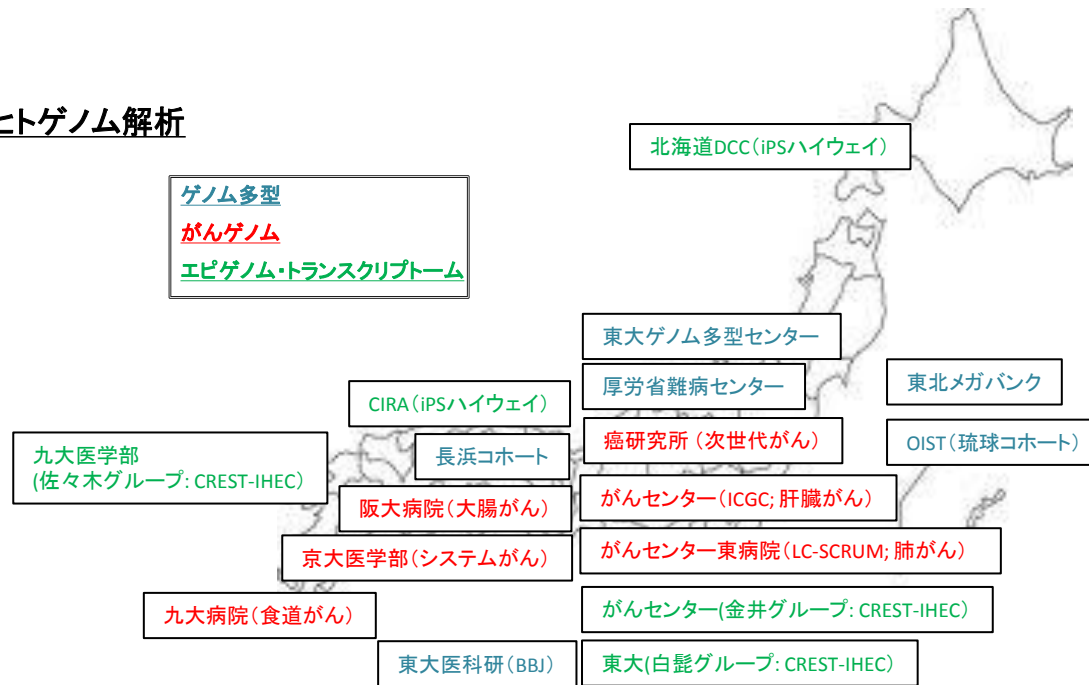


疾患ヒトゲノム変異の 生物学的機能注釈を目指した 多階層オミクスデータの統合

東京大学新領域創成科学研究科
菅野純夫



全国に展開するヒトゲノム解析



ゲノムデータは急速に蓄積している

ヒトオミクスデータ推定蓄積量

ゲノム多型(WGS/WES): >2000人

がんゲノム(WGS/WES/Target Seq): >1000症例

トランスクリプトーム(RNA Seq): >1000例

エピゲノム(BS/ChIP Seq): <100例

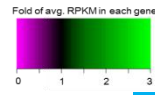
+培養細胞+PDX+モデル系:>5000例

+マウス等モデル生物: ???例

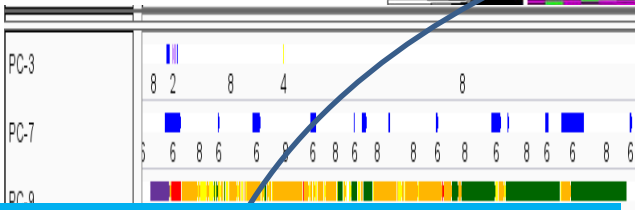
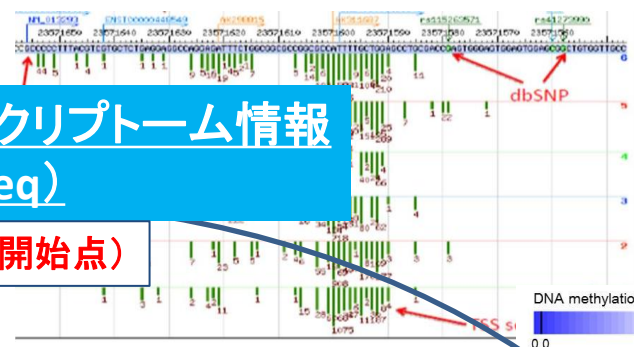
+個別研究者の蓄積するオミクス情報: ???例



ヒト応用研究を志向したオミクス情報の統合 (EGFR遺伝子を例に)

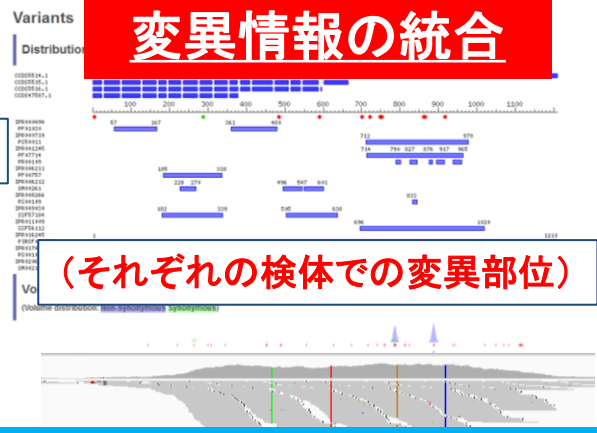


転写開始点/トランスクリプトーム情報 (TSS/RNA Seq)
(発現量と転写開始点)



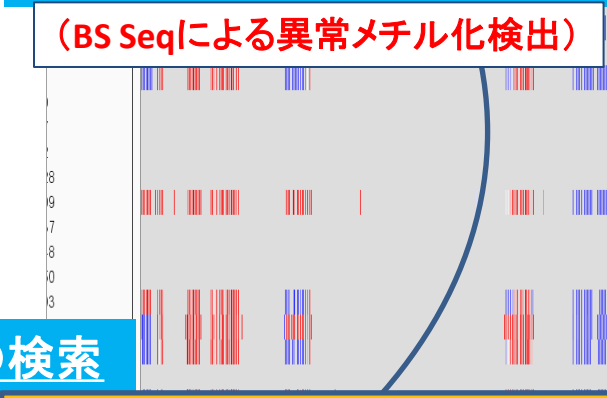
クロマチン情報 (ChIP Seq)
(ChrHMMパターンで示すヒストン修飾)

ヒトゲノム 変異情報の統合



(それぞれの検体での変異部位)

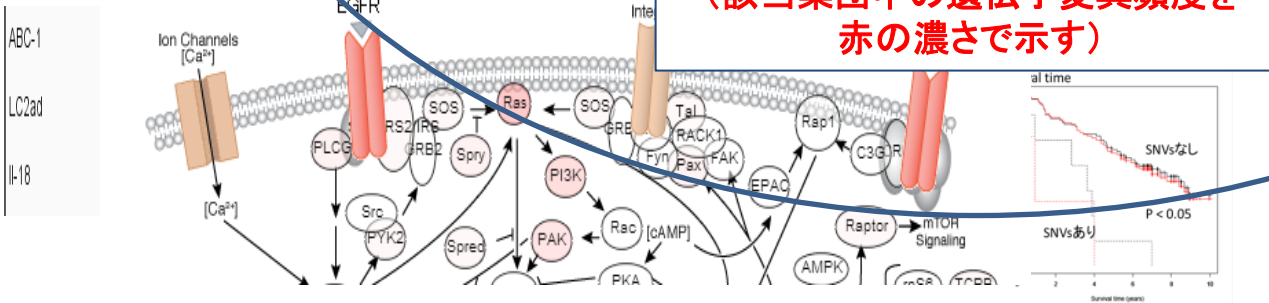
DNAメチル化情報 (BS Seq)
(BS Seqによる異常メチル化検出)



- 1 Active promoter
- 2 Weak/poised promoter
- 3 Strong enhancer
- 4 Weak enhancer
- 5 Transcriptional elongation
- 6 Inactive region
- 7 Inactive region/heterochromatin
- 8 Low/no signal

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)



