

統合的な転写制御データ基盤の構築

粕川 雄也

理化学研究所 生命医科学研究センター (IMS)

生命医科学大容量データ技術研究チーム

チームリーダー

takeya.kasukawa@riken.jp

統合的転写制御データ基盤

fanta.bio シスエレメント

シスエレメント
リファレンス
(ヒト・マウス・
非ヒト霊長類)

細胞種ごとの
シスエレメント活性

近接する変異
(ゲノムバリエーション)



ChIP-Atlas トランス因子・エピゲノム

トランス因子・エピゲノム

転写因子
(トランスエレメント)
結合領域の統合データ

クロマチン状態・
DNAメチル化の情報

エピゲノム
機能情報



バルク
5'RNA配列
データ

一細胞遺伝子
発現データ

ゲノム
バリエーション
データ

SRA
ChIP-Seqデータ

SRA
ATAC-Seq
Bisulfite-Seq

ENCODE
IHEC
Roadmap

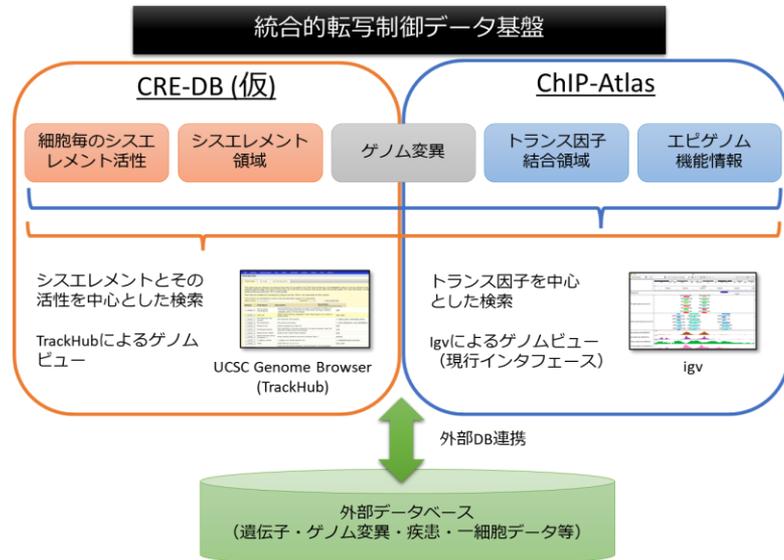
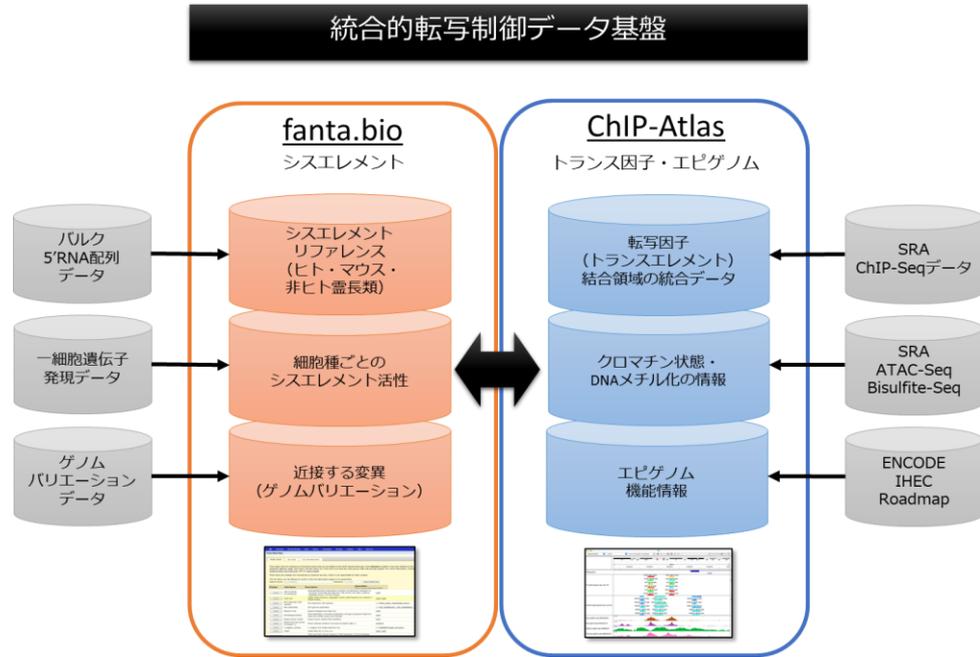
統合的転写制御データ基盤

