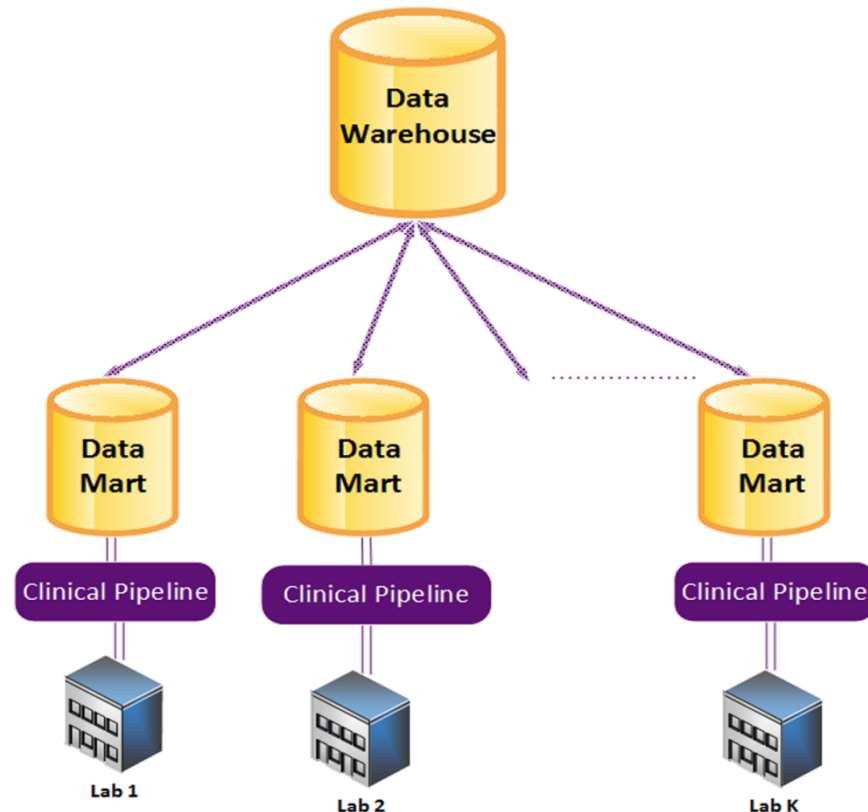


# 臨床ゲノムデータウェアハウスを用いた データシェアリング

フィルジエン株式会社 バイオサイエンス部  
(biosupport@filgen.jp)

- ゲノムデータの医療への応用には、臨床サンプルのゲノム解析により得られたデータや知見を、研究機関内または機関同士で、相互に共有するシステムの構築が必要。
- 次世代シーケンサーから得られた、臨床サンプルの遺伝子変異データを、各種アプリケーションリソースと共に保存・公開するデータウェアハウスサーバーを構築し、組織間でのデータの共有が可能となる。



