



## Golden Helix社製品カタログ

遺伝統計解析ソフトウェア

 SNP & VARIATION SUITE

臨床シーケンス解析ソフトウェア

 varSEQ®

次世代シーケンサー2次解析ソフトウェア

 Sentieon

フィルジェン株式会社

# INDEX

<b>Golden Helix社について</b> .....	P. 2
--------------------------------	------



<b>SNP &amp; Variation Suite</b> .....	P. 3
--	------

✓ 遺伝統計学 .....	P. 4
✓ SNP & Variation Suite概要 .....	P. 5
✓ Core Features .....	P. 7
✓ Genotype Analysis .....	P. 8
✓ Agrigenomics Analysis .....	P. 9
✓ CNV Analysis .....	P. 9
✓ DNA-Seq Analysis .....	P. 10
✓ RNA-Seq Analysis .....	P. 10
✓ 有償アドオン .....	P. 11
✓ ケーススタディ .....	P. 12



<b>VarSeq®</b> .....	P. 13
----------------------	-------

✓ 次世代シーケンサーによる遺伝子診断 .....	P. 14
✓ VarSeq®概要 .....	P. 15
✓ VSReports® .....	P. 20
✓ VSPipeline® .....	P. 21
✓ VS-CNV® .....	P. 22
✓ VSWarehouse® .....	P. 23
✓ VSClinical® .....	P. 24
✓ ケーススタディ .....	P. 26



<b>Sentieon Genomics Tools</b> .....	P. 27
--------------------------------------	-------

✓ Sentieon Genomics Tools概要 .....	P. 28
✓ 解析パイプライン .....	P. 29

## Golden Helix社について

Golden Helix社（米国）は、ゲノムデータ解析用のソフトウェアの開発と販売を行っている企業です。とくにマイクロアレイや次世代シーケンサーのような、大量データの処理技術に優れており、世界中に数百の大学・政府研究機関などのクライアントを保有しています。同社ソフトウェアを使用した論文は、すでに1300報以上にのぼり、ゲノム解析のリーダー的企業となっています。

- Founded in 1998
- Genetic association software
- Analytic services
- Hundreds of users worldwide
- Over 800 customer citations in scientific journals



### 【製品ラインナップ】

Golden Helix社では、遺伝統計解析ソフトウェア「SNP & Variation Suite」と、遺伝子診断支援ソフトウェア「VarSeq®」の2種類のソフトウェアパッケージの製品ラインナップがあります。またこれら以外にも、Sentieon社の次世代シーケンサー2次解析パイプラインソフトウェア「Sentieon Genomics Tools」と、無償で使用可能な高機能なゲノムブラウザー「GenomeBrowse®」を提供しており、お客様のゲノム解析を強力にサポートします。



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



- Variant Annotation & Filtering
- ACMG & AMP Guideline Workflow
- Clinical Reporting
- CNV Calling
- High-throughput NGS Testing

### 【サポートフィールド】

Golden Helix社の製品は、医学・生物学研究はもちろん、家畜や作物の品種改良などの農学研究にも使用可能です。また研究以外にも、疾患の診断や最適な治療オプションの決定における医療分野、遺伝学的解析などの教育現場などでも活躍します。



Agrigenomic  
Research



Genetics in the  
Classroom



Human Genetic  
Research



Pharma R&D



Testing  
Laboratories



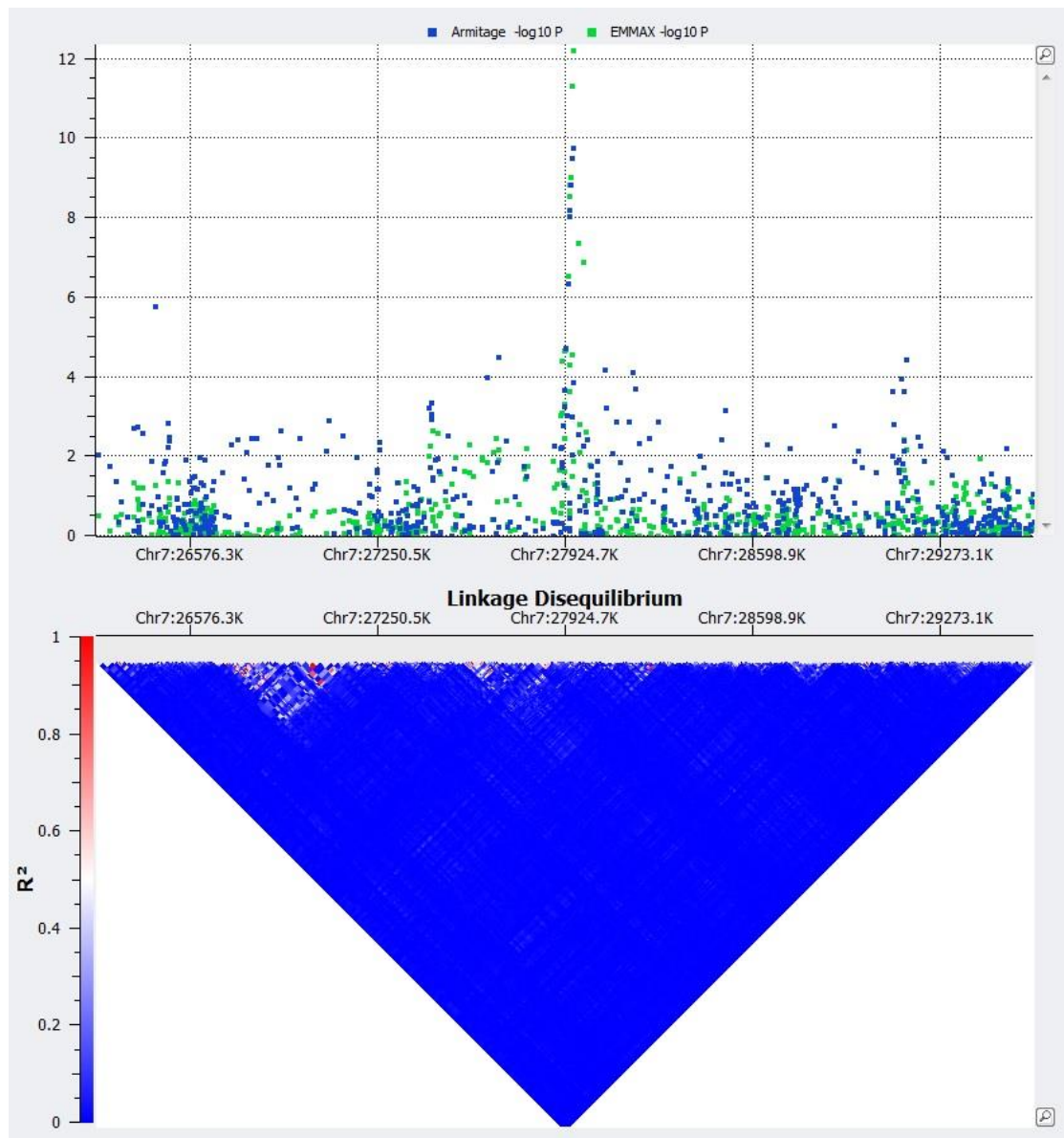
# SNP & VARIATION SUITE

A powerful analytic tool created specifically to empower biologists and other researchers to easily perform complex analyses and visualizations on genomic and phenotypic data.

## 遺伝統計学

マイクロアレイや次世代シーケンサーを利用することで、多数のサンプルについて膨大な量のゲノムデータを得られるようになりましたが、この大量のデータの中から有用な情報を抽出するためには、統計学を駆使したデータ解析手法を利用する必要があります。

そこで注目されているのが遺伝統計学です。遺伝統計学は、SNPなどの遺伝子多型情報をはじめとした遺伝情報と、疾患の有無や身体的特徴などの表現型情報のかかわりを調べる学問で、ハイスループットに得られたゲノムデータに対して、統計学および情報学的なアプローチを行うことによって、ゲノムデータのより明確な意味付けを行います。



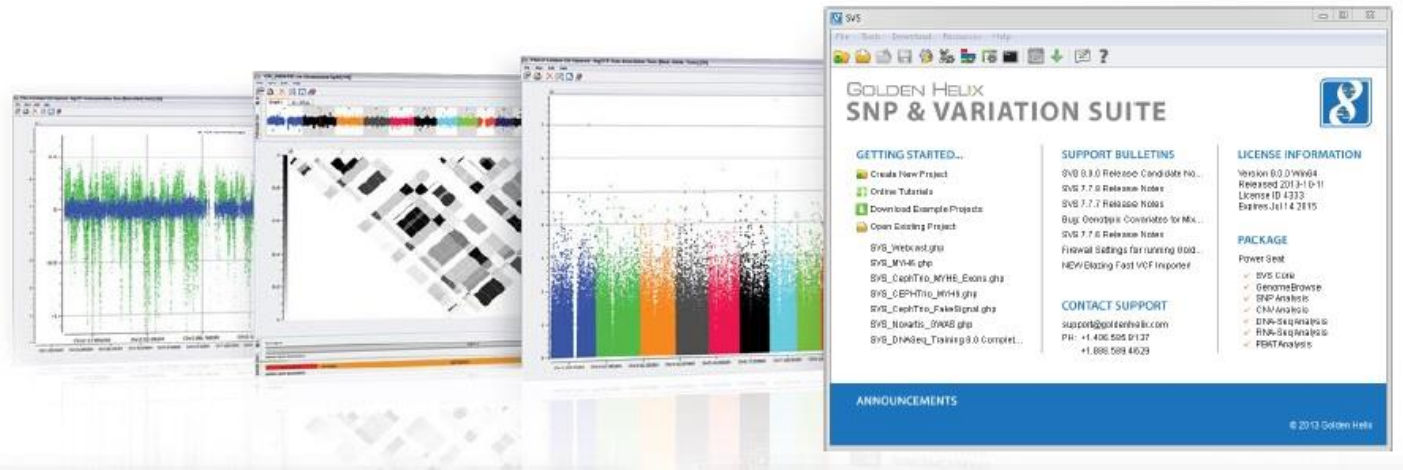
例えば、遺伝統計学で代表的なゲノムワイド関連解析（GWAS）では、マイクロアレイなどを使用してゲノム全域にわたってタイピングを行ったSNPデータを用いて、アレルの疾患発症リスクなどを統計的手法を用いて検定を行います。また農学分野では、家畜や農作物の育種において、同じくSNPデータを用いた遺伝能力の評価などに応用されています。その他、CNV（コピー数多型）データを用いたゲノム上の構造異常領域の検出や、次世代シーケンサーの登場により取得が可能となった、集団中の稀少変異（レアバリエント）や網羅的な遺伝子発現量データを用いた、表現型との関連の評価など、様々な研究に応用されています。

そしてこれらを行うにあたっては、大量のゲノムデータを効率的に処理するためのコンピューティングシステム、統計学的な解析アルゴリズム、さらに解析結果を視覚的に表示を行うグラフ機能などを備えた、非常に高機能なデータ処理用のソフトウェアが必要となります。

# SNP & Variation Suite概要

Golden Helix社SNP & Variation Suite (SVS)は、次世代シーケンサーやSNP/CNVマイクロアレイ解析によって出力された膨大な遺伝子変異データや表現型データを管理・分析・可視化することを目的に開発された、遺伝統計解析用の統合解析ソフトウェアです。

ゲノムワイド関連解析 (GWAS) などの遺伝統計学に必要な統計解析機能を多数搭載し、また多種多様なデータベースの管理ツールも内蔵しています。そのため、これらを統合的に利用したデータ解析が可能で、これにより膨大なデータから生物学的に重要なデータを効率よく見つけることができます。さらに、グラフ作成機能や高機能なゲノムブラウザ「GenomeBrowse®」を内蔵しており、データの全体像や把握や多角的な解釈が可能です。



## Core Features

- Powerful Data Management
- Rich Visualizations (GenomeBrowse)
- Robust Static
- Flexible

## Applications

- Genotype Analysis
- Agrigenomics Analysis
- DNA/RNA Sequence Analysis
- CNV Analysis

### 【有償アドオン】

SVSでは、別途有償のアドオンを追加することで、新たなアノテーションリソースを使用できるようになったり、またはソフトウェア本体に、新たな機能を追加することができます。

#### 機能モジュール

**PBAT** - 家系データに基づく関連解析

**SVS-CNV** - NGSデータからのCNV検出

**SVS-CVS** - バリアントの機能予測

#### アノテーションリソース

**OMIM** - 遺伝子・疾患情報データベース

**CADD** - バリアントの有害スコアデータベース

### 【ライセンス形態】

SVSでは、使用するユーザー数やコンピュータの種類にあわせて、様々なライセンスを用意しています。

シングルユーザーライセンス	1ユーザーが1台のコンピュータで使用する場合のライセンスです。
サーバーライセンス	ソフトウェアを1台のサーバーにインストールし、複数のユーザーで使用できます。2、3人であれば、複数のユーザーが同時に使用することもできますが、大きな負荷がかかる処理を行う場合は、1人のみの使用が推奨されます。また、インビュテーション機能が使用可能です。
ラボライセンス	1台のコンピュータを複数のユーザーで使用するライセンスです。

## 【生物種の対応】

SVSはヒト、マウスなどの哺乳動物はもちろん、植物や魚類、さらに昆虫や細菌など様々な生物種のデータ解析に対応しているため、医学研究だけでなく農学研究などにも活用できます。

### 哺乳動物

- Bos taurus (ウシ)
- Canis familiaris (イヌ)
- Capra hircus (ヤギ)
- Cricetulus griseus (チャイニーズハムスター)
- Equus caballus (ウマ)
- Felis catus (ネコ)
- Gallus gallus (ニワトリ)
- Heterocephalus glaber (ハダカデバネズミ)
- Homo sapiens (ヒト)
- Macaca mulatta (アカゲザル)
- Meleagris gallopavo (シチメンチョウ)
- Mus musculus (マウス)
- Nomascus leucogenys (ホロジロテナガザル)
- Oryctolagus cuniculus (アナウサギ)
- Ovis aries (ヒツジ)
- Pan troglodytes (チンパンジー)
- Rattus norvegicus (ラット)
- Sus scrofa (ブタ)
- Vicugna pacos (アルパカ)

### その他

- Apis mellifera (セイウミツバチ)
- Anopheles gambiae (ハマダラカ)
- Caenorhabditis elegans (線虫)
- Drosophila melanogaster (ショウジョウバエ)
- E.coli (大腸菌)
- Leishmania infantum JPCM5 (リーシュマニア寄生虫)
- Mycobacterium tuberculosis H37Rv (結核菌)
- Plasmodium falciparum 3D7 (マalaria)
- Saccharomyces cerevisiae (出芽酵母)
- Schizosaccharomyces pombe (分裂酵母)
- Staphylococcus aureus (黄色ブドウ球菌)

### 植物

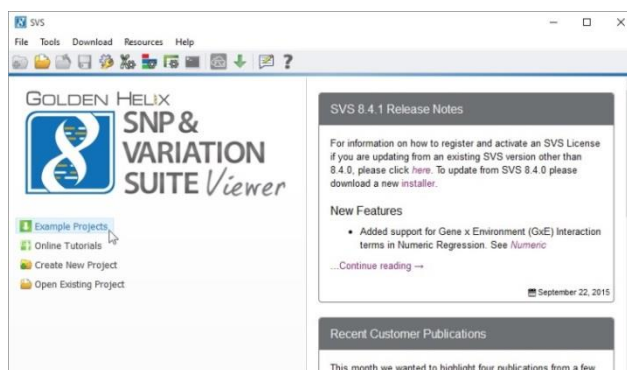
- Arabidopsis thaliana (シロイヌナズナ)
- Brassica rapa (アブラナ)
- Capsicum annuum (トウガラシ)
- Carica papaya (パパイヤ)
- Cicer arietinum (ヒヨコマメ)
- Citrullus lanatus (スイカ)
- Cucumis melo (メロン)
- Cynara scolymus (アーティチョーク)
- Eucalyptus grandis (ユーカリ)
- Glycine max (ダイズ)
- Gossypium raimondii (ワタ)
- Oryza sativa (イネ)
- Prunus persica (モモ)
- Setaria italica (アワ)
- Solanum lycopersicum (トマト)
- Solanum tuberosum (ジャガイモ)
- Sorghum bicolor (モロコシ)
- Theobroma cacao (カカオ)
- Zea mays (トウモロコシ)

