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

鈴木　穣
東京大学新領域創成科学研究科

http://dbkero.hgc.jp/

DBKERO
Database of Kerkom Encyclopedia 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. 56 or above) for the DBKERO browsing. We do not support Internet Explorer any m

Viewer [How to use]

Multi-Omics-Viewer [Help video1,2,3 (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 sits 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.

Featured Dataset

Single cell dataset [Download processed data]:
Single cell dataset

Cancer SV dataset New
Japanese lung adenocarcinoma sequenced by PromethION(Oxford Nanopore Technologies).

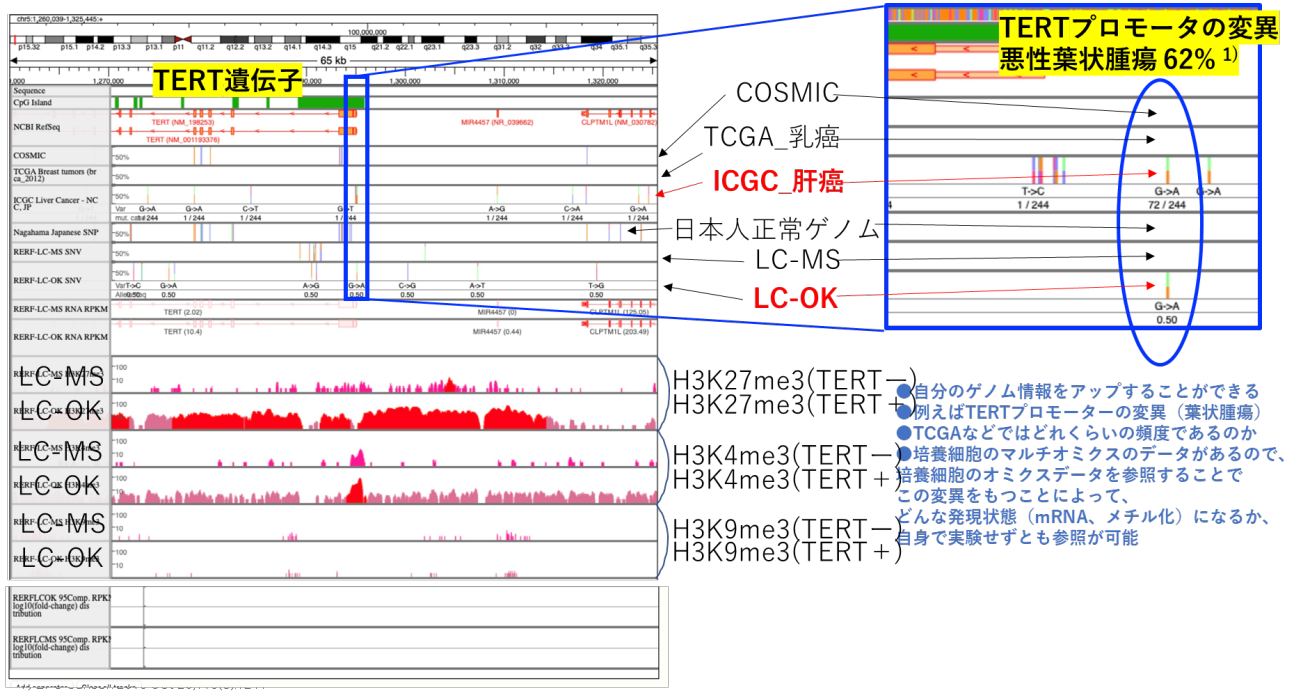
Visium dataset New
Spatial Gene Expression dataset (10x Genomics Visium system).
Spaceranger output files for loupe browser and Seurat analysis.

Cancer LongRNA Dataset New
Aberrant splicing isoforms in non-small cell lung cancer sequenced by MinION (Oxford Nanopore Technologies)

Download

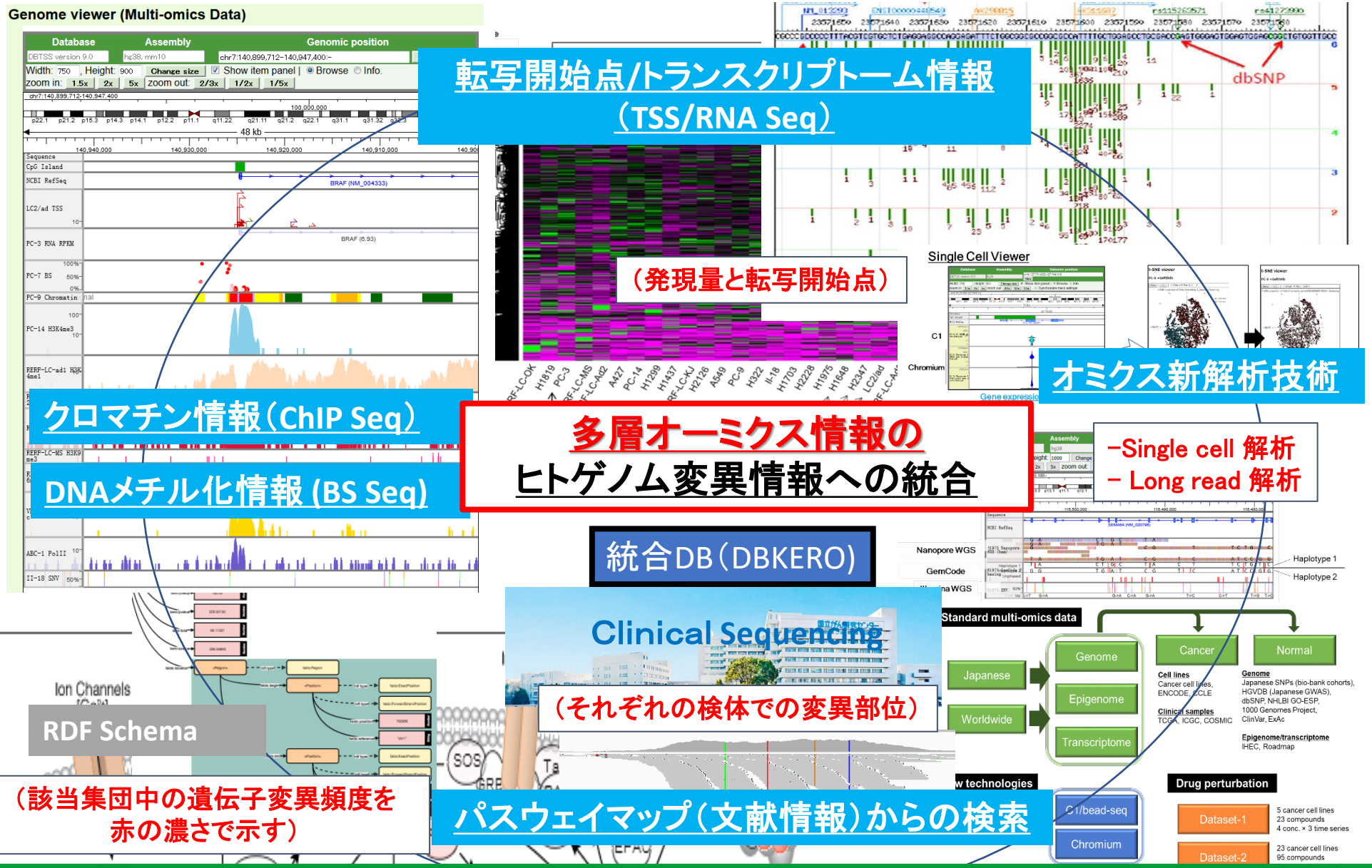
- Data Portal [Help video (Japanese)]
- Data Portal PAGS:Genome Shien(ゲノム支援)