## SNP & Variation Suite



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

## VarSeq<sup>®</sup>



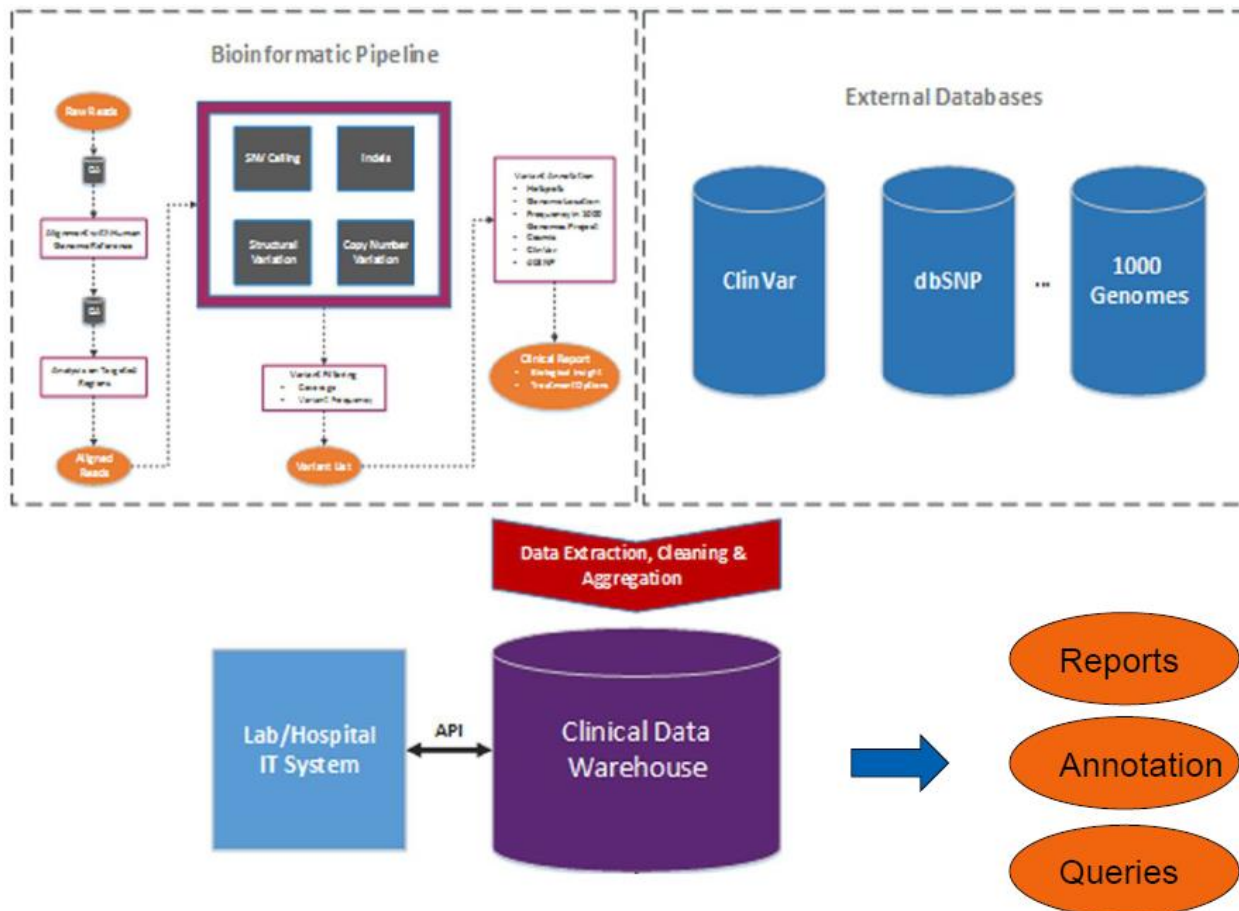
- Variant Interpretation
- Cancer Diagnostics
- CNV Calling
- Clinical Reporting
- High-throughput NGS Testing

## VSWarehouse



- Fully Integrated with VarSeq Workflows
- Scalable Technology
- Organize Samples into Projects
- Centralized Clinical Report Hosting
- Create Variant Assessment Catalogs

- VSWarehouseでは、次世代シーケンスデータのバイオインフォマティクス解析パイプラインで得られた、サンプルの遺伝子変異データと、外部データリソースのアノテーションデータを統合して保存し、クライアントはVarSeq<sup>®</sup>を使用して、データを引き出す。



- 次世代シーケンサーから得られた遺伝子変異データを、Webサーバーに保存し、データベースとして公開するためのソフトウェア
- Webサーバーへのデータのアップロードは、VarSeq<sup>®</sup>より行う。逆に、サーバーに保存されたデータをダウンロードし、VarSeq<sup>®</sup>上でアノテーションデータとして使用することも可能
- 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
- 発見した臨床的に重要な変異情報を、カタログデータとして保存し、公開することが可能
- ユーザーごとのアクセス権の設定



warehouse.goldenhelix.com

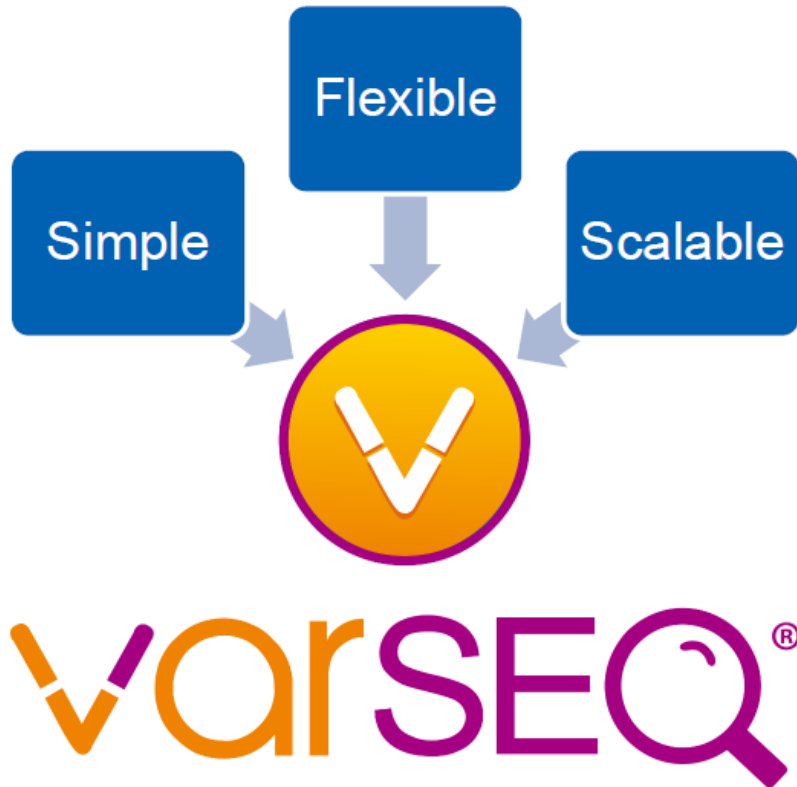
warehouse.goldenhelix.com/variants/Cancer%20Gene%20Panel/1/results/