### 魚類

- Ictalurus punctatus (アメリカナマズ)
- Danio rerio (ゼブラフィッシュ)
- Medaka (メダカ)
- Oncorhynchus mykiss (ニジマス)
- Oreochromis niloticus (ナイルティラピア)
- Salmo salar (タイセイヨウサケ)

## 【専用ビューワー】

Golden Helix社では、SVSで解析したデータを閲覧するためのビューワーソフトウェアを無償で提供しております。SVSの解析結果を、複数の研究者間で共有を行ったり、ゲノムブラウザーなどのビジュアライズ機能の利用が可能です。



## 【システム要件】

### OS

- Windows 7 or later (64-bit recommended), Linux Ubuntu 14.04 or later (64-bit only), Linux RHEL 7 or later, or equivalently CentOS 7 or later (64-bit only), Mac OS X 10.11 or later

### Standard

- 4GB of RAM
- Multicore CPU
- 100GB of space available for annotations and projects

### Advanced

- 16GB+ of RAM (32GB for Servers)
- 8+ CPU Cores
- 1TB of space available for annotations and projects

# Core Features

## 【データ管理】

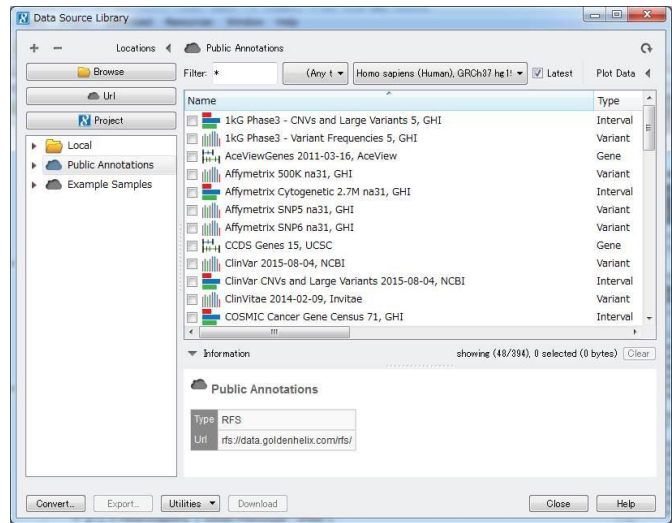
優れたデータ処理能力により、数千サンプルのマイクロアレイや全エクソームデータも、シームレスに解析を行うことが可能です。また、専用のデータベース管理ツールを用いて、各種アノテーションデータを容易にダウンロードできます。さらにプロジェクトデータをパスワードで保護することで、セキュリティを確保することが可能です。

Chromosome	Position	SNP_A-1909444	SNP_A-4303947	SNP_A-1886933	SNP_A-2283559	SNP_A-2205441	SNP_A-2161090	SNP_A-4261020	SNP_A-1861458
1	752566	1	1	1	1	1	1	1	1
rs3094315	rs4048617	rs2880300	rs2005036	rs4245756	rs4075116	rs442385	rs1890775	rs1130727	rs1130727
Associated Gene	?	?	?	?	?	?	?	?	TTLL10
Cytoband	p36.33	p36.33	p36.33	p36.33	p36.33	p36.33	p36.33	p36.33	p36.33
Reference Alleles A/B	[C/T]	[A/G]	[A/G]	[C/T]	[C/T]	[A/G]	[A/G]	[G/T]	[A/C]
Top Alleles	[G/A]	[A/G]	[T/C]	[C/T]	[C/T]	[A/G]	[A/G]	[G/T]	[A/C]
Bottom Alleles	[C/T]	[T/C]	[A/G]	[G/A]	[G/A]	[A/G]	[A/G]	[C/A]	[T/G]
Strand	-	+	-	+	+	-	+	+	+
Strand Versus dbSNP	same	same	reverse	reverse	same	same	same	same	same

## 【サポートするファイルフォーマット】

Thermo Fisher Scientific社 (Affymetrix)、illumina社のSNP/CNVマイクロアレイデータファイル、および次世代シーケンサーの解析パイプラインより取得した変異データファイルなどをインポート可能です。

- Text files
- Excel XLS and XLSX
- Affymetrix CEL, CHP, CNT, CNCHP, TXT files
- illumina Final Report Text files, Matrix Text files, iControlDB Data
- Plink PED, TPED and BED files with supporting files
- Agilent files
- NimbleGen Data Summary Files
- VCF files version 4.0+
- Complete Genomics Var Files
- Impute2 GWAS files
- HapMap format
- MACH output
- RNA-Seq Tabularized Quantification
- その他50以上のファイルフォーマットをサポート



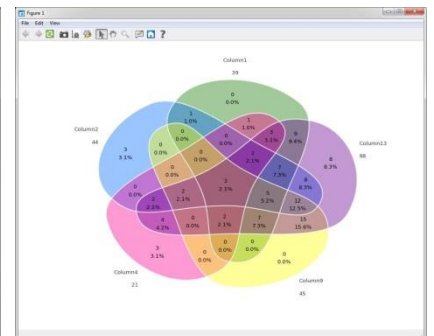
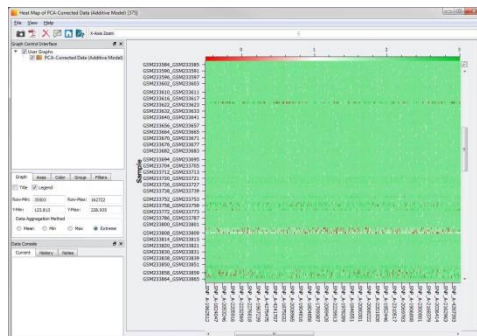
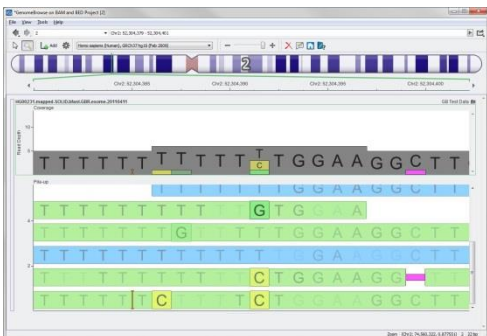
## 【Pythonスクリプトによる機能拡張】

ユーザー独自作成、または Golden Helix社から無償で提供されているPythonスクリプトを組み込み、ソフトウェア機能の拡張を行うことが可能です。

## 【ビジュアライゼーション】

Golden Helix社が開発した高機能なゲノムブラウザー「GenomeBrowse®」を内蔵し、さらに解析データの視覚化を行うための様々なグラフ表示機能が搭載されており、多角的な視点からのデータの解釈をサポートします。

- ヒストグラム
- XYスキッタープロット
- パイチャート
- 連鎖不平衡 (LD) プロット
- ヒートマップ
- ボックスプロット
- リードアライメントデータ (BAMファイル) プロット
- 数値データプロット
- ヘン図 など





# Genotype Analysis

## 【クオリティーチェック】

様々な手法を用いて、遺伝子型データのクオリティーチェックを行います。

- Call rate, HWE, MAPに基づくSNPフィルタリング
- 連鎖不平衡を示すSNPの除去
- Cryptic relatednessの検証
- メンデルエラーの検証
- 常染色体のヘテロ接合性
- Allele intensitiesの目視検査
- 集団の階層化 (Population stratification)
- SNP concordance
- 性別誤認 (Gender misidentification) の検出
- Identity by Descent (IBD) の計算
- 多次元解析による異常値検出
- 染色体異常スクリーニング など

## 【ゲノムワイド関連解析 (GWAS)】

遺伝子型と表現型の関連解析に必要な、遺伝モデルや統計アルゴリズムを多数搭載しています。表現型データには、ケースコントロールの様なバイナリデータ、または身長・体重の様な量的データが使用可能です。また、主成分分析 (PCA) による集団の階層化の補正、Bonferroni法やFDRを用いた多重検定の補正にも対応しています。

- Correlation/Trend Test
- Chi-Squared Test
- Odds Ratio
- Fisher's Exact Test
- Armitage Trend Test
- Analysis of Deviance
- linear or logistic regression

## 【連鎖不平衡 (LD) 解析】

連鎖不平衡 (LD) をゲノムレベル、染色体レベル、または遺伝子特異的に探索することができます。また、LDプロットを他のプロット図に加えて相関を確認したり、複数のLDプロットをゲノムブラウザを使って、同時に確認することができます。

## 【ハプロタイプ解析】

ハプロタイプ解析では、LDプロットやスプレッドシートからハプロタイプブロックの検出や調整、ハプロタイプおよびディプロタイプ頻度の計算、そしてハプロタイプ関連解析やトレンド回帰解析の実行が可能です。

## 【回帰解析 (Regression Analysis)】

線形およびロジスティック回帰解析、ステップワイズ回帰解析 (変数増加法および変数減少法)、遺伝子 - 環境相互作用回帰解析、および並べ替え検定を実行することも可能です。単一の従属変数に対し、数量またはカテゴリカルな共変数とともに、moving windowを使用できます。回帰は全ての変数および共変数と一緒に実施したり (フルモデル)、または減少法でグループ化された変数と共に実施することが可能です。

## 【インピュテーション】

BEAGLE4.0, 4.1のImputationアルゴリズムを実装し、遺伝子型データの密度を上げたり、プラットフォームが異なるSNPマイクロアレイのデータを同時に解析することが可能です。リファレンスパネルとして、1000 Genome Project の公開データを容易に利用できる様になっており、またカスタムリファレンスパネルデータを作成することも可能です。

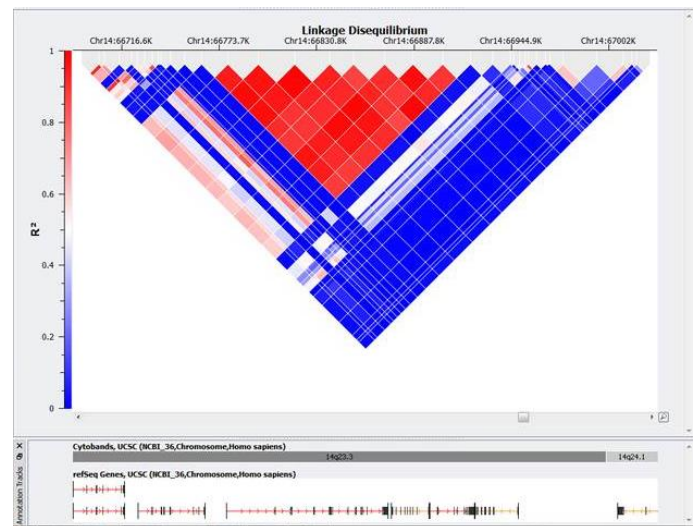
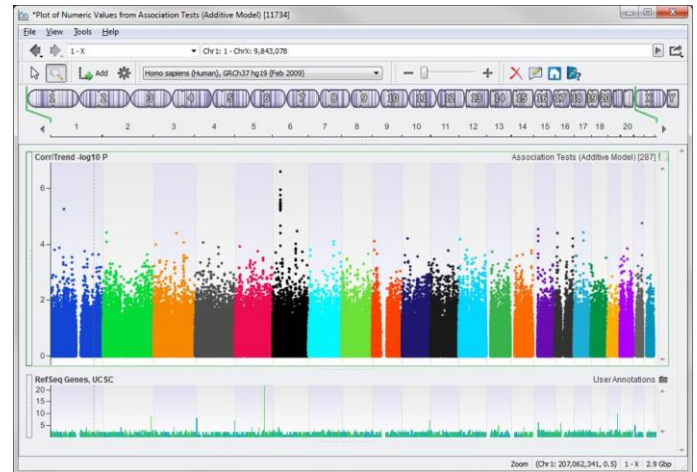
**\* インピュテーション機能を使用するには、サーバーライセンスが必要です。**

## 【メタアナリシス】

ある薬剤に関して効果があるとする報告がある一方、無いとする報告も存在するなど、同一の事象を対象としているにもかかわらず、研究結果が一致しないことがしばしばあります。メタアナリシスは、こうした複数の研究を統合し、1つの大きな知見を導きます。

## 【LD Score Regression】

GWAS結果データとLDスコアを統合し、遺伝率や遺伝共分散、遺伝相関の計算を行います。



# Agrigenomics Analysis

## 【線形混合モデル解析 (Mixed Linear Model Analysis)】

線形混合モデル解析は、Cryptic relatednessやPedigree structureを補正しつつ、遺伝子型データに対して回帰分析を実行するだけでなく、遺伝子型データに基づく回帰分析をより簡単に行うための新しい能力を提供します。

- Mixed Model GWAS using a single locus (EMMAX)
- Multi-locus mixed model GWAS (MLMM)

## 【Genomic Prediction】

データセットから、遺伝子型と表現型の関連を最も正しく説明する予測モデルを構築し、ゲノム育種価の推定 (gEBV) や、表現型未知サンプルの遺伝子型データから、表現型の推定を行うことが可能です。従来の育種価計算と比較し、遺伝子型データさえあれば予測が可能のため、表現型データを得るのに時間がかかる場合などにおける早期の個体の選抜が可能になり、コスト削減とスピードアップに効果的です。

- Genomic Best Linear Unbiased Predictors (GBLUP)
- Bayes C and C-pi

## 【K-Fold Cross Validation】

交差検証法 (Cross Validation) を用いて、各予測手法 (GBLUP, Bayes C, Bayes C-pi) のパフォーマンスの評価を行い、同時に表現型予測に使用するモデルの計算を行います。



# CNV Analysis

## 【コピー数異常 (CNV) の検出】

各社CNVマイクロアレイデータのシグナル値データを基に、ダイナミックプログラミングを利用したセグメント化を行うアルゴリズムによって、サンプルごとのCNV領域を決定します。

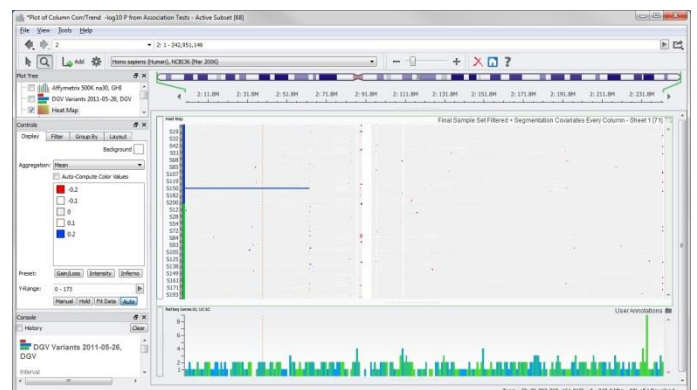
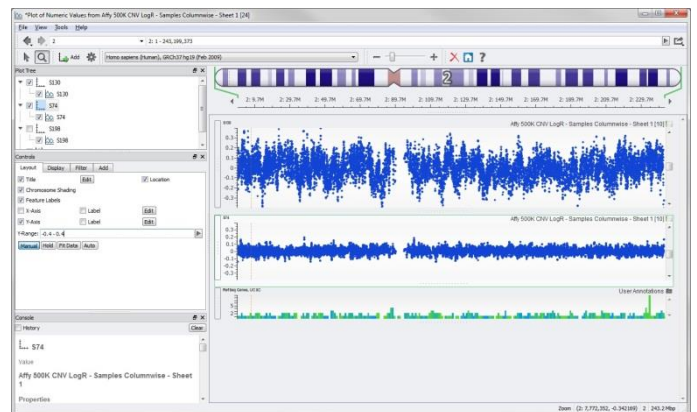
## 【CNVデータ補正・クオリティコントロール】

