

研究開発実施報告

□概要

研究開発課題名	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ムクスデータの統合
開発対象データベースの名称 (URL)	DBKERO (https://kero.hgc.jp/)
研究代表者氏名	菅野 純夫
所属・役職	東京医科歯科大学 難治疾患研究所 非常勤講師 (2021年3月時点)

□目次

§1. 研究実施体制	2		
§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	17
(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)	17
(2) ツール等一覧	3	① 出願件数	17
§3. 実施内容	4	② 一覧	17
(1) 本年度の研究開発計画と達成目標	4	(5) 受賞・報道等	17
(2) 進捗状況	8	① 受賞	17
§4. 成果発表等	16	② メディア報道	17
(1) 原著論文発表	16	③ その他の成果発表	17
① 論文数概要	16	§5. 研究開発期間中に主催した活動 (ワークショップ等)	18
② 論文詳細情報	16	1. 進捗ミーティング	18
(2) その他の著作物 (総説、書籍など)	16	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
(3) 国際学会および国内学会発表	16		
① 概要	16		
② 招待講演	16		
③ 口頭講演	16		



§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
菅野グループ(医科歯科)	菅野 純夫	東京医科歯科大学・非常勤講師	DBKERO開発の総括と計算機基盤の整備
菅野グループ(東大)	菅野 純夫	東京大学・客員共同研究員	DBKEROとツールの開発
土原グループ	土原 一哉	国立がん研究センター・分野長	がんゲノム多層オーミクスDBの開発と運用。
徳永グループ	徳永 勝士	東京大学・教授	ヒトゲノム多型・変異データベースの拡充。
井ノ上グループ	井ノ上 逸朗	国立遺伝学研究所	ヒトゲノム多型・変異データベースの拡充。

§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	DBKERO : 疾患ゲノム多層オーミクス機能注釈データベース	DBKERO	http://kero.hgc.jp/

【その他のデータベース】

No.	名称	別称・略称	URL
1	BEACON SCAN	なし	http://dbtss.hgc.jp/cgi-bin/ucscBeacon/query
2	DBKERO RDF viewer	なし	http://kero.hgc.jp/rdf/
3	DBKERO RDF sparql endpoint	なし	http://kero.hgc.jp/rdfs/services/kero/sparql.htm
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

予算額の縮小に対応して、当初計画について項目ごとに優先順位をつけて実施する。上位の項目を達成し、その成果について評価委員の確認を得た後に順次、次項目へと着手するものとする。特に項目 1-3 については第四年次上半期に完了するものとする。

1 ダウンロードポータルサイトの作成

DBKERO に掲載されるデータの利活用を促進するために、データ取得の利便を図る。データポータルを作成して、簡便にかつ一括して Raw データを所得できる枠組みを構築する。想定するサイトの設計を以下に示す(図1)。

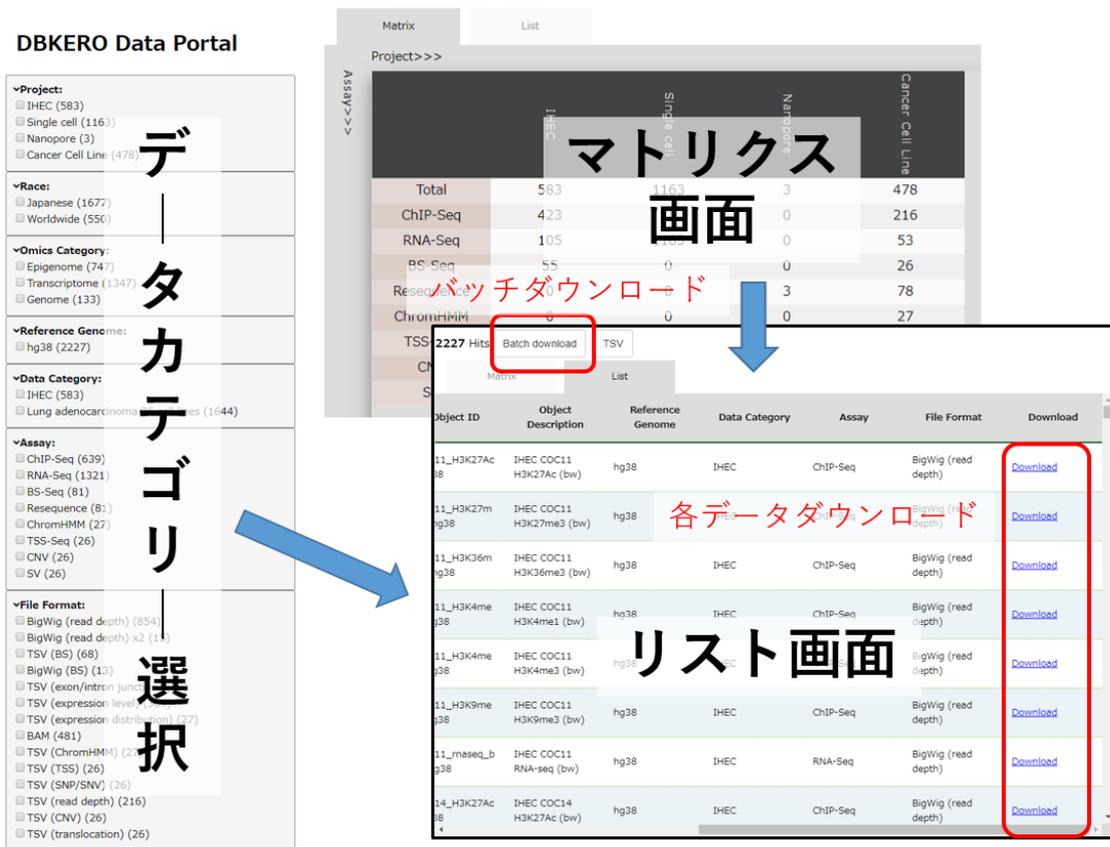


図1 データポータルのイメージ

目的のデータをデータマトリクスや、データカテゴリーのチェックボックスから絞り込み、リスト画面や Batch download ボタンから搭載データをダウンロード可能とする。

