

ライフサイエンスデータベース統合推進事業  
(統合化推進プログラム 2017年採択課題)

## 研究開発終了報告書

「エピゲノミクス統合データベースの開発と機能拡充」

沖 真弥 | 京都大学 大学院医学研究科 特定准教授

研究開発期間： 2017年4月～2022年3月



©2022 沖 真弥(九州大学) Licensed under CC BY 4.0

## §1. 研究開発実施の概要

ChIP-Atlas は、論文などで報告された世界中の ChIP-seq データを全て統合したデータベースであり、さらにその膨大なデータの利活用や統合解析をサポートするウェブツールとして 2015 年 12 月に公開された。

本報告書の研究開発は、Bisulfite-seq データを追加することで ChIP-Atlas をエピゲノミクス統合データベースとして拡充し、なおかつ持続可能性と利活用実績の拡大を目的として 2017 年 4 月に発足した。以来5年間の研究開発により、まず Bisulfite-seq データを計算処理するための解析パイプラインを構築した。これにより所与のゲノム領域におけるメチル化率と高/低メチル化領域がさまざまな細胞タイプに対して可視化された。また ATAC-seq データの解析も追加課題として遂行し、オープンクロマチン領域の可視化も実現された。最新のゲノムアセンブリへの対応やラットゲノムデータの追加により、ユーザの利便性が大幅に向上された。ChIP-Atlas のデータを多角的に活用するため、jPOST や UCSC ゲノムブラウザなどとの連携が構築された。さらに他のデータベースとの連携をはかるため、ChIP-Atlas のデータが全て RDF 化され、横断的なデータ抽出が可能となった。ChIP-Atlas を絶え間なく更新するために最も手間のかかる作業がサンプルメタデータのキュレーションであったが、その作業を機械学習によって半自動化させることに成功し、ChIP-Atlas の持続可能性が飛躍的に向上した。さらなる利用者の増加をはかるため、国内外での学会や研究集会、大学や研究機関での講演やハンズオンセミナーなどの広報活動が展開され、結果的に毎月 4-5 千人のユニークユーザが訪問し、20 万ページが閲覧されるサイトに成長した。また、ChIP-Atlas は 2018 年 12 月に論文として発表されたが、以来 300 報以上の学術論文によって引用された。

したがって本研究開発により、ゲノム-タンパク質相互作用 (ChIP-seq)、オープンクロマチン情報 (DNase-seq, ATAC-seq)、メチローム情報 (Bisulfite-seq) を統合的に俯瞰し利活用できるデータベースおよびウェブツールが開発された。さらにその利便性や持続可能性が向上されたことにより、世界的に活用され多くの論文に引用される生命科学系インフラとして発展を遂げた。

## §2. 研究開発実施体制

### 1. 研究グループ

#### (1) 「沖」グループ(研究代表者グループ)

- 担当項目  
データ更新、Bisulfite-seq と ATAC-seq データ、生物種、最新ゲノムアセンブリデータの追加

#### (2) 「三浦」グループ(主たる共同研究者グループ(1))

- 担当項目  
Bisulfite-seq データの追加

#### (3) 「浜本」グループ(主たる共同研究者グループ(2))

- 担当項目  
キュレーション

#### (4) 「川路」グループ(主たる共同研究者グループ(3))

- 担当項目  
データ閲覧機能の拡充

#### (5) 「竹本」グループ(主たる共同研究者グループ(4))

- 担当項目  
疾患関連 SNP の実験的検証

#### (6) 「沢津橋」グループ(主たる共同研究者グループ(5))

- 担当項目  
疾患関連 SNP の実験的検証

### 2. 有識者会議等

該当無し

### §3. 研究開発の目的、実施内容及び成果

#### 1. 研究開発の背景

近年の次世代シーケンス技術の発展により、ゲノム上の転写因子やヒストンの結合分布、およびゲノム DNA のメチル化状態が網羅的に解析できるようになった。それぞれ ChIP-seq や Bisulfite-seq と呼ばれるこの実験手法は 2016 年度末時点で8万以上もの報告件数があり、今後もさらなるデータ数の増加が見込まれる。ともにエピゲノミクス研究において中核をなす研究手法であり、細胞分化や遺伝子発現などの基礎研究だけでなく、診断やリスク予測など医療的な応用も進められている。その重要性は ENCODE、Roadmap Epigenomics や、我が国も参画する IHEC などの巨大プロジェクトが進行していることから裏付けられる。近年、実験プロトコルの安定化やシーケンスの価格が下落したことにより、中小規模のプロジェクトやラボ単位の研究データ収集が進んでいる。その数は上記の大規模プロジェクトで収集された数の 10 倍近くも存在するが、それらを統合的に活用するためのデータベースは存在しない。本研究ではこれらを網羅的に収集し、個別データや統合的な解析結果を提供することをめざす。

ChIP-seq データを網羅的に収集し、それらを容易に閲覧することが可能なデータベースはすでに存在しており、Cistrome DB (<http://cistrome.org/db>) や GTRD (<https://gtrd.biouml.org>) はその代表例である。しかしこれは個別の ChIP-seq データを閲覧するためのサービスであるため、膨大な量の ChIP-seq データの中から知識発見を導くことは困難である。

#### 2. 研究開発対象のデータベース・ツール

##### (1) データベース

- ・ 主要なもの

正式名称	略称	概要
ChIP-Atlas	ChIP-Atlas	エピゲノミクス統合データベースおよびデータ解析ツール

- ・ 上記以外のもの

該当無し

##### (2) ツール等

該当無し

#### 3. 達成目標及び実施計画

##### (1) 当初の実実施計画・達成目標

###### ■ Bisulfite-seq データの追加

Bisulfite-seq はメチル化シトシン (mC) の分布をゲノムワイドに定量するための実験手法であり、エピゲノム研究においてもはや無くてはならない実験手法として確立している。ChIP-seq と同様、そのシーケンス生データは NCBI SRA より自由にダウンロードできるが、それをレファレンスゲノムにマッピングしてメチル化コールを行う必要があり、情報解析技術なしにはその利活用が難しい。NGSmethDB や MethBase などのデータベースでは SRA を事前処理したデータを公開しているが、数百程度の実験データしか収集しておらず、約2万件もの Bisulfite-seq データのごく一部しかカバーしていない。そこで本研究では SRA として公開

されている Bisulfite-seq データのうち、データ数が上位3つの生物種(ヒト、マウス、シロイヌナズナ)の SRA を網羅的に収集し、事前処理したデータの公開をめざす。

Bisulfite-seq の解析パイプラインは、ChIP-seq のそれとよく類似しており、これまで開発したコードを一部改変して構築する。まず SRA ファイルを FASTQ 形式に復号し、レファレンスゲノムへのアライメントと mC-calling をおこなう。得られたテキスト形式の mC-call データは BigWig 形式に変換し、ゲノムブラウザでメチル化率を閲覧するためのファイルを作成する。また、統計学的有意にメチル化が亢進/低下している領域(hyper- / hypo-MR)をコールし、BED4 形式のファイルを作成する。さらにこれらすべてのデータを単一のファイル (BED9 形式) に統合する。これをゲノムブラウザで可視化することにより、どの細胞で、どのゲノム領域に、どの程度メチル化しているかが、一目で理解できるようにする。

#### ■生物種の追加

ChIP-Atlas は現在、ヒト、マウス、ショウジョウバエ、線虫、出芽酵母の5生物種を扱っている。しかし、Bisulfite-seq データではシロイヌナズナのデータが多いことや、ChIP-Atlas にラットやシロイヌナズナの追加を望む意見が学会発表などで得られているため、第3年次までにそれら2生物種の公開をめざす。まず初年次より、沖グループで両生物種を追加するための解析パイプラインを構築させ、第2年次より、植物を専門とする工藤、川勝研究参加者とともに既存データ全てのキュレーションをおこなう。また、これまでインターフェースの作成を担当した DBCLS の大田達郎 特任研究員の協力を仰ぎ、Web インターフェースの刷新をおこない、第3年次までにユーザが閲覧できるようにするとともに、毎月更新を軌道に乗せる。

#### ■キュレーションの効率化

毎月おこなわれる ChIP-Atlas の更新でもっとも労力を要するのが、メタデータのキュレーションである。これは厳格なルールに従っておこなわれ、以下の4つの属性値をマニュアルで記入している。

- ・抗原クラス (Histone, Transcription factor など)
  - └ 抗原名 (H3K4me3, Pou5f1, Trp53 など)
- ・細胞クラス (Blood, Liver, Pancreas など)
  - └ 細胞名 (K-562, Hepatocytes, PANC-1 など)

また、転写因子名は、HUGO や MGI などが定める official gene symbol に統一しており、synonymous な名称は使用しない。細胞名については official symbol のような国際的ルールがないため、Table 2 の表記法に従って記述されている。

