


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)

研究開発中間報告書

「疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層
オーミクスデータの統合」

菅野 純夫 | 東京医科歯科大学 難治疾患研究所
非常勤講師

 ©2019 菅野 純夫(東京医科歯科大学) Licensed under CC BY 4.0

§1. 研究開発実施の概要

本データベースは、ヒトゲノム多型・変異に生物学的機能注釈を与えるべく、ゲノム変異位置の近傍の多層オーミクスデータ(エピゲノム:ヒストン修飾、DNA メチル化パターン;トランスクリプトーム:発現量、スプライスパターン)をあたう限り広範囲で取得し、日本人ゲノム情報への統合を試みたものである。

これまでにヒトゲノム多型情報については、ヒトゲノムバリエーションデータベース(HGVDB)と、実体としての統合を完了した。統合対象となった HGVDB データには種々の疾患ゲノムの GWAS 解析結果および HLA 領域関連データが含まれている。特に HLA データは昨今、がん免疫の分野においてその整備が喫緊の課題となっているもののひとつである。また、東北メディカル・メガバンク、BioBank Japan とも連携を図り、ゲノム情報についても本邦で産生されたデータの横断的な 2 次利用を促進するべくデータの統合を開始した。また健常者多層オーミクスデータについては、CREST-IHEC データの収載を完了している。

特に、申請者らは個別研究あるいは製薬企業との共同研究において、ヒトがん細胞薬剤応答オーミクス摂動大規模データ、および各種プラットフォーム(フリューダイン C1、10X Genomics Chromium、独自 Nx1 システムによる)を用いたシングルセルデータを集積している。これらに、第二期が開始された新学術領域「先進ゲノム支援」から産出されるマウス等モデル生物種データを加え、上記のブラウザで統合表示することを可能とした。全体として、がんを主標的として、臨床検体では直接の集積が困難な多層オーミクス情報を充実させ、さらには国内研究で見出されつつある種々の日本人疾患ゲノム情報に生物学的機能解析を与える際に、実験的検証の場としてのモデル系におけるオーミクス情報の整備を試行している。

§2. 研究開発実施体制

1. 各グループの担当項目

(1) 「菅野(医科歯科)」グループ(研究代表者グループ)

データベースについて研究の統括を行った。各グループ間での進度の調整を行い、連携を促進した。また各グループで共通して必要となる計算機環境を整備し、その環境維持に努めた。

(2) 「菅野(東大)」グループ(主たる共同研究者グループ(1))

各種のデータセットの収集と加工を行い、データベースとして公開する主体を担っている。先進ゲノム支援、IHEC データについては、すでに内部的に格納されているものを用いた。d bSNP、ExAC 等の外部参照データについては、前期開発期間内に開発したツール群を用いて、定期的、半自動的作業でデータをできる限り、最新の情報に保っている。また、シングルセル、薬剤摂動オーミクスデータについては、関連解析ツール、ブラウザを開発した。前半期開発期間におこなった開発を発展的に継続する形でDBCLSの担当者と連絡を密にして、多層オーミクスデータの RDF 化スキーマの試用を進めている。特に、今後、実験条件、試料臨床情報(非公開)に関するメタデータオントロジーの整備と RDF 化は集中的に行う予定である。

(3) 土原グループ(主たる共同研究者グループ(2))

培養がん細胞多層オーミクスデータ、薬剤摂動データ、シングルセルデータの収載を行った。またがん臨床データについての整備と秘匿検索についての技術開発を模索した。今後、公開が予定されるがんゲノムデータについても二次利用としてデータベースに収載すべく検討を行う。

(4) 徳永グループ(主たる共同研究者グループ(3))

統合データベースプロジェクトおよび統合化推進プログラムにおいて構築してきた「ヒトゲノムバリエーションデータベース(HGVDB)」(日本人の健常者 SNP-DB、GWAS-DB、健常者 CNV-DB、疾患関連 CNV-DB、HLA-DB、統合化変異 DB を含む)を継承、移管し、登録データを一層充実させるとともに、より活用しやすいデータベースを目指して機能改善を行った。

(5) 井ノ上グループ(主たる共同研究者グループ(4))

統合データベースプロジェクトおよび統合化推進プログラム「ヒトゲノムバリエーションデータベース(HGVDB)」、特に国立遺伝学研究所で1次データ産生が行われてきたその HLA 多型部分について、徳永グループと共同して DBKERO へとデータを継承、移管し、登録データを一層充実させるとともに、より活用しやすいデータベースを目指して機能改善を行った。

2. 有識者会議等

(1) 会議概要

名称	DBKERO 産学利用促進委員会
目的	データコンテンツの充実と閲覧手法の改善をし、DBKERO の産学の利用を促

	進する。
委員数	6人

(2) 開催歴

年月日	場所	主な議題・指摘事項等
2019年 1月24日	東京大学医科学研究所・ 総合研究棟 8F 会議室	データコンテンツの充実と他のデータベースとの連携について
2019年 9月25日	東京大学大学院新領域創 成科学研究科情報生命科 学実験棟 2F 会議室	DBKERO を利用した創薬シーズ探索の可能性について

§3. 研究開発の目的、実施内容及び成果

1. 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
DBKERO: 疾患ゲノム多層オーミクス機能注釈データベース	DBKERO	本データベースではヒトゲノム多型・変異に生物学的機能注釈を与えるべく、ゲノム変異位置、近傍のエピゲノム(ヒストン修飾、DNAメチル化パターン)、トランスクリプトーム情報(発現量、スプライスパターン)をヒトゲノム情報に統合する。

上記以外のもの

該当なし

(2) ツール等

正式名称	略称	概要
BEACON SCAN	なし	ゲノムデータの統合検索システム
DBKERO RDF viewer	なし	RDF 検索による統合検索を受入れ、検索結果を表示する枠組み
Transcription Factor Binding Site Search	なし	ユーザーが指定するゲノム上のターゲット領域から、転写因子の結合領域を、ChIP-Seq の結果や配列パターンから予測するプログラム

2. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

初年度

- ・IHEC-CREST において、現在までに日本チームが集積、JGA からその配列情報が公開されているエピゲノム情報について、そのシグナル強度情報を全件収載、即時、公開する。その実務の大半を初年度に完了する。
- ・HGVDDB については、現在までに徳永グループから公開されているデータの全てを本データベースに移管する。これには遺伝研担当分である HLA 領域データ、現在格納されている希少疾患変異データセットを含む。初年度は HLA 領域データを中心にその 2/3 の移設を行う。これを検討材料に DBKERO との接続部のパイプライン、ブラウザ等についての関連プログラム群の整備を行う。

～第3年次末

- ・IHEC-CREST データについて、その公開データベースの動作確認および種々の最終確認

を行う。この時点でのアップデートを行うことで CREST-IHEC の最終成果とする。また IHEC の他国産出関連データについては IHEC データポータルについてリンクを整備する。データの具体的な統合についてもその実務を開始する。

・HGVDDB について、残り 2/3 のデータを移設し、全データについて最終確認を行う。データベースを公開して最終動作確認を行う。

・東北メディカル・メガバンクと、データ公開条件等についての付帯条件を整理し、本データベースに収載する。

(2) 期間中に追加・削除・変更した実施計画・達成目標

学会ブース出展

臨床系、ユーザー系の学会での発表を行うことでユーザーを発掘する取り組みを強化する。

- ・日本がん学会(9月28～30日:横浜)
- ・日本人類遺伝学会(11月15～18日:神戸)
- ・Human Genome Meeting, Yokohama(3月13～15日:横浜)

の各学会において本データベースに関する一連の発表を行い、実務担当者が同行してデータベースのさらに詳細な説明、データベースを活用した共同研究について提案を行う。それぞれの学会に派遣する3名の随行員(発表を行う鈴木、土原の旅費は別の研究費で負担される)は若手実務担当者から選抜する。研究室内部において、本データベースを活用して独自の研究を遂行する大学院学生についても、その同行させ、その活用例について具体例を提示する。その準備として、各学会趣旨に呼応した形で、データベースの切り分けを行い、広く簡易的な模擬検索が可能な形でデータベースサブディビジョンを構成する。これにより現在、煩雑となっているデータベース全体についての説明の必要性を回避し、利用者の利用開始への障壁を低減する。

3. 実施内容

(1) 実施内容と達成状況

上記に掲げた研究開発予定項目については概ね当初の目標を達成できたと考えている。

2017年度

本データベースは、ヒトゲノム多型・変異に生物学的機能注釈を与えるべく、変異近傍の多層オームクスデータの統合を目指したものである。特に、培養細胞やマウスなどの動物モデル系から得られたデータに特色を求める。これにより臨床検体で乏しいエピゲノム情報を充実、同時に生物学的機能解析の場としてのモデル系のオームクス情報の整備を行うものである。

初年度に掲げた達成目標は:

- ・IHEC-CREST において、現在までに日本チームが集積、JGA からその配列情報が公開されているエピゲノム情報について、そのシグナル強度情報を全件収載、即時、公開する。その実務の大半を初年度に完了する。
- ・HGVDDB については、現在までに徳永グループから公開されているデータの全データを本データベースに移管する。これには遺伝研担当分である HLA 領域データ、現在格納されている希少疾患変異データセットを含む。初年度は HLA 領域データを中心にその 2/3 の

移設を行う。これを検討材料に DBKERO との接続部のパイプライン、ブラウザ等についての関連プログラム群の整備を行う。

という2点であった。これらについては当初の目標を完遂することができた。

上記の対象となる全てのデータセットをまとめる形で、データベースを更新、一般公開を行った。データセットの詳細は申請書に記載したものである。またこれらのデータセットの対照となる公共参照データベースについても最新のものに更新した。日本人臨床検体データコンテンツに加えて、主として培養細胞株から取得された、シングルセル解析、ロングリード解析、薬剤摂動データからなる新規技術応用データセットについても公開を行った。新規公開データセットの概要をまとめた論文を作成、発表した(Suzuki et al NAR DB issue 2018)。また関連学会において、データベースを周知、データ活用例の紹介を行った(下記)。

2018年度

・IHEC-CREST データについて、データベースの動作確認および種々の最終確認を行った。また、IHEC-CREST チームより最終年度に補完的に産出されたデータセットについて、アップデートを行うことで CREST-IHEC の最終成果とした。IHEC の他国産出関連データについては IHEC データポータルについてリンクを整備した。DBKERO の他の部分、特にデータの整備に目途が立った HGVDB における日本人 SNP のデータとの統合についてもその実務を開始した。収載を予定、完了した IHEC データセットは表 1 の通りである。ゲノムブラウザでの表示例を図 1 に示す。

表 1 IHEC の追加予定データセット

器官	サンプル数 (検体)	データセット数
胃	14	RNA-seq: 11, BS-seq: 14, ChIP-seq: 47
子宮内膜	1	ChIP-seq: 1
血管内皮	14	ChIP-seq: 53

