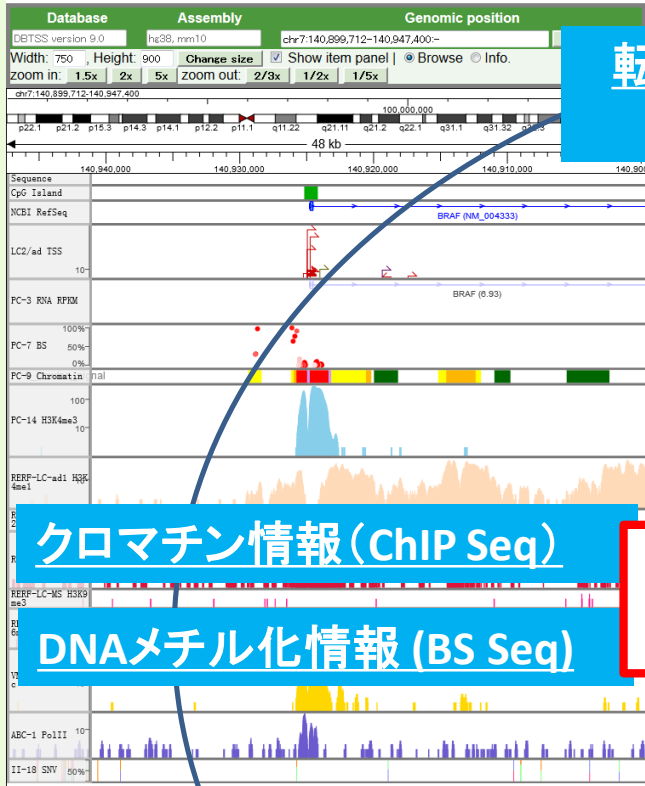


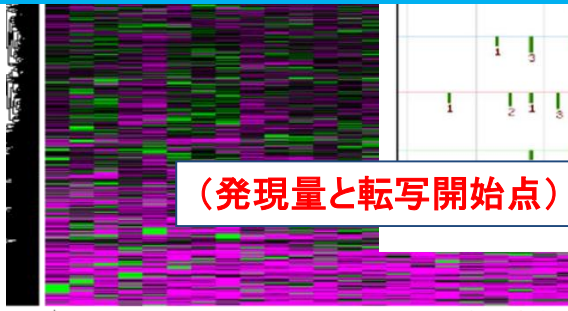
多階層オームクスデータ DBKERO: Kashiwa Encyclopedia for Researchers of multi-Omics analysis)

○鈴木穰、菅野純夫
東京大学新領域創成科学研究科
メディカル情報生命専攻

Genome viewer (Multi-omics Data)

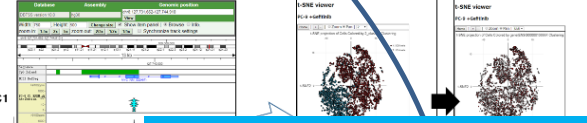


転写開始点/トランスクリプトーム情報
(TSS/RNA Seq)



(発現量と転写開始点)

Single Cell Viewer



オミクス新解析技術

- Single cell 解析
- Long read 解析

多層オミクス情報の
ヒトゲノム変異情報への統合

統合DB(DBKERO)

Clinical Sequencing

(それぞれの検体での変異部位)

パスウェイマップ(文献情報)からの検索

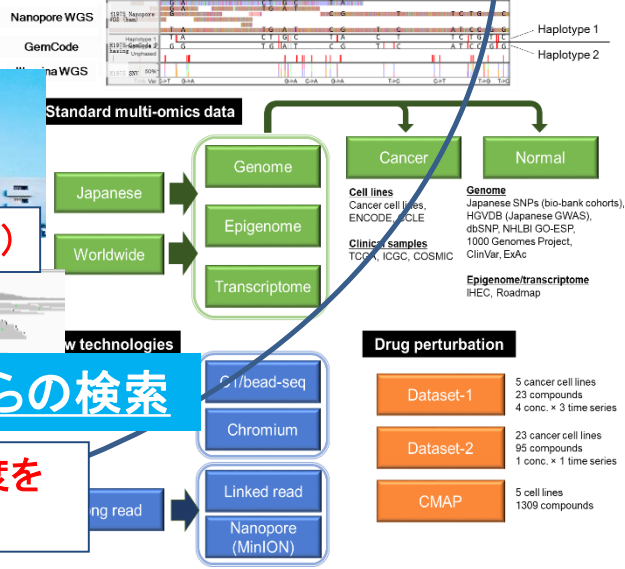
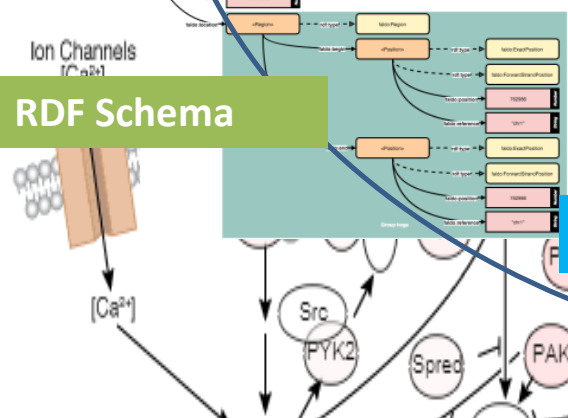
(該当集団中の遺伝子変異頻度を
赤の濃さで示す)

クロマチン情報(ChIP Seq)

DNAメチル化情報(BS Seq)

RDF Schema

Ion Channels (Ca²⁺)



http://kero.hgc.jp

DBKERO

Database of Korean Genotypes for Researcher of multi-Omics data

Quick-start: For the beginners

使い方の説明

to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any

Featured Dataset

[Single cell dataset](#) [[Download processed data](#)]:
Single cell dataset

[Cancer SV dataset](#) ^{New}
Japanese lung adenocarcinoma sequenced by PromethION(Oxford Nanopore Technologies).

[Visium dataset](#) ^{New}
Spatial Gene Expression dataset (10x Genomics Visium system).
Spaceranger output files for Ioupe browser and Seurat analysis.

Documents

- [Experimental Procedures](#)
- [Data Contents](#)
- [Tutorial](#)
- [Download bulk data](#)
- [References](#)

Movie (For Japanese)

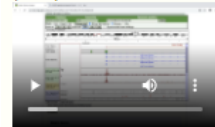
Overview movie 1 in Japanese (9 min.)



Overview movie 2 in Japanese (53 min.)



How to use genome browser (Basic)



How to use genome browser (Track selection)



How to use genome browser (Comparative viewer)



How to use data portal



Download

- [Data Portal](#) [[Help video](#) (Japanese)]

Tools

[TSS-Viewer](#) [[Help video](#) (Japanese)]:

Find out transcriptional start sites and compare promoter usage.

[Multi-Omics-Viewer](#) [[Help video](#) 1,2,3 (Japanese)], [[GitHub](#)]:

Browse mutations, transcripts and epigenetic modifications along the genome coordinates.

[Mutation Enriched Genes](#) [[Help video](#) (Japanese)]:

Find out which gene is mutated most in our data.

[TF Binding Site Search](#) [[Help video](#) (Japanese)]:

Find out TF binding sites in a region of genome of various organism.
Powered by [ChIP-Atlas*](#) and [GGGenome*](#).

[Pathway Map](#) [[Help video](#) (Japanese)]:

Find out the level of expression or modification of genes within a Pathway

[Chromatin-status Data Summary](#) [[Help video](#) (Japanese)]:

Get an overall view of expression and modification of a genome region among our entire data sets.

- [Search from Genomic Position](#)
- [Search from SNP \(dbSNP rsID\)](#)
- [Search from SNV \(COSMIC: somatic mutation\)](#)

[SNV Summary in Cancers](#) [[Help video](#) (Japanese)]:

Find out mutation frequency of a gene among our cancer data sets.

RDF gate (Trial)

Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq

- [RDF Schema](#)
- [RDF Browser](#)
- [SPARQL Endpoint](#)

Other tools

- [Single cell viewer](#) [[Help video](#) (Japanese)]

Tutorialの動画
(それぞれの
ブラウザについて)

何ができるか(1)

日本人正常オーミクスデータ (IHEC)

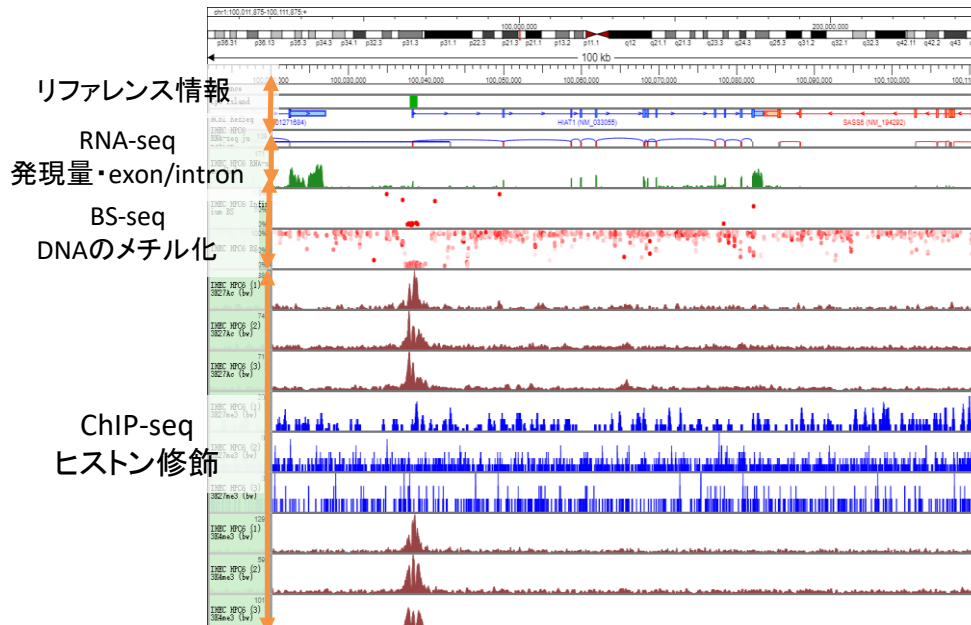
国際エピゲノムコンソシアム (IHEC)

で日本CRESTチームが収集した日本人標準エピゲノムが検索できます。

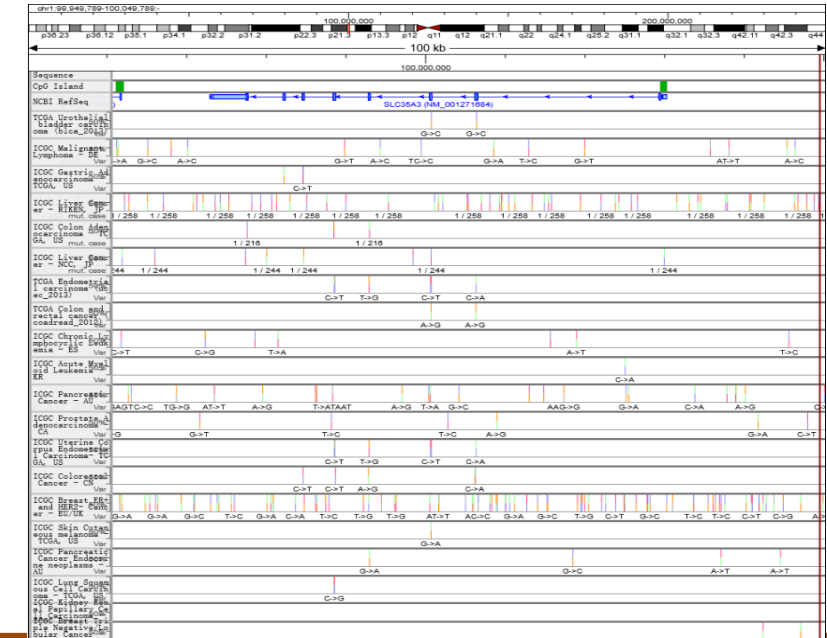
CREST-IHECデータの統合(日本人標準エピゲノム)

