

統合化推進プログラム 統合データ解析トライアル
平成 27 年度 キックオフミーティング

ChIP-seq SRA の統合的可視化と バイオデータベースとの連携

沖 真弥

九大・院・医・発生再生医学分野

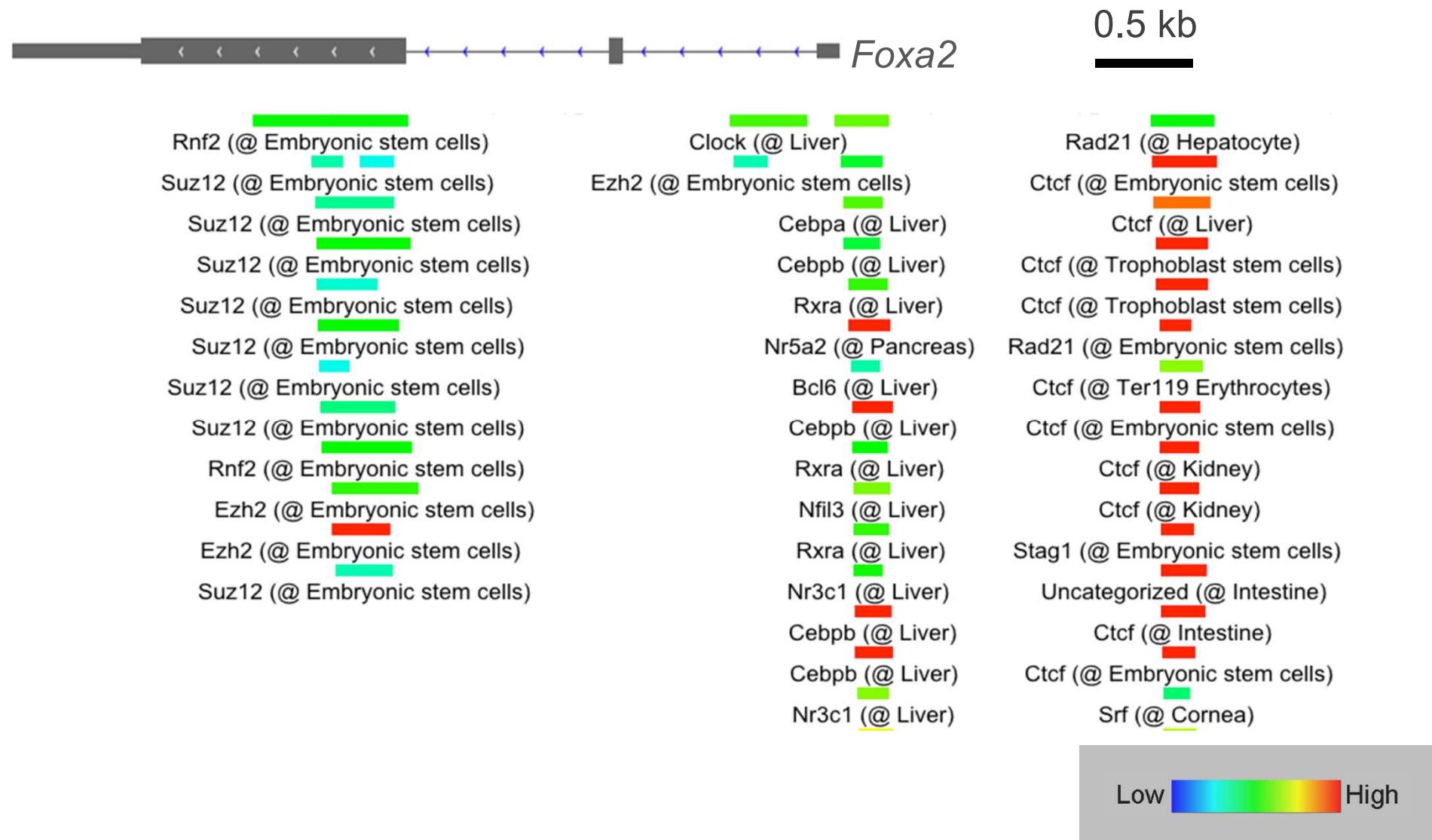


九州大学



©2015 沖 真弥(九州大学大学院) licensed under CC表示2.1日本

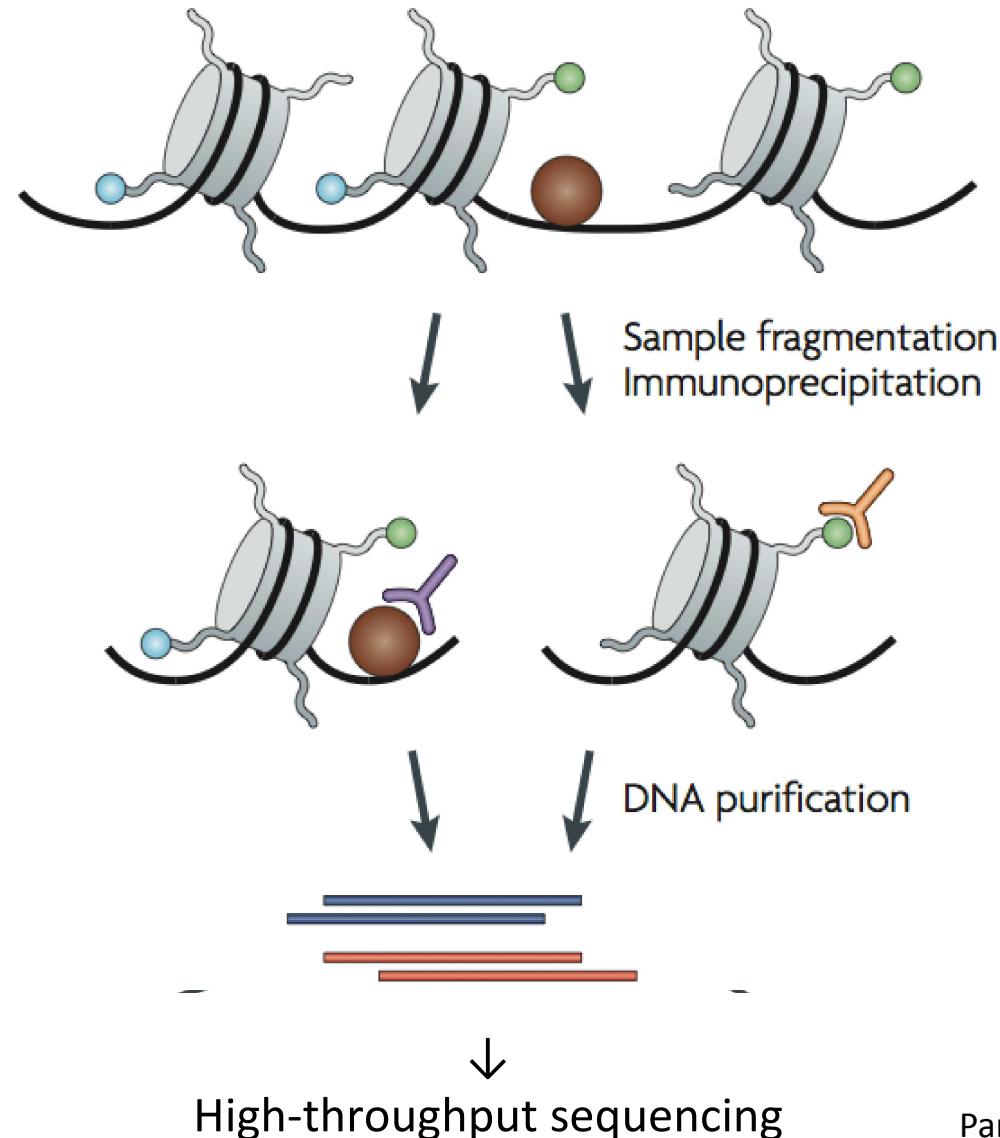
概要: ChIP-seq の統合的データベース



背景

背景: ChIP-seq とは

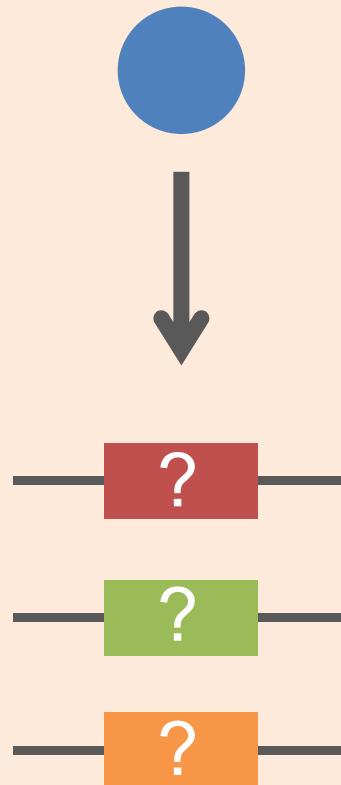
ChIP-seq = Chromatin ImmunoPrecipitation with Sequencing



