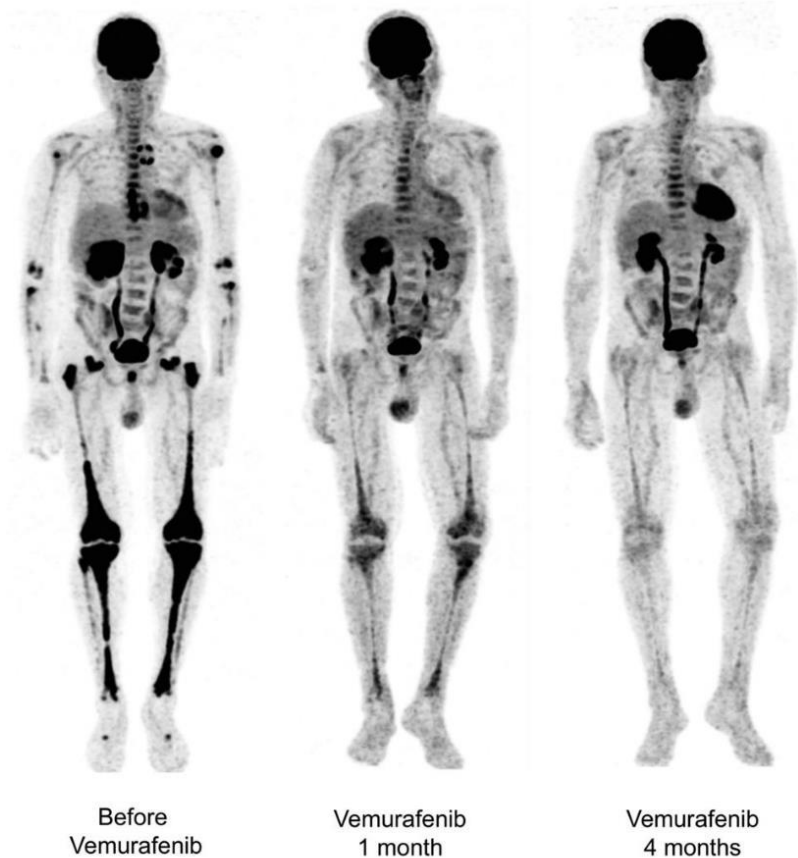


知識データベースを用いたがんゲノム解析

フィルジェン株式会社 バイオインフォマティクス部
(biosupport@filgen.jp)

- がんゲノム医療における、次世代シーケンサーを用いた遺伝子パネル解析では、治療薬情報など、臨床的意義のある解析結果を得るために、様々な知識データベースを活用して解析を行う必要がある。
- Golden Helix社VarSeq®では、次世代シーケンサーより得られた遺伝子変異データに対して、様々な臨床データベースの情報をアノテーションとして簡単に加えることができ、これらアノテーションを用いたバリエーションフィルタリングのワークフローを作成することによって、解析の効率化を行うことができる。



Haroche J. et al. Dramatic efficacy of vemurafenib in both multisystemic and refractory Erdheim-Chester disease and Langerhans cell histiocytosis harboring the *BRAF* V600E mutation. *Blood* 2013 121

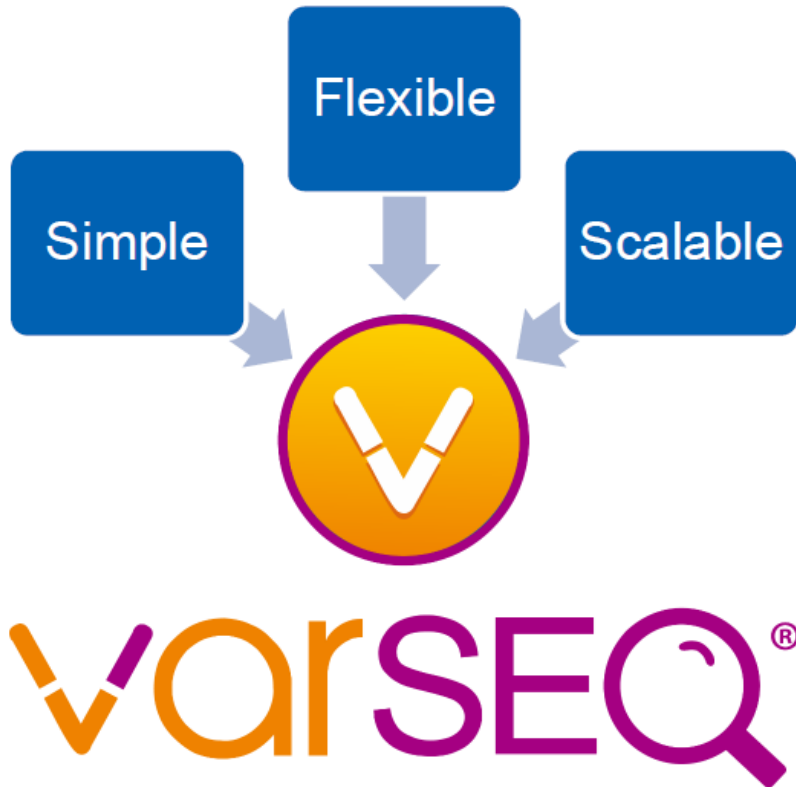


- GWAS & SNP Analysis
- Large-N DNA-Seq Analysis
- Genomic Prediction
- Copy Number Analysis
- RNA-Seq Analysis



- Cancer Diagnostics
- Gene Testing & Rare Disease Diagnosis
- CNV Calling
- Clinical Reporting
- High-throughput NGS Testing

- Golden Helix社では、遺伝統計解析ソフトウェア「SNP & Variation Suite」と、遺伝子診断支援ソフトウェア「VarSeq[®]」の2種類のソフトウェアパッケージを販売
- 医学・生物学研究や、家畜や作物の品種改良などの農学研究、さらに疾患の診断や最適な治療オプションの決定における医療分野、遺伝学的解析などの教育現場などで利用される



- キュレーションされた様々なデータリソースを使用し、変異データへアノテーション付けを実行
 - dbSNP
 - RefSeq Genes
 - 1000 Genome
 - NHLBI 6500 Exomes
 - ExAC Variant
 - gnomAD Exomes
 - SIFT and PolyPhen
 - dbNSFP Functional Predictions
 - ClinVar
 - CIVic
 - ICGC Simple Somatic Mutation
 - 各種遺伝子パネルのターゲットデータ
...など
- VCFファイルに含まれる変異データから、任意の検索条件でデータのフィルタリングを行うワークフローを作成
- カバレッジ計算やトリオ解析、表現型情報に基づく遺伝子ランキングなどの解析アルゴリズムを搭載
- ゲノムブラウザを搭載し、BAMファイルデータや各種アノテーションデータをグラフ表示
- 無償提供のビューワーソフトウェアが利用でき、解析結果を容易にシェアすることが可能