モデル系とのさらなる統合

研究開発体制と役割

菅野: 研究の統括/医科研スパコンの運用

鈴木: データの加工とデータベース設計

土原: ヒト疾患応用研究を志向した検索システムの実装

河野: データベースの構築、オントロジーの整備

+ 研究員 3名
+ 技術補佐員 2名
(雇用予定)

鈴木(東大・柏キャンパス)

土原(がんセンター東病院)

河野(DBCLS柏オフィス)

菅野(東大・医科研)

地理的に密接したエリア:
データ交換の利便性

データ統合対象

	名称	件数 (保有しているもの)	概要
1	ヒトがんゲノムデータ	300	肺腺・小細胞がん、食道がん、胃がん、大腸がんの全ゲノム/エキソームデータ
2	ヒト正常ゲノムデータ	500	申請者らの収集した正常日本人全ゲノム/エキソームデータの頻度情報
3	ヒトエピゲノム、トランスクリプトームデータ	300	肺腺がん、大腸がん: 正常肝臓(予定)、正常大腸(予定)、iPS/ES細胞
4	ヒト培養細胞オミクスデータ	30 (300計測点)	肺腺がん、大腸がん(予定)、胃がん(予定)、初代培養正常細胞・PDX(予定)
5	マウスオミクスデータ	10 (400計測点)	正常組織、初代培養マクロファージ、B細胞等の免疫担当細胞(FIRST審良P)
6	* その他のモデル動物データ	100 (500計測点)	「ゲノム支援」で収集、公開予定のデータ

用いるすべてのデータは申請者らがデータ産出、1次データ加工を実際に行ったものである。

九州大学病院別府病院
Keio University Beppu Hospital

難病情報センター
次世代遺伝子解析装置を用いた、難病の原因究明、治療法開発プロジェクト

文部科学省科学研究費新学術領域研究
ゲノム支援

国際ヒトエピゲノムコンソーシアム 日本チーム

最先端研究開発支援プログラム Funding Program for World-Leading Innovative R&D on Science and Technology
免疫ダイナミズムの統合的理解と免疫制御法の確立

ゲノム支援
(DBJより公開になり次第、収録許可をデータ所有者に打診)

- 支援課題数 129
- ライブラリ内訳
- ChIP : 753
- BS : 256
- RNA : 1705
- smallRNA : 40
- その他 (DNA, TSSなど) : 164
- 生物種内訳
- human : 185
- mouse : 1179
- その他 動物 : 790
- 植物・微生物・細菌 : 764

設備

メインサーバー: 医科研ヒトゲノムセンタースパコンシステム

* モデル系から得られたオミクスデータに焦点。

特に培養細胞40種類:
ゲノム-エピゲノム-トランスクリプトームのオミクスデータが同一の材料から得られたもの

比較ゲノムブラウザー
によりヒトゲノムと対応づけ

データ統合

+ 各サイトに購入予定のサーバー

構成名	機種名	Rpeak (Gflops)/node	CPU	OS	CPUコア数 (1nodeあたり)	メモリ /node	ディスク容量/node
大規模データ解析サーバ	DELL PowerEdge R720XD		インテル® Xeon® プロセッサ E5-2650 シュ、8.0GT/s QPL、ターボ、8C、95W) x2	RedHat Enterprise Linux	16core	128GB	27TB

既存のサーバーの利用: 最低限の設備投資での計画遂行

平成26年度の成果

- ・データベースの公開、論文化
(ヒト培養細胞の多階層オミクスデータに焦点)
- ・ヒトゲノム多型データの充実
(特に日本人5000人分)
 - ・ToMMo
 - ・PGx
 - ・ExAc

データベースの公開

- DataBase of Transcriptional Start Sites -
DBTSS

Release 9.0 Updated (September 15 2014)
Based on UCSC hg38, mm10

The screenshot displays the DBTSS website interface. On the left, there are several search panels: 'Database Search' with a keyword search box and species dropdown; 'Human Chromatin Features' with a search box for genomic position; 'Search from SNP (dbSNP)' with a search box for rs375229889; 'Search from somatic mutation' with a search box for BRAF; 'Search from SNV-enriched Gene in Cancers' with a dropdown menu for Lung adenocarcinoma. The main content area is titled 'About this database' and contains text describing the database's purpose and content. A 'News' section on the right mentions a September 15, 2014 update. Overlaid on the bottom half of the screenshot is a browser window showing a genomic track visualization with various data layers and a search bar. A large blue arrow points from the search bar area towards the right side of the image.

Genome

Transcriptome

Epigenome

DNA methylation

Histone modifications

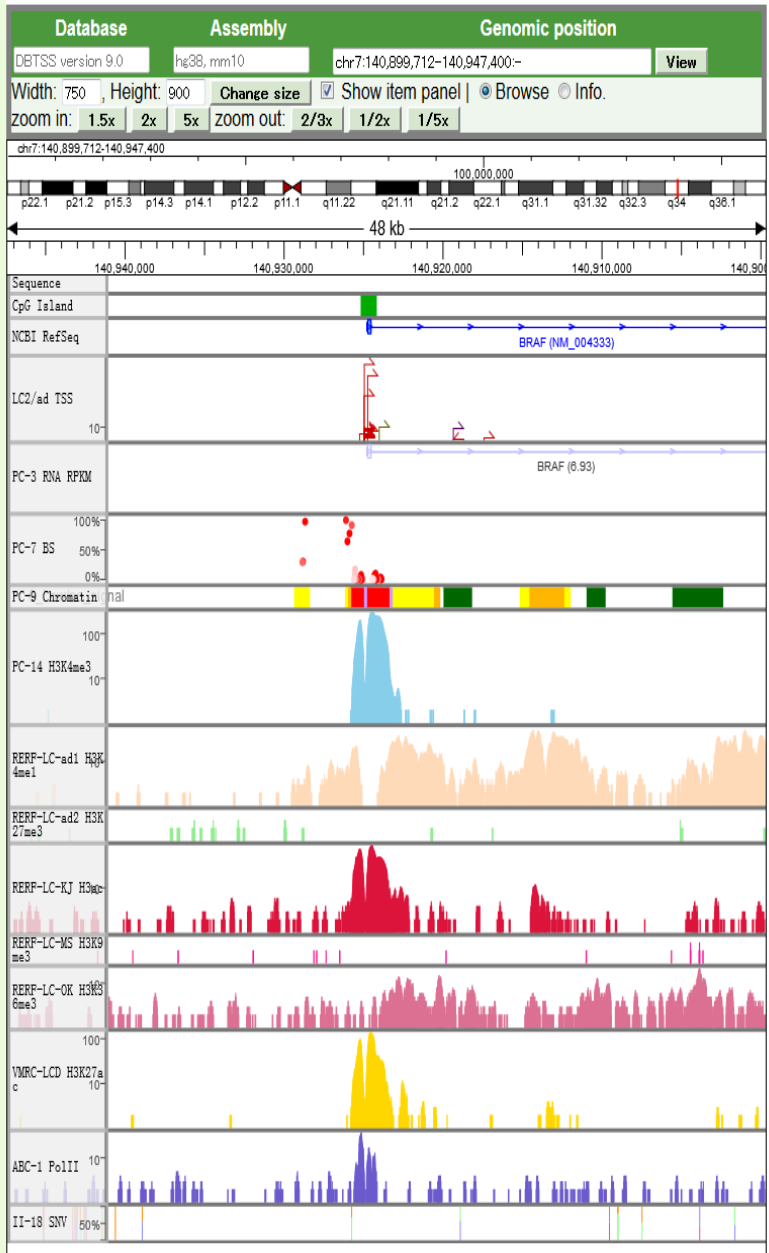
Further integration with

Image data

Clinical information

Search for novel/more useful biomarkers at "bulk" level

Database for multi-omics analysis

C**Genome viewer (Multi-omics Data)****D****Viewer contro**

- Gene model**
- TSS-Seq**
- RNA-Seq**
- BS-Seq**
- ChromHMM**

- H3K4me3**
- H3K27me3**
- H3K9Ac**
- H3K27Ac**
- H3K4me**
- H3K9me3**
- H3K36me3**
- Pol II**
- SNV**

ChIP-Seq

Database Search

Keyword Search

Species:
H. sapiens

Keyword:
NM_*

Search

Lift over:

hg38

chr1:99,950,000-100,050,000

Search

Human Chromatin Features

Search

Search from Genomic Position:

chr1:75,787,000

Search

Search from SNP (dbSNP rsID):

rs375229889

Search

Search from SNV (COSMIC: somatic mutation):

BRAF

Search

Search from SNV-enriched Gene in Cancers:

Lung adenocarcinoma ...