Park PJ (2009) *Nat Rev Genet*

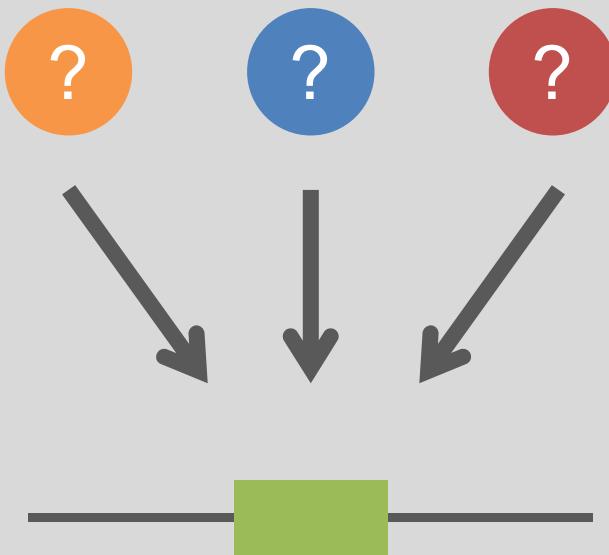
背景: 遺伝子制御ネットワークを知る

下流を知りたい



KO, KD 解析
個々の ChIP-seq データ

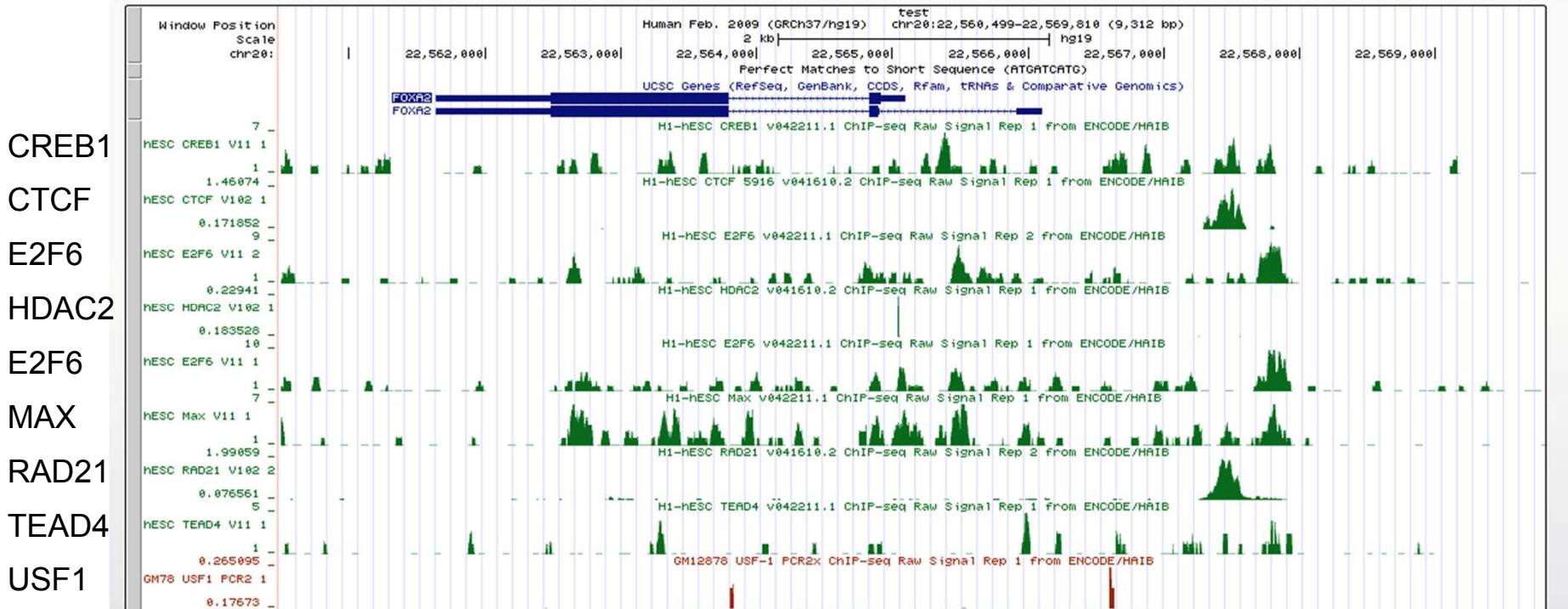
上流を知りたい



cis-element 解析, EMSA
多数の ChIP-seq データ

背景: ChIP-seq を見る。

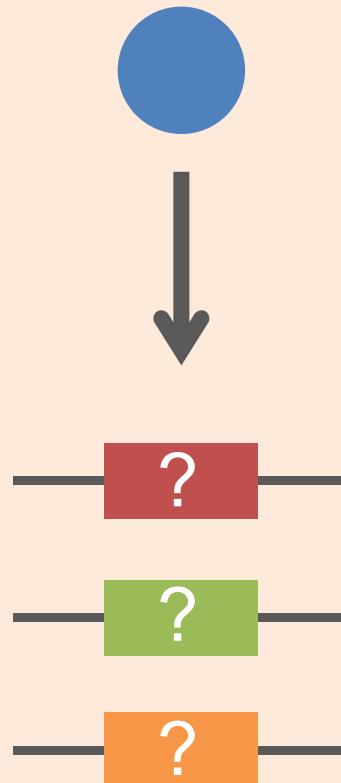
処理済みの公開データ (ENCODE データなど)