日本人多層オーミクス
日本人由来細胞株多層オーミクス
新規技術データ



ご質問は：
ysuzuki@hgc.jp
まで。

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）

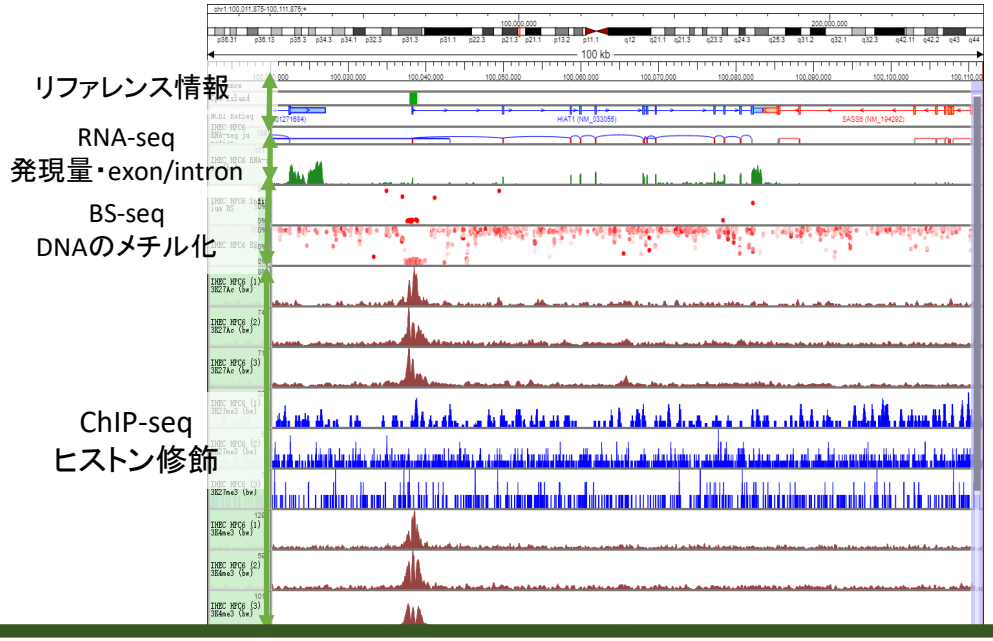
日本人正常オーミクスデータ （IHEC）

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。

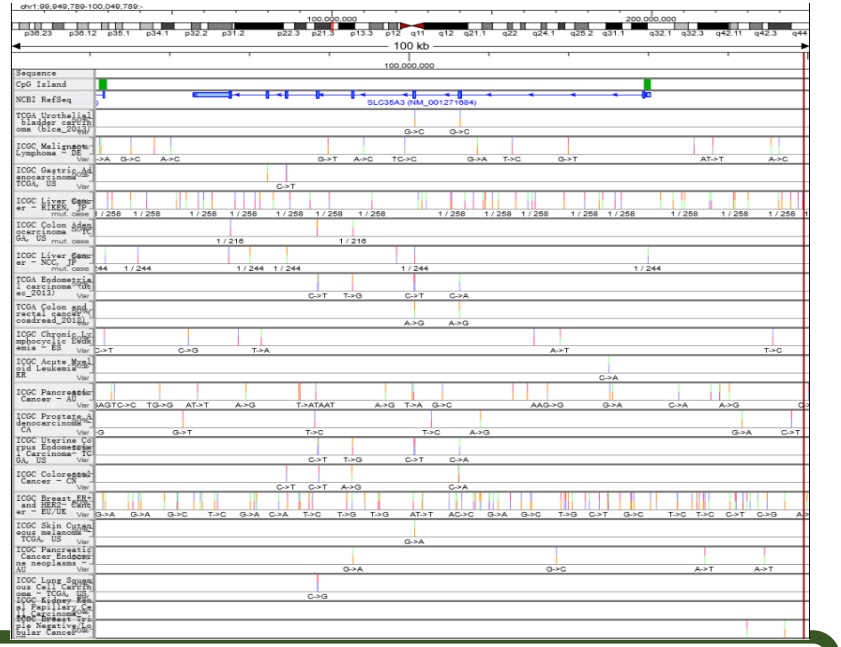
CREST-IHECデータの統合（日本人標準エピゲノム）

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

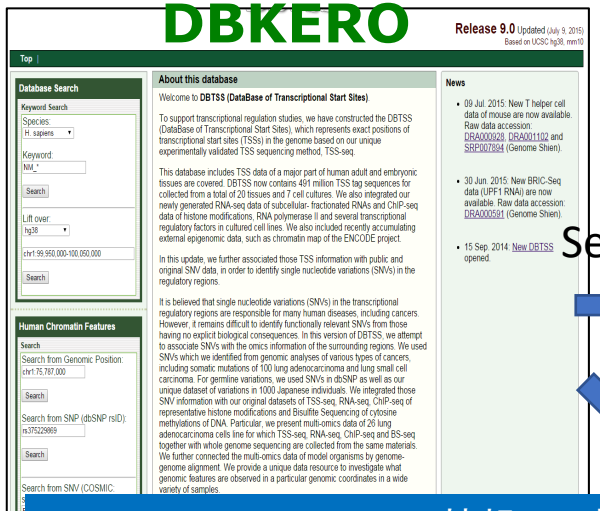


IHECデータとしてゲノムブラウザから閲覧可能なオープンデーター覧

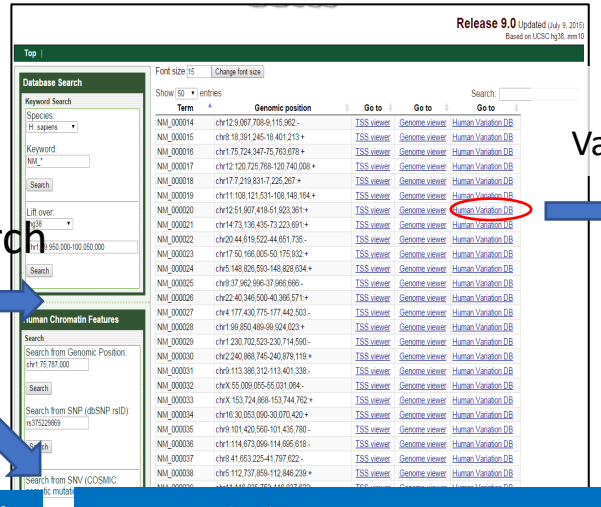
Sample	individuals	Expression		DNA methylation	Histone modifications					
		RNA-seq		BS-seq	H3K4me1	H3K4me3	H3K27ac	H3K27me3	H3K36me3	H3K9me3
		Read depth	Exon/intron junction							
Liver	8	8	8	8	13	13	13	13	13	13
Colon	11	11	11	11	11	11	11	11	11	11
Endometrial	15	30	0	13	14	15	15	15	15	15
Vascular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