これまで毎月、追加される約 1,300 件の ChIP-seq SRA のキュレーションをおこなってきたが、本申請では Bisulfite-seq データや他の生物種の追加も予定しているため、さらに作業負担の増大が予想される。そこで、沖グループと浜本グループで一連の作業を明確に記したマニュアルを作成し、データの入力方法や上述のルールを習熟させる。また、キュレーション作業には一定の生物学的知識が必要なため、教科書や論文の輪読会などを定期的におこなう。これまで、実験に用いられた細胞で特に多いのが、がん由来の臨床サンプルや細胞株であるため、がんなどの疾患を専門とする浜本グループの知識や経験を結集させてその教育と実行にあたる。また浜本グループには、本研究開始の当初より、全てのキュレー

ションログの精査を依頼する。

ChIP-Atlas のキュレーションログは全て保存しているため、現在ではその履歴を参考にしながらデータ入力をおこなっている。しかし、この作業を機械学習によって自動化できれば、作業効率の飛躍的な改善が期待できる。実例をあげると、SRA 投稿者が記述したメタ情報の中の「K562」というキーワードがあったとき、細胞クラスを「Blood」、細胞株名を「K-562」(ハイフンあり)とした履歴が複数ある。このような履歴を学習させることにより、更新によって新たに「K562」を含む SRA メタデータが追加されれば、瞬時にその正解データが得られる。機械学習のためのアルゴリズムの選定は江原遥研究参加者や瀬々潤アドバイザーに協力を仰ぎ、沖グループでその実装を進める。研究開始当初よりアルゴリズムの選定を進め、第2年次は新たに追加されるメタデータのキュレーションに試験的に導入する。キュレータによる結果との一致率を高めるよう、プログラムを改善し、第3年次より本格運用できることをめざす。

#### ■他のデータベースとの連携

ChIP-Atlas が公開しているデータは全て固有の URL を持ち、なおかつ IGV や UCSC ゲノムブラウザなどの汎用ツールによって可視化できるようなフォーマット(BigWig、BED 形式)で作成されている。また任意のデータを抽出できるような工夫をしており、任意の抗原名や細胞名によって対応するファイル名や URL が一意に引き出せるような対応表を作成しているため、抗原名や細胞名をキーワードとするその他のデータベースとの有機的な連携が可能である。そこで ChIP-Atlas に収蔵された全てのデータファイルを RDF 化し、その連携が円滑に行えるようにする。ChIP-Atlas の共同開発者である DBCLS の大田達郎氏に RDF 化のノウハウを提供していただき、沖グループでその作業を行う。追加生物種や Bisulfite-seq データが出揃い始める第2年次より RDF 化を開始し、それが完了したのち、多階層オーミクスデータベース DBKERO (<http://kero.hgc.jp>) に ChIP-Atlas データが表示できることをめざす。またそのノウハウを生かし、その他のデータベースとも連携を図る。

#### ■利用者増加のための広報活動

ChIP-Atlas は 2015 年 12 月に一般公開されたが、現在までの約1年間の利用回数は 15,000 回を超えている(データの閲覧やダウンロードなどのサービス利用回数。単純訪問者などは除く)。これまでに、学会発表や参加者へのチラシの配布をおこなうことで、着実に利用回数が増加しているため、今後さらに広報活動をおこなう。本研究代表者グループはそれぞれ分野の異なる学協会やデータ産生型プロジェクトに所属しているため、学会発表やチラシの配布を積極的におこなうことで、生命科学の幅広い分野に普及させ、さらに追加すべき機能や生物種などのニーズを吸収する。

また、論文執筆による ChIP-Atlas の普及にも努める。現在のところ、ChIP-Atlas を利用し GWAS データを解析した結果を投稿準備中であるため、初年次内にアクセプトされることをめざす。また、今後の ChIP-Atlas の機能拡充(生物種や Bisulfite-seq の追加)に合わせて、Nucleic Acid Research 誌の Database Issue にも投稿し、拡張された機能について紹介する。

このような広報活動を積極的におこなうことにより、第3年次末までには年間のサービス利用回数を 3 倍(45,000 回)に、第5年次末までには 5 倍(75,000 回)を獲得できるようめざす。

## (2) 期間中に追加・削除・変更した実施計画・達成目標

### ■データ閲覧機能の拡充(2017年度追加実施項目)

川路グループにおいて、データ閲覧機能拡充のための開発を実施する。ChIP-Atlasはクライアント・ソフトウェア IGV を用いたデータ閲覧機能を既に備えているが、これに加えて、専用ソフトウェアを用いずに ChIP-Atlas データを閲覧する環境を構築する。一般的なウェブ・ブラウザのみを利用することでゲノミクスデータを閲覧できる UCSC Genome Browser Data base や ENSEMBL Genome Browser などのゲノムブラウザー・データベースに対応するため、ChIP-Atlas データへ Track Hub 形式でアクセスする環境を整備する。Track Hub 形式の必要性や RDF 化への影響については「04\_付録\_trackhub」に補足説明。

### ■ATAC-seq データの追加(2019年度より実施計画に追加)

研究代表者の冲らはさまざまな学会や研究期間で ChIP-Atlas の広報活動を行なっているが、寄せられる要望の中で特に目立つのが ATAC-seq データの追加と、最新ゲノムアセンブリへの対応であり、当初計画にあったシロイヌナズナデータの追加や Bisulfite-seq データの追加よりもはるかに多い。したがって当初計画のデータ追加を進めつつも、ユーザからの需要をいち早く満たせるような研究開発が必要である。

現行の ChIP-Atlas では ChIP-seq データだけでなく、DNase-seq データもカバーしている。いっぽうで後発技術の ATAC-seq 法 (Assay for Transposase-Accessible Chromatin Sequencing) もオープンクロマチン領域を網羅的に解析できるが、DNase-seq に比べて操作が簡便で再現性も高いことで知られる。また数百個程度の少数細胞でもオープンクロマチン領域を検出できるため、世界中で急速に普及している。もはや ATAC-seq は DNase-seq よりもはるかに多用されており、その年間報告件数は Bisulfite-seq を超え ChIP-seq に迫るほどである。したがって「エピゲノミクス統合データベース」の開発を目指す本研究課題において、ATAC-seq データの整備は急務といえる。そこで本年度より ATAC-seq データを収集・解析し、2020年度内までにその公開をめざす。解析パイプラインは DNase-seq と同じもので良いが、ATAC-seq 専用の統合解析が必要なため、そのための解析パイプラインを構築する。また ATAC-seq データはこれまで3万件近く報告されているため、サンプルメタデータのキュレーションを浜本グループとの共同でおこなう。

### ■最新ゲノムアセンブリへの対応(2019年度より実施計画に追加)

ChIP-Atlas の開発を開始した 2014 年において、ヒトゲノムアセンブリの最新版は hg38 であったが、その一つ前のバージョン (hg19) で得られた知識があまりに多かったため、hg38 による解析はあまり好まれなかった。そのため ChIP-Atlas においても hg19 での開発を進めていったが、現在では hg38 の方が主流となっているため、その改訂を求めるユーザの声が増えている。また、マウス、ハエ、線虫などについても最新アセンブリに対応しておらず、とくにマウス研究者からは「mm9 は古すぎる」との声が多い。そこで本年度より最新ゲノムアセンブリによる再解析を開始し、2021年度内までにその公開をめざす。解析パイプラインそのものは既存のものと同じで良く、必要なのはアライメントのためのゲノムアセンブリを最新版に変更するだけである。またサンプルメタデータのキュレーションも必要ない。

### ■疾患の原因 SNP の予測と実験による検証(2021年度追加実施項目)

申請者は Genome-wide association study (GWAS) で同定された疾患と関連する no

ncoding SNP の解析に ChIP-Atlas を活用したところ、lead SNP の LD-block 内に、様々な転写因子を結合する hotspot が存在することを見出した。またそれらのいくつかは既知の causal SNP 座位と完全に一致していたため、転写因子結合 hotspot は causal SNP を特定するための指標となりうる可能性が示唆された。そこで上記手法で予測された causal SNP の疾患リスクを実験的に検証し、その発症プロセスの解明をめざす。

これまでに申請者は徳島大学の竹本龍也 教授と沢津橋俊 特任准教授との共同研究で、causal SNP 候補領域のマウス相同部位を特定し、ゲノム編集によるヘテロ変異マウスを作製している。本追加実施ではそれらマウスのホモ個体を作製し、表現型解析をおこなう。外見的な所見(下痢、皮膚の炎症)に加え組織学的なマーカー染色をおこない、症状の程度や病態の違いなどを分析し、経時的な発症プロセスを解明する。顕著な発症が見られない場合は、炎症性腸疾患を誘発する dextran sulfate sodium (DSS) の飲水投与や、乾癬を誘発する imiquimod (IMQ) の皮膚への塗布を低濃度でおこない、発症リスクの底上げを図る。