- アノテーション・フィルタリング

COSMIC

- 体細胞変異データベース

OMIM

- 遺伝子・疾患情報データベース

CADD

- 変異スコアデータベース



- 臨床レポートの作成



- コマンドライン型インターフェース



- CNV (コピー数多型) 解析



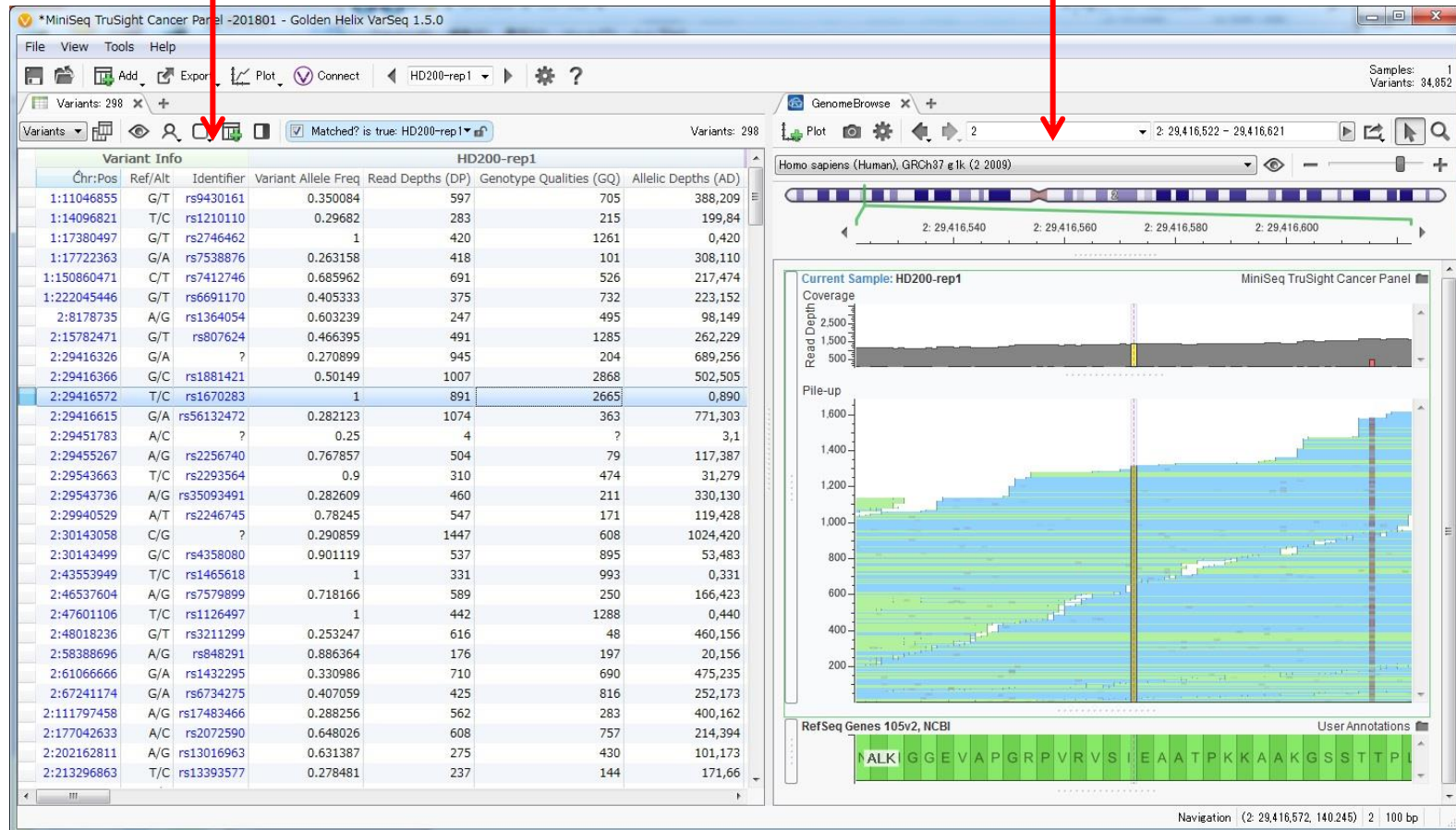
- データシェアリング用Webサーバー



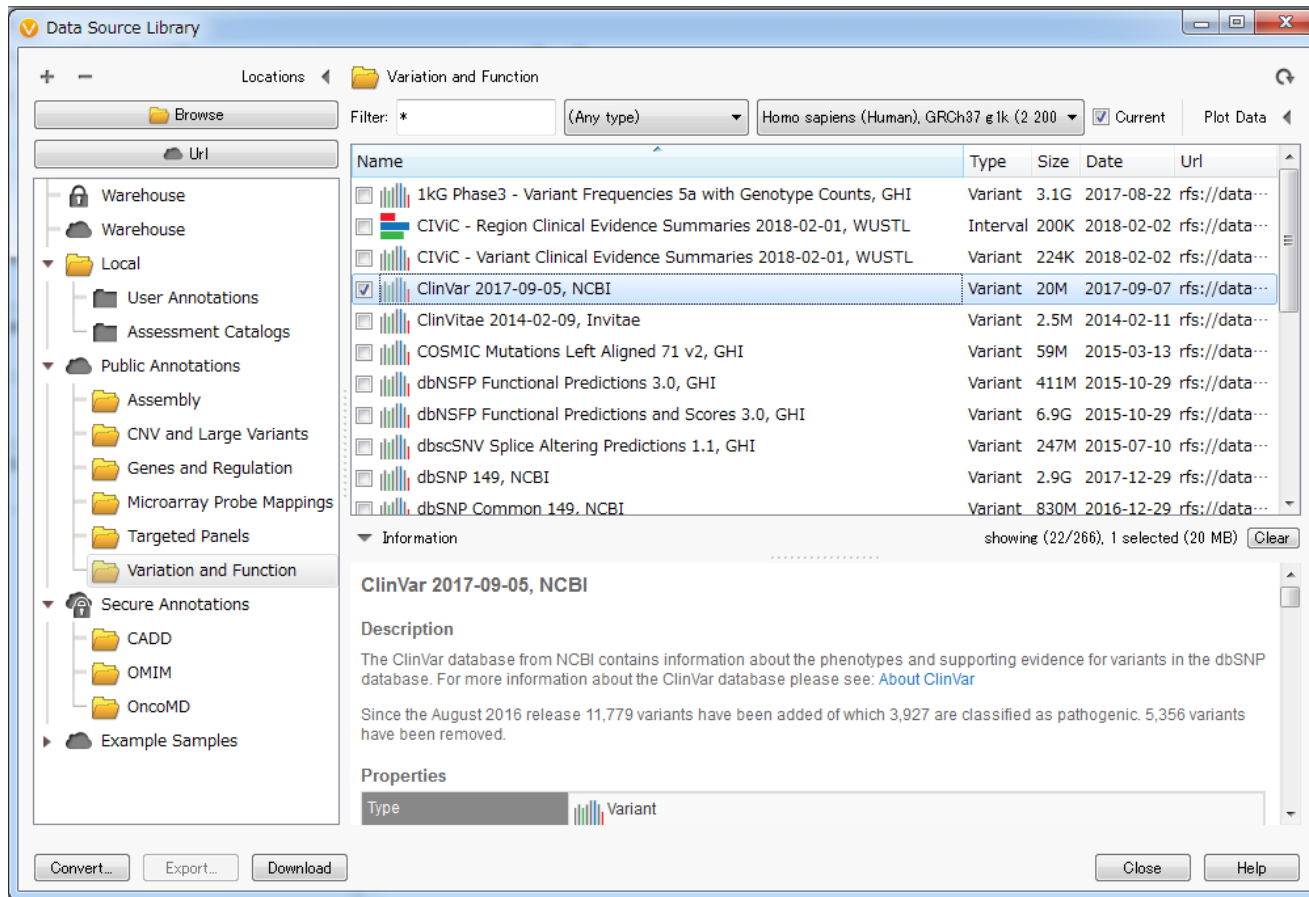
- ACMG/AMPガイドラインによる変異評価

使用するデータ

MiniSeq TruSight Cancer Panel - HD200-rep1.bam
MiniSeq TruSight Cancer Panel - HD200-rep1.vcf.gz



- 次世代シーケンサーの変異解析パイプラインで作成したVCFファイル（変異データ）とBAMファイル（リードアライメントデータ）を使用して、各種データ解析を実行



- Golden Helix社によってデータの精査・メンテナンスが行われている各種アノテーションリソースを、Data Source Libraryより自由にダウンロードし、変異データへのアノテーション付けに使用可能
- 日本人変異データベースHGVDなどのカスタムアノテーションデータをインポートし、アノテーションリソースとして使用が可能

Assemblies

- GRCh 37 (hg19)
- GRCh 38 (hg38)

Genes

- RefSeq
- Ensembl

Clinical Annotations

- ClinVar
- OMIM

Cancer Annotations

- COSMIC
- CIViC
- Cancer Hotspots
- ICGC Simple Somatic Mutations
- TCGA Variants
- MSK Impact

* 赤字は有償アドオン

Functional Annotations

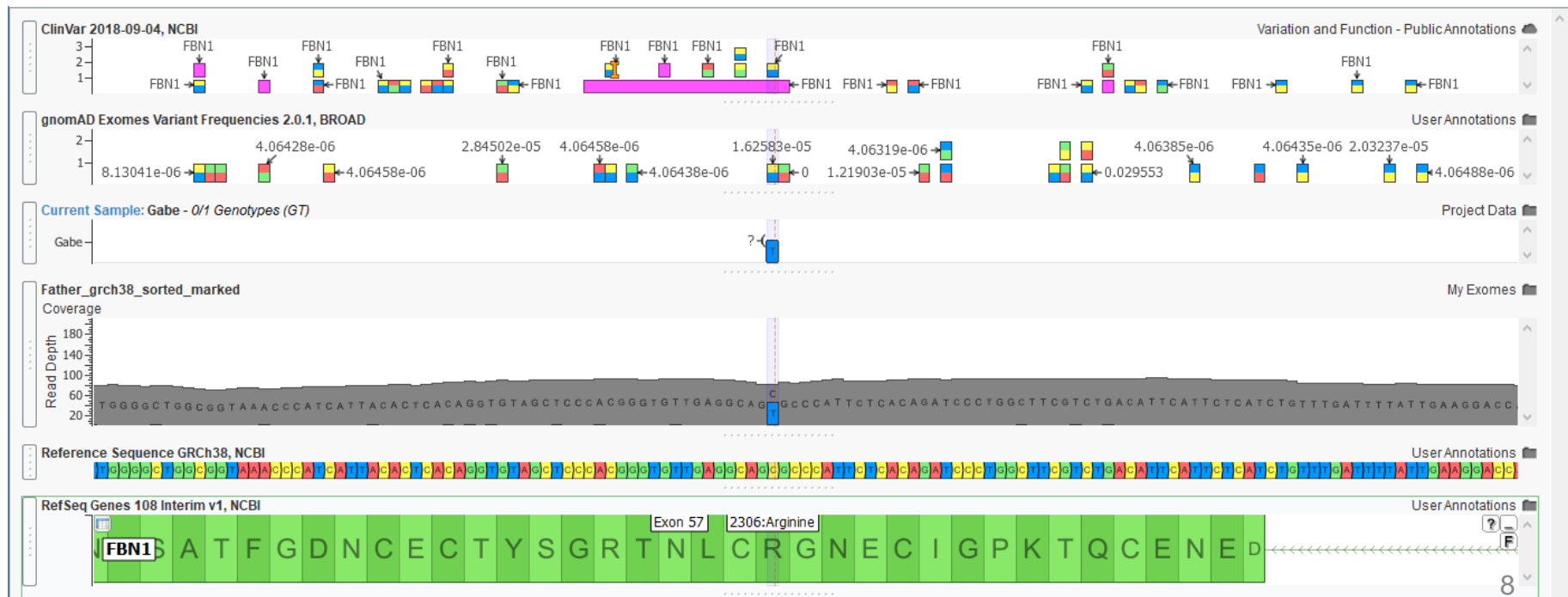
- dbNSFP (SIFT, PolyPhen, MutationTaster...)
- CADD