表1ダウンロードポータル搭載済みデータセット数

Project	Assay type								Total
	ChIP-Seq	RNA-Seq	BS-Seq	Resequencing	ChromHMM	TSS-Seq	CNV	SV	
IHEC	423	105	55	0	0	0	0	0	583
Single cell	0	1,163	0	0	0	0	0	0	1,163
Nanopore	0	0	0	3	0	0	0	0	3
Cancer Cell Line	216	53	26	78	27	26	26	26	478
Total	639	1,321	81	81	27	26	26	26	2,227

2 RDF 化の完了とNBDC RDF ポータルへの完全移行

RDF 化について、全データセットについて完了する。昨年度までに完成した RDF スキーマを活用して、トランスクリプトーム、エピゲノムのデータについてデータを成形し、NBDC が運用する RDF ポータルを通じて公開する。実験条件等についての詳細はメタデータについては、直近のスキーマの確立に時間を要すると思われるため、こちら自由記載の項に記述を転記することとするが、可及的速やかな RDF 化に向けて協議を進める。速やかに公開可能なデータセットは 200 件程度を想定している。現在、作成しているトランスクリプトーム、エピゲノムのスキーマ(図2)、および対象となるデータセットの概略を以下に示す(表1)。

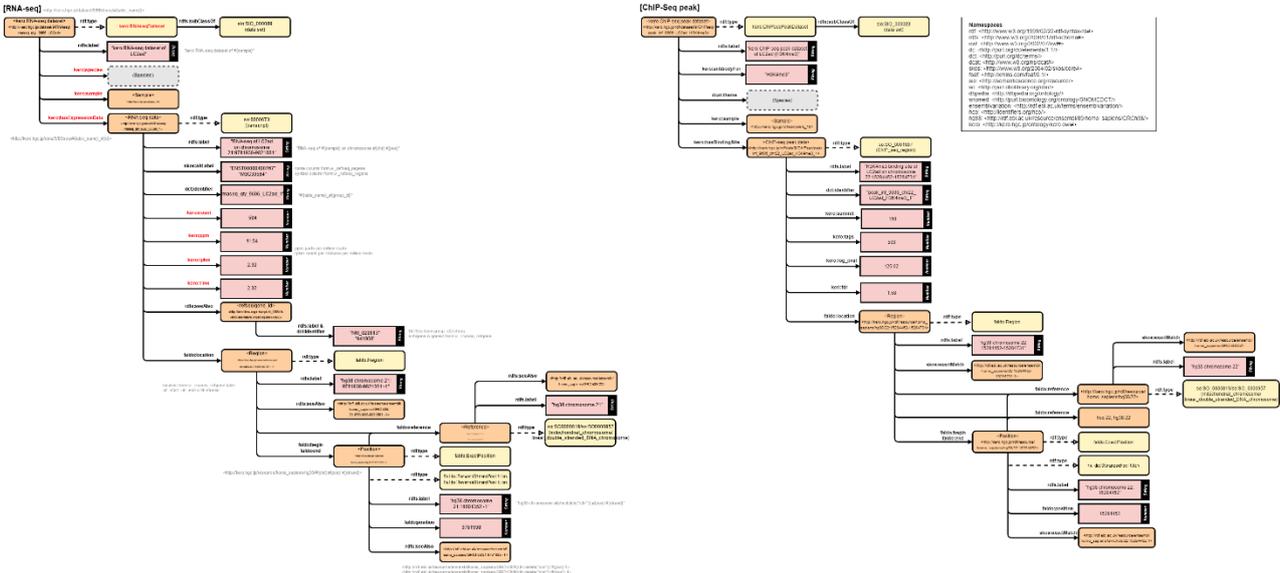


図2 RNA-seq 発現量(左)及び、ChIP-seq ピーク情報(右)のスキーマ図

表2 第4年次 RDF 追加登録予定データセット数

チーム	器官	登録名	検体数	データセットサンプル数		
				RNA-Seq	ChIP-Seq	合計
金井班	肝臓	HPC	8	8	48	56
金井班	腸管	COC	11	11	66	77
金井班	胃	GAC	8	8	36	44
白髭班	血管内皮	JTK	33	26	-	26
佐々木班	子宮内膜	JKU	15	15	-	15
合計			75	68	150	218

3 講習会の開催

以下のような講習会を開催して KERO コンテンツおよびデータベース利用法の周知に努める。

- がんセンターでの講習会(6月ごろ;がんセンター東病院)
 - 文科省データサイエンス教育プログラム D-DRIVE の全国ネットワークを通じた講習会(9/19;東大・本郷)
 - 医科研講習会(10月ごろ:東大・医科研)
 - ゲノム支援での講習会(12月ごろ;場所未定)
- において、単独での講習会を企画する。

その他、開催時期は未定であるが、

- NBDC の主宰する JST における合同データベース講習会(東京・時期未定)
- NBDC の主宰する AJACS による定例講習会(時期・場所未定)

