

## 研究開発実施報告

### □概要

研究開発課題名	エピゲノミクス統合データベースの開発と機能拡充
開発対象データベースの名称 (URL)	ChIP-Atlas ( <a href="https://chip-atlas.org/">https://chip-atlas.org/</a> )
研究代表者氏名	沖 真弥
所属・役職	京都大学 大学院医学研究科 特定准教授 (2021年3月時点)

### □目次

§ 1. 研究実施体制 .....	2	③ 口頭講演 .....	8
§ 2. 研究開発対象とするデータベース・ツール等 .....	3	④ ポスター発表 .....	8
(1) データベース一覧 .....	3	(4) 知的財産権の出願 (国内の出願件数のみ公開) .....	8
(2) ツール等一覧 .....	3	① 出願件数 .....	8
§ 3. 実施内容 .....	4	② 一覧 .....	8
(1) 本年度の研究開発計画と達成目標 .....	4	(5) 受賞・報道等 .....	9
(2) 進捗状況 .....	4	① 受賞 .....	9
§ 4. 成果発表等 .....	7	② メディア報道 .....	9
(1) 原著論文発表 .....	7	③ その他の成果発表 .....	9
① 論文数概要 .....	7	§ 5. 研究開発期間中に主催した活動(ワークショップ等) .....	9
② 論文詳細情報 .....	7	1. 進捗ミーティング .....	9
(2) その他の著作物(総説、書籍など) .....	7	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 .....	9
(3) 国際学会および国内学会発表 .....	7		
① 概要 .....	7		
② 招待講演 .....	8		



## §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
沖グループ	沖 真弥	京都大学・特定准教授	ChIP-Atlas の継続的運営と機能拡充
三浦グループ	三浦 史仁	九州大学・講師	Bisulfite-seq データの追加
浜本グループ	浜本 隆二	国立がん研究センター研究所・分野長	サンプルメタデータのキュレーション

## §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	ChIP-Atlas		<a href="https://chip-atlas.org">https://chip-atlas.org</a>

#### 【その他のデータベース】

No.	名称	別称・略称	URL
1			
2			
3			
4			
5			

### (2) ツール等一覧

No.	名称	別称・略称	URL
1			
2			
3			

### §3. 実施内容

#### (1) 本年度の研究開発計画と達成目標

##### ■Bisulfite-seq データの追加

三浦グループが開発したマッピングツール BMap を中心としたプログラム群を用いて Bisulfite-seq データを解析し、サンプルメタデータのキュレーションを行なった上で 2021 年度内の公開をめざす。

##### ■ATAC-seq データの追加

DNase-seq データの解析パイプラインを転用し、同じくオープンクロマチン領域を同定するための実験手法である ATAC-seq データを収集・解析し、サンプルメタデータのキュレーションを行なった上で 2021 年度内の公開をめざす。

##### ■最新ゲノムアセンブリへの対応

ヒトゲノムアセンブリを hg19 だけでなく hg38 にも対応させるなど、最新ゲノムアセンブリによる解析データを作成し、2021 年度内の公開をめざす。

##### ■キュレーション履歴の再検討

これまでに行われた ChIP-Atlas のキュレーション履歴を医学的観点より再検討する。

##### ■毎月のデータ更新

毎月行われるデータソース (NCBI SRA) の更新に合わせ、ChIP-Atlas データを更新する。また Bisulfite-seq や ATAC-seq データが追加されるため、それらのキュレーションも遂行する。

##### ■他のデータベースとの連携

ChIP-Atlas データの RDF 化を進めることにより、ChIP-Atlas 以外のバイオデータベースとの連携を進める。

##### ■利用者増加のための広報活動

学会や研究集会において、ChIP-Atlas やその利活用事例の発表とともに、参加者へのチラシの配布をおこなうことで、周知活動をおこなう。

#### (2) 進捗状況

##### ■Bisulfite-seq データの追加

これまでの開発において、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プログラム群を用いた解析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、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進めるとともに、NBDC 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進めるとともに、NBDC 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。