WAREHOUSE Workspace Query Results Logout rudy@goldenhelix.com

1.3 K variants

- 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
- 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
1:115256537	T/A	NRAS	NP_002515.1.p.Gln64His



- VCFファイルに含まれる変異データから、任意の検索条件で、データのフィルタリングを行うワークフローを作成
- キュレーションされた様々なデータリソースを使用し、変異データへアノテーション付けを実行
  - dbSNP
  - RefSeq Genes
  - COSMIC
  - 1000 Genome
  - NHLBI 6500 Exomes
  - ExAC Variant
  - gnomAD Exomes
  - SIFT and PolyPhen
  - dbNSFP Functional Predictions
  - ClinVar
  - CIVic
  - ICGC Simple Somatic Mutation
  - 各種遺伝子パネルのターゲットデータ ...など
- HGVDなどの、独自定義データもインポートし、アノテーションデータとして利用可能
- ゲノムブラウザーを搭載し、BAMファイルデータや各種アノテーションデータをグラフ表示
- 遺伝子パネルのターゲットデータを利用し、カバレッジレポートの作成や、有償アドオンによるCNV解析を実行

- VarSeq<sup>®</sup>でフィルターした変異データを、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、レポートとして出力
- 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、レポートに記載が可能
- 変異データセットを、Primary Findings（主要所見）とIncidental Findings（偶発所見）に分けて記載が可能
- がん関連変異データベースOncoMDと連携し、臨床試験情報とFDA承認薬情報、薬剤感受性データをレポート出力

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

**Results**

**Positive:** Mutations with an establish somatic link detected.

Affected Genes

<i>ABL1</i> (0)	<i>ASXL1</i> (0)	<i>BRAF</i> (1)	<i>BRCA1</i> (0)	<i>BRCA2</i> (0)	<i>BUB1B</i> (0)	<i>CALR</i> (0)	<i>CBL</i> (0)	<i>CDH1</i> (0)	<i>CDK4</i> (0)	<i>CDKN1C</i> (0)
<i>CDKN2A</i> (0)	<i>CEBPA</i> (0)	<i>CEP57</i> (0)	<i>CHEK2</i> (0)	<i>CUX1</i> (0)	<i>CYLD</i> (0)	<i>DDB2</i> (0)	<i>DIGER1</i> (0)	<i>DIS3L2</i> (0)	<i>KRAS</i> (1)	<i>NRAS</i> (1)

Primary Findings

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

Interpretation Summary

Mutations in three known oncogene were detected.

Recommendations

Enrollment in a clinical trial testing drugs targeting the mutation should be considered.

R&D Genetic Test Workflow

Cataloging of Tested Samples



Test Launched

Ongoing Data Science, Re-Annotation, Medical Archiving

## Annotate, Filter, Interpret Workflow

**VCF**  
Variant Call File



### Annotated Variants: Marked for Reporting

**Patient**

- Phenotype
- Lab Info
- Referring Info

Name

D.O.B.

Pathogenic

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut ...

Incidental

Ut enim ad minim veniam, quis

Structured

**VSReport**

**Pathogenic**

Although VSReport is most commonly associated with medical diagnosis, Lee et al. (2008) showed that VSReport is increasingly utilized in academia. As the subject interest across domains and/or modern approaches with variants was found in the VSReport, an annotation approach has advantage of broader drug-targeting, medicine in the gene.

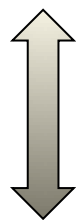
**Recommended Reading**

This is a VSReport report based on the VSReport.

