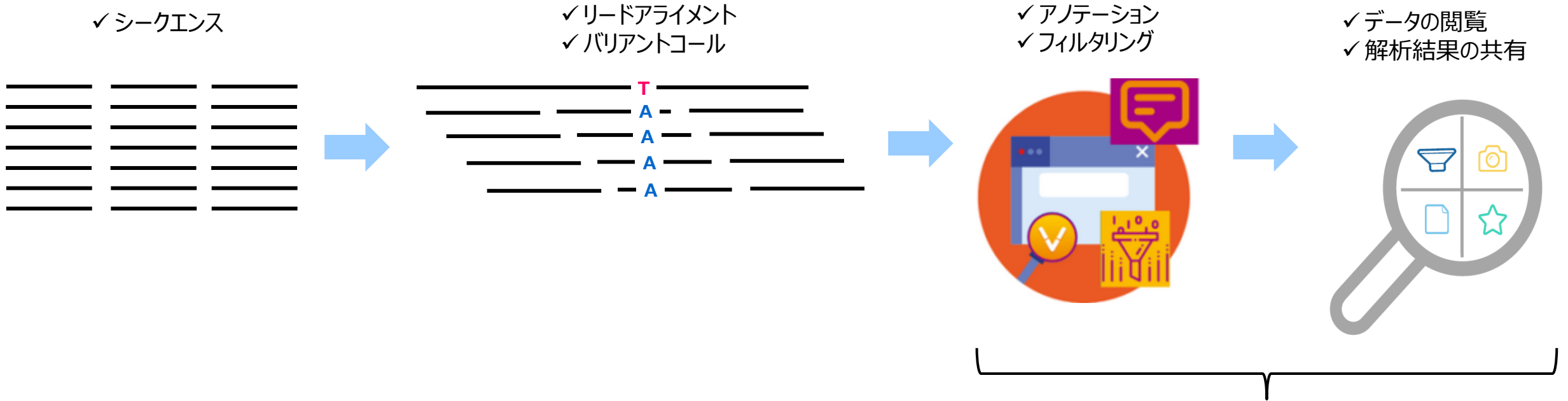
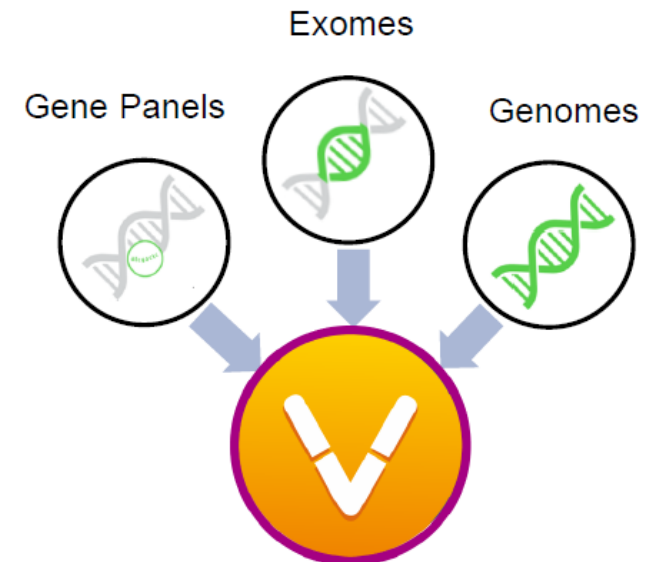


遺伝子バリアントの効率的なフィルタリング ワークフロー作成

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



- 次世代シーケンサーより出力された遺伝子バリエーションを評価し、疾患との関連を調べるには、バリエーションの効率的なフィルタリングと高品質なアノテーションが必要
- Golden Helix社VarSeq®ソフトウェアでは、遺伝子パネルや全エクソーム／全ゲノムシーケンスで得られたバリエーションデータに対して、様々なデータベースを利用したアノテーション付けと柔軟なフィルタリングワークフロー作成により、病原性バリエーションの同定が可能



Annotation

Variant Filtering

The screenshot displays the VarSeq software interface. At the top, a yellow callout box labeled 'Annotation' points to the 'Reference Sequence GRCH37 g1k, 1000Genomes' track. Another yellow callout box labeled 'Variant Filtering' points to the 'Variants - NA12878' track. The main window shows a genomic browser view for chromosome 16, with a zoomed-in view of a region around 16:2,495,482. Below the browser, a table lists 36 variants with columns for Variant Sites, Genotypes, Classification, and Compound Het Variants.

Variant Sites	Genotypes	Classification	Compound Het Variants
Chr_Pos Ref/Alt Identifier	Proband (NA12878) Mother (NA12891) Father (NA12892)	Sequence Ontology	Is CH? Inherited From
11:108183167 A/G rs659243	G,G G,G	missense_variant	False NA
13:49033835 G/A rs20211...	A,G A,G	missense_variant	False NA
14:24567498 A/C rs30211...	C,C C,C	missense_variant	False NA
14:73664751 T/G rs19972...	G,T G,T	missense_variant	False NA
14:106208082 G/T rs11621...	G,T T,T	missense_variant	False NA
16:2495482 T/G rs20154...	G,T T,T	missense_variant	True Father

Data Analysis

Genome Browser

- 様々なデータベースを用いて、バリエントデータ（VCFファイル）へアノテーション付けを実行

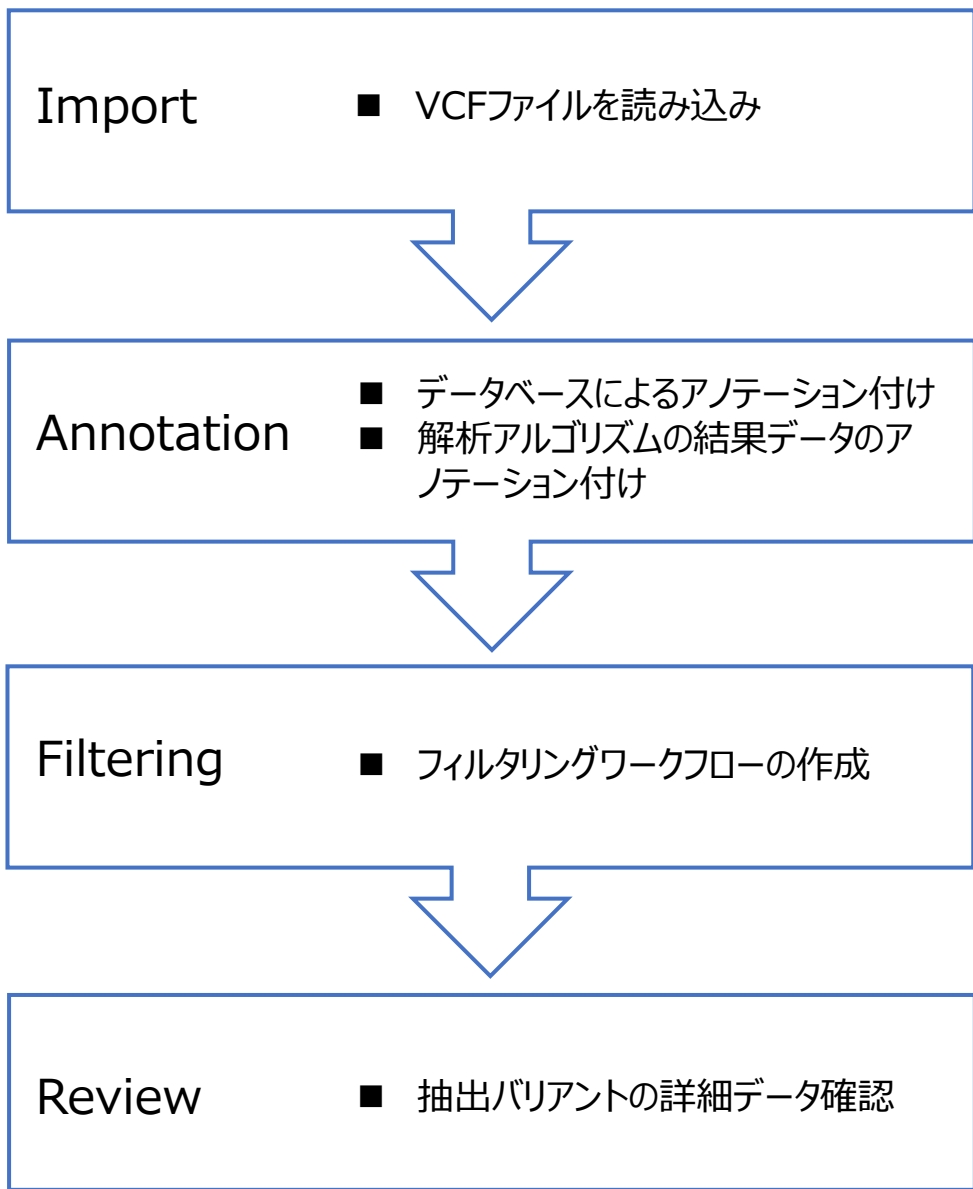
- RefSeq Genes
- dbSNP
- ClinVar
- OMIM
- CIViC
- ICGC / TCGA
- MSK Impact
- Orphanet
- BRCA Exchange
- PMKB
- dbNSFP
- REVEL
- CADD
- 1000 Genomes
- NHLBI 6500 Exomes
- ExAC Variant
- gnomAD Exomes/Genomes
- GenomeAsia 100K
- 各種遺伝子パネルのターゲットデータ ...など

