

2021年1月13日 NBDC  
AJACS オンライン5

多階層オーミクスデータ  
DBKERO:

Kashiwa Encyclopedia for Researchers of multi-Omics analysis

# 最近のオーミクスデータ产生のトピック

○鈴木穰、菅野純夫  
東京大学新領域創成科学研究所  
メディカル情報生命専攻

# ゲノム関連研究の国内外情勢

## 医学生物学のパラダイム・シフト



- 生体内の分子の時空間的状態のすべてを計測。
- 情報 (Big Data) として抽出して生命像を解明。
- ヒト疾患の解明による予防・診断・治療に直結。

## 米国: Precision Medicine Initiative (PMI): 110億円

- すべての疾患を対象に、住民コホート研究、電子カルテ情報を統合。
- オープンなデータ共有により、新しい研究パラダイムの実現を目指す。



## 英国: Genomics England: 330億円

- 500万人の全ゲノム配列解析を実施。
- 民間企業のクラスター形成ビジネスモデルの提出



"UK BioBank"

## 中国: 十三五計画: 総額1兆円(?)

## 日本 (準備段階: データ解析環境整備が喫緊の課題)

- ゲノム医療実現推進プラットフォーム事業が疾患別ゲノム研究を推進。
- 臨床ゲノム情報統合データベース整備事業が疾患中心に展開し、データ共有を推進。
- がんゲノム医療が開始され、がんゲノム情報管理センター(C-CAT)を設立。
- バイオバンク、コホートは高品質。

国名	プロジェクト	規模(人數)	予算	年
米国	Precision Medicine Initiative	100万人	110億円	2018
	Cancer Moonshot	300万人	330億円	2018
英国	Genomics England	500万人	総額470億円	2012-2018
フランス	Genomic Medicine France	100万人	総額870億円	2015-2020
EU全般	Horizon 2020	>1000万人	総額9700億円	2014-2020
中国	十三五	>1000万人	総額1兆円	2016-2020
日本	厚労省(毎日新聞報)	10万人?	550億?	2020-?
	AMED・文科省	1万人	累計100億円程度	2011-

表:  
各国のゲノム関連  
データ整備事業の  
概略を記す

# 医学生物学研究のパラダイム・シフト

## ゲノム解析がもたらした基礎医学研究のパラダイムシフト

病原体 → 疾患遺伝子 → Big Data

-がんを始め多くの疾患について有効な治療法が見つかった

-多くのがんについて薬剤療法での根治は依然として困難

-未だ何千もの疾患、特に多因子疾患は、依然として病態機序が不明

## アカデミアにおける次世代ゲノム研究の開始：配列から生物学データへ

- 生体内の遺伝子像の時空間的状態のすべてを計測
- (50兆個のすべての細胞は無理でも数万個は可能)
- 情報 (Big Data) として抽出して生命像を解明
- ヒト疾患の解明による予防・診断・治療に直結

従来型生物学

ゲノム医学へ

次世代ゲノム医学へ

比較的小規模な個別解析

シークエンサー・配列情報の集約化

先端計測・解析ネットワークの整備

アカデミアでは日本の誇る個々の科学者の強みを加速する好機

# 計測技術の進展

## 先端計測機器整備

-先端計測機器の読み出しツールとしての次世代シークエンサー

-1細胞、時空間オーミクス解析機器

(データ収集例1)

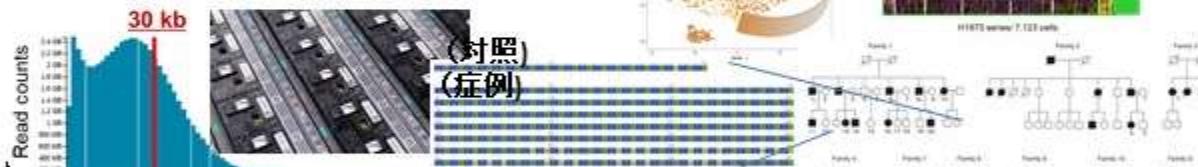
がんの転移・薬剤抵抗性等のまれな細胞集団、あるいは他の疾患で特定の細胞種にのみ起こる病因の解明



(Kashima et al, in preparation)

-長鎖DNA解読機器

がんの空間トランскルプトーム解析へ向けて



(Sone et al Nat Genet 2019)

(データ収集例2)

これまでの次世代シークエンサーで見つ  
からなかつた変異/多型の同定と解析

## 多因子疾患の解析

- WGS、WES

- 遺伝統計学解析法の開発・改良(GWAS、RVAS)による多因子疾患関連遺伝子座の解析

- オーミクスデータ統合によるファインマッピング

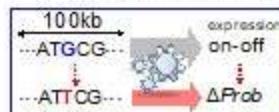
(原因遺伝子解明)



(Akiyama et al  
Nat Genet 2017)



(Kanai et al Nat  
Genet 2018)



Allによる遺伝的多様性の  
細胞学的影響予測(Koido  
et al. in preparation)

RC2:  
空間トランскриプトーム解析用マシン

**FISSEQ (fluorescent in situ sequencing)**  
**(Lee JH et al. 2014 Science)**

Resolution: **Nanometer-scale; sub-cellular**  
Gene expression: **200 genes**  
Sample: **Frozen or FFPE**

Sample prep kit  

- Oncology & immunology
- Neurobiology
- Infectious disease
- Custom

Multi-omics  

- RNA
- Genome
- Protein (DNA barcoded antibody)



## Single Tube Long Fragment Reads (stLFR)

Long range sequencing information  
from short reads

- ✓ Haplotyping phasing
- ✓ De novo assembly

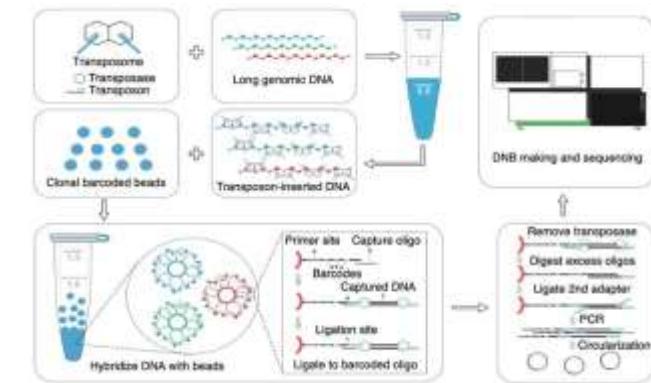
## シークエンサー

### DNBSEQ-G400

### DNBSEQ-T7

### DNBSEQ-Tx

- Build to order for million genome scale projects
- \$100 Genome Real
- Open slide, 500 nm pitch (not a flow cell)
- Robotic & fluidics
- An integrated system with 4 imagers
- Tx output: 20 Tb/day
- Tx Slide: up to 150 genomes



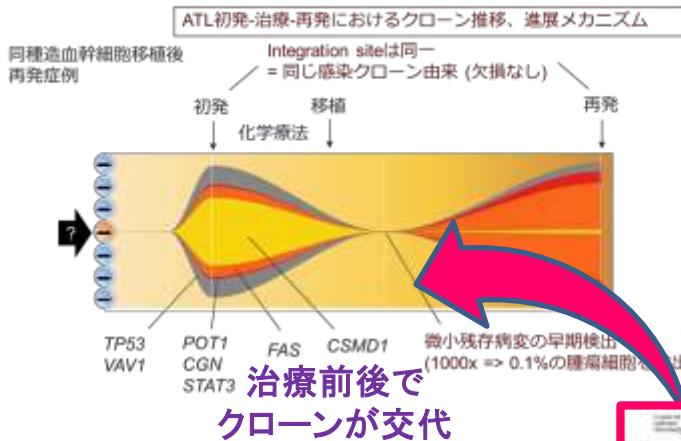
Wang O et al. 2019 *Genome Research*



# 様々な疾患解明の基礎、基盤としてのゲノム技術

## がんのクローン進化(血液腫瘍)

### 成人T細胞白血病 (ATL) 治療/移植と腫瘍細胞のクローン進化の推定

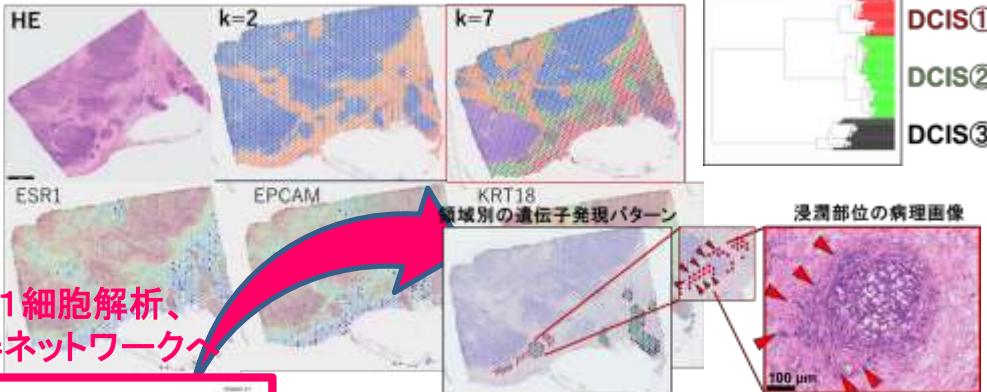


## がん化初期の微小病変 (固形がん)

設計図としてのゲノム  
読み出しとしてのオーミクス

### 疾患の理解と制御

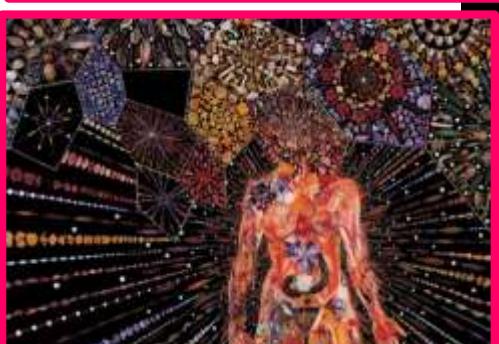
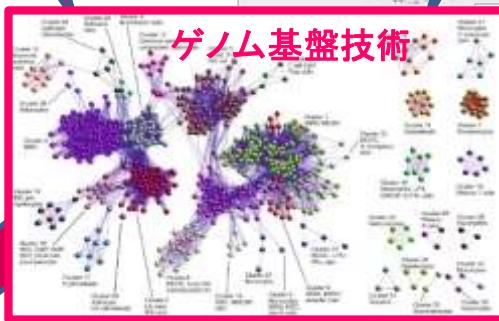
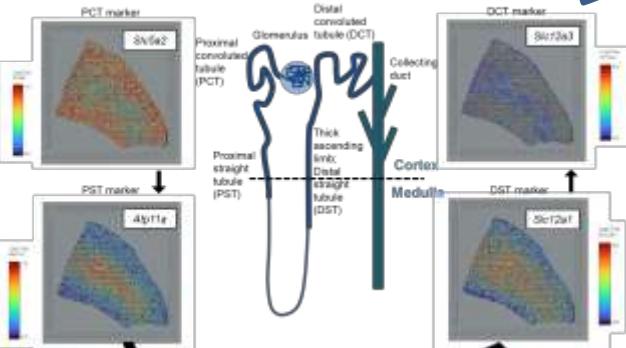
#### 乳癌の空間トランск립トーム解析



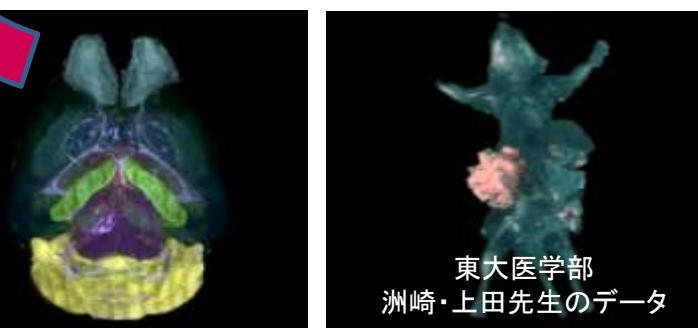
ゲノムから1細胞解析、  
臓器、臓器ネットワークへ

## 腎臓の局所微小病変解析 (Common Disease)

### ネフロンセグメント



## マウス全脳/全細胞の透明化 イメージング解析



# ゲノム解析技術/データの COVID-19 対策への応用

## 宿主ヒトゲノムの解析

- ・感染患者 vs 健常・非感染者
- ・重症患者 vs 軽症・無症候感染者

## ゲノム網羅的関連解析

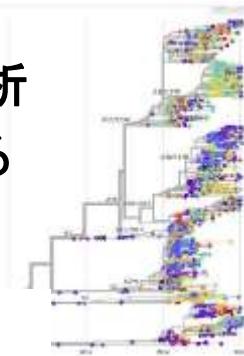
## HLA解析

## TCR/BCR解析



## ウイルスゲノム解析

- ・病原性変異の解析
- ・系統樹解析による伝播解析

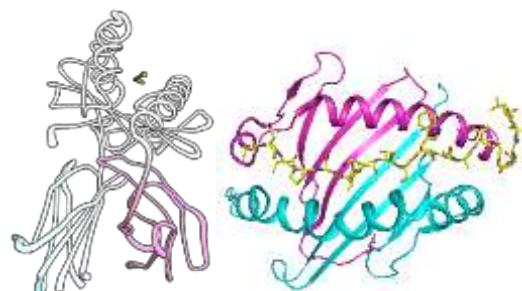


SARS-CoV-2 感染患者検体の解析については  
国立国際医療センター/国立感染研で大規模計画が進行中



計測プラットフォーム、情報解析技術/計算機資源/バイオバンク:  
一体となって、感染実態の把握、ワクチン・治療薬等の治療関連研究の加速を図る

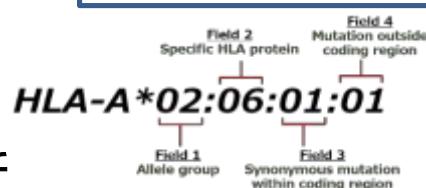
## HLA (ヒト白血球抗原)



