

がん研究に関連した バイオインフォマティクス入門

研究所 生物情報学分野

研究所 バイオインフォマティクス部門

がんゲノム情報管理センター (C-CAT) ゲノムデータ管理室

西野 穰

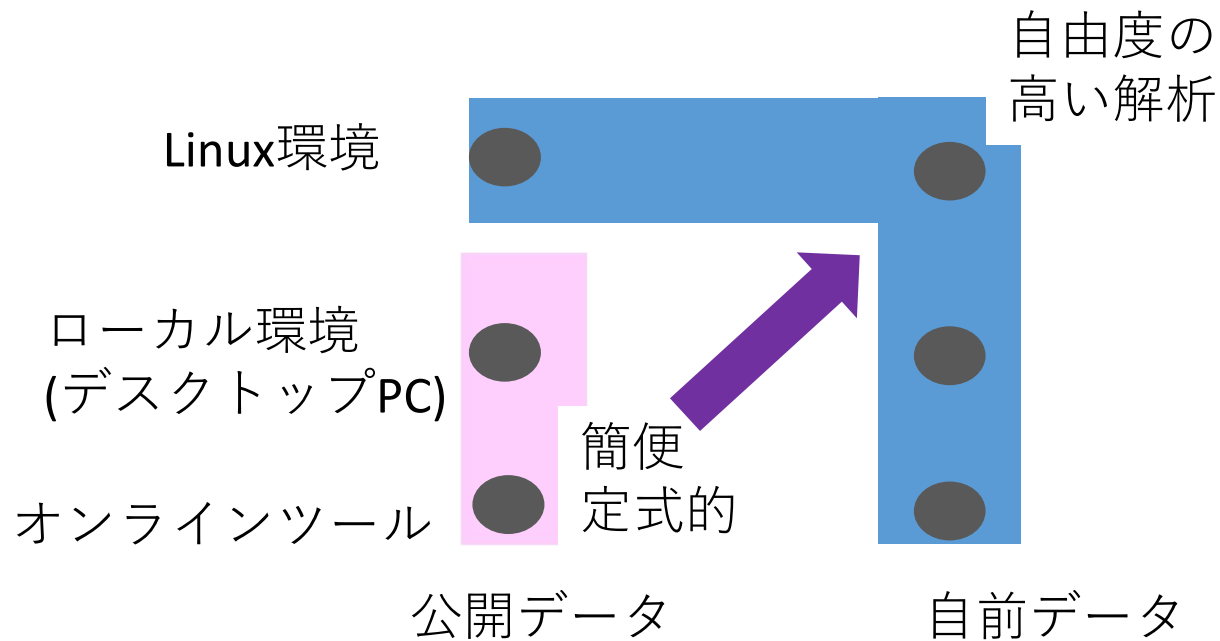
生物統計セミナー【発展編】

2023.12.18 (月)

本セミナーについて

- バイオインフォマティクス(BI)に馴染みのない方を主な対象として、BIのがん研究に関連した概要や入門的な内容を説明します
- Linuxコマンドを使わない解析
- ツールやデータベース毎に、
- 出来るだけ「解析の目的」を示しそれに対する「手順」を提示する形で進めます

解析環境の分類



解析環境	スキルレベル	対象データ	解析自由度・出力自由度
オンラインツール	スキルに依存せず、利用者に優しい	主に内蔵データ（公開データ）を使用し、特定のタスクに特化	使える解析は限られている。出力は定型的(きれいなことも多い)
ローカル環境	中程度	中小規模のデータ向き。自前データも扱える	中程度の自由度。BIツールのごく一部が使える
Linux環境	敷居が高い	大規模データも扱える。自前データも扱える	高い自由度。BIツールの大部分が使える

BI解析の目的、解析環境、ツールの例

□ がんに関係のあることが知られている遺伝子に体細胞変異を検出した。病原性(pathogenicity)を評価したい

⇒ 病原性のある体細胞変異は、生殖細胞系列の変異として集団中で低頻度に保たれているという仮定の下、gnomADを用いてこの変異の集団頻度を確認する。また、これまでがん組織の体細胞変異として報告のある変異か(COSMIC)、変異の有無で予後の違いがあるか(TCGA)知る

□ 研究仮説を思いついたが、データを取る前に、公開データで仮説の一部を検証できないか

⇒ 今回の解析は、オンラインの解析システムcBioPortalで解析できそうだ


□ ゲノムデータで予後予測モデルを作成した。結果をバリデートしたい

⇒ RパッケージTCGAbiolinksを用いてTCGAデータをローカルにダウンロードして解析

対象とするデータベースやツール

- gnomAD
 - 正常な個体の生殖細胞系列の変異をまとめたデータベース
疾患関連の遺伝子変異フィルタリングに使用
- COSMIC
 - がん組織の体細胞変異に関する詳細な情報を提供するがんゲノム変異カタログ
- OncoKB
 - がん関連の体細胞性変異に対する治療への影響や臨床的な重要性を提供するがんの遺伝子変異知識ベース
- TCGA
 - がんゲノムや遺伝子発現などの包括的な情報を提供するプロジェクト。GDC Data Portalを通じてTCGAデータの検索、フィルタ、ダウンロード、簡単な分析ができる
- cBioPortal
 - TCGAなどの多数のがんゲノム研究のデータをキュレート。がんゲノムデータの視覚化と解析をサポートするオープンソースのプラットフォーム
- TCGAbiolinks
 - TCGAデータなどへのアクセスを容易にし、データのダウンロードや解析、視覚化をサポートするRパッケージ

gnomAD



Genome Aggregation Database

gnomAD v4.0.0 ▾ Search by gene, region, or variant

Or

- [Download gnomAD data](#)
- [Read gnomAD publications](#)
- [Find co-occurrence of two variants](#)

Please note that the gnomAD v3 genomes are now part of gnomAD v4. For more information, see ["Should I switch to the latest version of gnomAD?"](#)

Examples

- Gene: [PCSK9](#)
- Transcript: [ENST00000302118](#)
- Variant: [1-55051215-G-GA](#)
- Structural variant region: [19-11078371-11144910](#)
- Copy number variant region: [19-11078371-11144910](#)
- Mitochondrial variant: [M-8602-T-C](#)

- gnomADは、健康な個人の生殖細胞系列変異をまとめたデータベースであり、各遺伝子座や変異に関する広範な情報を提供する
- 異なる人種や集団の遺伝子情報が含まれ、次世代シーケンシング技術を駆使して多様なデータが収集されている
- 主たる目的は、正常な個体の遺伝子座や変異のバリエーションを理解し、それに基づいて疾患との関連性を評価することである。疾患関連の変異を特定するための重要な基準として研究者や臨床医が活用している
- gnomADのデータは、遺伝学やゲノム解析に従事する研究者、臨床医、遺伝カウンセラーにとって、遺伝的な変異の頻度や分布に関する貴重な情報源となっている

BRAF V600Eの頻度

gnomAD
Genome Aggregation Database

gnomAD v4.0.0 | BRAF

- Download gnomAD data
- Read gnomAD publications
- Find co-occurrence of two variants

①検索窓にBRAfと入れ検索

Please note that the gnomAD v3 genomes are now part of gnomAD v4. For more information, see "Should I switch to the latest version of gnomAD?"

csv形式でBRAf全体の変異をダウンロードすることもできる

②Val600Gluを検索

Export variants to CSV | Configure table

Only variants located in or within 75 base pairs of a coding exon are shown here. To see variants in UTRs or introns, use the [region view](#).

The table below shows the HGVS consequence and VEP annotation for each variant's most severe consequence across all transcripts in this gene. Cases where the most severe consequence occurs in a non-MANE Select transcript (or non-canonical transcript if no MANE Select transcript exists) are denoted with †. To see consequences in a specific transcript, use the [transcript view](#).

Variant ID	Source	HGVS Consequence	VEP Annotation	LoF Curation	Clinical Significance	Flags	Allele Count	Allele Number	Allele Frequency	Number of Homozygotes
7-140749119-T-G	gnomAD	c.2112+168A>C	intron				2	594864	3.36e-6	0
7-140749120-A-T	gnomAD	c.2112+167T>A	intron				1	145584	6.87e-6	0
7-140749122-C-T	gnomAD	c.2112+165G>A	intron				35	766296	4.57e-5	0
7-140749123-G-A	gnomAD	c.2112+164C>T	intron				20	774278	2.58e-5	0
7-140749123-G-C	gnomAD	c.2112+164C>G	intron				2	622154	3.21e-6	0
7-140749123-G-T	gnomAD	c.2112+164C>A	intron				2	622154	3.21e-6	0
7-140749126-T-C	gnomAD	c.2112+161A>G	intron				1	663124	1.51e-6	0
7-140749213-T-C	gnomAD	c.2112+74A>G	intron				3	1590714	1.89e-6	0
7-140749214-A-G	gnomAD	c.2112+73T>C	intron				1	1438600	6.95e-7	0
7-140749218-C-T	gnomAD	c.2112+69G>A	intron				1	620664	1.61e-6	0
7-140749220-T-TA	gnomAD	c.2112+66dup	intron				1	826008	1.21e-6	0

③v4.0.0だとヒットしない

④gnomAD v2.1.1で検索してみる

SNV: 7-140453136-A-T(GRCh37) | Copy variant ID | Gene page | Dataset: gnomAD v2.1.1

Filters: Allele Count, Allele Number, Allele Frequency, Grpmax Filtering AF (95% confidence), Number of homozygotes, Mean depth of coverage

Exomes: 1 | Genomes: No variant | Total: 1

External Resources: dbSNP (rs113488022), UCSC, ClinVar (13961), ClinGen Allele Registry (CA123643)

Feedback: Report an issue with this variant

Genetic Ancestry Group Frequencies

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
South Asian	1	30612	0	0.00003267
African/African American	0	16252	0	0.00000000
Admixed American	0	34528	0	0.00000000
Ashkenazi Jewish	0	10076	0	0.00000000
East Asian	0	18392	0	0.00000000
European (Finnish)	0	21638	0	0.00000000
European (non-Finnish)	0	113638	0	0.00000000
Remaining Individuals	0	6124	0	0.00000000
XX	0	115468	0	0.00000000
XY	1	135792	0	0.000007364

Export variants to CSV | Configure table

Only variants located in or within 75 base pairs of a coding exon are shown here. To see variants in UTRs or introns, use the [region view](#).

The table below shows the HGVS consequence and VEP annotation for each variant's most severe consequence across all transcripts in this gene. Cases where the most severe consequence occurs in a non-canonical transcript (or non-canonical transcript if no MANE Select transcript exists) are denoted with †. To see consequences in a specific transcript, use the [transcript view](#).

Variant ID	Source	HGVS Consequence	VEP Annotation	LoF Curation	Clinical Significance	Flags	Allele Count	Allele Number	Allele Frequency	Number of Homozygotes
7-140453087-G-T	gnomAD	p.Ser616Ser	synonymous		Likely benign		5	282642	1.77e-5	0
7-140453098-A-G	gnomAD	p.Leu613Leu	synonymous		Likely benign		2	251288	7.96e-6	0
7-140453101-G-T	gnomAD	p.Gln612Lys	missense				1	251260	3.98e-6	0
7-140453102-T-C	gnomAD	p.Glu611Glu	synonymous				1	251260	3.98e-6	0
7-140453108-C-A	gnomAD	p.Gln609His	missense				1	251260	3.98e-6	0
7-140453115-G-C	gnomAD	p.Ser607Cys	missense				1	251260	3.98e-6	0
7-140453118-C-G	gnomAD	p.Gly606Ala	missense				1	251260	3.98e-6	0
7-140453118-C-T	gnomAD	p.Gly606Glu	missense				1	251260	3.98e-6	0
7-140453134-T-C	gnomAD	p.Lys601Glu	missense				1	251260	3.98e-6	0
7-140453136-A-T	gnomAD	p.Val600Glu	missense				2	251194	7.96e-6	0
7-140453138-T-C	gnomAD	p.Thr599Thr	synonymous				1	251260	3.98e-6	0
7-140453146-G-A	gnomAD	p.Leu597Leu	synonymous				1	251260	3.98e-6	0
7-140453170-T-C	gnomAD	p.Thr589Ala	missense				2	251194	7.96e-6	0

⑤v2.1.1だとヒットするIDをクリック

※ v2.1.1とv4.0.0は独立サンプル

gnomADで遺伝子型頻度も知ることができる

rs334を検索

SNV: 11-5227002-T-A(GRCh38) Copy variant ID Gene page Dataset: gnomAD v4.0.0

Filters
Allele Count: 2335
Allele Number: 1458356
Allele Frequency: 0.001601
Grpmax Filtering AF (95% confidence): 0.05474
Number of homozygotes: 31
Fraction of individuals with >20x coverage: 0.9

Exomes
Pass: 2335

Genomes
Pass: 1937

Total
4272

External Resources
• dbSNP (rs334)
• UCSC
• ClinVar (15333)
• All of Us

Feedback
[Report an issue with this variant](#)

Genetic Ancestry Group Frequencies

Note: Local ancestry inference for this variant is not available.

gnomAD | HGDP | 1KG | Local Ancestry

変異型アレルの数 (Allele Count) 全アレル数 (Allele Number) 変異型ホモ接合体の数 (Number of Homozygotes)

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African/African American	3707	74908	36	0.04949
Middle Eastern	23	6052	1	0.003800
Remaining	216	62392	2	0.003462
Admixed American	199	60018	1	0.003316
South Asian	81	91040	0	0.0008897
European (non-Finnish)	46	1176872	0	0.00003909
European (Finnish)	0	64018	0	0.000
Ashkenazi Jewish	0	29584	0	0.000
East Asian	0	44854	0	0.000
Amish	0	912	0	0.000

ヘモグロビンS (HbS) は、HBB遺伝子のミスセンス変異 (rs334; HBB c.20A>T; p.Glu6-Val) による異常なヘモグロビン。ホモ接合の状態では重度の鎌状赤血球貧血を引き起こす一方マラリア耐性がある

※ HbSを例に挙げたが、変異型ホモ接合ががん発症に関わる場合も多いと考えられる

全個体数 = Allele Number / 2 = 37,454

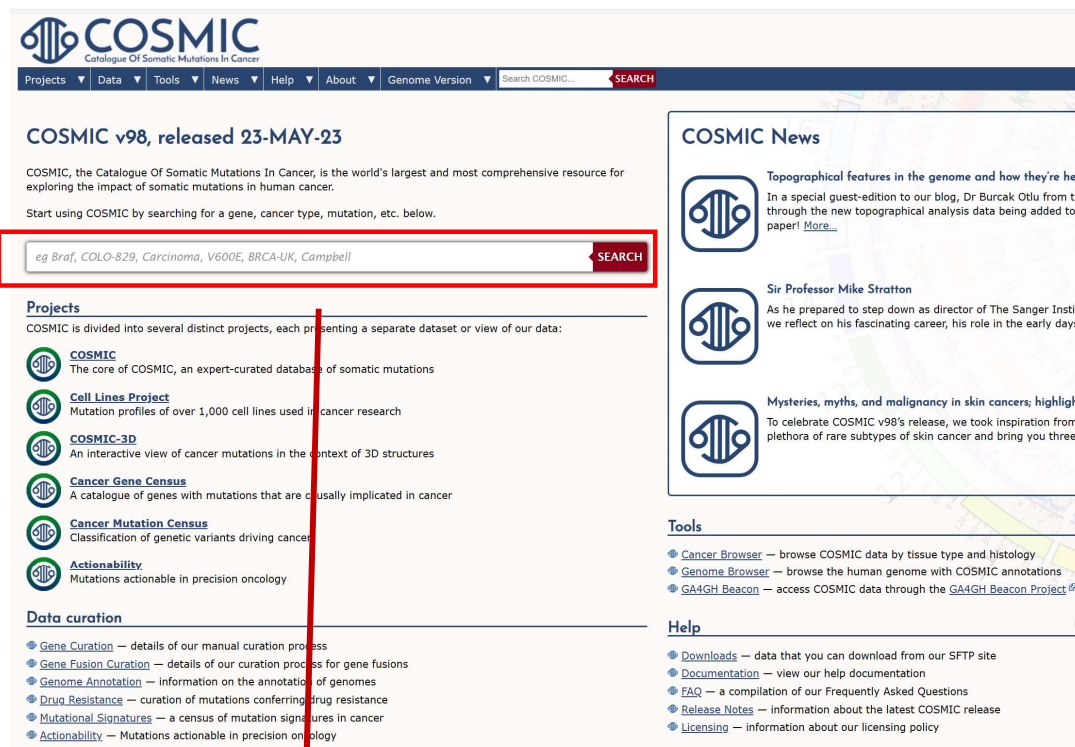
ヘテロ数 = Allele Count - 2 x Number of Homozygotes = 3,635

正常型ホモ接合体の数 = 全個体数 - ヘテロ数 - Number of Homozygotes = 33,783

遺伝子型	TT	TA	AA
数	33,783	3,635	36

ハーディー・ワインベルグ平衡下での期待値92人よりかなり小さい

COSMIC



COSMIC
Catalogue Of Somatic Mutations In Cancer

Projects Data Tools News Help About Genome Version Search COSMIC

COSMIC v98, released 23-MAY-23

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 SEARCH

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
- Cancer Mutation Census**
Classification of genetic variants driving cancer
- Actionability**
Mutations actionable in precision oncology

Data curation

- Gene Curation** — details of our manual curation process
- Gene Fusion Curation** — details of our curation process for gene fusions
- Genome Annotation** — information on the annotation of genomes
- Drug Resistance** — curation of mutations conferring drug resistance
- Mutational Signatures** — a census of mutation signatures in cancer
- Actionability** — Mutations actionable in precision oncology

COSMIC News

- Topographical features in the genome and how they're help**
In a special guest-edition to our blog, Dr Burcak Otlu from the through the new topographical analysis data being added to C paper! [More](#)
- Sir Professor Mike Stratton**
As he prepared to step down as director of The Sanger Institut we reflect on his fascinating career, his role in the early days c
- Mysteries, myths, and malignancy in skin cancers; highlight**
To celebrate COSMIC v98's release, we took inspiration from o plethora of rare subtypes of skin cancer and bring you three ti

Tools

- Cancer Browser** — browse COSMIC data by tissue type and histology
- Genome Browser** — browse the human genome with COSMIC annotations
- GA4GH Beacon** — access COSMIC data through the [GA4GH Beacon Project](#)

Help

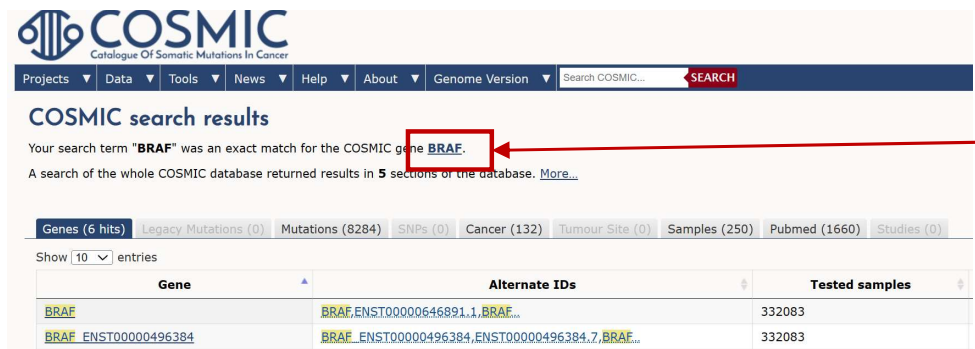
- Downloads** — data that you can download from our SFTP site
- Documentation** — view our help documentation
- FAQ** — a compilation of our Frequently Asked Questions
- Release Notes** — information about the latest COSMIC release
- Licensing** — information about our licensing policy

検索窓に

- 遺伝子名
 - 細胞株名(COLO-829)
 - 変異
 - 研究プロジェクト
- など入力して検索できる

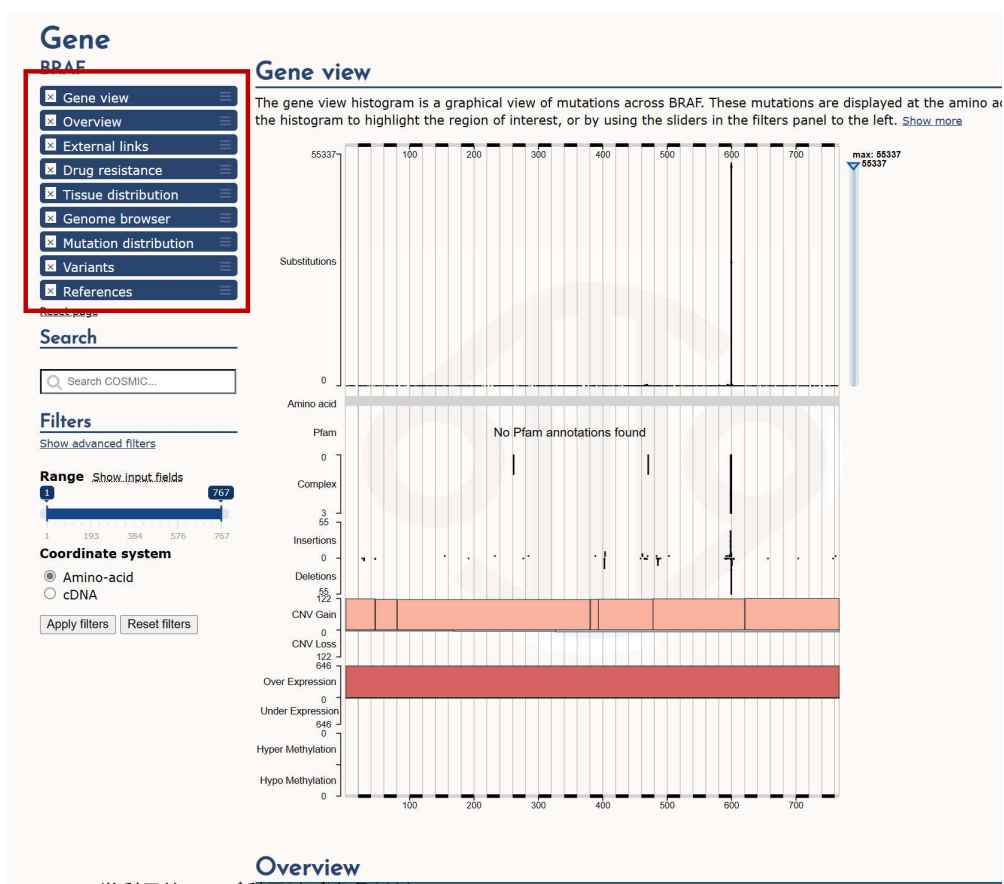
- COSMICは、がんにおける体細胞変異の影響を包括的に提供するデータベース
- COSMIC v86（2018年8月）には、26,000以上の文献から収集された1.4百万以上の腫瘍サンプルにわたり、約6百万のコーディング変異が含まれている
- 手作業でのキュレーションにより品質と精度が確保され、非コーディング変異や遺伝子融合など、がんを促進する多様なメカニズムも網羅
- COSMIC-3Dにおいては、タンパク質構造内の変異や薬物耐性の意義を探索する新機能が提供されている
- Cancer Gene Census（CGC）では、719のがんのドライバー遺伝子とその機能的説明をまとめ、がん研究に不可欠なツールとなっている

①前項の検索窓で”BRAF”を検索



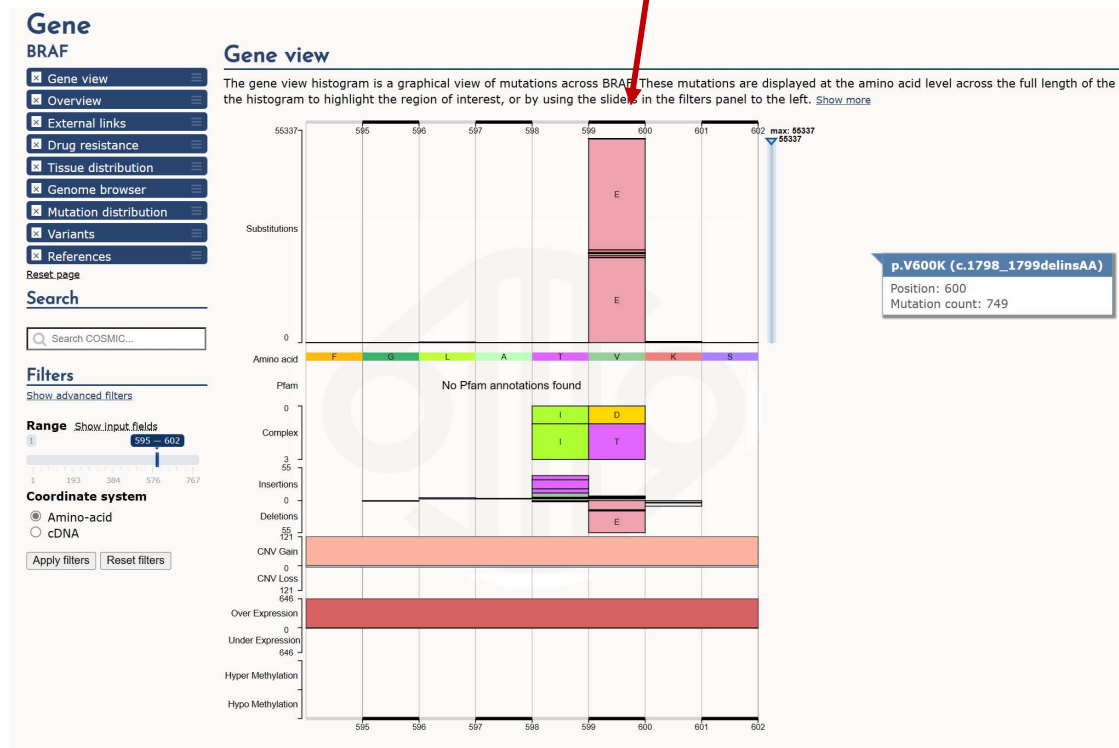
②6つヒットするが
上部の
“exact match for the COSMIC gene BRAF”
をクリック

③BRAFの詳細ページに移る



項目	説明
Gene View	各タイプの変異パターンやその頻度を示す
Overview	ゲノム位置、サンプル数、転写産物等の基本情報 Census gene、Curated gene、Hallmark geneであるか否か 3D構造と詳細へのリンク、パートナー遺伝子、遺伝子変異が感受性に影響する薬剤
External Links	他のバイオインフォマティクスの情報へのリンク
Drug Resistance	耐性変異に関する情報
Tissue Distribution	遺伝子が異なる組織でどの程度発現しているか示す
Genome Browser	ゲノム上の位置情報を視覚的に表示
Mutation Distribution	突然変異の発生頻度や分布に関する情報。
Variants	変異型に関する情報
References	利用された文献や研究の引用

①興味ある領域をドラッグし拡大



②BRAFはCensus gene, Curated gene, Hallmark geneに該当する

Overview

This section gives an overview of BRAF, along with links to any related data and resources.