IHEC(国際エピゲノムコンソシアム)データの公開

マルチオミクス情報の表示



TCGA/ICGC変異情報の表示



IHECデータとしてゲノムブラウザから閲覧可能なオープンデータ一覧

Sample	individuals	Expression		DNA methylation	Histone modifications					
		RNA-seq		BS-seq	H3K4me1	H3K4me3	H3K27ac	H3K27me3	H3K36me3	H3K9me3
		Read depth	Exon/intron junction							
Liver	8	8	8	8	13	13	13	13	13	13
Colon	11	11	11	11	11	11	11	11	11	11
Endometrial	15	30	0	13	14	15	15	15	15	15
Vascular endothelial	33	26	0	0	19	27	27	19	27	26

IHEC日本チームの作成した正常肝臓、血管内皮、胎盤組織におけるエピゲノムカタログを標準エピゲノムデータとして収載した。上の表は既にKEROから閲覧可能なデータセットの数。ゲノムブラウザ下部のTracksにある「CREST - IHEC (The International Human Epigenome Consortium (AMED-CREST, Japan))」の一覧表から、任意のデータ結果をゲノムブラウザ上に追加可能である(上図)。

Human Genome Variation DB (旧徳永グループ)との統合

日本人ゲノム多型

HGVDBへのリンク機能

DBKERO

Release 9.0 (Updated July 9, 2015)
Based on UCSC hg19 build

About this database

Welcome to DBTSS (Database of Transcriptional Start Sites)

To support transcriptional regulation studies, we have constructed the DBTSS (Database of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-seq.

This database includes TSS data of a major part of human adult and embryonic tissues are covered. DBTSS now contains 481 million TSS sequences for collected from a total of 20 tissues and 7 cell cultures. We also integrated our newly generated RNA-seq data of subcellular-fractionated RNAs and ChIP-seq data of histone modifications, RNA polymerase II and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external epigenomic data, such as chromatin map of the ENCODE project.

In this update, we further associated those TSS information with public and original SNV data, in order to identify single nucleotide variations (SNVs) in the regulatory regions.

It is believed that single nucleotide variations (SNVs) in the transcriptional regulatory regions are responsible for many human diseases, including cancers. However, it remains difficult to identify functionally relevant SNVs for those having no explicit biological consequences. In this version of DBTSS, we attempt to associate SNVs with the genes information of the surrounding regions. We used SNVs which we identified from genomic analyses of various types of cancers, including somatic mutations of 100 lung adenocarcinoma and lung small cell carcinoma. For germline variations, we used SNVs in dbSNP as well as our unique dataset of variations in 1000 Japanese individuals. We integrated those SNV information with our original datasets of TSS-seq, RNA-seq, ChIP-seq of representative histone modifications and BioRad Sequencing of cytosine methylations of DNA. Particular, we present multi-omics data of 50 lung adenocarcinoma cells line for which TSS-seq, RNA-seq, ChIP-seq and BS-seq together with whole genome sequencing are collected from the same materials. We further connected the multi-omics data of model organisms by genome-genome alignment. We provide a unique data resource to investigate what genomic features are observed in a particular genomic coordinates in a wide variety of samples.

News

- 09 Jul. 2015: New T helper cell data of mouse are now available. Raw data accession: GSE600028, GSE600102 and SRR007084 (Genome Shown)
- 30 Jun. 2015: New BRIC-Seq data (UPF1 RNA) are now available. Raw data accession: GSE600052 (Genome Shown)
- 15 Sep. 2014: New DBTSS opened

Search

Release 9.0 (Updated July 9, 2015)
Based on UCSC hg19 build

Database Search

Keywords Search

Species: H. sapiens

Keyword: NA

Search

Front size: 15

Show: 15 entries

Genomic position

NA_000014 chr12:9,067,708-9,115,962 TSS viewer Genome viewer Human Variation DB

NA_000015 chr10:39,246,761-40,121,213 TSS viewer Genome viewer Human Variation DB

NA_000016 chr1:19,724,347-75,763,079 TSS viewer Genome viewer Human Variation DB

NA_000017 chr12:129,725,768-130,740,038 TSS viewer Genome viewer Human Variation DB

NA_000018 chr17:2,219,871-2,225,367 TSS viewer Genome viewer Human Variation DB

NA_000019 chr11:108,121,531-108,148,164 TSS viewer Genome viewer Human Variation DB

NA_000020 chr12:71,907,418-51,923,361 TSS viewer Genome viewer Human Variation DB

NA_000021 chr14:73,136,435-73,223,681 TSS viewer Genome viewer Human Variation DB

NA_000022 chr20:44,610,522-44,651,735 TSS viewer Genome viewer Human Variation DB

NA_000023 chr17:50,160,005-50,175,932 TSS viewer Genome viewer Human Variation DB

NA_000024 chr5:148,626,593-148,628,634 TSS viewer Genome viewer Human Variation DB

NA_000025 chr3:37,962,866-37,966,666 TSS viewer Genome viewer Human Variation DB

NA_000026 chr22:40,348,500-40,396,571 TSS viewer Genome viewer Human Variation DB

NA_000027 chr17:418,775-177,442,603 TSS viewer Genome viewer Human Variation DB

NA_000028 chr1:99,650,489-99,624,023 TSS viewer Genome viewer Human Variation DB

NA_000029 chr1:230,702,523-230,714,500 TSS viewer Genome viewer Human Variation DB

NA_000030 chr2:460,960,745-240,879,119 TSS viewer Genome viewer Human Variation DB

NA_000031 chr1:113,386,312-113,401,338 TSS viewer Genome viewer Human Variation DB

NA_000032 chr3:55,009,055-55,031,064 TSS viewer Genome viewer Human Variation DB

NA_000033 chr1:153,724,868-153,744,702 TSS viewer Genome viewer Human Variation DB

NA_000034 chr16:30,050,060-30,020,420 TSS viewer Genome viewer Human Variation DB

NA_000035 chr6:101,420,550-101,435,780 TSS viewer Genome viewer Human Variation DB

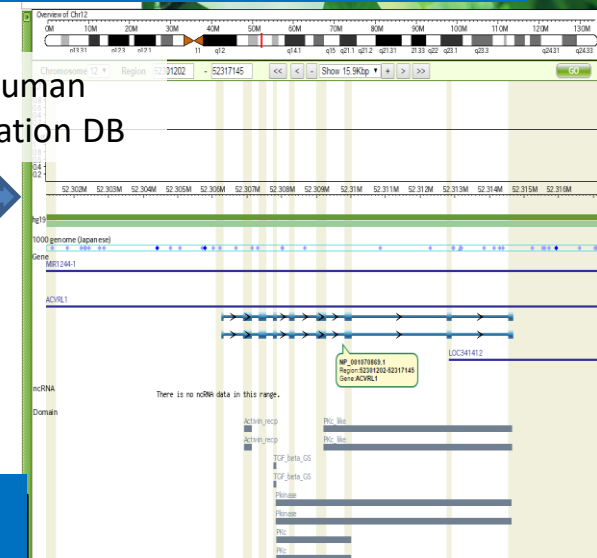
NA_000036 chr1:114,673,088-114,686,618 TSS viewer Genome viewer Human Variation DB

NA_000037 chr41:653,225-41,707,622 TSS viewer Genome viewer Human Variation DB

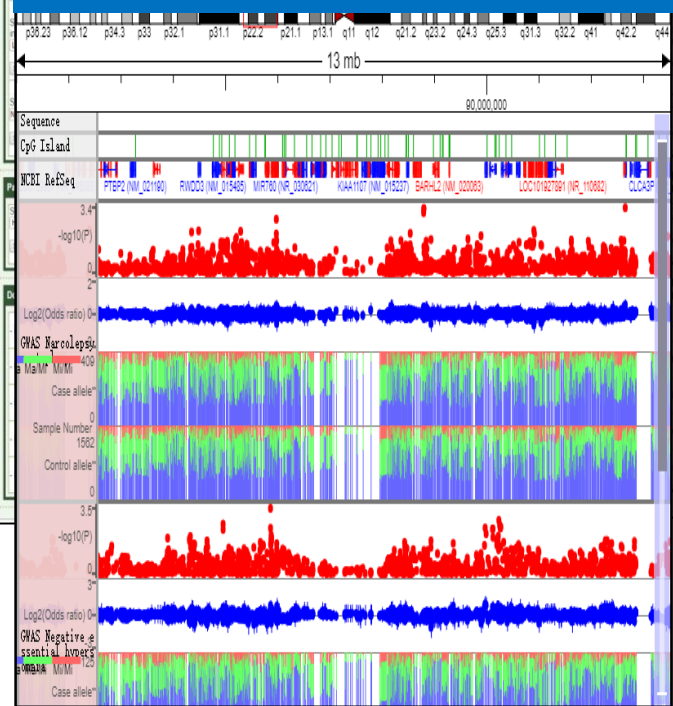
NA_000038 chr5:112,737,858-112,846,239 TSS viewer Genome viewer Human Variation DB

NA_000039 chr5:94,692,763-148,937,025 TSS viewer Genome viewer Human Variation DB

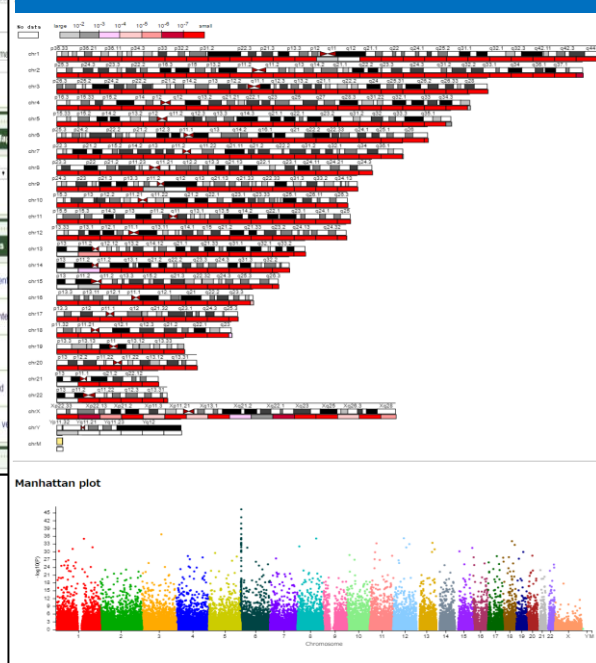
Human Variation DB



Genome browserへGWAS情報の可視化



Whole genome viewer, Manhattan viewer



NBDC ヒトDBとの連携へ

先進ゲノム支援(文科省新学術領域)