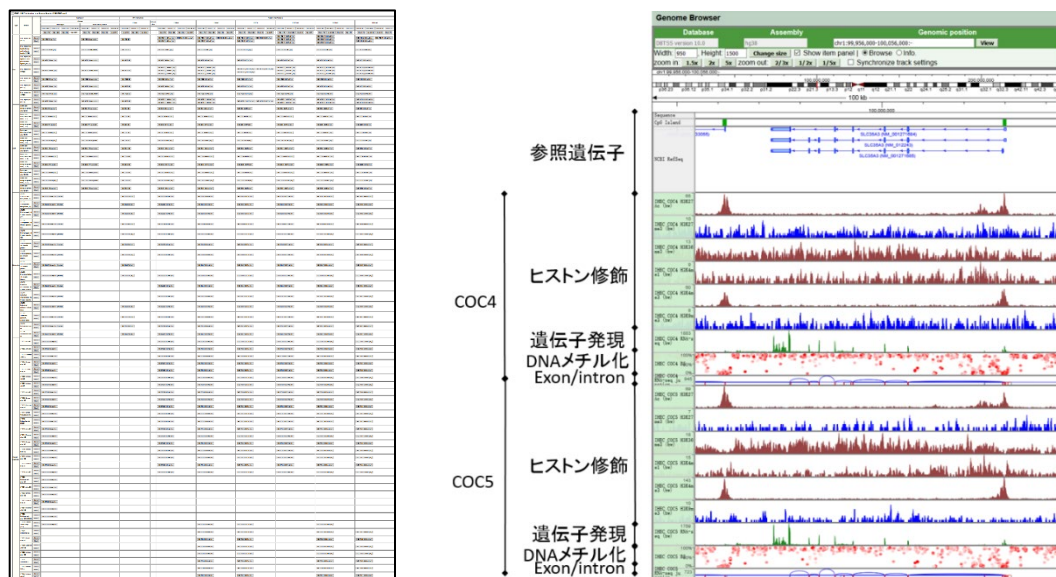


図 1 IHEC の最終確定ブラウザ(左図はデータの一覧図、右図は表示例)

- 九州大学で運用する ChIP-Atlas とのデータリンクを実装した。DBKERO 側で、ChIP-Atlas 側のデータを表示する形での試作版を作成、公開している。
- 第二期が開始された新学術領域「先進ゲノム支援」からもデータ産出が開始されている。これらについてマウス等モデル生物において、ヒトの病態関連解析、生物学的特性解析を指向したものについて、ヒトデータと関連付けて公開する。格納データについては、データ産生者が本支援活動の公開ポリシーに従ってデータを一般公開する際に、それと連動させる形で公開する。現在までにデータの1次公開が完了、DBKERO から追加公開したデータセットは表 2 の通りである。

表 2 ゲノム支援データの収載

生物種・データ内容	DRA Accession #	解析の種類
CLOCK related data of mouse liver	DRA001050	RNA-seq, ChIP-seq
T helper cell data of mouse	DRA001278	RNA-seq
	GSE51079	ChIP-seq
	GSE28292	ChIP-seq
Reg1[-/-],Roquin[San/San] data of mouse embryonic fibroblasts	DRA003234	RNA-seq
granulocyte macrophage progenitors (GMPs) data of mouse	DRA000485	ChIP-seq
	DRA000486	ChIP-seq
	DRA000487	ChIP-seq
	DRA000488	ChIP-seq
	DRA000858	ChIP-seq
Germ Cell Methylation data of mouse	DRA000484	RNA-seq, MethylC-seq, WBA-seq
	DRA000607	BS-seq (PBAT)
CD4Tcell/Bcell/EScell data of mouse	GSE73825	RNA-seq
HBZ over-expressed mouse	DRA003229	RNA-seq, ChIP-seq
	DRA003744	ChIP-seq

- 先進ゲノム支援関連データを模擬データに、これまでに開発の遅れていたメタデータについての RDF 化の枠組みを構築している。
- 大規模薬剤刺激オーミクスデータについては、新規のブラウザを開発し、その公開を行った。第1期の公開は、申請者らが取得した肺腺がん細胞株パネルに対する96化合物刺激条件で取得された RNA Seq/ATAC Seq データセット合計 1 万セットよりなる。詳細を表 3 に、ゲノムブラウザでの表示例を図 2 に示した。

表 3 大規模薬剤摂動オーミクスデータセット

データセット	
95 化合物	肺腺癌 23 セルライン (各 RNA-seq, ATAC-seq)
23 化合物	肺腺癌 5 セルライン×4 濃度×3 タイムコース (各 RNA-seq, ATAC-seq)

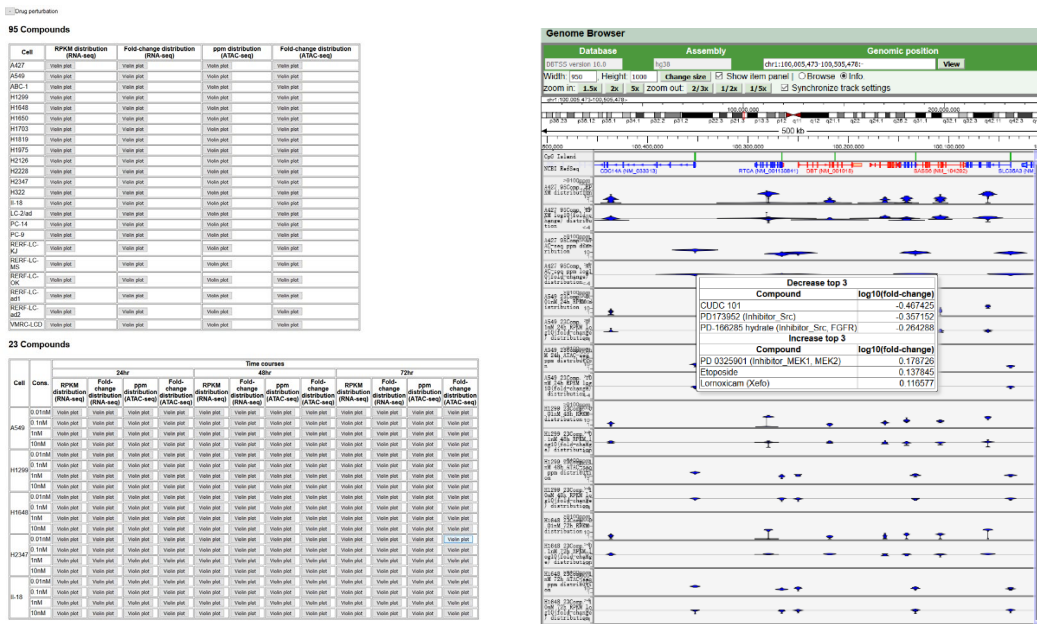


図 2 薬剤摂動データセットの公開(左図は搭載データの一覧図、右図は表示例)

