

2023年度 研究開発実施報告

概要

研究開発課題名	統合的な転写制御データ基盤の構築
開発対象データベースの名称(URL)	INTRARED (INtegrated TRAnscriptional REgulation Data platform) (https://www.intrared.org/)
研究代表者氏名	粕川 雄也 (10304031)
所属・役職	理化学研究所 生命医科学研究センター チームリーダー (2024年3月時点)



目次

概要	1
目次	2
§1. 研究実施体制	4
§2. 研究開発対象とするデータベース・ツール等	4
(1) データベース一覧	4
(2) ツール等一覧	4
§3. 実施内容	5
(1) 本年度の研究開発計画と達成目標	5
【研究開発項目(1) シスエレメントデータの構築と維持】	5
【研究開発項目(2) ChIP-Atlas 高度化によるエピゲノミクス・トランス因子データの構築と維持】	6
【研究開発項目(3) データベース連携】	7
(2) 進捗状況	7
【研究開発項目(1) シスエレメントデータの構築と維持】	7
【研究開発項目(2) ChIP-Atlas 高度化によるエピゲノミクス・トランス因子データの構築と維持】	9
【研究開発項目(3) データベース連携】	12
§4. 成果発表等	13
(1) 原著論文発表	13
① 論文数概要	13
② 論文詳細情報	13
(2) その他の著作物(総説、書籍など)	14
(3) 国際学会および国内学会発表	14
① 概要	14
② 招待講演	14
③ 口頭講演	15
④ ポスター発表	15
(4) 知的財産権の出願(国内の出願件数のみ公開)	16
出願件数	16
(5) 受賞・報道等	16
① 受賞	16
② メディア報道	16
③ その他の成果発表	16
§5. 主要なデータベースの利活用状況	17
(1) アクセス数	17
① 実績	17
② 分析	17
(2) データベースの利用状況を示すアクセス数以外の指標	17
(3) データベースの利活用により得られた研究成果(生命科学研究への波及効果)	17
(4) データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産業や科	

学技術への波及効果).....	17
§6. 研究開発期間中に主催した活動(ワークショップ等)	18
(1) 進捗ミーティング	18
(2) 主催したワークショップ、シンポジウム、アウトリーチ活動等.....	18

§1. 研究実施体制

グループ名	研究代表者または主たる共同研究者氏名	所属機関・役職名	研究題目
粕川グループ	粕川 雄也	理研IMS・チームリーダー	研究開発の取りまとめ、ヒト・マウスシスエレメントデータの作成、インターフェース開発
川路グループ	川路 英哉	東京都医学総合研究所・副センター長	シスエレメント同定パイプラインの構築と適用
柘屋グループ	柘屋 啓志	理研BRC・室長	ゲノム変異とシスエレメントの対応づけ
沖グループ	沖 真弥	京都大学・特定准教授 (2024年3月時点)	トランス因子・エピゲノミクスデータの収集

§2. 研究開発対象とするデータベース・ツール等

(1) データベース一覧

【主なデータベース】

No.	名称	別称・略称	URL
1	INTRARED		https://www.intrared.org/

【その他のデータベース】

No.	名称	別称・略称	URL
1	fanta.bio		https://fanta.bio/
2	ChIP-Atlas		https://chip-atlas.org/
3	SCPortalen		https://single-cell.riken.jp/
4	MoG+		https://molossinus.brc.riken.jp/mogplus/

(2) ツール等一覧

No.	名称	別称・略称	URL
1			

§3. 実施内容

(1) 本年度の研究開発計画と達成目標

【研究開発項目(1) シスエレメントデータの構築と維持】

研究開発項目(1)-1: シスエレメント領域データの作成

シスエレメント領域データの作成のため、以下の2つを行う。

1. 一細胞の 5'RNA 配列データセットのクラスタリングによる細胞種ごとのデータセット作成

一細胞の 5'RNA 配列データセットを公共リポジトリより取得し、クラスタリングにより細胞種ごとに分割したデータセットを作成するためのパイプラインを実装し、データを作成する。

(当初計画より予定変更を実施) 一細胞データのデータ処理については本年度の検討により、クラスタリング後の後処理、特に細胞腫の同定まで行わないとユーザにとって利便性のあるデータが提供できないことがある点や、正式版リリースを今年度前倒しで実施することにしたため、3年目に計画を移動させた。よって本年度は検討とデータ作成の準備のみを実施することとした。

2. バルクならびに一細胞の5'RNA 端配列データセットによる 8 版シスエレメント領域の同定

バルクの 5'RNA 配列データセットを公共リポジトリより取得し、上記で同定した一細胞由来の細胞種ごと 5'RNA 端配列データセットと組み合わせてシスエレメント領域を同定し、シスエレメント領域の 8 版データを作成する。