Census gene **Curated gene** **Mouse gene** **Hallmark gene**

Genomic coordinates

Synonyms

COSMIC-3D

Number of samples

332083 unique samples

58472 unique samples with mutations

Alternative transcripts

BRAF_ENST00000644969, BRAF_ENST00000288602, BRAF_ENST00000496384, BRAF_ENST00000469930

Sequences

You can see various sequences for this gene:

cDNA (ENST00000646891.1)

Protein (BRAF)

[Transcript and protein aligned](#) (ENST00000646891.1+BRAF)

Gene fusions

BRAF is involved in 17 fusions, with the following genes:

KIAA1549_ENST00000440172 (613 mutations in 1395 samples)

FAM131B (6 mutations in 107 samples)

SND1 (6 mutations in 11 samples)

AKAP9 (4 mutations in 292 samples)

RNF130 (2 mutations in 96 samples)

TRIM24 (1 mutation in 549 samples)

CLCN6 (1 mutation in 96 samples)

GNAI1 (1 mutation in 96 samples)

MKRN1 (1 mutation in 96 samples)

CEP89 (1 mutation in 38 samples)

LSM14A (1 mutation in 38 samples)

GATM (1 mutation in 11 samples)

HERPUD1_ENST00000300302 (1 mutation in 11 samples)

ZSCAN30_ENST00000333206 (1 mutation in 11 samples)

Drug sensitivity data

Mutations in BRAF are associated with altered sensitivity to the following 31 drugs:

FR-180204, VX-11e, Camptothecin

[Show all](#)

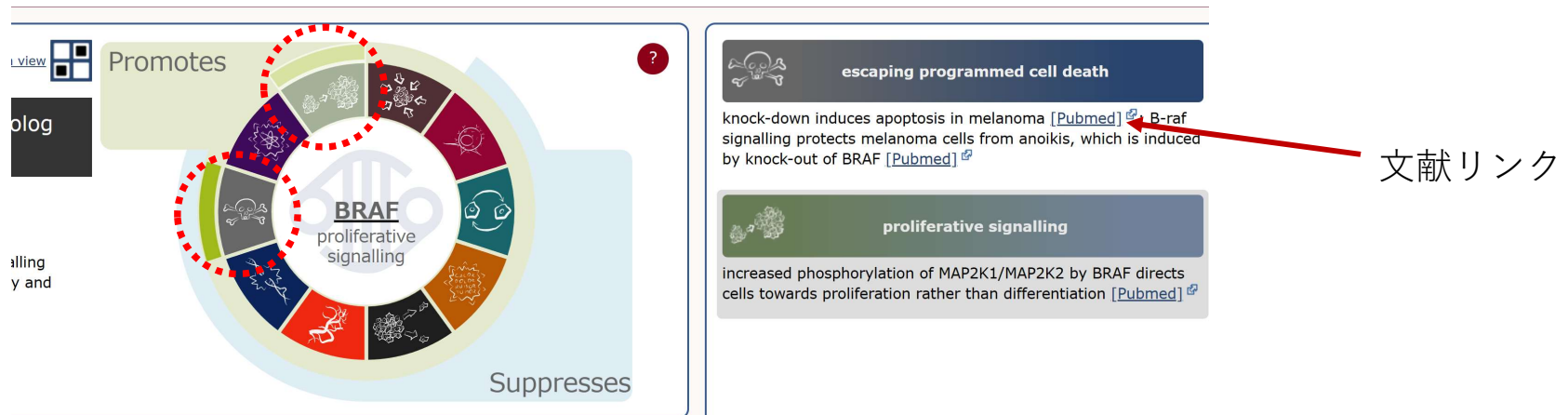
See all [drug sensitivity data](#) for BRAF.

③Hallmark geneをクリック

パートナー遺伝子

遺伝子変異が感受性に影響する薬剤

タイプ	説明
Census gene	Cancer Gene Census (CGC) に登録の遺伝子
Curated gene	COSMICのキュレーターにより登録された遺伝子. 該当しない遺伝子是他DBやゲノムワイド研究に由来
Mouse gene	マウスの実験でがん遺伝子とされた遺伝子
Hallmark gene	がんの進展に関する10個の特性(Hallmark)に当たる遺伝子



BRAFは「proliferative signalling」と「escaping programmed cell death」に該当

他のバイオインフォマティクスの情報へのリンク

External links

Links to bioinformatics resources that are related to BRAF.

OMIM	164757
Transcript	ENST00000646891.1
Genome Browsers	Ensembl , UCSC
Copy Number	CONAN
NCBI Entrez Gene	673
CCDS	CCDS5863.1
UniProt	n/a
Pfam	n/a
Atlas Genetic Oncology	BRAF
HGNC	1097

Drug resistance

This section shows the drugs associated with **BRAF** resistance mutations. In the tabs below you can see any other genes that have resistance mutations

Alternative transcripts are also displayed here for genes where reported resistant mutations are not located on the canonical transcript but are on the alt genomic position on both the canonical and alternative transcripts or on overlapping genes and/or fusions and share a COSM id.

You can change the list of drugs that are used to filter data in the panels below; click the name of a drug to toggle it on or off, then click "Update drugs".

Dabrafenib Vemurafenib

Update drugs

薬剤や遺伝子を選択し直してUpdate drugs, Update genesをクリックすると下のグラフが更新される

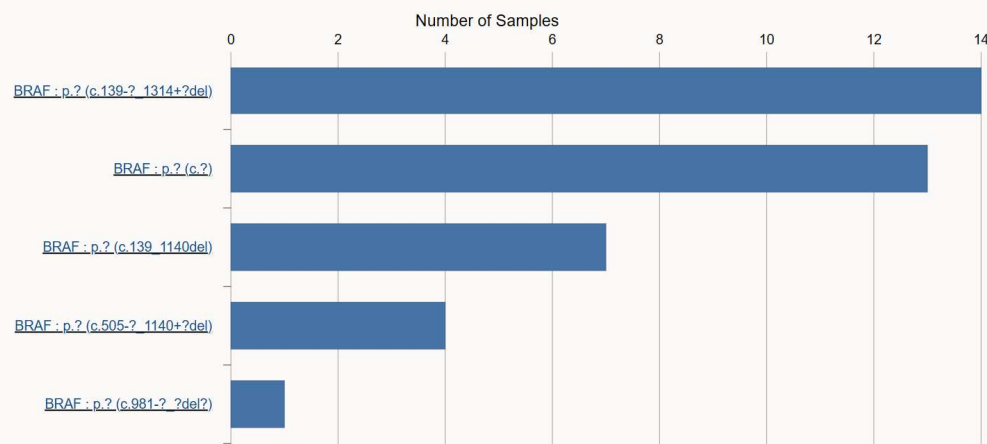
Mutations

Genes

For each of the selected genes and selected drugs, this histogram shows the number of samples with a particular resistant mutation. You can change the to toggle it on or off, then click "Update genes".

BRAF BRAF_ENST00000288602 CDKN2A_ENST00000304494 KRAS KRAS_ENST00000311936 MAP2K1 MAP2K2 MAP2K2_ENST0000039

Update genes



左図では
Dabrafenibと
Vemurafenib
に対するBRAFの変異に関して、
抵抗性変異とその観察数が示
されている

原発組織ごとに、BRAFの各変異タイプの分布を示す

Tissue distribution

The table shows the distribution of mutations across the primary tissue types that are curated by COSMIC. Histograms show the percentage of mutated samples for point mutations, CNV data and gene expression data. Moving your mouse over the histograms will show additional data. The number of samples tested on this page include samples from the targeted and whole genomes/exome resequencing where all the protein coding genes have been screened for mutations.

You can see additional information about the data presented here in the [help pages](#).

Show entries

Search:

Tissue	Point Mutations		Copy Number Variation		Gene Expression		Methylation	
	% Mutated	Tested	Variant %	Tested	% Regulated	Tested	% Diff. Methylated	Tested
Adrenal gland		1174		267		79		-
Autonomic ganglia		1620		-		-		-
Biliary tract		3335		-		-		-
Bone		1310		-		-		-
Breast		10855		1492		1104		-
Central nervous system		9609		1035		697		-
Cervix		1756		299		307		-
Endometrium		5081		586		602		-
Eye		1621		-		-		-
Fallopian tube		8		-		-		-
Female genital tract (site indeterminate)		54		-		-		-
Gastrointestinal tract (site indeterminate)		1094		-		-		-
Genital tract		530		-		-		-
Haematopoietic and lymphoid		25191		661		221		-
Kidney		3716		995		600		513
Large intestine		95723		718		610		281
Liver		3270		663		373		-
Lung		30897		1006		1019		717

Variants

Mutations Fusions CNV & Expression Methylation

This tab displays a table of mutations for the selected gene. You can see more information in our [help pages](#).

BRAF変異の一覧

Show 10 entries

Export: CSV TSV Search:

Position (AA)	Mutation (CDS)	Mutation (Amino Acid)	Legacy Mutation ID	Count	Mutation Type
592	c.1773_1793dup	p.I592_A598dup	COSM6979209	1	Insertion - In frame
594	c.1779_1780delinsGA	p.D594N	COSM211600	1	Substitution - Missense
594	c.1780G>A	p.D594N	COSM27639	80	Substitution - Missense
594	c.1780G>C	p.D594H	COSM144576	3	Substitution - Missense
594	c.1780G>T	p.D594Y	COSM6928839	3	Substitution - Missense
594	c.1781A>C	p.D594A	COSM1583010	5	Substitution - Missense
594	c.1781A>G	p.D594G	COSM467	141	Substitution - Missense
594	c.1781A>T	p.D594V	COSM466	6	Substitution - Missense
594	c.1781_1798dup	p.D594_T599dup	COSM26504	1	Insertion - In frame
594	c.1782T>A	p.D594E	COSM1121	4	Substitution - Missense

Showing 1 to 10 of 164 entries

First Previous 1 2 3 4 5 ... 17 Next Last

References

This section shows publications associated with BRAF. You can see more information in our [help pages](#).

BRAF変異に関するリファレンス

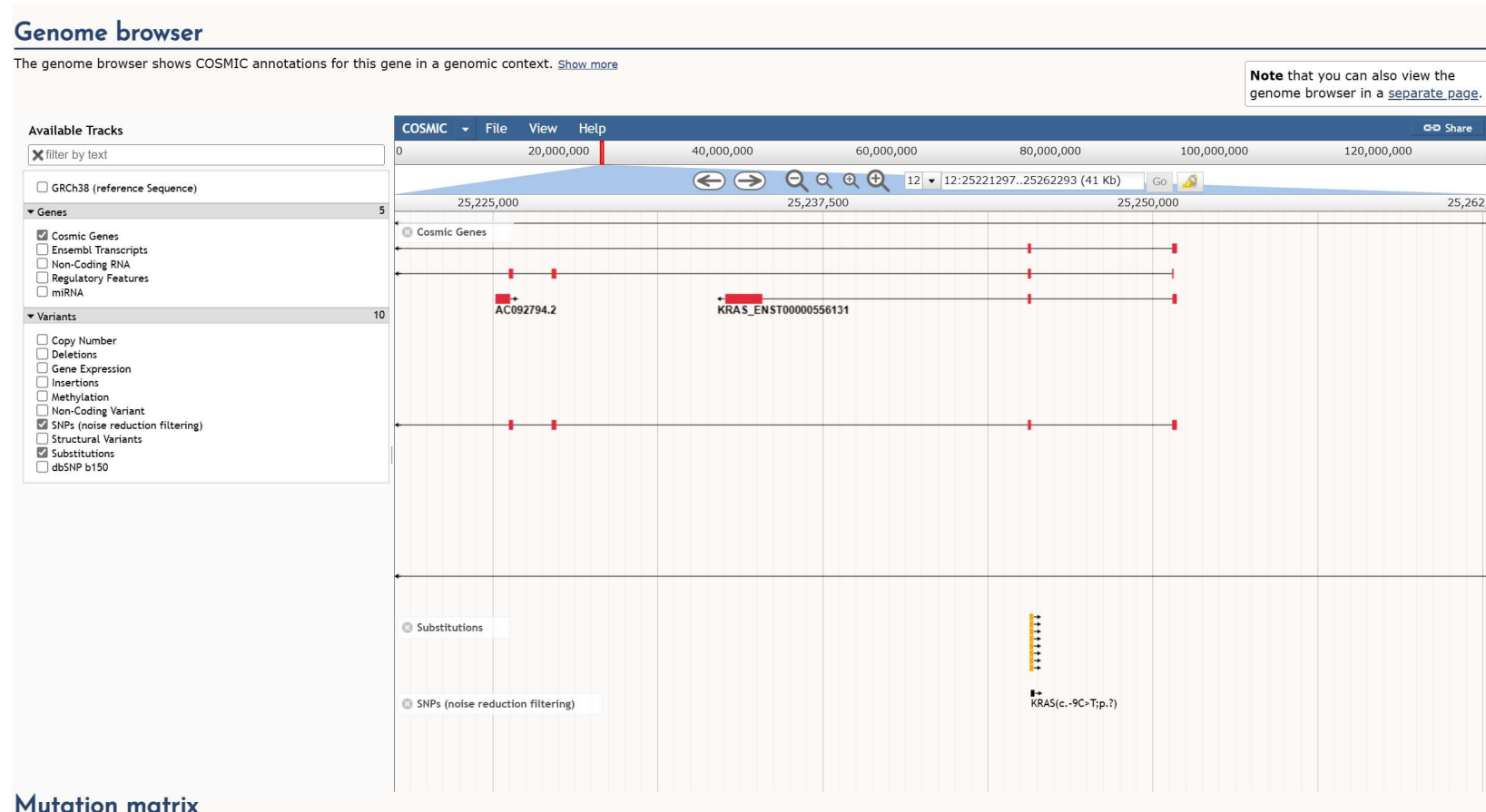
Show 10 entries

Reference Title	Author	Year	Journal	Status	COSMIC	Pubmed
BRAF and RAS mutations in human lung cancer and melanoma	Brose MS et al	2002	Cancer research;62(23):6997-7000	Curated	COSP5705	12460918
Missense mutations of the BRAF gene in human lung adenocarcinoma	Naoki K et al	2002	Cancer research;62(23):7001-3	Curated	COSP5706	12460919
Mutations of the BRAF gene in human cancer	Davies H et al	2002	Nature;417(6892):949-54	Curated	COSP5661	12068308
Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia	Yuen ST et al	2002	Cancer research;62(22):6451-5	Curated	COSP5697	12438234
Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status	Rajagopalan H et al	2002	Nature;418(6901):934	Curated	COSP5680	12198537
A genome-based strategy uncovers frequent BRAF mutations in melanoma	Pollock PM and Meltzer PS	2002	Cancer cell;2(1):5-7	Review	COSP7984	12150818
Absence of BRAF and NRAS mutations in uveal melanoma	Cruz F et al	2003	Cancer research;63(18):5761-6	Curated	COSP7978	14522897
Absence of BRAF gene mutations in uveal melanomas in contrast to cutaneous melanomas	Edmunds SC et al	2003	British journal of cancer;88(9):1403-5	Curated	COSP7983	12778069
Absence of exon 15 BRAF germline mutations in familial melanoma	Lang J et al	2003	Human mutation;21(3):327-30	Curated	COSP5725	12619120
Absence of mutations of the BRAF gene and constitutive activation of extracellular-regulated kinase in malignant melanomas of the uvea	Weber A et al	2003	Laboratory investigation; a journal of technical methods and pathology;83(12):1771-6	Curated	COSP8379	14691295

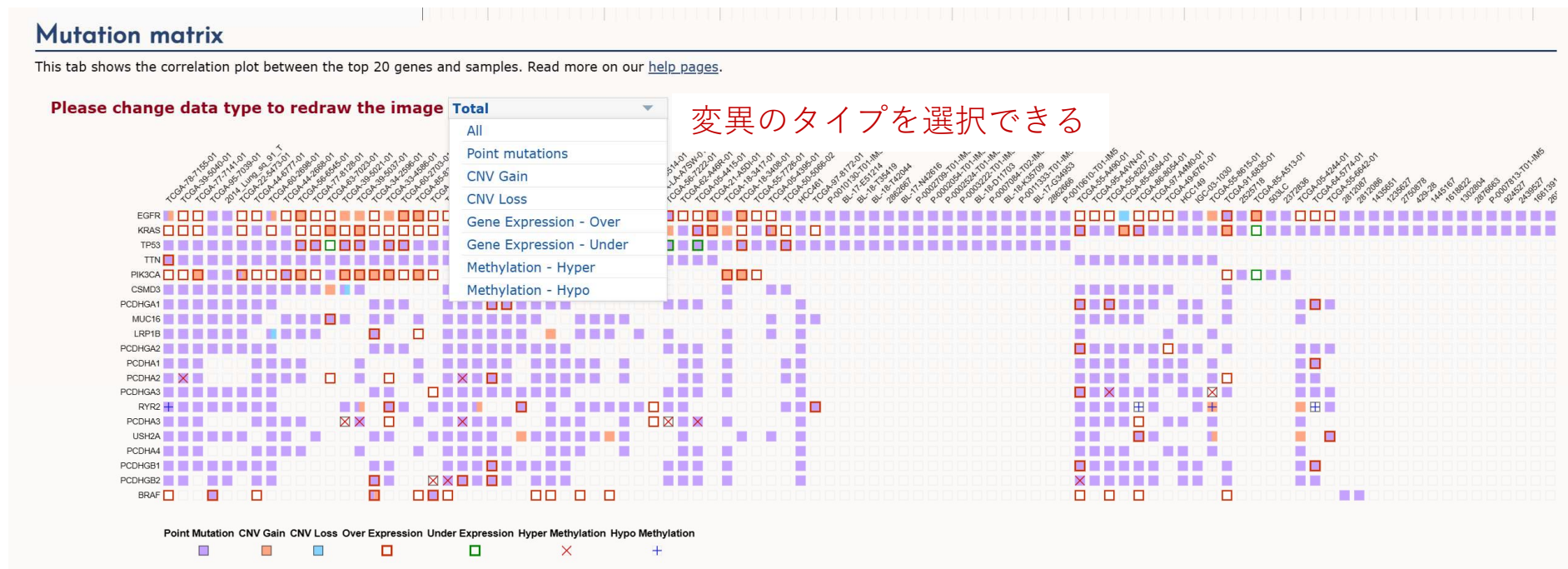
Showing 1 to 10 of 5,968 entries

First Previous 1 2 3 4 5 ... 597 Next Last

Genome browserでSNPと置換変異を表示させた例



サンプルごとに変異遺伝子を示す。変異タイプも分かる



OncoKB

Welcome to OncoKB™

MSK's Precision Oncology Knowledge Base

An FDA-Recognized Human Genetic Variant Database*

820

Genes

7536

Alterations

136

Cancer Types

136

Drugs

Search Gene / Alteration / Cancer Type / Drug / Genomic Variant 

Therapeutic Levels

Diagnostic Levels

Prognostic Levels

FDA Levels

1 Level 1

FDA-approved drugs

51 Genes

2 Level 2

Standard care

24 Genes

3 Level 3

Clinical evidence

35 Genes

4 Level 4

Biological evidence

27 Genes

R1/R2 Level R1/R2

Resistance

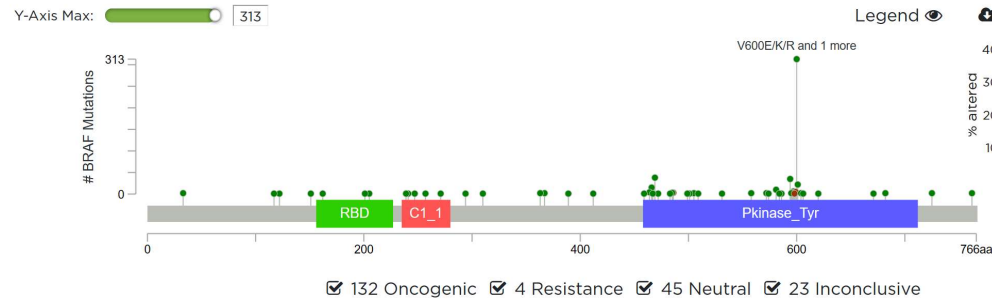
11 Genes

治療効果に関する
エビデンスレベル

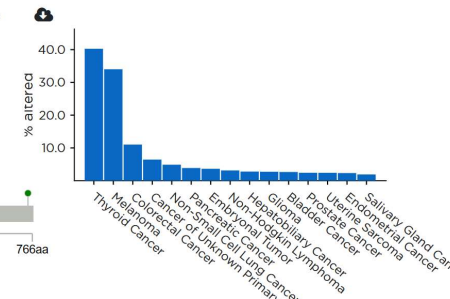
- OncoKBは、がん治療において特定の遺伝子変異に関する臨床的な情報を提供する精密医療の知識ベース
- 患者の腫瘍の分子特性に基づいて、特定の遺伝子変異がどの程度治療に影響を与えるかを評価し、FDA承認薬や臨床試験の情報を含めて医師や研究者に提供する
- OncoKBは、患者の治療戦略をサポートし、がん治療の個別化と最適化を目指している

OncoKB: トップページ画面でBRAFで検索

Annotated Mutations in MSK-IMPACT™ Clinical Sequencing Cohort (Zehir et al., Nat Med 2017)



Cancer Types with BRAF Mutations



Annotated Alterations Therapeutic Diagnostic FDA-Recognized Content

A list of the cancer type-specific BRAF alterations that may predict response to a targeted drug and the corresponding OncoKB™ level of evidence assigning their level of [clinical actionability](#).

If you notice any mistakes or omissions, please reach out to us. [✉](#)

Level	Alterations	Level-associated cancer types	Drugs	Citations	Description
1	V600	Erdheim-Chester Disease	Vemurafenib	2	
1	V600	Melanoma	Vemurafenib + Cobimetinib + Atezolizumab		
1	V600E	All Solid Tumors (excluding Colorectal Cancer)	Dabrafenib + Trametinib		
1	V600E	Anaplastic Thyroid Cancer	Dabrafenib + Trametinib		
1	V600E	Biliary Tract Cancer, NOS	Dabrafenib + Trametinib		

Citations列をマウスオーバーするとエビデンスの出典が出現する

エビデンスレベル = Level1 (FDAが承認した薬剤に対するFDAが認めたバイオマーカー)で、V600変異陽性のメラノーマに対して「Vemurafenib + Cobimetinib + Atezolizumab」が有効

OncoKB: Cancer Gene List

OncoKBの”Cancer Genes”リストはダウンロードできる

①クリック



OncoKB™ Cancer Gene List

1130 genes, last update 11/13/2023

The following genes are considered to be cancer genes by OncoKB™, based on their inclusion in various different sequencing panels, the Sanger Cancer Gene Census, or [Vogelstein et al. \(2013\)](#).