- アノテーション付けされたバリエントデータより、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

- ゲノムブラウザにより、サンプルのバリエントデータやリードアライメントデータ（BAM/CRAMファイル）、また各種アノテーションを可視化

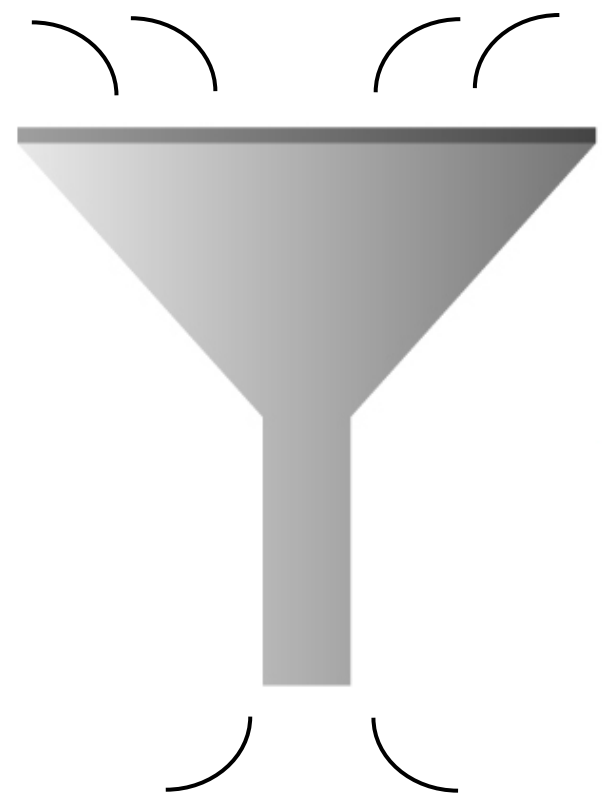
- 有償アドオンによる機能拡張で、CNV検出や臨床的意義の自動評価、パイプライン機能などが利用可能









```

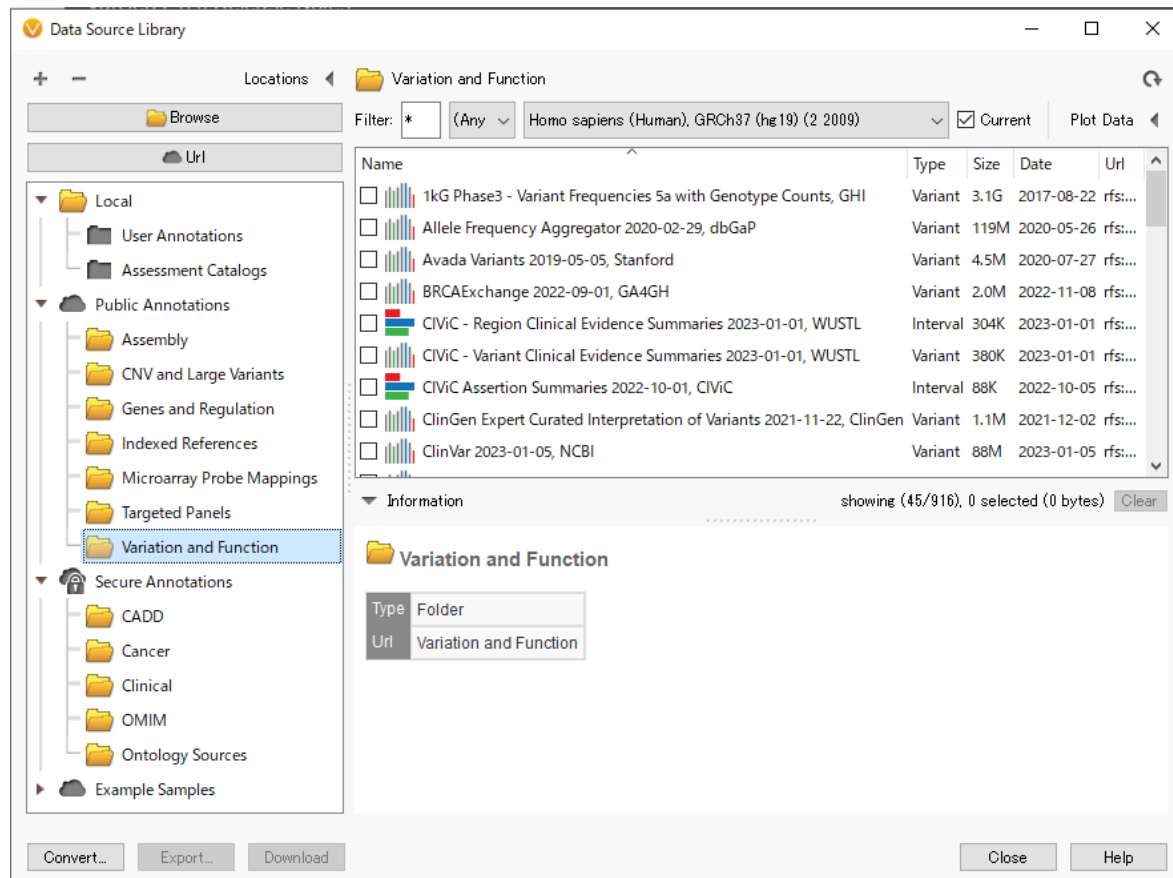
ACTGC GGGGAGACCTAGGCGGCTCTGCGGAC
GCAGCTCCTTCGC CGCCTTCCCCCTCCCGTCC
AGTGCC CAGGCGGCTCCTGGCTGCGACGCTG
CCCTGGGTGGGAGGCGCGGC ACCGCGGCAG
CTGAGCC CTCTGCGCGGCGCAGCCAGCTCTC
CCGCTCGCGCGGCGCCGTGACAGGTGCAGGG
TCCCCGCCCG AGACCCACCTGCAGGGGCTGT
  
