

# ヒト統合オーミクスデータベース DBKERO update 2019

○Yutaka Suzuki<sup>1</sup>, Hiroyuki Wakaguri<sup>2</sup>, Riu Yamashita<sup>3</sup>, Shin Kawano<sup>4</sup>, Katsuya Tsuchihara<sup>5</sup>, Kenta Nakai<sup>6</sup> and Sumio Sugano<sup>1</sup>

<sup>1</sup>Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, <sup>2</sup>Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, <sup>3</sup>Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan, <sup>4</sup>Database Center for Life Science, Research Organization of Information and Systems, Chiba, <sup>5</sup>Division of TR, The Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Chiba, <sup>6</sup>Human Genome Center, The Institute of Medical Science, The University of Tokyo

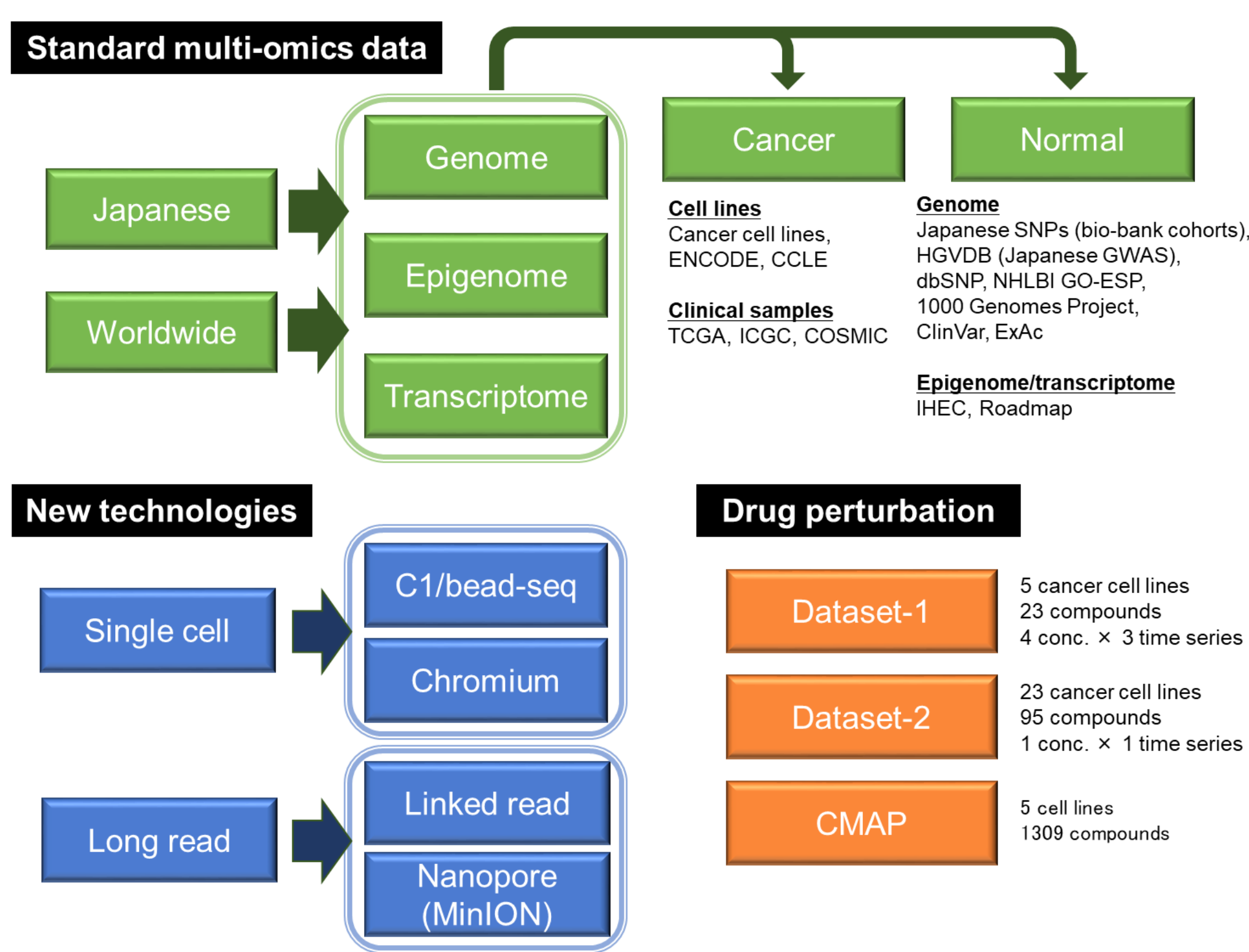
## ABSTRACT

我々のグループで開発するデータベースDBKERO (<http://kero.hgc.jp/>)では、ヒトゲノム多型・変異に生物学的機能注釈を与えるべく、臨床検体、培養細胞モデルの両面から多層オーミクスデータの統合を行っている。臨床検体データについては、今期、本データベースではCRESTが終了するIHEC(国際エピゲノムコンソシアム)データの本邦での受け皿となるべく、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(94検体からなるヒストン修飾、DNAメチル化、RNASeqの合計621データセット)を標準エピゲノムデータとして収録、ブラウザを可能とした。これに加えて東大徳永らの運営するヒトゲノム多型データベースとも融合し、HLA領域を含む多型情報さらに17のGWAS解析より得られた日本人ゲノム多型を正常検体8,757検体、疾患検体6,816検体について公開した。

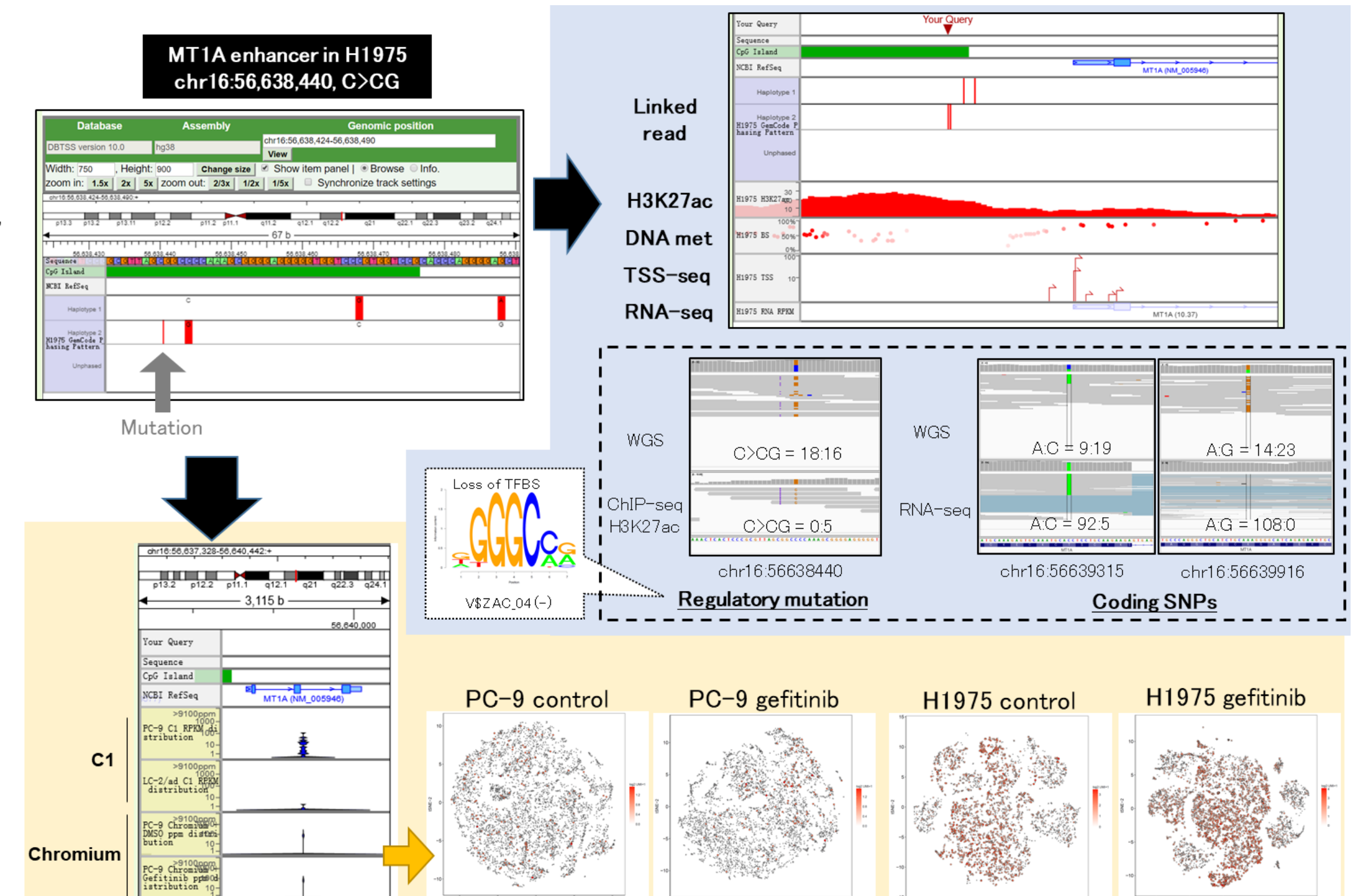
細胞モデルデータについては、約50種類のシングルセルデータ及びMinIONを用いて収集したヒトがんゲノムロングリードを公開した。また、これらのデータについて、統合オーミクス解析を可能とする一連のツール群についても公開を開始している。特に筆者らが今回公開した発現モジュール解析例について、肺腺癌細胞モデルに詳解する。

