全ゲノム／全エクソームシークエンスにおけるCNV解析

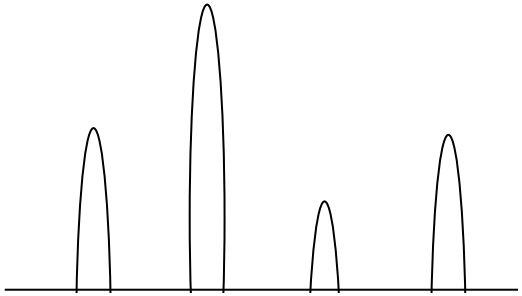
フィルジェン株式会社　バイオインフォマティクス部
(biosupport@filgen.jp)

はじめに

次世代シークエンサー

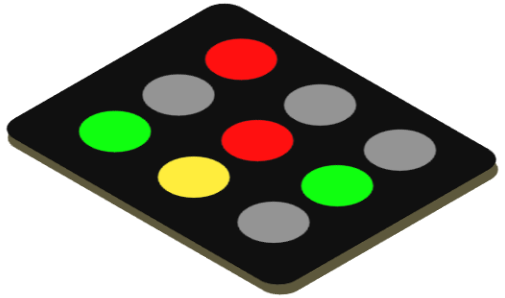
SNV, Small InDel

MLPA



Small CNV

マイクロアレイ



Large CNV

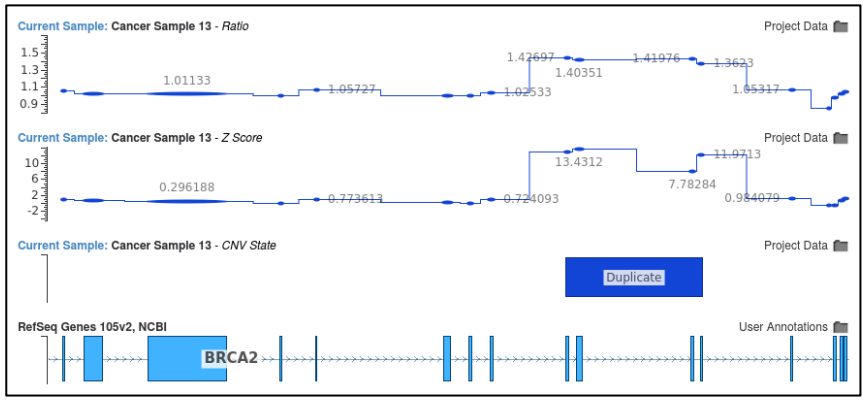
プラットフォームの一本化

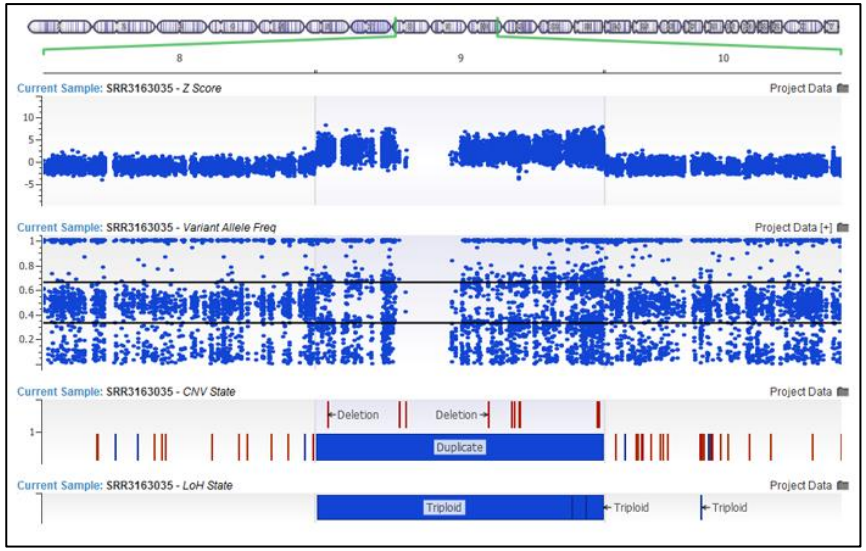
次世代シークエンサー

- ✓ SNV, Small InDel
- ✓ Small CNV
- ✓ Large CNV

- ■ 従来CNVの検出には、サイズに合わせて複数のプラットフォームを使い分ける必要があり、コストや実験手技の習熟などの問題を抱えていた
- ■ すでにSNVの検出には、次世代シークエンサーを用いることがスタンダードになっているが、同じくCNV検出にも使用できるようになれば、プラットフォームの一本化によって、これらの問題をクリアできる

VS-CNV





- BAMファイルのリードカバレッジデータをもとに、CNV検出を行う機能を付加する、VarSeq®用の有償アドオン
- 遺伝子パネルや全エクソーム、全ゲノムシークエンスデータに対応
- 検出されたCNVは、ゲノムブラウザーでグラフィカルに表示
- 検出CNVに対して、公共データベースの情報を用いたアノテーション付けや、ACMGガイドラインに基づいた臨床的意義の評価も可能

