CLC Genomics Workbench を使った 変異解析

株式会社CLCバイオジャパン シニアフィールドバイオインフォマティクスサイエンティスト 宮本真理 Ph.D.





アジェンダ

- データについて
- 変異解析フロー
- ・ ハンズオン
 - インポート
 - QC、トリミング
 - マッピング
 - ローカルリアライメント
 - 変異検出
 - 変異の比較、アノテーション







LocationとFolder



- Genomics Workbenchではデータを以下
 のような階層構造で保存可能です。フォ
 ルダの一番上位の階層を「Location」と呼び、その下の階層を「Folder」と呼びます。
- データの保存場所はロケーション毎に設 定可能です。たとえばあるデータはCドラ イブに保存し、あるデータはDドライブに 保存するといった事が可能です。
- ロケーション、フォルダの作成は以下の アイコンから作成できます。





ダウンロードしたデータを、解凍せずにImport>Standard Import からインポートください。右図のようなファイルがインポートされま す。





Standard Importは、サンガーシーケンサー、次世代シーケンサー以外のファイルのインポートに利用します。





• Importからインポートしたいリードのシーケンサータイプを選択。

File Edit \	/iew [Down	nload Too	olbox W	orksp	ace H	elp					
\rightarrow [+	\leftarrow	P	ЭĠ	Ŵ		\bigcirc	G	8			\sim	
Show New	Save	Impo	ort Export	Graphics	Print	Undo	Redo	Cut	Сору	Paste	Delete	
		凶	Standard	Import		Ctrl+I						
		62	Tracks									
	ſ	6	Roche 454	ł								
		່ 💕	Illumina									
		6	SOLID									
		6	Fasta Rea	d Files								
		6	Sanger									
		6	Ion Torre	nt								
		Solution	SAM/BAM	Mapping	Files							





Illuminaデータのインポート •

Gx Illumina		×	
1. Import files and options	Select files of types Illumina fil 参照: 🚺 Illumina	es (txt/ fastq/ fq/ gseg) _Data	General options Paired reads:ペアかどうか。
	また使った項 また使った項 またします またします またします またします またします	sequence.txt	 Discard reads names:リードについている名前を捨てる かどうか。デフォルトでは捨てるとなっていますが、マッ ピング後、SAMにてExportした際など、リード名で確認し たい場合があるため、最初は保存しておきましょう。 Discard quality scores:Quality Scoreが必要ない場合は このオプションにチェック。通常、インポート後にクオリ ティスコアが必要な事が多いです。
	ファイル名: スットワーク ファイルタイプ:	"s_1_1_sequence.txt" "s_1_2_sequence.txt" Illumina files (txt/_fastq/_fq/_qseq)	Paired read orientation:ペアの距離と向きを指定。
Marine Ma Ana ana ana ana ana ana ana ana ana ana	General options Paired reads Discard read names Discard quality scores Illumina options Remove failed reads Misco de-multipleving	Paired read information Paired-end (forward-reverse) Minimum distance 180 Maximum distance 250 Quality scores NOBL/Sanger or Illumina Pipeline 1.8 and later	 Illumina options Remove failed reads:シーケンサーでfailとマークされた リードを除去するかどうか。 Miseq de-multiplexing:MultiplexingされたデータをDe- multiplexingするかどうか。
70	Trim reads	← Previous → Next ✓ Finish X Cancel	■ Quality Score:使用するQuality Scoreのバージョンの選 択。



リードデータインポート

1. Import files and options Result handling 2. Result handling 1. Import files and options 3. Save location for new elements 2. Result handling 3. Save location for new elements 3. Save location for new elements 4. Import files and options 2. Result handling 5. Save location for new elements 3. Save location for new elements 6. Deter Update All 4. Deter Update All 6. Deter Update All 5. Save location for new elements 7. Import files and options 5. Save location for new elements 8. Save location for new elements 6. Deter Update All 6. Deter Update All 6. Deter Update All 7. Import files and options 7. Import files and options 8. Save location for new elements 6. Deter Update All 7. Import files and options 7. Import files and options 8. Save location for new elements 6. Deter Update All 8. Save location for new elements 6. Deter Update All 8. Save location for new elements 6. Deter Update All 8. Save location for new elements 6. Deter Update All 9. Deter Update All 6. Deter Update All 9. Deter Update All 6. Deteeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee	Gx Illumina		Gx Illumina	×
Image: Solution Separate Tolders Log handling Image: Open log I	Illumina I. Import files and options Result handling	Result handling Open Save Into separate folders Log handling Open log	Illumina I. Import files and options Result handling Save location for new elements	Save location for new elements
Result handling	Result handling	← Previous → Next Finish X Cancel]	

- データを開くか、保存の選択
- Into separate folders では、別々のフォルダへ保存するかどう かを選択できます。バッチ処理を行う際に便利です。

保存先の指定





Ion Torrentのインポート(.fastq または .sff)

9. Ion Torrent	×
1. Import files and options Select files of types 1on Torrent Flowgram (.sff) 1on Torrent (.fastq/ .fq) 参照: STO-409_FASTQ_BAM_REPORTS ・ 多 P :::: == Recent Items Desktop Documents ファイル名: STO-409.fastq ファイルタイプ: Ion Torrent (.fastq/ .fq)	 General options Paired reads:ペアかどうか。 Discard reads names:リードについている名前を捨てる かどうか。デフォルトでは捨てるとなっていますが、マッ ピング後、SAMにてExportした際など、リード名で確認し たい場合があるため、最初は保存しておきましょう。 Discard quality scores:Quality Scoreが必要ない場合は このオプションにチェック。通常、インポート後にクオリ ティスコアが必要な事が多いです。
General options Paired reads Discard read names Discard quality scores Discard quality scores Ion Torrent options Use clipping information	 Paired read orientation:ペアの距離と向きを指定。 Ion Torrent options Use clipping information:.sffファイルに含まれるクリッピングの情報を利用するとき(.sffファイルを選択した場合)。



この次のウィザードの画面はイルミナと共通です。

リードデータインポート

Gx Illumina	x	Gx Illumina	×
Illumina I. Import files and options Result handling	Result handling Open Open Save Into separate folders Log handling Open log	Illumina I. Import files and options Result handling Save location for new elements	Save location for new elements
Result handling]	

- データを開くか、保存の選択
- Into separate folders では、別々のフォルダへ保存するかどう かを選択できます。バッチ処理を行う際に便利です。

保存先の指定





Unmapped BAM

(※TorrentSuiteから作成される圧縮されたリードファイル。BAMファイル はマッピングした状態のものと、マッピングしていない状態のBAMファイ ルがある)

File Edit View	Download Toolbox Workspace Help		
Show New Save	Import Export Graphics Print Undo Redo	Cut Copy F	Paste Delete
	Standard Import Ctrl+I		
	Roche 454 Illumina PacBio		にImport>Standard Import からインポートするとリードファイ ルがインポートされます。
	SOLID Fasta Read Files		※Mapped BAMは、Import > SAM/BAM Mapping Files からインポートします。Mapped BAMをStandard Import でインポート
	Sanger		すると、リードノアイルとしてインボートされます。



- ゲノムはダウンロードアイコン より、生物種を指定してアノ テーションと共にインポートす ることが可能です。
- ゲノム配列とともに、アノテーションファイルをダウンロードすることも可能です。
- すでにGenomics Workbenchへ 取り込んでいるゲノム配列に ついて、アノテーションを付加 することも可能です。





Gx Download Refere	ence Genome Data		x	Gx Download Refer	rence Genome Data	
Download Refere	Genomes Genomes Animal (mammals) Animal (others) Plants Other	Homo sapiens (hg19) Caenorhabditis elegans Arabidopsis thaliana (Col-0 TA Escherichia coli (DH10B)	▼ IR10) ▼ ▼	 Download Refer Select organism Select sequence Do グロ Us つこ るさ 	rence Genome Data Select sequence Peterence genome Download genome sequence Use existing genome sequence: ウンロードする場合。 e exsting genome sequence ンロードしたゲノムにアノテー 場合。以下のようにトラックの	新規にゲノムを track: すでにダーションを追加す Dフォーマットに
The second secon				\ <u>\</u>	うているクラムを選択。	
? 5	← Pre	vious Arext	✓ Finish X Cancel	Reference genome	: sequence	
∎ ۲ï	コップダウン	リストから生物	物種を選択。	Homo sapiens	s (hg 19) sequence (Genome)	ର୍ଭ



Select organiam	Select annot	ations			
Select organism	Download	Name	Version	Provider	Size (in Mb)
Select sequence	V	Sequence	74	Ensembl	83
Select annotations	V	Gene annotation	74	Ensembl	
	V	Dbsnp (common) variants	137	UCSC	5
	V	Dbsnp variants	137	UCSC	14
	V	COSMIC	v67_241013	SANGER	
	V	Clinical variants in dbSNP		NCBI	
	V	HapMap Variants		Ensembl	4
	V	1000genomes	phase1	Ensembl	19

 希望するアノテーションにチェックを入れる。ゲノム配列をダウンロード するときは、Sequences にもチェックを入れる。
 選択した生物種により、表示されるアノテーションの種類は異なります。



アノテーションインポート

- Download Genome 以外にも、アノテーションファイルをインポート可能です。
- アノテーションとして取り込めるファイルは以下のフォーマットです。
- アノテーションファイルをインポートする際には、<u>対象となるゲノム配列が</u> すでにインポートされ、Trackのフォーマットになっていることが前提です。
 - VCF
 - GFF/GTF/GVF
 - BED
 - Wiggle
 - Complete Genomics Var file
 - UCSC Variation table damp
 - COSMIC variation database

※変異のデータについても、アノテーションとして自分の変異へアノテーションとして情報の追加を比較ができるため、アノテーションのインポート可能フォーマットに含めています。



アノテーションインポート

File Edit View Download Toolbox Workspace Help	Gx Tracks
Show New Save Import Expert Graphics Print Lindo Redo Out Conv. Paste	1. Select files to import File import Type of files to import: GFF/GTF/GVF
Show New Save Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export Graphics Frint Onto Redu Cut Copy Paste Import Export State Import Export State	Files to import: 参照: Annotations 参照: Mus musculus sequence.gff 最近使った項 Mus musculus sequence.gff デスクトップ アイドキュメント
	Image: Second Secon

- Type of files to importを選択
- インポートするファイルを選択

l C bio

17

■ Reference Track を選択

Homo sapiens (hg 19) sequence
 Homo sapiens (hg 19)_CDS
 Homo sapiens (hg 19)_Exon
 Homo sapiens (hg 19)_Gene
 Homo sapiens (hg 19)_mRNA
 Homo sapiens (hg 19)_Transcript
 Homo sapiens (hg 19)_COSMIC

TrackにはGenomics Workbenchが認識しているタイプによりアイコン がそれぞれ異なります。解析によっては、Trackのタイプを認識して、 解析に利用できるかどうか区別するため、データがどのようなタイプ で認識されているか、アイコンで確認するようにしてください。



トラックフォーマットとスタンドアロンフォーマット



トラックとスタンドアロンフォーマット

- Genomics Workbenchはビューアにスタンドアロンフォーマットとトラック フォーマットがあります。
- スタンドアロンフォーマットでは、1つのデータに配列情報、アノテーション がセットになっています。





- トラックフォーマットでは、リードやゲノム配列、アノテーションがばらばらのファイルになっており、好きに組み合わせて表示が可能です。
 - 🚟 reads (Reads)







• 複数のトラックを組み合わせることで好きなビューを作成できます。



s_1_1_sequence (paired) (Reads)-2



トラックとスタンドアロンフォーマット

スタンドアロンフォーマット

- reads mapping 🖕 リードマッピング

トラックフォーマット

青いヒストグラムが目印

Homo sapiens (hg 19) sequence 「ゲノムTrack
 Homo sapiens (hg 19)_CDS
 アノテーションTrack
 Homo sapiens (hg 19)_Exon
 Homo sapiens (hg 19)_Gene
 Homo sapiens (hg 19)_mRNA
 Homo sapiens (hg 19)_Transcript
 Homo sapiens (hg 19) COSMIC
 変異Track