### ■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 (61,679 件)

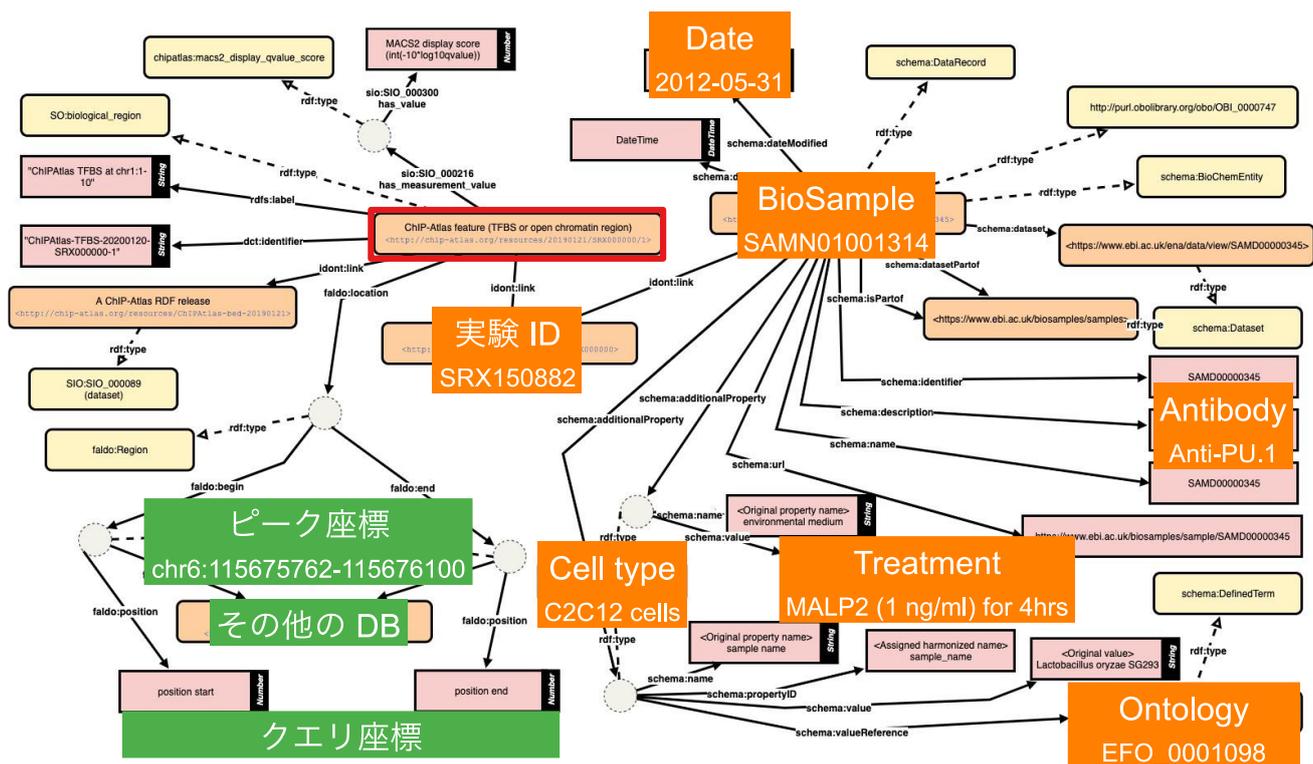
マウス：mm9 → mm10 (63,337 件)

ハエ：dm3 → dm6 (6,787 件)

線虫：ce10 → ce11 (3,253 件)

### ■他のデータベースとの連携

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。



### ■キュレーション履歴の再検討

浜本グループが中心となり、既存の ChIP-seq と DNase-seq 実験のサンプルメタデータのキュレーション履歴の見直しをすべて完了した。また現在、ヒト、マウス、ラット、ハエ、線虫、酵母の ATAC-seq データ (107,884 件) と Bisulfite-seq データ (60,034 件) のキュレーションを沖グループと共同で進めている。

