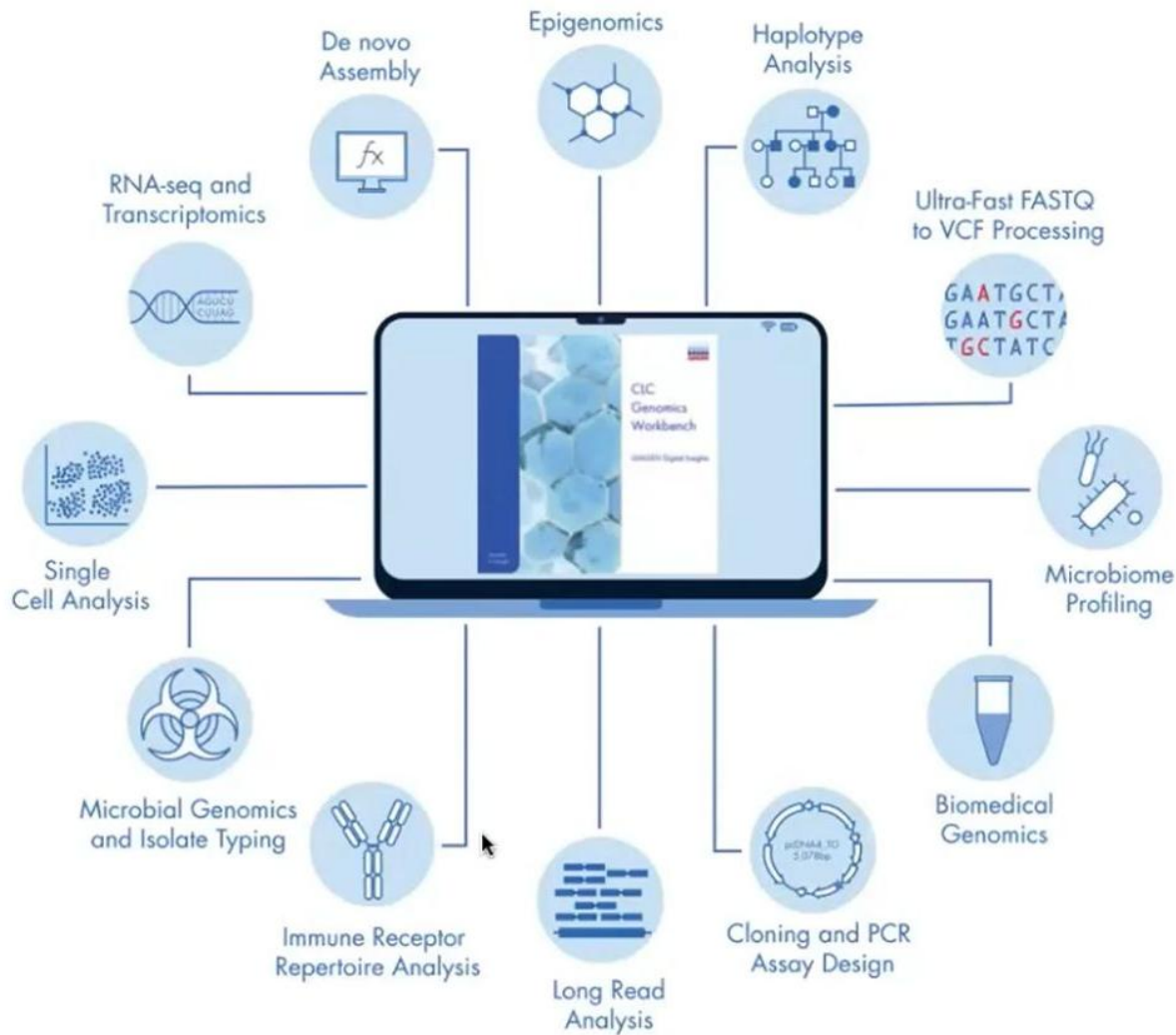


全ゲノム・全エクソーム・ターゲットシーケンス のバイオインフォマティクス解析

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

- 次世代シーケンサーによるゲノムリシーケンス解析では、シーケンサーより得られたリードデータのクオリティコントロールやリファレンスゲノムへのマッピング、変異の検出といった複数のステップを経て解析を実行します。
- キアゲン社CLC Genomics Workbenchの無償プラグイン Biomedical Genomics Analysisには、全ゲノムや全エクソーム、ターゲットシーケンス解析用のワークフローが搭載されていて、これらのデータ解析を簡単に実行することができます。





■ 対応プラットフォーム

- Illumina
- Ion Torrent
- MGI
- PacBio/Nanoporeなど

■ サポートOS

- Windows
- Mac
- Linux

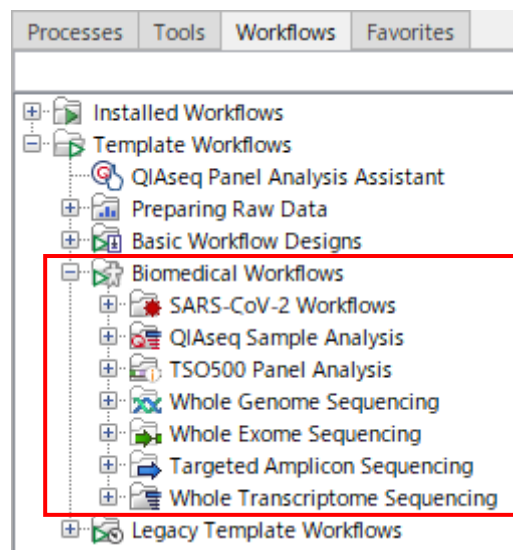
■ NGS用のアプリケーションの他、サンガーシーケンサー用のアプリケーションも搭載

- 波形データファイル（.abiファイルなど）のインポート
- アセンブルによるコンティグ作成
- マルチプルアライメント
- 系統樹作成
- BLAST
- プライマー設計
- 分子クローニング