解析によって必要とするフォーマットが異なります。スタンドアロン⇔トラックの変換は自由に行えます。



トラックとスタンドアロンフォーマット

注意点:Download Genome 以外でゲノム配列を取得した場合



Search for Sequences at NCBI 出検索してダウンロードした場合やImport > Fasta などでインポートを行うと、以下のようなアイコンのスタンドアロン フォーマットでインポートされます。解析によりトラックフォーマットのゲノムが 必要な場合は、次のページからの方法で変換をあらかじめ行ってください。

---- XX AE005174 🔹 1本の染色体





 トラックフォーマットからスタンドアロンフォーマット、またスタンドアロン フォーマットからトラックフォーマットへはGenomics Workbench の Toolbox > Track tools の中のツールを使って変換可能です。





スタンドアロンフォーマットへ変換する場合、スタンドアロン内に含めるアノ テーショントラックを含めて変換するようにしてください。

Select one sequence or	Select one sequence or reads track and optionally ann	otation tracks	📰 Homo sapiens 🗙	
eads track and pptionally annotation racks	Navigation Area CLC_Data CLC_D	Selected Elements (3) Homo sapiens (hg 19) sequence (Homo sapiens (hg 19)_Exon Momo_sapiens GRCh37.70_Gene		Sequence List Sequence List Sequence layout Annotation layout Annotation layout Annotation types Comparison of the second Market All Deselect All Deselect All Deselect All Deselect All Restriction sites Motifs Residue coloring Nucleotide info Find Text format
?	KC_chr2 (Genome) M Center search term> A Center search term> A Detch Center search term> A Center search term> Center search term> A Center search term> Centerm> Center search term> Centerm> Centerm> Center	ext Finish X Cancel	 スタンドアロンフォーマットでは、Setting Annotation Type からどういったアノテー 付属しているか確認できます。 	Panel の ションが



クオリティチェック



クオリティチェック流れ

- Quality Report作成: Create Sequencing QC Report
 - インポートしたリードのクオリティがどのぐらいか、その後のトリミングや、PCR Duplicate の状況などを確認するためにレポートを作成。
- PCR Duplicate の除去: Remove Duplicate Reads
 - フラグメント作成の過程でPCRが異常にかかってしまったものを補正。
- トリミング: Trim Sequences
 - アダプターの除去、クオリティスコアによる除去、長さを指定した除去 などを選択・組み合わせてトリミング。

上記処理の後に再度Quality Reportを作成すると処理前と処理後での リードのクオリティを比較でき、便利です。





- Navigation Areaから使用するリードデータを選択。
 Toolboxから NGS Core Tools > Create Sequencing QC Report を選択、 ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。



Gx Create Sequencing QC	Report	×
1. Select sequencing reads 2. Settings	Settings	
	Additional reporting	
	 ✓ Quality analysis ✓ Over-representation analysis 	
Carling Store Market	← Previous → Next ✓ Finish X Cano	cel

 Quality analysis: クオリティスコアに関する解析。
 Over-representations analysis: 過度に現れているような 塩基配列などの解析。

Gx Create Sequencing QC	Report 💦 🔀
 Select sequencing reads Settings Result handling 	Result handling
	Output options Image: Create graphical report Image: Create supplementary report Image: Create duplicated sequence list
	Result handling Open © Save
	Log handling
The second secon	
?	← Previous → Next ✓ Finish X Cancel

- Create graphical report: グラフィカルなレポート作成。
- Create supplementary report: 数値のレポート作成。
- Create duplicated sequence list: 重複のあった配列のリ スト作成。





3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.

AGTCがどういった割合で現れているか、チェック。

3.5 Quality distribution



Base-quality distribution along the base positions.

リードのポジションごとにクオリティスコアがどうなっているか。



Manager and the second second
reads - grannical UU report
📲 reads – supplementary QC report
🚟 📻 reads – duplicated sequences QC repor

3.5 Quality distribution

Summarizes the base-quality distribution along the base positions. base position: coverage: total number of quality values observed at that base position median & percentiles: median & percentiles of quality scores observed at that base position

base position	coverage	5%ile	25%ile	Median	75%ile	95%ile	
1	339,368	11	23	26	32	33	
2	339,368	11	24	26	31	34	
3	339,368	11	24	26	31	34	
4	339,368	12	25	26	31	33	
5	339,368	11	24	26	31	33	
6	339,368	12	24	26	31	33	
7	339,368	11	24	26	31	33	
8	339,368	11	24	26	31	33	
9	339,368	11	24	26	31	33	
10	339,368	12	24	26	31	33	
11	339,368	12	24	26	31	33	
12	339,368	12	24	26	31	33	
13	339,368	12	24	26	30	33	
14	339,368	12	24	26	30	33	
15	339,368	12	24	26	30	33	
16	339,368	12	24	26	30	33	
17	339,368	12	24	26	30	33	
18	339,368	12	24	26	30	33	
19	339,368	12	24	26	30	33	
20	339,368	12	24	26	30	33	
21	339,368	12	24	26	29	33	
22	339,368	12	23	26	29	33	
23	339,368	12	23	26	28	33	
24	339,368	12	23	26	28	33	

No name GTGTGTCCTGTGATCCGCCAGAGCACGTGTGTCCTGTGAGATCCGCCAGA Noname GCCAGAGCACGTGTGTCCTGTGAGATCCGCCAGAGCACGTGTGTCCTGTG No name GTGTGTCCTGTGAGATCCGCCAGAGCACGTGTGTCCTGTGAGATCCGCCA No name GGATCTCACAGGACACACGTGCTCTGGCGGATCACAGGACACACGTGCTC 20 No name GTGTGTCCTGTGAGATCCGCCAGAGCACGTGTGTCCTGTGATCCGCCAGA No name GGACACACGTGCTCTGGCGGATCTCACAGGACACGTGCTCTGGCGGAT Noname AGGACACACGTGCTCTGGCGGATCACAGGACACGTGCTCTGGCGGATC 20 No name CAGGACACACGTGCTCTGGCGGATCTCACAGGACACACGTGCTCTGGCGG No name ATCACAGGACACACGTGCTCTGGCGGATCTCACAGGACACACGTGCTCTG No name GTGCTCTGGCGGATCTCACAGGACACACGTGCTCTGGCGGATCACAGGAC 20 No name AGAGCACGTGTGTCCTGTGATCCGCCAGAGCACGTGTGTCCTGTGAGATC No name ATCCGCCAGAGCACGTGTGTCCTGTGAGATCCGCCAGAGCACGTGTGTCC 20 No name GTGCTCTGGCGGATCACAGGACACACGTGCTCTGGCGGATCTCACAGGAC No name AGAGCACGTGTGTCCTGTGAGATCCGCCAGAGCACGTGTGTCCTGTGATC 20 No name GATCCGCCAGAGCACGTGTGTCCTGTGAGATCCGCCAGAGCACGTGTGTC











クオリティトリミング原理

- Trimming ではQuality Score を使い、累積のQuality Score がある一定の 値より大きいものが続いた場合に、その箇所を取り除く、という処理を行 います。
- 具体的には以下:
 - 1. Phred Score をp値へ変換
 - 2. Trimming 中に設定するパラメータ(Limit)とp値の差を計算
 - 3. 差の累積和を計算。このとき、0以下の値は0とする
 - Trimming後のリード開始点は累積和がはじめて0以上になった点。Trimming後のリード終了点は累積和が最大の点





リード配列	G	С	С	С	Α	Т	G	Т	Т	С	G	Α	Т	G	С
Phred score	4	8	15	30	32	23	10	31	31	20	15	11	10	10	9
p値	0.40	0.16	0.03	0.00	0.00	0.01	0.10	0.00	0.00	0.01	0.03	0.08	0.10	0.10	0.13
Limit - p值 (D)	-0.35	-0.11	0.02	0.05	0.05	0.04	-0.05	0.05	0.05	0.04	0.02	-0.03	-0.05	-0.05	-0.08
<u>(</u> D)の累積和	0.00	0.00	0.02	0.07	0.12	0.16	0.11	0.16	0.21	0.25	0.27	0.24	0.19	0.14	0.06





Phred score の棒グラフ



グラフより、ある程度クオリティが高くなった場所からリードを使い、クオリティが 連続して悪くなっている箇所からリードをトリムしていることがわかる。 ※途中、1塩基のみクオリティが低いような場合は、必ずしもトリムされない。 これはできるだけリードを長く保とうとするため。




トリミング



Navigation Areaから使用するリードデータを選択。
 Toolboxから NGS Core Tools > Trim Sequences を選択、ダブルクリック。

■ ウィザードが起動し、選択したデータが選ばれていることを確認。





トリ	ミング
----	-----

G Trim Sequences

Gx Trim Sequences	X
1. Select sequencing data 2. Quality trimming	Set parameters
	Quality trimming Trim using quality scores Limit: 0.05 Trim ambiguous nucleotides Maximum number of ambiguities: 2
1 1000 mm	← Previous → Next ✓ Finish X Cancel

1. Select sequencing data	Adapter trimm	ing			
2. Quality trimming					
3. Adapter trimming					
	Trim adar	ter list			
	Searc	h on both strands			
		in on bour suands			
	Preview				
	Number of rea	ds 1,000 Nun	ber of nucleotides	76,011 Ave	length 76
	Name	Matches found	Reads discarded	Nucleotides re	Avg. length
Colores and					
O P					
Jan					
017 000 017					
10011000000000000000000000000000000000					
10000000000000000000000000000000000000					

X

- Trim using quality scores :トリミングに使用するLimitパラメータを決定
- Trim ambiguous nucleotides:N表示される塩 基について、最大何塩基まで保持させるか。





トリ	ミン	ッグ

Gx Trim Sequences	X
Select sequencing data Quality trimming Adapter trimming Sequence filtering	Set parameters
00	Trim bases Remove 5' terminal nucleotides Remove 3' terminal nucleotides Image: start sta
Contraction of the second seco	← Previous → Next ✓ Finish X Cancel

- Trim bases:リード配列の5'末、3'末から指定数の塩
 基を除去
- Filter on length:リード配列の5'末、3'末から指定数 の塩基を除去

Gx Trim Sequences	×
Select sequencing data Quality trimming Adapter trimming Sequence filtering S. Result handling	Result handling
	Output options Save discarded sequences Save broken pairs Oreate report
COLUMN AND A COLUM	Log handling
?	← Previous → Next ✓ Finish X Cancel

- Save discarded sequences: トリミングにより除去された配列の保存。
- Save broken pairs: ペアのリードでトリミングによりペア でなくなったリードを保存。
- Create report: レポートの作成。









2 Read length before / after trimming



 トリミング結果のデータはファイル名の 後に trimmed という名前が付いてい ます。ファイル内容はインポート後の データ同様に、配列と、クオリティスコ アを含んだファイルとなっています。

トリミング後は、トリムされたリードと、
 レポートを作成した場合は、そのレポートが作成されます。



QCレポート 再作成による比較

 トリミングされたリードを使って、QCレポートを再度作ることで、トリミング 前後の比較が行えます。

Before

3.5 Quality distribution



Base-quality distribution along the base positions.