マイクロアレイとaCGHの結果は、バッチ効果 (使用機器・実験時の環境など)、Genomic wave、集団の階層化など、多くの要因によって、様々なバイアスを受けます。SVSは、これらクオリティの問題を検出・補正する様々な機能を搭載しています。

- Derivative log ratio spread (DLRS)
- 極値分布
- 性別誤認 (Gender misidentification)
- SNPコール比
- Principal component analysis(PCA)
- Genomic waveの検出と補正 など

## 【CNV関連解析】

生データまたはPCA補正log ratioデータ、CNV区間平均値、loss・neutral・gainの3または2状態に基づいた離散数を使用して関連解析を行うことができます。また、CNVデータのみ、または他の遺伝マーカーと表現型データと組み合わせたデータに対しても、関連解析や線形・ロジスティック回帰解析を実行できます。



# DNA-Seq Analysis

## 【アノテーションリソース】

全ゲノムや全エクソーム、ターゲットリシーケンスなど、次世代シーケンサーを用いたデータ解析パイプラインから出力された遺伝子変異データ（VCFファイル）をインポートし、各種アノテーションリソースによるアノテーション付けを行うことが可能です。解析に必要なアノテーションリソースは、専用のデータベース管理ツールより、容易にダウンロードできます。

- dbSNP
- RefSeq Genes
- 1000 Genomes
- NHLBI 6500 Exomes
- ExAC Variant
- gnomAD Exomes/Genomes
- dbNSFP Functional Predictions
- ClinVar
- CIVic
- ICGC Simple Somatic Mutation
- 各種遺伝子パネルのターゲットデータ
- カスタムアノテーションデータ ...など

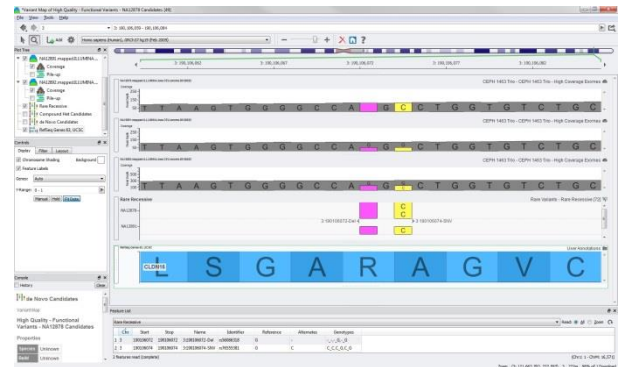
## 【フィルタリングと分類】

VCFファイルに含まれる付加情報（変異のリード深度、クオリティーなど）、または各種アノテーションリソースの情報を利用して、ノイズやCOMMON SNPの除去および疾患関連変異の絞り込みなどを行うことが可能です。

## 【家族性サンプルのデータ解析】

サンプルの家系情報やアノテーションデータのアレル頻度情報に基づき、任意の遺伝様式を示した変異の抽出を行うことが可能です。

- de Novo Mutations
- Compound Heterozygous Polymorphisms
- Rare Recessive Homozygous Polymorphisms



## 【レアバリエント関連解析】

低頻度変異（レアバリエント）の関連解析を行うために、複数のレアバリエントを遺伝子ごとにまとめて解析を行う「Collapse Methods」が搭載されています。その際に、集団階層化やサンプル間の関連性の補正を行う、線形混合モデルも同時に利用可能です。

- Count Variants per Gene
- CMC with Hotelling T Squared Test
- CMC with Regression
- KBAC with Regression
- Mixed-Model KBAC
- SKAT-O

# RNA-Seq Analysis

## 【DESeq Analysis】

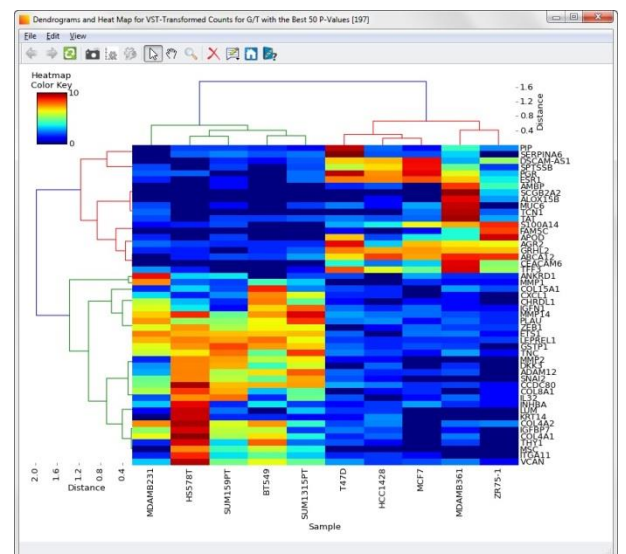
次世代シーケンサーによるRNA-Seq遺伝子発現解析パイプラインより得られた、遺伝子ごとのリードカウントデータを用いて、遺伝子発現量解析を行います。Anders and Huber (2010)により開発されたDESeqアルゴリズムを搭載し、リードカウントデータの平均分散の依存関係の推定や、2項分布モデルを利用して、発現変動遺伝子を検出することができます。その他にもグループ内の平均値や平方変動係数の計算なども行うことができ、計算結果のP-valueやFold Changeの値でフィルタリングを行うことで、統計的に信頼性のある発現変動遺伝子を簡単にみつけることができます。

## 【NormalizationとLog Transformation】

データの正規化（Normalization）を行うことによって、サンプル内、またはサンプル間のデータの分散を揃え、データ間の比較を行える様になります。これらのNormalizationやTransformationを行ったデータを使って、T-Testなどの各種統計解析を行うことができます。

## 【グラフ作成機能】

各データから様々なグラフを作成することができ、解析結果を様々な視点から確認することができます。ボルケーノプロットや階層型クラスタリングの解析結果のヒートマップ、主成分分析（PCA）プロット、ボックスプロットやスキャッタープロットなど、様々なグラフ作成機能を搭載しています。



# 有償アドオン

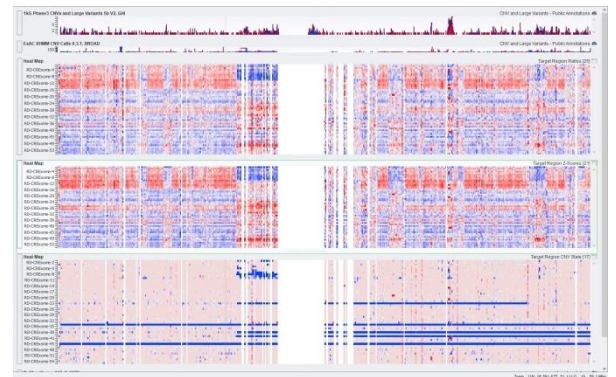
## 【PBAT】

PBATアドオンでは、Family-based、及びPopulation-basedの研究デザインにおいて、検出力の計算（Power calculation）を行います。またサンプルの家系情報に基づいて、Transmission Disequilibrium Test (TDT) 法によるSNPデータの関連解析、及びCNVデータの関連解析を行うことができます。

Map	Catalog ID	Family ID	Patient ID	Father ID	Mother ID	Sex	Affection Status
1	NA06984	1328	NA06984	?	?	0	0
2	NA06989	1328	NA06989	?	?	1	1
3	NA12344	13281	NA12344	NA12347	NA12348	0	0
4	NA12347	13281	NA12347	?	?	0	1
5	NA12348	13281	NA12348	?	?	1	0
6	NA06986	13291	NA06986	?	?	0	0
7	NA06995	13291	NA06995	NA07435	NA07037	0	0
8	NA06997	13291	NA06997	NA06986	NA07045	1	1
9	NA07037	13291	NA07037	?	?	1	1
10	NA07045	13291	NA07045	?	?	1	1
11	NA07435	13291	NA07435	?	?	0	0
12	NA07014	13292	NA07014	NA07051	NA07031	1	1
13	NA07031	13292	NA07031	?	?	1	1

## 【SVS-CNV】

SVS-CNVアドオンでは、遺伝子パネルや全エクソーム、さらに全ゲノムなどの次世代シーケンサーを用いたシーケンスデータから、コピー数異常（CNV）の検出を行います。検出したCNVデータは、SVSに標準搭載の各種統計解析機能を使用し、サンプルの表現型データとの関連解析などに使用することが可能です。



## 【SVS-CVS】

SVS-CVSアドオンでは、次世代シーケンサーで検出されたヒトのバリエーションに対する機能予測・スプライス部位予測を実行します。予測には以下のアルゴリズムが用いられます。

- Splice Site Predictions
- Multiple Sequence Alignment SIFT/Polyphen2
- Conservation Scores: GERP++/PhyloP

**Data Source Library**

Locations: VSclinical

Filter: (Any) Homo sapiens (Current) Plot Data

Name	Size
<input checked="" type="checkbox"/> Conservation Scores Exonic, GHI	1.4G
<input type="checkbox"/> Multiple Sequence Alignments of 100 Vertebrates, UCSC	22G
<input type="checkbox"/> SIFT and PolyPhen2 Missense Predictions, GHI	1.1G

Information: showing (3/6), 1 selected (1.4 GB)

**Conservation Scores Exonic, GHI**

Description: For the +/- 15bp region around exons, the Golden Helix GERP++ / phyloP algorithm was run on different species subsets of the Multiple Sequence Alignments (MSA) of 100 vertebrates species against the human genome.

Buttons: Convert..., Export..., Utilities, Download, Close, Help

## ケーススタディ



### **Dr. Fielding Hejtmancik**

**Senior Investigator, National Eye Institute - NIH**

Dr. James Fielding Hejtmancikと彼のチームは、遺伝性の眼疾患における遺伝メカニズムを理解するために、連鎖解析と関連解析を行っています。彼のチームは、網膜変性や近視および角膜ジストロフィーの特徴に注目し、メンデル性疾患と複合疾患だけでなく、機能的な仕組みも研究しています。SVSを使用することで、現在どの手法を用いて解析を行っているのか透明性を保ちつつ、ユーザーフレンドリーなインターフェースで解析を行うことが可能になりました。



### **Dr. Matt Mclure**

**Lead Geneticist, Irish Cattle Breeding Federation**

Dr. Matthew McClureの所属するIrish Cattle Breeding Federation (ICBF)では、牛肉産業における畜牛の遺伝的改良に取り組んでいます。次世代シーケンサーによるDNA-Seq解析を行うにあたり、SVSを使用することで、シーケンスデータを直接閲覧し、バリエーションの存在やアライメントの確認をしつつ、SNPデータを用いた遺伝統計解析を実行することが可能になりました。



### **Dr. Hela Azaiez**

**Associate Research Scientist in the Molecular Otolaryngology  
and Renal Research Laboratories, University of Iowa**

Hela Azaiezの所属するラボでは、遺伝性の難聴と腎疾患について、その分子遺伝学的なメカニズムと治療方法の研究を行っています。SVSではクリック1つで直ちに解析を実行することができるため、研究者が立てた様々な仮説の検証や、新しいアイデアを試す場合でも、多くの処理を迅速に実行することができるようになりました。



varSEQ®

Simple, fast and repeatable variant analysis software for gene panels,  
exomes and whole genomes.

# 次世代シーケンサーによる遺伝子診断

次世代シーケンサーの普及により、個人ごとの様々な病気と関連する遺伝子変異（バリエント）データが容易に取得できるようになり、遺伝子診断への応用が可能となりました。

しかしながら次世代シーケンサーを用いたゲノム解析では、シーケンスによって得られる大量の遺伝子配列データを処理する必要があり、データ解析のプロセスが複雑かつ長い時間を要します。またバリエントデータが得られた後も、膨大なデータの中から臨床的意義のあるバリエントの抽出や、それらデータから導き出された臨床的な解釈、そして最終的にそれらの情報をまとめたレポートの作成などの作業を実行する必要があります。



Sample Prep



Sequencing



Align & Call



Annotate & Filter



Variant Interpretation



Report

一般的にこれらのプロセスの実行には、標準的なデータ解析を行うためのバイオインフォマティクススキルはもちろんのこと、医学・生物学情報の高精度なデータベースも必要とされ、さらに大量検体を扱うための解析作業の効率化や、解析結果をできる限り分かりやすく視覚表示することにも気を配る必要があります。

また究極の個人情報であるゲノム情報を扱うためには、データの外部への流出を防ぐために、サンプルデータの外部サーバーへのアップロードを行わないセキュアなシステムであること、また逆に、医療施設や研究施設内、あるいは外部の共同研究者とのデータの共有を行うデータシェアリングにも対応したシステムである必要があります。

### 1 Filtering & Annotations CNVs, SNVs, Indels

- Gene Panels**
- Exomes**
- Genomes**

**Quality Metrics**

- DP, GQ, VAF

**Annotations**

- Frequency catalogs
- Gene tracks
- In-silico predictions
- Clinical submission databases

**Algorithms**

- Targeted phenotypes
- OMIM/HPO
- PhoRank
- Match gene list
- Count alleles
- Copy number variants

### 2 VSclinical: ACMG/AMP Autoclassification

**Germline**

ACMG

Probability of Pathogenic given scored Criteria:

Status: Probability of 88.7% predicting classification of Pathogenic

**Somatic**

AMP

Oncogenicity classification given Scored Criteria: Oncogenic (+10)

Scored Criteria: SC+3 CE+3 IP+1 AR+1 HR+1 NP+1

Oncogenicity Scale and Source:

**Process via Automated ACMG/AMP Guidelines  
Create interpretation & classification**

### 3 Clinical Assessment and Reporting

Generate ACMG/AMP clinical report

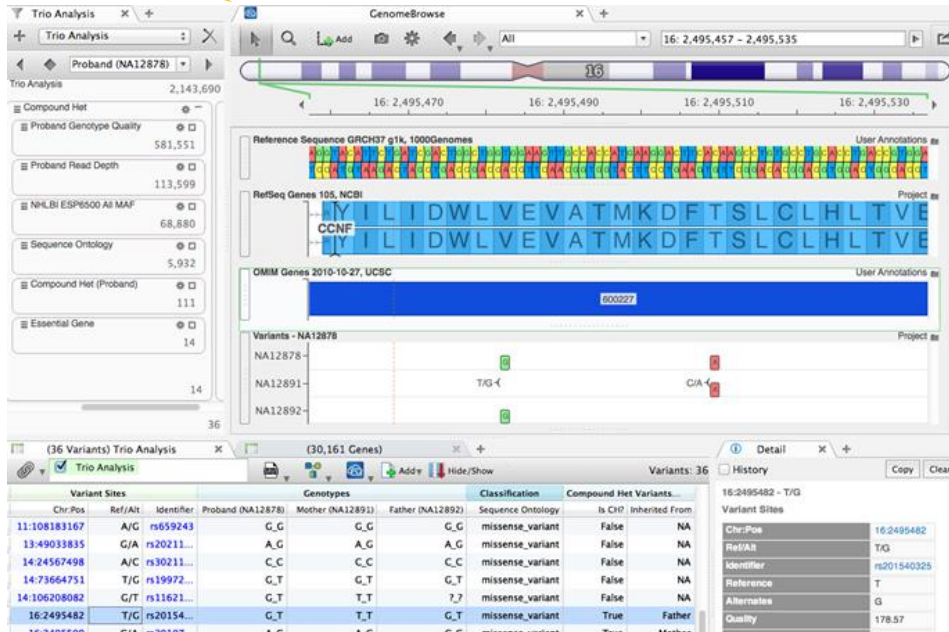
Store interpretation in assessment catalogs

