

統合的な転写制御データ基盤 INTRARED の構築について

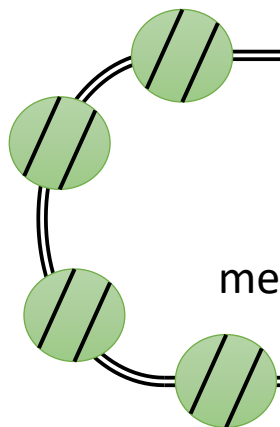
理化学研究所 生命医科学研究センター
粕川雄也

概要：転写制御に関わる様々な要素

転写制御に関わるゲノム領域

シスエレメント (CRE)

nucleosome



methylation



Silencer

Enhancer



Promoter

ゲノム等に結合し転写を制御する転写因子等

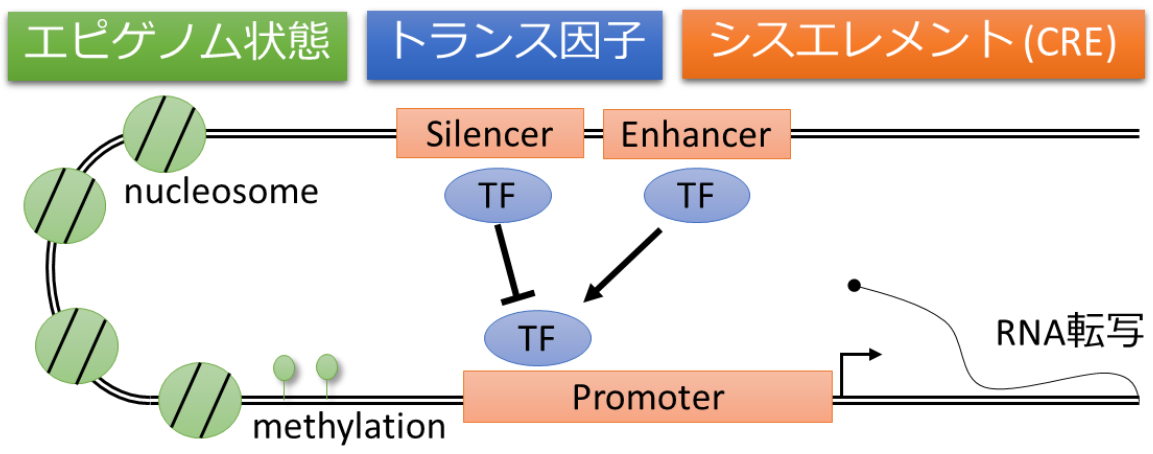
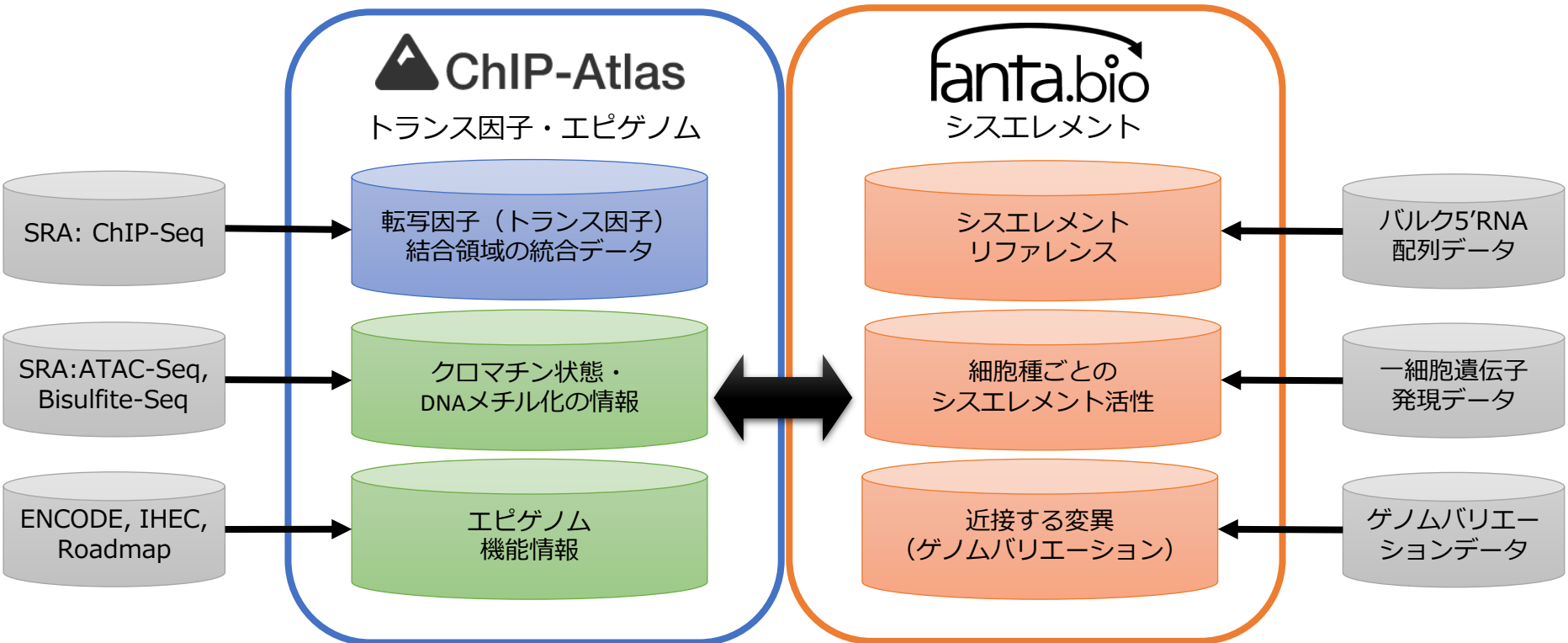
トランス因子

RNA転写

エピゲノム状態

転写制御に影響を与えるメチル化やヌクレオソーム位置等の状態

統合的転写制御データ基盤 (INTRARED: Integrated TRANscriptional Regulation Data platform)