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

日本人ゲノム多型

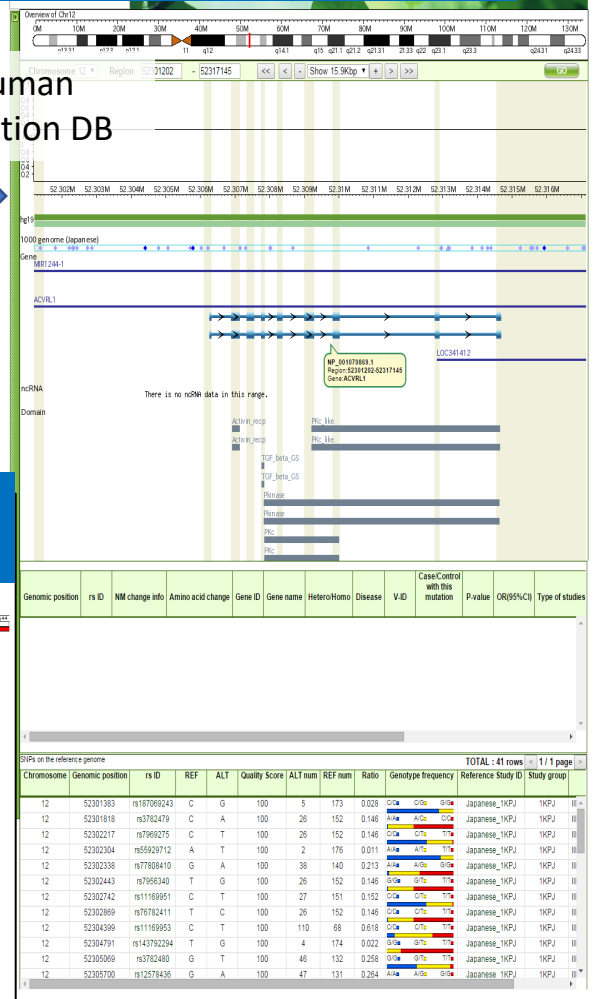


Search

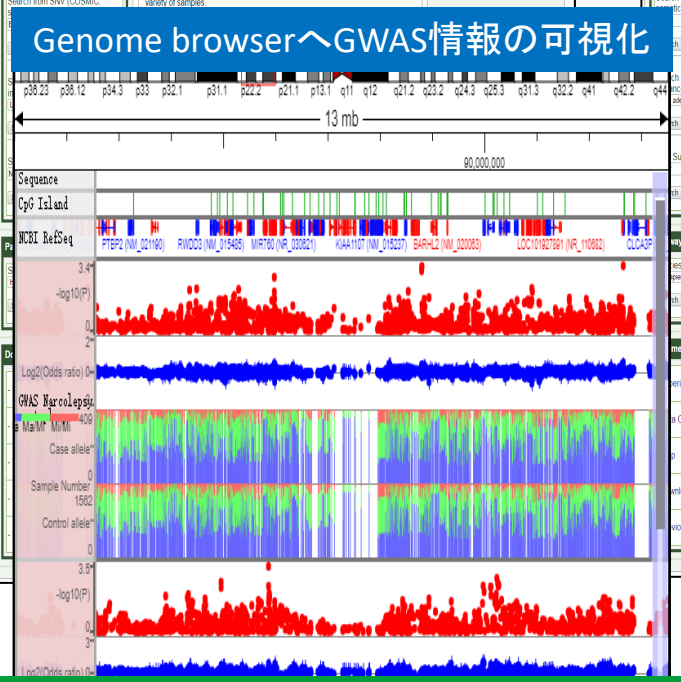


HGVDBへのリンク機能

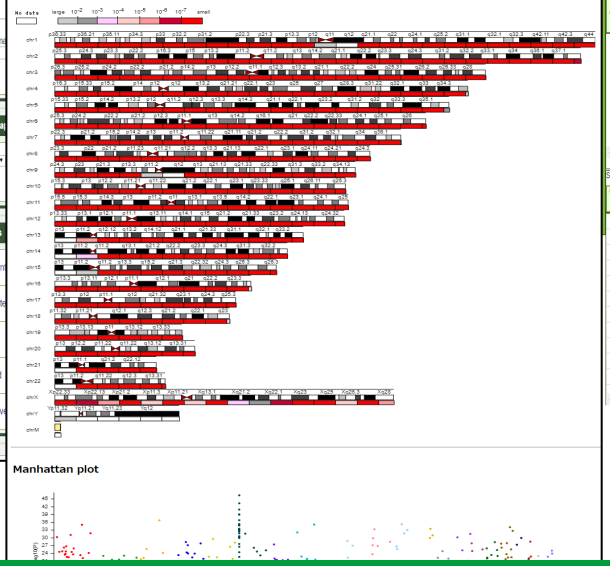
Human Variation DB



Genome browserへGWAS情報の可視化



Whole genome viewer,
Manhattan viewer



NBDC ヒトDBとの
連携へ

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

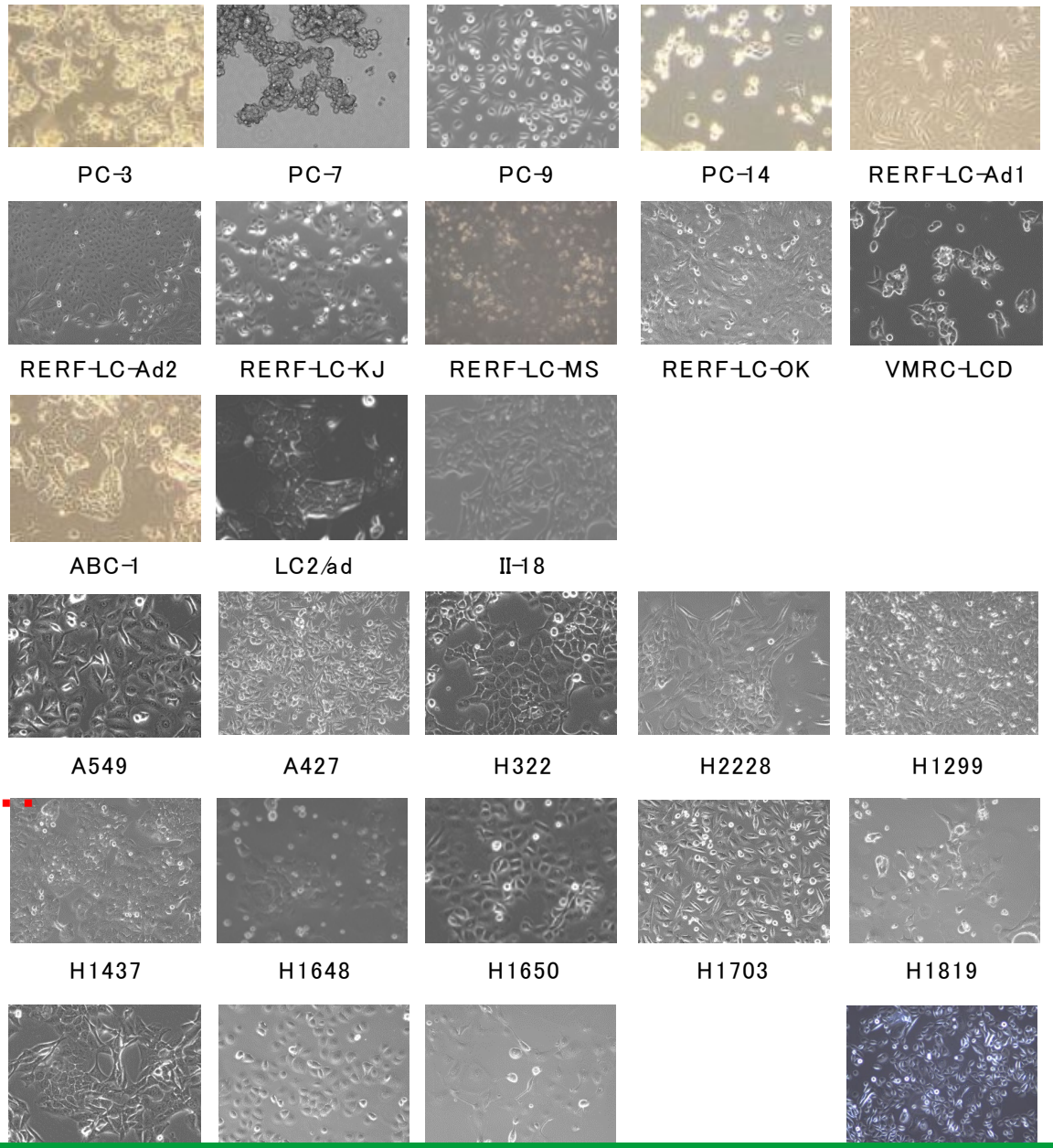
肺腺癌細胞株等で日本人由来のオーミクスデータが充実しています。

培養細胞オーミクス

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	

もちろん、ENCODE/CCLEにもありますが・・・



Multi-omics sequencing data of lung cancer cell lines

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

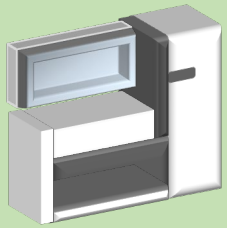
- Whole-genome sequencing
- Long-read sequencing (MinION/PromethION)