## VarSeq®概要

Golden Helix社VarSeq®は、クリニカルシーケンスによる精密医療（Precision Medicine）を実現するためのソフトウェアパッケージです。次世代シーケンサーによる全ゲノムや全エクソーム、遺伝子パネル解析などによって検出さ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によるアノテーション付けや、それらアノテーションリソースに基づいたフィルタリング、さらに高度な計算アルゴリズムを利用した様々なデータ解析、ゲノムブラウザーを用いたシーケンスデータの可視化などを行います。

Annotation

Filtering



Data Analysis

Genome Browser

### 【有償アドオン】

VarSeq®では、別途有償のアドオンを追加することで、新たなアノテーションリソースを使用できるようになったり、またはソフトウェア本体に、新たな機能を追加することができます。これらを利用することで、より質の高い情報が得られたり、遺伝子診断やゲノム研究用の強力な機能や解析アルゴリズムが使用できるようになります。

**varSEQ®**  
- アノテーション・フィルタリング

**OMIM**  
- 遺伝子・疾患情報データベース

**CADD**  
- バリエーションの有害スコアデータベース

**VS Reports®**  
- 臨床レポートの作成

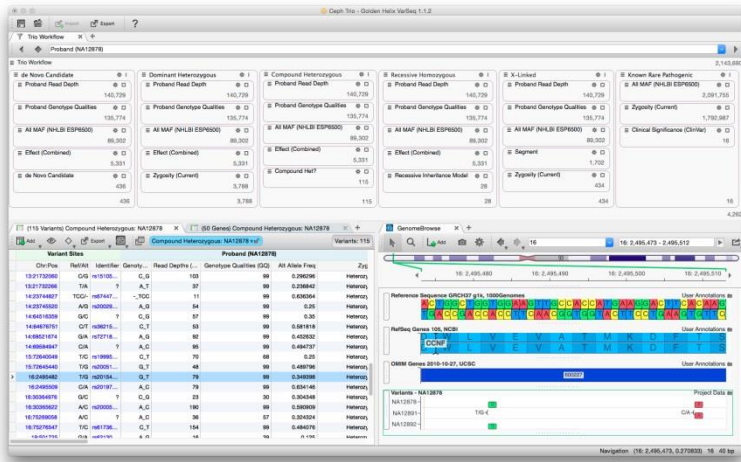
**VS Pipeline®**  
- コマンドライン型インターフェース

**VS-CNV®**  
- CNV解析

**VS Warehouse®**  
- データシェアリング用Webサーバー

**VS Clinical®**  
- ACMG/AMPガイドラインによるバリエーションの評価





### 【基本操作】

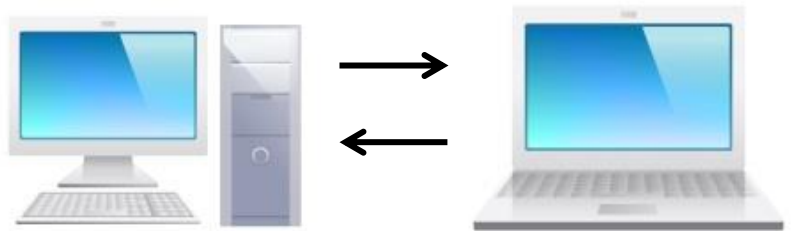
VarSeq®は、遺伝子検査ラボなどでシーケンスからバリエーション（SNV, InDels, CNV, SV）の検出までを行ったデータ（VCFファイル）を使用して解析を実行します。VCFファイルに含まれるバリエーションデータに対して様々なアノテーションの付加を行い、これらアノテーションなどに基づき病的バリエーションのフィルタリングを行います。

すべての処理をローカル環境で実行するため、セキュリティ対策に悩まされることもなく、また直感的な操作が可能なユーザーフレンドリーなインターフェースを搭載しているため、使用方法の習得も容易です。

### 【インストール可能なコンピュータ】

VarSeq®は、1ユーザーにつき2台までのコンピュータにインストールして使用することができます。データ解析用コンピュータとラップトップコンピュータというように、用途に応じてコンピュータを使い分けることが可能で、またOSも別ものを選択することができます。

\*同時に起動できるのは、1ライセンスにつき1台のみです。



### 【専用ビューワー】

Golden Helix社では、VarSeq®で解析したデータを開覧するためのビューワーソフトウェアを無償で提供しております。VarSeq®の解析結果を、複数の研究者間で共有を行ったり、ゲノムブラウザーなどのビジュアライズ機能の利用が可能です。

- VIEW FILTERED  
View tables  
Investigate results
- VISUALIZE  
GenomeBrowse  
BAM and VCF  
Save Screenshots
- EXPORT  
Excel Rich Reports  
Text and VCF
- SHARE  
Relocatable Projects  
Efficient Storage  
Anybody can Open



### 【システム要件】

#### OS

- 64-bit Windows 10, 11, 2019 Server, 2022 Server
- Linux Ubuntu 20.04, or 22.04 (64-bit only)
- Linux RHEL 7, or equivalently CentOS 7 (64-bit only)

- Rocky Linux 8, successor to CentOS 8 (64-bit only)
- Mac OS X 10.13 or later

#### Small Workflow Hardware Recommendations

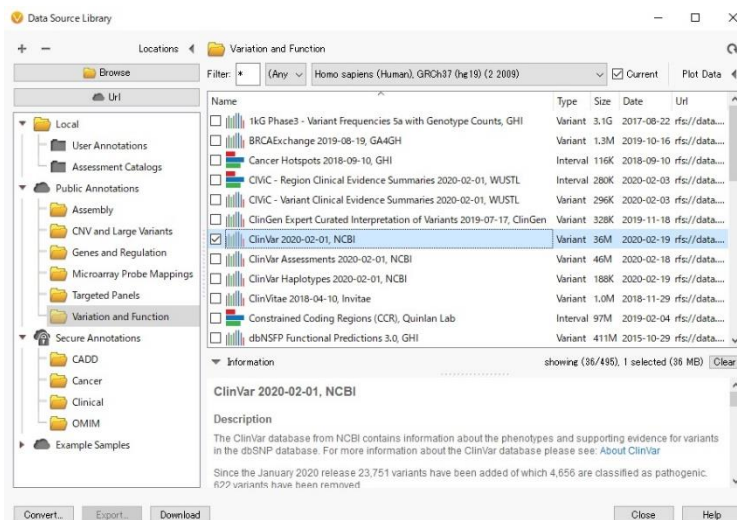
- 8GB of RAM
- 4+ CPU cores
- 100GB of space available for annotations and projects

#### Large Workflow Hardware Recommendations

- 32GB+ of RAM
- 12+ CPU Cores
- 1TB of space available for annotations and projects

## 【アノテーションリソース】

Golden Helix社によってメンテナンスが行われている高品質なアノテーションリソースを、Golden Helix社の専用サーバーからダウンロードマネージャーを通じてダウンロードし、バリエーションデータへのアノテーション付けを行います。また、アノテーションリソースのデータ以外にも、インハウスのカスタムアノテーションファイルをインポートして使用することもできます。ファイルフォーマットは、一般的に使用されているTSV、VCF、BED、GTPなどをサポートしています。



## 利用可能なアノテーションリソース例

### Assemblies

- GRCh37 (hg19)
- GRCh38 (hg38)
- T2T-CHM13

### Genes

- RefSeq
- Ensembl

### Clinical Annotations

- ClinVar
- ClinGen
- **OMIM**
- PharmaGKB
- BRCAExchange
- Orphanet

### Cancer Annotations

- CIViC
- Cancer Hotspots
- ICGC Simple Somatic Mutations
- TCGA Variants
- MSK Impact
- Precision Medicine Knowledgebase
- Cancer Genome Interpreter Biomarkers

### Functional Annotations

- dbNSFP
- REVEL
- **CADD**
- Conservation Scores

### Population Catalogs

- dbSNP
- 1000Genome
- NHLBI 6500 Exomes
- gnomAD
- TOPMed
- UK10K
- GenomeAsia 100K

### CNV and Large Variants

- ClinVar CNV
- ClinGen Gene Dosage Sensitivity
- DGV CNVs
- DECIPHER
- gnomAD High Frequency CNV
- gnomAD Structural Variants

\* 赤字はプレミアムアノテーションのライセンスが必要

## 【プレミアムアノテーション】

VarSeq®に標準搭載のアノテーションリソースの他に、有償のプレミアムアノテーションライセンスを購入することにより、さらに高品質なアノテーションデータを使用して、解析を行うことが可能になります。

## OMIM

ヒトの遺伝子と遺伝病の情報をまとめたデータベースです。遺伝病と関連する遺伝子名や遺伝形式、バリエーション情報、さらにデータベースのIDや出展論文の情報を、バリエーションデータのアノテーションとして利用できるようになります。VSReports®機能を使用する場合は、レポートに遺伝子やバリエーションの臨床情報の説明を含めることができます。

- OMIM Genes
- OMIM Phenotypes
- OMIM Variants
- OMIM Gene with Details
- OMIM Phenotypes with Details

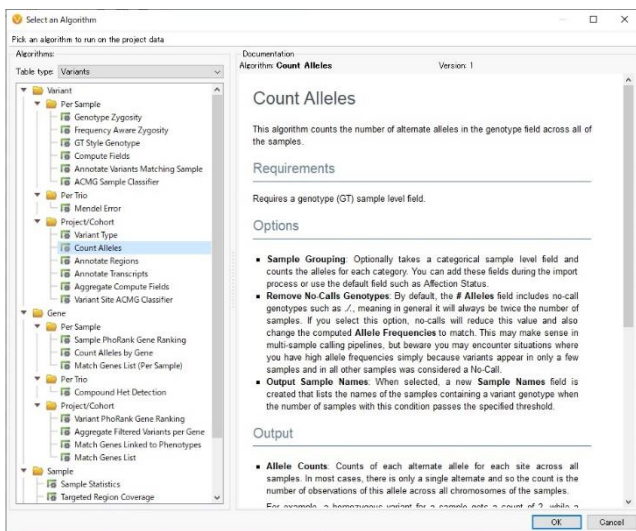
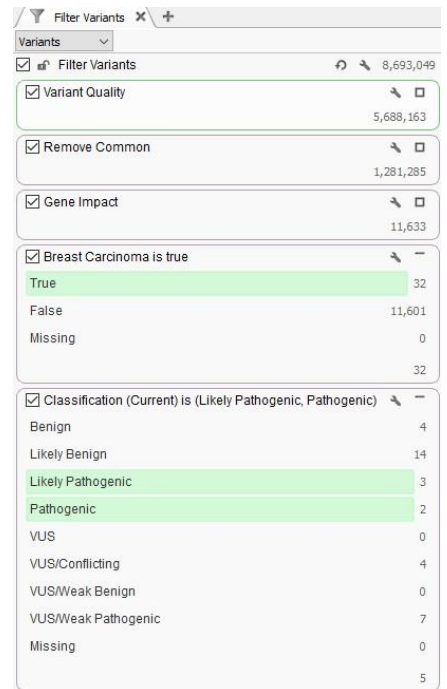
## CADD

CADD (Combined Annotation Dependent Depletion) は、ヒトゲノム上で取り得る全パターンのバリエーションの有害性を数値データとしてスコア化したデータベースです。SNVだけではなく、Insertion/Deletionについても、シミュレーションによりスコアを自動的に計算します。

## 【フィルタリングワークフロー】

VarSeq®では、VCFファイルよりインポートしたバリエーションデータに対して、各種アノテーションを利用したバリエーションフィルタリングのワークフローを作成することができます。作成したワークフローはテンプレートとして保存でき、別サンプルのVCFファイルのインポート時に使用することで、バリエーションデータのインポートからアノテーション付け・フィルタリングまでを、自動で実行させることができます。

- バリエーションクオリティ（VAF, Read Depthなど）によるフィルタリング
- 一般集団アレル頻度データによるコモンバリエーションの除去
- ホモ接合、ヘテロ接合バリエーションの抽出
- ミスセンスバリエーション、ナンセンスバリエーション、スプライシング異常バリエーションなどの抽出
- HPOタームを利用した、特定の疾患関連遺伝子の抽出、または疾患関連度によるランク付け
- アノテーションリソースに登録されている病的バリエーションなどの抽出
- 親子間の比較を行うトリオ解析、腫瘍・正常組織間の比較を行うペア解析



## 【データ解析アルゴリズム】

VarSeq®では、データベースによるアノテーションリソース以外に、VarSeq®に搭載されている解析用アルゴリズムを使用し、バリエーションや遺伝子、あるいはサンプルのデータテーブルに対し、様々な情報をアノテーションとして付加することができます。また、これで付加したアノテーションは、フィルタリングワークフローに組み込むことが可能です。

- トリオ解析（de Novoバリエーション、複合ヘテロ接合体の検出など）
- 集団サンプル内の各バリエーションのカウントや頻度計算
- サンプルの親縁関係、性別のチェック
- リードアライメントデータ（BAM/CRAMファイル）のカバレッジ計算
- 疾患関連遺伝子情報によるバリエーションのランク付け

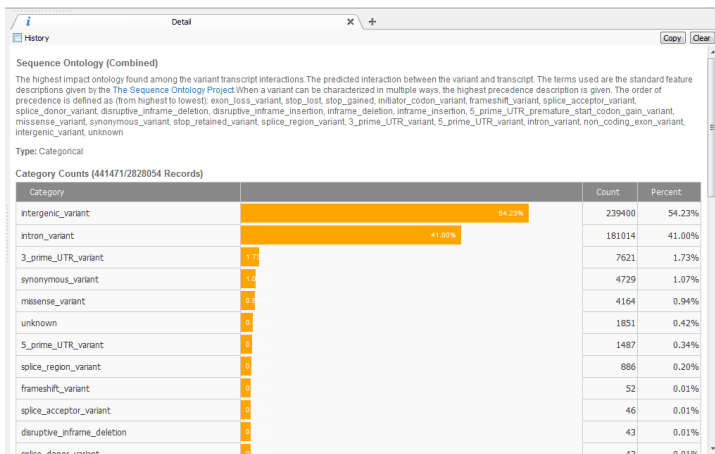
## 【バリエーションテーブルの表示】

VarSeq®にインポートされたVCFファイルのデータは、バリエーションデータのテーブルとして表示され、アノテーションリソースやデータ解析アルゴリズムによるアノテーション付けが可能です。また、ワークフローによるフィルタリングを実行すると、直ちにフィルタリング結果のバリエーションのみが表示に画面がアップデートされるため、フィルタリング結果の確認を迅速に行うことができます。