に参加する。

さらに、分子生物学会、癌学会、人類遺伝学会等の関連学会においての何らかの形で同様のイベントの開催を検討する。

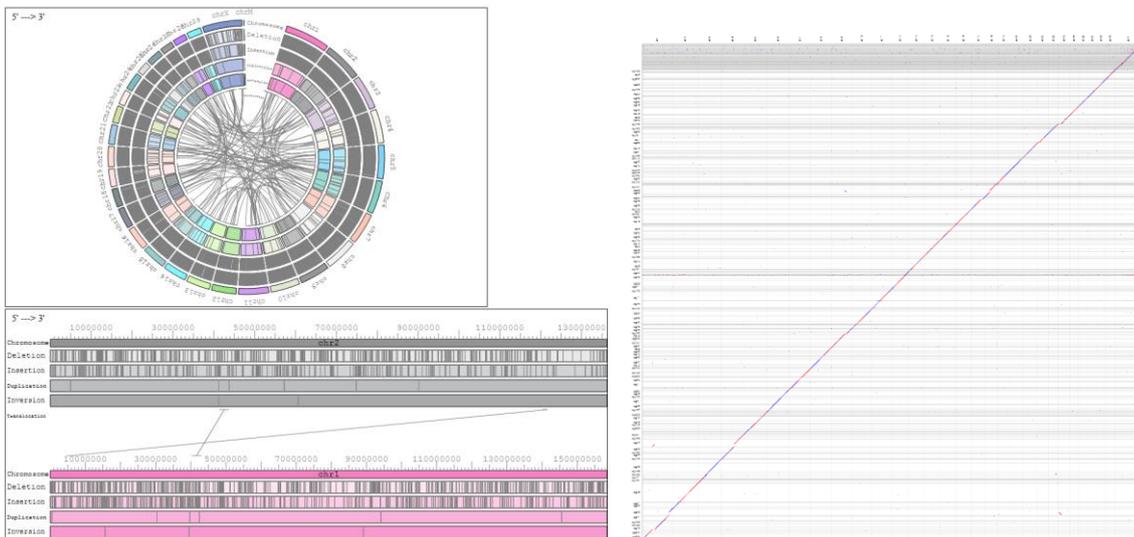
4 利用者の拡大

上記 1-3 の活動を通じて、利用者の全体的な拡大を図る。

項目 1-4 における目標達成後、下半期での開始になるが、第四年次計画に掲げた以下の項目についても随時実施する。

5 長鎖 DNA ブラウザの開発と実装

がんゲノム、正常ゲノムについて、申請者の研究室を始め、多くの研究室でロングリードシーケンスデータの産出が本格化している。これらのデータの取載をいち早く実現することは本データベースの責務であると考えている。これらについて申請者ら自身が長鎖 DNA シーケンサー PromethION を用いて産出したデータを模擬データセットに、ブラウザの開発・更新を行う。開発が完了次第、即時公開が可能な申請者らが保有する日本人ゲノム 20 検体の長鎖 DNA データについて、公開を行うものとする(図3)。



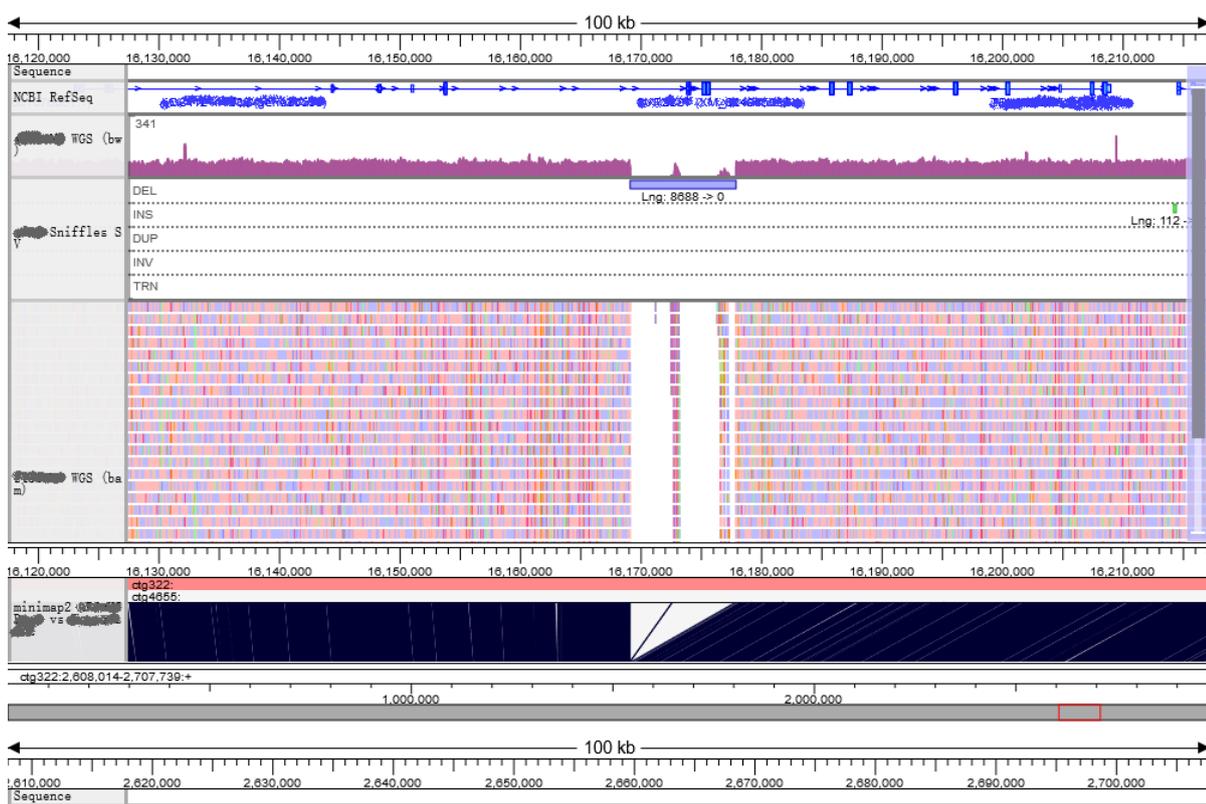


図3 長鎖ゲノムブラウザのイメージ

ゲノム構造ブラウザ(左上)、参照ゲノムとの俯瞰比較(ドットプロット図)(右上)、比較ブラウザ(下)

