

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

鈴木 穰

東京大学/新領域・メディカル情報生命専攻
/生命データサイエンスセンター
/東京大学統合ゲノム医科学機構

DBKERO : <http://dbkero.hgc.jp/>

DBKERO
Database of Korean Genotypes for Researches of multi-Omics data

Quick-start: For the beginners

mend to use Edge (ver. 40 or above), Google Chrome (ver. 61 or above) or Firefox (ver. 55 or above) for the DBKERO browsing. We do not support Internet Explorer any more.

Viewer [How to use]

- Multi-Omics-Viewer** [Help video 1, 2, 3 (Japanese)], [GitHub]: Browse mutations, transcripts and epigenetic modifications along the genome coordinates.
- TSS-Viewer** [Help video (Japanese)]: Find out transcriptional start sites and compare promoter usage.
- 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.

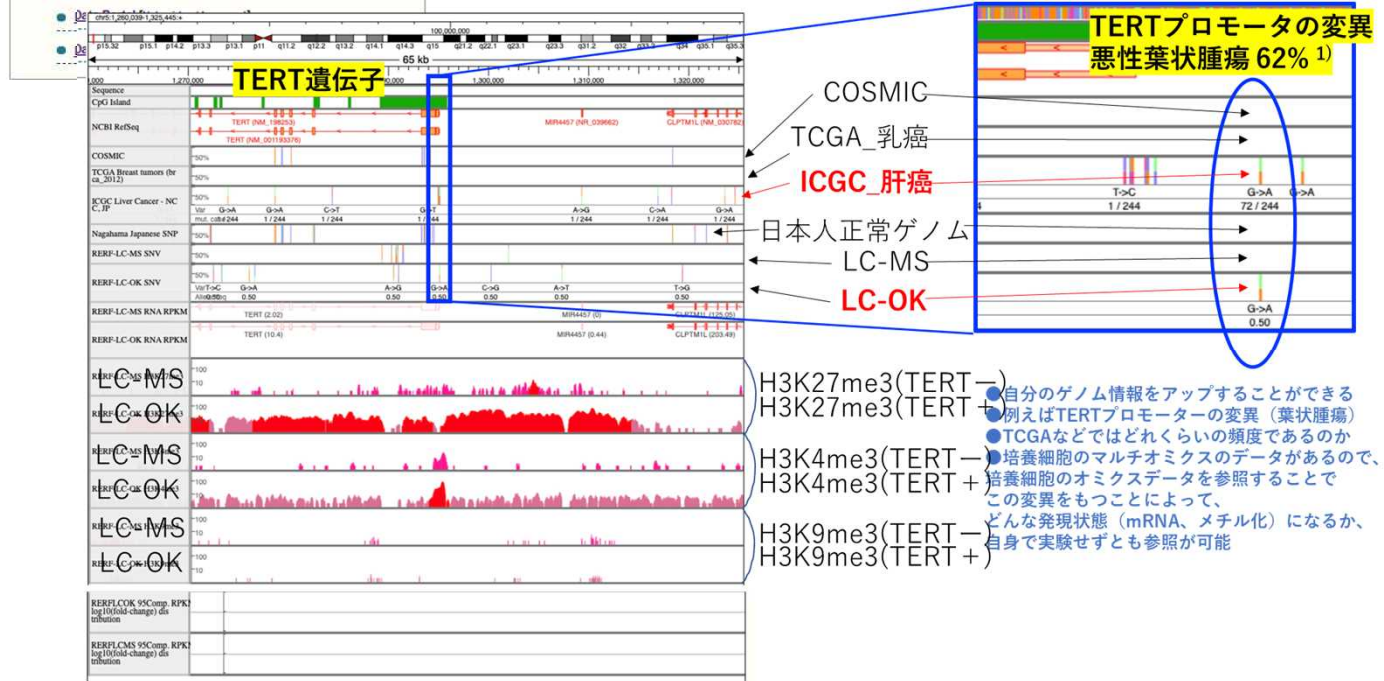
● [Search from Genomic Position](#)

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 loupe browser and Seurat analysis.
- Cancer LongRNA Dataset** ^{New}: Aberrant splicing isoforms in non-small cell lung cancer sequenced by MinION (Oxford Nanopore Technologies).

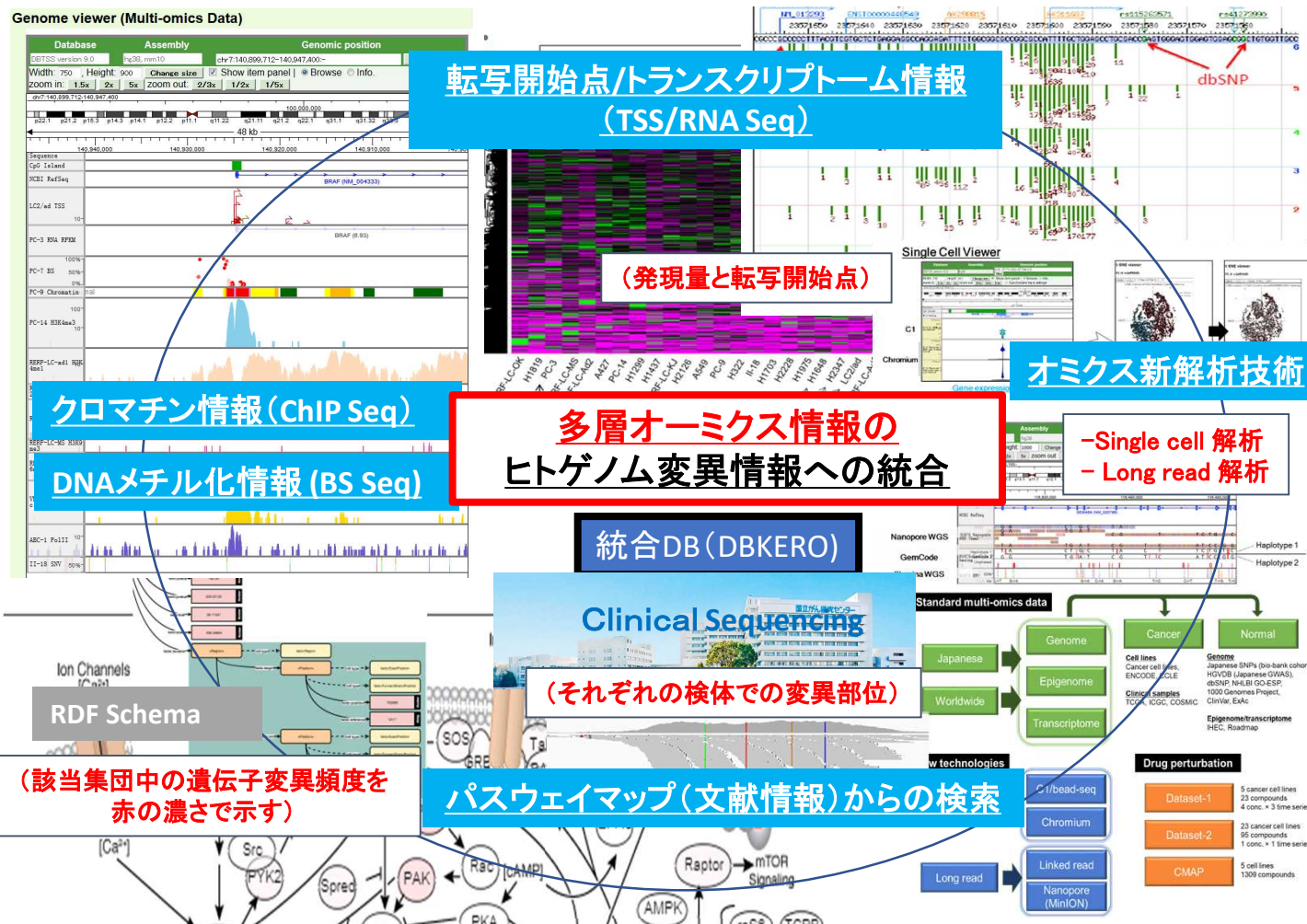
Download

日本人多層オームクス 日本人由来細胞株多層オームクス 新規技術データ



全体のねらい

ヒト応用研究を志向したオミクス情報の統合



何ができるか(1)

日本人正常オーミクスデータ (IHEC)