- 抗原提示（免疫反応制御に関与）
- 自己認識のための分子
- 移植・組織適合性に重要
- さまざまな病気の発症と関連  
(感染症、自己免疫疾患、がん  
アレルギー、薬剤過敏症など)

- 全ての HLA 遺伝子群の型  
の種類数合計 : 26,899
- 遺伝型の組み合わせ (ハプロ  
タイプ) はさらに膨大
- HLA 型の分布に著しい集団差  
200 以上の疾患と関連

カギとなるHLAタイプの網羅的理



HLA分子と T細胞側から見た  
抗原ペプチド HLA分子と抗原  
ペプチドの複合体

# 新しいオーミクス解析の形

健常者の生活習慣・環境に関する新しいプラットフォームの構築  
(ゲノム+精密オーミクスデータ:多層オーミクスによるリアルワールドの理解)

## 生体試料

バイオバンク

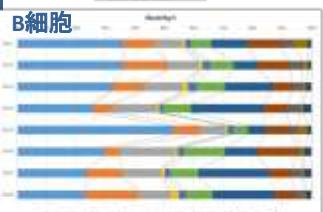
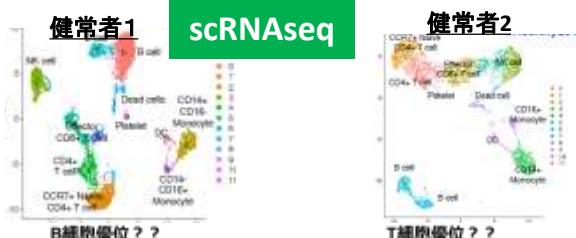
## 試料計測

NGSシークエンス拠点

## 計算機環境

ヒトゲノム解析センター

### ゲノム情報+遺伝子発現



1ヶ月間の変動

### 集団免疫の分子実態



### 環境ゲノム測定

#### 環境ゲノム解析

ショットガン  
メタゲノム解析  
長鎖DNA解析

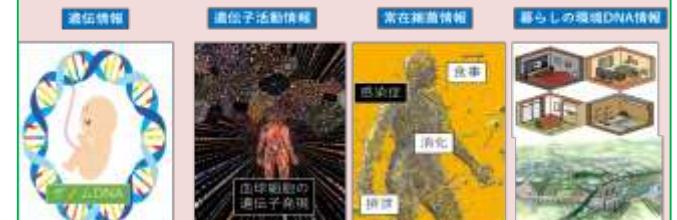


有害微生物の検出  
微生物叢の動態解析

ゲノム解析による  
“街の免疫力”的解析

→ 柏地区で試行中

様々なゲノムの解析を通じてヒト・街・暮らしを計測して、  
より安心で快適な生活環境の提供につとめる。

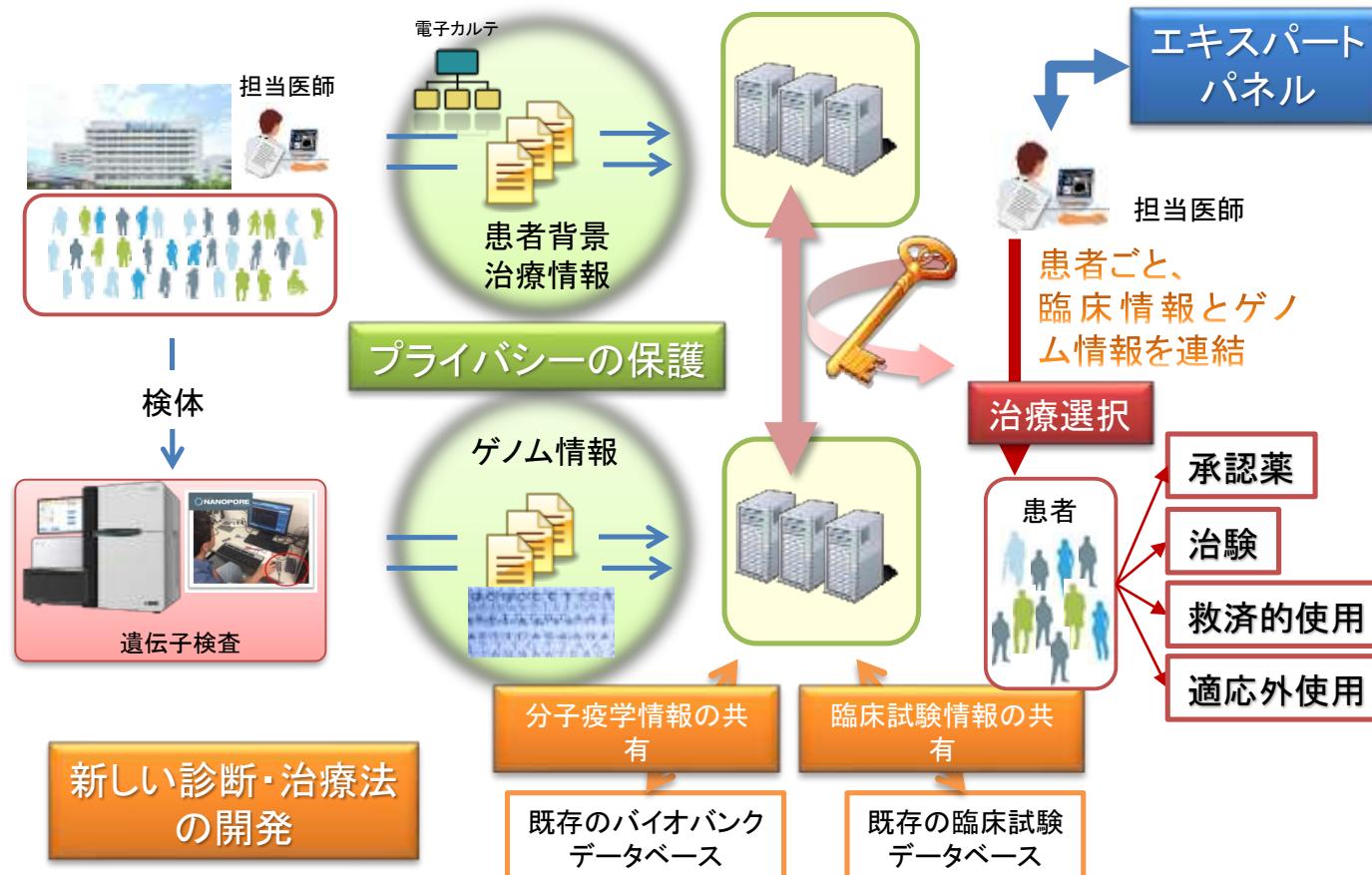


# 様々な疾患への臨床応用に必須な個人情報保護との両立

## ～安全なゲノム情報保管システムの構築が必要

### 倫理研究の必要性

※感染症研究、遺伝性疾患  
では特に急務



# 高速情報ネットワーク網 SINET の活用



SINET5: 日本全国の大学、研究機関等の学術情報基盤ネットワーク

-先進性・超高速性・高信頼性・国際性・高機能性ネットワークを実現

-機微性を担保した生体関連大規模データ共有が可能

-ハブ拠点(国立情報学研究所)が東大柏IIキャンパスに2020年秋開業

# 学術的な意義と社会的価値

## これまでの成果



GWASによる1000以上の  
感受性遺伝子座の同定

(GWAS:ゲノム網羅的関連解析)



バイオバンク・  
ジャパン(BBJ) X 医科研スーパー  
コンピュータ =  
51疾患・44万例

オーミクス統合解析に  
による病態の解明

国際共同研究による  
世界的なゲノム医療  
推進への貢献

## 波及効果

高品質臨床データ、関連リソース・試料のネットワーク強化

→ 製薬・バイオテク・情報に関するアカデミア研究の振興

→ 情報、計測、疾患研究の世界拠点へ

## 公民学連携

産業基盤の構築、雇用創成

本事業の公民学連携による事業化

→ 本事業終了後の恒久的維持

## 人材育成

先進データを用いたOJT

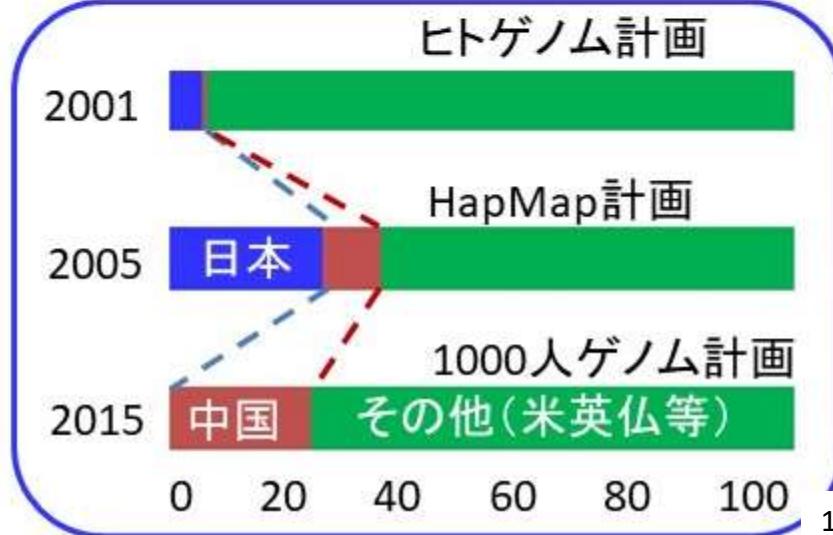
博士課程学生教育

社会人recurrent教育

→ 将来の該当分野を担う人材の創出と循環

ゲノム医科学研究基盤  
(ゲノム研究の力ミオカンデ)

国際ゲノム研究への  
日本の貢献の復活へ

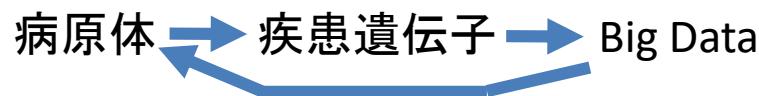


# 高高度解析技術の応用による次世代ゲノム科学基盤の創出



# 生物医学・情報学から発する人類社会のパラダイムシフト

## 医学生物学のパラダイム・シフト



- 生体内の分子の時空間的状態のすべてを計測。
- 情報 (Big Data) として抽出して生命像を解明。
- ヒト疾患の解明による予防・診断・治療に直結。



遺伝子/ゲノム



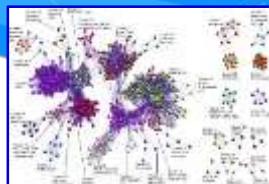
分子



細胞/組織



臓器ネットワーク



個体



社会



すべての分子群の位置、反応、相互作用の  
計測、情報化、解析



計算機技術/AIを駆使したシミュレーションと再構築



生物としてのヒトの完全なる理解

# http://kero.hgc.jp

DBKERO  
Database of Kuroda Encyclopedia for Researchers of multi-Omics data

## Quick-start: For the beginners

使い方の説明

The screenshot shows the DBKERO homepage with a red box highlighting the 'Quick-start: For the beginners' section. This section contains links to help videos for various tools: Data Portal, TSS-Viewer, Multi-Omics-Viewer, Mutation Enriched Genes, TF Binding Site Search, Pathway Map, Chromatin-status Data Summary, and RDF gate (Trial). Below this, there are sections for Featured Dataset (Single cell dataset, Cancer SV dataset, Visium dataset), Download, Tools, Documents (Experimental Procedures, Data Contents, Tutorial, Download bulk data, References), and Movie (Overview movie 1 and 2 in Japanese, How to use genome browser (Basic), How to use genome browser (Track selection), How to use genome browser (Comparative viewer), and How to use data portal).