→ 個々の転写因子の下流遺伝子はわかる。
しかし、上流を知ることは難しい。

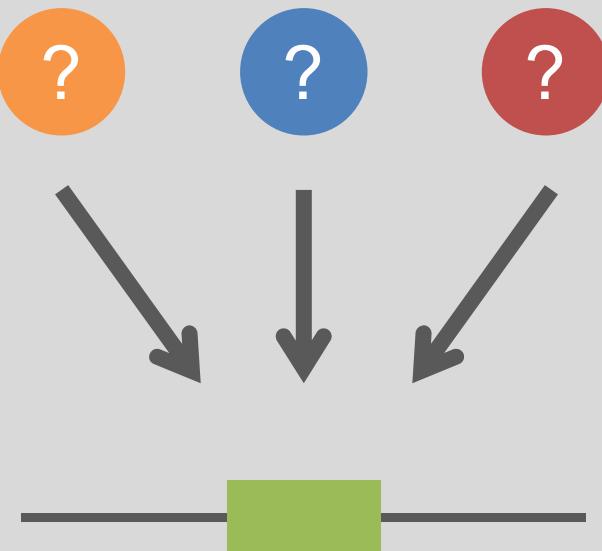
背景: 上流を知ることは難しい。

下流を知りたい



KO, KD 解析
個々の ChIP-seq データ

上流を知りたい

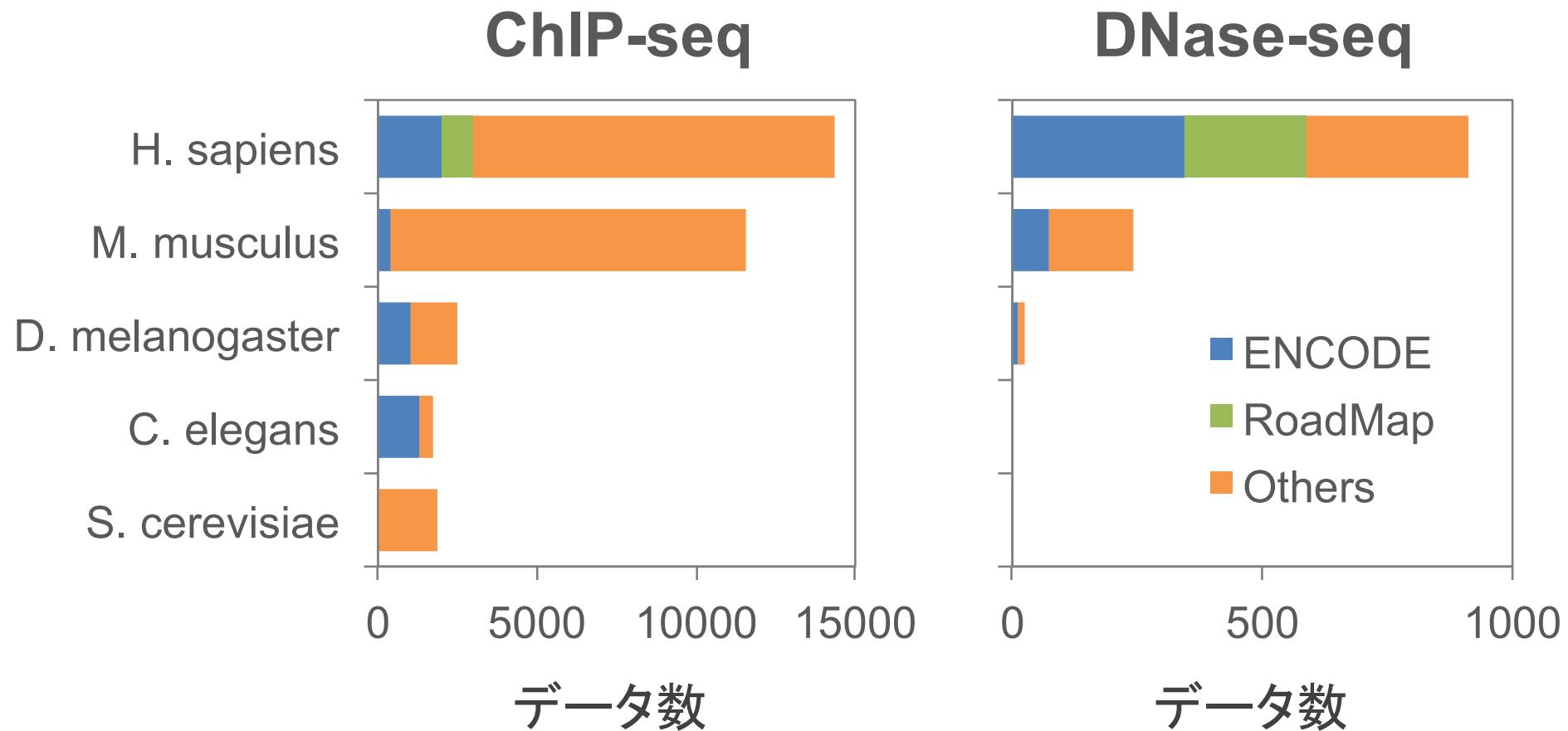


cis-element 解析, EMSA
多数の ChIP-seq データ

準備狀況

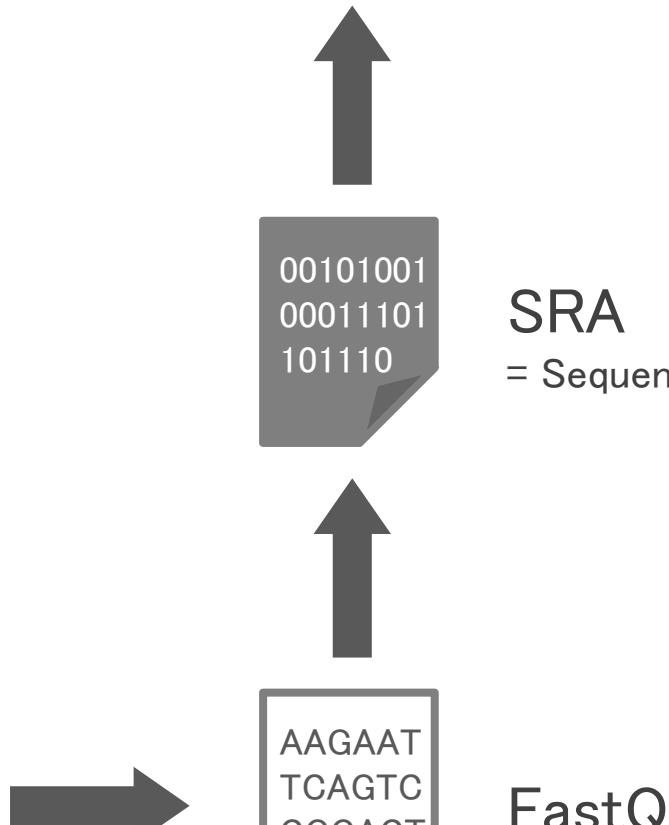
準備状況: 全ての ChIP-seq SRA を統合する。

(1) Data source



準備状況: 全ての ChIP-seq SRA を統合する。

(2) Primary data = SRA

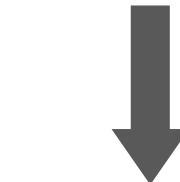


SRA
= Sequenced Read Archive

FastQ

準備状況: 全ての ChIP-seq SRA を統合する。

(2) Primary data = SRA



00101001
00011101
101110

SRA
= Sequenced Read Archive



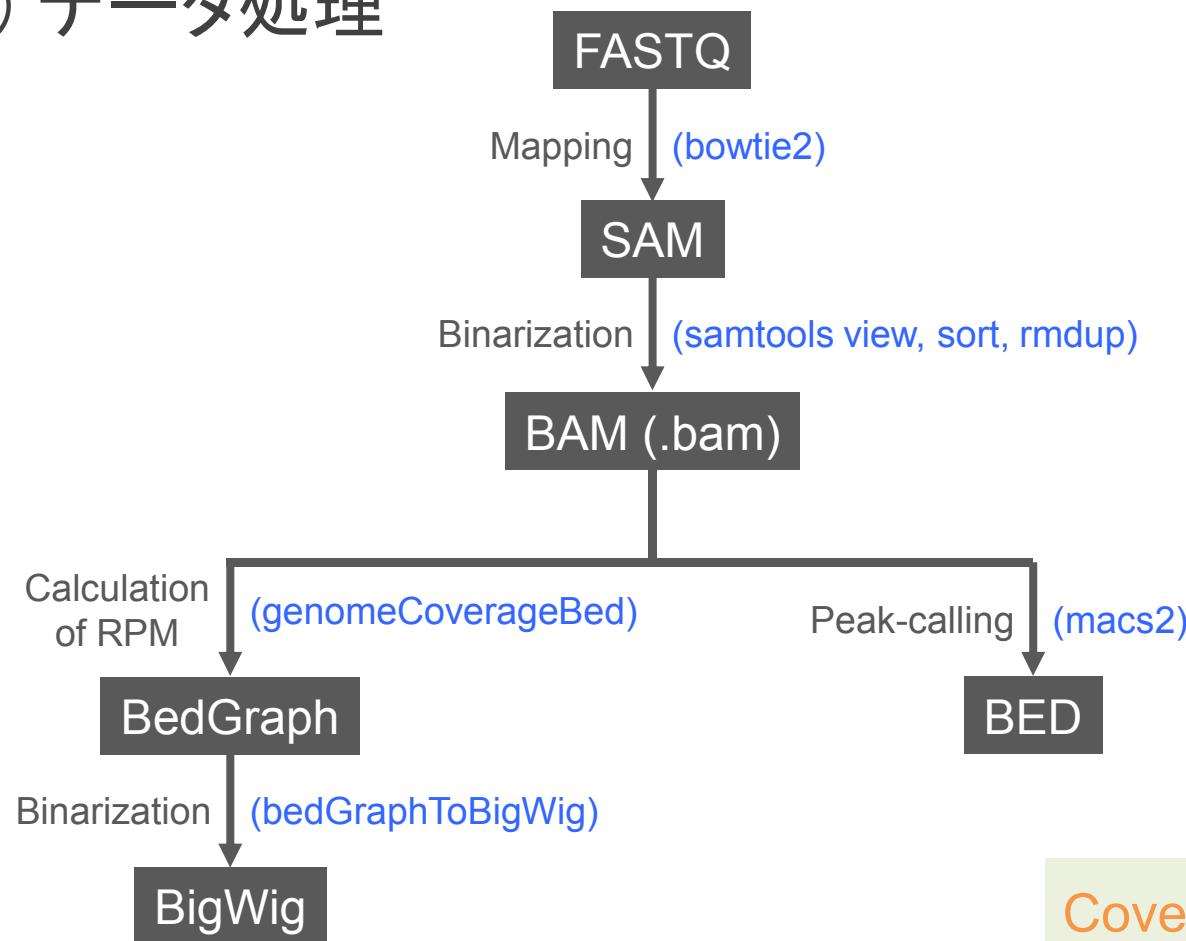
AAGAAT
TCAGTC
CGGACT
GAAT

FastQ

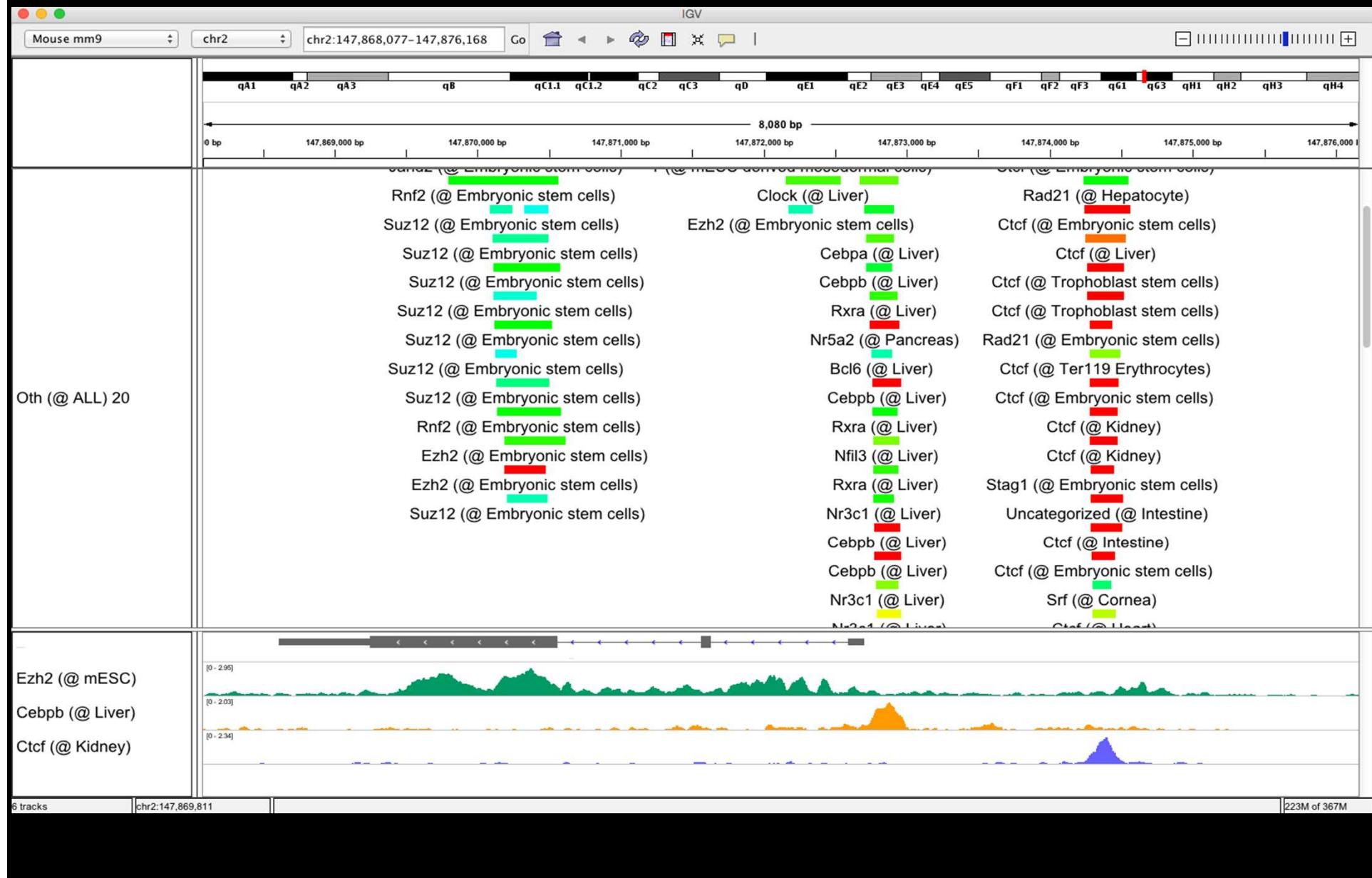


準備状況: 全ての ChIP-seq SRA を統合する。

(3) データ処理



準備状況: 全ての ChIP-seq SRA を統合する。



準備状況: 全ての ChIP-seq SRA を統合する。

(4) 表記ゆれの統一

抗原名

OCT4
Oct4
Oct4(N-19)(sc-8628), Santa Cruz
anti-Oct4 (Abcam, ab19857)
POU5F1_(SC-9081)

細胞名

ES cell culture
Embryonic Stem Cell
Embryonic Stem Cells (ESCs)
Embryonic Stem cells
Embryonic stem cell
Embryonic stem cells
Embyonic stem cells
V6.5 mES cells