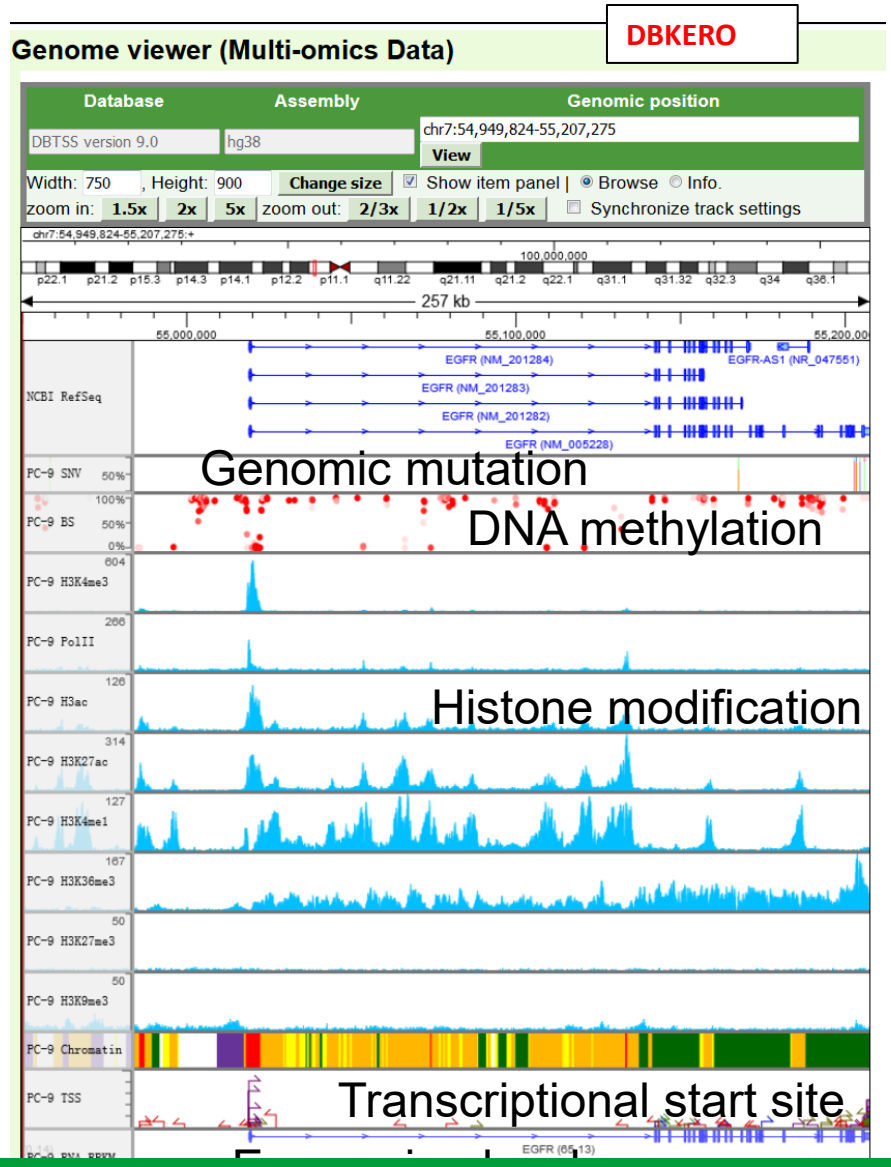
Epigenome

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

Transcriptome

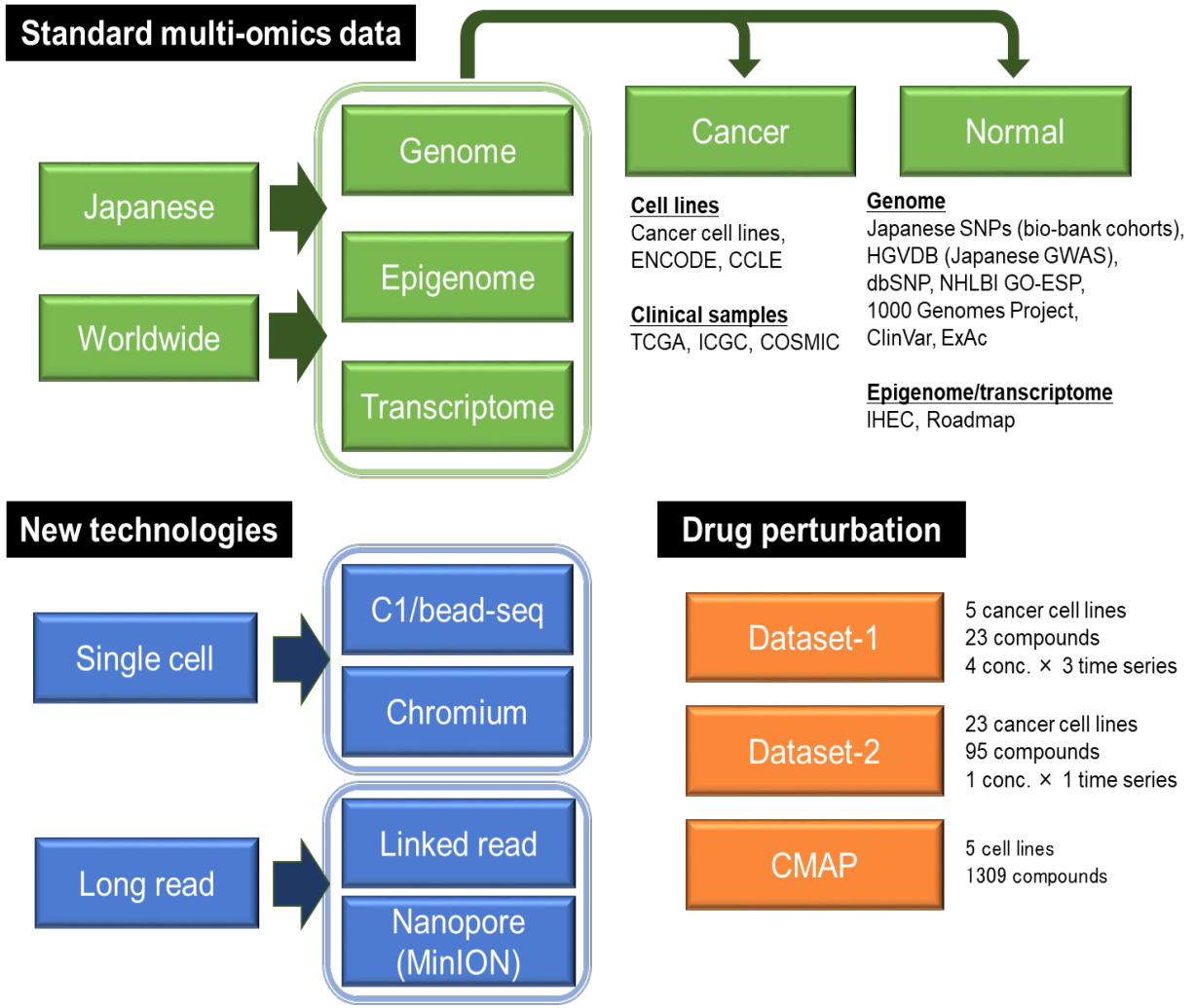
- TSS-Seq
- RNA-Seq
- miRNA-seq





データコンテンツの選択と表示

KEROデータコンテンツ

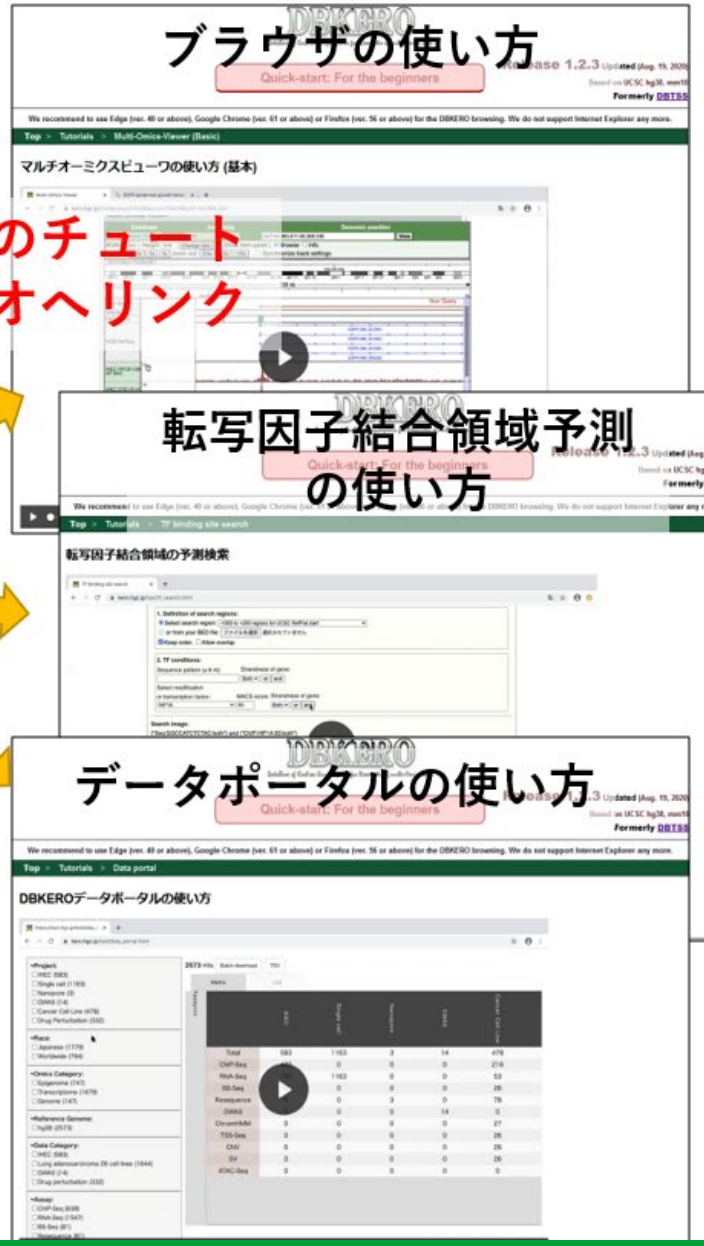


DBKEROのデータ構造の全体像．日本人の臨床オミックス情報がモデルシステムからの包括的オミックス情報とどのように関連しているかを示す．

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ーミ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．