②ダウンロード →

Cancer Gene List

OncoKBによる分類

Cancer Gene Census
Tier1遺伝子

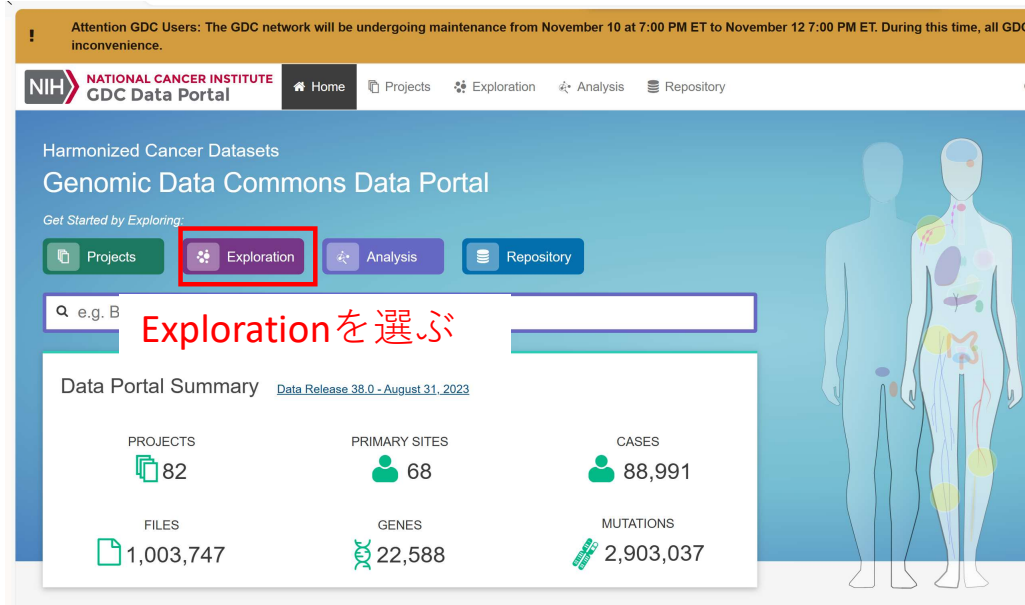
Gene	OncoKB™ Annotated	OncoKBによる分類 OncoGene/TSG	MSK-IMPACT™	MSK-IMPACT™ Heme	Foundation One CDx	Foundation One Heme	Vogelstein et al. 2013	Cancer Gene Census Tier 1	# of Sources
ABL1	✓	Oncogene	✓	✓	✓	✓	✓	✓	7
AKT1	✓	Oncogene	✓	✓	✓	✓	✓	✓	7
ALK	✓	Oncogene	✓	✓	✓	✓	✓	✓	7
AMER1	✓	TSG	✓	✓	✓	✓	✓	✓	7
APC	✓	TSG	✓	✓	✓	✓	✓	✓	7
AR	✓	Oncogene	✓	✓	✓	✓	✓	✓	7
ARID1A	✓	TSG	✓	✓	✓	✓	✓	✓	7
ASXL1	✓	TSG	✓	✓	✓	✓	✓	✓	7
ATM	✓	TSG	✓	✓	✓	✓	✓	✓	7

各種がんゲノムプロファイリング
検査での掲載状況

Vogelstein et al.(2013)
で挙げられた遺伝子

変異と治療の対応を示すリストをダウンロードするには、登録(審査あり)が必要 (実際に試すとごく簡単な申請で1営業日で困難なく承認された)

TCGA



- The Cancer Genome Atlas (TCGA) は、2006年から始まったNCI（National Cancer Institute）とNHGR（National Human Genome Research Institute）の共同プロジェクト
- このプログラムでは、33種類に及ぶがん種の20,000以上の原発がんと対応する正常サンプルからのゲノム、エピゲノム、トランスクリプトームなどのデータが解析されている
- このデータは既のがんの診断、治療、予防能力の向上につながっており、誰でも研究コミュニティで利用できるようになっている

あるがん種について、 各遺伝子の変異保有率をヒストグラムで表す

TCGA-LUAD (Lung Adenocarcinoma) と TCGA-COAD (Colon Adenocarcinoma) について調べてみる

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis Repository

Cases Clinical Genes Mutations

Search Cases
e.g. TCGA-A5-A0G2, 432fe4a9-2...

Upload Case Set

TCGA 563

Project

- CPTAC-3 1,191
- TCGA-BRCA 1,079
- MMRF-COMMPASS 953
- TARGET-ALL-P2 628
- TCGA-OV 570
- ☒ TCGA-LUAD 563
- TCGA-HNSC 523
- TCGA-KIRC 523
- TCGA-LGG 509
- TCGA-LUSC 504

Clear Project Id IS TCGA-LUAD AND Is Cancer Gene Census IS true

Cases (563) Genes (714) Mutations (11,824) OncoGrid

Gene ② Genesタブをクリック

Distribution of Most Frequently Mutated Genes

③ PNG形式でダウンロードして保存

SVG PNG JSON

% of Cases Affected

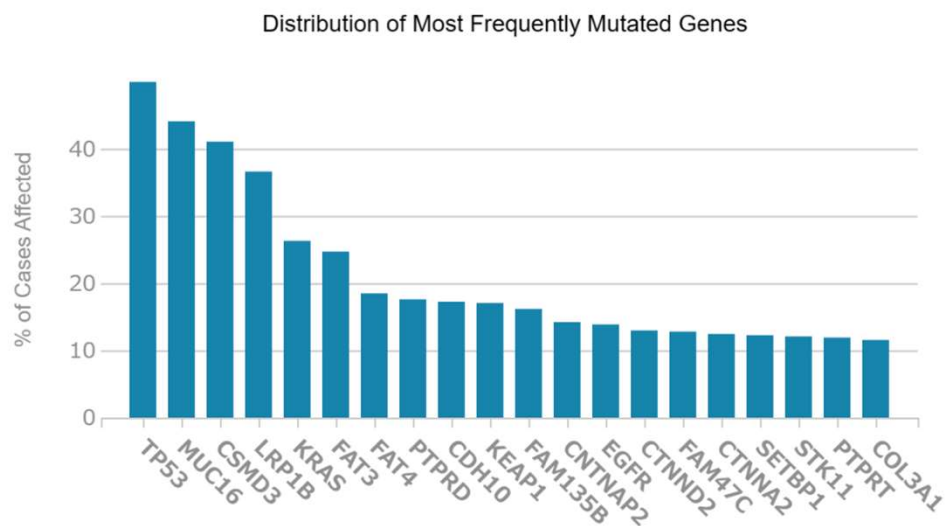
Survival Rate

Showing 1 - 10 of 714 genes

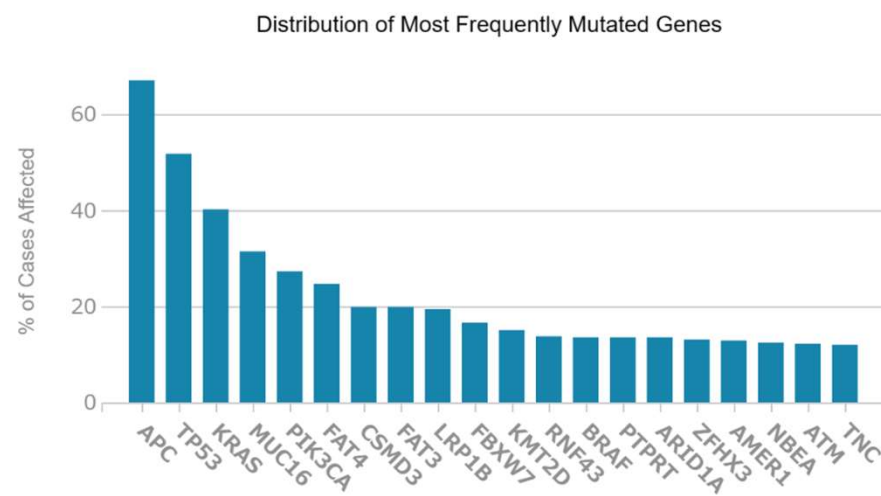
Gene	% of Cases Affected
TP53	45
MUC16	42
CSMD3	40
LRP1B	35
KRAS	25
FAT3	24
FAT4	18
PTPRD	17
CDH10	16
KEAP1	15
FAM135B	14
CNTNAP2	13
EGFR	12
CTNND2	11
FAM47C	11
CTNNA2	11
SETBP1	11
STK11	11
PTPRT	11
COL3A1	11

※ TCGA-COADについても同様に操作する

TCGA-LUAD



TCGA-COAD



3つのがん種の変異遺伝子の集合関係を視覚化

TCGA-LUAD (Lung Adenocarcinoma) と TCGA-COAD (Colon Adenocarcinoma)
TCGA-STAD (Stomach Adenocarcinoma) を対象とする

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

② Genesタブをクリック

① ProjectからTCGA-LUADを選ぶ

TCGA-LUAD

Distribution of Most Frequently Mutated Genes

Overall Survival Plot

503 Cases with Survival Data

Showing 1 - 10 of 714 genes

Symbol	Name	# SSM Affected Cases in Cohort	# SSM Affected Cases Across the GDC
TP53	tumor protein p53	281 / 553 (50.81%)	4,928 / 16,405
MUC16	mucin 16, cell surface associated	248 / 553 (44.85%)	2,887 / 16,405
CSMD3	CUB and Sushi multiple domains 3	231 / 553 (41.77%)	1,984 / 16,405
LRP1B	LDL receptor related protein 1B	206 / 553 (37.25%)	1,780 / 16,405

③ Save as new gene setを選択
Save Top:20を指定し、
「TCGA-LUAD-TOP20genes」として保存

TCGA-COADとTCGA-STADも同様に操作する

NIH NATIONAL CANCER INSTITUTE
GDC Data Portal

Home Projects Exploration **Analysis** Repository

Launch Analysis Results

① Analysisを選択

Set Operations
Display Venn diagram and find intersection or union, etc. of your sets of the same type.

Select Demo

② Selectをクリック

Clinical Data Analysis
Display basic statistical analyses for the selected case set.

Select Demo

Launch Analysis Results

Set Operations
Display Venn diagram and find intersection or union, etc. of your sets of the same type.

Back Demo

Select 2 or 3 of the same set type
You can create and save case, gene, and mutation sets from the [Exploration Page](#).

Select	Type	Name	Items
<input type="checkbox"/>	Genes	TCGA-LUAD_mt5%gene	89
<input type="checkbox"/>	Genes	TCGA-COAD_mt5%gene	177
<input type="checkbox"/>	Genes	TCGA-STAD_mt5%gene	164
<input checked="" type="checkbox"/>	Genes	TCGA-LUAD-TOP20genes	20
<input checked="" type="checkbox"/>	Genes	TCGA-COAD-TOP20genes	20
<input checked="" type="checkbox"/>	Genes	TCGA-STAD-TOP20genes	20
<input type="checkbox"/>	Cases	TCGA-LUAD-TP53	281
<input type="checkbox"/>			148
<input type="checkbox"/>			39
<input type="checkbox"/>			78
<input type="checkbox"/>	Cases	TCGA-LUAD-TOP20genes	20
<input type="checkbox"/>	Cases	TCGA-COAD-TOP20genes	20

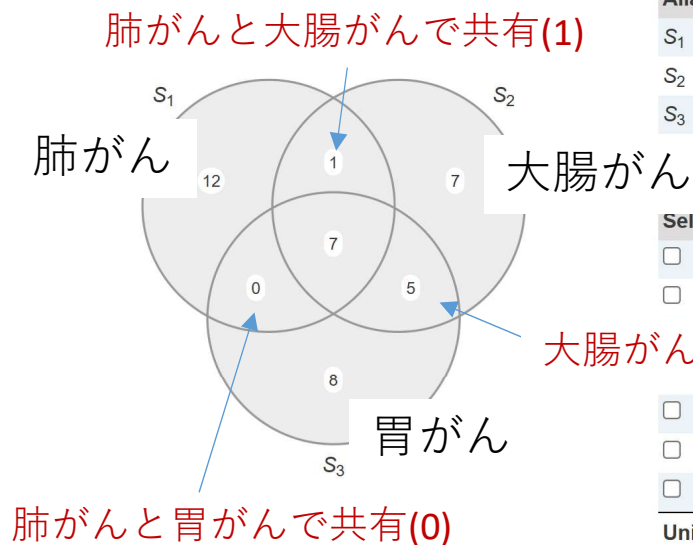
③作成したセットを選択

Run

④ Run

Set Operations

Click on the areas of the Venn diagram to include them in your result set.



Alias	Item Type	Name	# Items
S ₁	Genes	TCGA-LUAD-TOP20genes	20
S ₂	Genes	TCGA-COAD-TOP20genes	20
S ₃	Genes	TCGA-STAD-TOP20genes	20

Select	Set Operation	# Items	Save
<input type="checkbox"/>	$(S_1 \cap S_2 \cap S_3)$	7	
<input type="checkbox"/>	$(S_1 \cap S_2) - (S_3)$	1	
<input type="checkbox"/>	$(S_1 \cap S_3) - (S_2)$	0	
<input type="checkbox"/>	$(S_2 \cap S_3) - (S_1)$	5	
<input type="checkbox"/>	$(S_1) - (S_2 \cup S_3)$	12	
<input type="checkbox"/>	$(S_2) - (S_1 \cup S_3)$	7	
<input type="checkbox"/>	$(S_3) - (S_1 \cup S_2)$	8	

Union of selected sets

クリックすると該当する遺伝子を確認できる

全がん種で変異している7遺伝子：TP53, MUC16, CSMD3, LRP1B, FAT4, FAT3, PTPRT

大腸がん胃がんでは変異している5遺伝子：PIK3CA, KMT2D, ARID1A, NBEA, RNF43

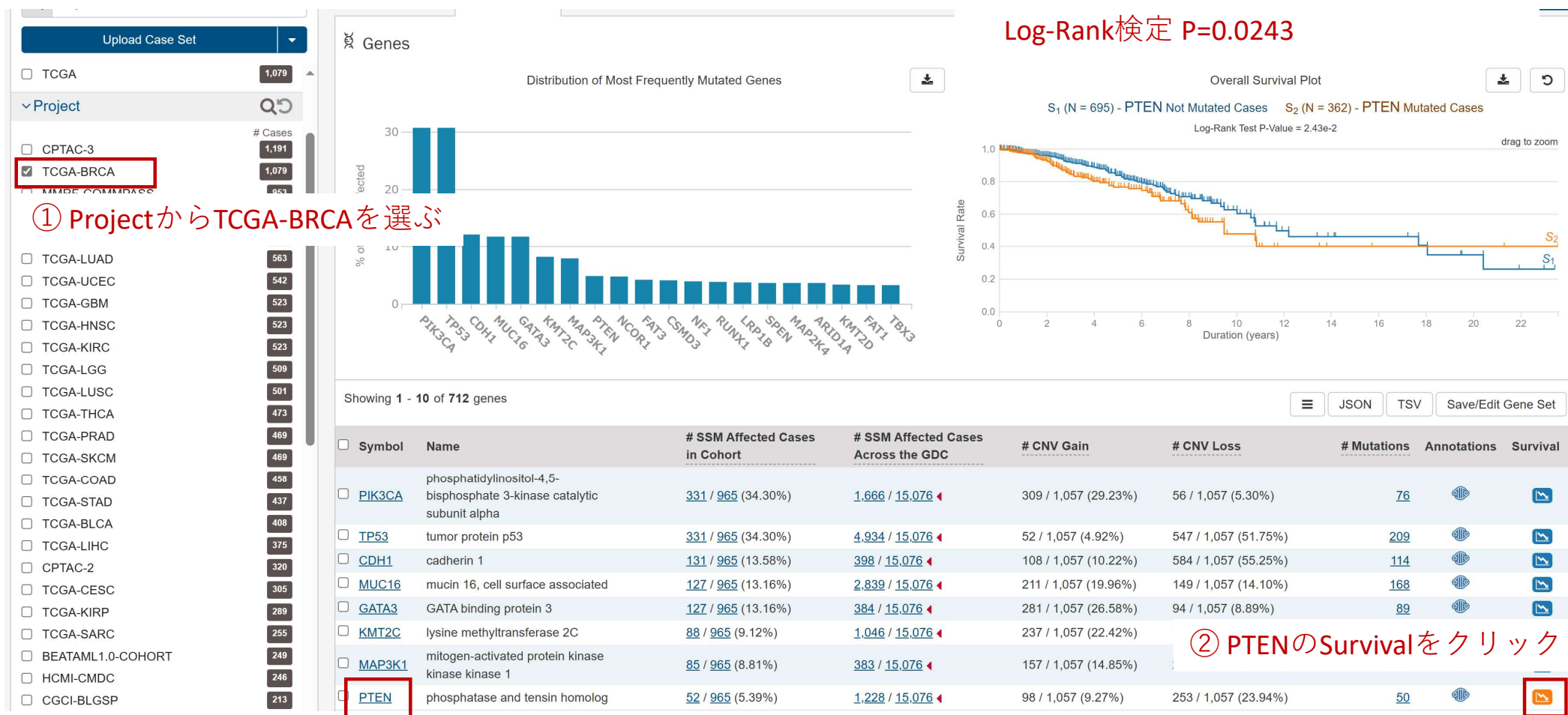
肺がん大腸がんでは変異している1遺伝子：KRAS

乳がんで予後因子として報告されているPTEN変異について、TCGAデータで予後との関連を見る

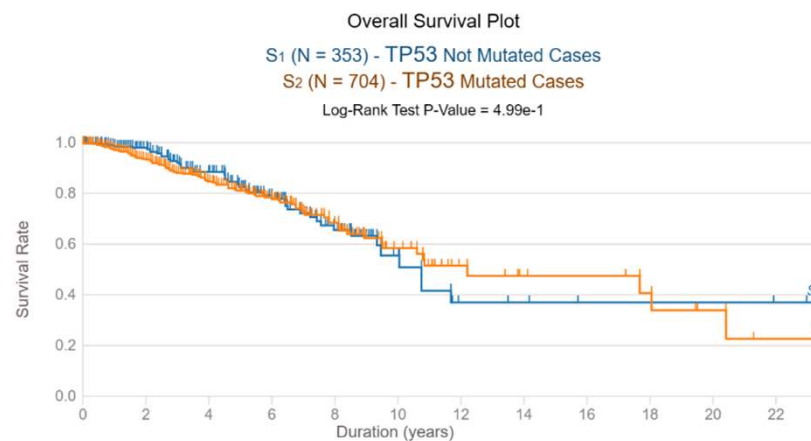
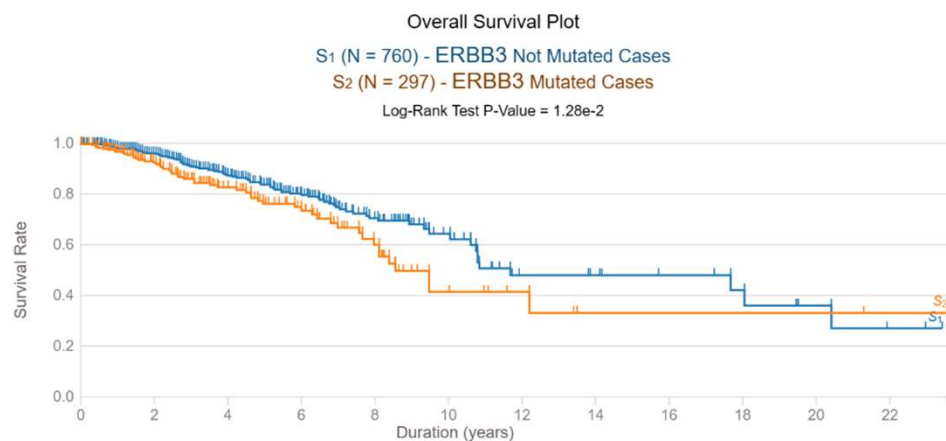
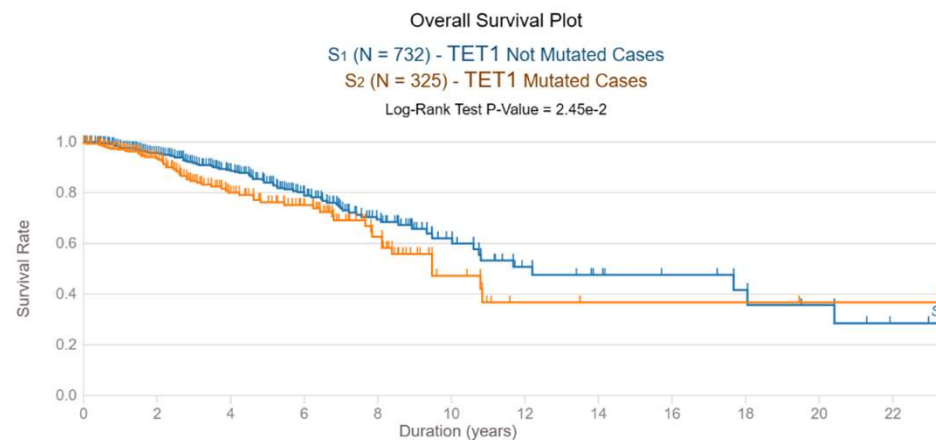
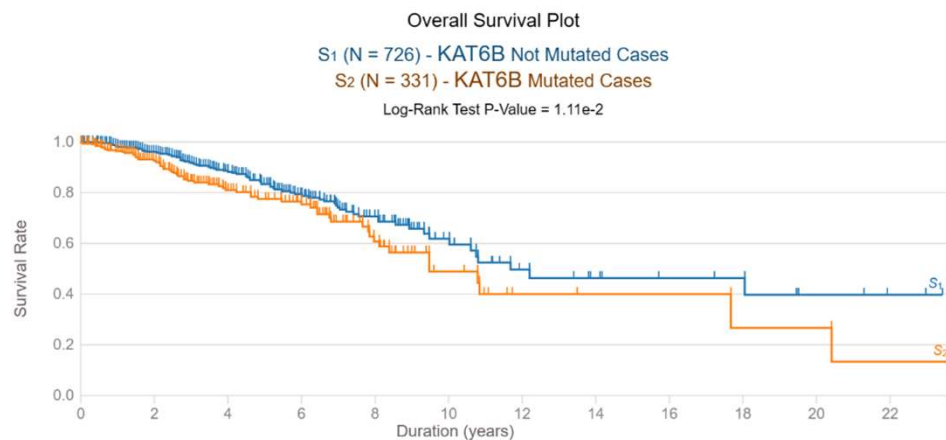
PTEN変異と予後との関連に関する参考文献：

Zhang, Hong-Yan, et al. "PTEN mutation, methylation and expression in breast cancer patients." *Oncology letters* 6.1 (2013): 161-168.

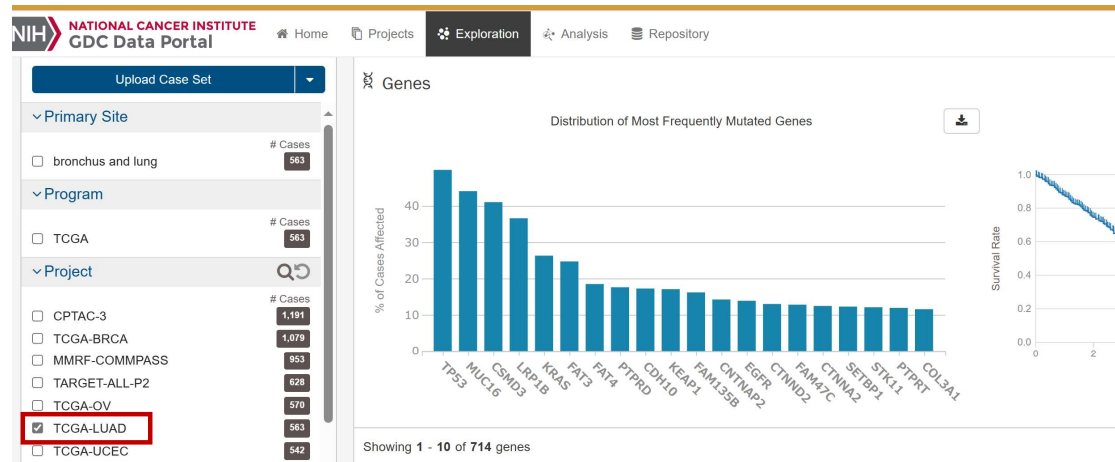
③ PTEN変異の有無で生存曲線が描かれる
変異無し 695例 vs. 有り 362例
Log-Rank検定 P=0.0243



他の遺伝子についても確認してみる(対象: TCGA-BRCA)



肺腺がんサンプルではKRAS/BRAF/EGFRは相互排他的に変異しているか.TCGAサンプルで確かめる

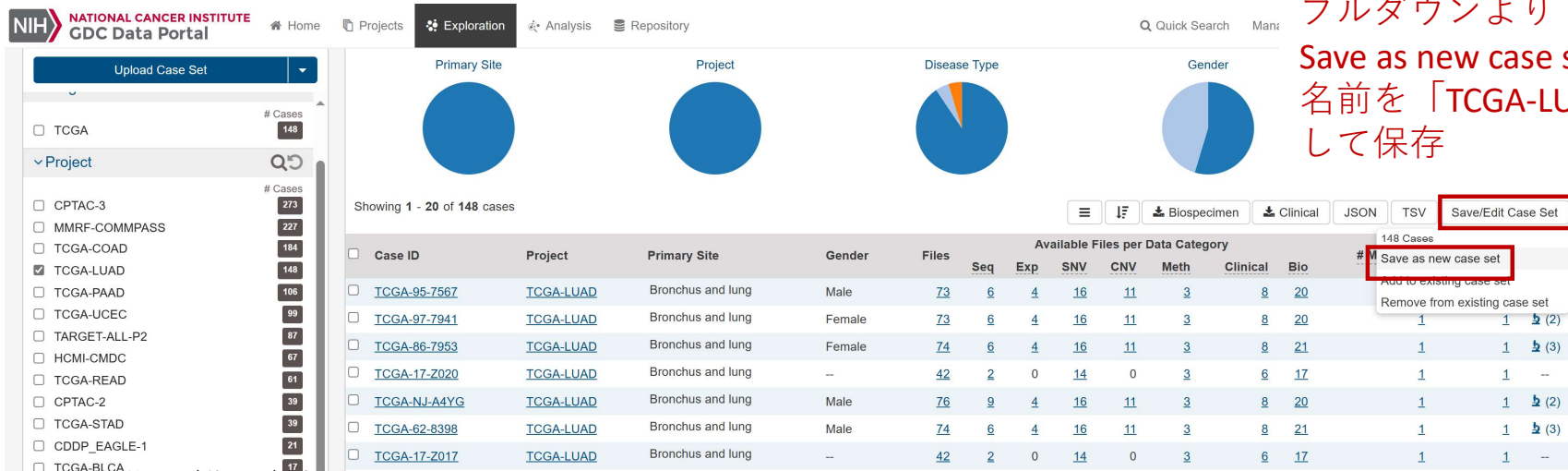


① TCGA-LUADを選ぶ

② KRAS 変異ありの148例をクリック

③ Save/Edit Case Setをクリック
プルダウンより

Save as new case setを選択
名前を「TCGA-LUAD-KRAS」として保存



同様にして「TCGA-LUAD-BRAF」、 「TCGA-LUAD-EGFR」 を作成して保存

その後、以下のとおり進める

NIH NATIONAL CANCER INSTITUTE
GDC Data Portal

Home Projects Exploration **Analysis** Repository

Launch Analysis Results

④ Analysisを選択

Set Operations
Display Venn diagram and find intersection or union, etc. of your sets of the same type.