3.5 Quality distribution





PCR Duplicate 除去 ツールの使い分け



2つのPCR Duplicate 除去

- Genomics Workbench には、2つのPCR Duplicate 除去の方法があります。
- Duplicate Read Removal:マッピング前にPCR Duplicate を除去。De Novo アセンブリなど、参照配列がわからない場合に利用します。
- Duplicate Mapped Read Removal:マッピング後のデータに対してPCR Duplicate 除去を行います。リードがセンス鎖、アンチセンス鎖のどちらへ 張り付いたかという事も考慮できますので、参照配列が分かる場合は、こ ちらをお使いください。









Mapping

特徴

- Suffix Array を使い参照配列をインデックス化し、高速なマッピングを可能にしています。
- ローカルアライメント・グローバルアライメントによるスコア計算が可能。
- 異なるシーケンステクノロジー、ペアエンド、シングルエンド、をあわせて マッピング可能。
- カラースペースによる配列のエラー補正も可能。







インデックスファイル作成







フィルタリング





- ?インデックス

Genome



Genomeの端から端まで順番に調べていては、膨大な時間がかかる





• Genomeにインデックスと言う辞書の索引のようなものを作成し、検索効率を上げる。

Genome





• Suffix Array







i	1	2	3	4	5	6	7
A[i]	7	1	5	6	2	4	3
1	\$	А	С	G	G	Т	Т
2		G	G	\$	Т	С	Т
3		Т	\$		Т	G	С
4		Т			С	\$	G
5		С			G		\$
6		G			\$		
7		\$					

もとの配列S と、Suffix Array, Aを使って高速に検索できる。







スコアリング 最適なマップ場所をLocal Alignmentで探索

Match = 1, Mismatch cost = 2

リード配列(20bp)が全て一致した場合







CGTATCAATCGATTACGCTATGAATG IIIIIIIIIIIIIII 19 TTCAATCGATTACGCTATGA

CGTATCAATCGATTACGCTATGAATG IIIIIIIIIIIIIII TTCAATCAATTACGCTATGA

CGTATCAATCGATTACGCTATGAATG IIIIIIIIIIIIII TTCAATCAATTGCGCTATGC





フィルタリング



リード配列と、参照配列がどの程度一致しているものを残すかを 決める。



Length Fraction と Similarity

- Length FractionとSimilarity パラメータを使って、どの程度アライメントされたリードを、マッピングされたものとして保持するか、決定します。
- Length Fraction とSimilarity は2つのパラメータの組み合わせで使用されます。
- Length fraction: フィルターをかける際に、考慮する長さ
- Similarity: Length Fraction で指定した長さのうち、どの程度類似しているものを 残すか。





Length Fraction と Similarity

• デフォルトの設定、Length Fraction = 0.5, Similarity = 0.8 の場合



50塩基中、40塩基一致している場合にマップしているとして残す。



Length Fraction と Similarity

・ どうして2つのパラメータが必要か?



- リードの一部は似ているけれども、大きな挿入や、欠失によりリードの一部が 参照配列と一致しない可能性がある場合
- トリミングが完全にできなかったクオリティの低い配列が末端部にある場合

(Length Fraction を小さくすることで、リードの一部に限定してアライメントの類 似度を設定できる)

Reference



- 参照配列とほぼ一致するが、所々、1塩基の変異があると想定 される場合



Mapping Parameters

Gx Map Reads to Reference	ie X
1. Select sequencing reads	Mapping options
2. References	
3. Mapping options	Read alignment Mismatch cost 2 Insertion cost 3 Deletion cost 3 Length fraction 0.5 Similarity fraction 0.8 Global alignment フィルター Color space alignment Color error cost Auto-detect paired distances Non-specific match handling Map randomly
T Provinces	
?	← Previous → Next ✓ Finish X Cancel



Mapping







Navigation Areaから使用するデータを選択。
 Toolboxから NGS Core Tools > Map Reads to Reference を選択、ダブルクリック。
 ウィザードが起動し、選択したデータが選ばれていることを確認。





		Select single genome track or reference seque	ences
	Y	Navigation Area	Selected elements (1)
Map Reads to Reference Select sequencing reads References	References References References References References No masking Exclude annotated Include annotated Include annotated Include annotated Include annotated Include annotated Include annotated Include annotated Include annotated Include annotated Include annotated Include Inc	CLC_Data Training_DataSet RAAseq ChIPseq SmallRNA ChIPseq PPAR Roche454 UluminaExome Fireads without duplicates Fireads without dupli	Cancel
	Trevious 7 Next Finish Cancel		

Reference:使用する参照配列を選択。
 Reference masking

 Exclude annotated:あるアノテーションを除外したい場合。
 Include annotated only:あるアノテーションのみ含みたい場合。

■ Referenceに使用するデータを選択。



Mapping

G Map Reads to Referen	-e X	
Map Reads to Reference 1. Select sequencing reads 2. References 3. Mapping options	Read alignment Mismatch cost 3 Insertion cost 3 Deletion cost 3 Length fraction 1.0 Global alignment Ø Color space alignment Oclor error cost 3 Auto-detect paired distances Non-specific match handling Ø Map randomly Ignore	 Mismatch cost:アライメントにマッチしないものがあった場合のコスト Insertion cost:アライメントに挿入がある場合のコスト Deletion cost:アライメントに欠失がある場合のコスト Length fraction:リードの長さのどの程度がマッピングされているべきか。 Similarity:どの程度類似しているべきか。 Global alignment: Global alignment: Global alignment: Color space alignment:カラースペースのデータかどうか、その場合にカラーによるエラー補正を行うかどうか。 Auto-detect paired distances: 自動でペアの距離を決めるかどうか。
1 JALON 107 97 JULIN VOLUMENT	← Previous → Next ✓ Finish X Cancel	 Non-specific match handling:同一ス コアでマップされる箇所がある場合の 対処。



Mapping

Gx Map Reads to Reference	e	×
1. Select sequencing reads	Result handling	
2. References		
3. Mapping options		
4. Result handling		
	Output options	
	Oreate reads track	
	Oreate stand-alone read mappings	
	Create report	
	Collect un-mapped reads	
	Besult handling	
	© Open	
	Save	
0		
and a comment	Log handling	
(EP	V Open log	
and and a second second		
And I Carling		
HEALD TREAM		
THE CONTRACTOR OF THE PARTY OF		
All I		
?	← Previous → Next ✓ Finish × Cano	cel

 Create reads track: 結果をトラックとして作成する場合。
 Create stand-alone read mappings: 結果をstand-aloneフォーマット(参照)

配列、リードマッピング、アノテーショ ンが一つになったファイル)で作成す るか。

- Create report:マッピング結果のレ ポート作成。
- Collect un-mapped reads:マップされ なかったリードをリストとして作成す るかどうか(リスト化することにより、 De Novoなど、別の解析へ利用可 能)





マッピング:結果

結果(トラック)

reads trimmed (Reads) ≓F reads trimmed un-mapped reads [no read group] (single) ∭ reads trimmed mapping summary report











• 基本の Report は「Summary Report」という名前で保存されています。

1 Summary mapping report

1.1 Summary statistics

	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	4,686,137.00	4,686,137	-
Mapped reads	5,078,221	97.95%	34.00	172,651,355	97.92%
Not mapped reads	106,167	2.05%	34.62	3,675,884	2.08%
Reads in pairs	4,987,262	96.20%	215.38	170,198,371	96.52%
Broken paired reads	90,959	1.75%	26.97	2,452,984	1.39%
Total reads	5,184,388	100.00%	34.01	176,327,239	100.00%



1.4 Distribution of read length



1.5 Distribution of matched read length



1.6 Distribution of non-matched read length





• 参照配列の追加

- リードマッピングの結果に参照配列を追加しましょう。







Local Realignment ツール





 マッピングのプロセスでは、各リードがもっとも高いアライメントスコア(参照配列との一致度 を示すスコア)を示す場所にマッピングをしています。しかしながら、時には近傍のリードの マッピングの状況から、最も高いアライメントスコアではなくとも、もっともらしいマッピング結 果が考えられる場合があります。

CTTTAGTTTCTTTTGCCGCTTTCTTTCTTTCTTTTTTTTAAGTCTCCCTCTGT



- たとえば上記例では、GCCGは左横にずれることで、他のリードのマッピングとも一致し もっともらしいマッピングになると考えられます。マッピングの段階では、各々のリードのアラ イメントスコアのみを考えているため、このような状況が発生します。
- さらにこの状況で変異やInsertion、Deletionの検出を行うと、正しく検出できないものも発生します。特にInsertionやDeletionが影響をうけると考えられています。



Local Realignment では、このような状況を修正するため、マッピングを部分的にやり直します。この際、通常のマッピングの段階とは異なり、他のリードのマッピング状況を考慮するため、先ほどのマッピングは以下のように変化します。



先ほどのマッピングよりも、こちらの方がもっともらしい結果であることが直感的に分かります。



Local Realignment 原理





Homer, N. & Nelson, S. F. Improved variant discovery through local re-alignment of shortread next-generation sequencing data using SRMA. *Genome biology* **11**, R99 (2010).

Local Realignment 原理

• グラフにして書き直し、それぞれのパスを通るリードのカバレッジを記入すると以下のように なる。このグラフを解く事で、Local Realignment は実行されている。

(e)



Homer, N. & Nelson, S. F. Improved variant discovery through local re-alignment of short-read next-generation sequencing data using SRMA. *Genome biology* **11**, R99 (2010).



Local Realignment



2種類のLocal Realignmentsがありま す。さらにGuided にはNo forceと Forceの2種類があります。

- Non guided
- Guided _
 - No force
 - Force


Local Realignment

- Guided Local Realignment
 - ガイドとなるような変異(InsertionやDeletion)の情報をあらかじめ与えておく
 ことで、その領域のInsertion、Deletionを考慮してリアライメントを行う。
 - ガイドとなる変異情報がない場合、Local Realignment では、少なくとも1本の リードがInsertionやDeletionを支持している必要がある。このような場合、ガイ ドとなる変異情報を与えることで、InsertionやDeletionを効率的に検出できる ようになる。
- Guided Local Realignment が有効な例



Local Realignment

- Guided Local Realignment の注意点
 - dbSNPなど、入力マッピングデータとは異なるデータからの変異を使う場合
 - Force realignment のチェックを必ずはずす。

Guidance-variant settings	
Guidance-variant track 🐂 dbSNP variants	
Force realignment to guidance-variants	

- 入力マッピングデータを基にした変異データを使う場合
 - Force realignment にチェックを入れる

Guidance-variant settings	
Guidance-variant track 🐂 reads (InDel)	6
Force realignment to guidance-variants	







Local Realignment 実行方法



Navigation Areaから使用するマッピングデータを選択。
 Toolboxから NGS Core Tools > Local Realignment を選択、ダブルクリック。
 ウィザードが起動し、選択したデータが選ばれていることを確認。





Local Realignment実行方法

Gx Local Realignment	×
 Select read mapping Realignment settings 	Realignment settings
Marine and Control of	Alignment settings Image: Realign unaligned ends Multi-pass realignment Guidance-variant settings Guidance-variant track Image: Porce realignment to guidance-variants
210 - 111 2	← Previous → Next ✓ Finish X Cancel

Realign unaligned ends: マッピングの際に マップされなかった末端(soft clipping)を Local Realignment の際に利用するかど うか。アダプターの一部のようなものが 残っていない限り、ここはチェックを入れ る。

Guidance-variant settings:ガイダンスあり、 なしの設定

- Guidance-variant track: ガイダンスに使用するトラックを選択。
 - Force realignment to guidancevariants: ガイダンスが、使用する マッピングデータから作成された場 合は、ここにチェックを入れることで、 より積極的にRealignmentを行える。
 - 注意:ガイダンスに選択するトラックが使用するマッピングデータ揺らいでない場合は、チェックを外してください。そうしないと正確なRealignmentが行えません。



Local Realignment実行方法

 Local Realignment 1. Select read mapping 2. Realignment settings 3. Result handling 	Result handling Output options Image: Create reads track Image: Create stand-alone read mappings Image: Output track of realigned regions Result handling Image: Open Image: Save	 Output options アウトプットの選択 Create reads track: トラックフォーマットでの作成。 Create stand-alone read mappings: スタンドアロンフォーマットでの作成。 Output track of realigned regions: Realignmentされた 個所をトラックとして保存す るかどうか。確認に便利。
2000 - 100 -	✓ Open log ← Previous Next ✓ Finish X Cancel	 Result handling Open: 実行後すぐに開く。 Save: 実行後一旦保存。 Log handling Make log: ログを作成するかどうか。



Local Realignment:結果

- 結果はマッピングのファイルとして作成され、名前の最後に locally realigned として作成されます。
 - スタンドアロンフォーマットで作成した場合

🚽 🚟 chr2 Big selection (Reads) - locally realigned

- トラックフォーマットで作成した場合

この後、通常と同じ方法で変異やInsertion, Deletionの検出を行います。



ターゲット領域の統計値計算



ターゲット領域の統計値計算



- Navigation Areaからマッピングデータを選択。
- Toolboxから Resequencing Analysis > Create Statistics for Target Regions を 選択、ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。





Gx Create Statistics for Ta	rget Regions	x
1. Select read mapping	Set target regions track	
2. Set target regions track		
	Truck actions have	
	Target regions track	
	Track of Target Regions 🎀 hg 19_chr20 (Gene)	0
	Report type	
	Report type 1x, 5x, 10x, 20x, 40x, 80x, 10 💌	
	Coverage	
	Minimum coverage 30	
	Read filters	
2	☑ Ignore non-specific matches	
and and	Ignore broken pairs	
(USP		
Star and Star and Star		
107		
Teres .		
TO COMPANY AND		
1		
? 🥱	← Previous → Next ✓ Finish X Car	ncel