<http://www.genome-sci.jp/>

文部科学省科学研究費助成事業「新学術領域研究『学術研究支援基盤形成』」

先進ゲノム解析研究推進プラットフォーム

Google “ゲノム支援”

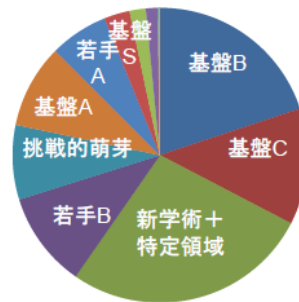
先進ゲノム支援

ホーム 先進ゲノム支援とは 公募要項 支援申請の流れ 支援申請 公募支援に関するFAQ お問い合わせ

新学術領域研究「先進ゲノム解析システム」の提供することにより様々な

この過程でゲノム解析システムを用いては生命科学のさまざまな研究が進展します。この支援にふさわ

② 支援採択課題の元科研費

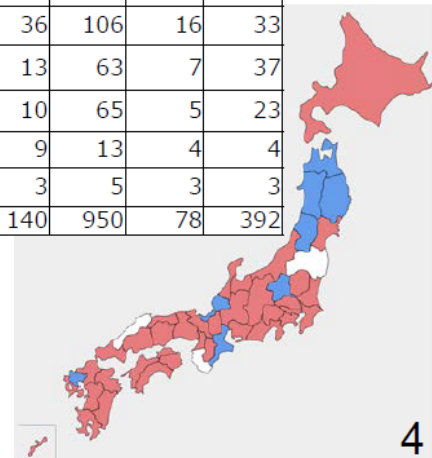


全期間の応募・採択数

年度	H22	H23	H24	H25	H26	H27
応募数	166	237	200	172	161	211
採択数	69	62	81	88	91	74

③ 支援依頼者所在機関 (H26まで)

	応募		採択	
	機関	件数	機関	件数
国立大学	53	662	38	282
公立大学	16	36	5	10
私立大学	36	106	16	33
国立研究所	13	63	7	37
独法研究所	10	65	5	23
公立研究所	9	13	4	4
財団研究所	3	5	3	3
合計	140	950	78	392



ysuzuki@hgc.jp

何ができるか(2)

どの培養細胞をモデルにするのか？ を選ぶオーミクスデータ

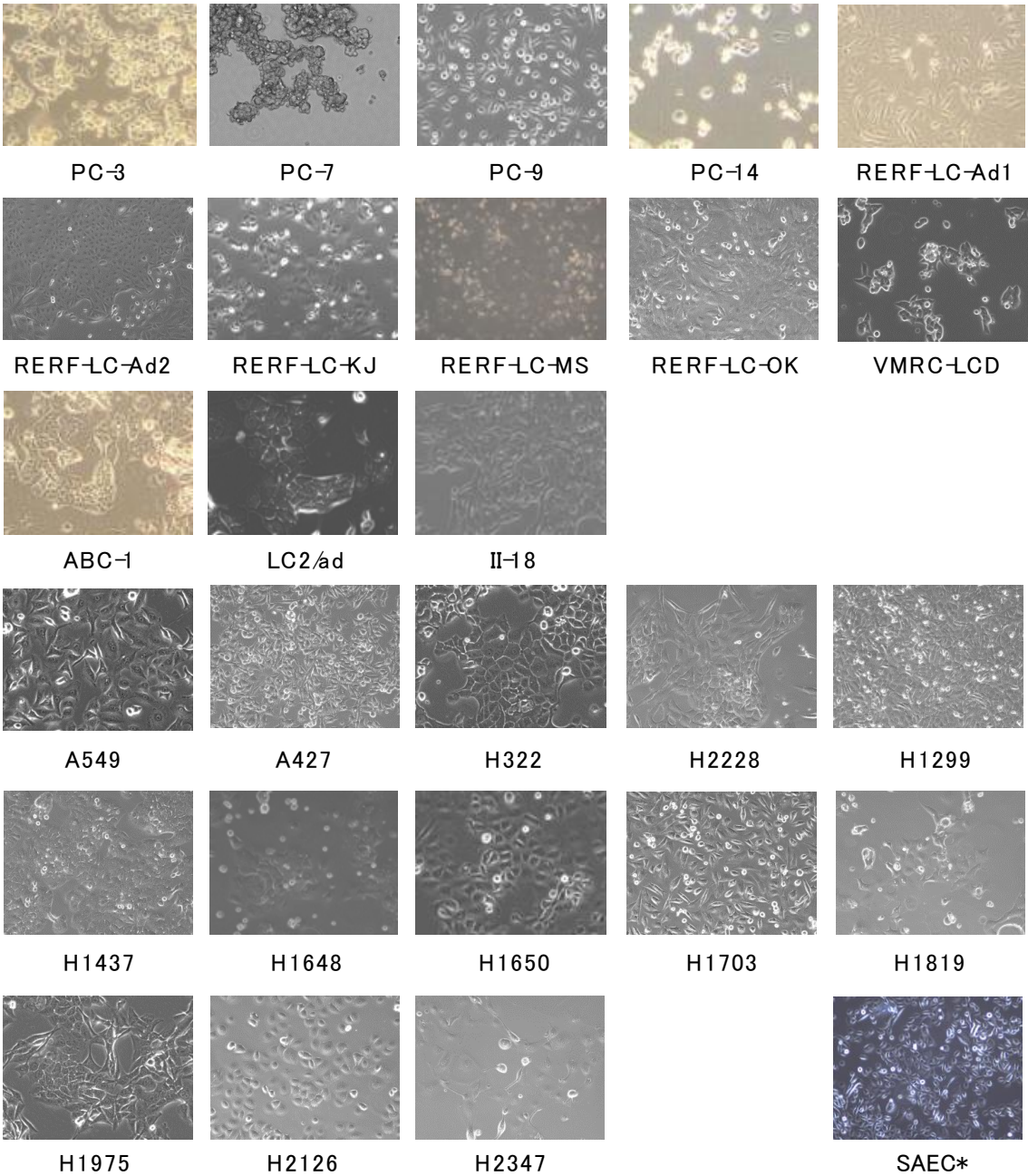
肺腺癌細胞株等で日本人由来のオーミクスデータが充実しています。

※実は公開に向けての倫理の問題もあります。

培養細胞オーミクス

26 lung adenocarcinoma cell lines

Cell line	Origin
PC-3	日本人由来
PC-7	
PC-9	
PC-14	
RERF-LC-Ad1	
RERF-LC-Ad2	
RERF-LC-KJ	
RERF-LC-MS	
RERF-LC-OK	
VMRC-LCD	
ABC-1	Non-Japanese
LC2/ad	
II-18	
A427	
A549	
H322	
H2228	
H1299	
H1437	
H1648	
H1650	
H1703	
H1819	
H1975	
H2126	
H2347	



もちろん、ENCODE/CCLEにもありますが...

*Normal small airway epithelial cells

Multi-omics sequencing data of lung cancer cell lines

26 NSCLC cell lines

13 Japanese
13 non-Asian
+ normal (SAEC)

Genome

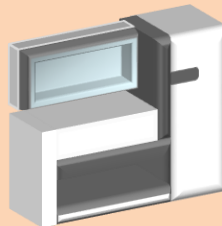
- Whole-genome sequencing
- Long-read sequencing (MinION/PromethION)

Epigenome

- Bisulfite seq/**EM seq** (not yet)
- ChIP-Seq
- ATAC-Seq

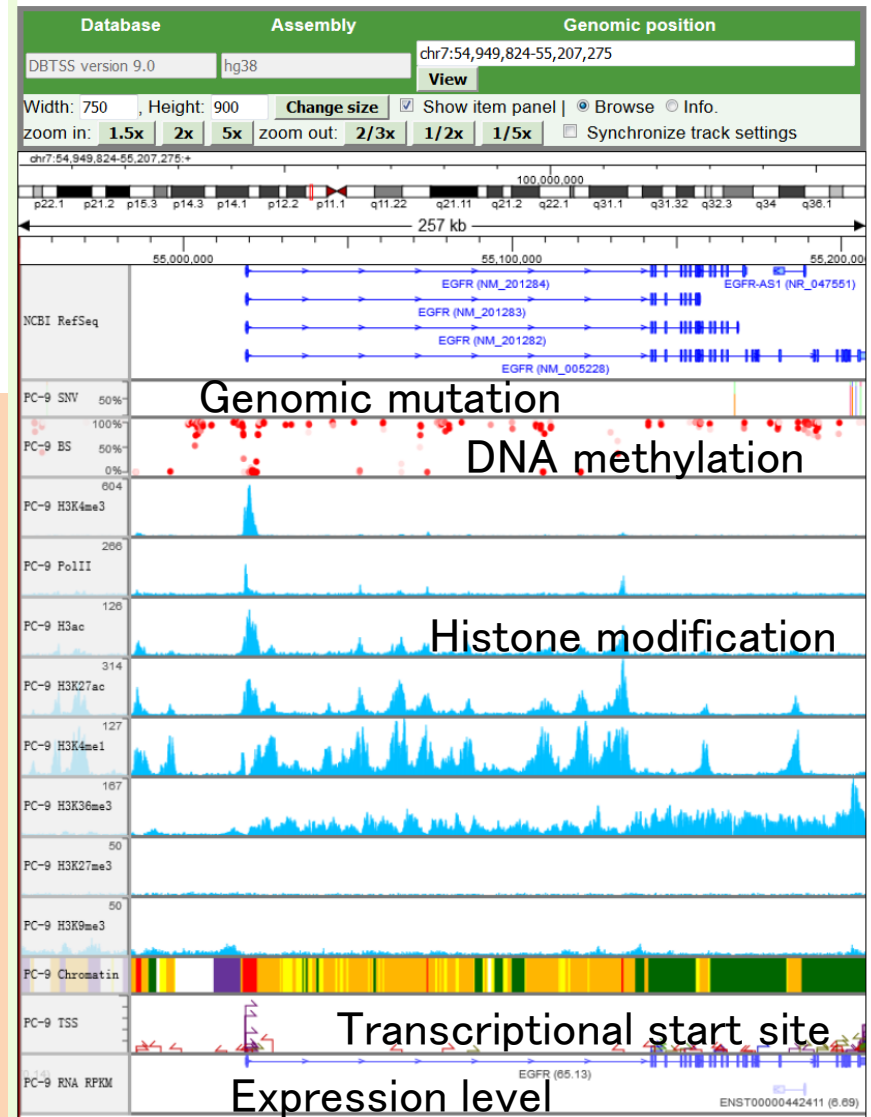
Transcriptome

- TSS-Seq
- RNA-Seq
- miRNA-seq



Genome viewer (Multi-omics Data)