- HGVDDB については、残りのデータを移設し、全データについて最終の形式、品質確認を行った。データベースを公開して最終動作確認を行った(図 3)。

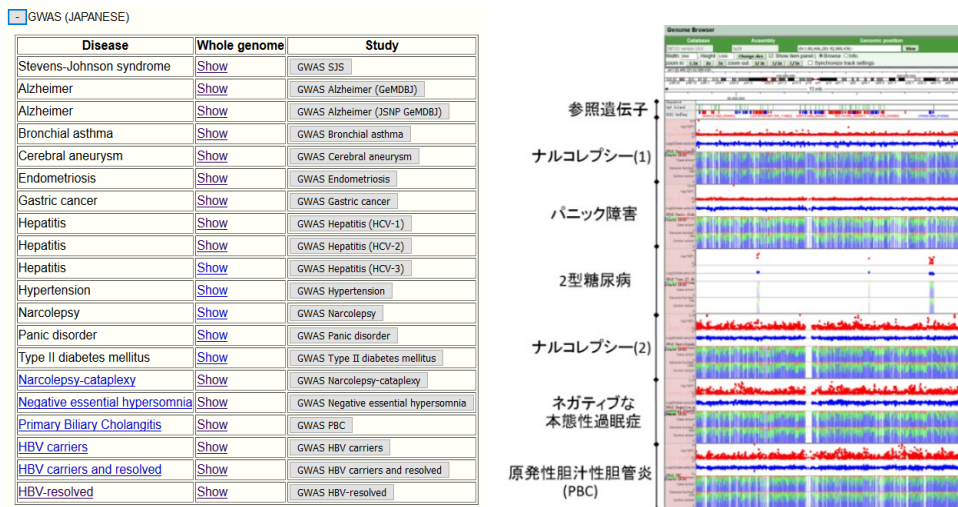


図 3 HGVDDB の最終確認ブラウザ(左図は搭載データの一覧、右図は表示例)

- 東北メディカル・メガバンクと、データ公開条件等についての付帯条件を整理し、本データベースへの取載を進めている。現在までに、同データベースとの相互参照はゲノム座標を起点にしたリンク形式で行っている(図 4)。今後のより抜本的な形での連携の可能性について協議を継続している。

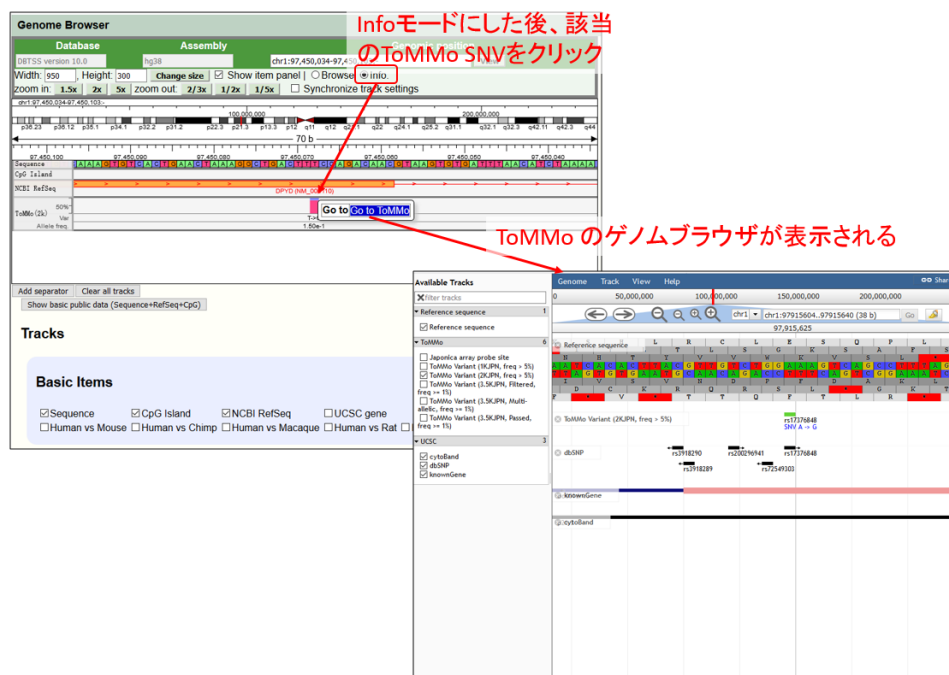


図 4 東北メディカル・メガバンクとの相互データ参照

※東北メディカル・メガバンクは公開の許諾が得られないため内部閲覧のみ

付加的課題

また、製薬会社等への民間利用を促進すべく、社内秘匿情報に考慮した独自データサーバーを構築した。現在、上記メインデータベースのデータのミラーリングは完了しており、内部機微データの格納を行っている。次年度以降、これらをモデルケースに集積情報の民間活用への筋道を模索している。

全体を通じて、最大の懸念は、公開の結果として利用者数の拡大が十分に期待した成果を満たすものでなかった点である。そこで後記(§ 4. 主要なデータベースの利活用状況 3. その他)のような第三者によるユーザビリティ調査の追加対応を行った。この点に関しては、引き続き検討を行っているものである。

(2) 「対象とするデータベース条件」のうち採択時に未達成であった項目の対応状況
該当なし。

(3) 統合化推進プログラムの他のチームや DBCLS との連携

- ・九州大学で運用する ChIP-Atlas とのデータリンクの実装を行った。DBKERO 側で、ChIP-Atlas 側のデータを表示する形で公開している(図 5)。