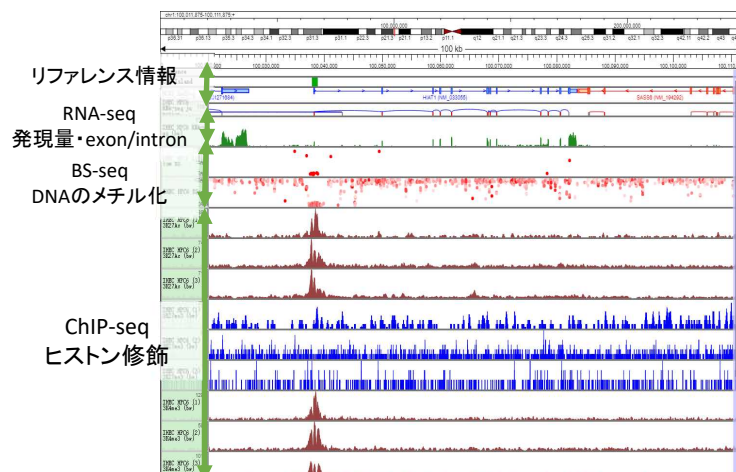
国際エピゲノムコンソシアム (IHEC)

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。

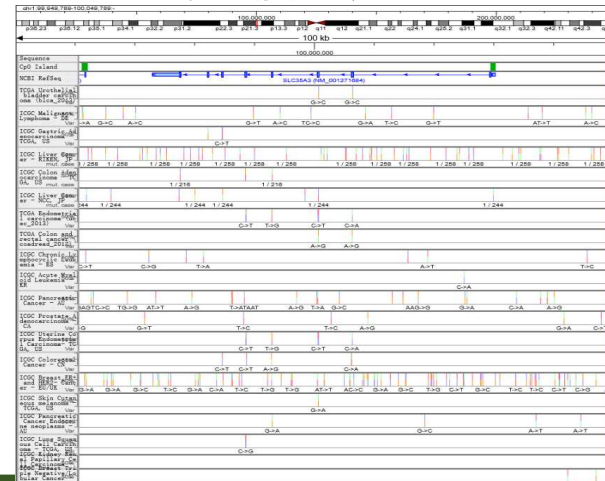
CREST-IHECデータの統合(日本人標準エピゲノム)

IHEC (国際エピゲノムコンソシアム) データの公開

マルチオーミクス情報の表示



TCGA/ICGC変異情報の表



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

日本人ゲノム多型

NBDC ヒトDBとの 連携へ

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

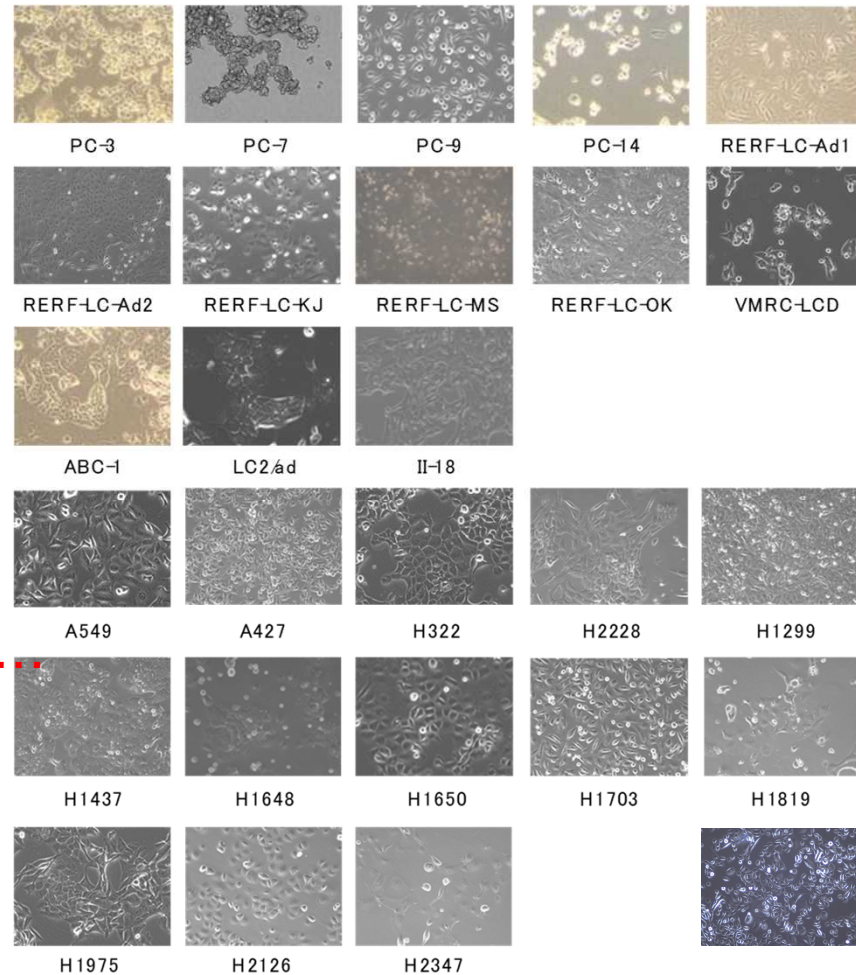
肺腺癌細胞株等で日本人由来のオーミクスデータが充実しています。

培養細胞オーミクス

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	
LC2/ad	
II-18	
A427	Non-Japanese
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...



Suzuki et al. 2014 *Nucleic Acids Research*
Suzuki et al. 2015 *Nucleic Acids Research DB*
issue

26 NSCLC cell lines

13 Japanese

13 non-Asian

+ normal (SAEC)

Genome

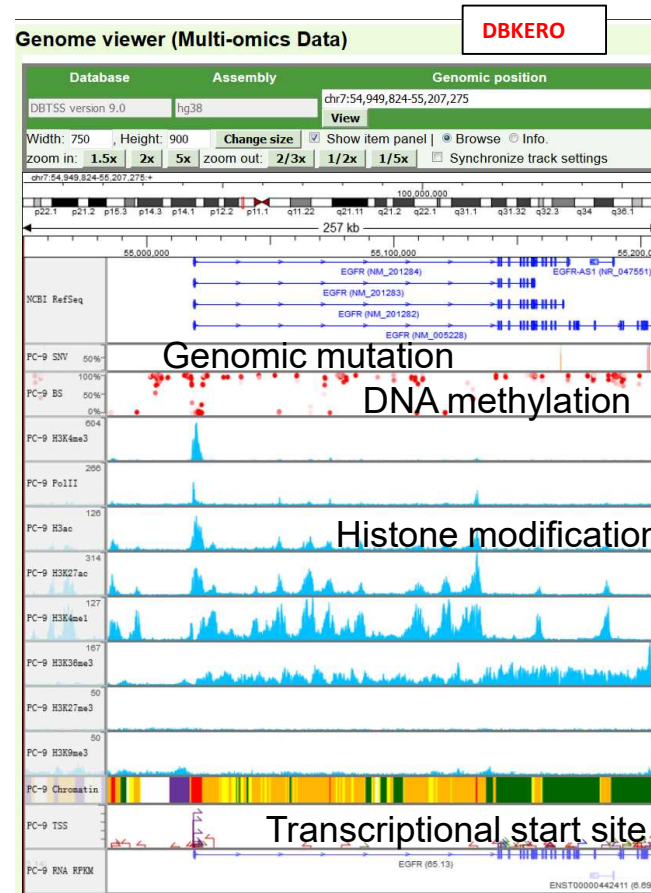
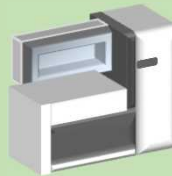
- Whole-genome sequencing
- Long-read sequencing (MinION/PromethION)

Epigenome

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

Transcriptome

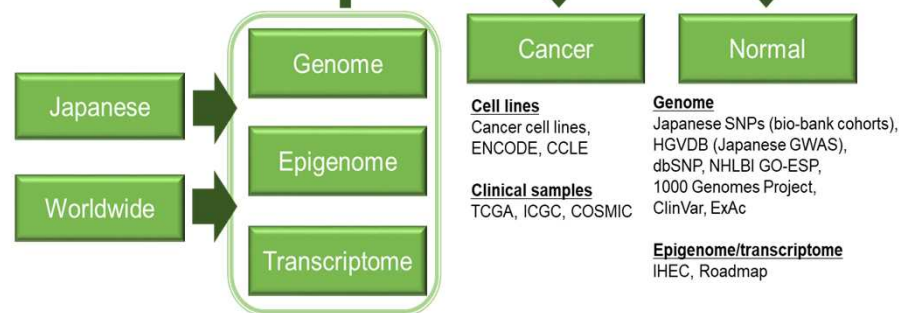
- TSS-Seq
- RNA-Seq
- miRNA-seq



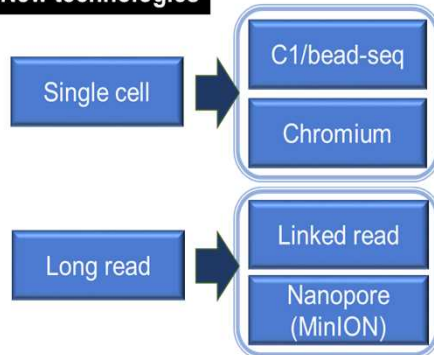
データコンテンツの選択と表示

KEROデータコンテンツ

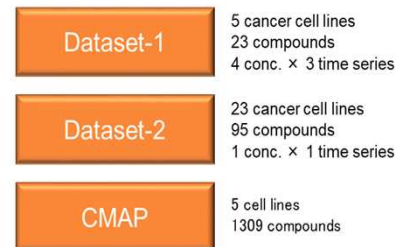
Standard multi-omics data



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.