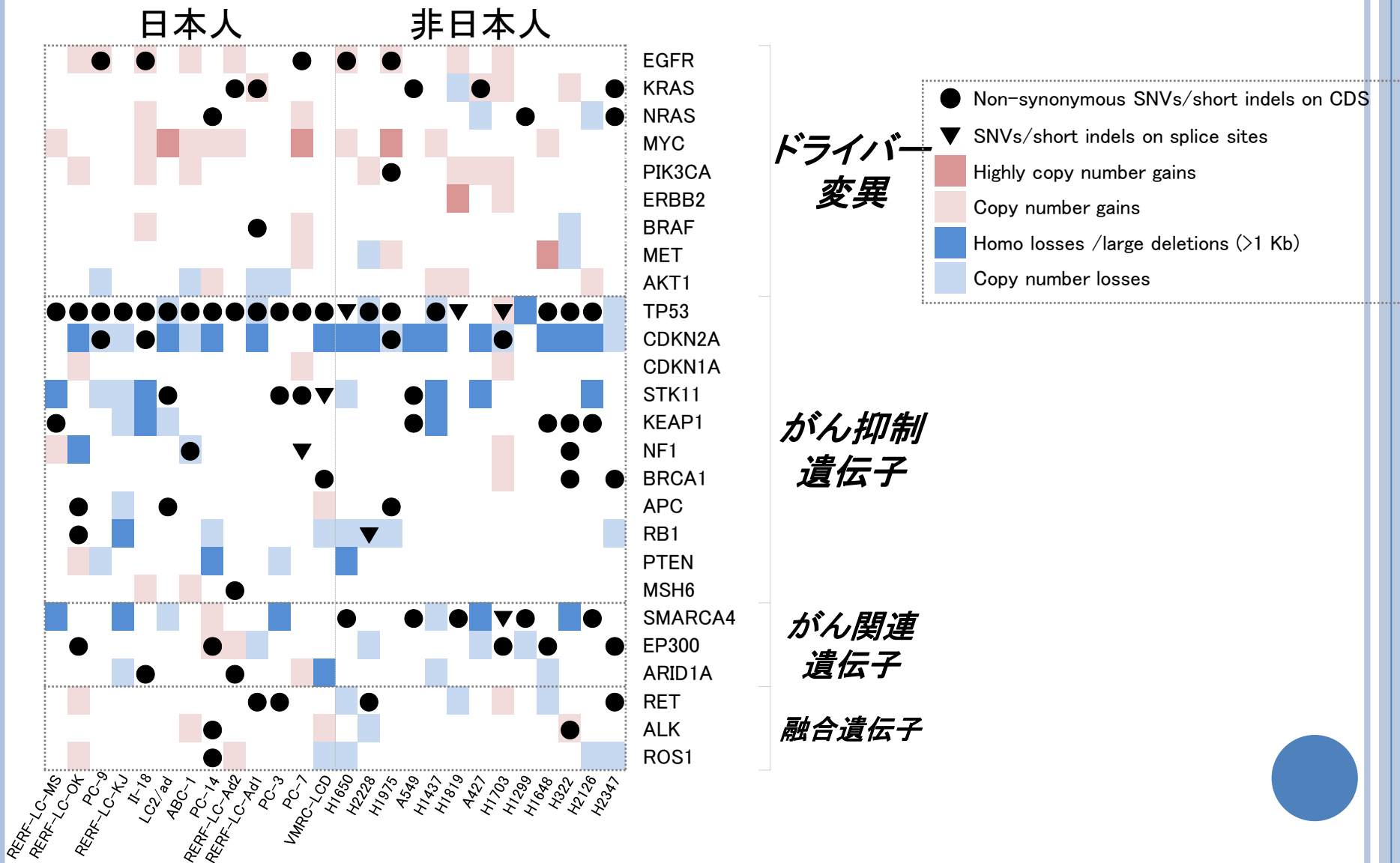
DBKERO



Suzuki et al. 2014 Nucleic Acids Research

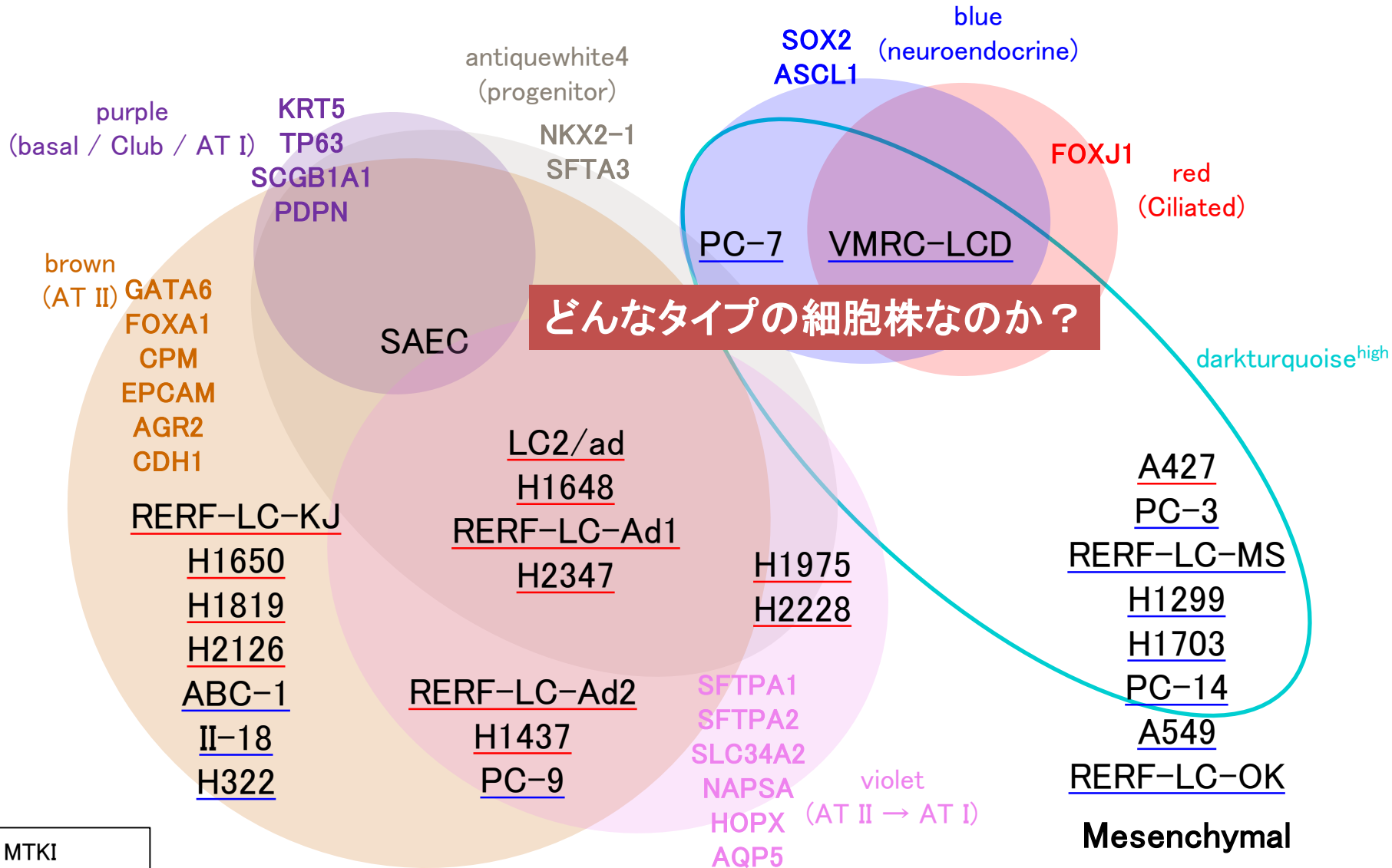
Suzuki et al. 2015 Nucleic Acids Research DB issue

どんなゲノム変異をもっている細胞株なのか



肺腺癌細胞株のTranscriptome/Epigenome層別化

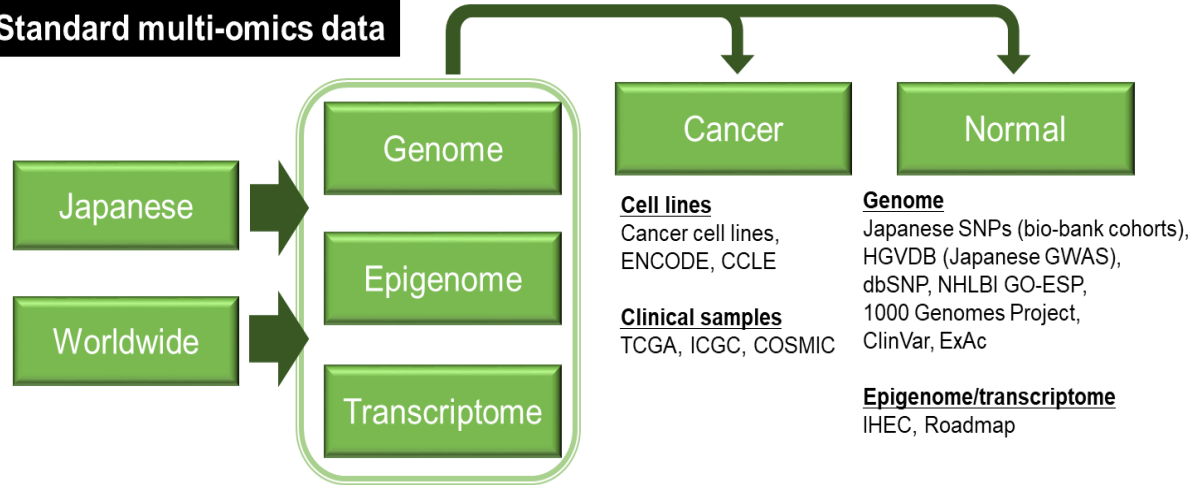
Pan-modules



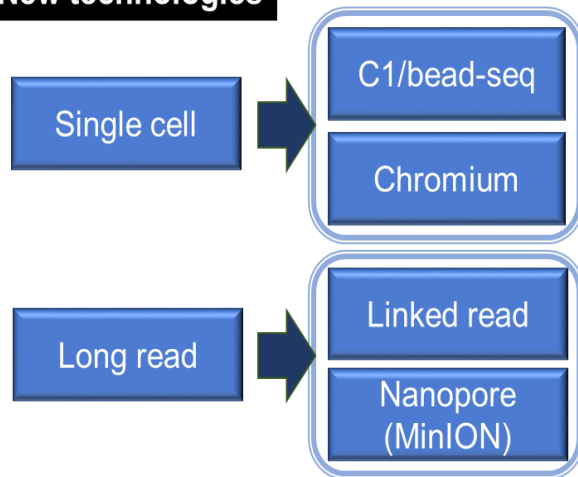
データコンテンツの選択と表示

KEROデータコンテンツ

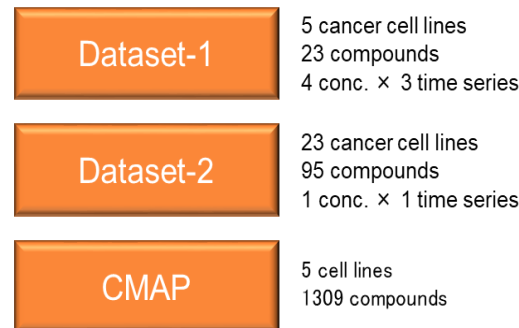
Standard multi-omics data



New technologies



Drug perturbation



DBKEROのデータ構造の全体像。日本人の臨床オミックス情報がモデルシステムからの包括的オミックス情報とどのように関連しているかを示す。

New technologiesとして、近年新たに産出可能となったシングルセルとロングリードのデータやDrug perturbationとして、化合物によるマルチオミックス摂動の情報を含んでいる。

異なるカテゴリーのデータセットは、異なる色のボックスとして示した。

We recommend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Top | Multi-Omics-Viewer | Tutorial

Tools

TSS-Viewer [Video (Japanese)]:

Find out transcriptional start sites and compare promoter usage.

