独立行政法人理化学研究所 HPCI計算生命科学推進プログラム



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

## 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)
  - <u>http://sign.hgc.jp/signbn/index.html</u>

\*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。
  - <u>http://www.csml.org</u>





## 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。
  - <u>https://cionline.hgc.jp/cionlineserver/apps/usersman/main</u>

- 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 オーム。GUIを用いて、遺伝子ネットワークやNGSの解析 を行える。
  - <u>http://nagasakilab.csml.org/ja/xip</u>





### 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); <u>http://gndb.cell-innovator.com</u>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。

## 1.1 入力データの準備(ファイルの整形)

		sample1	sample2	 sampleN
から1000遺伝子	\$version	1		
	geneA	100	200	 400
	geneB	300	300	 1000
数百	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)<u>http://sign.hgc.jp/formats.html</u>

──── 数十から数百サンプル ───→

- 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。ProbelD 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

	<pre>sign-doi@scls2 ~]\$ mkd [sign-doi@scls2 ~]\$ cd [sign-doi@scls2 apoptos [sign-doi@scls2 apoptos [sign-doi@scls2 apoptos [sign-doi@scls2 apoptos bs/ signhcbs.sh* sign [sign-doi@scls2 apoptos</pre>	<pre>ptosis ir apoptosis apoptosis is]\$ cp /home/sign-doi/t is]\$ mkdir bs is]\$ ls proc.sh* is]\$ [</pre>	mp/*.sh ./
(1)	作業用ディレクトリ	を作成。	mkdir apoptosis
(2)	作業用ディレクトリ	に移動。	cd apoptosis
(3)	2つのファイルをコヒ	<u>゚</u> ーする。	cp /home/sign-doi/tmp/*.sh ./
(4)	ブートストラップ用き	ディレクトリを作成。	mkdir bs

## 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 <u>sign-doi@hpci-scls.riken.jp</u>:~/

\*実習用のデータをコピー

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

### 2.1 SiGN-BN (bootstrap)



Tips



## 2.2 ジョブの確認

ACO	CEPT QUE	EUED ST	GIN R	EADY RUN	NING RUN	NOUT STO	GOUT	HOLD	ERROR	TOTAL
-	0	1	0	0	0	0	0	0	0	1
5	V	T	V	V	V	V	V	V	V	Ť
JOB_]	ID J	IOB_NAME	E MD	ST USEF	R ST	TART_DA	TE	ELAP	SE_LIM	NODE_REQUIRE
1141(	0 s	signhcbs	s.s NM	QUE sigr	n-doi (1	1/19 18	8:30)	0000	:10:00	12
				🔺 🗌			-			
sigr	n-doi@sc	cls apop	otosis]	\$			-			
[sigr	n-doi@sc	ls apop	otosis]	\$			-			
[sigr	n-doi@sc	ls apop	otosis]	\$						
[sigr	n-doi@sc	ls apop	otosis]	\$						
[sigr	n-doi@sc	ls apop	otosis]	\$						
[sigr	n-doi@sc	ls apop	otosis]	\$						
[sign	n-doi@so	こ <sup>s</sup> apop	otosis]	\$ [ 兄を確詞	忍でき	3.				
[sign	n-doi@so	こ <sup>s</sup> apop	otosis] で状況	\$ [] 兄を確詞	忍でき	3.				
[sigr ojsta 終了	n-doi@so at で、	:ls apop ジョブ	otosis] で状況 t bs/re	\$ [] 兄を確詞	忍でき o で結	る。 果を確				
[sigr	n-doi@sc	:ls apop	otosis]	\$			-			
[sigr	n-doi@sc	ls apop	otosis]	\$						
[sigr	n-doi@sc	ls apop	otosis]	\$						
[sign	n-doi@sc	こs apop	`の状?	\$ [] 兄を確言	忍でき	3.				
[sigr	n-doi@so	こ <sup>s</sup> apop	otosis] の状況	\$ [] 兄を確詞	忍でき	3.				
_sigr	n-doi@so	こ <sup>s</sup> apop	otosis] で状況	\$ ] 兄を確詞	忍でき	3.				
[sigr ojsta 終了	n-doi@so at で、	こち apop ジョブ	otosis] で状況 t bs/re	\$ ] 兄を確詞 sult.loo	忍でき g で結	る。 果を確	三汉。			