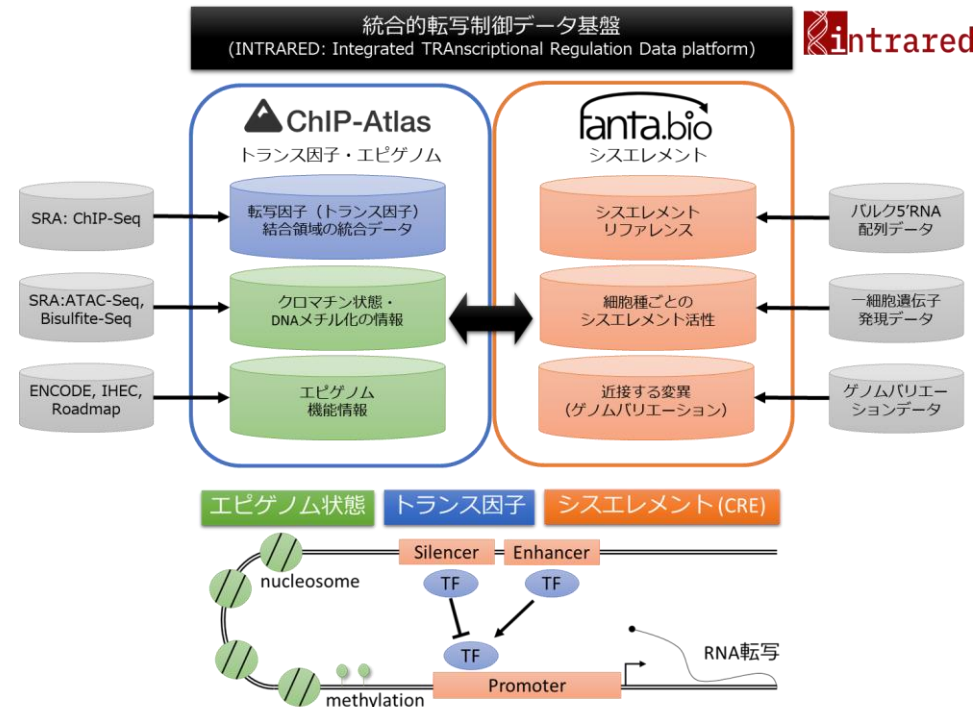
統合的転写制御データ基盤

- 転写制御に関わる
 - 「シスエレメント」
 - 「トランス因子（転写因子）」
 - 「エピゲノム状態」の情報を集約したデータ基盤

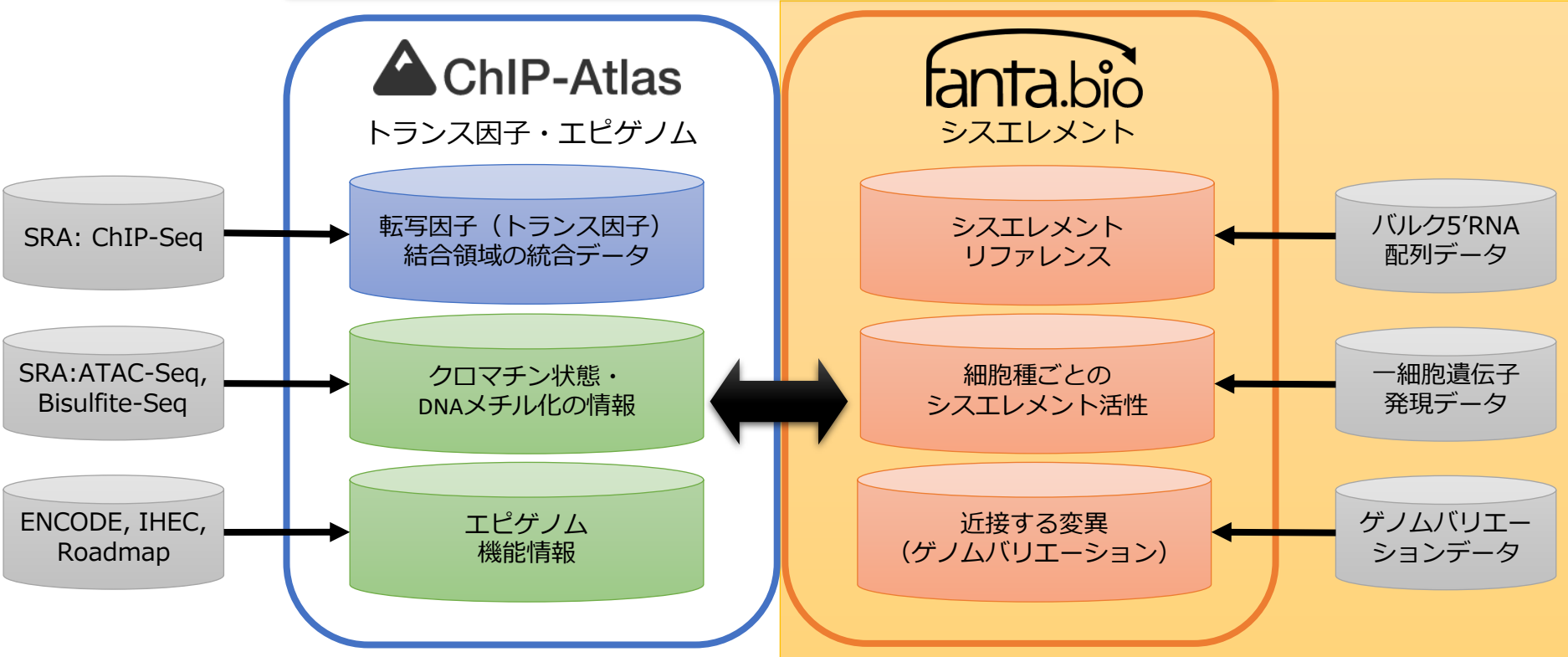
- すでに存在する「ChIP-Atlas」と新規作成する「fanta.bio」で構成される