他手法との比較

Filgen® biosciences & nanosciences

検出可能なCNV　　データタイプ

	Small: 150bp+	Medium: 1 – 10kb	Large: 10kb+	Gene Panel	Whole Exome	Whole Genome
MLPA	✓			✓		
Microarray			✓			✓
VS-CNV	✓	✓	✓	✓	✓	✓

VS-CNV導入のメリット

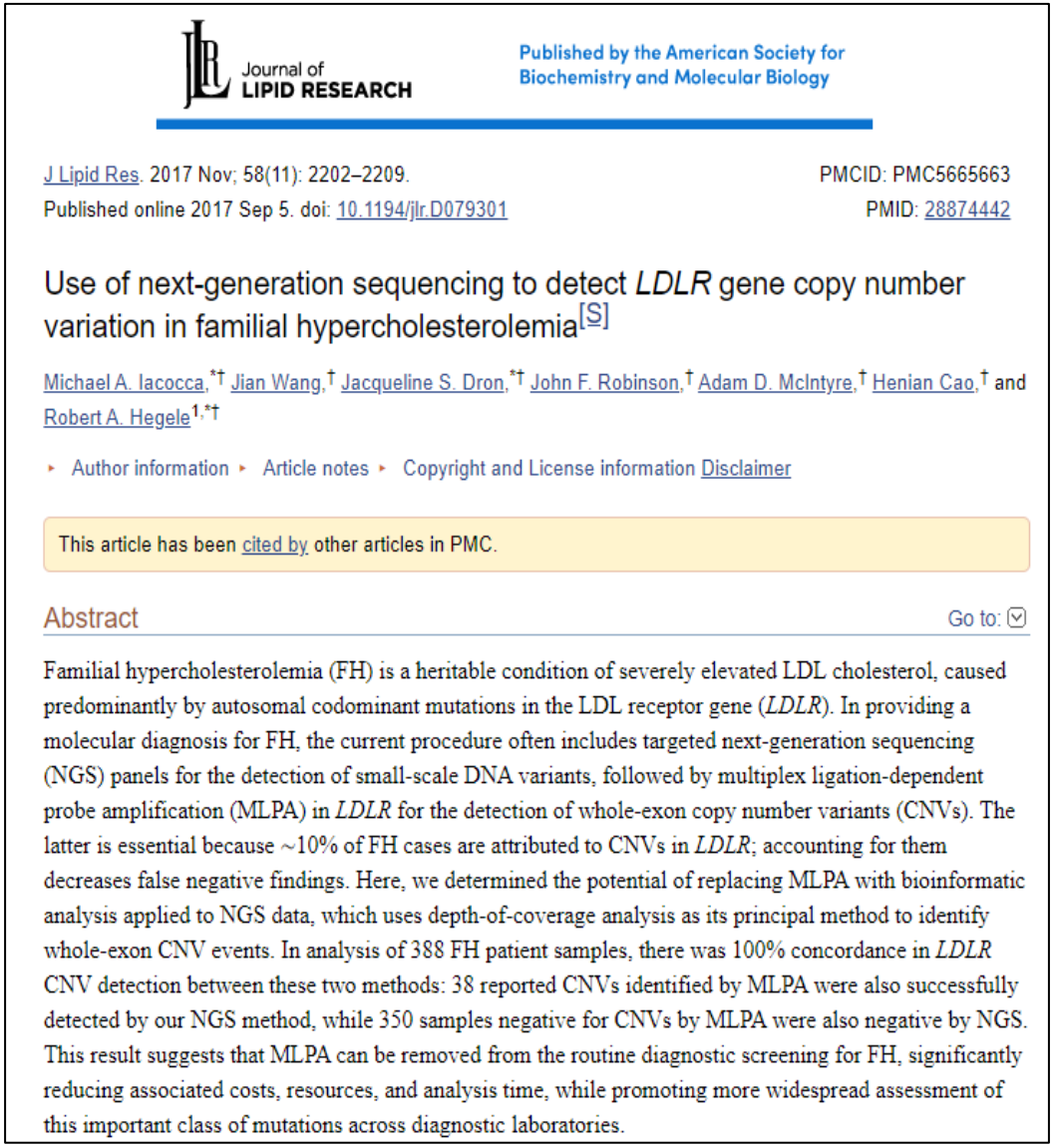
Filgen® biosciences & nanosciences

■ **実績**

- 15報以上の論文実績
- MLPA法との100%の相関
  (Iacocca et al. 2017)

■ **メリット**

- 実験コスト・時間の節約
- ワークフローのシンプル化



Journal of LIPID RESEARCH

Published by the American Society for Biochemistry and Molecular Biology

J Lipid Res. 2017 Nov; 58(11): 2202–2209.
Published online 2017 Sep 5. doi: 10.1194/jlr.D079301

PMCID: PMC5665663
PMID: 28874442

Use of next-generation sequencing to detect *LDLR* gene copy number variation in familial hypercholesterolemia[S]

Michael A. Iacocca,*† Jian Wang,† Jacqueline S. Dron,*† John F. Robinson,† Adam D. McIntyre,† Henian Cao,† and Robert A. Hegele[1,*†]

▸ Author information ▸ Article notes ▸ Copyright and License information Disclaimer

This article has been cited by other articles in PMC.