研究開発項目(1)-2: 細胞の種類・状態毎のシスエレメント活性データの作成

シスエレメントの活性データの作成のため、以下を行う。

1. シスエレメント活性データの 8 版データを作成

シスエレメント領域データ作成に用いたデータセットを用いて、細胞種ごとの活性データの 8 版を作成する。

研究開発項目(1)-3: シスエレメント近傍のゲノム変異情報の収集と維持

本課題で収集・整理するシスエレメント領域について、疾患モデル動物として用いられる実験動物マウスの各系統における配列の多様性(ゲノムバリエーション)を各シスエレメントのメタ情報としてデータ化する。

1. ゲノム解読データに基づくマウス変異情報の収集とシスエレメントの対応づけ

昨年度整備した 46 系統のマウス短鎖解読由来のゲノムバリエーション情報(SNP や Short-Indel)について、ゲノムバリエーションのデータフォーマットである VCF 化を行った。この成果に基づいて、今年度は、短鎖解読では検出できないゲノム多型が検出できる長鎖解読由来のゲノムバリエーション情報(約 5 系統)のデータ整備化と可視化を進める。特に今年度は、すでに行った VCF 形式を基盤として、変異情報を記述するととも

に、長鎖解析由来のゲノムバリエーション情報(SVなど)のVCF化における問題点と、バリエーション記述とシスエレメントとの対応づけに関する課題を明確化する。

2. 公共データベースや文献に基づくマウス変異情報およびシスエレメント情報の収集

上記に加え、実験的に確定されたマウスの変異およびシスエレメントの情報について、マウスゲノム統合データベースである MGI、表現型統合データベースである IMPC 等から収集する。さらに、MGI にリストされている論文についてキュレーションを行い、未だデータベース化されていない変異及びシスエレメントについて、文献からの情報収集を行う。収集された情報は VCF 化を行う。これに加え、MGI および IMPC でデータベース化されている変異がもたらす表現型について、VCF 化を通じたゲノムブラウザ上での可視化ができないかを試みる。

3. 対応データ、メタデータの作成

VCF で整備したゲノムバリエーションデータに関して、DBCLS、NBDC、遺伝研が策定を開始している VCF の RDF 版である形式の作成を行い、他のデータベースとのデータの相互乗り入れを試みる。ヒトの公開データベースからゲノムバリエーション情報を収集し、ヒトとマウスで可視化による比較ができるか調査を行う。

4. 運用するデータベースのミドルウェアおよび使用ライブラリの保守点検および更新

上記により得られる結果をプロジェクト継続期間に安定して提供するため、運用するデータベースのミドルウェアおよび使用ライブラリの保守点検および更新を行う。

【研究開発項目(2) ChIP-Atlas 高度化によるエピゲノミクス・トランス因子データの構築と維持】

研究開発項目(2)-1 エピゲノミクス・トランス因子データの新規コンテンツの追加

前年度に続き、下記の機能注釈データを準備する。

1. 発現制御に関わる機能性領域(ChromHMM, FANTOM enhancer)

最近報告された 100 種類の細胞タイプの ChromHMM データ(H. Vu, & J. Ernst. Genome Biol. 2022) を入手し、Strong Enhancer, Heterochromatin など 15 種類の分類群をゲノムブラウザ(IGV)で色分け表示できるような BED ファイルを作成する。FANTOM enhancer データは、FANTOM5 プロジェクトの Web サイトから入手し、IGV 上では細胞名とエンハンサーの強度に応じて濃淡表示できるような BED ファイルを作成する。

2. 疾患関連 SNPs(GWAS Catalog, ClinVar)

GWAS Catalog と ClinVar より種々の形質や疾患と関連する SNP 座位を収集し、IGV 上では形質名と P-value に応じて濃淡表示できるような BED ファイルを作成する。

3. 量的形質 SNPs(eQTL)

GTE_x より遺伝子の発現変動と関連する SNP 座位を収集し、IGV 上では P-value に応じて濃淡表示できるような BED ファイルを細胞名ごとに作成する。

研究開発項目(2)-2 ChIP-Atlas データ更新

ChIP-Atlas は最近のアップデートで Bisulfite-seq と ATAC-seq データを追加したが、その計算に多大な時間とリソースを費やしたため、2020 年 8 月以降のデータに対応できていなかった。前年度は 2022 年 6 月までのデータ更新を終えた。今年度はさらにその遅れを取り戻し、毎月、最新データを公開できるように進める。おもな実施項目は下記の通り。

- 1) SRA メタデータのダウンロードと更新データのチェック
- 2) 一次解析(シーケンスデータのダウンロード、マッピング、ピークコール)
- 3) サンプルメタデータのキュレーション
- 4) データ統合、データマイニング結果の作成
- 5) 公開サーバへのデータ転送

【研究開発項目(3) データベース連携】

前年度に引き続き fanta.bio で用意するシスエレメントデータならびに ChIP-Atlas で用意するトランス因子のデータを一括表示するための手法検討と実装を行う。