Select Demo

⑤ Selectをクリック

Clinical Data Analysis
Display basic statistical analyses for the selected case set.

Select Demo

NIH NATIONAL CANCER INSTITUTE
GDC Data Portal

Home Projects Exploration **Analysis** Repository

Launch Analysis Results

Set Operations
Display Venn diagram and find intersection or union, etc. of your sets of the same type.

Back Demo

Select 2 or 3 of the same set type

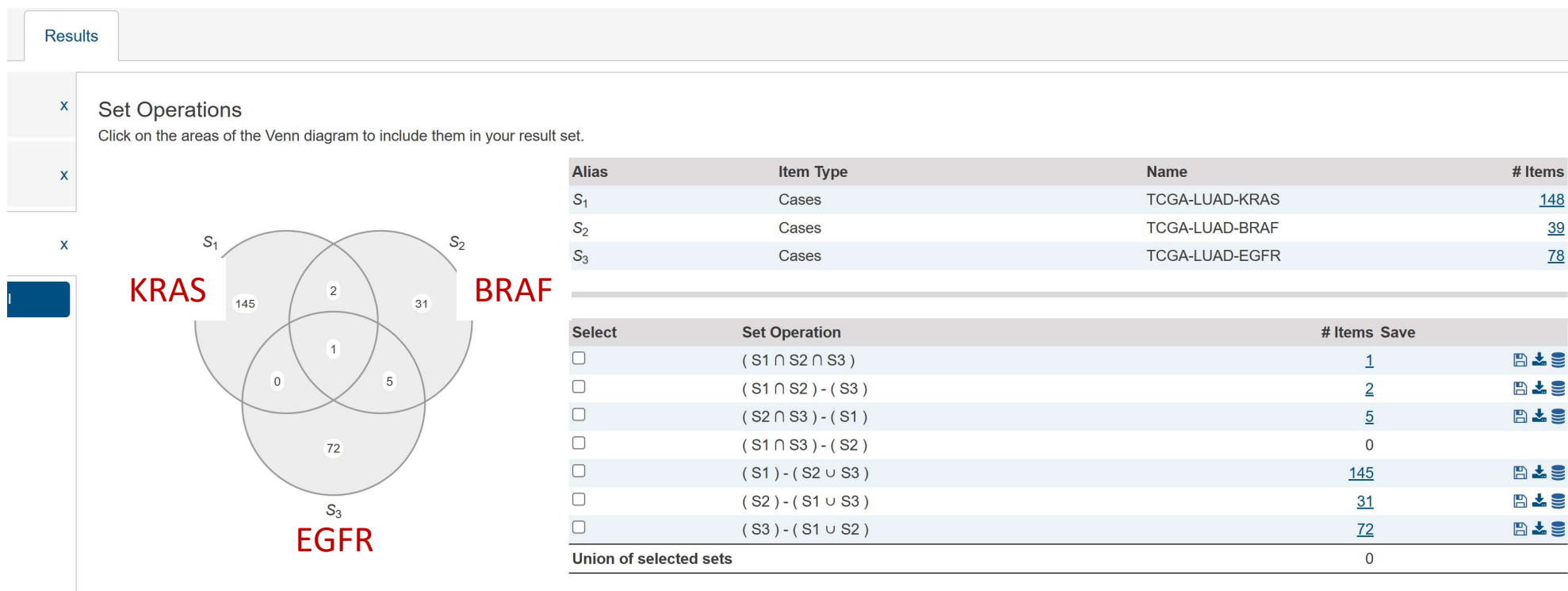
You can create and save case, gene, and mutation sets from the [Exploration Page](#).

Select	Type	Name	Items
<input type="checkbox"/>	Genes	TCGA-LUAD_mt5%gene	89
<input type="checkbox"/>	Genes	TCGA-COAD_mt5%gene	177
<input type="checkbox"/>	Genes	TCGA-STAD_mt5%gene	164
<input type="checkbox"/>	Genes	TCGA-LUAD-TOP20genes	20
<input type="checkbox"/>	Genes	TCGA-COAD-TOP20genes	20
<input type="checkbox"/>	Genes	TCGA-STAD-TOP20genes	20
<input type="checkbox"/>	Cases	TCGA-LUAD-TP53	281
<input checked="" type="checkbox"/>	Cases	TCGA-LUAD-KRAS	148
<input checked="" type="checkbox"/>	Cases	TCGA-LUAD-BRAF	39
<input checked="" type="checkbox"/>	Cases	TCGA-LUAD-EGFR	78
<input type="checkbox"/>	Cases	TCGA-LUAD-TOP20genes	20
<input type="checkbox"/>	Cases	TCGA-COAD-TOP20genes	20

⑥ 作成した3つのCases setsを選択

Run

⑦ Run



KRAS変異有のサンプル、BRAF変異有のサンプル、EGFR変有のサンプルはあまり重なっていない
ほぼ、排他的に変異していることがわかった

TCGA-COAD (大腸がん)ではどうか 同様にして調べた結果は以下のとおり

Set Operations

Click on the areas of the Venn diagram to include them in your result set.



Alias	Item Type	Name	# Items
S ₁	Cases	TCGA-COAD-APC	309
S ₂	Cases	TCGA-COAD-TP53	233
S ₃	Cases	TCGA-COAD-KRAS	174

Select	Set Operation	# Items	Save
<input type="checkbox"/>	$(S_1 \cap S_2 \cap S_3)$	81	
<input type="checkbox"/>	$(S_1 \cap S_2) - (S_3)$	109	
<input type="checkbox"/>	$(S_2 \cap S_3) - (S_1)$	12	
<input type="checkbox"/>	$(S_1 \cap S_3) - (S_2)$	65	
<input type="checkbox"/>	$(S_1) - (S_2 \cup S_3)$	54	
<input type="checkbox"/>	$(S_2) - (S_1 \cup S_3)$	31	
<input type="checkbox"/>	$(S_3) - (S_1 \cup S_2)$	16	
Union of selected sets		0	

Set Operations

Click on the areas of the Venn diagram to include them in your result set.



Alias	Item Type	Name	# Items
S ₁	Cases	TCGA-COAD-TP53	233
S ₂	Cases	TCGA-COAD-KRAS	174
S ₃	Cases	TCGA-COAD-PIK3CA	124

Select	Set Operation	# Items	Save
<input type="checkbox"/>	$(S_1 \cap S_2 \cap S_3)$	31	
<input type="checkbox"/>	$(S_1 \cap S_2) - (S_3)$	62	
<input type="checkbox"/>	$(S_2 \cap S_3) - (S_1)$	39	
<input type="checkbox"/>	$(S_1 \cap S_3) - (S_2)$	27	
<input type="checkbox"/>	$(S_1) - (S_2 \cup S_3)$	113	
<input type="checkbox"/>	$(S_2) - (S_1 \cup S_3)$	42	
<input type="checkbox"/>	$(S_3) - (S_1 \cup S_2)$	27	
Union of selected sets		0	

TCGAデータのダウンロード

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects **Exploration** Analysis Repository

Quick Search Manage Sets Login Cart 0 GDC Apps

Cases Clinical Genes Mutations

① Explorationを選択

Search Cases

e.g. TCGA-A5-A0G2, 432fe4a9-2...

Upload Case Set

☐ CPTAC-3 1,191

☐ TCGA-BRCA 1,079

☐ MMRF-COMMPASS 953

☐ TARGET-ALL-P2 626

☐ TCGA-OV 570

☒ TCGA-LUAD 563

☐ TCGA-UCEC 542

☐ TCGA-CM 563

TCGA-LUAD

② TCGA-LUADをチェック

View Files in Repository

② View Files ...をクリック

Cancer Gene Census IS true

Cases (563) Genes (714) Mutations (11,824) OncoGrid

Primary Site Project Disease Type Gender

Showing 1 - 20 of 563 cases

Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category							# Mutations	# Genes	Slides
					Seq	Exp	SNV	CNV	Meth	Clinical	Bio			
TCGA-17-Z031	TCGA-LUAD	Bronchus and lung	--	42	2	0	14	0	3	6	17	163	108	--
<input type="checkbox"/> TCGA-55-8506	TCGA-LUAD	Bronchus and lung	Female	74	6	4	16	11	3	8	21	111	92	b (3)
<input type="checkbox"/> TCGA-05-4382	TCGA-LUAD	Bronchus and lung	Male	78	10	4	16	11	3	9	21	117	86	b (3)
<input type="checkbox"/> TCGA-69-7979	TCGA-LUAD	Bronchus and lung	Female	66	6	4	16	5	3	8	20	111	85	b (2)
<input type="checkbox"/> TCGA-78-7155	TCGA-LUAD	Bronchus and lung	Male	76	8	4	16	11	3	8	21	108	82	b (3)

Files Cases

[Add a File Filter](#)

Search Files ?

Q e.g. 142682.bam, 4f6e2e7a-b...

Data Category

	# Files
<input type="checkbox"/> simple nucleotide variation	9,855
<input type="checkbox"/> copy number variation	6,044
<input type="checkbox"/> sequencing reads	4,279
<input type="checkbox"/> biospecimen	2,687
<input type="checkbox"/> structural variation	2,388
<input type="checkbox"/> transcriptome profiling	2,322
<input type="checkbox"/> dna methylation	1,914
<input type="checkbox"/> clinical	1,133
<input type="checkbox"/> proteome profiling	363

Less...

Data Type

	# Files
<input type="checkbox"/> Annotated Somatic Mutation	4,976
<input type="checkbox"/> Aligned Reads	4,279
<input type="checkbox"/> Raw Simple Somatic Mutation	2,488
<input type="checkbox"/> Transcript Fusion	2,388
<input type="checkbox"/> Slide Image	1,592
<input type="checkbox"/> Gene Level Copy Number	1,548
<input type="checkbox"/> Masked Intensities	1,276
<input type="checkbox"/> Raw Intensities	1,167
<input type="checkbox"/> Simple Germline Variation	1,167
<input type="checkbox"/> Copy Number Segment	1,143
<input type="checkbox"/> Masked Copy Number Segment	1,143
<input type="checkbox"/> Biospecimen Supplement	1,095
<input type="checkbox"/> Allele-specific Copy Number Segment	1,043
<input type="checkbox"/> Methylation Beta Value	638
<input type="checkbox"/> Clinical Supplement	616
<input type="checkbox"/> Aggregated Somatic Mutation	612
<input type="checkbox"/> Masked Somatic Mutation	612
<input type="checkbox"/> Gene Expression Quantification	597
<input type="checkbox"/> Splice Junction Quantification	597
<input type="checkbox"/> Isoform Expression Quantification	564
<input type="checkbox"/> miRNA Expression Quantification	564
<input type="checkbox"/> Pathology Report	517
<input type="checkbox"/> Protein Expression Quantification	363

Less...

Experimental Strategy

	# Files
<input type="checkbox"/> WXS	9,961
<input type="checkbox"/> Genotyping Array	7,211
<input type="checkbox"/> RNA-Seq	6,373
<input type="checkbox"/> Methylation Array	1,914
<input type="checkbox"/> miRNA-Seq	1,692
<input type="checkbox"/> Tissue Slide	1,056
<input type="checkbox"/> WGS	629
<input type="checkbox"/> Diagnostic Slide	536
<input type="checkbox"/> Reverse Phase Protein Array	363
<input type="checkbox"/> ATAC-Seq	22

Less...

Workflow Type

	# Files
<input type="checkbox"/> DNACopy	2,268
<input type="checkbox"/> BWA with Mark Duplicates and BQSR	1,924
<input type="checkbox"/> SeSAMe Methylation Beta Estimation	1,914
<input type="checkbox"/> MuSE Annotation	1,244
<input type="checkbox"/> MuTect2 Annotation	1,244
<input type="checkbox"/> Pindel Annotation	1,244
<input type="checkbox"/> VarScan2 Annotation	1,244
<input type="checkbox"/> Aliquot Ensemble Somatic Variant Merg...	1,224
<input type="checkbox"/> Arriba	1,194
<input type="checkbox"/> STAR - Counts	1,194
<input type="checkbox"/> STAR-Fusion	1,194
<input type="checkbox"/> Birdseed	1,167
<input type="checkbox"/> BCGSC miRNA Profiling	1,128
<input type="checkbox"/> ASCAT2	1,084
<input type="checkbox"/> ASCAT3	1,002
<input type="checkbox"/> MuSE	622
<input type="checkbox"/> MuTect2	622
<input type="checkbox"/> Pindel	622
<input type="checkbox"/> VarScan2	622
<input type="checkbox"/> STAR 2-Pass Chimeric	597
<input type="checkbox"/> STAR 2-Pass Genome	597
<input type="checkbox"/> STAR 2-Pass Transcriptome	597
<input type="checkbox"/> BWA-aln	564
<input type="checkbox"/> ABSOLUTE LiftOver	505

Less...

Data Format

	# Files
<input type="checkbox"/> txt	5,637
<input type="checkbox"/> vcf	4,976
<input type="checkbox"/> tsv	4,924
<input type="checkbox"/> bam	4,279
<input type="checkbox"/> maf	3,712
<input type="checkbox"/> sv	1,592
<input type="checkbox"/> idat	1,276
<input type="checkbox"/> bedpe	1,194
<input type="checkbox"/> cel	1,167
<input type="checkbox"/> bcr xml	1,079
<input type="checkbox"/> pdf	517
<input type="checkbox"/> bcr ssf xml	516
<input type="checkbox"/> bcr omf xml	94
<input type="checkbox"/> bcr biotab	22

Less...

Platform

	# Files
<input type="checkbox"/> affymetrix snp 6.0	6,044
<input type="checkbox"/> illumina	4,279
<input type="checkbox"/> illumina human methylation 450	1,512
<input type="checkbox"/> illumina human methylation 27	402
<input type="checkbox"/> rppa	363

Access

	# Files
<input type="checkbox"/> controlled	17,674
<input type="checkbox"/> open	13,311

① 左側のウィンドウでデータの属性を指定

② 該当するファイル(1ファイル1例分)が出現するので、ひとつずつCartに入れて、最後にCartをクリックしてダウンロード

※ ダウンロードは後に示すTCGAbiolinksを用いる方が簡単である

※ Controlled access のファイルを利用したいときはGDC に利用申請が必要である

cBioPortal

The screenshot shows the cBioPortal website. At the top, there is a navigation bar with the cBioPortal logo and links for Data Sets, Web API, Tutorials/Webinars, FAQ, News, and Visualize Your Data. Below the navigation bar, there is a search bar with a 'Query' button and a 'Quick Search Beta!' button. The main content area is titled 'Select Studies for Visualization & Analysis:' and shows '0 studies selected (0 samples)'. On the left, there is a list of study categories with their respective sample counts: PanCancer Studies (10), Pediatric Cancer Studies (14), Immunogenomic Studies (8), Cell lines (3), Adrenal Gland (2), Ampulla of Vater (1), Biliary Tract (15), Bladder/Urinary Tract (21), Bone (2), Bowel (19), Breast (25), CNS/Brain (25), Cervix (3), Esophagus/Stomach (20), and Eya (5). On the right, there is a 'Quick select:' section with two buttons: 'TCGA PanCancer Atlas Studies' and 'Curated set of non-redundant studies'. Below this, there is a text box stating 'Looking for AACR Project GENIE, the largest public clinicogenomic cancer data'. The 'PanCancer Studies' section lists several studies with checkboxes: MSK-IMPACT Clinical Sequencing Cohort (MSK, Nat Med 2017), Metastatic Solid Cancers (UMich, Nature 2017), MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018), SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018), TMB and Immunotherapy (MSK, Nat Genet 2019), Tumors with TRK fusions (MSK, Clin Cancer Res 2020), Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020), China Pan-cancer (Origimed, Nature 2022), Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020), and MSK MetTropism (MSK, Cell 2021). The 'Pediatric Cancer Studies' section lists several studies with checkboxes: Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019), Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018), Pediatric Rhabdoid Tumor (TARGET, 2018), Pediatric Wilms' Tumor (TARGET, 2018), Pediatric Acute Myeloid Leukemia (TARGET, 2018), Pediatric Neuroblastoma (TARGET, 2018), Pediatric Pan-Cancer (DKFZ, Nature 2017), and Pediatric Pan-cancer (Columbia U, Genome Med 2016).

cBioPortalは、がんゲノムデータを視覚化および解釈するためのオンラインのツールおよびリソースである：

- **多様ながんゲノムデータの統合：** cBioPortalは、TCGAなどがんゲノム研究の遺伝子発現、変異、コピー数異常、臨床データをキュレートして集積
- **直感的なデータ視覚化：** がんゲノムデータを直感的・効果的に視覚化でき、複雑ながんゲノムプロファイルを理解しやすくなる
- **変異とその臨床意義の解釈：** 個々のがんゲノム変異やパターンの臨床的な意義を理解するためのツールを提供する
- **相関解析：** 相関解析を行う機能を提供し、特定の変異や発現パターンが他の遺伝子や臨床的な特徴とどのように関連しているかを探索できる
- **公開データセットのアクセス：** 多くの公共データベースからデータを引用し、ユーザーは研究のためにこれらにアクセスできる

TMB and Immunotherapyのデータを見る

cBioPortalトップ画面

analysis: 1 study selected (1661 samples) Deselect all Data type Search...

Quick select: TCGA PanCancer Atlas Studies Curated set of non-redundant studies Help

Looking for AACR Project GENIE, the largest public clinicogenomic cancer dataset? It's available here.

PanCancer Studies

- ☐ MSK-IMPACT Clinical Sequencing Cohort (MSK, Nat Med 2017) 10945 samples
- ☐ Metastatic Solid Cancers (UMich, Nature 2017) 500 samples
- ☐ MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018) 249 samples
- ☐ SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2019) 141 samples
- ☒ TMB and Immunotherapy (MSK, Nat Genet 2019) 1661 samples
- ☐ Tumors with TRK fusions (MSK, Clin Cancer Res 2020) 109 samples
- ☐ Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020) 24146 samples
- ☐ China Pan-cancer (Origimed, Nature 2022)
- ☐ Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)
- ☐ MSK MetTropism (MSK, Cell 2021)

①チェック

マウスオーバーするとデータの概要が見れる

出典へのリンク

Genomic and survival data from 1661 tumor-normal pairs from 1661 patients with various cancer types sequenced with the MSK-IMPACT assay

Pediatric Cancer Studies

- ☐ Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019)
- ☐ Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018)
- ☐ Pediatric Rhabdoid Tumor (TARGET, 2018)
- ☐ Pediatric Wilms' Tumor (TARGET, 2018)
- ☐ Pediatric Acute Myeloid Leukemia (TARGET, 2018)
- ☐ Pediatric Neuroblastoma (TARGET, 2018)
- ☐ Pediatric Pan-Cancer (DKFZ, Nature 2017)
- ☐ Pediatric Pan-cancer (Columbia U, Genome Med 2016)
- ☐ Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)
- ☐ Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)
- ☐ Pediatric Ewing Sarcoma (DFCI, Cancer Discov 2014)

1 study selected (1661 samples) Deselect all

Query By Gene OR Explore Selected Studies

②クリック

研究全体の特徴をみる

遺伝子を指定して情報を得る

TMB and Immunotherapyデータの出典論文

LETTERS

<https://doi.org/10.1038/s41588-018-0312-8>

nature
genetics

Tumor mutational load predicts survival after immunotherapy across multiple cancer types

Robert M. Samstein^{1,2,11}, Chung-Han Lee^{3,4,11}, Alexander N. Shoushtari^{3,4,11}, Matthew D. Hellmann^{3,4,11}, Ronglai Shen⁵, Yelena Y. Janjigian^{3,4}, David A. Barron^{1,2}, Ahmet Zehir⁶, Emmet J. Jordan², Antonio Omuro⁷, Thomas J. Kaley⁷, Sviatoslav M. Kendal^{2,8}, Robert J. Motzer^{3,4}, A. Ari Hakimi⁹, Martin H. Voss^{3,4}, Paul Russo⁹, Jonathan Rosenberg^{3,4}, Gopa Iyer^{3,4}, Bernard H. Bochner⁹, Dean F. Bajorin^{3,4}, Hikmat A. Al-Ahmadie⁶, Jamie E. Chaff^{3,4}, Charles M. Rudin^{3,4}, Gregory J. Riely^{3,4}, Shrujal Baxi^{3,4}, Alan L. Ho^{3,4}, Richard J. Wong⁹, David G. Pfister^{3,4}, Jedd D. Wolchok^{3,4}, Christopher A. Barker¹, Philip H. Gutin⁹, Cameron W. Brennan⁹, Viviane Tabar⁹, Ingo K. Mellinghoff⁸, Lisa M. DeAngelis⁸, Charlotte E. Ariyan⁹, Nancy Lee¹, William D. Tap^{3,4}, Mrinal M. Gounder^{3,4}, Sandra P. D'Angelo^{3,4}, Leonard Saltz^{3,4}, Zsolt K. Stadler^{3,4}, Howard I. Scher^{3,4}, Jose Baselga^{3,4}, Pedram Razavi^{3,4}, Christopher A. Klebanoff^{3,4}, Rona Yaeger^{3,4}, Neil H. Segal^{3,4}, Geoffrey Y. Ku^{3,4}, Ronald P. DeMatteo⁹, Marc Ladanyi^{2,6}, Naiyer A. Rizvi¹⁰, Michael F. Berger^{3,6}, Nadeem Riaz^{1,2,8,12}, David B. Solit^{2,3,12*}, Timothy A. Chan^{1,2,8,12*} and Luc G. T. Morris^{2,8,9,12*}

Immune checkpoint inhibitor (ICI) treatments benefit some patients with metastatic cancers, but predictive biomarkers are needed. Findings in selected cancer types suggest that tumor mutational burden (TMB) may predict clinical response to ICI. To examine this association more broadly, we analyzed the clinical and genomic data of 1,662 advanced cancer patients treated with ICI, and 5,371 non-ICI-treated patients, whose tumors underwent targeted next-generation sequencing (MSK-IMPACT). Among all patients, higher somatic TMB (highest 20% in each histology) was associated with better overall survival. For most cancer histologies, an association between higher TMB and improved survival was observed. The TMB cutpoints associated with improved survival varied markedly between cancer types. These data indicate that TMB is associated with improved survival in patients receiving ICI across a wide variety of cancer types, but that there may not be one universal definition of high TMB.