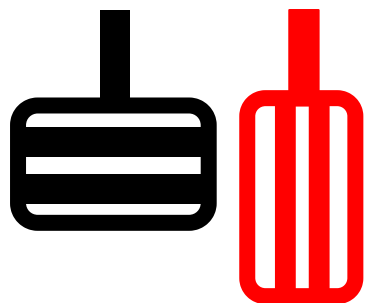
Population Catalogs

- dbSNP
- 1000 Genome
- NHLBI 6500 Exomes
- ExAC Variant
- gnomAD
- TOPMed

Targeted Panels

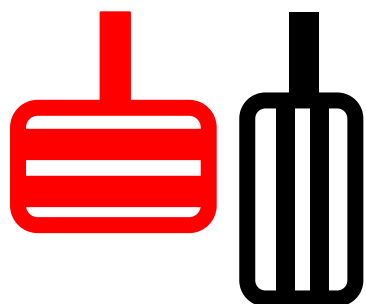
- TruSight
- Ion AmpliSeq





がん遺伝子 (Oncogene)

- 細胞増殖を促進させる遺伝子
- **機能獲得型**の遺伝子変異によって、細胞増殖が亢進する
- 変異が遺伝子上の特定領域（ホットスポット）に集中することがある
- 代表的な遺伝子：KRAS, BRAFなど



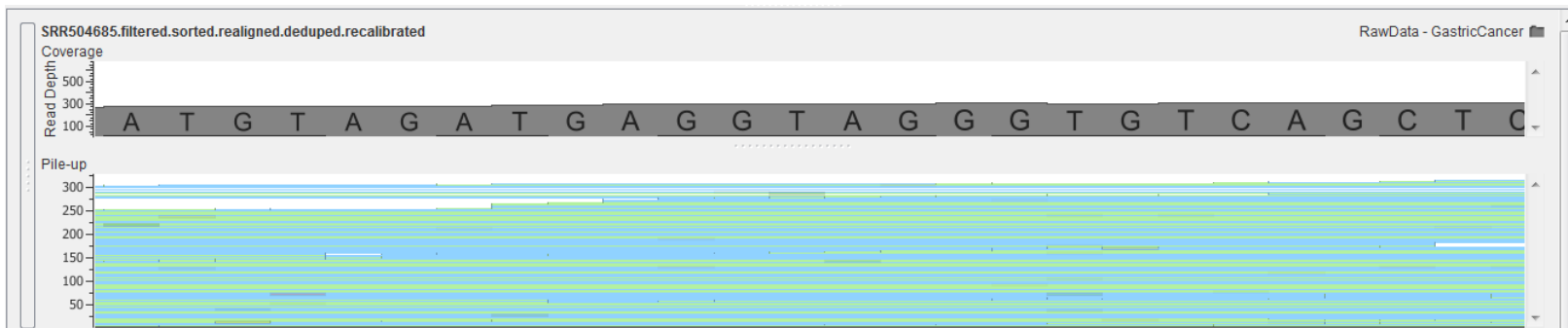
がん抑制遺伝子 (Tumor Suppressor Gene)

- 細胞増殖を抑制する遺伝子
- **機能喪失型**の遺伝子変異によって、細胞増殖が亢進する
- 遺伝性腫瘍の原因遺伝子であることが多い
- 代表的な遺伝子：BRCA1/2, APCなど

Hallmarks for PTEN:

	P	S
proliferative signalling		
suppression of growth	■	
escaping immune response to cancer		■
cell replicative immortality		■
tumour promoting inflammation		
invasion and metastasis		■
angiogenesis		
genome instability and mutations		■
escaping programmed cell death		■
change of cellular energetics	■	

Normal

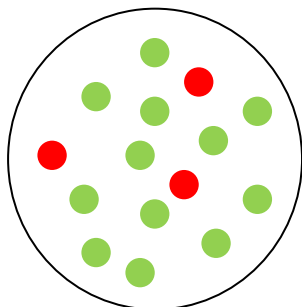


Tumor



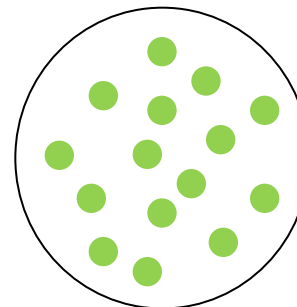
体細胞変異

- 後天的に取得する遺伝子変異で、変異が起こった細胞にのみ存在する
- 次世代に遺伝しない
- サンプル中にごく微量にのみ存在することがあるため、高感度な検出法が必要
- 検出の際は、生殖細胞系列変異と区別するために、**腫瘍サンプルと正常サンプルの両方**が必要



生殖細胞系列変異

- 個人が生まれながらに持っている、生涯変化しない
- すべての細胞で共通のものが存在する
- 次世代に遺伝する
- 血縁者間で一部共有されている



COSMIC

- Mutations: がん関連変異のカタログ（臓器情報やサンプルカウント情報含む）
- Cancer Gene Census: がん関連遺伝子のカタログ
- Fusions: 融合遺伝子の情報
- Hallmarks: がん特質に関する情報
- Transcript Counts: 転写物ごとに登録されているサンプル数や変異数

サンプルカウントデータベース

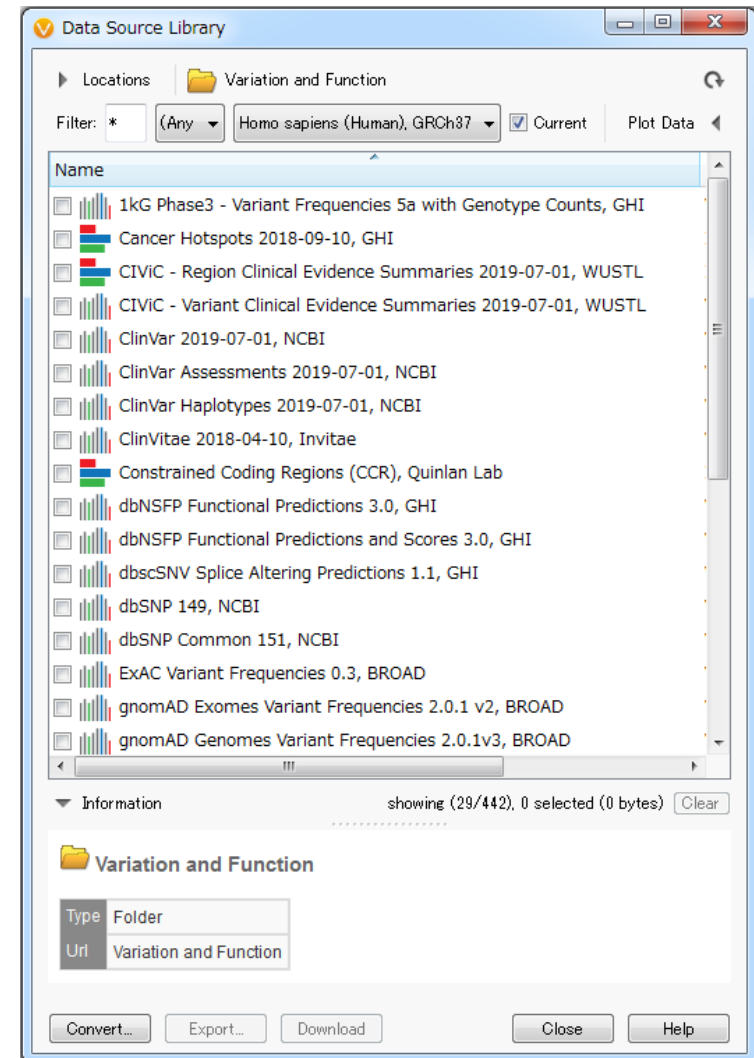
- ICGC Simple Somatic Mutations
- TCGA Variants
- MSK-Impact

知識データベース

- ClinVar
- CIViC
- The Precision Medicine Knowledgebase (PMKB)

その他

- Cancer Hotspots
- パスウェイデータベース
- 各種遺伝子パネルのターゲット領域データ（TruSight, Ion AmpliSeqなど）
- 各種アレル頻度データベース（1000 Genomes, gnomADなど）



