がん研究に関連した
バイオインフォマティクス入門

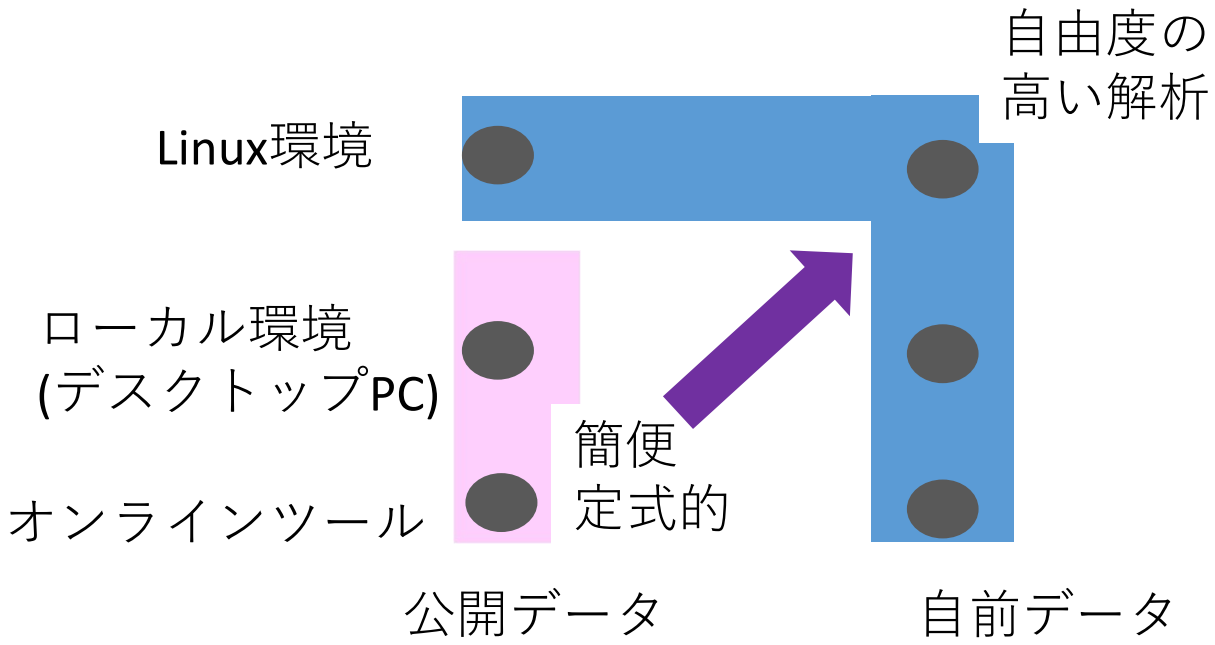
研究所 生物情報学分野
研究所 バイオインフォマティクス部門
がんゲノム情報管理センター（C-CAT）ゲノムデータ管理室
西野 穣

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

- がんと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

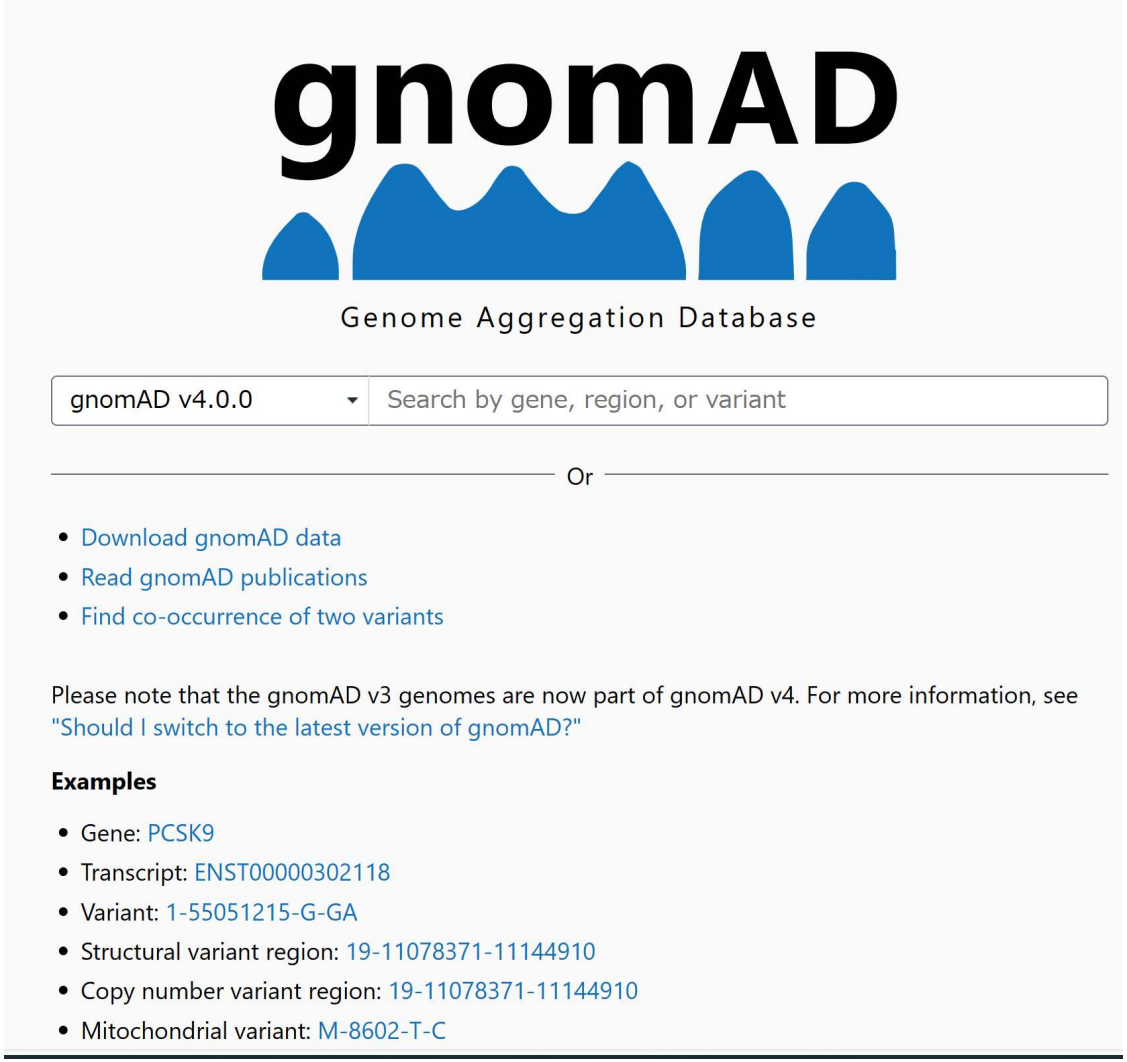
- ゲノムデータで予後予測モデルを作成した。結果をバリデートしたい

⇒ 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

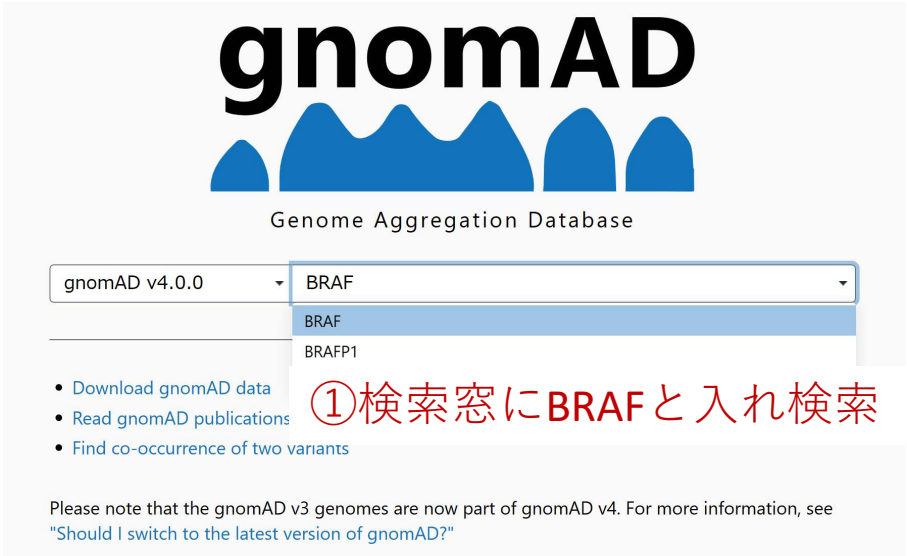
- gnomAD
  - 正常な個体の生殖細胞系列の変異をまとめたデータベース
    疾患関連の遺伝子変異フィルタリングに使用
- COSMIC
  - がん組織の体細胞変異に関する詳細な情報を提供するがんゲノム変異カタログ
- OncoKB
  - がん関連の体細胞性変異に対する治療への影響や臨床的な重要性を提供するがんの遺伝子変異知識ベース
- TCGA
  - がんゲノムや遺伝子発現などの包括的な情報を提供するプロジェクト。GDC Data Portalを通じてTCGAデータの検索、フィルタ、ダウンロード、簡単な分析ができる
- cBioPortal
  - TCGAなどの多数のがんゲ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



- 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

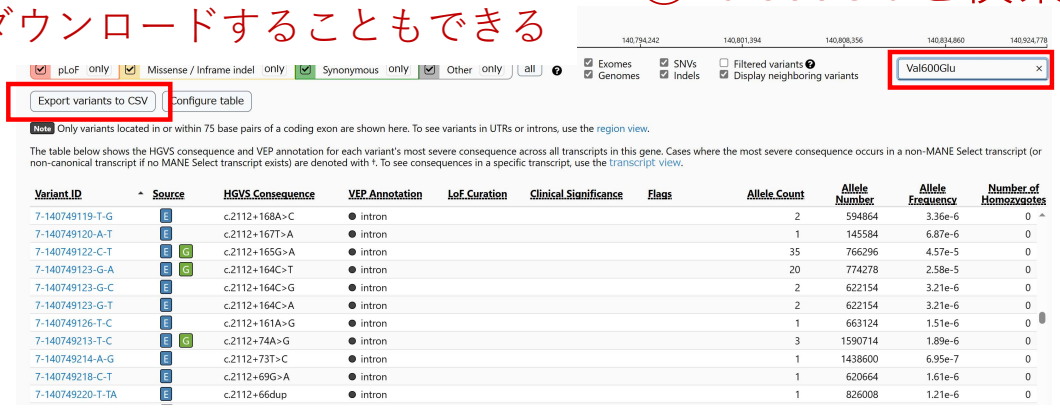
BRAF V600Eの頻度



①検索窓にBRAFと入れ検索

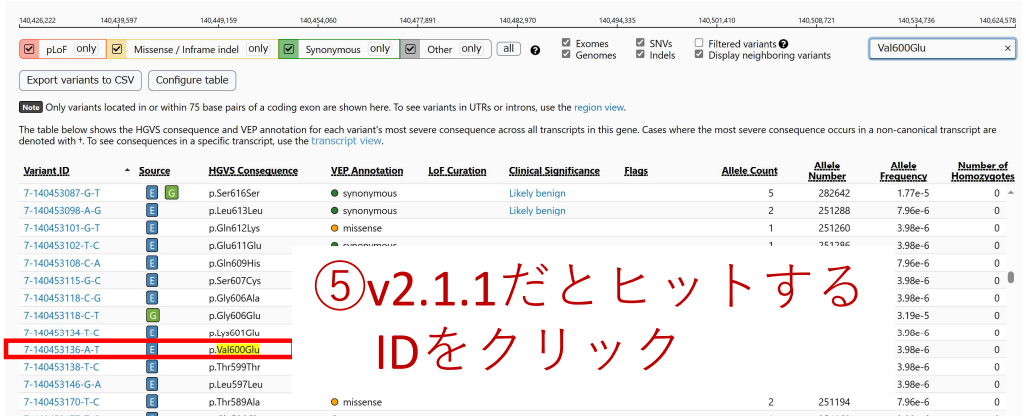
csv形式でBRAF全体の変異を
ダウンロードすることもできる

②Val600Gluを検索

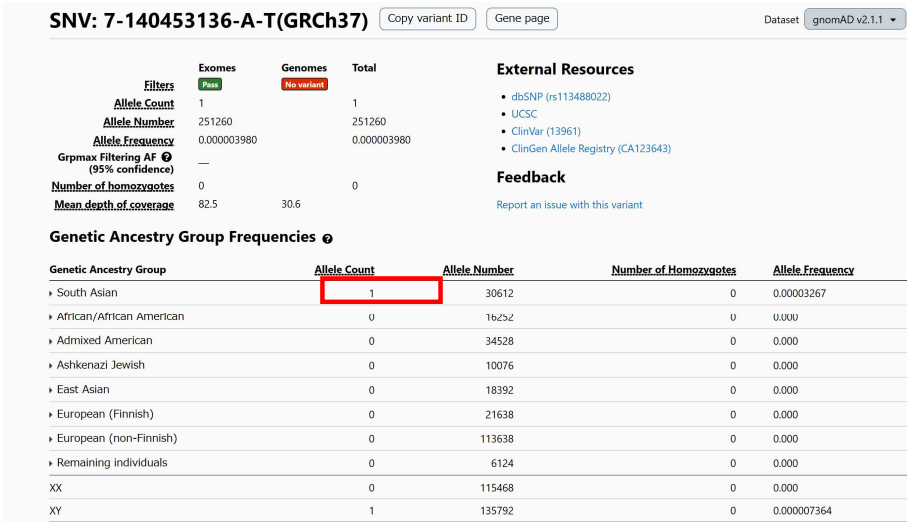


③v4.0.0だとヒットしない

④gnomAD v2.1.1で検索してみる



⑤v2.1.1だとヒットする
IDをクリック

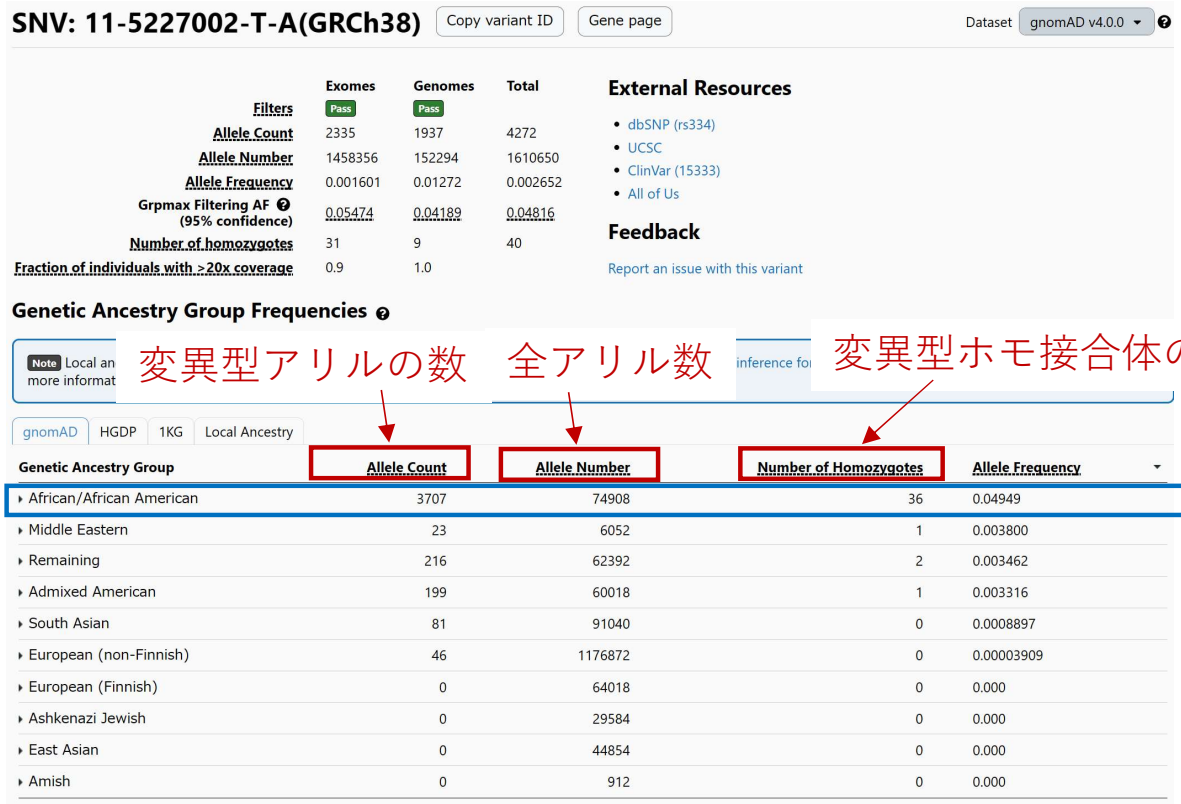


南アジアの個体に1コピー

※ 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

	Exomes	Genomes	Total
Filters	Pass	Pass	
Allele Count	2335	1937	4272
Allele Number	1458356	152294	1610650
Allele Frequency	0.001601	0.01272	0.002652
Grpmax Filtering AF (95% confidence)	0.05474	0.04189	0.04816
Number of homozygotes	31	9	40
Fraction of individuals with >20x coverage	0.9	1.0	

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

Feedback
Report an issue with this variant

Genetic Ancestry Group Frequencies

変異型アリルの数　全アリル数　変異型ホモ接合体の数

gnomAD | HGDP | 1KG | Local Ancestry

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

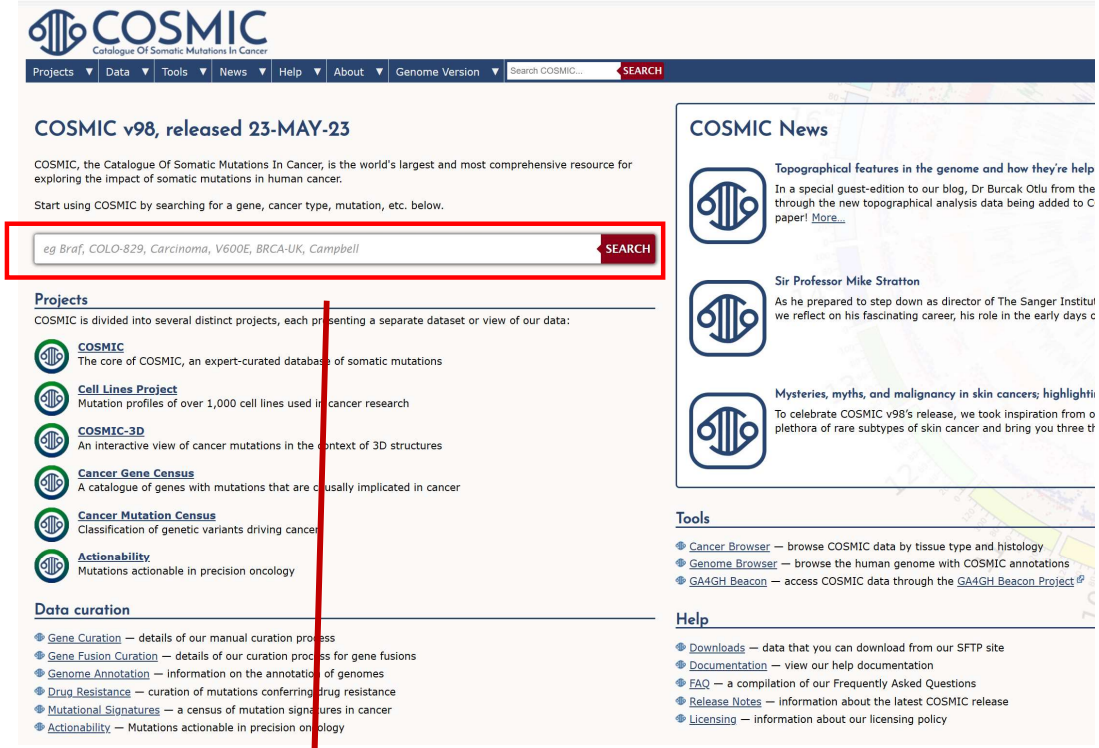
※ HbSを例に挙げたが、変異型ホモ接合ががん発症に関わる場合も多いと考えられる