```

VCF



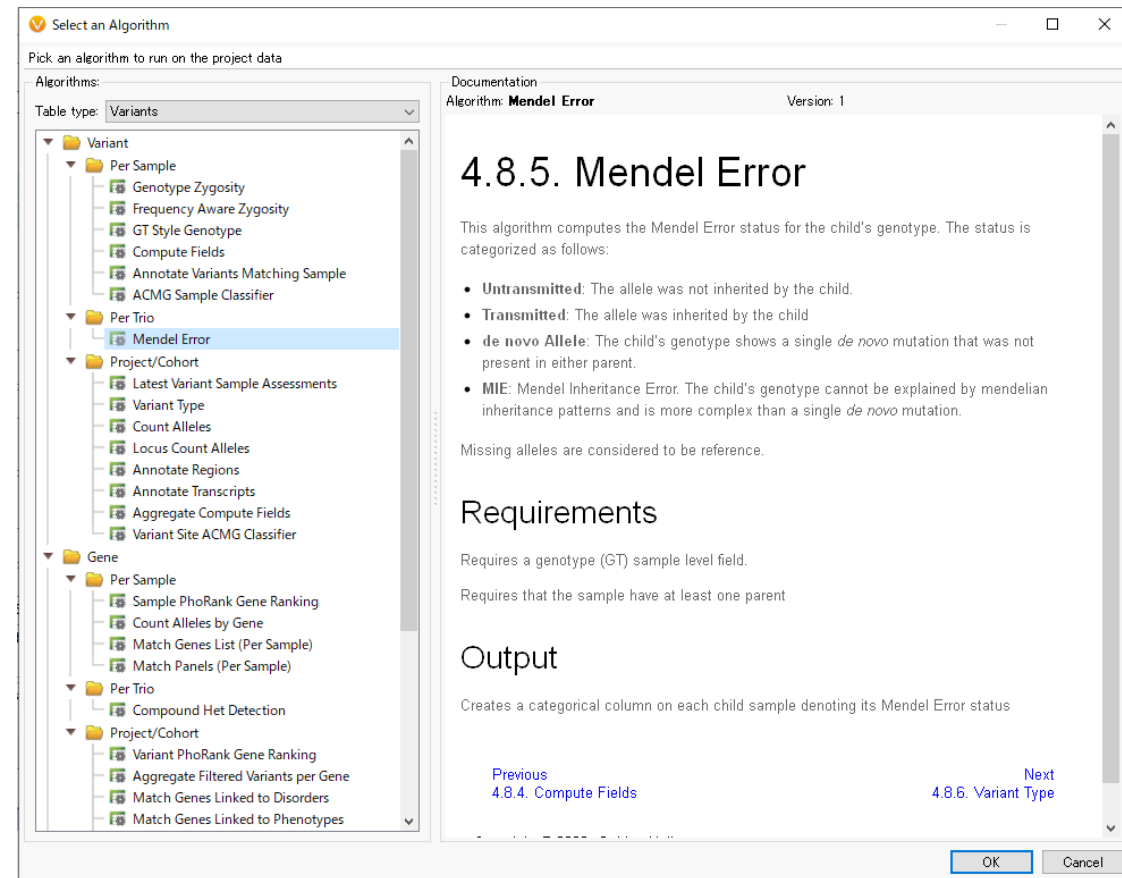
GCACC
病原性バリアント

遺伝子データベース 	アレル頻度データベース 	臨床情報データベース 
表現型ランキング 	トリオ解析 	



アノテーションリソースライブラリー

- アノテーションに用いるデータベースは、最新のデータをアノテーションリソースライブラリーよりダウンロード
- 外部VCFやBEDファイルなども使用可能



解析アルゴリズム選択画面

- 解析アルゴリズムは一覧より選択するだけで実行可能

バリエーションテーブルとフィルタリングワークフロー作成

Variant Info		HD200-rep1		RefSeq Genes 105.20220307, NCBI				gnomAD Exomes Variant		
Chr:Pos	Ref/Alt	Variant Allele Freq	Read Depths (DP)	Gene Names	SequenceOntology...	Effect (Combined)	HGVS p. (Clinically Relevant)	Alt Allele Freq (AF)	Alt Allele Count (AC)	#Alleles (AN)
17:29508775	G/A	0.602041	294	NF1	synonymous_variant	Other	NP_000258.1:p.Leu234=	0.627192	157498	251116
17:29553485	G/A	0.376699	516	NF1	synonymous_variant	Other	NP_000258.1:p.Pro678=	0.377777	94603	250420
17:36098040	G/A	0.589091	275	HNF1B	intron_variant	Other	?	?	?	?
17:36101156	T/C	0.560897	312	HNF1B	intron_variant	Other	?	?	?	?
17:41234451	G/A	0.241218	427	BRCA1	stop_gained	LoF	NP_009225.1:p.Arg1443Ter	2.38698e-05	6	251364
17:59760996	A/G	0.996387	1107	BRIP1	synonymous_variant	Other	NP_114432.2:p.Tyr1137=	0.5993	149716	249818
17:59763347	A/G	1	772	BRIP1	missense_variant	Missense	NP_114432.2:p.Ser919Pro	0.601855	151298	251386
17:59763465	T/C	1	494	BRIP1	synonymous_variant	Other	NP_114432.2:p.Glu879=	0.717913	180347	251210
17:74475024	G/A	0.945205	219	RHBDF2	missense_variant	Missense	NP_078875.4:p.Pro208Leu	0.885456	211523	238886
18:43316966	T/C	0.895189	582	SLC14A1	intron_variant	Other	?	?	?	?
18:46453463	T/C	0.253846	650	SMAD7	intron_variant	Other	?	?	?	?
19:30296853	T/C	0.271583	556	?	intergenic_variant	Other	?	?	?	?
19:45867259	C/T	0.36534	427	ERCC2	missense_variant	Missense	NP_000391.1:p.Asp312Asn	0.290034	41034	141480
19:45868309	T/G	0.482085	614	ERCC2	synonymous_variant	Other	NP_000391.1:p.Arg156=	0.579625	144875	249946
19:51364623	A/G	0.87061	541	KLK3	downstream_gene...	Unknown	?	?	?	?

アノテーション付けされたバリエーションテーブル

- アノテーション付けを実行すると、VCFファイルのデータ（バリエーションテーブル）にデータフィールドが追加される
- 追加した任意のフィールドを使用し、フィルタリングのためのワークフローを作成

The screenshot shows a filtering workflow interface with three filter steps. Each step has a checkbox, a filter name, a value input field, and a bar chart showing the distribution of variants.

- Step 1:** Read Depths (DP) (Current) >= 100. Value: 100. Distribution: Less than 100 (33,823), Equal to 100 (6), Greater than 100 (1,074), Missing (105,652).
- Step 2:** Alt Allele Freq (AF) < 0.01 OR missing. Value: 0.01. Distribution: Less than 0.01 (15), Equal to 0.01 (0), Greater than 0.01 (278), Missing (787).
- Step 3:** Effect (Combined) is (LoF, Missense). Distribution: Invalid (0), LoF (1), Missense (12), Other (768), Unknown (21), Missing (0).

フィルタリングワークフロー

Variant Info

Chr.Pos: 17:41234451
Ref/Alt: G/A

Show 19 hidden fields

Sample Fields

Sample: HD200-rep1
Variant Allele Freq: 0.241218
Read Depths (DP): 427

Show 7 hidden fields | Show 5 hidden samples

RefSeq Genes 105.20220307, NCBI

Gene Names	BRCA1
Sequence Ontology (Combined)	stop_gained
Effect (Combined)	LoF
N of 4 Predicted Splicing Disrupted (Combined)	?
Predicted Splicing Disrupted (Combined)	?
Splice Site Effect (Combined)	?
Transcript Name (Clinically Relevant)	NM_007294.4
HGVS c. (Clinically Relevant)	NM_007294.4:c.4327C>T
HGVS p. (Clinically Relevant)	NP_009225.1:p.Arg1443Ter

Show 24 hidden fields

ClinVar 2023-01-05, NCBI

Ref/Alt	1 G/A
Variant ID	17675
Classification	Pathogenic
Clinical Significance	Pathogenic
Aggregate of Interpretations from Submissions	Pathogenic (42)
Review Status	(3 Stars) Reviewed By Expert Panel



リードアライメント表示

- 詳細表示エリアにて、1バリアントごとに全アノテーションフィールドの詳細データを確認
- BAM/CRAMファイルがあれば、ゲノムブラウザーにリードアライメントを表示させ、バリアントの目視チェックなども可能

アノテーション詳細表示

シーケンスクオリティーによるフィルター

- ✓ Read Depth
- ✓ Variant Allele Frequency (VAF)
- ✓ Genotype Quality

遺伝子機能に影響を及ぼすバリアントの抽出

- ✓ Missense Variant
- ✓ Nonsense Variant
- ✓ Splice Region Variant

表現型と関連する遺伝子のバリアントの抽出

- ✓ PhoRank Gene Ranking
- ✓ Genes Linked to Phenotypes

The screenshot shows a 'Filter Variants' window with a list of filters. The filters are: Filter Variants (7,034), Variant Quality (2,709), Remove Common (341), Gene Impact (116), Remove Benign (113), and Human Phenotype (3). The 'Remove Benign' filter is highlighted with a red border.