(2) 進捗状況

【研究開発項目(1) シスエレメントデータの構築と維持】

研究開発項目(1)-1: シスエレメント領域データの作成

昨年度使用した NET-CAGE 法論文(Hirabayashi S et al., 2019. <https://doi.org/10.1038/s41588-019-0485-9>)で公開された NET-CAGE/CAGE データに加えて、BioProject/GEO/SRA より抽出した CAGE を中心とした 5'-end RNA-seq 由来の公開データセット、FANTOM プロジェクトから公開された FANTOM5 CAGE、FANTOM6 CAGE データを利用して、新たに CRE 領域の同定を行った。その結果、ヒト (GRCh38/hg38)で 447,315 個、マウス(GRCm38/mm10)で 201,420 個の CRE 領域を同定した。

さらに同定した CRE 領域については、その特性や機能の推定につながるアノテーション情報の付与を行った。具体的には、近傍の転写開始点情報との対応づけによるプロモーター名の付与、FANTOM、ENCODE SCREEN、refTSS 等の既知のエンハンサー領域、プロモーター領域との対応づけ、ChIP-Atlas で提供されている ChIP-seq データ由来の転写因子結合領域との対応づけを行った。これらの結果についてまとめた統計値は以下の表のとおりである。

表1: 同定した CRE 領域とそのアノテーション情報の統計値

Items	Human	Mouse
fanta.bio CREs	447,316	201,421
Promoter annotation	118,200	65,946
Overlapped FANTOM5 CAGE Peaks (human: 209,911, mouse: 164,672)	95,142	62,109
Overlapped FANTOM5 Enhancers (human: 63,285, mouse: 49,797)	47,636	25,445
Overlapped ENCODE SCREEN cCRE (human: 926,536, mouse: 339,815)	312,056	103,790
Overlapped refTSS TSSs (human: 241,049)	103,772	(in prep.)
Overlapped ChIP-Atlas TF and Others binding sites (peak cutoff: Q-score < 500)	384,591	157,735

研究開発項目(1)-2: 細胞の種類・状態毎のシスエレメント活性データの作成

研究開発項目(1)-1 で同定した各 CRE 領域について活性量を示すデータを作成した。活性量の定量に用いたデータは NET-CAGE 法論文の掲載 CAGE/NET-CAGE データ、FANTOM5 CAGE データ、FANTOM6 CAGE データのうちノックダウンを伴わないコントロール条件のものうち代表サンプルを用いた。まず、これらの元サンプルの情報を提供元より取得し、Primary cell 由来のものについて Cell Type 情報を記述するための Cell Ontology との対応づけ、Cell line 由来のものについて、Cellosaurus (<https://www.cellosaurus.org/>)との対応づけ、Tissue 由来のものについては Uberon オントロジーへの対応づけを行い、分類を行った。次にそれぞれのシーケンスデータのゲノムマッピング結果から各 CRE にマッピングされたタグ数をカウントし、最後にサンプル間の差を補正するための正規化を行った活性量テーブルのデータを作成した。最終的にヒトは 1,911 サンプル、マウスは 1,051 サンプルの活性量データとなった。

これら研究開発項目(1)-1、研究開発項目(1)-2 で作成した CRE データは昨年度構築した fanta.bio データベース (<https://fanta.bio/>) に格納し正式版(version 1.0.0)として 2024 年 2 月 26 日に公開した。本データベースでは個々の CRE についての詳細情報やアノテーション情報を検索・参照するための in house のインターフェースと、TrackHub 経由で UCSC Genome Browser 上で表示し他のゲノム領域情報と比較可能とするような仕組みを提供している。また in house インターフェースでは CRE 領域内に存在する、ChIP-Atlas で提供されている転写因子結合領域や、TogoVar (<https://togovar.org/>) ならびに MoG+で提供されているゲノム変異情報も参照できるようになっている。また、UCSC Genome Browser でもこれらの情報が参照可能となっており、さらに ChIP-Atlas については Cell line ごとに抽出して表示させることも可能となっている。

研究開発項目(1)-3 シスエレメント近傍のゲノム変異情報の収集と維持

1. ゲノム解読データに基づくマウス変異情報の収集とシスエレメントの対応づけ

昨年度整備したマウス 46 系統の VCF を Fanta.bio (<https://fanta.bio/>)から公開するため、研究グループと関連データの共有作業を行った。この結果、2023 年度に、短鎖解読由来の 46 種のゲノムバリエーション情

報をfanta.bioに搭載して公開し、シスエレメント情報と、マウス多型情報をブラウザ上で一括して表示し、ユーザが自分の興味に従ってゲノムの機能領域を探索することが可能になった。長鎖配列解読由来ゲノム多型データの整備については、別プロジェクトで進行中の5系統の解析結果をfanta.bioに搭載するための策定を進め、短鎖解読では検出できない長鎖解読由来のゲノムバリエーションデータの整備と可視化のための検討作業を進めた。特にVCF形式を用いて、50bp以上の構造多型(SV)を記述する方法の検討を行い、問題点の明確化するとともに、ゲノム多型情報と、シスエレメントとの対応づけに関する課題を検討した。データ属性(由来する生物、組織、細胞株およびその表現型情)をいかに記述するかの方針についても継続して進めた。