In recent years, ICI therapy has revolutionized the treatment of patients with advanced-stage cancers. These agents include antibodies that target CTLA-4 or PD-1/PD-L1. Durable benefit, however,

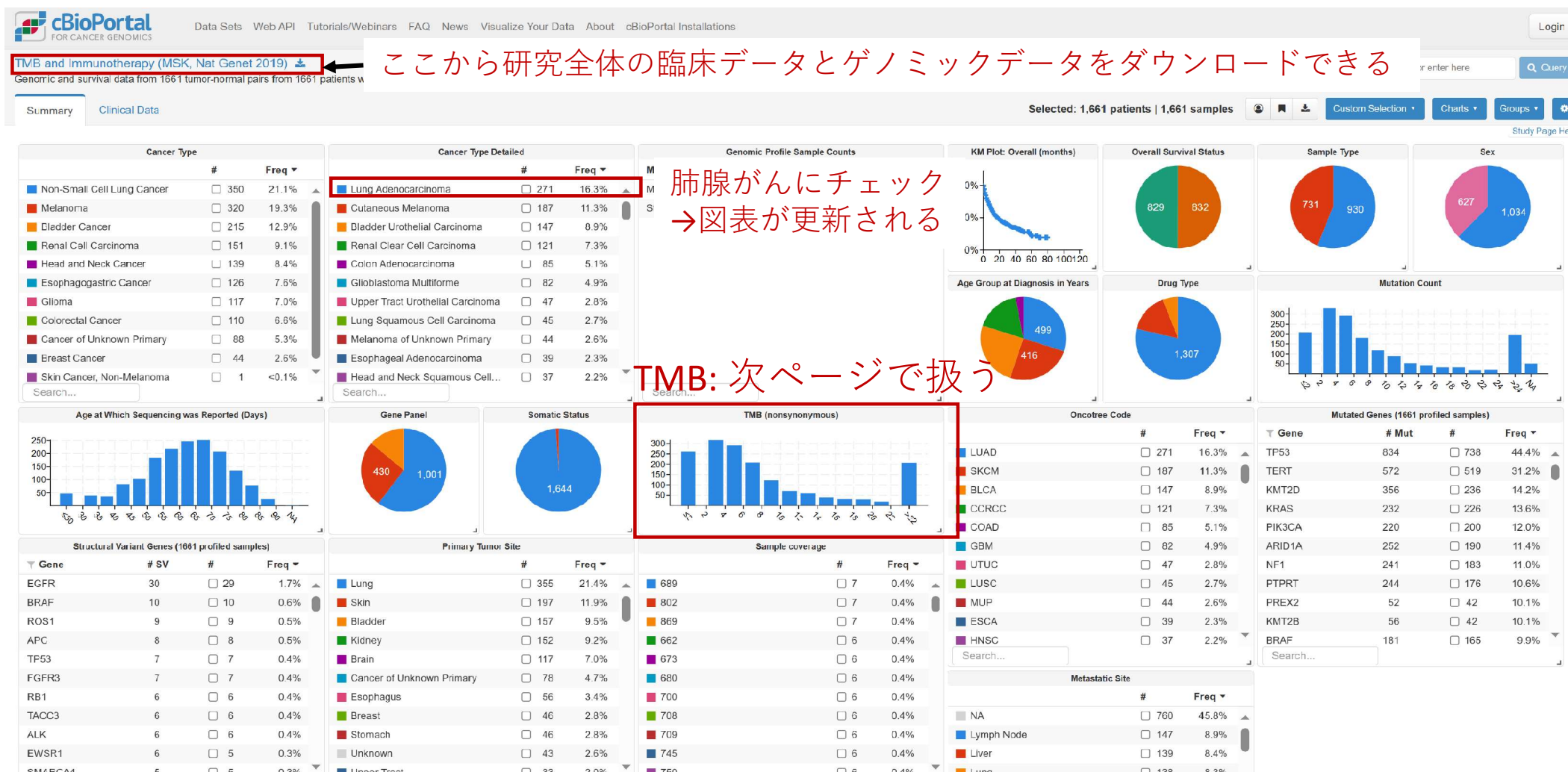
clonality, gene expression signatures and peripheral blood markers may correlate with clinical response¹. Additionally, an association between high mutational load and clinical benefit was observed in small cohorts of patients with melanoma treated with CTLA-4 blockade^{2,3}, and non-small cell lung cancer (NSCLC), patients with melanoma and bladder cancer treated with PD-1/PD-L1 inhibitors⁴⁻¹¹. However, it is unclear whether TMB is robustly predictive of clinical benefit across diverse human cancers, or outside of these specific clinical trial populations.

In previous studies, mutation load was determined by using whole-exome sequencing, which is not widely utilized in routine clinical care. Currently, the majority of precision oncology platforms use next-generation sequencing of targeted gene panels. At Memorial Sloan Kettering Cancer Center (MSK), as part of clinical care, patients undergo genomic profiling with the Food & Drug Administration (FDA)-authorized Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) assay¹². This test is performed in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory environment and identifies somatic exonic mutations in a predefined subset of 468 cancer-related genes

- 免疫チェックポイント阻害剤（ICI）治療は転移性がん患者の一部に有益で、バイオマーカーの予測が必要
- TMBがICI治療への臨床応答を予測する可能性があり、1,662人のICI治療患者および5,371人の非ICI治療患者のデータを分析
- 高いTMB（各組織について最上位20%）は全体生存期間の改善と関連し、高いTMBで生存期間の改善が観察された
- がん種によって有効なTMBの閾値が異なった

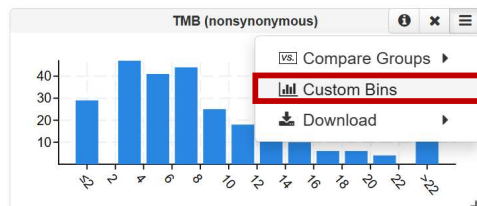
Explore selected study

がん種、背景因子、変異遺伝子に加えて、試験独自の変数(今回TMB)の分布を見れる

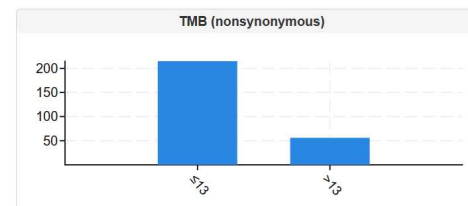


肺腺がんについて、TMBと予後の関係を見ることにする
⇒ TMB 上位20%超 vs. 20%以下 でKM-plotを作成する

① Custom Binsをクリック



⑤ $TMB \leq 13$ と >13 の分布に変わる



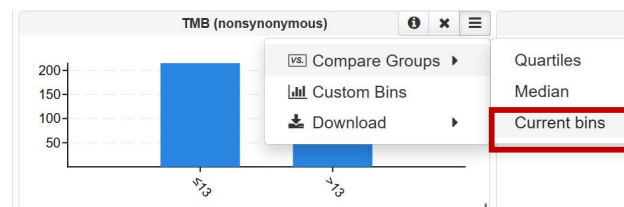
② ポップアップが出る

A dialog box titled "Custom Bins" with a close button (X) in the top right corner. It contains four radio button options: "Quartiles", "Median split", "Generate bins", and "Custom bins". The "Custom bins" option is selected. Below the options, there is a text input field with the placeholder text "Please specify bin boundaries of the x axis" and the value "13" entered. A red box highlights the input field. At the bottom right, there is an "Update" button, also highlighted with a red box.

③ 上位20%超とそれ以下に区切るような値である13を入力

④ クリック

⑥ Compare Groups → Current binsをクリック

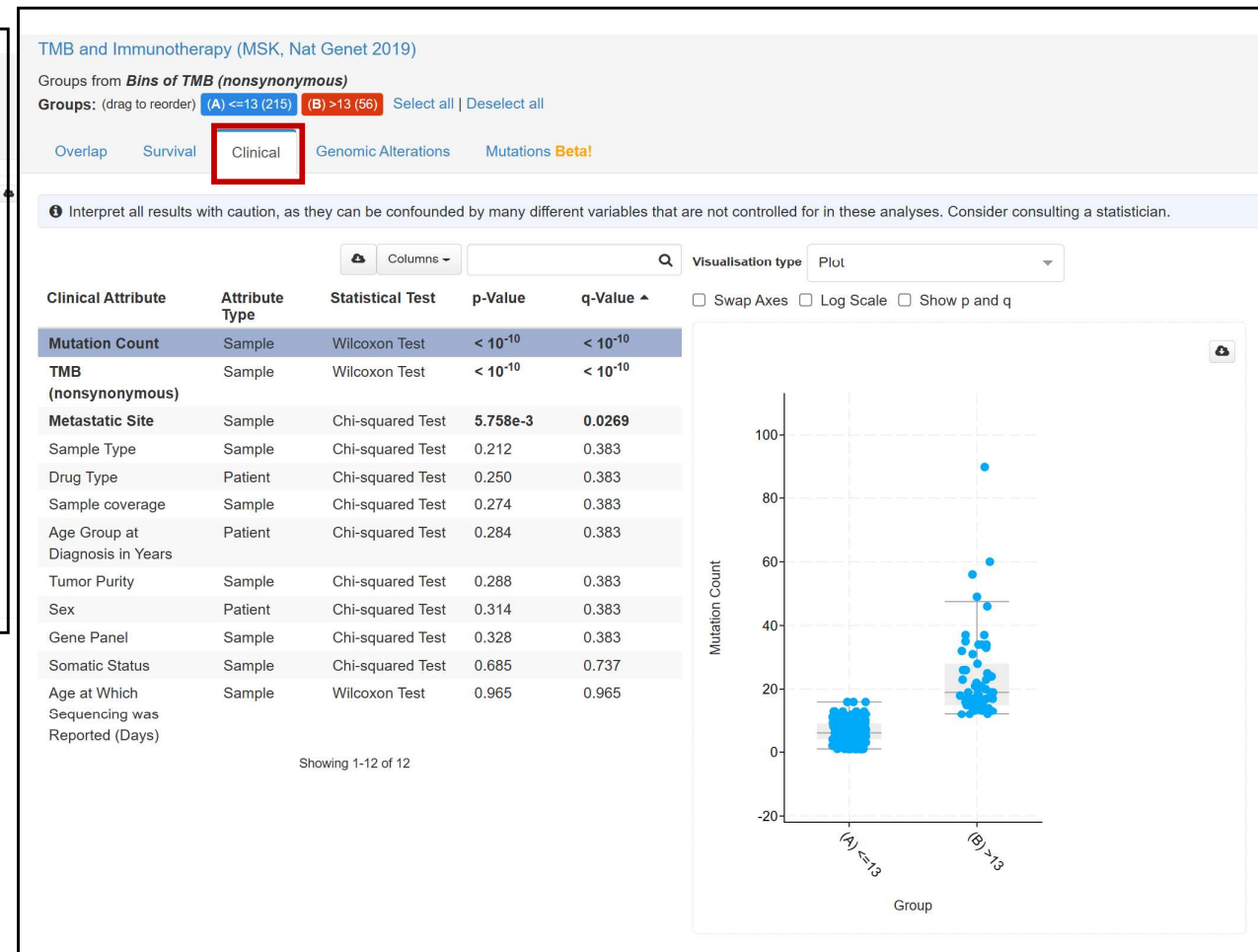


ケースの集合関係と背景の確認



群間のケースに重複がないことが分かる

※重複があった場合、重複を除いた生存時間解析がなされる



Metastatic Site以外の背景因子は、群間で偏りがないことが確認できた

※TMBで群分けしたので当然TMBは群間差あり

生存時間解析

TMB and Immunotherapy (MSK, Nat Genet 2019)

Groups from **Bins of TMB (nonsynonymous)**

Groups: (drag to reorder) (A) ≤13 (215) (B) >13 (56) Select all | Deselect all

Overlap Survival Clinical Genomic Alterations Mutations **Beta!**

Interpret all results with caution, as they can be confounded by many different variables that are not controlled for in these analyses. Consider consulting a statistician.

The log-rank test is used to test the null hypothesis that there is no difference between the groups in the probability of an event at any time point. Hazard ratios are derived from the log-rank test.

Overall

Overall patient survival status.

☒ Calculate hazard ratios

≤13

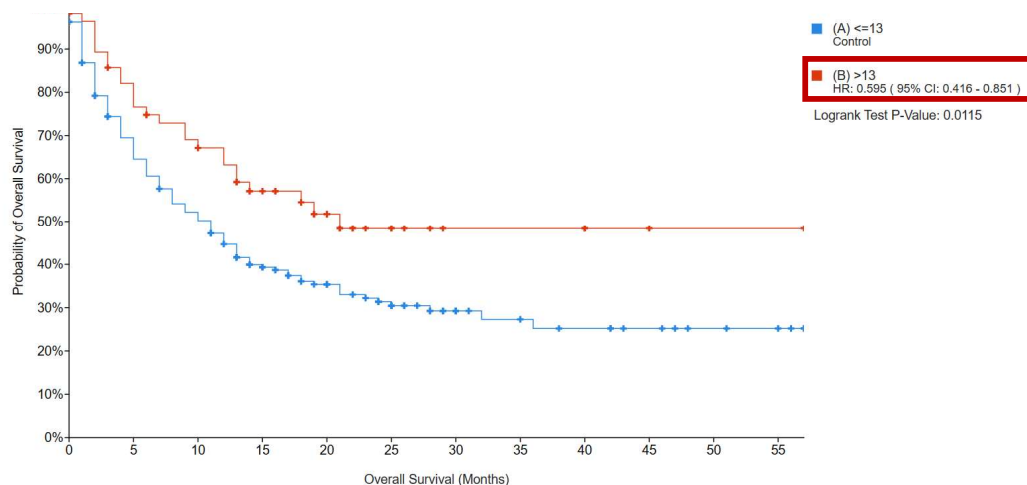
?

☐ Add landmarks

Add landmark values

?

チェックするとハザード比が表示される



Number at risk (n)

(A) ≤13 215 141 105 67 50 34 19 14 11 8 5 4
(B) >13 56 45 36 25 17 9 4 4 4 2 1 1

Hazard ratio

hazard ratio (±95% confidence interval)		(A) ≤13	(B) >13
(A) ≤13		1.000 (0.781-1.281)	0.595 (0.416-0.851)
(B) >13		1.681 (1.176-2.404)	1.000 (0.643-1.555)

Survival plot summary

	Number of Cases, Total	Number of Events	Median Months Overall (95% CI)
(A) ≤13	215	139	11.00 (8.00 - 13.00)
(B) >13	56	26	21.00 (13.00 - NA)

営利目的でのご利用はご遠慮ください

Nat. Genet. 2019 Fig2

LETTERS

NATURE GENETICS

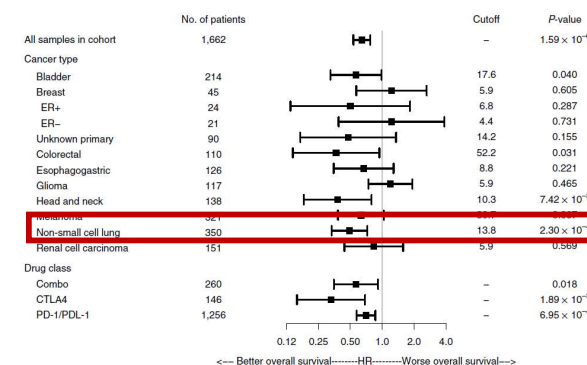


Fig. 2 | Effect of nonsynonymous mutational load on overall survival after ICI treatment, by cancer subtype and drug class. Forest plot for all patients in the identified cohort or individual cancer subtypes. Indicated are the number of patients and HR comparing overall survival after ICI in patients in the highest twentieth-percentile TMB within each histology. Bars represent the 95% CI. The cutoff defining the top 20% of normalized mutational burden from MSK-IMPACT for each cancer type is shown, as well as the two-sided log-rank *P* value for the comparison of high and low mutational burden survival curves. ER, estrogen receptor. All cancer types in analysis are displayed.

元の論文のNon-small cell LungのHRと大きな違いはないことが確認できる

一方の群でより変異がエンリッチしている遺伝子を調べる



※ 今回はTMBで群分けをしていて、(B)TMB>13群でエンリッチしている遺伝子が多くなるのは当然であり解釈には注意が必要

(A)TMB<=13群でエンリッチしている遺伝子を探す

(A)TMB<=13群でエンリッチしている遺伝子だけを表示する

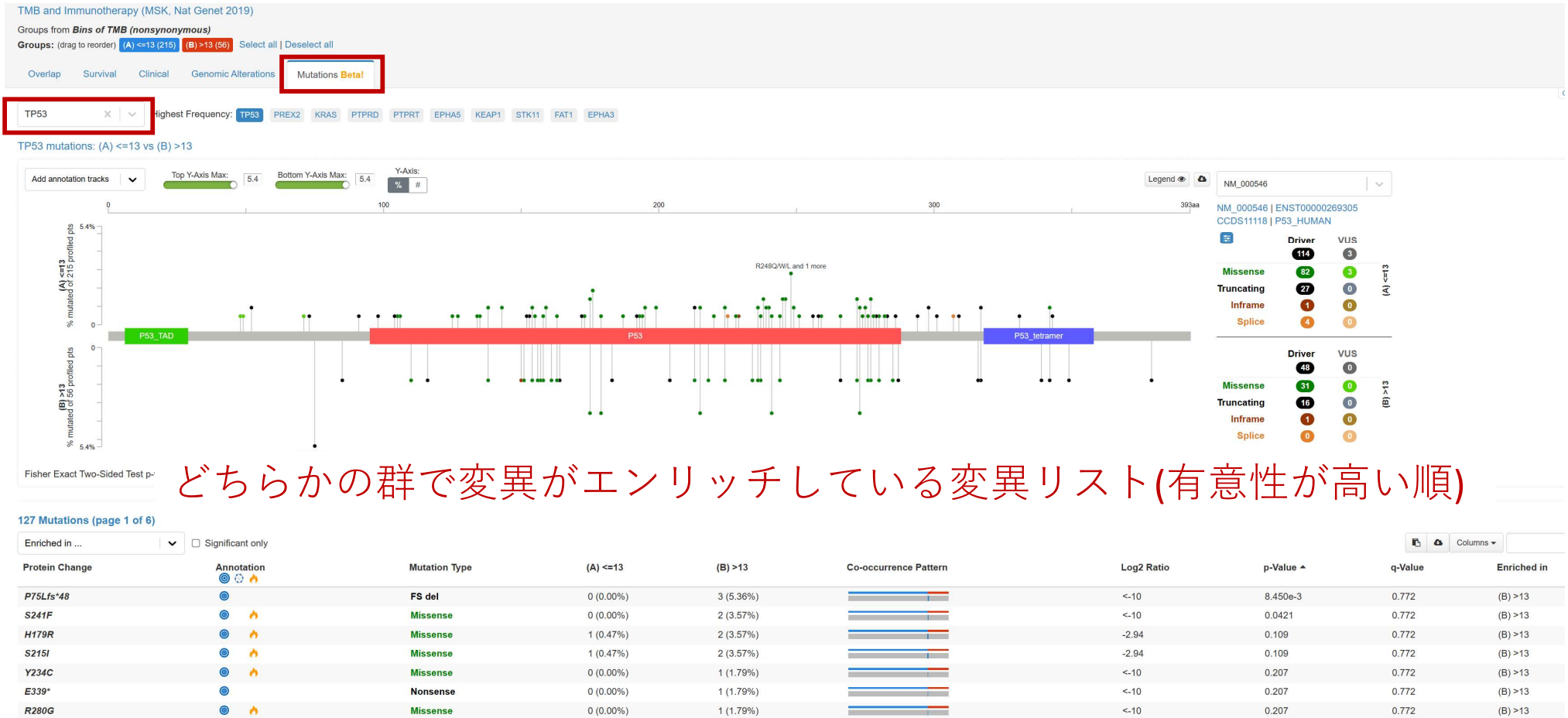


The screenshot shows the 'Genomic Alterations' interface. A dropdown menu 'Enriched in ...' is open, showing two options: '(A) <=13' (selected) and '(B) >13'. Below the dropdown, a table lists genomic alterations. The table has columns: Gene, Cytoband, (A) <=13, (B) >13, Co-occurrence Pattern, Log2 Ratio, p-Value, q-Value, and Enriched in. The 'Enriched in' column shows '(A) <=13' for all listed genes.


Gene	Cytoband	(A) <=13	(B) >13	Co-occurrence Pattern	Log2 Ratio	p-Value	q-Value	Enriched in
EGFR	7p11.2	38 (17.67%)	6 (10.71%)		0.72	0.308	0.477	(A) <=13
MAX	14q23.3	6 (2.79%)	0 (0.00%)		>10	0.351	0.509	(A) <=13
CD74	5q33.1	4 (1.86%)	0 (0.00%)		>10	0.584	0.742	(A) <=13
MEN1	11q13	4 (1.86%)	0 (0.00%)		>10	0.584	0.742	(A) <=13
PMS2	7p22.1	4 (1.86%)	0 (0.00%)		>10	0.584	0.742	(A) <=13
PTEN	10q23.31	5 (2.33%)	1 (1.79%)		0.38	1.00	1.00	(A) <=13
CTCF	16q22.1	4 (1.86%)	1 (1.79%)		0.06	1.00	1.00	(A) <=13
GATA1	Xp11.23	4 (1.86%)	1 (1.79%)		0.06	1.00	1.00	(A) <=13
AGAP3	7q36.1	1 (0.47%)	0 (0.00%)		>10	1.00	1.00	(A) <=13
AGK	7q34	1 (0.47%)	0 (0.00%)		>10	1.00	1.00	(A) <=13

(A)TMB<=13群でエンリッチしている遺伝子はなかった

特定の遺伝子(TP53)での変異のプロット



全ての研究のダウンロードページ



Data Sets

Web API

Tutorials/Webinars

FAQ

News


Visualize Your Data

About

cBioPortal Installations

Query

Quick Search **Beta!**



[Data Sets](#) [Web API](#) [Tutorials/Webinars](#) [FAQ](#) [News](#) [Visualize Your Data](#) [About](#) [cBioPortal Installations](#)

































Login

Datasets

The table below lists the number of available samples per cancer study and data type. It also provides links to download the data for each study. For alternative ways of downloading, see the [Download Documentation](#).