DBKERO
Database of Genomic Encyclopedia for Researcher of multi-Omics data

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](#) | [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 site 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 in 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

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

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](#) | [Tutorial](#) | [Multi-Omics-Viewer \(Basic\)](#)

マルチオミクスビューワの使い方 (基本)

転写因子結合領域予測の使い方
Quick-start: For the beginners

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](#) | [Tutorial](#) | [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

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](#) | [Tutorial](#) | [Data portal](#)

DBKEROデータポータルの使い方

Category	Cell	RNA-seq	ChIP-seq	BS-seq	TSS-seq
Human	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes
Mouse	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes
Human	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes
Mouse	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes	1000 Genomes

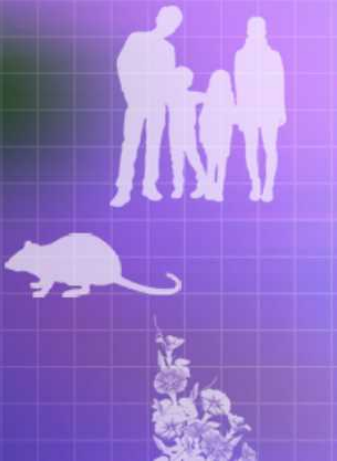
先進ゲノム支援(文科省新学術領域):
"PAGS" = Platform for Advanced Genome Sciences

<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課題の解決が期待されます。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分野で応用が期待されます。この支援にふさわしい研究を支援します。

② 支援採択課題の元科研費

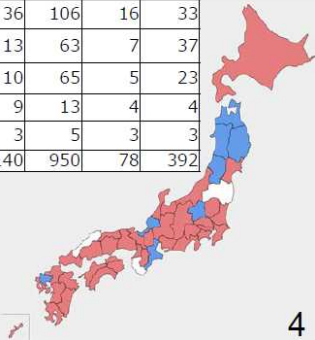


③ 支援依頼者所在機関 (H26まで)

	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

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



DBKEROデータポータル

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 adenocarcinomas (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) (127)
- TSV (expression level) (127)
- TSV (expression distribution) (127)
- BAM (481)
- TSV (ChromHMM) (26)
- 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>>>

Matrix

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	0	78
ChromHMM	0	0	0	27

マトリクス画面

バッチダウンロード

Batch download

TSV

Matrix

List

Object ID	Object Description	Reference Genome	Data Category	Assay	File Format	Download
11_H3K27Ac	IHEC COC11 H3K27Ac (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download
11_H3K27m	IHEC COC11 H3K27me3 (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download
11_H3K36m	IHEC COC11 H3K36me3 (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download
11_H3K4me	IHEC COC11 H3K4me1 (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download
11_H3K4me	IHEC COC11 H3K4me3 (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download
11_H3K9me	IHEC COC11 H3K9me3 (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download
11_mseqq_b	IHEC COC11 RNA-seq (bw)	hg38	IHEC	RNA-Seq	BigWig (read depth)	Download
14_H3K27Ac	IHEC COC14 H3K27Ac (bw)	hg38	IHEC	ChIP-Seq	BigWig (read depth)	Download

各データダウンロード

リスト画面

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。

ブラウザはなくとも、まずはデータだけでも・・・

DBKERO
Database of Genomic Knowledge for Research 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程度の多層オミクスデータセット; 細胞株が多いです)

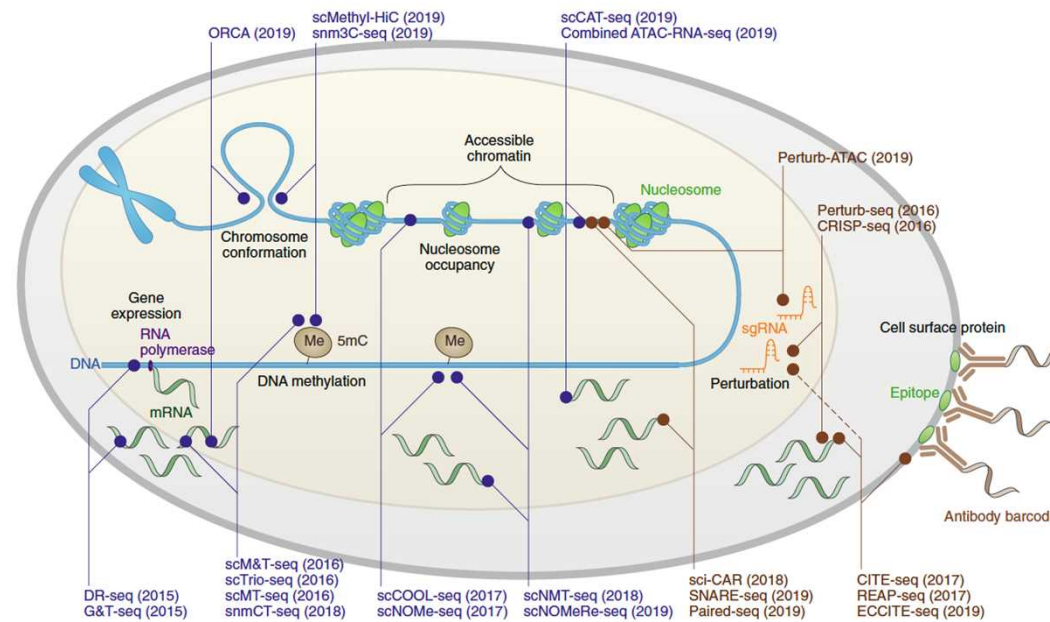
Longリードデータ (日本人肺腺がんのSV)

The screenshot shows the DBKERO website with several key sections and annotations:

- Download**: A blue arrow points to the "Data Portal [Help video (Japanese)]" link.
- Documents**: A blue arrow points to the "Tutorial" link.
- Movie (For Japanese)**: A blue arrow points to the "Overview movie 1 in Japanese (9 min.)" video thumbnail.
- Longリードデータ**: A blue arrow points to the "Mutation Enriched Genes [Help video (Japanese)]" section.

The website interface includes sections for "Download Dataset", "Download", "Documents", "Movie (For Japanese)", and "Other tools". The "Download" section lists various data types and links to help videos. The "Documents" section includes links to "Experimental Procedures", "Data Contents", "Tutorial", "Download bulk data", and "References". The "Movie" section features two overview movies in Japanese. The "Other tools" section includes links to "Single cell viewer [Help video (Japanese)]".