Tutorialの動画  
(それぞれの  
ブラウザについて)



# 全体のねらい

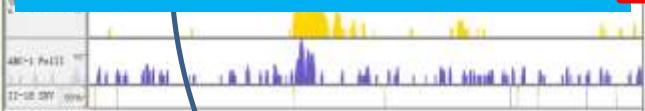
# ヒト応用研究を志向したオミクス情報の統合

## 転写開始点/トランскリプトーム情報 (TSS/RNA Seq)



## クロマチン情報(ChIP Seq)

## DNAメチル化情報 (BS Seq)



## 多層オーミクス情報の ヒトゲノム変異情報への統合

## オミクス新解析技術

- Single cell 解析
- Long read 解析

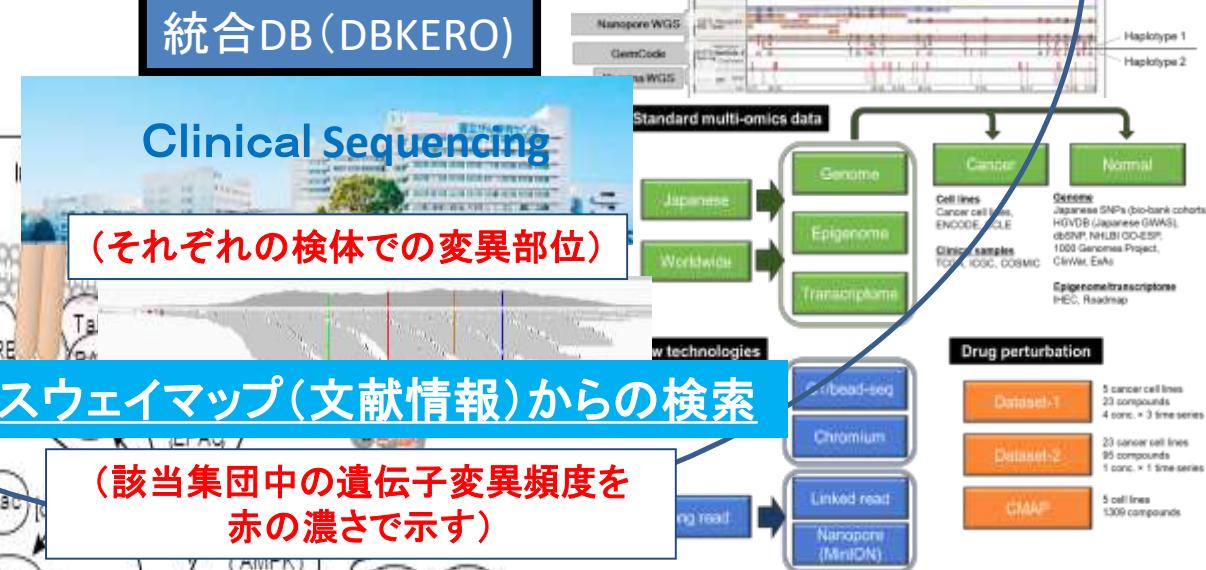
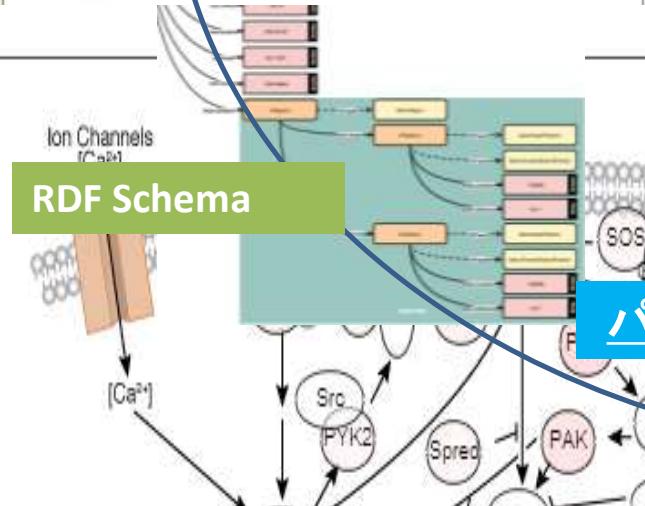
## 統合DB (DBKERO)

## 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)



何ができるか(1)

日本人正常オーミクスデータ (IHEC)

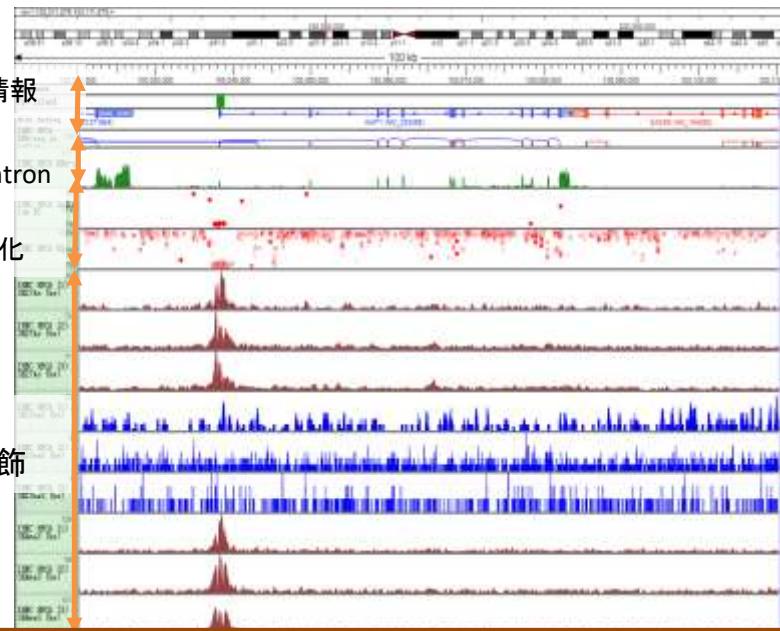
国際エピゲノムコンソシアム (IHEC)

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。

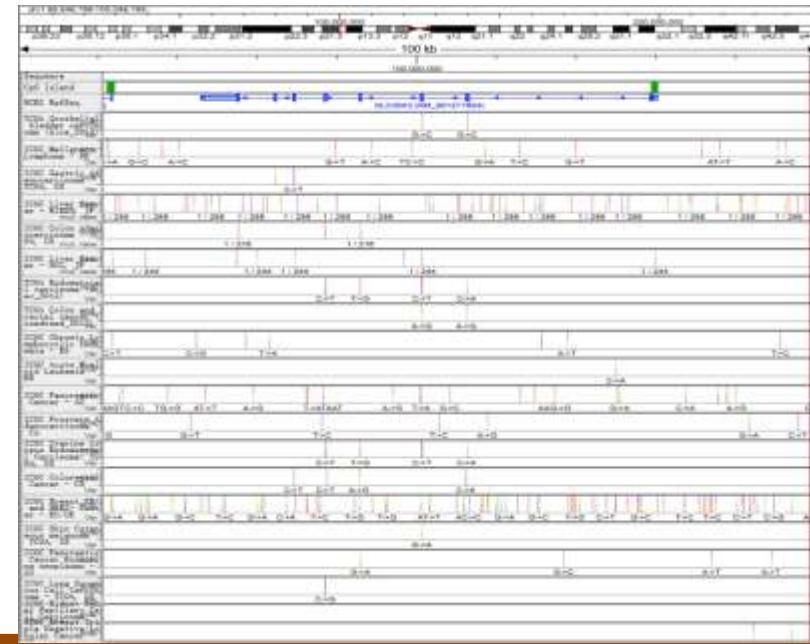
# CREST-IHECデータの統合(日本人標準エピゲノム)

## 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

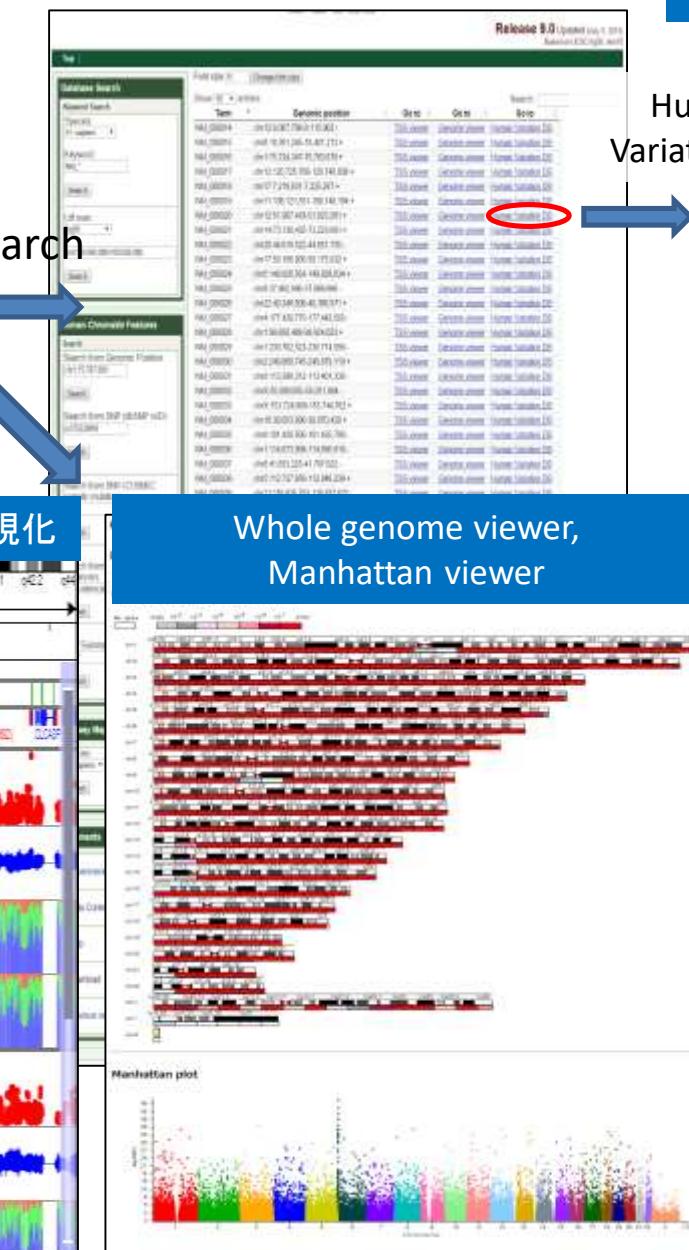
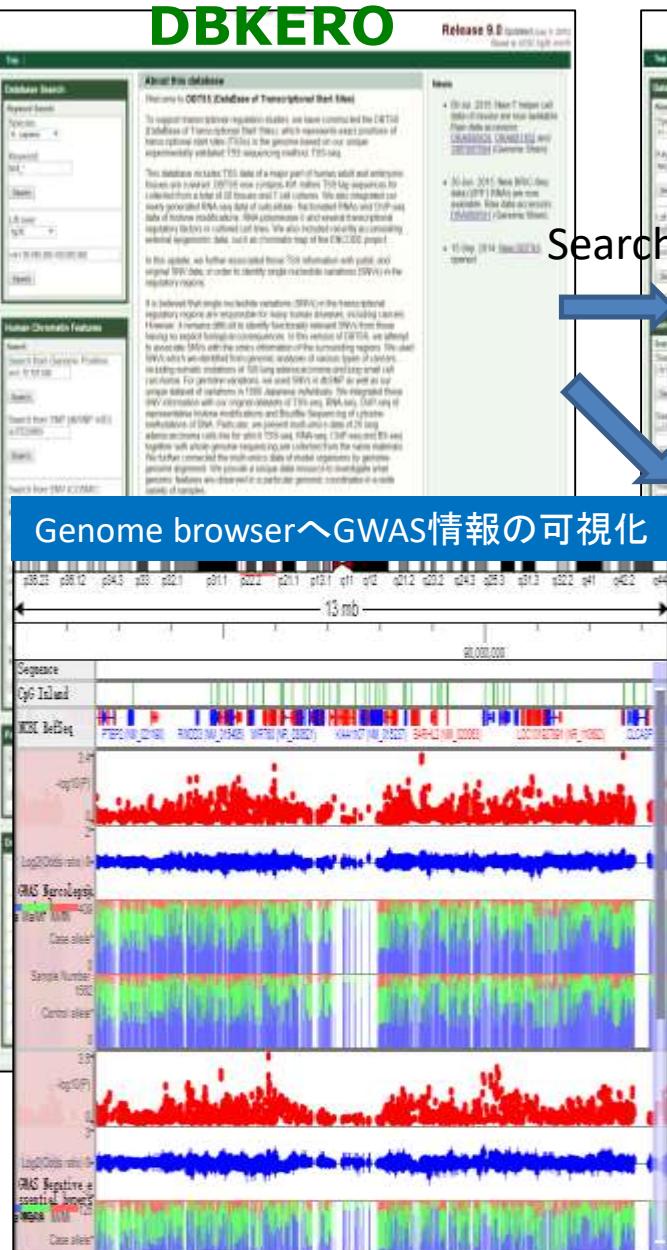
Sample	individuals	Expression		DNA methylation	Histone modifications					
		RNA-seq		BS-seq	H3K4me1	H3K4me3	H3K27ac	H3K27me3	H3K36me3	H3K9me3
		Read depth	Exon/intron junction							
Liver	8	8	8	8	13	13	13	13	13	13
Colon	11	11	11	11	11	11	11	11	11	11
Endometrial	15	30	0	13	14	15	15	15	15	15
Vascular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

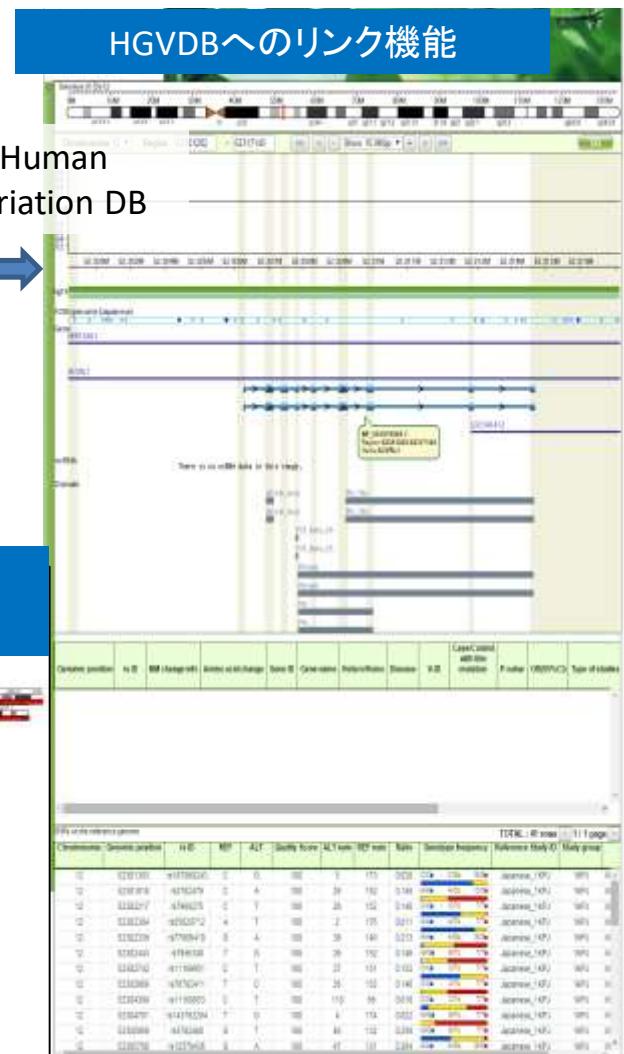
# Human Genome Variation DB (旧徳永グループ)との統合