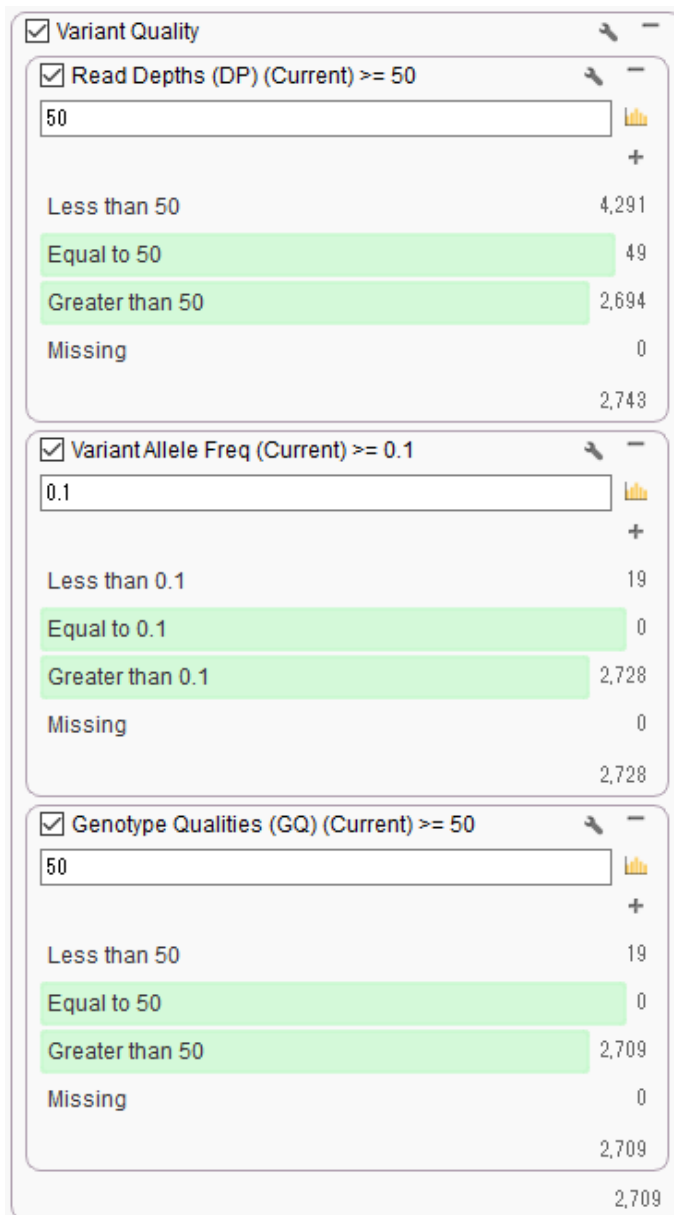
Filter	Count
Filter Variants	7,034
Variant Quality	2,709
Remove Common	341
Gene Impact	116
Remove Benign	113
Human Phenotype	3

コモンバリアントの除去

- ✓ gnomAD
- ✓ 1000 Genomes

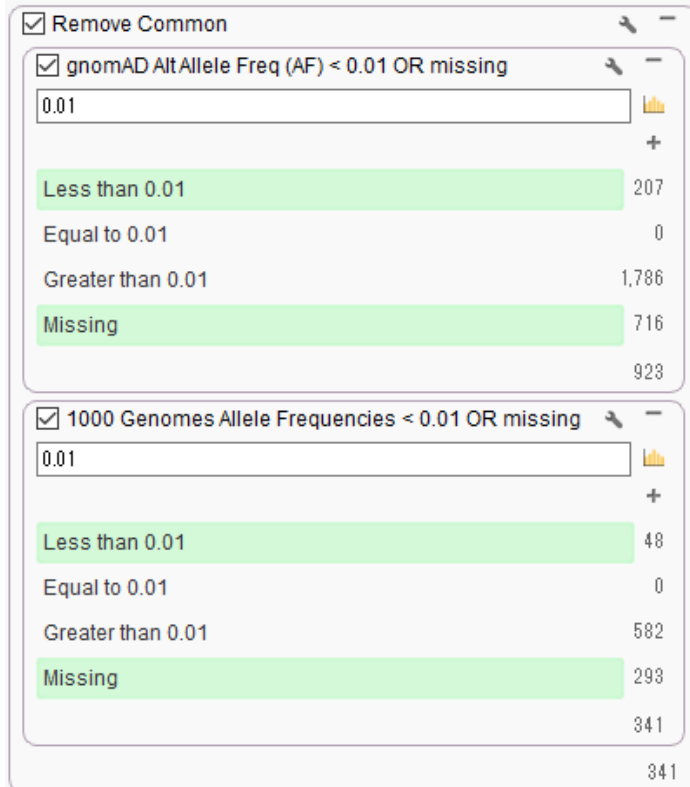
良性バリアントの除去

- ✓ ClinVar
- ✓ ACMG Classifier



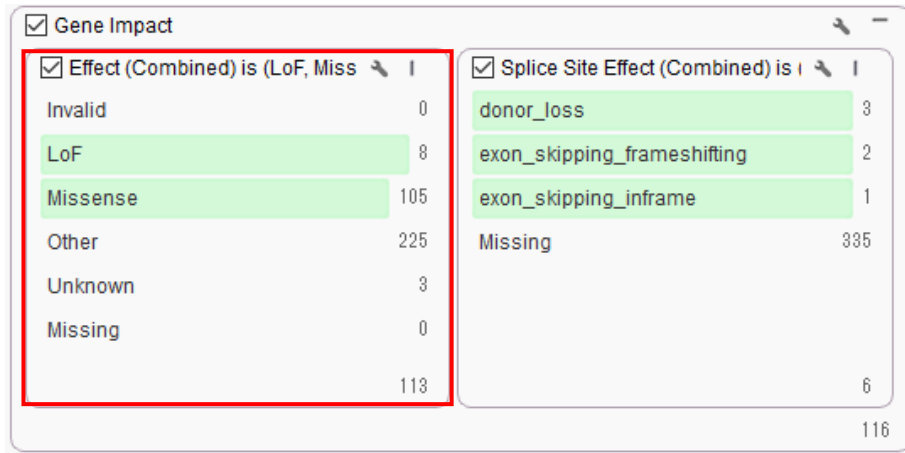
- VCFファイルに含まれているバリエーションごとのシーケンスクオリティの値でフィルターを行う
 - ✓ リード深度 (Read Depth)
 - ✓ バリエーションアレル頻度 (Variant Allele Frequency)
 - ✓ 遺伝型クオリティ (Genotype Quality)
- シーケンス実験のノイズデータなどの除去に使用
- カットオフ値はシーケンスプラットフォームやデータ量、サンプルの種類により異なる

Variant Info			P12878		
Chr:Pos	Ref/Alt	Identifier	Variant Allele Freq	Read Depths (DP)	Genotype Qualities (GQ)
6:564567	C/G	?	0.393443	122	99
6:7405508	G/A	rs56067778	0.457143	70	99
6:8428249	A/C	?	0.505155	194	99
6:10626704	T/G	rs200336999	0.409836	122	99
6:10756728	C/T	rs72821581	0.126437	87	99
6:12124988	T/C	?	0.433862	189	86.81



- バリエントの一般集団内のアレル頻度でフィルターを行う
- 一般集団内のアレル頻度が高いバリエントは良性とみなし、集団内に存在しないものや頻度が低いものを病理性バリエントとして抽出
- VarSeqでは様々な種類のデータベースが用意されていて、複数のデータベースを組み合わせてフィルタリングに使用することが可能
 - ✓ gnomAD Exome/Genome
 - ✓ 1000 Genomes
 - ✓ ExAC
 - ✓ NHLBI ESP6500
 - ✓ TOPMed
 - ✓ GenomeAsia 100K

gnomAD Exomes Variant Frequencies 2.1.1, BROAD				
Ref/Alt	Filter	Alt Allele Freq (AF)	Alt Allele Count (AC)	
A/T	PASS	0.999482	251070	
?	?	?	?	
A/G	PASS	0.999094	250271	
G/A	InbreedingCoeff	0.367252	81029	
?	?	?	?	
?	?	?	?	
G/A	PASS	0.063143	15859	



