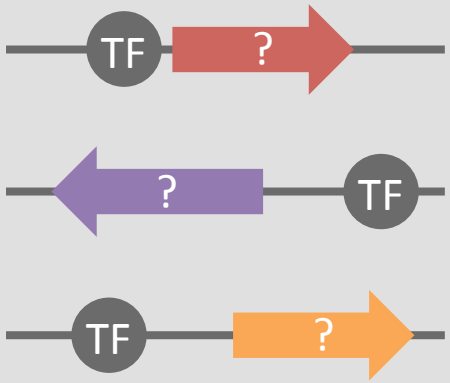
ChIP-seq データをまるわかりにする

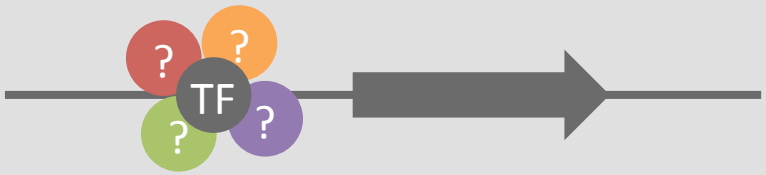
① どこに何が結合する？

② 標的遺伝子は？

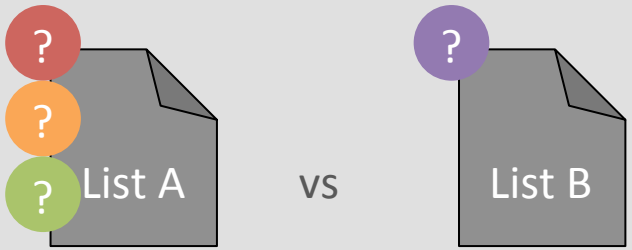


ChIP-Atlas

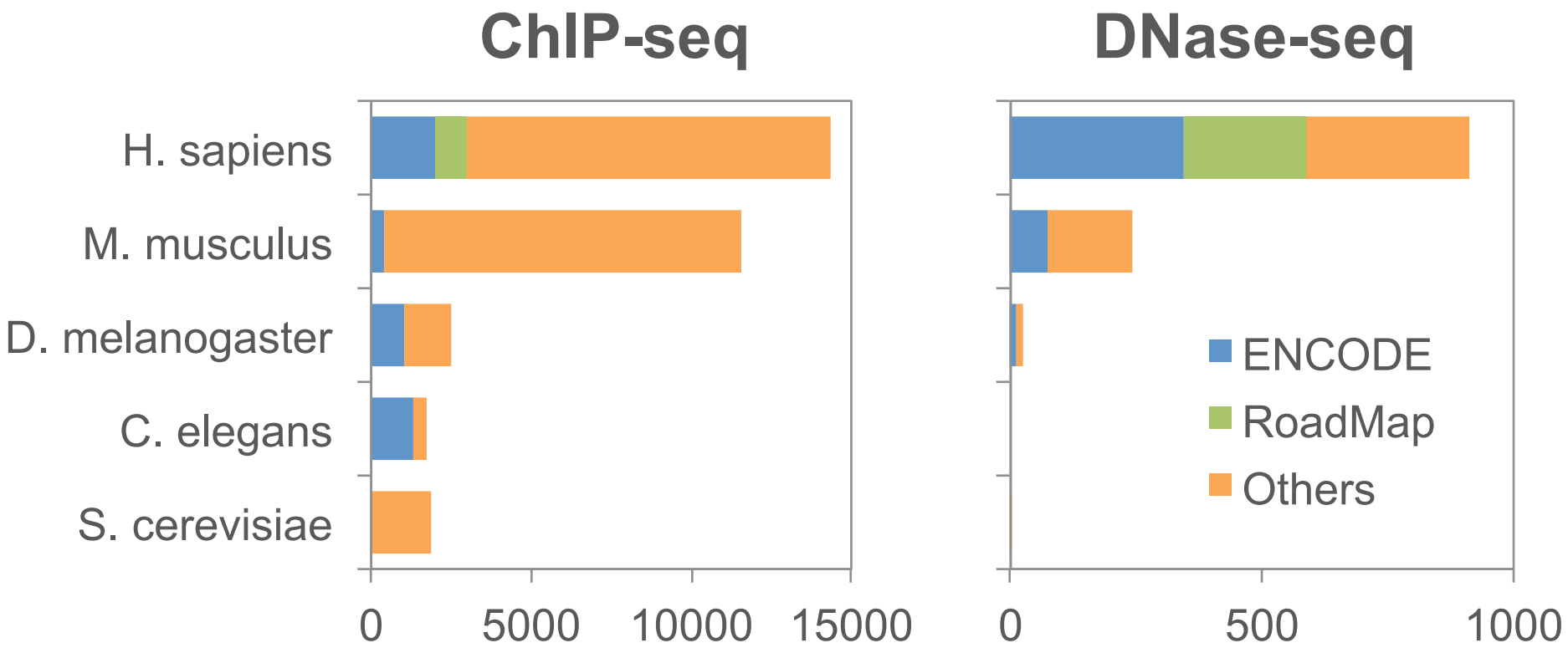
③ 共局在パートナーは？



④ ユーザデータの解析

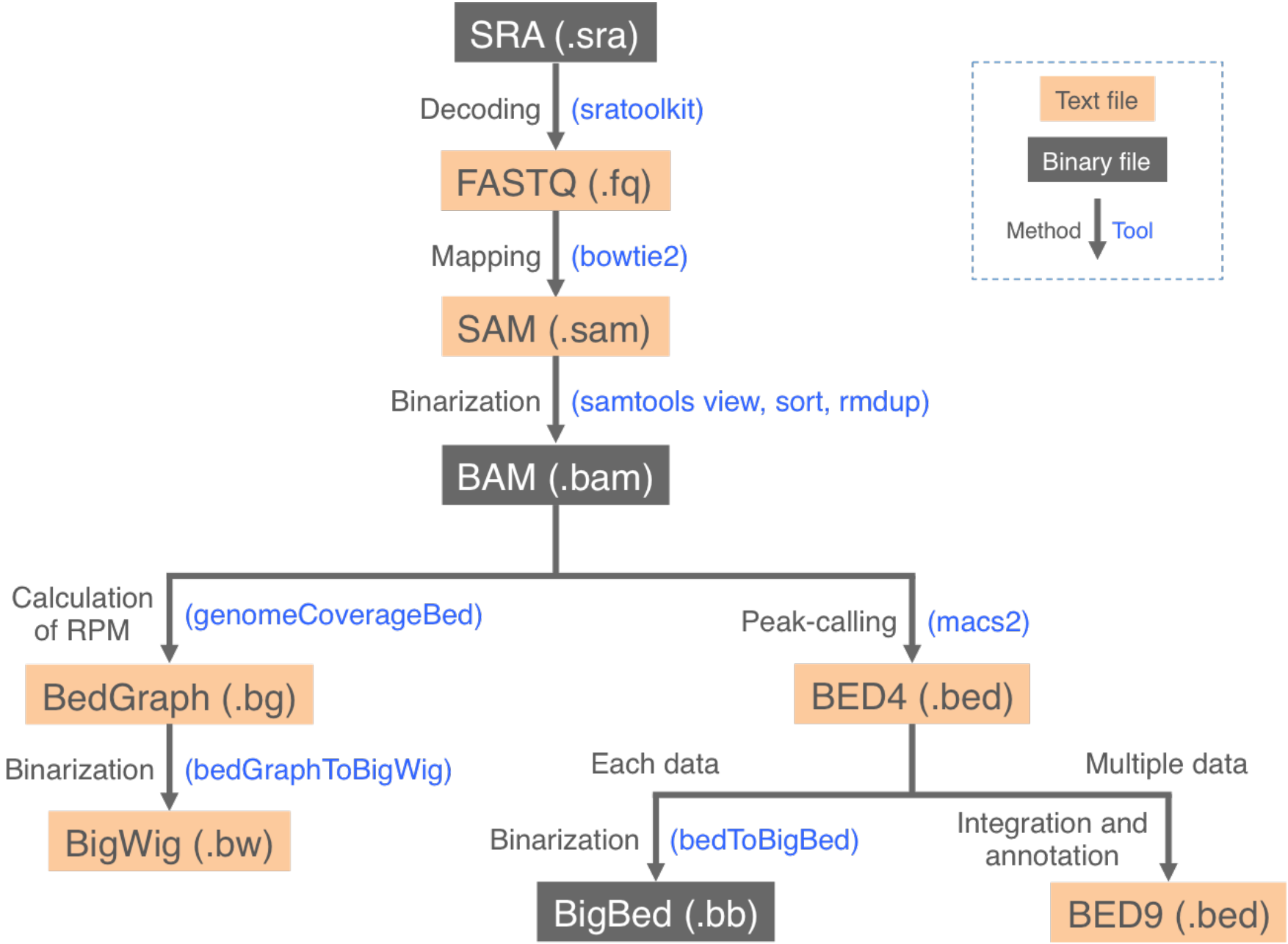


使用データ

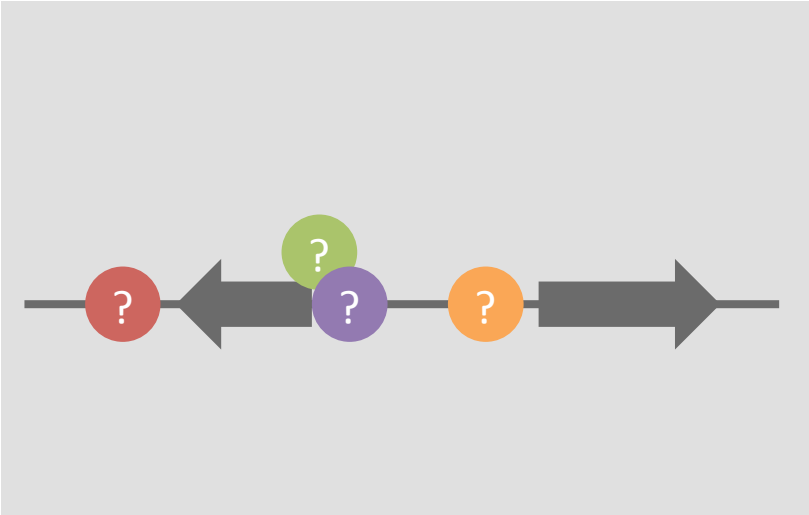


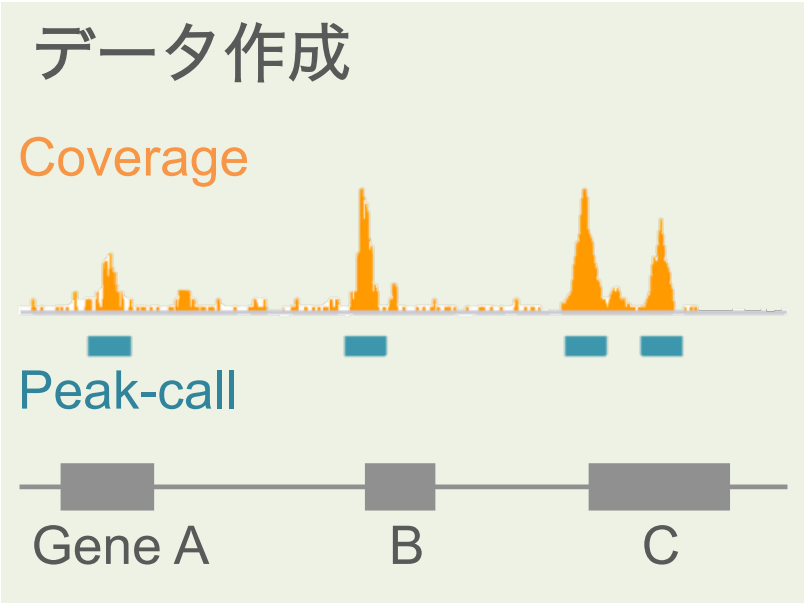
NCBI ENA DDBJ 2015 5月現在