## 日本人ゲノム多型

DBKERO



HGVDBへのリンク機能



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

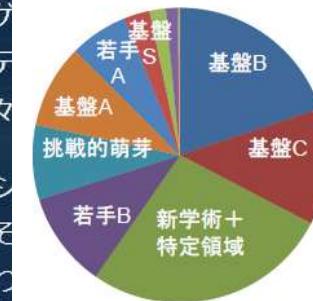
ホーム 先進ゲノム支援とは 公募要項 支援申請の流れ 支援申請 公募支援に関するFAQ お問い合わせ



[ysuzuki@gc.hgc.jp](mailto:ysuzuki@gc.hgc.jp)

新学術領域研究「先進ゲノム支援」は、既存の研究枠組みを超えた、多様な研究者による、多角的なアプローチで、新しい知識や技術を開拓するための研究を促進することにより、生命科学の発展に貢献することを目指す。この過程でゲノム解析システムの開発、データの蓄積、分析手法の確立など、生命科学の基礎となる技術の確立を目指します。この支援にふさわしい、多角的・総合的な研究が実現されるよう、支援体制を構築してまいります。

②支援採択課題の元科研費

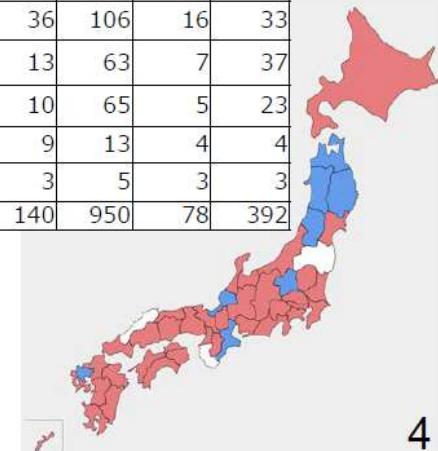


③支援依頼者所在機関 (H26まで)

	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



## 何ができるか(2)

どの培養細胞をモデルにするのか？  
を選ぶオミクスデータ

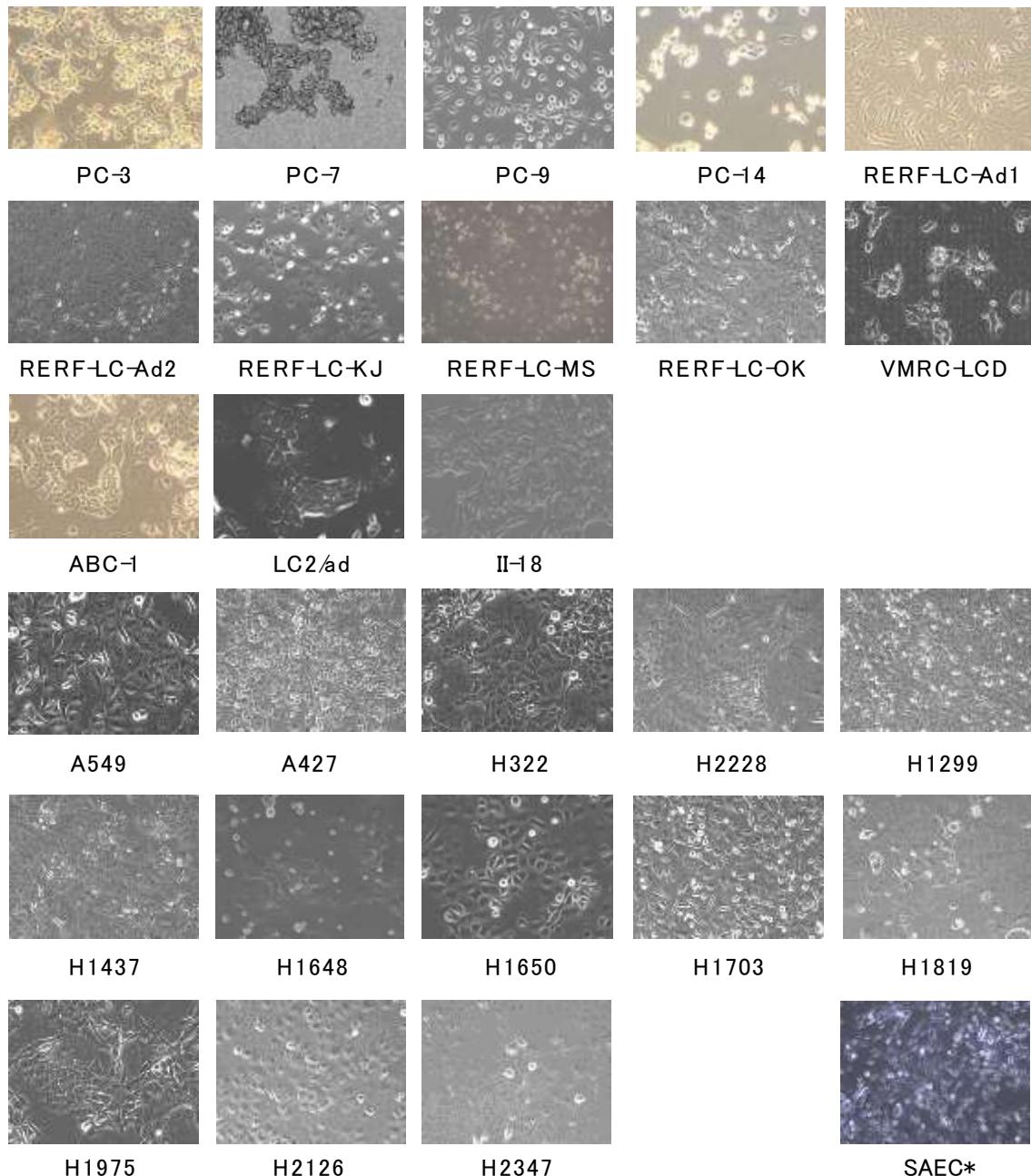
肺腺癌細胞株等で日本人由来のオミクスデータが充実しています。

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



もちろん、ENCODE/CCLEにもあります

\*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

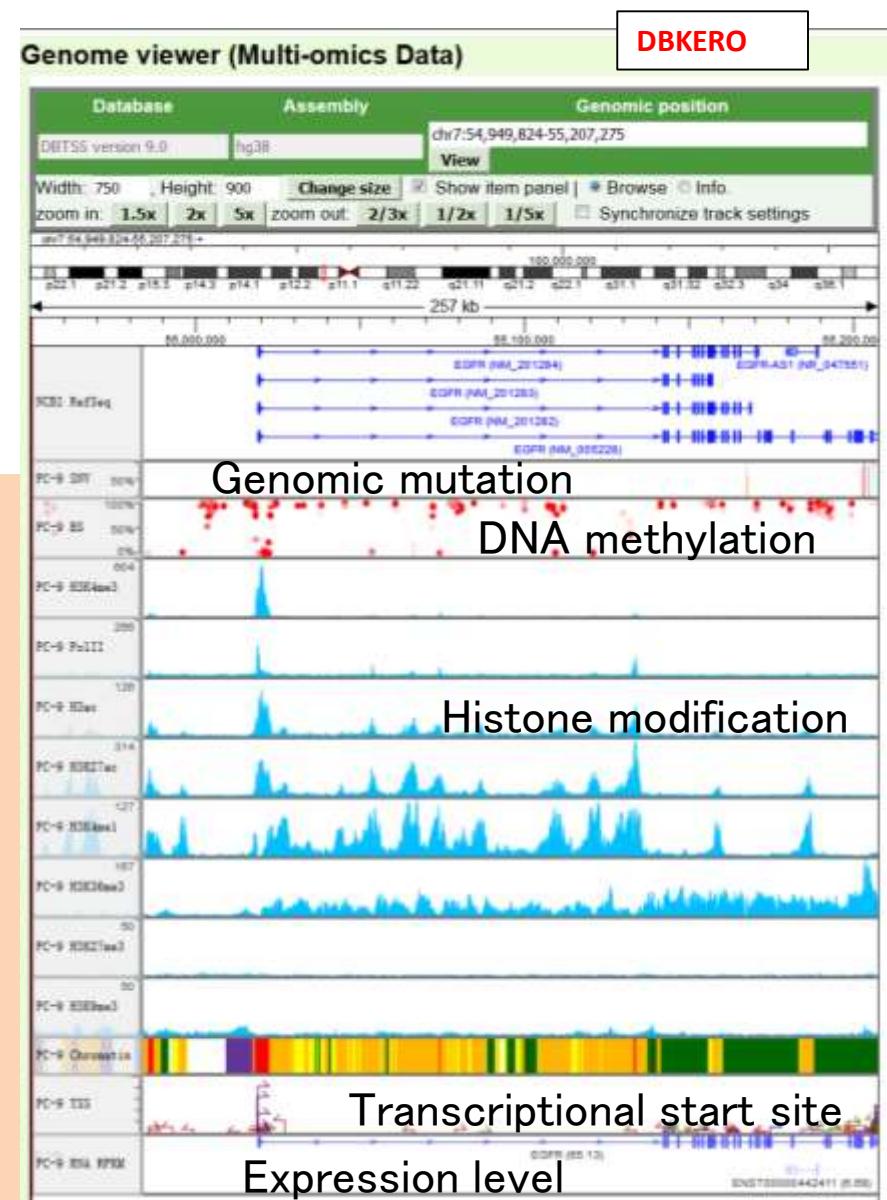
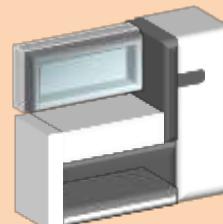
- Whole-genome sequencing
- Long-read sequencing  
(MinION/PromethION)

### Epigenome

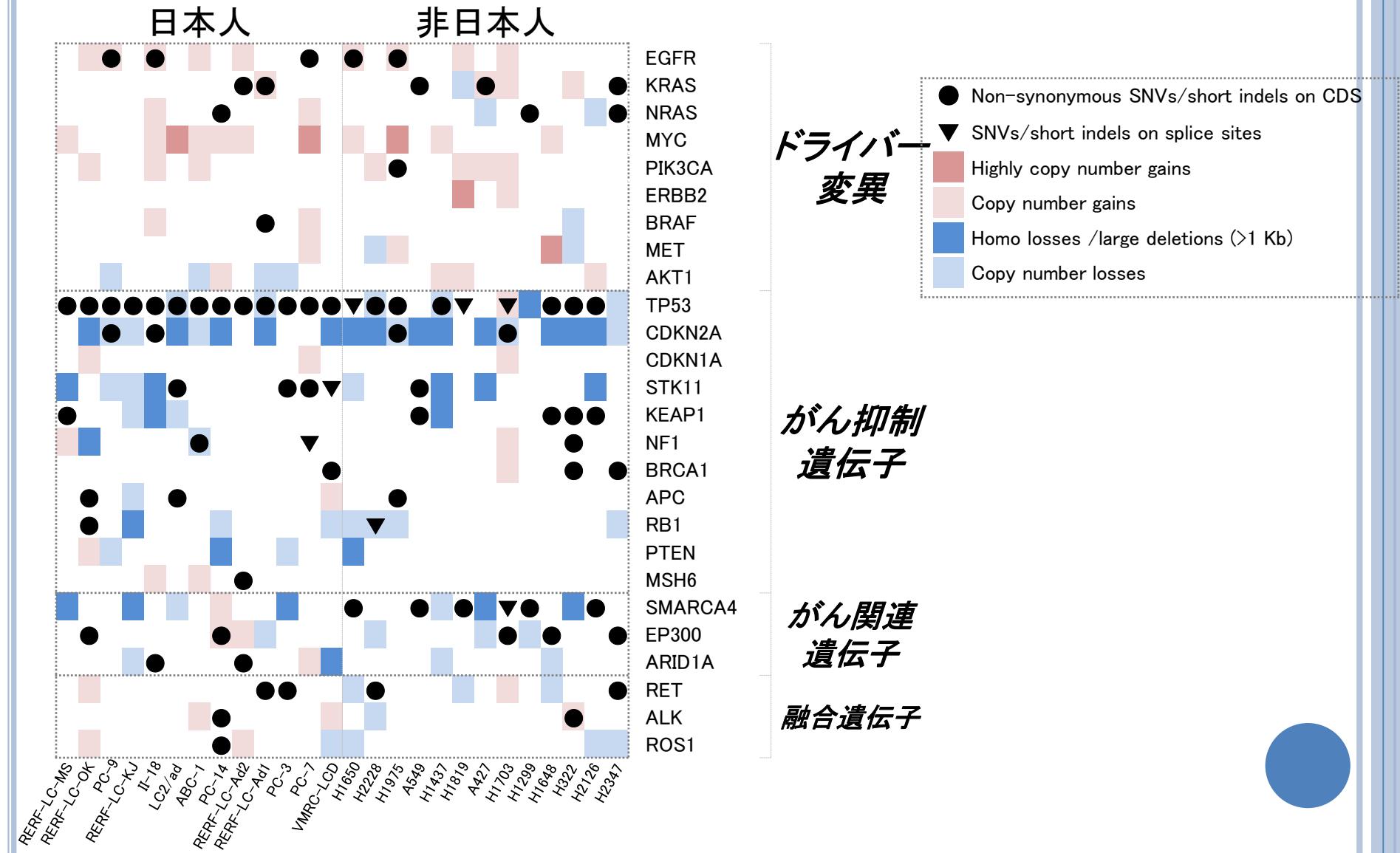
- Bisulfite seq/[EM seq](#) (not yet)
- ChIP-Seq
- ATAC-Seq