Rendered



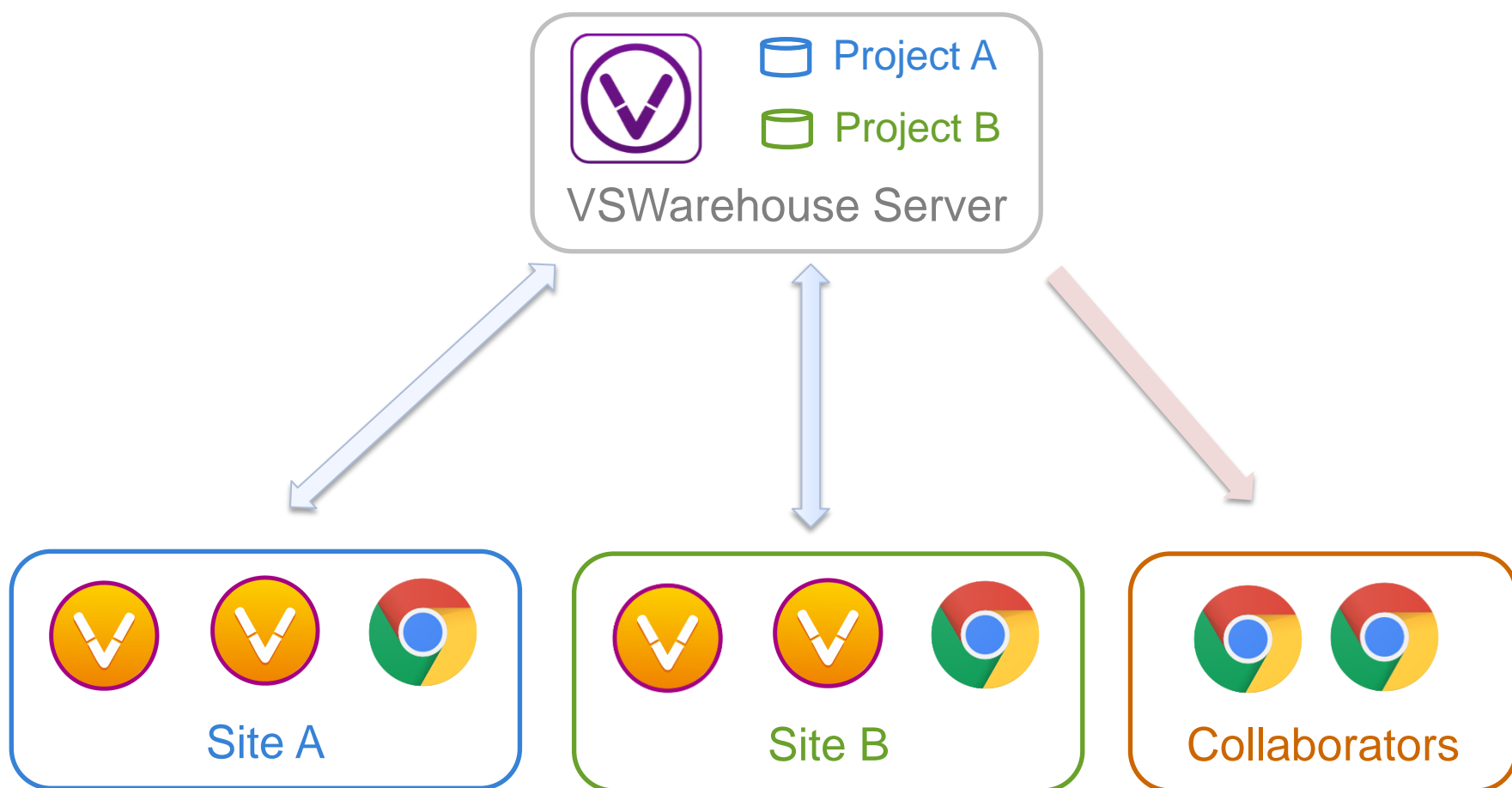




varSEQ™

- 研究プロジェクトごとにデータを管理し、VarSeq®からサンプルの変異データをアップロード
- プロジェクト内サンプルデータから、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
- 独自で集計した臨床上重要な変異データ、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
- VSReportで作成したレポートを保存することができ、変異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

- クライアント用ソフトウェアとしてVarSeq®を利用する場合は、自身のデータのアップロードや、カタログデータをアノテーションとして利用し、自身のデータの詳細解析が可能。
- Webブラウザを使用する場合は、保存されているサンプルデータの閲覧・フィルタリング・ファイル出力が可能。