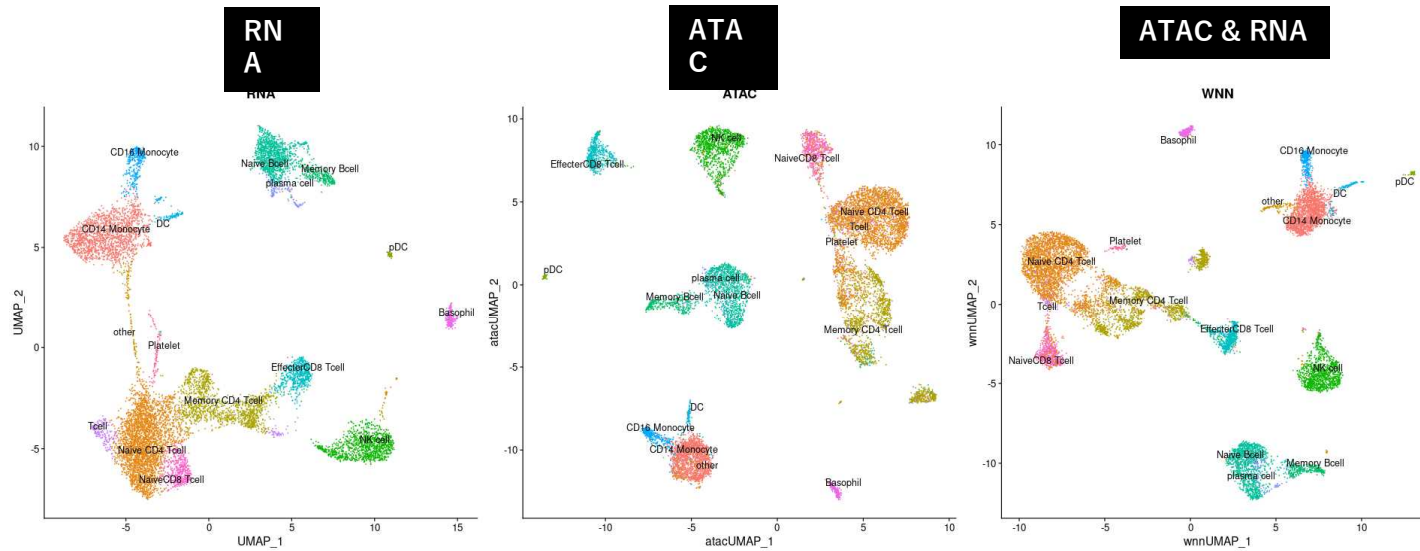
Nature Methods;Method of the Year 2019: Single-cell multimodal omics



Zhu C, Preissl S, Ren B. **Nat Methods**. 17:11-14. 2020

scMulti-Ome

UMAP Analysis



“柏の葉住人コホート”のFeasibility Study

三井不動産との共同研究

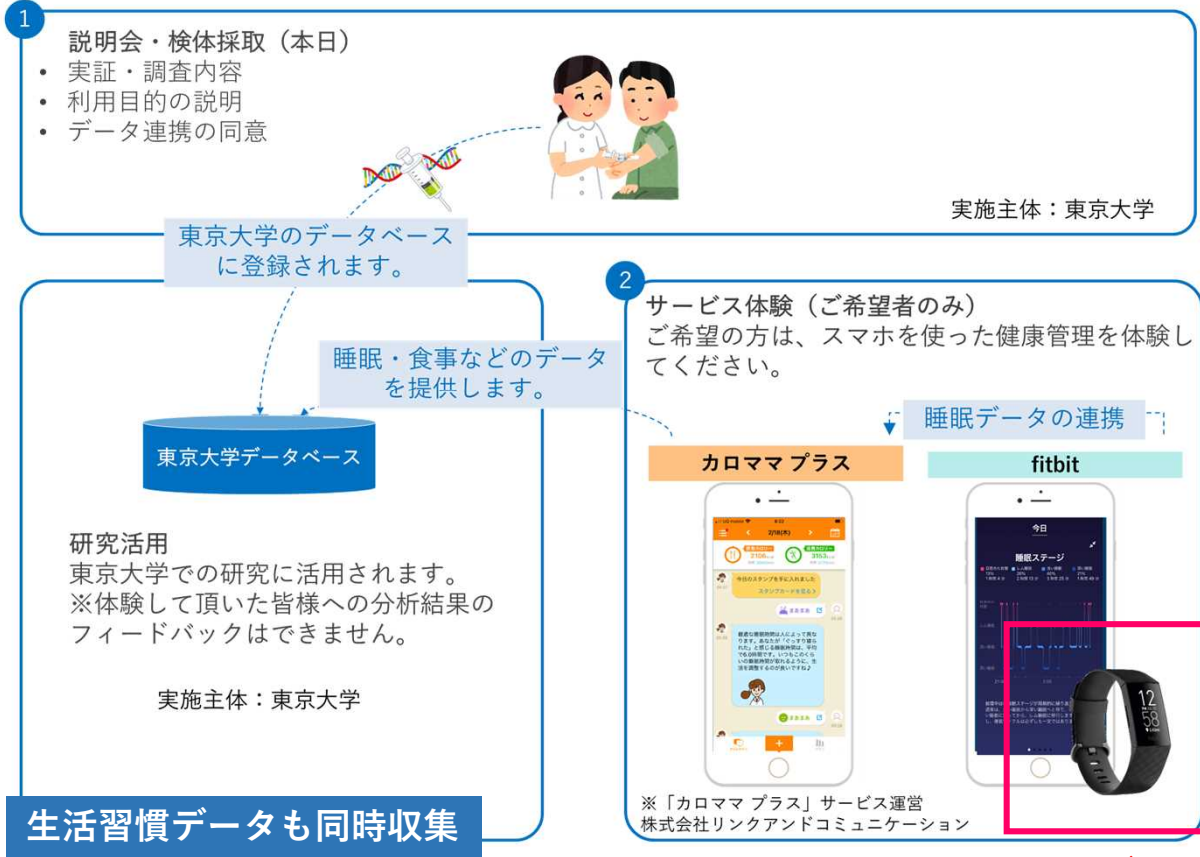


50人の健常者の採血
(ゲノム+scRNA)

<50歳, 50歳台, 60歳台, 70歳台, >80歳
をそれぞれ10人づつ



全体の流れ



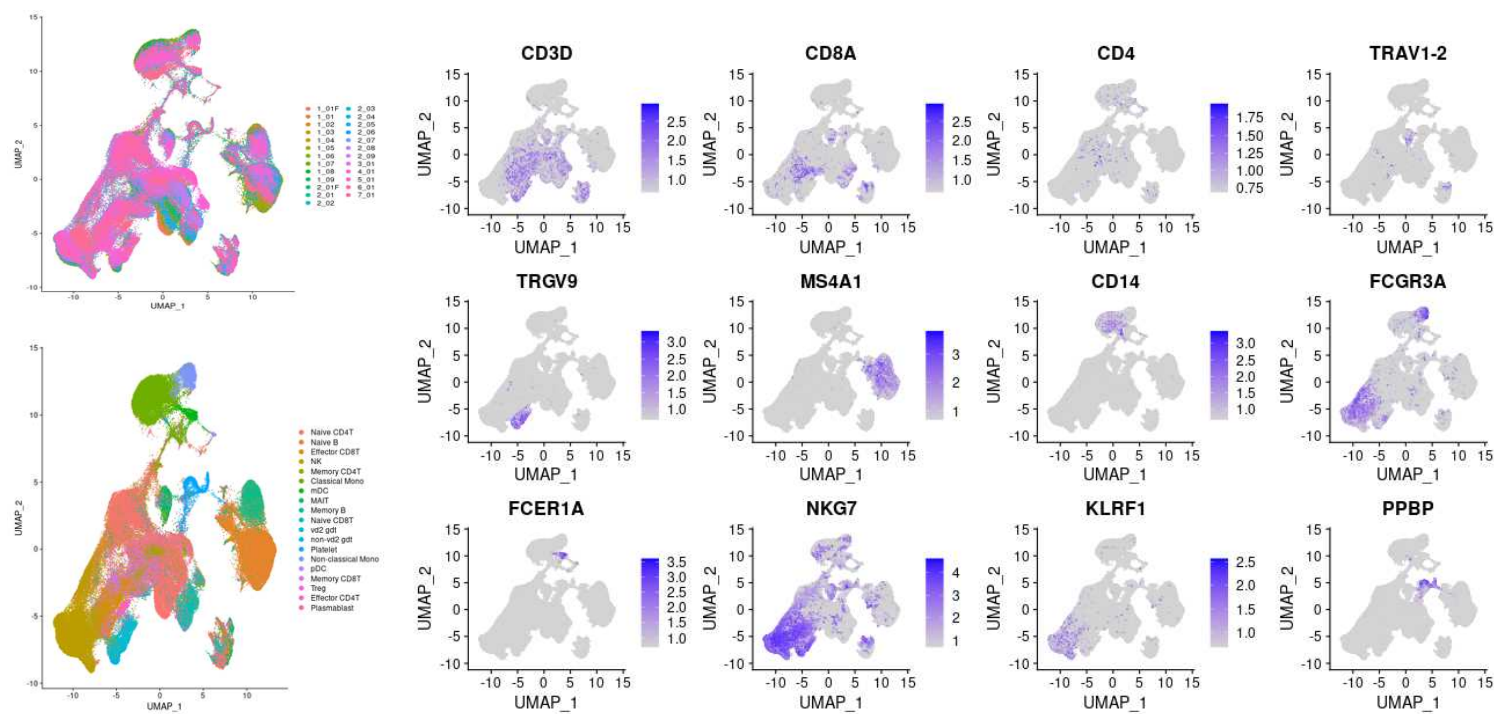
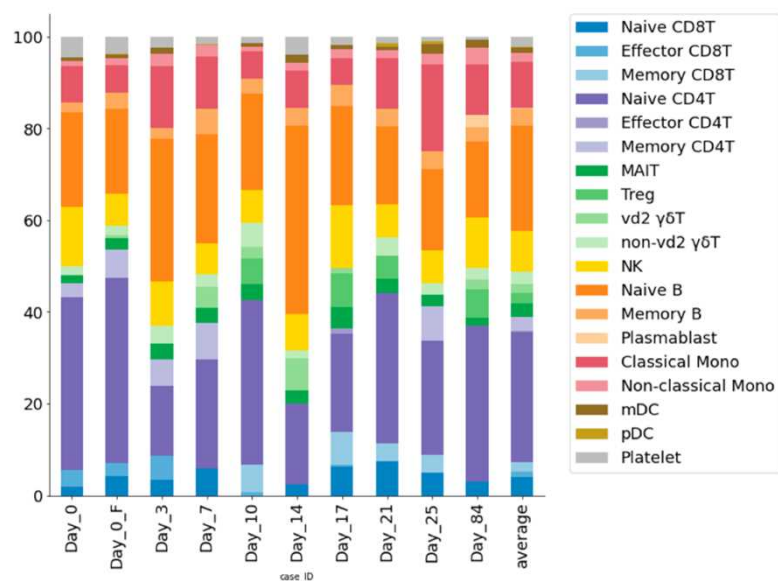