2. 公共データベースや文献に基づくマウス変異情報およびシスエレメント情報の収集

上記に加え、実験的に確定されたマウスのシスエレメントおよびその変異ともたらされた表現型について、Mouse Genome Informatics (MGI: <https://www.informatics.jax.org/>)、International Mouse Phenotyping Consortium (IMPC: <https://www.mousephenotype.org/>)等、マウス研究分野において、国際的に主要な役割を果たすゲノムおよび表現型のデータベースや公開論文情報から、ゲノム座標およびゲノム配列データを抽出するための調査を行い、データ収集を開始した。2023年度は、MGIにリストされている21,826件の論文を対象にして、未だデータベース化されていない変異及びシスエレメントについて、まずはシス制御領域に表現関連情報のある200件程度の論文を対象にしたキュレーションを行った。さらに、MGIの関係者と連携して、MGIが保有する変異アレルのゲノム位置のアノテーション情報(例、<https://www.alliancegenome.org/downloads/>)について内容の確認を開始するとともに、各種方法により収集した情報について、VCF化を通じたゲノムブラウザ上での可視化手法の検討を行った。

3. 対応データ、メタデータの作成

ゲノムバリエーションデータに関して、RDFによる他のデータベースとのデータの相互乗り入れを試みた。日本人ゲノム多様性統合データベース TogoVar (<https://togovar.org/>)との連携により、ヒト・マウスで同一のSNVを持つ例を探索しCDS上のPathogenicなヒト変異約13万座位を探索し、のべ207の同一SNVが野生由来系統を中心に存在することを確認した。また、利用可能なシス領域のゲノム多型情報の収集法の検討も行なった。これらに関して、一部の成果を研究会等で報告した。

4. 運用するデータベースのミドルウェアおよび使用ライブラリの保守点検および更新

上記により得られる結果をプロジェクト継続期間に安定して提供するため、使用ライブラリの保守点検および脆弱性の監視を常時行った。これ以外に、当該年度内にミドルウェアのアップデート3回(Apache Tomcat)、JavaErrorの修正1回を行った。

【研究開発項目(2) ChIP-Atlas 高度化によるエピゲノミクス・トランス因子データの構築と維持】

研究開発項目(2)-1 エピゲノミクス・トランス因子データの新規コンテンツの追加

1. アノテーションデータの追加

UCSC goldenPathを含む外部サーバから取得した、6つの生物を網羅する1,000以上のゲノムおよびエ

ピゲノム関連トラックを組み込んだ(表2)。具体的には下表のように、染色体コンフォメーション(ENCODE Hi-C、GTEx eQTL)、転写制御エレメント(ChromHMM、FANTOM5 エンハンサー、JASPAR TF motif)、ゲノムワイド関連研究(GWAS)で同定された SNP(GWAS Catalog)や遺伝子バリエーション(ClinVar)、遺伝子疾患/表現型関連データ(Orphanet および MGI Phenotype)、進化的に保存された領域(PhastCons)、反復配列(RepeatMasker)、RNA-seq に基づくトランスクリプトーム(ENCODE、GTEx、FlyAtlas2、modENCODE)、遺伝子モデル(Ensembl、GENCODE)、ENCODE Blacklist、CpG アイランドのデータを収集・加工し、公開サーバにデータを転送した。ChIP-Atlas ウェブサイトの Peak Browser の UI に「Annotation tracks」という項目を追加し、そこから上記のデータをさらに絞り込めるような UI を千葉大学の 大田達郎氏とともに作成した。

表2:外部サーバから取得したエピゲノム関連アノテーションデータ

Genome	hg38	hg19	mm10	mm9	rn6	dm6	dm3	ce11	ce10	sacCer3
ENCODE Hi-C	○	○	○	○						
GTEx eQTL	○	○								
ChromHMM	○	○	○	○						
FANTOM5 enhancers	○	○								
JASPAR TF motif	○	○	○	○		○	○	○	○	○
GWAS Catalog	○	○								
ClinVar	○	○								
Orphanet	○	○								
MGI Phenotype			○	○						
PhastCons	○	○	○	○	○	○	○	○	○	○
RepeatMasker	○	○	○	○	○	○	○	○	○	○
RNA-seq	○	○	○	○		○	○			
Ensembl genes	○	○	○	○	○	○	○	○	○	○
GENCODE genes	○	○	○	○						
ENCODE Blacklist	○	○	○	○		○	○	○	○	
CpG Islands	○	○	○	○	○	○	○	○	○	