ターゲット領域の統計値計算

→ reads (Reads) - locally realigned (coverage) Image: reads (Reads) - locally realigned (coverage report) Image: reads (Reads) - locally realigned (coverage table)

 Genome, Gene, Mapping, Coverage トラックを使ってト ラックリストを作成してみま しょう。





SNV検出





- Basic Variant Detection : クオリティと、バリアントの見られる頻度からバリアントのサイトを検出(version 7.5以前のQuality-Based Variant Detection)。
- Fixed Ploidy Variant Detection:確率モデルを使い、バリアントのサイトを 検出(version 7.5以前のProbabilistic Variant Detection)。
- Low Frequency Variant Detection: 低頻度で見られるバリアントの検出 ツール。倍数性を指定しないでバリアントの検出が行える。

使い分け:

バリアントの見られる頻度が、その領域において15%以下のような場合は、Basic Variant Detection, それよりも多い場合は、Fixed Ploidy Variant Detection をご利用ください。バリアントの見られる頻度が低い場合や、倍数性を指定できない場合などは、Fixed Ploidy Variant Detection をご利用ください。



Quality	
Neighborhood radius 5	
Maximum gap and mismatch count 2	
Minimum neighborhood quality 15	5
Minimum central quality 20)

Mapping後のデータに対し、を設定し、許容するミスマッ チや、gap、またQuality ScoreによりSNP detectionに含 めるデータのフィルタリングを行う。

Low-qu	ality redus are ignored in	
	Minimum coverage	10
Minimu	m variant frequency (%)	35.0
Adv	vanced	
	Maximum coverage	99999
	Required variant count	1
	Sufficient variant count	5

SNP とCallするために、最低必要なカバレッジや、SNPの 頻度を設定する。

CCCACG





- Navigation Areaからマッピングデータを選択。
- Toolboxから Resequencing Analysis > Quality-based Variant Detection を選 択、ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。



Gx Basic Variant Detection	×	
1. Select read mappings	Basic Variant Parameters	
2. Basic Variant Parameters		
	Basic Variant Parameters Ploidy 2	■ Ploidy: 参照配列の倍数性
Of Comment		
144 144 144 144 144 144 144 144 144 144		
LEWING		
AND MARKING		
? 4	← Previous → Next ✓ Finish X Cancel	





Gx Basic Variant Detection	×	Reference masking
 Select read mappings Basic Variant Parameters 	General filters 共通フィルター	 Ignore positions with coverage above:カバレッジが指定した 数字以上のバリアントについてリストに含めない
3. General filters	Reference masking Ignore positions with coverage above 100000 Restrict calling to target regions	 Restrict calling to target regions:バリアントを検出したい領域の指定(アノテーショントラックで指定)
	Read filters Ignore broken pairs Ignore non-specific matches Regions Minimum read length 20 Coverage and count filters Minimum coverage 10 Minimum frequency (%) 35.0	 Ignore broken pairs:ペアエンドのリードでペアと認識されな かったリードをバリアント検出の計算に含めるかどうか Ignore non-specific matches:「Reads」を選択すると、non- specificなマッチのリードを計算に含めなくなり、「Regions」を選 択すると、1本でもnon-specificなリードが含まれる場合、その 領域のバリアントを検出しません。 Minimum read length: Ignore broken pairとIgnore non-specific regions が指定された場合、このフィルターの対象となる最小 のリードの長さの設定が必要です。これは非常に短いリードは、 その短さからnon-specificになる可能性があるためです。
	← Previous → Next ✓ Finish X Cancel	Coverage and count filters Minimum coverage:最小カバレッジ Minimum count:バリアントを支持するリードの最低カウント数 Minimum frequency (%):最小頻度



	Basic Vari	ant Detection
		Quality filter ■ Base quality filter:塩基のクオリティに関するフィルター ■ Neighborhood radius:クオリティフィルターの対象とす る横方向の塩基数(奇数)
Gx Basic Variant Detection 1. Select read mappings	Noise filters Quality filters 共通フィルター	 Minimum central quality:縦方向の数(リード数) Minimum neighborhood quality:Neighborhood radiusで ビマレた範囲の是低クオリティ(Phrod score)
 Basic Variant Parameters General filters Noise filters 	Base quality filter Neighborhood radius 5 Minimum central quality 20 Minimum neighborhood quality 15 Direction and position filters	 Direction and position filters:リードの方向(ForwardとReverse)とポジションを使ったフィルター Read direction filter:どちらか一方の方向のリードが多数見られる場合にそれを排除(ただし、アンプリコンには適していませ
	Read direction filter Direction frequency (%) Relative read direction filter Significance (%) Read position filter Significance (%) 1.0	 ん)。 Relative read direction filter:リードの方向が一方のみに偏り すぎていないか、全体のForwardとReverseのバランスを見て 統計検定を行う。Significanceで閾値を入力。 Read position filter:システマティックなエラーを取り除くために 用いるツールでハイブルダイゼーションを行った場合のデータ
1997 10 000 000 000 000 000 000 000 000 000	Technology specific filters Remove pyro-error variants In homopolymer regions with minimum length With frequency below	に有効。リードを5つのセグメントに分割し、バリアントの見られるポジションの5つのセグメントに分割されたリードの分布が全体のそれと似ているかどうか検定を行う。Significanceで閾値を入力。
S C	← Previous → Next ✓ Finish X Canc	 Technology specific filters Remove pyro-error variants:ホモポリマー領域に対するエラーの除去 In homopolymer regions with minimum length:指定した長さのホモポリマー領域のInDelを取り除く。 With frequency below:指定した頻度以下のものについてのみフィルターを適用。



Gx Quality-based Variant	Detection
1. Select read mappings	Result handling
2. Set quality filters	
 Set significance thresholds 	
4. Set genome information	
5. Result handling	
	Output options
	Create track
	Create annotated table
	Result handling
	O Open
	Save
	Log handling
and and	Open log
(EP	
State Contraction of State and State	
1 free	
TOWNER CONTRACTOR	
IL PRIMI	
?	← Previous → Next ✓ Finish X Cancel

 Create track: トラックの作成
 Create annotated table: アノテーション付の テーブルの作成





新しいフィルター

7.5より3つのバリアント検出に共通のフィルターが搭載され、いくつか新しいフィルターも加わりました。

General filters 共通フィルター	 Reference masking Ignore positions with coverage above:カバレッジが指定した 数字以上のバリアントについてリストに含めない Restrict calling to target regions:バリアントを検出したい領域の指定(アノテーショントラックで指定)
Ignore positions with coverage above 100000 Restrict calling to target regions Image: Coverage and count filters Ignore non-specific matches Reads< ▼ Minimum read length 20 Coverage and count filters 10 Minimum coverage 10 Minimum count 2 Minimum frequency (%) 1.0	 Read filters Ignore broken pairs:ペアエンドのリードでペアと認識されな かったリードをバリアント検出の計算に含めるかどうか Ignore non-specific matches:「Reads」を選択すると、non- specificなマッチのリードを計算に含めなくなり、「Regions」を選 択すると、1本でもnon-specificなリードが含まれる場合、その 領域のバリアントを検出しません。 Minimum read length:Ignore broken pairとIgnore non-specific regions が指定された場合、このフィルターの対象となる最小 のリードの長さの設定が必要です。これは非常に短いリードは、 その短さからnon-specificになる可能性があるためです。
← Previous → Next ✓ Finish X Cancel	Coverage and count filters Minimum coverage:最小カバレッジ Minimum count:バリアントを支持するリードの最低カウント数 Minimum frequency (%):最小頻度



	新しい	フィルター
 Select read mappings Low frequency variant parameters General filters Naise filters 	Noise filters Quality filters Base quality filter Neighborhood radius 5 Minimum central quality 20	Quality filter Base quality filter:塩基のクオリティに関するフィルター Neighborhood radius:クオリティフィルターの対象とす る横方向の塩基数(奇数) Minimum central quality:縦方向の数(リード数) Minimum neighborhood quality:Neighborhood radiusで 指定した範囲の最低クオリティ(Phred score)
4. Noise hiters	Minimum neighborhood quality 15 Direction and position filters Read direction filter Direction frequency (%) 5.0 Relative read direction filter Significance (%) 1.0 Read position filter Significance (%) 1.0 Technology specific filters Remove pyro-error variants In homopolymer regions with minimum length 3 With frequency below 0.8	 Direction and position filters:リードの方向(ForwardとReverse)とポジションを使ったフィルター Read direction filter:どちらか一方の方向のリードが多数見られる場合にそれを排除(ただし、アンプリコンには適していません)。 Relative read direction filter:リードの方向が一方のみに偏りすぎていないか、全体のForwardとReverseのバランスを見て統計検定を行う。Significanceで閾値を入力。 Read position filter:システマティックなエラーを取り除くために用いるツールでハイブリダイゼーションを行った場合のデータに有効。リードを5つのセグメントに分割し、バリアントの見られるポジションの5つのセグメントに分割されたリードの分布が全体のそれと似ているかどうか検定を行う。Significanceで閾値を入力。
		 Technology specific filters Remove pyro-error variants:ホモポリマー領域に対するエラーの除去 In homopolymer regions with minimum length:指定した長さのホモポリマー領域のInDelを取り除く。 With frequency below:指定した頻度以下のものについてのみフィルターを適用。

フィルターによる除去例



Figure 27.22: The same data as in figure 27.21, now with the 'Show quality scores' option in the reads track switched on.

Basic quality filter 適用例:マッピングしたリードをクオリティで表示。クオリティの低いリードがマップされている箇所がバリアントのリストからはずされます。



フィルターによる除去例



Read direction filter 適用例:リードの色は緑(Forward)、赤(Reverse)、黄色(non-specific)を示しており、緑のリードが大部分のバリアントをサポートしていることがわかる。こういったアンバランスな箇所で検出されたバリアントが取り除かれる。





フィルターによる除去例





もしリードが理想的な均一なカバレッジであれば、検出されるバリアントをサポートする塩基のリード中の位置は、さまざまになるはずです。



 これを使い、リードをForward、Reverseの向きを 考慮して、それぞれ5分割、計10個の領域に分 断し、変異が見つかった箇所がリードのどの領 域に属するか、それらの分布が全体と大きく差 がないかを検定しています。





in reads trimmed. ×
 for show more tracks together, create a track list. ① Create Track List
 10,000,000
 20,000,000
 40,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000
 60,000,000



Rows: 231	Table vie	w: Genor	ne																			Filter]₹
Chrom Region	Туре	Refe	Allele	Ref	Len	Zygosity	Count	Cov	Freq	For	Rev	For	Averag	Rea	Rea	# u	# u	Bas	Rea	Rea	Нур	Но	
IC_010 1115086	SNV	A	G	No	1	Homozy	22	24	91.67	12	10	0.45	31.82	22	24	19	19		0.99	0.95	no	No	
IC_010 1152101	SNV	G	С	No	1	Heteroz	5	13	38.46	5	0	0.00	10.00	5	13	5	5	2.47	0.17	1.00	no	No	
C_010 1152101	SNV	G	G	Yes	1	Heteroz	8	13	61.54	8	0	0.00	28.38	8	13	7	7		0.38	1.00	no	No	
C_010 1152139	SNV	Α	Α	Yes	1	Heteroz	7	11	63.64	7	0	0.00	18.00	7	11	6	6		0.73	1.00	no	No	
C_010 1152139	SNV	Α	C	No	1	Heteroz	4	11	36.36	4	0	0.00	16.25	4	11	4	4	0.00	0.45	1.00	no	No	
C_010 1154738	SNV	Α	G	No	1	Homozy	34	34	100.00	23	11	0.32	30.53	34	34	25	25		1.00	1.00	no	No	
C_010 1166395.	MNV	TT	CC	No	2	Homozy	35	35	100.00	18	17	0.49	33.50	35	35	26	26		1.00	1.00	no	No	
C_010 1178224	SNV	Т	C	No	1	Homozy	38	38	100.00	19	19	0.50	33.58	38	38	27	27		1.00	1.00	no	No	
C_010 1186219	SNV	Т	G	No	1	Heteroz	10	27	37.04	0	10	0.00	10.40	10	27	9	9	2.63	0.02	0.15	no	No	
C_010 1186219	SNV	Т	Т	Yes	1	Heteroz	17	27	62.96	8	9	0.47	25.65	17	27	14	14		0.23	0.50	no	No	E
IC_010 1224504	SNV	G	Α	No	1	Homozy	33	33	100.00	13	20	0.39	32.73	33	33	24	24		1.00	1.00	no	No	
0 010 1044640	CAN	<u>_</u>		Me		Homory	24	24	100.00	10	14	0.47	22.01	24	24	26	26		1 00	1.00		Me	