Search

MAPK/Erk in Growth and Differentiation

Cancer type: Lung adenocarcinoma 26 cell lines (each)

Cell: LC2/ad

Unit: RPKM

- RPKM log2 fold against normal cell
- H3K4me3 (proximal)
- H3K4me1 (distal)
- H3K27ac (distal)
- PoIII (proximal)
- H3K36me (gene body)
- H3K27me3 (gene body)
- H3K27me3 (distal)

Tint control: light deep

Coloring: RPKM: 0 255<

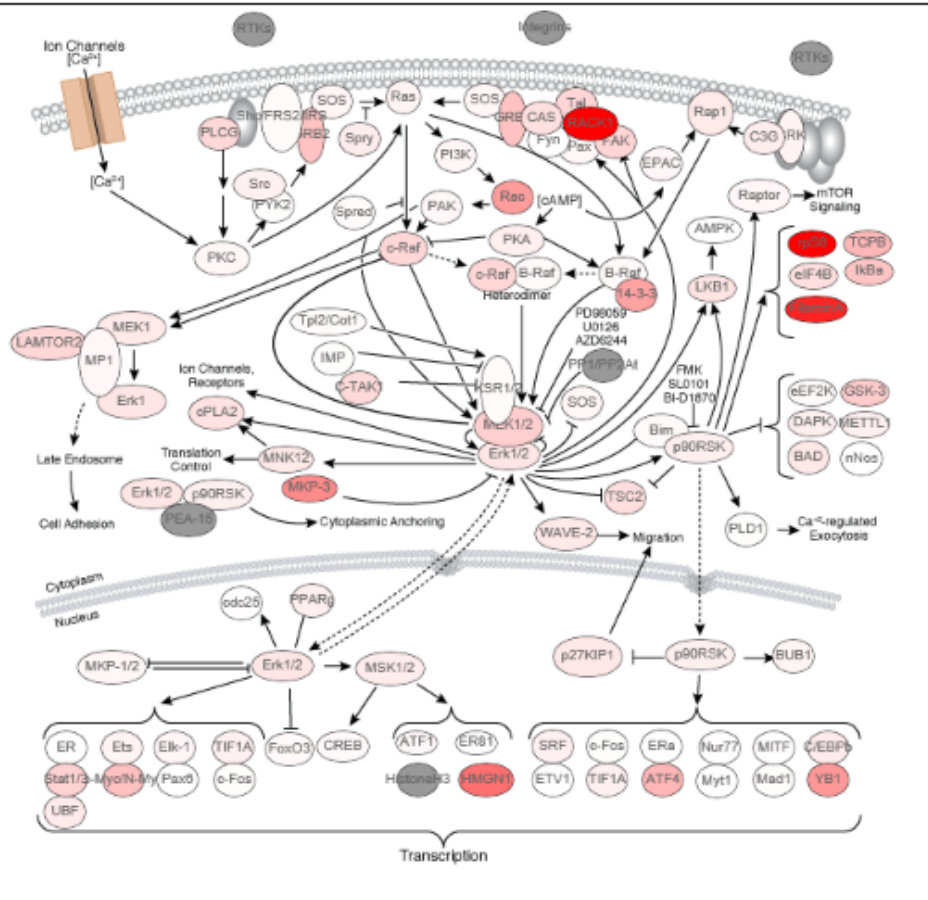


Illustration reproduced courtesy of Cell Signaling Technology, Inc.

DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data

Ayako Suzuki¹, Hiroyuki Wakaguri², Rlu Yamashita³, Shin Kawano⁴, Katsuya Tsuchihara⁵, Sumio Sugano¹, Yutaka Suzuki^{2,*} and Kenta Nakai^{6,*}

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

Received September 15, 2014; Revised October 16, 2014; Accepted October 16, 2014

ABSTRACT

DBTSS (<http://dbtss.hgc.jp/>) was originally constructed as a collection of uniquely determined transcriptional start sites (TSSs) in humans and some other species in 2002. Since then, it has been regularly updated and in recent updates epigenetic information has also been incorporated because such information is useful for characterizing the biological relevance of these TSSs/downstream genes. In the newest release, Release 9, we further integrated public and original single nucleotide variation (SNV) data into our database. For our original data, we generated SNV data from genomic analyses of various cancer types, including 97 lung adenocarcinomas and 57 lung small cell carcinomas from Japanese patients as well as 26 cell lines of lung cancer origin. In addition, we obtained publically available SNV data from other cancer types and germline variations in total of 11,322 individuals. With these updates, users can examine the association between sequence variation pattern in clinical lung cancers with its corresponding TSS-seq, RNA-seq, ChIP-seq and BS-seq data. Consequently, DBTSS is no longer a mere storage site for TSS information but has evolved into an integrative platform of a variety of genome activity data.

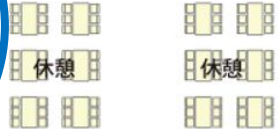
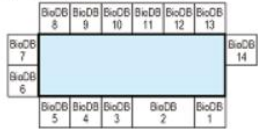
To know accurate TSS positions is valuable as it could lead to more accurate characterization of its upstream transcriptional regulatory region. Thus, we constructed a database containing such information of mostly human genes in 2002. Since then, its updates have been regularly reported in the *Nucleic Acids Research* database issues (2004, 2006, 2008, 2010 and 2012 (4)). With the advances in sequencing technologies, we have developed TSS-seq, where the oligo-capping technique is applied to next generation sequencing (NGS), allowing even more accurate genome-wide determination of TSSs (5). The NGS sequencers are not only suited for determining genomic DNA sequences but also for transcriptome analysis (RNA-seq (6)) and epigenome analysis (ChIP-seq (7) and bisulfite sequencing (BS-seq; (8))) Since such additional data enable further biological characterization of transcriptional regulatory regions, we have also incorporated transcriptomic/epigenomic data of various tissues/cell cultures in DBTSS. In the latest update, Release 9, it contains 1257 million TSS tag sequences collected from 24 tissues and 33 cell cultures (see Table 1). It also contains the data of subcellularly fractionated RNAs as well as the ChIP-seq data of various histone modifications, binding sites of RNA polymerase II and several transcription factors, mainly, in cultured cell lines.

In this report, we introduce Release 9 of DBTSS, where we significantly enlarged the number of incorporated single nucleotide variation (SNV) data, which were both collected from publically available databases and generated from our own experiments (see below). The association of such large-