**Database of Kashiwa Encyclopedia for human genome mutations in Regulatory regions and their Omics contexts (KERO@<http://kero.hgc.jp/>)**

## KEROデータコンテンツのスキーマ(A)



## シングルセルとロングリードのデータ閲覧(B)

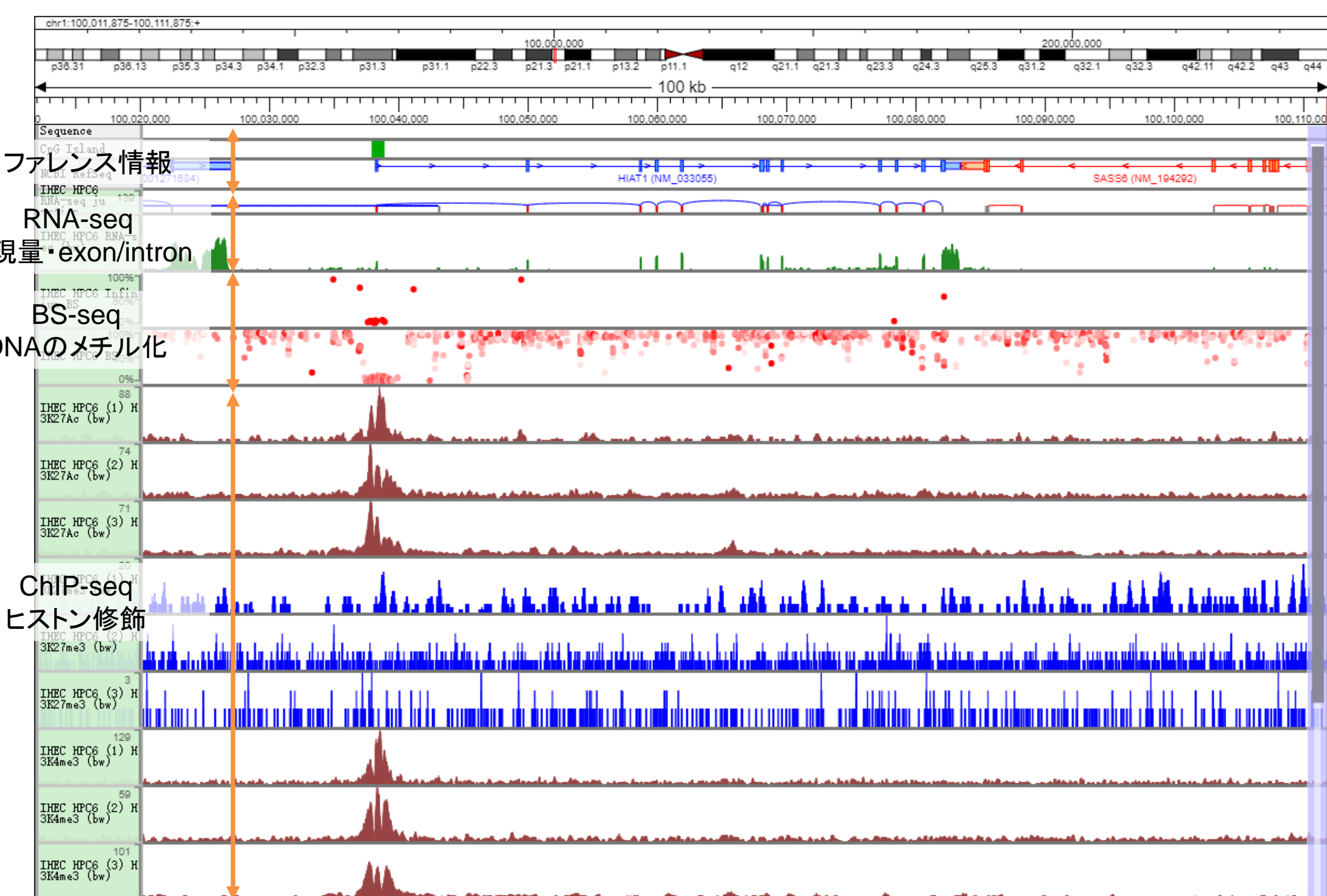


(A) DBKEROのデータ構造の全体像。日本人の臨床オミクス情報がモデルシステムからの包括的オミクス情報とどのように関連しているかを示す。New technologiesとして、近年新たに産出可能となったシングルセルとロングリードのデータやDrug perturbationとして、化合物によるマルチオミクス摂動の情報を含んでいる。異なるカテゴリーのデータセットは、異なる色のボックスとして示した。

(B) 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閲覧することができる。ヘルプ詳細は次のURLをから ([http://kero.hgc.jp/?doc:help\\_2017.html](http://kero.hgc.jp/?doc:help_2017.html)) のHow To Use Example-2: A MT1A promoter mutation represented by new genomic technologiesの項)。