Variant Info		Sample Variants		RefSeq Genes 105.20201022 v2, NCBI				ClinVar 2021-10-07, NCBI		
Chr.Pos	Ref/Alt	Variant Allele Freq	Read Depths (DP)	Gene Names	Sequence Ontology (Combined)	Effect (Combined)	HGVS p. (Clinically Relevant)	Variant ID	Classification	Review Status
2:132905792	G/C	0.346154	78	ANKRD30BL	missense_variant	Missense	NP_001345345.1:p.Ser230Cys	?	?	?
3:75787646	AA/-	0.389961	259	ZNF717	frameshift_variant	LoF	NP_001277137.1:p.Phe376Serfs*3	?	?	?
3:178917478	G/A	1	73	PIK3CA	missense_variant	Missense	NP_006209.2:p.Gly118Asp	156446	Likely Pathogenic	(0 Stars) No Assertion Criteria Pro...
6:31238859	G/C	0.365079	189	HLA-C	missense_variant	Missense	NP_002108.4:p.Gln204Glu	?	?	?
6:31238866	C/G	0.416268	209	HLA-C	missense_variant	Missense	NP_002108.4:p.Glu201Asp	?	?	?
7:55249071	C/T	0.714286	84	EGFR,EGFR...	missense_variant	Missense	NP_005219.2:p.Thr790Met?	16613	Other	(3 Stars) Reviewed By Expert Panel
7:55259515	T/G	0.726744	172	EGFR	missense_variant	Missense	NP_005219.2:p.Leu858Arg	16609	Other	(3 Stars) Reviewed By Expert Panel
8:10467571	C/T	0.415584	309	RP1L1	missense_variant	Missense	NP_849188.4:p.Gly1346Glu	?	?	?
8:10467673	C/T	0.494949	297	RP1L1	missense_variant	Missense	NP_849188.4:p.Gly1312Glu	?	?	?
8:133826909	A/-	1	51	PHF20L1	frameshift_variant	LoF	NP_057102.4:p.Asn322Metfs*42	?	?	?
9:172167	C/T	0.405063	79	CBWD1	missense_variant	Missense	NP_060961.3:p.Ser115Asn	?	?	?
9:100092969	-/GAGGAG	0.55814	92	CCDC180,SU...	inframe_insertion	Missense	NP_065944.3:p.Glu735_Glu736dup.?	?	?	?
10:46999592	-/ATGAGGG...	0.722772	101	GPRIN2	inframe_insertion	Missense	NP_001372216.1:p.Met262_Glu264d...	?	?	?
11:48387946	-/CA	0.322314	121	OR4C5	frameshift_variant	LoF	NP_001335152.1:p.Glu25Trpfs*4	?	?	?
12:124824722	-/GCCGCTGCT	1	50	NCOR2	inframe_insertion	Missense	NP_006303.4:p.Ser1837_Gly1839dup	?	?	?
14:20666176	-/A	1	82	OR11G2	frameshift_variant	LoF	NP_001372962.1:p.Gly196AArgfs*74	?	?	?
16:70896016	A/-	0.461929	197	HYDIN	frameshift_variant	LoF	NP_001257903.1:p.Gln390AArgfs*5	?	?	?
17:7577120	C/T	1	213	TP53	missense_variant	Missense	NP_000537.3:p.Arg273His	12366	Pathogenic	(3 Stars) Reviewed By Expert Panel
17:39346593	CCT/-	0.52381	84	KRTAP9-1	disruptive_inframe_deletion	Missense	NP_001177389.1:p.Thr152_Cys153de...	?	?	?
17:45219332	T/G	0.558376	197	CDC27	missense_variant	Missense	NP_001247.3:p.Asn480His	?	?	?
19:51499377	C/T	0.449438	89	KLKB1	missense_variant	Missense	NP_653088.1:p.Val286Leu	?	?	?
19:58352661	T/A	0.544828	145	ZNF587B	missense_variant	Missense	NP_001363152.1:p.Phe207Ile	?	?	?
20:17639786	T/G	0.269504	282	RRBP1	missense_variant	Missense	?	?	?	?
20:17639846	T/G	0.231618	272	RRBP1	missense_variant	Missense	?	?	?	?

## 【アノテーションの詳細表示と集計グラフ】

バリエーションテーブルでは、詳細データ表示用タブより、各バリエーションの全アノテーションデータを確認することができるため、テーブル上には表示しきれないアノテーションの詳細なテキストなども確認することができます。また、テーブルに含まれる全バリエーションに対する、任意のアノテーションの集計グラフなども表示させることが可能です。

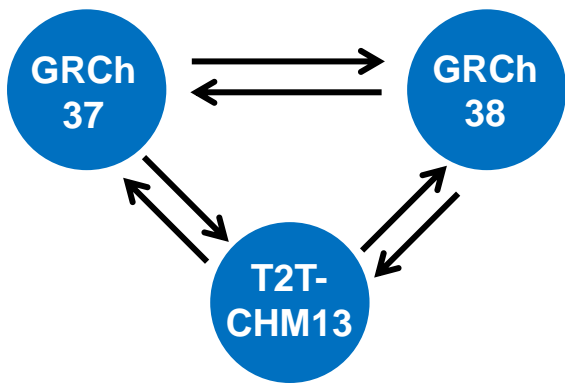


ClinVar 2021-06-03, NCBI

Ref/Alt	C/T
Variant ID	12582
Classification	Pathogenic
Review Status	(2 Stars) Criteria Provided, Multiple Submitters, No Conflicts
RSID	rs121913529
Gene Names	KRAS
Gene IDs	3845
HGVSc Name	NM_004985.5:c.35G>A
HGVSp Name	NP_004976.2:p.Gly12Asp

ClinVar Assessments 2021-06-03, NCBI

Assessment Source	1	2	3	4	5
Assessment Source	Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine	Invitae	OMIM	GeneDx	Arin Greene Laboratory, Boston Children's Hospital, Harvard Medical School
Classification	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic
Interpretation	Somatic KRAS variants have been identified in up to 15% of cases of ovarian carcinoma, and Gly12Asp [more]	This sequence change replaces glycine with aspartic acid at codon 12 of the KRAS protein (p. [more])	Using exome DNA sequencing and droplet digital PCR analysis, Nikolaev et al. (2018) identified a [more]	The G12D pathogenic variant in the KRAS gene has been reported previously with somatic mosaicism, [more]	?



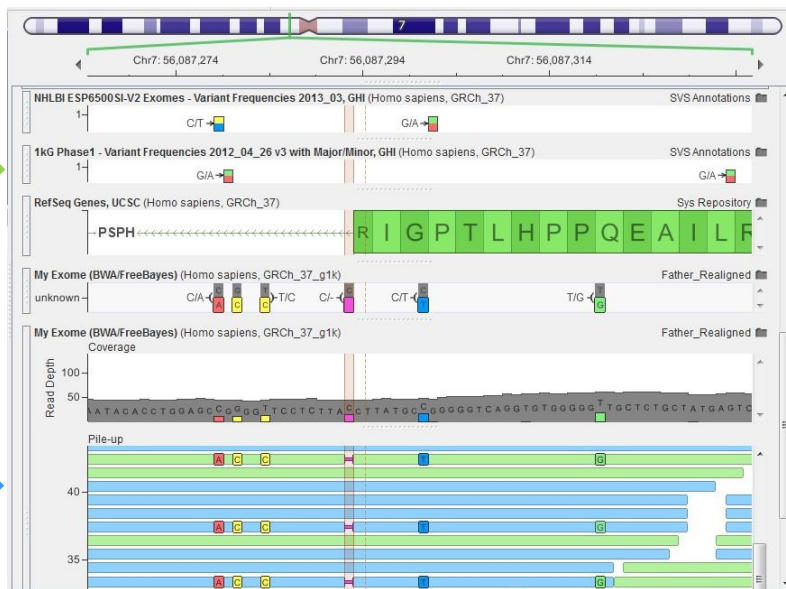
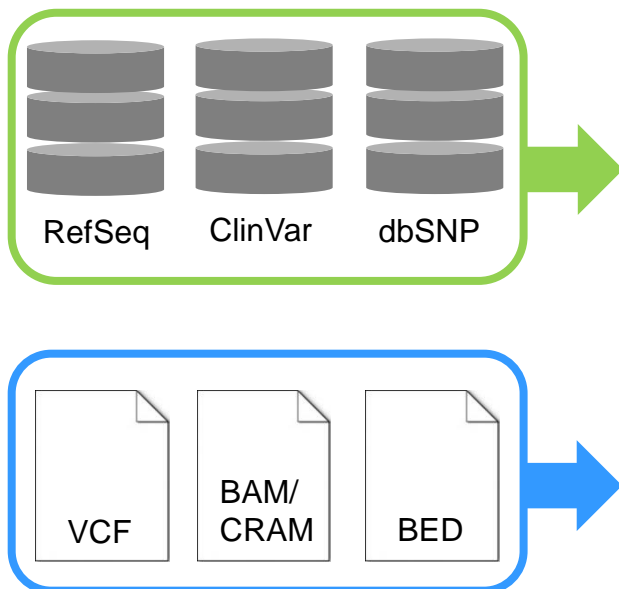
## 【ゲノムビルドのLiftover】

VCFファイルなどの各種データソースのインポート時に、ゲノムビルド GRCh37, GRCh38さらにT2T-CHM13間のLiftover機能を利用可能です。異なるゲノムビルドに基づいたデータ同士を対応させたり、一方のゲノムビルドでのみ提供されているアノテーションリソースを用いた解析などに使用できるようになります。


## 【ゲノムブラウザー】

VarSeq®には、Golden Helix社が開発した高機能なゲノムブラウザー「GenomeBrowse®」が組み込まれており、VCFファイルのバリエーションデータの他に、BAM/CRAMファイルによるリファレンスゲノムのリードアライメントデータを表示可能です。アライメントデータは、リード深度を表すカバレッジグラフと、リード配列のPile-upデータの表示が可能で、バリエーションの目視チェックなどを行う際に役立ちます。

またゲノムブラウザーには、VCFファイルやBAM/CRAMファイル以外に、ダウンロードマネージャーから取得したアノテーションリソースや、カスタムアノテーションデータ、さらにVAFなどの数値データもプロットでき、これら様々なデータを統合して表示することが可能です。



有償アドオン「VSReports®」を利用することで、VarSeq®でフィルターしたバリエーションデータに対して、OMIMデータベースに登録されている疾患情報とリンクさせたレポートを出力することが可能になります。レポートには、ラボ情報やサンプル情報、さらに自身のコメントや臨床的解釈の情報なども含めることが可能で、バリエーションデータセットを、Primary Findings（主要所見）とIncidental Findings（偶発所見）に分けて記載することができます。レポートには、疾患情報の出典論文のリストも含まれ、根拠となる論文を容易に探し出すことができます。


Golden Labs  
203 Enterprise Blvd  
Bozeman, 59718  
Phone: 406-587-8137  
Fax: 406-555-5555
Provider Information  
Physician Dr. James McCoy  
Institution Acme General Labs  
Case Id 1234

Patient Information		Sample Information	
Name	Arthur Dent	Sample Site	Blood
Gender	Male	Sample Type	Blood
Date of Birth	10/11/1985	Collection Met...	Peripheral Draw
Id	42	Panel Coverage	85.96%
Avg. Read Depth		3850x	
Collection Date		11/1/2016	
Receipt Date		11/3/2016	
Report Date		11/9/2016	

**Results**

Positive: Mutations with an established somatic link detected.

**Affected Genes**

ABL1 (0)	ASXL1 (0)	BRAF (1)	BRCA1 (0)	BRCA2 (0)	BUB1B (0)	CALR (0)	CBL (0)	CDH1 (0)	CDK4 (0)	CDKN1C (0)
CDKN2A (0)	CEBPA (0)	CEP57 (0)	CHEK2 (0)	CUX1 (0)	CYLD (0)	DDB2 (0)	DICER1 (0)	DIS3L2 (0)	KRAS (1)	NRAS (1)

**Primary Findings**

Gene	Zygoty	Variant	Exon	Pathogenicity
NRAS	Heterozygous	NM_002524.4:c.181C>A(NP_002515.1:p.Gln61Lys)	3	Pathogenic
BRAF	Heterozygous	NM_004333.4:c.1799T>A(NP_004324.2:p.Val600Glu)	15	Likely Pathogenic
KRAS	Heterozygous	NM_033360.2:c.38G>A(NP_203524.1:p.Gly13Asp)	2	Pathogenic

**Interpretation Summary**  
Mutations in three known oncogenes were detected.

## GOLDEN HELIX

Patient Information											
Name	Jane Doe	Gender	Female	Date of Birth	November 10, 2010	Id	1234				
Mother Information											
Name	Jane Doe	Date of Birth	September 28, 1988	Id	1235						
Father Information											
Name	John Doe	Date of Birth	November 10, 1988	Id	1236						
Reference Information											
Physician	Dr. Hydad	Institution	Golden Labs	Case Id	GL-1234						
Sample Information											
Sample Site	Blood	Sample Type	Blood	Collection Method	Venipuncture	Collection Date	January 30, 2017	Receipt Date	January 30, 2017	Report Date	January 31, 2017

### Results - Positive

Mutations with an established link detected.

#### Primary Findings

Gene	Exon	Variant	Zygoty	Pathogenicity
LPN2	20	NM_014846.2:c.*3C>T	Heterozygous	Uncertain Significance

#### Interpretation Summary

Patient possess a a single Pathogenic de novo variants in the LPN2 gene.

#### Recommendations

The variant could be associated with Chronic Recurrent Multifocal Osteomyelitis (CRMO). Treatment of CRMO typically involves administration of nonsteroidal anti-inflammatory drugs (NSAIDs) and physical therapy to avoid disuse atrophy of muscles and contractures.

#### Incidental Findings

Gene	Exon	Variant	Zygoty	Pathogenicity
BRCA2	2	NM_000059.3:c.6513G>C	Heterozygous	Pathogenic

## 【OMIMとの連携】

遺伝子と表現型の関連情報をまとめたデータベース「OMIM」のアノテーションリソースを使用し、データベースに含まれるメンデル性疾患と15,000以上の遺伝子情報を、遺伝子レベル、バリエーションレベルでレポートに記載することができます。

## 【レポートテンプレートのカスタマイズ】

レポートは、ラボで使用している他の文書とマッチするように、デザインを変更することができ、さらにバリエーションデータにアノテーション付けされた、任意のアノテーション項目をレポートに含めるようにカスタマイズを行うことも可能です。

## 【ACMGガイドラインのサポート】

VSReports®には、ACMGガイドラインに基づいたレポートテンプレートが搭載されています。テンプレートは、遺伝性疾患とがんのものがあり、両方とも無料で使用できます。

## 【ファイル出力】

レポートは、HTMLドキュメントとして作成され、1クリックでPDFファイルとして出力が可能です。

**Primary Findings**

Variant: 7:117199645 TCT/- (*CFTR*) ⓘ

Classification: Pathogenic ⓘ

Interpretation: **B I U** ⓘ

accompanied the migrations of early farmers standing from the Middle East and slowly progressing toward the northwest of Europe. The diffusion of the gene may have been favored by the selective advantage conferred by the gene. Strong association with the so-called haplotype B was detected. The possibility of false-positive results is the

**Incidental Findings**

Variant: 16:81942028 C/G (*PLCG2*) ⓘ

\*Classification: Uncertain Significance ⓘ

Interpretation: **B I U** ⓘ

This is a Missense Variant located in the PLCG2 gene. Enzymes of the phospholipase C family catalyze the hydrolysis of phospholipids to yield diacylglycerols and water-soluble phosphorylated derivatives of the lipid head.