embryonic stem cells
mouse embryonic stem cells

Pou5f1

Embryonic stem cells

實施項目

実施項目(1) 表記ゆれの統一



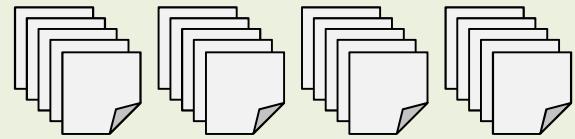
Google Refine
a power tool for working with messy data

old	new
h3k4me3	His@ H3k4me3
DMC1 (C-20) sc 8973	Oth@ Dmc1
PU1 sc-352	Oth@ Spi1
RPB2 sc-67318	Pol@ Polr2b
Abcam, ab8895	His@ H3K4me1
H4K8ac (Millipore 07-328)	His@ H4K8ac
H3K36me3 ab9050	His@ H3K36me3
BMAL1	Oth@ Arntl
H3K27ac (Abcam ab4729)	His@ H3K27ac
Ab3261 Abcam	His@ H3K79me1

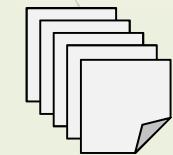
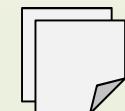
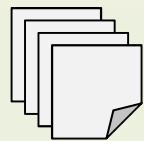
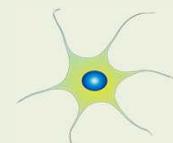
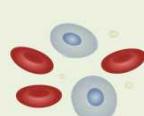
現在までに全体の7割まで終了。あと3割を済ませる。

実施項目(2) 統合データの作成・公開

All data



↓ 仕分け



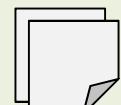
Histone



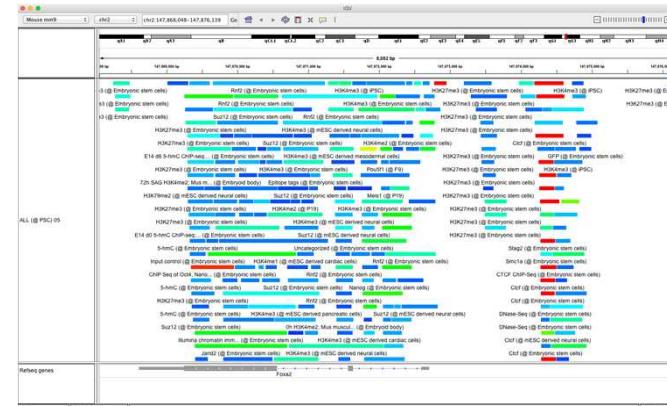
TFs



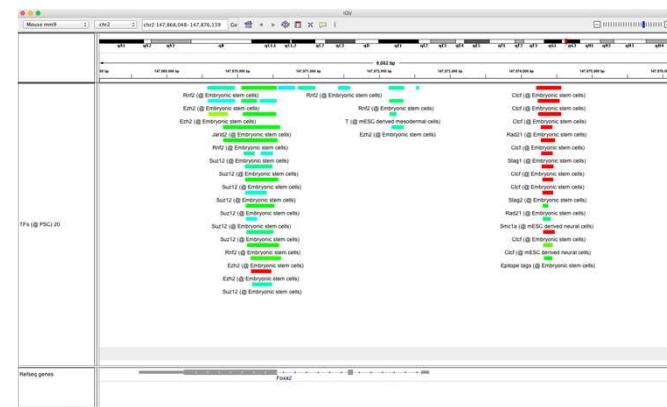
DNase



All data



TFs in ES cells



Human Mouse Fly Nematode Yeast

Antigen type:

- All
- DNase-seq
- Histone
- Input control
- RNA polymerase
- TFs and others

Cell type:

- All
- Adipocytes
- Blood
- Breast
- :
- Pluripotent SC

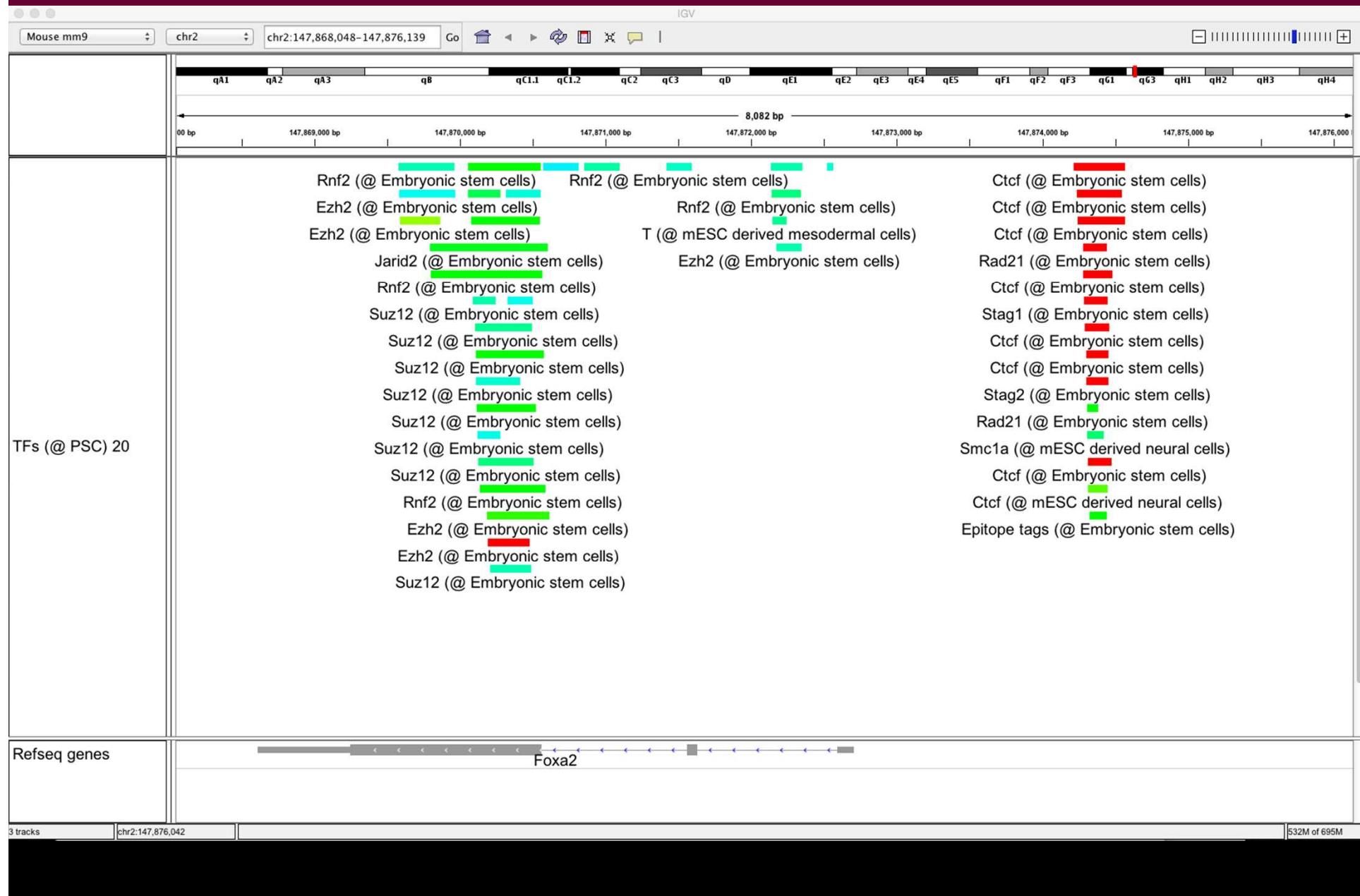
▼

- PML
- POU2F1
- POU2F2
- POU5F1
- PPARA
- PPARG

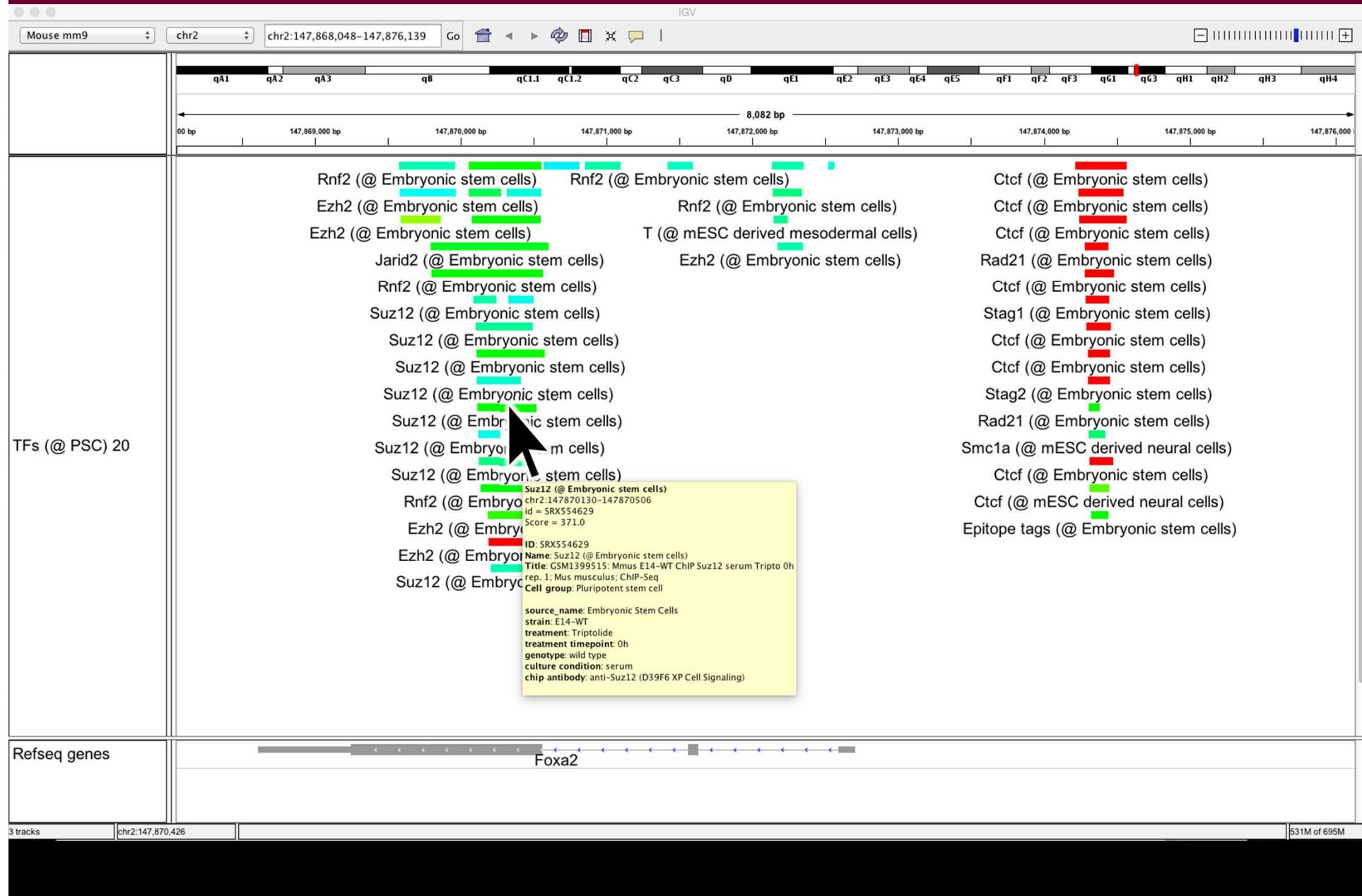
Show peaks

Collaboration with Tazro Ohta (DBCLS)

実施項目(2) 統合データの作成・公開



実施項目(2) 統合データの作成・公開



Accession: SRX554629

Title: E14-WT ChIP Suz12 serum Tripto 0h rep. 1; *Mus musculus*; ChIP-Seq

Cell class: Pluripotent stem cells

Cell subclass: Embryonic stem cells

Antigen class: Transcription factors

Antigen subclass: Suz12

Organism: *Mus musculus*

strain: E14-WT

treatment: Triptolide

genotype: wild type

culture condition: serum

chip antibody: anti-Suz12 (D39F6 XP Cell Signaling)



Accession: SRX554629

[BigWig on IGV](#)



[DBCLS SRA](#)

Title: E14-WT ChIP Suz12 serum Tripto 0h rep. 1; *Mus musculus*; ChIP-Seq

Cell class: Pluripotent stem cells

Cell subclass: Embryonic stem cells



Antigen class: Transcription factors

Antigen subclass: Suz12



Organism: *Mus musculus*

strain: E14-WT

treatment: Triptolide

genotype: wild type

culture condition: serum

chip antibody: anti-Suz12 (D39F6 XP Cell Signaling)

Accession: SRX554629

[BigWig on IGV](#)  NCBI [DBCLS SRA](#)

Title: E14-WT ChIP Suz12 serum Tripto 0h rep. 1; Mus musculus; ChIP-Seq

Cell class: Pluripotent stem cells

Cell subclass: Embryonic stem cells

Antigen class: Transcription factors

Antigen subclass: Suz12

統合化推進プログラム
のデータベース

Organism: Mus musculus

strain: E14-WT

treatment: Triptolide

genotype: wild type

culture condition: serum

chip antibody: anti-Suz12 (D39F6 XP Cell Signaling)

Accession: SRX554629

BigWig on IGV ENA DBCLS SRA

Strain: E14 WT ChIP Suz12 serum Trypto Oh rep. 1; Mus musculus; ChIP-Seq

Cell class: Pluripotent stem cells

Cell subclass: Embryonic stem cells

<http://www2.brc.riken.jp/lab/cell/list.cgi?sk=Embryonic+stem+cells>

細胞の詳細情報や入手先

Antigen subclass: Suz12

<http://omicspace.riken.jp/PosMed/search?acti onType=searchexec&objectSet=gene&specie s=Mm&keyword=Suz12>

<http://pdbj.org/mine/search?query=Suz12>

Strain: E14 WT

treatment: Tryptoid

genotype: wild type

culture condition: serum

chip antibody: anti-Suz12 (D39F6 XP Cell Signaling)

抗原の詳細情報

Collaboration with Tazro Ohta (DBCLS)



Accession: SRX554629

BigWig on IGV



DBCLS SRA

Title: E14-WT ChIP Suz12 serum Tripto 0h rep. 1; Mus musculus; ChIP-Seq

Cell class: Pluripotent stem cells

Cell subclass: Embryonic stem cells



Antigen class: Transcription factors

Antigen subclass: Suz12



Organism: Mus musculus

strain: E14-WT

treatment: Triptolide

genotype: wild type

culture condition: serum

chip antibody: anti-Suz12 (D39F6 XP Cell Signaling)

Accession: SRX554629

BigWig on IGV NCBI DBCLS SRA

Mile: E14 WT ChIP Suz12 Serum TripN Oh rep. 1; Mus musculus, ChIP-Seq

BigWig の閲覧 (IGV)