## IHEC(国際エピゲノムコンソシアム)データの公開

HPC6肝臓細胞のマルチオミクス情報の表示



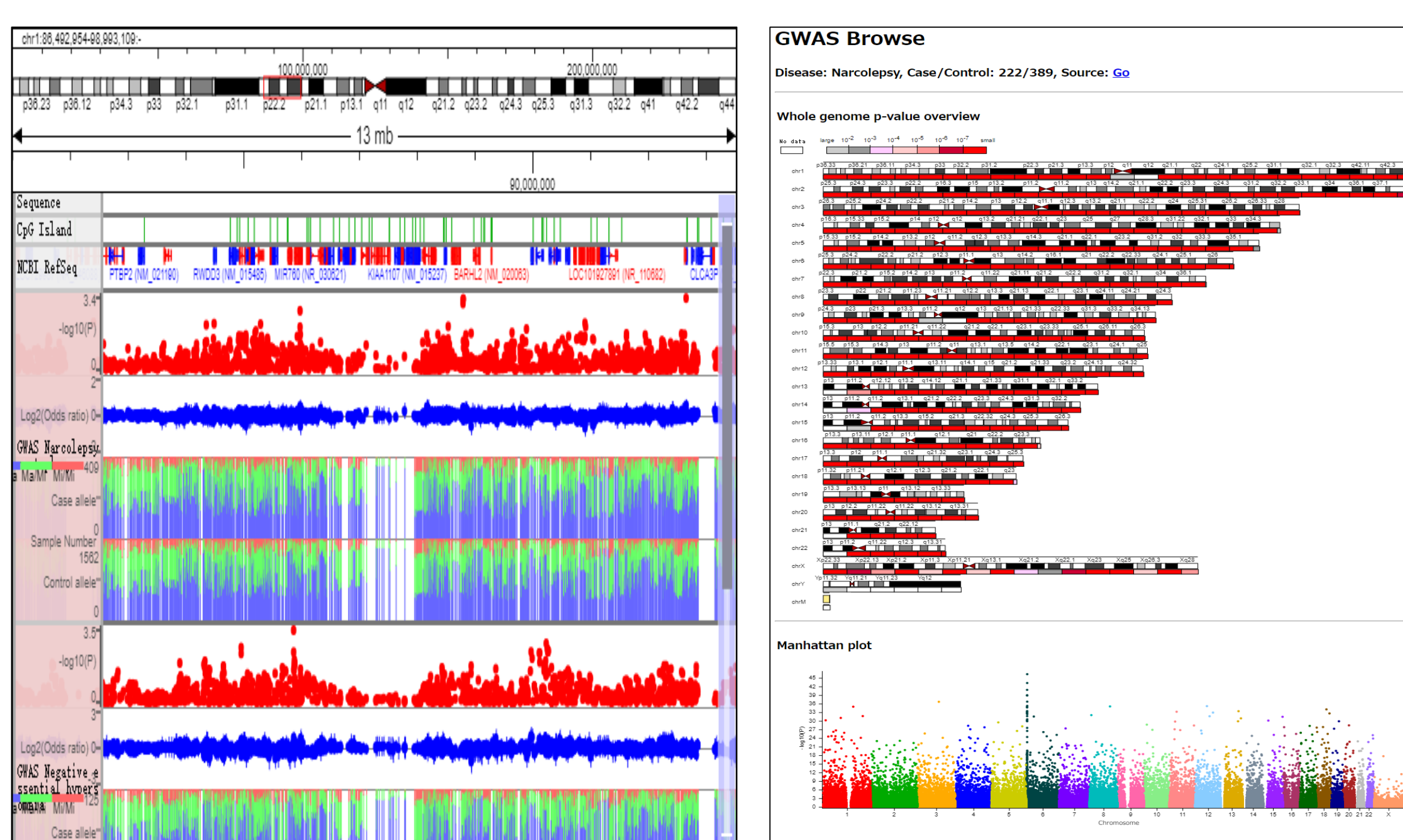
IHECデータとしてゲノムブラウザから閲覧可能なオープンデータ一覧

Sample	individuals	Expression		DNA methylation		Histone modifications					
		RNA-seq	DNA methylation	BS-seq	H3K4me1	H3K4me3	H3K27ac	H3K27me3	H3K36me3	H3K9me3	
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録した。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(左図)。