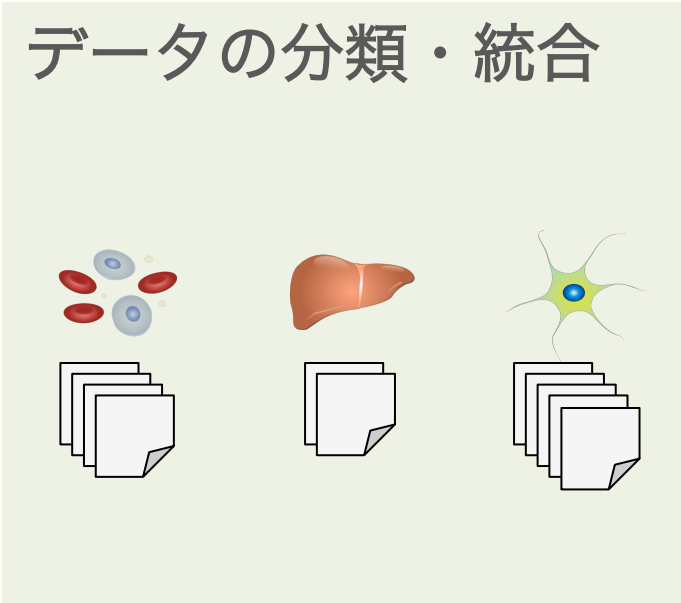
計算処理


SRA (.sra)
Decoding
(sratoolkit)
FASTQ (.fq)
Mapping
(bowtie2)
SAM (.sam)
Binarization
(samtools view, sort, rmdup)
BAM (.bam)
Calculation of RPM
(genomeCoverageBed)
BedGraph (.bg)
Binarization
(bedGraphToBigWig)
BigWig (.bw)
Peak-calling
(macs2)
BED4 (.bed)
Each data
Binarization
(bedToBigBed)
BigBed (.bb)
Multiple data
Integration and annotation
BED9 (.bed)
Text file
Binary file
Method
Tool