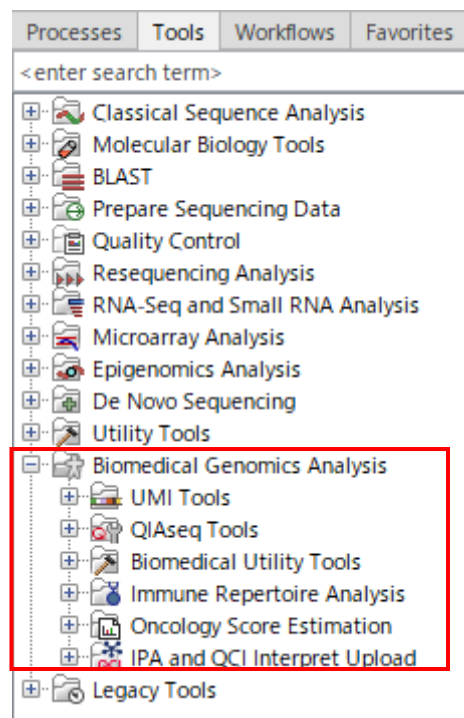
■ 無償利用可能な、ビューワー版ソフトを提供

■ バッチ解析による、大量サンプルの一括処理

■ ワークフロー作成

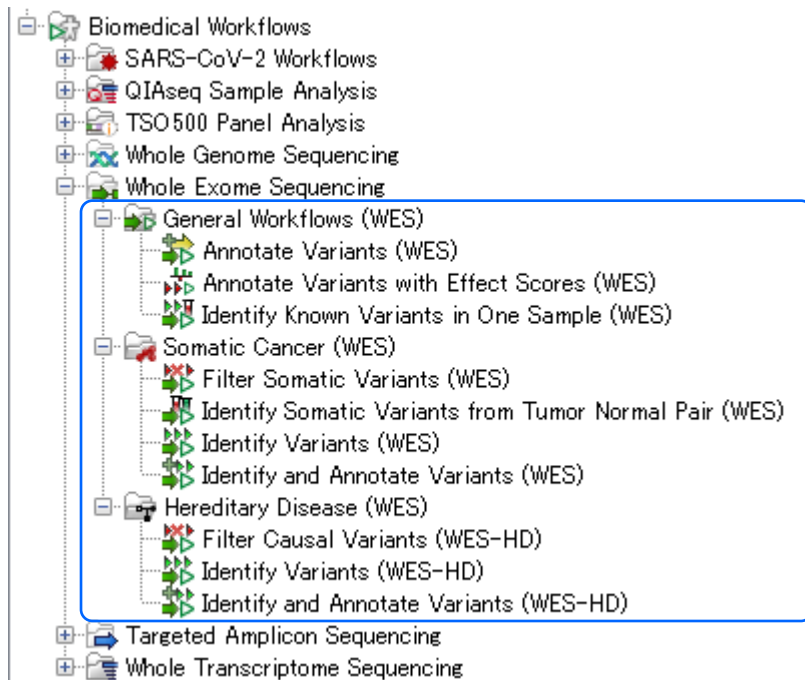


ワークフロー



ツール

- おもにヒト検体におけるリシーケンス解析機能を拡張するための無償プラグイン
- Biomedical Genomics Analysisプラグインには、医学研究用の変異解析ワークフローが標準で搭載されており、これらワークフローを使用するだけで、各解析を一気に実行することが可能
- ワークフローの他、おもにキアゲン社QIASeqパネル解析に使用する分子バーコード解析やレパトア解析用のツールも搭載



Whole Genome / Whole Exome / Target Amplicon Sequencing:

General Workflow

- 検出した変異に対するアノテーション(遺伝子、アミノ酸置換等)
- ユーザーのカスタムデータの変異がサンプルに含まれるか否か検証

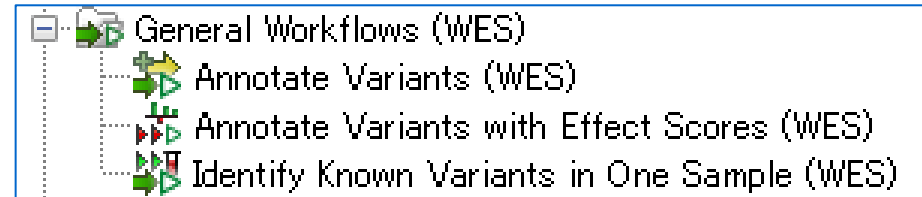
Somatic Cancer

- がんサンプル解析を主としてデザインされた、体細胞変異検出のためのワークフロー
- 1サンプルやペアサンプル(正常・病変)を用いた変異検出・アノテーション

Hereditary Disease

- 遺伝性疾患サンプル解析を主としてデザインされた、生殖細胞系列変異検出のためのワークフロー

- ✓ General Workflowは、がん／遺伝性疾患のサンプルの種類に関わらず使用することのできるワークフロー
- ✓ ワークフローは2または3種類あり、Whole Genome、Whole Exome、Target Amplicon Sequencingそれぞれの、General Workflowsフォルダにある



Annotate Variants:

- 変異に遺伝子名、保存スコア、アミノ酸置換、臨床関連変異データベースの情報を付加

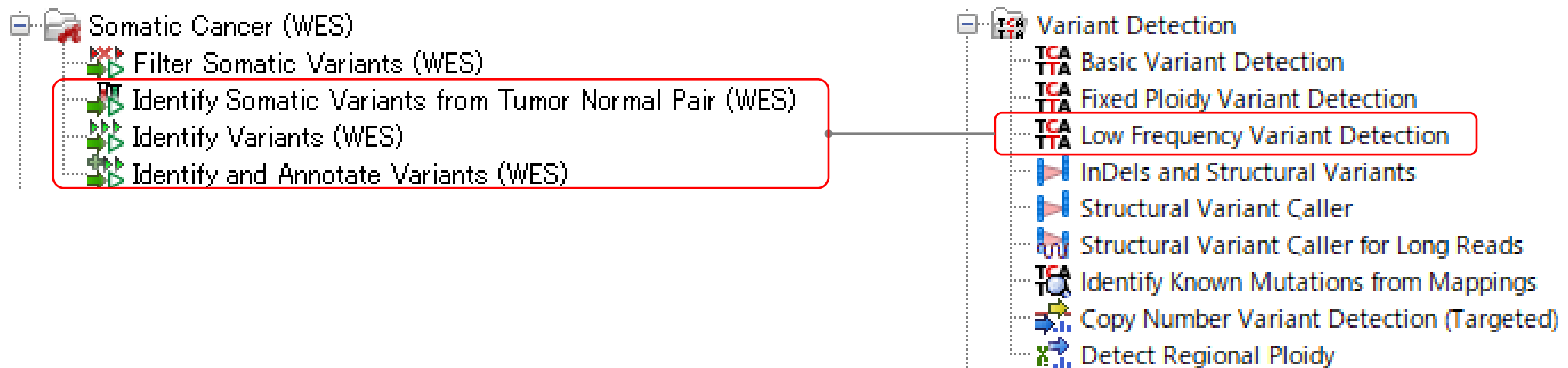
Annotate Variants with Effect Scores (WESのみ):

- 変異にSIFT effect score の情報を付加

Identify Known Variants in One Sample:

- シークエンスリードをリファレンスにマップし、そのマッピングデータにおいて、ユーザー定義の変異の有無を検証
(ユーザー定義以外の変異は検出されない)

- ✓ Somatic Cancerはがんサンプルの解析用にデザインされたワークフローを収録している
- ✓ ワークフローは、Whole Genome、Whole Exome、Target Amplicon Sequencingそれぞれの、Somatic Cancerフォルダにある
- ✓ ワークフローに搭載されている変異検出ツールとして、体細胞変異検出のために低頻度変異の同定に対応した、Low Frequency Variant Detectionを採用している



Filter Somatic Variants:

- 変異データから、SNPなどのコモンバリエントを除外することで、体細胞変異候補を取得
- 検出変異に対して遺伝子名や配列保存スコア、アミノ酸置換、臨床関連変異データベースの情報などをアノテーション

Identify Somatic Variants from Tumor Normal Pair:

- がんサンプルとコントロールサンプルのペア解析を行い、体細胞変異候補を検出
- 検出変異に対して遺伝子名や配列保存スコア、アミノ酸置換、臨床関連変異データベースの情報などをアノテーション

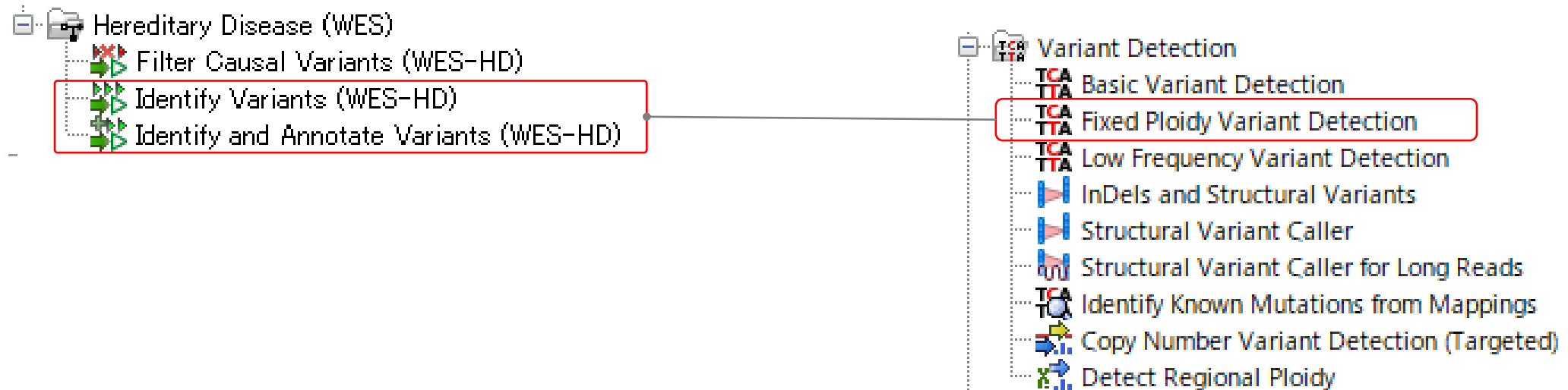
Identify Variants:

- リードをリファレンスにマッピングして変異を検出
- 検出した変異から偽陽性やターゲット領域外のものを除外

Identify and Annotate Variants:

- 上記Identify Variantsと、General WorkflowのAnnotate Variantsの2つを組み合わせたワークフロー

- ✓ Hereditary Diseaseは遺伝性疾患サンプル解析にデザインされたワークフローを収録してる
- ✓ ワークフローは、Whole Genome、Whole Exome、Target Amplicon Sequencingそれぞれの、Hereditary Diseaseフォルダにある
- ✓ ワークフローに搭載されている変異検出ツールとして、生殖細胞系列変異の検出のためにゲノムの倍数性を指定して実行する、Fixed Ploidy Variant Detectionを採用している



Filter Causal Variants (HD):

- 変異データから、SNPなどのコモンバリエントを除外
- 検出変異に対して遺伝子名や配列保存スコア、アミノ酸置換、臨床関連変異データベースの情報などをアノテーション

Identify Variants (HD):

- リードをリファレンスにマッピングして変異を検出
- 検出した変異から偽陽性やターゲット領域外のものを除外

Identify and Annotate Variants (HD):

- 上記Identify Variantsと、General WorkflowのAnnotate Variantsの2つを組み合わせたワークフロー

- ✓ Manage Reference Dataの「QIAGEN Sets」タブより、ワークフローに必要なリファレンスゲノムデータや、QIAseqパネル用プライマー情報データなどをダウンロード可能

Manage Reference Data

Download Genomes (Public Repositories) | **QIAGEN Sets (Reference Data Library)** | Custom Sets (Reference Data Sets) | Imported Data (Imported Reference Data)

Manage Reference Data: Locally

Free space in CLC_References location: 100.91 GB
Free space in temporary folder location: 100.91 GB