Abstract

Go to:

Familial hypercholesterolemia (FH) is a heritable condition of severely elevated LDL cholesterol, caused predominantly by autosomal codominant mutations in the LDL receptor gene (*LDLR*). In providing a molecular diagnosis for FH, the current procedure often includes targeted next-generation sequencing (NGS) panels for the detection of small-scale DNA variants, followed by multiplex ligation-dependent probe amplification (MLPA) in *LDLR* for the detection of whole-exon copy number variants (CNVs). The latter is essential because ~10% of FH cases are attributed to CNVs in *LDLR*; accounting for them decreases false negative findings. Here, we determined the potential of replacing MLPA with bioinformatic analysis applied to NGS data, which uses depth-of-coverage analysis as its principal method to identify whole-exon CNV events. In analysis of 388 FH patient samples, there was 100% concordance in *LDLR* CNV detection between these two methods: 38 reported CNVs identified by MLPA were also successfully detected by our NGS method, while 350 samples negative for CNVs by MLPA were also negative by NGS. This result suggests that MLPA can be removed from the routine diagnostic screening for FH, significantly reducing associated costs, resources, and analysis time, while promoting more widespread assessment of this important class of mutations across diagnostic laboratories.

データ解析の流れ

ターゲット法
（遺伝子パネル、全エクソームシークエンス）

ターゲット（エクソンなど）ごとにカバレッジを算出

Region	Name	Mean Depth	Min Depth	Max Depth
3:10183433-10183891	Gene=VHL,RNA=NM_000551.3,Ex=1	33.7582	22	43
3:10188179-10188340	Gene=VHL,RNA=NM_000551.3,Ex=2	12.7963	10	16
3:10191452-10191669	Gene=VHL,RNA=NM_000551.3,Ex=3	37.9587	20	53
3:37034920-37035174	Gene=MLH1,RNA=NM_000249.3,Ex=1	40.5569	26	50
3:37038091-37038220	Gene=MLH1,RNA=NM_000249.3,Ex=2	19.0308	12	22
3:37042427-37042564	Gene=MLH1,RNA=NM_000249.3,Ex=3	15.3986	7	20
3:37045873-37045985	Gene=MLH1,RNA=NM_000249.3,Ex=4	13.7168	9	19
3:37048463-37048574	Gene=MLH1,RNA=NM_000249.3,Ex=5	16.6607	12	21
3:37050286-37050416	Gene=MLH1,RNA=NM_000249.3,Ex=6	15.4275	8	20