アノテーションデータ

Variant Info		HD200-rep1		RefSeq Genes 105 Interim v1, NCBI			ClinVar 2019-07-01, NCBI				
Chr:Pos	Ref/Alt	VAF	DP	Gene Names	EffectCombined	HGVS p. (Clinically Relevant)	Classification	Review Status	Conditions	In COSMIC?	MutationIDUni...
2:29416326	G/A	0.270899	945	ALK	Missense	NP_004295.2:p.Pro1543Ser	?	?	?	True	COSM2941442
2:48018236	G/T	0.253247	616	MSH6	Missense	NP_000170.1:p.Ser144Ile	Benign	(3 Stars) Reviewed ...	not provided,Lynch syndro...	True	COSM4988649
3:10138019	G/A	0.305785	847	FANCD2,FA...	Missense	NP_149075.2:p.Asp1350Asn,?	?	?	?	True	COSM2915697,...
4:55604693	C/A	0.34	300	KIT	Other	NP_000213.1:p.Ser967=	?	?	?	True	COSM3301473
5:112179431	C/T	0.280665	482	APC	Missense	NP_000029.2:p.Arg2714Cys	Uncertain Significance	(1 Stars) Criteria Pr...	Hereditary cancer-predispo...	True	COSM2991126
8:145739087	T/A	0.270751	1014	RECQL4	Missense	NP_004251.3:p.Thr690Ser	Uncertain Significance	(1 Stars) Criteria Pr...	Baller-Gerold syndrome	True	COSM2871905
9:98239971	C/T	0.298957	865	PTCH1	Missense	NP_000255.2:p.Cys454Tyr	?	?	?	True	COSM2733335,...
9:98278975	C/T	0.259594	1331	PTCH1	Missense	?	?	?	?	True	COSM199000,C...
9:100437823	A/G	0.285714	280	XPA	Other	NP_000371.1:p.Ile240=	?	?	?	True	COSM2793140
10:43604493	C/T	0.273738	1170	RET	Missense	NP_006124.1:p.Arg360Trp	?	?	?	True	COSM33011,C...
13:32912750	G/T	0.336232	345	BRCA2	Missense	NP_000050.2:p.Asp1420Tyr	Benign	(3 Stars) Reviewed ...	not provided,Hereditary br...	True	COSM3736087,...
13:103515036	G/A	0.269939	978	BIVM-ERC...	Missense	NP_001191354.1:p.Ala967Thr,N...	?	?	?	True	COSM696103
15:89849327	G/A	0.2875	1200	FANCI	Missense	NP_001106849.1:p.Glu1147Lys	?	?	?	True	COSM2015328,...
16:2134450	G/A	0.286654	1064	TSC2	Other	NP_000539.2:p.Arg1409=	?	?	?	True	COSM4656201,...
16:3658433	C/A	0.28629	744	SLX4	Missense	NP_115820.2:p.Arg178Ile	?	?	?	True	COSM2919796
17:41234451	G/A	0.241218	427	BRCA1	LoF	NP_009225.1:p.Arg1443Ter	Pathogenic	(3 Stars) Reviewed ...	Breast-ovarian cancer, fami...	True	COSM979730,C...

Variants: 16

17:41234451 - G/A (1bp sub)

Chr:Pos: [17:41234451](#)

[rs41293455](#)

In COSMIC? is true: HD200-rep1

ClinVar Assessments 2019-07-01, NCBI									
	1	2	3	4	5	6	7	8	9
Ref/Alt	G/A	G/A	G/A	G/A	G/A	G/A	G/A	G/A	G/A
Variant ID	17675	17675	17675	17675	17675	17675	17675	17675	17675
Assessment Source	Counsyl	Michigan Medical Genetics Laboratories, University of Michigan	Baylor Miraca Genetics Laboratories,	Genetic Services Laboratory, University of Chicago	Fulgent Genetics	Integrated Genetics/Laboratory Corporation of America	OMIM	Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine	Genome Diagnostics Laboratory, University Medical Center Utrecht
Date	2014-10-02	2016-04-21	2017-02-23	2016-11-29	2017-05-18	2016-08-15	1994-12-01	2016-10-28	2014-10-08
Classification	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic
Conditions	Breast-Ovarian Cancer, Familial 1	Not Provided	Familial Cancer Of Breast	{Breast-Ovarian Cancer, Familial, 1}	Not Provided	Not Provided	Breast-Ovarian Cancer, Familial, Susceptibility	Not Provided	Not Provided

- アノテーション情報は変異データテーブルにそのまま追加され、データベースへのハイパーリンクもつく
- 変異データテーブル上の任意のデータをクリックすると、アノテーションの詳細情報も閲覧可能

Filter Variants 139,073

- Filter Variants 139,073
- Matched? is true 142
- Read Depths (DP) (Current) >= 30 124
- Variant Allele Freq (Current) >= 0.3 115
- HGVD Allele Freq 38
- Clinical Significance is (Likely Pathogenic, Pathogenic) 4
 - Affects 0
 - Association 0
 - Benign 16
 - Benign/Likely Benign 0
 - Conflicting Interpretations Of Pathogenicity 0
 - Drug Response 2
 - Likely Benign 7
 - Likely Pathogenic 4
 - Not Provided 1
 - Other 8
 - Pathogenic 2
 - Protective 0
 - Risk Factor 0
 - Uncertain Significance 2
 - Missing 14

Variants: 4

Clinical Significance is (Likely Pathogenic, Pathogenic): Sample3_Variants

Variant Info		Sample3_Variants			Summary of TruSight ...		RefSeq Ge	
Chr:Pos	Ref/Alt	Variant Allele Freq	Read Depths (DP)	Allelic Depths (CLCAD2)	Zygosity	Matched?	Gene Names	Effect (Combined)
7:55249071	C/T	0.714286	84	24,60	Heterozygous	True	EGFR	Missense
7:55259515	T/G	0.726744	172	47,125	Heterozygous	True	EGFR	Missense
9:21971153	C/A	1	32	0,31	Homozygous Variant	True	CDKN2A	LoF
17:7577120	C/T	1	213	0,207	Homozygous Variant	True	TP53	Missense

フィルタリングワークフロー

- 付加したアノテーション情報などを用いて、変異データに対する任意のフィルタリングワークフローを作成
- ワークフローの各項目ごとに、フィルタリング結果の変異データテーブルを確認可能で、項目を変更するとフィルタリング結果の変異データテーブルも自動的にアップデートされる

COSMIC Mutation

- がんに関連した体細胞変異を網羅的に収集したカタログで、アノテーションとして以下の情報をもつ
 - 変異の各種分類
 - 論文リスト
 - 変異が認められたサンプル数
 - 原発部位およびそれぞれのサンプル数
 - がんサブタイプおよびそれぞれのサンプル数
- Sample Countでフィルタリングをかけることで、高頻度に存在する変異のみを抽出することが可能
- ICGCなどの他のサンプルカウントデータベースも、ほぼ同じ種類のアノテーションをもつ
- Cancer Gene CensusやFusionsなどのCOSMICの他のデータベースは、基本的に遺伝子レベルのカタログとなっており、各遺伝子ごとの機能やがんの発生における役割、関連するがんの種類や融合遺伝子情報、論文リンクなどをアノテーションとしてもつ

In COSMIC?	True
Ref/Alt	C/T
Mutation ID	COSM1135366, COSM521
Mutation CDS	c.35G>A, c.35G>A
Mutation AA	p.G12D, p.G12D
Ensembl Transcript ID	ENST00000256078, ENST00000311936
RefSeq Transcript ID	?, ?
CDS Length	570, 567
Gene Name	KRAS
HGNC ID	6407
Pubmed ID	22696596, 26806338, 28481359, 25344691, 29360550, 26164066, 24705251, 23103869, 27245685, 24944587, [more]
Study ID	328
Sample Count	1590
Samples	2439721, 2439722, 2196603, 2196600, 2549400, 2549404, 2549391, 2549396, 2724529, 2549398, 2725034, [more]
Sample Primary Sites	Pancreas, Lung, Pancreas, Pancreas, Pancreas, Lung, Pancreas, Large intestine, Large intestine, [more]
Sample Tissue Types	Pancreas, Lung, Pancreas, Pancreas, Pancreas, Lung, Pancreas, Bowel, Bowel, Lung, Pancreas, Lung, [more]
Oncotree Tissue Type	Pancreas, Lung, Bowel, Lymphoid, Biliary Tract, Uterus, Esophagus/Stomach, Bladder/Urinary Tract, [more]
Oncotree Tissue Type Counts	524, 407, 398, 46, 40, 34, 26, 20, 14, 14, 13, 10, 9, 8, 7, 7, 5, 3, 3, 1, 1
Primary Site	Pancreas, Lung, Large intestine, Haematopoietic and lymphoid tissue, Biliary tract, Endometrium, [more]
Primary Site Counts	524, 407, 390, 46, 40, 34, 17, 17, 14, 13, 10, 9, 9, 8, 8, 7, 7, 7, 5, 3, 3, 1, 1
Site Subtype	Colon, Rectum, Appendix, Bile duct, Caecum, Bladder, Ampulla of Vater, Brain, Head neck, Duodenum, [more]
Site Subtype Counts	134, 45, 44, 40, 35, 14, 13, 7, 6, 6, 4, 3, 3, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1
Primary Histology	Carcinoma, Haematopoietic neoplasm, Other, Lymphoid neoplasm, Glioma, Pancreatic intraepithelial [more]
Primary Histology Counts	1473, 36, 27, 10, 7, 7, 6, 4, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1