<enter search term>

Reference Data Sets

- hg38 (Ensembl)
Ensembl v113, dbSNP v151, ClinVar 20241103
- hg38 (Refseq)
RefSeq GRCh38.p14 with IPA transcript priorities, dbSNP v151, ClinVar 20241103**
- hg38 (Ensembl MANE)
Ensembl MANE v1.3, dbSNP v151, ClinVar 20241103
- hg38 (Refseq MANE)
Refseq MANE v1.3, dbSNP v151, ClinVar 20241103
- Single Cell hg38 (Ensembl)
Cell Type Classifier v1.2, Ensembl v113, IMGT reference sequences, Peak Shape Filter v1
- hg19 (Ensembl)
Ensembl v110, dbSNP v151, ClinVar 20241103
- hg19 (Refseq)
RefSeq GRCh37.p13, dbSNP v151, ClinVar 20241103
- QIAseq Multimodal Panels hg38

hg38 (Refseq)
Version: 4.1, Reference Data Set

Size on disk: 9.74 GB

Download Delete Create Custom Set ...

Reference Data included:

Role	Version	Download Size	On Disk Size
1000_genomes_project	phase_3_ensembl_v113_hg38_no_alt_analysis_set	1.65 GB	1.65 GB
cds	refseq_GRCh38.p14_no_alt_analysis_set	12.6 MB	12.6 MB
clinvar	20241103_hg38_no_alt_analysis_set	176.9 MB	176.9 MB
dbSNP_common	151_ucsc_hg38_no_alt_analysis_set	1.37 GB	1.37 GB
gene_ontology	20231009_hg38_no_alt_analysis_set	6.2 MB	6.2 MB
conservation_scores_phastcons	hg38_no_alt_analysis_set	4.64 GB	5.28 GB
genes	refseq_GRCh38.p14_no_alt_analysis_set	3.2 MB	3.2 MB
hapmap	phase_3_ensembl_v97_hg38_no_alt_analysis_set	482.6 MB	572.5 MB
mma	refseq_GRCh38.p14_no_alt_analysis_set	11.8 MB	11.8 MB

Help Close

- ✓ Whole Exome・Target Amplicon Sequencing用ワークフローではシーケンスキャプチャーキットのターゲット領域データが必要
- ✓ ターゲット領域データはBED形式またはGFF形式のファイルで、多くの場合、各キットのメーカーより取得する(メーカーによってはホームページ上で公開されている)

ターゲット領域データ例: Illumina社のNextera Rapid Capture Exomeの場合

[nexterarapidcapture_exome_targetedregions_v1.2.bed](#)

- [Nextera Rapid Capture Exome v1.2 Product Files](#)

This download contains the Nextera Rapid Capture Exome v1.2 release notes, targeted regions manifest files (.txt and .bed), and the exome probe manifest. The readme document describes the Nextera Rapid Capture Exome v1.2 manifest files. The targeted regions manifest files list the 214,126 targeted exonic regions with start and stop chromosome locations in GRCh37/hg19.

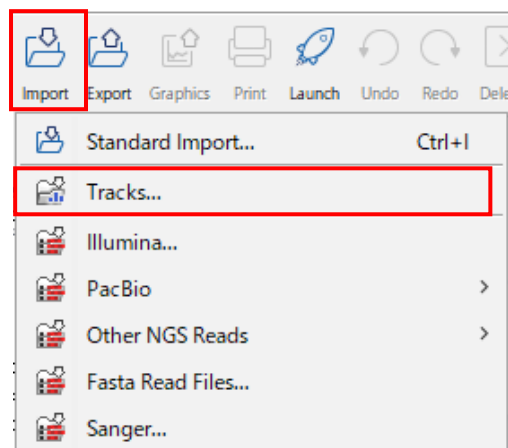
The Nextera Rapid Capture Exome v1.2 Targeted Regions Manifest and the Nextera Rapid Capture Exome Probe Manifest files are in .txt format. This format is required by the HiSeq Analysis Software.

» [Nextera Rapid Capture Exome Targeted Regions Manifest v1.2 \(BED Format\)](#)

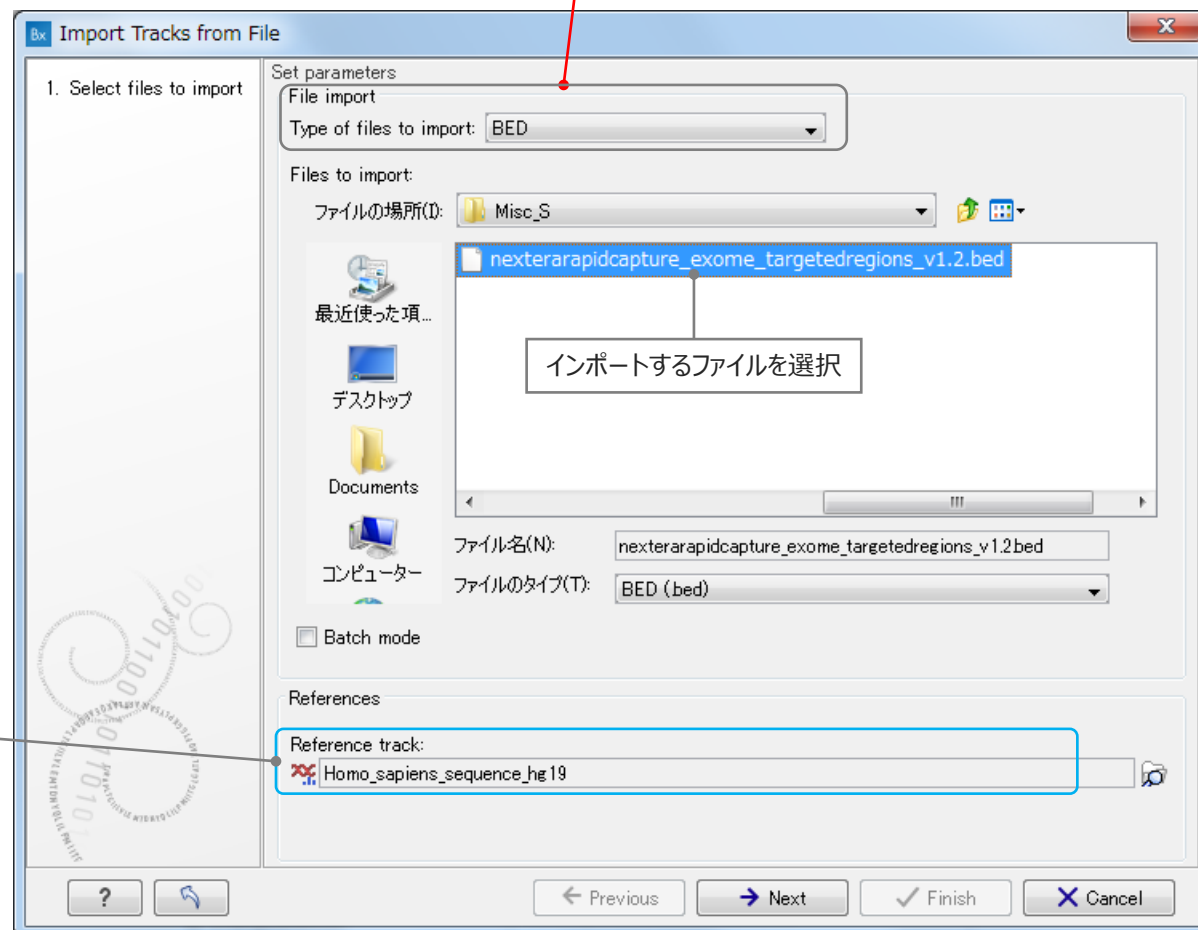
BED (5 MB)