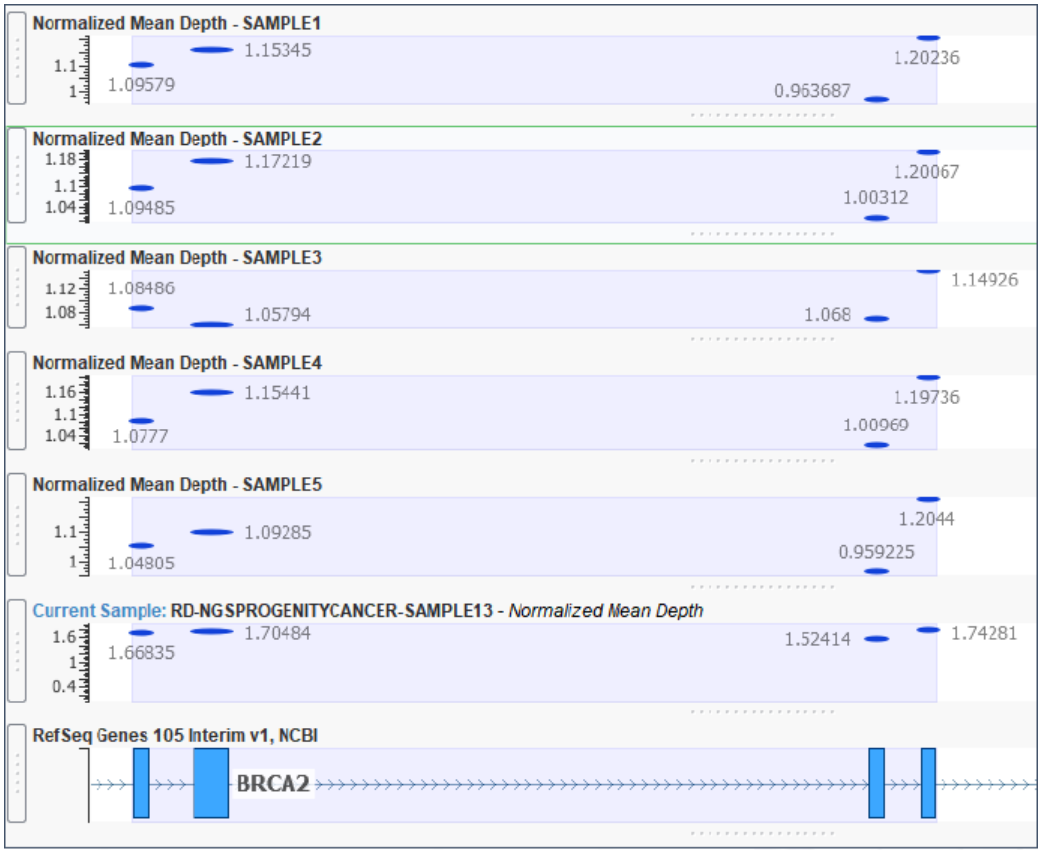
ビン法
（全ゲノムシークエンス）

任意指定のビンサイズごとにカバレッジを算出

Region	Mean Depth	Mean Forward Depth	Mean Reverse Depth
1:1-1000000	0.0155643	0.00798095	0.00758333
1:1000001-2000000	0.010238	0.005363	0.004875
1:2000001-3000000	0.0208937	0.0104758	0.0104179
1:3000001-4000000	0.0146835	0.00723882	0.00744471
1:4000001-5000000	0.035065	0.017518	0.017547
1:5000001-6000000	0.047041	0.023451	0.02359
1:6000001-7000000	0.026751	0.013479	0.013272
1:7000001-8000000	0.05058	0.024961	0.025619

CNVコール

カバレッジの正規化



✓ リファレンス用サンプル群の正規化済み
カバレッジの平均値

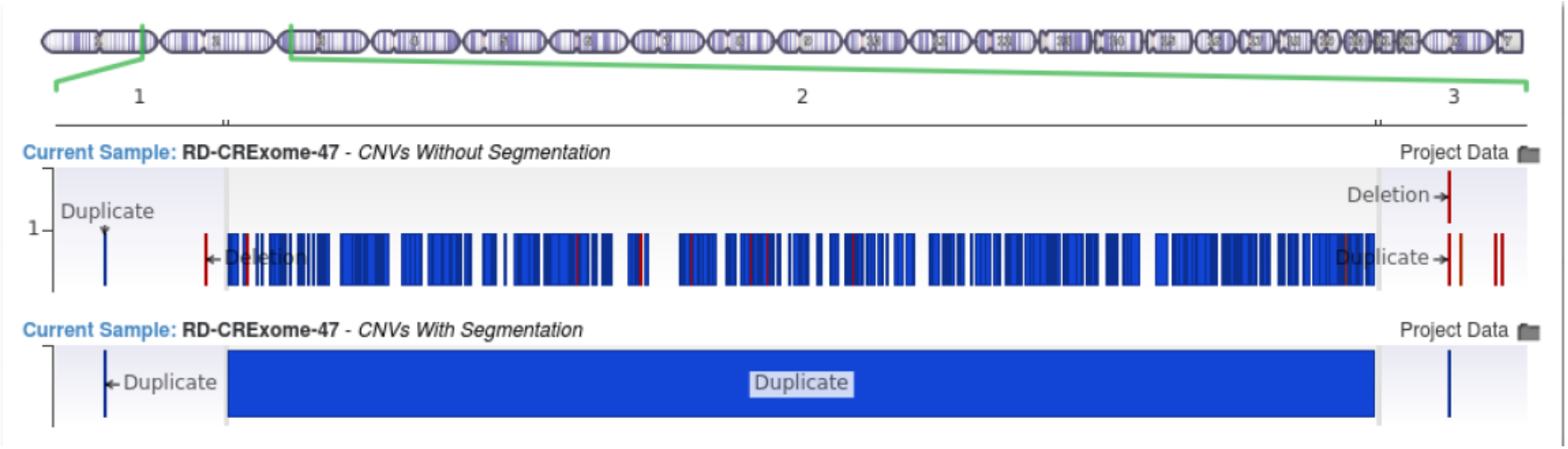
✓ 解析対象サンプルの正規化済み
カバレッジ

✓ CNV検出結果

Coverage Region I...	Overlapp...	RD-NGSP...	Target Copy Number State for RD-NGSPROGENITYCANCER-SAMPLE13		
Region	Gene Names	Mean Depth	CNV State	Normalized Mean Depth	Avg. Normalized Control Depth
13:32936641-32936850	BRCA2	810.281	Duplicate	1.66835	1.17383
13:32937297-32937690	BRCA2	828.008	Duplicate	1.70484	1.20154
13:32944520-32944714	BRCA2	740.241	Duplicate	1.52414	1.02256
13:32945074-32945257	BRCA2	846.446	Duplicate	1.74281	1.22956