### 3.1 SiGN-BN (proc)



## 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

## 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、見づらいため、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



### 4.2 XiP の起動



• 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



## 4.4 input コンポーネントを配置



• Input CSML コンポーネントを、ドラッグ&ドロップで配置。

0:0 亩 394M のうちの 236M

## 4.5 display コンポーネントを配置



0:0 盲 394M のうちの 276M

## 4.6 コンポーネントを接続



### 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

0			eXtensible Integrative Pipe
Edit View Server (	Component Job Select Windo	w Help	
I 🔍 🖿 🗖 📮	k 🖉 🖉 👗 🗎 🌔 🌳 🍕	100% 🔻 🔍 🗖	
O Components	pipeline1 ×		
splay 🔨	$\bigcirc \bigcirc \bigcirc$	Save	
IL			
View CSML Model	Save As: ninel	ine1 xml	
0111 June 17	ダイアログを表示		
Display as element lists	ALL HACKIN		
Display on CSML viewer	SCLS	* *	
		Name 🔺	Date Modified
	Password Handling	130905SiGN講	2013年9月11日 17:12
	Encount	apoptosis	2013年11月28日 17
		apoptosis.tar.gz	2013年11月22日 14
	Password	apoptosis400	2013年11月19日 18
	O Don't save	apoptosis_tree.c	2013年11月20日 9:52
	O Don't save	doc	2013年11月22日 17
	<ul> <li>Save directly</li> </ul>	images	2013年11月28日 16
	ファイルタを入力し	7 保友GN-BN実習テ	2013年11月28日 16
	27 - 170-11 C/(J) C	<b>● MALION</b> -BN実習テ	2013年11月20日 9:40
		signbn-hcbs.1.4	2013年11月26日 17
	File Formati	Pineline Files (* vml * nn	•
	File Format.	ripenne rites ( .xini, .pp	• •
	New Folder	(	Cancel Save

- 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



 ・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

	🗮 XiP	<b>ク</b> !	リックしてダイアロ	コグを表示				
	File Edit	View Serv	er Component Job Sel	ect Window Help				
	000	Components	Open					
	Q- display		practice	*				
	View CS	ML Model	Name	Date Modified				
	Disp	lay as element	apoptosis.csml	2013年11月28日 17:18				
	Disp	lay on CSML vie	File Format: Pipeline	2013年11月28日 17:08 <b>イルを選択。</b> Files (*.xml, *.pp キ Cancel Open				
Open 0	)アイコ	ンをク	リックして、ダ-	イアログを表示す	「る。			
ファイ	し名を選	<b>蟚択して</b>	、既存のパイプ	ラインを読み込る	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



### 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。

## 5.1 Cell Illustrator Player をダウンロード



## 5.2 Cell Illustrator Player を起動



launchClOPlayer.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

000	Cell Illustrator Player 5.0	
File Edit Element Replay View Analyze	Window Help	
Navigator クリックしてダイア	ログを表示	이 이 Element Settings 한 한 문화
	Open Open	
₽	practice ‡	
<b>\$</b>	Name Date Modified	
T	apoptosis.csml 2013年11月28日 17:18	
11	副 apoptosis.png         2013年11月28日 17:31           副 apoptosis_tree.csml         2013年11月28日 17:31	
	1 layout_and_color.xml 2013年11月28日 17:08	
크	レイアウト済みのファイルを選択	
1.5		
+-	File Format: CSML Files (*.csml, *.gon +	
+		
1	Cancel Open	
N		(名前) (設明)
		(00-73)
プリケーション	のダウンロードが終わると CIPI	averが記動する。
1 1 1 1 1 1		
non アイコンを・	クリックして ダイアログを表示さ	、廿 レイアウト溶
heir la la la		
OSMLファイル	を選択して読み込む。	
Element Lists 🛛 🕫 Biological Properties 🛄 Exter	nal References   🥗 Path Search Result	

## 5.4 Cl Player で表示した遺伝子ネットワーク



# 5.5 テーブルによるハブ遺伝子の選択(ソート)

RPS3A 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。							>			
 Entity Pro	cess   Connector	Fact Edge act Verte	ex Group	All Edges 🔻 Type	▼ Variable	▼ Initial Value ▼	Eval Once	• Log •	Clobal	
	mRNA	171	8	179 Doubl	a m938			Log	Global	
NGFRAP1	mRNA	150	6	156 Doubl	e m730	0	<b>V</b>			
XRCC5	mRNA	122	4	126 Doubl	e m1207	0	<ul> <li>Image: A start of the start of</li></ul>		ŏ	
UBC	mRNA	119	10	129 Doubl	e m1167	0	$\checkmark$			
TAF9B	mRNA	104	6	110 Doubl	e m1059	0	$\checkmark$			
MTCH1	mRNA	84	5	89 Doubl	e m696	0	$\checkmark$			
CFL1	mRNA	61	5	66 Doubl	e m260	0	$\checkmark$			
PDCD6IP	mRNA	58	2	60 Doubl	e m794	0				
Element Lists Biological Properties External References 🛹 Path Search Result										
Mo	ouse position: 39	98 : 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



# 5.6 既存の遺伝子ネットワークの利用

00	Apoptosis   GNDB	
▶   [2]   ♪   考   + (● gndb.cell-i メインページ – XiP-ja	nnovator.com/?page_id=30 Cell Illustrator Online (CIO) Apoptosis   GN	DB
	GNDB Gene Networks Data Base	
	Image: Second	
	ABOUT US CONTENTS REFERENCE	
	Apoptosis Gene Network image:	
	apoptosis クリックして、CSMLファイルをダウンロード。 CIO Player で閲覧可能。	
	CSML file: apoptosis.csmt To view a CSML file, please download and install Cell Illustrator or Cell Illustrator Player.	

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