

2018.12.05 愛知県がんセンター
統合データベース講習会 : AJACS名古屋2

ヒト統合オームクスデータベース DBKEROとがんゲノム解析

東京大学 大学院新領域創成科学研究科
メディカル情報生命専攻
鈴木 紗子

大規模シークエンス技術による 肺がんの多層オミクスシークエンス解析をやっています



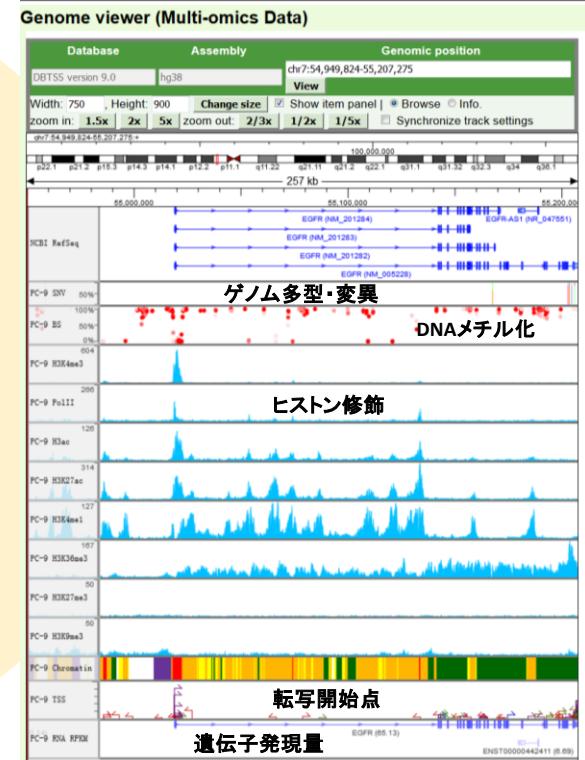
ゲノム
エピゲノム
トランスクリプトーム

多層オミクス解析

シングルセル解析

ロングリード解析

2010年からずっと
千葉県柏市



教育プログラム(生命データサイエンス人材育成)

DSTEP: data scientist training/education program

実践的オンザジョブトレーニングを核とした“問題解決型”教育
大学の先端技術をキャッチアップする人材育成の協創

2018年度開始 生命データサイエンスプログラム (DSTEP): 企業協創教育プログラム

東大+12社のコンソーシアム



社会連携講座 メディカル情報データサイエンス講座
1) OTJ教育プログラム課題の探索とプログラム運営
2) 国際インターンシップ・社会人大学院生の支援とアドバイス

学外機関との連携

がんセンター
東病院

NBDC柏(柏FC)

臨床応用

情報解析

連携講座

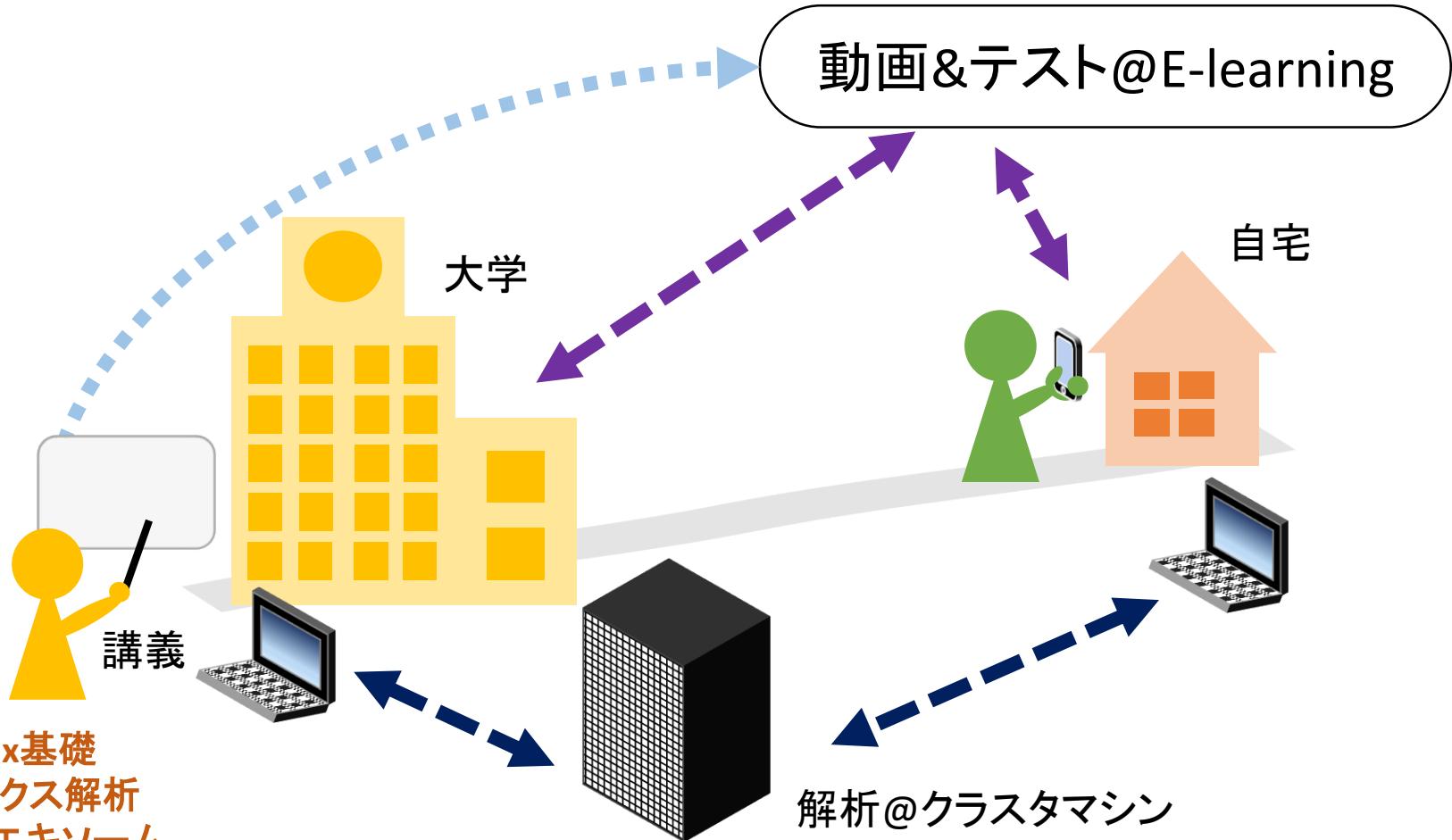
連携講座

東大からの人材輩出/循環実績



企業と一体となっての教育プログラムの実装: 社会人再教育

バイオデータプログラミング演習



Linux基礎
オミクス解析
• エキソーム
• RNA-seq
• ChIP-seq
シングルセル解析
ロングリード解析...

東京大学大学院新領域創成科学研究科 大学院科目等履修生
来年度も募集予定

ヒト統合オーミクスデータベース

DBKERO

データベースDBKERO

DBKERO: DataBase of Kashiwa Encyclopedia for human genome mutations in Regulatory regions and their Omics contexts

The screenshot shows the DBKERO homepage in a web browser. The title "DBKERO" is at the top center, with the subtitle "DataBase of Encyclopedia of Regulatory Omix" below it. A banner indicates "Release 1.1 Updated (Dec. 21, 2017) Based on UCSC hg38_mm10 Formerly DBTSS". A note at the top recommends using Edge (V40), Google Chrome (V61), or Firefox (V56). The main content area is divided into sections: "Tools" (Genome Browser [GitHub], Pathway Map, RDF for NGS Analysis Results (Trial), Chromatin Features), "News" (a list of recent updates), and a large red box at the bottom right containing the URL "DBKERO: <https://kero.hgc.jp/>".

We recommend to use Edge (V40 above), Google Chrome (V61 above) or Firefox (V56 above) for the DBKERO browsing. Internet Explorer has not been supported.

Top |

Tools

- Genome Browser [[GitHub](#)]
 - Search from Keyword or Genomic Position (human and mouse)
 - Search from SNV-enriched Gene in Cancers
- Pathway Map
 - Human Pathway Map
- RDF for NGS Analysis Results (Trial) (Lung adenocarcinoma 26 cell lines: RNA-seq, ChIP-seq, SNV, BS-seq, TSS-seq)
 - RDF Schema
 - RDF Browser
 - SPARQL Endpoint
- Chromatin Features (for Lung adenocarcinoma 26 cell lines)
 - Search from Genomic Position
 - Search from SNP (dbSNP rsID)
 - Search from SNV (COSMIC: somatic mutation)
- Overview of mutation frequency in patients

News

- 27 Jun. 2018: The function of "TF binding site search" was released [here](#).
- 21 Dec. 2017: New KERO (Release 1.1) opened
- 1 Sep. 2017: New C1 data of human lung adenocarcinoma cell line (LC-2/ad: (replicate)) are now available (See browser: Single cell -> C1 -> LC-2/ad (replicate)).
- 25 Aug. 2017: New C1 cell line (LC-2/ad: +Va) browser: Single cell ->
- 18 Aug. 2017: The whole depth data of Lung adenocarcinoma 26 cell lines were updated as bigWig files (See browser: Lung cancer omix -> [cell line (such as LC2/ad)] -> [cell line] wgs (bw)).

DBKERO: <https://kero.hgc.jp/>

ヒト応用研究を志向したオミクス情報の統合

がん細胞株(モデル)の多層オミクスデータと 臨床検体のゲノム変異・エピゲノム異常をリンクさせたデータベース

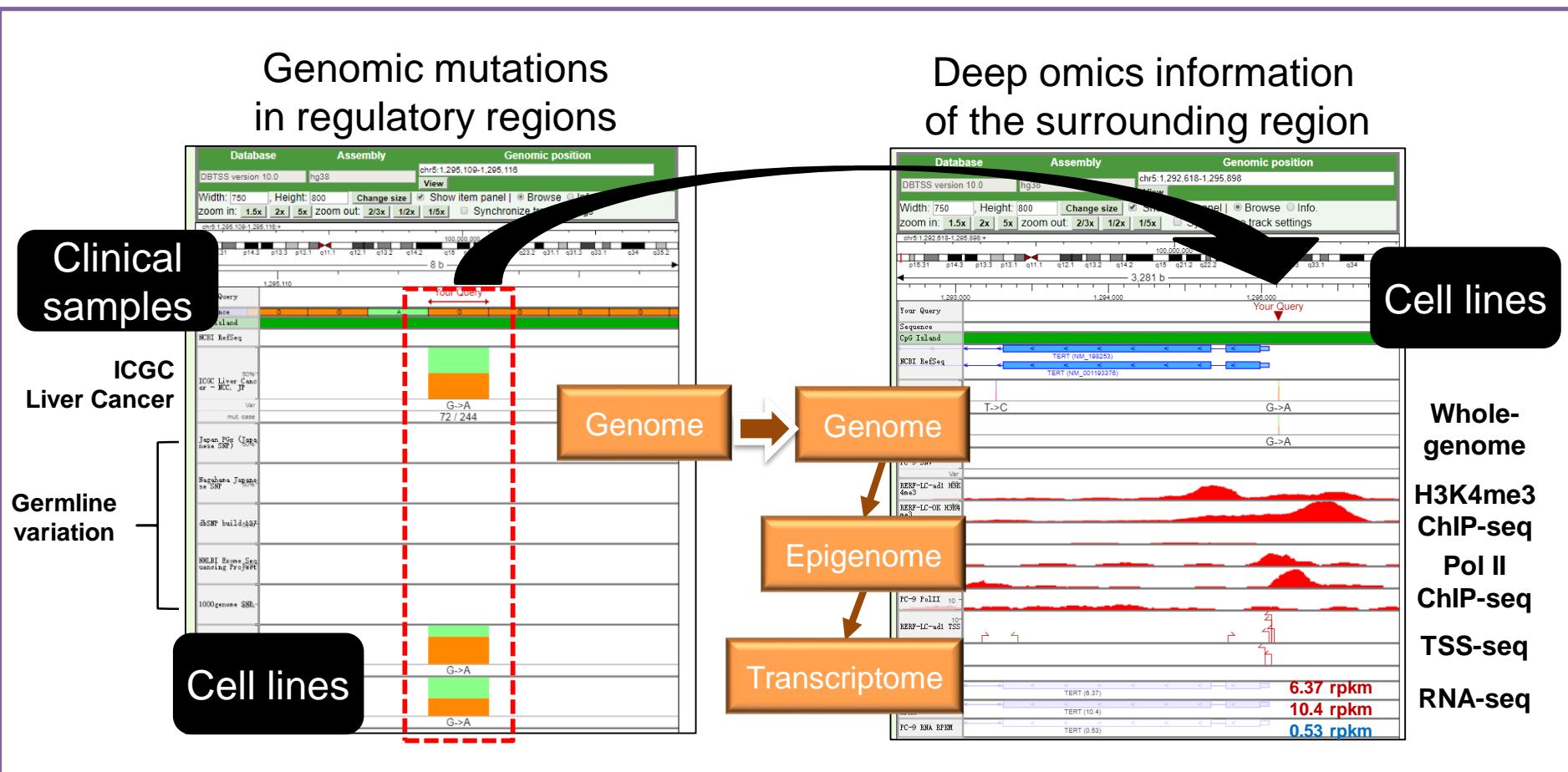


Figure 3 改変: Suzuki, Kawano and Mitsuyama et al. 2018 *Nucleic Acids Research*

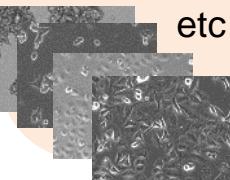
Contents of DBKERO

Standard multi-omics data

Model system

Cancer cell lines

- Whole-genome sequencing
- ChIP-seq/ATAC-seq
- BS-seq
- TSS-seq/RNA-seq
- etc ...