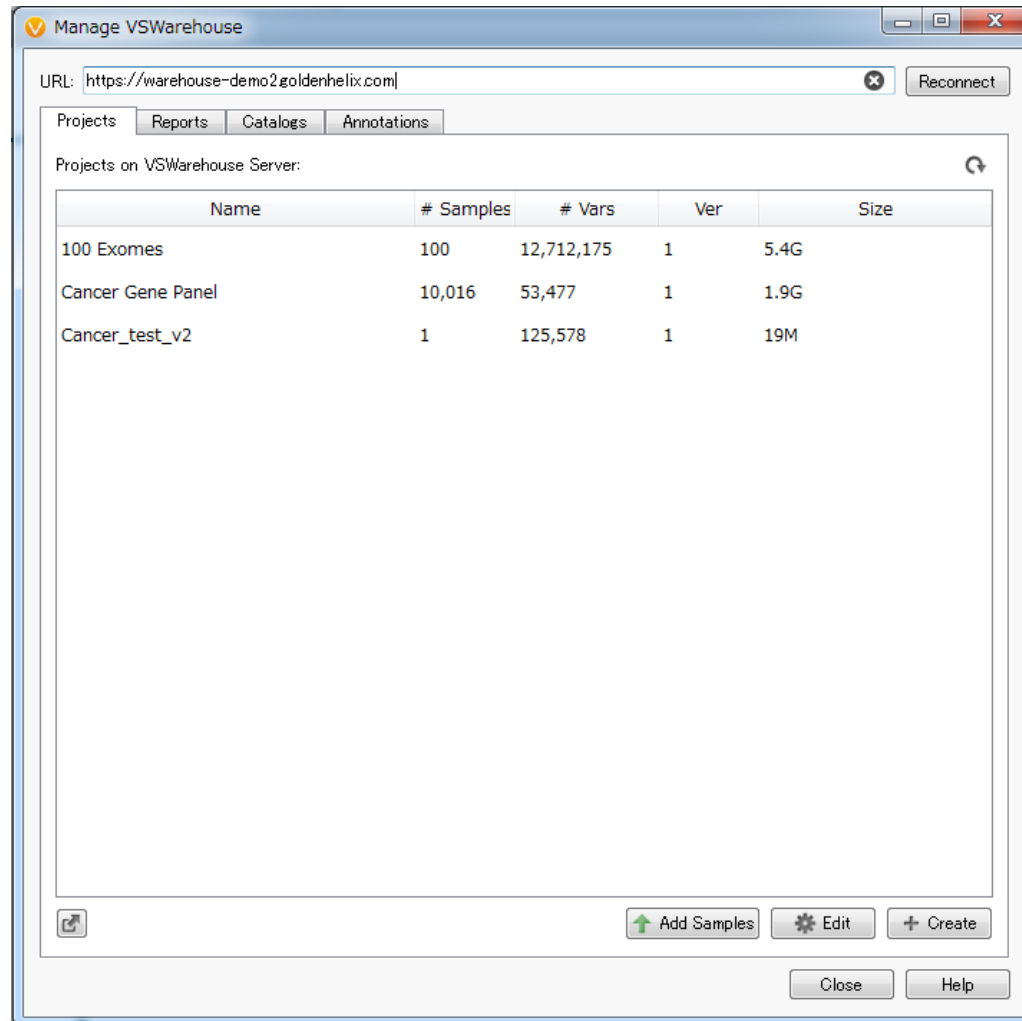


The screenshot shows the Golden Helix VarSeq 1.4.4 interface. The 'Connect' button in the top toolbar is highlighted with a red box. The main window displays the following information:

- Filter Chain:** 35,905 variants. Active filters include:
  - Read Depths (DP) (Current)  $\geq 30$ : 1,526
  - Variant Allele Freq (Current)  $> 0.2$ : 1,522
  - HGVD Filter: 1,359
  - Clinical Significance is (Likely Pathogenic, Pathogenic): HD200-rep1: 101 Benign, 1 Benign/Likely Benign, 0 Drug Response, 18 Likely Benign, 1 Likely Pathogenic, 10 Not Provided, 9 Other, 1 Pathogenic, 3 Risk Factor, 1 Uncertain Significance, 2 Missing.
- Variants Table:**

Chr:Pos	Ref/Alt	Identifier	Filter	Variant Allele Freq	Allelic Depths (AD)	Read Depths (DP)	Genotype Qualities (GQ)
6:151948366	G/A	rs2046210	PASS	0.612903	180,285	465	77
17:41234451	G/A	rs41293455	LowGQX	0.241218	324,103	427	
- GenomeBrowse:** Shows a genomic track for Homo sapiens (Human), GRCh37 g1k (2 2009) at position 17:41,234,912 - 41,234,989. A zoomed-in view shows read alignments for the variant at 17:41,234,986.
- ClinVar 2017-05-04, NCBI:** Shows three BRCA1 gene annotations with arrows pointing to the variant position.
- RefSeq Genes 105v2, NCBI:** Shows the BRCA1 gene structure with exons and introns.