6 臨床試験付随研究データの収載

土原グループががんセンターにおいて関連する臨床試験付随研究において、いくつかのがんゲノムデータが集積している。これらは民間企業との共同研究として実施してきたものであるが、当初の計画と知財化が終了し、一般に公開が可能になるものである。逆に、研究班が解散することから、公的なデータベースに登録しなければ、データ自体が散逸化する可能性がある。これらは JST A-Step により平成 23 年に開始され令和 2 年 4 月に最終的に終了する「網羅的ゲノム解析に基づく抗 EGFR 抗体薬治療効果予測診断キット」(通称 BREAC 試験)による大腸がん検体約 100 症例を含む。これらについては詳細な患者情報が付帯しているが、多くの患者が現在までに死亡しているために、その情報も併せて収載可能であるか、その技術的・法的な実現可能性についても検討を開始するものとする。

7 空間トランスクリプトーム解析データの収載

昨年度になって、空間トランスクリプトーム解析データの集積が開始され始めた。これは従来のシングルセル解析技術をさらに発展させ、組織の位置情報を保持した形で遺伝子発現情報を計測するものである(データ例を以下に図示する)。これらのデータは病理とオミクスデータの融合として近い将来に大きくデータ規模が拡大する可能性があるものであると考えている。今年度、そのデータ収載の枠組みについて検討を開始する。

(2)進捗状況

1 ダウンロードポータルサイトの作成

(ア)「ダウンロードポータルを2020年度第1四半期までに開設する」

DBKERO に掲載されるデータの利活用を促進するために、データ取得の利便を図った。データポータルを作成し、簡便にかつ一括して Raw データを取得できる枠組みを構築した(https://kero.hgc.jp/tool/data_portal.html) (図4)。

The screenshot shows the DBKERO Data Portal interface. On the left, there are filter sections for Project, Race, Omics Category, Reference Genome, Data Category, and Assay. The main area displays a table with 2573 hits, categorized by Assay. The table has columns for Assay, IHEC, Single cell, Nanopore, GWAS, Cancer Cell Line, and Drug Perturbation. The data is as follows:

Assay	IHEC	Single cell	Nanopore	GWAS	Cancer Cell Line	Drug Perturbation
Total	583	1163	3	14	478	332
ChIP-Seq	423	0	0	0	216	0
RNA-Seq	105	1163	0	0	53	226
BS-Seq	55	0	0	0	26	0
Resequencing	0	0	3	0	78	0
GWAS	0	0	0	14	0	0
ChromHMM	0	0	0	0	27	0
TSS-Seq	0	0	0	0	26	0
CNV	0	0	0	0	26	0
SV	0	0	0	0	26	0
ATAC-Seq	0	0	0	0	0	106

The screenshot shows the DBKERO Data Portal interface in list view. It displays a table with 2573 hits, categorized by Track ID, Project, Race, Omics Category, Object ID, Object Description, Reference Genome, and Data Category. The data is as follows:

Track ID	Project	Race	Omics Category	Object ID	Object Description	Reference Genome	Data Category
1	IHEC	Japanese	Epigenome	COC11_H3K27Ac_hg38	IHEC COC11 H3K27Ac (bw)	hg38	IHEC
2	IHEC	Japanese	Epigenome	COC11_H3K27me3_hg38	IHEC COC11 H3K27me3 (bw)	hg38	IHEC
3	IHEC	Japanese	Epigenome	COC11_H3K36me3_hg38	IHEC COC11 H3K36me3 (bw)	hg38	IHEC
4	IHEC	Japanese	Epigenome	COC11_H3K4me1_hg38	IHEC COC11 H3K4me1 (bw)	hg38	IHEC
5	IHEC	Japanese	Epigenome	COC11_H3K4me3_hg38	IHEC COC11 H3K4me3 (bw)	hg38	IHEC
6	IHEC	Japanese	Epigenome	COC11_H3K9me3_hg38	IHEC COC11 H3K9me3 (bw)	hg38	IHEC
7	IHEC	Japanese	Transcriptome	COC11_rnaseq_bw_hg38	IHEC COC11 RNA-seq (bw)	hg38	IHEC
8	IHEC	Japanese	Epigenome	COC14_H3K27Ac_hg38	IHEC COC14 H3K27Ac (bw)	hg38	IHEC

図4 DBKERO データポータルサイト

目的のデータをデータマトリクスや、データカテゴリーのチェックボックスから絞り込み、リスト画面や Batch download ボタンから搭載データをダウンロード可能とした。

(イ)「上記ダウンロードポータルから、研究倫理上、再利用可能な全データをダウンロード可能にする。ダウンロード可能にするため、必要に応じてデータ提供者と交渉し、承諾を得る」

DBKERO のデータベースに掲載されている全データセットは 5,510 件で、これらのデータはすべてゲノムブラウザ等を通して可視化されている。このうち倫理規定がクリアされているもの、あるいはその対象外であって、二次配

布が可能であるデータ(多くは培養細胞のデータセットからなる)はダウンロードポータル(Data Portal)に掲載を完了した。具体的には Single cell, IHEC, GWAS, Drug Perturbation, Cancer Cell Line, Long read に関連するデータの、2,573 件(47%)である(表3)。

残りのデータのうち ICGC, TCGA, ENCODE 等外部のデータベースから得たデータ 2,723 件(49%)については Data Portal には公開できないが、ヒト正常組織細胞 TSS-Seq(倫理規定の対象外であるもの)等の残り 203 件(4%)については引き続き公開の準備を進めており、第 5 年次に完了予定である。