## 4. 実施内容

### (1) 実施内容

#### ■ Bisulfite-seq データの追加

三浦グループが中心となり、株式会社 Rhelixa との協同で Bisulfite-seq のためのアライメントツールの評価をおこなった。Tsuji J, et al. Brief Bioinform. 2016 にある約 20 のアルゴリズムについて、計算時間、マルチスレッドの可否、メモリ使用量、マッピング率や、NIG supercomputer への移植性について検討した結果、Bismark が最も高いパフォーマンスを示したため、これが公開ツールの中では最もパイプライン構築に適していると判断した。これとは別に、三浦がこれまで自身で開発・運用を行ってきた BMap をはじめとする一連のマッピングツールは、論文未発表であるため見送る予定であったが、株式会社 SRA との協同開発によりパフォーマンスが向上したため、これを採用することとした。とくに、NIG supercomputer でも実行できるように改変し、また高速かつ低容量で計算できることを目的に、ネットワーク負荷や処理の冗長性を減少させ、gzip で圧縮されたファイルを直接入出力するための機能拡張を行った。

この NIG supercomputer に実装された解析パイプラインを使用し、NCBI-SRA から SRA のダウンロード、Fastq への展開、レファレンスゲノムへのアライメント、メチル化率データの作成、などの解析をすべて完了した(ヒト、マウス、ラット、ハエ、線虫、酵母のデータ合わせて 60,034 件)。また、高・低メチル化領域や、部分メチル化ドメインの抽出を行うためのツール methpipe を NIG supercomputer に実装し、これらのゲノム領域情報の BED ファイルを作成した。現在、サンプルメタデータのキュレーションを完了し、NBDC の公開サーバにデータを転送している。2021 年度中に ChIP-Atlas の Web サービスとして解析データの公開をめざす。

#### ■ ATAC-seq データの追加

沖グループが中心となり、NCBI から公開されている NCBI SRA Metadata から ATAC-seq データの実験 ID を抽出するためのプログラムを作成した。これを ChIP-Atlas の月例更新プログラムへの組み込み、自動的にデータを収集・解析するためのパイプラインを構築した。この解析パイプラインを NIG supercomputer に実装し、NCBI-SRA から SRA のダウンロー



ド、Fastq への展開、レファレンスゲノムへのアライメント、ピークコールなどの解析をすべて完了した(ヒト、マウス、ラット、ハエ、線虫、酵母のデータ合わせて 107,884 件)。現在、サンプルメタデータのキュレーションを完了し、NBDC の公開サーバにデータを転送している。2021 年度中に ChIP-Atlas の Web サービスとして解析データの公開をめざす。

#### ■生物種の追加

沖グループが中心となり、ラットの ChIP-seq データの追加を図った。おもな開発項目は以下の通り。

1. レファレンスゲノムの選定:rn6
2. マッピングやその他の計算のためのライブラリファイル作成。
3. 計算パイプラインの刷新。
4. 更新された SRA データの自動収集プログラムの作成。
5. キュレーションのための controlled vocabulary の収集。
6. 転写因子名を official gene symbol に統一。
7. キュレーションのための補助ツールの改訂。
8. NBDC のファイルサーバへの転送スクリプトの刷新。
9. Web インターフェースの刷新(DBCLS の大田達郎氏に依頼)

このように大幅な改訂や刷新を勧め、2018 年 1 月よりラットの ChIP-seq データを一般公開した。

また、シロイヌナズナのデータ収集を行うため、レファレンスゲノムを TAIR 10 に設定し、上記ステップの 4 まで進めた。

#### ■キュレーションの効率化

浜本グループが中心となり、これまでのキュレーション履歴の見直しを進めた。特に、どのレベルまで書き下すかという問題について議論を進めた。例えば白血病は病態の違いから [急性/慢性] [骨髄性/リンパ性] までの分類が必要であるいっぽう、乳がんのサブタイプは明確でないことが多いため「Breast cancer」までにとどめるべきなど、ある程度の基準が得られた。分類クラスの見直しもおこない、たとえば “Bone marrow cells” は当初 “Bone” クラスにしていたが、血球関連の研究に使用されることのほうが多いため “Blood” クラスに変更した。また、ヒト ES 細胞は株間における性質の差異が大きいことが知られているため、H1, H7, や H9 などの株名を明記することとした。これまで Embryonal carcinoma cells は「Embryo」という大分類群に入れていたが、ES/iPS 細胞と同様に多分化能細胞として使用されることが多いため「Pluripotent stem cell」という大分類群に変更した。Cardiomyocyte は「Muscle」大分類群としていたが、心筋は骨格筋と大きく性質が異なることから、「Cardiovascular」大分類群に変更した。また、細かいスペルミスや明らかな間違い、文字化けなどの修正を進めた。2019 年度内に本研究開始時(2017 年4月)までのキュレーション履歴の見直しをすべて完了した。さらに追加された ChIP-seq データについては 2020 年度内に見直しをすべて完了し、その後も継続的にキュレーションを進めている。また新たに追加した Bisulfite-seq と ATAC-seq データのキュレーションは沖・浜本グループの協働で 2021 年に全て完了した。

また、瀬々潤アドバイザーに協力を仰ぎ、機械学習によるキュレーション作業の自動化プログラムを開発していただいた。データ投稿者が記載した sample metadata の文字列を文

字 n-gram 法 (n=3~7) で分割したものを入力データとし、キュレータが修正した結果の文字列を正解データとして与える。これをサポートベクターマシン (SVM) 法で分類することにより学習させた。これを 10-fold cross validation により評価したところ、10 回以上の頻出歴のある正解データについては約9割以上の正答率が得られた。10 回未満のデータについては正答率が極端に下がるためさらなる精度向上の余地はあるが、まずはその実用を優先すべく、株式会社 Rhelixa との協同開発により、NIG supercomputer へ実装した。また毎月の更新に合わせて、キュレーション履歴を再学習し、新規のサンプルメタデータに対する予測結果を返すようなプログラムを開発した。これによりキュレーションの時間が短縮され、なおかつ入力ミスなども低減できた。

#### ■最新ゲノムアセンブリへの対応

ヒト、マウス、ハエ、線虫について、ChIP-seq、DNase-seq、ATAC-seq、Bisulfite-seq の最新のゲノムアセンブリに対するアライメントやその他のデータ解析を完了した。このうち、ChIP-seq と DNase-seq データについて ChIP-Atlas の Web サービスより解析データを一般公開した。

ヒト:hg19 → hg38

マウス:mm9 → mm10

ハエ:dm3 → dm6

線虫:ce10 → ce11

#### ■他のデータベースとの連携

DeepBlue Epigenomic Data Server (<http://deepblue.mpi-inf.mpg.de>; Albrecht, et al. *Nucleic Acids Res.* 2016) との連携を構築した。これはドイツ Max Planck Institute の Felipe Albrecht 氏が開発した、様々なオミクスデータを Web 上で横断的に統合解析できるサービスであるが、ChIP-seq データについては ENCODE や Roadmap Epigenomics などの大型プロジェクトのデータ(しかしながら全ての public ChIP-seq データの2割ほどにすぎない)しか利用できなかった。Albrecht 氏が福岡に訪問した際に共同プロジェクトを提案したところ、ChIP-Atlas に収録される全てのデータが DeepBlue サーバにインポートされ、これらのデータをそのサービス内で利用可能となった。これにより、例えば特定のゲノム領域に結合する転写因子を ChIP-Atlas データから引き出し、それらを発現する細胞を transcriptome データから抽出する、といったオミクスデータ横断的な再解析が可能となっている。

京都大学の石濱泰 教授らによるプロテオームデータベース jPOST では異なる細胞間で発現量が異なるタンパク質を抽出することができる。そのようなタンパク質をコードする遺伝子群の周辺に結合するような転写因子は、それら遺伝子群のマスター制御因子と考えられるが、その探索は ChIP-Atlas の Enrichment Analysis にこのような遺伝子群を submit すれば得られる。この一連の作業をよりシンプルに行えるようにすべく、jPOST で得られたタンパク質名を ChIP-Atlas の Enrichment Analysis へと自動入力するようなシステムの構築を進めた (DBCLS 守屋勇樹 特任助教との共同開発)。また、以前までは ChIP-Atlas の Enrichment Analysis では official gene symbol しか対応できていなかったが、jPOST との連携に伴い、UniProt ID や RefSeq gene ID も受け入れられるよう、システムの変更を進めた。

DBCLS の大田達郎 特任助教と池田秀也 学術支援技術専門員に協力いただき、ChIP-Atlas に収録されている ChIP-seq データの細胞名を controlled vocabulary (CV) に変換