- データ作成を各々のDBで分担し一つのまとまったデータリソースとして公開予定

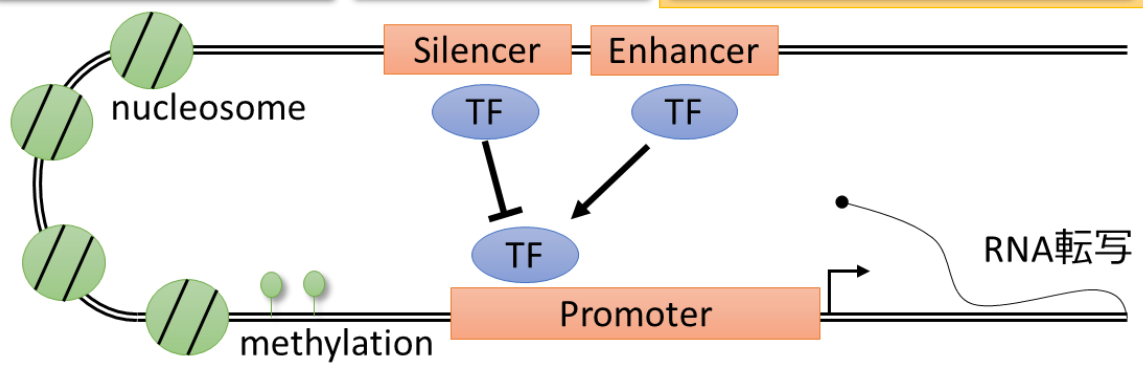
- インタフェースは参照・検索目的ごとに各々作成する



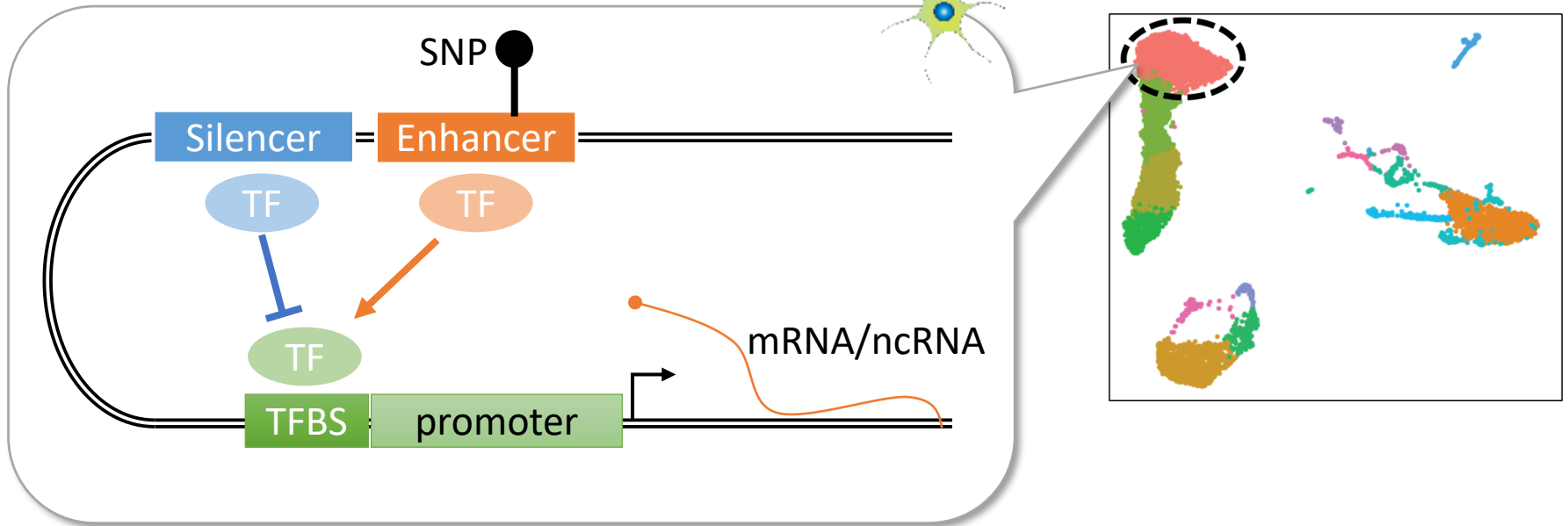
統合的転写制御データ基盤 (INTRARED: Integrated TRANscriptional Regulation Data platform)



エピゲノム状態 トランス因子 シスエレメント (CRE)

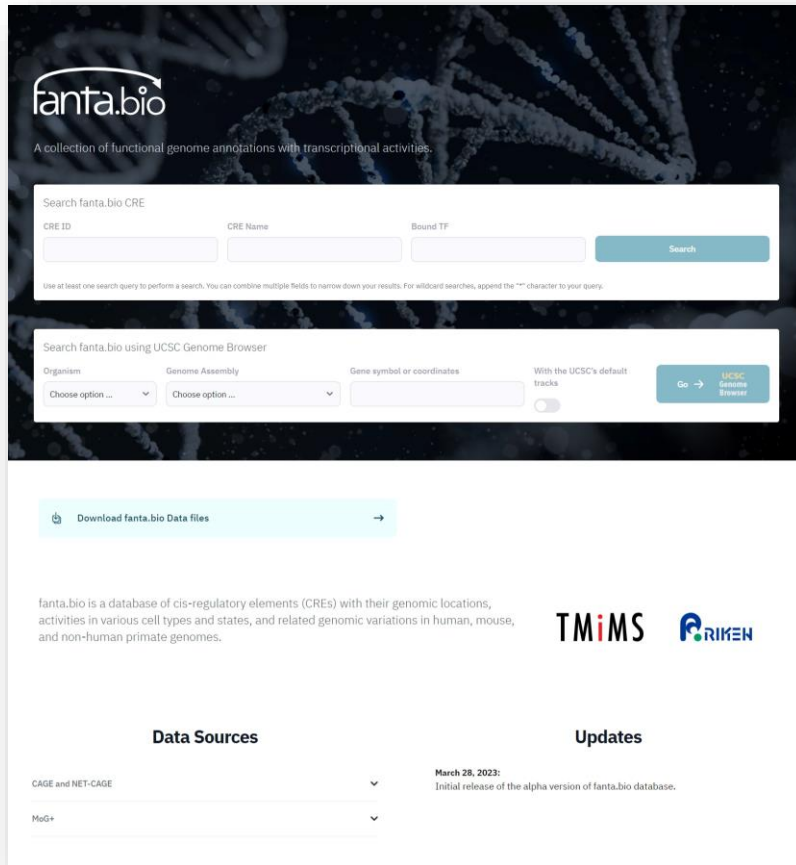


シスエレメントとは



- 遺伝子発現の調節に関わるゲノム領域
- 「プロモーター」・「エンハンサー」など非コード領域にあるものが一般的
- 細胞種ごとの特異的な発現を調節
- 表現型に関係する多型・変異が頻出する