- Count: クオリティのフィルターをパスしたリードの数
- Coverage: クオリティのフィルターをパスしたリードの数
- Frequency: バリアントが見られた頻度
- Probability: バリアントのアレルの事後確率(そのアレルが尤もであるとする確率。高い方がより確 度が高いという事。)
- Forward reads: その領域に見られたForwardリードの数
- Reverse reads:その領域に見られたReverseリードの数
- Forward/reverse: Forward/Total reads または Reverse/Total reads のうち小さい方の値。 ForwardとReverseが同じなら、0.5となる。
- Average quality: 該当する領域の平均リードクオリティ。
- # unique start positions:バリアントコールに使われたリードのうちスタートポジションにあるリードの数
- # unique end positions:バリアントコールに使われたリードのうち最後の箇所にあるリードの数
- BaseQRankSum:クオリティスコアについて、参照配列と同じアレルとバアリアントのアレルについてマンホイットニーU検定を行い計算されたZスコア。これが高いほど参照配列の塩基とバリアントの塩基に差がある。
- Hyper-alleic:想定されるアレルよりも頻度が高いかどうか
- Homopolymer:ホモポリマー領域かどうか

how column	-
🔽 Chromosome	
📝 Region	
📝 Туре	
Reference	
Allele	
Reference allele	
🔽 Length	
Linkage	
V Zygosity	
Count	
Coverage	
Frequency	
Probability	
Forward read count	
Reverse read count	
Forward/reverse balance	
📝 Average quality	
📝 Read count	
📝 Read coverage	
📝 # unique start positions	
📝 # unique end positions	
📝 BaseQRankSum	
📝 Read position test probability	
📝 Read direction test probability	
V Hyper-allelic	
🚺 Homopolymer	
Select All	
[] 레	7



Basic Variant Detection:トラックリスト作成





Toolboxから ResequTrack Tools > Create Track List を選択、ダブルクリック。

■ ウィザードが起動し、選択したデータが選ばれていることを確認。





テーブルが現れます。テーブルの行と、マッピングのビューア は対応しているので、テーブルで指定したポジションに自動的 にビューアが移動します。



Fixed Ploidy Variant Detection 詳細

Probabilistic Variant Detection

• 確率モデル(Bayes model)を使ったバリアント検出





与えられるリードから、そのポジションのSite Typeを推定 Reference と推定したSite typeが異なる場合、バリアントとして結果 返す。



Fixed Ploidy Variant Detection 詳細

Bayes model



 $P(A \cap B) = P(B \mid A)P(A)$ $P(A \cap B) = P(A \mid B)P(B)$ $P(B \mid A)P(A) = P(A \mid B)P(B)$



Reference ? : Site type (ex) A/A, A/T, A/C ... ? S:Site type $P(S \mid R) = \frac{P(R \mid S)P(S)}{P(R)}$ *R* : Reads

Fixed Ploidy Variant Detection

P(R|S): Error Model を使って推定 P(S): Genome Model を使って推定



Fixed Ploidy Variant Detection 詳細

Genome Model

- Reference がAのとき、Readの大部分はAになると仮定し、初期の確率を以下のように 設定し、EMアルゴリズムを使ってそれぞれの確率を推定する。
 - EMアルゴリズム(Expectation Maximization algorithm)は、得られたデータから推定したい現象が観察できない場合に、その確率を推定する、一般的な統計の手法。

Site Type	Initial Probability
A/A	0.2475
A/C	0.001
A/G	0.001
A/T	0.001
T/C	0.001
T/G	0.001
T/T	0.2475
G/C	0.001
C/C	0.2475
G/G	0.2475
G/-	0.001
A/-	0.001
C/-	0.001
T/-	0.001



Fixed Ploidy Variant Detection 詳細

- Error Model
 - リードに含まれるエラーを考慮するため、尤度のところにエラーを考慮した確率を推定する。初期値を以下のように設定し、EMアルゴリズムにて確率を推定する。

	Reference	А	С	G	т	_
Reads						
А		0.90	0.025	0.025	0.025	0.025
С		0.025	0.90	0.025	0.025	0.025
G		0.025	0.025	0.90	0.025	0.025
Т		0.025	0.025	0.025	0.90	0.025
-		0.025	0.025	0.025	0.025	0.90



Fixed Ploidy Variant Detection







- Navigation Areaからマッピングデータを選択。
- Toolboxから Resequencing Analysis > Variant Detectors > Fixed Ploidy Variant Detection を選択、ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。



Fixed Ploidy Variant Detection

Gx Fixed Ploidy Variant	Detection	
 Select read mappings Fixed ploidy variant parameters 	Fixed ploidy variant parameters	
Constant of the second	Fixed ploidy variant parameters Ploidy 2 Required variant probability (%) 90.0	 Ploidy:参照配列の倍数性 Required variant probability:バリアントが参照 配列と異なる確率(想定で入力)。この値を低く すると、検出されるバリアントが多くなります。
?	← Previous → Next ✓ Finish X Cancel	


Fixed Ploidy Variant Detection

Gx Fixed Ploidy Variant	Detection	Reference masking
 Select read mappings Fixed ploidy variant parameters General filters 	General filters	 Ignore positions with coverage above: カハレッシか指定した 数字以上のバリアントについてリストに含めない Restrict calling to target regions:バリアントを検出したい領域の指定(アノテーショントラックで指定)
3. General filters	Reference masking Ignore positions with coverage above Restrict calling to target regions Read filters Ignore broken pairs Ignore non-specific matches Reads Minimum read length 20 Coverage and count filters Minimum coverage 10 Minimum count 2 Minimum frequency (%) 20.0	 Read filters Ignore broken pairs:ペアエンドのリードでペアと認識されな かったリードをバリアント検出の計算に含めるかどうか Ignore non-specific matches:「Reads」を選択すると、non- specificなマッチのリードを計算に含めなくなり、「Regions」を選 択すると、1本でもnon-specificなリードが含まれる場合、その 領域のバリアントを検出しません。 Minimum read length:Ignore broken pairとIgnore non-specific regions が指定された場合、このフィルターの対象となる最小 のリードの長さの設定が必要です。これは非常に短いリードは その短さからnon-specificになる可能性があるためです。
17 Charles (Barring (Charles (Barring (Charles (Barring (Charles (Barring (Charles (Cha	← Previous → Next ✓ Finish X Cancel	Coverage and count filters Minimum coverage:最小カバレッジ Minimum count:バリアントを支持するリードの最低カウント数 Minimum frequency (%):最小頻度



Fixed Ploidy Variant Detection

		Quality filter
Gx Fixed Ploidy Variant I	Detection	■ Base quality filter:塩基のクオリティに関するフィルター
 Select read mappings Fixed ploidy variant parameters General filters 	Noise filters Quality filters Base quality filter Neighborhood radius Minimum central quality 20	 Neighborhood radius:クオリティフィルターの対象とす る横方向の塩基数(奇数) Minimum central quality:縦方向の数(リード数) Minimum neighborhood quality:Neighborhood radiusで 指定した範囲の最低クオリティ(Phred score)
4. Noise filters	Minimum central quanty 20 Minimum neighborhood quality 15 Direction and position filters Image: Constraint of the constr	 Direction and position filters:リードの方向(ForwardとReverse)とポジションを使ったフィルター Read direction filter:どちらか一方の方向のリードが多数見られる場合にそれを排除(ただし、アンプリコンには適していません)。 Relative read direction filter:リードの方向が一方のみに偏りすぎていないか、全体のForwardとReverseのバランスを見て統計検定を行う。Significanceで閾値を入力。 Read position filter:システマティックなエラーを取り除くために用いるツールでハイブリダイゼーションを行った場合のデータに有効。リードを5つのセグメントに分割し、バリアントの見られるポジションの5つのセグメントに分割されたリードの分布が全体のそれと似ているかどうか検定を行う。Significanceで閾値を入力。 Technology specific filters
?	← Previous → Next ✓ Finish	 Remove pyro-error Variants:ホモホリマー領域に対するエラーの除去 In homopolymer regions with minimum length:指定した長さのホモポリマー領域のInDelを取り除く。 With frequency below:指定した頻度以下のものについてのみフィルターを適用。

Fixed Ploidy Variant Detection

Gx Fixed Ploidy Variant I	Detection	
 Select read mappings Fixed ploidy variant parameters 	Result handling	
 General filters Noise filters 	Output options	
5. Result handling	Create annotated table Create report Result handling Open	 Create track: トラックの作成 Create annotated table: アノテーション付の テーブルの作成
	 Save Log handling Open log 	
Proprietor	← Previous → Next ✓ Finish X Cancel	









	Bowe	590 Table v	iew: Genome			Filter	Č.						Table Settings
	Tioms.	rabic i	ion conome			T Inter-							· · · · · · · · · · · · · · · · · · ·
	Туре	Reference	Allele	Reference	Zyzosity	Count	Coverage	Frequency	Forward read count	Reverse read count	Forward/reverse	Average quality	Column width
	SNV	A	C	No	Homozverous		75	75 100.0	35	3 4:	0.44	24.39	Show column
	SNV	C	T	No	Heteroz vizous		31	72 43.01	1	7 14	0.45	23.68	Chramasana
	SNV	С	С	Yes	Heterozygous		41	72 56.94	18	3 23	0.44	24.15	Chromosome
	SNV	С	т	No	Homozygous		57	57 100.01	16	5 4:	. 0.26	24.46	Region
	SNV	C	G	No	Homozygous		39	39 100.00	24	1	i 0.38	24.00	V Type
	SNV	Т	G	No	Homozygous		70	70 100.00	36	5 34	0.49	29.30	Reference
	SNV	A	G	No	Homozygous		69	69 100.00) 80) 3!	0.43	29.74	
	SNV	G	A	No	Homozygous		44	44 100.00) 8	3 31	i 0.18	30.80	V Allele
	SNV	C	Т	No	Heterozygous		7	12 58.3) E	5 :	. 0.29	24.43	Reference allele
	SNV	C	С	Yes	Heterozygous		5	12 41.6	1 1	1	0.20	29.80	🕅 Linkage
/	SNV	Т	C	No	Homozygous		11	11 100.00) 8	3 1	0.27	28.00	7 Zurocitu
	SNV	Т	C	No	Homozygous		18	41 43.9	15	7	0.06	24.22	V Zygosity
	SNV	A	C	No	Heterozygous		10	21 47.63	2 9	3	0.10	27.90	Count
	SNV	A	A	Yes	Heterozygous		11	21 52.3	10)	0.09	29.00	Coverage
	SNV	C	Т	No	Homozygous		9	20 45.0	1 8	3	0.11	24.56	Frequency
	SNV	A	Т	No	Heterozygous		27	53 50.94	1 2	2 21	i 0.07	24.26	
	SNV	A	A	Yes	Heterozygous		25	53 47.13	1 2	2 23	80.08	23.56	Probability
	SNV	C	Т	No	Heterozygous		7	19 36.84	1 8	3 4	0.43	25.71	Forward read count
	SNV	C	С	Yes	Heterozygous		12	19 63.11	i t	5	0.42	24.50	Reverse read count
	SNV	т	С	No	Homozygous		14	14 100.01	18	3	0.07	24.21	
	SNV	A	G	No	Homozygous		14	15 93.3	1) ,	0.29	23.79	Porward/reverse bala
	SNV	Т	G	No	Homozygous		7	11 63.64	1	1	i 0.14	25.29	Average quality
	SNV	A	G	No	Homozygous		8	21 38.11) 7	1	0.12	25.62	Select All
	SNV	A	G	No	Heterozygous		5	10 50.01) (ŧ –	0.20	29.80	Decelect All
	SNV	A	A	Yes	Heterozygous		5	10 50.01)	1	0.20	26.20	Deselect Mil
	5NV	C	Т	No	Homozygous		56	56 100.00	41	1 11	i 0.27	31.11	