NBDCバイオDBコーナー

データベースのデモをやっています。

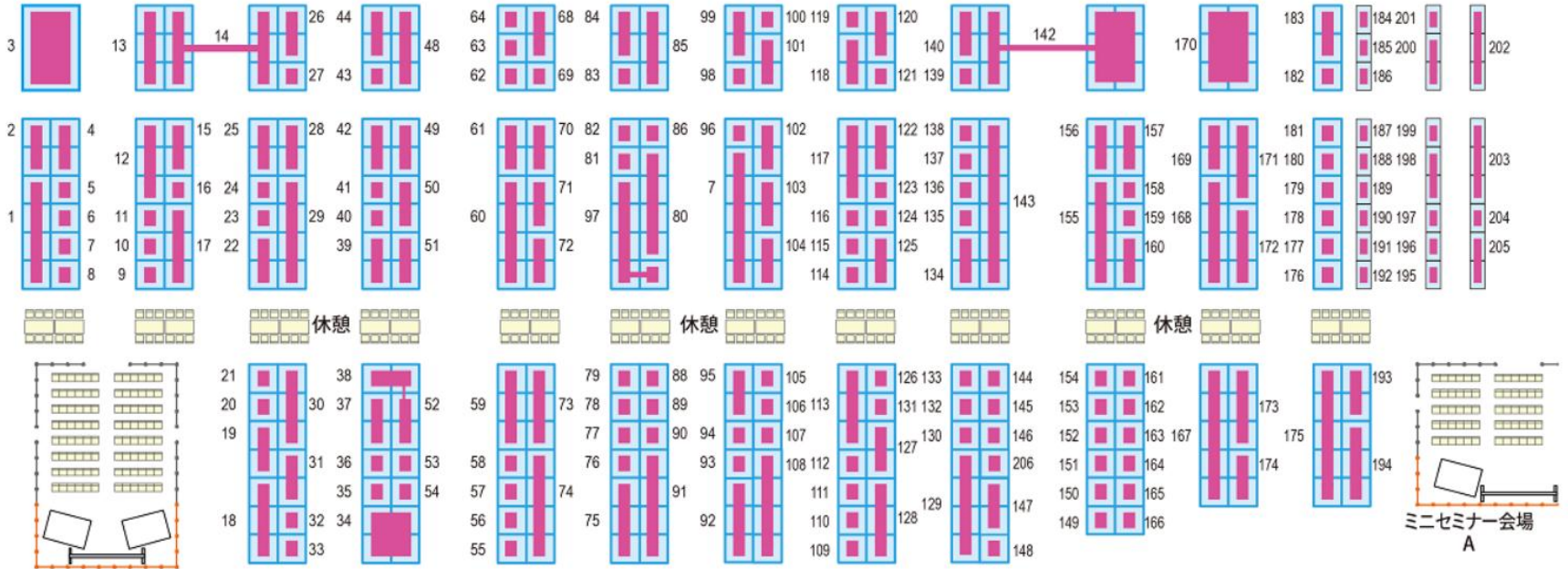
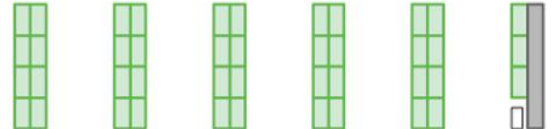
特別企画：
バイオデータベースコーナー
(バイオDBコーナー)



フード・ドリンク
コーナー



特別企画：
ナショナルバイオリソースプロジェクト (NBRP) コーナー



Data contents (ChIP Seq and others)

Cell lines

Data contents

Cell line			General information*						Cell culture		Sequencing dataset					Memo
Name	Type	in-house analysis number	Distributor	Catalogue number	Ethnicity	Gender	Age	Smoking status	Medium	Dish	Whole-genome	RNA-Seq	BS-Seq	ChIP-Seq**	TSS-Seq	
SAEC (control)	Small airway epithelial cell	s_35	TAKARA (Lonza)	-	-	M	-	-	-	collagen Type I-coated	-	+	-	+	+	Normal
PC-9	lung adenocarcinoma	s_9	RIKEN BRC	RCB4455	Japanese	-	-	-	RPMI	-	+	+	+	+	+	-
PC-14	lung adenocarcinoma	s_10	IBL	-	Japanese	-	-	-	RPMI	-	+	+	+	+	+	-
RERF-LC-KJ	lung adenocarcinoma	s_13	RIKEN BRC	RCB1313	Japanese	M	78	-	RPMI	-	+	+	+	+	+	-
RERF-LC-Ad1	lung adenocarcinoma	s_11	JCRB	JCRB1020	Japanese	M	70	-	RPMI	-	+	+	+	+	+	-
RERF-LC-Ad2	lung adenocarcinoma	s_12	JCRB	JCRB1021	Japanese	M	-	-	RPMI	-	+	+	+	+	+	-
LC2/ad	lung adenocarcinoma	s_18	RIKEN BRC	RCB0440	Japanese	F	51	-	DMEM	collagen Type I-coated	+	+	+	+	+	-
RERF-LC-MS	lung adenocarcinoma	s_14	JCRB	JCRB0081	Japanese	-	-	-	EMEM	-	+	+	+	+	+	-
VMRC-LCD	lung adenocarcinoma	s_16	JCRB	JCRB0814	Japanese	M	-	-	EMEM	-	+	+	+	+	+	-
ABC-1	lung adenocarcinoma	s_17	JCRB	JCRB0815	Japanese	M	47	-	EMEM	-	+	+	+	+	+	-
PC-7	lung adenocarcinoma	s_8	IBL	-	Japanese	-	-	-	RPMI	-	+	+	+	+	+	Non-adherent
PC-3	lung adenocarcinoma	s_7	JCRB	JCRB0077	Japanese	F	48	-	RPMI	collagen Type I-coated	+	+	+	+	+	-
II-18	lung adenocarcinoma	s_19	RIKEN BRC	RCB2093	Japanese	-	-	-	RPMI	-	+	+	+	+	+	-
RERF-LC-OK	lung adenocarcinoma	s_15	JCRB	JCRB0811	Japanese	-	-	-	RPMI	-	+	+	+	+	+	-
A549	lung adenocarcinoma	s_1	ATCC	CCL-185	Caucasian	M	58	-	DMEM	-	+	+	+	+	+	-
A427	lung adenocarcinoma	s_20	ATCC	HTB-53	Caucasian	M	52	-	RPMI	-	+	+	+	+	+	-
H322	lung adenocarcinoma	s_21	ATCC	CRL-5806	Caucasian	-	-	-	RPMI	-	+	+	+	+	+	-
H1648	lung adenocarcinoma	s_25	ATCC	CRL-5882	Black	M	39	Y	RPMI	collagen Type I-coated	+	+	+	+	+	-
H1650	lung adenocarcinoma	s_26	ATCC	CRL-5883	Caucasian	M	27	Y	RPMI	-	+	+	+	+	+	-

Human Genomes

Data provider	Cancer type	Number of samples	Reference
National Cancer Center Hospital East and University of Tokyo	Lung adenocarcinoma		97PLoS One. 2013 Sep 12;8(9):e73484.
National Cancer Center Hospital East	Small cell lung cancers		57J Thorac Oncol. 2014 Sep;9(9):1324-31.
ICGC	43 of ICGC DCC Project Codes	6,590	https://dcc.icgc.org/
Dr. Meyerson's Lab.	Lung adenocarcinoma	183	Cell. 2012 Sep 14;150(6):1107-20.
Dr. Ogawa's Lab.	Myelodysplasia	29	Nature. 2011 Sep 11;478(7367):64-9
	Clear-cell renal cell carcinoma	106	Nat Genet. 2013 Aug;45(8):860-7
TCGA	Gastric adenocarcinoma	295	Nature. Published online 23 July 2014
	Urothelial bladder carcinoma	131	Nature. 507 (7492):315-22.
	Glioblastoma	291	Cell. 155 (2):462-477.
	Clear cell renal cell carcinoma	446	Nature. 499 (7456):43-49.
	Endometrial carcinoma	373	Nature. 497 (7447):67-73.
	Acute myeloid leukemia	200	NEJM. 368:2059-2074.
	Breast tumors	507	Nature. 490 (7418):61-70.
	Squamous cell lung cancers	178	Nature. 489 (7417):519:525.
	Colon and rectal cancer	224	Nature. 487 (7407):330-337.
	Ovarian carcinoma	316	Nature. 474 (7353):609-615.
	Glioblastoma	91	Nature. 455 (7216):1061-1068.
HGVD	Normal (Japanese)	1,208	URL: http://www.genome.med.kyoto-u.ac.jp/SnpDB
ToMMo	Normal (Japanese)	1,070	URL: http://humandbs.biosciencedbc.jp/en/hum0015-v1
JPDSC (Japan PGx Data Science Consortium)	Normal (Japanese)	2,994	URL: http://humandbs.biosciencedbc.jp/hum0013-v1
Total		15,386	

Exon Archive Project at MGH

ExAC release0.3; Total: 60,706 samples

Population	Male Samples	Female Samples	Total
African/African American (AFR)	1,888	3,315	5,203
Latino (AMR)	2,254	3,535	5,789
East Asian (EAS)	2,016	2,311	4,327
Finnish (FIN)	2,084	1,223	3,307
Non-Finnish European (NFE)	18,740	14,630	33,370
South Asian (SAS)	6,387	1,869	8,256
Other (OTH)	275	179	454
Total	33,644	27,062	60,706

DBに登録されているSNP(SNV)のサイト数は 全人種トータルで10,185,731箇所

-Database of Transcriptional Start Sites-
DBTSS
Release 9.0 (Updated July 9, 2015)
Based on UCSC hg19, mm10

Database Search

Species: H. sapiens

Keyword:

Search

Lift over: hg18

Search

Human Chromatin Features

Search

Search from Genomic Position: chr1:75,787,000

Search

Search from SNP (dbSNP rsID): rs3752989

Search

Search from SNV (COSMIC somatic mutation): BRAF

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma

Search

SNV Summary in Cancers: NM.