する作業を行なった。とくに EMBL-EBI が運営する ZOOMA を利用することで、迅速に CV へのマッピングを行なった。またこれらのデータを RDF 化することにより、収録される全てのデータを SPARQL 言語でデータの抽出や絞り込みができるようになった(下図)。これにより、例えば任意の ChIP-seq データ ID (SRX150882) の BioSample データ (SAMN01001314) に紐づく抗体情報 (Anti-PU.1)、細胞情報 (C2C12 cells) とその cell ontology (EFO\_0001098)、介入処理 (MALP2 (1 ng/ml) for 4hrs) が抽出できる。また、ゲノム座標をクエリとした時、SRX150882 のピーク情報 (例 chr6:115675762-115676100) も抽出できるようになった。

#### ■データ閲覧機能の拡充(2017 年度追加実施項目)

川路グループが中心となり、ChIP-Atlas データを UCSC genome browser でも閲覧するための開発を実施した。ChIP-Atlas の Peak Browser のデータは別のゲノムブラウザ (IGV) に特化した BED9+GFF3 フォーマットであることから、それを全て BigBed 形式へと変換した。また、組み合わせる見ることの多いデータ (特定の因子に関する様々な臓器・細胞でのプロファイル等) について単一のデータファイルとして編成しなおすことで、より円滑な閲覧を実現した。つぎにこれを UCSC data hub 用のファイルサーバにインポートし、Custom tracks として閲覧する為の設定を整えた。これにより、ChIP-Atlas の Peak Browser のデータを IGV だけでなく UCSC genome browser でも閲覧できるようになった。また同様に、東京大学の鈴木穰 教授と共同し、DBKERO のゲノムブラウザでも Peak Browser のデータが閲覧できるようになった。

#### ■疾患の原因 SNP の予測と実験による検証(2021 年度追加実施項目)

2021 年度 10 月より、追加実施研究を開始する。

#### ■利用者増加のための広報活動

2018 年 11 月、EMBO Reports 誌に ChIP-Atlas の初報論文を報告し、九州大学/ROIS/JST/AMED で共同プレスリリースをおこなった。沖グループが中心となり本研究開発開始以降、国内外で 51 件の学会、研究セミナーや NBDC が主催する AJACS 講習会やトーゴの日シンポジウムなどで ChIP-Atlas に関する発表を行った。また ChIP-Atlas のハンズオンセミナーも好評であり、参加者の持参したパソコンで ChIP-Atlas を実際に使ってもらいながら、新規ユーザの獲得に努めた。ChIP-Atlas を紹介するためのパンフレットをアダチデザイン研究所に作成してもらい、上記のセミナーや学会の展示ブースなどでこれまでに 1,000 部以上を配布した。

## (2) 実施内容のうちの特定項目の詳細

### ① 研究コミュニティを含むデータ提供者との連携・協業

#### ■AMED 老化メカニズムの解明・制御プロジェクト 研究推進・支援拠点

代表: 鍋島陽一 (公益財団法人先端医療振興財団 先端医療センター センター長)

期間: 2017 年度～2021 年度

本プロジェクトは老化メカニズムに関する研究の推進を支援する拠点であり、沖は分担研究者として参画し、ChIP-Atlas を活用したデータ解析を担当した。

■ 科研費・新学術領域研究 多様な「個性」を創発する脳システムの統合的理解

代表: 大隅典子 (東北大学大学院医学系研究科 教授)

期間: 2017 年度～2020 年度

本領域は個性が生み出されるしくみを解明することを目的とし、沖は分担研究者として参画し、ChIP-Atlas を活用したデータ解析支援を担当した。

- 東北大学大学院医学系研究科 大隅典子との共同研究で、オス加齢マウスに由来する精子の低メチル化領域に転写因子 REST の結合がエンリッチすることを見出した (Yoshizaki K., et al., 2020 EMBO Rep)。

■ 科研費・基盤 B 顎顔面形成不全を伴う未診断稀少疾患の遺伝的原因の究明

代表: 黒坂 寛 (大阪大学歯学部附属病院 講師)

期間: 2019 年度～2021 年度

本研究は顎顔面形成不全が生み出されるしくみを解明することを目的とし、沖は分担研究者として参画し、ChIP-Atlas を活用したデータ解析支援を担当している。

■ その他、ChIP-Atlas を利用した研究の共同研究者 (順不同)

● 共著論文あり

望月 研太郎 (University of British Columbia; Mochizuki K., et al., 2018 *Cell Rep*)

魏范研 (熊本大; Hirayama M., et al., 2020 *Cell Rep*)

西原祥子 (創価大学; Pecori F., et al., 2021 *Sci Rep*)

神沼英里 (遺伝研; Kaminuma E., et al., 2020 *Genes Genet Sys*)

川路英哉 (理研; Lizio M., et al., 2019 *Nuc Acids Res*)

Han-Sung Jung (延世大学; Wu Z., et al., 2019 *Dev Dyn*)

日野信次朗 (熊本大; Anan K., et al., 2018 *Nuc Acids Res*)

野島 聡 (大阪大; Ohshima K., et al., 2017 *Sci Rep*)

宮本洋一 (医薬基盤研; Miyamoto Y., et al., 2021 *FASEB J*)

近藤寿人 (京都産業大; Matsuda K., et al., 2017 *Development*)

② データベース利用者への周知、利用者との連携・協業

§ 3.4(1)「利用者増加のための広報活動」に記載したように、ChIP-Atlas を紹介した論文を出版した (Oki S., et al., 2018 EMBO Rep)。またその使い方を詳述した和文 (沖真弥, 大田達郎. 2019 実験医学, 沖真弥, 大田達郎. 2019 THE LUNG perspectives) および英文書籍 (Practical Guide to Life Science Databases. Springer Nature. In press) を執筆した。§ 8.3 に示すように、学会発表のほか、大学などの研究機関や企業において ChIP-Atlas の使い方や利活用事例の紹介、さらにハンズオンセミナーを開催した。とくに分子生物学会では ChIP-Atlas のブースを出展し、パンフレットの配布、使い方の説明や個別相談などをおこなった。

③ 利用者にとって有用なデータ基盤の構築

§ 3.4(1)「他のデータベースとの連携」に記載したように、DeepBlue Epigenomic Data Server、jPOST、UCSC genome browser との連携を構築し、さらに全てのデータを RDF 化した。

④ 持続的なデータベース運用体制の構築に向けた取り組み

研究開発開始当初より、NCBI SRA の更新に伴う新規データの追加と自動解析を行うための解析パイプラインを構築している。また、§ 3.4(1)「Bisulfite-seq データの追加」「ATAC-seq データの追加」「生物種の追加」に記載したように、新たなプラットフォームのデータや生物種を柔軟に取り込めるように、解析パイプラインを改善した。さらに、持続可能性を高めるため § 3.4(1)「キュレーションの効率化」に記載したように、キュレーションの体系化と機械学習による半自動化を導入した。

⑤ 人材の育成

研究室での輪読会や講習会などを積極的に行い、研究員やキュレータの医学的知識を養成した。また学会などへの参加を促し、見聞を広めるとともに ChIP-Atlas の周知活動も分担させた。研究開発期間内にプロジェクト研究員として特定助教2名を雇用した。両者ともに「若手研究者の自発的な研究活動」に関する実施方針にしたがって、自身のキャリア向上のための研究活動も実施させた。また、代表者の沖は 2020 年度 4 月より京都大学大学院医学研究科で PI のポジションを得た。

⑥ 国際連携・国際貢献

§ 3.4(1)「他のデータベースとの連携」に記載したように、国際的に使用されているデータベース(jPOST、UCSC genome browser)との連携を構築した。また、§ 4.1(2)に後述のとおり、ユーザの約三分の二が海外からのアクセスであり、また引用論文の約三分の二が海外の研究グループである。

⑦ その他

特に無し

## §4. 主要なデータベースの利活用状況

### 1. アクセス数

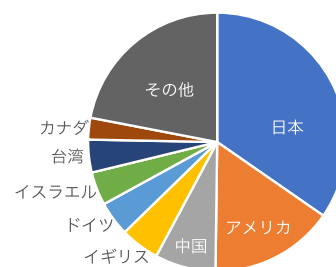
#### (1) 実績

表 1 研究開発対象の主要なデータベースの利用状況(月間平均)

名称	種別	2017年度	2018年度	2019年度	2019年度	2021年度 (10月時点)
CHIP-Atlas	訪問者数	2,589	694	1,452	4,360	4,613
	訪問数	762	794	3,545	13,459	12,934
	ページ数	79,474	97,174	126,004	143,627	210,951