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・ワインベルク平衡下での期待値92人よりかなり小さい

COSMIC

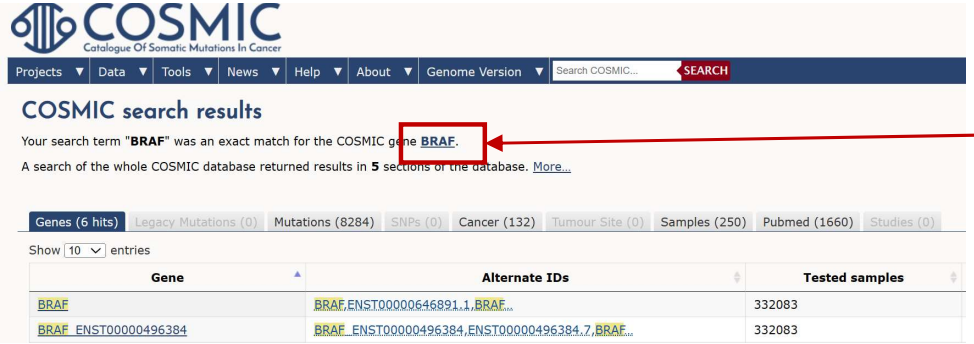


検索窓に
- 遺伝子名
- 細胞株名(COLO-829)
- 変異
- 研究プロジェクト

など入力して検索できる

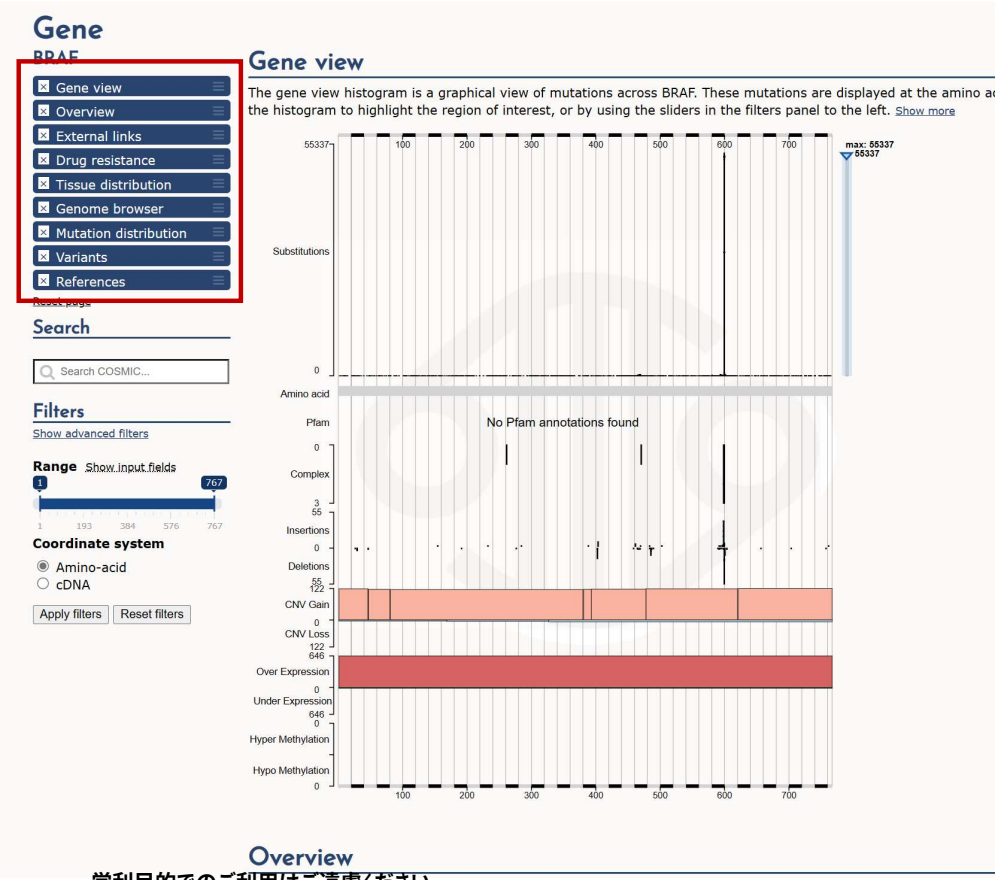
- COSMICは、がんにおける体細胞変異の影響を包括的に提供するデータベース
- COSMIC v86（2018年8月）には、26,000以上の文献から収集された1.4百万以上の腫瘍サンプルにわたり、約6百万のコーディング変異が含まれている
- 手作業でのキュレーションにより品質と精度が確保され、非コーディング変異や遺伝子融合など、がんを促進する多様なメカニズムも網羅
- COSMIC-3Dにおいては、タンパク質構造内の変異や薬物耐性の意義を探索する新機能が提供されている
- Cancer Gene Census（CGC）では、719のがんのドライバー遺伝子とその機能的説明をまとめ、がん研究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



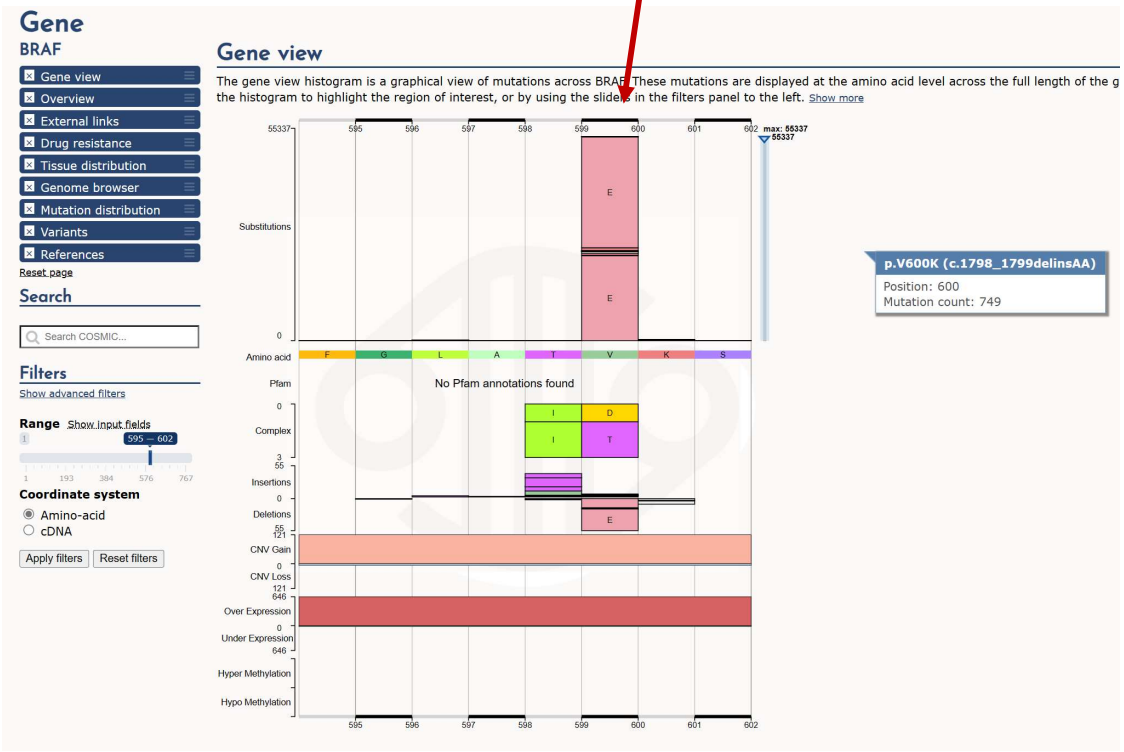
②6つヒットするが
上部の
“exact match for the COSMIC gene BRAF”
をクリック

③BRAFの詳細ページに移る



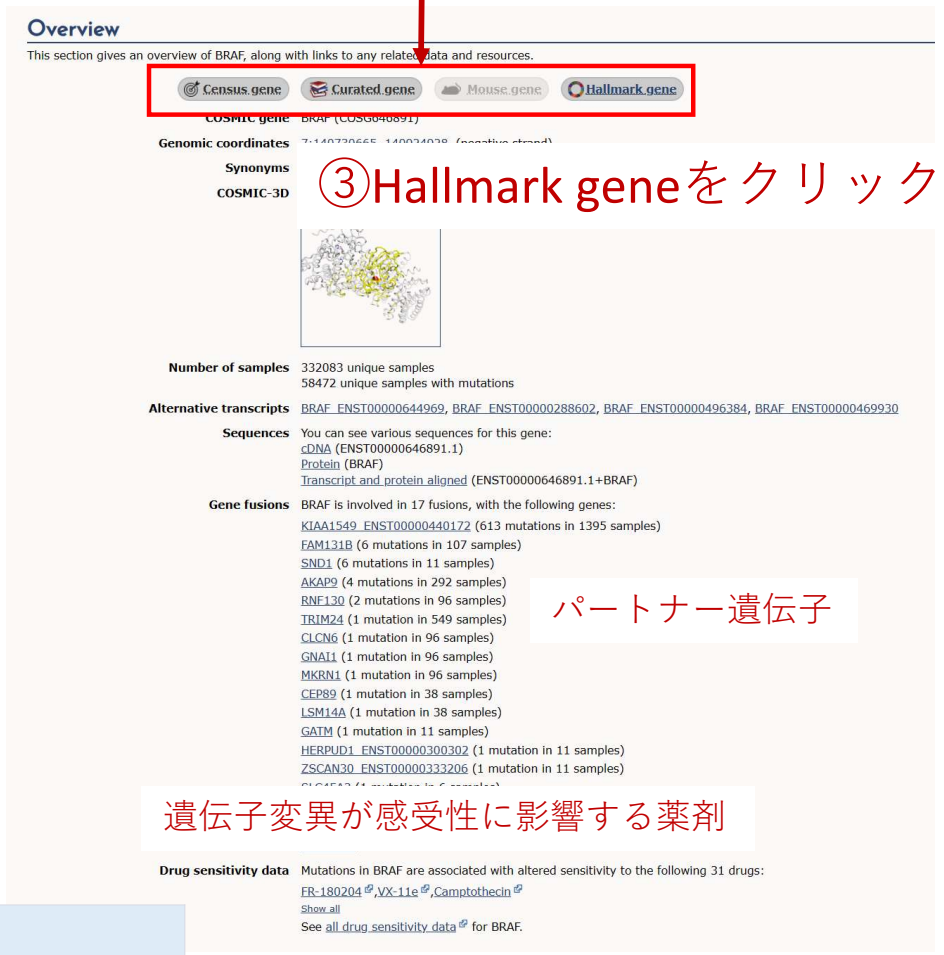
項目	説明
Gene View	各タイプの変異パターンやその頻度を示す
Overview	ゲノム位置、サンプル数、転写産物等の基本情報 Census gene、Curated gene、Hallmark geneであるか否か 3D構造と詳細へのリンク、パートナー遺伝子、遺伝子変異が感受性に影響する薬剤
External Links	他のバイオインフォマティクス的情報へのリンク
Drug Resistance	耐性変異に関する情報
Tissue Distribution	遺伝子が異なる組織でどの程度発現しているか示す
Genome Browser	ゲノム上の位置情報を視覚的に表示
Mutation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

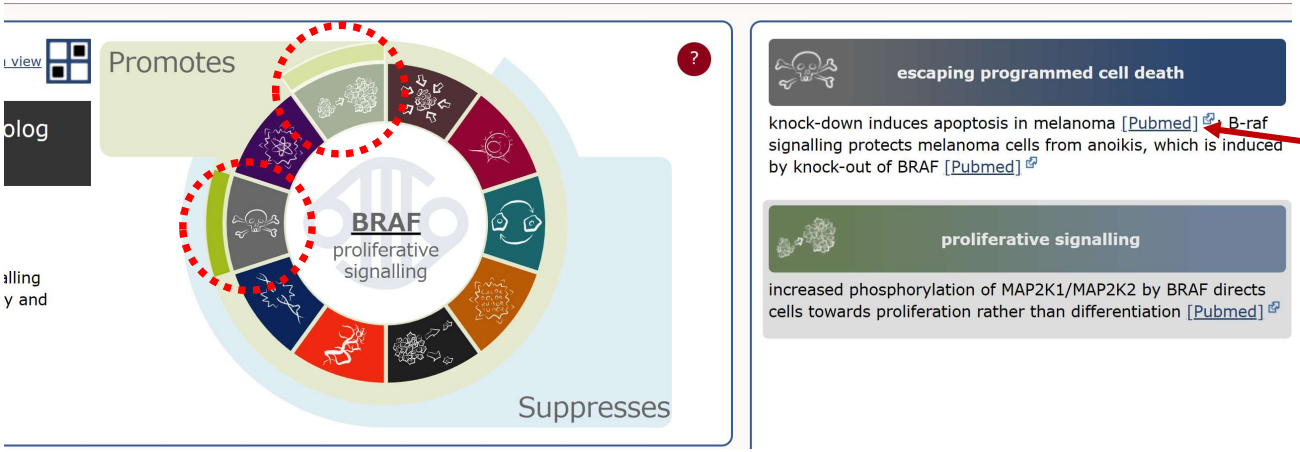


②BRAFはCensus gene, Curated gene, Hallmark geneに該当する

③Hallmark geneをクリック



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. 該当しない遺伝子は他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



文献リンク

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的情報へのリンク

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

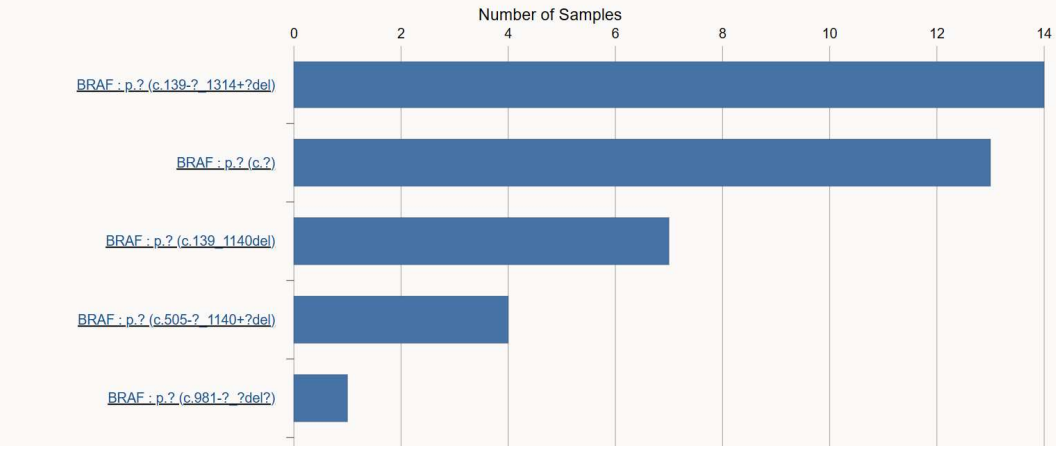
薬剤や遺伝子を選択し直してUpdate drugs, Update genesをクリックすると下のグラフが更新される

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_ENST000003

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 All 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

BRAF変異の一覧

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.

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

BRAF変異に関するリファレンス

This section shows publications associated with BRAF. You can see more information in our help pages.

Show 10 entries

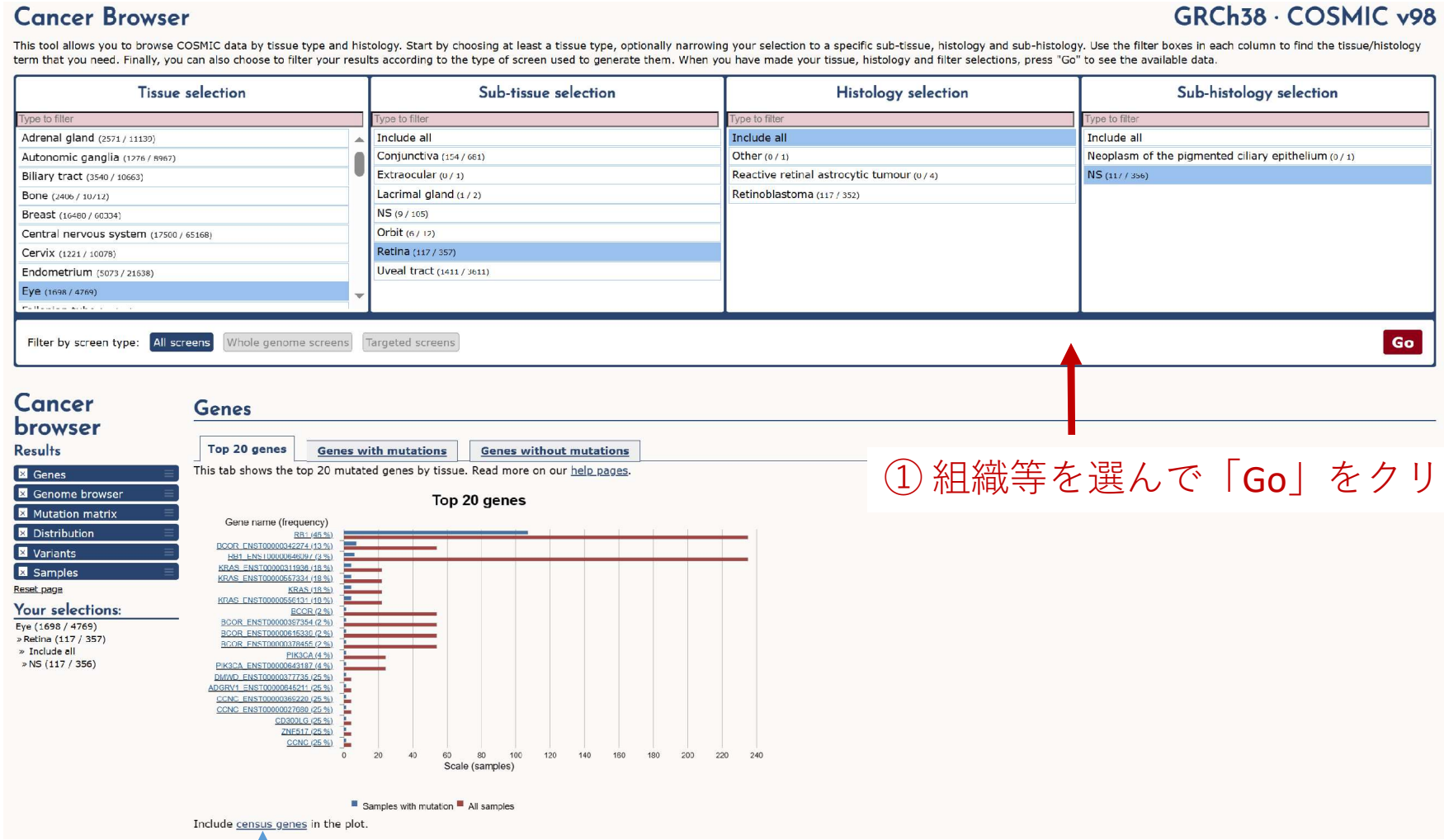
Reference Title	Author	Year	Journal	Status	COSMIC	Pubmed
BRAF and RAS mutations in human lung cancer and melanoma	Brose MS et al	2002	Cancer research;62(23):6997-7000	Curated	COSP5705	12460918
Missense mutations of the BRAF gene in human lung adenocarcinoma	Naoki K et al	2002	Cancer research;62(23):7001-3	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

COSMIC – Cancer Browser

組織 – histologyごとに、COSMICデータをゲノムブラウザ上で見る

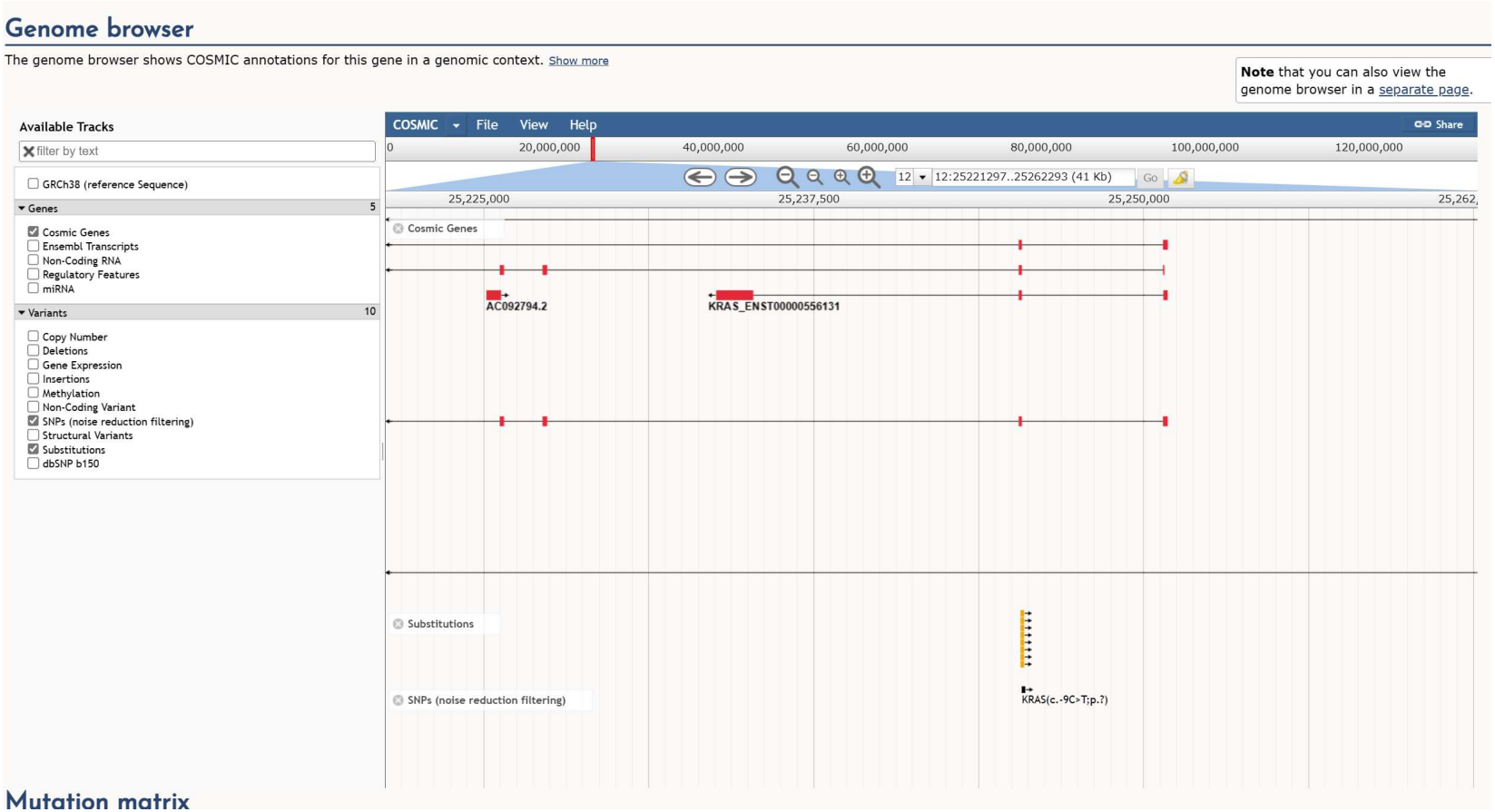
COSMICトップ画面のメニュー「Tools」→「Cancer Browser」を選択