Genome

Epigenome

Transcriptome

Clinical sample

HGVDB Japanese GWAS,
Japanese bio-bank cohorts,
dbSNP, 1000GP, ExAc

ICGC,
TCGA,
COSMIC

IHEC



New technologies

Single cell

Plate-based (C1/bead-seq)

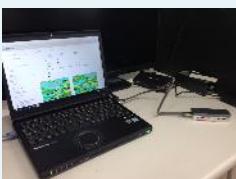


Droplet-based (10x Chromium)



Long read

Linked read (10x Chromium)

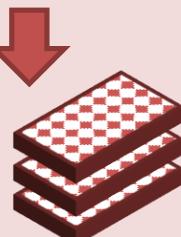
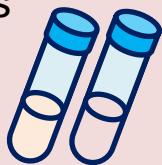


Nanopore (MinION)

Drug perturbation

95 well-annotated compounds

- Approved kinase inhibitors
- Cytotoxic drugs
- Epigenetic targeting drugs...



RNA-seq

ATAC-seq

23 lung cancer cell lines

How to use DBKERO: Genome Browser



We recommend to use Edge (V40 above), Google Chrome (V61 above) or Firefox (V56 above) for the DBKERO browsing. Internet Explorer has not been supported since 2015.

Top |

Tools

- Genome Browser [GitHub]
- Search from Keyword or Genomic Position (human and mouse)
- Search from SNV-enriched Gene in Cancers

Pathway Map

- Human Pathway Map

News

- 27 Jun. 2018: The function of "TF binding site search" was released [here](#).
- 21 Dec. 2017: New KERO (Release 1.1) opened
- 1 Sep. 2017: New C1 data of human lung adenocarcinoma

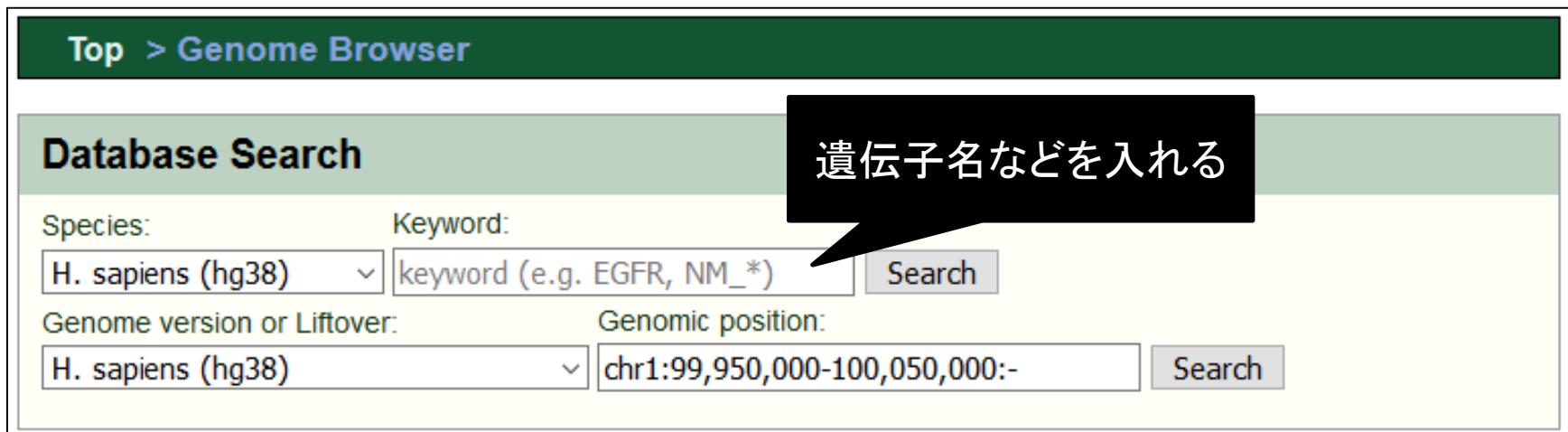
Release 1.1 Updated on 2018/06/27
Based on UCSC Genome Browser Form 1.1

Helpページに使い方が載っている



Documents

- [Experimental Procedures](#)
- [Data Contents](#)
- [Help](#)
- [Download](#)
- [Links](#)



Top > Genome Browser

Database Search

Species: H. sapiens (hg38) Keyword: keyword (e.g. EGFR, NM_*)

Genome version or Liftover: H. sapiens (hg38) Genomic position: chr1:99,950,000-100,050,000:-

遺伝子名などを入れる

How to use DBKERO: Genome Browser

“EGFR”を入れてSearch

Species: H. sapiens (hg38) Keyword: EGFR

Genome version or Liftover: H. sapiens (hg38) Genomic position: chr1:99,950,000-100,050,000:-

Font size: 16 Change font size

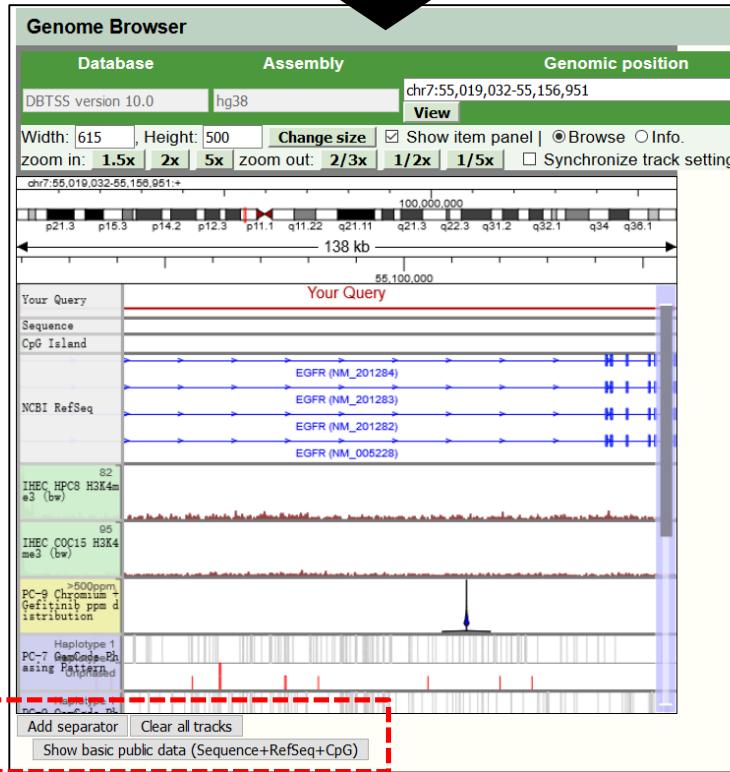
Show 50 entries

Term Genomic position Go to Go to Go to

EGFR	chr7:55,019,032-55,156,951:+	TSS viewer	Genome viewer	Human Variation DB (external site)
------	------------------------------	------------	-------------------------------	--

Showing 1 to 1 of 1 entries

Genome viewer



Genome Browser:
EGFR遺伝子領域を表示

デフォルトでいくつかコンテンツが
表示されます

真っ新にしたい場合は

1. Clear all tracks
2. Show basic public data を押すとよいです

How to use DBKERO: 表示データの選択

Tracks

Basic Items

Sequence CpG Island NCBI RefSeq UCSC gene
 Human vs Mouse Human vs Chimp Human vs Macaque Human vs Rat Human vs Zebrafish

Datasets

Standard multi-omics data

Ethnic	Omics Category	Tissue Type	Sample Type	Dataset	Track Type
<input checked="" type="radio"/> Japanese	<input checked="" type="radio"/> Genome	<input checked="" type="radio"/> Normal cell	<input checked="" type="radio"/> Cell lines		no data no data
<input type="radio"/> Worldwide	<input type="radio"/> Epigenome	<input type="radio"/> Cancer cell	<input type="radio"/> Clinical samples		
	<input type="radio"/> Transcriptome				

Check Track type

Drug perturbation

Dataset

- Dataset-1 (95 compounds)
- Dataset-2 (23 compounds)
- CMAP
 - (The Connectivity Map (Broad Institute))

New technologies

Type	Tech.	Track Type
<input type="radio"/> Single cell		no data no data
<input type="radio"/> Long read		

Check Track type

細胞株および臨床検体の公共データ

- ゲノム
- エピゲノム
- トランスクリプトーム

薬剤刺激後のトランスクリプトーム・エピゲノム
(肺腺癌細胞株)

肺腺癌細胞株のシングルセルトランスクリプトーム
およびロングリードデータ

How to use DBKERO: 表示データの選択

Track item details:

+ CREST - IHEC (The International Human Epigenome Consortium (AMED-CREST, Japan))

+ Single cell (lung cancer cells)

+ 10x GemCode (lung cancer cells)

+ GWAS (JAPANESE)

+ User data

+ Reference gene (common public data)

+ Comparative genome

+ SNP/SNV public data

+ Other multi-omics public data

+ Lung adenocarcinoma 26 cell lines

+ Other human cells

+ ENCODE

+ ChromHMM map (ENCODE)

+ Drug perturbation

直接個々のデータを選択
できる

Using your local PC files (without uploading)

参照... ファイルが選択されていません。 Add local BAM(.bam and .bai) or bigWig(.bw) track

* This is a test function. It may not work well.

* You need to prepare hg38 coordinated indexed BAM or BigWig files using chromosome names like chr1, chr2, chrX... (not supported 1, 2, X...).

ユーザーのファイルを表示する
こともできる
BAMファイル・bigWigファイル

公開されているがんゲノムデータとのリンク

ICGCやTCGAといった公開がんゲノムデータの頻度情報をリンク

<input checked="" type="checkbox"/> SNP/SNV public data
TCGA (SNV) TCGA Urothelial bladder carcinoma (blca_2013) TCGA Breast tumors (brca_2012) TCGA Colon and rectal cancer (coadread_2012) TCGA Glioblastoma (gbm_2008) TCGA Glioblastoma (gbm_2013) TCGA Clear-cell renal cell carcinoma (kirc_2013) TCGA Acute myeloid leukemia (aml_2012) TCGA Squamous cell lung cancers (lusc_2012) TCGA Ovarian carcinoma (ov_2011) TCGA stad (stad_2014) TCGA Endometrial carcinoma (ucec_2013) TCGA Lung adenocarcinoma (luad_2014) TCGA Chromophobe renal cell carcinoma (kich_2014) TCGA Papillary thyroid carcinoma (thca_2014) TCGA Head and neck squamous cell carcinomas (hnsc_2014) TCGA Diffuse lower grade gliomas (lgg_2015) TCGA Prostate adenocarcinoma (prad_2015) TCGA Adrenocortical carcinoma (acc_2016)
ICGC (SNV) ICGC Acute Myeloid Leukemia - TARGET, US ICGC Bladder Urothelial carcinoma - CN ICGC Bladder Urothelial Cancer - TGCA, US ICGC Bone Cancer - UK ICGC Breast Triple Negative/Lobular Cancer - UK ICGC Breast Cancer - TCGA, US ICGC Chronic Lymphocytic Leukemia - ES ICGC Chronic Myeloid Disorders - UK ICGC Colon Adenocarcinoma - TCGA, US ICGC Early Onset Prostate Cancer - DE ICGC Esophageal Adenocarcinoma - UK ICGC Esophageal Cancer - CN ICGC Gastric Cancer - CN ICGC Brain Glioblastoma Multiforme - TCGA, US ICGC Kidney Renal Clear Cell Carcinoma - TCGA, US ICGC Kidney Renal Papillary Cell Carcinoma - TCGA, US ICGC Acute Myeloid Leukemia - KR ICGC Brain Lower Grade Glioma - TCGA, US ICGC Liver Cancer - FR ICGC Liver Cancer - NCC, JP ICGC Liver Cancer - RIKEN, JP ICGC Lung Cancer - Squamous cell carcinoma - KR ICGC Lung Squamous Cell Carcinoma - TCGA, US ICGC Malignant Lymphoma - DE ICGC Neuroblastoma - TARGET, US ICGC Oral Cancer - IN ICGC Ovarian Cancer - AU ICGC Ovarian Serous Cystadenocarcinoma - TCGA, US ICGC Pancreatic Cancer - AU ICGC Pancreatic Cancer - CA ICGC Pancreatic Cancer Endocrine neoplasms - AU ICGC Pediatric Brain Cancer - DE ICGC Prostate Adenocarcinoma - CA ICGC Prostate Adenocarcinoma - UK ICGC Prostate Adenocarcinoma - TCGA, US ICGC Rectum Adenocarcinoma - TCGA, US ICGC Renal clear cell carcinoma - CN ICGC Renal Cell Cancer - EU/FR ICGC Skin Cutaneous melanoma - TCGA, US ICGC Gastric Adenocarcinoma - TCGA, US ICGC Thyroid Cancer - SA ICGC Head and Neck Thyroid Carcinoma - TCGA, US ICGC Uterine Corpus Endometrial Carcinoma- TCGA, US ICGC Acute Myeloid Leukemia - TARGET, US ICGC Bone Cancer - Ewing Sarcoma - FR ICGC Breast ER+ and HER2- Cancer - EU/UK ICGC Breast Cancer - FR ICGC Breast Cancer - KR ICGC Biliary tract cancer - JP ICGC Cervical Squamous Cell Carcinoma - TCGA, US ICGC Colorectal Cancer - CN ICGC Acute myeloid leukaemia and Chronic myelogenous leukaemia - CN ICGC Benign Liver Tumour - FR ICGC Liver Cancer - CN ICGC Liver Hepatocellular carcinoma - TCGA, US ICGC Liver Hepatocellular Macronodules - FR ICGC Lung Squamous cell carcinoma - CN ICGC Skin Cancer - AU ICGC Pancreatic Endocrine Neoplasms- IT ICGC Skin Adenocarcinoma - BR
Others 1000genome SNP Clear-cell renal cell carcinoma dbSNP build 137 NHLBI Exome Sequencing Project NCBI ClinVar COSMIC Meyerson's lung adenocarcinoma Myelodysplasia ICGC ALL Nagahama Japanese SNP Japan PGx (Japanese SNP) ExAC SCLC (Small Cell Lung Cancer) 26 lung adenocarcinoma cell lines Lung adenocarcinoma 97 clinical samples