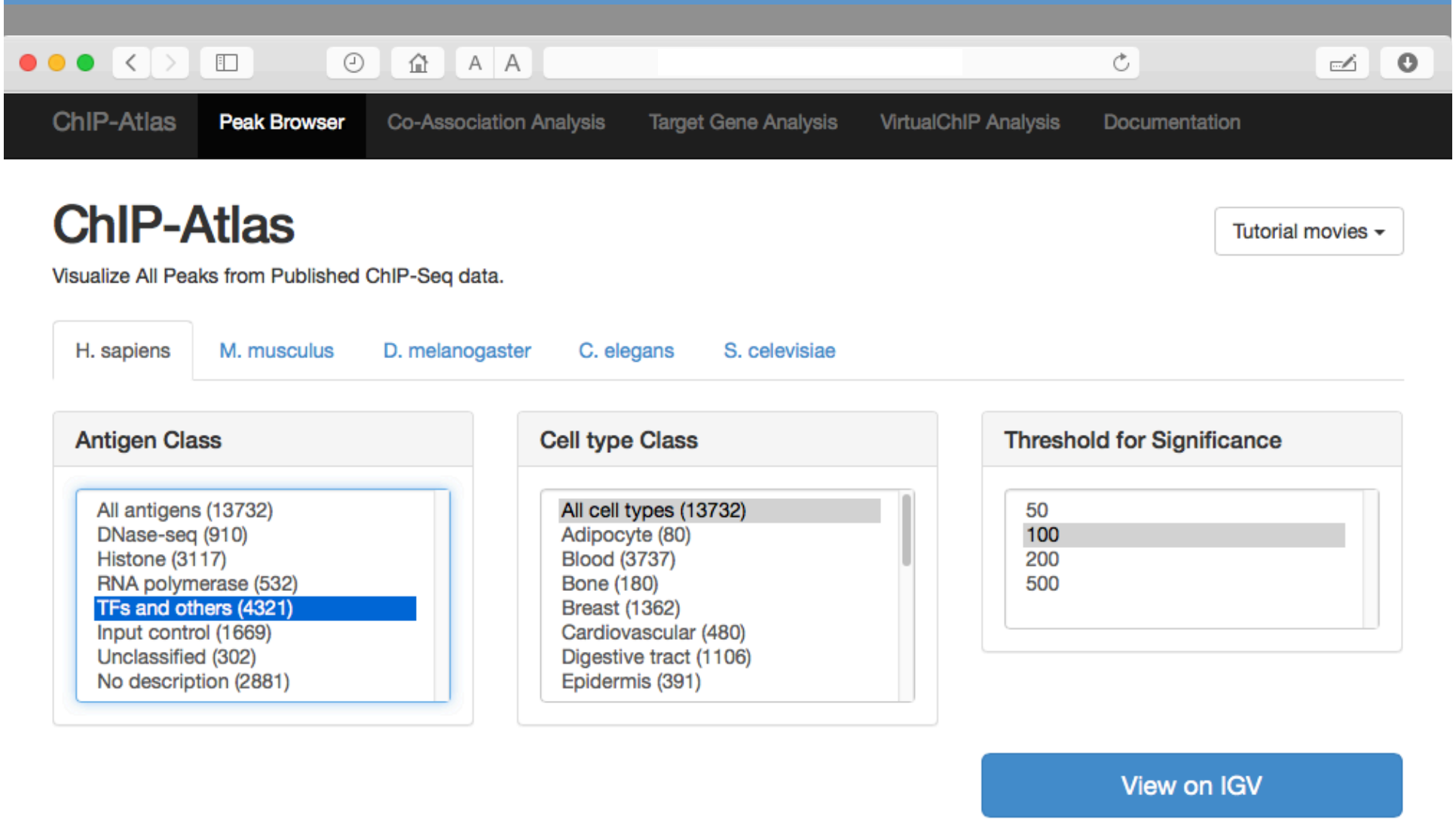
① どこに何が結合する？



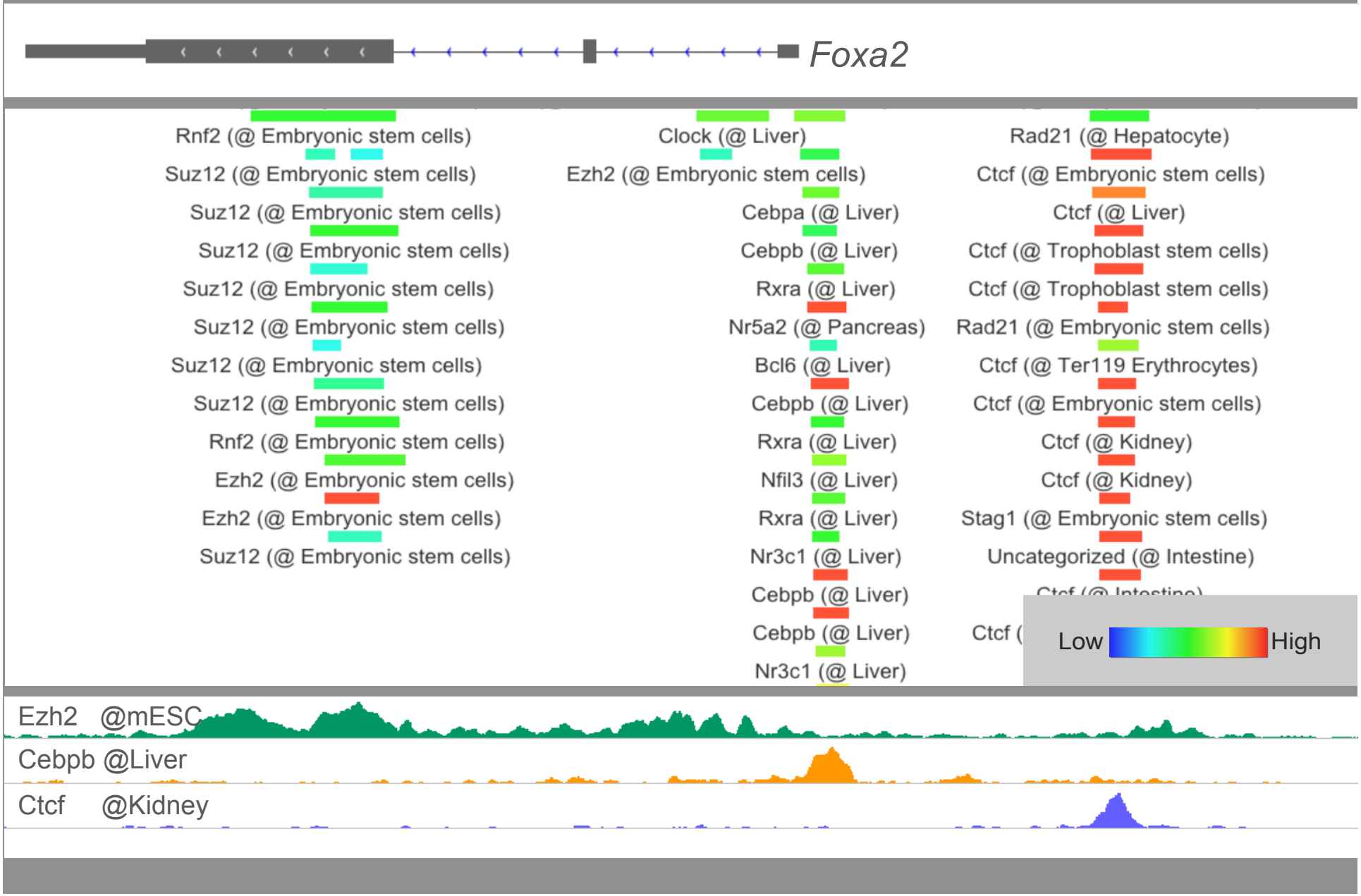




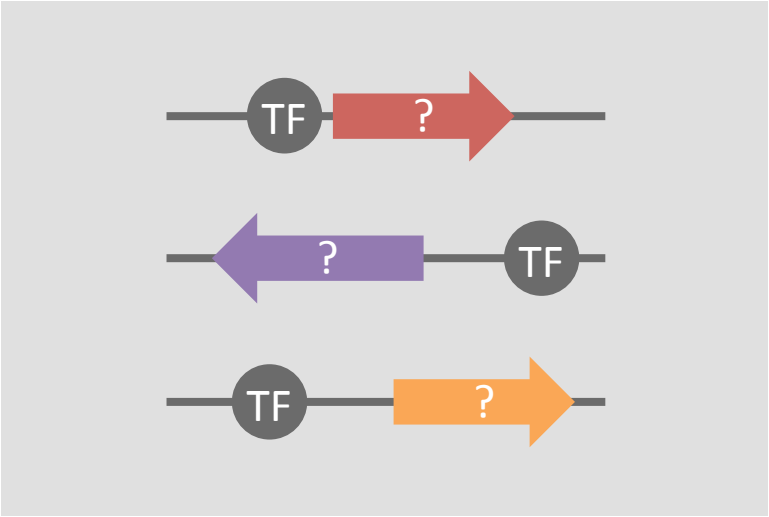
① どこに何が結合する？


ChIP-Atlas
Peak Browser
Co-Association Analysis
Target Gene Analysis
VirtualChIP Analysis
Documentation
ChIP-Atlas
Tutorial movies
Visualize All Peaks from Published ChIP-Seq data.
H. sapiens
M. musculus
D. melanogaster
C. elegans
S. celevisiae
Antigen Class
All antigens (13732)
DNase-seq (910)
Histone (3117)
RNA polymerase (532)
TFs and others (4321)
Input control (1669)
Unclassified (302)
No description (2881)
Cell type Class
All cell types (13732)
Adipocyte (80)
Blood (3737)
Bone (180)
Breast (1362)
Cardiovascular (480)
Digestive tract (1106)
Epidermis (391)
Threshold for Significance
50
100
200
500
View on IGV