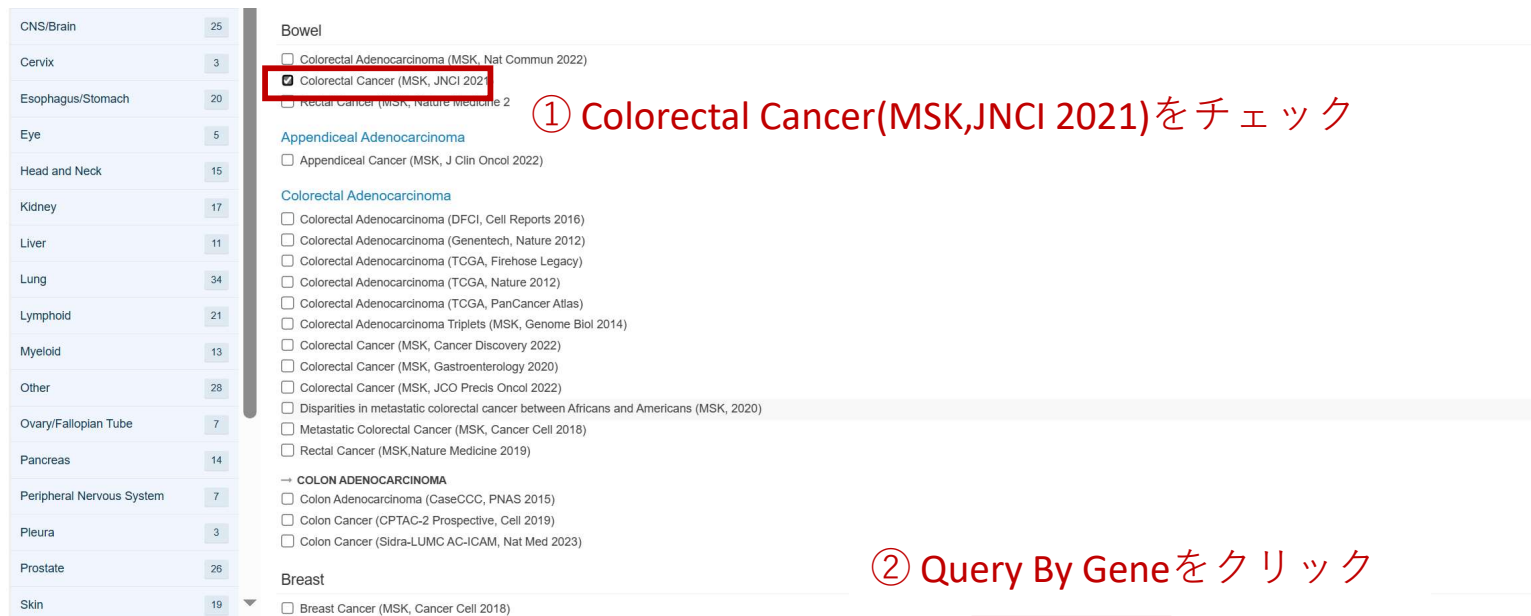
Q

Columns ▾

Name ▾	Reference	All	Mutations	CNA	RNA-Seq
Acinar Cell Carcinoma of the Pancreas (JHU, J Pathol 2014)	 Jial et al. J Pathol 2014	23	23	0	0
Acral Melanoma (TGEN, Genome Res 2017)	 Liang et al. Genome Res 2017	38	38	38	36
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)	 Andersson et al. Nat Genet 2015	93	93	0	0
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)	 Zhang et al. Nat Genet 2016	73	73	0	0
Acute Myeloid Leukemia (OHSU, Cancer Cell 2022)	 Bottomly et al. Cancer Cell 2022	942	903	0	698
Acute Myeloid Leukemia (OHSU, Nature 2018)	 Tyner et al. Nature 2018	672	622	0	451
Acute Myeloid Leukemia (TCGA, Firehose Legacy)		200	197	191	173
Acute Myeloid Leukemia (TCGA, NEJM 2013)		200	200	191	173
Acute Myeloid Leukemia (TCGA, PanCancer Atlas)	 TCGA, Cell 2018	200	200	191	173
Acute myeloid leukemia or myelodysplastic syndromes (WashU, 2016)	 Welch et al. N Engl J Med. 2016	136	136	0	0
Adenoid Cystic Carcinoma (FMI, Am J Surg Pathl. 2014)	 Ross et al. Am J Surg Pathl 2014	28	28	28	0
Adenoid Cystic Carcinoma (JHU, Cancer Prev Res 2016)	 Rettig et al. Cancer Prev Res 2016	25	25	0	0
Adenoid Cystic Carcinoma (MDA, Clin Cancer Res 2015)	 Mitani et al. Clin Cancer Res 2015	102	65	0	0
Adenoid Cystic Carcinoma (MGH, Nat Gen 2016)	 Drier et al. Nature Genetics 2016	10	10	0	0
Adenoid Cystic Carcinoma (MSK, Nat Genet 2013)	 Ho et al. Nat Genet 2013	60	60	60	0
Adenoid Cystic Carcinoma (Sanger/MDA, JCI 2013)	 Stephens et al. JCI 2013	24	24	0	0
Adenoid Cystic Carcinoma of the Breast (MSK, J Pathol. 2015)	 Martelotto et al. J Pathol 2015	12	12	12	0
Adenoid Cystic Carcinoma Project (J Clin Invest 2019)	 Allen et al. J Clin Invest 2019	1049	1049	928	0
Adrenocortical Carcinoma (TCGA, Firehose Legacy)		92	90	90	79
Adrenocortical Carcinoma (TCGA, PanCancer Atlas)		92	91	89	78
Adult Soft Tissue Sarcomas (TCGA, Cell 2017)	 TCGA, Cell 2017	206	206	206	206
Ampullary Carcinoma (Baylor College of Medicine, Cell Reports 2016)	 Gingras et al. Cell Rep 2016	160	160	0	0
Anaplastic Oligodendroglioma and Anaplastic Oligoastrocytoma (MSK, Neuro Oncol 2017)	 Thomas et al. Neuro Oncol 2017	22	22	22	0
Appendiceal Cancer (MSK, J Clin Oncol 2022)	 Michael B et al. J Clin Oncol 2022	273	273	273	0
Basal Cell Carcinoma (UNIGE, Nat Genet 2016)	 Bonilla et al. Nat Genet 2016	293	293	0	0
Bladder Cancer (Columbia University/MSK, Cell 2018)	 Lee, Suk Hyung et al. Cell 2018	130	130	130	0
Bladder Cancer (MSK, Cell Reports 2022)	 Clinton et al. Cell Reports 2022	1659	1659	1659	0
Bladder Cancer (MSK, Clin Cancer Research 2023)	 Guercio et al. Clin Cancer Res. 2023	526	526	526	0
Bladder Cancer (MSK, Eur Urol 2014)	 Kim et al. Eur Urol 2015	109	109	109	0
Bladder Cancer (MSK, J Clin Onco 2013)	 Iyer et al. J Clin Oncol 2013	97	97	97	0
Bladder Cancer (MSK, Nat Genet 2016)	 Al-Ahmadie et al. Nat Genet 2016	34	34	33	0
Bladder Cancer (MSK/TCGA, 2020)	 Pietzak et al. Eur Urol 2019	476	474	442	296

特定遺伝子を指定して情報検索(Query By Gene)

Colorectal Cancer(MSK,JNCI 2021)を対象とする



Bowel

- ☐ Colorectal Adenocarcinoma (MSK, Nat Commun 2022)
- ☒ Colorectal Cancer (MSK, JNCI 2021)
- ☐ Rectal Cancer (MSK, Nature Medicine 2)

Appendiceal Adenocarcinoma

- ☐ Appendiceal Cancer (MSK, J Clin Oncol 2022)

Colorectal Adenocarcinoma

- ☐ Colorectal Adenocarcinoma (DFCI, Cell Reports 2016)
- ☐ Colorectal Adenocarcinoma (Genentech, Nature 2012)
- ☐ Colorectal Adenocarcinoma (TCGA, Firehose Legacy)
- ☐ Colorectal Adenocarcinoma (TCGA, Nature 2012)
- ☐ Colorectal Adenocarcinoma (TCGA, PanCancer Atlas)
- ☐ Colorectal Adenocarcinoma Triplets (MSK, Genome Biol 2014)
- ☐ Colorectal Cancer (MSK, Cancer Discovery 2022)
- ☐ Colorectal Cancer (MSK, Gastroenterology 2020)
- ☐ Colorectal Cancer (MSK, JCO Precis Oncol 2022)
- ☐ Disparities in metastatic colorectal cancer between Africans and Americans (MSK, 2020)
- ☐ Metastatic Colorectal Cancer (MSK, Cancer Cell 2018)
- ☐ Rectal Cancer (MSK, Nature Medicine 2019)

→ COLON ADENOCARCINOMA

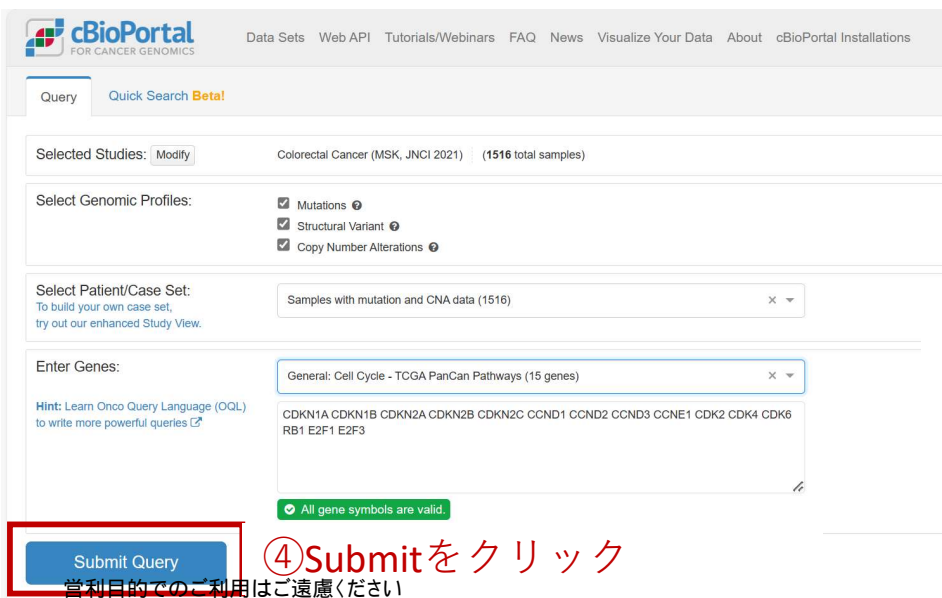
- ☐ Colon Adenocarcinoma (CaseCCC, PNAS 2015)
- ☐ Colon Cancer (CPTAC-2 Prospective, Cell 2019)
- ☐ Colon Cancer (Sidra-LUMC AC-ICAM, Nat Med 2023)

Breast

- ☐ Breast Cancer (MSK, Cancer Cell 2018)

① Colorectal Cancer(MSK,JNCI 2021)をチェック

② Query By Geneをクリック



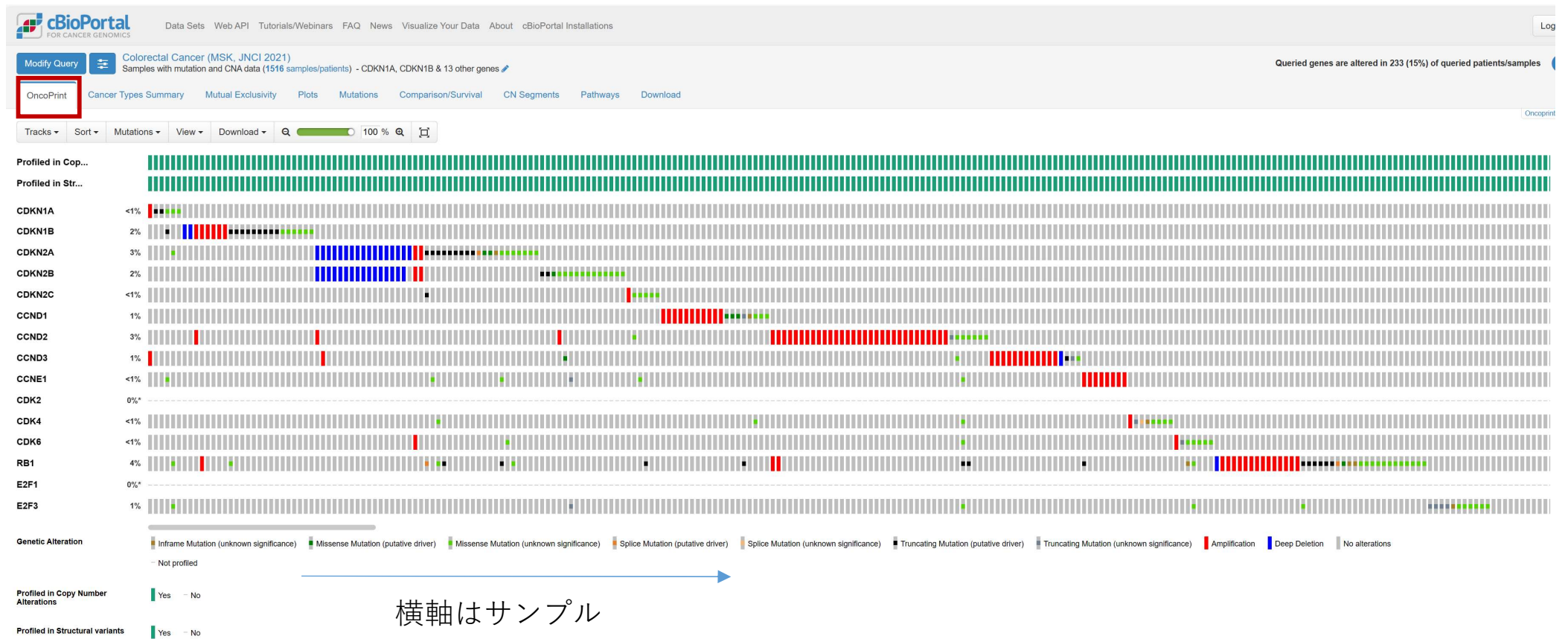
Submit Query

④ Submitをクリック



営利目的でのご利用はご遠慮ください

③ General: Cell Cycle- TCGA PanCan Pathways(15 genes)を選択⇒15遺伝子がボックスにセットされる

15遺伝子のOncoPrint



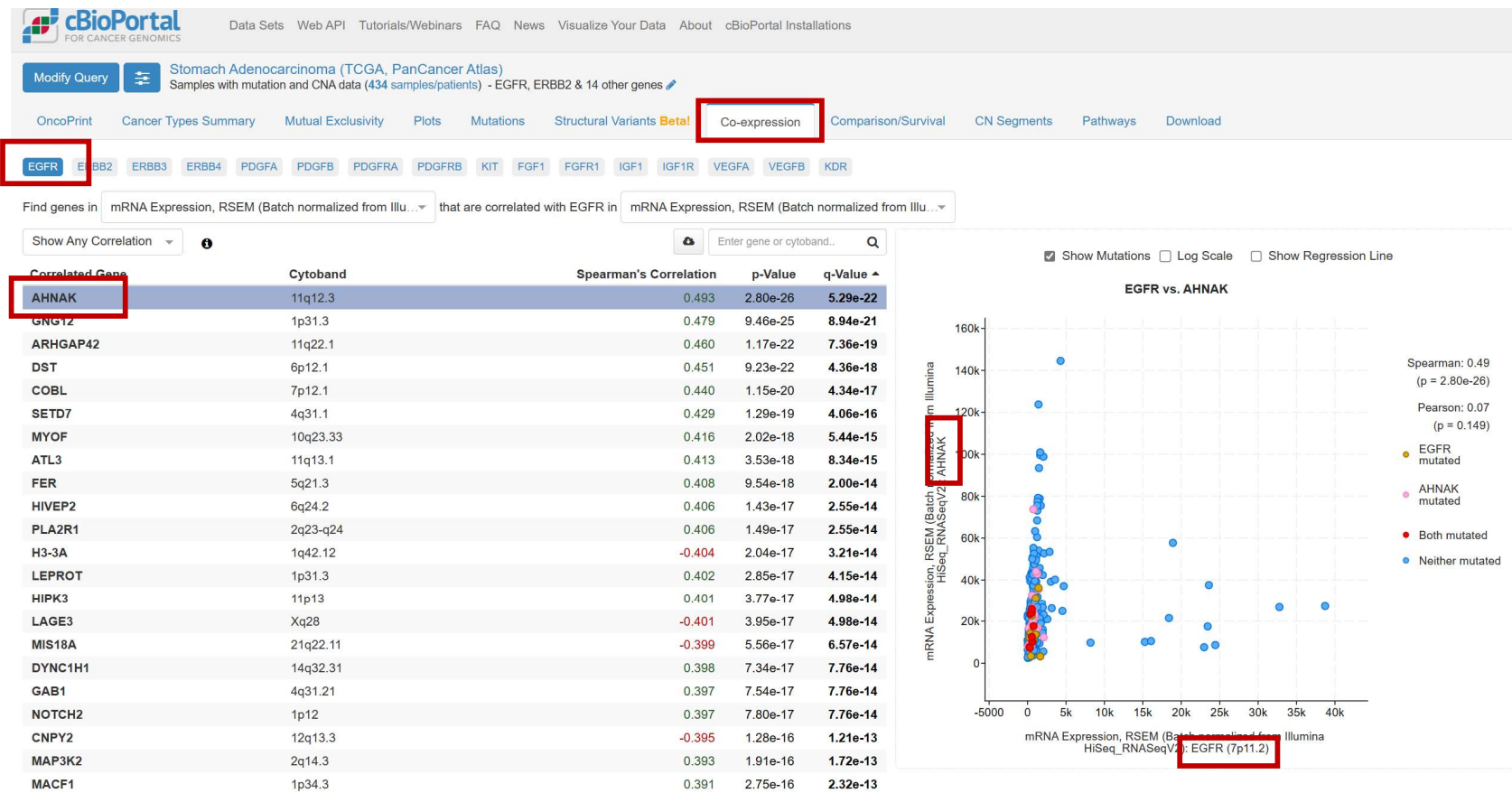
15遺伝子の各ペアのMutual Exclusivity(相互排他性)

<div>  Data Sets Web API Tutorials/Webinars FAQ News Visualize Your Data About cBioPortal Installations </div>									
P and q-values on this page have changed recently. The 1-sided Fisher exact test is now a 2-sided test (Read more).									
<div> <div> Modify Query  </div> <div> Colorectal Cancer (MSK, JNCI 2021) Samples with mutation and CNA data (1516 samples/patients) - CDKN1A, CDKN1B & 13 other genes </div> </div> <div> OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Comparison/Survival CN Segments Pathways Download </div> <div> Queried genes are altered in 23 </div>									
CDK2 and E2F1 are not profiled in any queried samples and therefore are excluded from this analysis.									
The analysis tested 78 pairs between the 13 tracks in the OncoPrint.									
<input checked="" type="checkbox"/> Mutual exclusivity <input checked="" type="checkbox"/> Co-occurrence <input type="checkbox"/> Significant only									
A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value [▲]	Tendency
CDKN2A	CDKN2B	1458	22	15	18	>3	<0.001	<0.001	Co-occurrence
RB1	E2F3	1447	50	12	4	>3	0.002	0.061	Co-occurrence
CDKN2A	RB1	1425	34	48	6	2.389	0.002	0.061	Co-occurrence
CDK6	RB1	1452	7	51	3	>3	0.004	0.070	Co-occurrence
CDK6	E2F3	1489	8	14	2	>3	0.004	0.070	Co-occurrence
CCNE1	E2F3	1485	12	14	2	>3	0.009	0.115	Co-occurrence
CCNE1	RB1	1448	11	51	3	2.953	0.012	0.132	Co-occurrence
CDKN2C	RB1	1454	5	52	2	>3	0.023	0.228	Co-occurrence
CDKN2A	CDK6	1485	38	8	2	>3	0.027	0.232	Co-occurrence
CDKN2A	CCNE1	1461	38	12	2	2.680	0.051	0.324	Co-occurrence
CDKN1A	CCNE1	1494	5	13	1	>3	0.054	0.324	Co-occurrence
CDK4	RB1	1450	9	52	2	2.631	0.056	0.324	Co-occurrence
CCND2	RB1	1421	38	50	4	1.581	0.059	0.324	Co-occurrence
CDKN1A	E2F3	1492	5	15	1	>3	0.062	0.324	Co-occurrence

CDKN2AとCDKN2Bはq-value<0.001でCo-occurrence (共起)

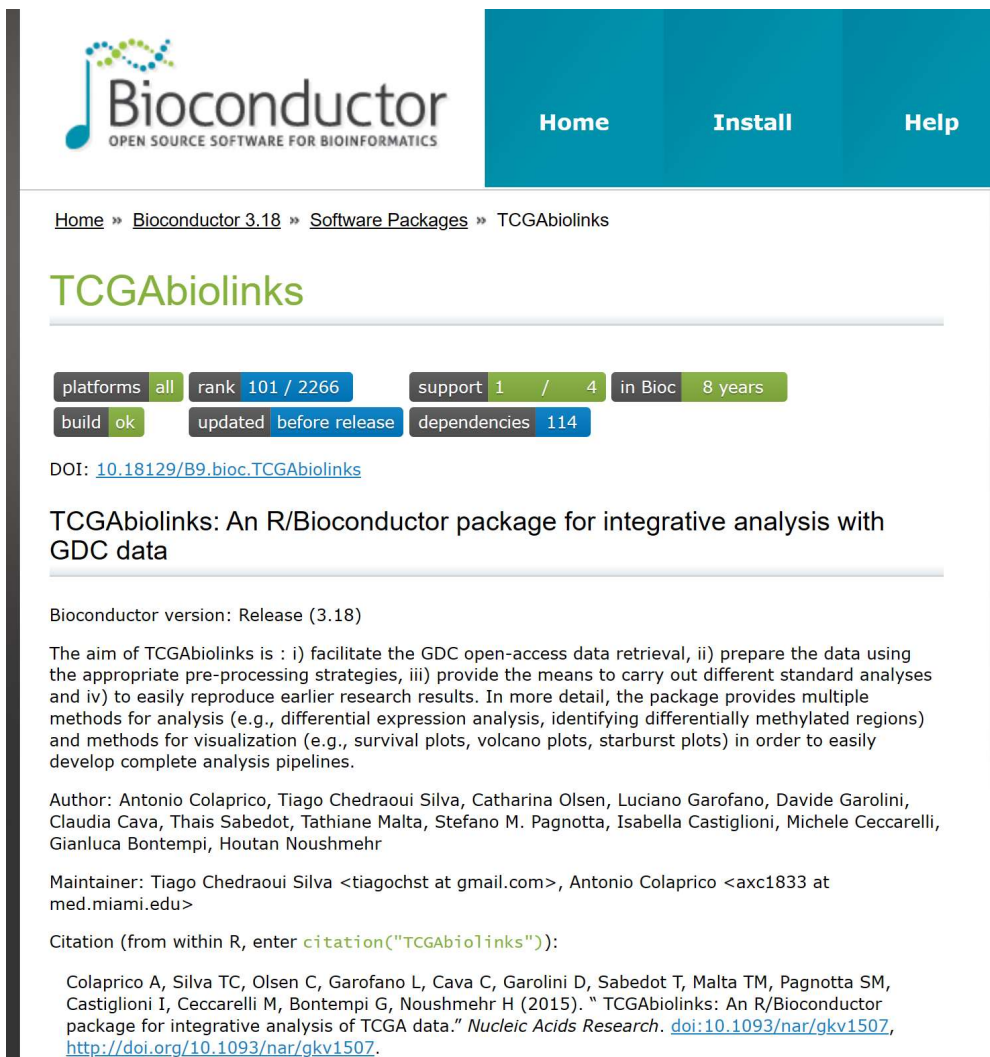
※ 発現データがある場合はCo-expressionも調べられる

Co-expression: 特定の遺伝子を固定して、それとその他の遺伝子の間での遺伝子発現の相関をみる



TCGAbiolinks

- TCGAbiolinkは、がんゲノムデータ解析を行うためのR言語のパッケージである
このパッケージは、TCGAなどGDCのデータから得られたがんゲノムデータへのアクセスと解析を容易にする
- 遺伝子発現、変異、臨床データなどを含む多様なデータにアクセスし、これらのデータを組み合わせて独自の解析を行うことができる



The screenshot shows the Bioconductor website for the TCGAbiolinks package. At the top, the Bioconductor logo is on the left, and navigation links for Home, Install, and Help are on the right. Below the navigation bar, the breadcrumb trail reads: Home » Bioconductor 3.18 » Software Packages » TCGAbiolinks. The package name 'TCGAbiolinks' is displayed in green. A series of status badges follows: 'platforms all', 'rank 101 / 2266', 'support 1 / 4', 'in Bioc 8 years', 'build ok', 'updated before release', and 'dependencies 114'. The DOI is listed as 10.18129/B9.bioc.TCGAbiolinks. The package description states: 'TCGAbiolinks: An R/Bioconductor package for integrative analysis with GDC data'. The Bioconductor version is noted as Release (3.18). The aim of the package is detailed: 'The aim of TCGAbiolinks is : i) facilitate the GDC open-access data retrieval, ii) prepare the data using the appropriate pre-processing strategies, iii) provide the means to carry out different standard analyses and iv) to easily reproduce earlier research results. In more detail, the package provides multiple methods for analysis (e.g., differential expression analysis, identifying differentially methylated regions) and methods for visualization (e.g., survival plots, volcano plots, starburst plots) in order to easily develop complete analysis pipelines.' The authors listed are Antonio Colaprico, Tiago Chedraoui Silva, Catharina Olsen, Luciano Garofano, Davide Garolini, Claudia Cava, Thais Sabedot, Tathiane Malta, Stefano M. Pagnotta, Isabella Castiglioni, Michele Ceccarelli, and Gianluca Bontempi. The maintainer is Tiago Chedraoui Silva. A citation instruction is provided: 'Citation (from within R, enter `citation("TCGAbiolinks")`):'. The full citation is: 'Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot T, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H (2015). "TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data." *Nucleic Acids Research*. doi:10.1093/nar/gkv1507, http://doi.org/10.1093/nar/gkv1507.'