International Cancer Genome Consortium (ICGC; <https://icgc.org/>)

The Cancer Genome Atlas (TCGA; <https://cancergenome.nih.gov/>)

Japanese lung adenocarcinoma (Suzuki and Mimaki et al. 2013 *PLoS ONE*)

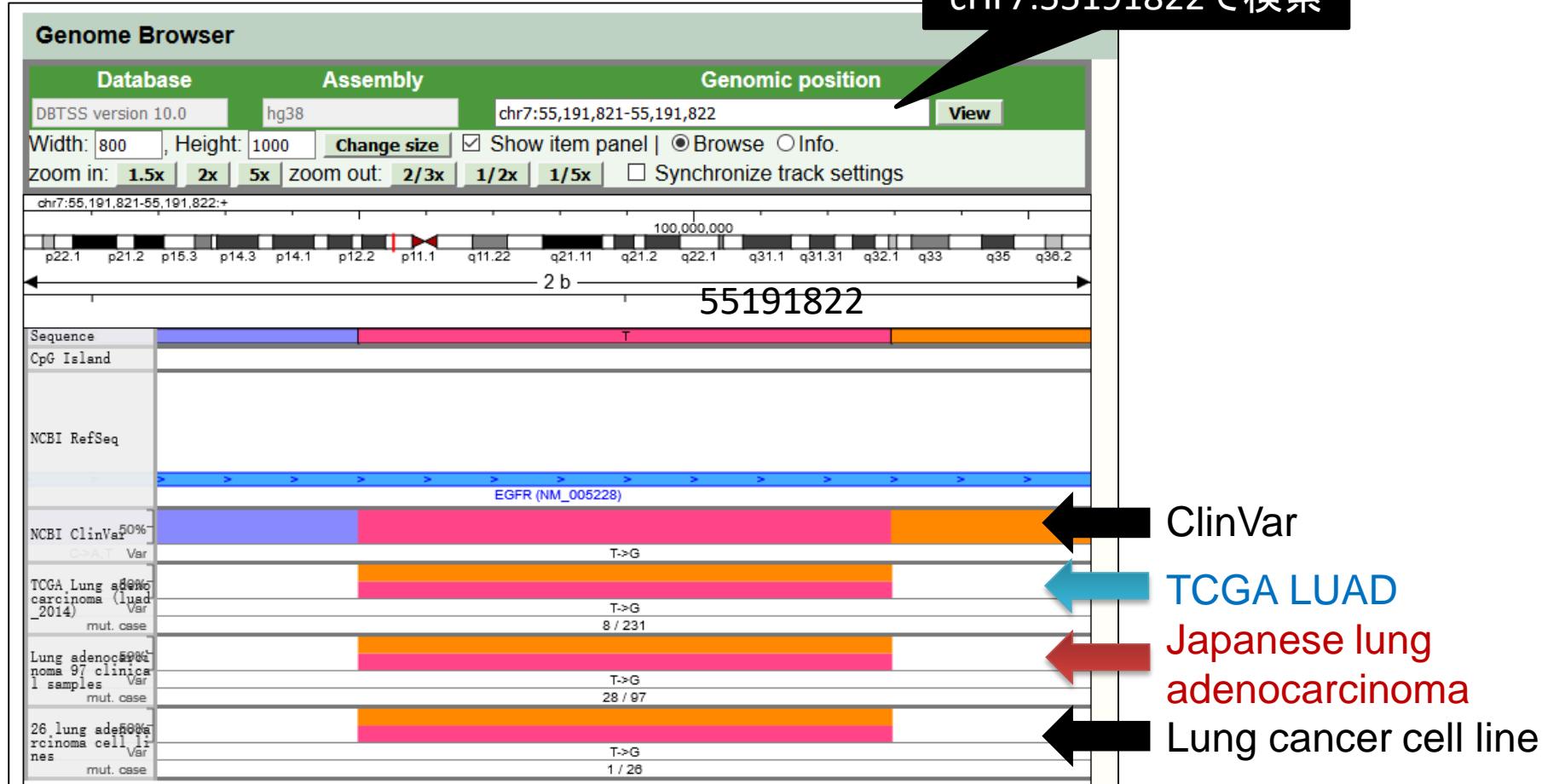
Japanese small cell lung cancer (Umemura et al. 2014 *J Thorac Oncol*) など

ゲノム変異検索

EGFR変異 (Lung adenocarcinoma)

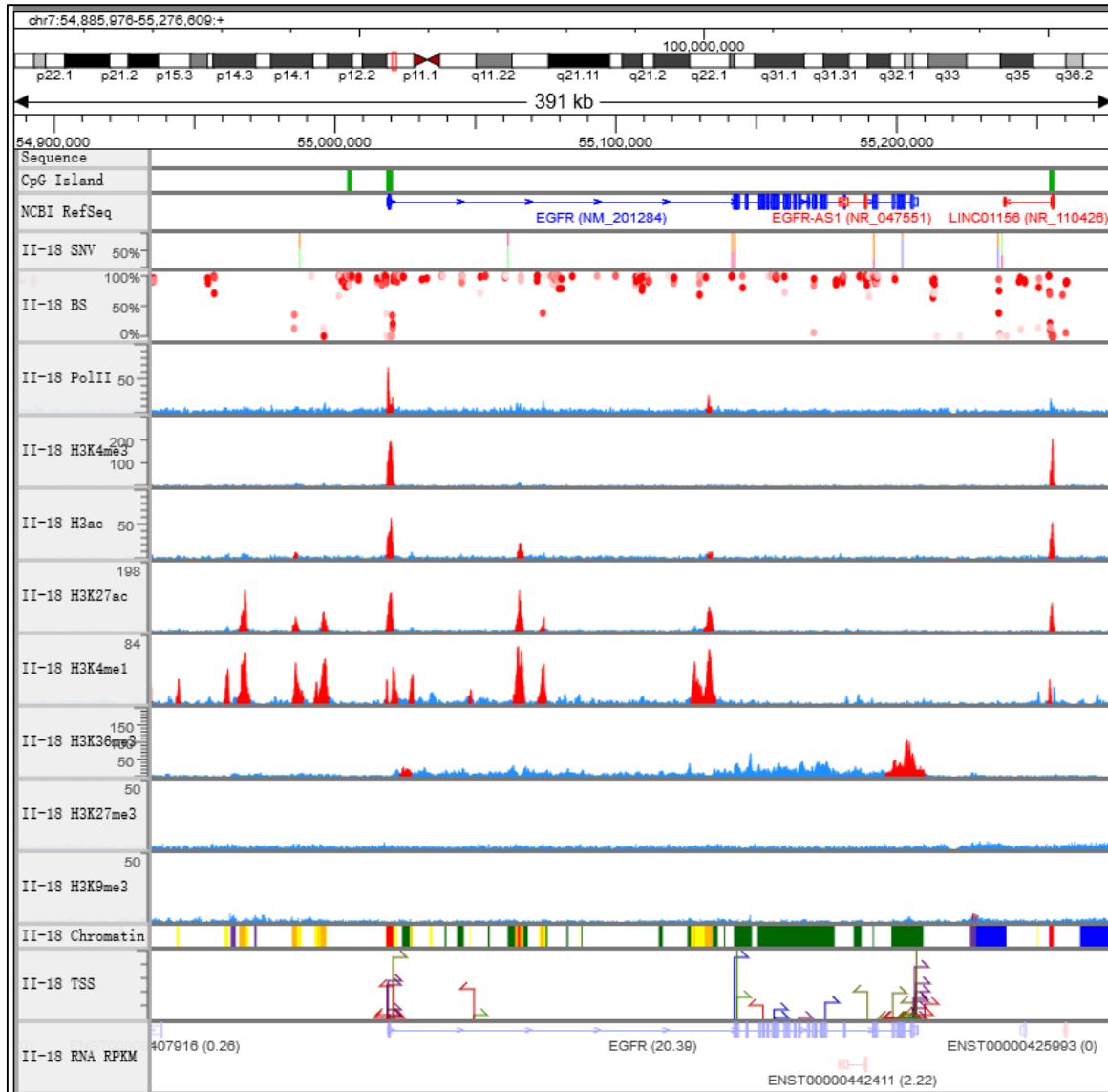
chr7:55191822, T>G, L858R

chr7:55191822で検索



ゲノム変異と多層オミクスデータ

EGFR遺伝子の多層オミクスステータス(細胞株モデル)



BS: DNA methylation

Pol II

H3K4me3

H3K9/14ac

H3K27ac

H3K4me1

H3K36me3

H3K27me3

H3K9me3

ChIP-seq

TSS

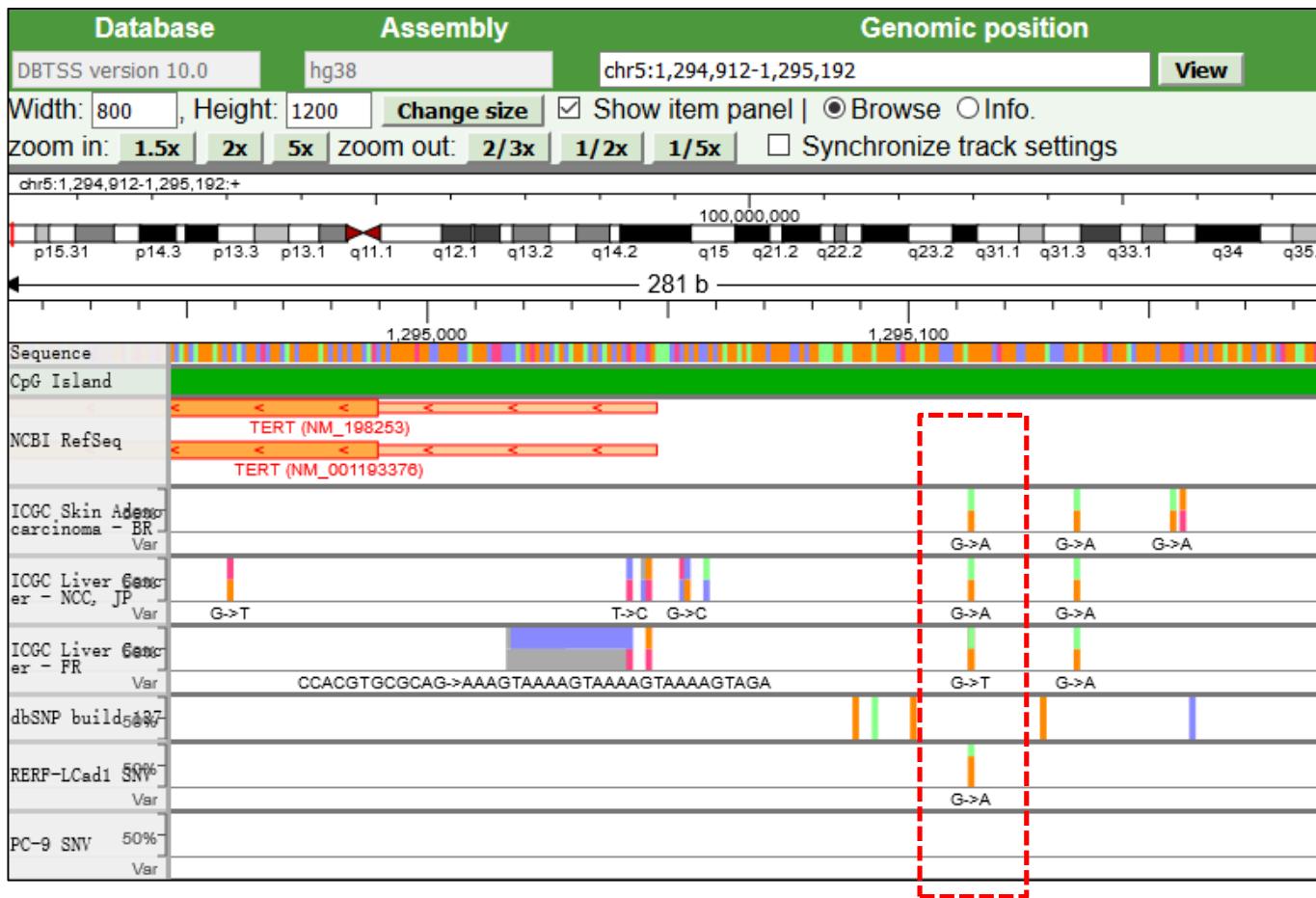
RNA-seq (rpkm)

Chromatin (ChromHMM)

非コード領域のゲノム多型・変異

*TERT*遺伝子のプロモーター変異

chr5:1295113, G>A



ICGC Skin Adenocarcinoma (BR)

ICGC Liver Cancer (JP)

ICGC Liver Cancer (FR)