2. Diff Analysis ツールの追加

ChIP-seq データセットから異なる結合領域 (DBR) や ATAC-seq および DNase-seq データセットから異なるアクセス可能領域 (DAR) を検出するための Diff Analysis ツールを開発した。本ツールの計算アルゴリズムは R パッケージ「DiffBind」を参考としているが、DiffBind は比較分析を行うために BAM ファイルを入力として必要とする。ChIP-Atlas サーバには BAM ファイルがないため、ChIP-seq、ATAC-seq、または DNase-seq ピーク内のアライメントリードをカウントするアルゴリズムを部分的に変更した。まず、ユーザのクエリーSRXに応じて、RPM (Reads Per Million) を示す BigWig フォーマットのアライメントデータを BedGraph 形式に変換する。次に、SRX のマッピングシーケンスリードの総数に関して RPM を整数値に変換する。これとは別にクエリーSRX からのピークコールデータに基づいて全ゲノムをフラグメント化しておく。各ゲノムフラグメントとアライメントの整数値を集計し、結果を $m \times n$ 行列として格納する (m はゲノムフラグメントの数、 n はクエリーSRX の数)。この行列を R パッケージ「edgeR」に入力し、標準アルゴリズムを使用して各ゲノムフラグメントに対する 2 つのクエリーSRX セット間のリードカウントの違いを評価する。最後にその結果を BED 形式で出力する。Bisulfite-seq データセットから異なるメチル化領域 (DMR) を検出する場合でも、まず各クエリーSRX のメチル化率を含む BigWig を BedGraph に変換する。これを DMR 解析ツール「metilene」に入力し、DMR と平均メチル化差や Q 値などの統計情報を含む BED 形式の結果を出力する。千葉大学の 大田達

郎氏とともに、Web ブラウザで入力された SRX を遺伝研スパコンに送信し、対応するデータの BigWig を呼び起こして上記の解析を行うための UI と API を作成した。

3. マニュアルの刷新

ChIP-Atlas のマニュアルは動画としてこれまで配信していた。しかし、動画では知りたい箇所をじっくり見たりすることができないため、必ずしもユーザにとって利便性が良いわけではなく、なおかつ更新作業に手間がかかるため 2016 年以來更新されていなかった。そこですべてのツールについて PDF 形式のマニュアルを作成した。スクリーンショットを豊富に用いた紙芝居形式であるため、じっさいの画面をページめくりで進行できるのが特長である。どのマニュアルも「要約→詳細→Tips」の順序で進行するため、とりあえず何ができるかを理解し、そのうえで詳細を十分に習得し、最後に便利な小技などを知ることができる。また、マニュアルページとは別に Docs ページも設けており、そこに「Annotation tracks」と「Diff Analysis」ツールのデータ収集や計算法などを追記した。

4. UI の刷新

ChIP-Atlas の既存ツール (Peak Browser, Target Genes, Colocalization, Enrichment Analysis) に、「Diff Analysis」ツールが加わり、さらにデータ検索のための「Dataset Search」もツールとして独立させ、計 6 ツールをアイコンとともにトップページに配置させた (千葉大学の 大田達郎氏との共同開発)。

研究開発項目(2)-2 ChIP-Atlas データ更新

2023 年度において ChIP-seq, WGBS, ATAC-seq, DNase-seq データ合わせて 37,124 件を ChIP-Atlas に追加し、398,889 件の実験データ (SRX 換算) を公開した。おもな実施項目は下記の通り。

1) 一次解析

- ・シーケンスデータのダウンロード
- ・レファレンスゲノムへのアライメント
- ・ピークコールデータの作成 (BED 形式; WGBS は高/低メチル化領域)
- ・カバレッジデータの作成 (BigWig 形式; WGBS はメチル化率データ)

2) サンプルメタデータのキュレーション

- ・既存データを学習セットとした機械学習モデルを作成
- ・新規 SRA metadata の入手、機械学習による予測
- ・人員によるチェック
- ・予測できなかったデータは文献などを参照し、人員によるキュレーション

3) 統合解析

- ・全てのピークコールデータの結合、ソート
- ・Peak Browser ツール用に、IGV 閲覧のための加工データとインデクス作成
- ・Target Genes ツール用に、各転写因子が結合する遺伝子の特定
- ・Colocalization ツール用に、共起的に結合する転写因子の特定
- ・Enrichment Analysis ツール用のライブラリファイルの作成

4) データ公開

- ・ 公開サーバへのアップロード

【研究開発項目(3) データベース連携】

fanta.bio の CRE データと ChIP-Atlas のデータを相互参照できるようにするため、今年度は両データベースで使用したサンプルのうち Cell line 情報について対応づけをキュレーションベースで行った。具体的には両者のサンプルを Cellosaurus に対応づけしたのち、共通の Cellosaurus エントリと対応づけられたサンプル間を同じ Cell line 由来のサンプルとした。これらの情報は fanta.bio のインターフェース等で用いている。

§4. 成果発表等

(1) 原著論文発表

① 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	1件
	国際(欧文)	8件
未発行論文 (accepted, in press 等)	国内(和文)	0件
	国際(欧文)	0件