表3 ダウンロードポータル搭載済みデータセット数

Project	Assay type										Total
	ATAC-Seq	ChIP-Seq	RNA-Seq	BS-Seq	Re-sequencing	Chrom HMM	TSS-Seq	CNV	SV	GWAS	
IHEC	0	423	105	55	0	0	0	0	0	0	583
Single cell	0	0	1,163	0	0	0	0	0	0	0	1,163
Nanopore	0	0	0	0	3	0	0	0	0	0	3
Cancer Cell Line	0	216	53	26	78	27	26	26	26	0	478
GWAS	0	0	0	0	0	0	0	0	0	14	14
Drug perturbation	106	0	226	0	0	0	0	0	0	0	332
Total	106	639	1,547	81	81	27	26	26	26	14	2,573

第 4 年次はさらに、先端ゲノム支援に関連する論文文化されたデータのうち、17 課題の 250 データセットを公開した。さらに、現在のところ約 20 課題のデータがデータベースへの収載は完了しているが、論文投稿が最終段階である。ゲノム支援については現在もデータ産生が引き続き行われている。残りの部分についても第 5 年次終了までの収載と公開を目指す(前項の表2に加え、さらに多数のデータが公開可能になると考えている)。

2 RDF 化の完了と NBDC RDF ポータルへの完全移行

RDF 化について、肺腺癌 26 細胞種 ChIP-Seq ピーク情報の全データセットについて計画通りに遂行した。2019 年度までに設計が完了した RDF スキーマ(図5)に従って、トランスクリプトーム、エピゲノムのデータを成形し、NBDC が運用する RDF ポータルを通じて公開を開始した(図5)。

第 5 年次に向けて、IHEC 日本担当分のデータの RDF 化も進めている。そのデータセット数を表4に示す。既に ChIP-Seq ピーク情報の一部のデータ(肝臓及び腸管)について、トリプルストアへの投入を行い 2020 年 10 月 1 日より RDF ポータルから公開を開始した (<https://integbio.jp/rdf/dataset/kero>)。これにより RDF 化に向けて喫緊の技術的障壁はある程度、解消されたと考えている。

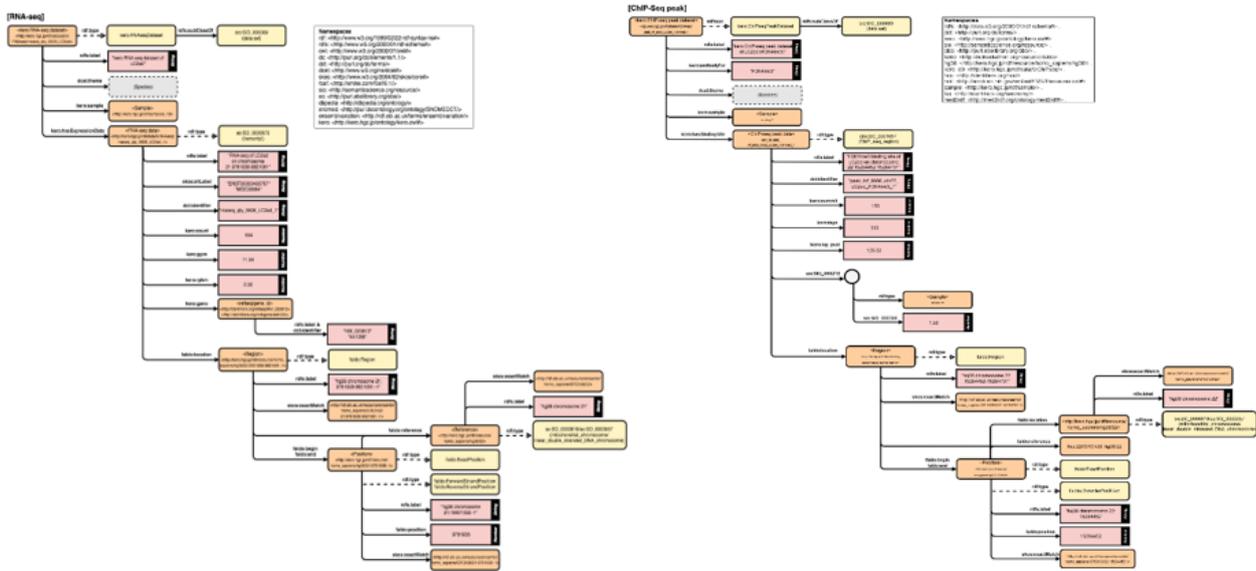


図5 RNA-seq 発現量(左)及び、ChIP-seq ピーク情報(右)のスキーマ図

26肺腺癌セルラインデータのRDF化	2020年4月～2020年9月	2020年10月～2021年3月	2021年4月～2021年9月	2021年10月～2022年3月
ChIP-Seq Peak	→			
IHECデータのRDF化				
COC RNA-Seq	→			
GAC RNA-Seq	→			
HPC RNA-Seq		→		
JTK RNA-Seq		→		
JKU RNA-Seq		→		
COC ChIP-Seq		→		
GAC ChIP-Seq		→		
HPC ChIP-Seq		→		
COC BS-Seq			→	
HPC BS-Seq			→	
JKU BS-Seq			→	
その他組織やその他の培養細胞における RDF化が可能なデータについてのRDF化				→
ロングリード・シングルセル結果について のRDF化				→

図6 第4年次以降のRDF化工程表(背景が赤はデータを変換中のデータ:灰色は公開の承諾を得る必要があるデータ)。第5年次までに現有するすべての該当データセットのRDF化を目指す。

表4 第4,5年次RDF追加登録予定データセット数

チーム	器官	登録名	検体数	データセットサンプル数		
				RNA-Seq	ChIP-Seq	合計
金井班	肝臓	HPC	8	8	48	56
金井班	腸管	COC	11	11	66	77
金井班	胃	GAC	8	8	36	44
白髭班	血管内皮	JTK	33	26	-	26
佐々木班	子宮内膜	JKU	15	15	-	15
合計			75	68	150	218