- 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（クロマチン状態）」
 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



開発体制

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
- 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

fanta.bio

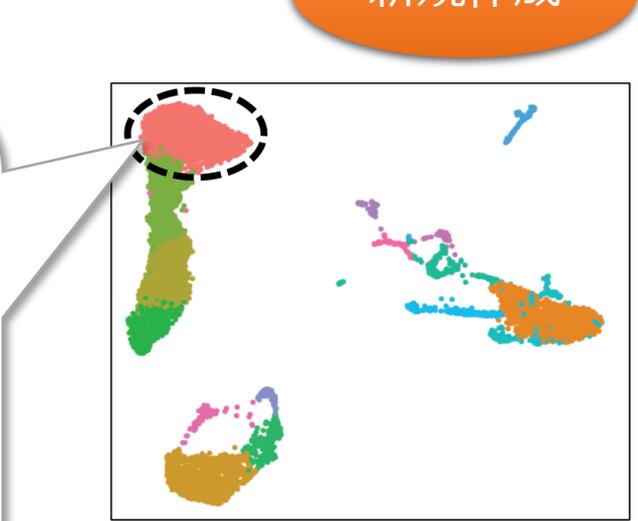
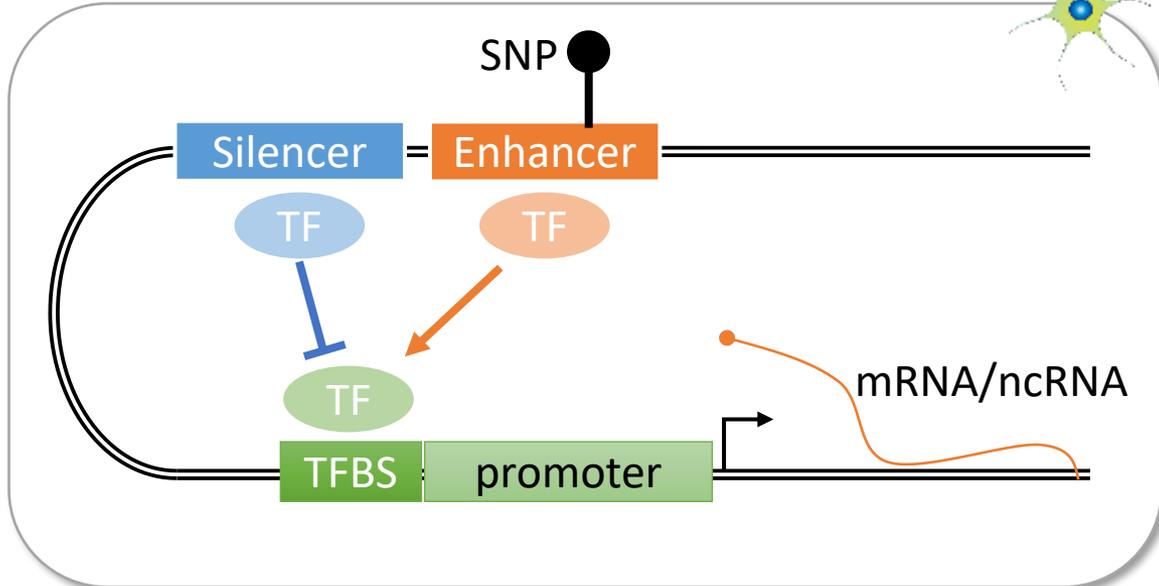
研究分担(3)：沖G（京都大）