### ■毎月のデータ更新

ChIP-Atlas は原則として毎月データを更新しており、配列生データの収集、アライメント、ピークコール、サンプルメタデータのキュレーション、そして全データの統合解析によるデータマイニングをおこなっている。2020 年度内は 12,889 件の ChIP-seq および DNase-seq データについてこのような作業を完遂した。

### ■利用者増加のための広報活動

これまで同様、さまざまな学会や研究集会で ChIP-Atlas に関する発表をおこなった。とくに研究代表者 (沖) は学会や研究会などにおいて 6 回の口頭発表をおこなった。このような地道な活動の結果、着実に知名度をあげることができており、昨年の約 2 倍以上となる 192 報の学術論文に引用された (<http://chip-atlas.org/publications> において被引用論文を公開)。

## §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)	0件
	国際(欧文)	5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

(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. 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.", Scientific Reports, vol. 11, No. 1, 2021 (DOI: 10.1038/s41598-020-79666-4).
2. 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 reports, vol. 22, No. 2, 2021 (DOI: 10.15252/embr.202051524).
3. Yoichi Miyamoto, Mitsuho Sasaki, Haruhiko Miyata, Yoko Monobe, Masahiro Nagai, Mayumi Otani, Penny A. F. Whaley, 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.", The FASEB Journal, vol. 34, No. 12, pp. 16224-16242, 2020 (DOI: 10.1096/fj.202000768rr).
4. 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 DNase-seq sequences.", Genes & genetic systems, vol. 95, No. 1, pp. 43-50, 2020 (DOI: 10.1266/ggs.19-00034, PMID:32213716).
5. 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 REPORTS, vol. 31, No. 1, pp. 107464-107464, 2020 (DOI: 10.1016/j.celrep.2020.03.028, Web of Science ID:WOS:000524976500007).

### (2) その他の著作物(総説、書籍など)

なし

### (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	6件
	国際	0件
口頭発表	国内	0件
	国際	0件
ポスター発表	国内	0件
	国際	0件

## ② 招待講演

〈国内〉

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

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、Web 発表、2020/12/3

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、Web 発表、2020/12/2

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、Web 発表、2020/9/10

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)、Web 学会、2020/9/3

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)、Web 学会、2020/9/2

〈国際〉

該当なし

## ③ 口頭講演

〈国内〉

該当なし

〈国際〉

該当なし

## ④ ポスター発表

〈国内〉

該当なし

〈国際〉

該当なし

## (4) 知的財産権の出願

### 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

該当なし

#### 2) 海外出願

該当なし

### 3) その他の知的財産権

該当なし

#### (5) 受賞・報道等

① 受賞

該当なし

② メディア報道

該当なし

③ その他の成果発表

該当なし

#### §5. 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
ほぼ毎月	チーム内ミーティング(非公開)	オンライン	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	ChIP-Atlas		<p>論文などで報告された ChIP-seq データの可視化と解析を行うサイトです。公開 NGS データレポジトリ (NCBI, EMBL-EBI, DDBJ) に登録されたほぼ全ての ChIP-seq データをデータソースとしています。</p> <p>ChIP-Atlasは以下の4つのサービスで構成されています。</p> <p>1) Peak Browser ChIP-seq データをゲノムブラウザ上に表示し、何処どこに結合しているかを一目で分かるようにしています。</p> <p>2) Target Genes 転写因子からターゲット遺伝子を予測します。</p> <p>3) Colocalization 転写因子から共局在する相手のタンパク質を予測します。</p> <p>4) Enrichment Analysis 既存データを使ってユーザデータの解析を行います。似たようなChIP-seq データの探索、指定したモチーフに結合するタンパク質の探索、指定した遺伝子に結合するタンパク質の探索などができます。</p>	<a href="https://chip-atlas.org/">https://chip-atlas.org/</a>	2015年	維持・発展	データベース等	Okii, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J., and Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep. e46255.