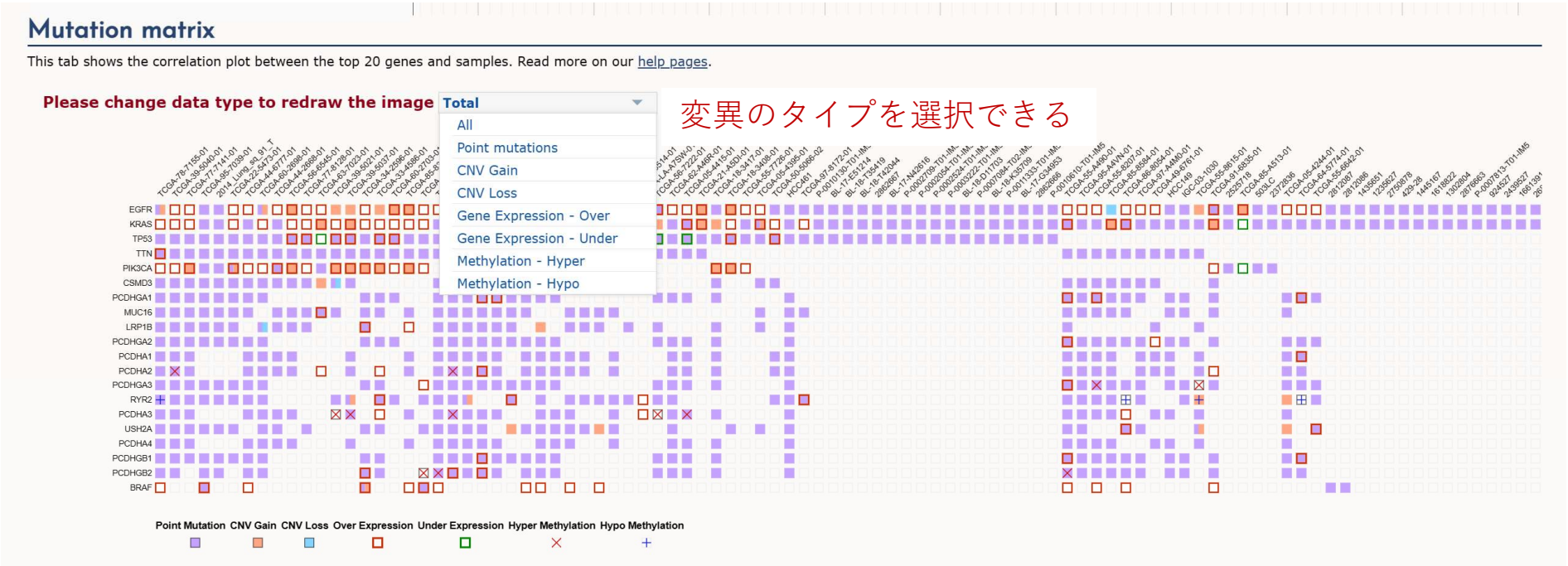
① 組織等を選んで「Go」をクリック

All genesとcensus genesの切替え

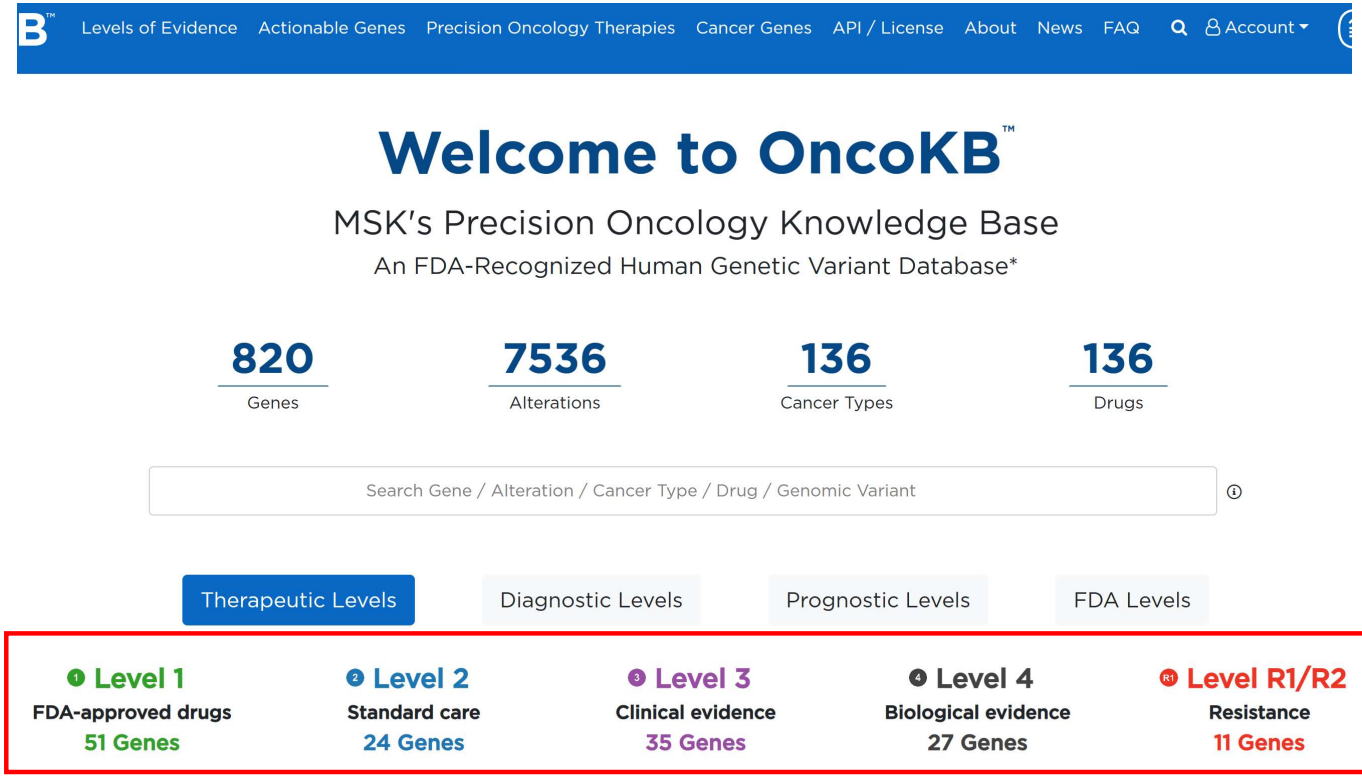
Genome browserでSNPと置換変異を表示させた例



サンプルごとに変異遺伝子を示す。変異タイプも分かる



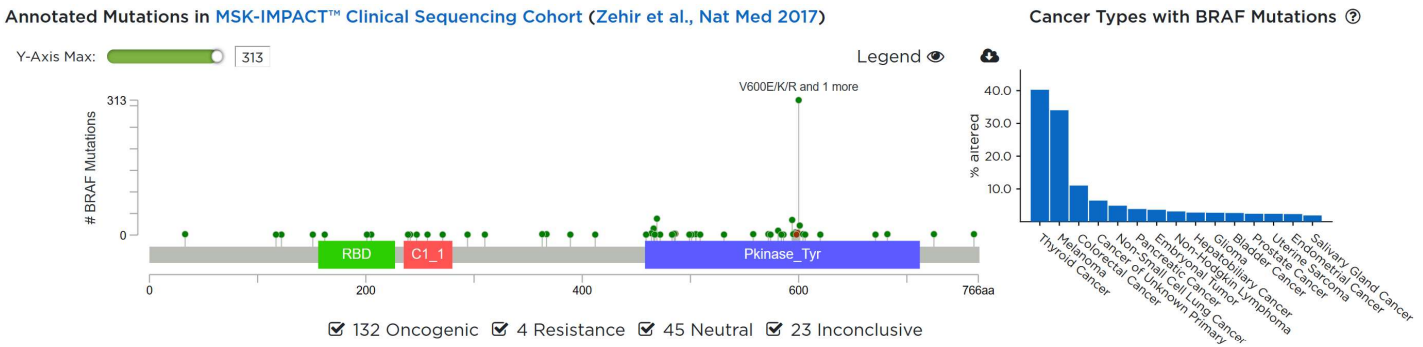
OncoKB



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: トップ画面でBRAFで検索



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.

Search ...

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		

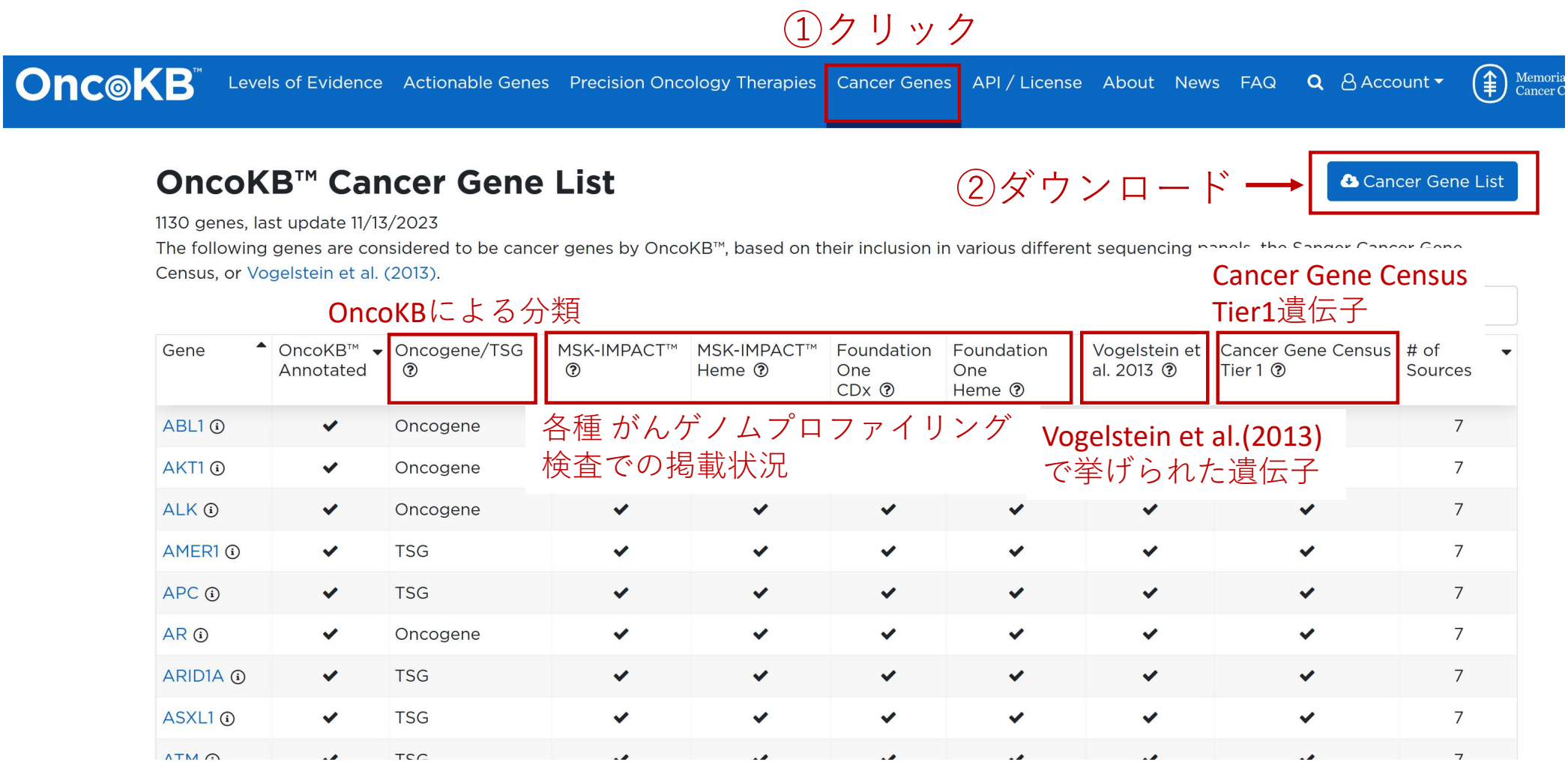
Atezolizumab, vemurafenib, and cobimetinib as first-line treatment for unresectable advanced BRAF(V600) mutation-positive melanoma (IMspire150): primary analysis of the randomised, double-blind, placebo-controlled, phase 3 trial.
Gutzmer R et al. Lancet. 2020 PMID: 32534646

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

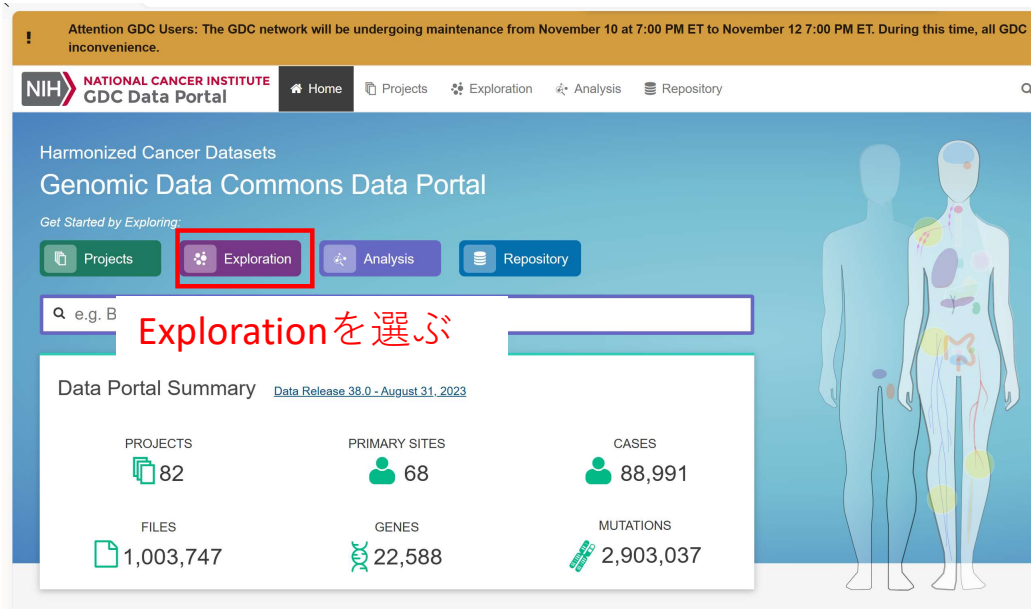
OncoKB: Cancer Gene List

OncoKBの”Cancer Genes”リストはダウンロードできる



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)

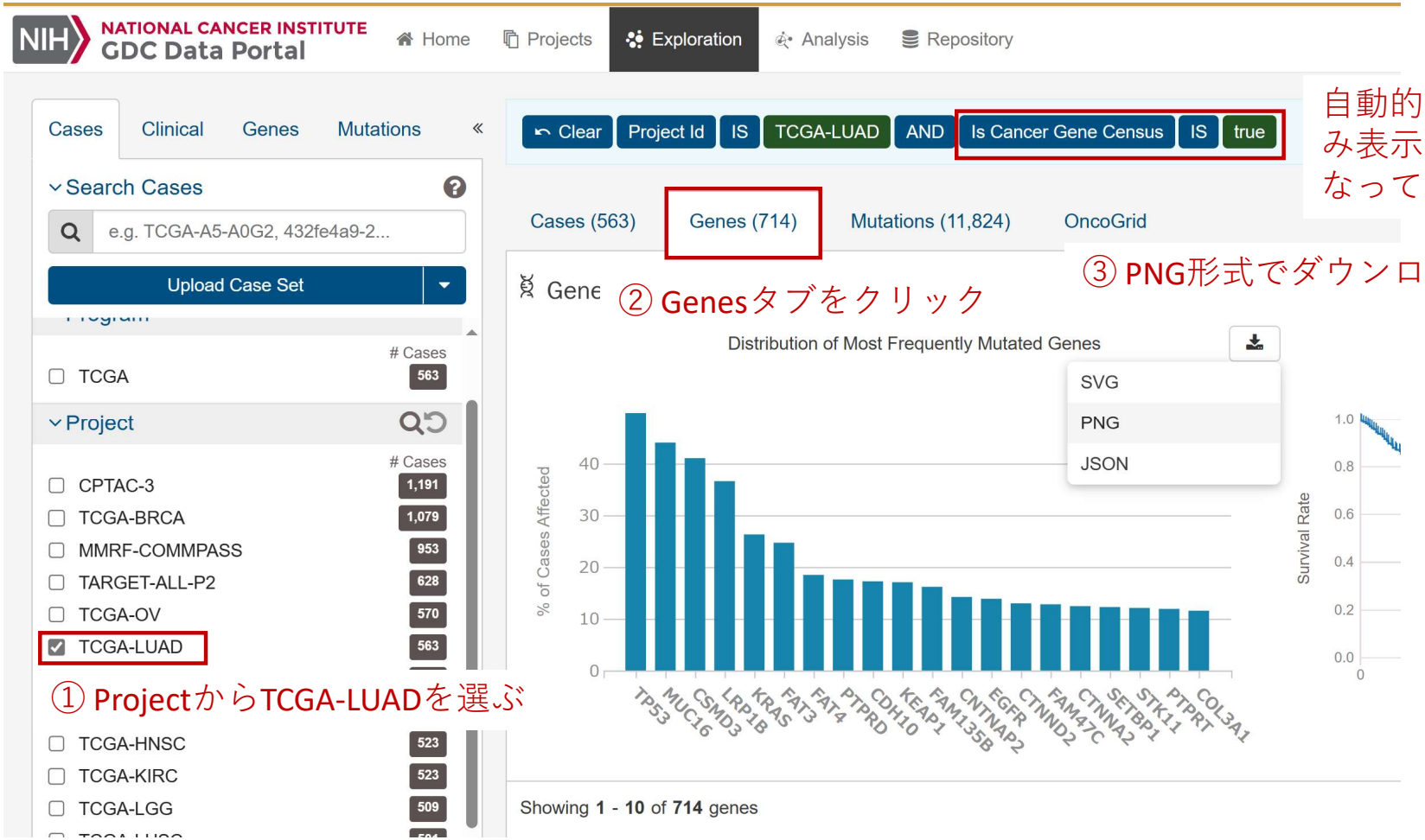
TCGA



- 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
- このデータは既にがんの診断、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