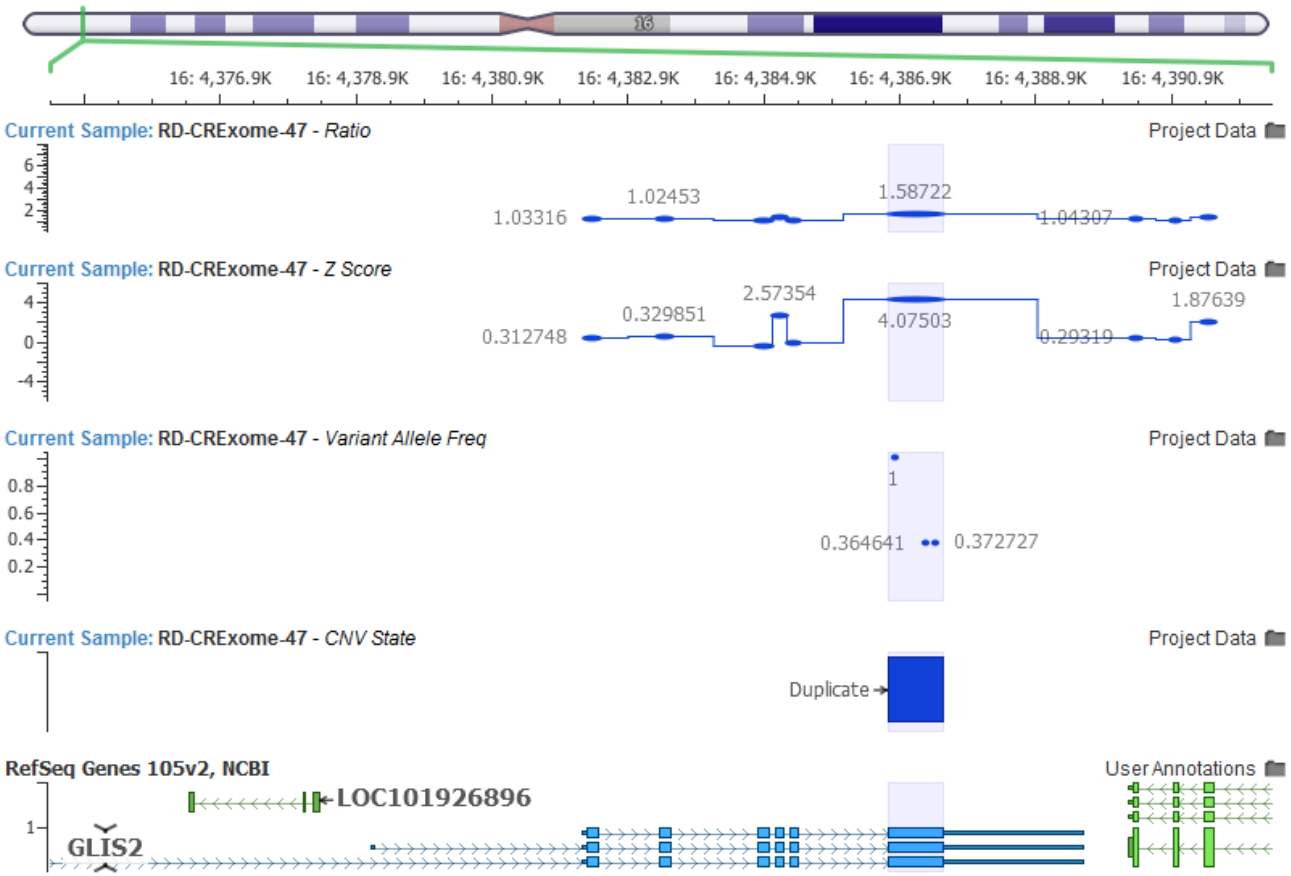
セグメンテーション

- ゲノム上で、連続するターゲットまたはビンでCNVが検出された場合は、それらを連結させて一つのCNVとする
- 外れ値などのノイズ除去のため、Circular Binary Segmentation (CBS) または CNAM Optimal Segmentationアルゴリズムを使用