- トランス因子・エピゲノムデータの収集
- ChIP-Atlas更新

ChIP-Atlas

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

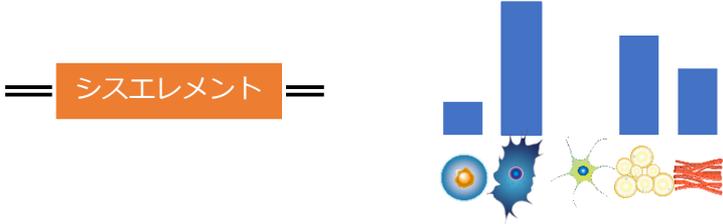
- 全貌解明はまだ途上である

シスエレメントの全貌解明に向けて

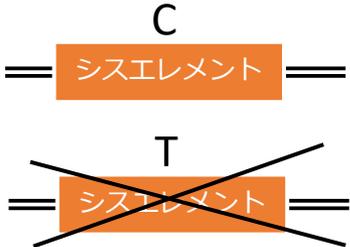
1) シスエレメントの位置 (静的な情報)



2) 各細胞種・状態におけるシスエレメントの活性 (動的な情報)



3) どの多型が転写制御活性を変えるのか (論理的な情報)



これらが網羅されたデータベースは存在しない

シスエレメント・データベース

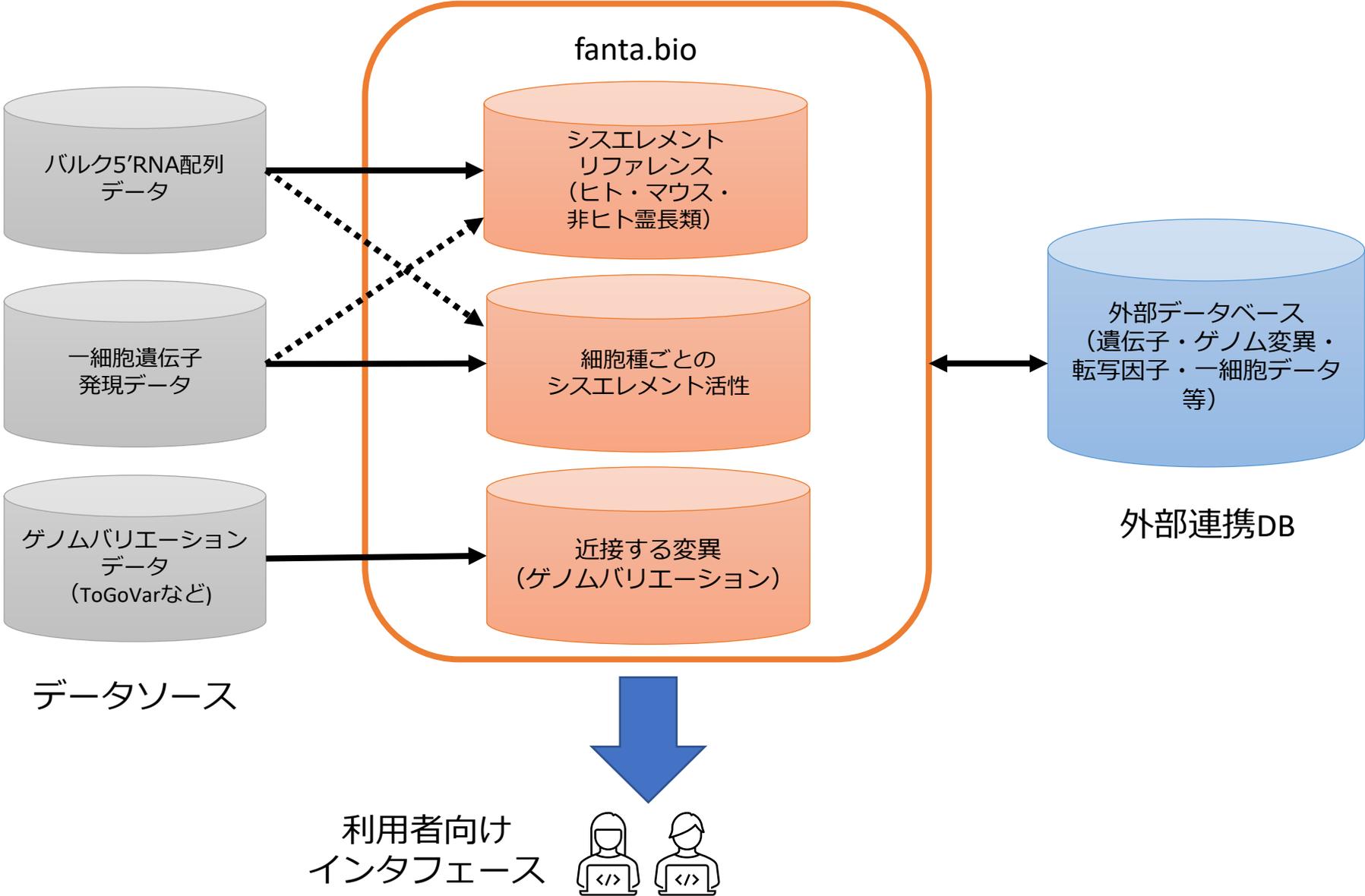


1) シスエレメントの位置 (静的な情報)

2) 各細胞種・状態におけるシスエレメントの活性 (動的な情報)

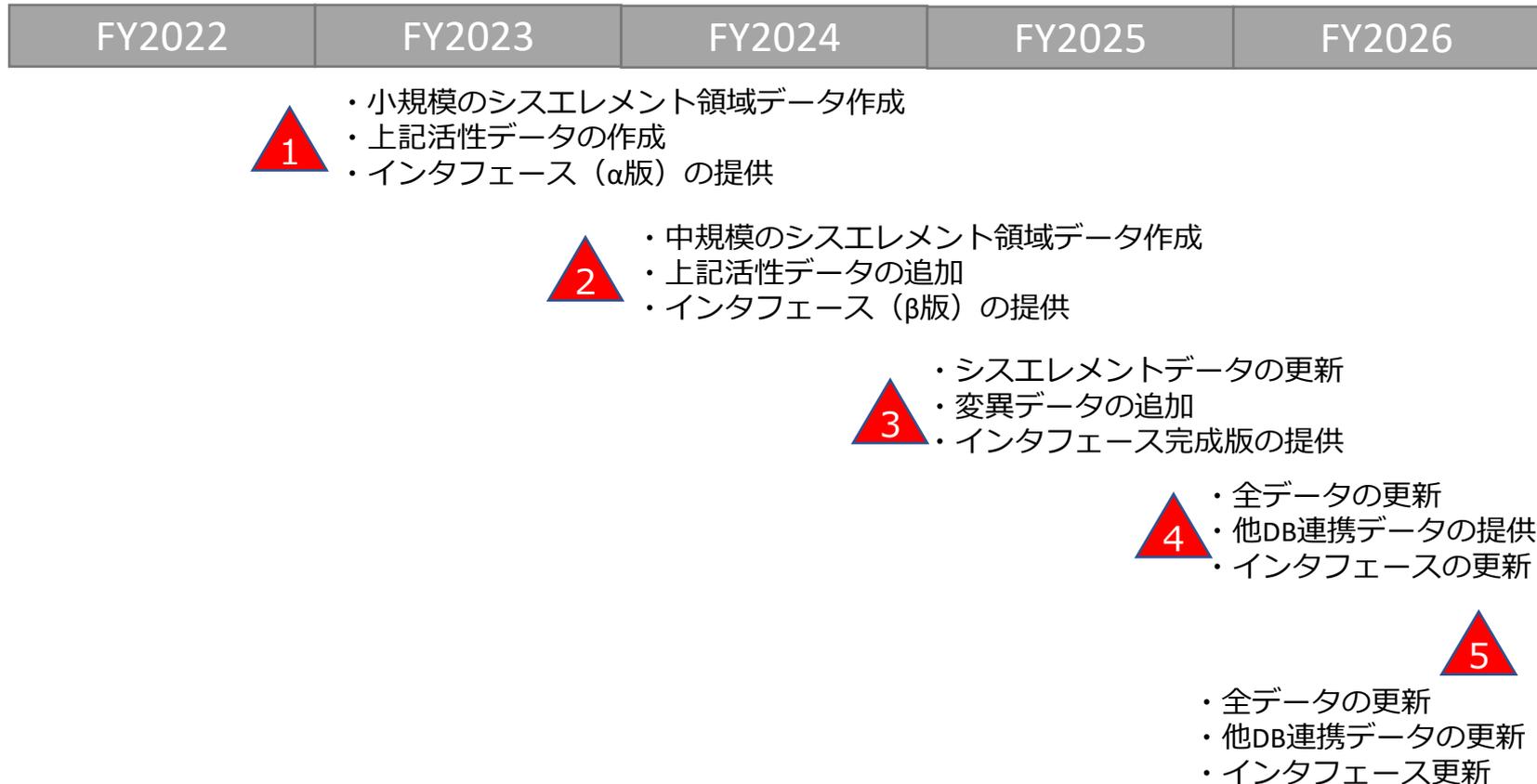
3) どの多型が転写制御活性を変えるのか (論理的な情報)