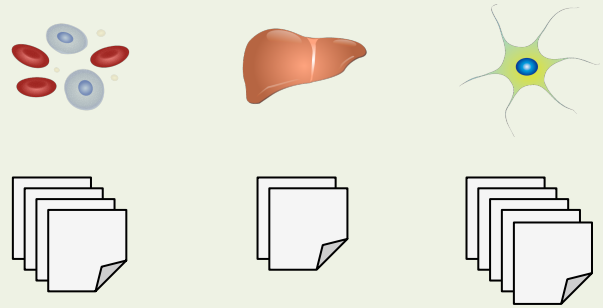
① どこに何が結合するか？



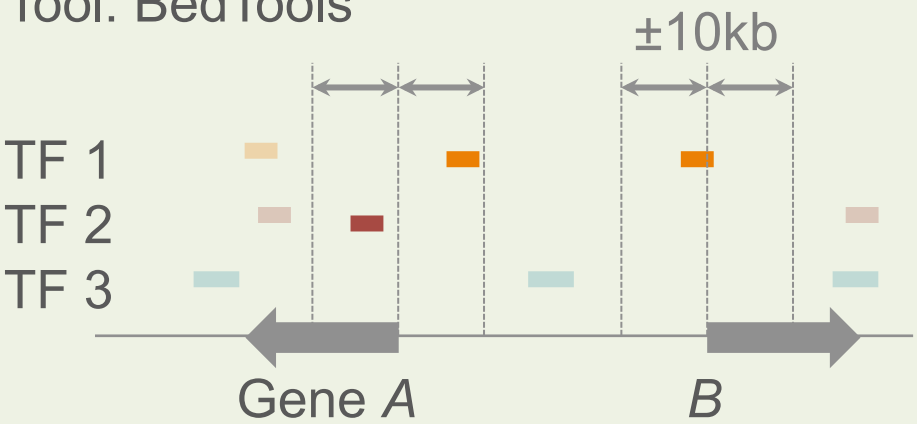
② 標的遺伝子は？



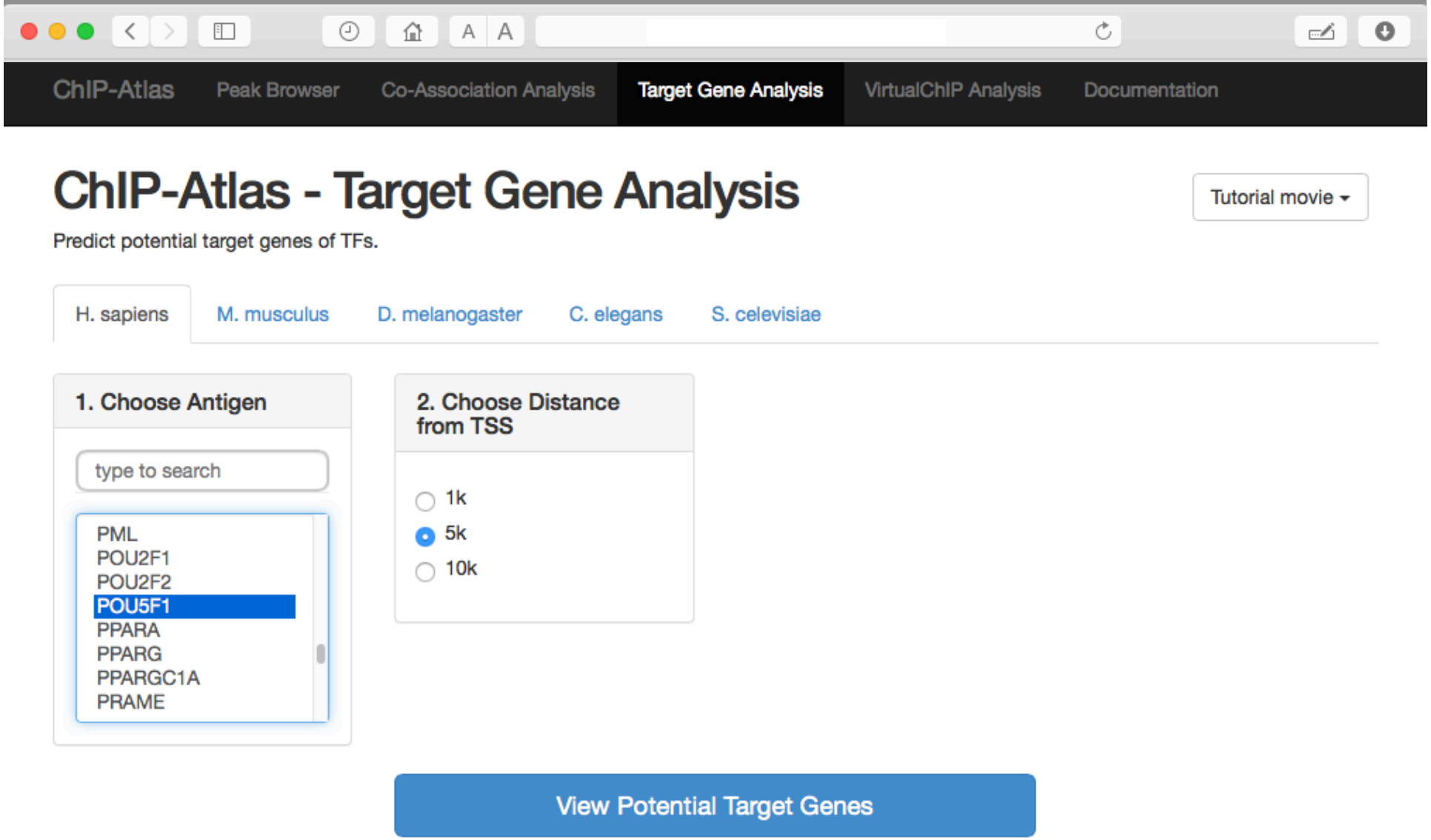
使用データ：
全 Peak-call データ



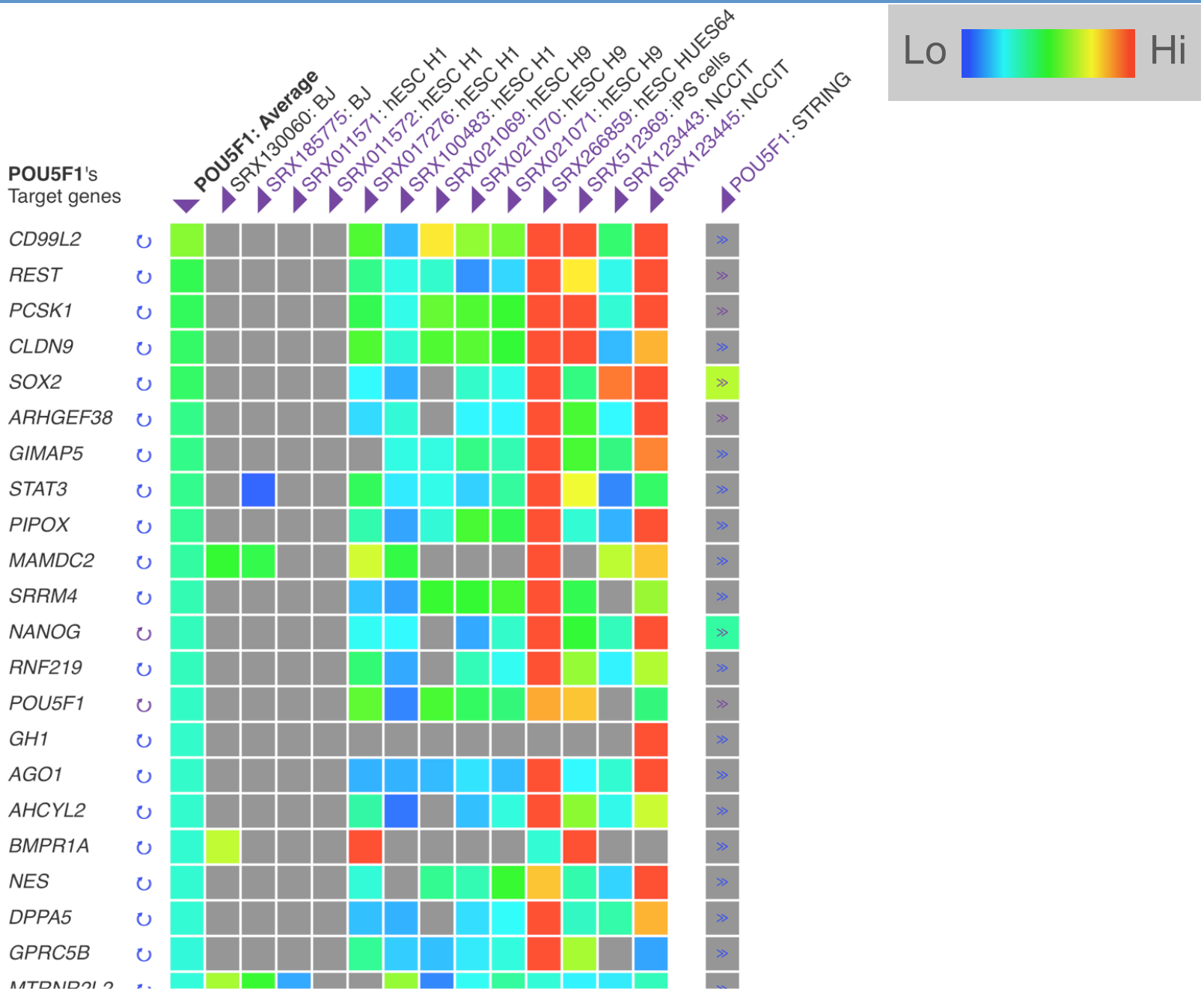
TSS 周辺の結合を探す
Tool: BedTools



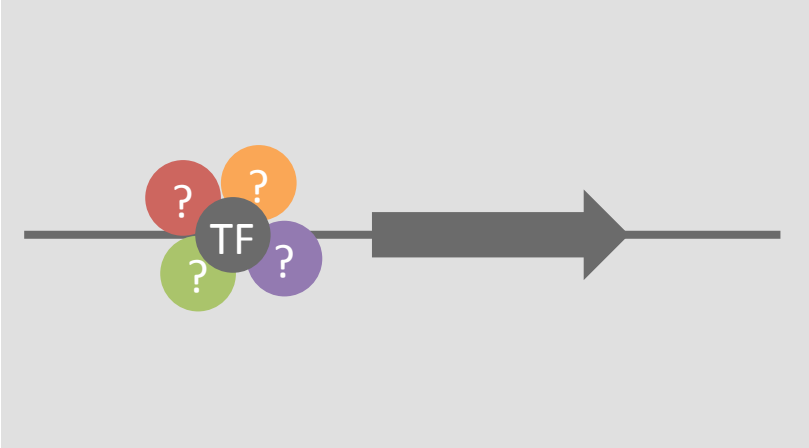
② 標的遺伝子は？


ChIP-Atlas
Peak Browser
Co-Association Analysis
Target Gene Analysis
VirtualChIP Analysis
Documentation
ChIP-Atlas - Target Gene Analysis
Tutorial movie
Predict potential target genes of TFs.
H. sapiens
M. musculus
D. melanogaster
C. elegans
S. celevisiae
1. Choose Antigen
type to search
PML
POU2F1
POU2F2
POU5F1
PPARA
PPARG
PPARGC1A
PRAME
2. Choose Distance from TSS
1k
5k
10k
View Potential Target Genes