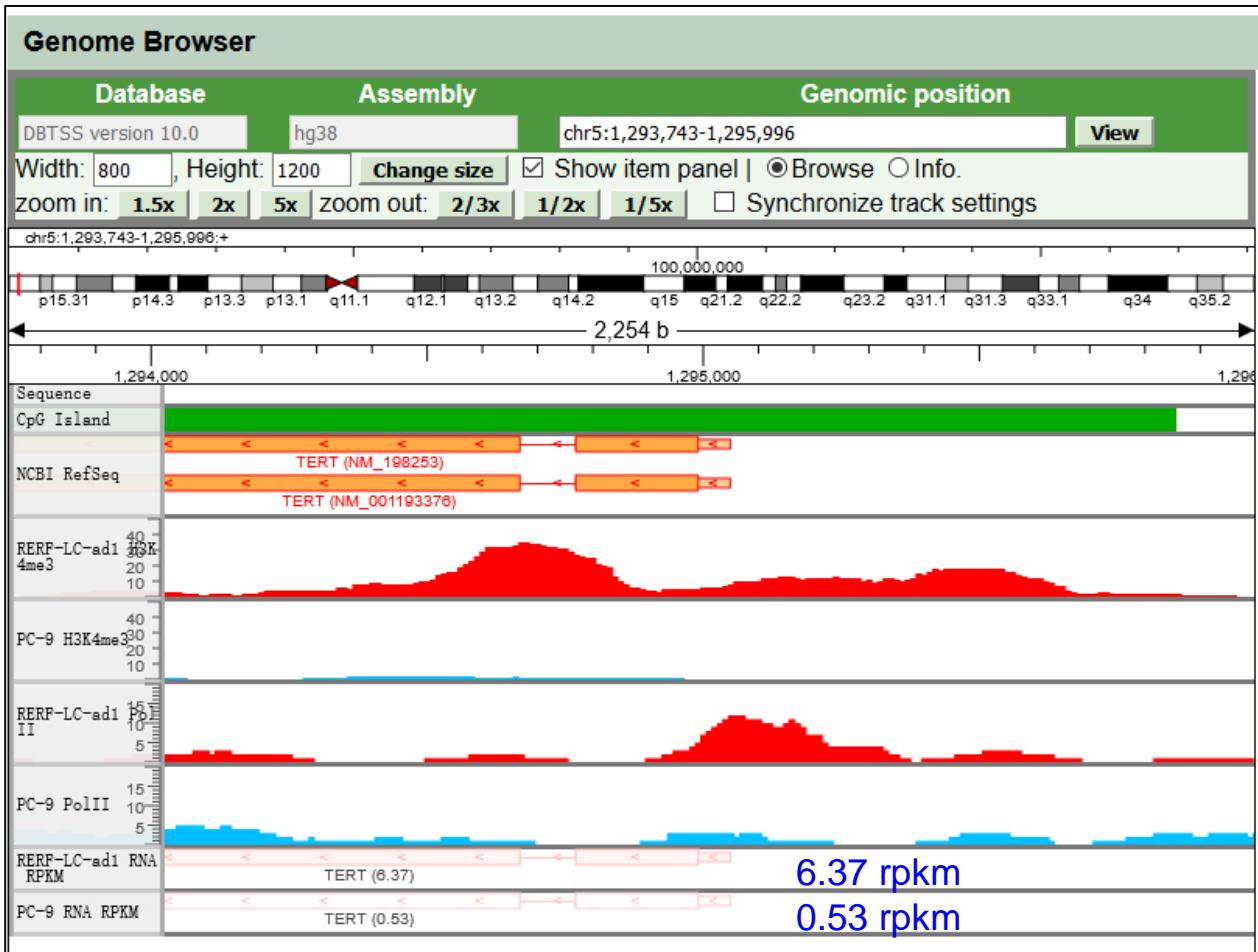
dbSNP (多型データベース)

RERF-LC-Ad1
PC-9

Lung cancer cell lines

非コード領域のゲノム多型・変異による 転写制御への影響

*TERT*遺伝子のプロモーター領域



RERF-LC-Ad1: 変異あり
PC-9: 変異なし

RERF-LC-Ad1 H3K4me3

PC-9 H3K4me3

RERF-LC-Ad1 Pol II

PC-9 Pol II

RERF-LC-Ad1 RNA
PC-9 RNA

CREST/IHEC日本チーム ヒト細胞のエピゲノムデータ

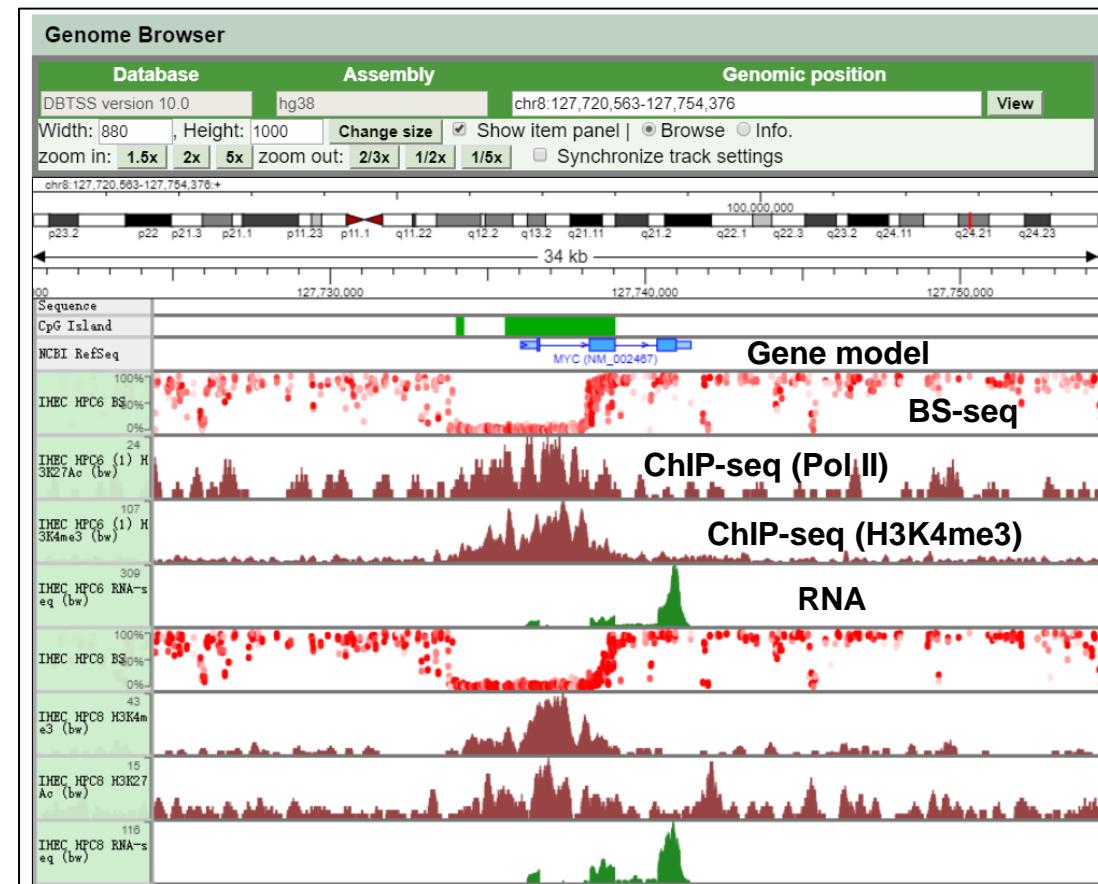


国際ヒトエピゲノムコンソーシアム
The International Human Epigenome
Consortium (IHEC)

<http://crest-ihec.jp/about/index.html>
<http://ihec-epigenomes.org/>

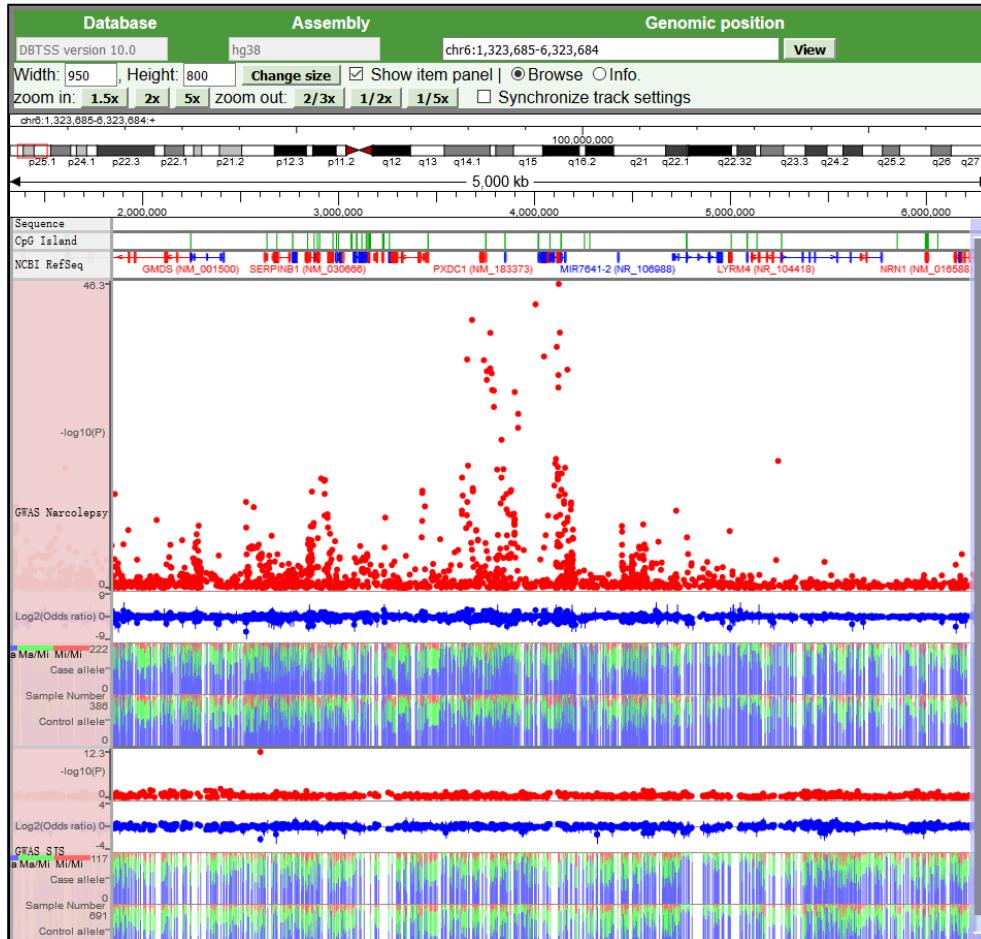
DBKEROからの公開

- Liver
- Colon
- Endometrial
- Vascular endothelial



ヒトゲノム多型データベース(HGVDB)との統合

東大徳永研GWASとの横断検索



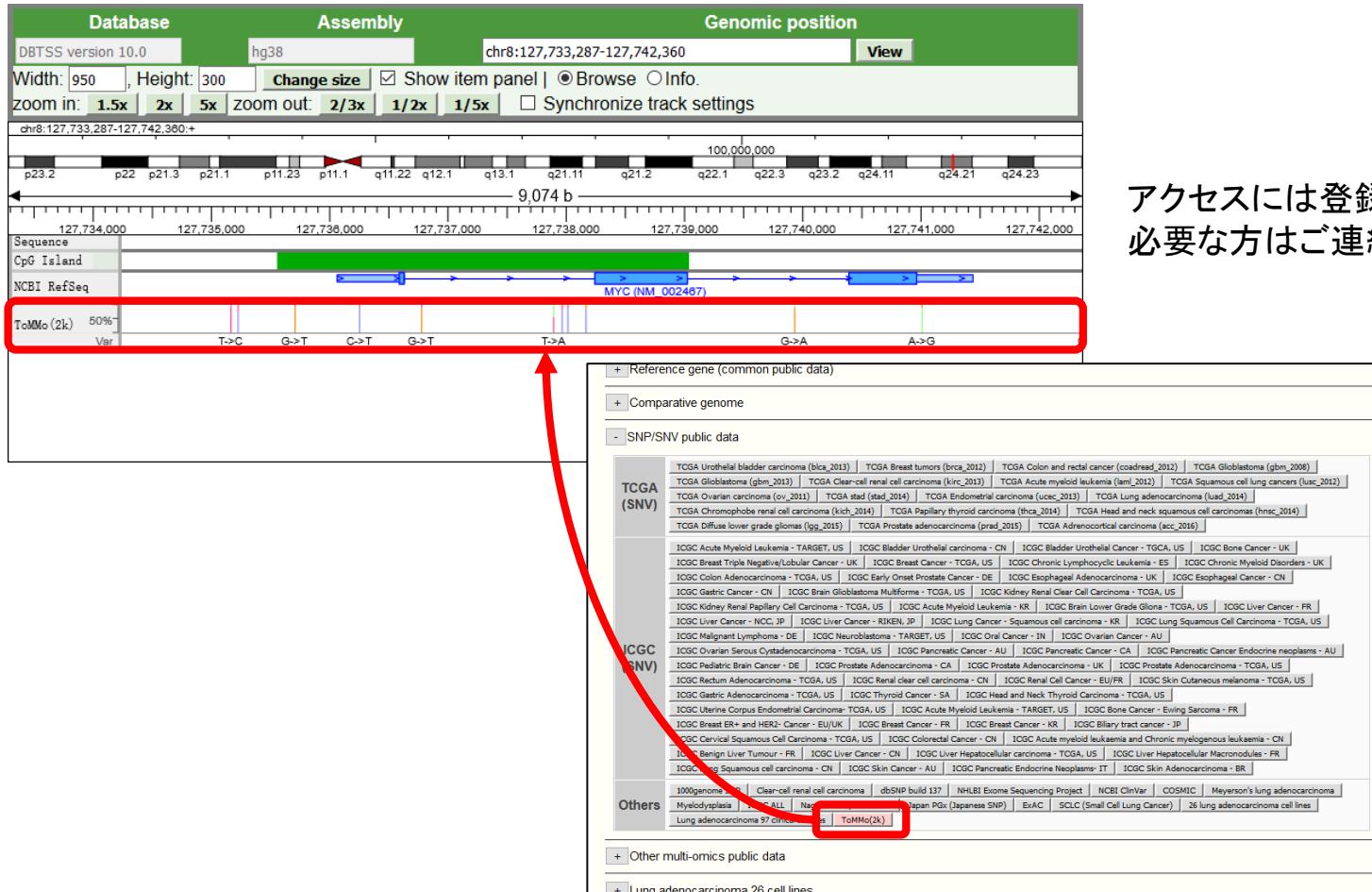
Disease	Result
Stevens-Johnson syndrome	GWAS SJS
Alzheimer	GWAS Alzheimer (GeMDBJ)
Alzheimer	GWAS Alzheimer (JSNP GeMDBJ)
Bronchial asthma	GWAS Bronchial asthma
Cerebral aneurysm	GWAS Cerebral aneurysm
Endometriosis	GWAS Endometriosis
Gastric cancer	GWAS Gastric cancer
Hepatitis	GWAS Hepatitis (HCV-1)
Hepatitis	GWAS Hepatitis (HCV-2)
Hepatitis	GWAS Hepatitis (HCV-3)
Hypertension	GWAS Hypertension
Narcolepsy	GWAS Narcolepsy
Panic disorder	GWAS Panic disorder
Type II diabetes mellitus	GWAS Type II diabetes mellitus