- VarSeq<sup>®</sup>にインポートを行ったデータを、VSWarehouseサーバーにアップロード。



- 新たにプロジェクトを作成し、サンプルデータの追加を行う。

Internal Clinical Classifications

Chr 18: 48604676 - A/G (Existing Record)

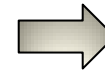
Classification: Likely Pathogenic

Notes: This is a Missense Variant located in the SMAD4 gene. In 5 unrelated patients with sporadic occurrence of Myhre syndrome (139210), Le Goff et al. (2012) identified a de novo heterozygous 1498A-G transition in the SMAD4 gene, resulting in an ile500-to-val (I500V) substitution in a highly conserved residue in the MH2 domain involved in transcriptional activation. The mutation was not found in 200 controls.

Editor: Gabe

Recent Assessments Using Current Schema

Date	User	Classification	Notes	Editor
2017-02-06 17:58	rudy@goldenhelix.com	Likely Pathogenic	This is a Missense Variant located in the SMAD4 gene. In 5 unrelated patients with sporadic occurrence of Myhre syndrome (139210), Le Goff et al. (2012) identified a de	Gabe



Variants) Clinical Significal

Genes) Clinical Significal

Clinical Significance is Pathogenic: Patho

Variants: 2

Internal Clinical Classifications				
Variant ID	Classification	Notes	Editor	
?	?	?	?	?
109	Likely Pathogenic	This is a Miss...	Gabe	

- OMIM/ClinVar類似の変異カタログのデータベースの作成。
- 変異の分類情報とコメント、表現型情報などの記載が可能。
- 登録した変異情報は、VarSeq<sup>®</sup>にダウンロードしてアノテーションとして利用が可能。

The screenshot displays the Golden Helix VSReport web interface. The top navigation bar includes 'File', 'View', 'Tools', and 'Help'. The main content area is divided into several sections:

- Patient Information:** Name: NA19240, Gender: Male, Date of Birth: 02/06/2017, Id: 1234.
- Mother Information:** Name: NA19238, Date of Birth: 02/06/2017, Id: 1235.
- Father Information:** Name: NA19239.
- Internal Clinical Classifications:** Chr 18: 48604676 - A/G (Existing Record). Classification: Likely Pathogenic. Notes: This is a Missense Variant located in the SMAD4 gene. In 5 unrelated patients with sporadic occurrence of Myhre syndrome (139210), Le Goff et al. (2012) identified a de novo heterozygous 1498A>G transition in the SMAD4 gene, resulting in an Ile500-to-val (I500V) substitution in a highly conserved residue in the... Editor: Gabe.
- Recent Assessments Using Current Schema:**

Date	User	Classification	Notes	Editor
2017-02-06 17:50	rud@goldenhelix.com	Likely Pathogenic	This is a Missense Variant located in the SMAD4 gene. In 5 unrelated patients with	Gabe
- Variant Table:**

Ref/Alt	Project ID	Allele Counts	Allele Frequencies	# Alleles	# Het	# HomoVar	Unaffected - Allele Counts
G/A	8619262	7	0.035	200	7	0	
?	?	?	?	?	?	?	

- サンプルごとのVSReportの保存を行い、Webブラウザからレポート内容の閲覧が可能。

The screenshot shows a web browser window displaying the Filgen Warehouse interface. The browser address bar shows the URL <https://warehouse-demo2.goldenhelix.com>. The page header includes the 'WAREHOUSE' logo, 'Manage' and 'Logout' buttons, a search bar, and the user email 'ozawa@filgen.jp'. The main content area is organized into several categories:

- Projects**
  - [100 Exomes](#): 100 samples, 12,712,175 variants, 1 available versions +
  - [Cancer Gene Panel](#): 10,016 samples, 53,477 variants, 1 available versions +
  - [Cancer\\_test\\_v2](#): 1 samples, 125,578 variants, 1 available versions +
- Reports**
  - [Filgen \(1 samples\)](#)
    - » [Primary Findings \(2 variants\)](#)
    - » [Incidental Findings \(2 variants\)](#)
- Assessment Catalogs**
  - [Clinical Significance](#): 1 variants, 1 available versions +
- Samples**
  - [Explore Samples](#): Query samples across all projects, 10018 samples
- ClinVar Changes**
  - [ClinVar 2017-04-06, NCBI](#)
  - [ClinVar 2017-05-04, NCBI](#)

- Webブラウザからアクセスした場合、「Projects」「Reports」「Assessment Catalogs」などのカテゴリーに分類されて表示される。

https://warehouse-demo2 x +  
https://warehouse-demo2.goldenhelix.com/project\_variant/1-100 Exomes/1/results/ 検索