② 論文詳細情報

1. Zhao Jin, Yoko Sato, Masayuki Kawashima, Minoru Kanehisa. KEGG tools for classification and analysis of viral proteins, *Protein Science*, vol. 32, No. 12, e4820, 2023 (DOI: 10.1002/pro.4820).
2. Mengying Cui, Koji Yamano, Kenichi Yamamoto, Hitomi Yamamoto-Imoto, Satoshi Minami, Takeshi Yamamoto, Sho Matsui, Tatsuya Kaminishi, Takayuki Shima, Monami Ogura, Megumi Tsuchiya, Kohei Nishino, Brian T. Layden, Hisakazu Kato, Hidesato Ogawa, Shinya Oki, Yukinori Okada, Yoshitaka Isaka, Hidetaka Kosako, Noriyuki Matsuda, Tamotsu Yoshimori, Shuhei Nakamura. HKDC1, a target of TFEB, is essential to maintain both mitochondrial and lysosomal homeostasis, preventing cellular senescence, *Proceedings of the National Academy of Sciences*, vol. 121, No. 2, 2024 (DOI: 10.1073/pnas.2306454120).
3. Zhaonan Zou, Yuka Yoshimura, Yoshihiro Yamanishi, Shinya Oki. Elucidating disease-associated mechanisms triggered by pollutants via the epigenetic landscape using large-scale ChIP-Seq data, *Epigenetics & Chromatin*, vol. 16, No. 1, 34, 2023 (DOI: 10.1186/s13072-023-00510-w).
4. Yuta Ueno, Emiko Yoshida, Shuko Nojiri, Tomoyasu Kato, Takashi Ohtsu, Toshiyuki Takeshita, Shunji Suzuki, Hiroshi Yoshida, Ken Kato, Masayoshi Itoh, Tsuguto Notomi, Kengo Usui, Takashi Sozu, Yasuhisa Terao, Hideya Kawaji, Hisamori Kato. Use of clinical variables for preoperative prediction of lymph node metastasis in endometrial cancer., *Japanese journal of clinical oncology*, vol. 54, No. 1, 38-46, 2024 (DOI: 10.1093/jco/hyad135).
5. Takuya Kuroda, Satoshi Yasuda, Satoko Matsuyama, Takumi Miura, Rumi Sawada, Akifumi Matsuyama, Yumiko Yamamoto, Masaki Suimye Morioka, Hideya Kawaji, Takeya Kasukawa, Masayoshi Itoh, Hidenori Akutsu, Jun Kawai, Yoji Sato. ROR2 expression predicts human induced pluripotent stem cell differentiation into neural stem/progenitor cells and GABAergic neurons., *Scientific reports*, vol. 14, No. 1, 690-690, 2024 (DOI: 10.1038/s41598-023-51082-4).
6. Akari Nakamura, Seiichiro Sakai, Yoshitaka Taketomi, Jun Tsuyama, Yoshimi Miki, Yuichiro Hara, Nobutaka Arai, Yuki Sugiura, Hideya Kawaji, Makoto Murakami, Takashi Shichita. PLA2G2E-mediated lipid metabolism triggers brain-autonomous neural repair after ischemic stroke., *Neuron*, vol. 111, No. 19, 2995-3010, 2023 (DOI: 10.1016/j.neuron.2023.06.024).
7. Takumi Miura, Tsukasa Kouno, Megumi Takano, Takuya Kuroda, Yumiko Yamamoto, Shinji Kusakawa, Masaki Suimye Morioka, Tohru Sugawara, Takamasa Hirai, Satoshi Yasuda, Rumi Sawada, Satoko Matsuyama, Hideya Kawaji, Takeya Kasukawa, Masayoshi Itoh, Akifumi Matsuyama, Jay W Shin, Akihiro Umezawa, Jun Kawai, Yoji Sato.

Single-Cell RNA-Seq Reveals *LRRC75A*-Expressing Cell Population Involved in VEGF Secretion of Multipotent Mesenchymal Stromal/Stem Cells Under Ischemia, *Stem Cells Translational Medicine*, vol. 12, No. 6, 379-390, 2023 (DOI: 10.1093/stcltm/szad029).

- Hideya Kawaji. Unraveling the genomic code in gene regulation, TMIMS International Symposium on Cells and Chromosomes, 2023.

(2) その他の著作物(総説、書籍など)

- 鄒 兆南, 沖 真弥. 「ChIP-Atlas ～転写制御ランドスケープの歩き方～」, *JSBi Bioinformatics Review*, vol. 4, No. 1, 1-9, 2023 (DOI: 10.11234/jsbibr.2023.primr1).