統合元: This human genome variation database

https://gwas.biosciencedbc.jp/cgi-bin/hvdb/hv_top.cgi

東北メディカルメガバンクデータベース(ToMMo)との統合

国内最大規模の日本人ゲノムデータ

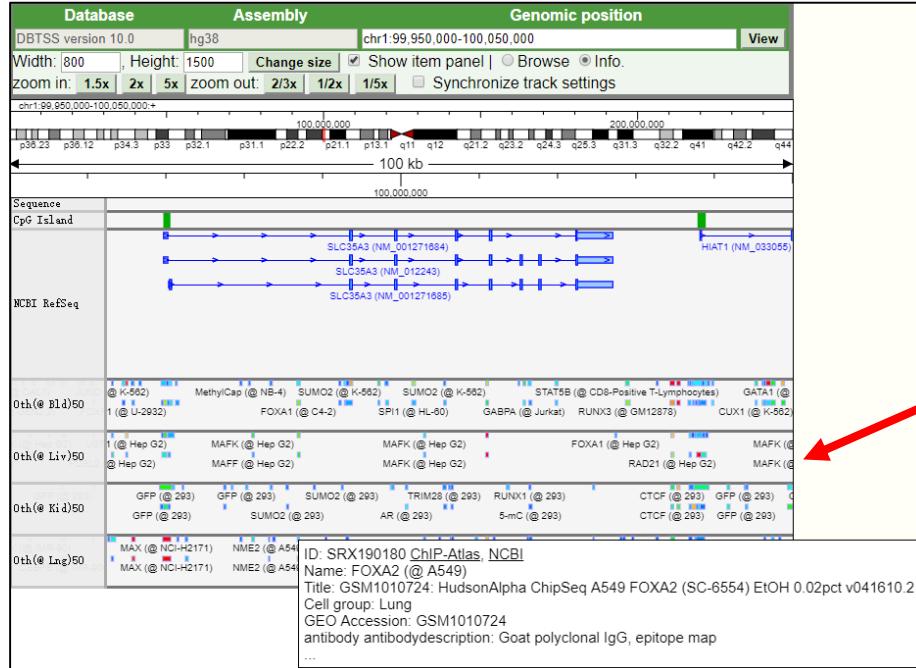


アクセスには登録が必要ですので、
必要な方はご連絡ください。

統合元: integrative Japanese Genome Variation Database (iJGVD)
<https://ijgvd.megabank.tohoku.ac.jp/>

ChIP-Atlasとの統合

ChIP-Atlasとの統合も準備中です



Cell type	Antigen						
	DNase-seq (all)	Histone (all)	RNA polymerase (all)	TFs and others (all)	Input control (all)	Unclassified (all)	No description (all)
Adipocyte (all)	DNS(@ Adp50)	His(@ Adp50)	Poi(@ Adp50)	Oth(@ Adp50)	InP(@ Adp50)	Unc(@ Adp50)	NoD(@ Adp50)
Blood (all)	DNS(@ Bls50)	His(@ Bls50)	Poi(@ Bls50)	Oth(@ Bls50)	InP(@ Bls50)	Unc(@ Bls50)	NoD(@ Bls50)
Bone (all)	DNS(@ Bon50)	His(@ Bon50)	Poi(@ Bon50)	Oth(@ Bon50)	InP(@ Bon50)	Unc(@ Bon50)	NoD(@ Bon50)
Breast (all)	DNS(@ Brs50)	His(@ Brs50)	Poi(@ Brs50)	Oth(@ Brs50)	InP(@ Brs50)	Unc(@ Brs50)	NoD(@ Brs50)
Cardiovascular (all)	DNS(@ CDV50)	His(@ CDV50)	Poi(@ CDV50)	Oth(@ CDV50)	InP(@ CDV50)	Unc(@ CDV50)	NoD(@ CDV50)
Digestive tract (all)	DNS(@ Dig50)	His(@ Dig50)	Poi(@ Dig50)	Oth(@ Dig50)	InP(@ Dig50)	Unc(@ Dig50)	NoD(@ Dig50)
Epidermis (all)	DNS(@ Epd50)	His(@ Epd50)	Poi(@ Epd50)	Oth(@ Epd50)	InP(@ Epd50)	Unc(@ Epd50)	NoD(@ Epd50)
Gonad (all)	DNS(@ Gon50)	His(@ Gon50)	Poi(@ Gon50)	Oth(@ Gon50)	InP(@ Gon50)	Unc(@ Gon50)	NoD(@ Gon50)
Kidney (all)	DNS(@ Kid50)	His(@ Kid50)	Poi(@ Kid50)	Oth(@ Kid50)	InP(@ Kid50)	Unc(@ Kid50)	NoD(@ Kid50)
Liver (all)	DNS(@ Liv50)	His(@ Liv50)	Poi(@ Liv50)	Oth(@ Liv50)	InP(@ Liv50)	Unc(@ Liv50)	NoD(@ Liv50)
Lung (all)	DNS(@ Lng50)	His(@ Lng50)	Poi(@ Lng50)	Oth(@ Lng50)	InP(@ Lng50)	Unc(@ Lng50)	NoD(@ Lng50)
Muscle (all)	DNS(@ Myo50)	His(@ Myo50)	Poi(@ Myo50)	Oth(@ Myo50)	InP(@ Myo50)	Unc(@ Myo50)	NoD(@ Myo50)
Neural (all)	DNS(@ Neu50)	His(@ Neu50)	Poi(@ Neu50)	Oth(@ Neu50)	InP(@ Neu50)	Unc(@ Neu50)	NoD(@ Neu50)
Pancreas (all)	DNS(@ Pan50)	His(@ Pan50)	Poi(@ Pan50)	Oth(@ Pan50)	InP(@ Pan50)	Unc(@ Pan50)	NoD(@ Pan50)
Placenta (all)	DNS(@ Pla50)	His(@ Pla50)	Poi(@ Pla50)	Oth(@ Pla50)	InP(@ Pla50)	Unc(@ Pla50)	NoD(@ Pla50)
Pluripotent stem cell (all)	DNS(@ Psa50)	His(@ Psa50)	Poi(@ Psa50)	Oth(@ Psa50)	InP(@ Psa50)	Unc(@ Psa50)	NoD(@ Psa50)
Prostate (all)	DNS(@ Psc50)	His(@ Psc50)	Poi(@ Psc50)	Oth(@ Psc50)	InP(@ Psc50)	Unc(@ Psc50)	NoD(@ Psc50)
Uterus (all)	DNS(@ Utr50)	His(@ Utr50)	Poi(@ Utr50)	Oth(@ Utr50)	InP(@ Utr50)	Unc(@ Utr50)	NoD(@ Utr50)
Others (all)	DNS(@ Oth50)	His(@ Oth50)	Poi(@ Oth50)	Oth(@ Oth50)	InP(@ Oth50)	Unc(@ Oth50)	NoD(@ Oth50)
Unclassified (all)	DNS(@ Unc50)	His(@ Unc50)	Poi(@ Unc50)	Oth(@ Unc50)	InP(@ Unc50)	Unc(@ Unc50)	NoD(@ Unc50)
No description (all)							NoD(@ NoD50)

TF Search機能搭載予定 → 転写因子結合部位の検索が可能に！

The form allows users to search for transcription factor binding sites across the ChIP-Atlas and GGGenome databases. It includes fields for Species (Homo sapiens), Definition of search regions (using UCSC RefFlat start or BED file), TF conditions (Sequence pattern, Strandness of gene, Transcription factor, MACS score), and a preview section showing a DNA sequence with a transcription factor binding site highlighted.

統合元: ChIP-Atlas
<https://chip-atlas.org/>

日本人ゲノム多様性統合データベースTogoVarとの統合

TogoVar(NBDC/DBCLS)との統合も準備中です

The screenshot shows two main pages of the TogoVar website.

Left Page (Search Results):

- Header: TOGO VAR Home Datasets Downloads Terms Contact About Help
- Left sidebar: Data set (checkboxes for All, JGA NGS, JGA SNP, 3.5KPN, HGVD, ExAC, ClinVar), Alternative allele frequency (checkbox for JGA NGS), Variant type (checkboxes for All, SNV, Insertion, Deletion, Indel, Substitution), Clinical significance (checkboxes for All, Not in ClinVar, Pathogenic, Likely pathogenic, Uncertain significance, Likely benign).
- Main content: Search bar (Search for disease or gene symbol or rs...), Example: Disease Breast-ovarian cancer, familial 2, Gene ALDH2, refSNP rs114202595, TogoVar Region(GRCh37/hg19) 10:73270743-73376976. Warning: Scroll function over 1,000,000 results is currently unavailable.
- Table: Showing 1 to 25 of 1,000,000 variants (filtered from 19,537,677 variants). Columns: ID, Position, Ref, Alt, Symbol, Frequency. Rows show variants like lgv21, lgv27, lgv29, lgv41, lgv42, lgv44, lgv45.

Right Page (Summary):

- Header: TOGO VAR Home Datasets Downloads Terms Contact About Help Language: en / ja
- Title: TogoVar収録データセット一覧
- Section: ヒトゲノムリファレンス配列 (GRCh37.p13 (06/28/2013))
- Section: Japanese Genotype-phenotype Archive (JGA)の個人ゲノムデータ由来のバリアントデータセット
- Table: データセット名 | バージョン/最終更新日 | サンプルサイズ | 検出バリアント数 | 除外後バリアント数 | 作成者

JGA-NGS	06/01/2018	125	13,338,968	4,679,025	NBDC
JGA-SNP	06/01/2018	183,884	1,966,919	1,249,724	NBDC

- Note: アレル数5以下のバリアントはJGA-NGS/JGA-SNPデータセットから削除しています。
- Section: 第三者が作成したバリアントデータセット
- Table: データセット名 | バージョン/最終更新日 | サンプルサイズ | バリアント数 | バリアント部位数 | 作成者

ClinVar	02/01/2017	—	198,707	—	NCBI
Exome Aggregation Consortium (ExAC)	Release 1 (02/27/2017)	60,706	10,195,872	9,362,319	Broad Institute
Human Genetic Variation Database (HGVD)	Version 2.30 (08/02/2017)	1,208	554,461	501,556	京都大学
Integrative Japanese Genome Variation Database (JGV)	3.5KJPN (09/28/2017)	3,554	—	7,931,579	東北メディカル・メガバンク機構

- Note: HGVD 3.5KJPNについては、DTAなしで閲覧可能なアレル頻度1%以上のバリアントのみを収録しております。詳細はData releaseをご覧ください。
- Section: バリアント以外のデータセット
- Table: データセット名 | バージョン/最終更新日 | 内容 | 作成者

Coll	01/29/2018	生命科学分野の文献間の引用関係の情報	DBCLS
PubTator	04/23/2018	バリアント名が出現する文献の情報	NCBI
TogoGenome	hg19+UniProt_201804	ゲノムに関連する多種多様な情報	DBCLS

TogoVar:日本人ゲノム配列の個人による違い(バリアント)と、関連する疾患情報などを収集・整理したデータベース

統合元: TogoVar <https://togovar.biosciencedbc.jp/>

新規技術：シングルセル解析

シングルセル解析

1細胞ごとのゲノム・エピゲノム・トランスクリプトームパターンを次世代シークエンス技術によって解析する

→ さまざまな細胞種・状態が混ざり合った不均一な細胞集団において、個々の1細胞の状態を詳細に解析できる

特に、シングルセルRNA-seq (scRNA-seq)については、細胞分離・逆転写・増幅を自動で行うプラットフォームがいくつか販売されている

Platform	反応系	サンプルあたりの 解析細胞数	特徴
C1 (Fluidigm)	マイクロフリューディクス	96および800	Full-length (800の場合3')
ddSEQ (BioRad/illumina)	マイクロドロップレット	数百	3' RNA-seq
ICELL8 (Takara)	マイクロウェル	~1800	3' RNA-seq 5~100μm
BD Rhapsody (BD)	マイクロウェル	~10,000	3' target sequencing
Chromium (10x Genomics)	マイクロドロップレット	500~10,000	3' RNA-seq

がん細胞のscRNA-seq解析

Kashima Y et al. 2018 *Scientific Reports*

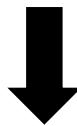
Combinatory use of distinct single-cell RNA-seq analytical platforms reveals the heterogeneous transcriptome response