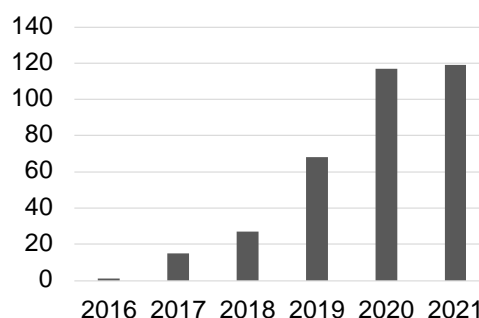
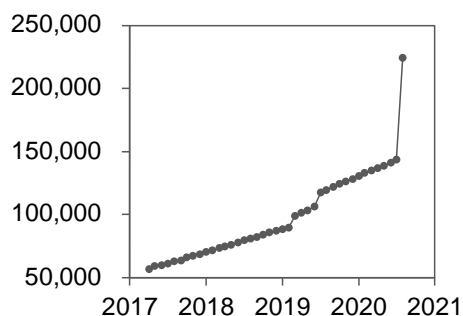
#### (2) 分析

- ChIP-Atlas のページ数は毎年約 1.2～1.4 倍の増加を示している。§ 3.4(1)「利用者増加のための広報活動」に記載したように、ChIP-Atlas 論文や和文総説の出版、また学会などの発表によって着実にユーザを増やしている。
- 右図に示すように、2021 年度の国別アクセス数(ページ数)を集計すると、およそ三分の二が海外からとなっているため、国際的に使用されていることがわかる。また、次項に示す引用論文も同様に、およそ三分の二が海外の研究グループである。とはいえ人口比率から考えるとまだ海外ユーザが少ないとも言えるため、今後は国際学会などでの発表や、続報論文による周知、さらに学術的インパクトの高い利活用事例の作成が必要と思われる。



#### 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

下左図のように ChIP-seq データは毎月約 2,000 件が追加されており、2020 年 8 月までは毎月データ更新した。その後、ATAC-seq と Bisulfite-seq データを追加し、総計 224,659 件のデータ(SRX 換算)となり、2021 年度内に公開する予定である。下右図は ChIP-Atlas の引用論文数を示しており、毎年ほぼ倍増していることがいえる(Google scholar で「ChIP-Atlas」を検索し、全文閲覧して確かに引用されていることを確認。Oki S., et al., 2020 EMBO Rep を引用していない論文も含む)。2021 年 9 月時点で 352 報の論文に引用されている。



### **3. 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）**

これまでに 352 報の論文に ChIP-Atlas が引用されており、おもにがん、免疫、老化、進化、分化、幹細胞などさまざまな分野の基礎研究に利用されている。産業や科学技術に対する直接的な波及効果は把握できていないが、企業の複数の社員より、ChIP-Atlas を研究開発に用いているという私信を得ている。

## §5. 今後の展開

ChIP-Atlas は、もはや研究インフラの一部として世界中の研究者に利用されているため、本プログラム終了後も新たな予算を獲得できれば継続的に運営したい。これまで同様、NCBI SRA の更新に合わせて ChIP-Atlas のデータをアップデートして公開することで、最新のエピゲノミクスデータを世界に提供していきたい。近年、ChIP-seq とは別のアプローチでタンパク質-ゲノム相互作用を解析する手法 (CUT&Tag や ChIL-seq など) も開発されている。手法の簡便さと少数細胞でも可能なため今後急速にデータが増加すると思われるため、それらのデータ解析と公開にも取り組んでいきたい。ChIP-Atlas の利活用事例を創出するためにも、次の統合化推進プロジェクトでは実験による検証や、ウェット研究者との共同研究も募集要項に入れていただきたい。とくに、徳島大学の竹本・沢津橋グループとの共同研究により、ChIP-Atlas が疾患ゲノムにおける仮説形成の好例となると期待される。また、代表者の沖が開発している Photo-Isolation Chemistry 技術は特定の組織の遺伝子発現やエピゲノム情報を正確に解析できるため、その結果を ChIP-Atlas で再解析することで、各組織のマスター制御因子の特定が可能となる。これらの情報はダイレクトプログラミング技術や組織工学などへ貢献できると期待される。



## §6. 自己評価

§3に記載したように、研究開発開始時の目標をおおむね達成できている。また§4に記載したように、ChIP-Atlas は世界中の研究者に使われ、論文に引用されているため、本プログラムの趣旨である「科学的知見を容易に閲覧・参照」「関連分野の有用情報の発見」「大規模解析によって新たな知見を見いだす」「科学技術イノベーションの創出」などに貢献できている。

## §7. 外部発表等

### 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)	30件
	国際(欧文)	4件
未発行論文	国内(和文)	0件
	国際(欧文)	0件

#### (2) 論文詳細情報

(直接的な成果論文のほかに、開発対象データベースを利用した間接的な成果論文を含む場合があります。)

1. Matsuda K, Mikami T, Oki S, Iida H, Andrabi M, Boss JM, Yamaguchi K, Shigenobu S, Kondoh H. 2017. ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. *Development* 144: 1948–1958. (DOI: 10.1242/dev.143479).
2. Ohshima K, Nojima S, Tahara S, Kurashige M, Hori Y, Hagiwara K, Okuzaki D, Oki S, Wada N, Ikeda JI, et al. 2017. Argininosuccinate Synthase 1-Deficiency Enhances the Cell Sensitivity to Arginine through Decreased DEPTOR Expression in Endometrial Cancer. *Sci Rep* 7: 45504. (DOI: 10.1038/srep45504).
3. Sanosaka T, Imamura T, Hamazaki N, Chai M, Igarashi K, Ideta-Otsuka M, Miura F, Ito T, Fujii N, Ikeo K, et al. 2017. DNA Methylation Analysis Identifies Transcription Factor-Based Epigenomic Signatures of Multilineage Competence in Neural Stem/Progenitor Cells. *Cell Rep* 20: 2992–3003. (DOI: 10.1016/j.celrep.2017.08.086).
4. Semba Y, Harada A, Maehara K, Oki S, Meno C, Ueda J, Yamagata K, Suzuki A, Onimaru M, Nogami J, et al. 2017. Chd2 regulates chromatin for proper gene expression toward differentiation in mouse embryonic stem cells. *Nucleic Acids Res* 45: 8758–8772. (DOI: 10.1093/nar/gkx475).
5. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, Kawaji H, Nakaki R, Sese J, Meno C. 2018. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* 19: e46255. (DOI: 10.15252/embr.201846255).  
【論文概要】ChIP-Atlas の初報論文として発表した。ChIP-Atlas がカバーする生物種やデータ数、公開データの解析方法、および使い方などを説明した。またその利活用事例として組織特異的に発現する遺伝子に対し、結合がエンリッチする転写因子の解析結果を示した。
6. Miura F, Fujino T, Kogashi K, Shibata Y, Miura M, Isobe H, Ito T. 2018. Triazole linking for preparation of a next-generation sequencing library from single-stranded DNA. *Nucleic Acids Res* (DOI: 10.1093/nar/gky452).
7. Miura F, Ito T. 2018. Post-Bisulfite Adaptor Tagging for PCR-Free Whole-Genome Bisulfite Sequencing. *Methods Mol Biol* 1708: 123–136. (DOI: 10.1007/978