Search

About this database

Welcome to DBTSS (Database of Transcriptional Start Sites)

To support transcriptional regulation studies, we have constructed the DBTSS (Database of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-seq.

This database includes TSS data of a major part of human adult and embryonic tissues are covered. DBTSS now contains 491 million TSS seq sequences for collected from a total of 20 tissues and 1 cell cultures. We also integrated our newly generated RNA-seq data of subcellular-fractionated RNAs and ChIP-seq data of histone modifications, RNA polymerase II and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external epigenomic data, such as chromatin map of the ENCODE project.

In this update, we further associated this TSS information with public and original SNV data, in order to identify single nucleotide variations (SNVs) in the regulatory regions.

It is believed that single nucleotide variations (SNVs) in the transcriptional regulatory regions are responsible for many human diseases, including cancers. However, it remains difficult to identify functionally relevant SNVs from those having no respect biological consequences. In this version of DBTSS, we attempt to associate SNVs with the omics information of the surrounding regions. We used SNVs which we identified from genomic analyses of various types of cancers, including somatic mutations of 100 lung adenocarcinoma and lung small cell carcinoma. For germline variations, we used SNVs in dbSNP as well as our unique dataset of variations in 1000 Japanese individuals. We integrated those SNV information with our original database of TSS-seq, RNA-seq, ChIP-seq of representative histone modifications and Bluefire Sequencing of cytosine methylations of DNA. Particular, we present multi-omics data of 26 lung adenocarcinoma cell lines for which TSS-seq, RNA-seq, ChIP-seq and seq together with whole genome sequencing are collected from the same materials. We further connected the multi-omics data of model organisms by genome-genome alignment. We provide a unique data resource to investigate what genomic features are observed in a particular genomic coordinate in a wide variety of samples.

These data can be browsed in our new viewer which also supports versatile search conditions of users. We believe new DBTSS is helpful to understand biological consequences of the massively identified TSSs and identify human genetic variations which are associated with diversified transcriptional regulations.

News

- 09 Jul 2015: New T helper cell data of mouse line now available. Raw data accession: [DBR000284](#), [DBR001102](#) and [SRR0070284](#) (Genome Shen)
- 30 Jun 2015: New BRIC-seq data (LIFT RNA) are now available. Raw data accession: [DBR000285](#) (Genome Shen)
- 15 Sep 2014: New DBTSS opened.

References

Sasaki A, Wakaguni H, Yamashita R, Kawase S, Tsuchihara K, Sugita Y, Naka K. DBTSS as an integrative platform for transcription, epigenome sequence variation data. *Nucleic Acids Res*. 2015; **43**(11):D1-G5.

Sasaki A, Mimaki S, Yamane Y, Kawase A, Mabushima K, Suzuki M, Sugita S, Fuzum H, Sasaki Y, Tsuchihara K. Identification and clinical correlations in Japanese lung adenocarcinoma without separate tissue counterparts. *PLoS One*. 2011; **6**:178,610.

Yamashita R, Sathira NP, Kiama A, Tanimoto K, Arauchi T, Tanaka Y, Sugano S, Naka K, Sasaki Y. (2011) Genome-wide characterization of transcriptional start sites in humans by integrative transcription analysis. *Genome Res*. 2011; **21**:189-203.

Tsuchihara K, Sasaki Y, Wakaguni H, the T. Tanimoto K, Hishimoto M, Mitsuhashi K, Akizawa S, Sugano J, Yamashita R, Naka K, Benito M, Sugano S. (2009) Massive transcriptional start site analysis in *in vitro* cells. *Nucleic Acids Res*. 2009; **37**:22.

[Japanese]
"Database Manual" (Yoshida)

Contact us

We welcome your comments and feedback about our database. Please feel free to contact us: yuzuzaki@ims.uioyok.ac.jp

Copyright © 2015

-Database of Transcriptional Start Sites-
DBTSS
Release 9.0 (Updated July 9, 2015)
Based on UCSC hg19, mm10

Database Search

Show 50 entries

Term	Genomic position	Go to	Go to	Search	Go to
chr12:007,709,815-952	chr12:007,709,815-952	TSS viewer	Genome viewer	Human Variation DB	
chr18:361,245-18,401,213	chr18:361,245-18,401,213	TSS viewer	Genome viewer	Human Variation DB	
chr1:75,724,347-76,875,678	chr1:75,724,347-76,875,678	TSS viewer	Genome viewer	Human Variation DB	
chr12:120,725,768-120,740,008	chr12:120,725,768-120,740,008	TSS viewer	Genome viewer	Human Variation DB	
chr17:219,831-225,207	chr17:219,831-225,207	TSS viewer	Genome viewer	Human Variation DB	
chr11:108,121-108,148,164	chr11:108,121-108,148,164	TSS viewer	Genome viewer	Human Variation DB	
chr12:91,907,418-91,923,981	chr12:91,907,418-91,923,981	TSS viewer	Genome viewer	Human Variation DB	
chr4:73,197,154-73,223,891	chr4:73,197,154-73,223,891	TSS viewer	Genome viewer	Human Variation DB	
chr12:91,907,418-91,923,981	chr12:91,907,418-91,923,981	TSS viewer	Genome viewer	Human Variation DB	
chr2:152,920,000-152,930,000	chr2:152,920,000-152,930,000	TSS viewer	Genome viewer	Human Variation DB	
chr1:75,787,000	chr1:75,787,000	TSS viewer	Genome viewer	Human Variation DB	

Human Chromatin Features

Search

Search from Genomic Position: chr1:75,787,000

Search

Search from SNP (dbSNP rsID): rs3752989

Search

Search from SNV (COSMIC somatic mutation): BRAF

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma

Search

SNV Summary in Cancers: NM.

Search

Human Variation DB

Chromosome 12 | Region: 52,301,202 - 52,317,145

Show 15 SNPs

Genomic position	rs ID	NI change info	Amino acid change	Gene ID	Gene name	Hetero/Homo	Disease	V-ID	Case/Ctrl with mutation	P-value	OR(95%CI)	Type of studies	
52,301,383	rs187098243	C	G	100	5	173	0.028	0	0	0	0	Japanese_HKPU	HKPU
52,301,818	rs3782479	C	A	100	26	152	0.148	0	0	0	0	Japanese_HKPU	HKPU
52,302,217	rs7989276	C	T	100	26	152	0.148	0	0	0	0	Japanese_HKPU	HKPU
52,302,304	rs5592912	A	T	100	2	176	0.011	0	0	0	0	Japanese_HKPU	HKPU
52,302,338	rs7989410	G	A	100	38	140	0.219	0	0	0	0	Japanese_HKPU	HKPU
52,302,443	rs7995340	T	G	100	26	152	0.148	0	0	0	0	Japanese_HKPU	HKPU
52,302,742	rs11699951	C	T	100	27	151	0.135	0	0	0	0	Japanese_HKPU	HKPU
52,302,969	rs76782411	T	C	100	26	152	0.148	0	0	0	0	Japanese_HKPU	HKPU
52,304,599	rs11699953	C	T	100	110	88	0.619	0	0	0	0	Japanese_HKPU	HKPU
52,304,791	rs143792324	T	G	100	4	174	0.023	0	0	0	0	Japanese_HKPU	HKPU
52,305,069	rs3782480	G	T	100	48	132	0.258	0	0	0	0	Japanese_HKPU	HKPU
52,305,700	rs12578436	G	A	100	47	131	0.254	0	0	0	0	Japanese_HKPU	HKPU

Search

TSS viewer

Genome viewer

Human Variation DB
(hg19 へ座標変換後リンク)

-Database of Transcriptional Start Sites-
DBTSS
Release 9.0 (Updated July 9, 2015)
Based on UCSC hg19, mm10

Database Search

Keyword:

Search

Lift over: hg18

Search

Human Chromatin Features

Search

Search from Genomic Position: chr1:75,787,000

Search

Search from SNP (dbSNP rsID): rs3752989

Search

Search from SNV (COSMIC somatic mutation): BRAF

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma

Search

SNV Summary in Cancers: NM.