肺腺癌細胞株における分子標的薬に対する発現応答を解析した研究

Bead-seq

- ・細胞数少
- ・細胞あたりのリード数・検出遺伝子数多

個々の細胞を詳しく解析



Chromium

- ・細胞数多
- ・細胞あたりのリード数・検出遺伝子数少

集団の特徴を解析

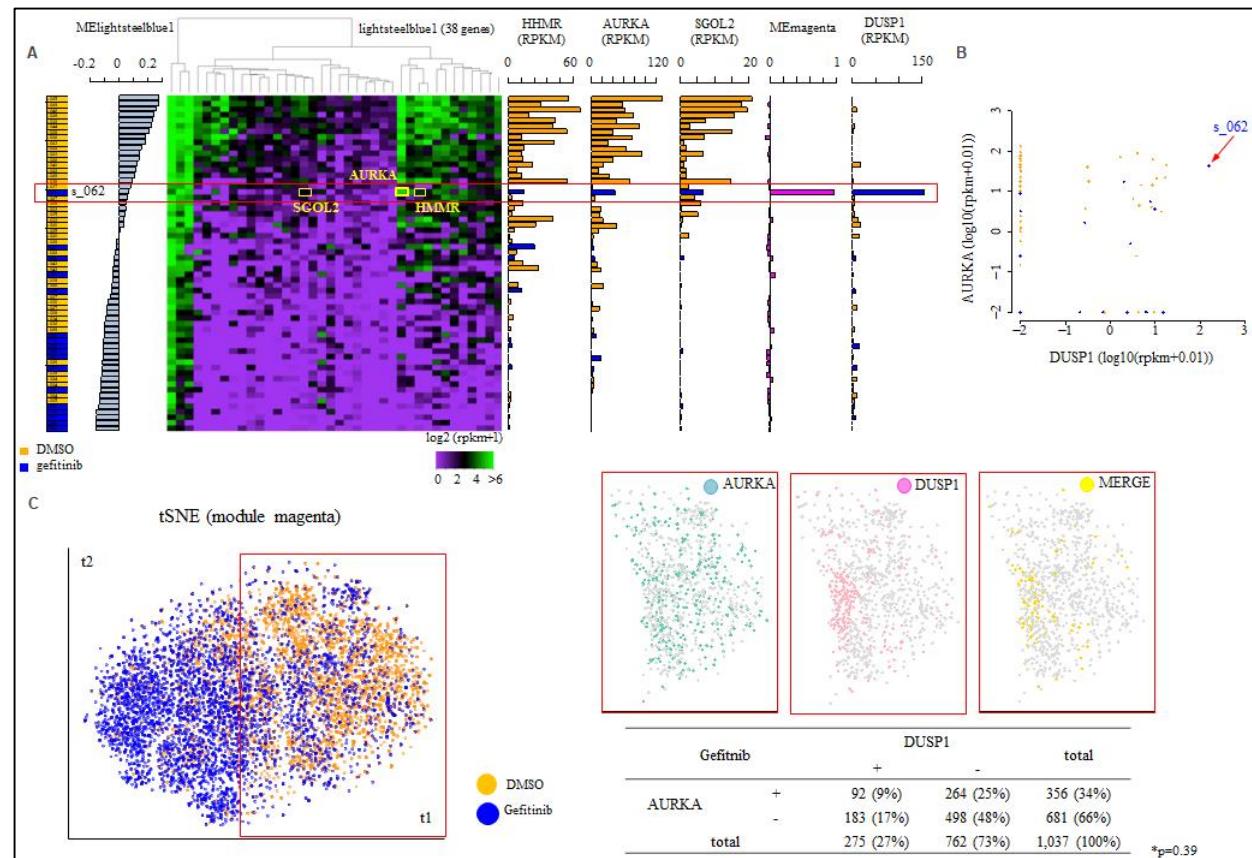


Figure 7; Kashima Y et al. 2018 *Scientific Reports*

DBKERO: 肺腺癌細胞株のscRNA-seqデータ

肺腺癌細胞株を分子標的薬で刺激したときの遺伝子発現変化の多様性を解析したデータを収載

New technologies

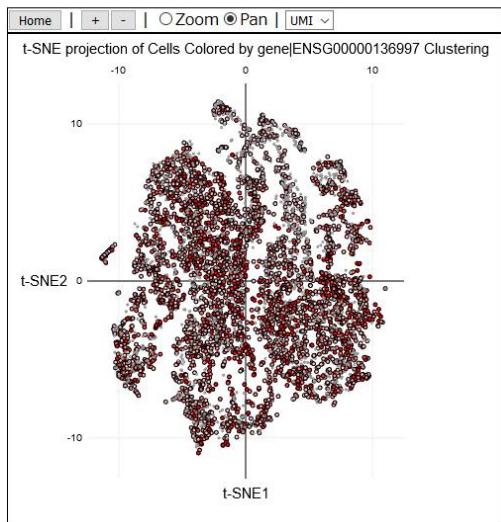
Type	Tech.	Track Type
<input checked="" type="radio"/> Single cell	<input type="radio"/> C1	<input type="checkbox"/> ppm_distribution
<input type="radio"/> Long read	<input checked="" type="radio"/> Chromium (t-SNE Summary)	<input type="checkbox"/> bigWig
	<input type="radio"/> Bead-seq	<input type="checkbox"/> BAM

Check Track type

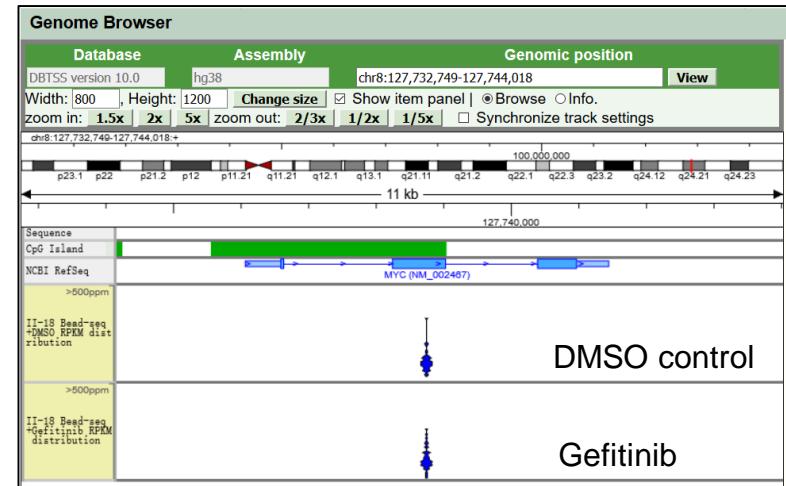
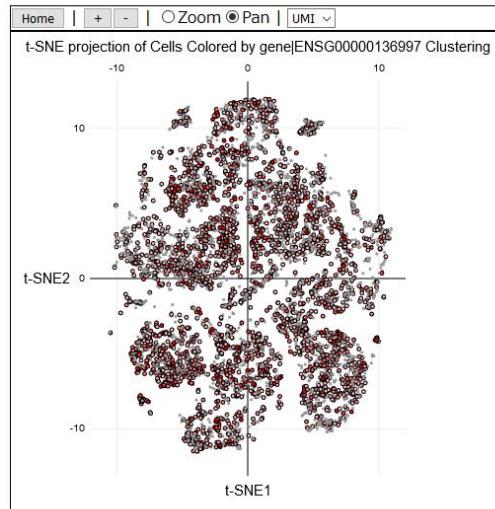


Chromium (t-SNE view)

II-18 Control



II-18 + Gefitinib



Bead-seq (rpkm分布)

MYC遺伝子の発現量
II-18細胞株(EGFR変異陽性株)

新規技術：ロングリード解析

Physical long-read sequencing technology

ナノポア型ロングリードシークエンサー(ONT)

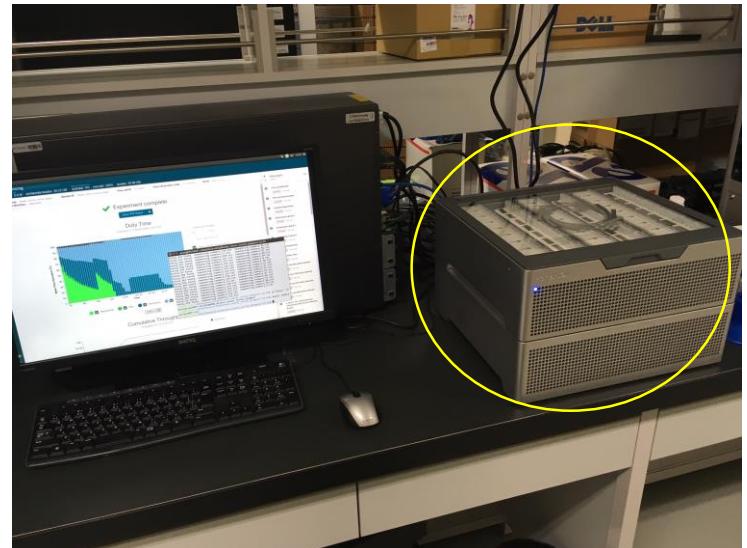
MinION

- 安価な初期投資 (\$1,000スターターキット)
- ポータブル
- 5 Gb / フローセル



PromethION

- 大型機
 - 48枚のフローセルがささる
 - 30-100 Gb / フローセル
- 1～数ランでヒトゲノムを解読



特徴

- ✓ 数十kbのロングリードを解読(ゲノム構造異常検出・フェージング解析が可能)
- ✓ ~48h シークエンスラン
- ✓ Direct RNA-seqが可能
- ✓ 修飾塩基(メチル化など)の解析が可能

MinIONを用いたがんゲノム解析

Suzuki A and Suzuki M et al. 2017 *DNA Research*
Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer
 MinIONを用いた変異検出およびフェーズ情報解析

Seki M and Katsumata E et al. 2018 *DNA Research*
Evaluation and application of RNA-Seq by MinION
 MinIONを用いた全長RNA-seq解析

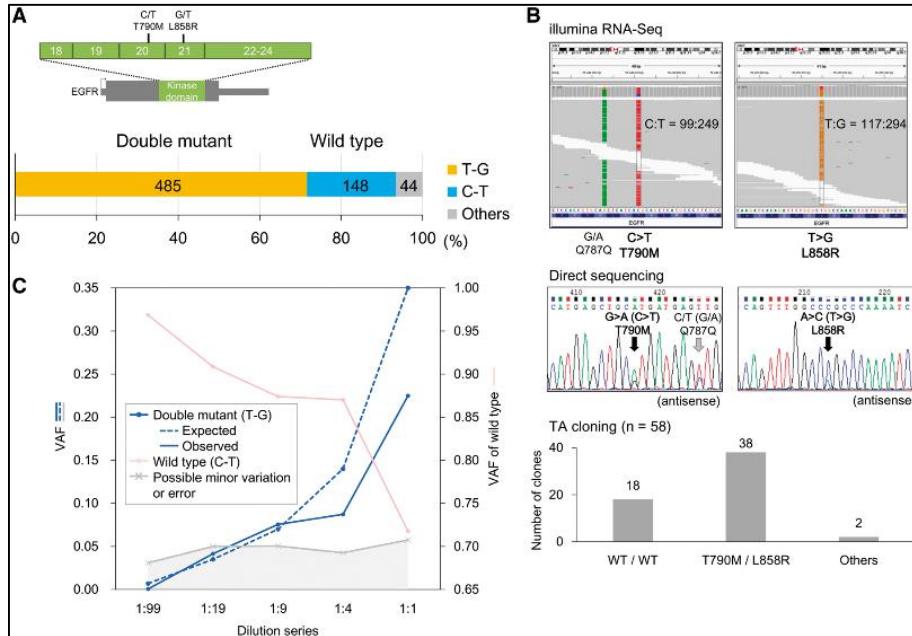


Figure 4; Suzuki A and Suzuki M et al. 2017 *DNA Research*

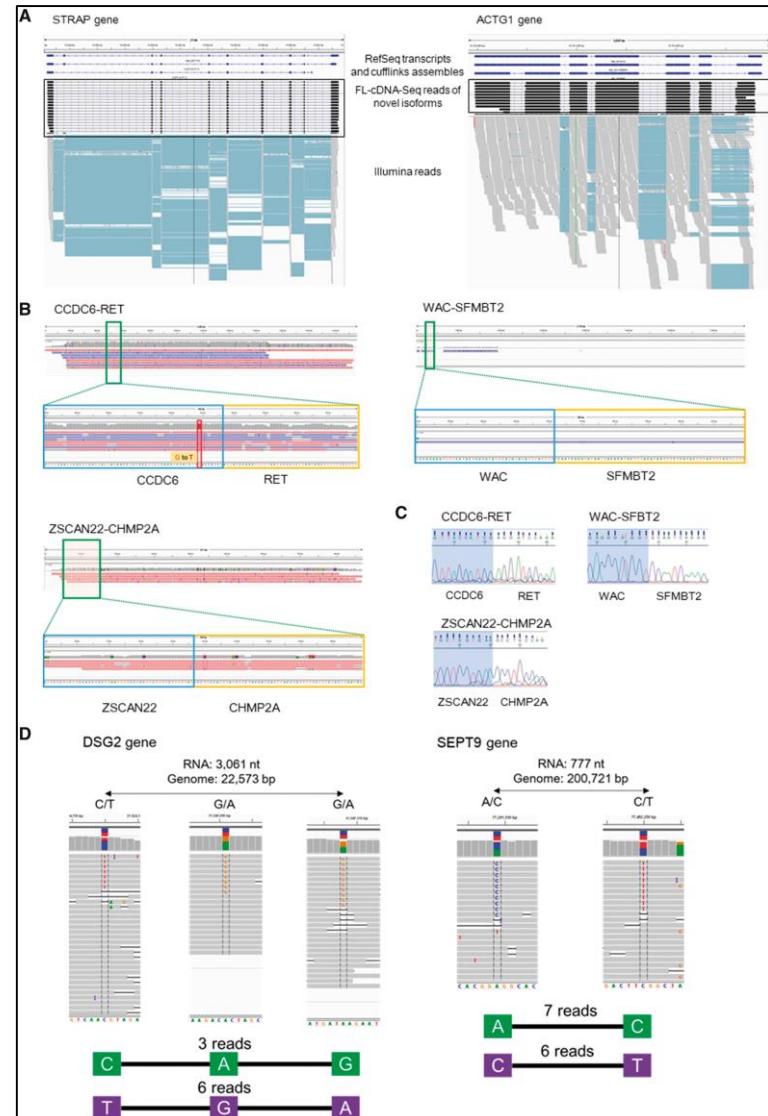


Figure 3; Seki M and Katsumata E et al. 2018 *DNA Research*

MinION/PromethIONを用いたがんゲノム解析

平成30年度「次世代がん医療創生研究事業」(2次公募)