シスエレメント・データベースの構築



シスエレメント・データベース（ロードマップ）

新規作成



概ね1年に1回程度のメジャーリリースを目指す

ChIP-Atlas

[ChIP-Atlas](#)
[Peak Browser](#)
[Target Genes](#)
[Colocalization](#)
[Enrichment Analysis](#)
[Publications](#)
[Docs](#)

 Find By ID: [Advanced](#)



ChIP-Atlas

An integrative, comprehensive database to explore public Epigenetic dataset, including ChIP-Seq, DNase-Seq, ATAC-Seq, and Bisulfite-Seq data: ChIP-Atlas covers almost all public data archived in Sequence Read Archive of NCBI, EBI, and DDBJ with over 224,000 experiments.

[Watch movie introduction](#)

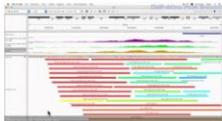
What's new

- Minor bug fix: Peak Browser / Enrichment Analysis UI updates the number of experiments by selecting experiment type (2022/03/01)
- Updated the order of the genome assembly tabs, now GRCh38 is default. And some minor fixes came together! (2022/02/08)
- Added **ATAC-Seq and Bisulfite-Seq**, together with UI improvement including 'peak' icon! (2021/10/04)
- Updated site design: (2021/01/04)
 - Dataset search is available as a standalone function
 - ID search at the top right corner now accepts GEO Sample ID (e.g. GSM469863) along with SRA Experiment ID (e.g. SRR018625)
 - New information layout for individual experiment list
- Added new genome assemblies: hg38, mm10, dm6, ce11. Happy holidays! (2020/12/01)

Peak Browser

Visualizes protein binding on given genomic loci with IGV genome browser

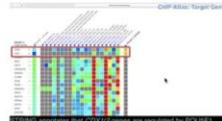
[Watch tutorial movie](#)



Target Genes

Predict target genes bound by given transcription factors

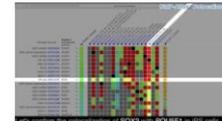
[Watch tutorial movie](#)



Colocalization

Predict partner proteins colocalizing with given transcription factors

[Watch tutorial movie](#)