リードクオリティ・引用論文

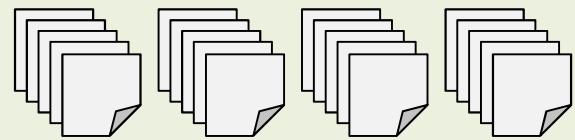
Organism: Mus musculus
Strain: E14
Treatment: ChIP-on-chip
Genotype: wild type
Culture condition: serum
ChIP antibody: anti-Suz12 (D39F6 XP Cell Signaling)

サンプル・実験情報

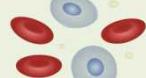
Collaboration with Tazro Ohta (DBCLS)

実施項目 (3) 統合解析

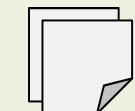
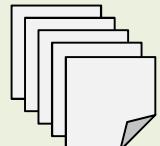
All data



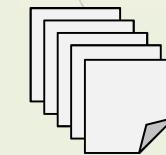
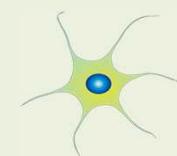
↓ 仕分け



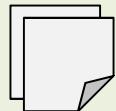
Histone



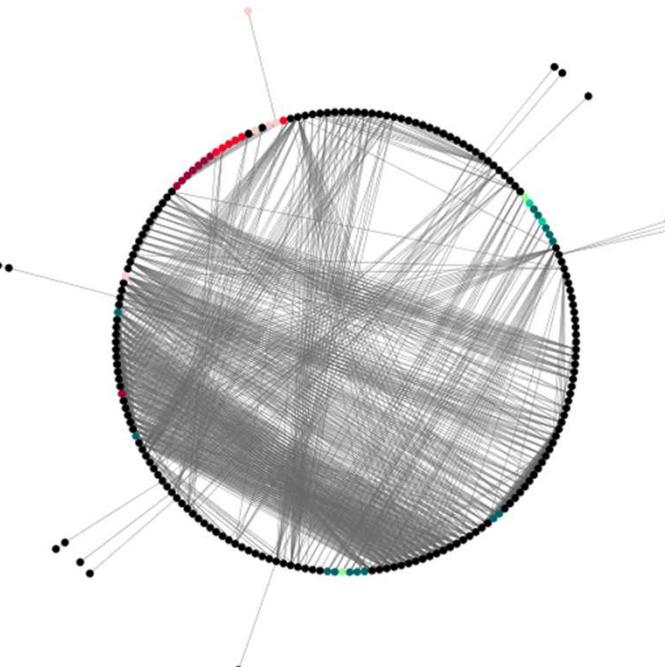
TFs



DNase



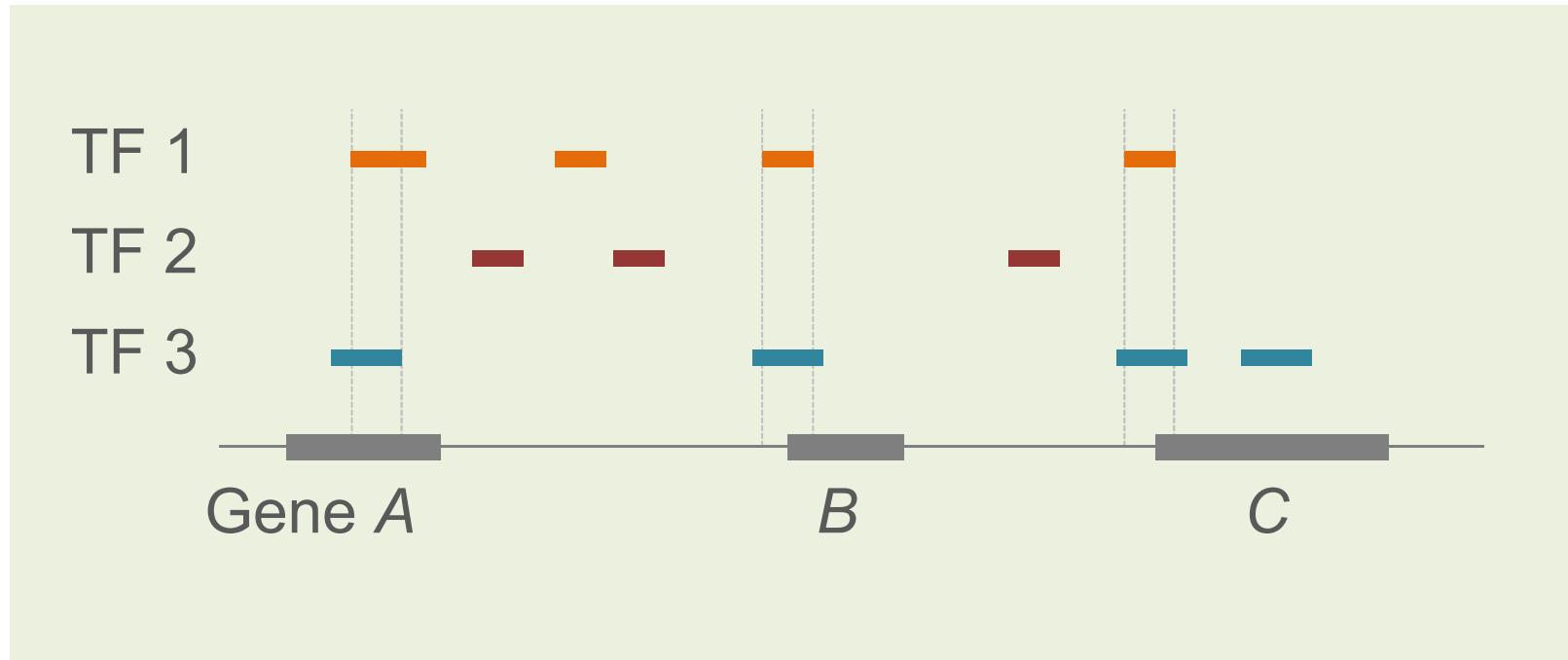
解析



実施項目 (3) 統合解析

共局在性解析

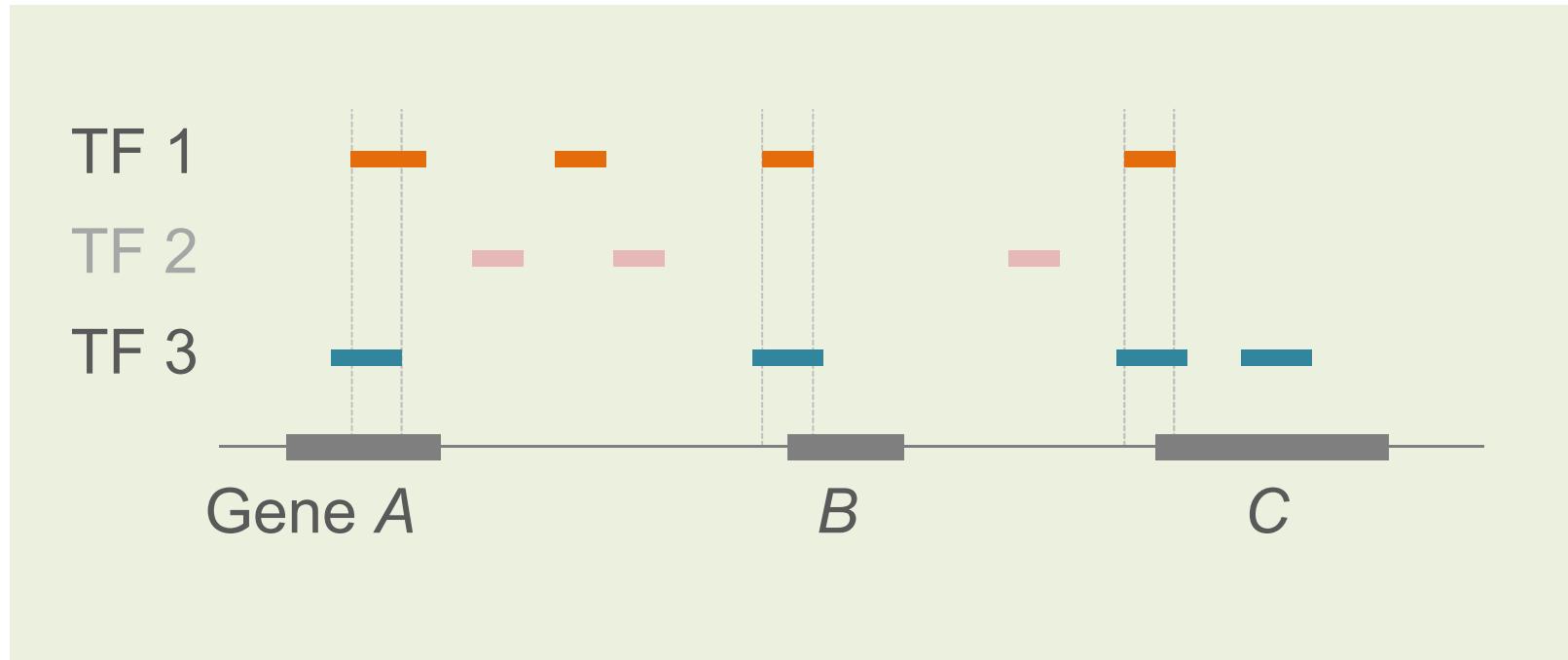
ツール: CoLo (東大・油谷研・仲木氏)