ClinVar

- 疾患と関連する変異を網羅的に収集したデータベースで、アノテーションとして以下の情報をもつ
 - 変異の病原性分類 (Classification)
 - キュレーション状況 (Review Status)
 - 変異が報告された研究レコード
- Classificationでフィルタリングをかけることで、病原性を持つ変異のみを抽出することが可能

Ref/Alt	C/T
Variant ID	12582
Classification	Pathogenic
Clinical Significance	Pathogenic
Aggregate of Interpretations from Submissions	Pathogenic (21), Likely pathogenic (3)
Review Status	(2 Stars) Criteria Provided, Multiple Submitters, No Conflicts
RSID	rs121913529
Gene Names	KRAS
Gene IDs	3845
HGVS g. Name	NC_000012.11:g.25398284C>T
HGVS g. Name (GRCh38)	NC_000012.12:g.25245350C>T
HGVS c. Name	NM_004985.4:c.35G>A
HGVS p. Name	NP_004976.2:p.Gly12Asp
Per Condition Record Accessions	RCV000013411, RCV000013412, RCV000022799, RCV000029214, RCV000029215, RCV000144969, RCV000144970, [more]
Conditions	Carcinoma of pancreas, Neoplasm of stomach, Epidermal nevus, Nevus sebaceous, Epidermal nevus [more]

ClinVar Assessments

- 変異評価のエビデンスとなる、実際にClinVarに登録された研究プロジェクトの情報で、アノテーションとして以下の情報をもつ
 - 研究プロジェクトのソース
 - 変異の病原性分類
 - 変異が報告された条件 (疾患名など)
 - 報告者による解釈
 - 論文リスト
- 疾患の種類ごとに、報告者による変異に対する解釈を確認できる

	1	2	3	4
Ref/Alt	C/T	C/T	C/T	C/T
Variant ID	12582	12582	12582	12582
Assessment Source	OMIM	Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine	Database of Curated Mutations (DoCM)	Yale Center for Mendelian Genomics, Yale University
Date	2012-06-10	2014-08-28	2016-03-10	2012-10-25
Classification	Pathogenic	Pathogenic	Not Provided	Pathogenic
Conditions	Gastric Cancer, Somatic	Not Provided	Not Provided	Nevus Sebaceous
Guidelines Used	?	?	?	?
Interpretation	Motojima et al. (1993) identified mutations in KRAS codon 12 in 46 of 53 pancreatic carcinomas ([more])	Somatic KRAS variants have been identified in up to 15% of cases of ovarian carcinoma, and Gly12Asp [more]	?	?

	1	2	3	4
Ref/Alt	G/A	G/A	G/A	G/A
Tumor Type	Colorectal Adenocarcinoma	Colorectal Adenocarcinoma	Breast	Breast
Oncotree Tissue Type	Bowel	Bowel	Breast	Breast
Gene Name	PIK3CA	PIK3CA	PIK3CA	PIK3CA
CIViC Gene ID	37	37	37	37
Entrez Gene ID	5290	5290	5290	5290
Representative Transcript	ENST00000263967.3	ENST00000263967.3	ENST00000263967.3	ENST00000263967.3
Variant Name	E542K	E542K	E542K	E542K
CIViC Variant ID	103	103	103	103
Variant Origin	Somatic Mutation	Somatic Mutation	Somatic Mutation	Somatic Mutation
Variant Summary	PIK3CA E545K/E542K are the second most recurrent PIK3CA mutations in breast cancer, and are highly [more]	PIK3CA E545K/E542K are the second most recurrent PIK3CA mutations in breast cancer, and are highly [more]	PIK3CA E545K/E542K are the second most recurrent PIK3CA mutations in breast cancer, and are highly [more]	PIK3CA E545K/E542K are the second most recurrent PIK3CA mutations in breast cancer, and are highly [more]
Phenotypes	?	?	?	?
Disease	Colorectal Cancer	Colorectal Cancer	Breast Cancer	Breast Cancer
Disease Ontology ID	9256	9256	1612	1612

Drugs	Regorafenib	Cetuximab, Panitumumab	CH5132799	Rapamycin (Sirolimus)
Drug Interaction Type	?	Substitutes	?	?
CIViC Evidence ID	83	256	310	311
Evidence Type	Predictive	Predictive	Predictive	Predictive
Clinical Significance	N/A	Resistance	Sensitivity/Response	Sensitivity/Response
Evidence Direction	Does Not Support	Supports	Supports	Supports
Evidence Level	B - Clinical	B - Clinical	D - Preclinical	D - Preclinical
Evidence Statement	PIK3CA mutation status was not predictive of response to regorafenib treatment in patients that had [more]	The presence of a PIK3CA mutation in exon 9 (such as E542K) or exon 20 were associated with lack of [more]	Breast cancer cell lines with E542K/E545K mutations showed increased sensitivity to CH5132799 than [more]	Rapamycin inhibits transformation induced by mutation in PIK3CA.
PubMed ID	24559322	19223544	21558396	15647370
Citation	Sartore-Bianchi et al., 2014, Expert Rev Anticancer Ther	Sartore-Bianchi et al., 2009, Cancer Res.	Tanaka et al., 2011, Clin. Cancer Res.	Kang et al., 2005, Proc. Natl. Acad. Sci. U.S.A.
Clinical Trial Registry Numbers	?	?	?	?
Trust Rating	4 out of 5 Stars	3 out of 5 Stars	4 out of 5 Stars	3 out of 5 Stars

CIViC

- がんに関連する遺伝子変異の臨床的解釈を集めたオープンアクセスのデータベースで、アノテーションとして以下の情報をもつ
 - 変異の各種分類
 - がんの種類や臓器
 - 変異のサマリー
 - 治療薬の種類
 - 臨床的意義
 - エビデンスレベルと関連する記述
 - 論文リスト

- 研究プロジェクトで用いられた治療薬とその効果などを、説明付きで確認できる
- Clinical SignificanceやEvidence Levelでフィルタリングを行うことで、任意の臨床的意義を持つ変異を抽出できる
- SNV, InDelの他、データベースにはコピー数異常に関する情報ももつ

The Precision Medicine Knowledgebase (PMKB)

- 病理学者によるデータの精査が行われている、がんに関連する遺伝子変異の臨床的解釈をまとめたデータベースであり、アノテーションとして以下の情報をもつ
 - 変異の各種分類
 - 変異の臨床的解釈
 - 治療薬の種類
 - がんの種類や臓器
 - 論文リスト
- 治療薬とその効果などを、説明付きで確認できる
- SNV, InDelの他、データベースにはコピー数異常に関する情報をもつ

	1	2	3
Interpretation ID	361	467	467
Gene	BRCA1	BRCA1	BRCA1
Entrez Gene ID	672	672	672
Mutation Type	Coding Mutation	Amplification	Deletion
Variant Name	BRCA1 any mutation	BRCA1 copy number gain	BRCA1 copy number loss
Interpretations	Inactivating mutations in BRCA1 may be associated with sensitivity to PARP inhibitors. DrugRucaparib Niraparib Olaparib	This gene is a known cancer gene.	This gene is a known cancer gene.
Drugs	Olaparib, PARP Inhibitor, Niraparib, Rucaparib	?	?
PMKB Tumor Types	Adenocarcinoma	Acinar Cell Carcinoma, Acinic Cell Carcinoma, Acute Myeloid Leukemia, Adenocarcinoma, Adenoid [more]	Acinar Cell Carcinoma, Acinic Cell Carcinoma, Acute Myeloid Leukemia, Adenocarcinoma, Adenoid [more]
PMKB Tissue Types	Breast, Ovary, Prostate	Adrenal Gland, Anus, Ampulla (Pancreaticobiliary Duct), Appendix, Bladder, Blood, Bone, Bone Marrow, [more]	Adrenal Gland, Anus, Ampulla (Pancreaticobiliary Duct), Appendix, Bladder, Blood, Bone, Bone Marrow, [more]
Oncotree Tissue Type	Breast, Ovary/Fallopian Tube, Prostate	Adrenal Gland, Bowel, Ampulla of Vater, Bowel, Bladder/Urinary Tract, Myeloid, Bone, Bone, CNS/ [more]	Adrenal Gland, Bowel, Ampulla of Vater, Bowel, Bladder/Urinary Tract, Myeloid, Bone, Bone, CNS/ [more]
Tier	1	2	2
PubMed IDs	26510020, 21862407, 28229583, 27717299, 20406929, 24607283, 27962476, 19553641, 25366685, 27908594	27899578	27899578
Prognostic	False	False	False

Reference Sequence GRCH37 g1k, 1000Genomes

Cancer Hotspots 2018-09-10, GHI

Current Sample: T990005 - 0/1 Genotypes (GT)

RefSeq Genes 105 Interim v1, NCBI

Extended Fields

Gene	PIK3CA
Residue	E542
Type	single residue
Amino Acids	K, A, Q, V, G
Amino Acid Counts	356, 6, 5, 3, 2
Q-value	0
Samples	372
Tumor Types	breast, bladder, bowel, lung, head_neck, uterus, cervix, stomach, brain, kidney, unk, prostate, skin, biliary_tract, soft_tissue, ovary, pancreas, testis, thyroid, ampulla_of_vater, liver, lymph, thymus, vulva
Tumor Type Counts	112, 46, 45, 32, 31, 22, 16, 16, 10, 8, 6, 5, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1
Style	Fixed Color #1245d6