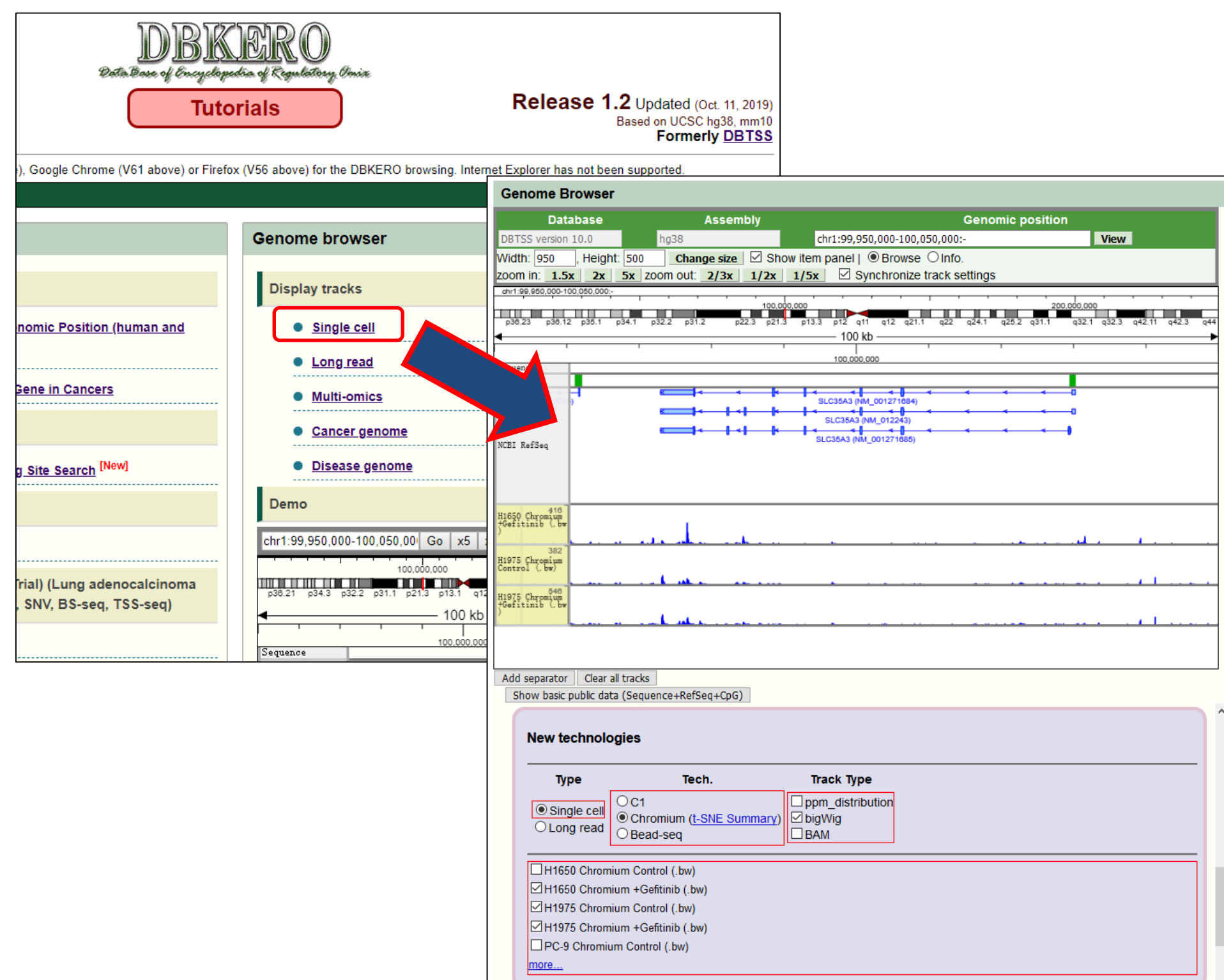
## GWASデータの可視化



日本人のGWASデータとして、ナルコレプシー、PBC(原発性胆汁性胆管炎)等のデータを搭載。マンハッタンplot viewからゲノム全体からの俯瞰的に閲覧できるブラウザを開発・公開した。

[https://dbtss.hgc.jp/wakaguri/whole\\_genome\\_index.html?SJS117CTRL691\\_44920](https://dbtss.hgc.jp/wakaguri/whole_genome_index.html?SJS117CTRL691_44920)  
[5SNP\\_info.hg38\\_cut.bed.gz](https://dbtss.hgc.jp/wakaguri/5SNP_info.hg38_cut.bed.gz)

## ユーザインターフェースの改良



トップページ(<http://kero.hgc.jp/>)からのユーザインターフェースの改良を行っている。今年度は、トップページにDisplay tracksのボックスを設けた。近年のデータカテゴリーの増大によって、DBKEROにどのようなデータが搭載されているのか、把握したり目的のデータにたどり着くのが困難になってきていた。今年度Display tracksを設けることで、興味対象の研究カテゴリーからデータベースに搭載されているデータの選択をしやすくした。

## その他、RDF化も順次進め、KEGGパスウェイAPIを使ったパスウェイマップ等も公開しています

What and Why RDF?

- RDF: Resource Description Framework
- Represented as "Subject-Predicate-Object" triple
- All resources except literals have URIs (URLs)
- Fundamental technology of Semantic Web together with OWL ontologies
- Standardized by W3C
- Query language is available (SPARQL)
- Federated query
- RDF resources can join in LOD network