Enrichment Analysis

Predict proteins bound to given genomic loci and genes

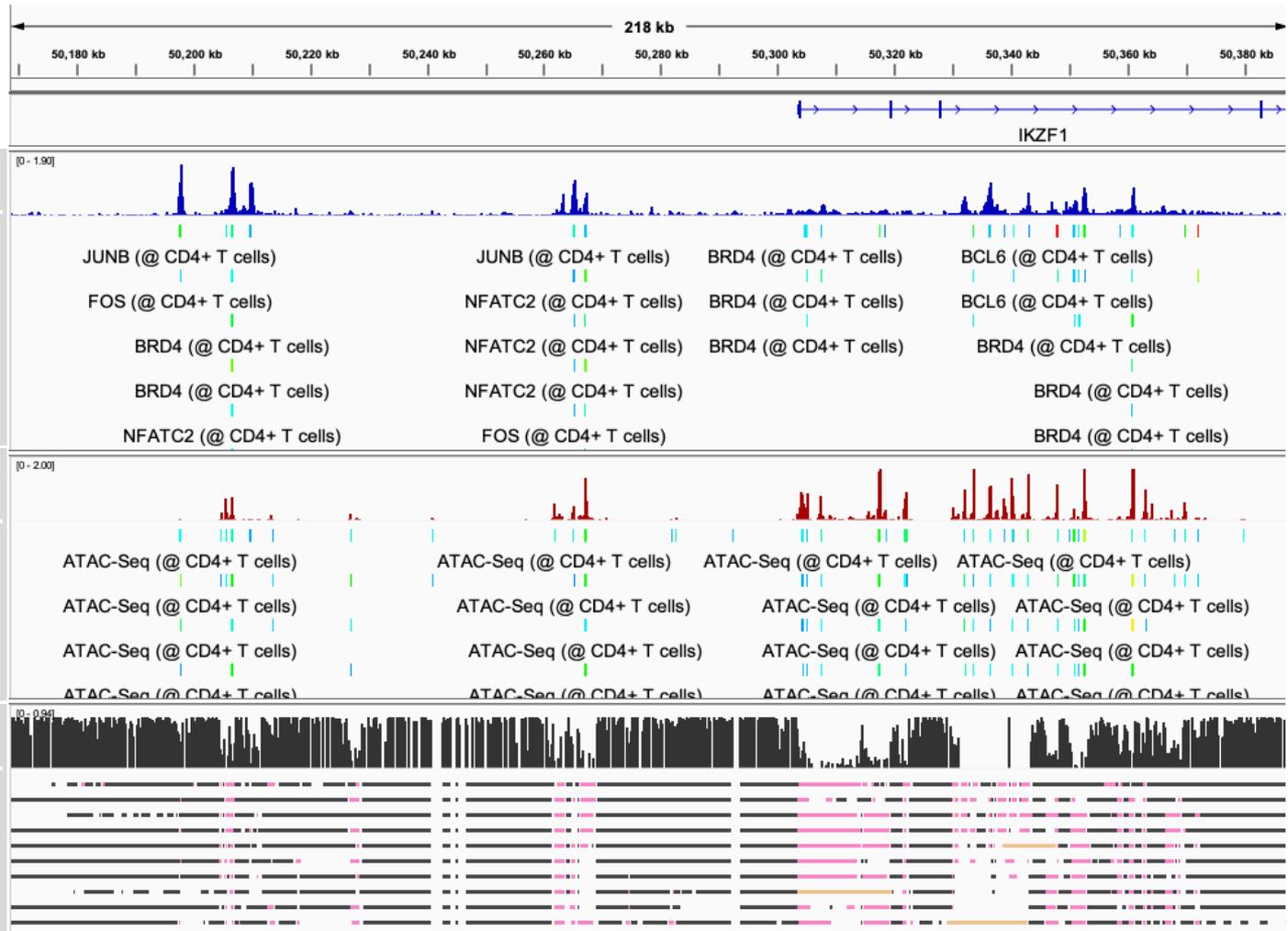
[Watch tutorial movie](#)





THIS WORK IS SUPPORTED BY NIG SUPERCOMPUTER SYSTEM AND NATIONAL BIOSCIENCE DATABASE CENTER.
 NEED HELP? CREATE AN ISSUE ON GITHUB OR CONTACT US

ChIP-Atlas 収録済データ



ChIP-Atlas高度化

新規アノテーションデータの追加

- 発現制御関連エピゲノムデータ (ChromHMM等)
- 疾患関連 SNPs (GWAS)データ
- 量的形質 SNPs (eQTL)データ
- TADs (Topologically Associating Domain) データ
- その他リファレンスデータ (RepeatMasker, phastCons, LD-blocks...)