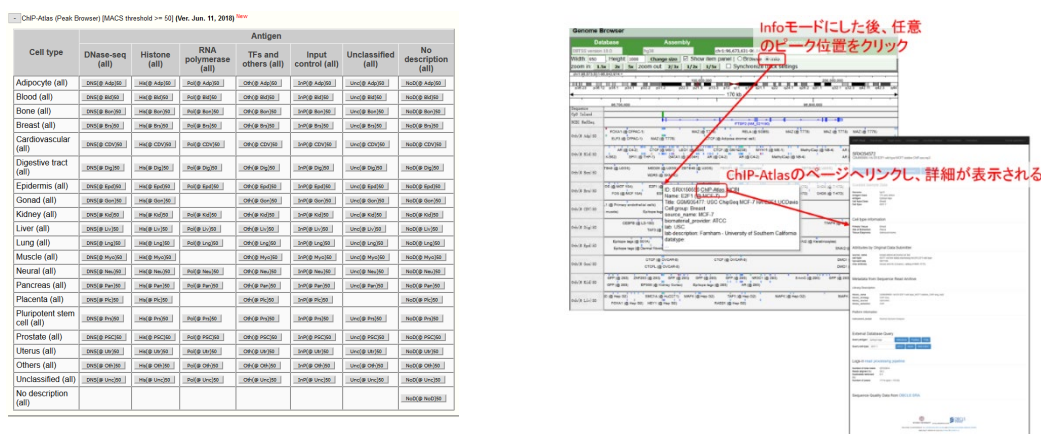


図 5 ChIP-Atlas との連携(左図は搭載データの一覧、右図は表示例)

•Transcription Factor Binding site search の ChIP-Atlas、GGGenome との連携。

当機能は、ユーザーの興味のあるゲノム領域や、遺伝子周辺領域について、既知の転写因子の ChIP-Seq の解析結果や配列パターンから、転写因子の結合領域を予測する機能である。この機能を実現するにあたり、転写因子の結合領域の予測データとして九州大運用の ChIP-Atlas のデータを、また配列パターンの高速検索を行うために DBCLS 内藤雄樹先生らの運用する GGGenome の REST API を利用させていただいている。

•BioBank Japan あるいは NBDC の主導のもと、いくつかの日本人ゲノム多型の大規模データの収集が開始されている。これらについては次年度以降のデータ統合の可能性をにらんで担当者との協議を開始している。特に TogoVar(<https://togovar.biosciencedbc.jp/>)では、今年度に入って、個々の position の variation データについて、アノテーション情報が急速に充実している。今年度、TogoVar に掲載されている variation 情報で、第三者が作成したデータセット(ClinVar, ExAC, HGVD, ToMMo)については、DBKERO にも既にすべて搭載した。

またこれに加えてその他の外部がんゲノムデータ、COSMIC、ICGC、TCGA 等のがん関連変異データも充実させている(図 6)。

- SNP/SNV public data	
TCGA (SNV)	TCGA Urothelial bladder carcinoma (blca_2013) TCGA Breast tumors (brca_2012) TCGA Colon and rectal cancer (coadread_2012) TCGA Glioblastoma (gbm_2008)
	TCGA Glioblastoma (gbm_2013) TCGA Clear-cell renal cell carcinoma (kirc_2013) TCGA Acute myeloid leukemia (aml_2012) TCGA Squamous cell lung cancers (lusc_2012)
	TCGA Ovarian carcinoma (ov_2011) TCGA stad (stad_2014) TCGA Endometrial carcinoma (ucec_2013) TCGA Lung adenocarcinoma (luad_2014)
	TCGA Chromophobe renal cell carcinoma (kich_2014) TCGA Papillary thyroid carcinoma (thca_2014) TCGA Head and neck squamous cell carcinomas (hnac_2014)
	TCGA Diffuse lower grade gliomas (lgg_2015) TCGA Prostate adenocarcinoma (prad_2015) TCGA Adrenocortical carcinoma (aco_2016)
ICGC (SNV)	ICGC Acute Myeloid Leukemia - TARGET, US ICGC Bladder Urothelial carcinoma - CN ICGC Bladder Urothelial Cancer - TCGA, US ICGC Bone Cancer - UK
	ICGC Breast Triple Negative/Lobular Cancer - UK ICGC Breast Cancer - TCGA, US ICGC Chronic Lymphocytic Leukemia - ES ICGC Chronic Myeloid Disorders - UK
	ICGC Colon Adenocarcinoma - TCGA, US ICGC Early Onset Prostate Cancer - DE ICGC Esophageal Adenocarcinoma - UK ICGC Esophageal Cancer - CN
	ICGC Gastric Cancer - CN ICGC Brain Glioblastoma Multiforme - TCGA, US ICGC Kidney Renal Clear Cell Carcinoma - TCGA, US
	ICGC Kidney Renal Papillary Cell Carcinoma - TCGA, US ICGC Acute Myeloid Leukemia - KR ICGC Brain Lower Grade Glioma - TCGA, US ICGC Liver Cancer - FR
	ICGC Liver Cancer - NCC, JP ICGC Liver Cancer - RIKEN, JP ICGC Lung Cancer - Squamous cell carcinoma - KR ICGC Lung Squamous Cell Carcinoma - TCGA, US
	ICGC Malignant Lymphoma - DE ICGC Neuroblastoma - TARGET, US ICGC Oral Cancer - IN ICGC Ovarian Cancer - AU
	ICGC Ovarian Serous Cystadenocarcinoma - TCGA, US ICGC Pancreatic Cancer - AU ICGC Pancreatic Cancer - CA ICGC Pancreatic Cancer Endocrine neoplasms - AU
	ICGC Pediatric Brain Cancer - DE ICGC Prostate Adenocarcinoma - CA ICGC Prostate Adenocarcinoma - UK ICGC Prostate Adenocarcinoma - TCGA, US
	ICGC Rectum Adenocarcinoma - TCGA, US ICGC Renal clear cell carcinoma - CN ICGC Renal Cell Cancer - EU/FR ICGC Skin Cutaneous melanoma - TCGA, US
	ICGC Gastric Adenocarcinoma - TCGA, US ICGC Thyroid Cancer - SA ICGC Head and Neck Thyroid Carcinoma - TCGA, US
	ICGC Uterine Corpus Endometrial Carcinoma - TCGA, US ICGC Acute Myeloid Leukemia - TARGET, US ICGC Bone Cancer - Ewing Sarcoma - FR
	ICGC Breast ER+ and HER2- Cancer - EU/UK ICGC Breast Cancer - FR ICGC Breast Cancer - KR ICGC Biliary tract cancer - JP
	ICGC Cervical Squamous Cell Carcinoma - TCGA, US ICGC Colorectal Cancer - CN ICGC Acute myeloid leukaemia and Chronic myelogenous leukaemia - CN
	ICGC Benign Liver Tumour - FR ICGC Liver Cancer - CN ICGC Liver Hepatocellular carcinoma - TCGA, US ICGC Liver Hepatocellular Macronodules - FR
ICGC Lung Squamous cell carcinoma - CN ICGC Skin Cancer - AU ICGC Pancreatic Endocrine Neoplasms - IT ICGC Skin Adenocarcinoma - BR	
Others	1000genome SNP Clear-cell renal cell carcinoma dbSNP build 137 NHLBI Exome Sequencing Project NCBI ClinVar COSMIC Meyerson's lung adenocarcinoma
	Myelodysplasia ICGC ALL Nagahama Japanese SNP Japan PGx (Japanese SNP) ExAC SCLC (Small Cell Lung Cancer) 26 lung adenocarcinoma cell lines Lung adenocarcinoma 97 clinical samples