RefSeq Genes 105.20220307, NCBI			
Gene Names	Sequence Ontology (Combined)	Gene Region (Combined)	Effect (Combined)
VNN1	missense_variant	exon	Missense
VNN3P	intron_variant	intron	Other
RPS12	intron_variant	intron	Other
EYA4	stop_gained	exon	LoF
EYA4,TARID	intron_variant	intron	Other
TCF21	5_prime_UTR_variant	UTR5	Other
SLC2A12	missense_variant	exon	Missense

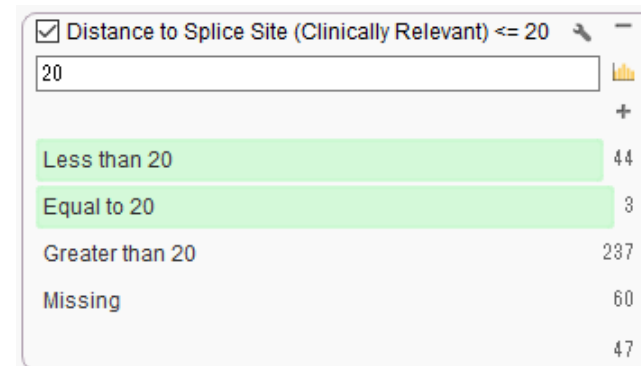
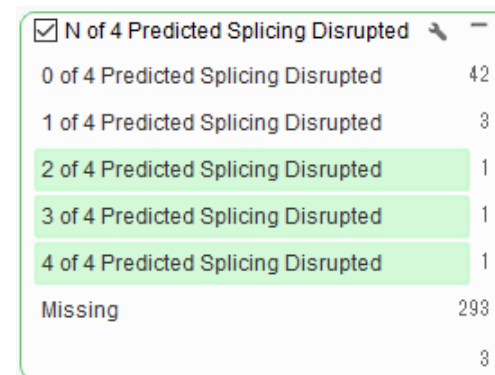
- タンパク質のアミノ酸配列の変化などの遺伝子機能に影響を及ぼすバリアントを抽出
- RefSeqによる遺伝子アノテーション付加により、ミスセンスバリアントや機能喪失バリアント（ナンセンス・フレームシフト・スプライシングドナー／アクセプター部位バリアントなど）を抽出可能
- 配列オントロジーや遺伝子領域に関するアノテーションも同時に付加され、フィルタリングに利用することも可能

Category	Count	Percent
intron_variant	1052	38.83%
missense_variant	596	22.00%
synonymous_variant	527	19.45%
intergenic_variant	235	8.67%
splice_region_variant	95	3.51%
non_coding_exon_variant	73	2.69%
3_prime_UTR_variant	55	2.03%
5_prime_UTR_variant	37	1.37%
stop_gained	10	0.37%



RefSeq Genes 105.20220307, NCBI			
N of 4 Predicted Splicing Disrupted (Combined)	Predicted Splicing Disrupted (Combined)	Splice Site Effect (Combined)	Distance to Splice Site...
4 of 4 Predicted Splicing Disrupted	GeneSplicer, MaxEntScan, NNSplice, PWM	donor_loss	3,1858
3 of 4 Predicted Splicing Disrupted	MaxEntScan, NNSplice, PWM	donor_loss	0
1 of 4 Predicted Splicing Disrupted	MaxEntScan	exon_skipping_frameshifting	4
1 of 4 Predicted Splicing Disrupted	NNSplice	exon_skipping_inframe	16
1 of 4 Predicted Splicing Disrupted	NNSplice	exon_skipping_frameshifting	9
2 of 4 Predicted Splicing Disrupted	MaxEntScan, PWM	donor_loss	0

- RefSeqでは、*In Silico*予測されたスプライシング異常を引き起こすバリエーションに対しても、アノテーション付けとフィルタリングが可能
- 以下4種類のプログラムで予測された、エクソンスキップなどのスプライシング異常や何種類のプログラムで予測が一致したかのデータでフィルタリングを実行
 - ✓ GeneSplicer
 - ✓ NNSplice
 - ✓ MaxEntScan
 - ✓ PWM
- スプライシング部位からの距離が近いバリエーションを抽出することで、未知のスプライシング異常の候補バリエーションを抽出することも可能



Remove Benign	
Benign	1
Conflicting	1
Likely Benign	2
Likely Pathogenic	0
Likely Pathogenic/Likely Risk Allele	0
Other	0
Pathogenic	1
Pathogenic/Likely Pathogenic/Likely Risk Allele	0
Pathogenic/Likely Risk Allele	0
Pathogenic/Pathogenic, Low Penetrance	0
Uncertain Risk Allele	0
Uncertain Significance	3
Uncertain Significance/Uncertain Risk Allele	0
Missing	108
	113

- 既知の良性（Benign）バリエントの除去を行う
- ClinVarに登録されているバリエントの分類情報で、「Benign」と「Likely Benign」と一致しないものだけをフィルター抽出する
- ClinVarの代わりにVSClinical ACMG Classifier（有償アドオンが必要）を用いることで、ACMGガイドラインで良性と判断されたバリエントを除去することも可能

ClinVar 2023-01-05, NCBI					
Ref/Alt	Variant ID	Classification	Clinical Significance	Review Status	
G/T	41549	Benign	Benign	(3 Stars) Reviewed By Expert Panel	
?	?	?	?	?	
G/A	1475824	Uncertain Significance	Uncertain significance	(1 Stars) Criteria Provided, Single Submitter	
?	?	?	?	?	
C/A	1486292	Uncertain Significance	Uncertain significance	(1 Stars) Criteria Provided, Single Submitter	
G/A	17675	Pathogenic	Pathogenic	(3 Stars) Reviewed By Expert Panel	

ACMG Sample Classifier for ACMG Sample 1					
Ref/Alt	Gene Name	Gene Inheritance	Sequence Ontology	Classification	ACMG Classification Criteria
T/C	PRDM16	Dominant	missense_variant	Benign	BA1,BP6
?	?	?	?	?	?
A/C	PRDM16	Dominant	missense_variant	VUS/Weak Pathogenic	PM2
A/G	PRDM16	Dominant	intron_variant	Benign	BA1,BS2,BP4,BP7
G/A	PRDM16	Dominant	intron_variant	Likely Benign	PM2,BP4,BP7
A/C	PRDM16	Dominant	intron_variant	Likely Benign	PM2,BP4,BP7

Human Phenotype

Less than 0.9 23

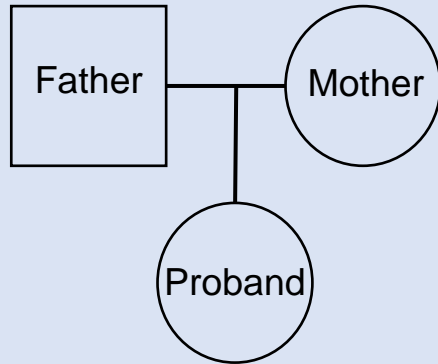
Equal to 0.9 0

Greater than 0.9 3

Missing 89

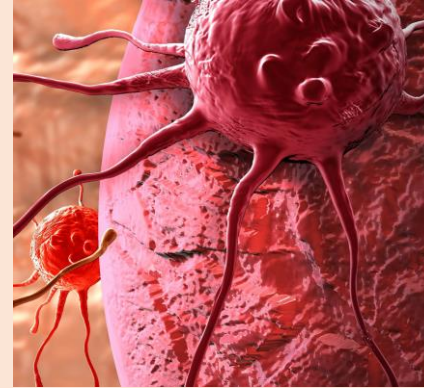
- サンプルの表現型を表すHPO (Human Phenotype Ontology) タームを入力し、入力タームと関連が強い遺伝子に優先順位付けを行う
- 順位は0~1の数値で付けられ、1に近いほど関連が強い遺伝子ということになる
- 順位付けまで行う場合はPhoRank Gene Rankingアルゴリズムを、関連する遺伝子を抽出するだけならGenes Linked to Phenotypesアルゴリズムを使う