## ChIP RDF Schema

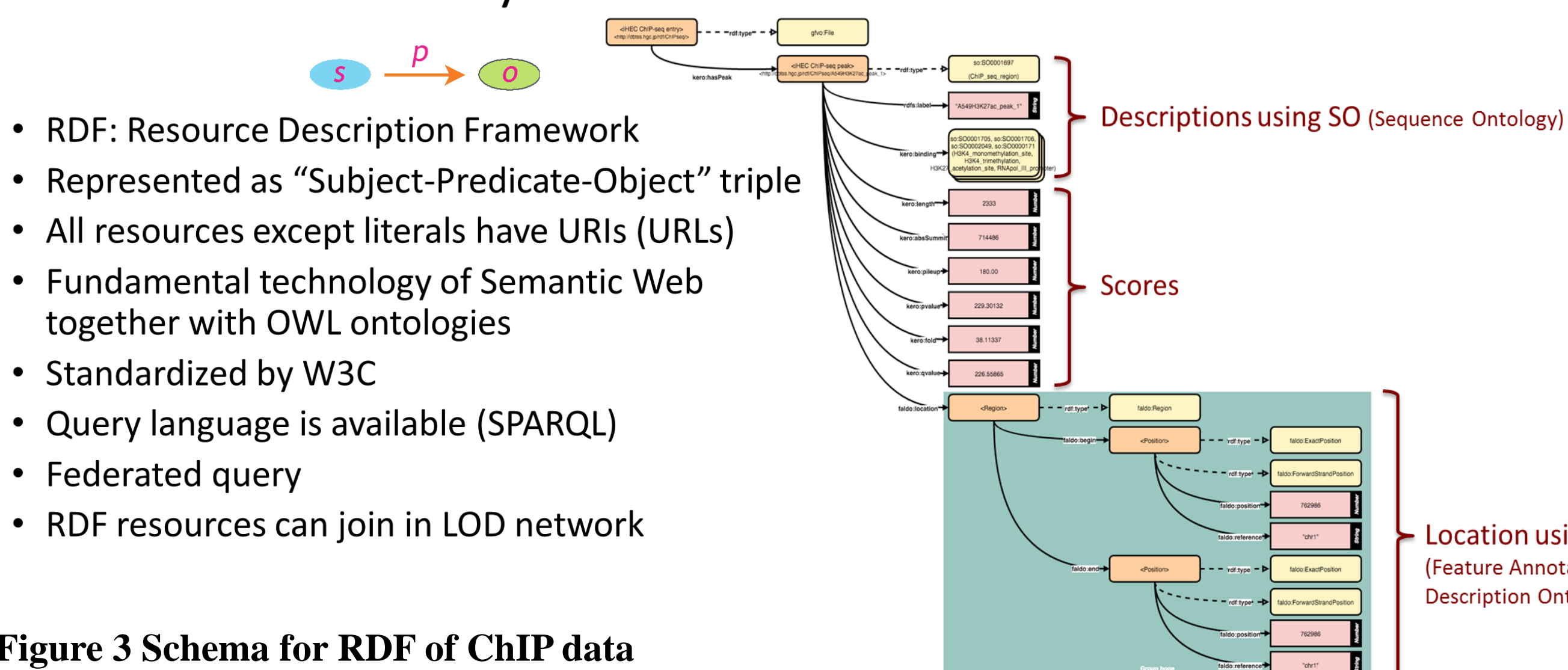