図 6 DBKERO に搭載される variation 関連データ

国内で大規模なデータベースのうち、東北メディカル・メガバンクのデータについては、内部閲覧のみしか実現されておらず、JGA のデータについては、現状では、当データベースに搭載・公開することができていない。既に TogoVar でダウンロードできるようになっているデータ(<https://togovar.biosciencedbc.jp/downloads>) については、次年度以降のデータ統合の可能性をにらみ DBKERO への搭載と公開に向けて、担当者との協議を開始している。協議をへて、今後、DBKERO から TogoVar のアノテーションデータへのリンクやゲノムブラウザの連携について充実させていく予定である(図 7)。

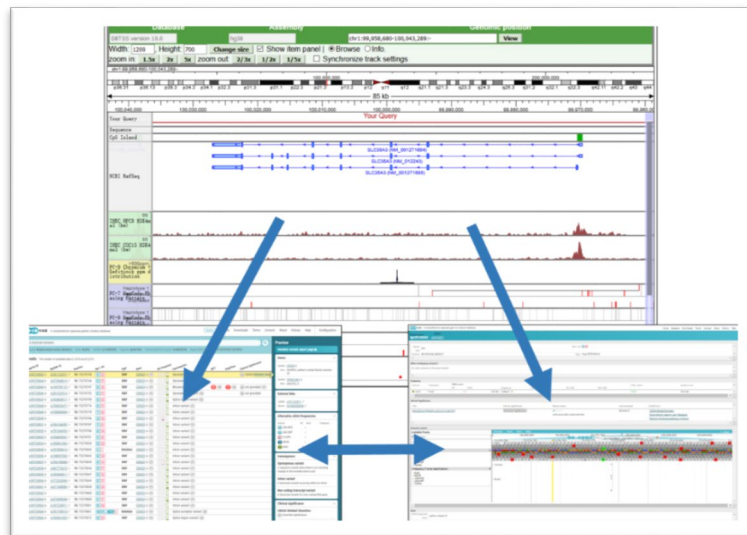


図 7 DBKERO と TogoVar 連携のイメージ

§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 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)

種別	2016年度	2017年度	2018年度	2019年度 (9月末時点)
ユニーク IP 数	42	171	243	147
ページ数	1,907	2,135	6,733	2,767

(2) 分析

2016年度はDBTSSから分割する形でDBKEROを開設した年のため、アクセス数が少ない。2018年の1月にNAR Database Issueに論文が掲載されたため、2018年度はユーザー数が伸びたと考えられる。特に2018年の6月～8月11月～12月に10,000ページビューを超えたが、その後は3,000ページビュー程度に落ち着いている。今後はユーザビリティを改善し、継続的にユーザー数が多く推移するような形に持っていきたい。

2. データベースを利用して得られた研究成果・産業応用の例

1. 上記と同様なデータ産生、情報解析を遂行する次世代の生命データサイエンス人材を育成する教育プログラム DSTEP を開始し(参加企業12社:協和キリン株式会社、富士フイルム株式会社、小野薬品工業株式会社、中外製薬株式会社、株式会社ダイナコム、ピッツ株式会社、アメリエフ株式会社、Genomedia 株式会社、合同会社みらか中央研究所、アジレント・テクノロジー株式会社、フリーダタイム株式会社、株式会社羊土社)、本データベースを教育用教材としても活用している。この枠組みにおいて本データベースは民間企業が大規模データ利活用を行う情報インフラの構築に向けて重要なモデルになっていると考えている。
2. MuNAGe(Mutated Nucleotide and Amino-acid sequence Generator)は、国立がんセンター研究所 服部恵美 先生らが開発する NGS 結果から質量分析同定用のタンパク質配列を生成するソフトウェア(論文投稿中)である。このツールの検証に DBKERO に収載されている3種の肺腺癌細胞株(ABC-1, PC-3, RERF-LC-KJ)の WGS 及び RNA-Seq のデータが活用された。Swiss-Prot で検出できないタンパク質が MuNAGe で検出可能なことが確認された。
3. がん特異的な脆弱性の分子基盤解明の研究として、国立がんセンター研究所の塩谷文章 先生は、DBKERO に収載されている肺腺癌培養細胞の WGS および RNA-Seq データを用いた。DNA 複製ストレス応答の中心的な制御因子である ATR キナーゼの阻害剤感受性因子の検索と、発現情報の取得のためにこのデータが活用されている。

