovator.com>
- ソフトウェア
 - CIO; <https://cionline.hgc.jp/cionlineserver/apps/usersman/main>
 - SiGN; <http://sign.hgc.jp/signbn/index.html>
 - XiP; <http://xip.hgc.jp>