### Transcriptome

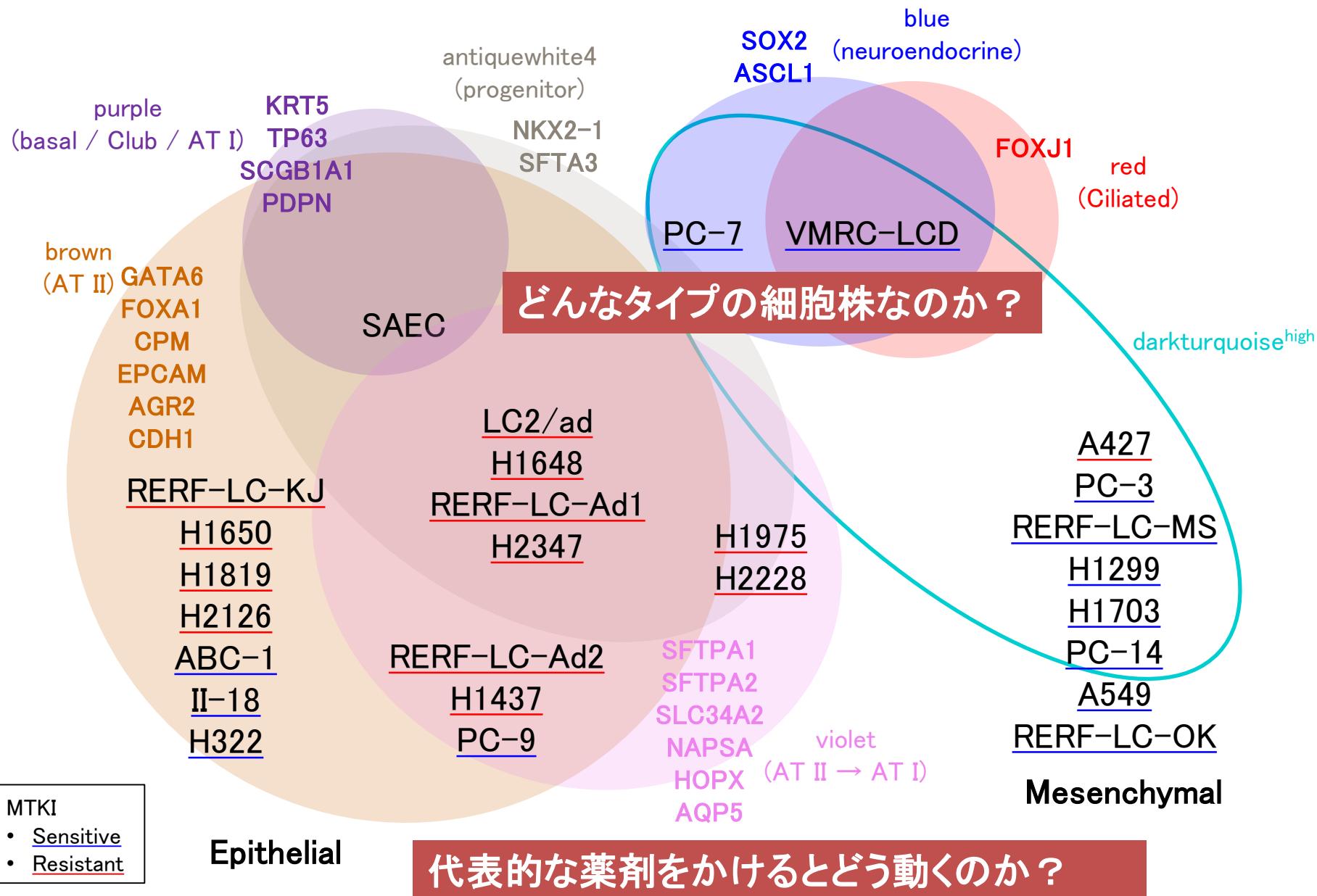
- TSS-Seq
- RNA-Seq
- miRNA-seq



# どんなゲノム変異をもっている細胞株なのか

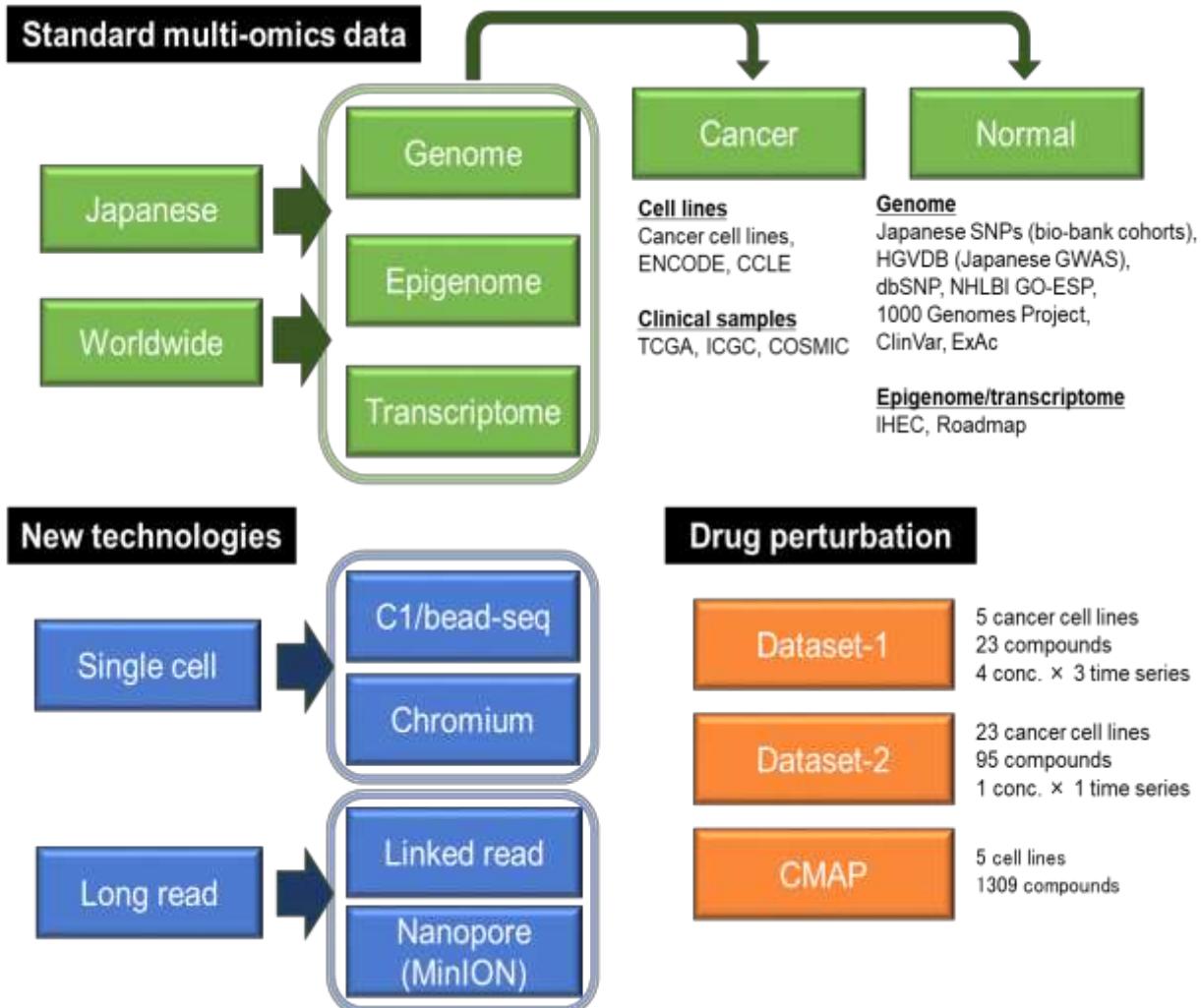


# 肺腺癌細胞株のTranscriptome/Epigenome層別化



# 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擾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.

## 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 sites in a region of genome of various organism. Powered by ChIP-Atlas\* and GIGGenome\*.

## 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

## 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 Overview movie 2 in Japanese (53 min.)



## How to use genome browser (Basic)



## How to use genome browser (Comparative viewer)



## 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  
リアルビデオへリンク

## DBKERO 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 | Tutorial | Multi-Omics-Viewer (Basic)

マルチオーミクスビューフの使い方 (Basic)

## 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

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

DBKEROデータポータルの使い方

Quick-start: For the beginners

Top | Tutorial | 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скри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



Kashiwa Campus,  
Univ Tokyo

Hiseq2500 x 6+ Hiseq3000 x 1  
NovaSeq x 2

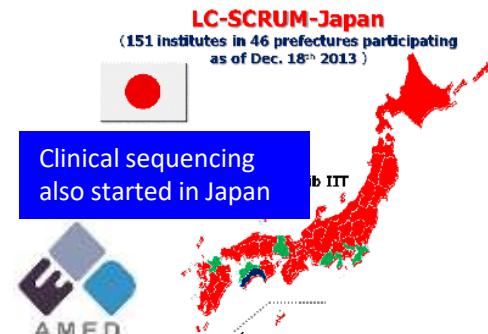


PI: Yutaka Suzuki.  
Univ. Tokyo



Operators:  
Technicians 5  
Programmers 4

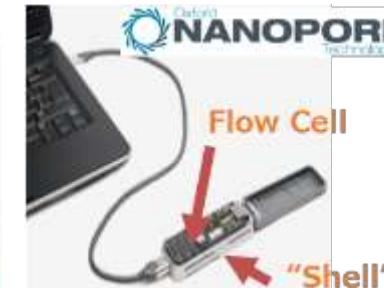
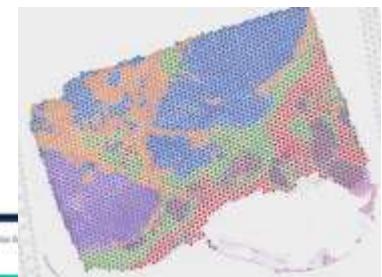
## On-going contributions as a sequence center



And serving as an  
Incubation center for new genome technologies)



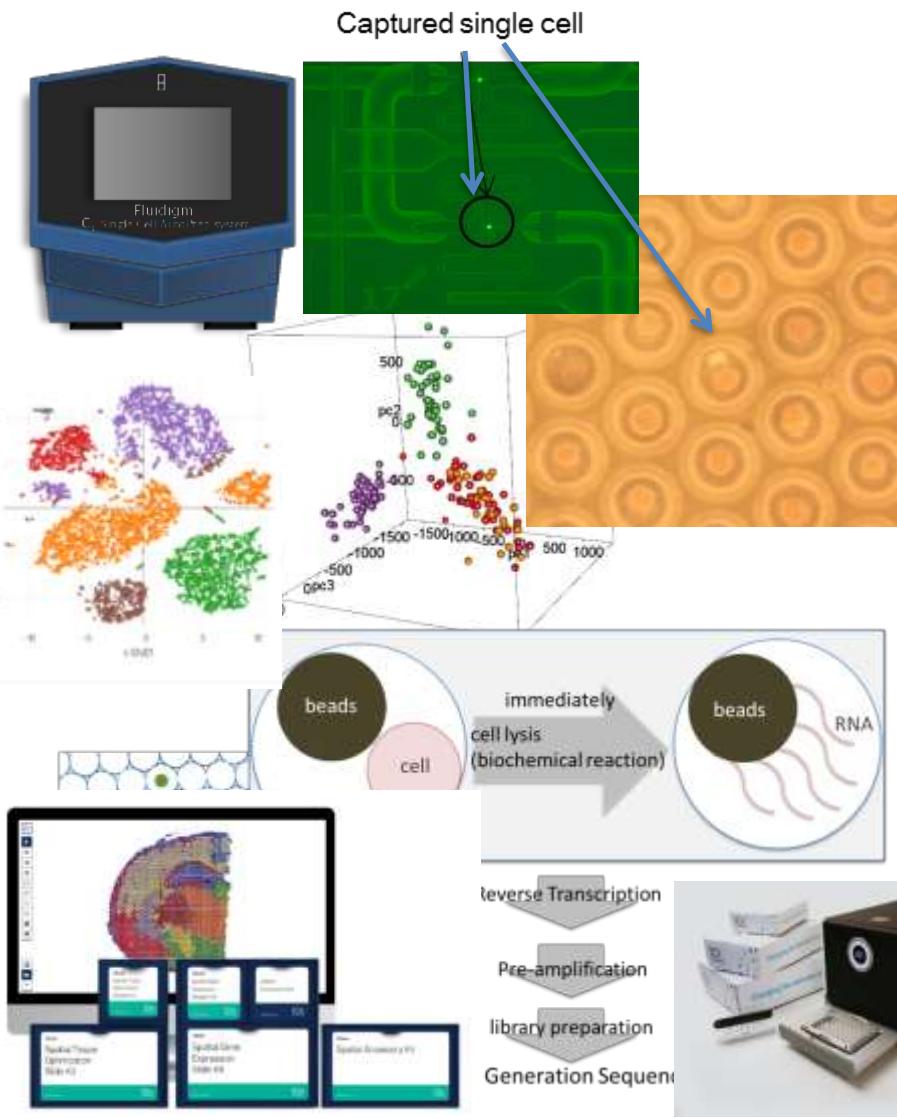
Single cell analyzer;  
C1 (Fluidigm)