実施項目 (3) 統合解析

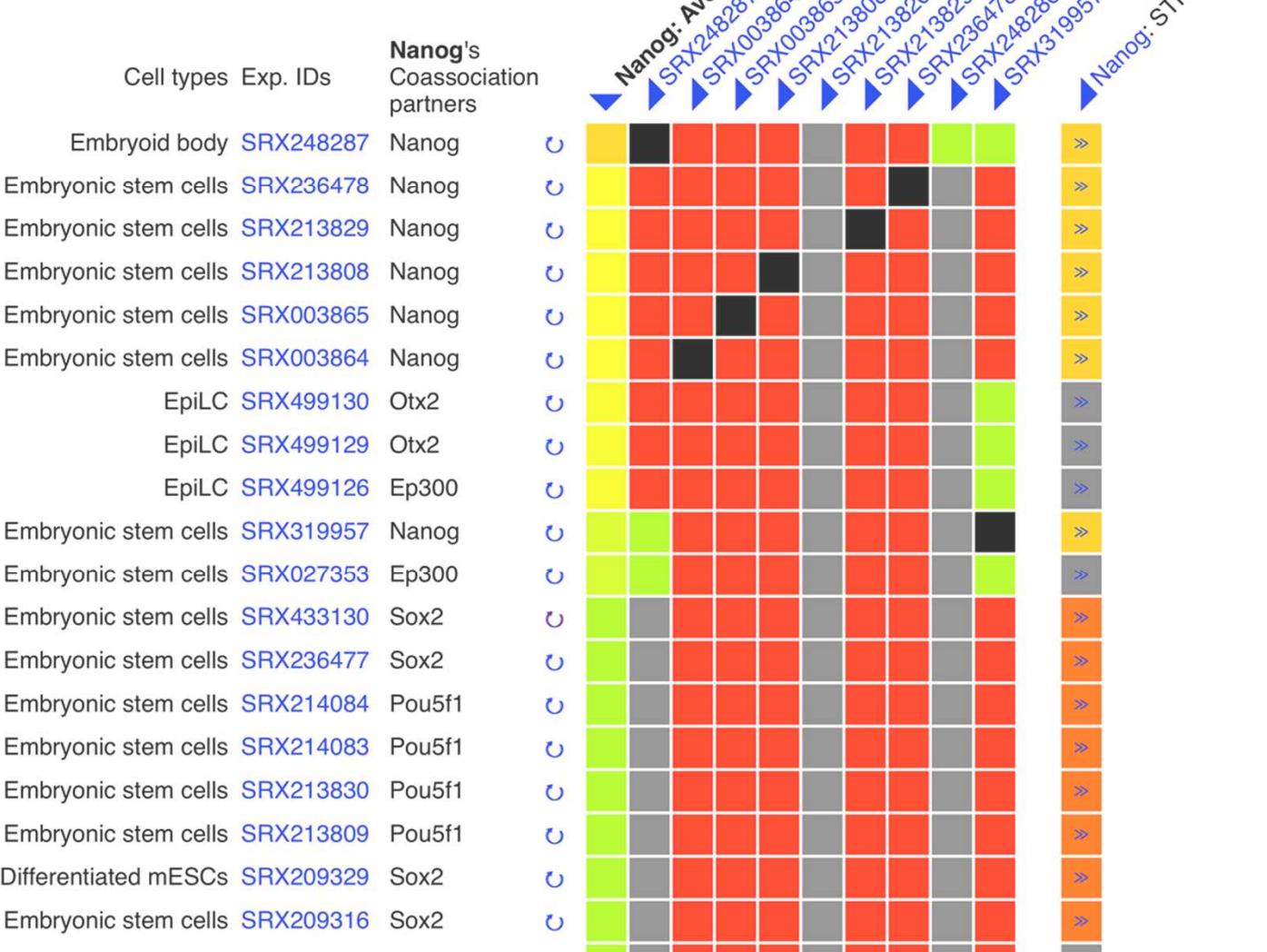
共局在性解析

ツール: CoLo (東大・油谷研・仲木氏)