03/12/2014

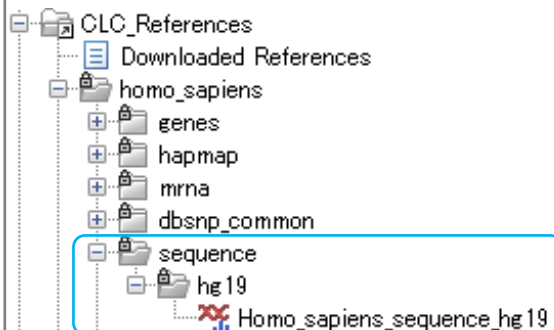
✓ ターゲット領域データ(e.g. .bedファイル)は、Import > Tracks...からインポート



ファイルタイプを選択(.bedならBED、.gtfならGFF2/GTF/GVFを選択)

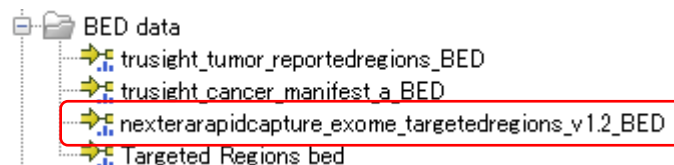


- 対象とするリファレンスゲノム配列を選択
- CLC_Referencesのデータを利用可能

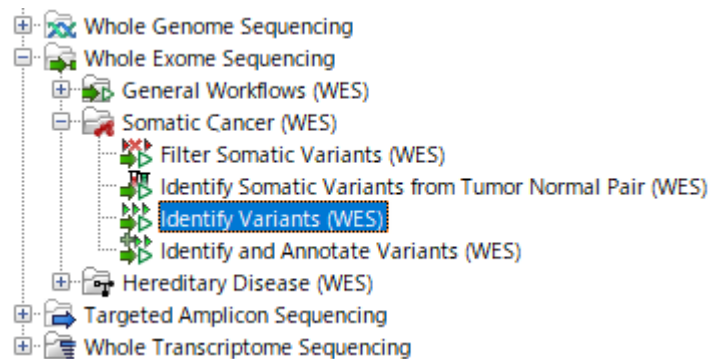


※上記はヒトhg19を例示

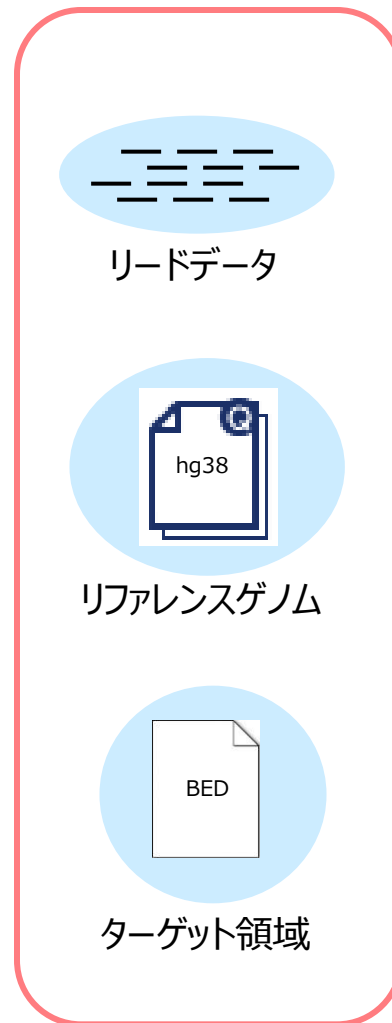
✓ インポートが完了すると、アノテーション形式のトラックデータが作成され、ワークフローに使用できるようになる