自動的にCensus geneのみ表示されるようになっている

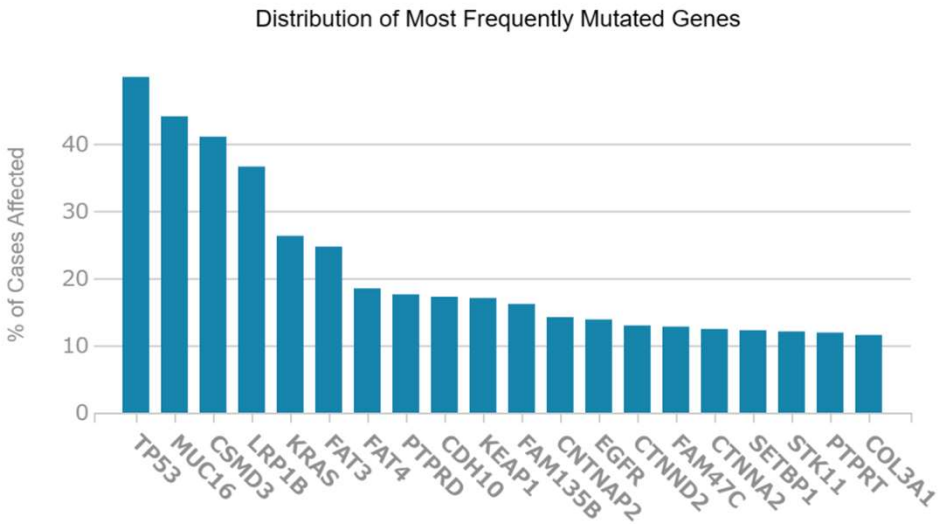
② Genesタブをクリック

③ PNG形式でダウンロードして保存

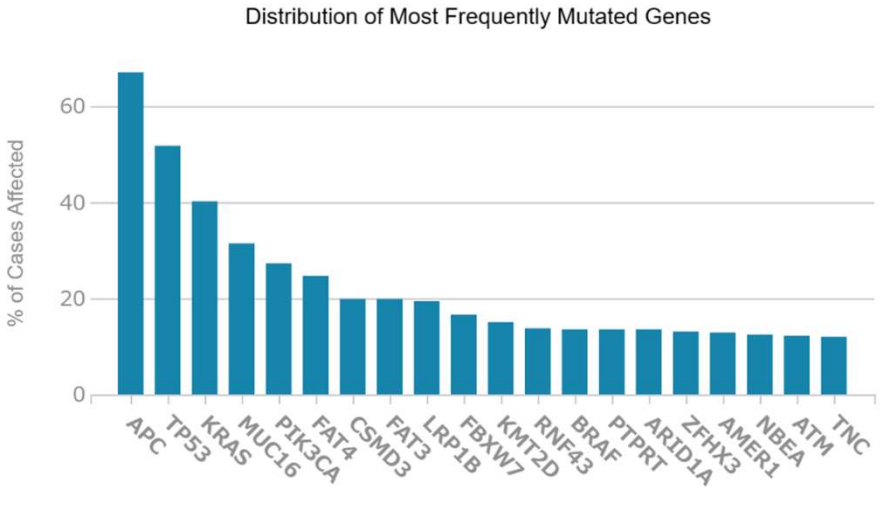
① ProjectからTCGA-LUADを選ぶ

※ TCGA-COADについても同様に操作する

TCGA-LUAD

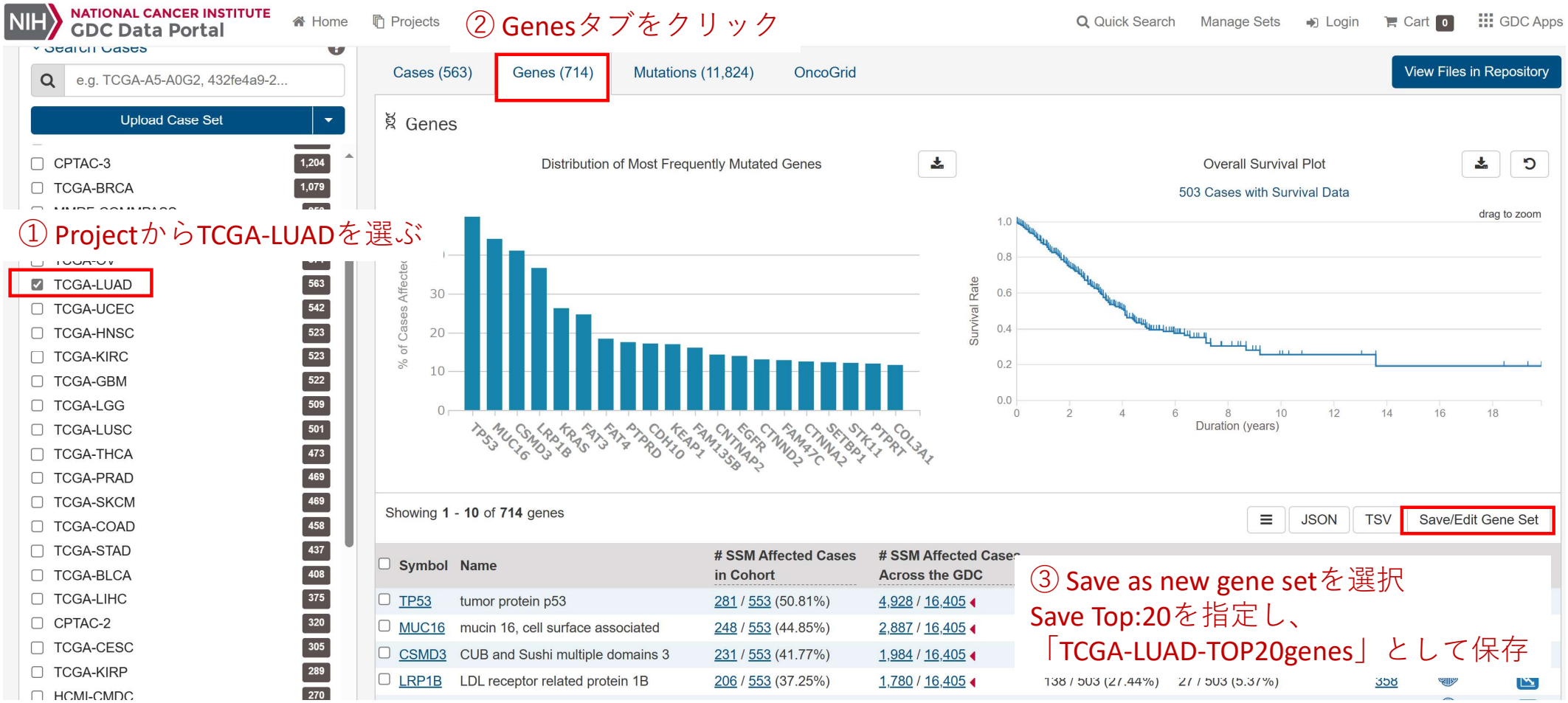


TCGA-COAD

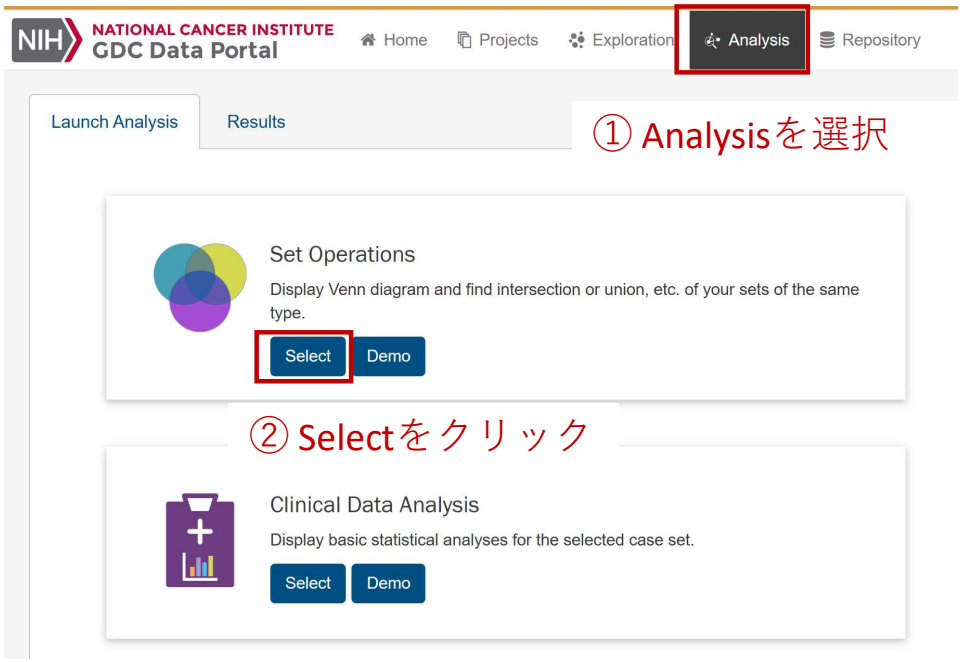


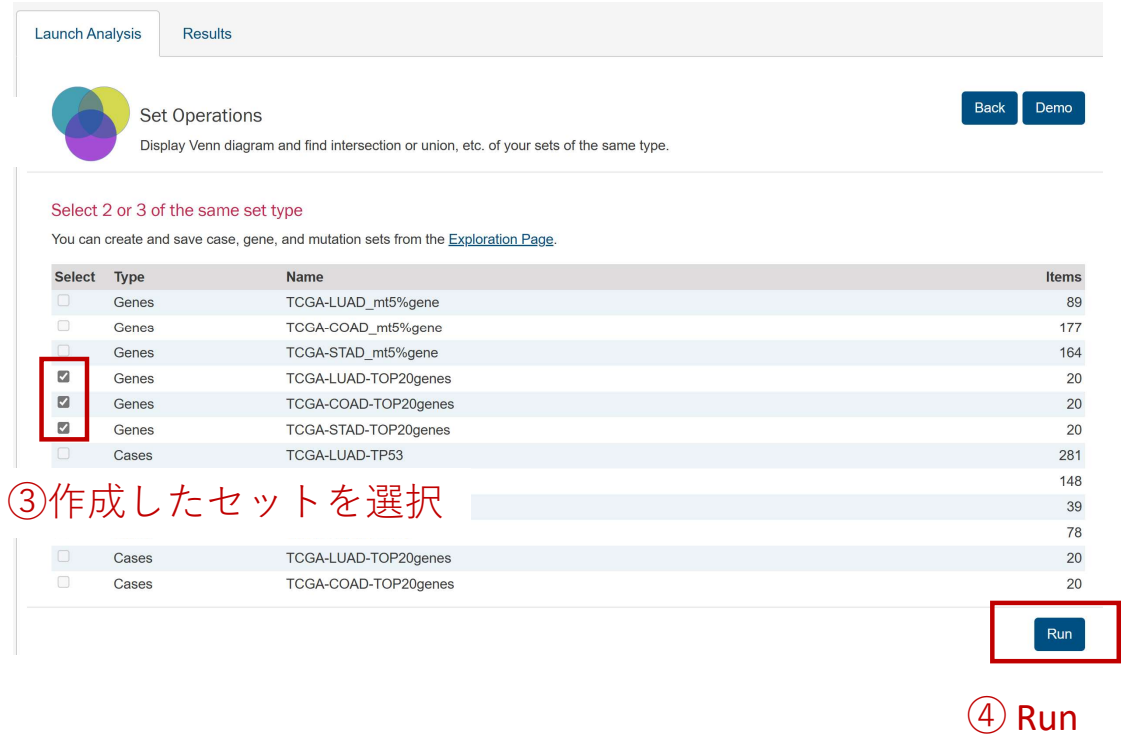
３つのがん種の変異遺伝子の集合関係を視覚化

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



TCGA-COADとTCGA-STADも同様に操作する

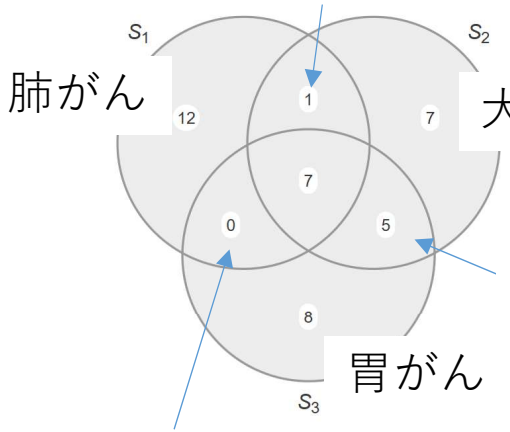




Set Operations

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

肺がんと大腸がんで共有(1)



大腸がんと胃がんで共有(5)

肺がんと胃がんで共有(0)

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
☐	(S1 ∩ S2 ∩ S3)	7	
☐	(S1 ∩ S2) - (S3)	1	
	S1)	5	
	S2)	0	
☐	(S1) - (S2 ∪ S3)	12	
☐	(S2) - (S1 ∪ S3)	7	
☐	(S3) - (S1 ∪ S2)	8	

Union of selected sets

クリックすると該当する
遺伝子が確認できる

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

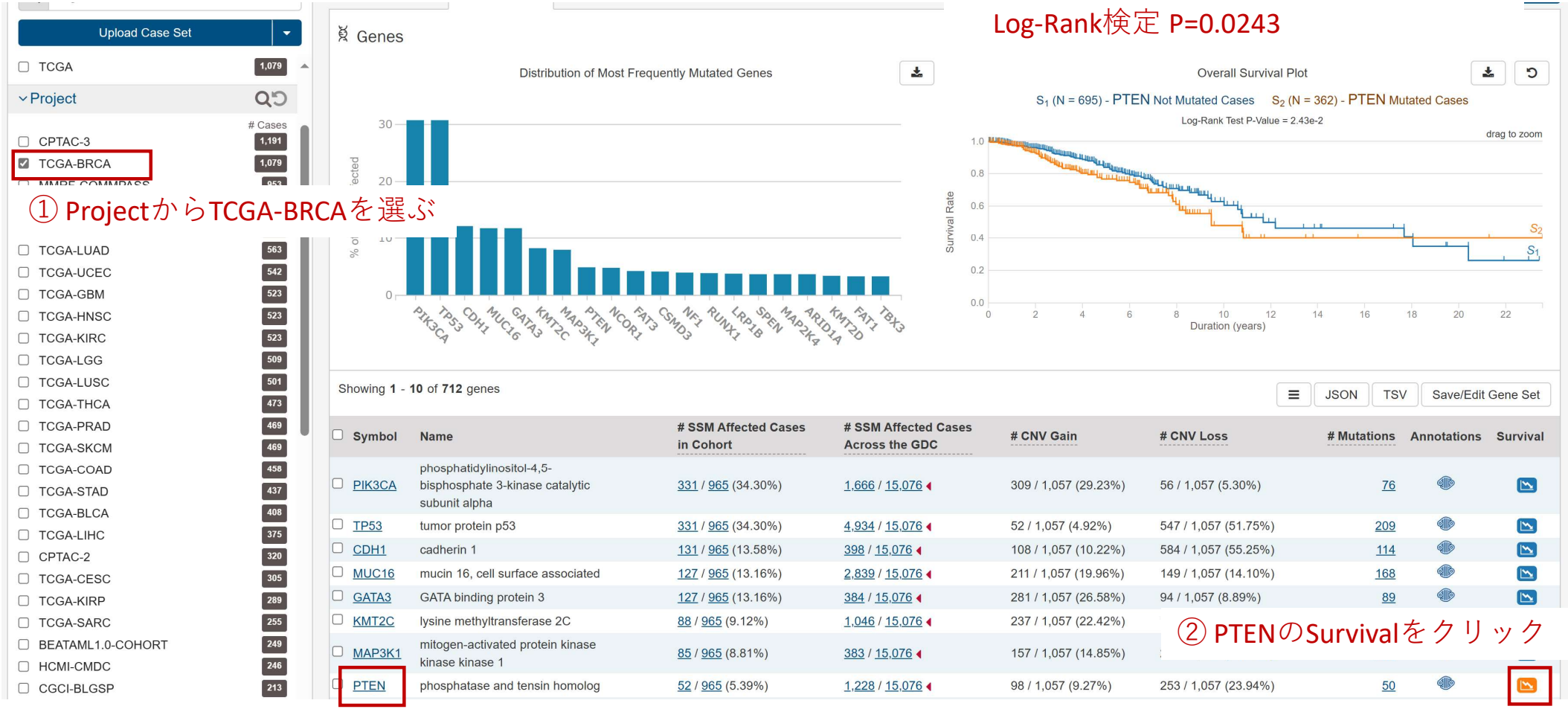
大腸がん胃がんで変異している5遺伝子： PIK3CA,KMT2D,ARID1A,NBEA,RNF43

肺がんと大腸がんで変異している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

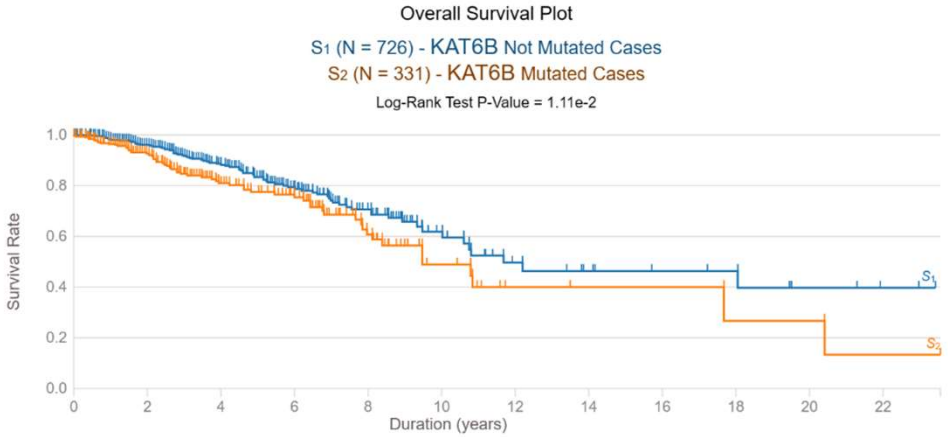


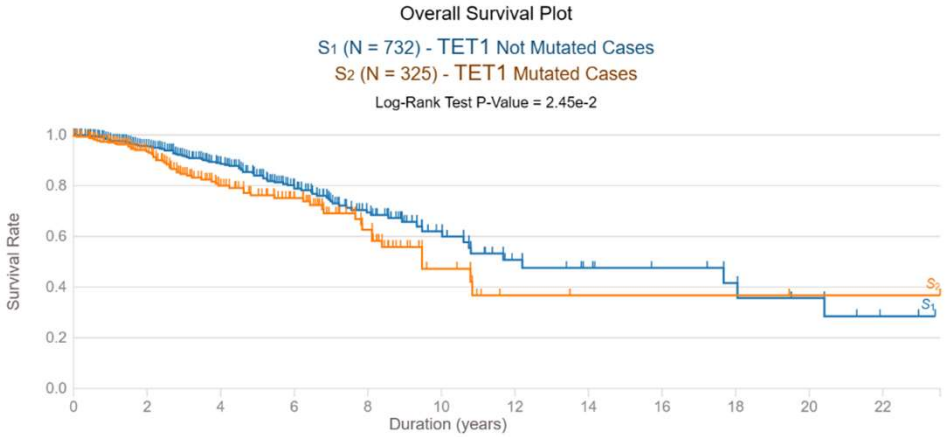
Symbol	Name	# SSM Affected Cases in Cohort	# SSM Affected Cases Across the GDC	# CNV Gain	# CNV Loss	# Mutations	Annotations	Survival
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	331 / 965 (34.30%)	1,666 / 15,076	309 / 1,057 (29.23%)	56 / 1,057 (5.30%)	76		
TP53	tumor protein p53	331 / 965 (34.30%)	4,934 / 15,076	52 / 1,057 (4.92%)	547 / 1,057 (51.75%)	209		
CDH1	cadherin 1	131 / 965 (13.58%)	398 / 15,076	108 / 1,057 (10.22%)	584 / 1,057 (55.25%)	114		
MUC16	mucin 16, cell surface associated	127 / 965 (13.16%)	2,839 / 15,076	211 / 1,057 (19.96%)	149 / 1,057 (14.10%)	168		
GATA3	GATA binding protein 3	127 / 965 (13.16%)	384 / 15,076	281 / 1,057 (26.58%)	94 / 1,057 (8.89%)	89		
KMT2C	lysine methyltransferase 2C	88 / 965 (9.12%)	1,046 / 15,076	237 / 1,057 (22.42%)				
MAP3K1	mitogen-activated protein kinase kinase kinase 1	85 / 965 (8.81%)	383 / 15,076	157 / 1,057 (14.85%)				
PTEN	phosphatase and tensin homolog	52 / 965 (5.39%)	1,228 / 15,076	98 / 1,057 (9.27%)	253 / 1,057 (23.94%)	50		

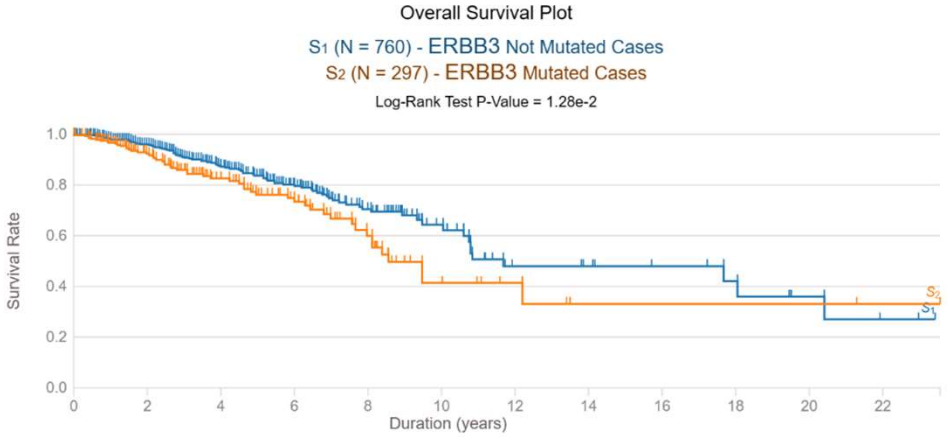
① ProjectからTCGA-BRCAを選ぶ

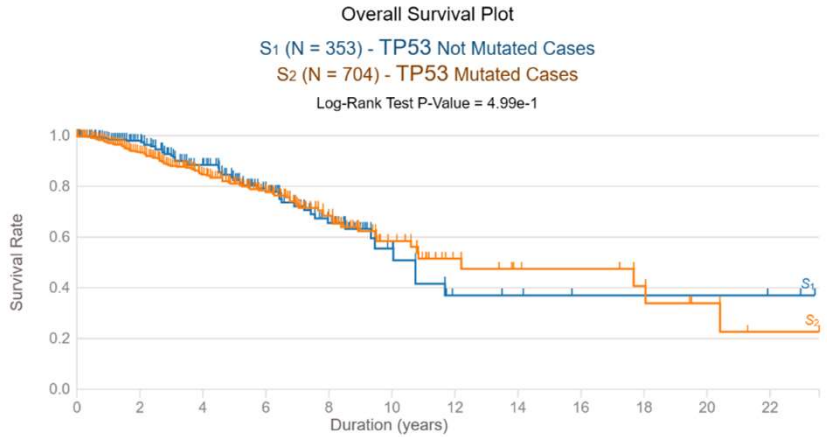
② PTENのSurvivalをクリック

他の遺伝子についても確認してみる(対象: TCGA-BRCA)