新規手法データの追加

- 新規ChIP-seq類似手法 (ChIL-seq, CUT&TAG)
- 新規Bisulfite-seq類似手法 (EM-seq, PBAT法)

ChIP-Atlasデータ更新

- SRAデータによる月次更新

ChIP-Atlas (ロードマップ)

更新・拡張

- 新規にデータベースを開発する本提案におけるロードマップや想定されるボトルネックなどをご説明いただきたい。



ChIP-Atlasの月次更新



新規アノテーションデータの追加

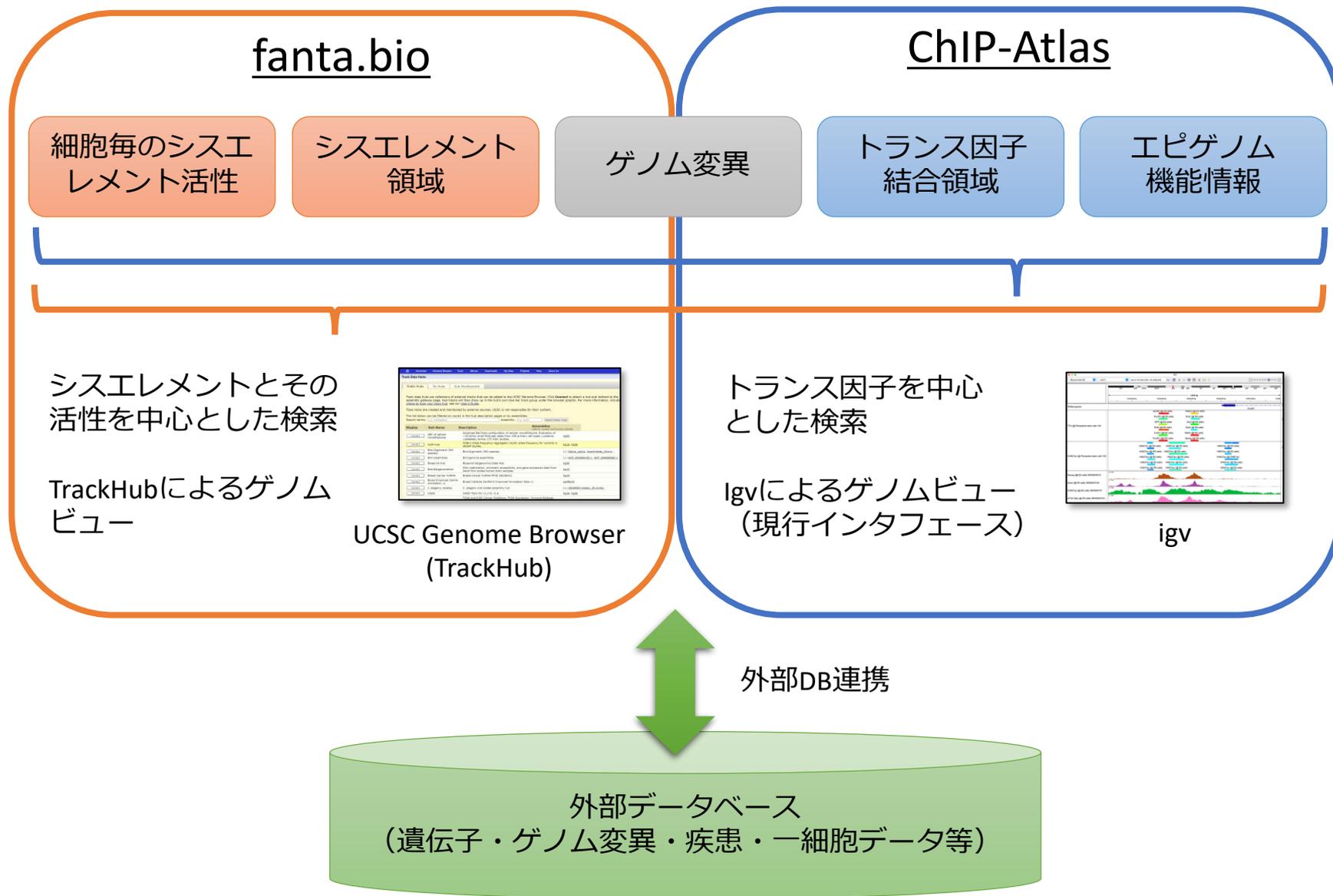


新規手法データの追加



機微データの追加

データ連携

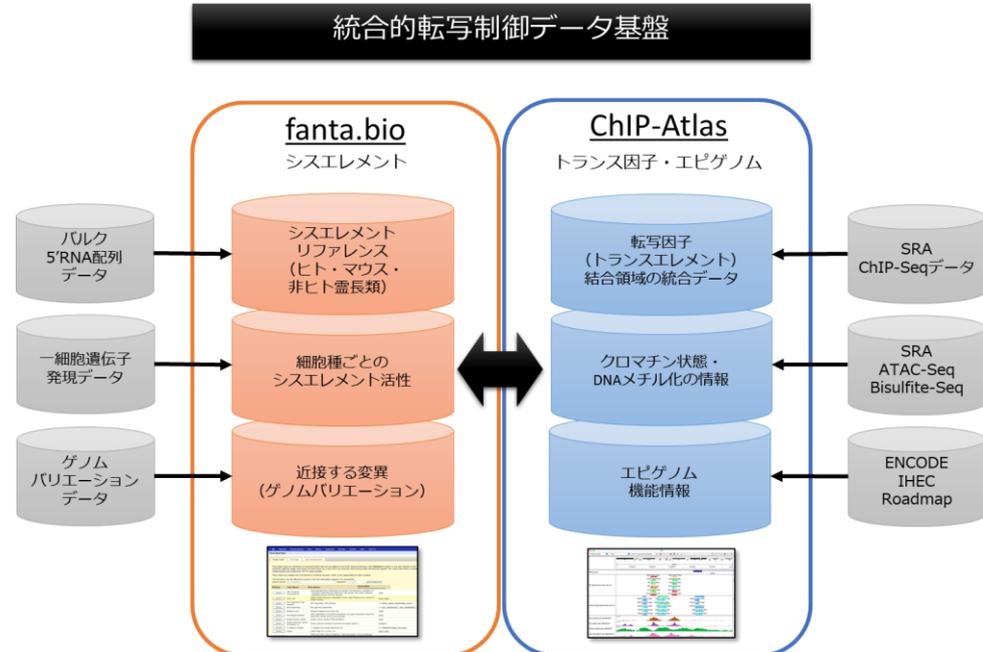


まとめ

- 「統合的転写制御データ基盤」の構築
 - 対象データ：「シスエレメント」「トランス因子」「エピゲノム情報」
 - DB枠組：「既存の ChIP-Atlasの拡張」「新規のシスエレメントデータベース」
- 新規構築のシスエレメントデータベース(fanta.bio)では
 - シスエレメントの位置（静的情報）
 - 細胞ごとのシスエレメント活性（動的情報）
 - シスエレメントと変異の関連（論理的情報）
- ChIP-Atlasの高度化として
 - 新規アノテーションデータの追加
 - 新規手法データの追加
 - データセットの追加
- 今年度中のα版公開を目指す

新規メンバー募集中です！

- 研究員 1 名
- テクニカルスタッフ 1 名



謝辞

- 理化学研究所IMS 生命医科学大容量データ技術研究チーム
 - 研究開発
 - 長谷川哲、Scott Walker、森岡勝樹、Imad Abugessaisa、Nishad Thalhath
 - 計算機インフラ
 - 武田伸幸、北倉輝昭、古川章
- 東京都医学総合研究所
 - 川路英哉
- 理化学研究所BRC 統合情報開発室
 - 柘屋啓志、高田豊行
- 京都大学医学研究科
 - 沖 真弥
- FANTOM Consortium members