Variant: 17:41243800 C/T (*BRCA1*) ⓘ

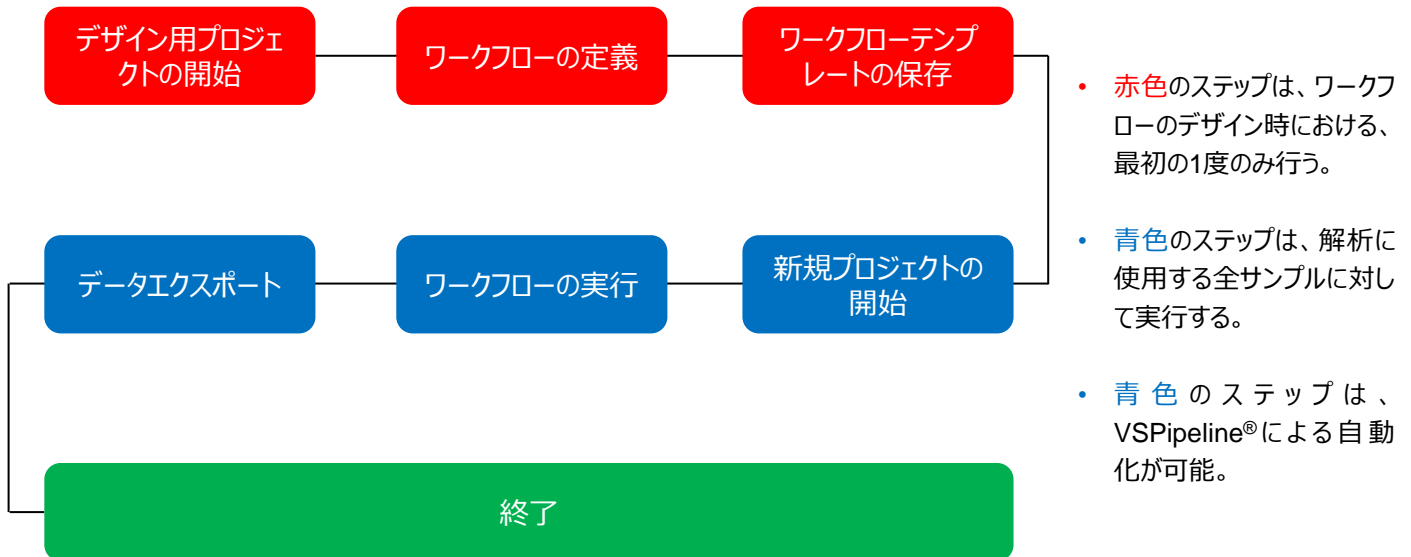
\*Classification: Likely Pathogenic ⓘ

Interpretation: **B I U** ⓘ

This is a Missense Variant located in the BRCA1 gene. [Castilla et al. \(1994\)](#) studied 50 probands with a family history of breast and/or ovarian cancer (604370) for germline mutations in the coding region of the BRCA1

# VSPipeline®

有償アドオン「VSPipeline®」では、VarSeq®標準のGUIインターフェースに加えて、コマンドライン型のインターフェースが利用可能になります。VarSeq®で作成したワークフローを、複数サンプルデータでバッチ処理を行ったり、ユーザー独自で構築したバイオインフォマティクス解析パイプラインに組み込んだ使用が可能となります。



## 【アノテーション付けの自動化】

VarSeq®は、Golden Helix社がメンテナンスとアップデートを行っている幅広いパブリックデータベースへのアクセスを提供し、同社による注釈とキュレーションのプロセスによって高品質なバリエーションアノテーションを得ることができます。またこれらのデータベースは常に最新バージョンが利用可能ですが、過去のバージョンも同様に利用できるようになっており、バージョン指定による結果の再現性も得ることが可能です。VSPipeline®によってアノテーション付けの工程を自動化することによって、これらの作業を効率的に行うことができます。

## 【解析パイプラインへのインテグレート】

VSPipeline®のコマンドライン型インターフェースを利用することによって、独自に構築したバイオインフォマティクスパイプラインにVarSeq®をそのまま組み込むことができます。これによりVarSeq®用プロジェクトファイルに加え、アノテーション・フィルタリング済みのバリエーションテーブルのExcelファイルやVCFファイル、サンプルやカバレッジレポートなどの自動出力が可能になります。

## 【ワークフローの反復実行】

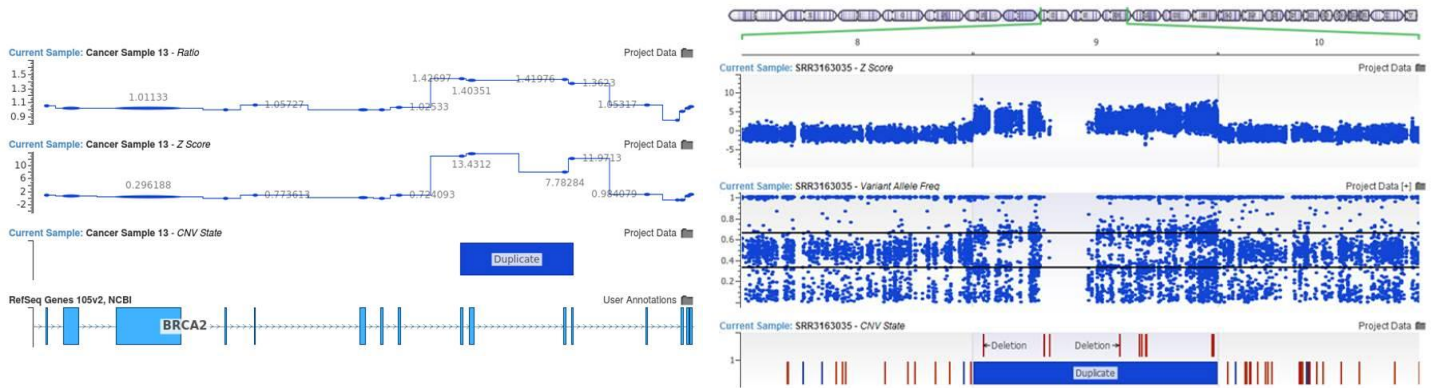
VSPipeline®の利用により、テンプレートとして保存しておいたアノテーション付けやフィルタリングのワークフローを反復して実行することが可能となり、複数サンプルにおけるバッチ処理などにおいて効力を発揮します。

```
grudy@linserverdev:~/scratch$ vspipeline -c batch file=panel_workflow.vsbatch
Loading...
Importing variants...
[0%] Starting...
[0%] Preparing source reader
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[0%] Writing new file
[15%] Writing new file - About 10 Seconds remaining
[16%] Writing new file - About 10 Seconds remaining
[22%] Writing new file - About 10 Seconds remaining
[27%] Writing new file - About 10 Seconds remaining
[29%] Writing new file - About 10 Seconds remaining
[34%] Writing new file - About 10 Seconds remaining
[35%] Writing new file - About 10 Seconds remaining
[40%] Writing new file - Less than 10 Seconds remaining
[43%] Writing new file - Less than 10 Seconds remaining
[45%] Writing new file - Less than 10 Seconds remaining
```

## 【クリニカルラボでの使用】

VSPipeline®は、全ゲノムや全エクソームまたは遺伝子パネルシーケンスを実施しているクリニカルラボ、あるいはバイオインフォマティクスのコアラボにおける、データ解析パイプラインの自動化に活用できます。

有償アドオン「VS-CNV®」の導入によって、VarSeq®においてゲノム上のCNV（コピー数異常）の検出が可能になります。CNV検出は、リードアライメントデータ（BAM/CRAMファイル）におけるカバレッジデータを基に計算され、遺伝子パネルによるターゲットシーケンスや全エクソームシーケンス、さらに全ゲノムシーケンスのデータも適応可能です。またLoH (Loss of Heterozygosity) 解析にも対応し、検出されたCNVは、ゲノムブラウザでグラフィカルに表示が可能です。



### 【CNVの検出】

VS-CNV®は、次世代シーケンサーのデータセットから、CNVを検出できるようにデザインされています。このアルゴリズムでは、リファレンスサンプルセットとの比較と、臨床サンプルによるバリデーションが行われたパラメータによって、高精度かつ高感度にCNVを検出することができます。

### 【NGSデータへの適用】

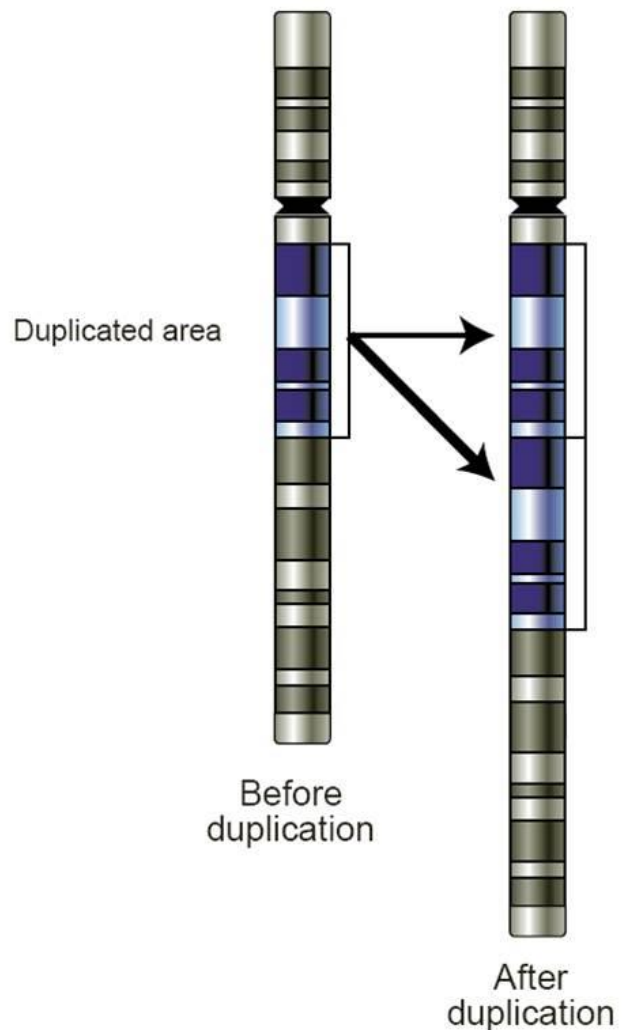
このCNV検出アルゴリズムは、既存の次世代シーケンサーを用いた遺伝子パネル解析、全エクソーム解析、全ゲノムシーケンスに適用可能です。検出されたCNVデータは、SNP/InDelsなどのスモールバリエーションと同様にアノテーション付けとフィルタリングを行い、さらにVSClinical®による評価も可能です。

### 【ビジュアライゼーション】

CNV検出を行うと、コピー数の変化が検出された各領域ごとのP-valueに加え、遺伝子パネルの各ターゲット領域やゲノム全体の指定されたウィンドウサイズごとのRatioとZ-scoreが算出されます。これらの値をゲノムブラウザにプロットすることで、ゲノム上のコピー数の変化をグラフィカルに表示できます。

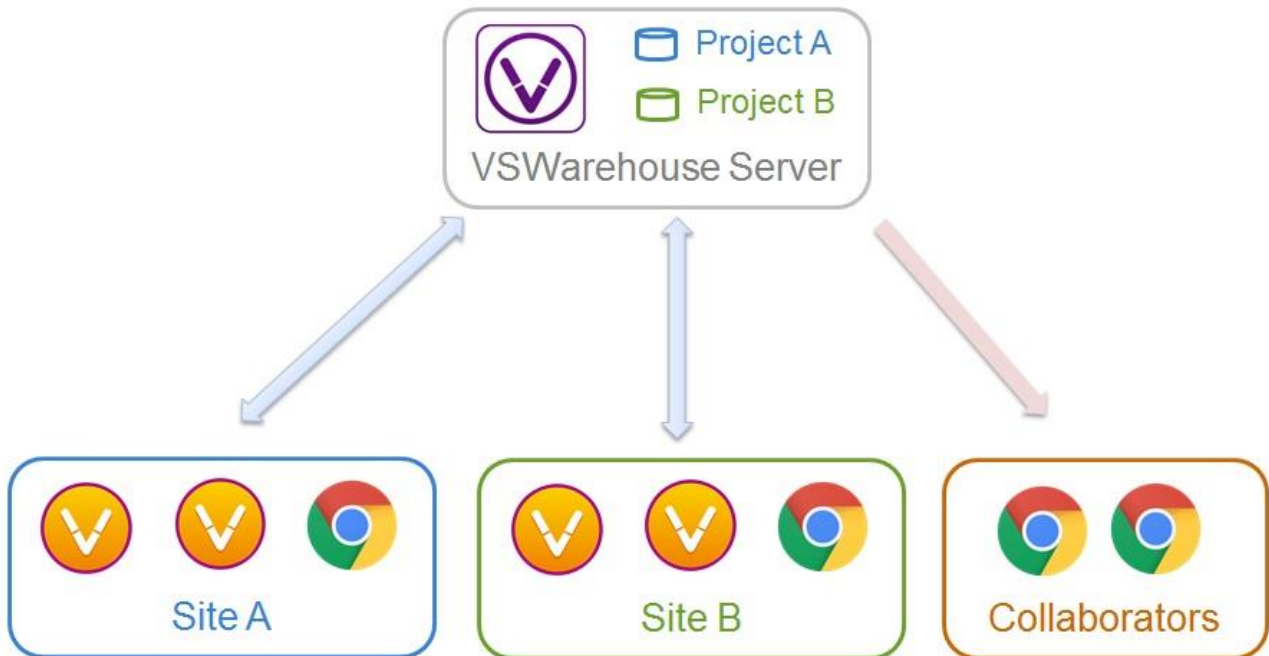
### 【生殖細胞・がん遺伝子検査】

VS-CNV®は生殖細胞・がんの遺伝子検査に適用可能です。このアドオンで検出可能なCNVは、200bp程度の1エクソン単位から、数Mbpの染色体の異数性にまでおよび、従来のMLPAやqPCR、またCMAsといったマルチプレックスアッセイによる検出系を、次世代シーケンサーによる既存のSNP/InDels解析パイプラインを利用した単一プラットフォームを使用するアッセイ系に置き換えることができます。またLoss of Heterozygosity (LoH) 情報とCNV情報を統合することで、コピー数変動がないLoHと、コピー数変動を伴うヘテロ接合性欠失とを区別することが可能です。



# VS Warehouse®

有償アドオン「VSWarehouse®」は、バリエーションデータをWebサーバーに保存しデータベースとして公開する、データシェアリングのためのシステムです。データベースに保存されたデータは、Webクライアントソフトウェアを使用して、自由にデータの閲覧・検索、さらにデータのファイル出力などが可能になり、研究施設内あるいは施設間でデータの共有が容易になります。



## 【プロジェクトごとのサンプル管理】

VSWarehouse®は、ハイパフォーマンスなストレージテクノロジーから構成されています。研究プロジェクトごとにデータを管理し、サンプルデータはVarSeq®よりアップロードを行います。

## 【スケーラブルテクノロジー】

VSWarehouse®は、完全にカスタマイズと最適化が行われたストレージと、クエリー実行レイヤーをもつデータベーステクノロジーから構成されており、従来型のデータベース管理用ソフトウェアと比較して、高速にデータの検索を行うことが可能です。

## 【バリエーションデータカタログ】

独自で集計した临床上重要なバリエーションデータ、または外部データリソースをカタログ化し、アノテーションデータとして公開が可能です。

## 【アレル頻度アノテーション】

プロジェクト内サンプルデータから、バリエーションごとのアレル数や頻度を自動的に集計し、アノテーションデータとして利用が可能です。

## 【レポートの保存】

VSReports®で作成したレポートを保存することができ、バリエーションデータとサンプル情報、さらにデータに対する自身の解釈などをまとめて保存しておくことが可能です。

The screenshot shows the VS Warehouse web interface. The main content is a table of variants with columns for Genomic Coordinate, Ref/Alt, Gene Names, and HGVS p. (Clinically Relevant). The table shows 1.3 K records. On the right side, there are filter options for '1.3 K variants', including 'In COSMIC? is True', 'Read Depths (DP) is greater than or equal to 100.0 or Read Depths (DP) is null for Sample1', 'Genotype Qualities (GQ) is greater than or equal to 80.0 or Genotype Qualities (GQ) is null for Sample1', and 'Effect (Combined) is either Missense or LoF'.