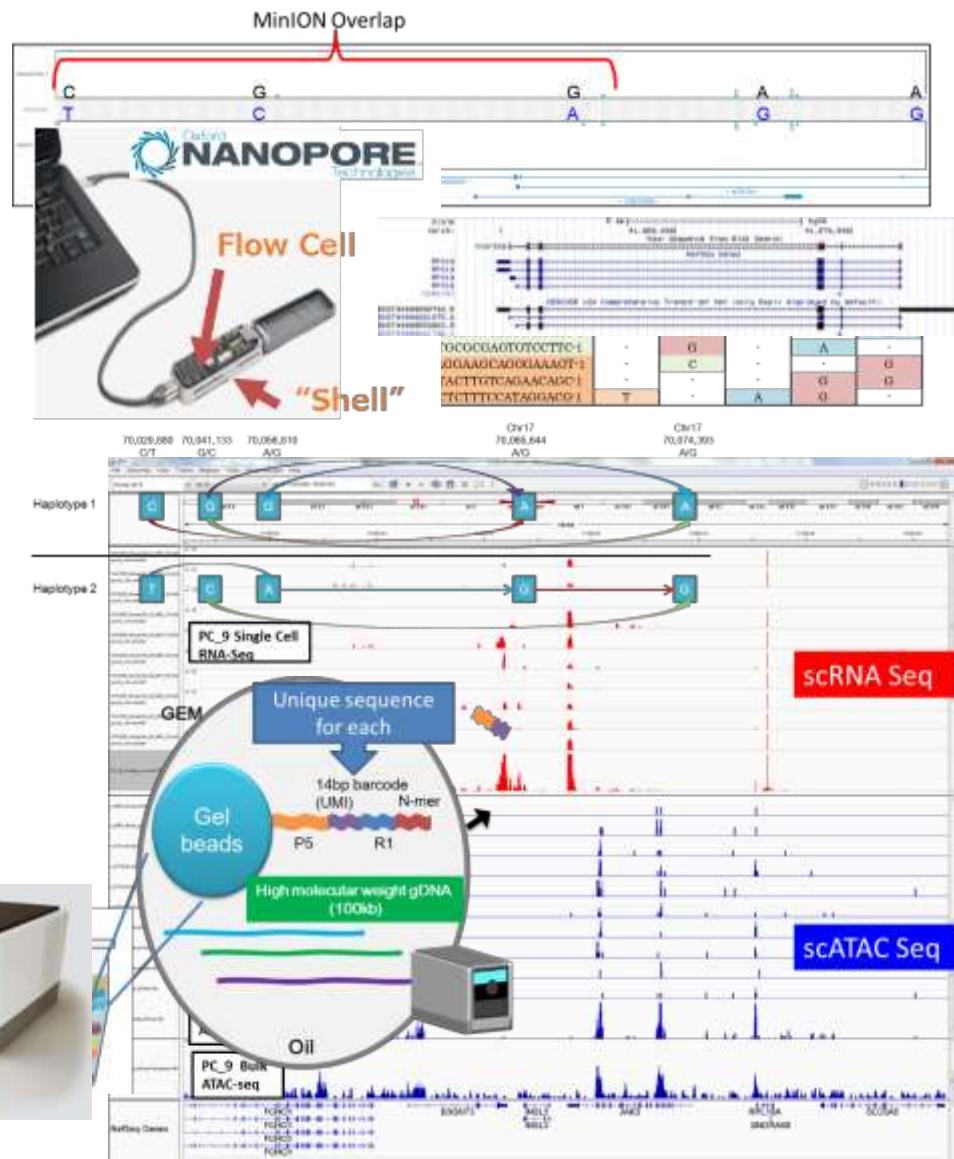
ysuzuki@hgc.jp

# 新しい手法でどんなデータが出てくるの？

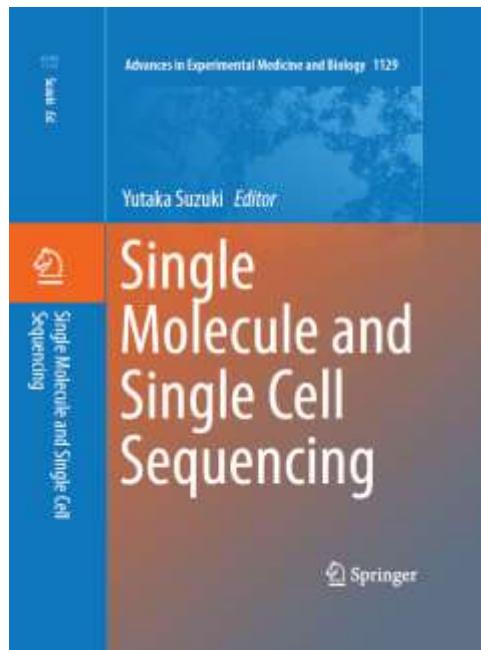
## シングルセル解析(C1, Chromium) + 空間トランскriプトーム解析



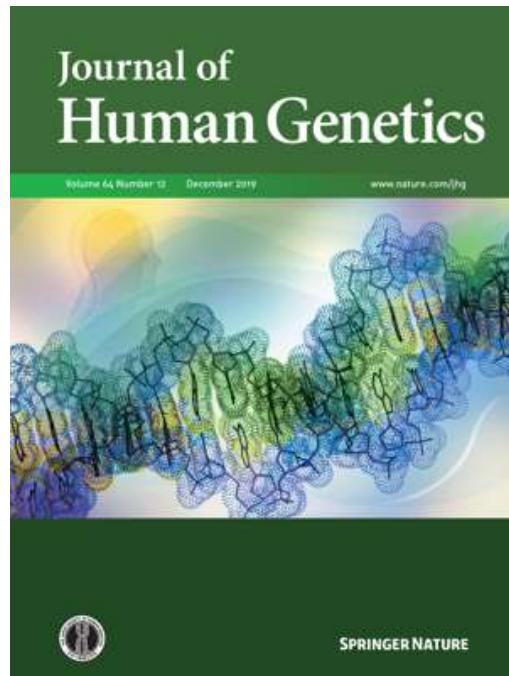
## Long read解析(SV/Phasing)



## 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.)

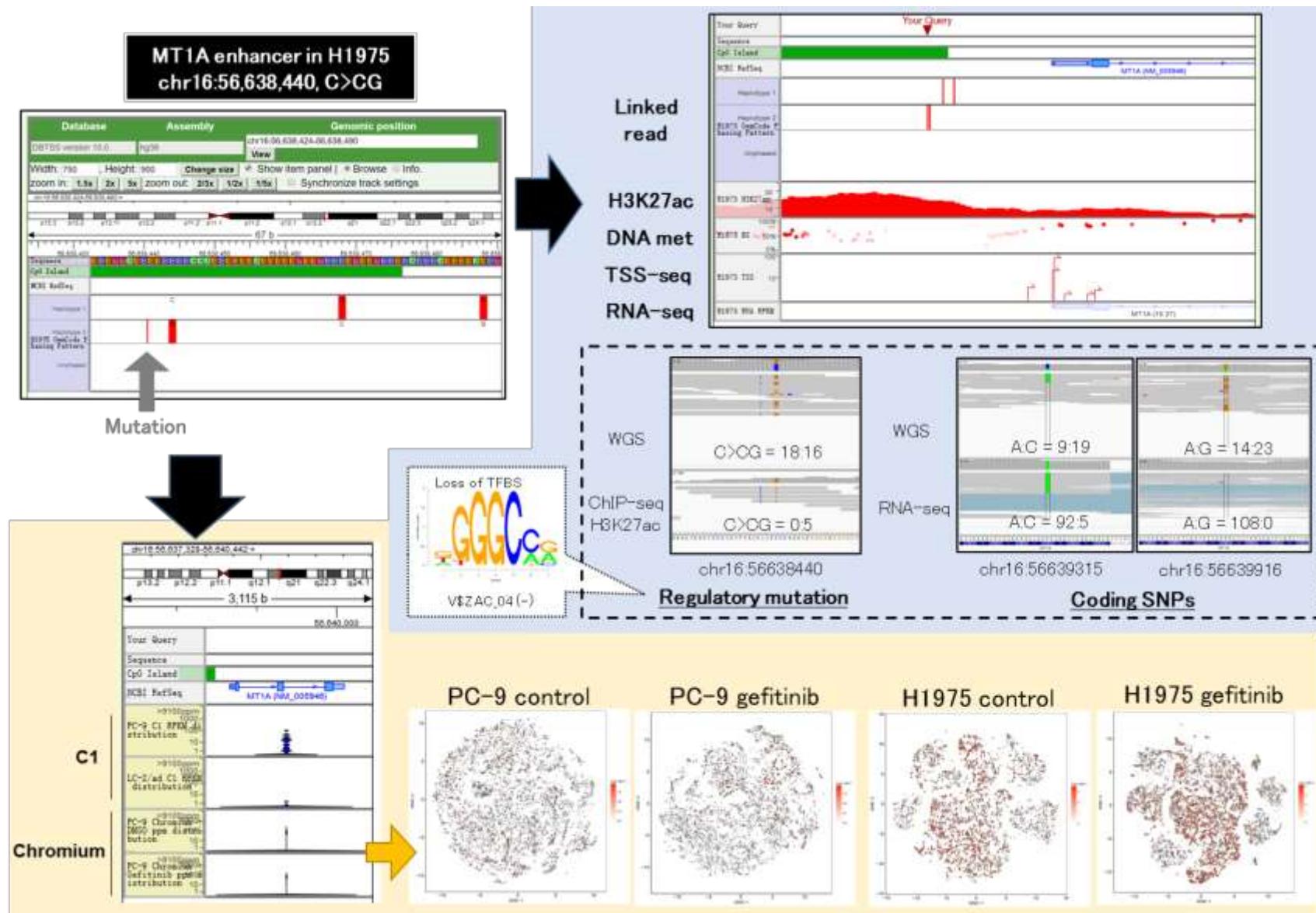


Journal of Human Genetics  
Special issue for “New type Sequencer”  
Jan (2020)  
(Yutaka Suzuki ed.)



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 Keio 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 Explorer any more.

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)

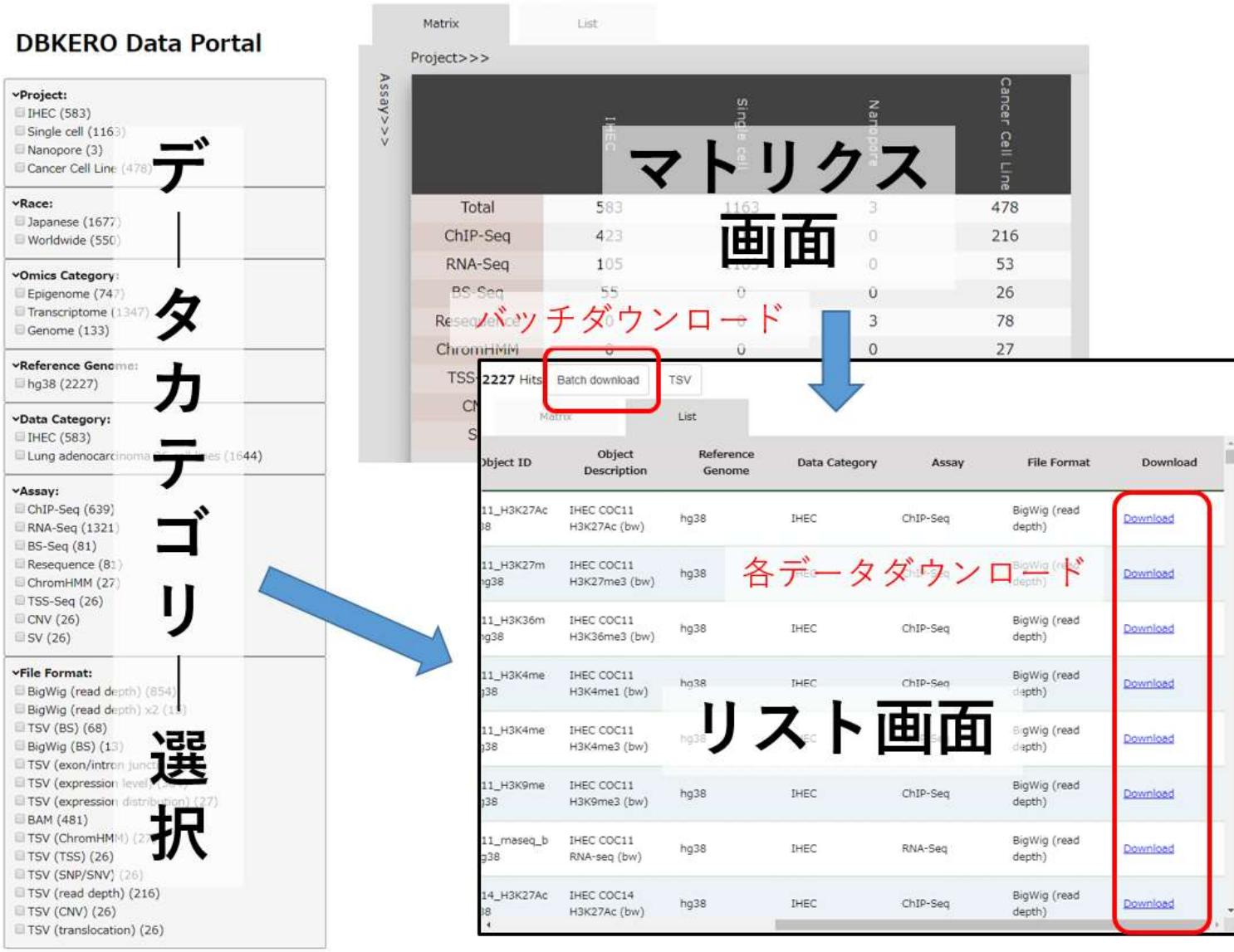
The screenshot shows the DBKERO homepage with several sections:

- Top Bar:** Quick-start: For the beginners
- Left Sidebar:** Single Cell Dataset, Download, Documents (Experimental Procedures, Data Contents, Tutorial, Download bulk data, References).
- Middle Section:** 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.
- Right Sidebar:** Long Read Data (Japanese), Mutation Enriched Genes, TF Binding Site Search, Pathway Map, Chromatin-status Data Summary, SNV Summary in Cancers, RDF gate (Trial), Other tools (RDF Schema, RDF Browser, SPARQL Endpoint), Single cell viewer.