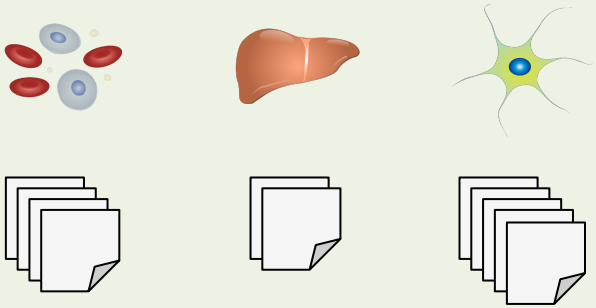
② POU5F1 の標的遺伝子は？



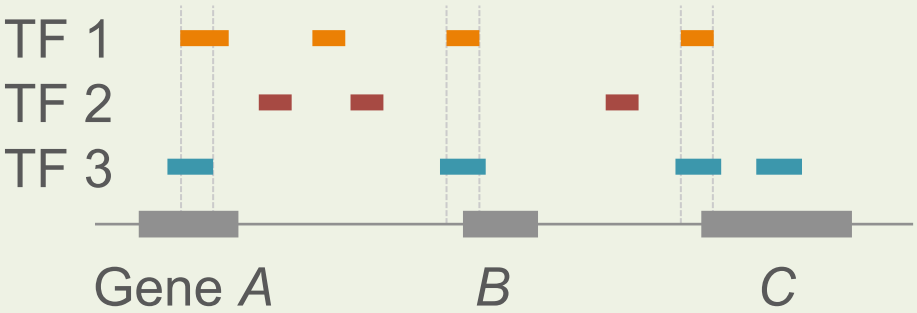
③ 共局在パートナーは？



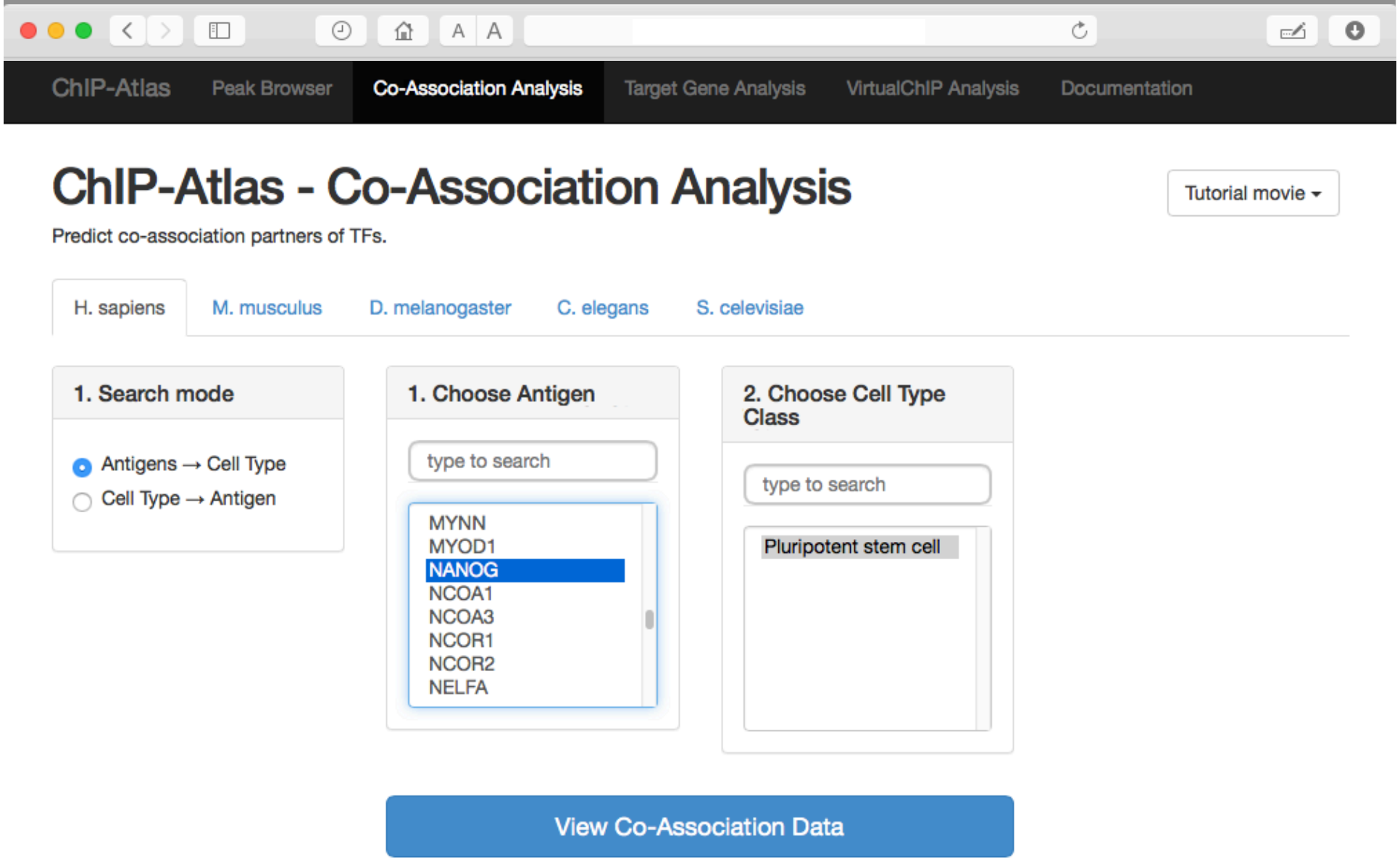
使用データ：
全 Peak-call データ



Similarity analysis
Tool: CoLo (東大・油谷研・仲木氏)

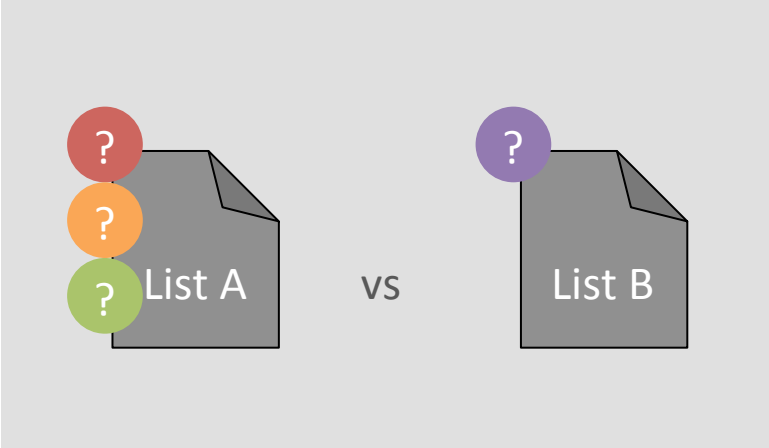


③ 共局在パートナーは？

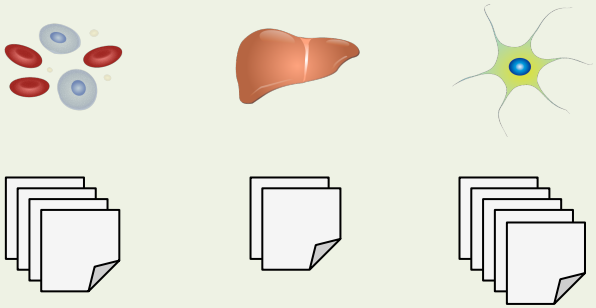

ChIP-Atlas
Peak Browser
Co-Association Analysis
Target Gene Analysis
VirtualChIP Analysis
Documentation
ChIP-Atlas - Co-Association Analysis
Tutorial movie
Predict co-association partners of TFs.
H. sapiens
M. musculus
D. melanogaster
C. elegans
S. celevisiae
1. Search mode
Antigens → Cell Type
Cell Type → Antigen
1. Choose Antigen
type to search
MYNN
MYOD1
NANOG
NCOA1
NCOA3
NCOR1
NCOR2
NELFA
2. Choose Cell Type Class
type to search
Pluripotent stem cell
View Co-Association Data

③ NANOG の共局在パートナーは？

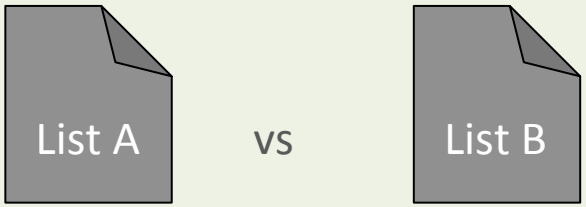
④ ユーザデータの解析



使用データ：
全 Peak-call データ



Overlap enrichment を比較
Tool: BedTools



受け入れ可能なリスト:
Gene symbols, Bed, sequence motif

④ ユーザデータの解析

ChIP-Atlas | Peak Browser | Co-Association Analysis | Target Gene Analysis | VirtualChIP Analysis | Documentation

ChIP-Atlas - VirtualChIP

Tutorial movie

Analyze your data with public ChIP-seq data.