図1：健康者末梢血におけるscRNA seqプロファイル
 左：バッチ効果の除去により相互比較が可能になったUMAP図
 右：それぞれの細胞アノテーションに用いたマーカー

健常者 1



健常者 2

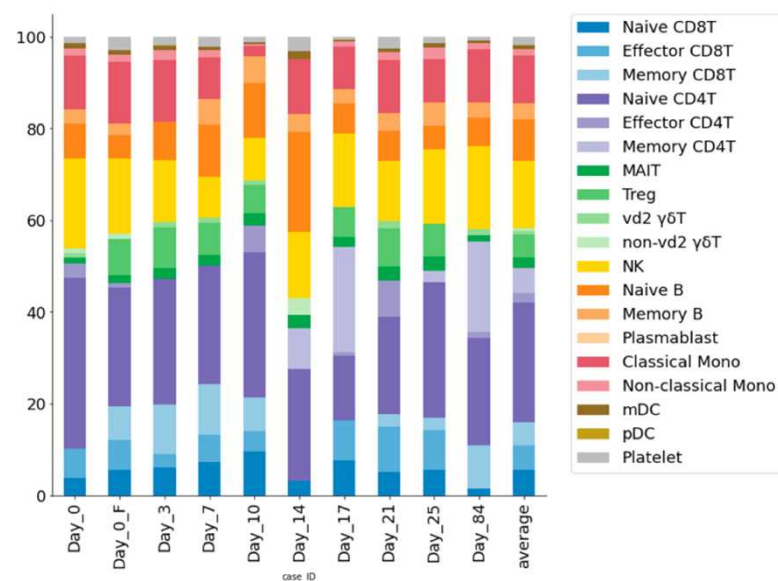


図 2：健常者末梢血におけるscRNA seqプロフィール
それぞれの個人について、1か月間にわたる日動変化を示す。

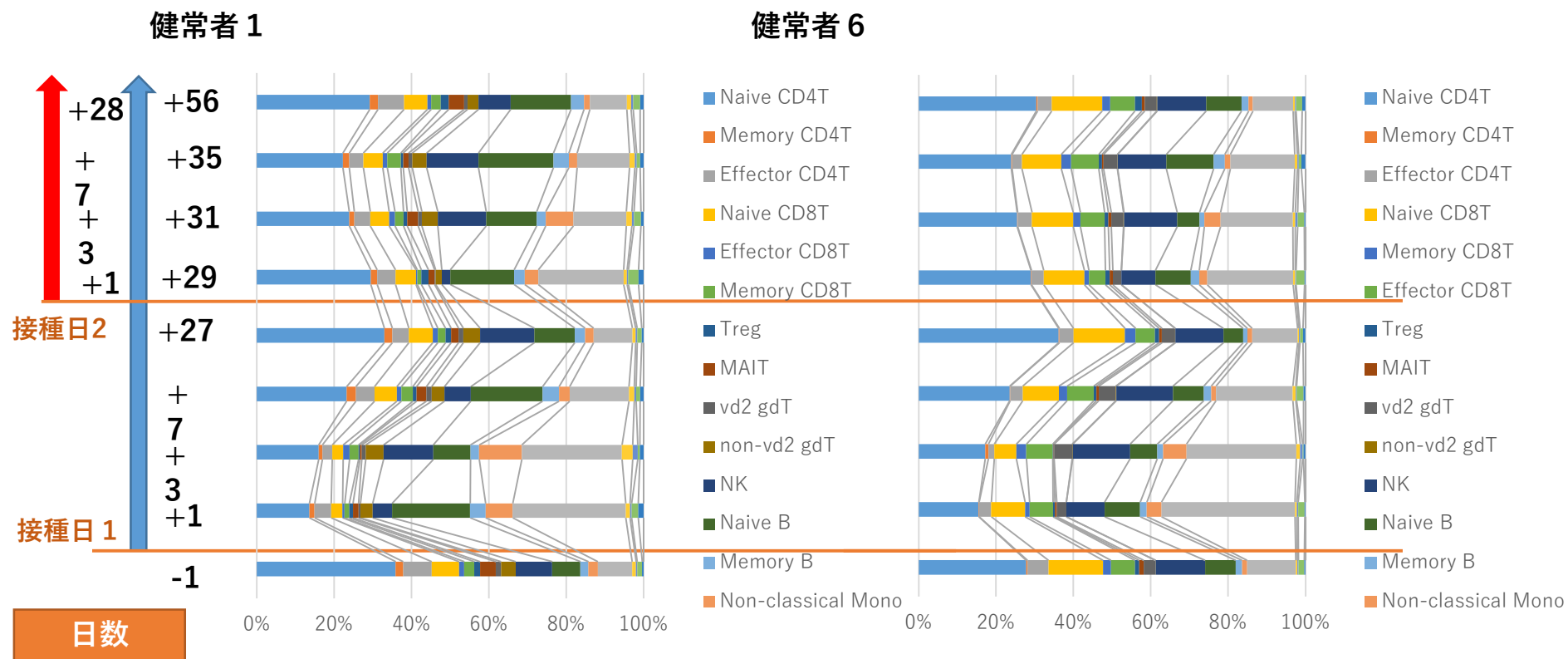


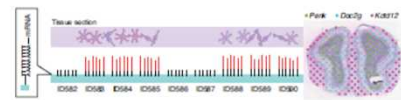
図 3： ワクチン接種前後でscRNA seqにより検出された各細胞種の割合を示す。

Nature Methods; Methods of the Year 2020

Method of the Year: spatially resolved transcriptomics

Nature Methods has crowned spatially resolved transcriptomics Method of the Year 2020.

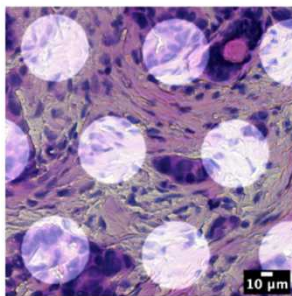
Vivien Marx



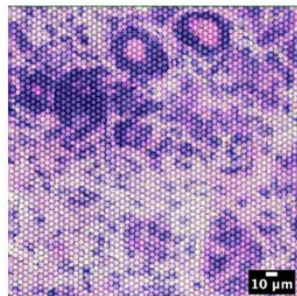
Researchers in Sweden developed an approach in which fixed, stained tissue is imaged, permeabilized and the mRNAs attach to an array of barcoded oligos. The RNAs are reverse-transcribed, the cDNAs are sequenced and yield spatially resolved transcriptomic information. Credit: Adapted with permission from ref. 1, AAAS

Visium

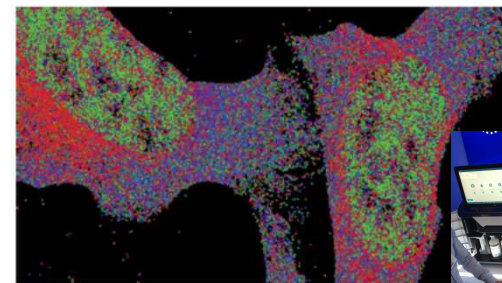
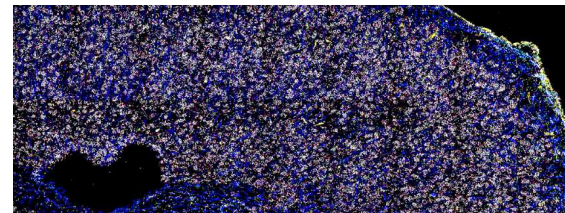
Visium HD



5000 spots per tissue section



~Millions of spots per tissue section



With MERFISH, the Zhuang lab captured the expression of 10,050 genes in individual cells in mouse brain. RNA molecules from different genes are shown in different colors. Credit: X. Zhuang laboratory, Harvard U./HHMI

シーケンスをリードアウトとしないプラットフォーム
(イメージングをリードアウトとする)

世界5か所テストユースが始まります。

Nagasawa et al Com Biol 2021

Onoda et al under review



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.