Multi-Omics-Viewer [Video 1.2.3 (Japanese)] [GitHub]:

Browse mutations, transcripts and epigenetic modifications along the genome coordinates.

Mutation Enriched Genes [Video (Japanese)]:

Find out which gene is mutated most in our data.

TF Binding Site Search [Video (Japanese)]:

Find out TF binding sites in a region of genome of various organism. Powered by ChIP-Atlas* and GGGenome*.

Pathway Map [Video (Japanese)]:

Find out the level of expression or modification of genes within a Pathway

Chromatin-status Data Summary [Video (Japanese)]:

Get an overall view of expression and modification of a genome region among our entire data sets.

- Search from Genomic Position
- Search from SNP (dbSNP rsID)
- Search from SNV (COSMIC: somatic mutation)

SNV Summary in Cancer [Video (Japanese)]:

Find out mutation frequency of a gene among our cancer data sets.

RDF gate (Trial)

Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq

- RDF Schema
- RDF Browser
- SPARQL Endpoint

Other tools

- Single cell viewer [Video (Japanese)]
- SV viewer new

About this database

Download

- Data Portal [Video (Japanese)]

Documents

- Experimental Procedures
- Data Contents
- Tutorial
- Download bulk data
- References

Movie (For Japanese)

Overview movie 1 in Japanese (9 min.)



Overview movie 2 in Japanese (53 min.)



How to use genome browser (Basic)



How to use genome browser (Track selection)



How to use genome browser (Comparative viewer)



How to use data portal



News

- 1 Oct. 2020: The mouse lung single cell data sets are now available [here](#).

各種使い方のチュートリアルビデオヘリンク

ブラウザの使い方

Quick-start: For the beginners

Release 1.2.3 Updated (Aug. 19, 2020)

Based on UCSC hg38, mm10

Formerly DBTSS

We recommend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Top | Tutorials | Multi-Omics-Viewer (Basic)

マルチオミクスビューの使い方 (基本)



転写因子結合領域予測の使い方

Quick-start: For the beginners

Release 1.2.3 Updated (Aug. 19, 2020)

Based on UCSC hg38, mm10

Formerly DBTSS

We recommend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Top | Tutorials | TF binding site search

転写因子結合領域の予測検索



データポータルの使い方

Quick-start: For the beginners

Release 1.2.3 Updated (Aug. 19, 2020)

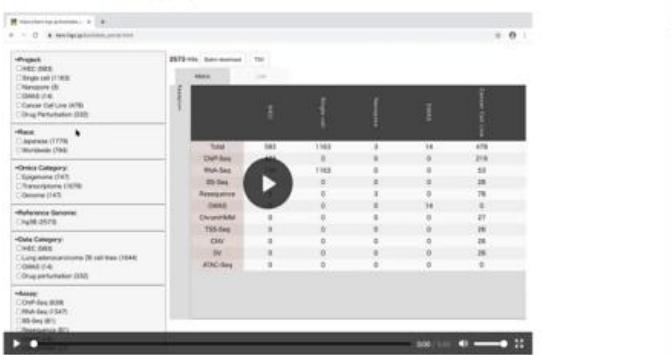
Based on UCSC hg38, mm10

Formerly DBTSS

We recommend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Top | Tutorials | Data portal

DBKEROデータポータルの使い方



何ができるか(3)

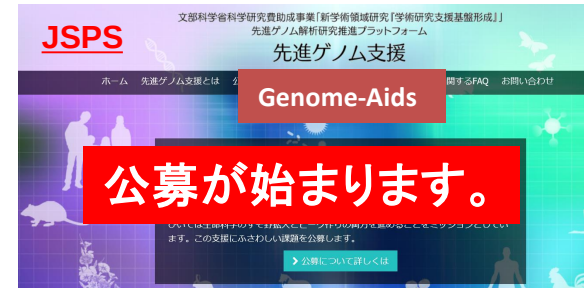
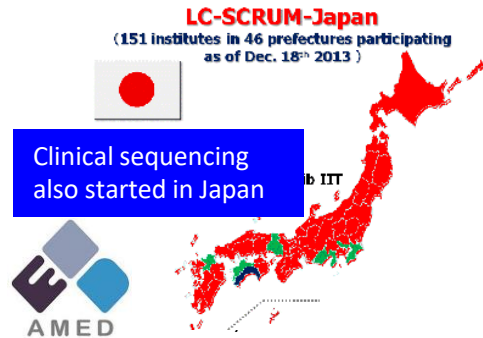
新しい手法でどんなデータが出てくるのか？

シングルセルデータ、空間トランスクリプトームデータ、
ナノポアデータ等がおちています。

東大・柏拠点



On-going contributions as a sequence center



Hiseq2500 x 6+ Hiseq3000 x 1
NovaSeq x 2



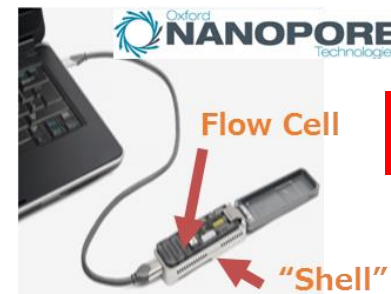
And serving as an Incubation center for new genome technologies)



Single cell analyzer;
C1 (Fluidigm)