實施項目(3) 統合解析

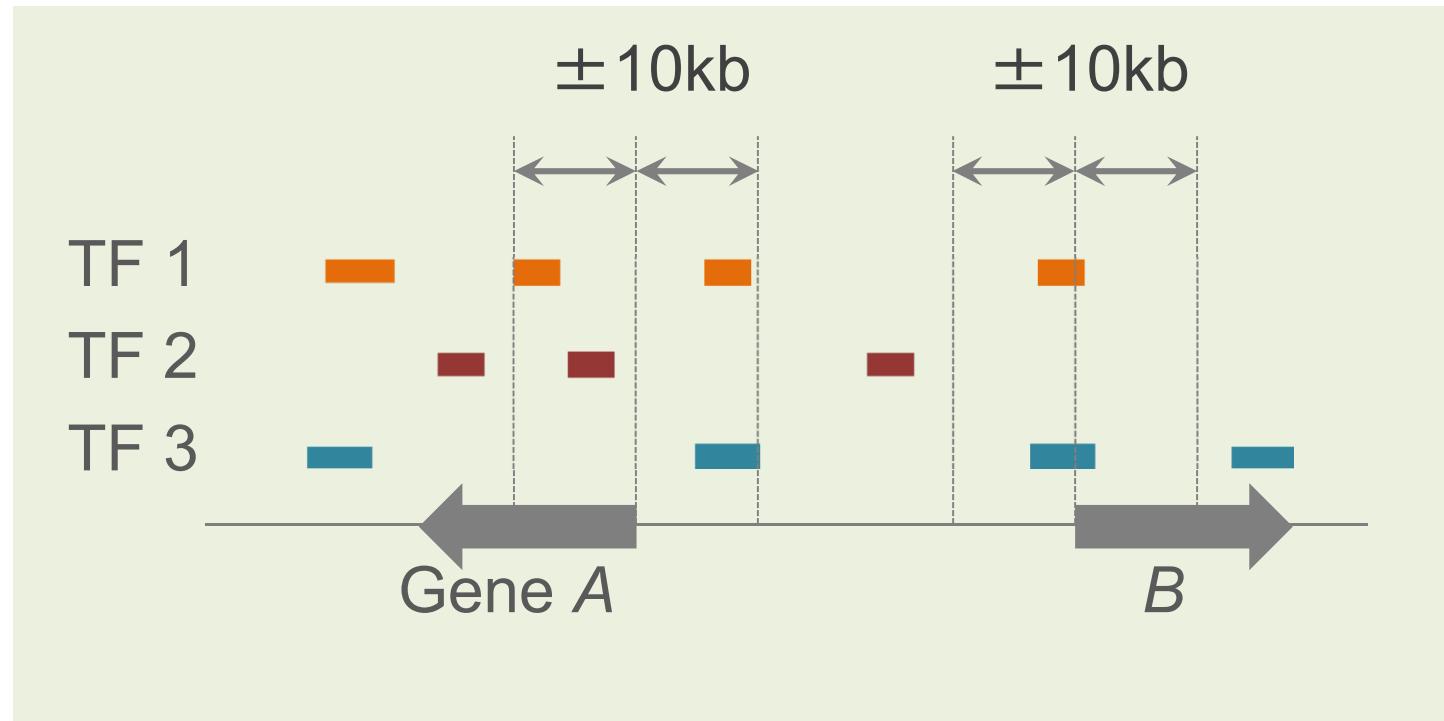
共局在性解析



Lo Hi

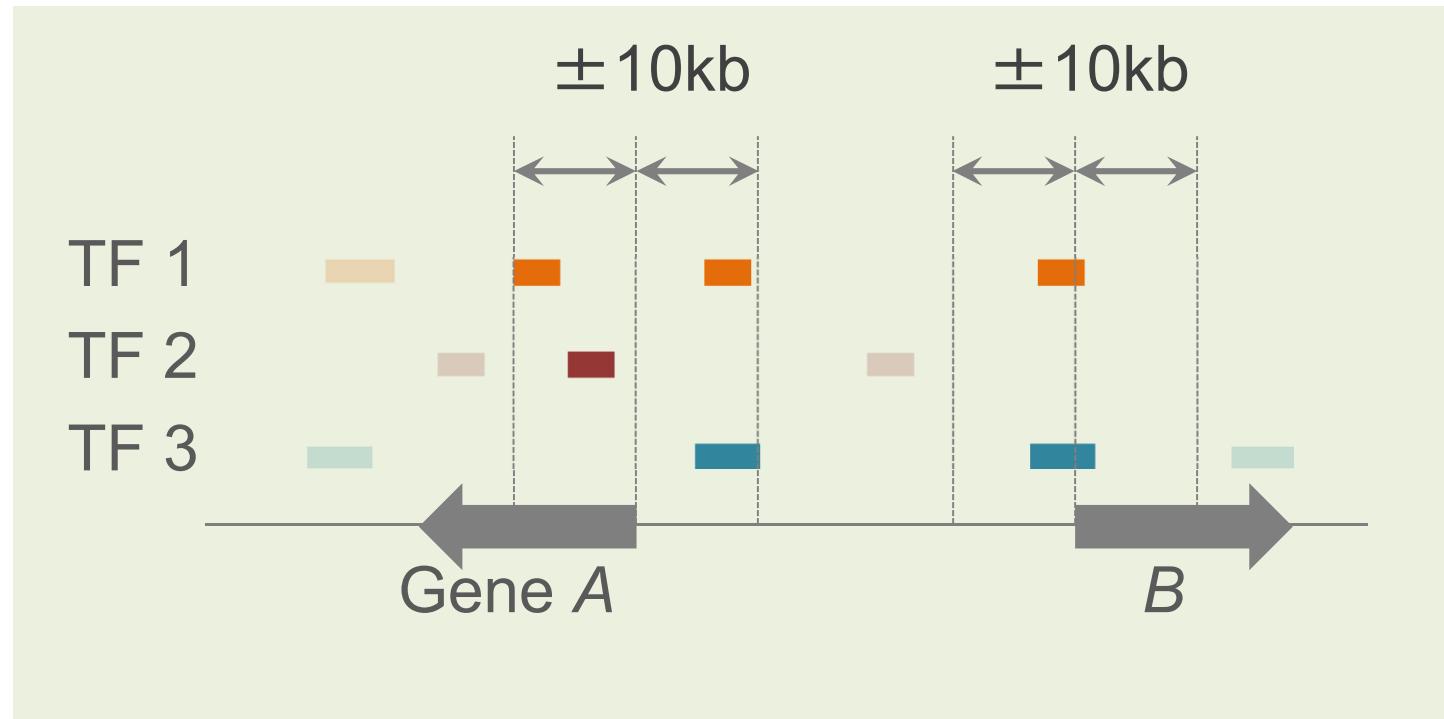
実施項目 (3) 統合解析

標的遺伝子解析
ツール: BedTools



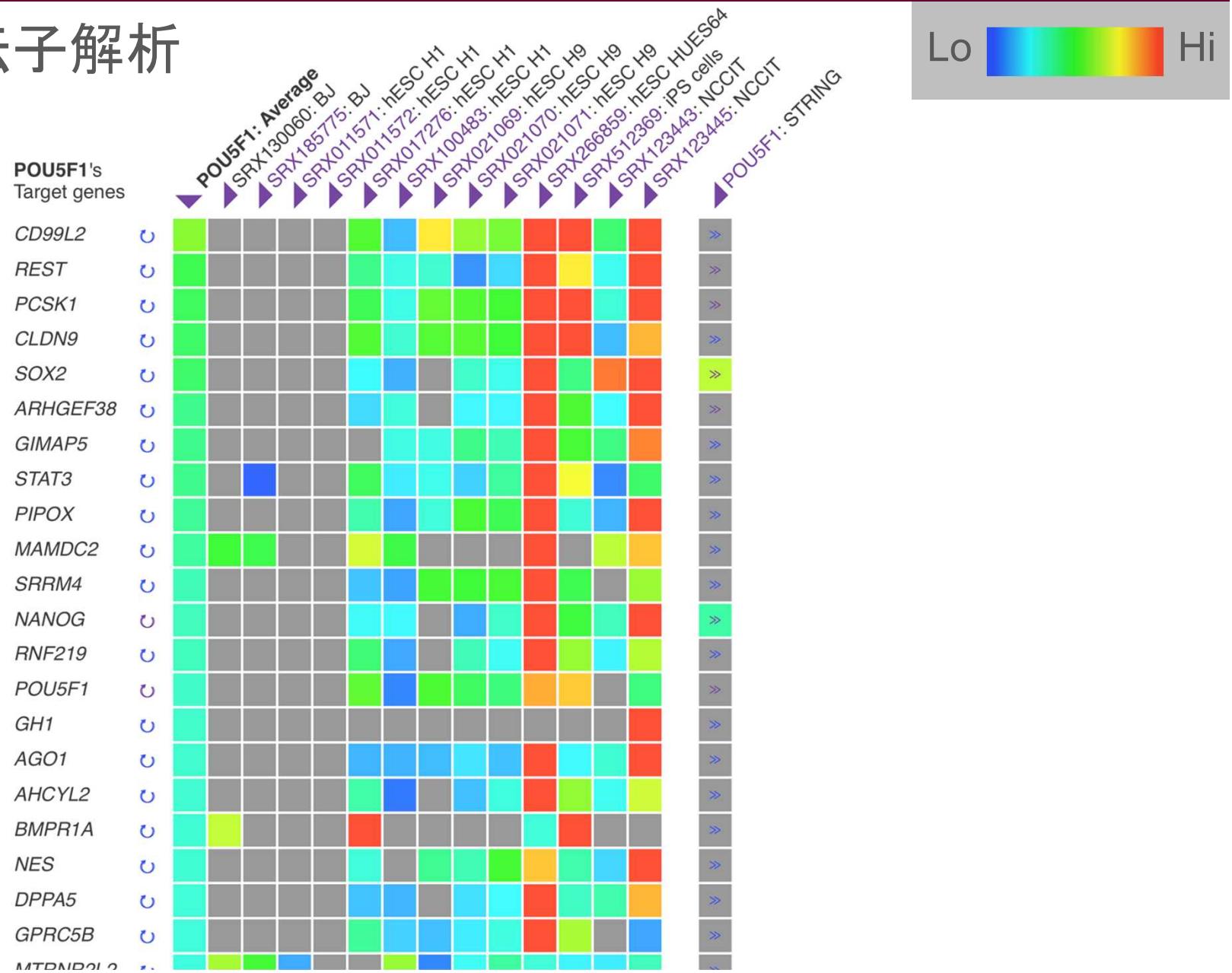
実施項目 (3) 統合解析

標的遺伝子解析
ツール: BedTools



實施項目 (3) 統合解析

標的遺伝子解析



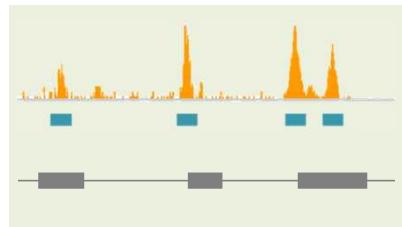
まとめ

SRA

AAGAAT
TCAGTC
CGGACT
GAAT

↓
Mapping
Peak-call

個別データ

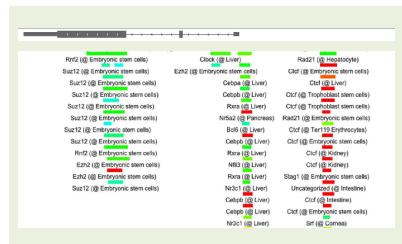


目的

個別データの閲覧

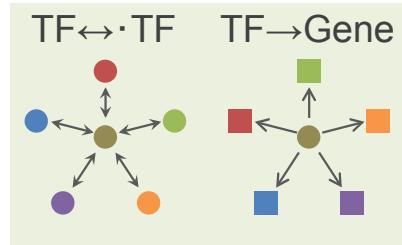
↓
表記ゆれ統一
データ統合

統合データ



↓
統合解析

統合的
解析データ



↑
転写因子の共局在性の予測
下流遺伝子の予測

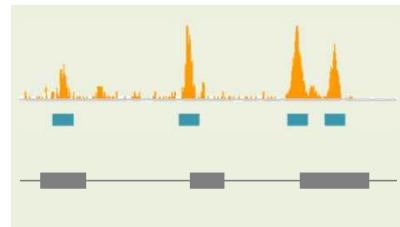
まとめ：今後の方針

SRA

AAGAAT
TCAGTC
CGGACT
GAAT

Mapping
Peak-call

個別データ
(~ 4TB)



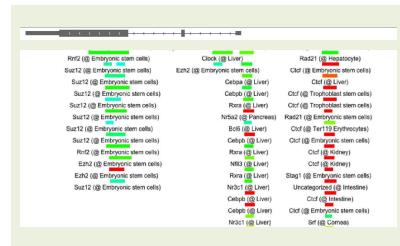
方針

QC データ
データ更新(毎月)

統合データ
(~ 6TB)

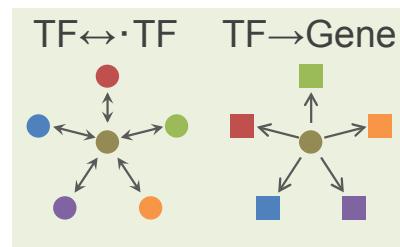
表記ゆれ統一
データ統合

残り3割を仕上げる。



共局在性
標的遺伝子
(~ 1TB)

統合解析



Web UI の作成 (大田氏)
既存データベースとの連携

Web UI の作成

サーバ確保・公開・論文投稿

謝辞



院・医
発生再生医学分野
目野主税



ライフサイエンス技術基盤研究センター
塩井 剛



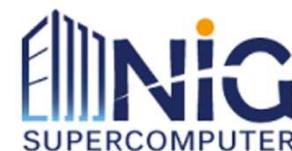
大田達郎



先端科学技術研究センター
仲木 竜



畠中秀樹



敬称略