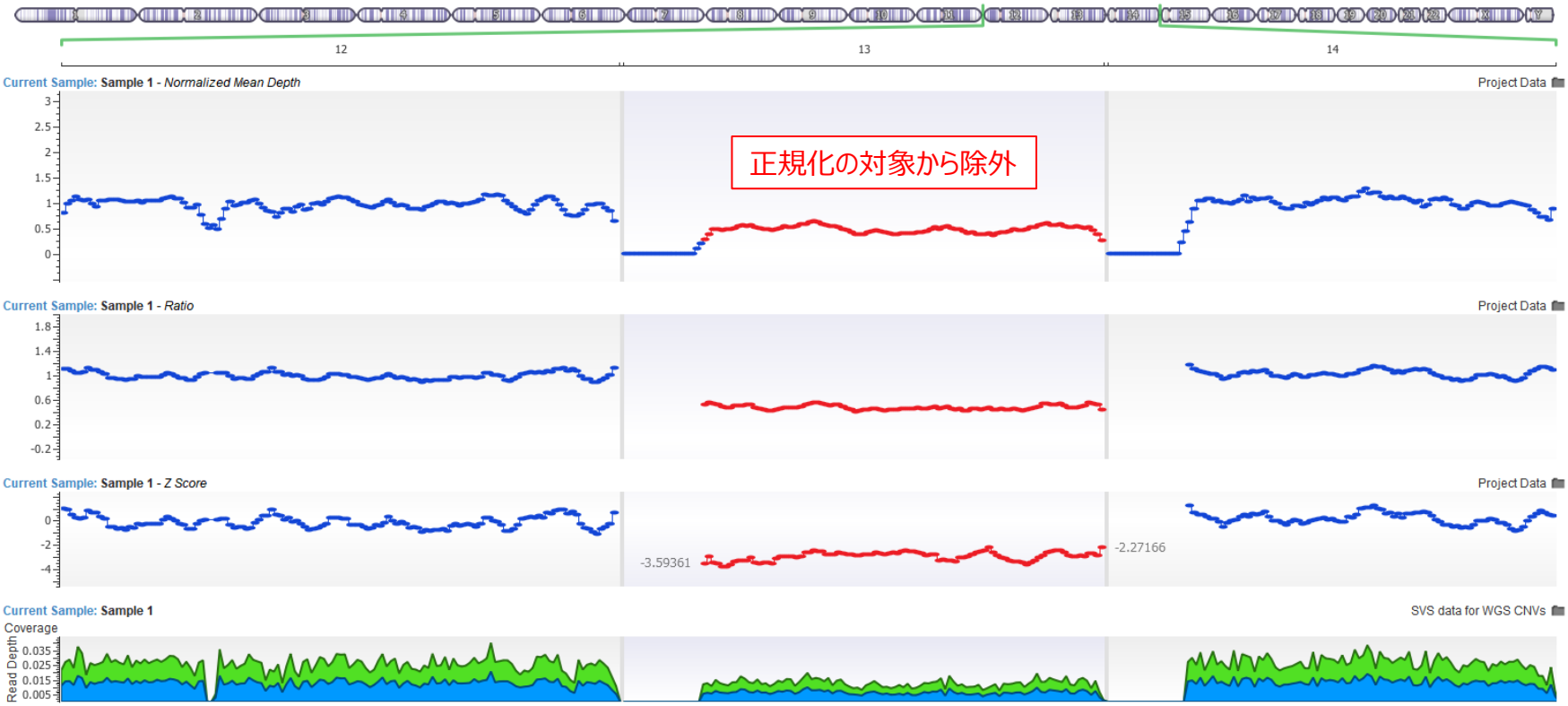
VAFによる検証

- 検出されたCNV上に存在する、SNVなどのVAF（Variant Allele Frequency）の値に基づき、CNVの信頼性を自動で検証
- VAFの値が1/3、2/3になっていれば、Duplicationが起こっている根拠とされる



LOH補正

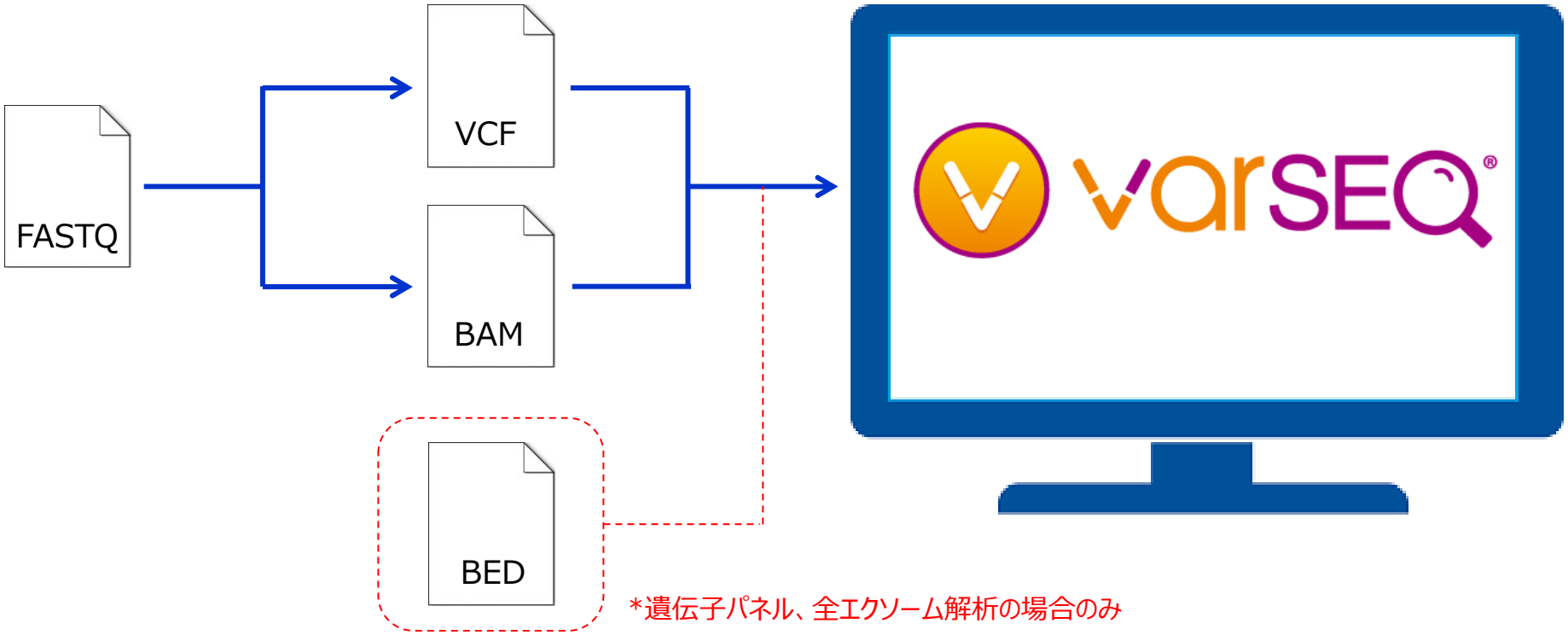
- 染色体上に、大規模なDeletionまたはDuplicationが存在する場合、サンプルの平均カバレッジが影響を受ける
- VAFの値を基にLOH (Loss of Heterozygosity) を自動で検出し、染色体上の2倍体以外の領域は、カバレッジの正規化の計算対象から除外する