(3) 国際学会および国内学会発表

① 概要

種別	国内外	件数
招待講演	国内	14 件
	国際	0 件
口頭発表	国内	4 件
	国際	0 件
ポスター発表	国内	8 件
	国際	2 件

② 招待講演

〈国内〉

- 沖 真弥. 空間的な遺伝子制御機構の解明とその解析支援. 第 20 回生命資源研究・支援センターシンポジウム. 2024/2/22.
- 沖 真弥. ChIP-Atlas ハンズオンセミナー. 第2回エビ研リソースセミナー. 2023/12/11.
- 沖 真弥. 空間的な遺伝子発現制御のしくみを探る. 高知大学医学専攻特別研究ゼミナール. 2023/11/27.
- 沖 真弥. ChIP-Atlas の使いかたと使われかた. 第 96 回日本生化学会大会. 福岡市, 2023/11/2.
- Scott Walker, 長谷川哲, Nishad Thalath, 高田豊行, 榊屋啓志, 川路英哉, 粕川雄也, 転写制御理解のためのゲノムワイドなシスエレメントデータベース fanta.bio, 第 96 回日本生化学会大会, 福岡市, 2023/11/1-3
- 粕川雄也, 転写制御理解のための統合的データ基盤の開発, 2023 年度理研ハッカソン 公開シンポジウム, 2023/10/24
- 川路英哉, Unraveling the genomic code in gene regulation, 第 25 回都医学研国際シンポジウム, 東京, 2023/10/16
- 粕川雄也, “統合的な転写制御データ基盤 INTRARED の構築について”, トーゴの日シンポジウム 2023, 東京, 2023/10/5
- 沖 真弥. 空間的な遺伝子発現制御のしくみを探る. 第 2 回 Philosophy 研究会. 2023/9/16.
- 沖 真弥. Photo-isolation chemistry による局所的高深度トランスクリプトーム解析. 第 9 回日本筋

学会学術集会. 2023/8/19.

11. Shinya Oki. Photo-isolation chemistry for high-resolution and deep spatial transcriptome with tissue sections. The 56th Annual Meeting of the Japanese Society of Developmental Biologists. 2023/7/25.
12. 沖 真弥. 空間的な遺伝子発現制御のしくみを探る. 学術変革領域研究(A)適応回路センサ第 3 回領域会議. 2023/6/22.
13. 沖 真弥. エピゲノム統合データベース、ChIP-Atlas の使い方と使われ方. 第 16 回日本エピジェネティクス研究会年会. 2023/6/19.
14. 沖 真弥. バイオのための情報学・ことはじめ. ACT-X「生命と化学」主催:第 1 回 研究者交流会議. 2023/5/10.

〈国際〉

該当なし

③ 口頭講演

〈国内〉

1. 高田豊行. マウス日本産疾患モデル系統のゲノム解析. 第 33 回モロシヌス研究会, 篠山, 2023/11/22-23
2. 高田豊行, 宮澤秀幸, 豊田 敦, 天野 孝紀, 田村 勝, 吉木 淳, 阿部訓也, 城石俊彦, 野口英樹, 榎屋啓志. 日本産マウス疾患モデル系統の長鎖解読を利用したゲノム解析. 日本遺伝学会第 95 回大会, 熊本市, 2023/9/6-8
3. 高田豊行, 臼田大輝, 榎田達矢, 城石俊彦, 榎屋啓志. マウスゲノム多型に基づいた遺伝子機能解析を支援する MoG+のアップデート. 第 70 回日本実験動物学会総会, つくば市, 2023/5/24-26
4. 榎田達矢, 臼田大輝, 高田豊行, 榎屋啓志. BRC バイオリソース検索システムにおける疾患検索機能の強化. 第 70 回日本実験動物学会総会, つくば市, 2023/5/24-26

〈国際〉

該当なし

④ ポスター発表

〈国内〉

1. 粕川雄也, Scott Walker, 長谷川哲, Nishad Thalath, 信定知江, 沖真弥, 高田豊行, 榎屋啓志, 川路英哉, INTRARED: シスエレメント、トランス因子、エピゲノム情報を統合した転写制御データプラットフォーム, 第 46 回分子生物学会年会, 神戸市, 2023/12/6-8
2. 齊藤紗希, 和田涼子, 川路英哉, ヒト肝癌細胞における薬剤誘導性エンハンサーRNA の機能解析, 第 46 回分子生物学会年会, 神戸市, 2023/12/6-8
3. 吉沢直子, 原雄一郎, 和田涼子, 夏目豊彰, 川路英哉, 転写終結点近傍領域を対象とした、転写開始制御活性の網羅的解析と標的領域の検証, 第 46 回分子生物学会年会, 神戸市, 2023/12/6-8
4. 高田豊行, 臼田大輝, 榎田達矢, 城石俊彦, 榎屋啓志. マウスゲノム多型データベース MoG+の機能拡張. 第 46 回分子生物学会年会, 神戸市, 2023/12/6-8

5. 櫛田達矢, 臼田大輝, 高田豊行, 山縣友紀, 榊屋啓志. ヒト・哺乳類表現型オントロジーマッピングを活用したヒト疾患モデル生物の推定方法の検討. 第 46 回日本分子生物学会年会, 神戸市, 2023/12/6-8
6. 齊藤紗希, 和田涼子, 夏目豊彰, 川路英哉, "Genome-wide profiling of enhancer RNA initiation events identified novel drug-responsive enhancers in human hepatocytes", 第 23 回 シトクロム P450 国際会議, 静岡市, 2023/9/25
7. 吉沢直子, 原雄一郎, 和田涼子, 夏目豊彰, 川路英哉, "Investigation of cis-regulatory activities in proximal regions to transcription termination.", 第 12 回生命医薬情報学連合大会, 柏市, 2023/09/07
8. 臼田大輝, 櫛田達矢, 高田豊行, 榊屋啓志. 理研 BRC におけるバイオリソース検索システムの運用と機能強化. 第 70 回日本実験動物学会総会, つくば市, 2023/5/24-26