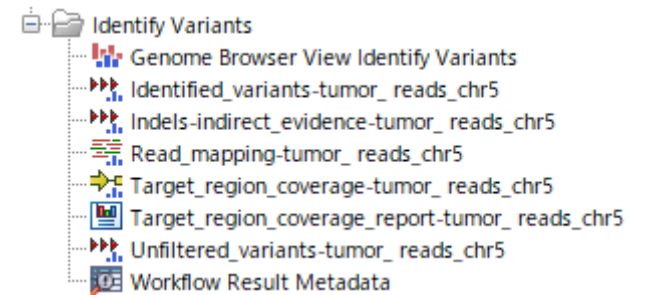
Chromosome	Region	Name
1	10385471..10385472	rs17401966.chr1.10385470.10385472
1	11046855..11046856	rs9430161.chr1.11046854.11046856
1	14096821..14096822	rs1210110.chr1.14096820.14096822
1	14804874..14804875	rs7555566.chr1.14804873.14804875
1	17345376..17345454	SDHB.chr1.17345375.17345454
1	17349103..17349226	SDHB.chr1.17349102.17349226
1	17350468..17350570	SDHB.chr1.17350467.17350570
1	17354244..17354361	SDHB.chr1.17354243.17354361
1	17355095..17355232	SDHB.chr1.17355094.17355232
1	17359555..17359641	SDHB.chr1.17359554.17359641
1	17371256..17371384	SDHB.chr1.17371255.17371384
1	17380443..17380515	SDHB.chr1.17380442.17380515
1	17722363..17722364	rs7538876.chr1.17722362.17722364
1	45794978..45795110	MUTYH.chr1.45794977.45795110
1	45796188..45796230	MUTYH.chr1.45796187.45796230
1	45796854..45797007	MUTYH.chr1.45796853.45797007
1	45797092..45797229	MUTYH.chr1.45797091.45797229
1	45797333..45797522	MUTYH.chr1.45797332.45797522
1	45797695..45797759	MUTYH.chr1.45797694.45797759
1	45797838..45797983	MUTYH.chr1.45797837.45797983
1	45798063..45798161	MUTYH.chr1.45798062.45798161
1	45798246..45798360	MUTYH.chr1.45798245.45798360
1	45798435..45798507	MUTYH.chr1.45798434.45798507
1	45798590..45798632	MUTYH.chr1.45798589.45798632
1	45798769..45798843	MUTYH.chr1.45798768.45798843
1	45798957..45798997	MUTYH.chr1.45798956.45798997
1	45799085..45799276	MUTYH.chr1.45799084.45799276
1	45800063..45800184	MUTYH.chr1.45800062.45800184
1	45805891..45805927	MUTYH.chr1.45805890.45805927



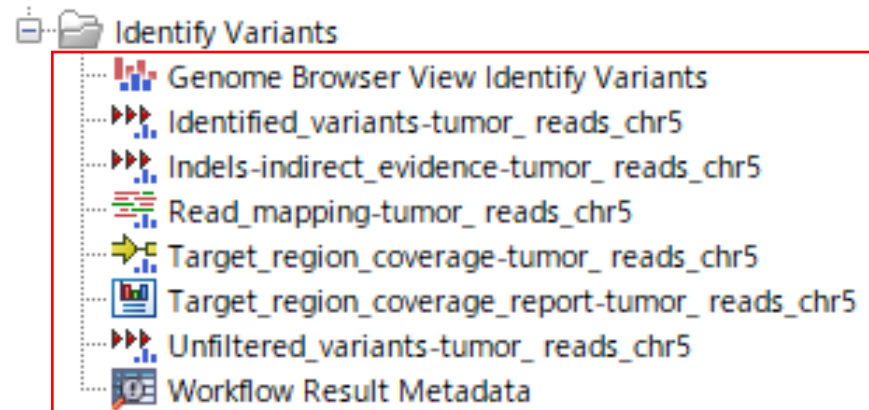
ワークフローを起動



解析に必要なデータを入力

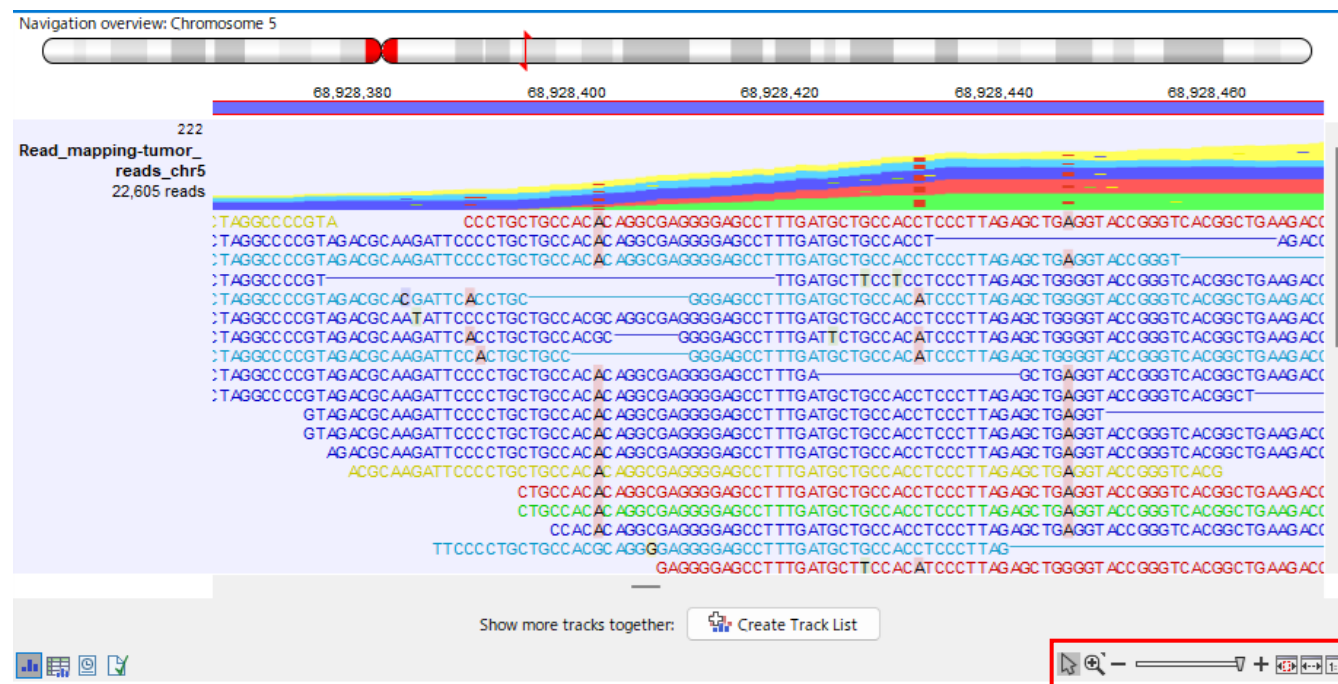
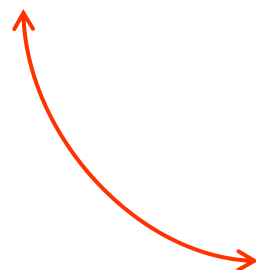
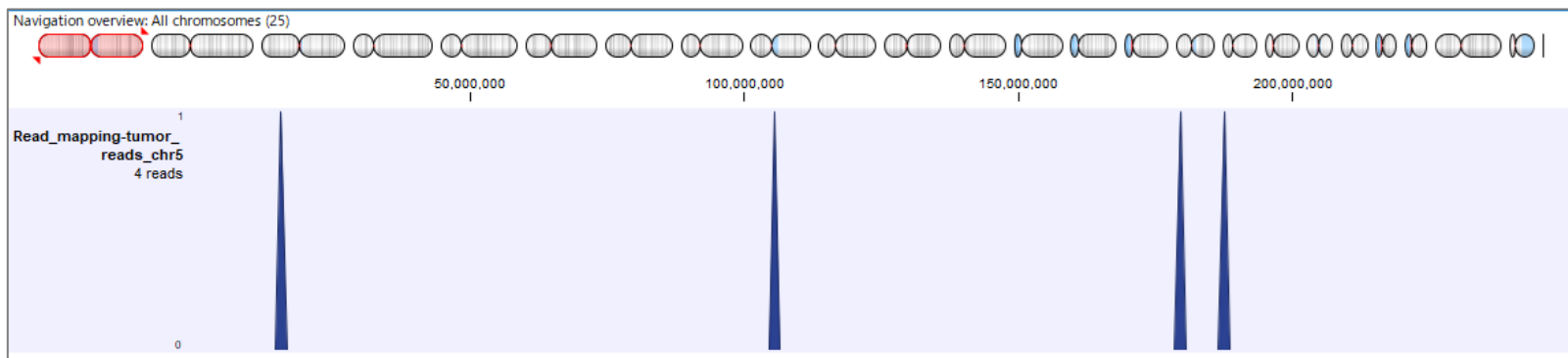