- 全貌解明はまだ途上である



<https://fanta.bio/>

本プログラムで新規作成

公共データを再処理し、ゲノム中の転写制御に関わる領域（シスエレメント/CRE）について

- ゲノム中のCREの位置
- 細胞種や細胞状態ごとのCREの活性
- CREと関係するゲノム変異の情報をまとめたデータベース

対象とする生物種

- ヒト、マウス、非ヒト霊長類

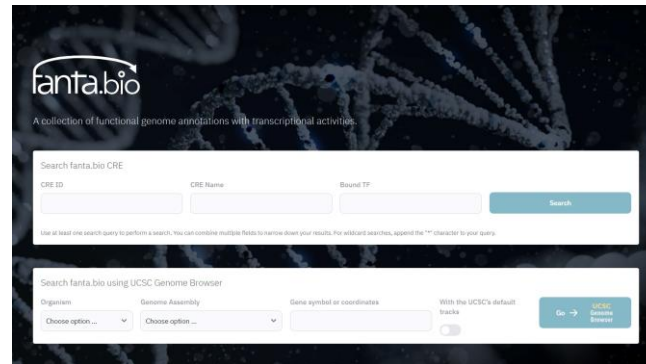
提供方法

- UCSC Genome Browser でのビュー (TrackHub)
- CREアノテーションのビューアー
- バルクデータ

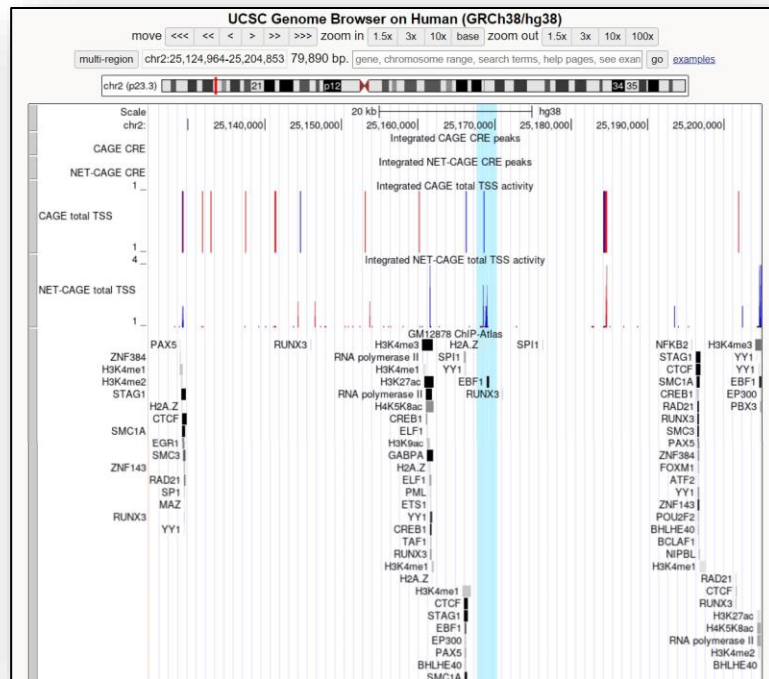
fanta.bio



fanta.bio 検索フォーム



UCSC Genome Browser
CREのゲノム上でのビュー



CRE annotation view
CREについての各種
アノテーション
情報のビュー



fanta.bio

CRE ID : flrtmp_hg38_11
CRE Name : cp9@SAM011

Genomic Position : chr1:923133 - 923733

Promoter/TSS annotation

| | |
|-------------------------------|--|
| Nearest transcript IDs | ENST00000616016.5, ENST00000618323.5 |
| Nearest transcript distance | 190 |
| Promoter GeneID | 148308 |
| Promoter HGNC ID/MGI ID | HGNC:28796 |
| Promoter UniProt ID | A0A0R79M17.4, A0A0R79W2.4 |
| Promoter Gene Names | sterile alpha motif domain containing 11 |
| Promoter Gene Symbols | SAMD11 |
| Promoter Gene Synonyms | MGC45873 |
| Overlapped FANTOM5 CAGE Peaks | id : hg29:chr1:859638_858448,+hg_21.1 name: hg_21.1 |
| Overlapped refTSS TSSs | rhg_69.1 |

Enhancer annotation

Overlapped FANTOM5 Enhancers : chr1:922829-923779

CRE annotation

Overlapped SCREEN cCREs : id : E08022115450
type: eELS.CTCF-bound

External Links : [Ensembl](#) [UCSC Genome Browser](#)

データベース連携

ChIP-Atlas Antigens
List of ChIP-Atlas antigens for this CRE.

Q Search...

| Antigen | Max Qscore | External Links |
|-----------|------------|--|
| ▼ HNS1 | 3034 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ ERG | 3482 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ OSER1 | 3424 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ RCOR1.1 | 3194 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ USF | 3156 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ SMY3 | 3127 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ NFKB1 | 3036 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ RFX2 | 3022 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ RFX6 | 3007 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ NFKB2 | 3003 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ ZNF42 | 3001 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ PRKSR1 | 895 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ HNF4B | 796 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ BAP1 | 713 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ LMO1 | 687 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ SPI1 | 685 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ SPAD1 | 677 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ CTCF | 659 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ ZNF708 | 625 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |
| ▼ KLF8 | 615 | ChIP-Atlas ENTP1 ENTP2 ENTP3 |

1-20 of 47 items 1 of 4 pages

CRE 領域に結合する
転写因子情報

(ChIP-Atlas)

TogoVar Variations
List of TogoVar variations for this CRE.

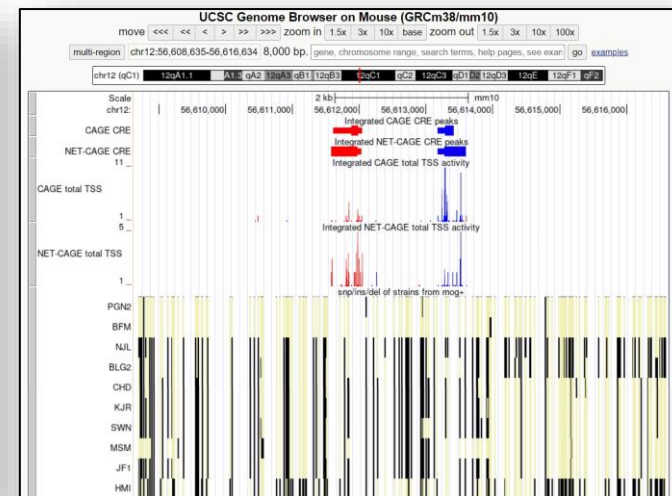
Q Search...

| TogoVar ID | RefSNP ID | Position | Reference | Alternate | Type |
|--------------|-------------|-----------|-----------|-----------|------|
| rs123066169 | -123066169 | 1,923,331 | C | G | SNV |
| rs92218751 | -92218751 | 1,923,336 | G | A | SNV |
| rs139853348 | -139853326 | 1,923,338 | C | G | SNV |
| rs997 | - | 1,923,353 | G | A | SNV |
| rs1230989132 | -1230989132 | 1,923,354 | G | A | SNV |
| rs143062745 | -143062745 | 1,923,355 | C | G | SNV |
| rs127214456 | -127214456 | 1,923,357 | C | T | SNV |
| rs38933270 | -38933270 | 1,923,358 | C | G | SNV |
| rs138070769 | -138070769 | 1,923,363 | G | A | SNV |
| rs104839416 | -104839416 | 1,923,367 | C | G | SNV |
| rs9230487 | -9230487 | 1,923,368 | C | T | SNV |
| rs133969733 | -133969733 | 1,923,368 | G | A | SNV |
| rs133756158 | -133756158 | 1,923,376 | G | A | SNV |
| rs96131766 | -96131766 | 1,923,375 | G | A | SNV |
| rs96131788 | -96131788 | 1,923,375 | G | T | SNV |
| rs79087627 | -79087627 | 1,923,376 | C | T | SNV |
| rs103936306 | -103936306 | 1,923,379 | G | A | SNV |
| rs109024279 | -109024279 | 1,923,387 | T | C | SNV |
| rs58773573 | -58773573 | 1,923,394 | G | C | SNV |
| rs138132235 | -138132235 | 1,923,396 | C | T | SNV |