Gene		Global Developmental Delay PhoRank Clinical				Match Genes Linked to Phenotypes	
Gene Names	Gene Rank	Gene Score	Path	Directly Associated Phenotypes	Directly Associated Phenotype Count	Global Developmental Delay Genes	Match
MICAL1	0.257334	0.278178	MICAL1 / HP...	?	0	False	?
FIG4	0.593668	0.533	FIG4 / HP:00...	Global developmental delay	1	True	FIG4
MFSD4B	?	?	?	?	0	False	?
LAMA4	?	?	?	?	0	False	?
FAM184A	?	?	?	?	0	False	?
TBC1D32	?	?	?	?	0	False	?
GJA1	0.423177	0.515996	GJA1 / HP:00...	Global developmental delay	1	True	GJA1
KIAA0408,SOGA3-KIAA04...	??	??	??	?	0,0	False	?
PTPRK	?	?	?	?	0	False	?
L3MBTL3	?	?	?	?	0	False	?
EPB41L2	?	?	?	?	0	False	?
MED23	0.975893	0.621361	MED23 / HP:...	Global developmental delay	1	True	MED23



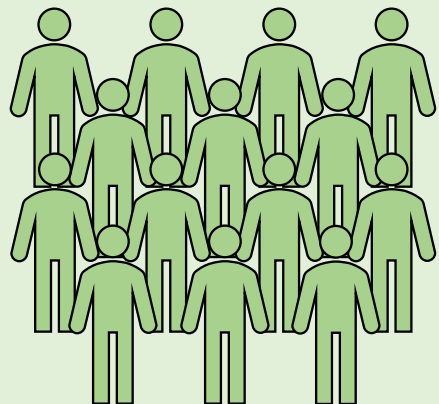
遺伝性疾患

- ✓ トリオ解析



がん

- ✓ 腫瘍 - 正常組織ペア解析
- ✓ がん専用データベースによるアノテーション



集団ゲノム解析

- ✓ 集団内のアレルカウント



バリエントの機能予測

- ✓ 病原性予測の分類
- ✓ 病原性スコアの計算

■ トリオ解析

- VCFファイルにGTフィールド（1/1, 0/1など）が書き込まれている場合は、バリエーションテーブルにHomozygous、Heterozygousなどの接合性データのフィールドを表示可能
- トリオ（患者と両親）などの複数サンプルのVCFファイルを読み込み、バリエーションテーブルに全サンプルのデータを表示させることで、各サンプルのZygosityフィールドごとにフィルター項目を作成できる

Zygosity (Current) is Homozygous Variant	
Heterozygous	544
Homozygous Variant	164
Reference	0
Missing	0
	164

← 患者の接合性

Zygosity (Mother) is Heterozygous	
Heterozygous	36
Homozygous Variant	107
Reference	0
Missing	21
	36

← 母親の接合性

Zygosity (Father) is Heterozygous	
Heterozygous	21
Homozygous Variant	14
Reference	0
Missing	1
	21

← 父親の接合性

Proband (P12878)			Mother (M12892)			Father (F12891)		
VAF	DP	Zygosity	VAF	DP	Zygosity	VAF	DP	Zygosity
1	56	Homozygous Variant	0.506024	250	Heterozygous	0.5	42	Heterozygous
1	86	Homozygous Variant	0.4	145	Heterozygous	0.5	88	Heterozygous
1	62	Homozygous Variant	0.454936	233	Heterozygous	0.491228	57	Heterozygous
1	64	Homozygous Variant	0.53012	250	Heterozygous	0.402778	72	Heterozygous
0.982759	59	Homozygous Variant	0.476923	65	Heterozygous	0.426667	75	Heterozygous
0.988889	90	Homozygous Variant	0.475728	103	Heterozygous	0.421687	83	Heterozygous
1	44	Homozygous Variant	0.48	77	Heterozygous	0.521127	71	Heterozygous
0.818182	33	Homozygous Variant	0.4	36	Heterozygous	0.54386	59	Heterozygous
1	67	Homozygous Variant	0.342466	75	Heterozygous	0.608108	74	Heterozygous
1	103	Homozygous Variant	0.326087	46	Heterozygous	0.495495	111	Heterozygous
1	250	Homozygous Variant	0.471311	244	Heterozygous	0.532	250	Heterozygous

- De Novoバリエントと複合ヘテロ接合体は、解析アルゴリズムを実行することによって抽出

Variant Info		Proband (HG02024)					Mother (HG02025)					Father (HG02026)				
Chr:Pos	Ref/Alt	VAF	DP	Mendel Error	Inherited From	Zygosity	VAF	DP	Mendel Error	Inherited From	Zygosity	VAF	DP	Mendel Error	Inherited From	Zygosity
1:1225553	C/T	0.666667	12	Transmitted	Father	Heterozygous	?	?	?	?	?	0.483871	31	?	?	Heterozygous
1:1225558	T/C	0.214286	14	de Novo Allele	?	Heterozygous	?	?	?	?	?	?	?	?	?	?
1:1225563	-/CC	0.25	12	de Novo Allele	?	Heterozygous	?	?	?	?	?	?	?	?	?	?
1:1225579	G/C	0.5	24	Transmitted	Mother	Heterozygous	1	73	?	?	Homozygous Variant	0.438596	57	?	?	Heterozygous
1:1225612	G/A	1	71	Transmitted	Both	Homozygous Variant	0.398496	133	?	?	Heterozygous	0.472222	109	?	?	Heterozygous
1:1225641	C/A	1	92	Transmitted	Both	Homozygous Variant	0.378698	169	?	?	Heterozygous	1	127	?	?	Homozygous Variant
1:1225959	C/G	0.4375	144	Transmitted	Father	Heterozygous	0.461279	298	?	?	Heterozygous	1	241	?	?	Homozygous Variant
1:1226102	G/A	?	?	Untransmitted	?	?	0.433526	173	?	?	Heterozygous	0.439716	143	?	?	Heterozygous
1:1226221	A/G	1	28	Transmitted	Both	Homozygous Variant	1	70	?	?	Homozygous Variant	1	70	?	?	Homozygous Variant

De Novoバリエント

Compound Het Genes for Proband (HG02024)				Variant Info		Proband (HG02024)			Mother (HG02025)			Father (HG02026)			Gene Names
Gene Names	Has Compound Het?	Inherited from Father	Inherited from Mother	Chr:Pos	Ref/Alt	VAF	DP	Zygosity	VAF	DP	Zygosity	VAF	DP	Zygosity	Gene Names
AHNAK	True	1	1	11:62289513	C/T	0.477212	373	Heterozygous	?	?	?	0.542403	569	Heterozygous	AHNAK
FLG	True	3	1	11:62299482	C/G	0.475	522	Heterozygous	0.469641	808	Heterozygous	?	?	?	AHNAK
KRT4	True	1	1												
MUC5AC	True	3	2												
MUC5B	True	1	1												
PCDHGA1	True	1	1												

複合ヘテロ接合体

Project Template

Folder: [Project Templates](#) Reset Browse...

*

- ACMG Guidelines Gene Panel
- ACMG Guidelines Trio
- Advanced ACMG Guidelines Gene Panel
- Cancer Gene Panel Starter Template
- Comprehensive Cancer Template
- Exome Trio Template**
- Hereditary Gene Panel Starter Template
- My Cancer Panel Template
- Tumor-Normal Template
- Empty Project

Exome Trio Template

A basic template for exome trio analysis.

Author
Golden Helix Inc.

Notes
The Exome Trio Template is a basic template for trio analysis. Included are filter chains for:
 * de Novo Variants
 * Dominant Heterozygous Variants
 * Compound Heterozygous Variants
 * Recessive Homozygous Variants
 * X-linked Variants (Excludes PAR)
 * Pathogenic in Incidental Finding Genes

Please note that variants passing the X-linked filter chain are only valid if the proband is male. If a female proband is being analyzed, we recommend deleting this filter chain.