必要なデータファイル

■ **必要ファイル**

- VCFファイル
- BAMファイル
- BEDファイル（オプション） ＊一部のパネルのBEDファイル（ターゲット領域データ）は、ソフトウェア上からダウンロードも可能



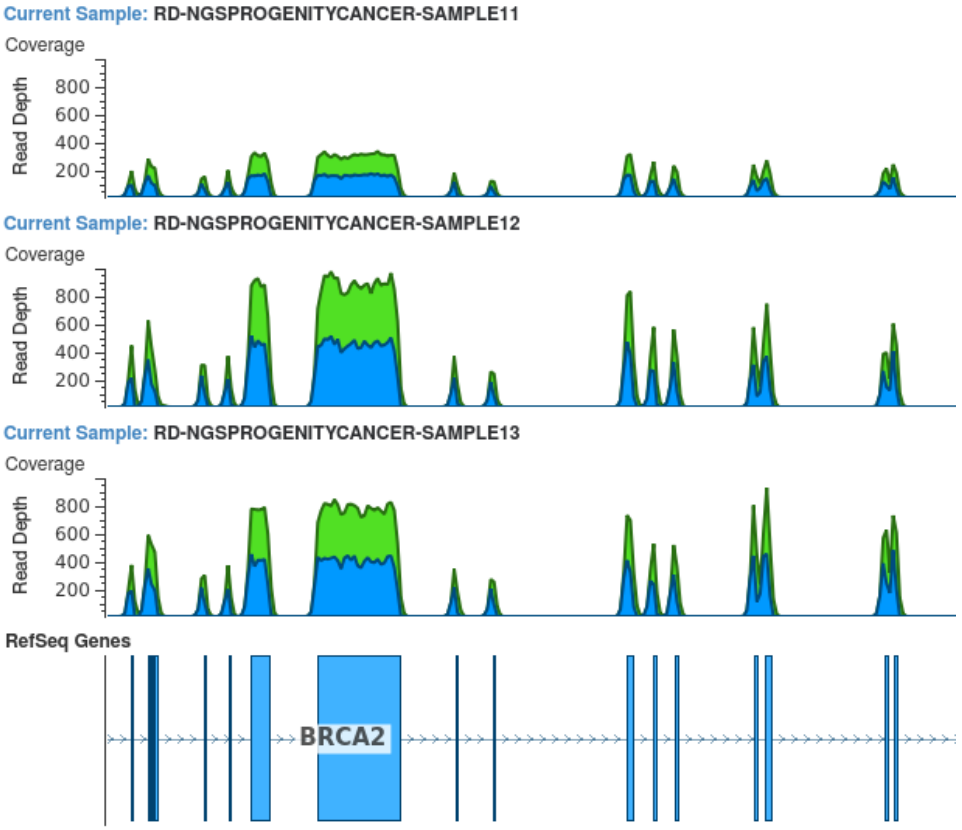
データとサンプルの条件

■ **カバレッジ（リード深度）**

- 遺伝子パネル、全エクソームシークエンスの場合は、100x 以上が必要
- 全ゲノムシークエンスの場合は、0.02x 以上が必要（ビンサイズ：1Mbpの場合）

■ **サンプル**

- 30以上のリファレンス用サンプル
- 同一の実験条件（パネルの種類、サンプル調整、シークエンスデータ量など）



出力データ

■ ターゲット（ビン）ごと

- Ratio: 解析サンプルとリファレンスサンプルのカバレッジ比

- Z-score: 解析サンプルとリファレンスサンプルのカバレッジの差を、標準偏差で割った値

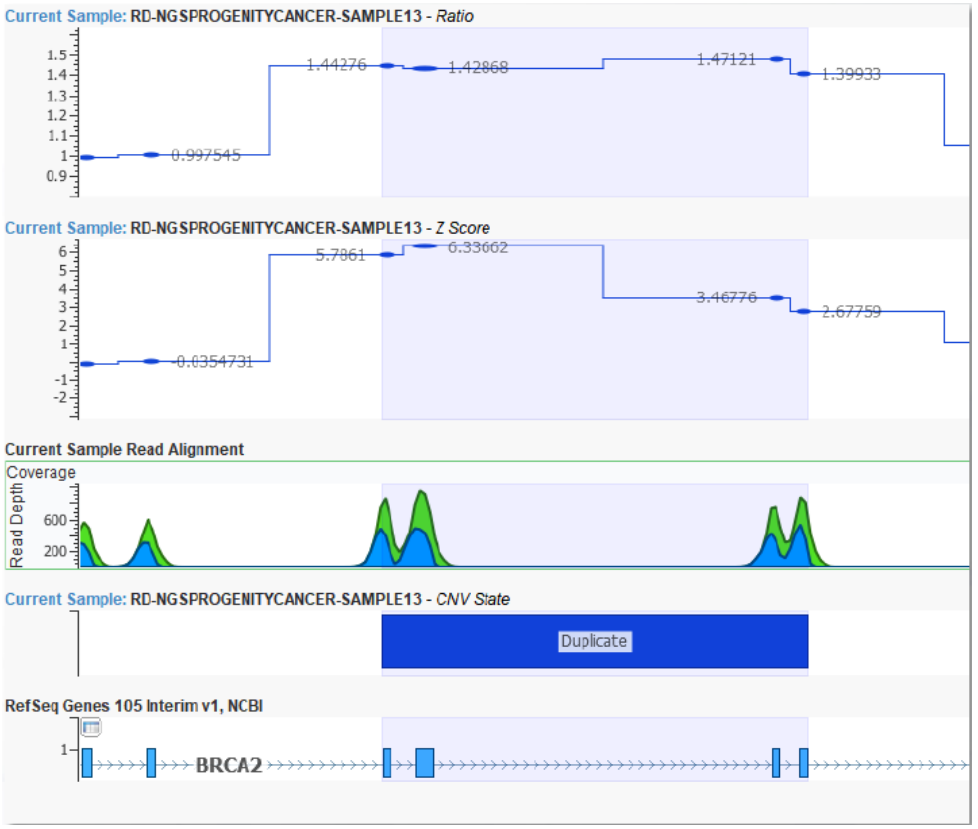
■ CNVごと

- P-value: 検出されたCNVの信頼値