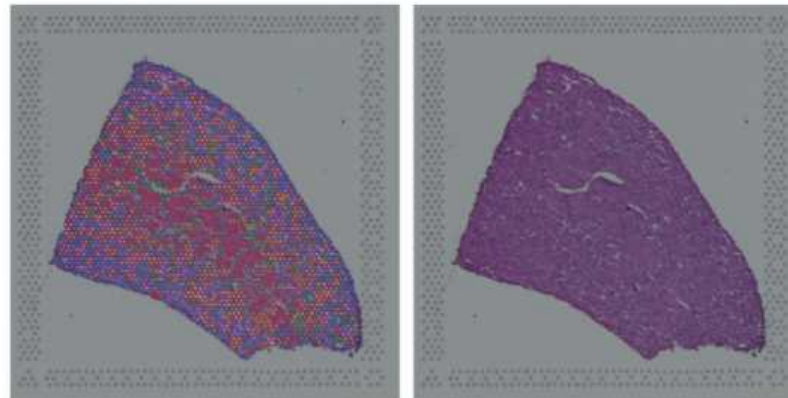
You can download the dataset from the link below.

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

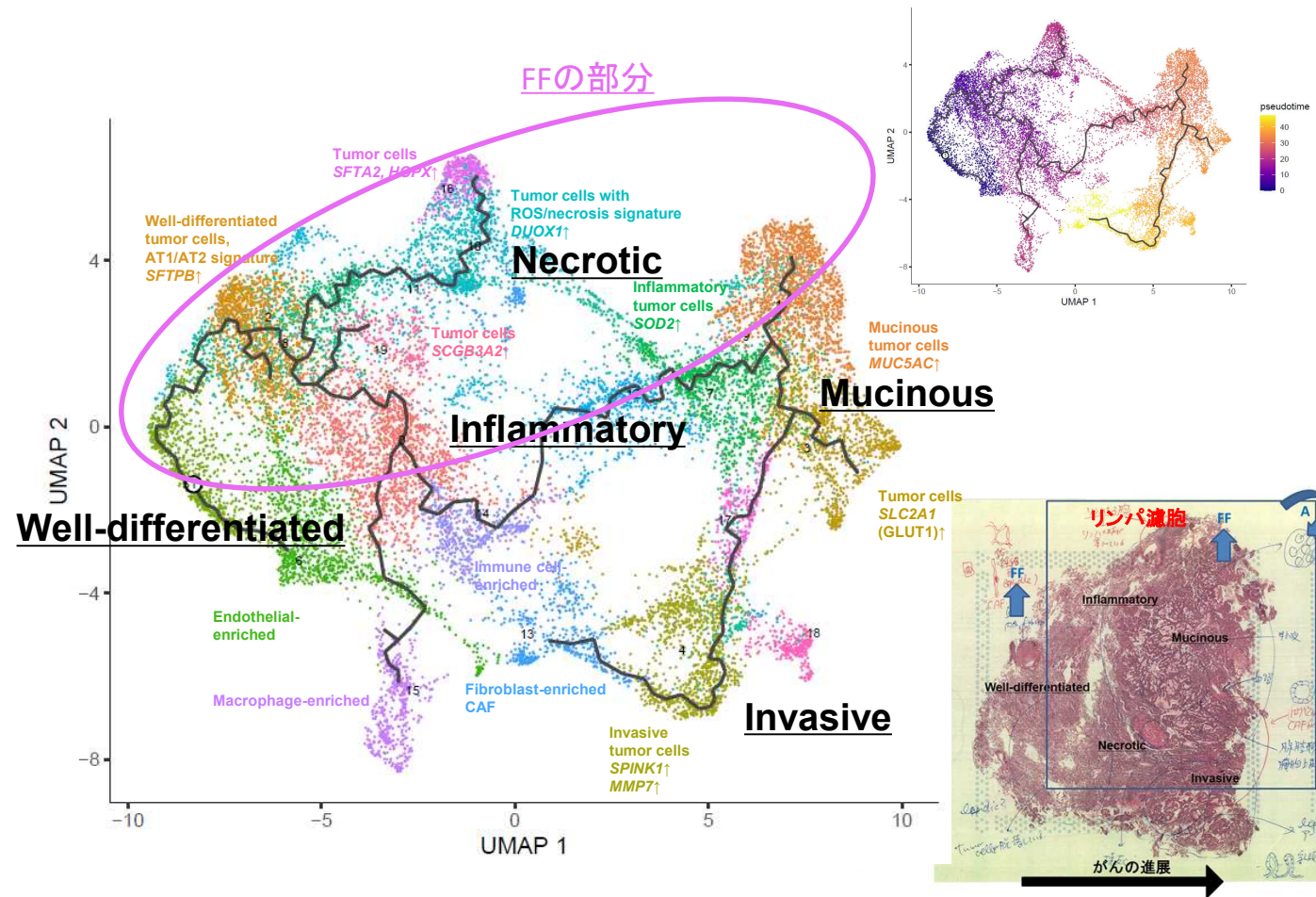
- [Summary.html \(by spaceranger\)](#)
- [Dataset for loupe* visualization](#)
- [Dataset for Seurat analysis](#)

*loupe is visualization software released by 10x genomics.

please download from [10x genomics web site](#) (Registration required).



Trajectory analysis of LUAD No.2 FF + FFPE



Nanopore PromethION

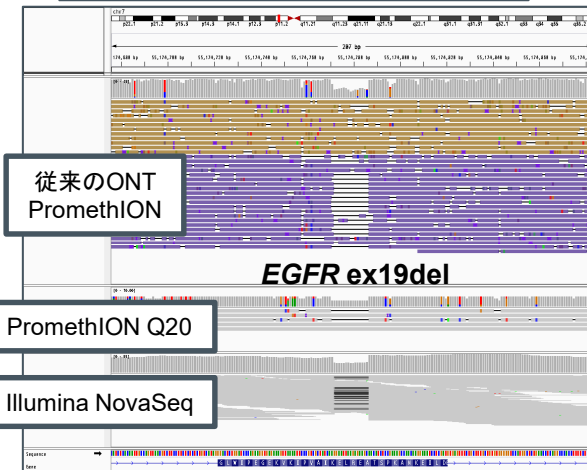
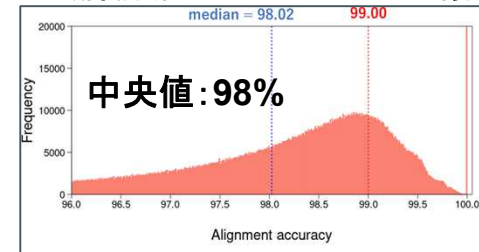
Q20 (99% Accuracy) の長鎖リード解析が可能となる

PromethION Q20 Early Access Programに参加

標準DNA(HG002)のBLAST結果



Sakamoto Y et al.の症例を再解析
S20 (肺腺癌 *EGFR* ex19del陽性)



より正確な変異検出、フェーズ情報解析が可能となる

25

Long read関連技術開発

肺がんゲノムのLong read解析

構造多型は時として非常に複雑：従来のシーケンサーでは検出困難

Research

Long-read sequencing for non-small-cell lung cancer genomes

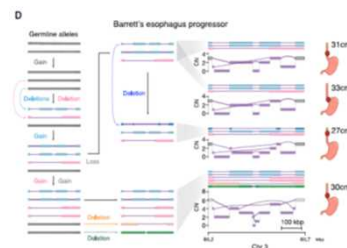
Yoshitaka Sakamoto,¹ Liu Xu,¹ Masahide Seki,¹ Toshiyuki T. Yokoyama,¹ Masahiro Kasahara,¹ Yukie Kashima,^{2,3} Akihiro Ohashi,³ Yoko Shimada,⁴ Noriko Motoi,⁵ Katsuya Tsuchihara,² Susumu S. Kobayashi,³ Takashi Kohno,⁴ Yuichi Shiraishi,⁶ Ayako Suzuki,^{1,2} and Yutaka Suzuki¹