H. sapiens | M. musculus | D. melanogaster | C. elegans | S. celevisiae

1. Select your data

○ Genomic regions (BED) or sequence motif ⓘ

◉ Gene list (Gene symbols) ⓘ

SERPINA11
SERPIND1
SLC17A2
SLC22A9
SMLR1
SOAT2
SULT2A1
TF
TM4SF5
TTR

肝臓特異的
遺伝子リスト

ファイルを選択 ファイル未選択

2. Select dataset to be compared

◉ Refseq coding genes (exluding user data) ⓘ

○ Gene list (Gene symbols) ⓘ

Click the info buttons above to show the description format.

その他の
遺伝子

ファイルを選択 ファイル未選択

3. Describe datasets

User data title: Sample A

Compared data title: Sample B

Project title: My project

submit

[illegible]D1 ChIP peaks.bed

ランダム.bed

T[illegible]GCGAGA motif

T[illegible]TCGAGA motif

④ ユーザデータの解析

肝臓特異的発現する遺伝子に enrich するタンパク質

TFs	Cells	$Log_{10}P$
EP300	Liver/Hep G2	−33.9
HNF4G	Liver/Hep G2	−31.5
HNF4A	Liver/Hep G2	−30.4
FOXA2	Liver/Hep G2	−23.7
FOXA1	Liver/Hep G2	−21.8
SP1	Liver/Hep G2	−21.6
RXRA	Liver/Hep G2	−21.2
NR2F2	Liver/Hep G2	−15.8
CEBPB	Liver/Hep G2	−13.5
NFIC	Liver/Hep G2	−12
TEAD4	Liver/Hep G2	−11.1
ARID3A	Liver/Hep G2	−10.2
TBP	Liver/Hep G2	−9.6
HEY1	Liver/Hep G2	−9.4
HDAC2	Liver/Hep G2	−8.4
JUND	Liver/Hep G2	−7.8
HNF4A	Digestive tract/Ca	−5.8
MYBL2	Liver/Hep G2	−5.7
TCF7L2	Liver/Hep G2	−5.7

C2C12 における[illegible]D1 ChIP peaks にenrich するタンパク質

TFs	Cells	$Log_{10}P$
Tcf12	Muscle/C2C12	−324
Myod1	Adipocyte/Brown	−324
Jund	Embryo/Embryoni	−324
Tcf3	Muscle/Myoblasts	−324
Ep300	Breast/3134	−324
Runx2	Bone/MC3T3-E1	−324
Myod1	Embryonic fibrobla	−324
Cebpb	Bone/MC3T3-E1	−314.6
Ep300	Embryonic fibrobla	−312.7
Runx2	Bone/MC3T3-E1	−311.4
Nr3c1	Breast/3134	−308.6
Runx2	Bone/MC3T3-E1	−301.5
Rxra	Bone/IDG-SW3	−269.8
Smarca4	Breast/3134	−264.4
Myod1	Muscle/C2C12	−254.4
Rxra	Bone/IDG-SW3	−253.3
Runx2	Bone/MC3T3-E1	−249.5
Rxra	Bone/IDG-SW3	−236.9
Ascl1	Neural/Neural pro	−236.8

T[illegible]GCGAGA モチーフに enrich するタンパク質

TFs	Cells	$Log_{10}P$
CHD2	Blood/K-562	−168.5
CHD2	Pluripotent stem c	−115.7
CHD2	Liver/Hep G2	−111.6
ZBTB33	Blood/GM12878	−108.8
ZBTB33	Lung/A549	−105.2
ZBTB33	Digestive tract/H(	−100.7
EP300	Kidney/293	−99.7
ZBTB33	Neural/SK-N-SH	−98.6
MAX	Lung/NCI-H128	−92.3
KDM5B	Breast/T-47D	−92.2
CHD2	Uterus/HeLa	−91
MAX	Neural/U-87 MG	−89.3
NOTCH1	Blood/T-ALL	−88.4
BRD4	Blood/MM.1S	−88.1
TCF7L2	Breast/MDA-MB-	−86.9
MAX	Lung/NCI-H2171	−85.9
STAT1	Blood/Monocytes	−84.2
MAX	Blood/P493-6	−82.8
RUNX1	Blood/CD34+	−79.9

メタデータのクレンジング

+ シーケンス 生データ

+ メタデータ

属性	属性値
ID	SRX213809
Title	AntiGFP KD Oct4; Mus musculus; ChIP-Seq
source_name	Anti-GFP KD mESCs
strain	129S4/Svjae
phenotype	agouti
gender	male
cell type	ESC
genotype	Anti-GFP shRNA KD
chip antibody	Oct4(N-19)(sc-8628), Santa Cruz Biotechnology

抗原：Pou5f1
細胞：Embryonic Stem Cells

メタデータのクレンジング

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)

↓

Pou5f1

細胞名

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

↓

Embryonic stem cells

メタデータのクレンジング

ID	Genome	抗原	細胞クラス	細胞
SRX030144	mm9	Pax3	Muscle	Myoblasts
SRX145064	mm9	Pax5	Blood	CH12
SRX030143	mm9	Pax7	Muscle	Myoblasts
SRX195321	mm9	Phf19	Pluripotent stem cell	Embryonic Stem Cells
SRX195355	mm9	Pitx1	Embryo	Embryonic limb
SRX170715	mm9	Pknox1	Embryo	Embryonic trunk
SRX003866	mm9	Pou5f1	Pluripotent stem cell	Embryonic Stem Cells
SRX118173	mm9	Ppara	Liver	Liver
SRX003137	mm9	Pparg	Embryonic fibroblast	3T3-L1
ERX204071	mm9	Rad21	Liver	Hepatocytes
SRX467472	mm9	H3K27me3	Blood	B-Lymphocytes
SRX023706	mm9	H3K36me3	Blood	B-Lymphocytes
SRX185797	mm9	H3K4me1	Adipocyte	Brown adipocytes
SRX341043	mm9	H3K4me2	Adipocyte	Brown preadipocytes
SRX185796	mm9	H3K4me3	Adipocyte	Brown adipocytes

任意の属性値でデータ抽出が可能。