3. その他

- ・DBKERO のユーザビリティ調査の実施

データベースの現在の問題点の抽出を行い、サイトの改善点の発見と今後の活用を活性化
する目的で、第三者目線での DBKERO の使い勝手の調査をポップインサイトに依頼し、目下
実施中である。調査方法は、関連メールマガジン等のアンケートから被験者を募り、被験者が
実際に使っているところを、マウス操作の追跡やマイクからの音声でモニタリングし、サイトの問題
点や改善点を抽出するものである。この結果を元にサイトの改善を行い、データベース利活
用の向上に繋げることをねらいとしている。

モニタリング調査は引き続き行っているが、現在までに以下のような意見がでている。

1. データの出所に関する情報が載っていない、がん細胞の情報が ID レベルの情報しか
なく内容が把握しにくい。
2. 他のデータベースにはないデータが収録されている点が良い。

これについて、以下のように改善を行っている。

1. については、現在様々なコンテンツのデータが雑多に存在し、データを分かりやすく
ユーザーに提供できている状態ではないと言えるかもしれない。こちらは、メタデータの
整備をすすめ、ユーザインターフェースの統一性の観点からも、ENCODE のデータマトリ
ックス(<https://www.encodeproject.org/matrix/?type=Experiment>)等の既存の有用サ
イトを手本に改善を図りたいと思う。
2. については、今後も外部のデータをさらに取り込んでいくことに加え、他のデータベー
スにはない、シングルセルやロングリード等のオリジナルデータをさらに充実していく必要
があると考えられる。

モニタリング調査については、現在も進行中であり、この結果を元に今後のデータベースの改
良と、研究者ユーザーのニーズをつかみ、利用者の拡大へ繋げたい。

§5. 今後の展開

近年、ヒトゲノム関連データの利活用に関する議論を耳にする機会が増えた。ただし、これらはがんを始めとする疾患ゲノムについて、その DNA 配列を1次元的に解析するものであることが多いように思える。実際、DNA 配列の異同と患者臨床情報の相関を、人工知能等の情報学的手法を駆使して、対応付けしようとする試みは数多い。しかし、いかにしてゲノム配列の変異がエピゲノム制御を介して、トランスクリプトームとして読みだされているのか、その生物学的根拠なしに DNA 配列と臨床像を直接結び付けようとするスキームは基礎生物学者の観点からはやや短兵急に過ぎるように思える。実際、依然として多くのがんについて変異情報に基づいた薬剤療法での根治は困難であり、他の疾患、特に多因子疾患は、依然として病態機序が不明である。GWAS で検出されたゲノム領域が遺伝子発現制御に関わるとされる遺伝子間領域に位置するケースが 90%を超える、とする推定もある。これらに関しては、より高次元のオーミクス空間において、より生物学的証拠に依拠した仮設構築を進める必要がある。そのためにも本データベースで試みたような多層オーミクスデータの補完的利用は必須に思える。今期の開発でもデータ収載を実践したように、シングルセル解析、大規模統合解析、長鎖リード解析等の新規の技術の進展も著しい。これらを臨床検体の解析へと直接に利用した例も急速に増えている。本データベース該当部分のさらなる拡充には製薬企業からの関心も高い部分である。さらに高い精度と密度でのデータの蓄積はより密接に分子機序に依拠した定量的なモデリングを可能とすると考えている。本研究開発終了後、“ゲノム配列以降”について、ますます急速に進展する“新規技術”より産出される大規模データを“モデル化”を志向する産官学の研究者に提供するデータベースのさらなる発展を企画したいと考えている。

§6. 自己評価

今期の開発項目については、当初の計画通りに概ね順調に推移している、と考えている。ただし、これは、投下予算に比して十分な成果が得られているという自己評価には直結しない。実際、研究期間開始以来の3年間に関連分野を取り巻く状況は一変している。日本人ゲノムについても、今年度に入って国規模での大規模解析計画が始められようとしている。前述の通り、これを真に生物学的な知識に昇華させるには多層オーミクス解析データのさらなる充実が必須になる。研究開始時には必ずしも予見できなかった付帯状況の変化にも関わらず、大規模データ利活用について、最新のデータを産官学の現場へ届け、その開発基盤として資すべく、より弾力的な開発計画とその迅速な実現が後半期において、最重要な課題になると考えている。

§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)	0 件
	国際(欧文)	1 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. Suzuki A, Kawano S, Mitsuyama T, Suyama M, Kanai Y, Shirahige K, Sasaki H, Tokunaga K, Tsuchihara K, Sugano S, Nakai K, Suzuki Y, "DBTSS/DBKERO for integrated analysis of transcriptional regulation.", *Nucleic Acids Res.*, 4;46(D1):D 229-D238, 2018 (DOI:10.1093/nar/gkx1001).

2. その他の著作物(総説、書籍など)

1. 鹿島幸恵、鈴木絢子、関真秀、鈴木穰、「シングルセル解析とがんの不均一性」、*実験医学* Vol.36 No.2 (増刊)がん不均一性を理解し、治療抵抗性に挑む、2018
2. 関真秀、鈴木穰、「シングルセル解析とナノポアシーケンス」、*医学のあゆみ*、266 巻 5 号、2018
3. 鹿島幸恵、鈴木絢子、関真秀、鈴木穰、「ゲノム医療における一細胞解析」、*実験医学増刊* 36 巻 15 号、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	3 件
	国際	8 件
口頭発表	国内	4 件
	国際	2 件
ポスター発表	国内	0 件
	国際	2 件

(2) 招待講演

〈国内〉

1. 鈴木穰、「シングルセル解析」、日本癌学会、パシフィコ横浜(神奈川)、2017 年 9 月 28 日-30 日
2. 鈴木穰、「ロングリード解析とシングルセル解析」、日本内分泌学会、フェニックス・シーガ