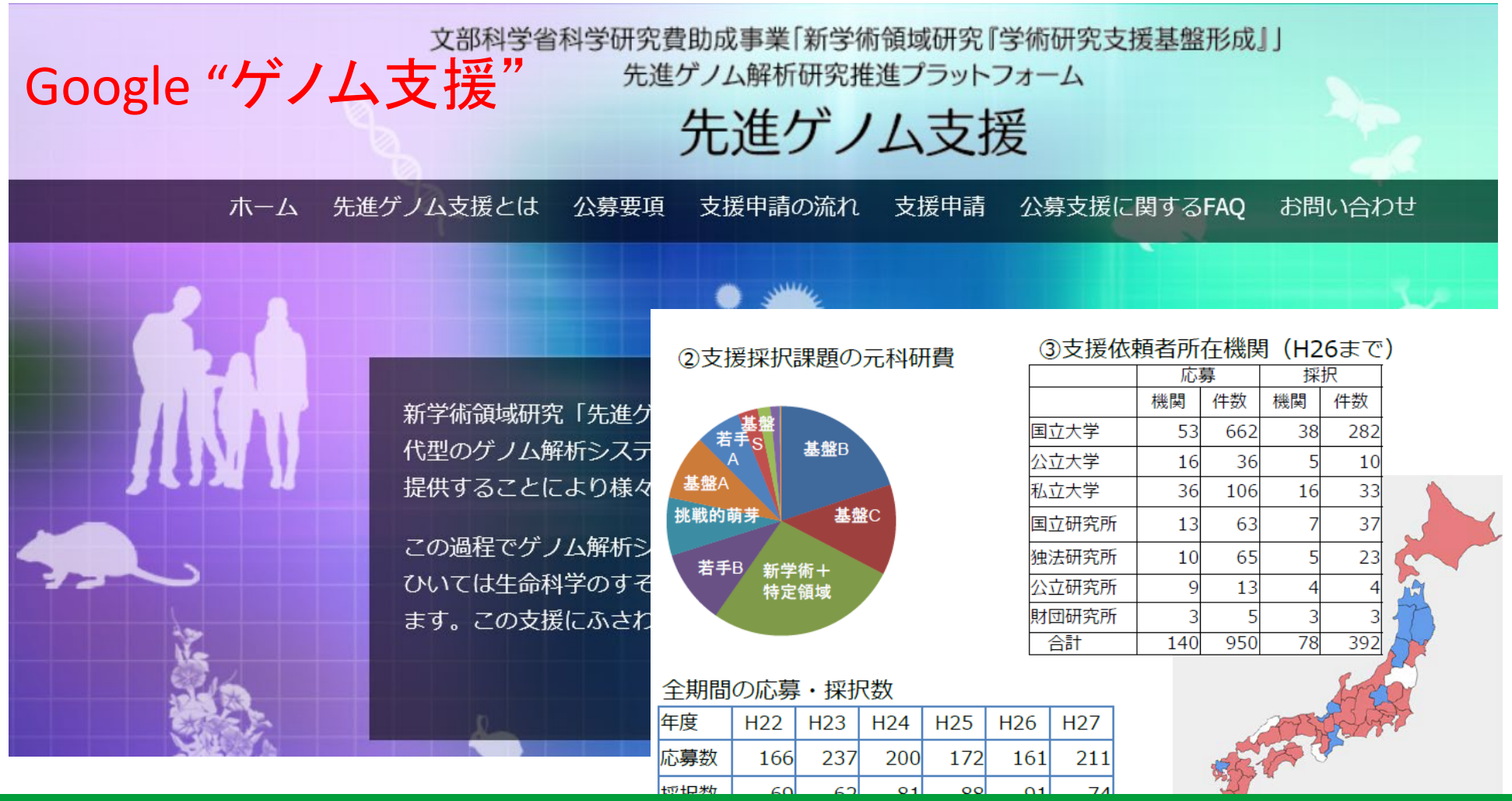
各種使い方のチュートリアルビデオへリンク



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/

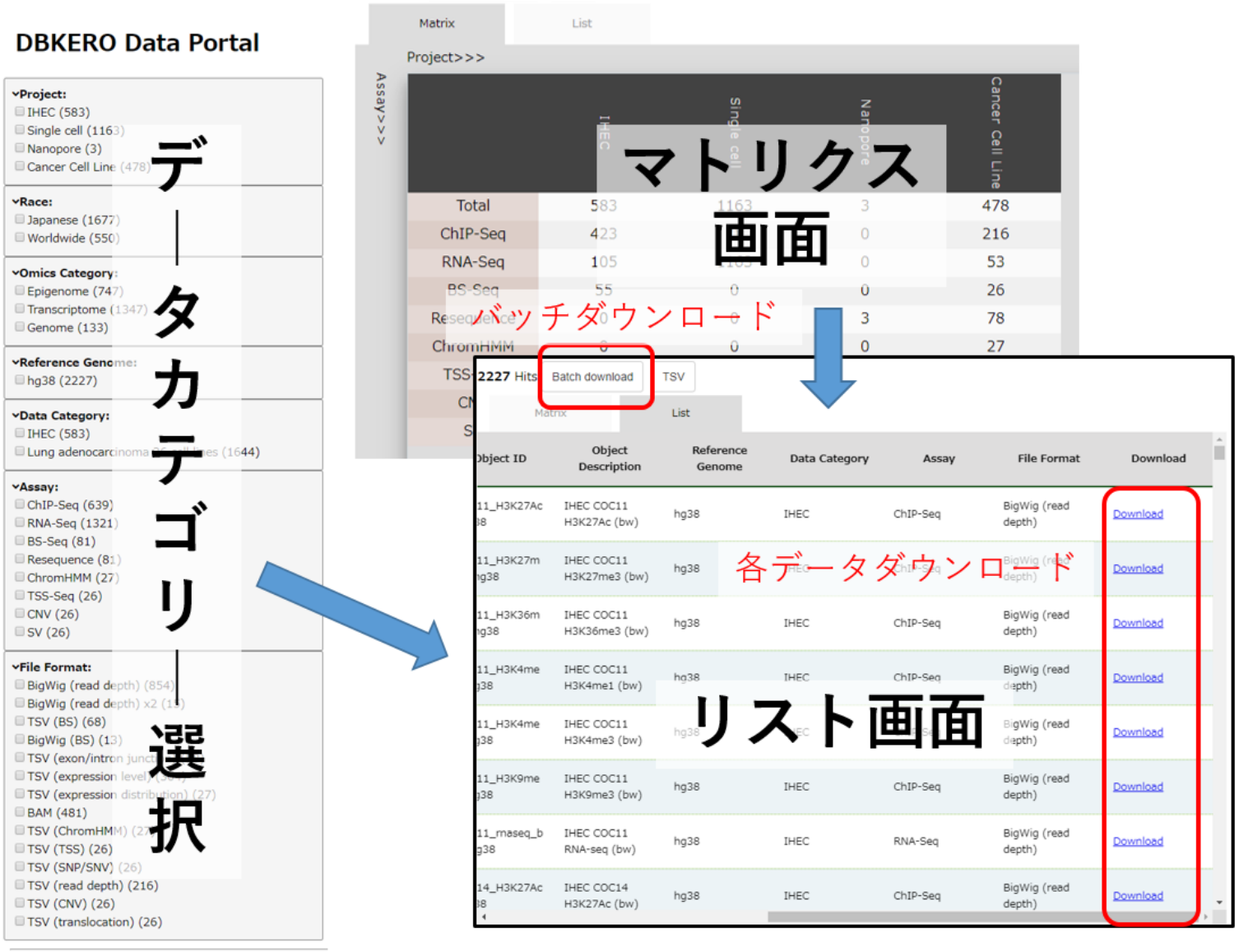
Google “ゲノム支援”



	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

年度	H22	H23	H24	H25	H26	H27
応募数	166	237	200	172	161	211

DBKEROデータポータルの構築



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・・・

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）

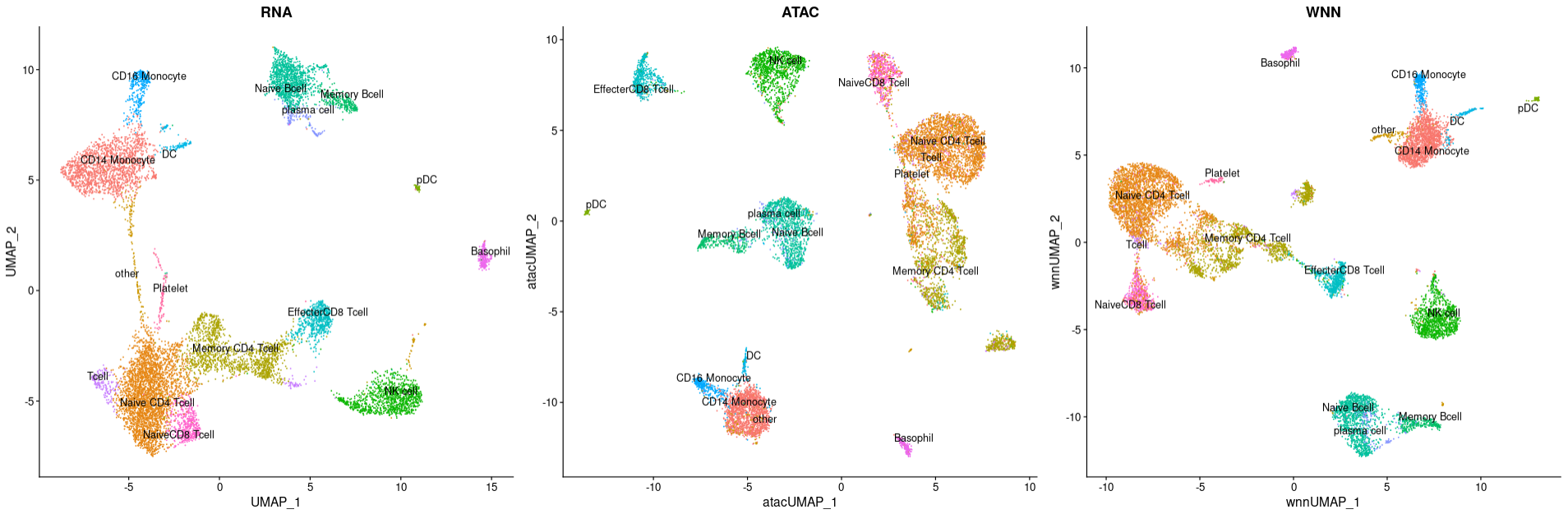
scMulti-Ome

UMAP Analysis

RNA

ATAC

ATAC & RNA



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_vmem=16G,mem_req=16G
```

Rをご使用する場合は、R/3.6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/3.6
```

リンク：シングルセル解析の様々なプラットフォーム

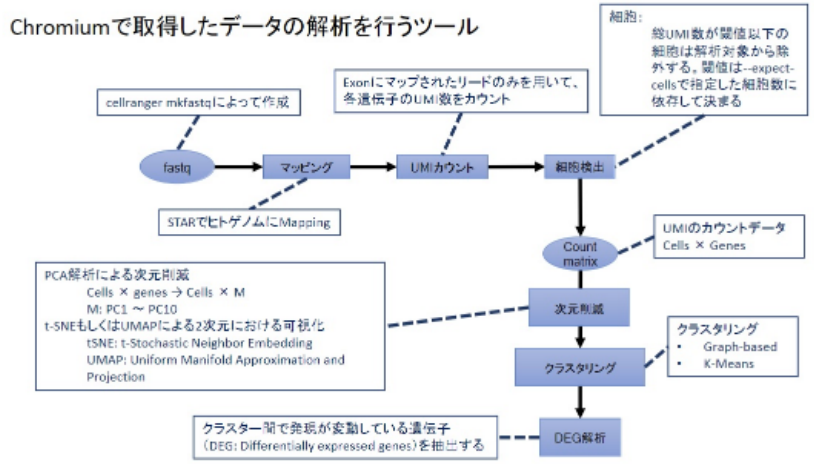
Chromiumを使ったscRNA-seqデータ解析

ここではマウス肺正常組織のscRNA-seqデータと、解析ツールとしては10x G

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



詳しくは以下 (英文)

シングルセル解析を、
コピペできるコマンドライン形式で詳解しています

```
  mkref           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
  sitecheck       Collect Linux system configuration information
  help            Prints this message or the help of the given subcommand(s)
```

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穣研究室でChromiumプラットフォームを用いて準備したマウス肺細胞scRNA-seqの配列データ(FASTQ)で解析を行い
データセットは https://kero.hgc.jp/tutorials/learning/data/10X_RNAv3_Lung.tar からダウンロードできます。

```
$ #マウス肺細胞 Chromium scRNA-seq配列テストデータセットの準備
$ cd ~/data/    #データをダウンロードするディレクトリに移動
$ wget https://kero.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。