結果の保存



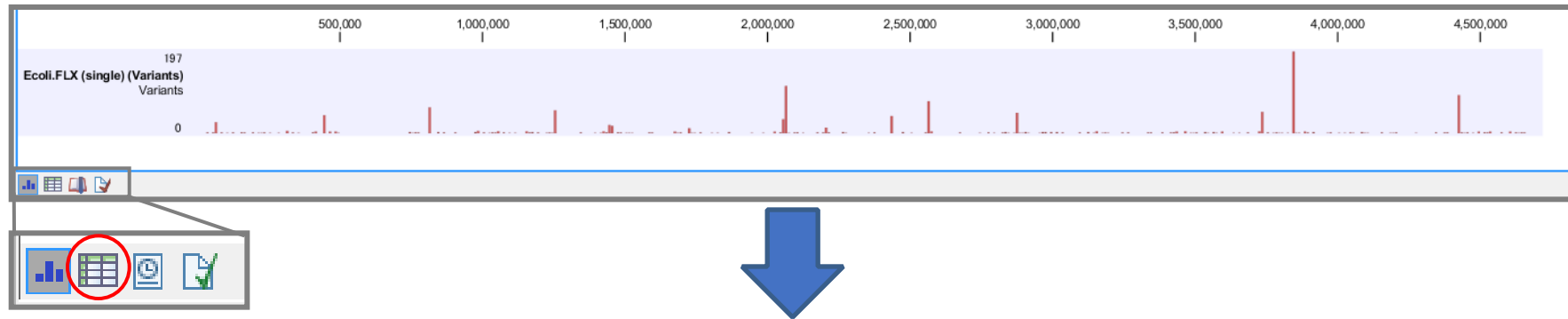
- **Read_mapping**
 - マッピング結果データ
- **Target_region_coverage**
 - ターゲット領域ごとのカバレッジテーブル
- **Target_region_coverage_report**
 - ターゲット領域のカバレッジ計算結果のサマリーレポート
- **Unfiltered_variants**
 - フィルタリング実行前の変異のリスト
- **Identified_variants**
 - フィルタリングをパスした変異のリスト
- **Indels-indirect_evidence**
 - サイズの大きいInsertion/Deletionのリスト
- **Genome_browser_view**
 - リファレンスゲノムとアノテーション、マッピング結果データ、変異リストデータを統合してゲノムブラウザー表示させたデータ