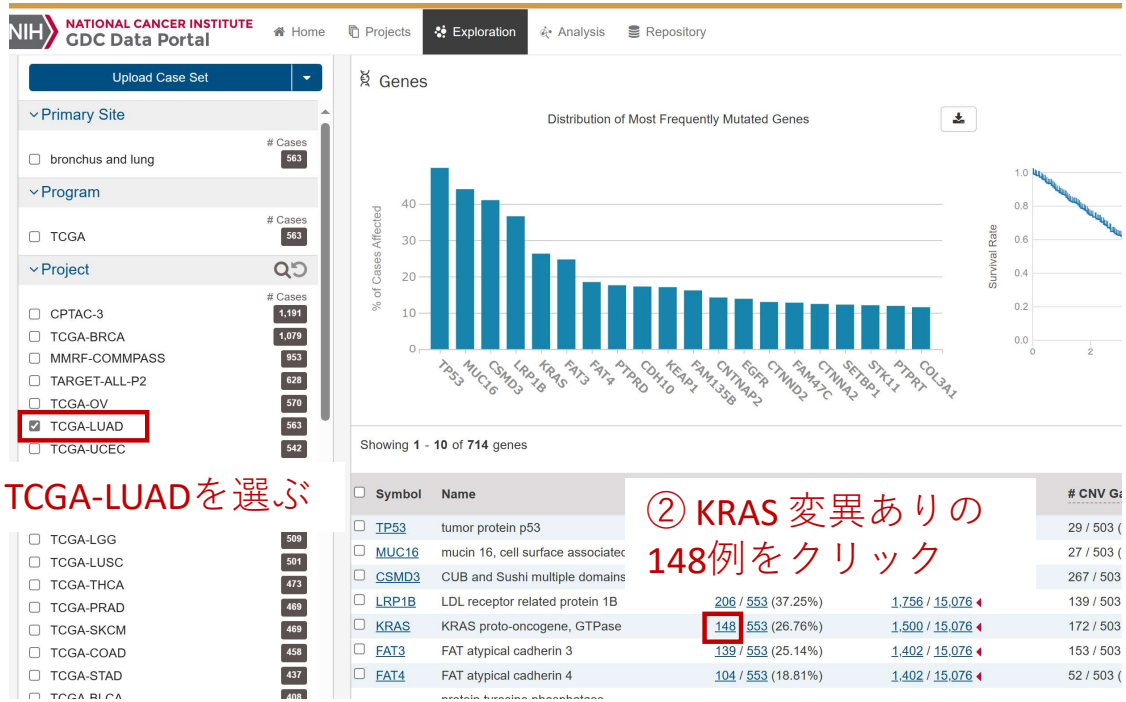






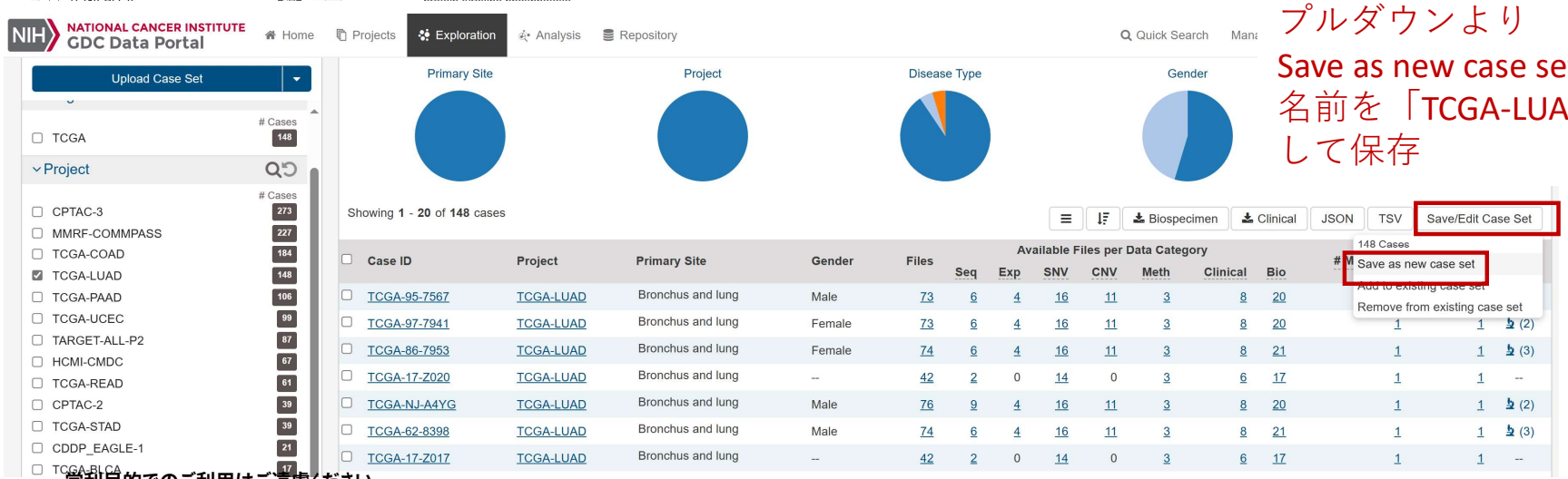


肺腺がんサンプルではKRAS/BRAF/EGFRは相互排他的に変異しているか. TCGAサンプルで確かめる



① TCGA-LUADを選ぶ

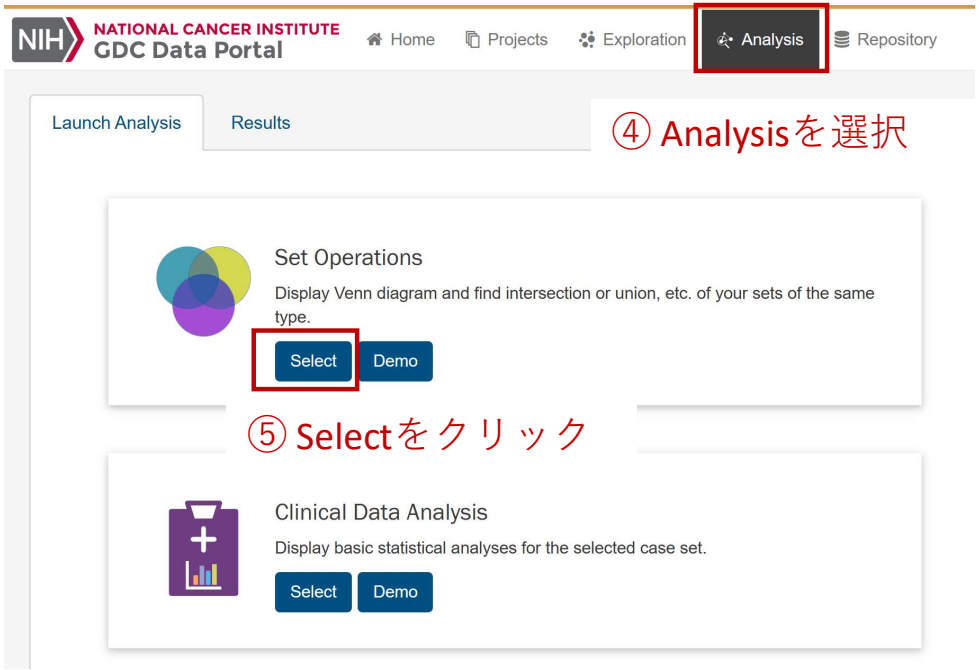
② KRAS 変異ありの148例をクリック

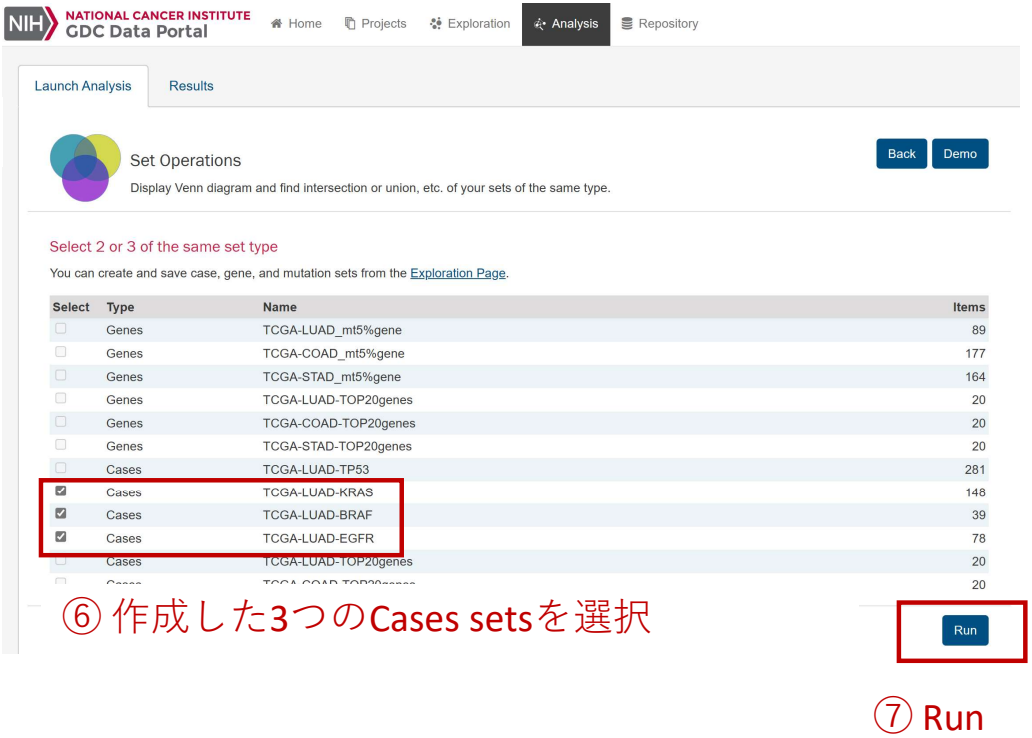


③Save/Edit Case Setをクリック
プルダウンより
Save as new case setを選択
名前を「TCGA-LUAD-KRAS」として保存

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



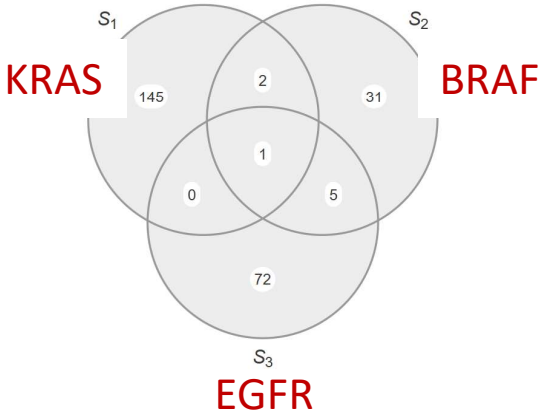


⑦ Run

Results

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-LUAD-KRAS	148
S_2	Cases	TCGA-LUAD-BRAF	39
S_3	Cases	TCGA-LUAD-EGFR	78

Select	Set Operation	# Items	Save
☐	(S1 ∩ S2 ∩ S3)	1	
☐	(S1 ∩ S2) - (S3)	2	
☐	(S2 ∩ S3) - (S1)	5	
☐	(S1 ∩ S3) - (S2)	0	
☐	(S1) - (S2 ∪ S3)	145	
☐	(S2) - (S1 ∪ S3)	31	
☐	(S3) - (S1 ∪ S2)	72	
Union of selected sets		0	

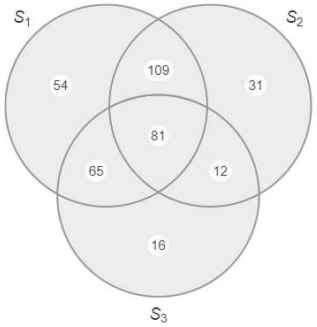
KRAS変異有のサンプル、BRAF変異有のサンプル、EGFR変有のサンプルはあまり重なっていない

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.

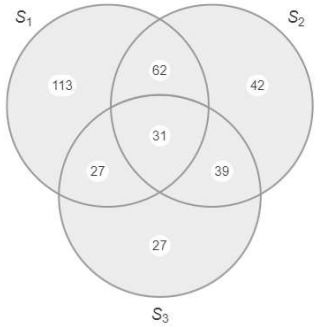


Alias	Item Type	Name	# Items
S_1	Cases	TCGA-COAD-APC	309
S_2	Cases	TCGA-COAD-TP53	233
S_3	Cases	TCGA-COAD-KRAS	174

Select	Set Operation	# Items	Save
☐	(S1 ∩ S2 ∩ S3)	81	
☐	(S1 ∩ S2) - (S3)	109	
☐	(S2 ∩ S3) - (S1)	12	
☐	(S1 ∩ S3) - (S2)	65	
☐	(S1) - (S2 ∪ S3)	54	
☐	(S2) - (S1 ∪ S3)	31	
☐	(S3) - (S1 ∪ S2)	16	
Union of selected sets		0	

Set Operations

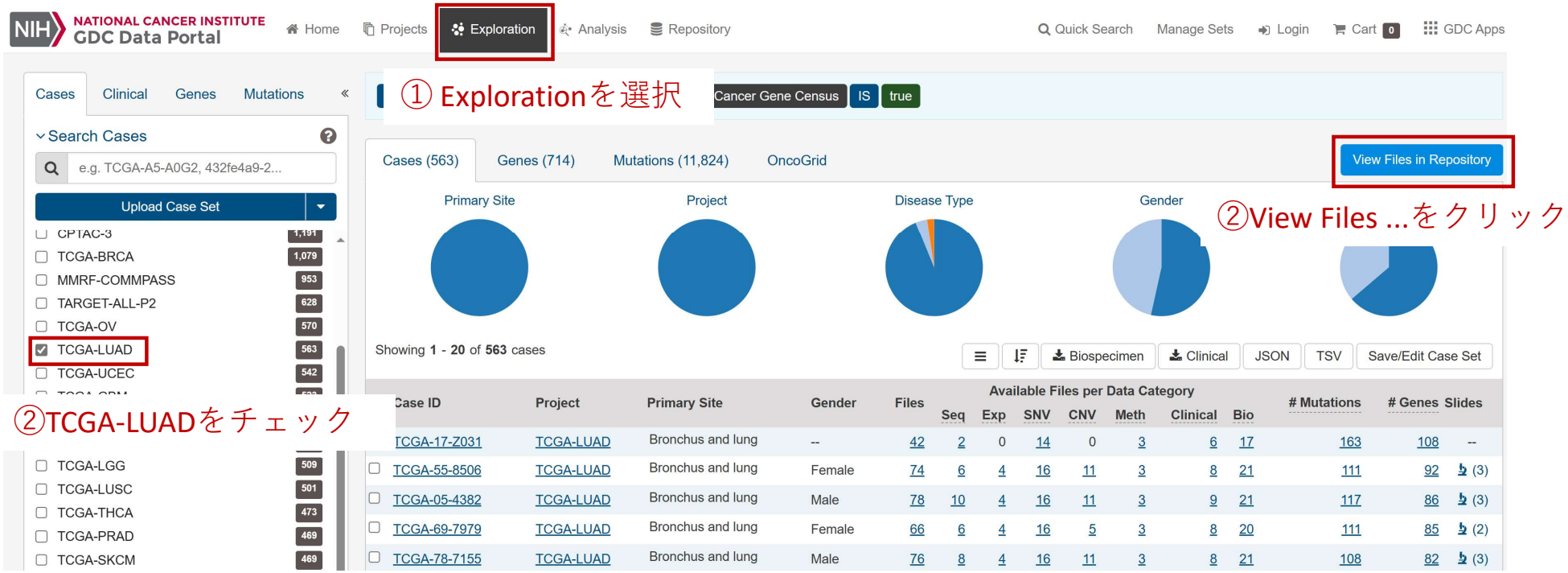
Click on the areas of the Venn diagram to include them in your result set.

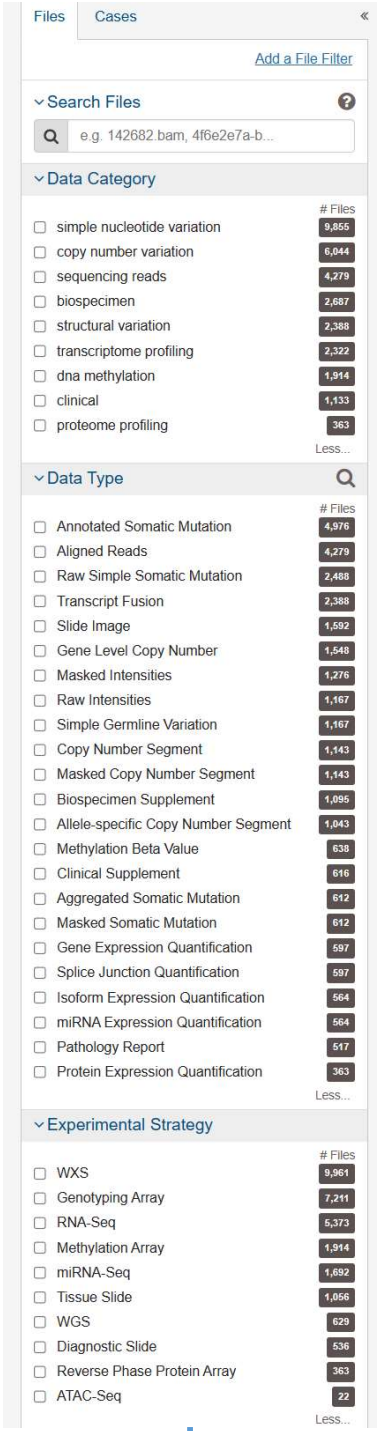


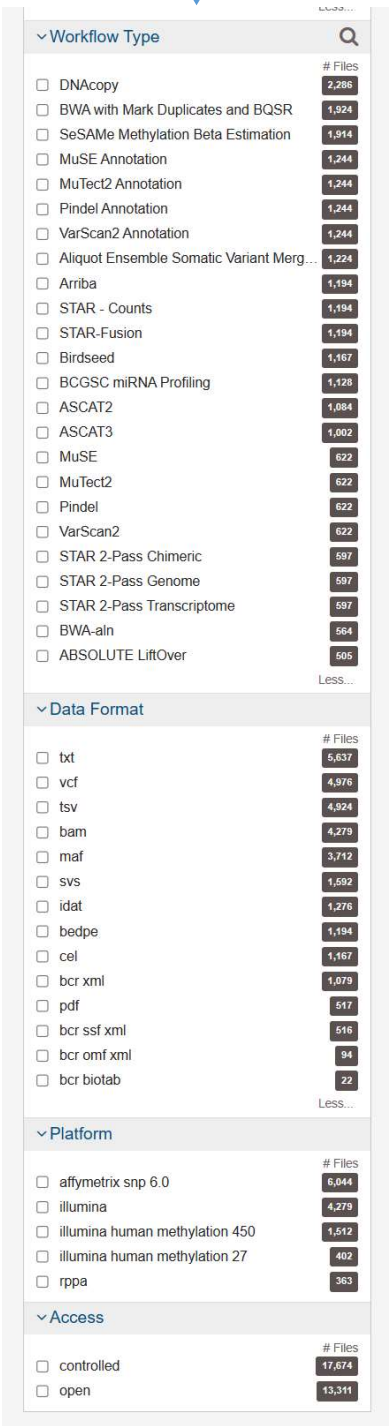
Alias	Item Type	Name	# Items
S_1	Cases	TCGA-COAD-TP53	233
S_2	Cases	TCGA-COAD-KRAS	174
S_3	Cases	TCGA-COAD-PIK3CA	124

Select	Set Operation	# Items	Save
☐	(S1 ∩ S2 ∩ S3)	31	
☐	(S1 ∩ S2) - (S3)	62	
☐	(S2 ∩ S3) - (S1)	39	
☐	(S1 ∩ S3) - (S2)	27	
☐	(S1) - (S2 ∪ S3)	113	
☐	(S2) - (S1 ∪ S3)	42	
☐	(S3) - (S1 ∪ S2)	27	
Union of selected sets		0	