```
$ #適当な作業ディレクトリを作成し移動
                                                                    ラスタリング，発現変動遺伝子(DEG)解析)コマンド)を実行し
                                                                    セットを展開したディレクトリを指定
$ #--expect-cells: 予想されるシングルセル数(=10000)
$ #--localcores: 解析に使うCPUコア数(ご自身の環境に合わせてなるべく大きな値にしてください)
$ #--localmem: 使用メモリ量(GB)(ご自身の環境に合わせてなるべく大きな値にしてください)
$ cellranger count --id=smn_rnaseq --transcriptome=../data/refdata-gex-mm10-2020-A/ --fastqs=../data/10X_RNAv3_Lung/ --expect-cel
$ #...解析完了までお待ちください(上のコマンドはhgcのスーパーコンピュータで処理した場合ですが、6~7時間を要しました)
```

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 ~/analysis_scRNA/smn_rnaseq/
$ ls
SC_RNA_COUNTER_CS  _filelist    _invocation  _log       _perf      _tags      _uuid     _versions          outs
_cmdline           _finalstate  _jobmode     _prosource  _sitecheck  _timestamp  _vdrkill  smn_rnaseq_mri.tgz
$ #outsに解析結果ファイルが出力されます
$ cd outs
$ ls
analysis                  filtered_feature_bc_matrix.h5  possorted_genome_bam.bam      raw_feature_bc_matrix.h5
cloupe.cloupe             metrics_summary.csv            possorted_genome_bam.bam.bai  web_summary.html
filtered_feature_bc_matrix  molecule_info.h5             raw_feature_bc_matrix
```

以下のような結果ファイルが出力されます。

File Name	Description

Nagasawa et al Com Biol 2021

DBKERO

Database of Kidney Encyclopedia for Researches of multi-Omics data

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 Explo

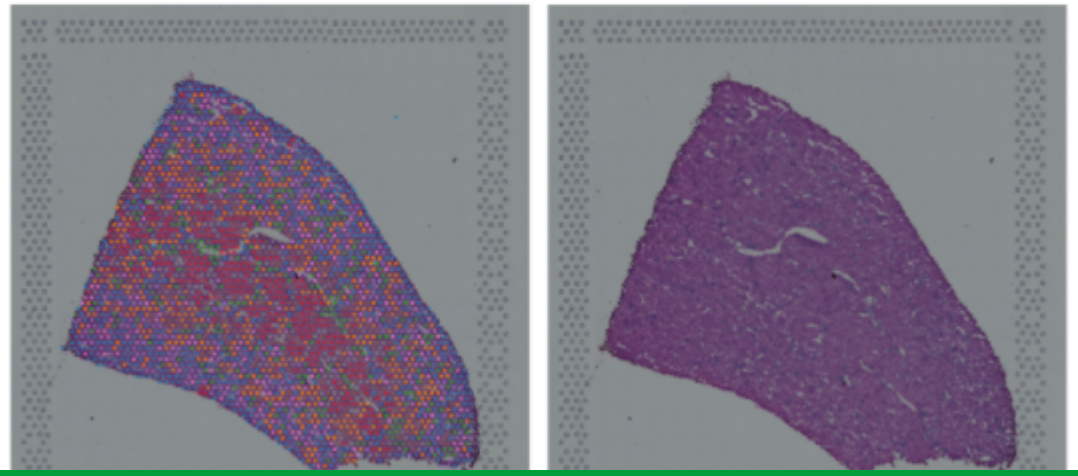
Mouse Visium Dataset

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).



独自にも進める技術開発

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：従来のシークエンサーでは検出困難

Research

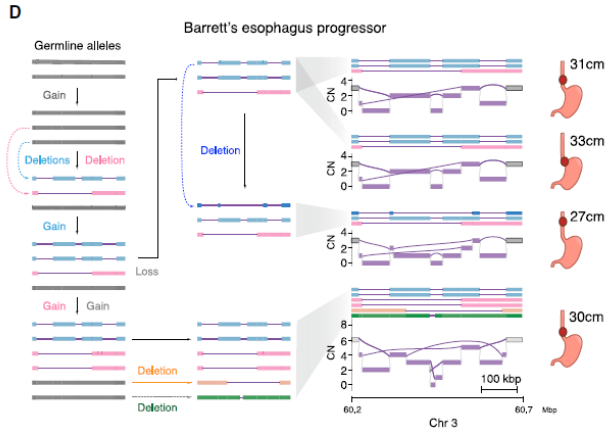
Long-read sequencing for non-small-cell lung cancer genomes

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 Suzuki[1]

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

鈴木絢子准教授
（CBMS,東大新領域）



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 Xu[1†], Toshihiro Suzuki[3,4], Toshiaki Yoshikawa[4], Hiromi Sakamoto[5], Hayato Uemura[6], Akiyasu C. Yoshizawa[6], Yutaka Suzuki[1*], Tetsuya Nakatsura[4], Yasushi Ishihama[6], Ayako Suzuki[1] and Masahide Seki[1*]

NMDのおこらないガンで多い？

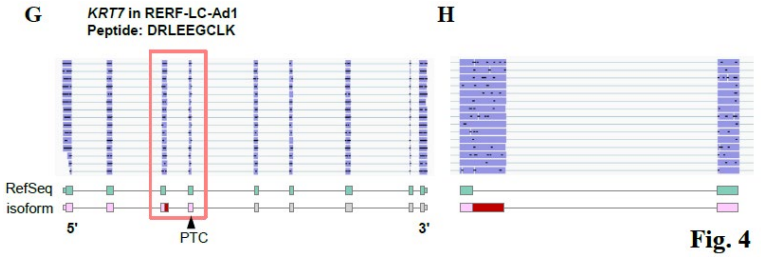
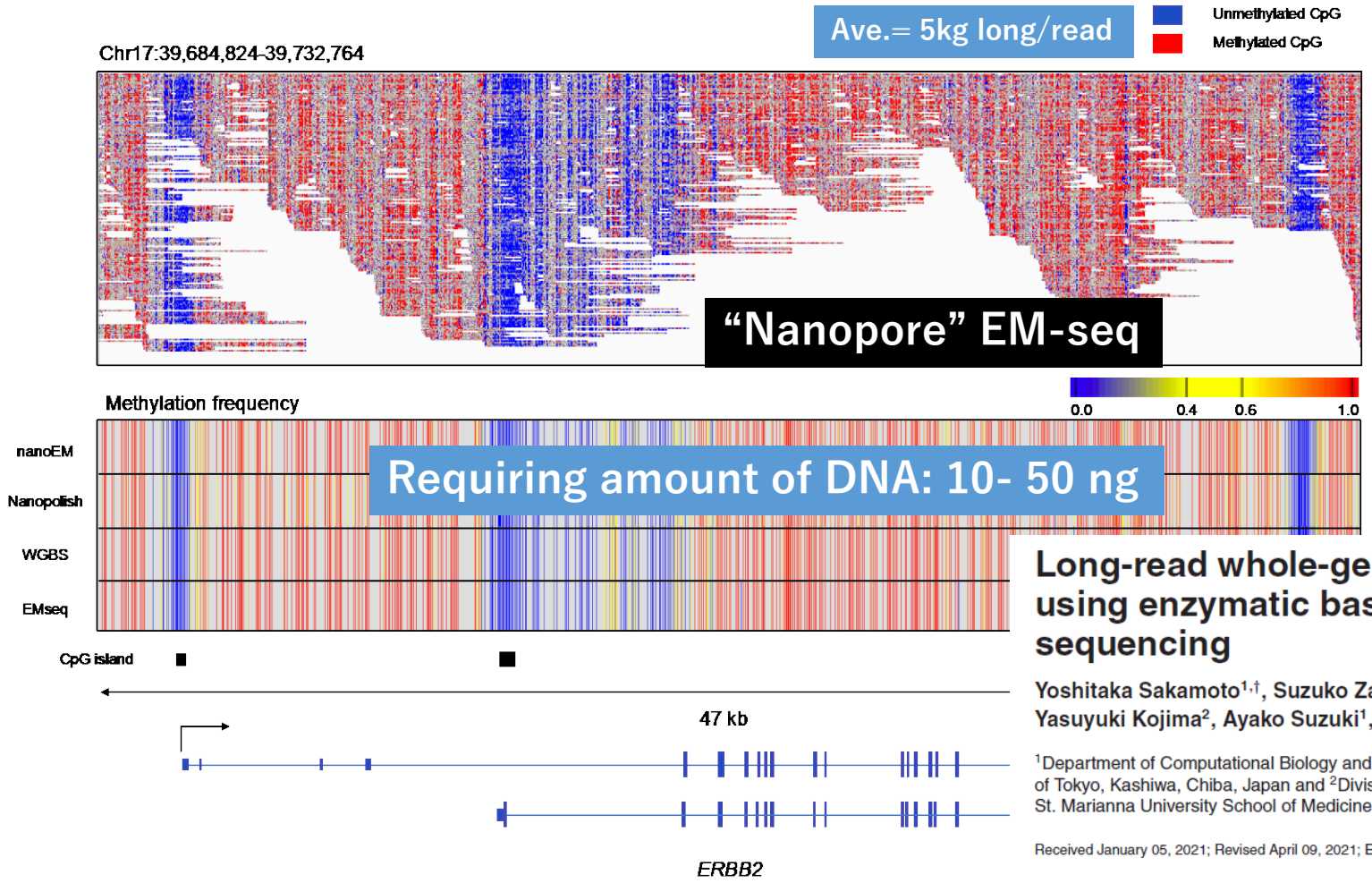