Read_mapping-tumor_reads_chr5



- 画面右下のZoom機能で拡大/縮小表示を調整可能

Identified_variants-tumor_reads_chr5

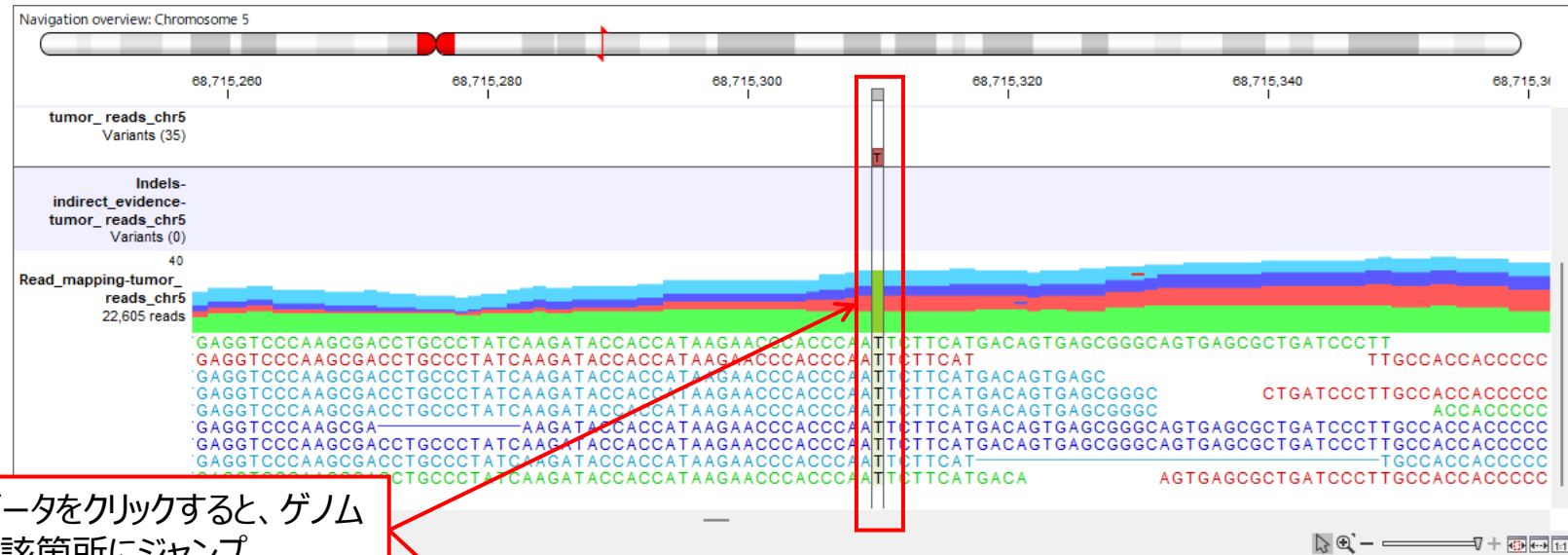


Rows: 35 Table view: Homo sapiens

Chromosome	Region	Type	Reference	Allele	Reference all...	Length	Zygosity	Count	Coverage	Frequency	Probability	Forward read...	Reverse read...
5	68396764^68396765	Insertion	-	G	No		1 Homozygous	13	14	92.86	1.00	4	9
5	68413043	Deletion	T	-	No		1 Homozygous	47	48	97.92	1.00	35	12
5	68419054	SNV	A	T	No		1 Homozygous	15	15	100.00	1.00	12	3
5	68463043	SNV	G	T	No		1 Heterozygous	9	47	19.15	1.00	4	5
5	68463043	SNV	G	G	Yes		1 Heterozygous	38	47	80.85	1.00	24	14
5	68463771	Deletion	A	-	No		1 Heterozygous	2	15	13.33	1.00	1	1
5	68463771	SNV	A	A	Yes		1 Heterozygous	13	15	86.67	1.00	6	7
5	68464234	SNV	A	G	No		1 Heterozygous	12	22	54.55	1.00	3	9
5	68464234	SNV	A	A	Yes		1 Heterozygous	10	22	45.45	1.00	5	5
5	68471176	SNV	A	G	No		1 Heterozygous	2	22	9.09	0.87	1	1
5	68471176	SNV	A	A	Yes		1 Heterozygous	20	22	90.91	1.00	16	4
5	68471247	SNV	G	A	No		1 Heterozygous	19	47	40.43	1.00	11	8
5	68471247	SNV	G	G	Yes		1 Heterozygous	28	47	59.57	1.00	14	14
5	68498319	SNV	T	C	No		1 Heterozygous	2	16	12.50	0.96	1	1
5	68498319	SNV	T	T	Yes		1 Heterozygous	14	16	87.50	1.00	8	6
5	68524006^68524007	Insertion	-	T	No		1 Homozygous	13	13	100.00	1.00	12	1

- ・ デフォルト表示では、リファレンス配列上の変異数をグラフ表示
- ・ テーブルアイコンをクリックすると、変異の詳細をテーブル形式で表示が可能

... Genome Browser View Identify Variants



- 変異リスト内のデータをクリックすると、ゲノムブラウザー上の当該箇所へジャンプ

Table view: Homo sapiens													Filter to Selection...	Filter
Chromosome	Region	Type	Reference	Allele	Reference alt...	Length	Zygosity	Count	Coverage	Frequency	Probability	Forward read...	Reverse read...	
5	68525027	SNV	C	C	Yes	1	Heterozygous	9	11	81.62	1.00	5		
5	68531253	SNV	C	T	No	1	Homozygous	15	15	100.00	1.00	7		
5	68578752	SNV	G	T	No	1	Heterozygous	2	19	10.53	1.00	1		
5	68578752	SNV	G	G	Yes	1	Heterozygous	17	19	89.47	1.00	8		
5	68609885	SNV	G	T	No	1	Homozygous	16	16	100.00	1.00	3		
5	68651637	Deletion	A	-	No	1	Heterozygous	2	26	7.69	0.99	1		
5	68651637	SNV	A	A	Yes	1	Heterozygous	24	26	92.31	1.00	11		
5	68661359	SNV	T	C	No	1	Heterozygous	2	56	3.57	0.98	1		
5	68661359	SNV	T	T	Yes	1	Heterozygous	54	56	96.43	1.00	32		
5	68661380	Deletion	T	-	No	1	Heterozygous	4	66	6.06	1.00	2		
5	68661380	SNV	T	T	Yes	1	Heterozygous	62	66	93.94	1.00	40		
5	68681954	SNV	A	G	No	1	Heterozygous	2	25	8.00	1.00	1		
5	68681954	SNV	A	A	Yes	1	Heterozygous	23	25	92.00	1.00	17		
5	68695940	SNV	T	G	No	1	Heterozygous	3	17	17.65	1.00	2		
5	68695940	SNV	T	T	Yes	1	Heterozygous	14	17	82.35	1.00	6		
5	68715310	SNV	C	T	No	1	Homozygous	32	32	100.00	1.00	20		
5	68728914	SNV	C	T	No	1	Heterozygous	2	74	2.70	1.00	1		
5	68728914	SNV	C	C	Yes	1	Heterozygous	72	74	97.30	1.00	26		

お問い合わせ先：フィルジェン株式会社

TEL: 052-624-4388 (9:00～18 : 00)

FAX: 052-624-4389

E-mail: support@filgen.jp