Genomic Coordinate	Ref/Alt	Gene Names	HGVS p. (Clinically Relevant)
1:2494329	G/A	TNFRSF14	NP_003811.2:p.Val241Ile
1:11190803	C/T	MTOR	NP_004949.1:p.Glu1799Lys
1:27105927	-/G	ARID1A	NP_006006.3:p.Gly1848fs
1:27105930	G/-	ARID1A	NP_006006.3:p.Asp1850fs
1:43814978	G/A	MPL	NP_005364.1:p.Ser505Asn
1:43814980	G/A	MPL	NP_005364.1:p.Ala506Thr
1:43815007	T/A	MPL	NP_005364.1:p.Trp515Arg
1:43815007	T/C	MPL	NP_005364.1:p.Trp515Arg
1:43815008	G/C	MPL	NP_005364.1:p.Trp515Ser
1:45797504	C/G	MUTYH	NP_001121897.1:p.Gln338His
1:65332610	C/T	JAK1	NP_002218.2:p.Val310Ile
1:97770919	C/T	DPYD	NP_000101.2:p.Val732Ile
1:97981394	T/C	DPYD	NP_000101.2:p.Ile543Val
1:115252203	C/T	NRAS	NP_002515.1:p.Ala146Thr
1:115256517	T/A	NRAS	NP_002515.1:p.Ser65Cys
1:115766637	T/A	NRAS	NP_002515.1:p.Gln61His

## 【様々なインターフェース】

VCFファイルに含まれる情報を損なわずに、アップロードされたすべてのバリエーションデータのマトリックスを作成し、ウェブインターフェースやVarSeq®などの複数のインターフェースを用いたアクセスが可能です。



有償アドオン「VS Clinical<sup>®</sup>」では、American College of Medical Genetics (ACMG) と Association for Molecular Pathology (AMP) のガイドラインに基づき、それぞれ生殖細胞系列バリエーションと体細胞バリエーションの臨床的意義の評価を行います。評価時には、レディーメイドのワークフローや専用の知識ベースなどのアノテーションリソースを活用し、自動的に処理が行われます。評価結果にはバリエーションの病原性分類情報に加え、臨床的意義の解釈や治療薬情報などと共にレポート出力が可能で、出典論文のリストなども含めることができます。

## ACMGガイドライン

### Evidence Summary

#### Variant Summary:

The missense variant NM\_001354689.3(RAF1):c.770C>T(p.S257L) is not observed in the large population cohorts of gnomAD, or 1kG (Genome Aggregation Database et al., 2020; 1000 Genomes Consortium et al., 2015). The variant was added to dbSNP as rs80338796 in version 131. This variant was found in ClinVar (Variant 13957) with a classification of Pathogenic and a review status of (3 stars) reviewed by expert panel. There is a large physicochemical difference between serine and leucine, which is likely to impact secondary protein structure as these residues differ in polarity, charge, size and/or other properties.

#### Recommended to Score Pathogenic

- PM2** The p.Ser257Leu variant is novel (not in any individuals) in gnomAD. The p.Ser257Leu variant is novel (not in any individuals) in 1kG.
- PM1** 30 variants within 6 amino acid positions of the variant p.Ser257Leu have been shown to be pathogenic, while none have been shown to be benign.
- PP2** The gene RAF1 has a low rate of benign missense variation as indicated by a high missense variants Z-Score of 2.49. The gene RAF1 contains 56 pathogenic missense variants, indicating that missense variants are a common mechanism of disease in this gene.
- PP3** The p.Ser257Leu missense variant is predicted to be damaging by both SIFT and PolyPhen2. The serine residue at codon 257 of RAF1 is conserved in all mammalian species. The nucleotide c.770 in RAF1 is predicted conserved by GERP++ and PhyloP across 100 vertebrates.
- PS1** The p.Ser257Leu variant is a missense mutation resulting in an amino acid change which is shared by the previously classified pathogenic variant p.S257L.
- PM5** The p.Ser257Leu variant is a missense mutation resulting in an amino acid change which occurs at the same amino acid position as 3 previously classified pathogenic variants.

### 【病原性の自動分類】

VS Clinical<sup>®</sup>に搭載のACMG auto-classifierツールを使用することで、高度にキュレーションされたアノテーションリソースやバイオインフォマティクスアルゴリズムを利用し、自動的にバリエーションの病原性分類を行うことが可能です。病原性分類結果はアノテーションとしてバリエーションテーブルに表示され、フィルタリングに用いることも可能です。

### 【ユーザーフレンドリーなインターフェース】

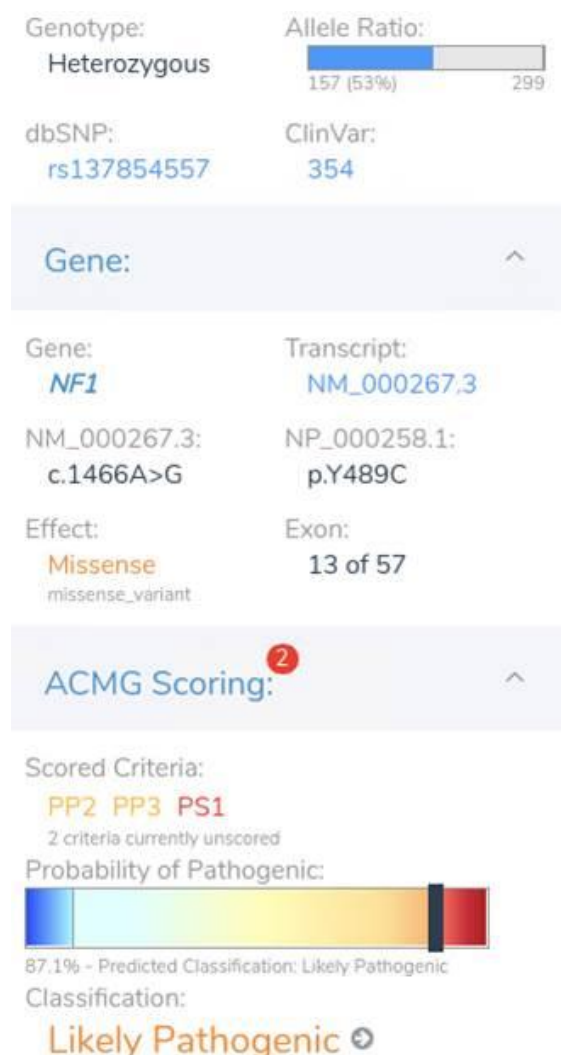
VS Clinical<sup>®</sup>のACMGガイドラインワークフローでは、評価項目のスコアリングをガイドするための専用のインターフェースが搭載されています。評価実施時には、各評価項目に関連した質問が表示され、ユーザーが回答を選択していくごとに結果がアップデートされます。また回答を行う画面では、VS Clinical<sup>®</sup>が判断した推奨の回答や、その根拠となる引用文献などの各種データソースも表示され、ユーザー自身による評価ステップを強力にサポートします。

### 【多種類のバリエーションのサポート】

VS Clinical<sup>®</sup>では、SNV/InDelsに加えCNV（コピー数異常）とSV（構造異常）も評価可能で、またVCFファイルよりインポートしたバリエーションだけでなく、HGVS表記などで手動で入力したバリエーションも使用可能です。

### 【ACMGガイドラインの実装】

ACMG/AMP共同ガイドラインに基づき、バリエーションの病原性をPathogenic、Benign、VUSなどの5段階に分類して評価します。このガイドラインによる評価では、バリエーションのアノテーションやゲノム上の位置、既報の臨床研究情報などを精査する必要がありますが、VS Clinical<sup>®</sup>はこれらの評価項目をスコアリングするためのワークフローを提供するとともに、スコアリングと解釈のプロセスを支援するバイオインフォマティクス手法、臨床知識ベースより引き出された文献やエビデンス情報なども提供します。



Drug Sensitivity Interpretations (6)	Drug Resistance Interpretations (2)	Drug Descriptions		
Drugs	Tier	Saved For	Variants	Clinical Trials
Bevacizumab, Ramucirumab, Regorafenib	Tier 1A	Colorectal Adenocarcinoma	Unspecified	0 Selected
Encorafenib + Cetuximab, Encorafenib + Panitumumab	Tier 1A	Colorectal Adenocarcinoma	BRAF V600E (Activating Mutation)	0 Selected
Entrectinib, Larotrectinib	Tier 1A	Solid Tumor Cancers	NTRK1 Fusion (Activating Mutation)	0 Selected
Pembrolizumab	Tier 1A	Solid Tumor Cancers	TMB High	0 Selected
Dostarlimab-gxly, Nivolumab, Nivolumab + Ipilimumab, Pembrolizumab	Tier 1A	Colorectal Adenocarcinoma	MSI High	0 Selected

### 【AMPガイドラインの実装】

がんゲノム検査における体細胞バリエーションの解釈に用いられるAMPガイドラインでは、治療の有効性などに基づいたエビデンスレベルを用いて、バイオマーカーを分類・評価します。VSClinical®のAMPガイドラインワークフローでは、SNV/InDelsに加えCNV（コピー数異常）、融合遺伝子、さらにTMBやMSIなどのゲノミックシグネチャーに対して、指定したがん種に有効な治療オプションなどを評価し、薬剤情報・臨床試験情報などとともにレポート出力が可能です。

### 【専用知識ベースへのアクセス】

Golden Helix CancerKBは、種々のがんや遺伝子における体細胞バリエーションやCNV、融合遺伝子、ゲノミックシグネチャーなどのバイオマーカーをはじめ、治療における薬剤感受性や抵抗性、予後・診断情報などのクリニカルエビデンスを網羅的に収集した、専門的なキュレーションデータベースです。VSClinical®では、このデータベースに収録されている遺伝的背景や確立済のクリニカルエビデンス、引用文献を含んだ臨床的意義の解釈などの充実したコンテンツにアクセスできます。

### 【腫瘍原性の分類】

新規の、またはまれに観察される体細胞バリエーションを評価するには、まず遺伝子に対するバリエーションの影響と、がん遺伝子の活性化またはがん抑制遺伝子の非活性化に対する能力を分類する必要があります。VSClinical®には体細胞バリエーションの腫瘍原性を評価するためのアルゴリズムが搭載されており、評価結果のスコアに応じて、バリエーションをOncogenic、Benign、VUSなどに自動で分類します。

### 【生殖細胞系列バリエーションの評価】

生殖細胞系列の起源が疑われる、または確認されたバリエーションに対しては、ACMGガイドラインに基づき病原性の評価を行い、レポート出力します。VSClinical®では、バリエーションのアノテーションなどをもとに自動で評価された病原性の分類結果に加え、評価結果の根拠となるデータソースも提示されます。

Drug Sensitivity Interpretation Editor

Tier 1A Drug Sensitivity for BRAF V600E in Colorectal Adenocarcinoma:

Sensitivity to Drugs:

- Encorafenib + Cetuximab ✕
- Encorafenib + Panitumumab ✕
- Add Drug +

Interpretation Saved For: Colorectal Adenocarcinoma ▼

Biomarker Scopes: + Add Scope...

BRAF V600E (Activating Mutation)

Clinical Evidence Tier: Tier I - Level A ▼

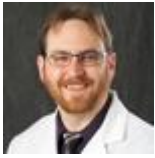
Interpretation:

Encorafenib is a BRAF inhibitor with similar efficacy towards wild-type BRAF and BRAF p.V600E mutation and minimum MAPK reactivation effect (PMID: 29356698). It was FDA-approved for the treatment of melanoma (PMID: 29573941) and colorectal cancer (PMID: 33503393) harboring BRAF p.V600E mutation. Cetuximab and panitumumab are monoclonal antibodies with antagonistic activity against EGFR (PMID: 21596817, 30218345, 31616627). They were approved by the FDA to treat patients with metastatic colorectal cancer with wild-type RAS (PMID: 17522246, 29083635).

In BRAF p.V600E positive metastatic CRC, several studies have shown that the combination therapy using encorafenib plus cetuximab improved overall survival, objective response rate and progression-free survival compared to standard chemotherapy (PMID: 33503393, 31566309). Based on this data, the FDA approved the use of encorafenib in combination with cetuximab to treat adult patients with metastatic CRC with a BRAF p.V600E mutation (PMID: 31566309, 30892987).

Inline References: View All 7 References ...

## ケーススタディ



### Dr. Benjamin Darbro

Director, Shivanand R. Patil Cytogenetics and Molecular Laboratory,  
University of Iowa

アイオワ大学のThe Shivanand R. Patil Cytogenetics and Molecular Labでは、45年間にわたって臨床試験を行ってきました。現在のはがん、出生前あるいは出生後の遺伝子検査に注力しています。次世代シーケンサーを用いた疾患遺伝子パネル解析を行うにあたって、Dr. Benjamin DarbroはVarSeq®を使用し、アノテーション付けやフィルタリング、さらにデータの閲覧も1つのソフトウェアのみで行えるようにし、さらにVSReports®を用いることで、レポート作成まで行える環境をラボ内で整えました。



### Dr. Laura Li

Associate Director, Clinical Genomics Laboratory,  
Children's Hospital Los Angeles

Children's Hospital Los Angeles(CHLA)におけるDr. Laura Liたちのチームでは、視神経形成不全の遺伝子的原因について研究を行っています。彼女のチームでは、25家系の全エクソームのトリオ解析を実施し、アライメントとバリエーションコールを行ったVCFファイルに対して、VarSeq®を用いてバリエーションデータのアノテーション付けとフィルタリングを行います。VarSeq®を使用することによって、興味のある遺伝子とバリエーションのリストを高速に取得できるようになりました。



### Dr. Chaim Jalas

Director, Genetic Resources & Services, The Center for Rare Jewish  
Genetic Disorders, Bonei Olam

1999年より、Bonei Olamでは不妊治療と研究のために、多額の資金を費やしてきました。同組織の遺伝子研究は、Center for Rare Jewish Genetic Disordersで実施され、稀少疾患や障害を引き起こす遺伝子の同定に焦点を当てています。Jalasと彼のチームは、次世代シーケンサーを用いた全エクソームのトリオ解析を行い、VarSeq®を用いてバリエーションデータのアノテーション付けとフィルタリングを実行しています。さらに彼らはVSWarehouse®を使用し、サンプル集団のカタログを作成しており、他の組織が同じくVSWarehouse®で作成したデータベースと統合させることによって、スケールの大きいコホート研究にも対応できるようになりました。



### Dr. Jeffrey Rosenfeld

Bioinformatics Scientist, Rutgers Cancer Institute of New Jersey

Rutgers Cancer Institute of New Jerseyの精密医療プログラムでは、患者ごとの治療オプションの選択のために、腫瘍組織の変異を迅速に同定する必要があり、バイオインフォマティクスプログラムが不可欠な役割を果たしています。Rosenfeldの所属するバイオインフォマティクスチームでは、次世代シーケンサーを用いてシーケンスされた50個のがん関連遺伝子パネルのデータからバリエーションの検出を行い、出力されたVCFファイルをVarSeq®にインポートし、独自に組み立てたワークフローを用いて、アノテーション付けとフィルタリングを迅速に実行します。ここで解析されたバリエーションのリストはXMLファイルで出力され、N-of-One社に送付されて同社のデータ解釈サービスを受けた後に、N-of-One社から返されたデータファイルをVSReports®にアップロードし、遺伝子やバリエーションの疾患情報、さらに治療方策などをまとめた、最終的なレポートファイルが作成されます。



# Sentieon

Bioinformatics secondary analysis tools that process genomics data with high computing efficiency, fast turnaround time, exceptional accuracy and 100% consistency.