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)

Sakamoto et al  
Genome Res 2029

# DBKEROデータポータルの構築



むしろ手元でブラウズしてもらうのがいいような…

# 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学研究科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

```
$ login -l suser@hgc.mrcs.ac.jp
```

Rをご使用する場合は、R/3.6以上と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

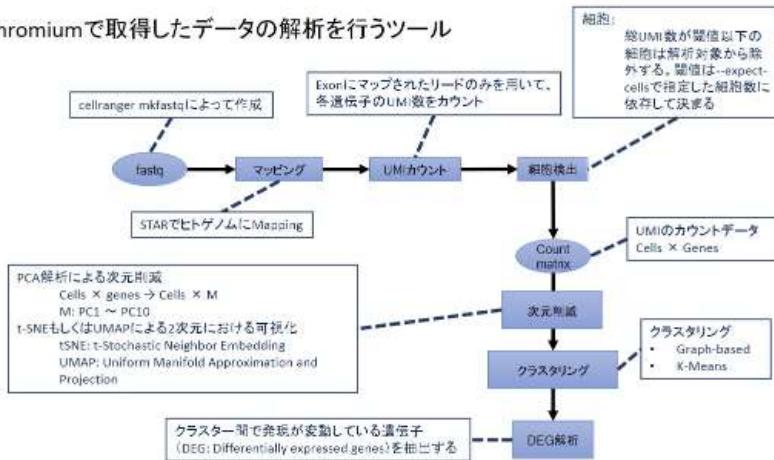
ここではマウス肺臓組織のscRNA-seqデータと、解析ツールとしては10x G

以下の操作について説明します。

1. データの準備・ソフトウェアの準備
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/whatis-cell-ranger>

## データの準備・ソフトウェアの確認

10x Genomics社のダウンロードのページより、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。

```
$ ringer
$ ringer
$ upload
$ upload
$ sitecheck
$ help
```