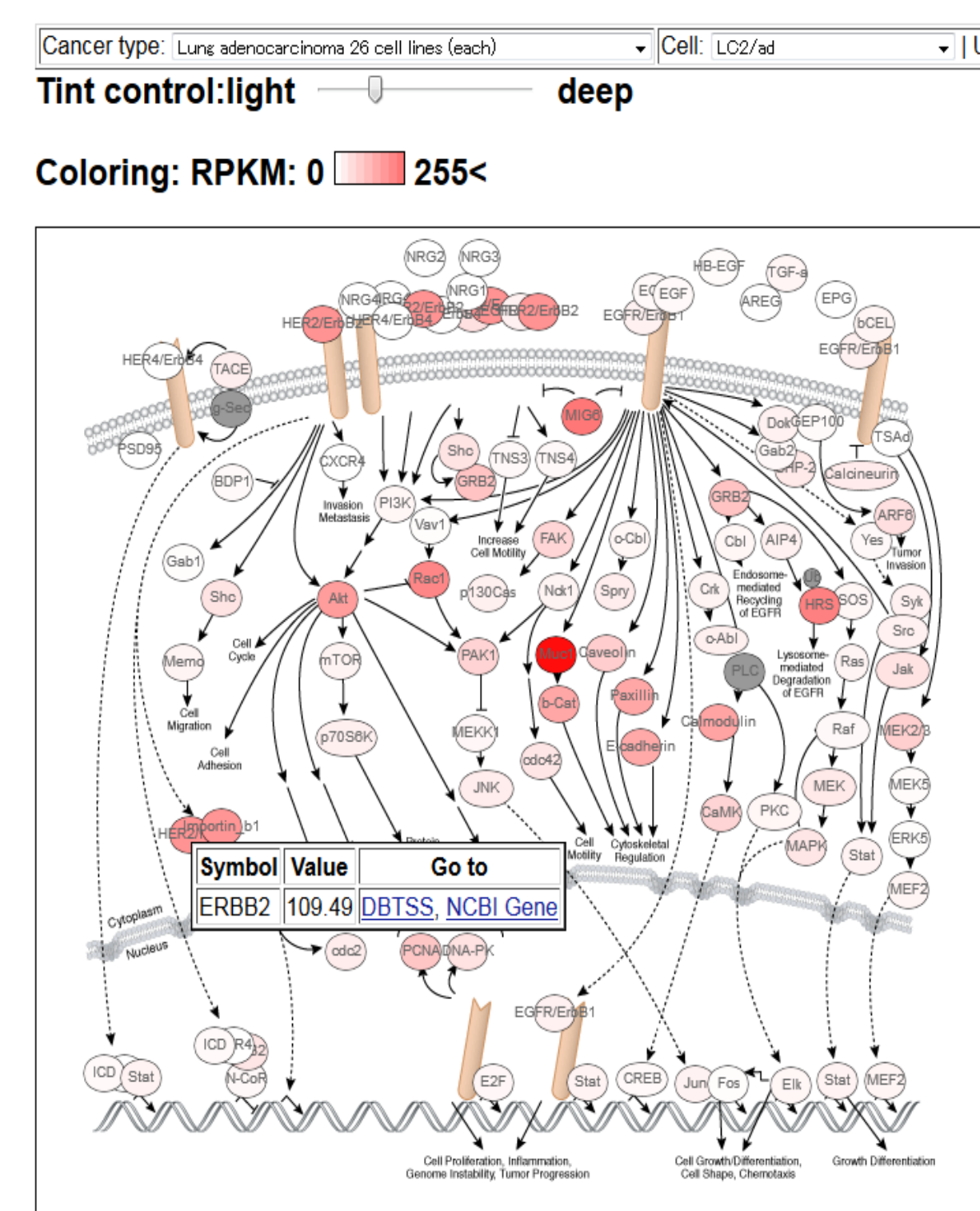