# Sentieon Genomics Tools概要

Sentieon Genomics Toolsは、次世代シーケンサーの遺伝子変異解析における2次解析用のソフトウェアです。遺伝子変異解析のゴールドスタンダードであるBWA-GATK、および体細胞バリエーション検出用のMuTect、MuTect2を高速化したパイプラインが使用可能です。いずれのパイプラインも、一般的なCPUベースのシステムで動作し、またFASTQファイルからVCFファイル作成までに必要とする計算時間の10倍以上の高速化を実現しています。また、コマンドライン型インターフェースを採用しており、複数サンプルデータの一括処理や、ユーザー独自の解析パイプラインの構築に適しています。



-Same mathematical models as the Broad Institute

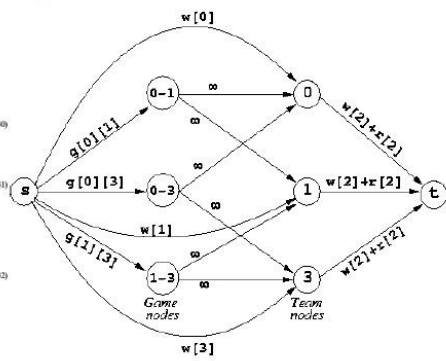
-more efficient compute algorithms

-Enterprise strength software implementation

$$\frac{\partial}{\partial \alpha} \ln \left( \sum_{i=1}^n e^{\alpha x_i} \right) = \frac{\sum_{i=1}^n x_i e^{\alpha x_i}}{\sum_{i=1}^n e^{\alpha x_i}}$$

$$\frac{\partial}{\partial \alpha} \ln \left( \sum_{i=1}^n e^{\alpha x_i} \right) = \frac{\sum_{i=1}^n x_i e^{\alpha x_i}}{\sum_{i=1}^n e^{\alpha x_i}}$$

$$\frac{\partial}{\partial \alpha} \ln \left( \sum_{i=1}^n e^{\alpha x_i} \right) = \frac{\sum_{i=1}^n x_i e^{\alpha x_i}}{\sum_{i=1}^n e^{\alpha x_i}}$$



<http://www.cs.princeton.edu/courses/archive/spr05/cos226/assignments/baseball/>

```
1 /* This line basically imports the "stdio" header file, part of
2 * the standard library. It provides input and output functional
3 * to the program.
4 */
5 #include <stdio.h>
6
7 /*
8 * Function (method) declaration. This outputs "Hello, world" to
9 * standard output when invoked.
10 */
11 void sayHello() {
12     // printf() in C outputs the specified text (with optional
13     // formatting options) when invoked.
14     printf("Hello, world!");
15 }
16
17 /*
18 * This is a "main function". The compiled program will run the
19 * defined here.
20 */
21 void main() {
22     // Invoke the sayHello() function.
23     sayHello();
24 }
```

[http://www.wikiwand.com/en/Programming\\_language](http://www.wikiwand.com/en/Programming_language)

## 【ライセンスの種類】

Sentieon Genomics Toolsでは、解析を行うサンプル数にあわせて3種類のライセンスタイプがあります。下記の表は、ライセンスごとの1年間に解析を行うサンプル数の目安です。

	WGS	WES	Gene Panel
Tier 1	500	2,500	25,000
Tier 2	2,000	8,000	75,000
Tier 3	4,000	16,000	N/A

## 【システム要件】

### ハードウェア

- 16GB+ of RAM (32GB for Servers)
- 8+ CPU Cores
- インターネット接続

### OS

- 64-bit Windows 10 or later (WSL2上で使用可)
- Linux Ubuntu 18.04 (64-bit only)
- RedHat/CentOS 7.x
- Debian 7.7
- OpenSUSE-13.2

## 【受賞実績】

Sentieon Genomics Toolsは、「ICGC-TCGA Dream Mutation Calling Challenge」「precisionFDA Consistency Challenge」「precisionFDA Truth Challenge」の3つのコンテストにおいて、解析結果データの一致率や正確性の評価で、トップの成績を取っています。

precisionFDA Consistency Challenge Reproducibility  
F1-score (%) between runs and between samples

	Garvan vs. Garvan rerun			Garvan vs HLI		
	FP	FN	F1-score	FP	FN	F1-score
Sentieon by UNM*	0	0	100	134633	315831	95.07
Sentieon (dual mapping)	0	0	100	107302	295568	95.53
ISAAC**	0	0	100	112255	266952	95.32
Genalica**	3621	3673	99.92	286306	433504	92.08
Edico Dragen**	3147	3216	99.93	161213	315611	94.87

precisionFDA Consistency Challenge Accuracy  
F1-score (%) to NIST truth set

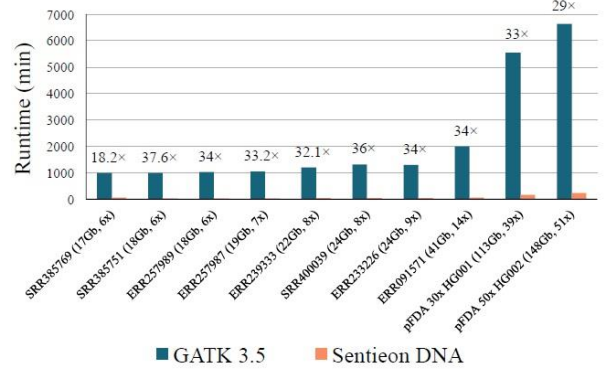
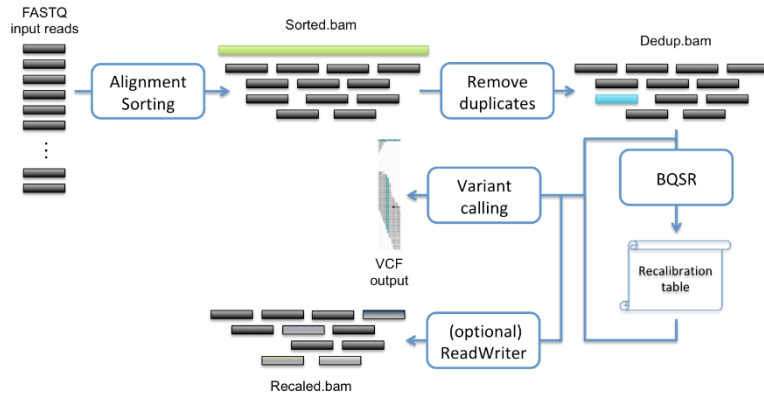
	Garvan sample			HLI sample		
	All	SNP only	Indel only	All	SNP only	Indel only
Sentieon by UNM*	99.39	99.86	95.85	98.97	99.73	92.94
Sentieon (dual mapping)	99.47	99.88	96.37	99.06	99.77	93.46
ISAAC**	97.29	98.57	86.66	96.34	97.99	82.10
Genalica**	98.04	99.20	89.10	97.25	98.83	84.66
Edico Dragen**	99.25	99.74	95.49	98.85	99.62	92.74

\* Sentieon standard pipeline results from Jeremy Edwards (University of New Mexico) submission  
 \*\* Edico, Genalica (MAP 2.2.0) and Isaac (aligner v01.14.07.14 & variant caller v2.0.13) run results from Changhoon Kim's (MacroGen Clinical Laboratory) submission

# 解析パイプライン

## 【Sentieon DNaseq】

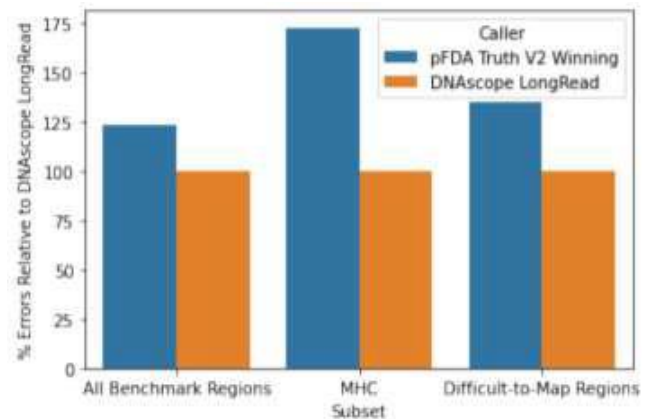
Sentieon DNaseqは、Broad InstituteのBWA-GATK HaplotypeCaller Best Practiceのバイオインフォマティクスパイプラインをもとに開発されており、同パイプラインと同じ数学モデルを用いて、リファレンスゲノム配列へのリード配列のマッピングからバリエーションコールまでを行うことが可能です。ただし計算アルゴリズムの効率化が行われており、FASTQファイルからVCFファイル作成の計算処理の場合は10倍、BAMファイルからVCFファイル作成の場合は20～50倍の高速化を実現しています。また、100,000ものサンプルデータのジョイントコールを行う場合でも、中間ファイルのマージをせずに実行できるため、時間と労力の節約が可能です。



\*Server specs: 32 core 2.4 GHz Intel Xenon server, 64 GB memory

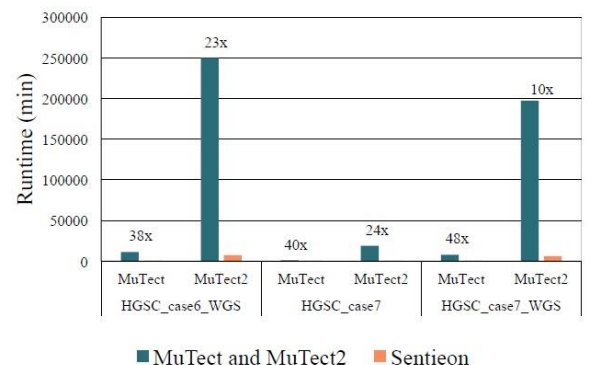
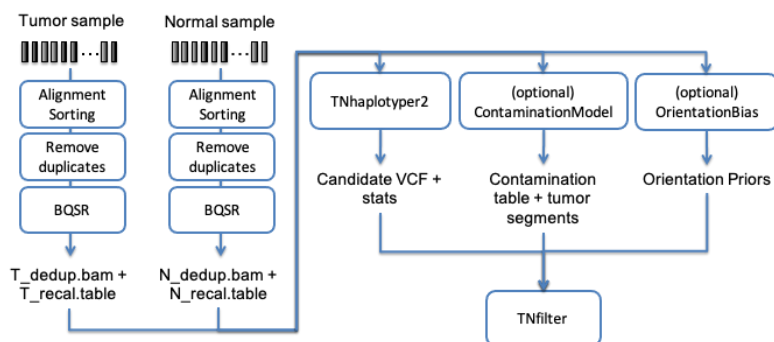
## 【Sentieon DNAscope】

Sentieon DNAscope は、生殖細胞系列のバリエーション検出用のパイプラインです。機械学習モデルを用いることにより、Sentieon DNaseq より高精度に SNV や Small InDels のバリエーションコールを行うことが可能で、さらに構造異常 (SV) の検出にも対応しています。また PacBio やOxford Nanopore のロングリード解析にも対応し、PacBio HiFiリードによる30xのヒト全ゲノムシーケンスを、16コアCPUマシンでおよそ4時間で解析可能で、また GIAB のベンチマークデータセットにおいて、99.83%以上の精度を達成しています。



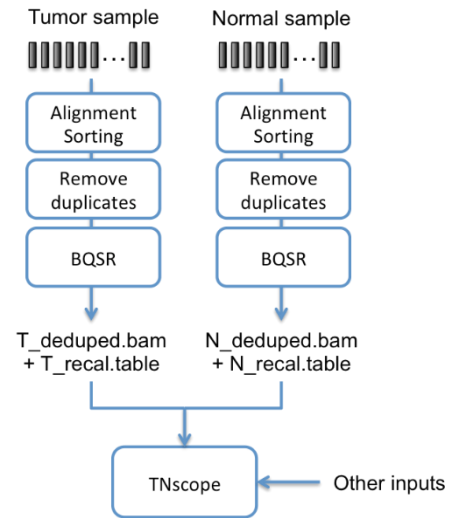
## 【Sentieon TNseq】

Sentieon TNseqは、Broad Instituteの体細胞バリエーション検出用パイプラインMuTect (SNV), MuTect2 (SNVとINDEL) と同じ数学モデルを使用した解析パイプラインです。腫瘍／正常の各サンプルのシーケンスデータを用いて、効率化された計算アルゴリズムで、MuTectおよびMuTect2の10倍以上の計算速度で、体細胞バリエーションの抽出を効率的に行います。



## 【Sentieon TNscope】

Sentieon TNscope は、体細胞バリエーション検出用のパイプラインです。MuTect および MuTect2 で使用されている数学モデルをさらに改良することで、より高い感度と特異度を実現しています。Tumor only サンプルと Tumor-Normal サンプルの両方の解析に対応し、SNV と Small InDels、および構造異常（SV）の検出を行います。



## 【サンプルスクリプト】

Sentieon Genomics Tools では、様々な種類の解析パイプラインを簡単に実行するためのサンプルスクリプトが提供されています。下記は、現在提供されているスクリプトの一例です。

### Germlineサンプル用パイプライン

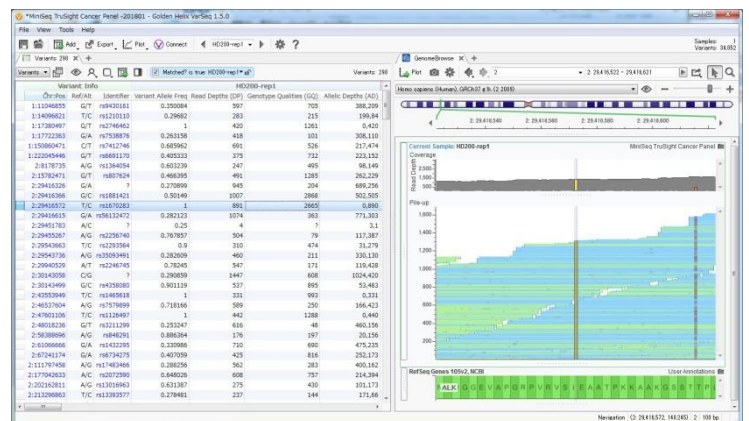
- Joint calling
- WES/WGS
- RNA-Seq calling
- PacBio HiFi reads

### Somaticサンプル用パイプライン

- Tumor only
- Tumor + Normal
- Hybridization panel
- Amplicon panel
- ctDNA with UMI
- ctDNA without UMI

## 【3次解析用ソフトウェアとの連携】

Sentieon Genomics Toolsから出力されたBAMファイルとVCFファイルは、VarSeq®などの3次解析用ソフトウェアにインポートし、データの閲覧やバリエーションデータへのアノテーション付加とフィルタリングといった、各種のバイオインフォマティクス解析に用いることが可能です。





# Your partner for **all** NGS workflows

Panels to Exomes to Genomes  
Labs to Regional Hospitals to Genome Centers  
Germline to Prenatal to Oncology

**GOLDEN HELIX**  
*Accelerating the Quest for Significance™*

## 輸入販売元



## フィルジェン 株式会社 バイオインフォマティクス部

【お問い合わせ】

〒459-8011 愛知県名古屋市緑区定納山1丁目1409番地

TEL : 052-624-4388 FAX : 052-624-4389

E-mail : support@filgen.jp URL : <https://filgen.jp/>

代理店

(Aug.,2023)