ナノポア型長鎖シークエンサーを駆使したがんゲノム異常における新規概念の創出
および患者層別化手法の開発

クリニカルシークエンスの実用開始

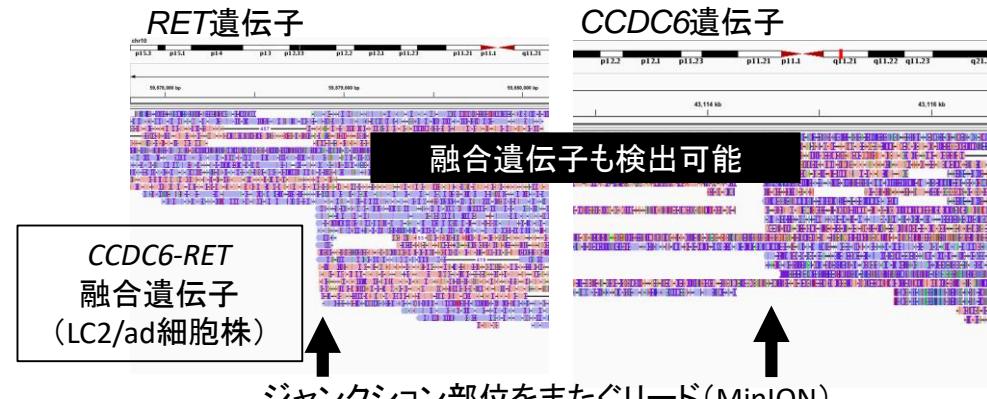
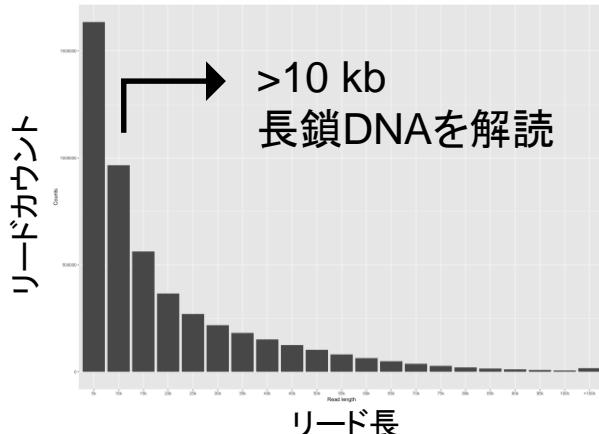
多くのがんは、ドライバー変異に対する分子標的薬によって個別化医療を実現してきた

問題点

- ドライバー変異不明症例
- 治療抵抗性の出現

→ 従来法(ショートリード解析)で見逃されてきた複雑なゲノム変異があるのではないか？

長鎖シークエンサーMinION/PromethIONでがんゲノムを再評価



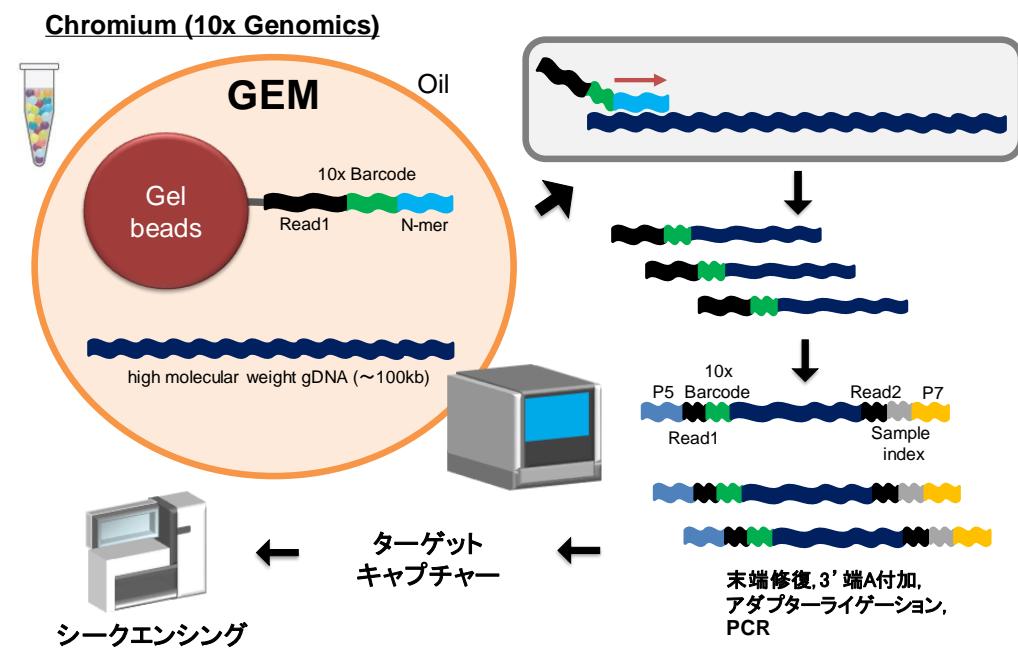
新規技術: ロングリード解析

Synthetic long-read sequencing technology
Chromium Genome/Exome (10x Genomics)

Linked Read

Synthetic long-read sequences
by molecular barcoding

- 750,000 beads
- Barcoding long genomic DNA
fragments (~ 100 kb)
- Short-read sequencing



Long-read reconstruction

- Phase information
- Complex genomic structures

Kitzman et al. 2016 *Nature Biotechnology*

Linked readを用いたがんゲノム解析

Sereewattanawoot S et al. 2018

Scientific Reports

Identification of potential regulatory mutations using multi-omics analysis and haplotyping of lung adenocarcinoma cell lines

プロモーターおよびエンハンサー領域の変異における転写制御・遺伝子発現への影響を解析

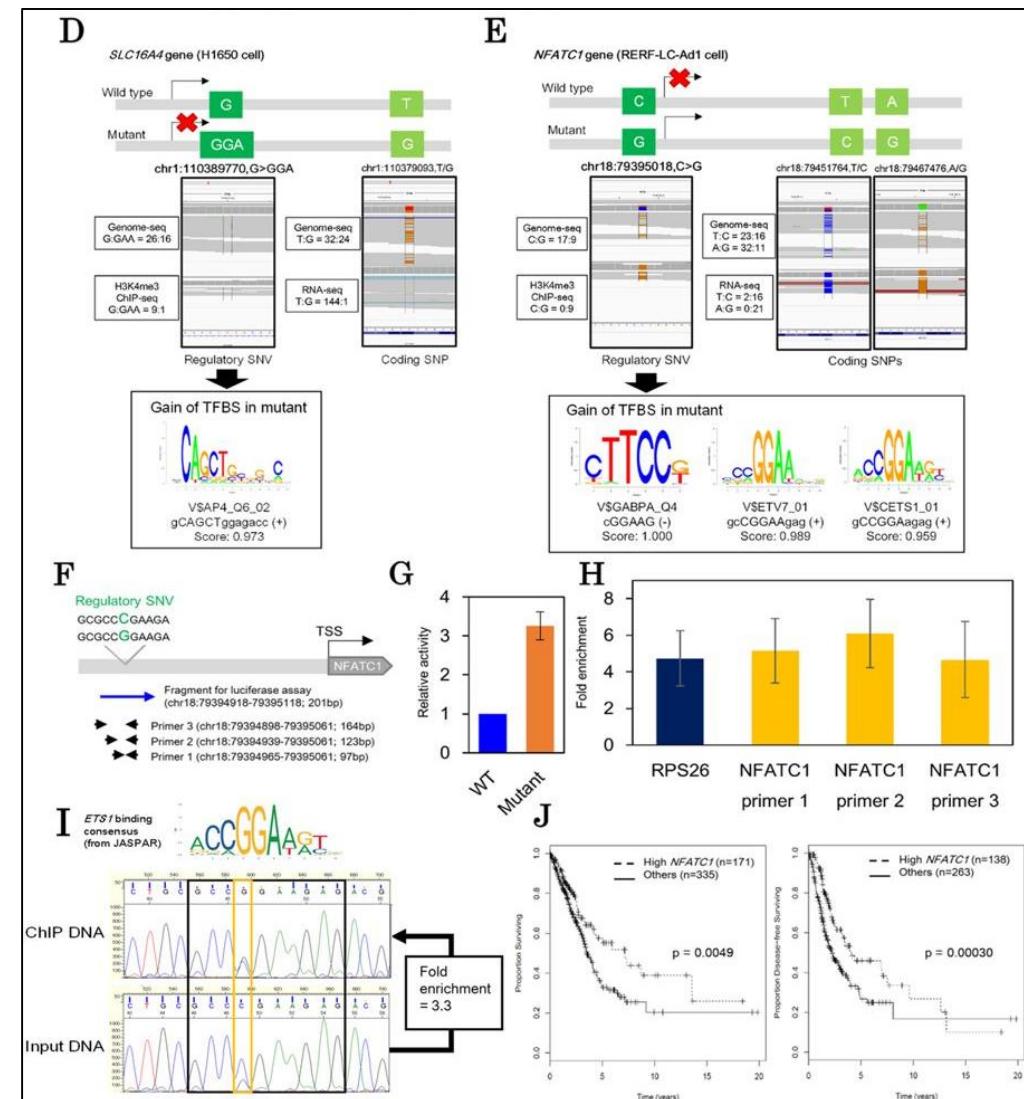
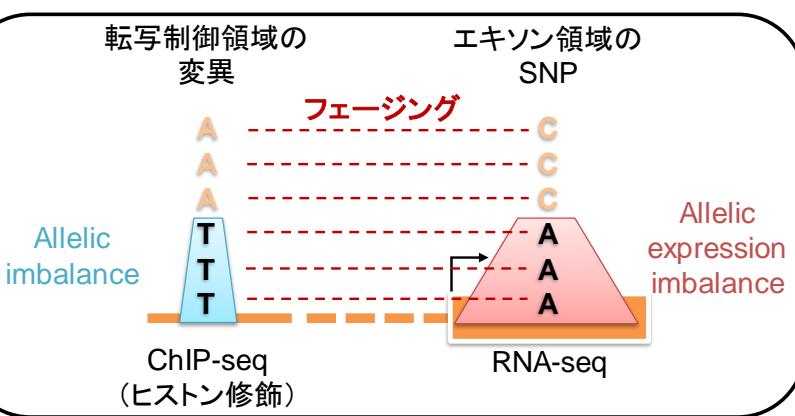


Figure 4; Sereewattanawoot S et al. 2018 *Scientific Reports*

DBKERO: 肺腺癌細胞株のロングリードデータ

New technologies

Type	Tech.	Track Type
<input type="radio"/> Single cell	<input type="radio"/> Linked read	<input type="checkbox"/> BAM
<input checked="" type="radio"/> Long read	<input checked="" type="radio"/> Nanopore (MinION)	

Check Track type



MinION

Linked read



がんゲノム解析関連データベース と解析ツール

がん体細胞突然変異データベースCOSMIC

COSMIC: Catalogue Of Somatic Mutations In Cancer

<https://cancer.sanger.ac.uk/cosmic>

COSMIC v87, released 13-NOV-18

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

eg Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell

Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

- COSMIC** The core of COSMIC, an expert-curated database of somatic mutations
- Cell Lines Project** Mutation profiles of over 1,000 cell lines used in cancer research
- COSMIC-3D** An interactive view of cancer mutations in the context of 3D structures
- Cancer Gene Census** A catalogue of genes with mutations that are causally implicated in cancer

COSMIC News

COSMIC Release v87

The November COSMIC release (v87) is now live! We have 4 newly expert-curated genes, a new fusion pair and a significant update to a further gene. There has been another disease focussed curation, this time on mesothelioma. [More...](#)

Mesothelioma Focus

As part of release v87 we have continued the disease curation focus trialled in the last release and this time we have focussed on updating the expert-curated mutation data for mesothelioma, a rare and malignant cancer. [More...](#)

In the literature

The COSMIC team have been busy writing papers, have a read of our articles in Nature Genetics, Nature Reviews Cancer and NAR to find out about the future of COSMIC. [More...](#)

Tools

- [Cancer Browser](#) — browse COSMIC data by tissue type and histology
- [Genome Browser](#) — browse the human genome with COSMIC annotations
- [CONAN](#) — the COSMIC copy number analysis tool
- [GA4GH Beacon](#) — access COSMIC data through the [GA4GH Beacon Project](#)
- [COSMIC in BigQuery](#) — search COSMIC via the [LSB Cancer Genomics Cloud](#)

DBKEROでもリンクさせていただいております。

Others	1000genome SNP	Clear-cell renal cell carcinoma	dbSNP build 137	NHBLI Exome Sequencing Project	NCBI ClinVar	COSMIC	Meyerson's lung adenocarcinoma
	Myelodysplasia	ICGC ALL	Nagahama Japanese SNP	Japan PGx (Japanese SNP)	ExAC	SCLC (Small Cell Lung Cancer)	26 lung adenocarcinoma cell lines
			Lung adenocarcinoma 97 clinical samples				

Genomic Data Commons (GDC) Data Portal

GDC: <https://portal.gdc.cancer.gov/>

The screenshot shows the GDC Data Portal homepage. At the top, there's a navigation bar with links for Home, Projects, Exploration, Analysis, Repository, Quick Search, Manage Sets, Login, Cart (1), and GDC Apps. Below the navigation is a section titled "Harmonized Cancer Datasets" with the "Genomic Data Commons Data Portal" logo. A "Get Started by Exploring:" section features four buttons: Projects, Exploration, Analysis, and Repository. There's also a search bar with placeholder text "e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2". A "Data Portal Summary" box displays metrics: 43 Projects, 69 Primary Sites, 33,096 Cases, 358,092 Files, 22,872 Genes, and 3,142,246 Mutations. To the right is a large graphic of a human figure with internal organs highlighted in various colors. Next to it is a bar chart titled "Cases by Major Primary Site" showing the count of cases for different cancer types. A "GDC Applications" section lists various tools: Data Portal, Website, Data Transfer Tool, API, Data Submission Portal, Documentation, and Legacy. A sidebar on the right is titled "Program" and lists several programs with counts: TCGA (33), TARGET (6), CTSP (1), FM (1), NCICCR (1), and VAREPOP (1). A "Less..." button is at the bottom right of the sidebar.