Cancer Hotspots

- がん関連遺伝子における、高頻度に変異が発生するゲノム領域（ホットスポット）データで、アノテーションとして以下の情報をもつ
 - 遺伝子名
 - 該当領域の野生型のアミノ酸配列
 - 変異ごとに観察されたアミノ酸配列とサンプル数
 - がんの種類とサンプル数
- バリエントフィルタリングを行う場合は、ホットスポット領域内に存在する変異のみを抽出することが可能
- コドン内の変異によって、高頻度に観察されたアミノ酸残基などを確認できる
- ゲノムブラウザにプロットさせることで、遺伝子上のホットスポットを視覚的に表示させることも可能

1kG Phase3 - Variant Frequencies 5a with Genotype Counts, GHI									
Ref/Alt	Identifier	Allele Counts	Allele Frequencies	AMR - Allele Counts	AMR - Allele Frequencies	AFR - Allele Counts	AFR - Allele Frequencies	SAS - Allele Counts	SAS - Allele Frequencies
A/G	rs1412829	923	0.184305	131	0.188761	18	0.0136157	265	0.270961
G/A	rs4977756	3564	0.711661	550	0.792507	899	0.68003	710	0.725971
G/A	rs357564	1987	0.396765	282	0.40634	298	0.225416	461	0.47137
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
A/G	rs965513	4005	0.79972	484	0.697406	1181	0.893343	795	0.812883
G/T	rs865686	3546	0.708067	474	0.682997	652	0.493192	859	0.878323
T/C	rs505922	1773	0.354034	152	0.21902	483	0.365356	384	0.392638
A/G	rs1800858	3774	0.753594	520	0.74928	1296	0.980333	679	0.694274
?	?	?	?	?	?	?	?	?	?
A/G	rs1800860	3882	0.77516	433	0.623919	1177	0.890318	737	0.753579
G/T	rs1800861	3568	0.71246	519	0.747839	1217	0.920575	585	0.598159
C/T	rs3123078	2629	0.52496	417	0.600865	700	0.529501	399	0.407975
T/C	rs10993994	2417	0.482628	418	0.602305	469	0.354766	374	0.382413

アレル頻度データベース

- 人種などの特定集団内のアレル頻度を変異ごとに計算したデータベースで、以下種類のデータベースが利用できる
 - dbSNP
 - 1000 Genome
 - NHLBI 6500 Exomes
 - ExAC Variant
 - gnomAD Exomes/Genomes
 - TOPMed
 - UK10K
- データ解析時には、任意のアレル頻度でフィルタリングを行うことで、ごくまれに存在する変異のみを抽出することが可能
- データベースの種類によっては、特定人種のアレル頻度データをフィルタリングに使用することも可能
- Human Genetic Variation (HGVD)などの日本人SNPデータベースを別途インポートすることによって、同様にフィルタリングに用いることも可能

解析サンプル：腫瘍細胞のみ

解析に用いるデータファイル：各サンプルのVCFファイル

解析の目的：がん遺伝子における機能獲得型変異と、がん抑制遺伝子における機能喪失型変異の検出

The screenshot displays the Golden Helix VarSeq 1.4.2 interface. The main window shows a list of 30 variants filtered by 'In COSMIC? is true: 10 Percent'. The variant list includes columns for Chr:Pos, Ref/Alt, Filter, Read Depth (DP), Variant Allele Freq, In COSMIC?, and Mutation ID (Unique). The right-hand panel shows the 'Effect (Combined)' analysis for the selected variants, including a 'Category Counts' table and a 'Categories of Effect (Combined)' section.

Category	Count	Percent
Misense	14	46.67%
Other	14	46.67%
LoF	2	6.67%
Total	30	100.0%

Categories of Effect (Combined)

- Misense:** The variant will cause at least one amino acid to change or cause a premature start codon in the UTR5. The ontologies included in this category are: disruptive_inframe_deletion, disruptive_inframe_insertion, inframe_deletion, inframe_insertion, 5_prime_UTR_premature_start_codon_gain_variant, missense_variant.
- Other:** The variant is likely to have a low or unknown effect on the transcript's functional product. These changes do not change the amino acid sequence of the protein. The ontologies included in this category are: synonymous_variant, stop_retained_variant, splice_region_variant, 3_prime_UTR_variant, 5_prime_UTR_variant, intron_variant, non_coding_exon_variant, intergenic_variant, unknown.
- LoF:** Loss of Function. The variant is likely to cause the transcript's product to lose function. The ontologies included in this category are: transcript_ablation, exon_loss_variant, stop_lost, stop_gained, initiator_codon_variant, frameshift_variant, splice_acceptor_variant, splice_donor_variant.

Filter Variants x +

Filter Variants 126,914

Matched? is true

True	125,773
False	1,141
Missing	0
125,773	

Read Depth (DP) (Current) >= 2,500

2,500

Less than 2,500	27,475
Equal to 2,500	1
Greater than 2,500	40,744
Missing	57,553
40,745	

Variant Allele Freq (Current) >= 0.01

0.01

Less than 0.01	0
Equal to 0.01	0
Greater than 0.01	210
Missing	40,535
210	

Matched?	Name
True	NRAS.exon.2.line.7.chr.1.115258670.1...
True	NRAS.exon.2.line.7.chr.1.115258670.1...
True	NRAS.exon.2.line.7.chr.1.115258670.1...
True	NRAS.exon.2.line.7.chr.1.115258670.1...
False	?
False	?
False	?

Variant Info			NA12877-50-Horizon_S1			
Chr:Pos	Ref/Alt	Identifier	Read Depth (DP)	Variant Allele Freq	Genotype Qualities (GQ)	Gene Names
11:118348810	T/C	?	9955	0.062695	100	KMT2A
11:118348811	G	?	10093	?	100	KMT2A
11:118348812	T	?	10059	?	100	KMT2A
11:118348813	G	?	10029	?	100	KMT2A
11:118348814	G	?	10035	?	100	KMT2A
11:118348815	G	?	10083	?	100	KMT2A
11:118348816	C	?	9479	?	100	KMT2A
11:118348817	A	?	9457	?	100	KMT2A
11:118348818	G	?	10099	?	100	KMT2A
11:118348819	T	?	10077	?	100	KMT2A
11:118348820	G	?	10066	?	100	KMT2A
11:118348821	T/A	?	10028	0.0620318	100	KMT2A
11:118348822	C	?	10062	?	100	KMT2A
11:118348823	C	?	9999	?	100	KMT2A
11:118348824	C/G	?	10058	0.0569778	100	KMT2A

- 遺伝子パネルデザインに用いられたターゲット領域、およびサンプルデータのリード深度 (Read Depths) とアレル頻度 (Variant Allele Freq) のフィルター設定を行う

Filter Variants x +

Variants

- Filter Variants 126,914
- Matched? is true 125,773
- Read Depth (DP) (Current) >= 2,500 40,745
- Variant Allele Freq (Current) >= 0.01 210
- Allele Frequencies < 0.01 OR missing
 - 0.01
 - Less than 0.01 6
 - Equal to 0.01 0
 - Greater than 0.01 43
 - Missing 161
- Allele Frequencies < 0.01 OR missing
 - 0.01
 - Less than 0.01 0
 - Equal to 0.01 0
 - Greater than 0.01 0
 - Missing 167
- In COSMIC? is true
 - True 32
 - False 135
 - Missing 0

- 続いて2種類のアレル頻度データベースを用いて、それぞれの集団内で頻度が小さい変異と、COSMIC Mutationデータベースに登録されている変異のフィルター設定を行う

1kG Phase3 - Variant Frequencies 5						
Ref/Alt	Identifier	Flags	Read Depth (DP)	Allele Counts	Allele Frequencies	# Alleles
A/G	rs1364054	EX_TARGET	21383	2445	0.488219	5008
G/T	rs807624	?	19967	2962	0.591454	5008
?	?	?	?	?	?	?
G/C	rs1881421	EX_TARGET	17724	2869	0.572883	5008

HGVD1210-V2_30						
Ref/Alt	Allele Frequencies	rsID	Ref	Alt	#Sample	
?	?	?	?	?	?	
A/G	0.295687	rs2292884	A	G	881	
G/A	0.328386	rs12330369	G	A	1078	
?	?	?	?	?	?	

Summary of COSMIC Mutations 88, GHI				
In COSMIC?	Mutation ID (Unique)	Gene Name (Unique)	Mutation CDS (Unique)	Mutation AA (Unique)
True	COSM3749653.C...	ALK	c.27C>G	p.L9L
False	?	?	?	?
False	?	?	?	?
True	COSM3927762	EPCAM	c.344T>C	p.M115T

Oncogene

Role in Cancer is Oncogene

Fusion 7

Oncogene 20

TSG 20

Missing 0

Matched? is true

True 6

False 14

Missing 0

Classification

Association Not Found 0

Benign 0

Conflicting 0

Likely Benign 0

Likely Pathogenic 1

Other 0

Pathogenic 3

Uncertain Significance 2

Missing 0

COSMIC Cancer Gene Census 89, GHI		
Molecular Genetics	Role in Cancer	Mutation Types
Dominant	Oncogene	Missense
Dominant	Oncogene	Missense,Other
Dominant	Oncogene	Missense,Other
Recessive	TSG	Missense,Nonsense,Fra...
Recessive	TSG	Missense,Nonsense,Fra...