Figure 3 Schema for RDF of ChIP data

Description	Mutation frequency		
	Upstream distal: -50k to -1k	Upstream proximal: -1k to -1	Gene body
Clear cell renal cell carcinoma by Dr. Ogawa's Lab.	0/100	0/100	0/100
K562 Acute Myeloid Leukemia - IS	1/78	1/78	1/78
K562 Acute Myeloid Leukemia - TARGET US	0/2	0/2	0/2
K562 Bladder Urothelial Cancer - TCGA US	1/128	1/128	1/128
K562 Bladder Urothelial Cancer - CN	1/103	1/103	1/103
K562 Bone Cancer - UK	0/95	0/95	0/95
K562 Brain Glioblastoma Multiforme - TCGA US	5/288	5/288	5/288
K562 Brain Lower Grade Glioma - TCGA US	1/288	1/288	1/288
K562 Breast Cancer - TCGA US	1/943	1/943	1/943
K562 Breast Triple Negative/Lobular Cancer - UK	1/117	1/117	1/117
K562 Chronic Lymphocytic Leukemia - ES	1/109	1/109	1/109
K562 Chronic Myeloid Leukemia - UK	0/129	0/129	0/129
K562 Colon Adenocarcinoma - TCGA US	25/216	25/216	25/216
K562 Early Onset Prostate Cancer - DE	1/11	1/11	1/11
K562 Esophageal Adenocarcinoma - UK	1/16	1/16	1/16
K562 Esophageal Cancer - CN	0/88	0/88	0/88
K562 Gastric Adenocarcinoma - TCGA US	3/299	3/299	3/299
K562 Gastric Cancer - CN	0/9	0/9	0/9
K562 Head and Neck Thyroid Carcinoma - TCGA US	232/361	232/361	232/361
K562 Kidney Renal Clear Cell Carcinoma - TCGA US	1/64	1/64	1/64

## ErbB / HER Signaling

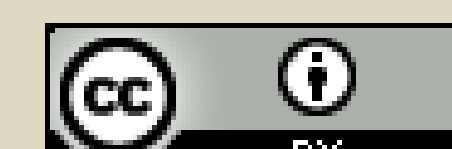


パスウェイマップ。肺腺癌26培養細胞において、発現情報を可視化したもの。各ノードが1つの遺伝子に対応し、色の濃淡でRNA-seq解析で得られた発現量を表現している。

各ノードをクリックすることで、その遺伝子周辺のゲノムブラウザやNCBIのGeneの情報へリンクし、詳細を確認できる。

パスウェイ図は Cell Signaling Technology社の図及び、KEGG Pathwayの図を利用している。

All contents and raw data freely downloadable from KERO at: <http://kero.hgc.jp/>



Licensed under a Creative Commons表示4.0国際ライセンス (c)2019 Yutaka Suzuki (Graduate School of Frontier Sciences, The University of Tokyo)