TCGAデータのダウンロード







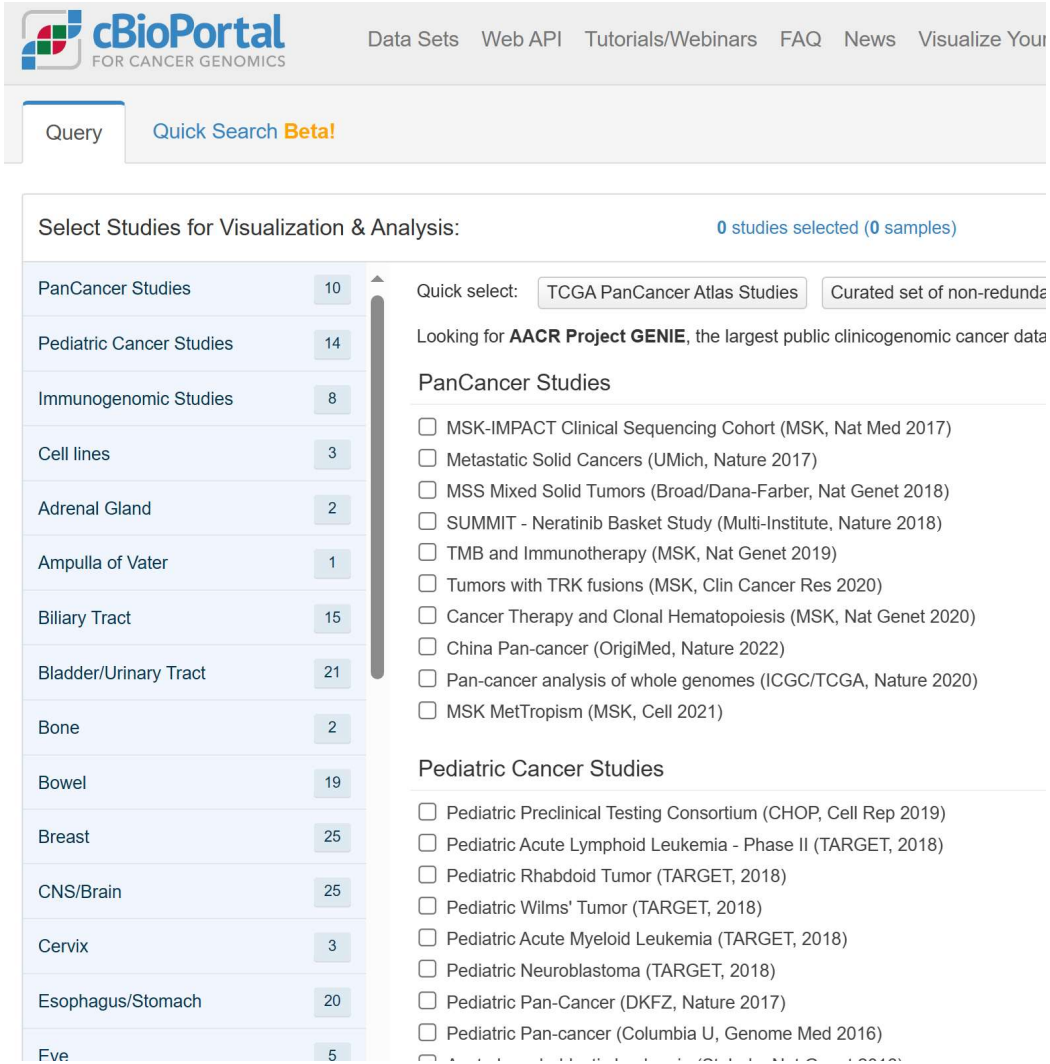
① 左側のウィンドウでデータの属性を指定

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

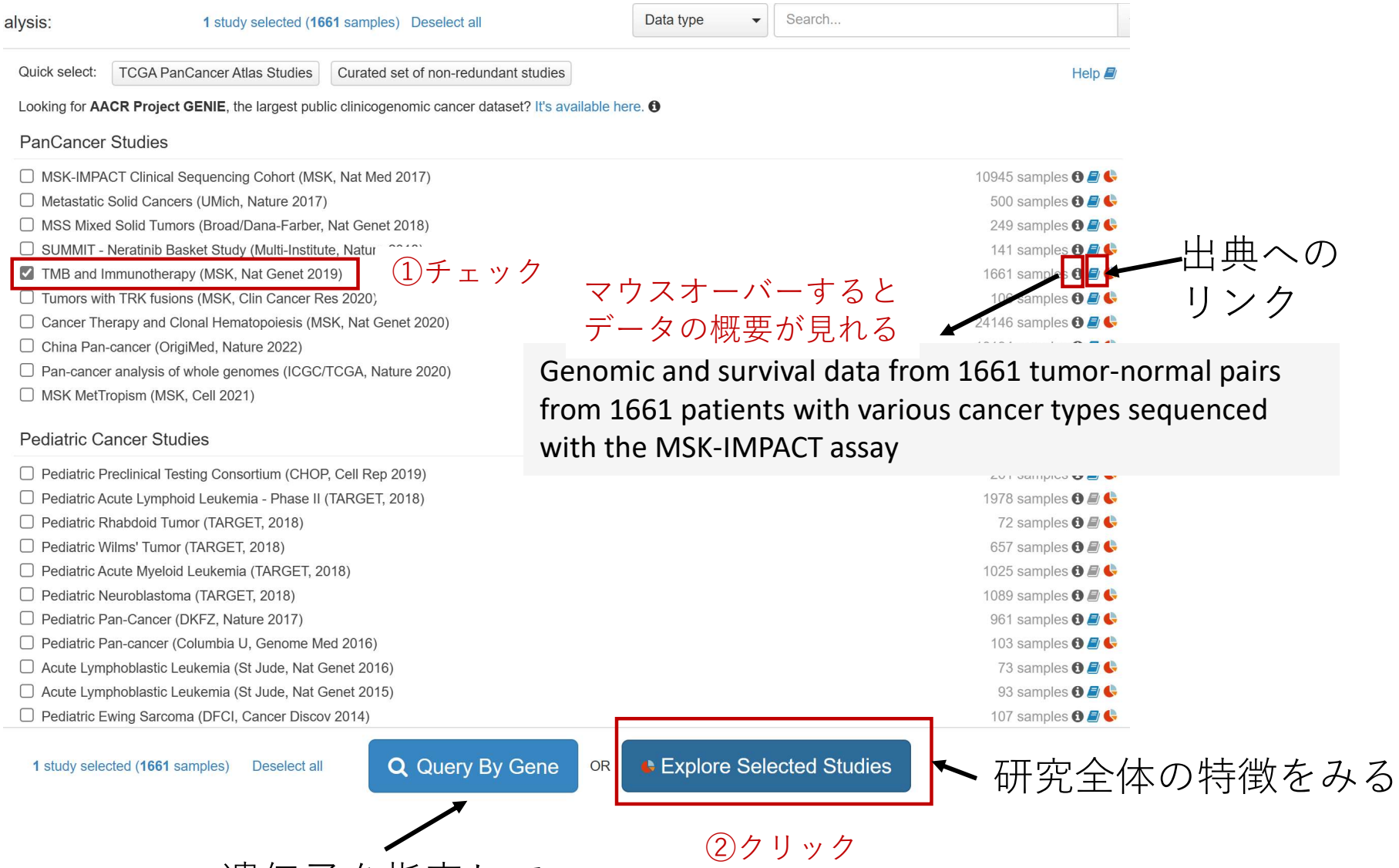


cBioPortalは、がんゲノムデータを視覚化および解釈するためのオンラインのツールおよびリソースである：

- **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



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 Shen[5], Yelena Y. Janjigian[3,4], David A. Barron[1,2], Ahmet Zehir[6], Emmet J. Jordan[2], Antonio Omuro[7], Thomas J. Kaley[7], Sviatoslav M. Kendall[2,8], Robert J. Motzer[3,4], A. Ari Hakimi[9], Martin H. Voss[3,4], Paul Russo[9], Jonathan Rosenberg[3,4], Gopa Iyer[3,4], Bernard H. Bochner[9], Dean F. Bajorin[3,4], Hikmat A. Al-Ahmadie[6], Jamie E. Chaft[3,4], Charles M. Rudin[3,4], Gregory J. Riely[3,4], Shrujal Baxi[3,4], Alan L. Ho[3,4], Richard J. Wong[9], David G. Pfister[3,4], Jedd D. Wolchok[3,4], Christopher A. Barker[1], Philip H. Gutin[9], Cameron W. Brennan[9], Viviane Tabar[9], Ingo K. Mellinghoff[8], Lisa M. DeAngelis[8], Charlotte E. Ariyan[9], Nancy Lee[1], William D. Tap[3,4], Mrinal M. Gounder[3,4], Sandra P. D'Angelo[3,4], Leonard Saltz[3,4], Zsofia 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. DeMatteo[9], Marc Ladanyi[2,6], Naiyer A. Rizvi[10], 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.

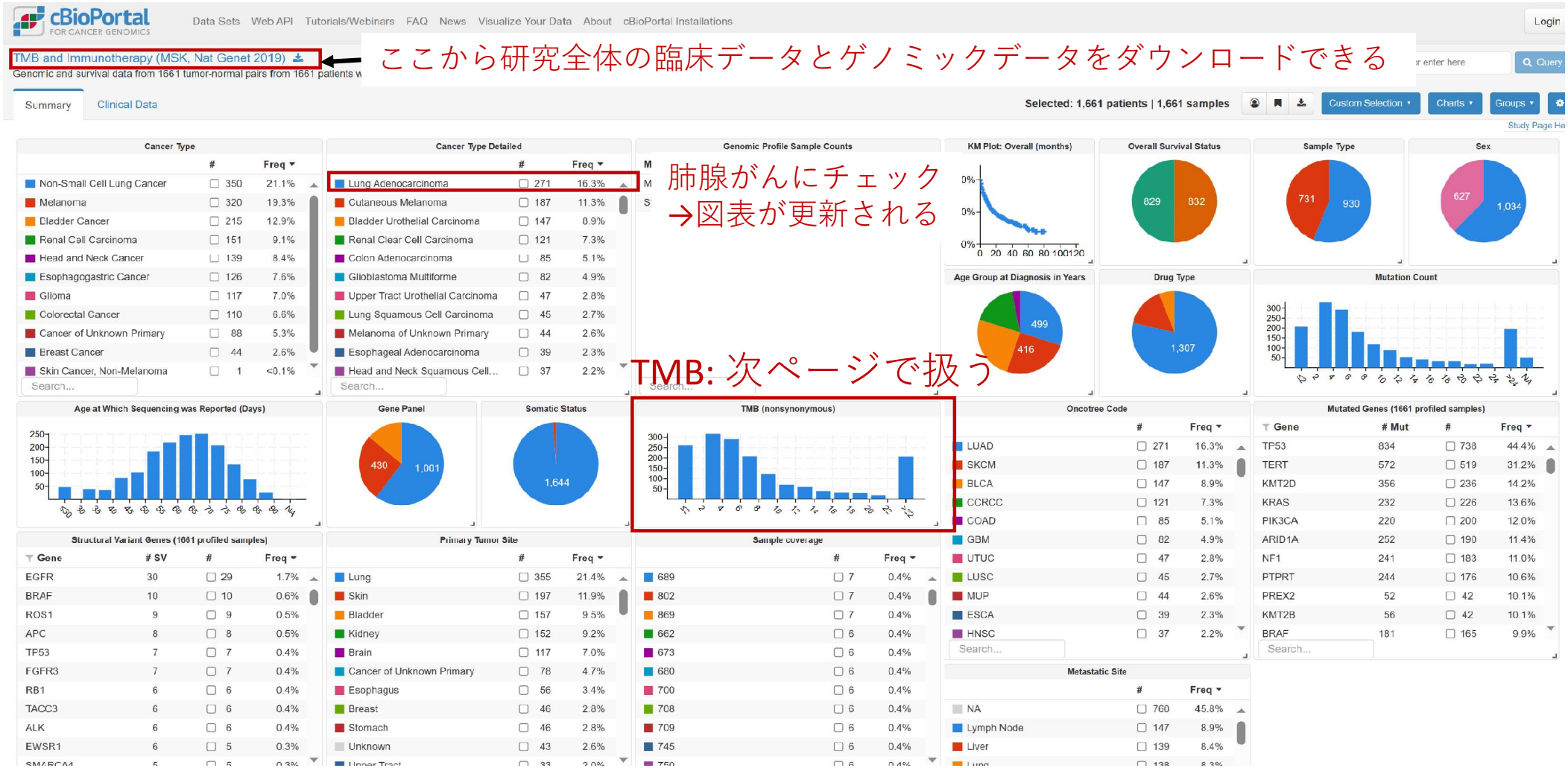
In recent years, ICI therapy has revolutionized the treatment of patients with advanced-stage cancers. These agents include antibodies that target CTLA-4 or PD-1/PD-L1[1]. Durable benefit, however, clonality, gene expression signatures and peripheral blood markers may correlate with clinical response[6]. Additionally, an association between high mutational load and clinical benefit was observed in small cohorts of patients with melanoma treated with CTLA-4 blockade[7,8], and non-small cell lung cancer (NSCLC), patients with melanoma and bladder cancer treated with PD-1/PD-L1 inhibitors[9–11]. 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) assay[12]. 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

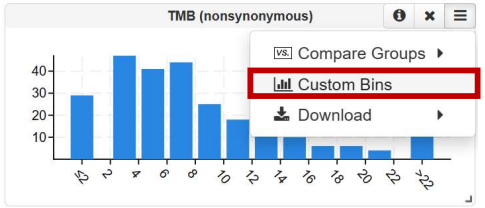
がん種、背景因子、変異遺伝子に加えて、試験独自の変数(今回TMB)の分布を見れる



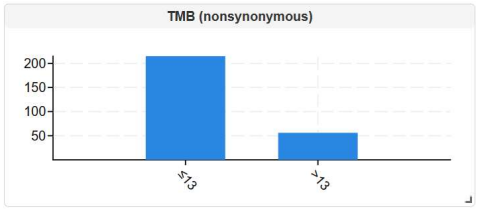
肺腺がんについて、TMBと予後の関係を見ることにする

⇒ TMB 上位20%超 vs. 20%以下　でKM-plotを作成する

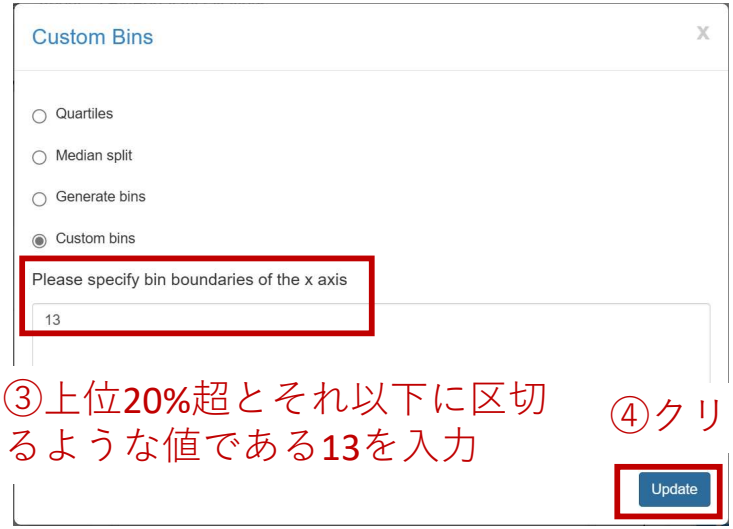
①Custom Binsをクリック



⑤ TMB ≦ 13 と >13の分布に変わる



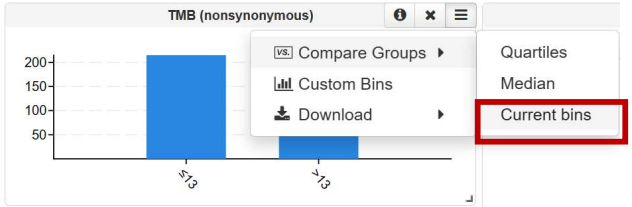
②ポップアップが出る



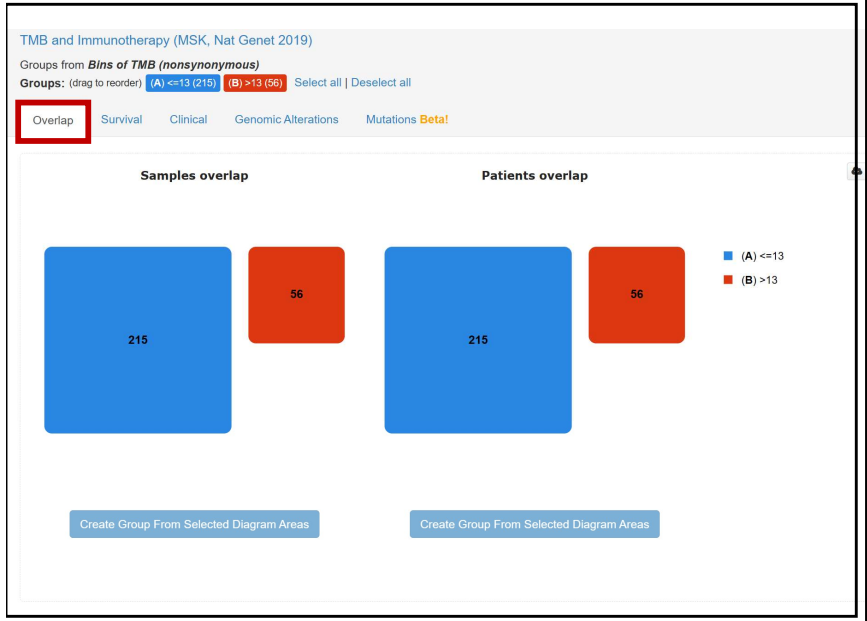
③上位20%超とそれ以下に区切るような値である13を入力

④クリック

⑥Compare Groups
→ Current binsをクリック

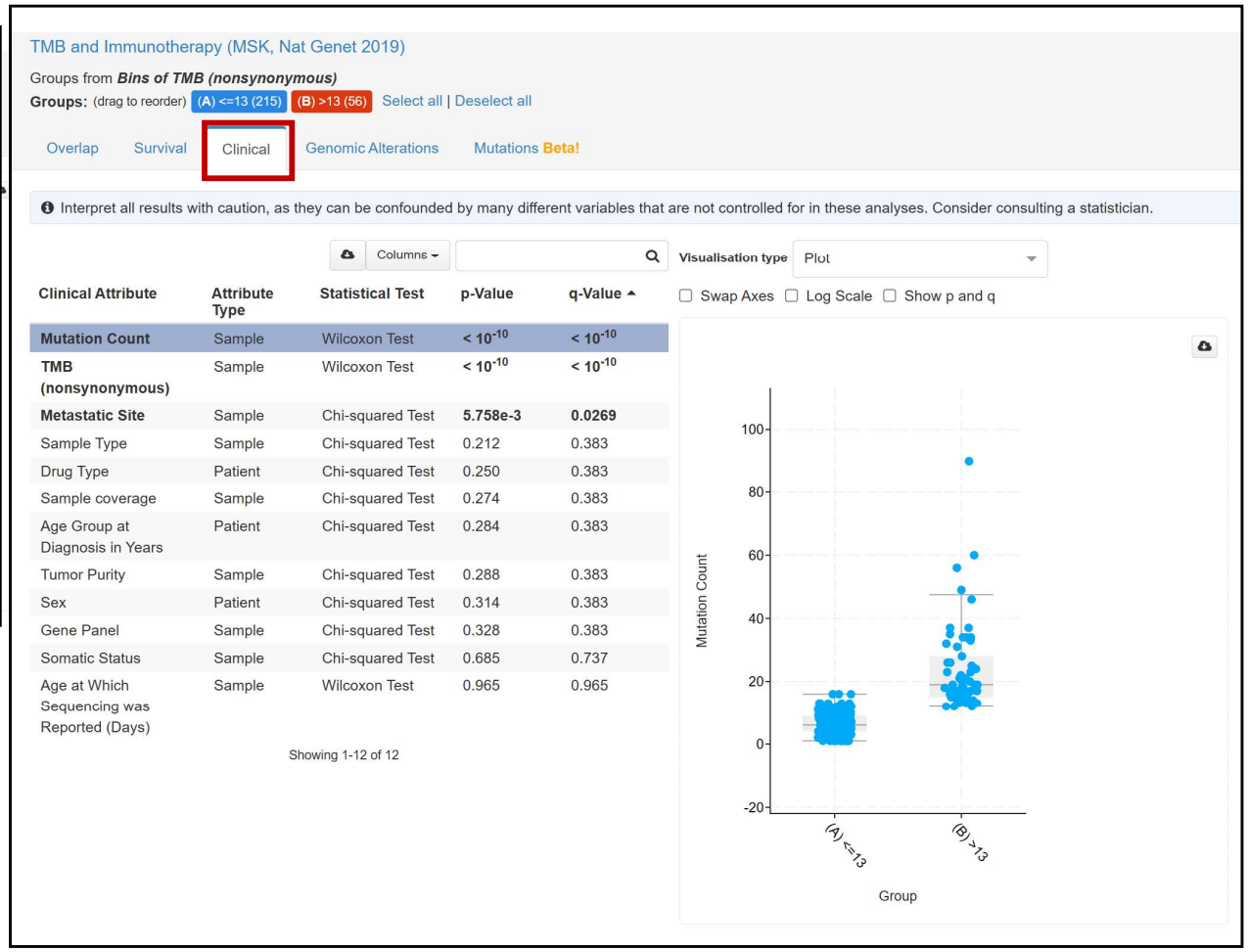


ケースの集合関係と背景の確認



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.

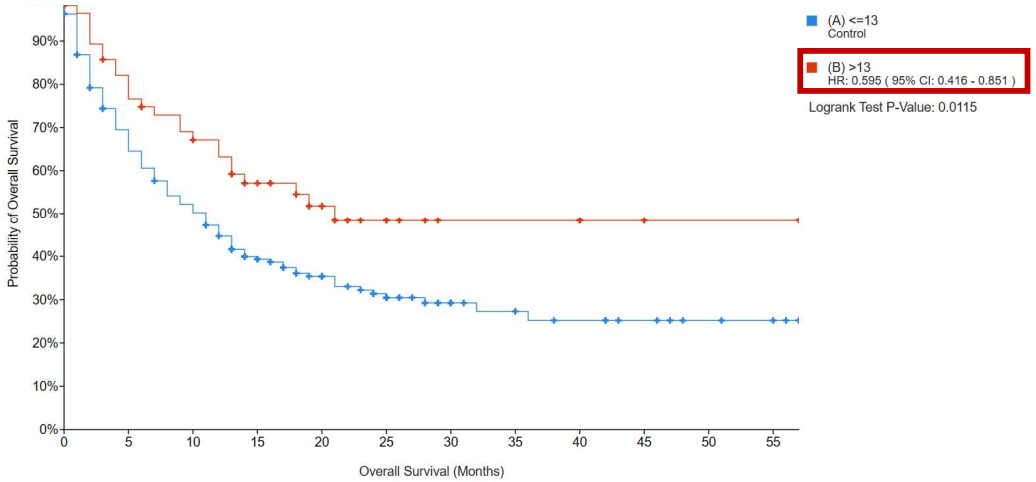
The log-rank test is used to test the null hypothesis that there is no difference between the groups in the probability of an event at any time point. Hazard ratios are derived from the log-rank test.