1-30 of 237 items 1 of 27 pages

CRE 領域内にあるヒト
変異情報

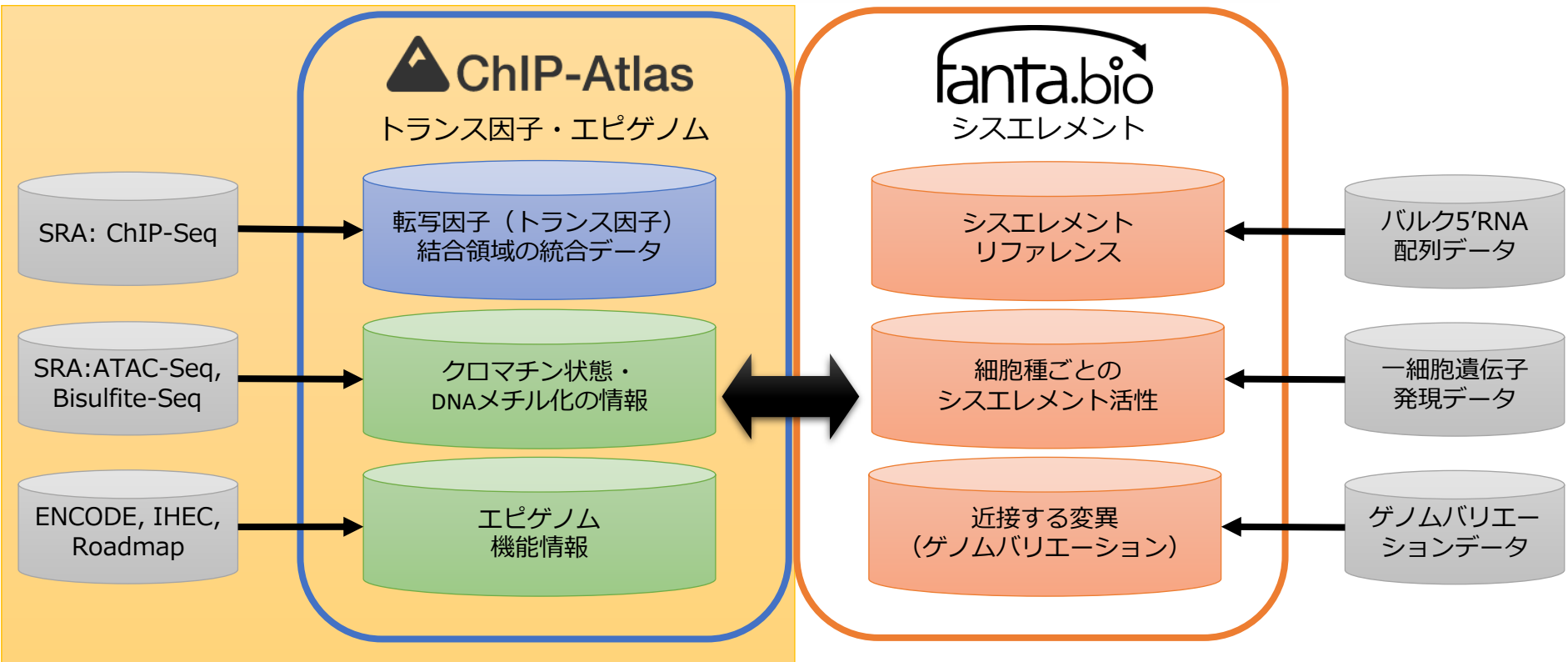
(TogoVar)



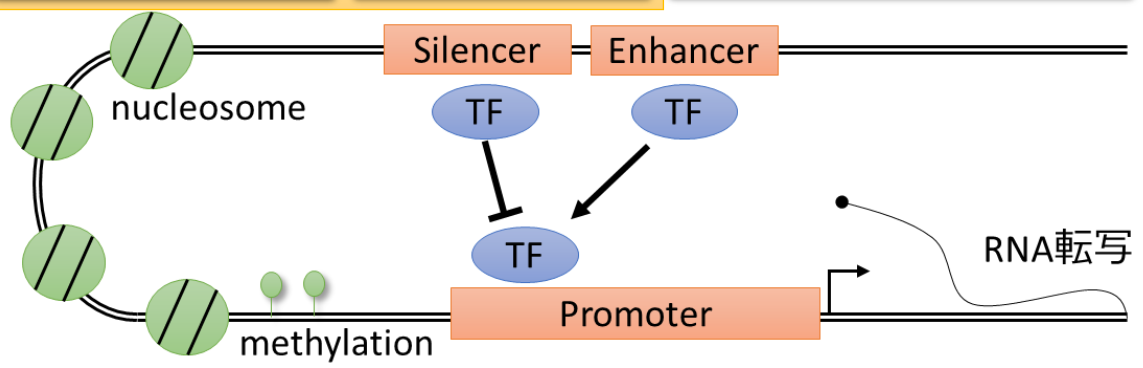
CRE 領域内にあるマウス
変異情報

(MoG+)

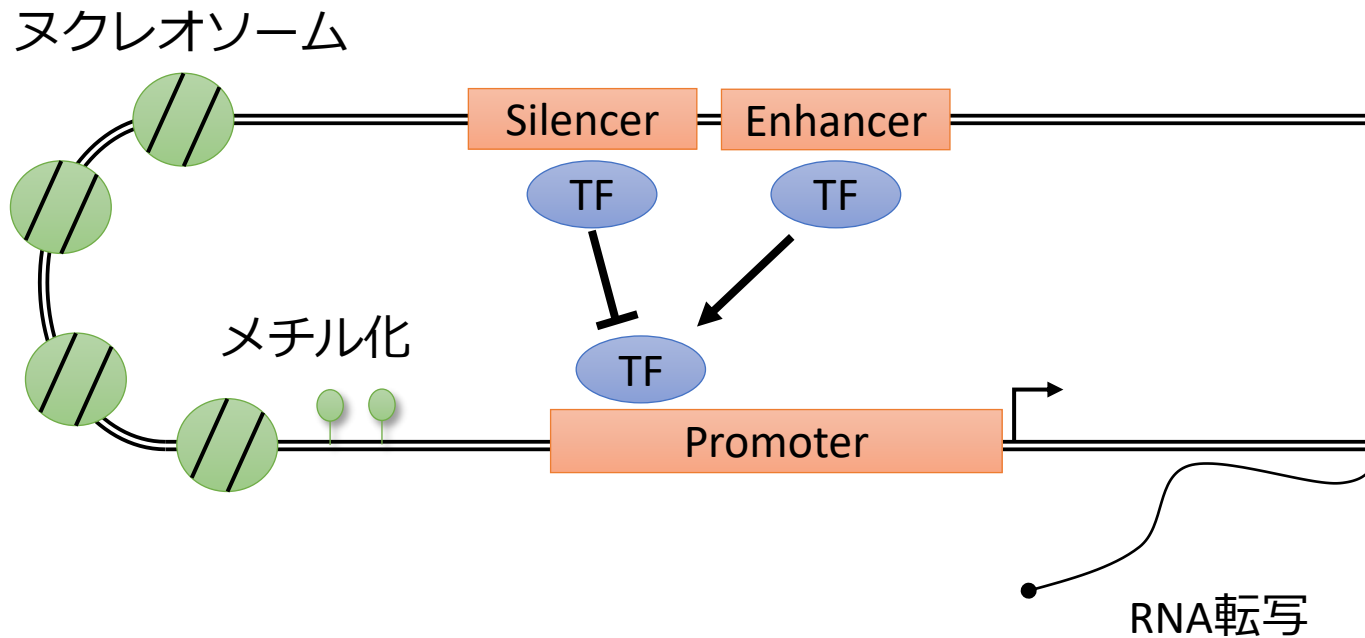
統合的転写制御データ基盤 (INTRARED: Integrated TRANscriptional Regulation Data platform)