WAREHOUSE 100 Exomes (1) Query Results Logout Search... ozawa@filgen.jp

12.7 million records < < > > 1 Go page 1 / 635609

Genomic Coordinate	Ref/Alt	Gene Names	Effect (Combined)	HGVS c. (Clinically Relevant)
1:10250	A/C	?	Other	?
1:10352	T/A	?	Other	?
1:10407	T/-	?	Other	?
1:10492	C/T	?	Other	?
1:13082	C/A	?	Other	?
1:13273	G/C	?	Other	?
1:13418	-/GAGA	?	Other	?
1:14464	A/T	?	Other	?
1:14610	T/C	?	Other	?
1:14653	C/T	?	Other	?
1:14671	G/C	?	Other	?
1:14673	G/C	?	Other	?
1:14677	G/A	?	Other	?

12.7 million variants Clear All

No filters are applied

Browse the data in the table to the left or click [Query](#) to create and add filters.

- 「Projects」では、アップロードされた全変異データのリストが表示され、「Query」よりフィルタリングの設定を行う。



The screenshot shows the Filgen Warehouse interface. The browser address bar indicates the URL: [https://warehouse-demo2.goldenhelix.com/project\\_variant/1-100 Exomes/1/query/](https://warehouse-demo2.goldenhelix.com/project_variant/1-100 Exomes/1/query/). The page title is "WAREHOUSE" and it shows "100 Exomes (1)" and "Query" results. The search bar contains "ozawa@filgen.jp".

On the left sidebar, the "Clinical Significance" filter is selected. The main content area displays "Clinical Significance - ClinVar 2016-05-02, NCBI" with a description: "Clinical Significance, unknown, untested, non-pathogenic, probable-non-p... read more". A list of clinical significance categories is shown, with "Likely Pathogenic" and "Pathogenic" selected.

On the right, a filter panel titled "30 variants" is visible. It contains four active filters, each with a red 'X' icon:

- Read Depths (DP) is greater than or equal to 50.0 and Read Depths (DP) is not missing for HG00096,HG00097,HG00099, and 97 others
- Affected - Allele Frequencies is greater than or equal to 0.3 or Affected - Allele Frequencies is missing
- Effect (Combined) is either 'LoF' or 'Missense'
- Clinical Significance is either 'Likely Pathogenic' or 'Pathogenic'

- 変異データとともにアップロードしたアノテーションデータを使用し、VarSeq<sup>®</sup>と同様に、任意のフィルターの設定を行う。

# フィルタリング結果

https://warehouse-demo2.x +  
https://warehouse-demo2.goldenhelix.com/project\_variant/1-100 Exomes/1/results/ 検索

WAREHOUSE 100 Exomes (1) Query Results Logout Search... ozawa@filgen.jp

30 records « First < Prev Next > Last » 1 Go page 1 / 2 Change Columns... Export... Hide Filters

Genomic Coordinate	Ref/Alt	Gene Names	Effect (Combined)	Clinical Significance	Disease Name
1:98348885	G/A	DPYD	Missense	Pathogenic, Not Provided	Dihydropyrimidine dehydrogenase deficiency, not provided
1:100672060	T/C	DBT	Missense	Pathogenic, Other	Intermediate maple syrup urine disease type 2, not specified
1:161599693	T/C	FCGR3B	Missense	Pathogenic	Neutrophil-specific antigens na1/na2
1:169519049	T/C	F5	Missense	Pathogenic, ?, Other, ?, Other, ?, Other, ?, Pathogenic, ?, Pathogenic, Risk Factor	Thrombophilia due to factor V Leiden, Ischemic stroke, susceptibility to, Budd-Chiari syndrome, susceptibility to, Recurrent abortion, Factor V deficiency, ?
1:171076966	G/A	FMO3	Missense	Pathogenic, Likely Pathogenic, Pathogenic	Trimethylaminuria, mild, Trimethylaminuria, Trimethylaminuria
1:196659237	C/T	CFH	Missense	Risk Factor, Pathogenic	Age-related macular degeneration 4, Basal laminar drusen
2:227892720	C/T	COL4A4	Missense	Likely Pathogenic	not provided
4:187113041	C/G	CYP4V2	Missense	Pathogenic, Benign, Likely Benign	Bietti crystalline corneoretinal dystrophy, not specified, not provided

30 variants Clear All

- Read Depths (DP) is greater than or equal to 50.0 and Read Depths (DP) is not missing for HG00096,HG00097,HG00099, and 97 others
- Affected - Allele Frequencies is greater than or equal to 0.3 or Affected - Allele Frequencies is missing
- Effect (Combined) is either 'LoF' or 'Missense'
- Clinical Significance is either 'Likely Pathogenic' or 'Pathogenic'

- 変異データリスト画面に戻り、フィルタリング結果の確認を行う。変異リストでは、アノテーションなどの表示フィールドの設定も可能。

## Export with Samples Choose export options

**Export Multi Sample File**

Visible columns will be used as INFO Fields. Select sample level FORMAT fields.