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)**

Nat. Genet. 2019 Fig2

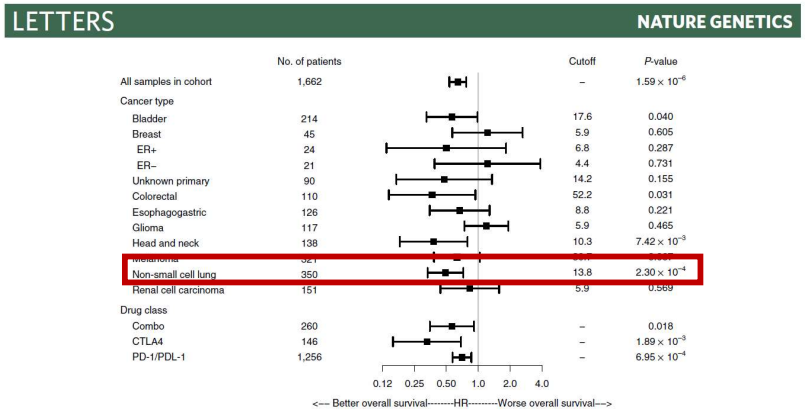
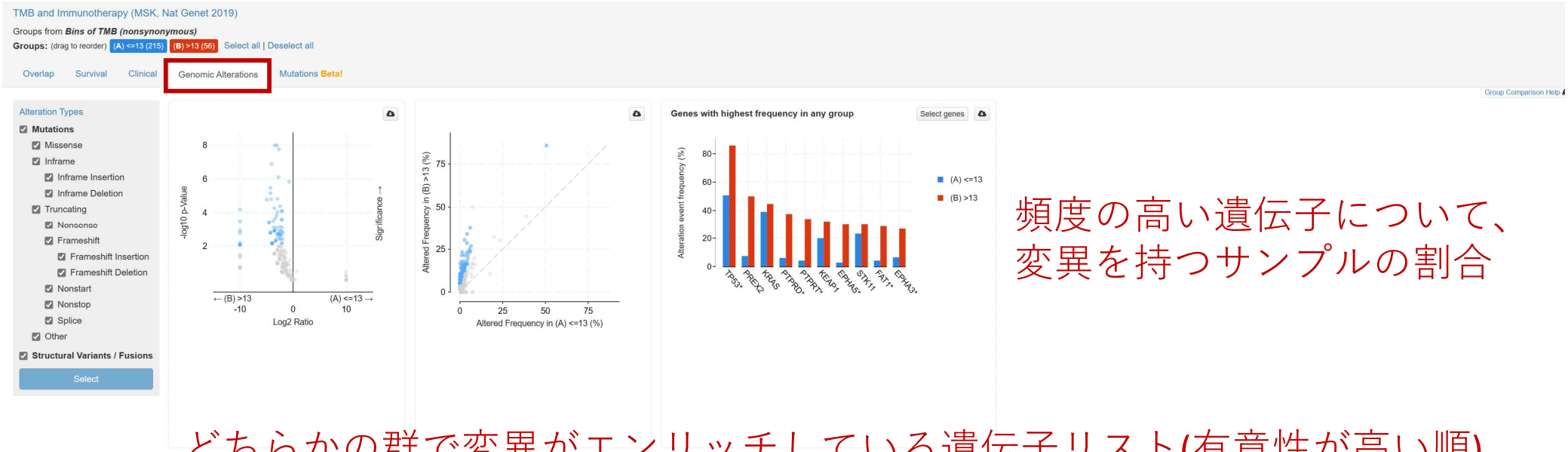


Fig. 2 | Effect of nonsynonymous mutational load on overall survival after ICI treatment, by cancer subtype and drug class. Forest plot for all patients in the identified cohort or individual cancer subtypes. Indicated are the number of patients and HR comparing overall survival after ICI in patients 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と大きく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



頻度の高い遺伝子について、
変異を持つサンプルの割合

どちらかの群で変異がエンリッチしている遺伝子リスト(有意性が高い順)

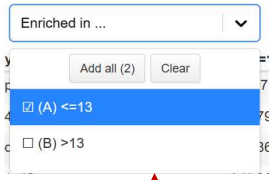
Genomic Alterations

Sample-level enrichments | Enriched in ... | Significant only | Columns

Gene	Cytoband	(A) <=13	(B) >13	Co-occurrence Pattern	Log2 Ratio	p-Value	q-Value	Enriched in
EPHA5	4q13.1-q13.2	6 (2.79%)	17 (30.36%)		-3.44	9.92e-9	2.024e-6	(B) >13
PTPRT	20q12-q13.11	9 (4.19%)	19 (33.93%)		-3.02	1.03e-8	2.024e-6	(B) >13
PTPRD	9p24.1-p23	13 (6.05%)	21 (37.50%)		-2.63	1.70e-8	2.234e-6	(B) >13
NTRK3	15q25.3	3 (1.40%)	13 (23.21%)		-4.06	1.25e-7	1.236e-5	(B) >13
FAT1	4q35.2	9 (4.19%)	16 (28.57%)		-2.77	7.70e-7	6.081e-5	(B) >13
TP53	17p13.1	109 (50.70%)	48 (85.71%)		-0.76	1.419e-6	9.342e-5	(B) >13
BRIP1	17q23.2	2 (0.93%)	10 (17.86%)		-4.26	3.329e-6	1.878e-4	(B) >13
PGR	11q22.1	2 (1.13%)	9 (21.43%)		-4.25	6.684e-6	3.300e-4	(B) >13

※ 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

Genomic Alterations

Sample-level enrichments | Enriched in ... | ☐ Significant only | Columns

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

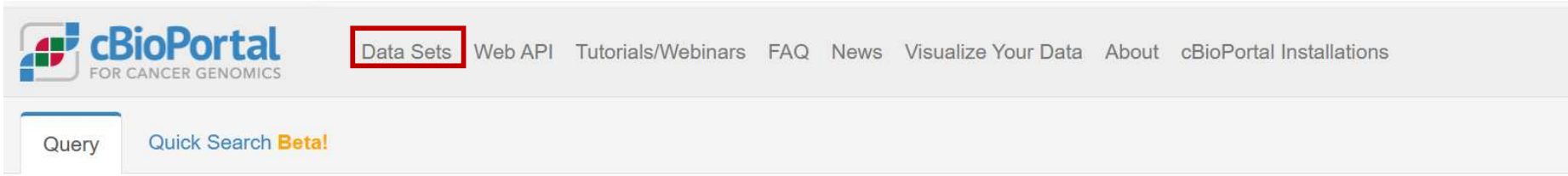
どちらかの群で変異がエンリッチしている変異リスト(有意性が高い順)

127 Mutations (page 1 of 6)

Enriched in ... Significant only Columns

Protein Change	Annotation	Mutation Type	(A) <=13	(B) >13	Co-occurrence Pattern	Log2 Ratio	p-Value	q-Value	Enriched in
*P75Lfs*48*		FS del	0 (0.00%)	3 (5.36%)		<-10	8.450e-3	0.772	(B) >13
S241F		Missense	0 (0.00%)	2 (3.57%)		<-10	0.0421	0.772	(B) >13
H179R		Missense	1 (0.47%)	2 (3.57%)		-2.94	0.109	0.772	(B) >13
S215I		Missense	1 (0.47%)	2 (3.57%)		-2.94	0.109	0.772	(B) >13
Y234C		Missense	0 (0.00%)	1 (1.79%)		<-10	0.207	0.772	(B) >13
*E339**		Nonsense	0 (0.00%)	1 (1.79%)		<-10	0.207	0.772	(B) >13
R280G		Missense	0 (0.00%)	1 (1.79%)		<-10	0.207	0.772	(B) >13

全ての研究のダウンロードページ



cBioPortal FOR CANCER GENOMICS | Data Sets | Web API | Tutorials/Webinars | FAQ | News | Visualize Your Data | About | cBioPortal Installations | Login

Datasets

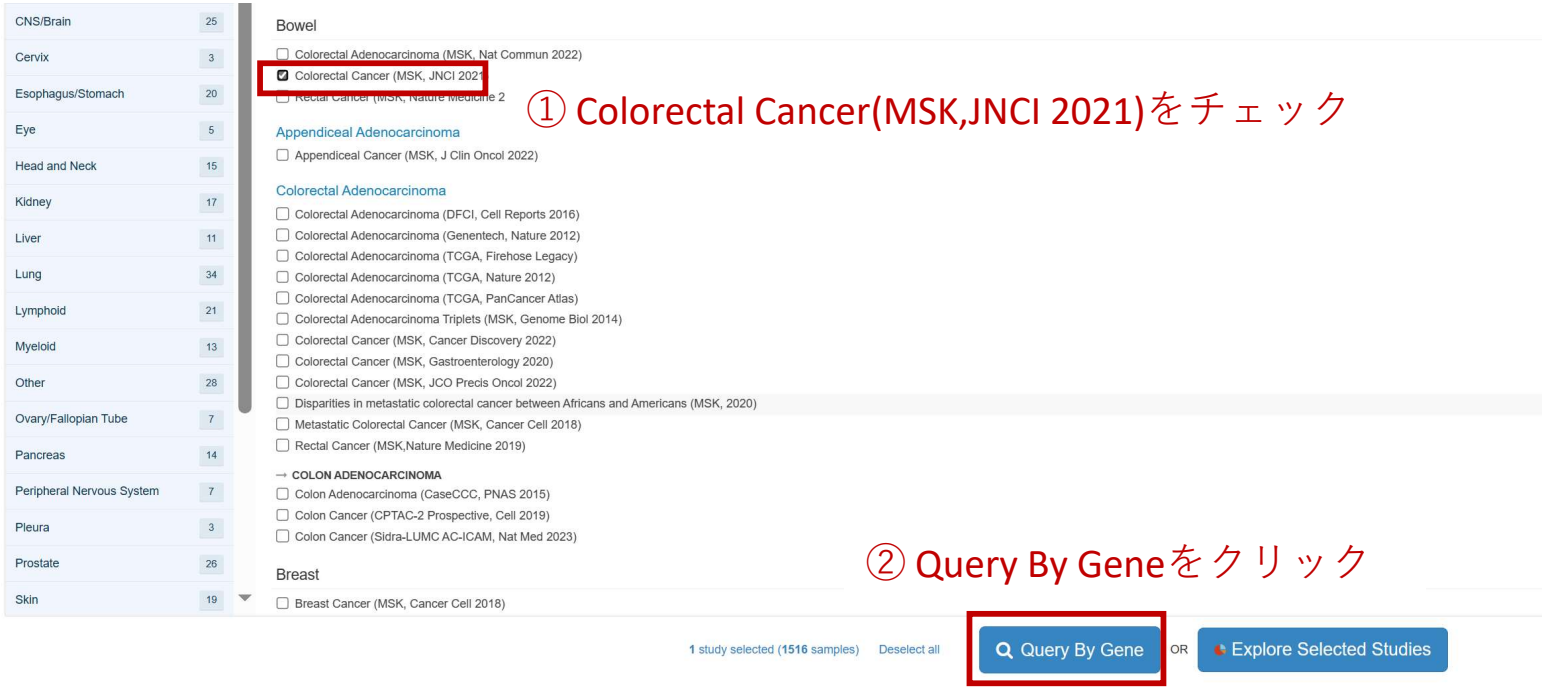
The table below lists the number of available samples per cancer study and data type. It also provides links to download the data for each study. For alternative ways of downloading, see the Download Documentation.

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)	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)	TCGA, Cell 2018	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

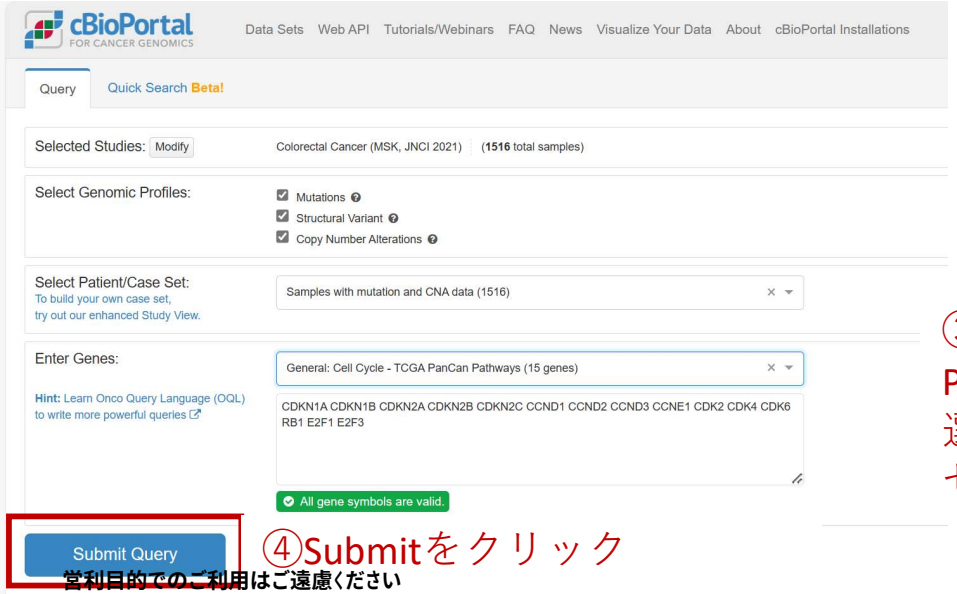
特定遺伝子を指定して情報検索(Query By Gene)

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



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

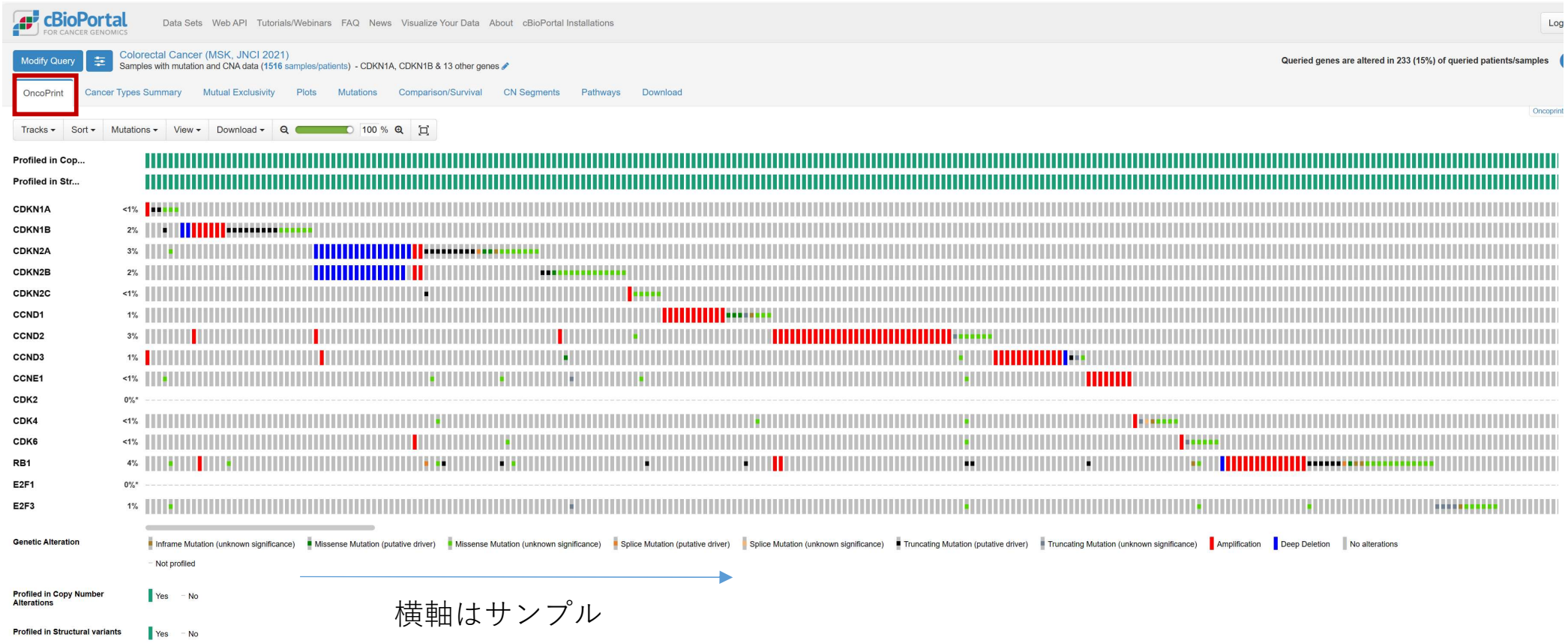
② Query By Geneをクリック



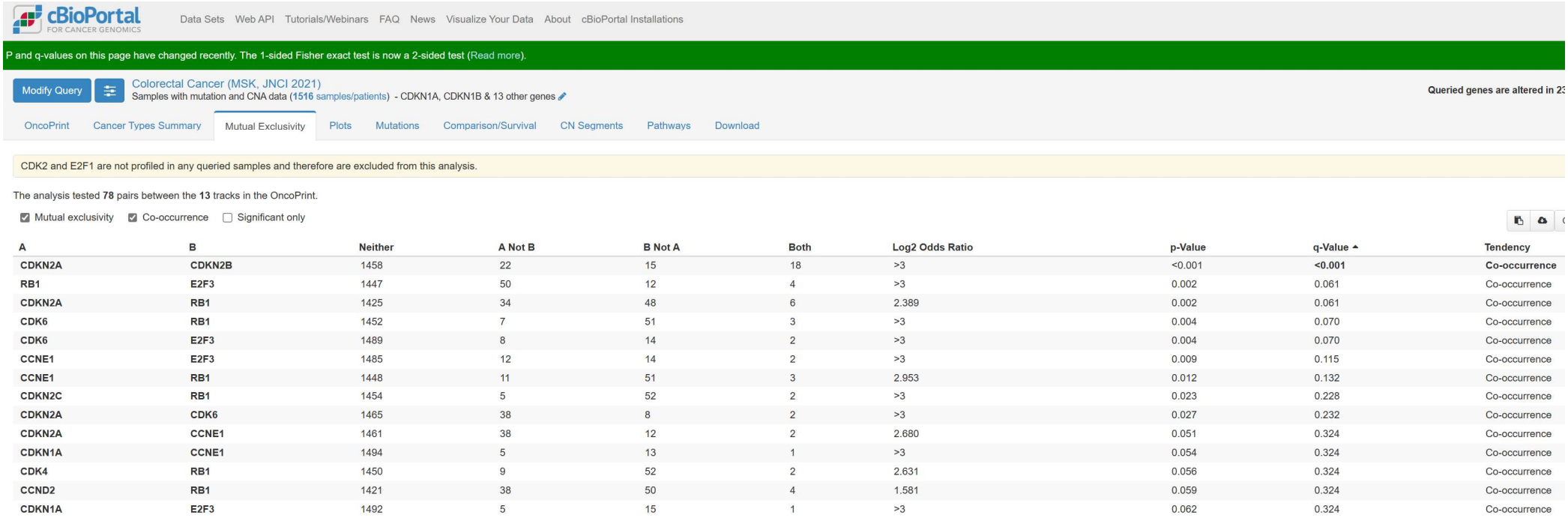
③ General: Cell Cycle- TCGA PanCan Pathways(15 genes)を選択⇒15遺伝子がボックスにセットされる

④Submitをクリック

15遺伝子のOncoPrint



15遺伝子の各ペアのMutual Exclusivity(相互排他性)



cBioPortal FOR CANCER GENOMICS

Data Sets Web API Tutorials/Webinars FAQ News Visualize Your Data About cBioPortal Installations

P and q-values on this page have changed recently. The 1-sided Fisher exact test is now a 2-sided test (Read more).

Modify Query | Colorectal Cancer (MSK, JNCI 2021)
Samples with mutation and CNA data (1516 samples/patients) - CDKN1A, CDKN1B & 13 other genes

Queried genes are altered in 23

OncoPrint | Cancer Types Summary | Mutual Exclusivity | Plots | Mutations | Comparison/Survival | CN Segments | Pathways | Download

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.

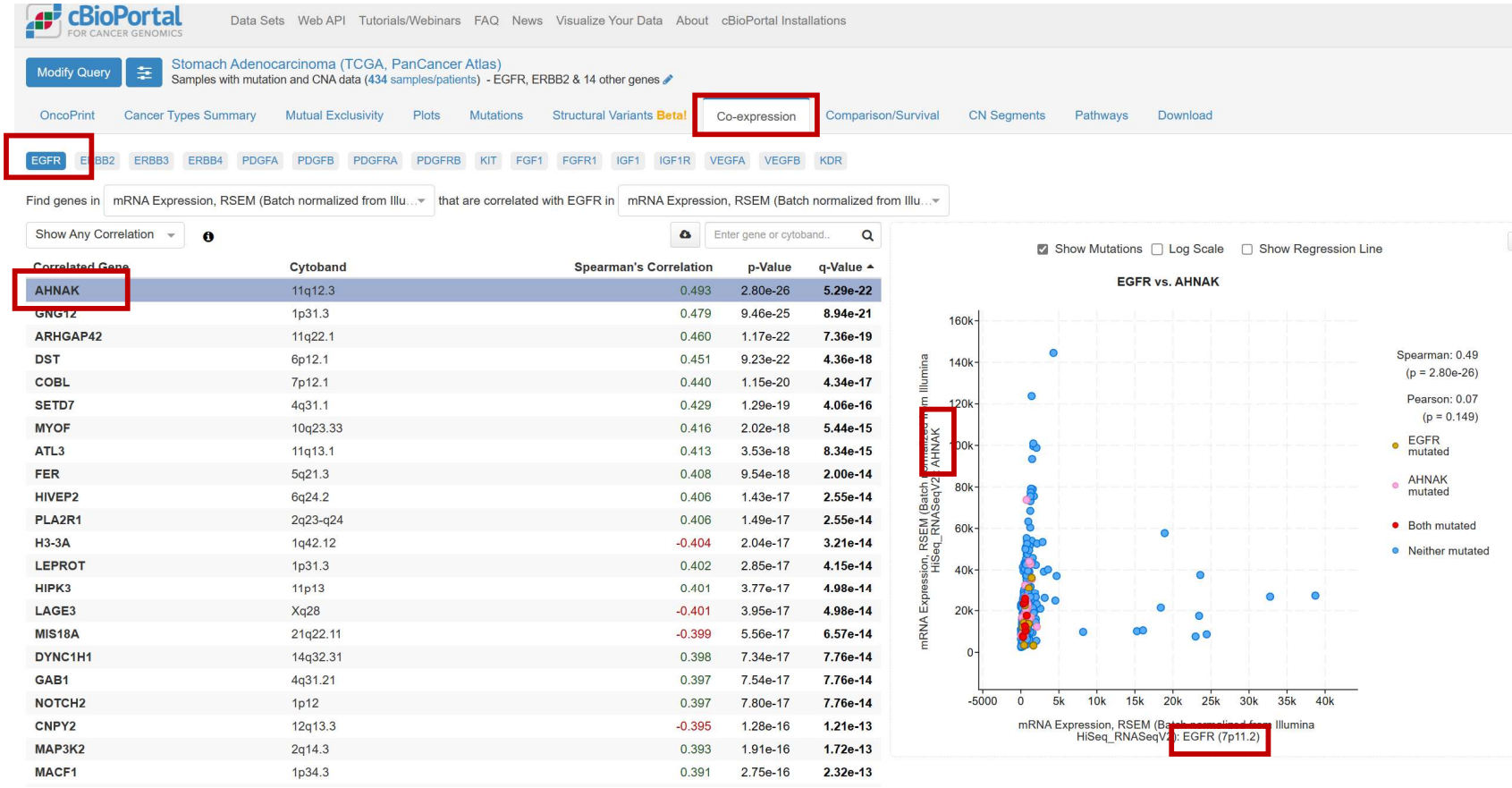
☑ Mutual exclusivity ☑ Co-occurrence ☐ Significant only

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	1465	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