Search

Genome viewer (Multi-omics Viewer)

Gene:

Genomic position:

Search

TSS viewer

Gene:

Genomic position:

Search

Genome viewer (Multi-omics Data)

Gene:

Genomic position:

Search

Human Chromatin Features

Search

Search from Genomic Position: chr1:75,787,000

Search

Search from SNP (dbSNP rsID): rs3752989

Search

Search from SNV (COSMIC somatic mutation): BRAF

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma

Search

SNV Summary in Cancers: NM.

Search

Details of the Promoter Region

Genomic position: chr1:75,787,000-150,000

Search

TSS Seq Data

Gene:

Genomic position:

Search

Human Chromatin Features

Search

Search from Genomic Position: chr1:75,787,000

Search

Search from SNP (dbSNP rsID): rs3752989

Search

Search from SNV (COSMIC somatic mutation): BRAF

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma

Search

SNV Summary in Cancers: NM.

Search

平成27年度の計画(経過)

- ・ゲノム支援データ(マウス)の収載
- ・RDF化の推進(TSSを端緒に)
- ・臨床ゲノムデータとのリンク方法の開発

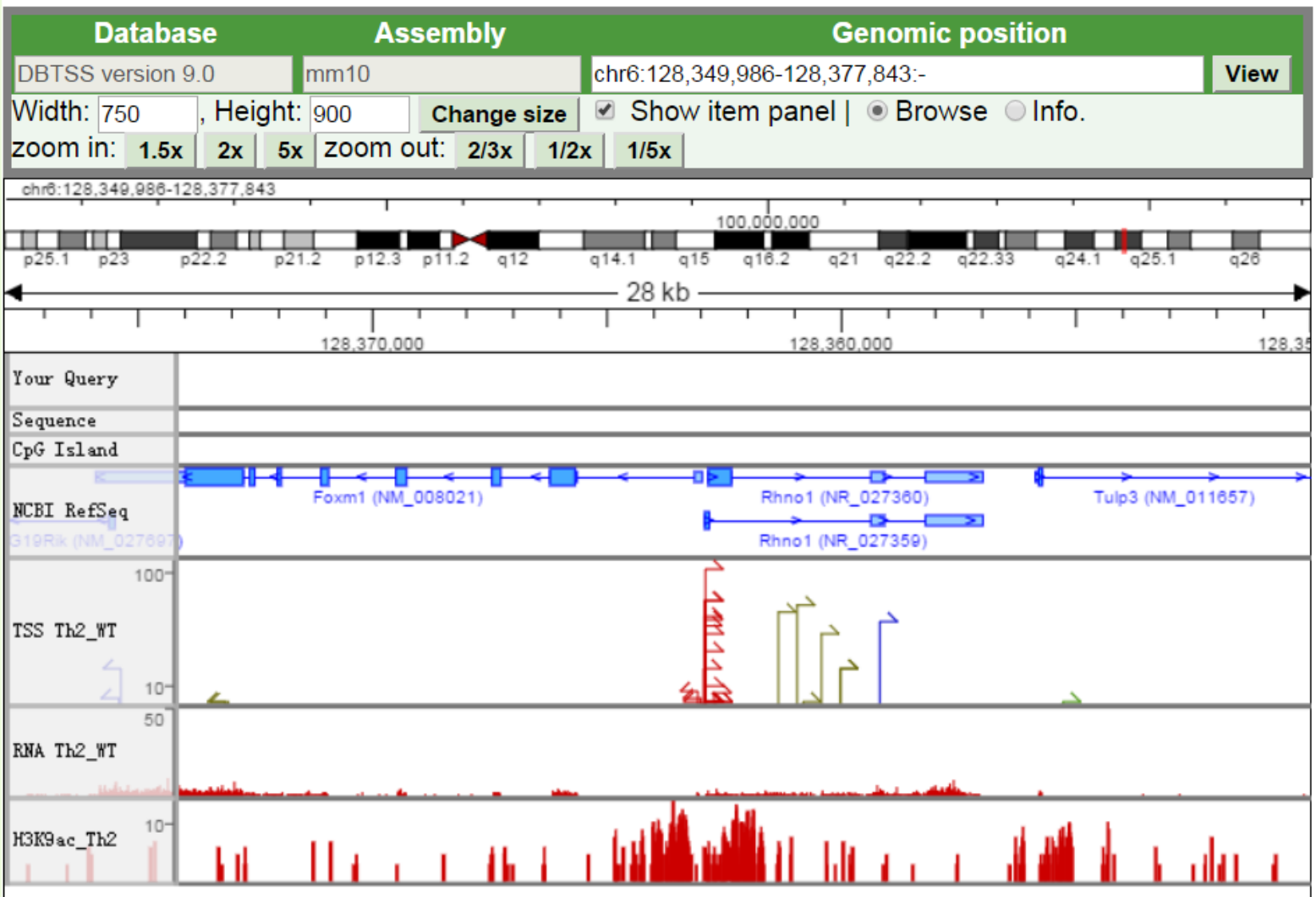
Other species

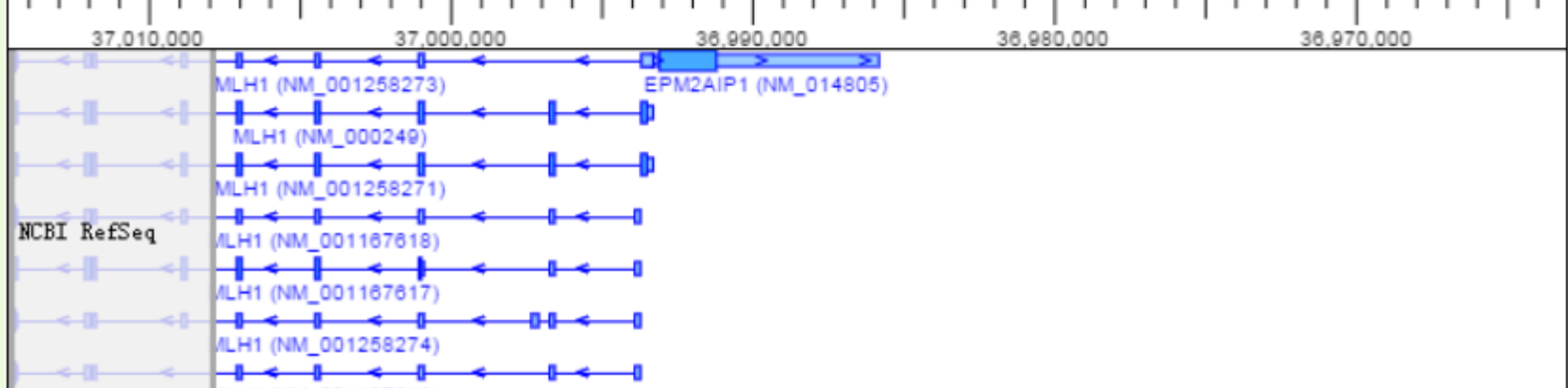
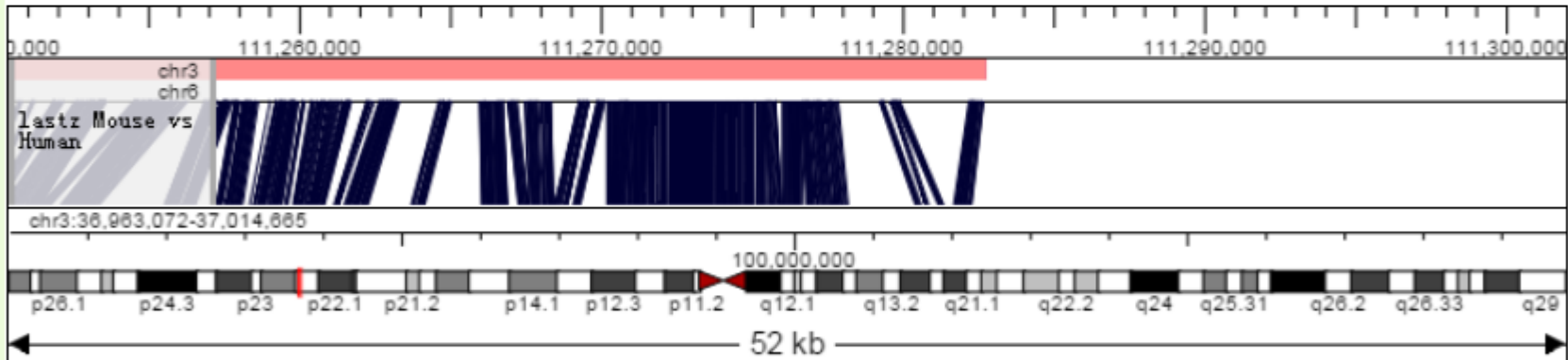