NIH NATIONAL CANCER INSTITUTE
GDC Data Portal

Home Projects Exploration Analysis Repository

Quick Search Manage Sets Login Cart 1 GDC Apps

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary Data Release 13.0 - September 27, 2018

Category	Value
PROJECTS	43
PRIMARY SITES	69
CASES	33,096
FILES	358,092
GENES	22,872
MUTATIONS	3,142,246

Cases by Major Primary Site

Cancer Type	Number of Cases
Adrenal Gland	~100
Bile Duct	~100
Bladder	~1,000
Blood	~1,000
Bone	~1,000
Bone Marrow	~100
Brain	~1,000
Breast	~3,500
Cervix	~100
Colon/Rectal	~2,500
Esophagus	~100
Eye	~100
Head and Neck	~1,000
Kidney	~2,000
Liver	~2,000
Lung	~4,000
Lymph Nodes	~100
Nervous System	~1,000
Ovary	~1,500
Pancreas	~100
Pleura	~100
Prostate	~1,000
Skin	~1,000
Soft Tissue	~100
Stomach	~1,000
Testis	~100
Thymus	~100
Thyroid	~100
Uterus	~1,000

GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

Data Portal Website Data Transfer Tool API Data Submission Portal Documentation Legacy

Site Home | Policies | Accessibility | FOIA
U.S. Department of Health and Human Services | National Institutes of Health | National Cancer Institute | USA.gov

Program

- TCGA 33
- TARGET 6
- CTSP 1
- FM 1
- NCICCR 1
- VAREPOP 1

Less...

統合TVに使い方が紹介されています
<https://tогotv.dbcls.jp/20171214.html>

cBioPortal for Cancer Genomics

cBioPortal: <http://www.cbioportal.org/>

Gao et al. Sci. Signal. 2013 & Cerami et al. Cancer Discov. 2012

The cBioPortal for Cancer Genomics provides **visualization**, **analysis** and **download** of large-scale **cancer genomics** data sets. Please cite Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012 when publishing results based on cBioPortal.

QUERY **DOWNLOAD DATA**

Select Studies:

PanCancer Studies	3	<input type="checkbox"/> Select all listed studies (233)
Cell lines	2	<input type="checkbox"/> MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017) 10945 samples
Adrenal Gland	2	<input type="checkbox"/> Pan-Lung Cancer (TCGA, Nat Genet 2016) 1144 samples
Ampulla of Vater	1	<input type="checkbox"/> Pediatric Mixed Tumors (PIP-Seq 2017) 103 samples
Biliary Tract	6	
Bladder/Urinary Tract	12	<input type="checkbox"/> Cancer Cell Line Encyclopedia (Novartis/Broad, Nature 2012) 1020 samples
Bone	2	<input type="checkbox"/> NCI-60 Cell Lines (NCI, Cancer Res. 2012) 67 samples
Bowel	7	
Breast	14	
CNS/Brain	15	

Enter Genes:
Advanced: Onco Query Language (OQL)

User-defined List
Enter HUGO Gene Symbols or Gene Aliases

Submit Query Please select one or more cancer studies.

What's New [@cbioportal](#)

cBioPortal @cbioportal

The ninth phase of the cBioPortal architectural upgrade is complete: the results page is now a single-page application with better performance. You can now also look at gene expression levels in the context of different mutation types, developed in collaboration with [@TheHyveNL](#)

Sign up for low-volume email news alerts

Subscribe

Cancer Studies
The portal contains 233 cancer studies ([details](#))

Cases by Top 20 Primary Sites

Primary Site	Number of Cases
Breast	10945
Lung	1144
CNS/Brain	103
Lymphoid	92
Prostate	92
Kidney	67
Bowel	56
Stomach	48
Bladder	40
Uterus	36
Head/Neck	32
Ovary	28
Myeloid	24
Thyroid	20
PNS	18
Skin	16
Liver	14
Pancreas	12
Soft Tissue	10

cBioPortal for Cancer Genomics

収載しているデータセット

<http://www.cbioportal.org/datasets>

- Mutation, Expression, Survivalデータなどが閲覧できる
- 任意の遺伝子・症例でOncoPrintやKaplan-Meier curveも描画できる

View study summaryを押すと
サマリーが見られる

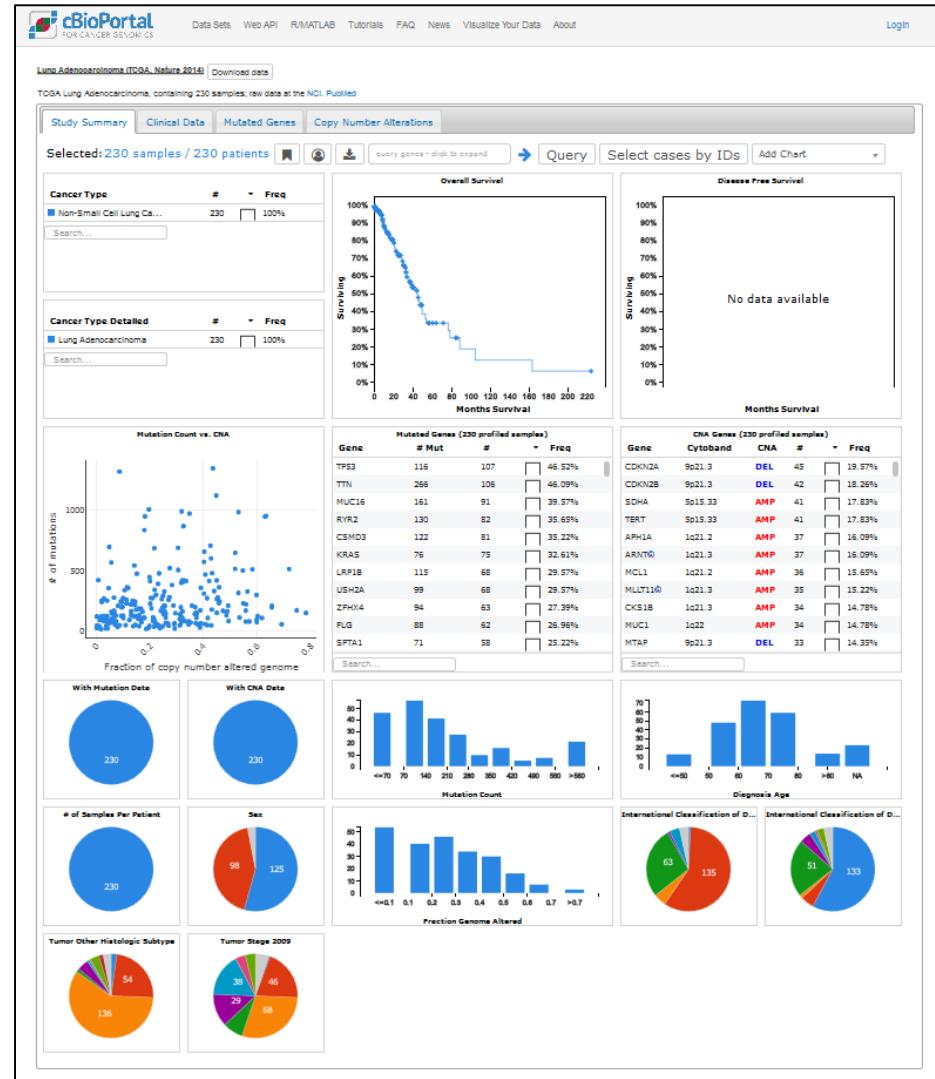
→ LUNG ADENOCARCINOMA

- Lung Adenocarcinoma (Broad, Cell 2012)
- Lung Adenocarcinoma (MSKCC 2015)
- Lung Adenocarcinoma (TCGA, Nature 2014)
- Lung Adenocarcinoma (TCGA, PanCancer Atlas)
- Lung Adenocarcinoma (TCGA, Provisional)
- Lung Adenocarcinoma (TSP, Nature 2008)
- MSK-IMPACT Clinical Sequencing Cohort for Non-Small Cell Cancer (MSK, Cancer Disc...

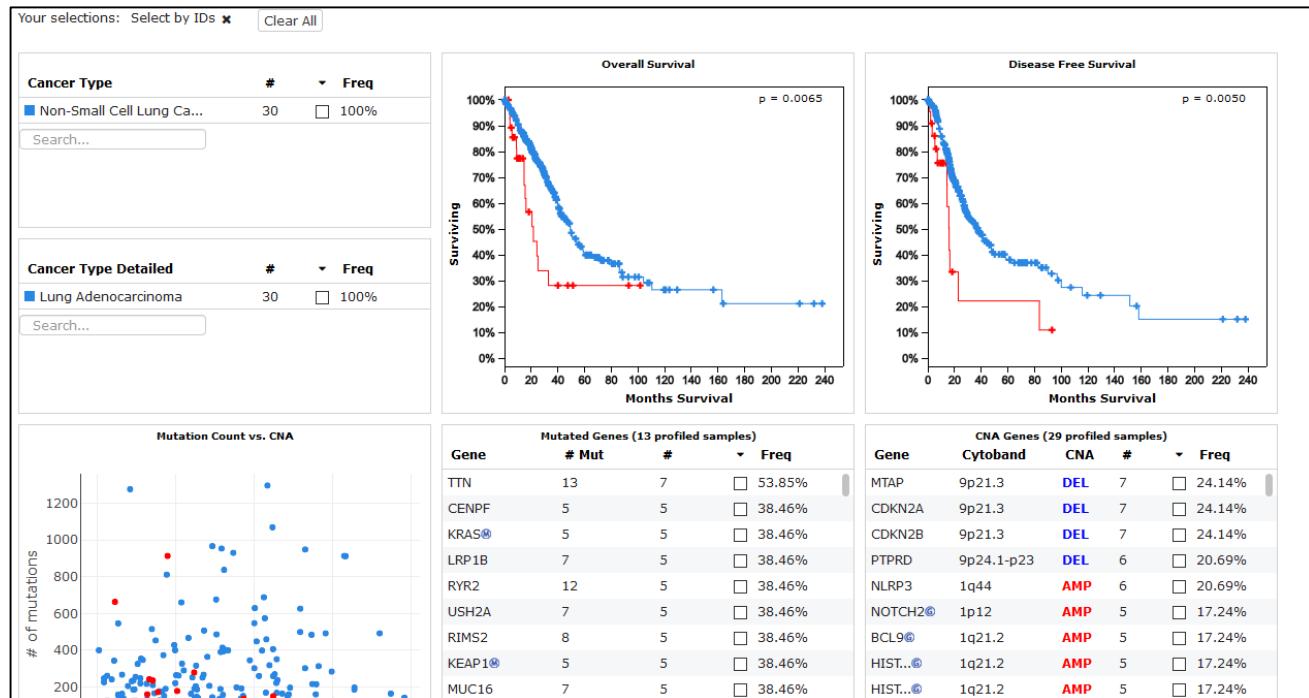
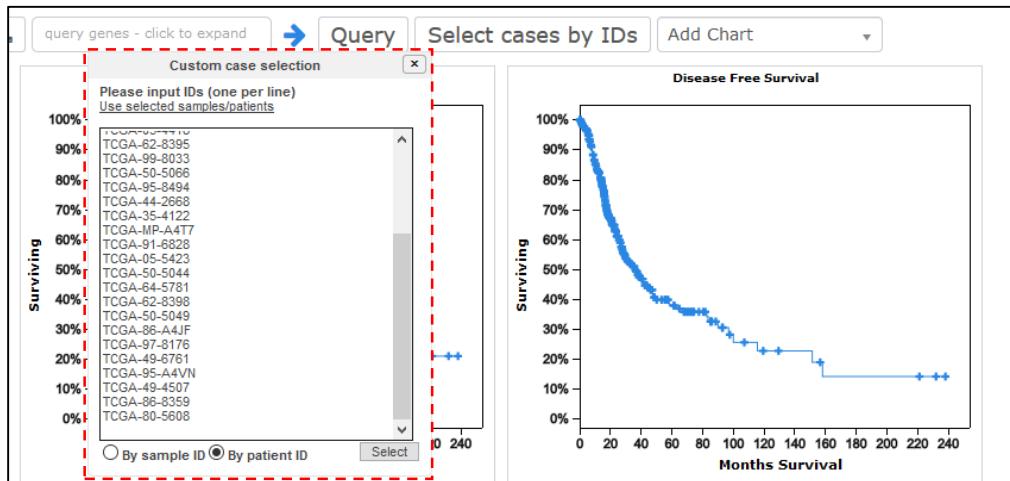
183 samples
35 samples

View study summary

230 samples (1 blue, 1 red, 1 grey)
566 samples (1 blue, 1 red, 1 grey)
586 samples (1 blue, 1 red, 1 grey)
163 samples (1 blue, 1 red, 1 grey)
915 samples (1 blue, 1 red, 1 grey)



cBioPortal for Cancer Genomics



謝辞

ライフサイエンスデータベース
統合推進事業(統合化推進プログラム)

菅野班

東京医科歯科大学

菅野 純夫

東京大学

徳永 勝士

鈴木 穂

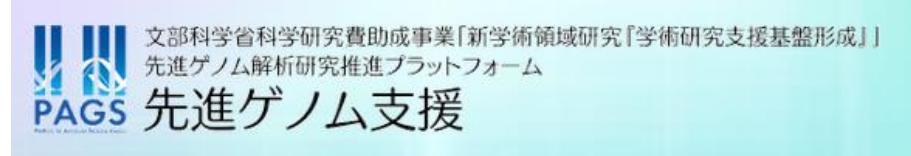
国立がん研究センター

土原 一哉

CREST/IHECメンバーの皆様

DBCLSスタッフの皆様

DBKEROは、東京大学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
システムより公開されています。



AMED-CREST



AMED次世代がん医療創生研究事業(P-CREATE)

ご清聴ありがとうございました