3 講習会の開催・学会年会でのワークショップやセミナーの企画と実施

第4年次にはコロナ感染状況の推移に鑑みつつオンラインでも講習会を開催した。第4年次の典型的なパターンを下記に示した。

講義時間は60分。

■講演(60分):がん遺伝子等、テーマを決めて、ヒトに関するデータベースを広く検索した。

■演習(60分):その中でマルチオミクスデータベースとしてのDBKEROの特色を解説し、そのウェブベースでの利用をガイドした。

■演習(120分):DBKEROデータポータルに掲載されるデータについて、その加工と利活用を推進する初等プログラミング講習を行った。一例として第4年次の医科研での講習会の告示を示した(図7)。

The poster is titled "Human Genome Center 講習会 「DBTSS/DBKERO 講習会・シングルセル解析入門」". It contains the following information:

- この度、「DBTSS/DBKERO講習会・シングルセル解析入門」と題しまして、転写開始点および疾患多層オームクスデータベースDBTSS/DBKEROの利用説明会と、DBKEROデータを用いたシングルセル解析の入門講習を開催いたします。DBTSSおよびDBKEROは、医科学研究所ヒトゲノム解析センタースーパーコンピュータSHIROKANEにより提供されています。
- 日程: 2020年12月9日(水)
- 時間: 13時00分～17時00分
- 会場: Webexによるオンライン開催
- 定員: 30名(アカウント配布)。聴講のみ場合は特に定員はなし。
- 対象: 多層オームクスデータの活用・シングルセル解析に興味のある方。初学者歓迎します。
- 参加費: 無料
- 本講習会は日本語で実施します。

タイムスケジュール

- 13:00 - 13:30 「1. 最近のオームクスデータ産生のトピック」(鈴木純子)
- 13:30 - 14:00 「2. シングルセル解析、空間トランスクリプトーム解析の概略」(鈴木純子)
- 休憩 --
- 14:10 - 14:30 「3. データベース DBTSS/DBKERO 講習会」(鹿島幸恵・若菜浩幸)
- 14:30 - 17:00 「4. DBKERO シングルセル解析講習会」(鹿島幸恵・若菜浩幸)

申し込み方法

下記URLのフォームからお申込みください。後日WebexのURLをご連絡します。
[URL貼付予定]

皆様のご参加をお待ちしています。

問合せ先: 鈴木純子(内線: 64076)

本講習会は、医科学研究所ヒトゲノム解析センターの主催で開催されます。また、大学院新領域創成科学研究科 生命データサイエンス教育プログラム DSTEP(<http://dstep.dems.k.u-tokyo.ac.jp/>)/セミナーシリーズの一部として行われます。

図7 講習会の告示(第4年次に東大・医科研において行ったものである)

人類遺伝学会、分子生物学会で口演し、データベースの紹介を行った。またNBDCの主催するAJACS講習会でも口演を行った。

(その他)

前述のように、第4年次には、コロナ禍の状況下、対面での説明会、ハンズオンでの講習会を開催することが容易ではなかった。上記のようにいくつかのオンライン講習会を開催するが、それを補完する意味で、DBKERO上にデータベース利用を解説する動画を加えた。これは各ブラウザのページごとその利用法を実際に解説したものである。5分程度のshort動画10本程度からなる(図8)。また、特にシングルセル解析については、実際にどのようにデータポータル上のデータを加工、自身の目的に利活用ができるのか、という質問が多数寄せられた。これに対応するために、DBKERO掲載のマウス肺のscRNAデータを材料にデータプロセス手法を解説したTutorial Pageを開設した。本ページでは、同じくマウス肺のscATACデータとのオームクス階層統合手法も解説されている(図9)。これらのウェブコンテンツの充実により、従来、ハンズオン講習会により行うべきであった新規利用者への周知と既存利用者の便宜のさらなる向上に努めた。