112

■ バリアントテーブルの行をクリックすると、

Fixed Ploidy Variant Detection:結果

Rows: 2	231 T	able vie	w: Genor	ne																			Filter]₹
Chrom	Region	Туре	Refe	Allele	Ref	Len	Zygosity	Count	Cov	Freq	For	Rev	For	Averag	Rea	Rea	# u	# u	Bas	Rea	Rea	Нур	Но	
IC_010	1115086	SNV	A	G	No	1	Homozy	22	24	91.67	12	10	0.45	31.82	22	24	19	19		0.99	0.95	no	No	
C_010	1152101	SNV	G	С	No	1	Heteroz	5	13	38.46	5	0	0.00	10.00	5	13	5	5	2.47	0.17	1.00	no	No	
C_010	1152101	SNV	G	G	Yes	1	Heteroz	8	13	61.54	8	0	0.00	28.38	8	13	7	7		0.38	1.00	no	No	
C_010	1152139	SNV	A	A	Yes	1	Heteroz	7	11	63.64	7	0	0.00	18.00	7	11	6	6		0.73	1.00	no	No	
C_010 :	1152139	SNV	A	С	No	1	Heteroz	4	11	36.36	4	0	0.00	16.25	4	11	4	4	0.00	0.45	1.00	no	No	
	1154738	SNV	Α	G	No	1	Homozy	34	34	100.00	23	11	0.32	30.53	34	34	25	25		1.00	1.00	no	No	
	1166395	MNV	TT	CC	No	2	Homozy	35	35	100.00	18	17	0.49	33.50	35	35	26	26		1.00	1.00	no	No	
	1178224	SNV	т	С	No	1	Homozy	38	38	100.00	19	19	0.50	33.58	38	38	27	27		1.00	1.00	no	No	
C_010 :	1186219	SNV	Т	G	No	1	Heteroz	10	27	37.04	0	10	0.00	10.40	10	27	9	9	2.63	0.02	0.15	no	No	
_010	1186219	SNV	Т	Т	Yes	1	Heteroz	17	27	62.96	8	9	0.47	25.65	17	27	14	14		0.23	0.50	no	No	=
	1224504	SNV	G	A	No	1	Homozy	33	33	100.00	13	20	0.39	32.73	33	33	24	24		1.00	1.00	no	No	
010	1044640	CAR/	<u>_</u>	Δ.	Ma		Homony	24	24	100.00	10	14	0.47	22.01	24	24	26	76		1.00	1.00		Me	

- Count: クオリティのフィルターをパスしたリードの数
- Coverage: クオリティのフィルターをパスしたリードの数
- Frequency: バリアントが見られた頻度
- Probability: バリアントのアレルの事後確率(そのアレルが尤もであるとする確率。高い方がより確 度が高いという事。)
- Forward reads: その領域に見られたForwardリードの数
- Reverse reads:その領域に見られたReverseリードの数
- Forward/reverse: Forward/Total reads または Reverse/Total reads のうち小さい方の値。 ForwardとReverseが同じなら、0.5となる。
- Average quality: 該当する領域の平均リードクオリティ。
- # unique start positions:バリアントコールに使われたリードのうちスタートポジションにあるリードの数
- # unique end positions:バリアントコールに使われたリードのうち最後の箇所にあるリードの数
- BaseQRankSum:クオリティスコアについて、参照配列と同じアレルとバアリアントのアレルについてマンホイットニーU検定を行い計算されたZスコア。これが高いほど参照配列の塩基とバリアントの塩基に差がある。
- Hyper-alleic:想定されるアレルよりも頻度が高いかどうか
- Homopolymer:ホモポリマー領域かどうか

how column	-
Chromosome	
Region	
🔽 Туре	
Reference	
V Allele	
Reference allele	
🔽 Length	
Linkage	
Zygosity	
Count	
🔽 Coverage	
V Frequency	
Probability	
Forward read count	
Reverse read count	
Forward/reverse balance	
🔽 Average quality	
🔽 Read count	
🔽 Read coverage	
📝 # unique start positions	
📝 # unique end positions	
📝 BaseQRankSum	
📝 Read position test probability	
📝 Read direction test probability	
V Hyper-allelic	
V Homopolymer	
Select All	
[] 리	



- Low frequency Variant Detection では、倍数性を仮定せず、対象となる 領域が、シーケンスエラーなのか、そうではない(=バリアント)なのかを検 定しています。
- Error モデルについては、Fixed Ploidy Variant Detection にて採用したエラーモデルを使い、計算し、尤度比検定を行っています。





- Navigation Areaからマッピングデータを選択。
 Toolboxから Resequencing Analysis > Variant Detector
- Toolboxから Resequencing Analysis > Variant Detectors > Low Frequency Variant Detection を選択、ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。



Gx Low Frequency Varian	nt Detection	×	
 Select read mappings Low frequency variant parameters 	Low frequency variant parameters		
	Previous → Next ✓ Finish	Required s	significance :シーケンスエラーかどうか、検定の際の閾値



Gx Low Frequency Varian	nt Detection
1. Select read mappings	General filters
2. Low frequency variant parameters	
3. General filters	Reference masking Ignore positions with coverage above 100000 Restrict calling to target regions
	Read filters ✓ Ignore broken pairs Ignore non-specific matches Reads Minimum read length 20
	Coverage and count filters Minimum coverage 10 Minimum count 2 Minimum frequency (%) 1.0
The second secon	
?	← Previous → Next ✓ Finish X Cancel

Reference masking

- Ignore positions with coverage above:カバレッジが指定した 数字以上のバリアントについてリストに含めない
- Restrict calling to target regions:バリアントを検出したい領域の指定(アノテーショントラックで指定)

Read filters

- Ignore broken pairs:ペアエンドのリードでペアと認識されな かったリードをバリアント検出の計算に含めるかどうか
- Ignore non-specific matches:「Reads」を選択すると、non-specificなマッチのリードを計算に含めなくなり、「Regions」を選択すると、1本でもnon-specificなリードが含まれる場合、その領域のバリアントを検出しません。
- Minimum read length: Ignore broken pairとIgnore non-specific regions が指定された場合、このフィルターの対象となる最小 のリードの長さの設定が必要です。これは非常に短いリードは、 その短さからnon-specificになる可能性があるためです。

Coverage and count filters

- Minimum coverage:最小カバレッジ
- Minimum count:バリアントを支持するリードの最低カウント数
- Minimum frequency (%):最小頻度



	Quality filter
 Low Frequency Variant Detection Select read mappings Low frequency variant parameters General filters Noise filters Noise filters Minimum central quality Minimum neighborhood quality Direction and position filters 	 Base quality filter:塩基のクオリティに関するフィルター Neighborhood radius:クオリティフィルターの対象とする横方向の塩基数(奇数) Minimum central quality:縦方向の数(リード数) Minimum neighborhood quality:Neighborhood radiusで指定した範囲の最低クオリティ(Phred score) Direction and position filters:リードの方向(ForwardとReverse)とポジ
Read direction filter Direction frequency (%) 5.0 Relative read direction filter Significance (%) 1.0 Read position filter Significance (%) 1.0 Technology specific filters Remove pyro-error variants In homopolymer regions with minimum length 3 With frequency below 0.8	 ションを使ったフィルター Read direction filter:どちらか一方の方向のリードが多数見られる場合にそれを排除(ただし、アンプリコンには適していません)。 Relative read direction filter:リードの方向が一方のみに偏りすぎていないか、全体のForwardとReverseのバランスを見て統計検定を行う。Significanceで閾値を入力。 Read position filter:システマティックなエラーを取り除くために用いるツールでハイブリダイゼーションを行った場合のデータに有効。リードを5つのセグメントに分割し、バリアントの見られるポジションの5つのセグメントに分割されたリードの分布が全体のそれと似ているかどうか検定を行う。Significanceで閾値を入力。
	 Technology specific filters Remove pyro-error variants:ホモポリマー領域に対するエラーの除去 In homopolymer regions with minimum length:指定した長さのホモポリマー領域のInDelを取り除く。 With frequency below:指定した頻度以下のものについてのみフィルターを適用。

Gx Low Frequency Variar	nt Detection	
1. Select read mappings	Result handling	
parameters		
3. General filters	Output options	
4. Noise filters	Create track	
5. Result handling	Create annotated table	
	Create report	 Create track: トラックの作成 Create appateted table: スノニーションけの
	Result handling	■ Create annotated table. アフリーション内の テーブルの作成
	 Save 	
	Log handling	
()	Open log	
Stational States		
L DALLAND		
TO LA PROVIDENTO LITTLE		
?	← Previous → Next ✓ Finish X Cancel	



Rows:	231 T	able vie	w: Genor	ne																			Filter]₹
Chrom	Region	Туре	Refe	Allele	Ref	Len	Zygosity	Count	Cov	Freq	For	Rev	For	Averag	Rea	Rea	# u	# u	Bas	Rea	Rea	Нур	Но	
IC_010	1115086	SNV	A	G	No	1	Homozy	22	24	91.67	12	10	0.45	31.82	22	24	19	19		0.99	0.95	no	No	
C_010	1152101	SNV	G	С	No	1	Heteroz	5	13	38.46	5	0	0.00	10.00	5	13	5	5	2.47	0.17	1.00	no	No	
C_010	1152101	SNV	G	G	Yes	1	Heteroz	8	13	61.54	8	0	0.00	28.38	8	13	7	7		0.38	1.00	no	No	
C_010	1152139	SNV	A	A	Yes	1	Heteroz	7	11	63.64	7	0	0.00	18.00	7	11	6	6		0.73	1.00	no	No	
C_010	1152139	SNV	A	С	No	1	Heteroz	4	11	36.36	4	0	0.00	16.25	4	11	4	4	0.00	0.45	1.00	no	No	
C_010	1154738	SNV	A	G	No	1	Homozy	34	34	100.00	23	11	0.32	30.53	34	34	25	25		1.00	1.00	no	No	
C_010	1166395	MNV	TT	CC	No	2	Homozy	35	35	100.00	18	17	0.49	33.50	35	35	26	26		1.00	1.00	no	No	
C_010	1178224	SNV	т	С	No	1	Homozy	38	38	100.00	19	19	0.50	33.58	38	38	27	27		1.00	1.00	no	No	
C_010	1186219	SNV	Т	G	No	1	Heteroz	10	27	37.04	0	10	0.00	10.40	10	27	9	9	2.63	0.02	0.15	no	No	
C_010	1186219	SNV	Т	Т	Yes	1	Heteroz	17	27	62.96	8	9	0.47	25.65	17	27	14	14		0.23	0.50	no	No	E
C_010	1224504	SNV	G	A	No	1	Homozy	33	33	100.00	13	20	0.39	32.73	33	33	24	24		1.00	1.00	no	No	
C 010	1044640	CAN	<u>c</u>	Δ.	Ma		Homory	24	24	100.00	10	14	0.47	22.01	24	24	26	26		1 00	1.00		Me	

- Count: クオリティのフィルターをパスしたリードの数
- Coverage: クオリティのフィルターをパスしたリードの数
- Frequency: バリアントが見られた頻度
- Probability: バリアントのアレルの事後確率(そのアレルが尤もであるとする確率。高い方がより確 度が高いという事。)
- Forward reads: その領域に見られたForwardリードの数
- Reverse reads:その領域に見られたReverseリードの数
- Forward/reverse: Forward/Total reads または Reverse/Total reads のうち小さい方の値。 ForwardとReverseが同じなら、0.5となる。
- Average quality: 該当する領域の平均リードクオリティ。
- # unique start positions:バリアントコールに使われたリードのうちスタートポジションにあるリードの数
- # unique end positions:バリアントコールに使われたリードのうち最後の箇所にあるリードの数
- BaseQRankSum:クオリティスコアについて、参照配列と同じアレルとバアリアントのアレルについてマンホイットニーU検定を行い計算されたZスコア。これが高いほど参照配列の塩基とバリアントの塩基に差がある。
- Hyper-alleic:想定されるアレルよりも頻度が高いかどうか
- Homopolymer:ホモポリマー領域かどうか

how column	-
Chromosome	
Region	
🔽 Туре	
Reference	
Allele	
Reference allele	
Length	
Linkage	
Zygosity	
Count	
Coverage	
Frequency	
Probability	
Forward read count	
Reverse read count	
Forward/reverse balance	
🚺 Average quality	
Read count	
Read coverage	
🚺 # unique start positions	
📝 # unique end positions	
BaseQRankSum	
Read position test probability	
Read direction test probability	
V Hyper-allelic	
V Homopolymer	
Select All	
	1