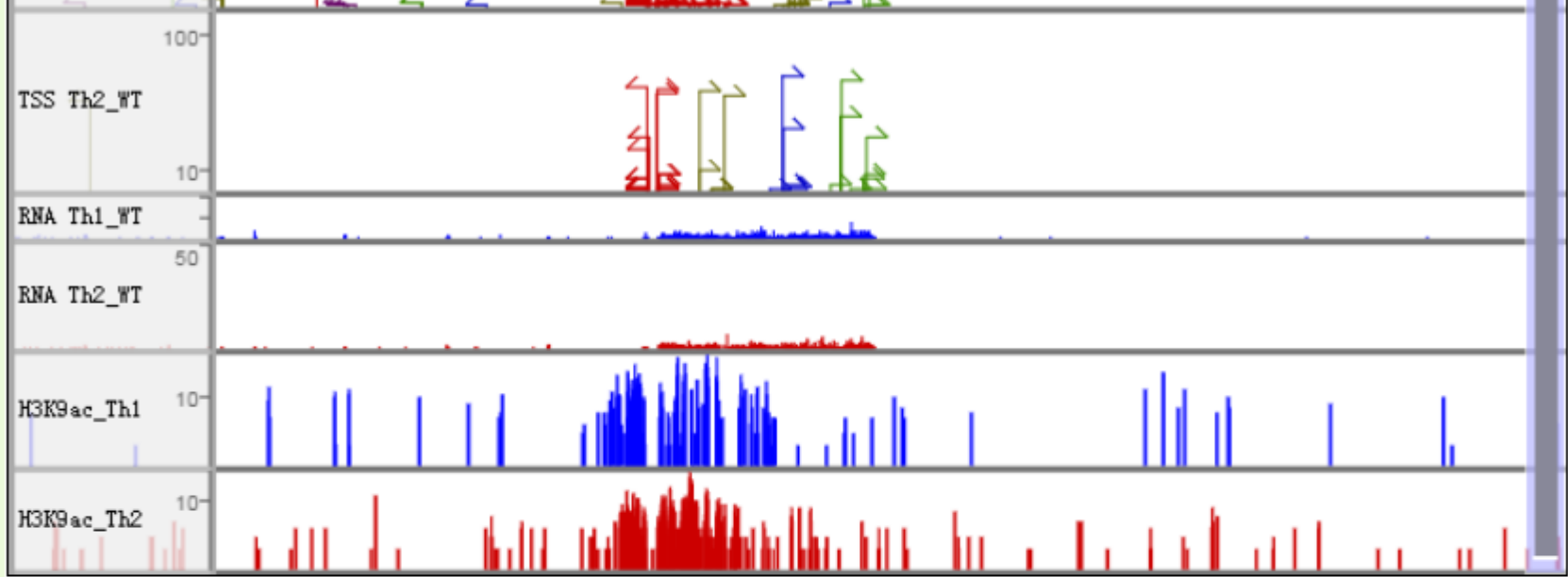
	Human	Mouse	Malaria	Chyzon	Rat	Chimp	Macaque
TSS-seq	73	7	1	1	1	1	2
RNA-seq	42	0	0	0	0	0	0
ChIP-seq	255	0	0	0	0	0	0
RIP-seq	12	0	0	0	0	0	0
BS-seq	26	0	0	0	0	0	0
ChromHM M	36	0	0	0	0	0	0
SNV	49	0	0	0	0	0	0

Integrating mouse data (ゲノム支援)

名前	所属	課題	Acc	データセット	配列数
田中 知明	千葉大学	転写因子p53による新たな代謝調節機能と代謝環境応答のエピジェネティクス制御	DRA000928 SRP007894 DRA001102	26	515,880,048
秋光 信佳	東京大学アイソトープ総合センター	RNA分解を介した核局在型高分子非コードRNA作用マシナリーの制御機構の解明	DRA000591	25	943,054,680

Genome viewer (Multi-omics Data)





Integrating mouse data (ゲノム支援; 作業中)

名前	所属	課題	Acc	データセット	配列数
三野 享史	京都大学ウイルス研究 所感染防御研究分野	Regnase-1の作用機構から探る自然免疫応答における転写後調節の解明	DRA003234	8	161,938,914
岩間 厚志	千葉大学大学院医学 研究院先端応用医学 講座細胞分子医学	エピジェネティック異常による造血器腫瘍の発症機構の解明	DRA000485 DRA000486 DRA000488 DRA000858	20	600,000,000
河野 友宏	東京農業大学生物応 用科学部	次世代シーケンサーを用いた生殖系列のエピゲノム修飾とトランスクリプトーム解析	DRA000484 DRA000607	133	8,900,000,000
深田 吉孝	東京大学大学院理学 系研究科	生存戦略としての体内時計システムの分子解剖	DRP001092 DRP001093 DRP001094 DRP001349	54	1,350,000,000
小野寺 淳	千葉大学・大学院医 学研究院・免疫発生 学	ポリコム/トライソラックス複合体による免疫記憶維持機構の解明	GSE51079	6	180,000,000
西中村 隆一	熊本大学	胎児型腎臓幹細胞の成体腎での再活性化	DRA000957	6	150,000,000