図8 チュートリアル動画の追加: 赤枠内等、各ページでそのページの使い方をガイドした動画を配した。

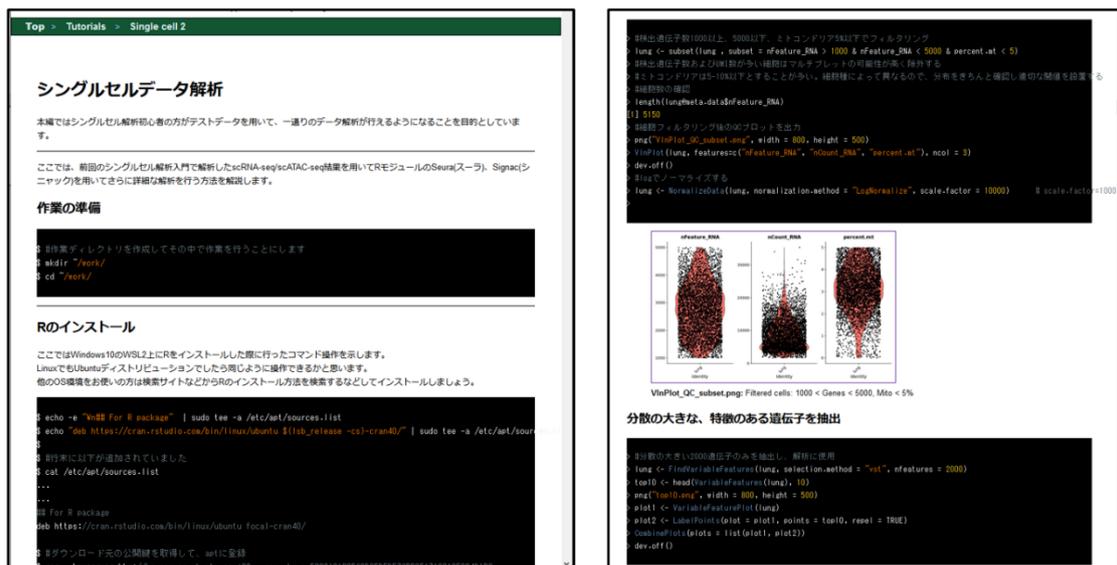


図9 シングルセルデータ解析処理手順の解説ページ
DBKERO 収載データについて、それぞれの加工手順を詳細にチュートリアルした。

4 利用者の拡大

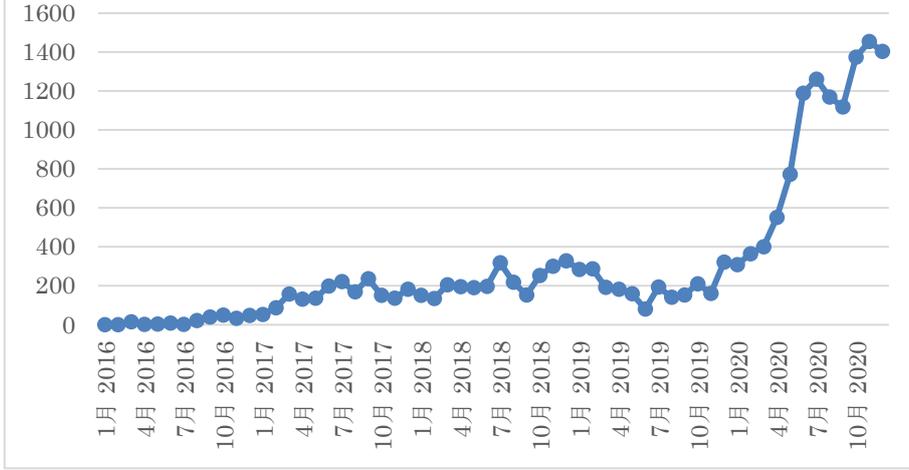
2017年1月以降のAWSTATSによるWebアクセス統計から、講習会等の実施に従い最近ではアクセス数、特に訪問者数(新規アクセス者)が増えつつある。ページビュー数については、アクセスするファイル数を減らす工夫が逆にページビュー数を減らす影響を与えることもあり、評価が難しく月によるばらつきも大きい、概ね上昇傾向にあると思われた(表5、図9)。今後もコンテンツの充実とともに広報活動をより活発にして、さらなる利用者の拡大に努める。

表5 AWSTATS による利用者統計

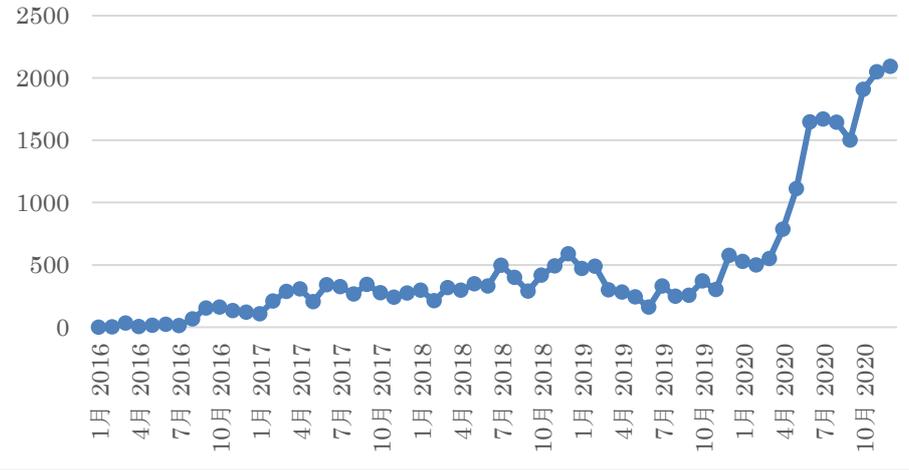
月	訪問者	訪問数	ページ	件数	バイト
1月2017	53	107	1,403	1,987	80.70 Mb
2月2017	88	210	1,500	2,845	103.73 Mb
3月2017	157	287	2,566	4,323	107.59 Mb
4月2017	131	306	1,259	3,119	117.51 Mb
5月2017	136	204	480	1,497	58.46 Mb
6月2017	199	339	1,393	3,190	167.40 Mb
7月2017	221	324	1,885	3,745	159.86 Mb
8月2017	169	267	702	1,518	71.74 Mb
9月2017	236	342	962	2,173	46.51 Mb
10月2017	151	276	361	361	114.54 Kb
11月2017	136	239	7,582	14,568	150.23 Mb
12月2017	183	273	2,464	3,215	56.38 Mb
1月2018	151	296	1,358	2,798	3.01 Gb
2月2018	134	213	2,558	6,512	146.53 Mb
3月2018	206	318	4,617	12,042	361.31 Mb
4月2018	196	296	4,036	7,564	503.15 Mb
5月2018	190	349	2,404	5,269	299.51 Mb
6月2018	197	330	4,833	10,395	1.83 Gb
7月2018	318	497	11,697	17,093	583.76 Mb
8月2018	218	398	16,337	20,676	935.36 Mb
9月2018	153	290	6,631	7,952	308.71 Mb
10月2018	252	417	5,537	8,653	360.37 Mb
11月2018	301	492	9,305	13,601	904.65 Mb
12月2018	328	588	12,176	16,108	807.95 Mb