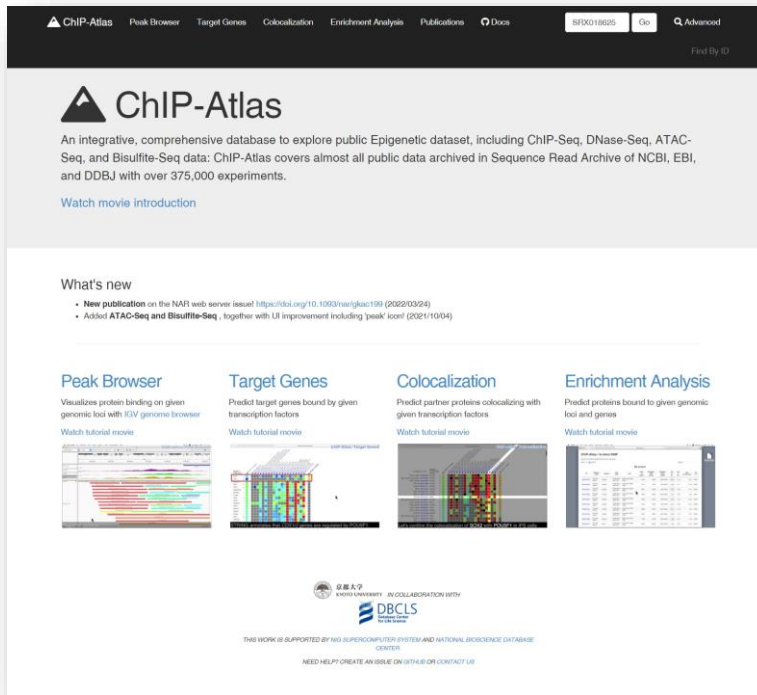
エピゲノム状態 トランス因子 シスエレメント (CRE)



トランス因子・エピゲノム状態



- トランス因子
 - ゲノムDNA（シスエレメント）に結合して、転写制御の調整を担う因子
 - 転写因子等
- エピゲノム（状態）
 - メチル化やクロマチンの開閉状態等、ゲノム配列以外のゲノム上の状態
 - 転写制御に強い関わり
 - ゲノムDNAのメチル化、ヒストンのメチル化、ヌクレオソームの位置・クロマチンの開閉状態等



公共のChIP-seqやその他エピゲノムデータを統合した、転写因子結合情報・エピゲノム状態についてのデータベース

- ChIP-seq
- ATAC-seq
- Bisulfite-seq
- DNase-seq
- 機能注釈データ

対象とする生物種

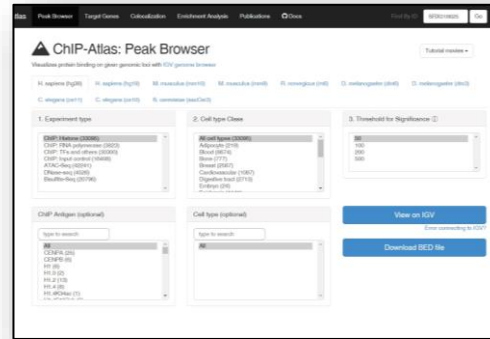
- ヒト、マウス、ラット、ショウジョウバエ、線虫、酵母

提供方法

- igvによるビュー
- 独自の検索・解析インタフェース
- バルクデータ

<https://chip-atlas.org/>

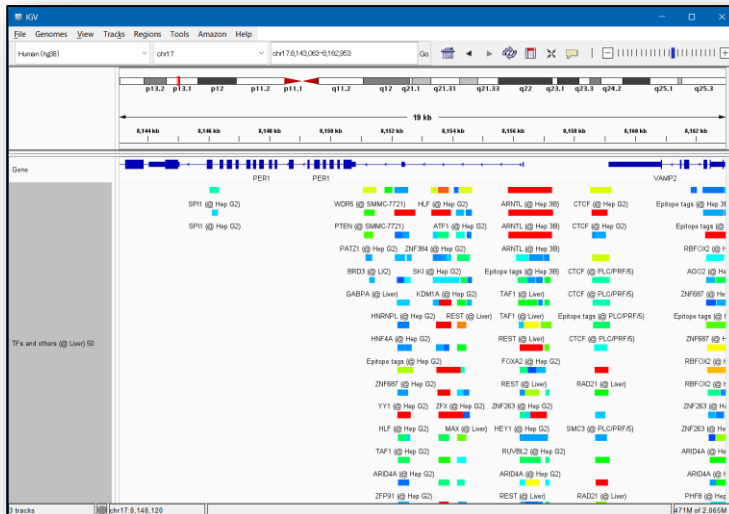
本プログラムで機能拡張・高度化



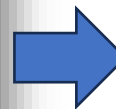
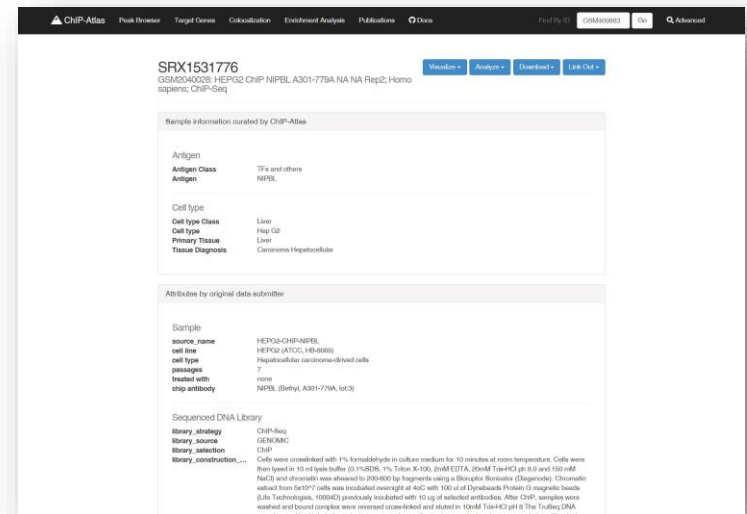
ChIP-Atlas 検索フォーム



IGV
トランス因子・エピゲノム状態の
ゲノム上でのビュー



各種シーケンス実験・データセッ
トについての情報

SRX1531776
GSE1204025: HepG2 ChIP NIPBL A301-779A NA NA Rep2; Homo sapiens; ChIP-Seq

Sample information curated by ChIP-Atlas

| | |
|------------------|--------------------------|
| Antigen | TFs and others |
| Antigen Class | NIPBL |
| Antigen | |
| Cell type | Liver |
| Cell type Class | Hep G2 |
| Cell type | Liver |
| Primary Tissue | Carcinoma Hepatocellular |
| Tissue Diagnosis | |

Attributes by original data submitter

| | |
|---------------|--|
| Sample | |
| source_name | HEPG2-CHIP-NIPBL |
| cell_line | HEPG2 (ATCC: HB-8002) |
| cell type | Hepatocellular carcinoma-derived cells |
| passages | 7 |
| treated with | none |
| chip antibody | NIPBL (Bethyl, A301-779A, lot3) |