Operators:
Technicians 5
Programmers 4

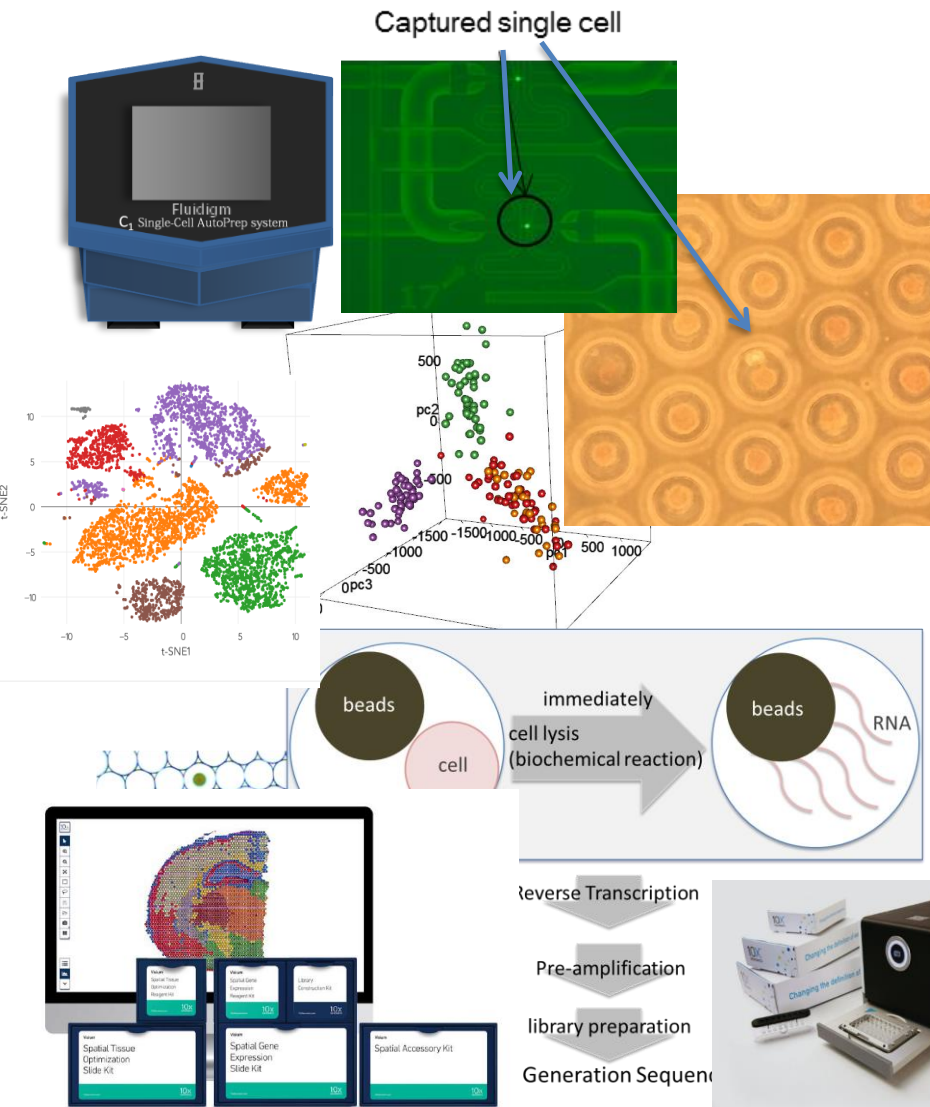


ysuzuki@hgc.jp

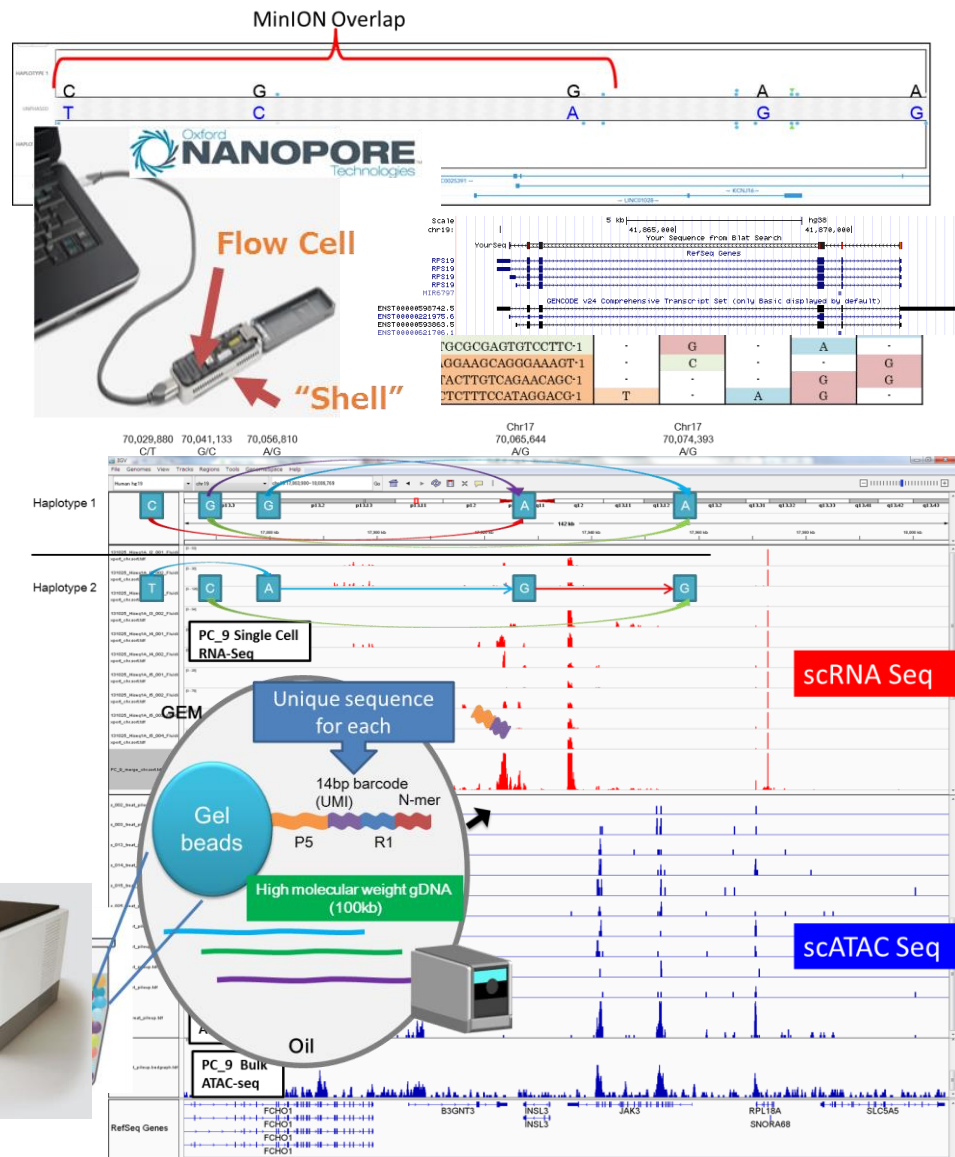
PI: Yutaka Suzuki.
Univ. Tokyo

新しい手法でどんなデータが出てくるの？

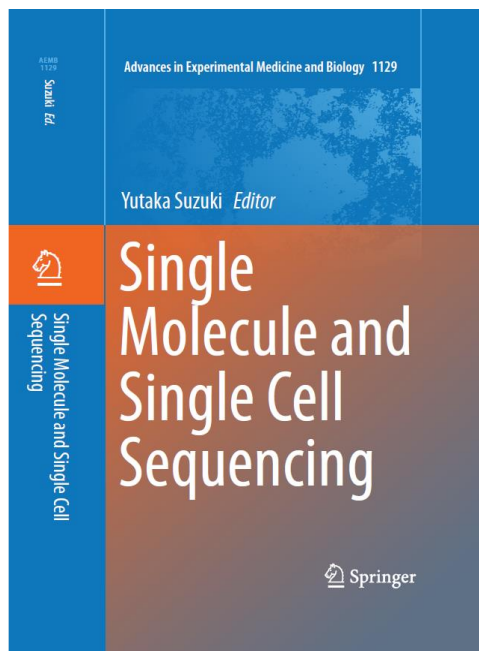
シングルセル解析 (C1, Chromium) + 空間トランスクリプトーム解析



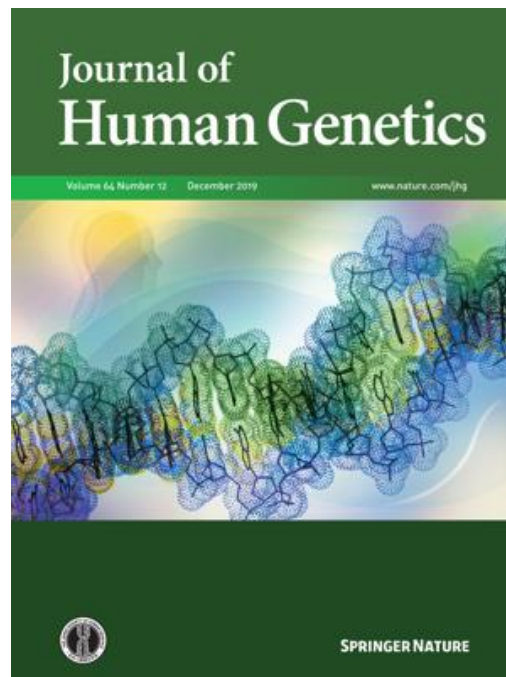
Long read解析 (SV/Phasing)



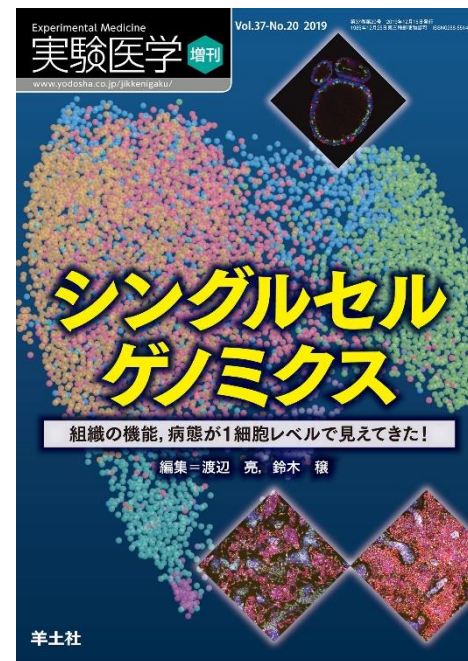
参考文献(シングルセル+ロングリード)



Nature Springer Book
Single Molecule and Single Cell Sequencing.
(Yutaka Suzuki ed.)



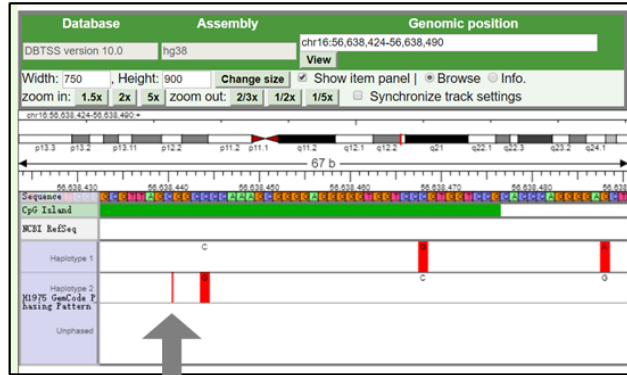
Journal of Human Genetics
Special issue for “New type Sequencer”
Jan (2020)
(Yutaka Suzuki ed.)



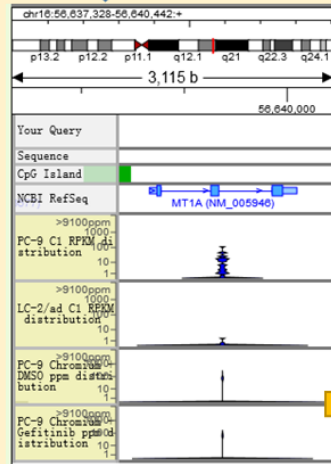
渡辺・鈴木編 (2019)

シングルセルとロングリードのデータ閲覧

MT1A enhancer in H1975
chr16:56,638,440, C>CG



Mutation



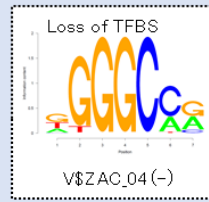
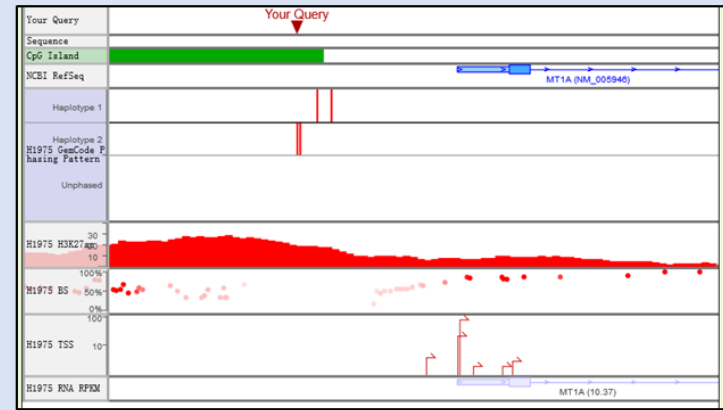
C1

Chromium



Linked read

H3K27ac
DNA met
TSS-seq
RNA-seq



WGS

C>CG = 18:16

C>CG = 0.5

ChIP-seq
H3K27ac

Regulatory mutation

WGS

A:C = 9:19

A:C = 92:5

RNA-seq

A:G = 14:23

A:G = 108:0

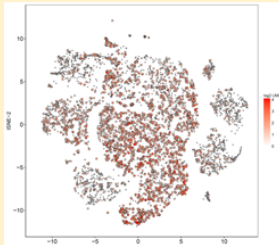
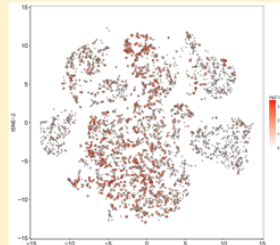
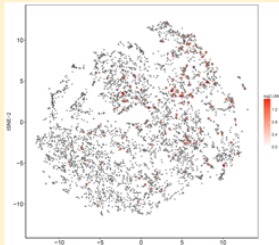
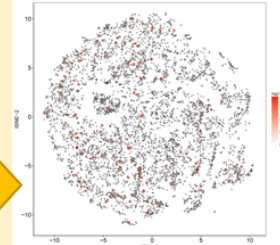
Coding SNPs

PC-9 control

PC-9 gefitinib

H1975 control

H1975 gefitinib



MT1A遺伝子での例. ゲノムブラウザを中心に目的の遺伝子に関連したプロモーター領域や関連のシングルセルのデータを確認することができる.

ブラウザはなくとも、まずはデータだけでも・・・

DBKERO

Database of Korean Genotypes for Researcher of multi-Omic data

Quick-start: For the beginners

シングルセルデータ

シングルセルデータ加工の
Tutorial動画

VISIUMのデータ
(空間トランス
クリプトーム解析

データポータル
(2500程度の多層
オームクスデータセット；
細胞株が多いです)

Longリードデータ
(日本人肺腺がんのSV)

Sakamoto et al
Genome Res 2029

Recommend to use Edge (ver. 40 or above), Google Chrome (ver. 81 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Selected Dataset

- Cell dataset [Download processed data]:
- Cell dataset

Download

- Data Portal [Help video (Japanese)]

Documents

- Experimental Procedures
- Data Contents
- Tutorial
- Download bulk data
- References

Movie (For Japanese)

Overview movie 1 in Japanese (9 min.)

Overview movie 2 in Japanese (53 min.)

How to use genome browser (Basic)

How to use genome browser (Track selection)

How to use genome browser (Comparative viewer)

How to use data portal

Mutation Enriched Genes [Help video (Japanese)]:
Find out which gene is mutated most in our data.

TF Binding Site Search [Help video (Japanese)]:
Find out TF binding sites in a region of genome of various organism.
Powered by ChIP-Atlas* and GGGGenome*.

Pathway Map [Help video (Japanese)]:
Find out the level of expression or modification of genes within a Pathway

Chromatin-status Data Summary [Help video (Japanese)]:
Get an overall view of expression and modification of a genome region among our entire data sets.