構造変異と大きなInDel



- Quality Based Variant Detection やProbabilistic Variant Detection では 変異やInDelを検出できました。
- しかしながら大きなInDelの検出や構造変異については、上記ツールでの
 検出は難しい場合があります。



- アルゴリズムにとっては、大きなInsertionやDeletionを受け入れるよりは、 Unaligned endとするほうがスコアを大きくできるからです。
- InDel and Structural Variants ツールでは、このUnaligned end に着目して、大きなInDelや構造変異を見つけます。
- Unaligned end が別の領域に十分な量マップすることができれば、そこまでの距離のInsertionやDeletion、構造変異と考えられます。
- ・ 注意:このツールでは、同一染色体内の構造変異のみが検出可能です。



Processes Toolbox Favorites

- Navigation Areaからマッピングデータを選択。
- Toolboxから Resequencing Analysis > InDels and Structual Variantsを選択、 ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。



Gx InDels and Structur	al Variants	
 Select read mappings Select settings 	Select settings	
	Significance of unaligned end breakpoints P-Value threshold 0.0001 Maximum number of mismatches 3 Filter variants Filter variants Minimum number of reads 2	 Significance of unaligned end breakpoints Unaligned end とするに十分なリードがあるかどうか、二項分布に当てはめて検定をしている。p値を小さく設定するほど、j 十分なリードがないと、Unaligned end とならない。
		 Filter variants 変異とするための最低必要なリード数を 指定するかどうか
1 1001 Control Value	← Previous → Next ✓ Finish X Ca	ancel



Gx InDels and Structura	l Variants	
1. Select read mappings	Result handling	
2. Select settings		
3. Result handling		
	Output options	
	✓ Create report	
	Create breakpoints	
	Create InDel variants	Output options
	Create structural variations	
		■ 出力するレポートの植類を選択
	Result handling	
	O Open	
	Save	
6	Log handling	
O C C	Open log	
(USP		
an amaring		
10 1 0 1		
Lever Contract		
THE O'S AND ROUTH		
and the second s		
?	← Previous → Next ✓ Finish	× Cancel



*** reads (Reads) - locally realigned (InDel)

✦reads (Reads) - locally realigned (SV)
iereads (Reads) - locally realigned structural variants report

📰 reads (Reads) 🤇	× 🖽 reads (I	Reads) 🗙 🖽 reads (Reads) 🗙					
Rows: 21 Ta	able view: Geno	me				Filter	Ŧ
Region	Туре	Reference	Allele	Reference a	Length	Zygos	
1382907213829075	Deletion	AAGA	-	No		4 Hetero	
1510091815100927	Deletion	GTGTGTGTGC	-	No		10 Homo:	
17472115^17472116	Insertion	-	CCCTTCATGGATGCTGTCGGGG	No		22 Hetero	
1800272418002737	Deletion	TAGGTGATGGGGTG	-	No		14 Hetero	
20335557^20335558	Insertion	-	GGAGGTGGTCACACTGTGAGAGGGGAGGTGGTCA	No		34 Hetero	1
2176686421766878	Deletion	ACTTGAGAACTGAGG	-	No		15 Homo:	
22137015^22137016	Insertion	-	CAGCATATAGCCTTGGCAGCCTATAGCCTTGG	No		32 Homos	
23615819^23615820	Insertion	-	CACACA	No		6 Hetero	
2418883024188883	Deletion	GTGTGCAATCGCCTGTGTGTGTGCATGCGTGTG	-	No		54 Hetero	
2418898424189025	Deletion	ATGTGTGTGCGCGTGTGTATGTGTAATCTCCTG	-	No		42 Homos	
2418912924189154	Deletion	CGTGTAATCTCCTGTGTGCGTGCGTG	-	No		26 Homo:	
24396453^24396454	Insertion	-	GTCCTGTGATCTCACTCTGCCCTC	No		24 Hetero	-

reads (Reads)... × 📰 reads (Reads)... ×

						_
Rows: 17	Table view: Genome				Filter	₹
Chromosome	Region	Name	Evidence	Length	Reference sequence	
20	complement(1616924116169432)	Complex	Cross mapped breakpoints (invalid orientation)	192		
20	1744915817449220	Replacement	Paired breakpoint	63	TCCTGCTGTAGGCGGGCAGCCTACCATAGG	p
20	1747205017472114	Complex	Multiple breakpoints	65		
20	1747207217472114	Complex	Multiple breakpoints	43		
20	complement(1990502019905133)	Complex	Cross mapped breakpoints (invalid orientation)	114		Ξ
20	2006507320065272	Replacement	Paired breakpoint	200	CTGCTGCTGCTGTCACTCCAGTCCTTAAGTG	л
20	complement(2033556020335591)	Complex	Cross mapped breakpoints (invalid orientation)	32		
20	2033712520338115	Complex	Multiple breakpoints	991		
20	21114722~21114733	Insertion	Close breakpoints	0		-
20	22452630~22452633	Insertion	Close breakpoints	0		
20	23923282~23923285	Insertion	Close breakpoints	0		
20	2403756024037568	Replacement	Paired breakpoint	26	GAGGGTCAA	-
1						

SVの結果に記載されているEvidenceの詳細については、以下を参照ください。 http://www.clcsupport.com/clcgenomicsworkbench/current/index.php?manual=Theoretically_expect

ed structural variant signatures.html



SNV比較とアノテーション付け



変異の比較とアノテーション付け

- 視覚的な比較:検出した変異を複数サンプル(又はアノテーションと)並べて比較することで、視覚的に確認する。
- 比較リストの作成:検出した変異を複数(又はアノテーションと)比較し、
 差があるものをリストアップ。
 - dbSNP
 - 1000genome
- アノテーション付け:検出した変異に既存のデータを使ってアノテーション 付け。
 - 遺伝子名をつける
 - Cosmicの情報を付ける
 - Conservation Score を付ける







 Navigation Areaから変異データを選択。
 Toolboxから Resequencing Analysis > Function consequence > Amino Acid Changeを選択、ダブルクリック。
 ウィザードが起動し、選択したデータが選ばれていることを確認。





Gx Amino Acid Changes	s X	
1. Select variant tracks	Set parameters	
2. Set parameters		
	Select CDS track CDS track 🍂 hg 19_chr20 (CDS)	■ Select CDS track:CDS情報
	Select transcript (mRNA) track mRNA track →the 19 chr20 (mRNA)	 Select transcript track:mRNAなど
	Select sequence track	■ Select sequence track: ゲノム情報
000	Filtering and annotation Filter synonymous Genetic code 1 Standard	 Filtering and annotation アミノ酸置換を引き起こさないものを取り除くかどうか
Contraction of the second		
? (%)	← Previous → Next ✓ Finish X Cancel	



🔤 👫 reads (Variants, AAC)



Reverse real Forward/rev. Average quality Coding region change Amino acid change Amino acid change in longest transcript Coding 59 018 2003 18.60 ENST00003411420c [1257C>T]: ENST0000217246c [272. ENST0000339312p [His400Gh]: ENSP00000339912p His400Gh ENST0000339912p His400Gh ENST00000310348c172. ENST0000339912p His400Gh<	Rows: 1,08	5 Table	view: Genome				Filter
59 018 2008 Provide the second se	Reverse rea	Forward/rev	Average quality	Coding region change	Amino acid change	Amino acid change in longest transcript	Coding region cha
16 2 19.60 ENST0000341420c_[1257C>T]: ENST0000217246c_[272 ENST00 21.05 21.05 21.05 21.05 ENST0000331420c_[1200C>A]: ENST00000310348c_[272 ENSP00000339912p_[His400Gin]: ENSP00000 ENSP00000339912p_His400Gin] ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72 ENSP00000310348c_[72	59	018	20.03				
2105 80.0 ENST00000341420c_[1200C>A]: ENST0000310348c_[272 ENSP00000339912p_[His400Gin]: ENSP00000339912p_His400Gin] ENSP00000339912p_His400Gin ENSP00000339912p_His400Gin 0 0.00 24.47 ENST00000310348c_[72 ENSP00000339912p_His400Gin ENSP00000339912p_His400Gin ENSP00000339912p_His400Gin 1 0.05 24.58 ENST00000217246c_[728-76A>G]; ENST00000310348c_[7 ENST00 0 0.00 20.91 ENST00000310348c_[7 ENST00 0 0.00 22.88 ENST00000378058c_[281-45A>G]; ENST00000310348c_[7 ENST00 0 0.00 23.74 ENST00000378058c_[281-45A>G]; ENST00000310348c_[1 ENST00 0 0.00 23.75 ENST00000402914c_[356-76A>G]; ENST00000310348c_[1 ENST00	10		19.60	ENST00000341420:c.[1257C>T]; ENST00000217246:c.[272			ENST000034142
International and the second			21.05				
0 0 2471 0 0.00 2243 ENST0000402914c_[23-80T>C]; ENST00000310348c_[72 ENST00 1 0.05 2453 ENST00000217246c_[728-76A>G]; ENST00000310348c_[7 ENST00 0 0.00 24.12 ENST00000217246c_[728-76A>G]; ENST00000310348c_[7 ENST00 0 0.00 29.91 ENST00 ENST00 0 0.00 22.88 ENST00000378058c_[221-45A>G]; ENST00000402914c_[2 ENST00 0 0.00 23.74 ENST00000402914c_[356-76A>G]; ENST00000310348c_[1 ENST00 0 0.00 23.75 ENST00000402914c_[356-76A>G]; ENST00000310348c_[1 ENST00			18.02	ENST00000341420:c.[1200C>A]; ENST00000310348:c.[272	ENSP00000339912:p.[His400Gln]; ENSP00000	ENSP00000339912:p.His400Gln	ENST0000034142
0 0.00 2243 ENST00000402914c[23-80T>C]: ENST00000310348c[72 1 0.05 2453 0 0.00 2412 ENST0000217246c[728-76A>G]: ENST0000310348c[7 0 0.00 2091 0 0.00 2288 ENST00000378058c[281-45A>G]: ENST00000402914c[2 0 0.00 2374 0 0.00 2375 ENST00000402914c[356-76A>G]: ENST00000310348c[1 ENST00 0 0.00 2189	0	0_0	24.7				
1 0.05 2453 0 0.00 24.12 ENST0000217246c.[728-76A>G]: ENST00000310348c.[7 ENST00 0 0.00 20.91 ENST00000217246c.[728-76A>G]: ENST00000402914c.[2 ENST00 0 0.00 22.88 ENST00000378058c.[281-45A>G]: ENST00000402914c.[2 ENST00 0 0.00 23.74 ENST00000310348c.[1 ENST00 0 0.00 23.75 ENST00000310348c.[1 ENST00 0 0.00 21.89 ENST00000310348c.[1 ENST00	0	0.00	22.43	ENST00000402914:c.[23-80T>C]; ENST00000310348:c.[72			ENST00000310348
0 0.00 24.12 ENST0000217246c[728-76A>G]; ENST000031034%c[7 0 0.00 20.91 0 0.00 22.88 ENST000037805&c[281-45A>G]; ENST0000402914c[2 0 0.00 23.74 0 0.00 23.75 ENST0000402914c[358-76A>G]; ENST0000031034%c[1 0 0.00 21.89 ENST00	1	0.05	24.53				
0 0.00 20.91 0 0.00 22.88 ENST000037805&c[281-45A>G]: ENST00000402914c.[2 0 0.00 23.74 0 0.00 23.75 ENST00000402914c.[356-76A>G]: ENST0000031034&c.[1 0 0.00 21.89	0	0.00	24.12	ENST00000217246:c.[728-76A>G]; ENST00000310348:c.[7			ENST0000031034
0 0.00 22.88 ENST0000037805&c [281-45A>G]; ENST00000402914 c [2 ENST00 0 0.00 28.74 0 28.75 ENST00000402914 c [356-76A>G]; ENST0000031034&c [1 ENST00 0 0.00 21.89 ENST00	0	0.00	20.9				
0 0.00 28.74 0 0.00 28.75 ENST00000402914∞[358-76A>G]: ENST00000310348∞[1 0 0.00 21.89	0	0.00	22.88	ENST00000378058:c.[281-45A>G]; ENST00000402914:c.[2			ENST0000031034
0 0.00 23.75 ENST00000402914c[356-76A>G]; ENST0000031034&c[1 0 0.00 21.89	0	0.00	23.74				
0 0.00 21.89	0	0.00	23.75	ENST00000402914:c.[356-76A>G]; ENST00000310348:c.[1			ENST00000310348
	0	0.00	21.89				
0 0.00 26.40	0	0.00	26.40				
0 0.00 21.11	0	0.00	21.1				