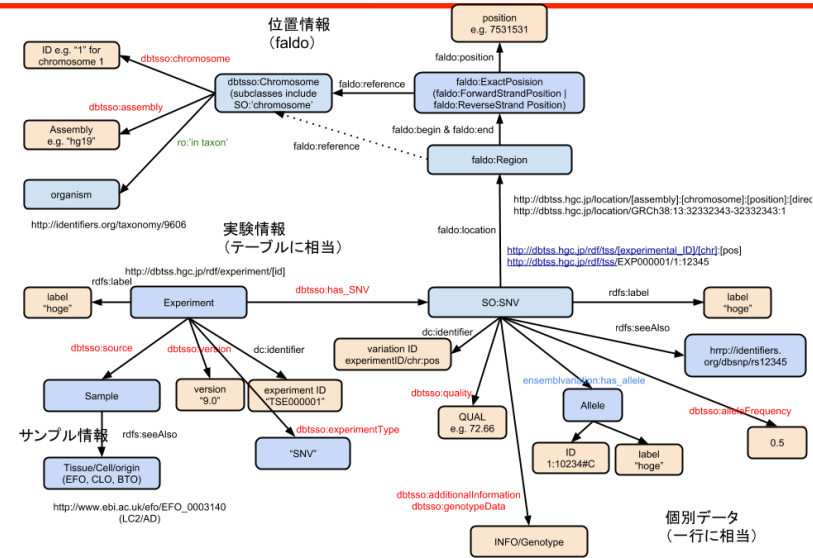
RDF化

Variationデータ用オントロジー

- variation: ENSEMBL Ontology
- location: FALDO
- 不足分は独自オントロジー定義

DBTSSオントロジー

RDF化スクリプト



DBTSS Variation RDF

```
<http://dbtss.hgc.jp/rdf/experiment/TSE000086> <http://dbtss.hgc.jp/rdf/ontology/has_SNV> <http://dbtss.hgc.jp/rdf/data/TSE000086/1:10234> a <http://purl.obolibrary.org/obo/SO_0001483>;
<http://www.w3.org/2000/01/rdf-schema#label> "variation on chr1:10234 from TSE000086";
<http://biohackathon.org/resource/faldo#location> <http://dbtss.hgc.jp/rdf/location/chromosome:GRCh38:1:10234-10234:1>;
<http://dbtss.hgc.jp/rdf/ontology/additionalInformation> "AC=1;AF=0.50;AN=2;BaseQRankSum=-0.777;DB=53;Dels=0.02;FS=1.862";
<http://dbtss.hgc.jp/rdf/ontology/alleleFrequency> 0.5;
<http://dbtss.hgc.jp/rdf/ontology/allelicDepths> "42,10";
<http://dbtss.hgc.jp/rdf/ontology/dbsnpID> "rs145599635";
<http://dbtss.hgc.jp/rdf/ontology/genotype> "0/1";
<http://dbtss.hgc.jp/rdf/ontology/genotypeData> "GT:AD:DP:GQ:PL|0/1:42,10:52:61.94:103,0,62";
<http://dbtss.hgc.jp/rdf/ontology/genotypeQuality> 61.94;
<http://dbtss.hgc.jp/rdf/ontology/quality> 72.66;
<http://purl.org/dc/terms/identifier> "1:10234";
<http://www.w3.org/2000/01/rdf-schema#seeAlso> <http://info.identifiers.org/dbsnp/rs145599635>,
<http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=rs145599635>;
<https://github.com/simonjupp/ensembl-rdf/blob/master/ontology/ensembl_variation_ontology.owl#has_allele> <http://dbtss.hgc.jp/rdf/data/TSE000086/1:10234#T> .

<http://dbtss.hgc.jp/rdf/data/TSE000086/1:10234#C> a <https://github.com/simonjupp/ensembl-rdf/blob/master/ontology/ensembl_variation_ontology.owl#ancestral_allele>;
<http://www.w3.org/2000/01/rdf-schema#label> "TSE000086 chr1:10234 allele C";
<http://purl.org/dc/terms/identifier> "1:10234#C" .
```

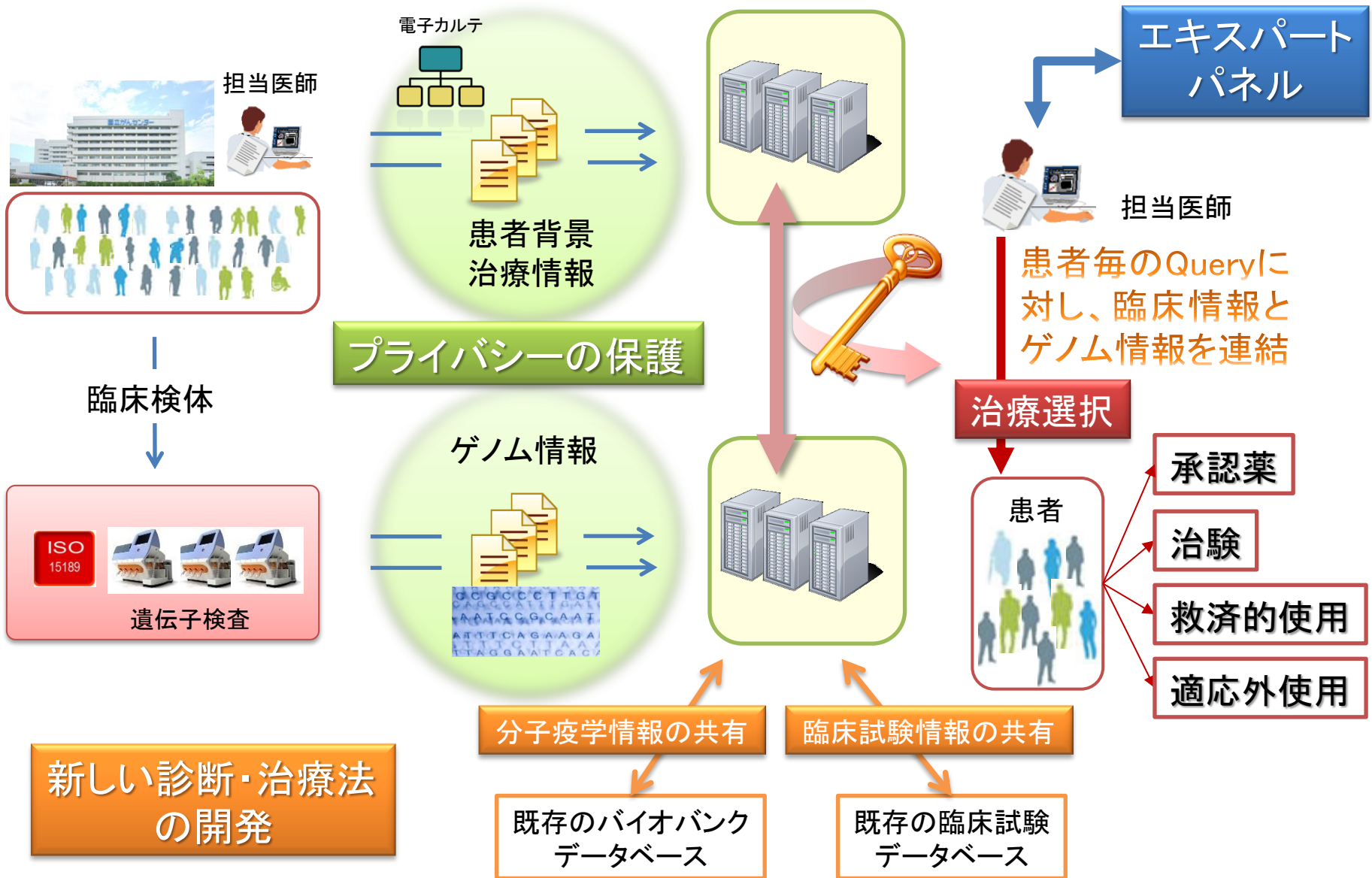
```
#####
#
# Data properties
#
#####

### http://dbtss.hgc.jp/rdf/ontology/dbtss.owl#additionalInformation
:additionalInformation rdf:type owl:DatatypeProperty ;
    rdfs:label "Additional Information" ;
    rdfs:comment "Values from INFO tag" .

### http://dbtss.hgc.jp/rdf/ontology/dbtss.owl#alleleFrequency
:alleleFrequency rdf:type owl:DatatypeProperty ;
    rdfs:label "Allele Frequency" ;
    rdfs:comment "Allele Frequency, for each ALT allele, in the same order as listed" ;
    rdfs:range xsd:float .

### http://dbtss.hgc.jp/rdf/ontology/dbtss.owl#allelicDepths
:allelicDepths rdf:type owl:DatatypeProperty ;
    rdfs:label "Allelic Depths" ;
```

がん“最適化創薬”の実現に向けた 臨床情報・ゲノム情報管理システムの構築



ヒト疾患ゲノム統合DB (DBTSSの拡張):

KERO(Kashiwa Encyclopedia of Regulatory Omics)



ヒトゲノム・エピゲノム・トランスクリプトームデータの統合

ヒト疾患ゲノム変異への機能的注釈

パターン検索システムの開発と実装

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



オミクスデータ統合が加速するヒトゲノム臨床応用研究

＝疾患ゲノムのその座標で“何が起きているのか”を網羅的に検索