- Search from Genomic Position
- Search from SNP (dbSNP rsID)
- Search from SNV (COSMIC: somatic mutation)

SNV Summary in Cancers [Help video (Japanese)]:
Find out mutation frequency of a gene among our cancer data sets.

RDF gate (Trial)
Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq

- RDF Schema
- RDF Browser
- SPARQL Endpoint

Other tools

- Single cell viewer [Help video (Japanese)]

DBKEROデータポータルの構築

DBKERO Data Portal

▼Project:

- ☐ IHEC (583)
- ☐ Single cell (1163)
- ☐ Nanopore (3)
- ☐ Cancer Cell Line (478)

▼Race:

- ☐ Japanese (1677)
- ☐ Worldwide (550)

▼Omics Category:

- ☐ Epigenome (747)
- ☐ Transcriptome (1347)
- ☐ Genome (133)

▼Reference Genome:

- ☐ hg38 (2227)

▼Data Category:

- ☐ IHEC (583)
- ☐ Lung adenocarcinoma (1644)

▼Assay:

- ☐ ChIP-Seq (639)
- ☐ RNA-Seq (1321)
- ☐ BS-Seq (81)
- ☐ Resequencing (81)
- ☐ ChromHMM (27)
- ☐ TSS-Seq (26)
- ☐ CNV (26)
- ☐ SV (26)

▼File Format:

- ☐ BigWig (read depth) (854)
- ☐ BigWig (read depth) x2 (14)
- ☐ TSV (BS) (68)
- ☐ BigWig (BS) (13)
- ☐ TSV (exon/intron junction) (27)
- ☐ TSV (expression level) (27)
- ☐ TSV (expression distribution) (27)
- ☐ BAM (481)
- ☐ TSV (ChromHMM) (27)
- ☐ TSV (TSS) (26)
- ☐ TSV (SNP/SNV) (26)
- ☐ TSV (read depth) (216)
- ☐ TSV (CNV) (26)
- ☐ TSV (translocation) (26)

データカテゴリ選択

Matrix List

Project>>>

Assay>>>

	IHEC	Single cell	Nanopore	Cancer Cell Line
Total	583	1163	3	478
ChIP-Seq	423	0	0	216
RNA-Seq	105	0	0	53
BS-Seq	55	0	0	26
Resequencing	0	3	78	0
ChromHMM	0	0	27	0
TSS	2227 Hits			

Batch download TSV

マトリクス画面

バッチダウンロード

各データダウンロード

リスト画面

むしろ手元でブラウザしてもらった方がいいような・・・

シングルセル解析入門

本編ではシングルセル解析初心者の方がテストデータを用いて、一通りのデータ解析が行えるようになることを目的としています。

ここでは、トランスクリプトーム解析としてscRNA-seq、エピゲノム解析としてscATAC-seq解析の方法を説明します。

リンク：シングルセル解析の必要性

リンク：シングルセル解析でできること・問題点

リンク：シングルセル解析の様々なプラットフォーム

ここでは、10x Genomics社のChromiumプラットフォームを用いた配列データの解析方法について説明します。

Cell Rangerによる処理は多くのコンピュータ資源を必要としますので、高性能計算機の使用をお勧めします。なおLoupeによる結果の閲覧は、通常のPCで多くのこの文書にある処理は東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータ(HGC)で動作を確認しています。HGCをご利用になる場合は、以下

```
$ qlogin -l s_memo=100,mem=seq=100
```

Rをご使用する場合は、R3.6以上のバージョンが必要で、HGCをご利用の場合は、以下のように使用するRを指定してください。

```
$ module load R/3.6
```

リンク：シングルセル解析の様々なプラットフォーム

Chromiumを使ったscRNA-seqデータ解析

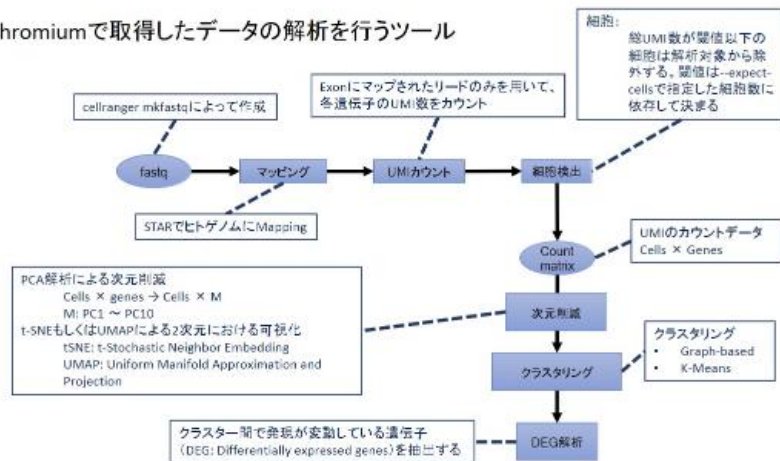
ここではマウス脳正常細胞のscRNA-seqデータと、解析ツールとしては10x Genomics社のCell Rangerを用いて解析を行います。

以下の操作について説明します。

1. データの準備・ソフトウェアの確認
2. Cell Rangerによる解析
3. Cell Ranger実行結果の確認
4. Loupe Cell Browserを用いたデータの確認

Cell Rangerとは

Chromiumで取得したデータの解析を行うツール



詳しくは以下 (英文)

<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>

データの準備・ソフトウェアの確認

10x Genomics社のダウンロードページより、最新版のCell Ranger(4.0.0)とreferenceデータセット(mouse mm10(2020/09/23執筆時)をダウンロードします。
※メールアドレスの登録が必要となります。

```
mkdir -p /data/10x
cd /data/10x
wget https://ker0.hgc.jp/tutorials/learning/data/10X_RNAv3_Lung.tar
tar xvf 10X_RNAv3_Lung.tar
```

テストデータの準備

今回はテストデータセットとして鈴木研究室でChromiumプラットフォームを用いて準備したマウス脳細胞scRNA-seqの配列データ(FASTQ)で解析を行い、データセットはhttps://ker0.hgc.jp/tutorials/learning/data/10X_RNAv3_Lung.tar からダウンロードできます。

```
$ mkdir -p /data/10x
cd /data/10x
wget https://ker0.hgc.jp/tutorials/learning/data/10X_RNAv3_Lung.tar
tar xvf 10X_RNAv3_Lung.tar
```

Cell Rangerによる解析

下記コマンドでCell Rangerを実行することが可能です。先ほど用意したfastqファイルを解析してみましょう。
※今回の解析では--expect-cells=10000 となります。

```
$ mkdir -p /data/10x
cd /data/10x
cellranger count --id=mm10 --transcriptomes=/data/refdata-gex-mm10-2020-04 --fastqs=/data/10X_RNAv3_Lung/ --expect-cells=10000
```

Cell Ranger 実行結果の確認

Cell Ranger の処理が完了したら、出力ファイルを確認しましょう。

```
$ cd /data/analysis/scRNA/001_rnaseq/
ls
SC_RNA_COUNTER_CS _filelist _limcation _log _perf _tags _uuid _versions _outs
_logline _finalstate _jobnode _mresource _sitecheck _timestamp _vckill 001_rnaseq_mri.tgz
$ ls
analysis filtered_feature_bo_matrix.h5 possorted_genome_bam.bam raw_feature_bo_matrix.h5
loupe.loupe metrics_summary.csv possorted_genome_bam.bai web_summary.html
filtered_feature_bo_matrix molecule_info.h5 raw_feature_bo_matrix
```

以下のような結果ファイルが出力されます。

File Name	Description
web_summary.html	HTML形式のサマリー。
metrics_summary.csv	CSV形式のサマリー。
possorted_genome_bam.bam	BAMファイル。IGVで閲覧可能。
possorted_genome_bam.bai	BAIファイル。
filtered_feature_bo_matrix.h5	Feature barcode matrix (H5)。
molecule_info.h5	Molecule information (H5)。
raw_feature_bo_matrix.h5	Raw feature barcode matrix (H5)。

Quick-start: For the beginners

recommend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer.

Mouse Visium Dataset

We prepared mouse kidney spatial gene expression dataset (10x Genomics visium system).
Sequenced by Novaseq 6000.

You can download the dataset from the link below.

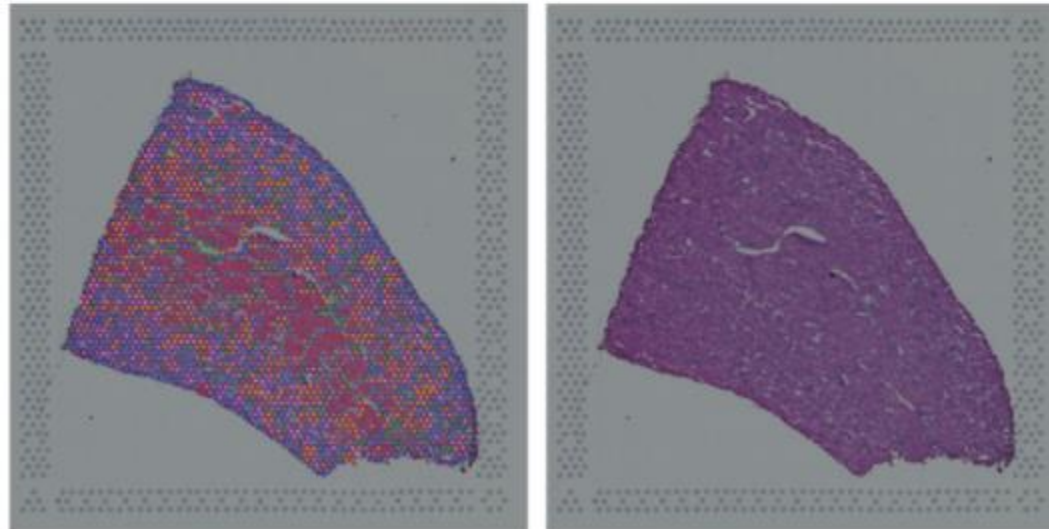
These datasets are processed by spaceranger v1.0.0 (refdata-cellranger-mm10-3.0.0).

- [Summary.html \(by spaceranger\)](#)
- [Dataset for loupe* visualization](#)
- [Dataset for Seurat analysis](#)

*loupe is visualization software released by 10x genomics.
please download from [10x genomics web site](#) (Registration required).

現在はマウスですが、ヒトも出します

(倫理もあるので) count matrix



Quick-start: For the beginners

We recommend to use Edge (ver. 40 or above), Google Chrome (ver. 51 or above) or Firefox (ver. 55 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Cancer SV dataset

Here, we report the application of a long-read sequencer, PromethION, for analyzing human cancer genomes. We first conducted whole-genome sequencing on lung cancer cell lines. We found that it is possible to genotype known cancerous mutations, such as point mutations. We also found that long-read sequencing is particularly useful for precisely identifying and characterizing structural aberrations, such as large deletions, gene fusions, and other chromosomal rearrangements. In addition, we identified several medium-sized structural aberrations consisting of complex combinations of local duplications, inversions, and microdeletions. These complex mutations occurred even in key cancer-related genes, such as STK11, NF1, SMARCA4, and PTEN. The biological relevance of these mutations was further revealed by epigenome, transcriptome, and protein analyses of the affected signaling pathways. Such structural aberrations were also found in clinical lung adenocarcinoma specimens. Those structural aberrations were unlikely to be reliably detected by conventional short-read sequencing. Therefore, long-read sequencing may contribute to understanding the molecular etiology of patients for whom causative cancerous mutations remain unknown and therapeutic strategies are elusive.