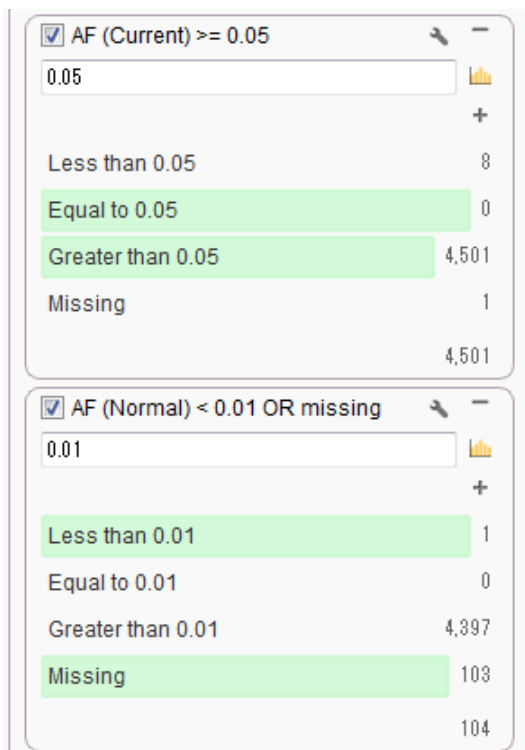
- ワークフローを分岐させることによって、複数の遺伝形式によるフィルタリングを同時に実行
- VarSeq標準搭載のレディメイドワークフローを使用することで、ユーザー自身でワークフローを構築する必要なく、De Novoバリエーションや優性ヘテロ・劣性ホモバリエーション、複合ヘテロ接合体やX連鎖バリエーションをまとめて抽出することも可能

Inheritance Model

<input checked="" type="checkbox"/> de Novo Candidate <input checked="" type="checkbox"/> Alt Allele Freq (AF) < 0.01 OR miss 3,754 <input checked="" type="checkbox"/> Effect (Combined) is (LoF, Missense) 149 <input checked="" type="checkbox"/> Updated inputs are <input checked="" type="checkbox"/> Mendel Error (Current) is de Novo. MIE 0 Transmitted 149 Untransmitted 0 de Novo Allele 0 Missing 0	<input checked="" type="checkbox"/> Dominant Heterozygous <input checked="" type="checkbox"/> Alt Allele Freq (AF) < 0.005 OR missing 3,710 <input checked="" type="checkbox"/> Effect (Combined) is (LoF, Missense) 132 <input checked="" type="checkbox"/> Updated inputs are available which may produc <input checked="" type="checkbox"/> Either Parent is Heterozygous <input checked="" type="checkbox"/> Father Het 132 <input checked="" type="checkbox"/> Mother Het 26 <input checked="" type="checkbox"/> Zygosity (Father) is Heterozygous 105 <input checked="" type="checkbox"/> Zygosity (Mother) is Heterozygous 105 <input checked="" type="checkbox"/> Zygosity (Mother) is (Reference) 105 <input checked="" type="checkbox"/> Zygosity (Father) is (Reference) 0 <input checked="" type="checkbox"/> Zygosity (Current) is Heterozygous 105	<input checked="" type="checkbox"/> Recessive Homozygous <input checked="" type="checkbox"/> Alt Allele Freq (AF) < 0.01 OR missing 3,754 <input checked="" type="checkbox"/> Effect (Combined) is (LoF, Missense) 149 <input checked="" type="checkbox"/> Updated inputs are av <input checked="" type="checkbox"/> Recessive Inheritance Model <input checked="" type="checkbox"/> Zygosity (Current) is Homozygous 0 <input checked="" type="checkbox"/> Zygosity (Mother) is Heterozygous 0 <input checked="" type="checkbox"/> Zygosity (Father) is Heterozygous 0	<input checked="" type="checkbox"/> Compound Heterozygous <input checked="" type="checkbox"/> Alt Allele Freq (AF) < 0.01 OR miss 3,754 <input checked="" type="checkbox"/> Effect (Combined) is (LoF, Missense) 149 <input checked="" type="checkbox"/> Updated inputs are <input checked="" type="checkbox"/> Compound Het? (Current) is true True 0 False 149 Missing 0	<input checked="" type="checkbox"/> X-Linked <input checked="" type="checkbox"/> Not In PAR regions 0 <input checked="" type="checkbox"/> Alt Allele Freq (AF) < 0.01 OR missing 0 <input checked="" type="checkbox"/> Zygosity (Current) is (Hemizygous, Homozygous Variant) Hemizygous 0 Heterozygous 0 Homozygous Variant 0 Reference 0 Missing 0
--	---	--	---	--

■ 腫瘍 – 正常組織ペア解析

- 腫瘍組織、正常組織の両サンプルのVCFファイルを読み込み、バリエントテーブルに両サンプルのデータを表示させることで、それぞれのバリエントアレル頻度フィールドごとにフィルター項目を作成できる
- 腫瘍組織サンプルではバリエントアレル頻度が大きく、かつ正常組織サンプルでは低い（または未検出）バリエントを抽出することで、体細胞バリエントの検出が可能



Variant Info		Tumor (T990005)		Normal (N990005)		GeneNames	SequenceOntologyCo...	EffectCom...
Chr:Pos	Ref/Alt	DP	AF	DP	AF			
3:9920151	G/A	200	0.100503	?	?	CIDEC	missense_variant	Missense
3:10088404	C/T	163	0.228395	?	?	FANCD2	synonymous_variant	Other
3:32031995	T/C	62	0.483871	?	?	ZNF860	missense_variant	Missense
3:41278119	C/A	249	0.325301	?	?	CTNNB1	missense_variant	Missense
3:46009157	C/A	69	0.130435	?	?	FYCO1	missense_variant	Missense
3:75714034	C/G	250	0.164	?	?	FRG2C	missense_variant	Missense
3:100170596	T/C	117	0.119658	?	?	LNP1	missense_variant	Missense
3:100170598	C/T	120	0.116667	?	?	LNP1	synonymous_variant	Other
3:100170600	A/T	122	0.139344	?	?	LNP1	missense_variant	Missense
3:148899778	G/C	194	0.238342	?	?	CP	intron_variant	Other

<input checked="" type="checkbox"/> Matched? is true	
True	5
False	6
Missing	0
	5

Summary of CIViC		CIViC - Region Clinical Evidence Summaries 2023-01-01, WUSTL			
Matched?	Mutation Type	Tumor Type	Oncotree Tissue Type	Gene Name	
True	Fusion,Fusion,Fusion,F...	Non-Small Cell Lung ...	Lung,Lung,Lung,Lung,L...	ALK,ALK,ALK...	
False	?	?	?	?	
True	Coding Mutation,Codi...	Bowel,Colorectal Ade...	Bowel,Bowel	APC,APC	
False	?	?	?	?	
True	Coding Mutation,Codi...	Medulloblastoma,?	CNS/Brain	PTCH1,PTCH1	
False	?	?	?	?	
True	Fusion,Fusion,Fusion,F...	Non-Small Cell Lung ...	Lung,Lung,Lung,Lung,L...	RET,RET,RET,...	

<input checked="" type="checkbox"/> Oncotree Tissue Type is (Bowel, Bre	
Adrenal Gland	0
Biliary Tract	0
Bladder/Urinary Tract	0
Bone	0
Bowel	3
Breast	2
CNS/Brain	1
Cervix	0

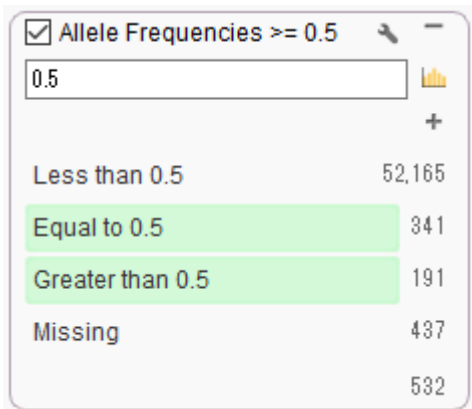
CIViC

The screenshot displays a genomic analysis interface. At the top, a DNA sequence is shown with a mutation highlighted in blue: G C C T A C G C C A C C A G C T C C A A C T A C C A C. Below the sequence, there are two question marks and two 'T' characters. A red box highlights three mutation annotations: G18, G12, and A11. To the right, the 'Extended Fields' panel shows details for the KRAS gene, including residue G13, amino acid counts, and tumor types.