¹Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8562, Japan; ²Division of Translational Informatics, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Chiba 277-8577, Japan; ³Division of Translational Genomics, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Chiba 277-8577, Japan; ⁴Division of Genome Biology, National Cancer Center Research Institute, Tokyo 104-0045, Japan; ⁵Department of Pathology, National Cancer Center Hospital, Tokyo 104-0045, Japan; ⁶Division of Cellular Signaling, National Cancer Center Research Institute, Tokyo 104-0045, Japan

Genome Res. published online September 4, 2020
Access the most recent version at doi:10.1101/gr.261941.120



鈴木 祐子准教授
(CBMS, 東大新領域)



Sakamoto et al Genome Res 2020

肺がんTranscriptomeのLong read解析

がんのスプライス異常がネオ抗原の母地になっている可能性

Oka et al. Genome Biology (2021) 22:9
https://doi.org/10.1186/s13059-020-02240-8

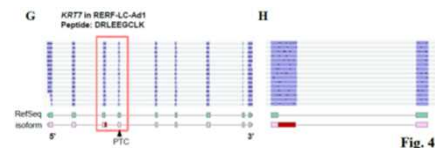
Genome Biology

RESEARCH Open Access

Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer

Miho Oka^{1,2}, Liu Xu¹, Toshihiro Suzuki³, Toshiaki Yoshikawa⁴, Hiromi Sakamoto⁵, Hayato Uemura⁶, Akiyasu C. Yoshizawa⁷, Yutaka Suzuki¹, Tetsuya Nakatsuna⁸, Yasushi Ishihara⁹, Ayako Suzuki¹ and Masahide Seki¹

NMDのおこらないガンで多い？

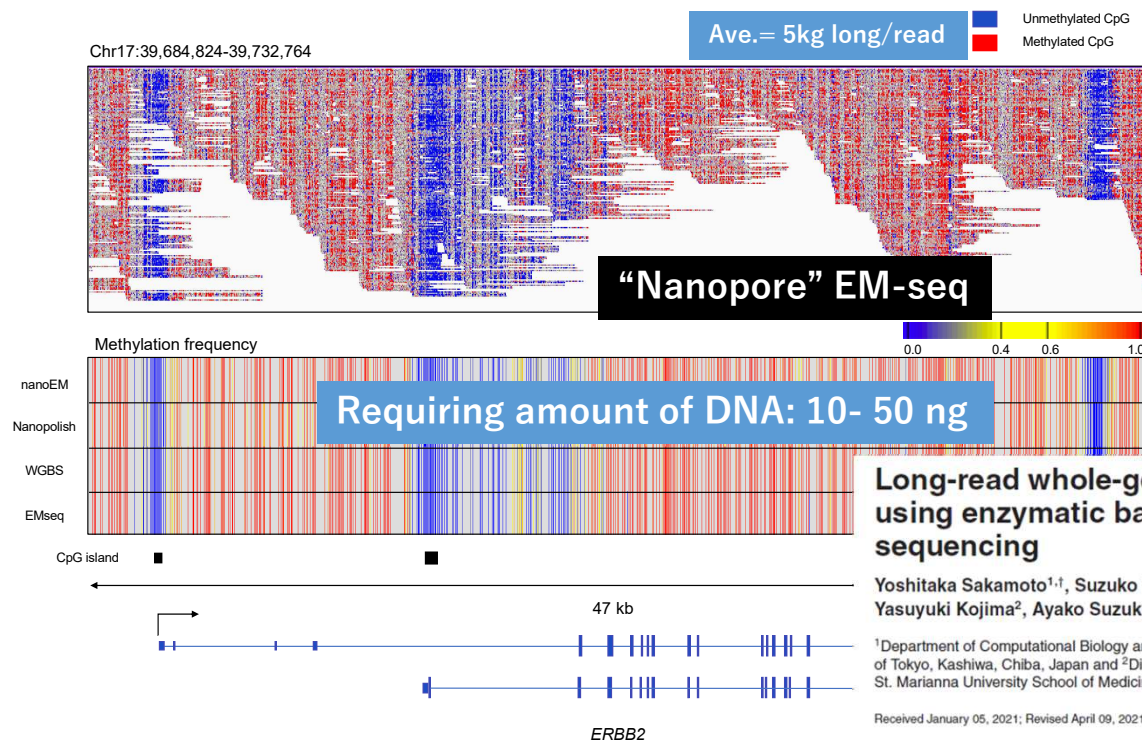


岡 実穂 DSTEP博士課程学生
(小野薬品工業/CBMS, 東大新領域)

Oka et al Genome Biol 2021

エピゲノム（DNAメチル化）の長鎖解析

Long read Methylation Analysis using nanoEM-seq



関真秀DSTEP特任准教授
(CBMS, 東大新領域)

Sakamoto et al. NAR 2021

Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing

Yoshitaka Sakamoto^{1,†}, Suzuko Zaha^{1,†}, Sato Nagasawa¹, Shuhei Miyake¹, Yasuyuki Kojima², Ayako Suzuki¹, Yutaka Suzuki^{1,†} and Masahide Seki^{1,†}

¹Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan and ²Division of Breast and Endocrine Surgery, Department of Surgery, St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan

Received January 05, 2021; Revised April 09, 2021; Editorial Decision April 23, 2021; Accepted April 30, 2021

本書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わかるようになることを目的としています。

リンク・シングルセル: 基物の必要性

リンク・スキャンルネサンスの様々なプロジェクト

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。

この文書にある結果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、以下の

```
qlogin -l s_vmem=16G,mem_req=16G
```

Rをご使用する場合は、R3.6以上のバージョンが必要です。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/3.8
```

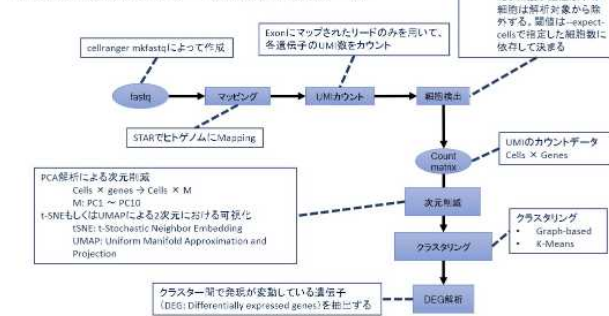
リンク：シングルセル解析の様々なプラットフォーム

ここではマウス肺がん細胞のncRNA-seqデータと、解析ツールとしては10x GenomicsのCell Rangerを用いた。

④ 既知の濃度の標準溶液を用いて検出限界を調べる。

- Cell Ranger実行結果の確認
- Loupe Cell Browserを用いたデータの確認

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>

※メールアドレスの登録が必須となります。

image	Provide a reference for use with 10x analysis software. Requires a GTF and FASTA.
kirif	Prepare a reference for use with 10x analysis software. Requires a GTF and FASTA.
upload	Upload a summary of an analysis pipeline job to 10x Genomics support.
sitcheck	Collect Linux system configuration information
help	Prints this message or the help of the given subcommand(s)

今回はテストデータセットとして熊本県立大学でChromiumプラットフォームを用いて準備したマウス肺組織scRNA-seqの単細胞データ(FASTQ)で解析を行います。

1. *Journal of the American Medical Association*, 1997; 277: 1033-1036.

```
od ~/data/ #データをダウンロードするディレクトリに移動
wget https://hseon.hgu.jp/tutorials/learning/data/10X_RNA_v3_Lung.tar
tar xvf 10X_RNA_v3_Lung.tar
```

下記コマンドでCell Ranges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。

イン形式で詳解しています

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。

... ..