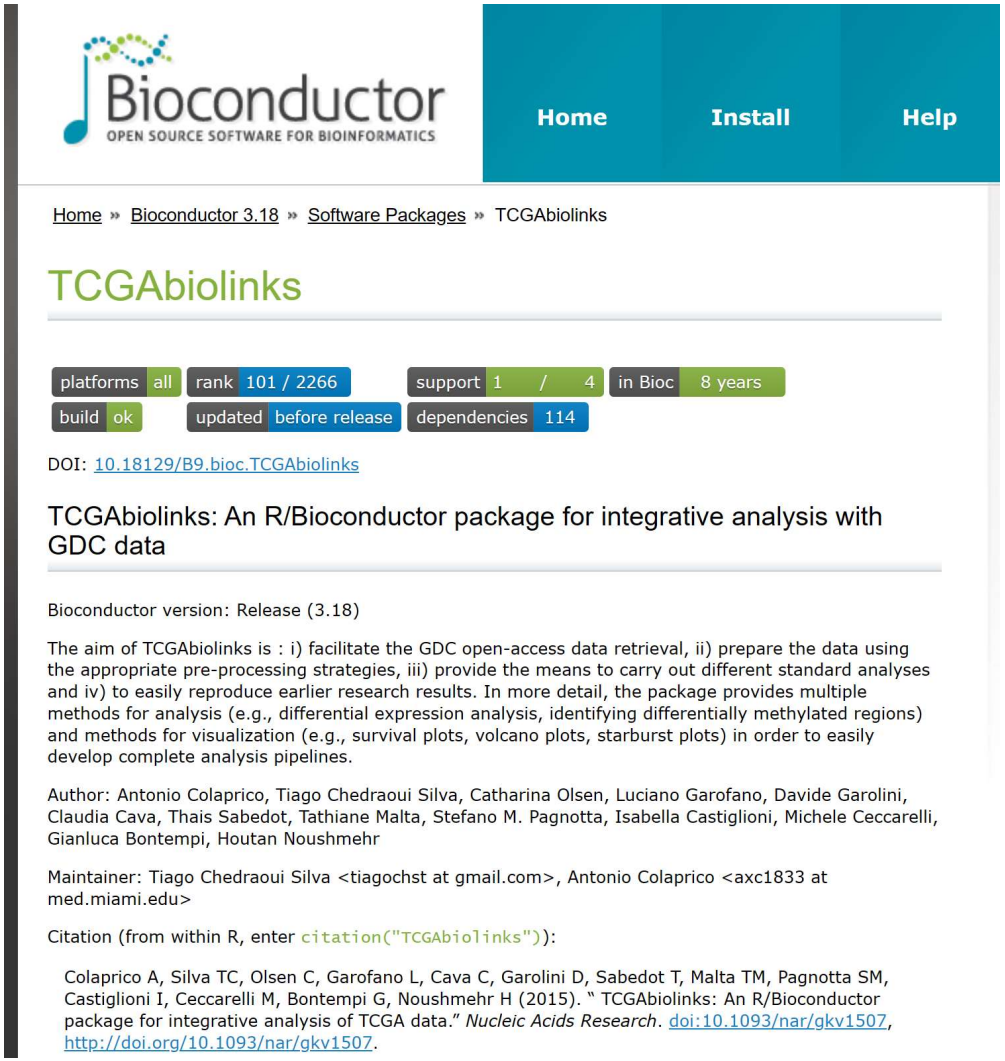
CDKN2AとCDKN2Bはq-value<0.001でCo-occurrence (共起)

※ 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

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

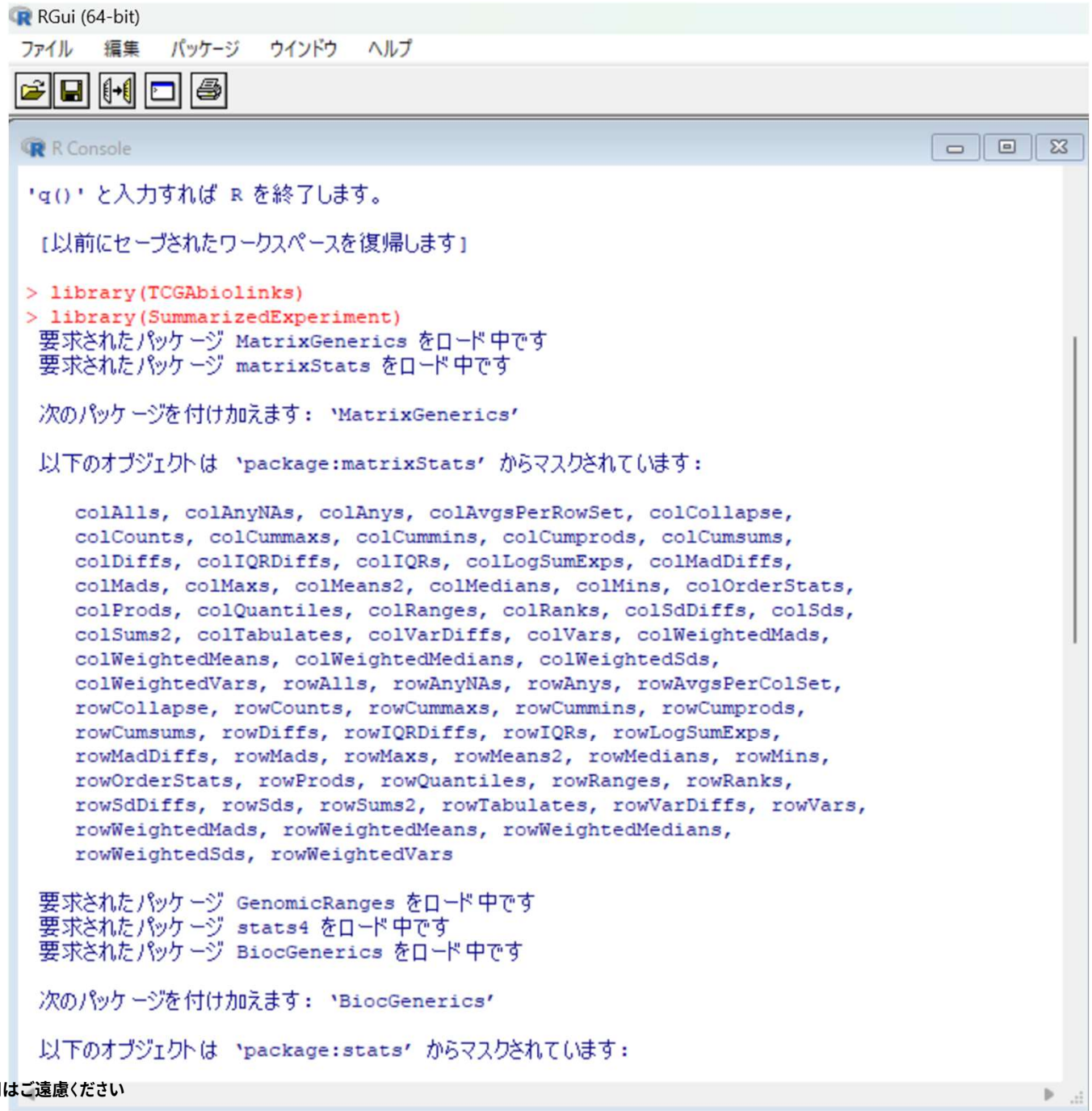
発現データからヒートマップを描く

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



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()

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....
11                   OHSU-CNL Hematopo....
12                  REBC-THYR Thyroid ....
13              TARGET-ALL-P1 Hematopo....
14              MMRF-COMMPASS Hematopo....
15                TARGET-CCSK       Kidney
16       ORGANOID-PANCREATIC      Pancreas
17               NCICCR-DLBCL  Lymph nodes
18                 TARGET-NBL Meninges....
19                  TCGA-CHOL Other an....
20                  TARGET-OS Bones, j....
21                 TARGET-AML Unknown,....
22                  TARGET-RT Lip, Liv....
23                  TARGET-WT       Kidney
24                  TCGA-SARC Meninges....
25                  TCGA-PCPG Other an....
26                  TCGA-COAD Rectosig....
27                   TCGA-ACC Adrenal ....
28                 WCDT-MCRPC Prostate....
29                  TCGA-UCEC Corpus u...
```

```
51                TCGA-THYM Thymic E....
52           VAREPOP-APOLLO Squamous....
53                 TCGA-UCS Complex ....
54                TCGA-SKCM Nevi and....
55                 TRIO-CRU Not Appl....
56                TCGA-HNSC Squamous....
57                TCGA-PAAD Ductal a....
58                TCGA-TGCT Germ Cel....
59                TCGA-CESC Cystic, ....
60                TCGA-ESCA Cystic, ....
61                TCGA-THCA Epitheli....
62                 TCGA-LGG      Gliomas
63                TCGA-LIHC Adenomas....
64                TCGA-PRAD Ductal a....
65                TCGA-READ Cystic, ....
66                  MATCH-I Gliomas,....
67                  MATCH-W Gliomas,....
68                  MATCH-B Transiti....
69                  MATCH-H Adenomas....
70                   TCGA-OV Not Repo....
71                 TCGA-UVM Nevi and....
72               MATCH-Z1A Mesothel....
73                  MATCH-U Mesothel....
74       BEATAML1.0-COHORT Myeloid ....
75                TCGA-BLCA Adenomas....
76                TCGA-LUSC Squamous....
77               CGCI-BLGSP Mature B....
78                HCMI-CMDC Soft Tis....
```

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")

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

```
> 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-VD-A8KE-01A-11R-A40B-13
6  152fdff5-c442-4384-a19a-8ccbe124ae20 2018-09-06T20:37:37.991443-05:00 50363092-20b0-4b7e
7  2373d6d8-d98b-4cc3-a5ee-a8da3bbde2                    data_type    state experime
8  fc0fe5f2-4d8c-4dcc-85c7-010e7e3e09 1    Splice Junction Quantification released
9  665251c9-f45f-4c4f-ad2a-a9fc7e17cb 2    Gene Expression Quantification released
10 ed79c017-e771-4455-89cb-1bbfb8a5da 3    Splice Junction Quantification released
11 ab9762e1-dc39-4dca-9262-e1e3cff5b8 4  Isoform Expression Quantification released
12 d3791987-53b8-444e-97c2-bc77f51613 5  Isoform Expression Quantification released
13 1d970f75-011a-4fc6-a151-1ad62e016d 6    miRNA Expression Quantification released
14 a75a4d9a-266e-4a0b-835a-6f49a25a86 7  Isoform Expression Quantification released
15 0747c448-d280-40bd-ab2a-08a6daa8f9 8    Splice Junction Quantification released
16 eb1ba0fe-22eb-48b8-b51f-6f2640cbd3 9    miRNA Expression Quantification released
17 f84b24e5-751f-4b17-a6dd-74b046844d 10 Isoform Expression Quantification released
18 b8059d49-e273-4c75-81e1-a805d0503c 11   miRNA Expression Quantification released
19 654d9d43-a561-45f9-bf51-620eb11a3a 12   miRNA Expression Quantification released
20 50363092-20b0-4b7e-bbe1-06344d13c9 13   Gene Expression Quantification released
       access                         14   Splice Junction Quantification released
                                      15 Isoform Expression Quantification released
                                      16 Isoform Expression Quantification released
                                      17   Gene Expression Quantification released
                                      18   Splice Junction Quantification released
                                      19   Splice Junction Quantification released
                                      20   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()

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-a151-1ad62e016c35         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  d148f481-a38c-481e-b0c0-cdad04cbf3ba         TSV TCGA-VD-A8KH-01A-11R-A405-07
7  b7ac31c0-1284
8  7a3399f3-a7c4
9  3ec5342f-740b
10 cb6cee1a-43a4
11 e1849623-895a
12 e77bc8dd-c55c
13 3c44e284-32e2
14 e34a4ad6-6c02
15 d48060d6-bea4
16 2c941f66-90ae
17 2e2874d4-4932
18 b5b0f524-d70e
19 d79c97d7-9fcd
20 a7ff797b-4601
   access
1    open
2    open
3    open
```

```
                           data_type    state experimental_strategy version
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
  Gene Expression Quantification released               RNA-Seq       1
0 Gene Expression Quantification released               RNA-Seq       1
1 Gene Expression Quantification released               RNA-Seq       1
2 Gene Expression Quantification released               RNA-Seq       1
3 Gene Expression Quantification released               RNA-Seq       1
4 Gene Expression Quantification released               RNA-Seq       1
5 Gene Expression Quantification released               RNA-Seq       1
6 Gene Expression Quantification released               RNA-Seq       1
7 Gene Expression Quantification released               RNA-Seq       1
8 Gene Expression Quantification released               RNA-Seq       1
9 Gene Expression Quantification released               RNA-Seq       1
0 Gene Expression Quantification released               RNA-Seq
```

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")
```

#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: 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")
```

UVM_mat2 <- UVM_mat[1:100,]

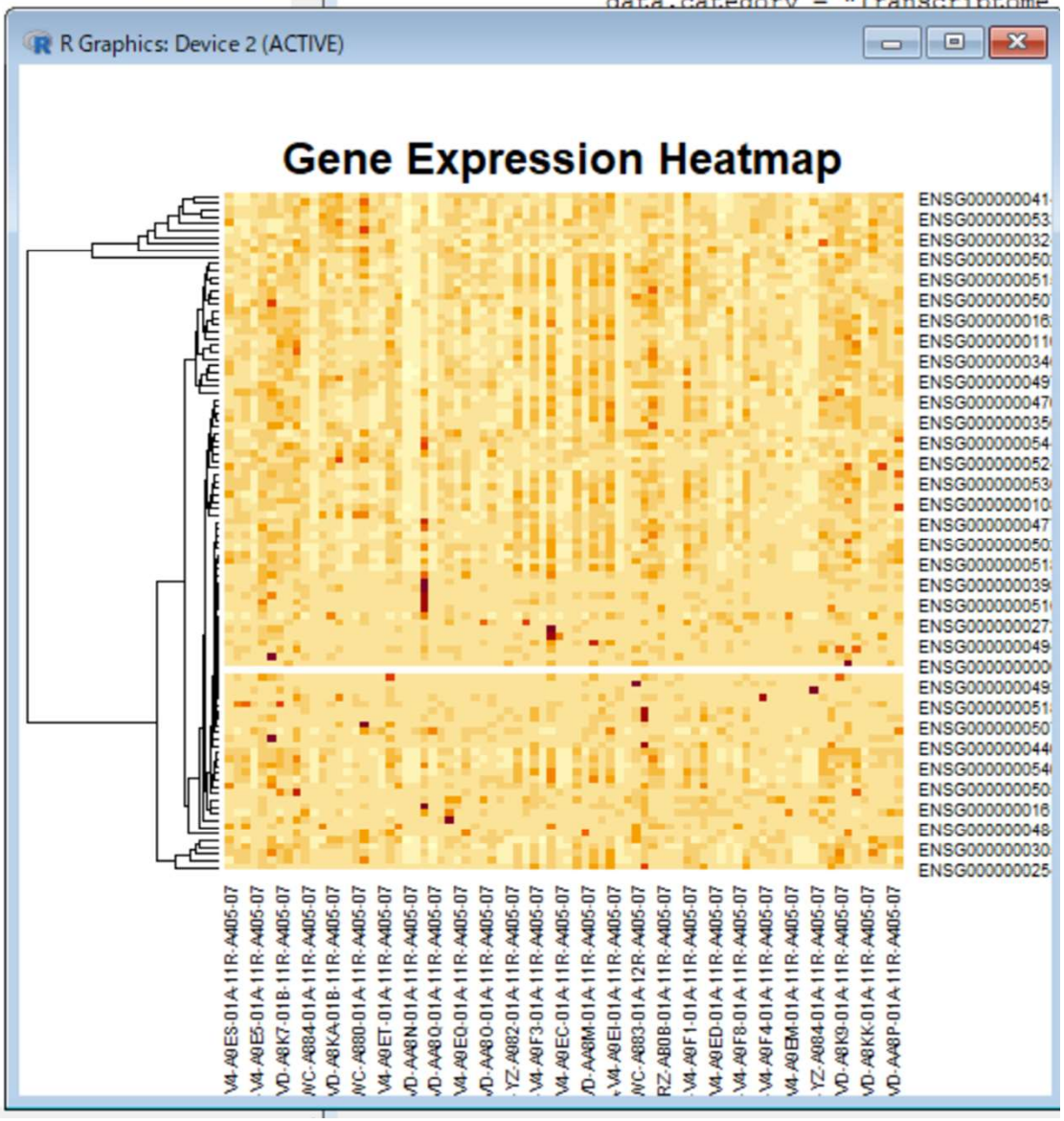
```
>
> UVM_mat2 <- UVM_mat[1:100,]
>
```

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: 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")
```

ヒートマップ

```
> 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

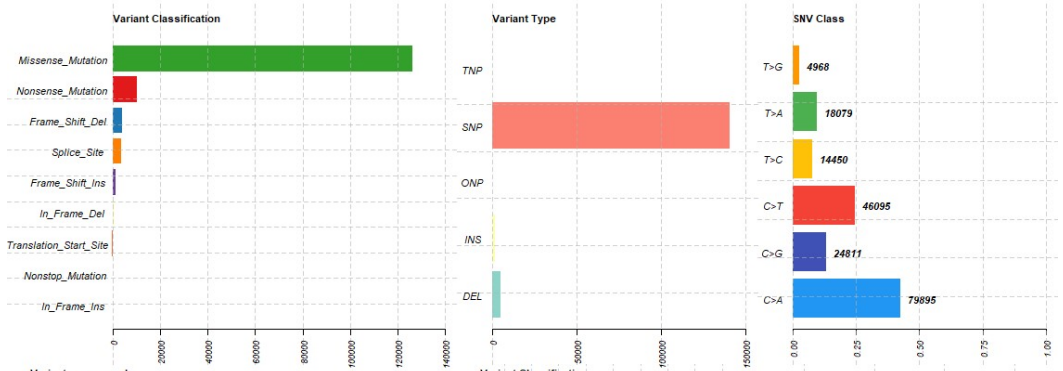
```
library(TCGAbiolinks)
library(maftools)

# 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

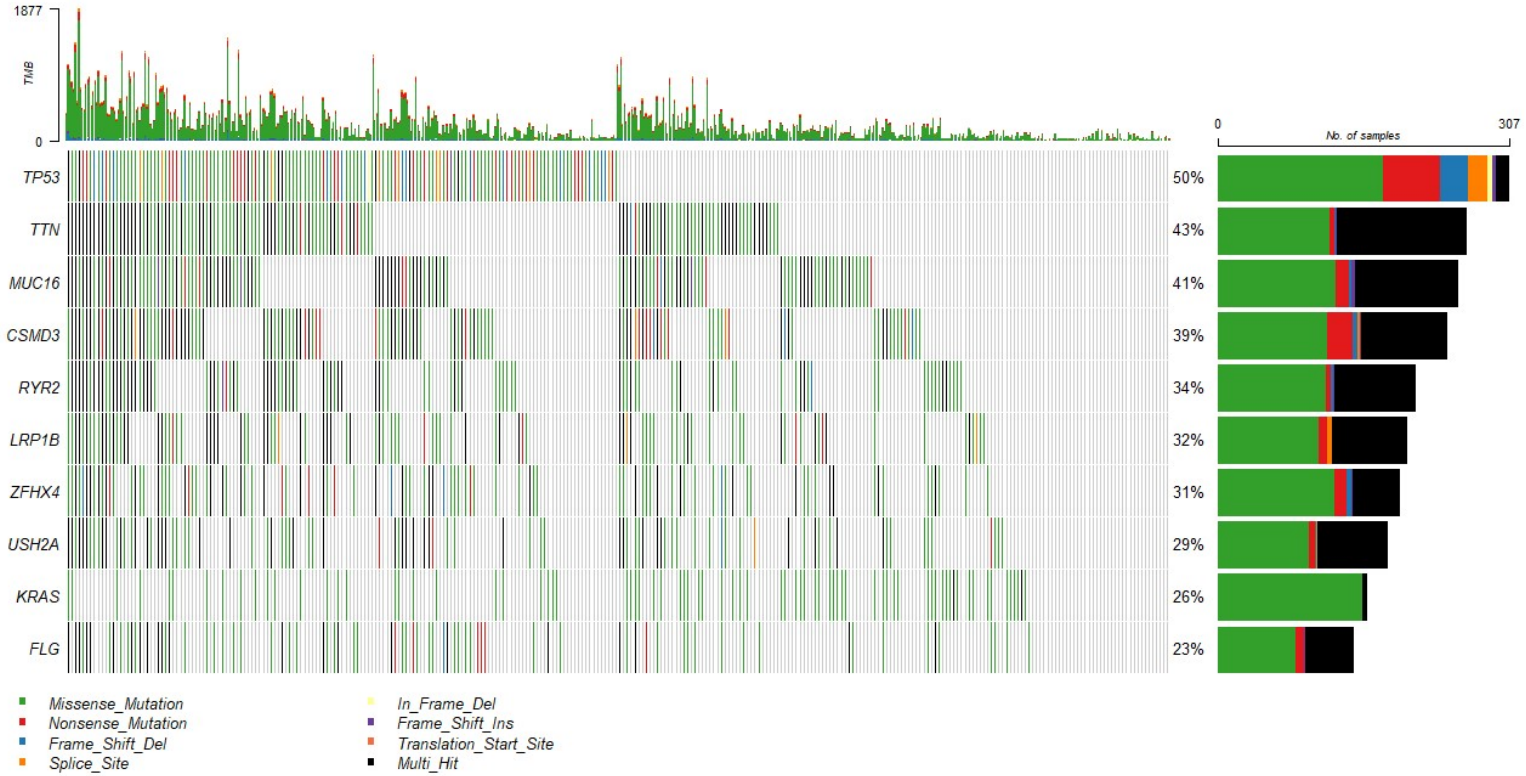
```
> # クエリを作成．データを検索するプロセスは省略している
> query <- GDCquery( project = "TCGA-LUAD",
+     data.category = "Simple Nucleotide Variation",
+     access = "open",    data.type = "Masked Somatic Mutation",
+     workflow.type = "Aliquot Ensemble Somatic Variant Merging and Masking")
----------------------------------
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()
```