月	訪問者	訪問数	ページ	件数	バイト
1月2019	284	471	3,014	4,682	223.66 Mb
2月2019	287	489	2,938	4,397	318.89 Mb
3月2019	192	300	1,886	3,539	161.82 Mb
4月2019	183	282	2,004	3,160	203.61 Mb
5月2019	159	244	1,232	2,482	166.73 Mb
6月2019	80	161	813	1,407	69.15 Mb
7月2019	194	330	1,656	3,145	175.75 Mb
8月2019	142	247	892	1,854	85.73 Mb
9月2019	153	256	2,904	5,065	3.07 Gb
10月2019	210	372	70,409	80,599	4.62 Gb
11月2019	161	303	9,102	13,309	945.52 Mb
12月2019	321	577	20,420	36,896	9.00 Gb
1月2020	308	528	9,920	19,330	743.78 Mb
2月2020	364	500	5,792	9,966	237.27 Mb
3月2020	400	550	4,735	10,407	1.38 Gb
4月2020	552	785	10,250	17,296	1.01 Gb
5月2020	773	1,110	11,189	19,434	94.46 Gb
6月2020	1,189	1,647	11,635	18,968	61.00 Gb
7月2020	1,261	1,669	15,767	23,366	427.33 Gb
8月2020	1,169	1,644	64,834	76,261	921.00 Gb
9月2020	1,118	1,500	13,741	22,964	899.08 Gb
10月2020	1,375	1,909	28,965	47,147	421.53 Gb
11月2020	1,454	2,049	23,795	46,805	645.87 Gb
12月2020	1,403	2,092	24,167	58,825	1319.81 Gb

訪問者数推移



訪問数推移



ページビュー数推移

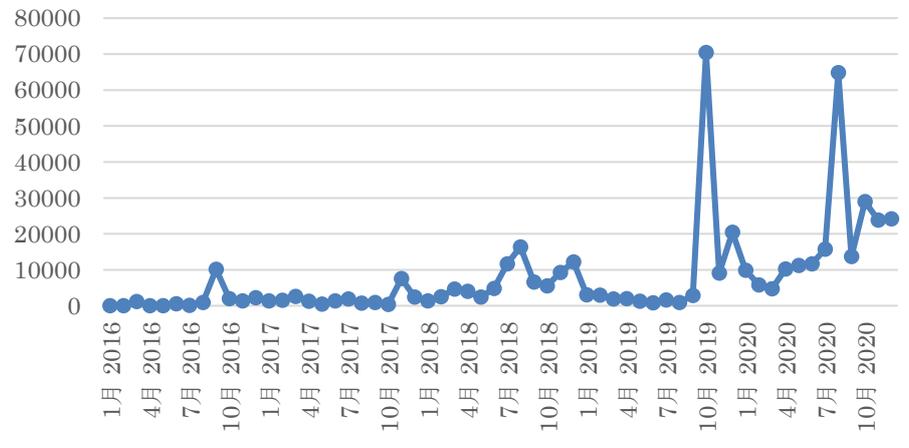




図9 訪問者数・訪問数・ページビュー数・ファイル件数の推移

※2019年10月にページビュー数、ファイル件数が増えているのは、脆弱性テストを行った影響である。2020年9月に訪問数・アクセス数が減少しているのはサーバメンテナンスで1週間の停止期間があった影響が考えられる。2020年12月にファイル件数が多いのはハンズオンセミナーを行ったため。

§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)	2 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. Oka M, Xu L, Suzuki T, Yoshikawa T, Sakamoto H, Uemura H, Yoshizawa AC, Suzuki Y, Nakatsuma T, Ishihama Y, Suzuki A, Seki M., "Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer", *Genome Biol.*, vol. 22, No. 9, 2020 (DOI: 10.1186/s13059-020-02240-8.).
2. Sakamoto Y, Xu L, Seki M, Yokoyama TT, Kasahara M, Kashima Y, Ohashi A, Shimada Y, Motoi N, Tsuchihara K, Kobayashi SS, Kohno T, Shiraiishi Y, Suzuki A, Suzuki Y., "Long-read sequencing for non-small-cell lung cancer genomes", *Genome Res.*, vol. 30, No. 9, pp. 1243-1257, 2020 (DOI: 10.1101/gr.261941.120.).

(2) その他の著作物(総説、書籍など)

該当なし。

(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	3 件
	国際	1 件
口頭発表	国内	1 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〉

1. 鈴木穰、がんのシングルセルおよび空間トランスクリプトーム解析、第 79 回日本癌学会学術総会、リーガロイヤルホテル広島/メルパルク広島(オンラインとのハイブリッド開催)、2020/10/1-3
2. 鈴木穰、シングルセル解析技術の進展、第 19 回日本再生医療学会総会、2020/5/18-29、2020/8/27-9/10、国内(オンライン)、口頭
3. Yutaka Suzuki, Introduction to Single Cell Analysis、第 84 回日本循環器学会学術総会、オンライン開催、2020/7/27-8/2

〈国際〉

1. Yutaka Suzuki, Characterization of complicated structural variants with local copy-number changes in cancer genomes using PromethION, the 19th International Symposium on Pediatric Neuro-Oncology 2020、軽井沢プリンスホテルウエスト、2020/12/13-16

③ 口頭講演

〈国内〉

1. 鈴木穰、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ミクスデータの統合、第 43 回日本分子生物学会年会、オンライン開催、2020/12/2-4

〈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
2020年 12月4日	チーム内ミーティング (非公開)	オンライン	5人	DBTSS/DBKERO 講習会・シングルセル解析入門のためのミーティング
2020年 12月18日	チーム内ミーティング (非公開)	オンライン	5人	DBTSS/DBKERO 講習会・シングルセル解析入門のためのミーティング
2021年 1月8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	進捗確認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

年月日	名称	場所	参加人数	目的・概要
2020年 12月9日	DBTSS/DBKERO 講習会・シングルセル解析入門	オンライン	100人	転写開始点および疾患多層オーミクスデータベース DBTSS/DBKERO の利用説明会と、DBKERO データを用いたシングルセル解析の入門講習
2021年 3月10日	空間トランスクリプトーム解析【Visium】講習会	オンライン	100人	空間トランスクリプトーム Visium データ解析の入門講習

以上