Cancer Hotspots

■ がん専用データベースによるアノテーション

- がん研究に特化したデータベースが多数用意されており、アノテーション付けやフィルタリングに活用できる
- CIViCでは、がんに関連した遺伝子の抽出や臓器・治療エビデンス情報、Cancer Hotspotsでは変異のホットスポット部位によるフィルタリングが可能



■ 集団内のアレルカウント

- 多数のサンプルVCFファイルを同時に読み込んだ場合は、解析アルゴリズムにてサンプル集団内のバリエーションごとのアレル数や頻度などを計算し、フィルタリングに利用することが可能
- サンプル集団を複数のグループに分け、各グループごとにカウントを行うことも可能

Variant Info		ERR1949557		ERR1949559		ERR1949561		ERR1949563		ERR1949563		ERR1949555		Count Alleles					
Chr:Pos	Ref/Alt	VAF	DP	VAF	DP	VAF	DP	VAF	DP	VAF	DP	VAF	DP	Allele Counts	Allele Frequencies	# Alleles	# Het	# HomoVar	# Samples
1:928475	G/-	0.975936	375	1	121	0.942623	367	1	340	0.959184	49	0.976471	171	17	0.566667	30	11	3	14
1:928478	G/C	0.970509	374	1	122	0.942308	367	0.964286	337	0.958333	48	1	169	19	0.633333	30	11	4	15
1:928501	G/-	0.46729	1285	0.452736	403	0.453074	1236	0.529138	1287	0.832998	497	0.599232	781	15	0.5	30	15	0	15
1:928504	G/C	0.46087	1270	0.446154	392	0.455907	1205	0.525547	1233	1	476	0.605938	741	15	0.5	30	15	0	15
1:928520	A/G	1	1305	1	395	1	1260	1	1327	1	507	1	808	29	0.966667	30	1	14	15
1:928578	G/A	1	1315	1	404	1	1261	1	1309	0.980507	515	1	829	28	0.933333	30	2	13	15
1:928836	C/T	1	21	1	17	1	39	1	67	1	14	1	30	20	0.666667	30	2	9	11
1:928983	-/C	0.903226	31	0.869565	23	0.940298	67	0.956522	70	1	38	0.87931	60	18	0.6	30	8	5	13
1:929190	A/G	1	16	1	20	1	67	1	57	1	36	1	46	24	0.8	30	2	11	13
1:929316	C/T	1	15	1	14	1	47	1	35	1	32	1	46	25	0.833333	30	1	12	13
1:929321	A/C	1	14	1	15	1	46	1	37	0.9375	32	1	47	24	0.8	30	2	11	13
1:929327	A/G	1	15	1	15	1	46	1	38	1	32	1	48	25	0.833333	30	1	12	13
1:931014	A/G	1	177	1	52	1	144	1	147	1	154	1	188	28	0.933333	30	2	13	15
1:931730	A/G	1	13	?	?	1	58	1	4	1	17	1	18	18	0.6	30	2	8	10
1:932111	A/C	0.736842	19	1	14	0.769231	169	0.769231	57	1	42	0.8	56	15	0.5	30	11	2	13

dbNSFP Functional Prediction	
<input checked="" type="checkbox"/> SIFT Pred is D	
.	1
D	2
T	1
Missing	0
	2

N of 6 Predicted Damaging is	
0 of 6 Predicted as Damaging	0
1 of 6 Predicted as Damaging	0
2 of 6 Predicted as Damaging	1
3 of 6 Predicted as Damaging	0
4 of 6 Predicted as Damaging	1
5 of 6 Predicted as Damaging	1
6 of 6 Predicted as Damaging	0
Missing	0
	2

dbNSFP Functional Predictions and Scores 4.2, GHI					
SIFT Score	SIFT Converted Rankscore	SIFT Pred	MutationTaster Score	MutationTaster Converted Rankscore	MutationTaster Pred
0.024	0.47745	D	1	0.08975	N
0.001	0.7849	D	0.999736	0.48635	D
0.001	0.7849	D	0.641085	0.32877	D
1	0.00964	T	1	0.08975	P
0.19	0.21144	T	0.98494	0.24755	N

dbNSFP 4.2

dbNSFP Functional Prediction Voting				
N of 6 Predicted Tolerated	N of 6 Predicted Damaging	SIFT Pred (C)	Polyphen2 HVAR Pred (C)	MutationTaster Pred (C)
0 of 6 Predicted as Tolerated	6 of 6 Predicted as Damaging	Damaging	Possibly damaging	Damaging
1 of 6 Predicted as Tolerated	5 of 6 Predicted as Damaging	Damaging	Probably damaging	Damaging
0 of 6 Predicted as Tolerated	6 of 6 Predicted as Damaging	Damaging	Possibly damaging	Damaging
0 of 6 Predicted as Tolerated	6 of 6 Predicted as Damaging	Damaging	Probably damaging	Damaging
2 of 6 Predicted as Tolerated	4 of 6 Predicted as Damaging	Damaging	Possibly damaging	Damaging
2 of 6 Predicted as Tolerated	4 of 6 Predicted as Damaging	Damaging	Probably damaging	Damaging
2 of 6 Predicted as Tolerated	4 of 6 Predicted as Damaging	Damaging	Possibly damaging	Damaging
1 of 6 Predicted as Tolerated	5 of 6 Predicted as Damaging	Damaging	Probably damaging	Damaging

dbNSFP 3.0

■ 病原性予測の分類

- バリエーションのタンパク質の機能に与える影響を、タンパク質立体構造やアミノ酸配列の進化的保存度を様々なプログラムで評価したdbNSFPデータベースでアノテーション付け
- バージョン4.2では30種類以上のアルゴリズムの評価結果を個別にフィルタリングし、バージョン3.0ではプログラムは6種類のみだが、6種類中のいくつかのプログラムで有害と判定されたかという条件でフィルタリングが可能

PHRED Score >= 20 OR REVEL >= 0.2

PHRED Score >= 20

- Less than 20 0
- Equal to 20 0
- Greater than 20 3
- Missing 0

REVEL >= 0.2

- Less than 0.2 0
- Equal to 0.2 0
- Greater than 0.2 3
- Missing 0

3

CADD Scores 1.6				REVEL Functional Predictions 2016-06-03, GHI			
Ref/Alt	Raw Score	PHRED Score	Estimated?	Ref/Alt	AA Ref	AA Alt	REVEL
T/G	3.85956	26.6	False	T/G	F	V	0.872
A/G	4.7157	32	False	A/G	E	G	0.736
C/G	2.85034	23.2	False	C/G	R	P	0.166
T/G	2.6005	22.7	False	T/G	H	P	0.374
A/G	1.01983	13.56	False	A/G	S	P	0.075
T/C	-0.703073	0.019	False	T/C	K	R	0.015
C/T	1.85472	18.61	False	C/T	R	K	0.153
G/C	2.76662	23.1	False	G/C	A	G	0.393

■ 病原性スコアの計算

- タンパク質への影響をスコア化したデータベースとして、CADD（有償アドオンが必要）とREVELが利用可能
- CADDはすべてのSNVと既知のInDelsに対し、ヒトリファレンスゲノム上で取り得る全パターンのバリエーションのランクを1 - 99の範囲でPHREDスコアとして算出（上位10%ではPHREDスコアが10、上位1%では20、上位0.1%では30）
- REVELはミスセンスバリエーションに対して、SIFTやPolyphen-2などの13個の機能予測プログラムからのスコアの組み合わせに基づき病原性を予測したもので、0 - 1の範囲のスコアが算出される

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