イア・リゾート(宮崎)、2018年4月26-28日

3. 鈴木穰、「がん細胞株のシングルセル解析:複数プラットフォーム間の横断解析」、日本分子腫瘍マーカー研究会、大阪国際会議場(大阪)、2018年9月26日

〈国際〉

1. Yutaka Suzuki, "Multi-omics analysis of cancer cells", 第12回国際ゲノム会議、学術総合センター 一橋講堂(東京)、2017年6月27日-29日
2. Yutaka Suzuki, "Single cell analytical platform", Human Cell Atlas - Asia Meeting、沖縄科学技術大学院大学(沖縄)、2017年11月30日・12月1日
3. Yutaka Suzuki, "Multi-omics analysis of cancer cells", Human Genome Meeting 2018、パシフィコ横浜(神奈川)、2018年3月12日-15日
4. Yutaka Suzuki, "Combinatory use of distinct single-cell RNA-seq analytical platforms reveals the heterogeneous transcriptome response", IUBMB SEOUL, COEX, Seoul, Korea, June 4-8, 2018
5. Yutaka Suzuki, "Single Cell Analysis—from Technical Evaluation to Clinical Application: Pipeline Construction and Data Analysis", 10X Genomics UGM, Matrix Building, Singapore, October 26, 2018
6. Yutaka Suzuki, "Single Cell and Long Read Sequencing of Cancer Cells", TBC/BI OINFORM, Hotel Samjung, Seoul, 韓国, October 31 - November 2, 2018
7. Yutaka Suzuki, "Single Cell Analysis Reveals the Heterogeneity of Cancer Cells: Lessons from the Analyses of Adult T Cell Leukemia and Lung Cancer Cells", The 28th KOGO Annual Conference 2019, Sejong Convention Center, Seoul, Korea, September 4-6, 2019
8. Yutaka Suzuki, "Identification and characterization of a new class of cancerous structural variations using PromethION", Nanopore Tech Tour, Beijing, China, September 23, 2019

(3) 口頭講演

〈国内〉

1. 菅野純夫、「疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オーミクスデータの統合」、トーゴーの日シンポジウム 2017、東京大学弥生講堂一条ホールアネックス(東京)、2017年10月4日・5日
2. 鈴木穰、「生命科学のデータベース活用法(フォーラム)」、日本分子生物学会、神戸ポートアイランド(兵庫)、2017年12月6日-9日
3. 菅野純夫、「疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オーミクスデータの統合」、トーゴーの日シンポジウム、日本科学未来館(東京)、2018年10月5日
4. Ayako Suzuki, Sarun Sereewattanawoot, Yutaka Suzuki, Katsuya Tsuchihara, "Multi-omics characterization of lung cancer cells based on gene co-expression modules", 日本癌学会、大阪国際会議場(大阪)、2018年9月27-29日

〈国際〉

1. Yutaka Suzuki, "Multi-omics analysis of cancer cells", The 1st International Sym

- posium for Trans-Omics、東京大学小柴ホール(東京)、2017年11月21日・22日
2. Yutaka Suzuki “Phasing regulatory SNVs to their regulating transcript regions in cancers by using MinION”, London Calling 2018, May 24-25, 2018, Old Billingsgate, London, UK, May 24, 2018

(4) ポスター発表

〈国内〉

該当なし

〈国際〉

1. Ayako Suzuki, Sarun Sereewattanawoot, Yutaka Suzuki, Katsuya Tsuchihara, “Characterization of aberrant transcriptional regulations in lung cancers”, Advances in Genome Biology and Technology 2018, Hilton Orlando Bonnet Creek, Orlando, Florida, USA, 2018年2月12日-15日
2. Ayako Suzuki, Masahide Seki and Yutaka Suzuki, “Identification and Characterization of a new class of cancerous structural variations using PromethION”, Advances in Genome Biology and Technology (AGBT) 2019 General Meeting, Marco Islands, FL, USA, February 27 - March 2, 2019

4. 知財出願

該当なし

5. 受賞・報道等

該当なし

§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
2017年 4月24日	キックオフミーティング (非公開)	東京大学医学部	15人	本研究開発費で行う研究の 役割分担の調整
2017年 4月28日	チーム内ミーティング(非 公開)	東京大学フュー チャーセン ター	5人	同上
2017年 5月15日	チーム内ミーティング(非 公開)	東京大学医学部	3人	進捗状況確認のためのミーテ ィング
2017年 6月19日	チーム内ミーティング(非 公開)	東京大学医学部	3人	同上
2017年 7月14日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	3人	同上
2017年 7月18日	沖真弥チーム、NBDCと のミーティング(非公開)	JST 東京本部	10人	追加実施策に関する打ち合 わせ
2017年 7月19日	チーム内ミーティング(非 公開)	東京大学医学部	3人	進捗状況確認のためのミーテ ィング
2018年 3月28日	チーム内ミーティング(非 公開)	東京大学医学部	3人	同上
2018年 8月30日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	5人	研究進捗報告のためのミーテ ィング
2018年 11月22日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	5人	研究進捗報告のためのミーテ ィング
2019年4 月26	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	進捗状況確認のためのミーテ ィング
2019年5 月29日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	同上
2019年6 月20日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	同上
2019年7 月26日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	同上
2019年8 月23日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	同上
2019年9 月25日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	同上
2019年10 月2日	チーム内ミーティング(非 公開)	東京大学柏 キャンパス	4人	同上

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
2017年 12月6日- 9日	生命科学系学会合同年次大会(ConBio2017)特別企画「使ってみようバイオデータベースーつながるデータ、広がる世界(BioDB)」	神戸ポートアイランド	30人	目的:データベースの一般への利用と普及の促進 概要:データベースの紹介
2018年 11月28日- 30日	日本分子生物学会特別企画「使ってみようバイオデータベースーつながるデータ、広がる世界(BioDB)」	パシフィコ横浜	30人	目的:データベースの一般への利用と普及の促進 概要:データベースの紹介
2018年 12月5日	統合データベース講習会:AJACS名古屋2	愛知県がんセンター 国際医学交流センター	100人	目的:生命科学系のデータベースやツールの使い方を紹介する(初心者向け) 概要:DBKERO およびその他がんゲノム解析関連データベースの紹介

以上