Summary of... Cancer Hotspots 2018-09-10, GHI						
Matched?	Gene	Residue	Type	Amino Acids	Amino Acid Counts	
True	NRAS	Q61	single residue	R,K,L,H,P,*	204,142,46,27,2,1	
False	?	?	?	?	?	
True	KIT	D816	single residue	V,H,Y,A	11,5,5,1	
False	?	?	?	?	?	

ClinVar 2019-07-01, NCBI			
Ref/Alt	Variant ID	Classification	Clinical Significance
G/T	73058	Uncertain Significance	Uncertain Significance
A/T	13852	Uncertain Significance	Uncertain Significance
A/T	13961	Pathogenic	Pathogenic
C/T	12580	Pathogenic	Pathogenic
C/T	12582	Pathogenic	Pathogenic

- Oncogeneの選択とホットスポット上の変異の抽出、加えて病原性のある変異の抽出のフィルター設定を行う

Tumor Suppressor Gene

Role in Cancer is TSG

Fusion	7
Oncogene	20
TSG	20
Missing	0
	20

(Classification OR Effect (Combined))

AND Results satisfy criteria of **all** contained filters

OR Results satisfy criteria of **any** contained filter

<input checked="" type="checkbox"/> Classification	<input checked="" type="checkbox"/> Effect (Combined)
Association Not Found	Invalid
Benign	LoF
Conflicting	Missense
Likely Benign	Other
Likely Pathogenic	Unknown
Other	Missing
Pathogenic	
Uncertain Significance	
Missing	

COSMIC Cancer Gene Census 89, GHI		
Molecular Genetics	Role in Cancer	Mutation Types
Dominant	Oncogene	Missense
Dominant	Oncogene	Missense,Other
Dominant	Oncogene	Missense,Other
Recessive	TSG	Missense,Nonsense,Fra...
Recessive	TSG	Missense,Nonsense,Fra...

ClinVar 2019-07-01, NCBI			
Ref/Alt	Variant ID	Classification	Clinical Significance
G/T	73058	Uncertain Significance	Uncertain Significance
A/T	13852	Uncertain Significance	Uncertain Significance
A/T	13961	Pathogenic	Pathogenic
C/T	12580	Pathogenic	Pathogenic
C/T	12582	Pathogenic	Pathogenic

OR

RefSeq Genes 105 Interim v1, NCBI		
Gene Names	Sequence Ontology (Combined)	Effect (Combined)
TET2	missense_variant	Missense
TET2	missense_variant	Missense
TET2	stop_gained	LoF
TET2	missense_variant	Missense

- Tumor Suppressor Geneの選択に続き、病原性のある変異または機能喪失型変異の抽出のフィルター設定を行う

Oncogene or Tumor Suppressor Gene

AND Results satisfy criteria of **all** contained filters

OR Results satisfy criteria of **any** contained filter

Oncogene

Role in Cancer is Oncogene

Fusion	7
Oncogene	20
TSG	20
Missing	0
Total	20

Matched? is true

True	6
False	14
Missing	0
Total	6

Classification

Association Not Found	0
Benign	0
Conflicting	0
Likely Benign	0
Likely Pathogenic	1
Other	0
Pathogenic	3
Uncertain Significance	2
Missing	0
Total	4

Tumor Suppressor Gene

Role in Cancer is TSG

Fusion	7
Oncogene	20
TSG	20
Missing	0
Total	20

(Classification OR Effect (Combined))

<input checked="" type="checkbox"/> Classification		<input checked="" type="checkbox"/> Effect (Combined)	
Association Not Found	0	Invalid	0
Benign	0	LoF	3
Conflicting	0	Missense	14
Likely Benign	0	Other	3
Likely Pathogenic	1	Unknown	0
Other	0	Missing	0
Pathogenic	0		
Uncertain Significance	0		
Missing	19		
Total	1	Total	3

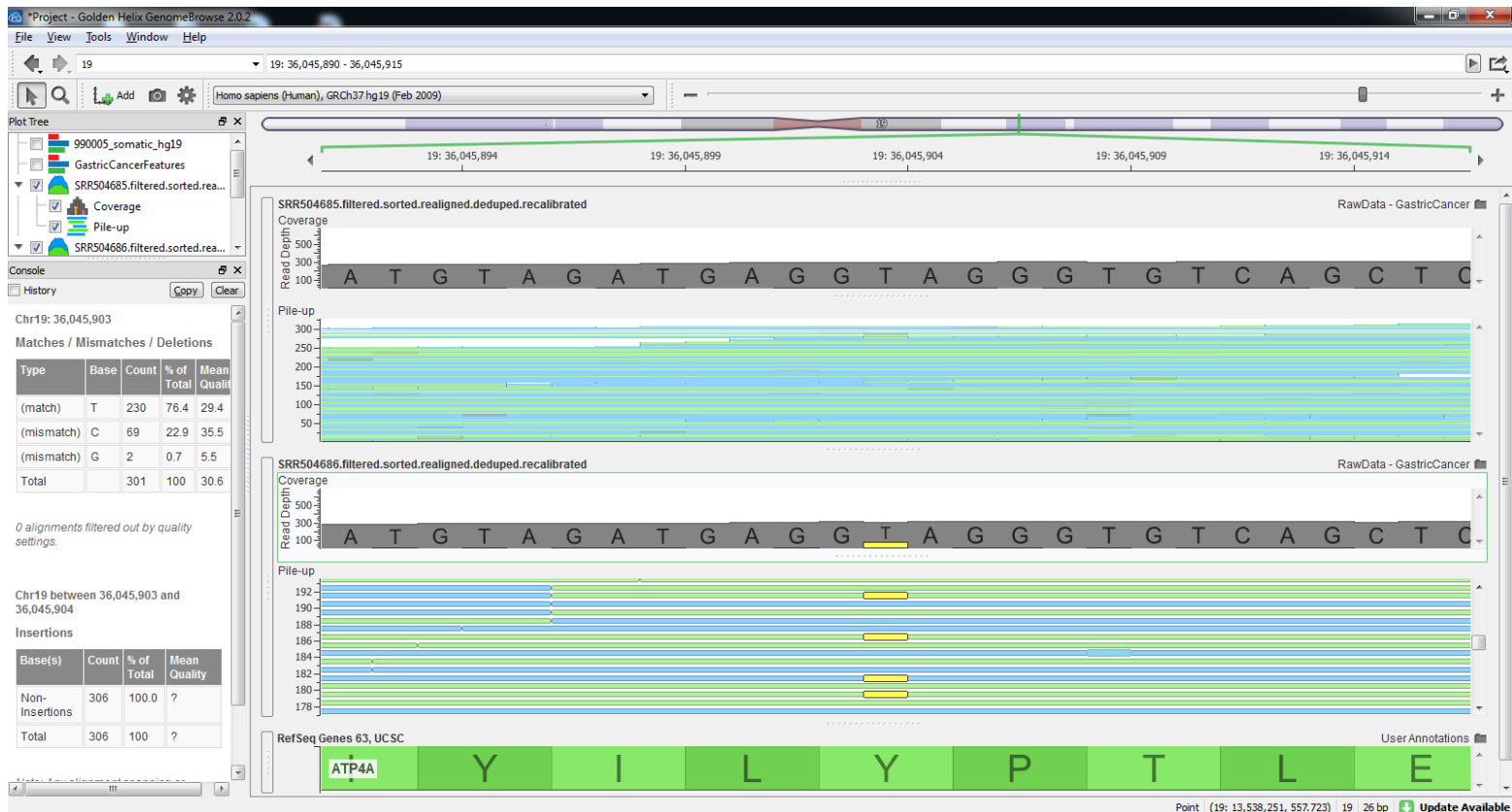
7

- OncogeneとTumor Suppressor GeneのフィルターコンテナをOR検索で繋げ、同時に検索を実行できるようにする
- このフィルター設定により、集団内にまれに存在するCOSMIC登録済みの変異より、がん遺伝子とがん抑制遺伝子においてがんの発生に関わるものを検出できる
- 各変異の臨床的意義や治療薬情報などは、各変異の詳細情報ウインドウより確認