Rows: 1,08	5 Table v	view: Genome					Filter
Reverse rea	Forward/rev	Average quality	Coding region change	Amino acid change	Amino acid chan	ge in longest transcript	Coding region change
59	0.48	20.03					
46	0.48	19.60	ENST00000341420:c.[1257C>T]; ENST00000217246:c.[272				ENST00000341420:c.1
32	0.48	21.05					
24	0.47	18.02	ENST00000341420:c.[1200C>A]; ENST00000310348:c.[272	ENSP00000339912:p.[His400Gln]; ENSP00000	ENSP0000033991	2:p.His400Gln	ENST00000341420:c.1
0	0.00	24.71					
0	0.00	22.43	ENST00000402914:c.[23-80T>C]; ENST00000310348:c.[72				ENST00000310348:c.7
1	0.05	24.53					
0	0.00	24.12	ENST00000217246:c.[728-76A>G]; ENST00000310348:c.[7				ENST00000310348:c.7
0	0.00	20.91					
0	0.00	22.88	ENST00000378058:c.[281-45A>G]; ENST00000402914:c.[2				ENST00000310348:c.9
0	0.00	23.74					
0	0.00	23.75	ENST00000402914:c.[356-76A>G]; ENST00000310348:c.[1				ENST00000310348:c.1
0	0.00	21.89					
0	0.00	26.40					
0	0.00	21.11					
	0.00	00.00					

■ アミノ酸置換に関する情報がテーブルに追加されました

Show column 😑
✓ Chromosome
🔽 Region
💟 Туре
📝 Reference
Allele
🔽 Reference allele
🔽 Length
🔲 Linkage
🔽 Zygosity
🔽 Count
Coverage
V Frequency
Probability
Forward read count
Reverse read count
V Forward/reverse balance
V Average quality
Coding region change
🔽 Amino acid change
👿 Amino acid change in longest transcript
🔽 Coding region change in longest trans
Other variants within codon
Von-synonymous
Select All
Deselect All

133

dbSNP比較とアノテーション付け



アノテーション付け・比較

アノテーション

- Track tools > Annotate and Filter
- Resequencing analysis > Annotate and Filter Variants

比較

Resequencing analysis > Compare Variants







dbSNPとの比較

■ dbSNPと比較し、dbSNPに含まれないものを取得する



- Navigation Areaから変異データを選択。
- Toolboxから Resequencing Analysis > Annotate and Filter Variants > Filter against Known Variants を選択、ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。





dbSNPとの比較

Gx Filter against Known	n Variants	x	
1. Select variant tracks 2. Set parameters	Set parameters		
	Track from a variant database Known variants track 🐂 hg19_dbsnp_common_variants_chr20	ିଇ	
	Auto join Join adjacent MNVs and SNVs	• k	「 (nown variants track: dbSNPを選択
anna la	Filter options Keep variants with exact match found in the track of known variants Keep variants with overlap found in the track of known variants Keep variants with no exact match found in the track of known variants 	■ / ¥	Auto join:連続する変異をMNVとして 結合させる
1000 Martin Martin Contraction		■ F	Filter options ■ アレルまで一致したものを残す ■ ポジションのみ一致したもの ■ アレルが一致しないものを残す
? %	← Previous → Next ✓ Finish X Ca	ncel	





dbSNPとの比較



Column wiath 👘
Manual 👻
Show column —
📝 Chromosome
🔽 Region
📝 Туре
🔽 Reference
Allele
📝 Reference allele
🔽 Length
🔲 Linkage
🔽 Zygosity
🔽 Count
Coverage
V Frequency
🔽 Probability
Forward read count
🔽 Reverse read count
Forward/reverse balance
🔽 Average quality
🔽 Overlap
Coding region change
🔽 Amino acid change
📝 Amino acid change in longest transcript
Coding region change in longest trans
Other variants within codon
Von-synonymous
Select All
Deselect All





遺伝子名をつける



- Navigation Areaから変異データを選択。
- Toolboxから Resequencing Analysis > Track Tools > Annotate with Overlap Information を選択、ダブルクリック。
- ウィザードが起動し、選択したデータが選ばれていることを確認。



遺伝子名をつける

Gx Annotate with Overla	p Information
 Select a variant track or an annotation track 	Overlap track
2. Overlap track	
	Select a track for overlap comparison
	Overlap track 🔩 hg 19_chr20 (Gene)
6	
Con and	
(JEP)	
State of the state	
Contractor	
TO TO THE ATENTO LITER	





遺伝子名をつける

🏪 👫 reads (Variants, AAC, KNOWN, OA).

		10,000,000 I	20,000	0,000 30	0.000,000 I	40,000,000 I
(Variants	259 , AAC, KNOWN, OA) Variants (637)					
	0	-				
📰 reads (Var	rian X				1 to 2, a	
Rows: 637	Table view: Genome					Filter
ranscript	Coding region change in longest transcript	Other varian	Non-synony	hg19_chr20 (Gene)	Gene Cards	ENSEMBL
	ENST00000399002:c.289A>C	No	No	SPTLC3	SPTLC3	ENSG00000172296
	ENST00000399002:c.303+84C>T	No	-	SPTLC3	SPTLC3	ENSG00000172296
		No	No	SPTLC3	SPTLC3	ENSG00000172296
	ENST00000399002:c.1467A>G	No	Yes	SPTLC3	SPTLC3	ENSG00000172296
		No	-			
		No	-	10141	10144	ENCONDECTOR STOR
	ENERT2000000000 000 00000	No	NO	ISM1	ISM1	ENSG00000101280
	ENS10000262487c.80C>G	No	Tes	ISM1 TASD1	ISM1 TASE1	ENSG00000101280
	ENET20000002240- 1105017	NO	NO No	TASPI	TASP1	ENSC00000089123
	ENG (00000337743C.1105G21	NO	NO	TASPI	TASPI	ENSC000000000123
	ENST00000227742-0 110045-0	No	NO Vec	TASP1	TASP1	ENSC00000089128
	E146100000007746.C.1100A2G	No	-	Mori	InorT	ENGG0000089128
		No	-			
		No	-			
		No	-			
		No	-			
		No	-			
	ENST00000284951:c.1570+8G>A	No	-	SEL1L2	SEL1L2	ENSG00000101251
	ENST00000284951:c.1570+8G>A ENST00000284951:c.1332T>C	No No	– No	SEL1L2 SEL1L2	SEL1L2 SEL1L2	ENSG00000101251 ENSG00000101251
	ENST00000284951:c.1570+8G>A ENST00000284951:c.1332T>C ENST00000310348:c.728-80T>C	No No No	- No -	SEL1L2 SEL1L2 MACROD2	SEL1L2 SEL1L2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264
	ENST0000284951c.1570+8G>A ENST0000284951c.1332T>C ENST00000310348c.728-80T>C ENST00000310348c.728-76A>G	No No No	- No -	SEL1L2 SEL1L2 MACROD2 MACROD2	SEL1L2 SEL1L2 MACROD2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264 ENSG00000172264
	ENST0000284951c.1570+8G>A ENST0000284951c.13327>C ENST00000810348c.728-80T>C ENST00000810348c.728-76A>G	No No No No	- No - -	SEL1L2 SEL1L2 MACROD2 MACROD2	SEL1L2 SEL1L2 MACROD2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264 ENSG00000172264
	ENST00000284951c.1570+93>A ENST00000284951c.1582T>C ENST00000310348&728-80T>C ENST00000310348c.728-75A>G	No No No No No	- No - - -	SEL1L2 SEL1L2 MACROD2 MACROD2	SEL1L2 SEL1L2 MACROD2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264 ENSG00000172264
	ENST00000284951:c.1570+963-A ENST00000284951:c.1382T>C ENST0000034386-728-80T>C ENST00000310348c-728-76A>G	No No No No No No	- No - - - -	SEL1L2 SEL1L2 MACROD2 MACROD2	SEL1L2 SEL1L2 MACROD2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264 ENSG00000172264
	ENST00000284951c.1570+6Q:A ENST00000284951c.1322T>C ENST00000284951c.2322T>C ENST00000310348c.728-80T>C ENST00000310348c.728-76A>G	No No No No No No No	- No - - - -	SEL 11.2 SEL 11.2 MACRO D2 MACRO D2	SEL1L2 SEL1L2 MACROD2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264 ENSG00000172264
	ENST00000284951c.1570+963-A ENST00000284951c.13827D-C ENST0000018486.728-80TDC ENST00000310348c.728-76A3-G	No No No No No No No No	- No - - - -	SEL 1L2 SEL 1L2 MACRO D2 MACRO D2	SEL1L2 SEL1L2 MACROD2 MACROD2	ENSG00000101251 ENSG00000101251 ENSG00000172264 ENSG00000172264

To show more tracks together, create a ti

XENNIX פכוז ויצמן למדע WEIZMANN INSTITUTE OF SCIENCE a division of Map ards® a <u>commercial license from LifeMap Sciences, In</u> The Human Gene Compandium User Feedback Home GeneCards Guide Suite Terms and Conditions About Us Mirror sites Set Analyses: GeneALaCarl) GeneDecks • Search Advanced Search keyword(s) Serine Palmitoyltransferase, Long Chain Base Subunit 3 SPTLC3 Gene (Previous names: chromosome 20 open reading frame 38, serine protein-coding GIFtS: 53 palmitoyltransferase,...) (Previous symbols: C20orf38, SPTLC2L) GCID: GC20P012938 Antibodies/ cDIU/ RIUAi Proteins & Enzymes Assays & Kits/ Pathways SABiosciences Gene Network GIAGEN A GAGEN Company TFBS PCR Arrays Primers: ChIP / RT² GenScript Genes See related diseases ORIGENE Antibodies Proteins at MalaCards Assays/ Genes/ shRNA/ Prim Antibodies Peptides Jump to Section ... ۲ Aliases Serine Palmitoyltransferase, Long Chain Base Chromosome 20 Open Reading Aliases Subunit 312 Frame 381 for SPTLC3 gene SPTLC2L1235 LCB2B² C20orf381 2 3 LCB3² (According to ¹HGNC, ²Entrez Serine Palmitoyltransferase, Long Chain Base Gene, Subunit 2-Like ³UniProtKB/Swiss-Prot, SPT3² (Aminotransferase 2)¹ ² ⁴<u>UniProtKB/TrEMBL</u>, ⁵<u>OMIM</u>, Long Chain Base Biosynthesis Protein 2b23 dJ718P11² 6GeneLoc, 7Ensembl, 8DME, 9miRBase, 10fRNAdb, 12H-InvDB, Long Chain Base Biosynthesis Protein 323 dJ718P11.12 13NCBI, 14NONCODE, and/or Serine-Palmitoyl-CoA Transferase 32 3 hLCB2b² ¹⁵RNAdb) LCB 3^{2 3} Serine Palmitoyltransferase 3² About This Section SPT 32 3 LCB2b³ EC 2.3.1.5038 EC 2.3.18 External Ids: HGNC: 162531 Entrez Gene: 553042 Ensembl: ENSG000001722967 OMIM: 6111205 UniProtKB: Q9NUV73

vww.genecards.org/cgi-bin/carddisp.pl?gene=SPTLC3#pathways_interactions



アノテーション付け・比較

アノテーション

- Track tools > Annotate and Filter
- Resequencing analysis > Annotate and Filter Variants

比較

Resequencing analysis > Compare Variants