Fig. 4

岡実穂 DSTEP博士課程学生
（小野薬品工業/CBMS,東大新領域）

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,†], Satoi Nagasawa[1], Shuhei Miyake[1], Yasuyuki Kojima[2], Ayako Suzuki[1], Yutaka Suzuki[1,*] and Masahide Seki[1,*]

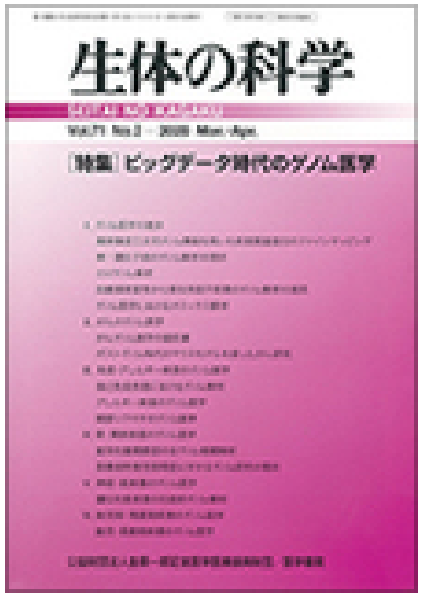
[1]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan and [2]Division of Breast and Endocrine Surgery, Department of Surgery, St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan

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

国内総説を用いた広報活動

遺伝子医学34号(2020)

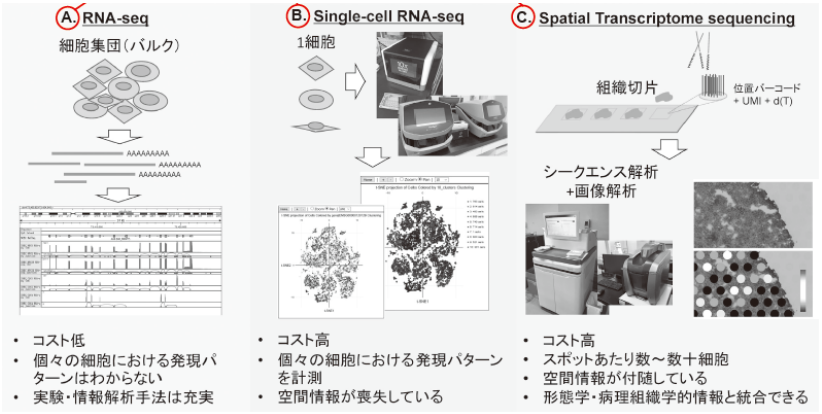
生体の科学（2020)



その他

遺伝子医学35号(2021)

和光純薬時報(2021)



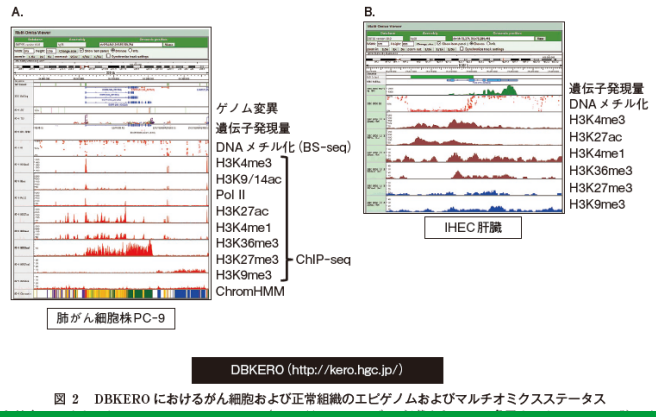


図 2　DBKERO におけるがん細胞および正常組織のエピゲノムおよびマルチオミクスステータス

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

http://dbkero.hgc.jp/

まずはデータだけでも見てみたい！

※近日公開

Yamagishi et al
Nat Com 2021

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

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

Sakamoto et al
Genome Res 2020

Longリードデータ
（日本人肺腺がんのSV）

Oka et al
Genome Biol 2021

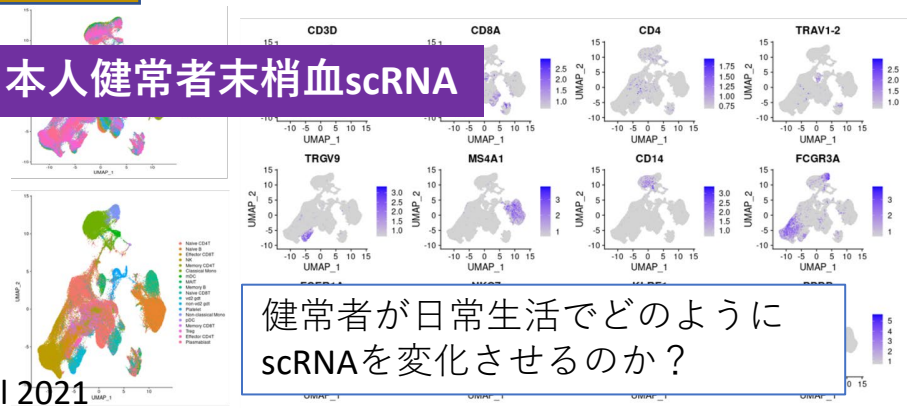
LongリードmRNA解析
（日本人肺腺がんの異常スプライス）

Sakamoto et al
NAR 2021

Long readメチル化解析

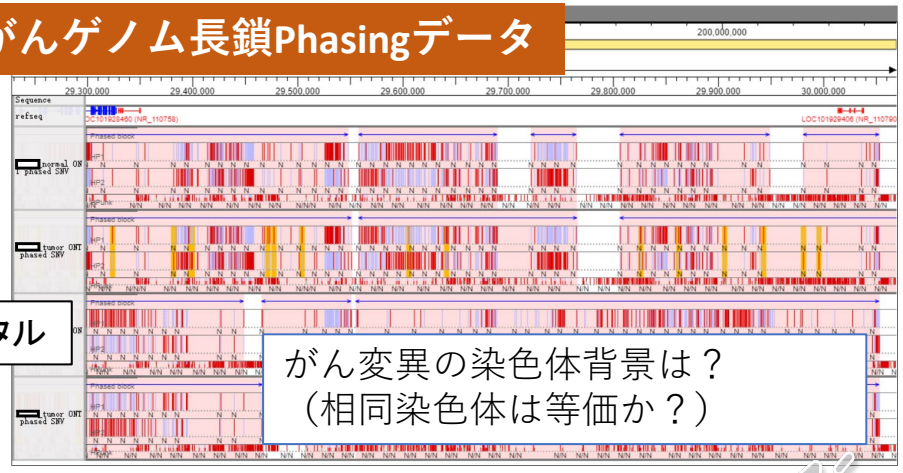
ゲノム支援データポータル

日本人健常者末梢血scRNA



健常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？）

ご質問は：
ysuzuki@hgc.jp
まで。