```

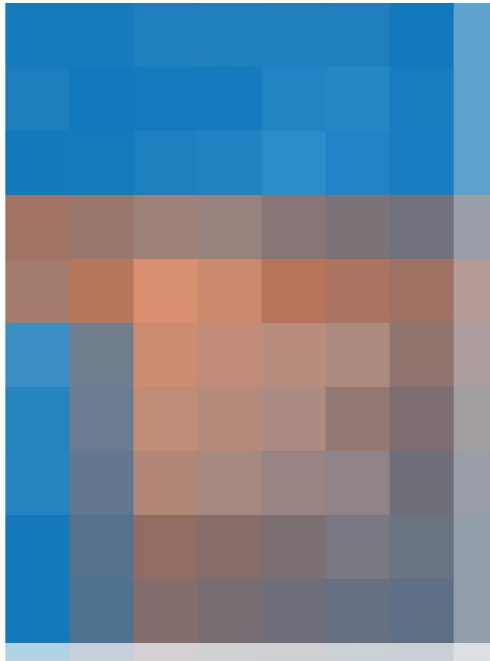
ls
ls_BNA_COUNTER_CS Filelist _limonote _log _perf _tags _uid _versions ou
limonote _finalstate _limonote _log _perf _tags _uid _versions ou
ls $Acctsに取得結果ファイルが出力されます
ls
analysis filtered_feature_bo_matrix.h5 posorted_genome_ban.ban raw_feature_bo_matrix.h5
couple.couple metrics_summary.osr posorted_genome_ban.ban 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。

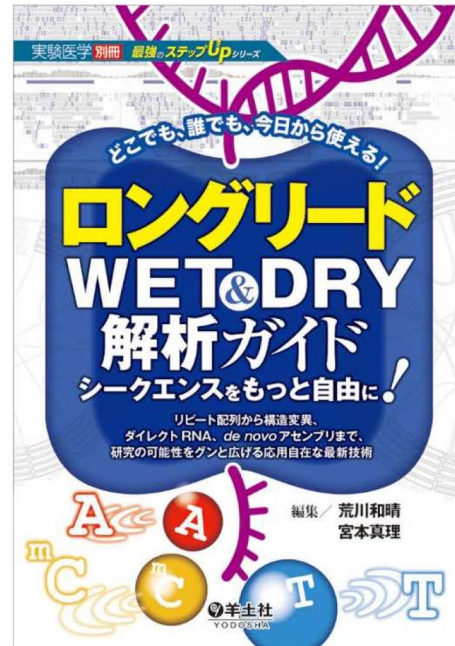
File Name	Description
web_summary.html	HTML形式のサマリー。
metrics_summary.csv	CSV形式のサマリー。
possorted_genome_bam.bam	BAMファイル。IGVで閲覧可能。
possorted_genome_bam.bai	BAMファイル。

参考文献(シングルセル・空間＋長鎖シーケンス技術)



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

英語総説書



実験医学増刊
「ロングリード」

2021年9月発売



羊土社
「東大式 生命データサイエンス
即戦力講座」

2021年11月発売

公的/民間企業保有の機微情報に配慮した個別DB空間の作成

ロングリード技術を駆使した非小細胞肺癌における新規ゲノム構造異常の抽出と意義付け

東京大学・国立がん研究センター
見出されたCLCLの例をDBKEROより公開

ナノポア型ロングリードシーケンサー PromethION

肺癌ゲノムにおいて、コピー数異常を有する小・中規模なゲノム構造異常 (CLCL) を発見

ショートリードデータを用いた確認

DBKEROの細胞株WGSデータ

DBKEROの細胞株多層オミクスデータ

※本研究のデータは、AMED次世代がん医療創生研究事業の助成を受けて取得されている。細胞株より得られたロングリードデータは、DBKEROより公開している。

Sakamoto Y et al. 2020 *Genome Research* in press
Long read sequencing for non-small cell lung cancer genomes

<https://kero.hgc.jp/examples/CLCL/hg38/>

製薬企業

Confidential

Extraction of co-expression modules: WGCNA weighted gene co-expression network analysis

WGCNA (Langfelder and Horvath et al. 2008 BMC Bioinformatics)

Similarity (Pearson correlation coefficient)

$$s_{ij} = \frac{1}{n} \sum_{k=1}^n x_{ik} x_{jk}$$

gene adjacency

Signed network $a_{ij} = |0.5 + 0.5 \cdot \text{cor}(x_i, x_j)|^p$

Scale independence

Mean connectivity

Activity of the module is represented by "Module Eigengene" (ME)

Expression levels of 13407 genes in 26 cells

73 modules (30-2835 genes)

Module "lightcyan"

Gene numbers

158 genes

共発現モジュール解析を基盤とした肺がんの新規層別化と治療標的の探索

東京大学

薬剤刺激RNA-seq/ATAC-seqデータのDBKERO Data Portalからの公開

DBKEROのオミクスデータから肺がんを特徴づけるネットワークモジュールの抽出

薬剤によるモジュールの制御可能性 → 治療

Suzuki A, Onodera K et al. 2019 *Scientific Reports* 9(1):19529.
Characterization of cancer omics and drug perturbations in panels of lung cancer cells.

DBKERO
Data Base of Encyclopedia of Regulatory Omics

肺がんロングリード・ショートリードRNA-seqデータからの新規ネオアンチジェン探索法の開発

小野薬品工業株式会社・東京大学

MinIONでがん特異的発現を示すアイソフォームの検出

Flow Cell

Tumor (MinION)

RefSeq配列

がん特異的アイソフォーム

DBKEROの細胞株オミクスデータ

- スプライスジャンクションの確認
- 発現量の確認
- ゲノム変異探索

新規ネオアンチジェン候補！

Oka M, Xu L et al. *Genome Biol* 2021
Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer.

プロテオゲノミクスのソフトウェア検証

服部恵美・塩澤久美子・近藤格 (国立がん研究センター 研究所 希少がん研究分野)

研究の概要: プロテオゲノミクスを用いたがんの変異タンパク質探索

- がんの変異タンパク質はバイオマーカー・創薬標的候補となりうる。がんの変異タンパク質を質量分析で同定したい。
- 質量分析結果をタンパク質配列DBに対して検索して同定する際、既存のDBでは未知の変異タンパク質を同定できない。
- サンプルごとに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。
- Mutated Nucleotide and Amino-acid sequence Generator, MuNAGEを開発した。
- DBKEROの肺腺癌細胞株のNGSデータ-MuNAGEに入力して細胞株ごとにタンパク質配列データベースを生成し、Swiss-Protで検索できなかったタンパク質を検出できることを確認した。

使用データ: DBKEROで公開されている肺腺癌細胞株データのうち3細胞株のRNA-seqとWGS

用途: 開発したソフトウェアの検証

つないで使える

MuNAGEの入力はNGSデータの一般的なフォーマットであるVCF(変異)とGTF(転写物) → 公共DBのデータを入力できる様々なサンプルに対応できる

公的機関

Confidential

腎特異的マーカー遺伝子発現のISH結果

近位尿細管

遠位尿細管

集合管

nphs1 ボドサイト

Slc12a3 遠位尿細管

des 血管平滑筋

tek 血管内皮

Slc5a2 S1, S2

AQP2 集合管

Henleループ下降脚

Henleループ上降脚

AQP1 近位全体

Slc12a1 遠位

これらのマーカー遺伝子を用いて細胞種を推定

© Kyowa Kirin Co., Ltd.

まずはデータだけでも見てみてください！！！！

<http://dbkero.hgc.jp/>

DBKERO
Database of Korean Genotypes for Researcher of multi-Omic data
Quick-start: For the beginners

Viewer [How to use]

Yamagishi et al Nat Com 2021
Multi-Om
Browse multi-omic data
ATLシングルセルデータ
シングルセルデータ加工のTutorial動画

Mutation Enriched Genes [Help video (Japanese)]:
Find out which gene is mutated most in our data
VISIUMのデータ
(空間トランスクリプトーム解析)

Nagasawa et al Com Biol 2021
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

Featured Dataset

Single cell dataset [Download processed data]:
Single cell dataset
Longリードデータ
(日本人肺腺がんのSV)

Cancer SV data
Japanese lung adenocarcinoma (Nanopore Technologies).
Oka et al Genome Biol 2021

Visium dataset New
Spatial Gene Expression
Spaceranger output
LongリードmRNA解析
(日本人肺腺がんの異常スプライス)

Cancer LongRNA Dataset New
Aberrant splicing isoforms in non-small cell lung cancer sequenced by MiniON (Oxford Nanopore Technologies).
Sakamoto et al NAR 2021

Long readメチル化解析

Download

Data Portal [Help video (Japanese)]
Data Portal PAGS: Genome Shien(ゲノム支援)
ゲノム支援データポータル

※追加開発

日本人健常者末梢血scRNA

健康者が日常生活でどのようにscRNAを変化させるのか?

がんゲノム長鎖Phasingデータ

がん変異の染色体背景は?
(相同染色体は等価か?)