〈国際〉

1. Toyoyuki Takada, Daiki Usuda, Tatsuya Kushida, Toshihiko shiroishi, Takeya Kasukawa, Hiroshi Masuya. Update of MoG+: a database of mouse genomic variations connecting bioresource and biomedical research. The Allied Genetics Conference (TAGC), Washington DC, United States, March 6–10 (2024)
2. Hiroshi Masuya, Toyoyuki Takada, Daiki Usuda, Naomi Yuhara, Keiko Kurihara, Yuri Namiki, Shigeru Iwase, Kenta Suzuki, Yuki Yamagata, Takeya Kasukawa, Hideya Kawaji, Shinya Oki, Tatsuya Kushida. Online databases and data integration in the RIKEN BioResource Research Center. The Allied Genetics Conference (TAGC), Washington D.C., Japan, March 6–10 (2024)

(4) 知的財産権の出願（国内の出願件数のみ公開）

出願件数

種別		件数
特許出願	国内	0 件

(5) 受賞・報道等

① 受賞

該当なし

② メディア報道

1. (ニュースリリース “遺伝子の転写制御に関わるゲノム領域、シスエレメントの情報を収載したデータベース「fanta.bio」が公開されました”, <https://biosciencedbc.jp/news/20240229-01.html>)

③ その他の成果発表

該当なし

§5. 主要なデータベースの利活用状況

(1) アクセス数

① 実績

表 1 研究開発対象の主要なデータベースの利用状況（月平均）

DB 名	種別	2023 年度
INTRARED	訪問者数	---
	訪問数	---
	閲覧ページ数	---
fanta.bio	訪問者数	---
	訪問数	---
	閲覧ページ数	---
ChIP-Atlas	訪問者数	---
	訪問数	---
	閲覧ページ数	---

② 分析

INTRARED, fanta.bio については、ウェブサイトを独自のサーバ内に構築・公開するのではなく、PasS 型のクラウド環境である CloudFlare 上で構築・公開しており、awstats を用いたログ解析を行うことはできない。そこで、Google Analytics 4 を用いたアクセス情報取得の仕組みを現在構築しており、2024 年度より利用可能となる予定である。よって本年度のアクセス数についての評価はできない。

2023 年度は ChIP-Atlas の利用状況が例年より極端に少なく、awstats の集計がうまくできていないと思われる。

以上の理由により、本研究課題で対象とするデータベースの今年度のアクセス数については評価できなかった。

(2) データベースの利用状況を示すアクセス数以外の指標

なし

(3) データベースの利活用により得られた研究成果(生命科学研究への波及効果)

2023 年に ChIP-Atlas を利用した論文が 220 報あった(Google Scholar による全文検索)。

(4) データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産業や科学技術への波及効果)

該当なし

§6. 研究開発期間中に主催した活動(ワークショップ等)

(1) 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2023年 9月8日	チーム内ミーティング(非公開)	オンライン	4人	研究進捗の共有ならびにサイトビジットの準備
2023年 9月21日	サイトビジット	理化学研究所横浜事業所	12人	NBDCによるサイトビジット
2023年 9月25日～ 9月29日	チーム内ミーティング(非公開)	和歌山	5人	DBCLS主催第14回国内版バイオハッカソンにおいて、進捗報告と、表現型情報の記述など、今後の進め方について確認した

(2) 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・概要

以上

別紙1 既公開のデータベース・ウェブツール等

No.	正式名称	別称・略称	概要	URL	公開日	状態	分類	関連論文
1	INTRARED		転写制御の総合的理解の推進を目的に、シスエレメント・トランス因子・エピゲノミクスデータを統合したデータベース。本データベースは fanta.bio と ChIP-Atlas の2つのデータベースで構成される。	https://www.intrared.org/	2022/11/29	維持・発展	データベース等	
2	ChIP-Atlas		SRAとして公開されている全てのChIP-seq, DNase-seq, ATAC-seq, Bisulfite-seqデータを統合したデータベースであり、またそれらを簡単に活用するためのWebサービスである。	https://chip-atlas.org/	2015/12/1	維持・発展	データベース等	Oki, et al., (2018) EMBO Rep Zou, et al., (2022) Nucleic Acids Res
3	fanta.bio		ヒト・マウス・非ヒト霊長類を対象に、ゲノム中の転写調節に関わるシスエレメント (cis-regulatory element / CRE) について、ゲノム中の位置、様々な細胞種や細胞状態における活性、シスエレメントと関連する変異についてまとめたデータベース	https://fanta.bio/	2023/3/27	維持・発展	データベース等	