TCGAbiolinks

TCGAbiolinksはGDCデータをRで利用可能な形にセットアップしてくれる

- GDCデータのダウンロードを容易にしてくれる
 - ◆ 必要なデータの検索
 - ◆ ダウンロード
- 取得したGDCプロジェクトデータをRオブジェクトやデータフレームに変換し、Rで扱いやすくしてくれる
- heatmapや生存時間解析など、統計解析や図の作成では、必ずしもTCGAbiolinksを使用する必要はなく、他のRパッケージを使った方がよいことが多い

※ 以降ではR言語(>= ver. 4.0)やTCGAbiolinksを含む
必要なRパッケージのインストールを前提としています

発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)
```

```
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認
```

```
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
```

```
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling")
```

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定しでさらに絞る.
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling",  
                  data.type = "Gene Expression Quantification")
```

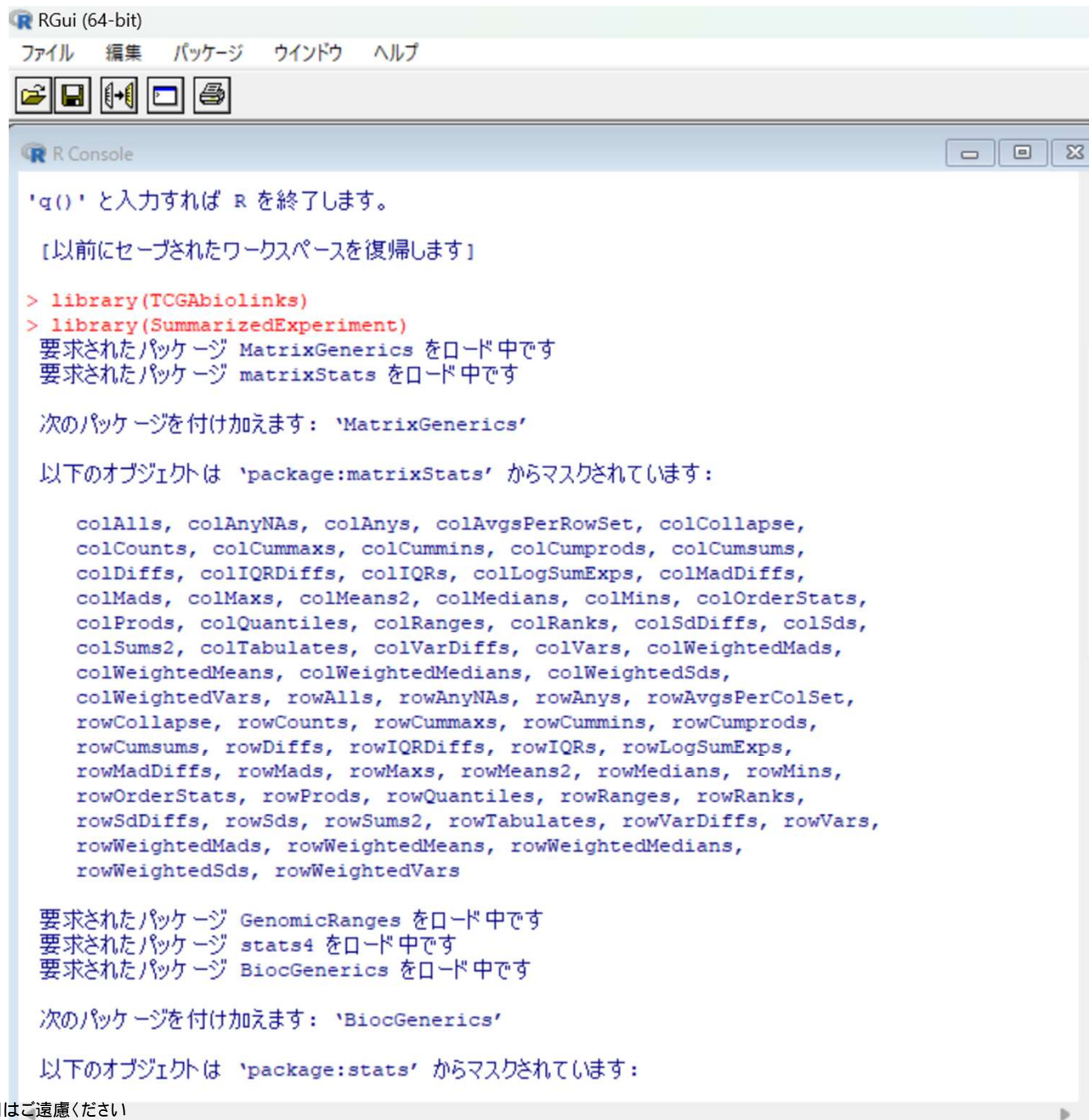
```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

データを検索する

⇒ ・ `getProjectSummary()`
・ `GDCquery()`

使用するパッケージの読み込み



```
RGui (64-bit)
ファイル 編集 パッケージ ウィンドウ ヘルプ

R Console

'q()' と入力すれば R を終了します。

[以前にセーブされたワークスペースを復帰します]

> library(TCGAbiolinks)
> library(SummarizedExperiment)
要求されたパッケージ MatrixGenerics をロード中です
要求されたパッケージ matrixStats をロード中です

次のパッケージを付け加えます: 'MatrixGenerics'

以下のオブジェクトは 'package:matrixStats' からマスクされています:

colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
colWeightedMeans, colWeightedMedians, colWeightedSds,
colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
rowWeightedSds, rowWeightedVars

要求されたパッケージ GenomicRanges をロード中です
要求されたパッケージ stats4 をロード中です
要求されたパッケージ BiocGenerics をロード中です

次のパッケージを付け加えます: 'BiocGenerics'

以下のオブジェクトは 'package:stats' からマスクされています:
```


発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る
query <- GDCquery(project = "TCGA-UVM",
                  data.category = "Transcriptome Profiling")
```

データを検索する

⇒ ・ `getProjectSummary()`
・ `GDCquery()`

```
output_query <- getResults(query)
head(output_query, 20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定してさらに絞る.
query <- GDCquery(project = "TCGA-UVM",
                  data.category = "Transcriptome Profiling",
                  data.type = "Gene Expression Quantification")
```

```
output_query <- getResults(query)
head(output_query, 20)
```

全 GDC プロジェクトを確認

```
> # 全 GDCプロジェクトを確認
> getGDCprojects()
```

	id	primary_site	dbgap_accession_number
1	TCGA-BRCA	Breast	<NA>
2	CPTAC-3	Kidney,	phs001287
3	TCGA-STAD	Stomach	<NA>
4	TCGA-LUAD	Bronchus....	<NA>
5	EXCEPTIONAL_RESPONDERS-ER	Other an....	<NA>
6	CGCI-HTMCP-LC	Bronchus....	phs000530
7	CPTAC-2	Other an....	phs000892
8	CMI-MBC	Breast	phs001709
9	TARGET-ALL-P3	Unknown,	<NA>
10	TARGET-ALL-P2	Hematopo....	<NA>
11	OHSU-CNL	Hematopo....	<NA>
12	REBC-THYR	Thyroid	<NA>
13	TARGET-ALL-P1	Hematopo....	<NA>
14	MMRF-COMMPASS	Hematopo....	<NA>
15	TARGET-CCSK	Kidney	<NA>
16	ORGANOID-PANCREATIC	Pancreas	<NA>
17	NCICCR-DLBCL	Lymph nodes	<NA>
18	TARGET-NBL	Meninges....	<NA>
19	TCGA-CHOL	Other an....	<NA>
20	TARGET-OS	Bones, j....	<NA>
21	TARGET-AML	Unknown,	<NA>
22	TARGET-RT	Lip, Liv....	<NA>
23	TARGET-WT	Kidney	<NA>
24	TCGA-SARC	Meninges....	<NA>
25	TCGA-PCPG	Other an....	<NA>
26	TCGA-COAD	Rectosig....	<NA>
27	TCGA-ACC	Adrenal	<NA>
28	WCDT-MCRPC	Prostate....	<NA>
29	TCGA-HCES	Corpus u...	<NA>
30	TCGA-KICH	Adenomas....	<NA>
31	TCGA-THYM	Thymic E....	<NA>
32	VAREPOP-APOLLO	Squamous....	<NA>
33	TCGA-UCS	Complex	<NA>
34	TCGA-SKCM	Nevi and....	<NA>
35	TRIO-CRU	Not Appl....	<NA>
36	TCGA-HNSC	Squamous....	<NA>
37	TCGA-PAAD	Ductal a....	<NA>
38	TCGA-TGCT	Germ Cel....	<NA>
39	TCGA-CESC	Cystic,	<NA>
40	TCGA-ESCA	Cystic,	<NA>
41	TCGA-THCA	Epitheli....	<NA>
42	TCGA-LGG	Gliomas	<NA>
43	TCGA-LIHC	Adenomas....	<NA>
44	TCGA-PRAD	Ductal a....	<NA>
45	TCGA-READ	Cystic,	<NA>
46	MATCH-I	Gliomas,....	<NA>
47	MATCH-W	Gliomas,....	<NA>
48	MATCH-B	Transiti....	<NA>
49	MATCH-H	Adenomas....	<NA>
50	TCGA-OV	Not Repo....	<NA>
51	TCGA-UVM	Nevi and....	<NA>
52	MATCH-Z1A	Mesothel....	<NA>
53	MATCH-U	Mesothel....	<NA>
54	BEATAML1.0-COHORT	Myeloid	<NA>
55	TCGA-BLCA	Adenomas....	<NA>
56	TCGA-LUSC	Squamous....	<NA>
57	CGCI-BLGSP	Mature B....	<NA>
58	HCM1-CMDC	Soft Tis....	<NA>

発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)  
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認  
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例  
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る  
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling")
```

```
output_query <- getResults(query)  
head(output_query, 20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定しでさらに絞る.  
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling",  
                  data.type = "Gene Expression Quantification")
```

```
output_query <- getResults(query)  
head(output_query, 20)
```

データを検索する

⇒ ・ getProjectSummary()
・ GDCquery()

“TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認

```
> # “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
> getProjectSummary("TCGA-UVM")
$file_count
[1] 4549

$data_categories
  file_count case_count data_category
1      1376         80 Simple Nucleotide Variation
2       724         80 Sequencing Reads
3       320         80 Biospecimen
4       171         80 Clinical
5       970         80 Copy Number Variation
6       320         80 Transcriptome Profiling
7       240         80 DNA Methylation
8        12         12 Proteome Profiling
9        96         48 Somatic Structural Variation
10      320         80 Structural Variation

$case_count
[1] 80

$file_size
[1] 3.963955e+13
```

発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)
```

```
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認
```

```
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
```

```
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る  
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling")
```

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定しでさらに絞る.
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling",  
                  data.type = "Gene Expression Quantification")
```

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

データを検索する

⇒ ・ getProjectSummary()
・ GDCquery()

GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る

```
> # GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る
> query <- GDCquery(project = "TCGA-UVM",
+                   data.category = "Transcriptome Profiling")
-----
o GDCquery: Searching in GDC database
-----
Genome of reference: hg38
-----
oo Accessing GDC. This might take a while...
-----
ooo Project: TCGA-UVM
-----
oo Filtering results
-----
-----
oo Checking data
-----
ooo Checking if there are duplicated cases
Warning: There are more than one file for the same case. Please verify query result$
ooo Checking if there are results for the query
-----
o Preparing output
-----
```


発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)
```

```
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認
```

```
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
```

```
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling")
```

データを検索する

⇒ ・ `getProjectSummary()`
・ `GDCquery()`

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定しでさらに絞る.
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling",  
                  data.type = "Gene Expression Quantification")
```

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る

```
> output_query <-getResults(query)
> head(output_query,20)
```

	id	data_format	cases
1	e3224205-7a8c-4de2-983c-ac2e9b592e61	TSV	TCGA-V4-A9ES-01A-11R-A405-07
2	99ce0e9b-b1c8-4adf-8529-393068fa8f69	TSV	TCGA-V4-A9ES-01A-11R-A405-07
3	0daede84-5a80-406c-8bf2-225df4f11d7c	TSV	TCGA-V4-A9EF-01A-21R-A405-07
4	4b4d4a58-ed2b-4d95-b89c-6dd1725414a1	TXT	TCGA-V4-A9EF-01A-21R-A40B-13
5	8a653ff6-f968-4f16-ac9d-2ae982d5c7e6	TXT	TCGA-V4-A9EF-01A-21R-A40B-13
6	152fdff5-c442-4384-a19a-8ccbel24ae	TXT	TCGA-V4-A9EF-01A-21R-A40B-13
7	2373d6d8-d98b-4cc3-a5ee-a8da3bbde2	Splice Junction Quantification	released
8	fc0fe5f2-4d8c-4dcc-85c7-010e7e3e09	Gene Expression Quantification	released
9	665251c9-f45f-4c4f-ad2a-a9fc7e17cb	Splice Junction Quantification	released
10	ed79c017-e771-4455-89cb-1bbfb8a5da	Isoform Expression Quantification	released
11	ab9762e1-dc39-4dca-9262-e1e3cff5b8	Isoform Expression Quantification	released
12	d3791987-53b8-444e-97c2-bc77f51613	miRNA Expression Quantification	released
13	1d970f75-011a-4fc6-a151-lad62e016c	Isoform Expression Quantification	released
14	a75a4d9a-266e-4a0b-835a-6f49a25a8e	Splice Junction Quantification	released
15	0747c448-d280-40bd-ab2a-08a6daa8f9	miRNA Expression Quantification	released
16	eb1ba0fe-22eb-48b8-b51f-6f2640cbd3	Isoform Expression Quantification	released
17	f84b24e5-751f-4b17-a6dd-74b046844c	miRNA Expression Quantification	released
18	b8059d49-e273-4c75-81e1-a805d0503c	miRNA Expression Quantification	released
19	654d9d43-a561-45f9-bf51-620eb11a3a	Gene Expression Quantification	released
20	50363092-20b0-4b7e-bbel-06344dl3c9	Splice Junction Quantification	released
	access	Isoform Expression Quantification	released
		Isoform Expression Quantification	released
		Gene Expression Quantification	released
		Splice Junction Quantification	released
		Splice Junction Quantification	released
		miRNA Expression Quantification	released

発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)
```

```
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認
```

```
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
```

```
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling")
```

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定してさらに絞る.
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling",  
                  data.type = "Gene Expression Quantification")
```

```
output_query <- getResults(query)
```

```
head(output_query, 20)
```

データを検索する

⇒ ・ getProjectSummary()
・ GDCquery()

最終的なクエリの作成. GDCquery関数で、 data.typeを指定してさらに絞る.

```
> # 最終的なクエリの作成. GDCquery関数で、 data.typeを指定してさらに絞る.  
> query <- GDCquery(project = "TCGA-UVM",  
+                   data.category = "Transcriptome Profiling",  
+                   data.type = "Gene Expression Quantification")
```

```
-----  
o GDCquery: Searching in GDC database  
-----
```

```
Genome of reference: hg38  
-----
```

```
oo Accessing GDC. This might take a while...  
-----
```

```
ooo Project: TCGA-UVM  
-----
```

```
oo Filtering results  
-----
```

```
ooo By data.type  
-----
```

```
oo Checking data  
-----
```

```
ooo Checking if there are duplicated cases
```

```
ooo Checking if there are results for the query  
-----
```

```
o Preparing output  
-----
```

発現データからヒートマップを描く

GDCデータベースを検索して対象データを絞り込む

```
library(TCGAbiolinks)
```

```
library(SummarizedExperiment)
```

```
# 全 GDCプロジェクトを確認
```

```
getGDCprojects()
```

```
# “TCGA-UVM”(Nevi and Melanomas)を選びデータの種類や症例数を確認⇒80例
```

```
getProjectSummary("TCGA-UVM")
```

```
# GDCquery関数で“Transcriptome Profiling”を指定し、該当するデータ見る
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling")
```

```
output_query <- getResults(query)
```

```
head(output_query,20)
```

```
# 最終的なクエリの作成. GDCquery関数で、 data.typeを指定しでさらに絞る.
```

```
query <- GDCquery(project = "TCGA-UVM",  
                  data.category = "Transcriptome Profiling",  
                  data.type = "Gene Expression Quantification")
```

```
output_query <- getResults(query)
```

```
head(output_query,20)
```

データを検索する

⇒ ・ `getProjectSummary()`
・ `GDCquery()`

Gene expressionだけに絞った場合

```
> output_query <-getResults(query)
> head(output_query,20)
```

	id	data_format	cases
1	99ce0e9b-b1c8-4adf-8529-393068fa8f69	TSV	TCGA-V4-A9ES-01A-11R-A405-07
2	1d970f75-011a-4fc6-a151-lad62e016c35	TSV	TCGA-YZ-A983-01A-11R-A405-07
3	f84b24e5-751f-4b17-a6dd-74b046844d0d	TSV	TCGA-VD-A8KO-01A-11R-A405-07
4	37b6b698-b9d0-4efd-b8fa-410803ff43f7	TSV	TCGA-V4-A9E5-01A-11R-A405-07
5	0db0e29f-bd16-4f9a-944a-9b2b73ae3231	TSV	TCGA-WC-A87Y-01A-11R-A405-07
6	dl48f481-a38c-481e-b0c0-cdad04cbf3ba	TSV	TCGA-VD-A8KH-01A-11R-A405-07
7	b7ac31c0-1284-4100-b2e6-052e0e4b000d	TSV	TCGA-VD-A8KH-01A-11R-A405-07

	data_type	state	experimental_strategy	version
8	Gene Expression Quantification	released	RNA-Seq	1
9	Gene Expression Quantification	released	RNA-Seq	1
10	Gene Expression Quantification	released	RNA-Seq	1
11	Gene Expression Quantification	released	RNA-Seq	1
12	Gene Expression Quantification	released	RNA-Seq	1
13	Gene Expression Quantification	released	RNA-Seq	1
14	Gene Expression Quantification	released	RNA-Seq	1
15	Gene Expression Quantification	released	RNA-Seq	1
16	Gene Expression Quantification	released	RNA-Seq	1
17	Gene Expression Quantification	released	RNA-Seq	1
18	Gene Expression Quantification	released	RNA-Seq	1
19	Gene Expression Quantification	released	RNA-Seq	1
20	Gene Expression Quantification	released	RNA-Seq	1

	access
1	open
2	open
3	open

対象データの確認、ダウンロード、変換、ヒートマップ描画

#データ数、症例数を確認しておく⇒80行、80症例

```
length(output_query[,1])
```

```
length(unique(output_query$cases.submitter_id)) #cases.submitter_id: 病例ID
```

ダウンロード

GDCdownload(query)

summarizedExperimentオブジェクトに変換

```
UVM <- GDCprepare(query)
```

assay関数で数値データ(行列)を抽出

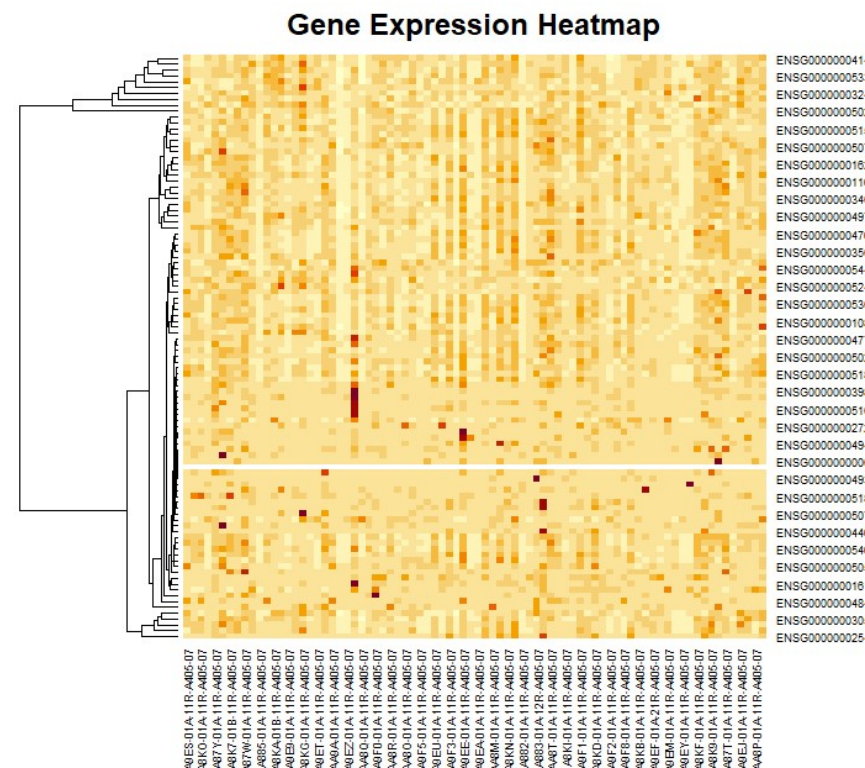
```
UVM mat <- assay(UVM)
```

行列の次元⇒ 60660 80

```
dim(UVM_mat)
```

```
UVM mat2 <- UVM mat[1:100,]
```

```
heatmap(UVM_mat2,  
        Colv = NA,  
        scale = "row",  
        main = "Gene Expression Heatmap")
```



#データ数、症例数を確認しておく⇒ 80行、80症例

```
>
>
> #データ数、症例数を確認しておく⇒ 80行、80症例
> length(output_query[,1])
[1] 80
> length(unique(output_query$cases.submitter_id)) #cases.submitter_id: 症例ID
[1] 80
>
~
```

対象データの確認、ダウンロード、変換、ヒートマップ描画

#データ数、症例数を確認しておく⇒ 80行、80症例

```
length(output_query[,1])
```

```
length(unique(output_query$cases.submitter_id)) #cases.submitter_id: 症例ID
```

ダウンロード

```
GDCdownload(query)
```

summarizedExperimentオブジェクトに変換

```
UVM <- GDCprepare(query)
```

assay関数で数値データ(行列)を抽出

```
UVM_mat <- assay(UVM)
```

行列の次元⇒ 60660 80

```
dim(UVM_mat)
```

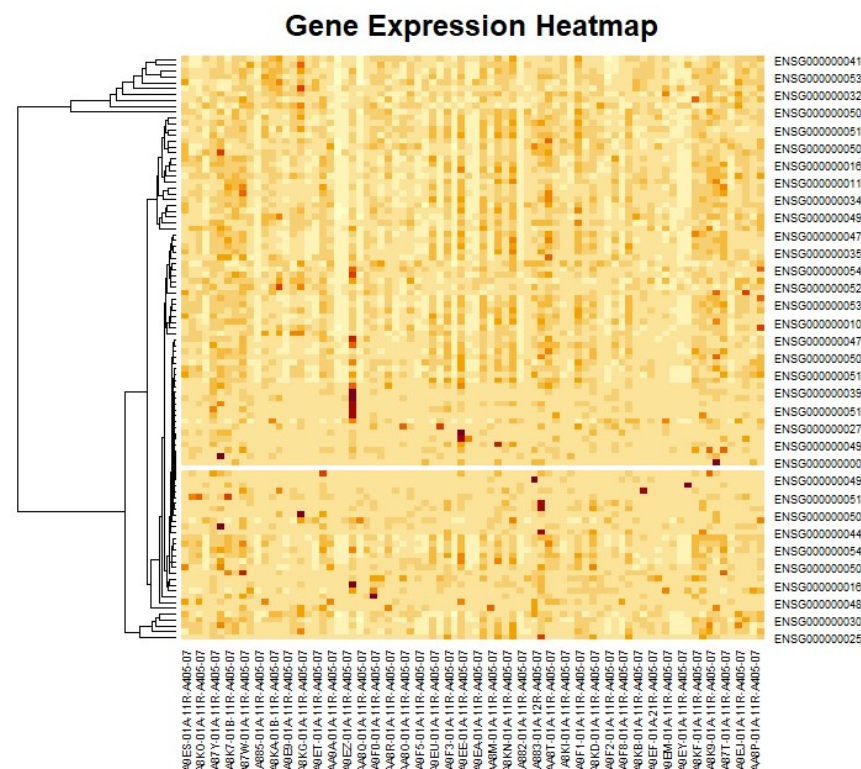
```
UVM_mat2 <- UVM_mat[1:100,]
```

```
heatmap(UVM_mat2,
```

```
Colv = NA,
```

```
scale = "row",
```

```
main = "Gene Expression Heatmap")
```



ダウンロード

```
>  
> # ダウンロード  
> GDCdownload(query)  
Downloading data for project TCGA-UVM  
GDCdownload will download 80 files. A total of 337.254347 MB  
Downloading as: Mon_Feb_12_22_32_15_2024.tar.gz  
Downloading: 77 MB      >  
> |
```

対象データの確認、ダウンロード、変換、ヒートマップ描画

#データ数、症例数を確認しておく⇒ 80行、80症例

```
length(output_query[,1])
```

```
length(unique(output_query$cases.submitter_id)) #cases.submitter_id: 症例ID
```

ダウンロード

```
GDCdownload(query)
```

summarizedExperimentオブジェクトに変換
UVM <- GDCprepare(query)

assay関数で数値データ(行列)を抽出

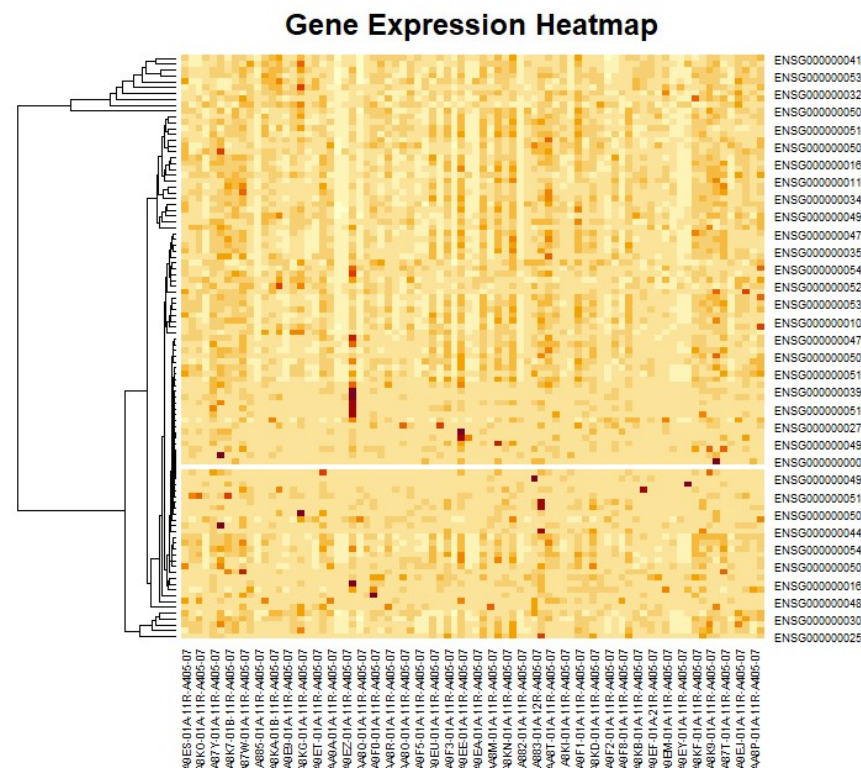
```
UVM_mat <- assay(UVM)
```

行列の次元⇒ 60660 80