Prepare a reference for use with 10x analysis software. Requires a GTF and FASTA  
Upload a summary of an analysis pipeline job to 10x Genomics support  
Collect Linux system configuration information  
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耕研究室でChromiumプラットフォームを用いて準備したマウス肺臓组织scRNA-seqの配列データ(FASTQ)で解析を行います。データセットは[https://keru.hgc.jp/tutorials/learning/data/10X\\_RNAv3\\_Lung.tar](https://keru.hgc.jp/tutorials/learning/data/10X_RNAv3_Lung.tar)からダウンロードできます。

```
$ #マウス肺臓細胞 Chromium scRNA-seq配列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://keru.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。

①適当な作業ディレクトリを作成し移動

ラストリング、発現度割合(%) (DEG) 前折りコマンド)を実行し

2. フォルダ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: 前折に使うCPUコア数に自身の環境に合わせてなるべく大きな数にしてください)
$ #--localmem: 使用メモリ量(GB)に自身の環境に合わせてなるべく大きな数にしてください)
$ cellranger count --id=scrnaseq --transcriptome ./data/refdata-gex-mm10-2020-A/ --fastqs ./data/10X_RNAv3_Lung/ --expect-cells 10000
$ #... 前折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/scrnaseq/
$ ls
SCRNA_COUNTER_CS_filelist _imputation _log _perf _tags _uid _versions _outs
imeline _finalstate _jobmode _resource _sitecheck _timestamp _wdkill emrusage.mri.tgz
$ #outsに実行結果ファイルが出来されます
$ cd outs
$ ls
analysis filtered_feature_bc_matrix.h5 posorted_genome_bam.bai raw_feature_bc_matrix.h5
oloupe.oloupe metrics_summary.csv posorted_genome_bam.bai.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output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。
posorted_genome_bam.bam	BAMファイル、IGVで閲覧可能。
posorted_genome_bam.bam	BAMファイル。
posorted_genome_bam.bam.ba	BAMファイル。

**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.

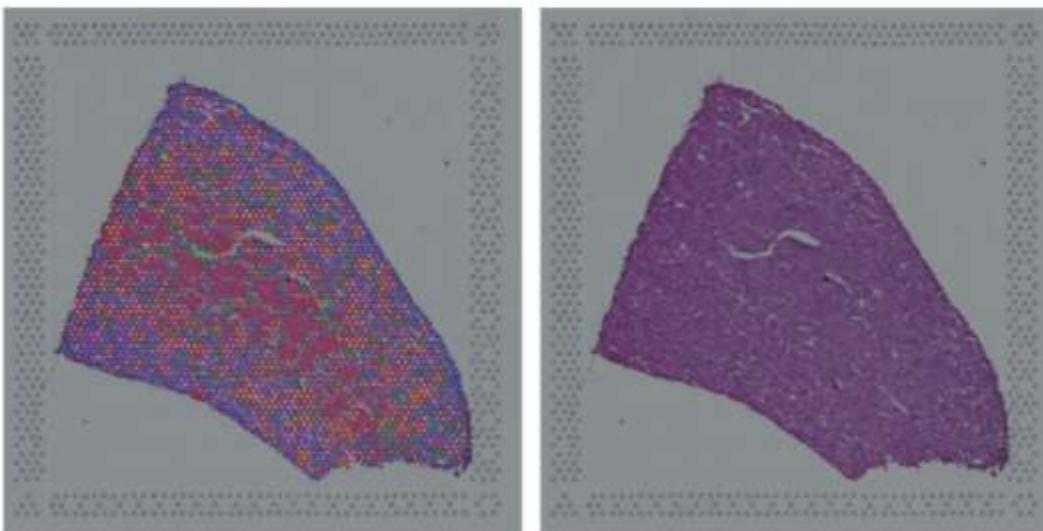
These datasets are processed by spaceranger v1.0.0 (refdata-cellranger-mm10-3.0.0).

- [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



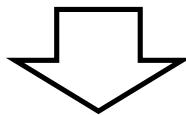
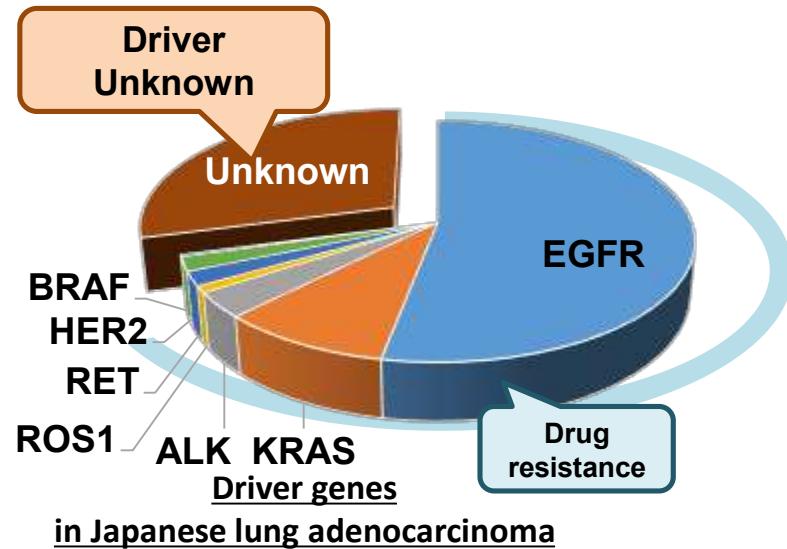
# Lung cancer

## Clinical Sequencing for cancer patients

Lung adenocarcinomas have been classified by so-called **driver genes** which can be targets for molecular targeted medicine.

### Problems

- Driver unknown cases (30%)
- Occurrence of resistance to drugs



**Long-read sequencing of cancer genomes to identify novel genomic alterations in cancer-related genes**

Quick-start: For the beginners

We recommend 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.

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.

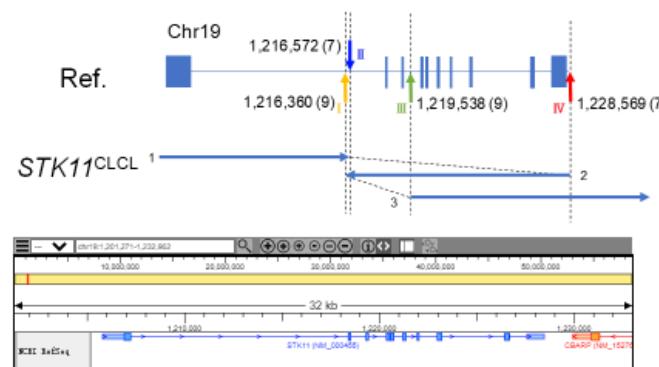
Reference: Long-read sequencing for non-small-cell lung cancer genomes

Accession: JGAS000000000065 (JGAD00000000252 / JGAD00000000253)\* Controlled Access.

sampleID	Tumor			Normal			Processed data
	Yields(Gb)	Number of reads	Coverage(x)	Yields(Gb)	Number of reads	Coverage(x)	
S1	99	14861240	33	57	11709609	19	S1_SV_gene_candidates.bedpe
S2	94	25379061	31	41	7016159	14	S2_SV_gene_candidates.bedpe
S3	77	15952312	25	48	12971173	15	S3_SV_gene_candidates.bedpe
S5	82	21640571	27	35	6503970	11	S5_SV_gene_candidates.bedpe
S6	76	11336329	25	48	7531078	16	S6_SV_gene_candidates.bedpe
S7	85	13880384	28	46	7555316	15	S7_SV_gene_candidates.bedpe
S8	100	16661003	33	54	8351411	18	S8_SV_gene_candidates.bedpe
S9	98	13179209	32	34	4483001	11	S9_SV_gene_candidates.bedpe
S10	85	12424087	28	42	7029224	14	S10_SV_gene_candidates.bedpe
S11	69	21140772	22	59	7832201	19	S11_SV_gene_candidates.bedpe
S12	73	15634853	24	84	12908975	27	S12_SV_gene_candidates.bedpe
S13	104	30719862	34	38	5506152	12	S13_SV_gene_candidates.bedpe
S14	74	14328017	24	56	6464828	18	S14_SV_gene_candidates.bedpe
S15	75	12278084	24	55	9585959	18	S15_SV_gene_candidates.bedpe
S16	74	20118226	24	55	7924757	18	S16_SV_gene_candidates.bedpe
S17	52	7066616	17	59	6766398	20	S17_SV_gene_candidates.bedpe
S18	62	7091142	20	44	5335320	15	S18_SV_gene_candidates.bedpe
S19	60	6352847	20	37	5025576	12	S19_SV_gene_candidates.bedpe
S20	58	5838744	19	47	5620346	15	S20_SV_gene_candidates.bedpe
S21	63	8985953	21	57	6166961	19	S21_SV_gene_candidates.bedpe

Viewer (Cell line data)

RERF-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

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

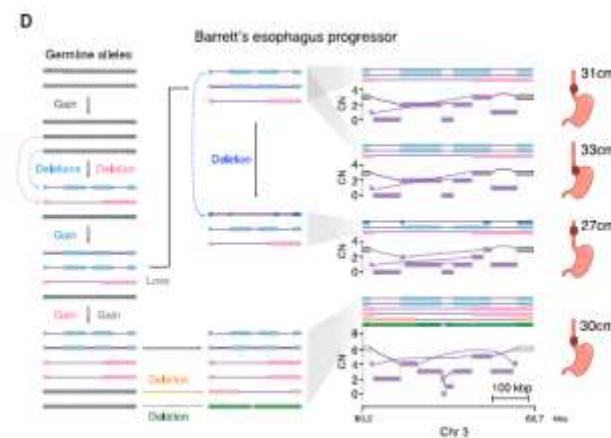
Yoshitaka Sakamoto,<sup>1</sup> Liu Xu,<sup>1</sup> Masahide Seki,<sup>1</sup> Toshiyuki T. Yokoyama,<sup>1</sup> Masahiro Kasahara,<sup>1</sup> Yukie Kashima,<sup>2,3</sup> Akihiro Ohashi,<sup>3</sup> Yoko Shimada,<sup>4</sup> Noriko Motoi,<sup>5</sup> Katsuya Tsuchihara,<sup>2</sup> Susumu S. Kobayashi,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

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



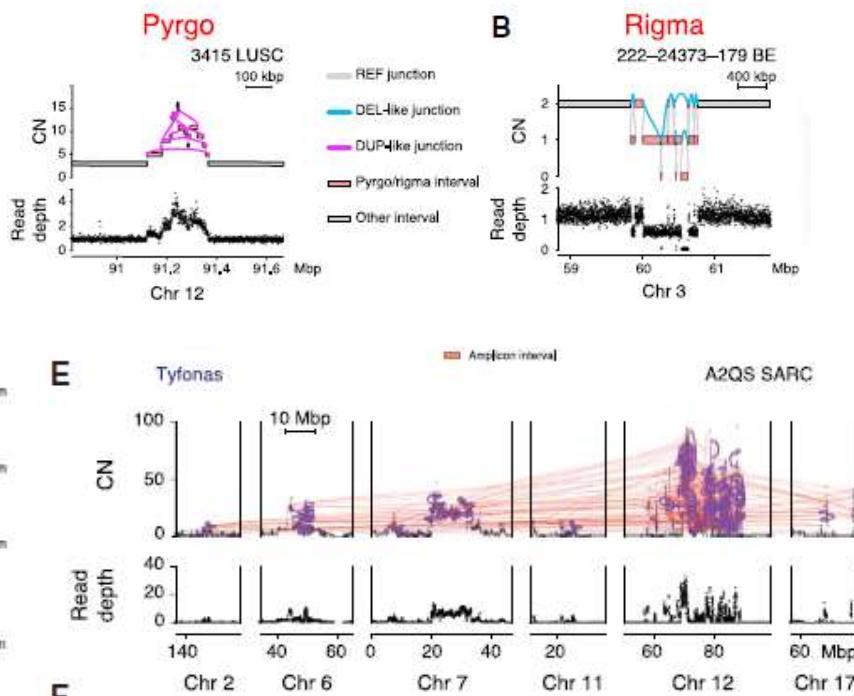
## Ayako Suzuki (Associate Prof.)



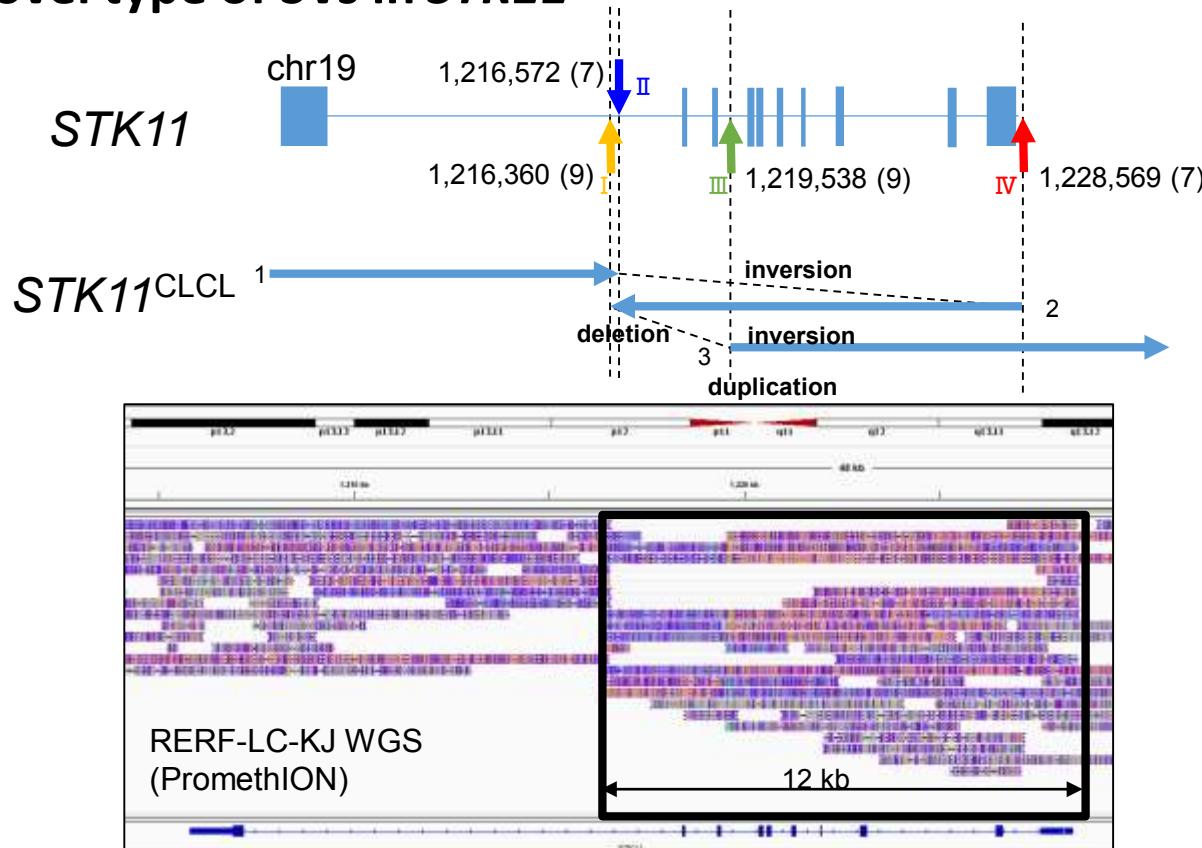
Cell

# **Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs**

Hadi et al., 2020, Cell 183, 197–210  
October 1, 2020 © 2020 Elsevier Inc.



## A novel type of SVs in *STK11*



Inversion + deletion + duplication → Partial copy number gain

**CLCL: Cancerous Local Copy-number Lesion**

# Phasing Analysis of Cancer Genomes

In tumor (small) HS158 PTENを含むPhase block

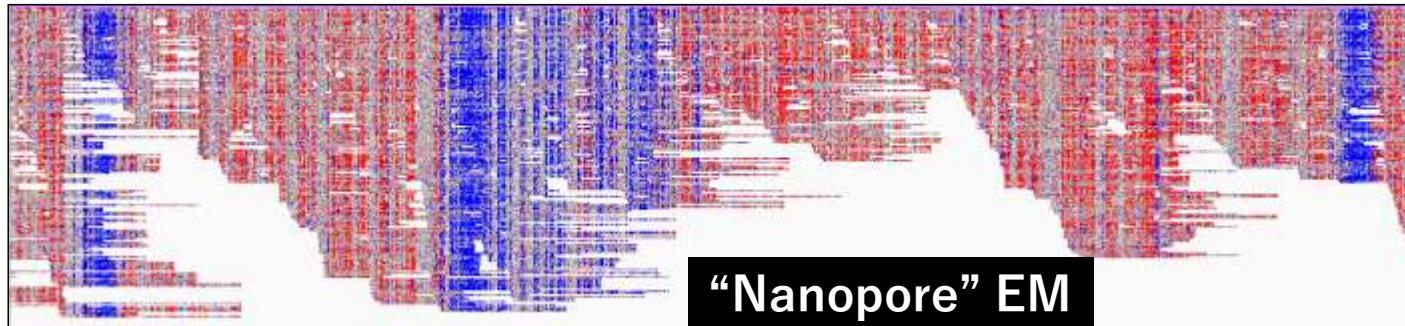


## Long read Methylation Analysis of the *ERBB2* gene

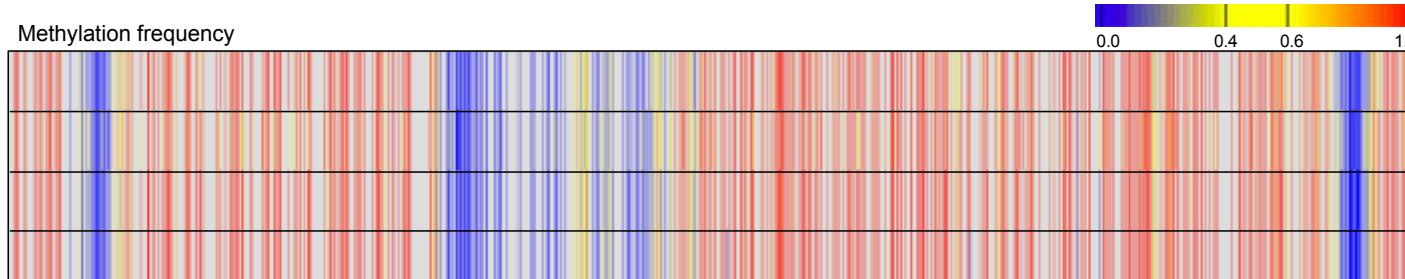
Chr17:39,684,824-39,732,764

Ave.= 5kg long/read

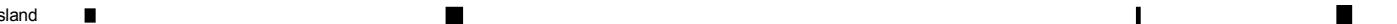
Unmethylated CpG  
Methylated CpG



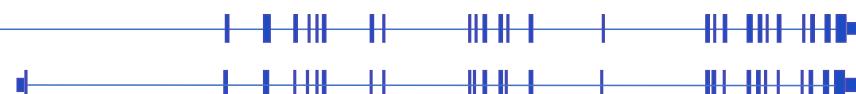
Masahide Seki  
(Assistant Prof.)



CpG island



47 kb



*ERBB2*

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](mailto: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



PAGS  
Platform for Advanced Genome Science

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/>

# 国内総説を用いた広報活動

遺伝子医学34号(2020)



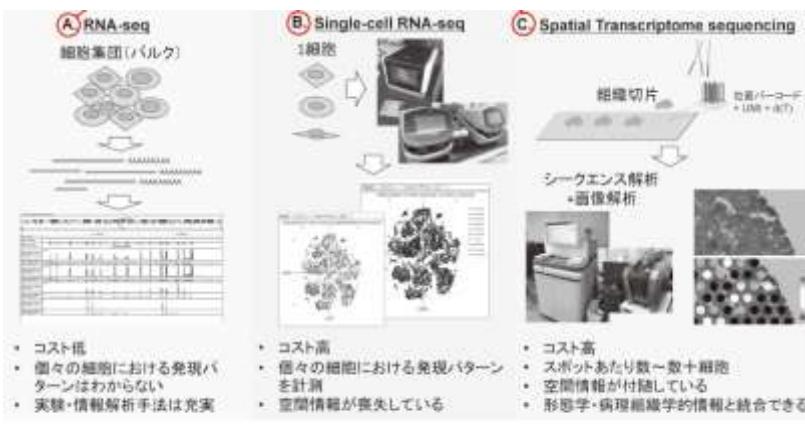
生体の科学(2020)



その他出版中

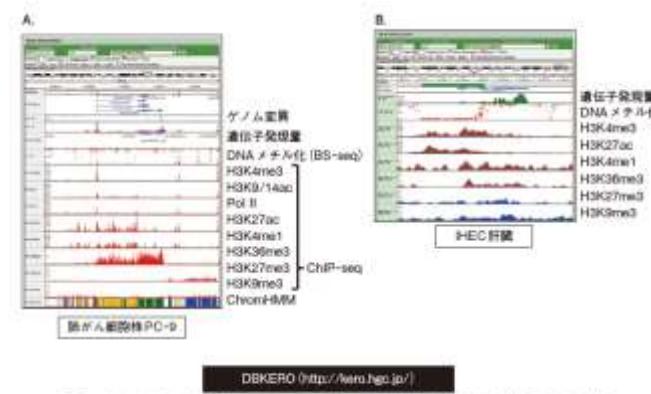
遺伝子医学35号(2021)

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скルプトーム解析の概要をまとめている。図中のゲノムデータおよび t-SNE プロットは [癌遺伝子オミクスデータベース DBKERO \(https://kerogenomics.jp/\)](https://kerogenomics.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テータスヒト結合マルチオミクスデータベース DBKERO (<http://kerogenomics.jp/>) に収載されている多層オミクスデータ。腫瘍細胞株 PC-9 における EGFR 遺伝子領域と、その周辺のオミクス状況を表示した(A)。また、HEK293T のエピゲノムデータ(肝臓)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>

初めての方は、 “For the first users”

そもそも  
データセット例  
(最新！)



The screenshot shows the DBKERO homepage with several sections:

- Featured Dataset:** Includes "Single cell dataset" (Download processed data), "Cancer SV dataset" (New), and "Visium dataset" (New).
- Documents:** Lists "Experimental Procedures", "Data Contents", "Tutorial", "Download bulk data", and "References".
- Movie (For Japanese):** Shows two video thumbnails: "Overview movie 1 in Japanese (9 min.)" and "Overview movie 2 in Japanese (53 min.)".
- Tools:** Includes "Data Portal" (Help video (Japanese)), "TSS-Viewer" (Help video (Japanese)), "Multi-Omics-Viewer" (Help video 1,2,3 (Japanese), GitHub), "Mutation Enriched Genes" (Help video (Japanese)), "TF Binding Site Search" (Help video (Japanese)), "Pathway Map" (Help video (Japanese)), "Chromatin-status Data Summary" (Help video (Japanese)), "SNV Summary in Cancers" (Help video (Japanese)), "RDF gate (Trial)" (Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq), and "Other tools" (RDF Schema, RDF Browser, SPARQL Endpoint, Single cell viewer).
- Tutorial:** Shows "How to use genome browser (Basic)", "How to use genome browser (Track selection)", "How to use genome browser (Comparative viewer)", and "How to use data portal".

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

(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。