- 1-4939-7481-8\_7).
8. Miura F, Shibata Y, Miura M, Sangatsuda Y, Hisano O, Araki H, Ito T, "Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging.", *Nucleic acids research*, vol. 47, No. 15, 2019. (DOI: 10.1093/nar/gkz435).
  9. Machiko Kojima, Kenbun Sone, Katsutoshi Oda, Ryuji Hamamoto, Syuzo Kaneko, Shinya Oki, Asako Kukita, Hidenori Machino, Harunori Honjoh, Yoshiko Kawata, Tomoko Kashiya, Kayo Asada, Michihiro Tanikawa, Mayuyo Mori-Uchino, Tetsushi Tsuruga, Kazunori Nagasaka, Yoko Matsumoto, Osamu Wada-Hiraike, Yutaka Osuga & Tomoyuki Fujii, "The histone methyltransferase WHSC1 is regulated by EZH2 and is important for ovarian clear cell carcinoma cell proliferation.", *BMC Cancer*, vol. 19, No. 455, 2019. (DOI: 10.1186/s12885-019-5638-9).
  10. Masayoshi Yamada, Yutaka Saito, Hitoshi Imaoka, Masahiro Saiko, Shigemi Yamada, Hiroko Kondo, Hiroyuki Takamaru, Taku Sakamoto, Jun Sese, Aya Kuchiba, Taro Shibata, Ryuji Hamamoto, "Development of a Real-Time Endoscopic Image Diagnosis Support System Using Deep Learning Technology in Colonoscopy", *Sci Rep*, vol. 9, No. pp. 14465, 2019 (DOI: 10.1038/s41598-019-50567-5).
  11. Matsuda T, Irie T, Katsurabayashi S, Hayashi Y, Nagai T, Hamazaki N, Adefuin AMD, Miura F, Ito T, Kimura H, Shirahige K, Takeda T, Iwasaki K, Imamura T, Nakashima K, "Pioneer Factor NeuroD1 Rearranges Transcriptional and Epigenetic Profiles to Execute Microglia-Neuron Conversion.", *Neuron*, vol. 101, No. 3, pp. 472-485.e7, 2019 (DOI: 10.1016/j.neuron.2018.12.010).
  12. Kazuma Kobayashi, Naoya Murakami, Kana Takahashi, Koji Inaba, Hiroshi Igaki, Ryuji Hamamoto, Jun Itami, "A Population-based Statistical Model for Investigating Heterogeneous Intraprostatic Sensitivity to Radiation Lizio M, Abugessaisa I, Noguchi S, Kondo A, Hasegawa A, Hon CC, de Hoon M, Severin J, Oki S, Hayashizaki Y, et al. 2019. Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res* 47: D752–D758. (DOI: 10.1093/nar/gky1099).
  13. Kazuma Kobayashi, Naoya Murakami, Kana Takahashi, Koji Inaba, Ryuji Hamamoto, Jun Itami, "Local Radiotherapy or Chemotherapy for Oligo-recurrent Cervical Cancer in Patients With Prior Pelvic Irradiation.", *In Vivo*, vol. 33, pp. 1659-1665, 2019. (DOI: 10.21873/invivo.11652).
  14. Toxicity After 125 I Seed Implantation.", *In Vivo*, vol. 33, No. pp. 2103-2111, 2019 (DOI: 10.21873/invivo.11710).
  15. Hiromitsu Araki, Fumihito Miura, Akira Watanabe, Chikako Morinaga, Fumiyo Kitaoka, Yuko Kitano, Noriko Sakai, Yumiko Shibata, Motoki Terada, So Goto, Shinya Yamanaka, "Base-Resolution Methylome of Retinal Pigment Epithelial Cells Used in the First Trial of Human Induced Pluripotent Stem Cell-Based Autologous Transplantation.", *Stem cell reports*, vol. 13, No. 4, pp. 761-774, 2019. (DOI: 10.1016/j.stemcr.2019.08.014).
  16. Asako Kukita, Kenbun Sone, Katsutoshi Oda, Ryuji Hamamoto, Syuzo Kaneko, Masaaki Komatsu, Miku Wada, Harunori Honjoh, Yoshiko Kawata, Machiko Kojima, Shinya Oki, Masakazu Sato, Kayo Asada, Ayumi Taguchi, Aki Miyasaka, Michihiro Tanikawa, Kazunori Nagasaka, Yoko Matsumoto, Osamu Wada-Hiraike, Y

- utaka Osuga, Tomoyuki Fujii, "Histone Methyltransferase SMYD2 Selective Inhibitor LLY-507 in Combination With Poly ADP Ribose Polymerase Inhibitor Has Therapeutic Potential Against High-Grade Serous Ovarian Carcinomas.", *Biochem Biophys Res Commun*, vol. 513, No. 2. pp. 340-346 (DOI: 10.1016/j.bbrc.2019.03.155).
17. Ryuji Hamamoto, Masaaki Komatsu, Ken Takasawa, Ken Asada, Syuzo Kaneko, "Epigenetics Analysis and Integrated Analysis of Multiomics Data, Including Epigenetic Data, Using Artificial Intelligence in the Era of Precision Medicine", *Biomolecules*, vol. 10, No. 1, pp. 62- , 2019. (DOI: 10.3390/biom10010062).
  18. Sangchul Kim, Amina Bolatkan, Syuzo Kaneko, Noriko Ikawa, Ken Asada, Masaaki Komatsu, Shinya Hayami, Hidenori Ojima, Nobutsugu Abe, Hiroki Yamaue, Ryuji Hamamoto, "Deregulation of the histone lysine specific demethylase 1 is involved in human hepatocellular carcinoma.", *Biomolecules*, vol. 9, No. 12, pp. 810-, 2019, (DOI: 10.3390/biom9120810).
  19. Suguru Yasutomi, Tatsuya Arakaki, Ryuji Hamamoto, "Shadow Detection for Ultrasound Images Using Unlabeled Data and Synthetic Shadows", *MIDL 2019, the 2nd International Conference on Medical Imaging with Deep Learning*, 2019.
  20. Wu Z, Rao Y, Zhang S, Kim E, Oki S, Harada H, Cheung M, Jung H. 2019. Cis-control of Six1 expression in neural crest cells during craniofacial development. *Dev Dyn* dvdy.109. (DOI: 10.1002/dvdy.109).
  21. Yusuke Matsuno, Yuko Atsumi, Atsuhiko Shimizu, Kotoe Katayama, Haruka Fujimori, Mai Hyodo, Yusuke Minakawa, Yoshimichi Nakatsu, Syuzo Kaneko, Ryuji Hamamoto, Teppei Shimamura, Satoru Miyano, Teruhisa Tsuzuki, Fumio Hanaka, Ken-ichi Yoshioka, "Replication stress triggers microsatellite destabilization and hypermutation leading to clonal expansion in vitro", *Nature Communications*, vol. 10, No. 2, pp. 3925, 2019 (DOI: 10.1038/s41467-019-11760-2).
  22. Zhaoming Wu, Yanxia Rao, Sushan Zhang, Eun-Jung Kim, Shinya Oki, Hidemitsu Harada, Martin Cheung, Han-Sung Jung, "Cis-control of Six1 expression in neural crest cells during craniofacial development.", *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 248, No. 12, pp. 1264-1272, 2019. (DOI:10.1002/dvdy.109).
  23. Eli Kaminuma, Yukino Baba, Masahiro Mochizuki, Hirotaka Matsumoto, Haruka Ozaki, Toshitsugu Okayama, Takuya Kato, Shinya Oki, Takatomo Fujisawa, Yasukazu Nakamura, Masanori Arita, Osamu Ogasawara, Hisashi Kashima, Toshihisa Takagi, "DDBJ Data Analysis Challenge: a machine learning competition to predict Arabidopsis chromatin feature annotations from DNA sequences.", *Genes & genetic systems*, vol. 95, No. 1, pp.43-50, 2020 (DOI:10.1266/ggs.19-00034).
  24. Hideaki Isago, Akihisa Mitani, Yu Mikami, Masafumi Horie, Hirokazu Urushiyama, Ryuji Hamamoto, Yasuhiro Terasaki, Takahide Nagase, "The Epithelial Expressions of YAP and TAZ Are Sequentially Required in Lung Development.", *Am J Respir Cell Mol Biol*, vol. 62, No. 2, pp.256-266, 2020. (DOI: 10.1165/rcmb.2019-02180C).
  25. Ken Asada, Kazuma Kobayashi, Samuel Joutard, Masashi Tubaki, Satoshi Takahashi, Ken Takasawa, Masaaki Komatsu, Syuzo Kaneko, Jun Sese, Ryuji Hamamoto, "Uncovering Prognosis-related Genes and Pathways by Multi-omics Analysis in Lung Cancer.", *Biomolecules*, vol. 10, No. 4, pp. 524-, 2020. (DOI: 10.3390/