```
dim(UVM_mat)
```

```
UVM_mat2 <- UVM_mat[1:100,]
```

```
heatmap(UVM_mat2,  
  Colv = NA,  
  scale = "row",  
  main = "Gene Expression Heatmap")
```



summarizedExperimentオブジェクトに変換

```
> # summarizedExperimentオブジェクトに変換
> UVM <- GDCprepare(query)
|=====|100%
Starting to add information to samples
=> Add clinical information to samples
=> Adding TCGA molecular information from marker papers
=> Information will have prefix 'paper_'
uvm subtype information from:doi:10.1016/j.ccell.2017.07.003
Available assays in SummarizedExperiment :
=> unstranded
=> stranded_first
=> stranded_second
=> tpm_unstrand
=> fpkm_unstrand
=> fpkm_uq_unstrand
>
```

\$

対象データの確認、ダウンロード、変換、ヒートマップ描画

#データ数、症例数を確認しておく⇒ 80行、80症例

```
length(output_query[,1])
```

```
length(unique(output_query$cases.submitter_id)) #cases.submitter_id: 症例ID
```

ダウンロード

```
GDCdownload(query)
```

summarizedExperimentオブジェクトに変換

```
UVM <- GDCprepare(query)
```

assay関数で数値データ(行列)を抽出

```
UVM_mat <- assay(UVM)
```

行列の次元⇒ 60660 80

```
dim(UVM_mat)
```

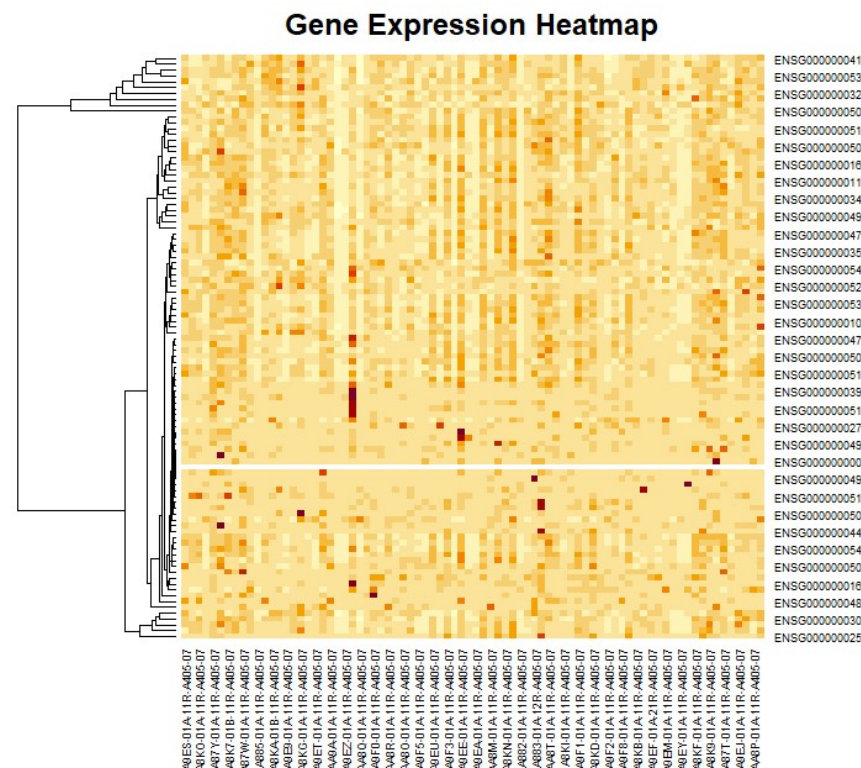
```
UVM_mat2 <- UVM_mat[1:100,]
```

```
heatmap(UVM_mat2,
```

```
Colv = NA,
```

```
scale = "row",
```

```
main = "Gene Expression Heatmap")
```



assay関数で数値データ(行列)を抽出
行列の次元⇒ 60660 80

```
>  
> # assay関数で数値データ(行列)を抽出  
> UVM_mat <- assay(UVM)  
>  
> # 行列の次元⇒ 60660 80  
> dim(UVM_mat)  
[1] 60660 80  
>  
>
```

対象データの確認、ダウンロード、変換、ヒートマップ描画

#データ数、症例数を確認しておく⇒80行、80症例

```
length(output_query[,1])
```

```
length(unique(output$query$cases.submitter id)) #cases.submitter id: 病例ID
```

ダウンロード

GDCdownload(query)

summarizedExperimentオブジェクトに変換

```
UVM <- GDCprepare(query)
```

assay関数で数値データ(行列)を抽出

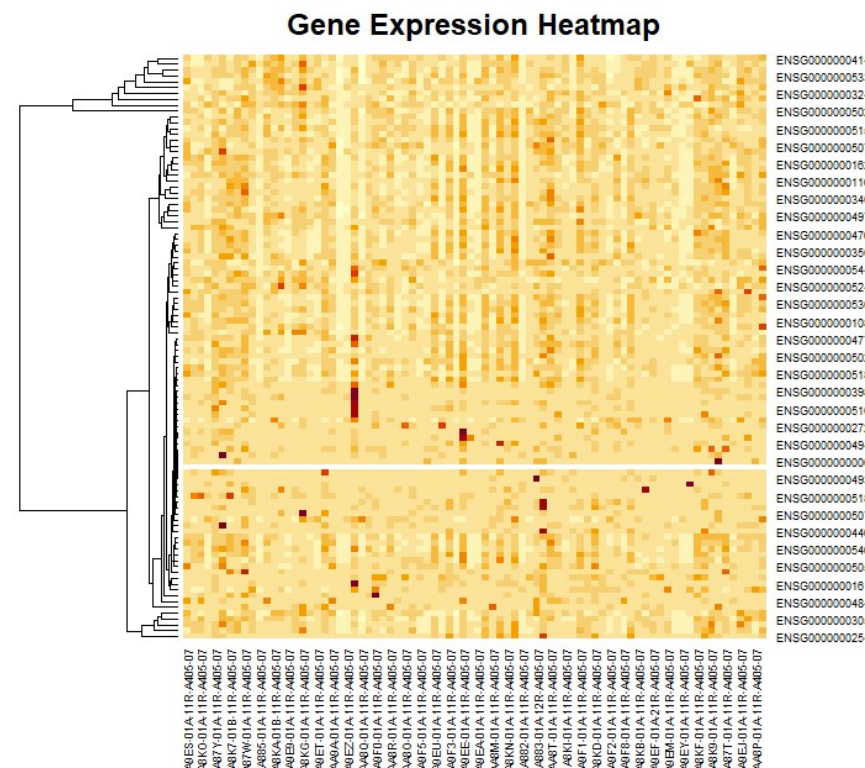
```
UVM mat <- assay(UVM)
```

行列の次元⇒ 60660 80

```
dim(UVM_mat)
```

```
UVM_mat2 <- UVM_mat[1:100,]
```

```
heatmap(UVM_mat2,
        Colv = NA,
        scale = "row",
        main = "Gene Expression Heatmap")
```



```
UVM_mat2 <- UVM_mat[1:100,]
```

```
>  
> UVM_mat2 <- UVM_mat[1:100,]  
>
```

対象データの確認、ダウンロード、変換、ヒートマップ描画

#データ数、症例数を確認しておく⇒80行、80症例

```
length(output_query[,1])
```

```
length(unique(output$query$cases.submitter id)) #cases.submitter id: 病例ID
```

ダウンロード

GDCdownload(query)

summarizedExperimentオブジェクトに変換

```
UVM <- GDCprepare(query)
```

assay関数で数値データ(行列)を抽出

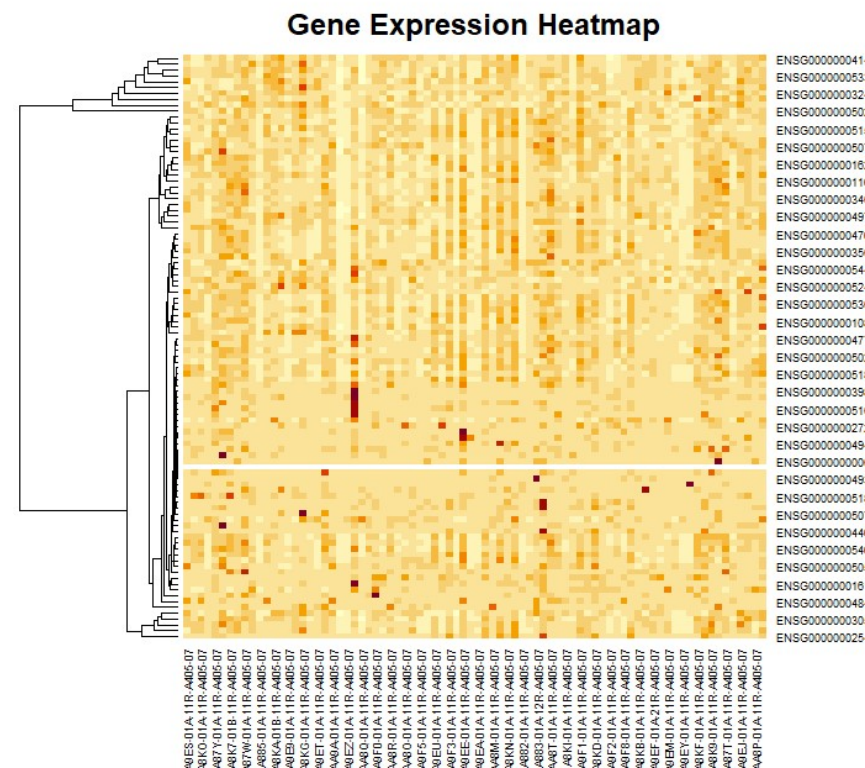
```
UVM mat <- assay(UVM)
```

行列の次元⇒ 60660 80

```
dim(UVM_mat)
```

```
UVM mat2 <- UVM mat[1:100,]
```

```
heatmap(UVM_mat2,
        Colv = NA,
        scale = "row",
        main = "Gene Expression Heatmap")
```

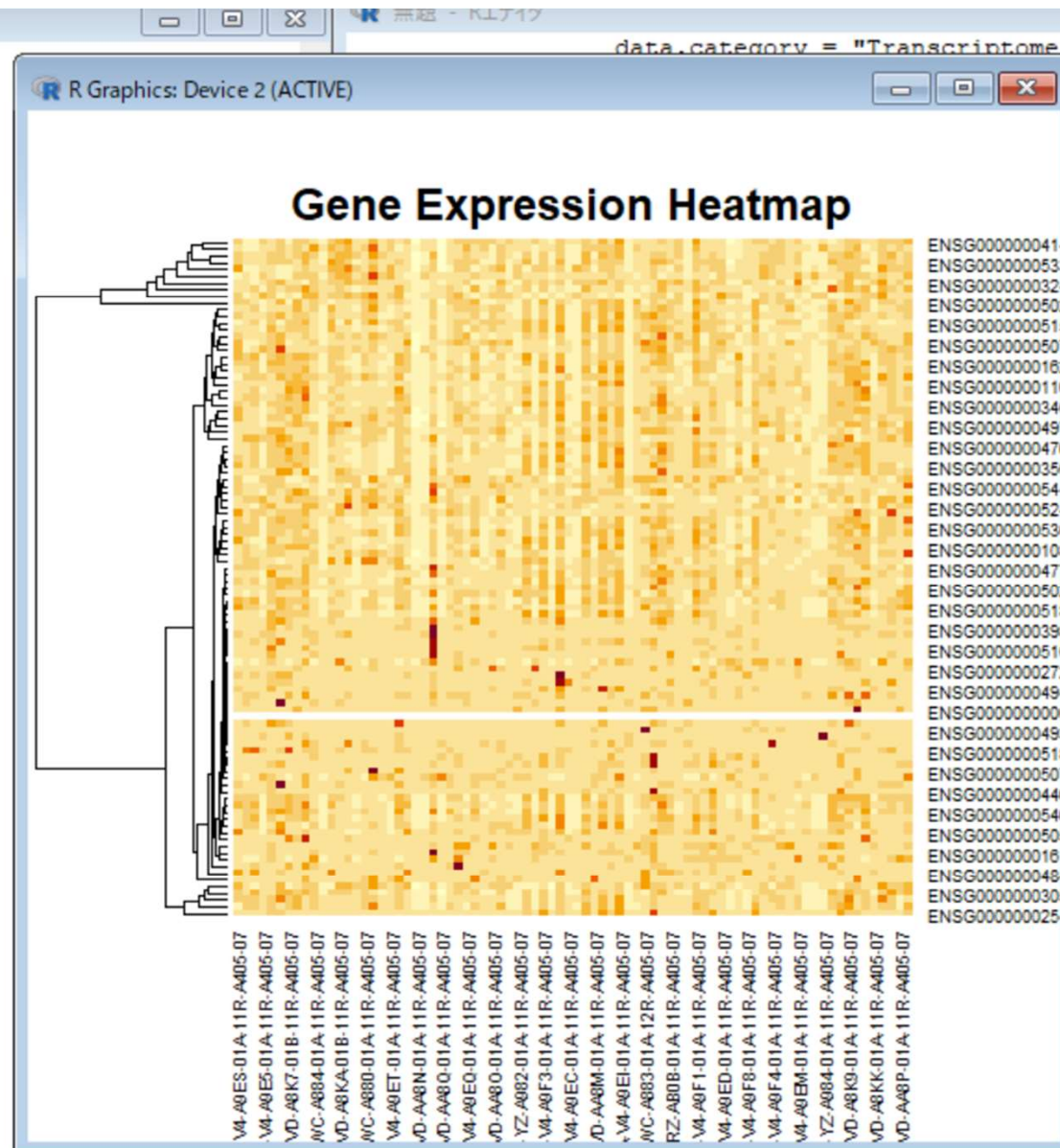


ヒートマップ

```

R Console
> GDCdownload(query)
Downloading data for project TCGA-UVM
GDCdownload will download 80 files. A total of 337.254347 MB
Downloading as: Mon_Feb_12_22_32_15_2024.tar.gz
Downloading: 77 MB
>
> # summarizedExperimentオブジェクトに変換
> UVM <- GDCprepare(query)
|=====|100%
Starting to add information to samples
=> Add clinical information to samples
=> Adding TCGA molecular information from marker papers
=> Information will have prefix 'paper_'
uvm subtype information from:doi:10.1016/j.ccell.2017.07.003
Available assays in SummarizedExperiment :
=> unstranded
=> stranded_first
=> stranded_second
=> tpm_unstrand
=> fpkm_unstrand
=> fpkm_uq_unstrand
>
> # assay関数で数値データ(行列)を抽出
> UVM_mat <- assay(UVM)
>
> # 行列の次元⇒ 60660      80
> dim(UVM_mat)
[1] 60660      80
>
> UVM_mat2 <- UVM_mat[1:100,]
>
> heatmap(UVM_mat2,
+         Colv = NA,
+         scale = "row",
+         main = "Gene Expression Heatmap")
>
>

```



SNVの変異のサマリーとオンコプロット

```
library(TCGAbiolinks)
```

```
library(maftools)
```

```
# クエリを作成. データを検索するプロセスは省略している
```

```
query <- GDCquery( project = "TCGA-LUAD",
```

```
  data.category = "Simple Nucleotide Variation",
```

```
  access = "open", data.type = "Masked Somatic Mutation",
```

```
  workflow.type = "Aliquot Ensemble Somatic Variant Merging and Masking")
```

```
# ダウンロード
```

```
GDCdownload(query)
```

```
# maf形式に変換. QueryがSNVのときGDCprepareの出力はmaf形式になる
```

```
LUAD <- GDCprepare(query)
```

```
LUAD1 = read.maf(maf = LUAD)
```

```
# MAFサマリー作成
```

```
plotmafSummary(maf = LUAD1, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = FALSE)
```

```
oncoplot(maf = LUAD1, top = 10)
```

クエリ作成とダウンロード

```
> # クエリを作成. データを検索するプロセスは省略している
> query <- GDCquery( project = "TCGA-LUAD",
+   data.category = "Simple Nucleotide Variation",
+   access = "open", data.type = "Masked Somatic Mutation",
+   workflow.type = "Aliquot Ensemble Somatic Variant Merging and Masking")
-----
o GDCquery: Searching in GDC database
-----
Genome of reference: hg38
-----
oo Accessing GDC. This might take a while...
-----
ooo Project: TCGA-LUAD
-----
oo Filtering results
-----
ooo By access
ooo By data.type
ooo By workflow.type
-----
oo Checking data
-----
ooo Checking if there are duplicated cases
ooo Checking if there are results for the query
-----
o Preparing output
-----
>
> # ダウンロード
> GDCdownload(query)
Downloading data for project TCGA-LUAD
GDCdownload will download 618 files. A total of 61.056443 MB
Downloading as: Mon_Feb_12_23_03_29_2024.tar.gz
Downloading: 61 MB      >
>
```

SNVの変異のサマリーとオンコプロット

```
library(TCGAbiolinks)
```

```
library(maftools)
```

```
# クエリを作成. データを検索するプロセスは省略している
```

```
query <- GDCquery( project = "TCGA-LUAD",  
  data.category = "Simple Nucleotide Variation",  
  access = "open", data.type = "Masked Somatic Mutation",  
  workflow.type = "Aliquot Ensemble Somatic Variant Merging and Masking")
```

```
# ダウンロード
```

```
GDCdownload(query)
```

```
# maf形式に変換. QueryがSNVのときGDCprepareの出力はmaf形式になる
```

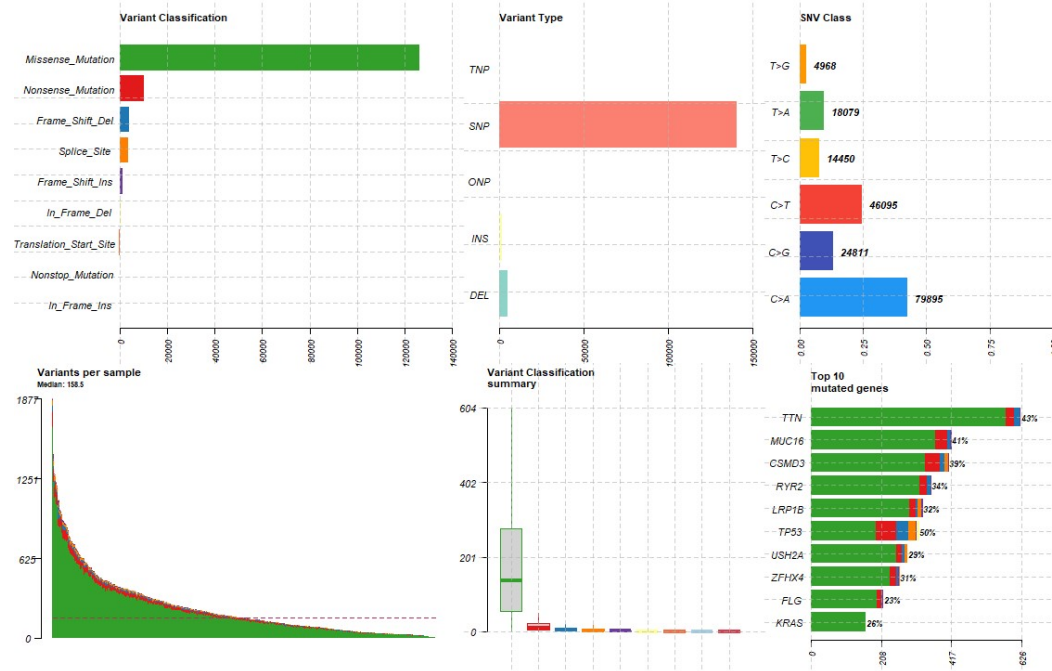
```
LUAD <- GDCprepare(query)
```

```
LUAD1 = read.maf(maf = LUAD)
```

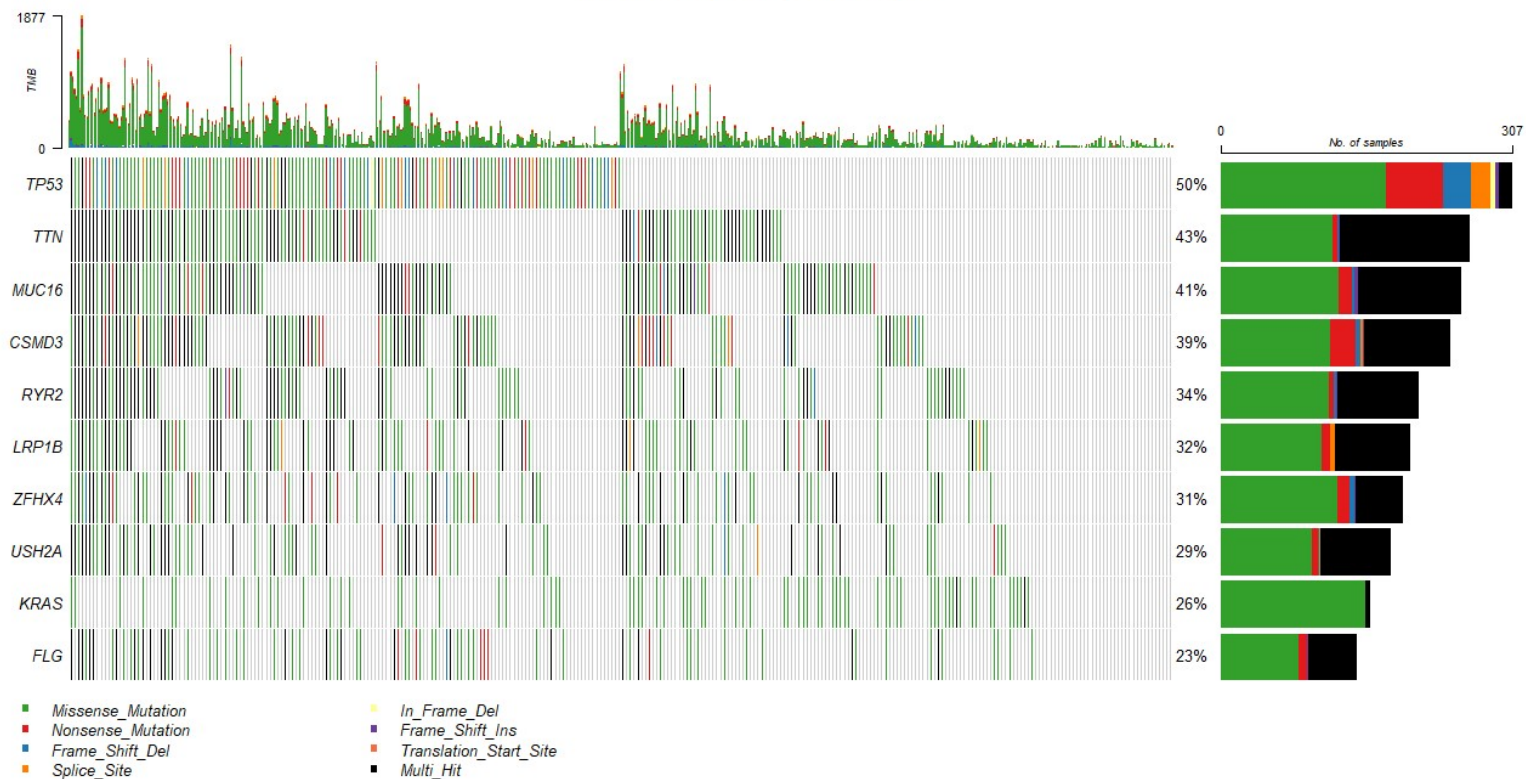
```
# MAFサマリー作成
```

```
plotmafSummary(maf = LUAD1, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = FALSE)
```

```
oncoplot(maf = LUAD1, top = 10)
```



Altered in 538 (87.34%) of 616 samples.



臨床データのダウンロードとKaplan-Meierプロット

```
library("survminer")
```

```
library("TCGAbiolinks")
```

```
# TCGA-LUADの全ての臨床データをダウンロード
```

```
clin <- GDCquery_clinic("TCGA-LUAD")
```

```
head(clin) #データフレーム形式であることが分かる
```

```
# KMプロット
```

```
TCGAanalyze_survival(clin, clusterCol="gender", risk.table = FALSE, conf.int = FALSE)
```

```
#現在のディレクトリ=保存場所
```

```
getwd()
```