解析サンプル：腫瘍細胞と正常細胞のペアサンプル

解析に用いるデータファイル：各サンプルのVCFファイル、BAMファイル

解析の目的：がん関連の体細胞変異と、病原性の生殖細胞系列変異の抽出



Filter Variants x +

Variants

Filter Variants 11,076

Somatic

Read Depths (DP) (Current) >= 50

50

Less than 50 5,392

Equal to 50 40

Greater than 50 4,470

Missing 1,174

4,510

AF (Current) >= 0.05

0.05

Less than 0.05 8

Equal to 0.05 0

Greater than 0.05 4,501

Missing 1

4,501

AF (Normal) < 0.01 OR missing

0.01

Less than 0.01 1

Equal to 0.01 0

Greater than 0.01 4,397

Missing 103

104

104

Variant Info		Tumor (T990005)		Normal (N990005)		GeneNames
Chr:Pos	Ref/Alt	Read Depths (DP)	AF	Read Depths (DP)	AF	
3:1339681	GTT...	60	0.293103	47	0.288889	CNTN6
3:1418753	G/A	247	0.40081	177	0.443182	CNTN6
3:1424718	G/A	250	0.473896	213	0.43128	CNTN6
3:1424745	C/T	226	0.451327	176	0.4375	CNTN6
3:1424850	T/G	182	0.5	99	0.585859	CNTN6
3:1637940	G/A	9	0.375	8	0.5	?
3:1771749	A/C	20	1	18	1	?
3:2404094	C/G	10	1	15	1	CNTN4
3:2404329	C/T	34	1	15	1	CNTN4
3:3076401	C/T	248	0.477733	233	0.420601	CNTN4
3:3076443	A/G	248	0.495968	249	0.48583	CNTN4
3:3081825	T/C	230	0.430435	193	0.435233	CNTN4
3:3098041	G/C	14	0.357143	7	0.857143	CNTN4
3:3108637	T/C	3	1	?	?	IL5RA
3:3109444	C/T	5	0.8	?	?	IL5RA

- 腫瘍細胞サンプルデータのリード深度 (Read Depths) とアレル頻度 (AF)、および正常細胞サンプルデータのアレル頻度のフィルター設定を行う

Filter Variants x +

Variants

Filter Variants 11,076

Somatic

Read Depths (DP) (Current) >= 50 4,510

AF (Current) >= 0.05 4,501

AF (Normal) < 0.01 OR missing 104

In COSMIC? is true

True 29

False 75

Missing 0

29

Sample Count >= 100

100

Less than 100 28

Equal to 100 0

Greater than 100 1

Missing 0

1

Summary of COSMIC Mutations 88, GHI				
In COSMIC?	Mutation ID (Unique)	Gene Name (Unique)	Mutation CDS (Unique)	Mutation AA (Unique)
True	COSM3749658,C...	ALK	c.27C>G	p.L9L
False	?	?	?	?
False	?	?	?	?
True	COSM3927762	EPCAM	c.344T>C	p.M115T

COSMIC Mutations 88, GHI					
Gene Name	HGNC ID	Pubmed ID	Study ID	Sample Count	Oncotree Tissue Type
CIDEA	24229	29681454	?	1	Lung
FANCD2	3585	26496030,...	589	12	CNS/Brain,Thyroid,B...
ZNF860	?	?	660	2	Liver
CTNNB1	2514	25759019	413	2	Bladder/Urinary Tra...
FYCO1	14673	25275298	?	42	Head and Neck,Thyr...
FRG2C	33626	26873401,...	646	13	Thyroid,Bowel,Liver,...

- 続いて、COSMIC Mutationデータベースに登録されている変異、および登録サンプル数のフィルター設定を行う

Filter Variants x +

Variants ▾

Filter Variants 11,076

Germline

Read Depths (DP) (Normal) >= 50

50

Less than 50 4,728

Equal to 50 40

Greater than 50 4,027

Missing 2,281

4,067

AF (Normal) >= 0.05

0.05

Less than 0.05 6

Equal to 0.05 0

Greater than 0.05 4,058

Missing 3

4,058

4,058

Variant Info		Tumor (T990005)		Normal (N990005)		
Chr:Pos	Ref/Alt	Read Depths (DP)	AF	Read Depths (DP)	AF	GeneNames
3:1339681	GTT...	60	0.293103	47	0.288889	CNTN6
3:1418753	G/A	247	0.40081	177	0.443182	CNTN6
3:1424718	G/A	250	0.473896	213	0.43128	CNTN6
3:1424745	C/T	226	0.451327	176	0.4375	CNTN6
3:1424850	T/G	182	0.5	99	0.585859	CNTN6
3:1637940	G/A	9	0.375	8	0.5	?
3:1771749	A/C	20	1	18	1	?
3:2404094	C/G	10	1	15	1	CNTN4
3:2404329	C/T	34	1	15	1	CNTN4
3:3076401	C/T	248	0.477733	233	0.420601	CNTN4
3:3076443	A/G	248	0.495968	249	0.48583	CNTN4
3:3081825	T/C	230	0.430435	193	0.435233	CNTN4
3:3098041	G/C	14	0.357143	7	0.857143	CNTN4
3:3108637	T/C	3	1	?	?	IL5RA
3:3109444	C/T	5	0.8	?	?	IL5RA

- 正常細胞サンプルデータのリード深度 (Read Depths) とアレル頻度 (AF) のフィルター設定を行う

Filter Variants x +

Variants ▾

Filter Variants 11,076

Germline

Read Depths (DP) (Normal) >= 50 4,067

AF (Normal) >= 0.05 4,058

Allele Frequencies < 0.01 OR missing

0.01

Less than 0.01 116

Equal to 0.01 0

Greater than 0.01 3,677

Missing 265

381

Alt_allele_freq < 0.01 OR missing

0.01

Less than 0.01 45

Equal to 0.01 0

Greater than 0.01 141

Missing 195

240

240

1kG Phase3 - Variant Frequencies 5

Ref/Alt	Identifier	Flags	Read Depth (DP)	Allele Counts	Allele Frequencies	# Alleles
A/G	rs1364054	EX_TARGET	21383	2445	0.488219	5008
G/T	rs807624	?	19967	2962	0.591454	5008
?	?	?	?	?	?	?
G/C	rs1881421	EX_TARGET	17724	2869	0.572883	5008

HGVD1210-V2_30

Ref/Alt	Allele Frequencies	rsID	Ref	Alt	#Sample
?	?	?	?	?	?
A/G	0.295687	rs2292884	A	G	881
G/A	0.328386	rs12330369	G	A	1078
?	?	?	?	?	?

- 続いて2種類のアレル頻度データベースを用いて、それぞれの集団内で頻度が小さい変異のフィルター設定を行う

Filter Variants x +

Variants

Filter Variants 11,076

Germline

Read Depths (DP) (Normal) >= 50 4,067

AF (Normal) >= 0.05 4,058

Allele Frequencies < 0.01 OR missing 381

Alt_allele_freq < 0.01 OR missing 240

Effect (Combined) is (LoF, Missense)

LoF 8

Missense 61

Other 171

Missing 0

69

Classification is (Likely Pathogenic, Pathogenic)

Association Not Found 0

Benign 0

Conflicting 2

Likely Benign 2

Likely Pathogenic 0

Other 0

Pathogenic 1

Uncertain Significance 0

Missing 64

1

RefSeq Genes 105 Interim v1, NCBI

Gene Names	Sequence Ontology (Combined)	Effect (Combined)
TET2	missense_variant	Missense
TET2	missense_variant	Missense
TET2	stop_gained	LoF
TET2	missense_variant	Missense

ClinVar 2019-07-01, NCBI

Ref/Alt	Variant ID	Classification	Clinical Significance
G/T	73058	Uncertain Significance	Uncertain Significance
A/T	13852	Uncertain Significance	Uncertain Significance
A/T	13961	Pathogenic	Pathogenic
C/T	12580	Pathogenic	Pathogenic
C/T	12582	Pathogenic	Pathogenic

- 非同義変異または機能喪失型変異の抽出に加えて、病原性のある変異の抽出のフィルター設定を行う

Filter Variants 11,076

AND Results satisfy criteria of **all** contained filters

OR Results satisfy criteria of **any** contained filter

Somatic

Read Depths (DP) (Current) ≥ 50 4,510

AF (Current) ≥ 0.05 4,501

AF (Normal) < 0.01 OR missing 104

In COSMIC? is true

True	29
False	75
Missing	0
Total	29

Sample Count ≥ 100

100

Less than 100 28

Equal to 100 0

Greater than 100 1

Missing 0

Total **1**

Germline

Read Depths (DP) (Normal) ≥ 50 4,067

AF (Normal) ≥ 0.05 4,058

Allele Frequencies < 0.01 OR missing 381

Alt_allele_freq < 0.01 OR missing 240

Effect (Combined) is (LoF, Missense)

LoF	8
Missense	61
Other	171
Missing	0
Total	69

Classification

Association Not Found	0
Benign	0
Conflicting	2
Likely Benign	2
Likely Pathogenic	0
Other	0
Pathogenic	1
Uncertain Significance	0
Missing	64
Total	1

- 体細胞と正常細胞のフィルターコンテナをOR検索で繋げ、同時に検索を実行できるようにする
- このフィルター設定により、COSMIC登録済みでサンプル数の多い体細胞変異、集団内にまれに存在するClinVar登録済みの病原性の生殖細胞系列変異を検出できる
- 各変異の臨床的意義や治療薬情報などは、各変異の詳細情報ウインドウより確認
- 必要に応じて、ゲノムブラウザーに表示させたBAMファイルデータより、リード配列上の変異の有無の目視チェックなどを行う

お問い合わせ先：フィルジエン株式会社

TEL: 052-624-4388 (9:00～18 : 00)

FAX: 052-624-4389

E-mail: biosupport@filgen.jp