- biom10040524).
26. Sou Hirose, Naoya Murakami, Kazuaki Takahashi, Ikumi Kuno, Daisuke Takayanagi, Yuka Asami, Maiko Matsuda, Yoko Shimada, Shotaro Yamano, Kuniko Sunami, Kazushi Yoshida, Takayuki Honda, Tomomi Nakahara, Tomoko Watanabe, Masaaki Komatsu, Ryuji Hamamoto, "Genomic Alterations in STK11 Can Predict Clinical Outcomes in Cervical Cancer Patients.", *Gynecol Oncol*, vol. 156, No. 1, pp. 203-210, 2020. (DOI: 10.1016/j.ygyno.2019.10.022).
  27. Yoshida K, Maekawa T, Ly NH, Fujita SI, Muratani M, Ando M, Katou Y, Araki H, Miura F, Shirahige K, Okada M, Ito T, Chatton B, Ishii S, "ATF7-Dependent Epigenetic Changes Are Required for the Intergenerational Effect of a Paternal Low-Protein Diet.", *Molecular Cell*, vol. 78 No.3, pp. 445-458 e6, 2020. (DOI:10.1016/j.molcel.2020.02.028).
  28. Federico Pecori, Ikuko Yokota, Hisatoshi Hanamatsu, Taichi Miura, Chika Ogura, Hayato Ota, Jun-ichi Furukawa, Shinya Oki, Kazuo Yamamoto, Osamu Yoshie, Shoko Nishihara, "A defined glycosylation regulatory network modulates total glycome dynamics during pluripotency state transition", *Sci Rep* 11(1), 2020, (DOI:10.1038/s41598-020-79666-4).
  29. Yoichi Miyamoto, Mitsuho Sasaki, Haruhiko Miyata, Yoko Monobe, Masahiro Nagai, Mayumi Otani, Penny A. F. Whiley, Akane Morohoshi, Shinya Oki, Junichiro Matsuda, Ken-ichi Akagi, Jun Adachi, Masaru Okabe, Masahito Ikawa, Yoshihiro Yoneda, Kate L. Loveland, Masahiro Oka, "Genetic loss of importin  $\alpha 4$  causes abnormal sperm morphology and impacts on male fertility in mouse", *FASEB J* 34(12) 16224-16242, 2020 (DOI:10.1096/fj.202000768RR).
  30. Mayumi Hirayama, Fan-Yan Wei, Takeshi Chujo, Shinya Oki, Maya Yakita, Daiki Kobayashi, Norie Araki, Nozomu Takahashi, Ryoji Yoshida, Hideki Nakayama, Kazuhito Tomizawa, "FTO Demethylates Cyclin D1 mRNA and Controls Cell-Cycle Progression", *Cell Rep* 31(1), 2020 (DOI:10.1016/j.celrep.2020.03.028)
  31. Kaichi Yoshizaki, Ryuichi Kimura, Hisato Kobayashi, Shinya Oki, Takako Kikkawa, Lingling Mai, Kohei Koike, Kentaro Mochizuki, Hitoshi Inada, Yasuhisa Matsui, Tomohiro Kono, Noriko Osumi, "Paternal age affects offspring via an epigenetic mechanism involving REST/NRSF", *EMBO Rep*, 22, e51524, 2021 (DOI:10.15252/embr.202051524).

## 2. その他の著作物(総説、書籍など)

1. 沖 真弥、大田達郎: ChIP-Atlas:既報の ChIP-seq データをフル活用するためのウェブサービス. *実験医学*, 37, 2763-2760, 2019
2. 沖 真弥、大田達郎: ChIP-Atlas:公共 ChIP-seq データを統合的に活用するためのウェブサービス. *THE LUNG perspectives*, 27, 243-249, 2019
3. 三浦史仁、伊藤隆司. バイオインフォメーションに向けて~バイオテクノロジーの新技術からの新しい視点~. *シングルセルエピゲノム解析*. 2019
4. 三浦史仁、伊藤隆司. エピゲノム解析手技の標準化:全ゲノムバイサルファイトシーケンシング. *実験医学増刊*. 34(10). 2018

## 3. 国際学会発表及び主要な国内学会発表

### (1) 概要

種別	国内外	件数
招待講演	国内	50 件
	国際	2 件
口頭発表	国内	10 件
	国際	0 件
ポスター発表	国内	4 件
	国際	3 件

## (2) 招待講演

〈国内〉

1. 沖 真弥「時空間的な遺伝子発現制御のしくみを探る」日本生物工学会東日本支部 生物工学フォーラム, オンライン, 2021/8/27
2. 沖 真弥「空間的な遺伝子発現制御のしくみを探る」第 85 回日本循環器学会学術集会, パシフィコ横浜, 2021/3/27
3. 沖 真弥「Data-driven and technical approaches to understand spatial gene regulation」第 43 回日本分子生物学会年会, オンライン, 2020/12/3
4. 沖 真弥「ChIP-Atlas: 公共 ChIP-seq データを利活用できる」第 43 回日本分子生物学会年会, オンライン, 2020/12/2
5. 沖 真弥「ChIP-seq ビッグデータを統合解析し、薬剤の作用機序解明に迫る」京都大学「医学領域」産学連携推進機構/一般社団法人芝蘭会 産学情報交流会, オンライン, 2020/9/10
6. 沖 真弥「ChIP-Atlas の使い方と使われ方」第 9 回生命医薬情報学連合大会 (IIBMP2020), オンライン, 2020/9/3
7. 沖 真弥「空間的な遺伝子発現制御のしくみを探る」第 9 回生命医薬情報学連合大会 (IIBMP2020), オンライン, 2020/9/2
8. 沖 真弥「遺伝子発現の時空間的な制御のしくみ」KBC 第 4 回勉強会, 九州大学 病院キャンパス, 2020/2/19
9. 沖 真弥「Data-driven and technical approaches to understand spatial gene regulation」WPI-IHIS Seminar, 筑波大学国際統合睡眠医科学研究機構, 2020/1/20
10. 沖 真弥「Data-driven and technical approaches to understand spatial gene regulation」RIKEN BDR seminar in Kobe, RIKEN BDR (神戸市), 2020/1/16
11. 沖 真弥「公共 ChIP-seq データをフル活用し、遺伝性疾患の解明や創薬に挑む」次世代脳プロジェクト冬のシンポジウム 2019, 一橋講堂 (東京都千代田区), 2019/12/19
12. 沖 真弥「ChIP-Atlas: 公共 ChIP-seq データを利活用できる」第 42 回日本分子生物学会年会, 福岡国際会議場, 2019/12/3
13. 沖 真弥「位置情報と遺伝子発現とその仕組み」新学術領域「個性創発脳」若手の会・技術支援講習会, 岡崎コンファレンスセンター, 2019/11/14
14. 沖 真弥「Annotathon 2019」Annotathon 2019, DBCLS 柏の葉キャンパス, 2019/11/12
15. 沖 真弥「ChIP-Atlas をつないで使う」トーゴーの日シンポジウム 2019, 日本科学未来館,

2019/10/5

16. 沖 真弥「ChIP-Atlas: 既報 ChIP-seq データの統合データベース」統合データベース講習会:AJACS 番町3, JST 東京本部, 2019/8/7
17. 沖 真弥「ChIP-Atlas: 既報 ChIP-seq データの統合データベース」日本プロテオーム学会 2019 年大会, シーガイア コンベンションセンター, 2019/7/26
18. 沖 真弥「公共 ChIP-seq データの利活用術」国立精神・神経医療研究センター セミナー, 国立精神・神経医療研究センター, 2019/6/25
19. 沖 真弥「公共 ChIP-seq データの統合解析」京都大学 セミナー, 京都大学大学院 医学研究科 創薬医学講座, 2019/6/12
20. 沖 真弥「ChIP-Atlas の使い方とその応用」資生堂 セミナー, 資生堂リサーチセンター, 2019/5/22
21. 沖 真弥「薬効の作用点となる転写因子の特定と創薬への応用」日本たばこ医薬総合研究所 セミナー, 日本たばこ医薬総合研究所, 2019/5/16
22. 沖 真弥「公共 ChIP-seq データをフル活用し、遺伝性疾患の解明や創薬に挑む」質量分析インフォマティクス研究会, JST 東京本部, 2019/3/19
23. 沖 真弥「公共 ChIP-seq データの利活用術」新潟大学 セミナー, 新潟大学, 2019/3/4
24. 沖 真弥「公共 ChIP-seq データの利活用術」産総研 セミナー, 産総研, 2019/2/20
25. 沖 真弥「公共 ChIP-seq データの利活用術」全薬工業 セミナー, 全薬工業株式会社, 2019/1/23
26. 沖 真弥「公共 ChIP-seq データの利活用術」創価大学 セミナー, 創価大学, 2019/1/18
27. 沖 真弥「公共 ChIP-seq データの利活用術」東京理科大学 セミナー, 東京理科大学, 2019/1/17
28. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」北里大学 セミナー, 北里大学, 2019/1/11
29. 沖 真弥「公共 ChIP-seq データをフル活用し、遺伝性疾患の解明に挑む」JSBi 九州部会, 九州大学, 2018/12/17
30. 沖 真弥「ChIP-Atlas: 既報の ChIP-seq データをフル活用できる」統合データベース講習会:AJACS 町田, 協和発酵キリン株式会社, 2018/12/14
31. 沖 真弥「医工学研究ビッグデータのデータベース構築と活用」東京大学 工学系研究科 医工学概論, 東京大学, 2018/12/14
32. 沖 真弥「公共 ChIP-seq データと疾患関連 SNP の統合解析」第 41 回日本分子生物学会年会, パシフィコ横浜, 2018/11/28
33. 沖 真弥「公共 ChIP-seq データをフル活用し、遺伝性疾患の解明に挑む」第 16 回 JCGG シンポジウム, 東京大学, 2018/11/27
34. 川路英哉「FANTOM5 で構築されたヒトのトランスクリプトーム・アトラス」第 9 回核酸医薬レギュラトリーサイエンスシンポジウム, 東京, 2018 年 1 月 29 日
35. 沖 真弥, 石濱泰「ChIP-seq とプロテオーム:公共データをつないで使う」トーゴの日シンポジウム 2018, 日本科学未来館, 2018/10/5
36. 沖 真弥「公共 ChIP-seq データをフル活用し、遺伝性疾患の解明に挑む」九州工業大学 セミナー, 九州工業大学, 2018/9/7
37. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」医薬基盤・健