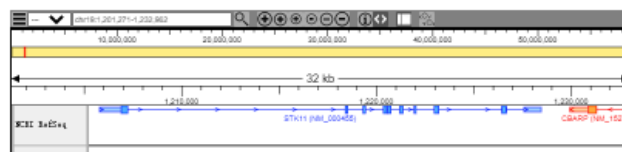
Reference: [Long-read sequencing for non-small-cell lung cancer genomes](#)

Accession: [JGAS000000000065](#) / [JGAD000000000252](#) / [JGAD000000000253](#)* Controlled Access.

sampleID	Tumor			Normal			Number of SVs	Processed data
	Yields(Gb)	Number of reads	Coverage(x)	Yields(Gb)	Number of reads	Coverage(x)		
S1	99	14861240	33	57	11709609	19	7	S1 SV gene candidates.bedpe
S2	94	25379061	31	41	7016159	14	120	S2 SV gene candidates.bedpe
S3	77	15952312	25	48	12971173	15	41	S3 SV gene candidates.bedpe
S5	82	21640571	27	35	6503970	11	11	S5 SV gene candidates.bedpe
S6	76	1136329	25	48	7531078	16	72	S6 SV gene candidates.bedpe
S7	85	13880384	28	46	7555316	15	91	S7 SV gene candidates.bedpe
S8	100	18661003	33	54	8351411	18	67	S8 SV gene candidates.bedpe
S9	96	13179209	32	34	4483001	11	48	S9 SV gene candidates.bedpe
S10	85	12424087	28	42	7029224	14	64	S10 SV gene candidates.bedpe
S11	89	21140772	22	59	7832201	19	42	S11 SV gene candidates.bedpe
S12	73	15634853	24	84	12908975	27	2	S12 SV gene candidates.bedpe
S13	104	30719862	34	38	5506152	12	31	S13 SV gene candidates.bedpe
S14	74	14329017	24	56	6464828	18	53	S14 SV gene candidates.bedpe
S15	75	12278084	24	55	9585699	18	7	S15 SV gene candidates.bedpe
S16	74	20118226	24	55	7924757	18	103	S16 SV gene candidates.bedpe
S17	52	7086616	17	59	6766396	20	3	S17 SV gene candidates.bedpe
S18	62	7091142	20	44	5335320	15	14	S18 SV gene candidates.bedpe
S19	60	6352847	20	37	5025576	12	10	S19 SV gene candidates.bedpe
S20	58	5836744	19	47	5620346	15	87	S20 SV gene candidates.bedpe
S21	63	8685953	21	57	6166961	19	19	S21 SV gene candidates.bedpe

Viewer (Cell line data)

REF-LC-KJ: STK11



日本人肺腺がん

ひとまず20人分

(倫理もあるので)SVのvcf graph

複雑なSVとブラウザ

Sakamoto et al Genome Res 2020

Human Genome Center 講習会 「DBTSS/DBKERO 講習会・シングルセル解析入門」

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ミクスデータベースDBTSS/DBKEROの利用説明会と、DBKEROデータベース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SHIROKAWAで提供しています。

- 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)、聴講のみの場合は特に定員を設けません
- 対象: 多層オミクスデータの活用・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. 最近のオミクスデータ産生のトピック」(鈴木穰)

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」(鈴木絢子)

-- 休憩 --

14:10 ~ 14:30

「3. データベース DBTSS/DBKERO 講習会」(鹿島幸恵・若栗浩幸)

14:30 ~ 17:00

「4. DBKERO シングルセル解析講習会」(鹿島幸恵・若栗浩幸)

申込み方法

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: 鈴木穰・鈴木絢子 (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.cbms.k.u-tokyo.ac.jp/>) セミナーシリーズの一部としても行われます。

HGC講習会(医科学研究所ヒトゲノム解析センター主催) 「DBTSS/DBKERO講習会・シングルセル解析入門」

- 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
- 申し込みURL
<https://www.at.hgc.jp/event-details/dbtss-dbkero-koshukai-shinguruserukaisekinyumon-2>

--DBKERO(多層オミクス解析DB)の利用説明

--シングルセルデータの加工基礎についての
ハンズオン講習

--アカウント付き参加登録(30名)はSold out

--聴講のみ(70名)は若干空席あり

Visiumデータプロセス講習会

新型コロナウイルスの影響で、延期となっていました。
年明け開催予定で調整中(オンライン)

先進ゲノム支援(文科省新学術領域)

<http://www.genome-sci.jp/>

Google “ゲノム支援”

ysuzuki@hgc.jp

基本、無償で。
(上限はありますが・・・)

「先進ゲノム支援」課題募集

1/8～1/28
(予定)



最終年度になります。

第1回公募は2021年1月開始



新学術領域研究「先進ゲノム解析研究推進プラットフォーム」(略称「先進ゲノム支援」)

2021年度
支援対象課題公募

- 第1回 / 1月(4月支援開始): 対象は2021年度科研費継続課題
- 第2回 / 5月(8月支援開始): 対象は2021年度科研費課題(新規採択を含む)

詳細は <https://www.genome-sci.jp/>

先進ゲノム支援事務局 大学共同利用機関法人 情報・システム研究機構 国立遺伝学研究所

国内総説を用いた広報活動

遺伝子医学34号(2020)



生体の科学(2020)



その他出版中

遺伝子医学35号(2021)

和光純薬時報(2021)

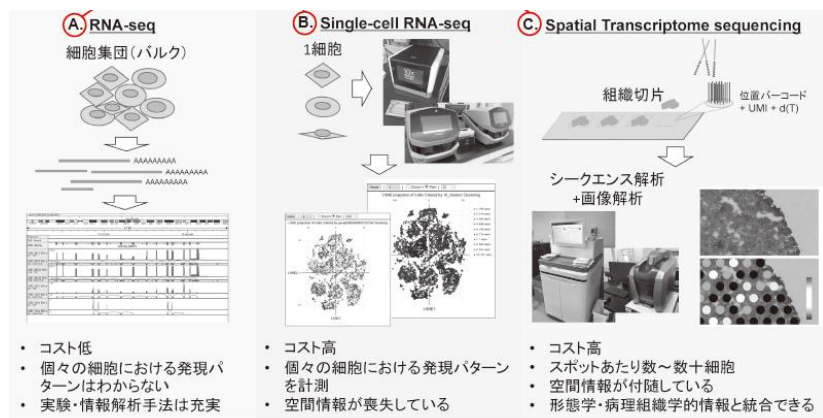


図1 RNA-seq, scRNA-seq, および Spatial Transcriptome sequencing
RNA-seq, single-cell RNA-seq (scRNA-seq) および空間トランスクリプトーム解析 (spatial transcriptome sequencing) について特徴をまとめている。図中のゲノムブラウザおよび t-SNE プロットは 疾患オミクスデータベース DBKERO (<https://kero.hgc.jp/>) に掲載され

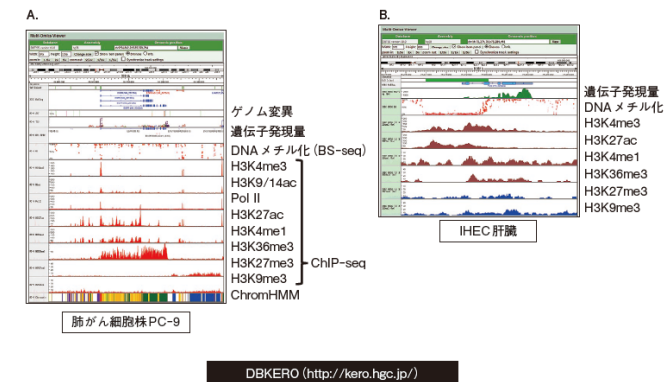


図2 DBKERO におけるがん細胞および正常組織のエピゲノムおよびマルチオミクスステータス
ヒト統合マルチオミクスデータベース DBKERO (<http://kero.hgc.jp/>) に収載されている多層オミクスデータ。肺がん細胞株 PC-9 における EGFR 遺伝子領域と、その周辺のオミクス状態を表示した (A)。また、IHEC のエピゲノムデータ (肝細胞) についても FOS 遺伝子領域について同様に表示した (B)。

http://kero.hgc.jp

DBKERO

日本人正常オーミクスデータ (IHEC)

がん培養細胞オーミクスデータ

シングルセル・ナノポアデータ

初めての方は、**“For the first users”**

mmend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any

Featured Dataset

[Single cell dataset \[Download processed data\]:](#)
Single cell dataset

[Cancer SV dataset ^{New}](#)
Japanese lung adenocarcinoma sequenced by PromethION(Oxford Nanopore Technologies).

[Visium dataset ^{New}](#)
Spatial Gene Expression dataset (10x Genomics Visium system).
Spaceranger output files for Ioupe browser and Seurat analysis.

Documents

- [Experimental Procedures](#)
- [Data Contents](#)
- [Tutorial](#)
- [Download bulk data](#)
- [References](#)

Movie (For Japanese)

Overview movie 1 in Japanese (9 min.)



Overview movie 2 in Japanese (53 min.)



How to use genome browser (Basic)



How to use genome browser (Track selection)



How to use genome browser (Comparative viewer)



How to use data portal



Download

- [Data Portal \[Help video \(Japanese\)\]](#)

Tools

[TSS-Viewer \[Help video \(Japanese\)\]:](#)
Find out transcriptional start sites and compare promoter usage.

[Multi-Omics-Viewer \[Help video 1,2,3 \(Japanese\)\], \[GitHub\]:](#)
Browse mutations, transcripts and epigenetic modifications along the genome coordinates.

[Mutation Enriched Genes \[Help video \(Japanese\)\]:](#)
Find out which gene is mutated most in our data.

[TF Binding Site Search \[Help video \(Japanese\)\]:](#)
Find out TF binding sites in a region of genome of various organism.
Powered by [ChIP-Atlas*](#) and [GCGenome*](#).

[Pathway Map \[Help video \(Japanese\)\]:](#)
Find out the level of expression or modification of genes within a Pathway

[Chromatin-status Data Summary \[Help video \(Japanese\)\]:](#)
Get an overall view of expression and modification of a gene among our entire data sets.

- [Search from Genomic Position](#)
- [Search from SNP \(dbSNP rsID\)](#)
- [Search from SNV \(COSMIC: somatic mutation\)](#)

[SNV Summary in Cancers \[Help video \(Japanese\)\]:](#)
Find out mutation frequency of a gene among our cancer data sets.

RDF gate (Trial)

Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq

- [RDF Schema](#)
- [RDF Browser](#)
- [SPARQL Endpoint](#)

Other tools

- [Single cell viewer \[Help video \(Japanese\)\]](#)

そもその
データセット例
(最新！)

(ほぼ) 全部のデータが
簡単に取れます。

各データが
ブラウズできます。

ご質問は：
ysuzuki@hgc.jp

とりあえず
Tutorialへ

他にも
・他の統合DBとの連携
・RDF化のスキーマの作成
も進めています。