**Types**

Excel

TSV

VCF

**Sample FORMAT Fields**

[Select All](#) [Unselect All](#)

Allelic Depths (AD)

Read Depths (DP)

Genotype Qualities (GQ)

0/1 Genotypes (GT)

MIN\_DP

PGT

PID

GT Likelihoods Phred (PL)

SB

Alt Allele Freq

**Select Samples**

Select a cohort

[Select All](#) [Unselect All](#) 100 of 100 selected

Export	Name	Renamed Samples	Sample Source File Name	Affected?	Up Sc
<input checked="" type="checkbox"/>	HG00096	HG00096	2500 Samples 400 Gene	Affected	?

[Export](#)

[Close](#)

- フィルタリング結果の変異データリストは、VCFなどのフォーマットでファイル出力が可能。

## Details

### Primary Findings

**4:55593464 A/C (KIT)**

Classification Pathogenic

Interpretation This is a Missense Variant located in the KIT gene.  
The tyrosine kinase receptor KIT and its ligand, KITLG ([184745](#)), function in hematopoiesis, melanogenesis, and gametogenesis ([Rothschild et al., 2003](#)).  
This gene has been observed to exhibit Autosomal

**7:55242466 GAATTAAGAGAAGCA/- (EGFR)**

Classification Pathogenic

Interpretation This is a Inframe Deletion located in the EGFR gene.  
EGFR and its ligands are cell signaling molecules involved in diverse cellular functions, including cell proliferation, differentiation, motility, and survival, and in tissue development ([Wang et al., 2004](#)).

### Incidental Findings

**17:7578195 CAC/- (TP53)**

Interpretation This is a Inframe Deletion located in the TP53 gene.  
The transcription factor p53 responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in

Exit

- 「Reports」では、アップロードした各サンプルのVSReportに含まれている変異の詳細情報を確認。

The screenshot displays the Filgen Warehouse web interface. The browser address bar shows the URL: [https://warehouse-demo2.goldenhelix.com/reports/6-Filgen/1/Sample1\\_Variants/view/frame/](https://warehouse-demo2.goldenhelix.com/reports/6-Filgen/1/Sample1_Variants/view/frame/). The page title is "Sample1\_Variants" and the user is logged in as "ozawa@filgen.jp".

**Provider Information:**

Filgen, Inc  
1409, Jonoyama 1-Chome, Midori-ku  
Nagoya, Aichi-pref. 459-8011  
Phone: +81-52-624-4388  
Fax: +81-52-624-4389

**Patient Information:**

Name: Sample1\_Variants  
Gender: Male  
Date of Birth: 5/15/2017  
Id: 1234

**Sample Information:**

Sample Site: [Blank]  
Sample Type: [Blank]  
Collection Method: [Blank]  
Panel Coverage: [Blank]  
Avg. Read Depth: [Blank]  
Collection Date: 5/15/2017  
Receipt Date: 5/15/2017  
Report Date: 5/15/2017

**Results:**

**Positive:** Mutations with an establish somatic link detected.

**Affected Genes:**

ABL1 (0)	ALK (0)	ASXL1 (0)	ATRX (0)	BCOR (0)	BCOR1 (0)	BRAF (0)	CALR (0)	CBL (0)	CBLB (0)	CBLG (0)
CDKN2A (0)	CEBPA (0)	CUX1 (0)	<b>EGFR (1)</b>	<b>ERCC2 (1)</b>	JAK2 (0)	JAK3 (0)	KDM6A (0)	<b>KIT (1)</b>	TET2 (0)	<b>TP53 (1)</b>

**Primary Findings:**

Gene	Zygoty	Variant	Exon	Pathogenicity
KIT	?	NM_000222.2:c.1621A>C(NP_000213.1:p.Met541Leu)	10	Pathogenic
EGFR	?	NM_005228.4:c.2236_2250delGAATTAAGAGAAGCA(NP_005219.2:p.Glu746_Ala750del)	19	Pathogenic

- VSReport形式でレポートを閲覧し、HTML形式でファイル出力が可能。

お問い合わせ先：フィルジエン株式会社  
TEL: 052-624-4388 (9:00～17:00)  
FAX: 052-624-4389  
E-mail: [biosupport@filgen.jp](mailto:biosupport@filgen.jp)