- 康・栄養研究所 セミナー, 医薬基盤・健康・栄養研究所, 2018/4/17
38. 沖 真弥, 大田 達郎, 塩井 剛, 畠中 秀樹, 小笠原 理, 奥田 喜広, 川路 英哉, 仲木 竜, 瀬々 潤, 目野 主税「公共 ChIP-seq データをフル活用し、遺伝性疾患の解明に挑む」第 17 回日本再生医療学会総会, パシフィコ横浜, 2018/3/21
  39. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」旭川医科大学セミナー, 旭川医科大学, 2018/3/12
  40. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」慶應義塾大学セミナー, 慶應義塾大学 医学部キャンパス, 2018/1/30
  41. 沖 真弥「公共 ChIP-seq データのフル活用術～組織特異的エンハンサーと遺伝性疾患の解析～」東京農業大学セミナー, 東京農業大学, 2018/1/23
  42. 沖 真弥「公共 ChIP-seq データのフル活用術～組織特異的エンハンサーと遺伝性疾患の解析～」岩手医科大学セミナー, 岩手医科大学, 2018/1/18
  43. 沖 真弥「生命科学のデータベース活用法フォーラム :ChIP-Atlas」ConBio 2017, 神戸国際会議場, 2017/12/9
  44. 沖 真弥「公共 ChIP-seq データのフル活用術」新学術領域「個性創発脳」若手の会・技術支援講習会, 京都大学薬友会館, 2017/11/21
  45. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」名古屋大学医学系研究科セミナー, 名古屋大学 医学系研究科, 2017/11/13
  46. 沖 真弥「エピゲノミクス統合データベースの開発と機能拡充」トーゴの日シンポジウム 2017, 東京大学弥生講堂, 2017/10/5
  47. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」大阪大学生命機能研究科セミナー, 大阪大学生命機能研究科, 2017/9/7
  48. 沖 真弥「公共 ChIP-seq データのフル活用術 組織特異的エンハンサーと遺伝性疾患の解析」徳島大学藤井節郎記念医科学センターセミナー, 徳島大学藤井節郎記念医科学センター, 2017/8/28
  49. 沖 真弥「公共 ChIP-seq データのフル活用術」新学術領域「個性創発脳」第2回領域会議, 静岡県御殿場市, 2017/7/7
  50. 沖 真弥「公共 ChIP-seq データのフル活用術」東京大学 医学系研究科セミナー, 東京大学 医学系研究科, 2017/7/6

〈国際〉

1. 沖 真弥「ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data」, Yonsei University, 2019/10/24
2. 沖 真弥「ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data」, Yonsei University, 2019/3/13

### (3) 口頭講演

〈国内〉

1. 新海 典夫、金子 修三、浅田 健、高澤 建、浜本 隆二:Oncogenic TF 情報を取り入れたオミックス解析プラットフォームの構築 第 3 回日本メディカル AI 学会学術集会(オンライン)2021 年6月 11 日
2. 新海 典夫、金子 修三、浅田 健、高澤 建、浜本 隆二:臨床応用を志向した ChIP-seq



データセットによる機械学習分析プラットフォーム開発 第80回日本癌学会学術総会(横浜)2021年9月30日

3. 三浦史仁「高効率な1本鎖 DNA ライゲーション技術を用いた高品質な全ゲノムバイサルファイトシーケンシング」シングルセルゲノミクス研究会, 柏の葉カンファレンスセンター, 2019/08/29
4. 浜本 隆二「臨床応用を志向した人工知能技術を活用した統合的ながん医療システムの開発」第1回日本メディカル AI 学会, 国立がん研究センター研究所, 2019/1/26
5. 沖 真弥「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」第4回 包括的神経グリア研究会, 熊本大学 医学部キャンパス, 2018/1/6
6. 沖 真弥「ChIP-Atlas: Make full use of public ChIP-seq data」CREST ミーティング「離散構造統計学の創出と癌科学への展開」, 名古屋工業大学, 2017/12/22
7. 沖 真弥, 大田 達郎, 塩井 剛, 畠中 秀樹, 小笠原 理, 奥田 喜広, 川路 英哉, 仲木 竜, 瀬々 潤, 目野 主税「公共 ChIP-seq データをフル活用し遺伝性疾患の解明に挑む」ConBio 2017, 神戸国際会議場, 2017/12/7
8. 沖 真弥「公共 ChIP-seq データのフル活用術」第2回次世代生命科学の研究会, 九州大学 病院キャンパス, 2017/7/13
9. Shinya Oki, Tazro Ohta, Go Shioi, Hideki Hatanaka, Osamu Ogasawara, Yoshihiro Okuda, Hideya Kawaji, Ryo Nakaki, Jun Sese, Chikara Meno「Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions」50th Annual Meeting of the Japanese Society of Developmental Biologists, タワーホール船堀, 2017/5/1
10. 川路英哉「国際的ゲノミクスデータ統合のアジア拠点: The Asian Mirror of the UCS C Genome Browser Database」2017 年度生命科学系学会合同年次大会, 神戸, 2017 年 12 月 9 日

〈国際〉

該当無し

#### (4) ポスター発表

〈国内〉

1. 三浦史仁, 柴田由希子, 三月田祐平, 三浦美希, 伊藤隆司、「改良型 PBAT 法による高品質で低コストなメチロームシーケンシング」, 第 12 回日本エピジェネティクス研究会年会, 札幌
2. 三浦史仁, 柴田由希子, 三浦美希, 三月田祐平, 久野修, 荒木啓充, 伊藤隆司、「1本鎖 DNA ライゲーション技術に基づく PBAT と両鎖混合ライブラリー法による高品質で低コストなメチローム解析」, 第 41 回日本分子生物学会年会, 横浜
3. 川路英哉「国際的ゲノミクスデータ統合のアジア拠点: The Asian Mirror of the UCS C Genome Browser Database」トーゴの日シンポジウム 2017, 東京, 2017 年 10 月 5 日
4. 沖 真弥, 大田 達郎, 塩井 剛, 畠中 秀樹, 小笠原 理, 奥田 喜広, 川路 英哉, 仲木 竜, 瀬々 潤, 目野 主税「公共 ChIP-seq データで遺伝性疾患の解明に挑む」NGS 現場の会第五回研究会, 仙台国際センター, 2017/5/22

〈国際〉

1. Shinya Oki, Mizuki Honda, Akihito Harada, Kazumitsu Maehara, Kaori Tanaka, and Yasuyuki Ohkawa「High resolution spatial transcriptomics method by photo-isolation chemistry」EMBO/EMBL Symposium: Multiomics to Mechanisms - Challenges in Data Integration, EMBL Heidelberg, Germany, 2019/9/11
2. Shinya Oki「Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions」Towards Understanding “INDIVIDUALITY”, 京都大学, 2018/7/24
3. Shinya Oki, Tazro Ohta, Go Shioi, Hideki Hatanaka, Osamu Ogasawara, Yoshihiro Okuda, Hideya Kawaji, Ryo Nakaki, Jun Sese, Chikara Meno 「Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions」Keystone Symposium; Gene Control in Development and Disease, Whistler Conference Centre (Whistler, British Columbia, Canada) 2018/3/23

#### 4. 知財出願

##### (1) 出願件数

種別	件数
特許出願	国内 0 件
	国外 0 件
その他の知的財産出願	0 件

##### (2) 一覧

###### ① 国内出願

該当無し

###### ② 海外出願

該当無し

###### ③ その他の知的財産権

該当無し

#### 5. 受賞・報道等

##### (1) 受賞

該当無し

##### (2) メディア報道

該当無し

##### (3) その他

該当無し

## §8. 研究開発期間中の活動

### 1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2018.07.18	チーム内ミーティング (非公開)	国立がんセンター	3人	研究進捗報告のためのミーティング
2019.01.25	チーム内ミーティング (非公開)	国立がんセンター	3人	研究進捗報告のためのミーティング
2020.01.31	チーム内ミーティング (非公開)	国立がんセンター	3人	研究進捗報告のためのミーティング

### 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

該当なし

以上

## 別紙 研究開発対象のデータベース等

No.	研究開発課題名	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDPポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	エピゲノミクス統合データベースの開発と機能拡充	ChIP-Atlas		ChIP-seq, ATAC-seq, DNase-seq, Bisulfite-seqデータを統合したエピゲノミクスデータベース、およびその閲覧や利活用を支援するウェブツール	<a href="https://chip-atlas.org">https://chip-atlas.org</a>	2015年12月1日	継続・発展	データベース兼ツール等	公開済	対象外	公開済	<a href="https://doi.org/10.15252/embr.201846255">https://doi.org/10.15252/embr.201846255</a>