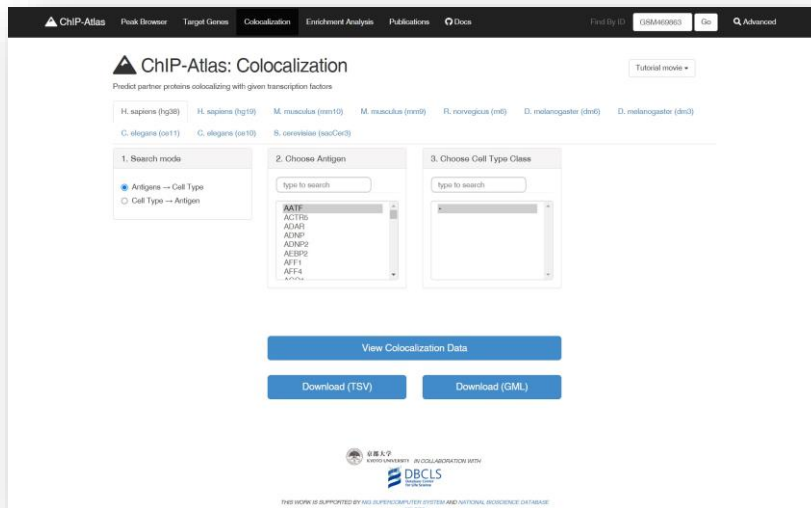
Sequenced DNA Library

| | |
|-------------------------|----------|
| library_strategy | ChIP-Seq |
| library_source | GENOMIC |
| library_selection | CHIP |
| library_construction... | |

Cells were crosslinked with 1% formaldehyde in culture medium for 10 minutes at room temperature. Cells were then lysed in 10 ml lysis buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.0 and 150 mM NaCl) and chromatin was allowed to crosslink for 10 minutes using a Bio-Rad Sonicator (Diagenode). Chromatin was cut into 100bp fragments and incubated overnight at 4°C with 100 µl of Dynabeads Protein G magnetic beads (Life Technologies, 100043) previously incubated with 10 µg of selected antibodies. After ChIP, samples were washed and bound complex were reversed cross-linked and eluted in 100µM Tris-HCl pH 8. The Tris-HCl DNA was purified using a spin column and sequenced on a HiSeq 2500 (Illumina) using TruSeq library preparation kit.

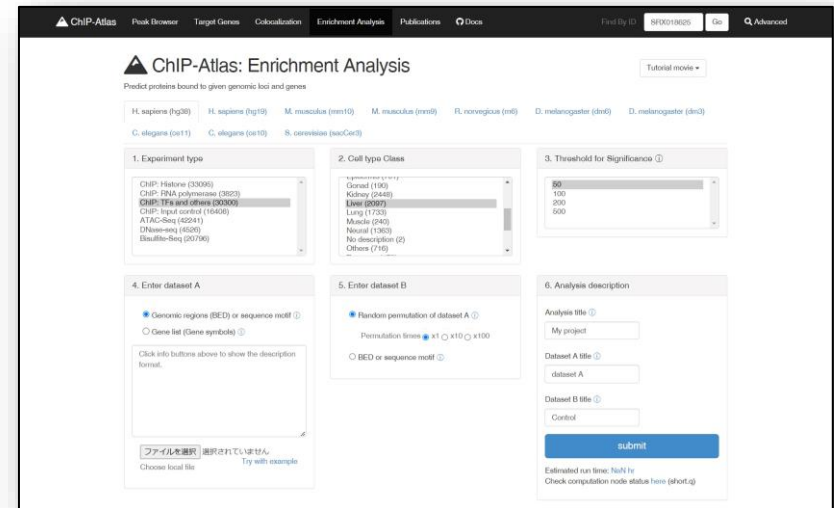
ChIP-Atlasデータの様々なウェブ解析ツール

Colocalization



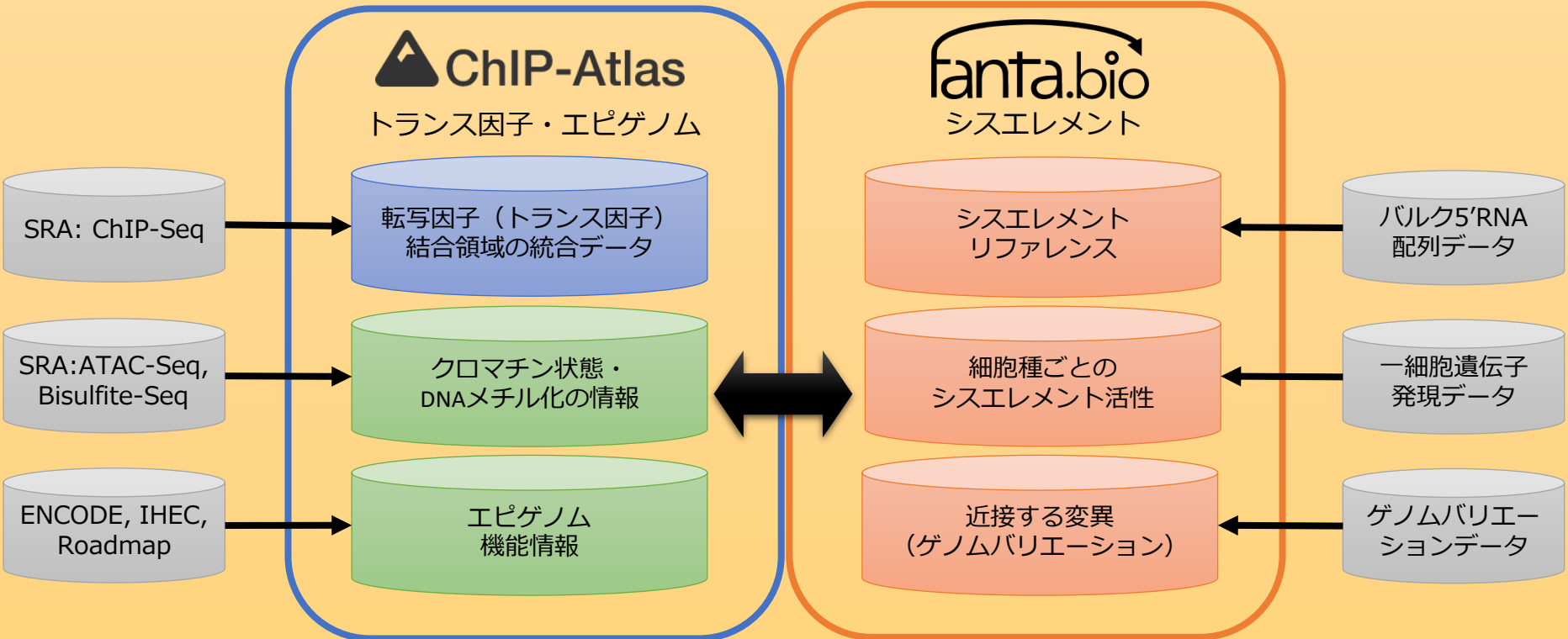
指定した転写因子とゲノム上で
共局在する転写因子の探索

Enrichment analysis

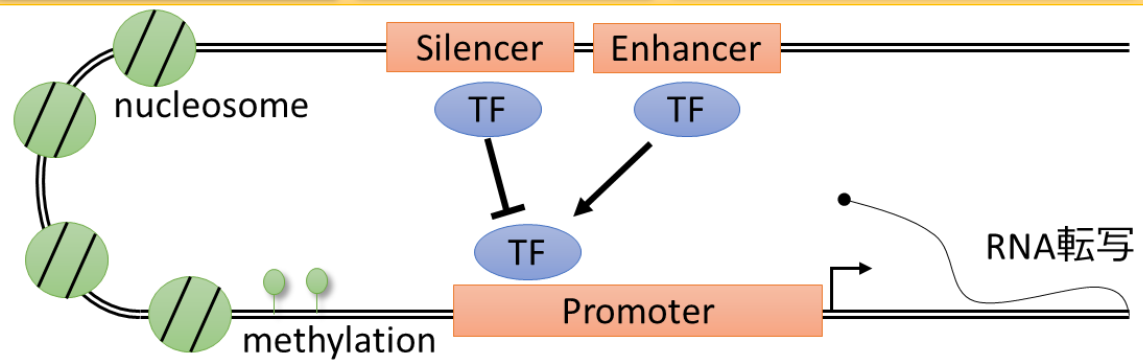


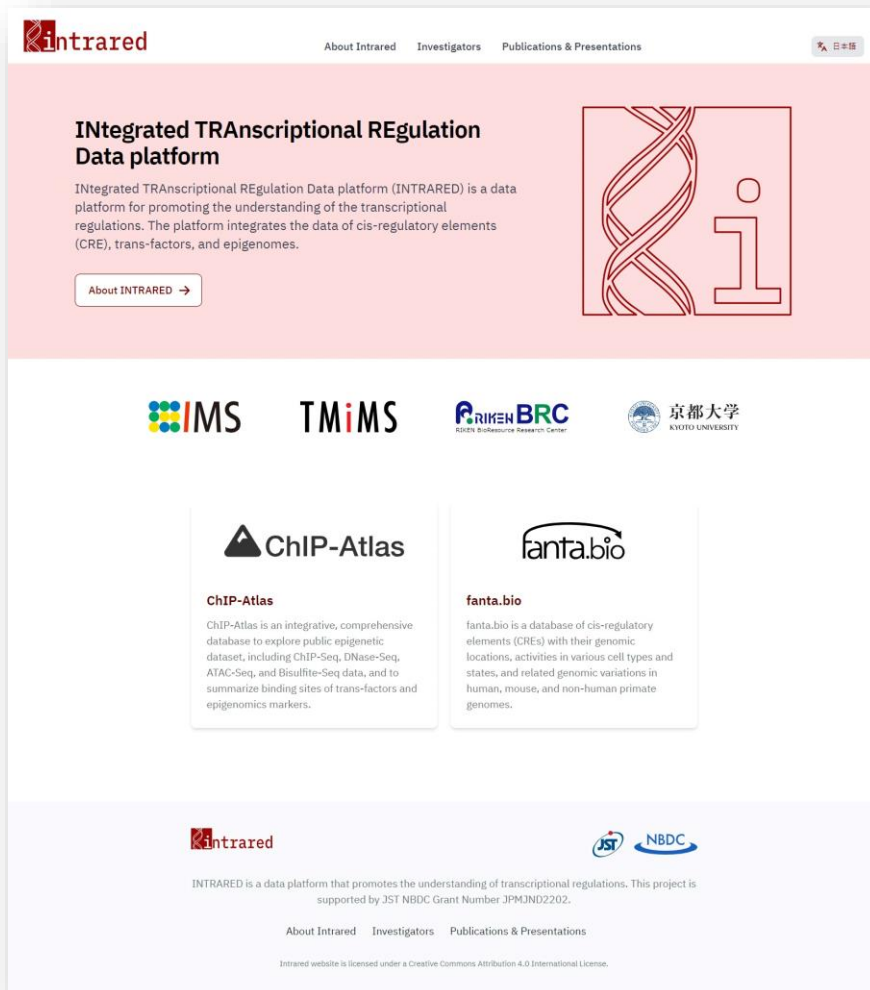
指定したゲノム領域や遺伝子群で
有意に結合する転写因子の探索

統合的転写制御データ基盤 (INTRARED: Integrated TRANscriptional Regulation Data platform)



エピゲノム状態 トランス因子 シスエレメント (CRE)





The screenshot shows the homepage of the INTRARED platform. At the top left is the INTRARED logo. Navigation links for 'About Intrared', 'Investigators', and 'Publications & Presentations' are in the top right. A search bar is also present. The main heading is 'INtegrated TRAnscriptional REGulation Data platform'. Below this is a brief description of the platform and a link to 'About INTRARED'. A large graphic of a DNA double helix and a stylized 'i' is shown. Below the main content are logos for partner institutions: IMS, TMiMS, RIKEN BRC, and Kyoto University. Two featured databases are highlighted: ChIP-Atlas and fanta.bio, each with a brief description of their content. At the bottom, there is a footer with the INTRARED logo, a paragraph of text about the project's support by JST/NBDC, and a Creative Commons license notice.

各DBを統合した転写制御データ基盤

現時点ではプロジェクトのポータルサイト

今後、バルクダウンロードのポータルページや、統合検索などを追加していく予定

<https://www.intrared.org/>

fanta.bio/ChIP-Atlas/INTRAREDの現在の状況（リリース計画）

- fanta.bio: <https://fanta.bio/>
 - 2023年3月にプレビュー用データによるアルファ版リリース (UCSC Genome Browser のTrackHub + バルクダウンロード)
 - 2023年10月にシスエレメントアノテーションの参照・検索対応
 - 年内に正式リリース版の公開予定

- ChIP-Atlas: <https://chip-atlas.org/>
 - 月次更新により逐次データベース更新中
 - 機能注釈データの拡充版を今後公開予定
 - 解析ツールの拡充準備中

- INTRARED: <https://www.intrared.org/>
 - ポータルサイトとして公開中
 - 今後コンテンツの拡充予定（ダウンロード、ユースケースの掲載等）



開発体制

研究分担(3) : 沖G (京都大)



- トランス因子・エピゲノムデータの収集
- ChIP-Atlas更新

研究代表 : 粕川G (理研IMS)



- 研究開発の取りまとめ
- ヒト・マウスのシスエレメントデータの作成
- インタフェース開発
- データベース運用

研究分担(1) : 川路G (都医学研)



- シスエレメント同定のパイプライン開発
- ゲノムビューアー可視化
- 非ヒト霊長類のシスエレメントデータの作成

研究分担(2) : 柘屋G (理研BRC)



- ゲノム変異とシスエレメントの対応づけ
- メタデータの作成

謝辞

- 理化学研究所IMS
 - 長谷川哲、Scott Walker, Nishad Thalhath, 信定知江
- 東京都医科学研究所
 - 川路英哉
- 理化学研究所BRC
 - 柘屋啓志、高田豊行
- 京都大学大学院医学研究科
 - 沖真弥、鄒兆南
- ライフサイエンス統合データベースセンター
 - 三橋信孝