- Flag: 検出されたCNVのQC情報（低カバレッジ、極端なGC含量など）

■ サンプルごと

- Flag: サンプルごとのQC情報（低カバレッジ、リファレンスサンプルとのミスマッチなど）



ビジュアライゼーション

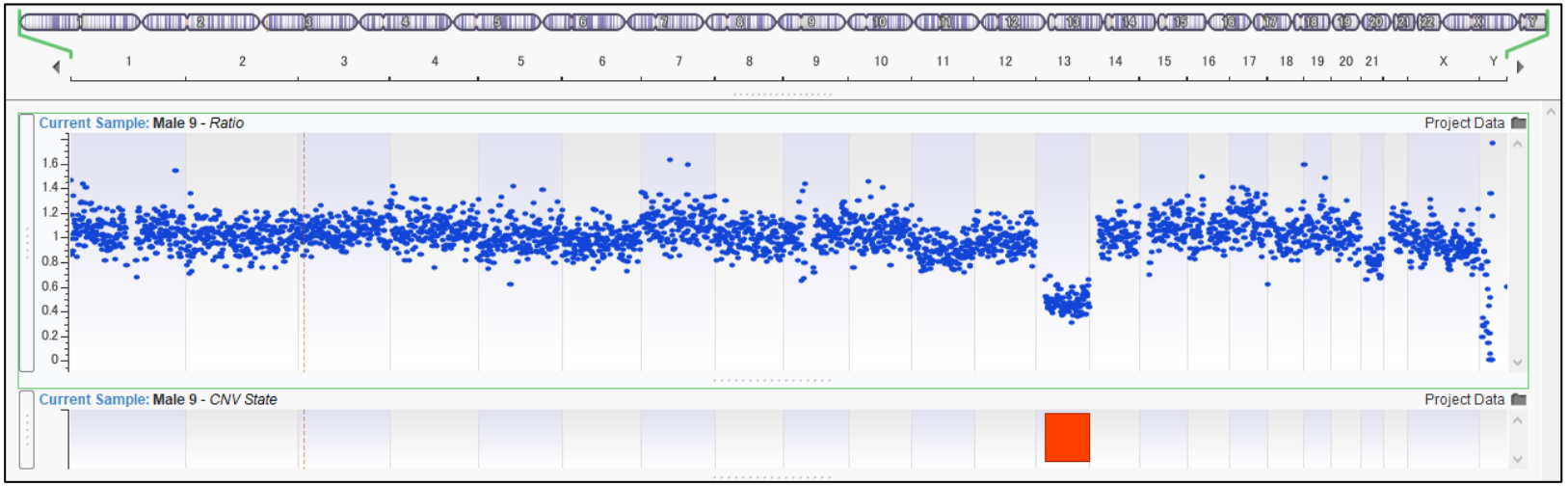
- 各出力データは、ワンクリックでゲノムブラウザーにプロットが可能

Coverage Region Info	Bin Copy Number State for Male 9		
Region	CNV State	Z Score	Ratio
1:1-1000000	Diploid	-0.533545	0.860434
1:1000001-2000000	Diploid	1.50802	1.46392
1:2000001-3000000	Diploid	0.623491	1.17986
1:3000001-4000000	Diploid	-0.573874	0.853885
1:4000001-5000000	Diploid	-0.133061	0.980613
1:5000001-6000000	Diploid	-0.774154	0.88389

Ratioと検出CNVをプロット

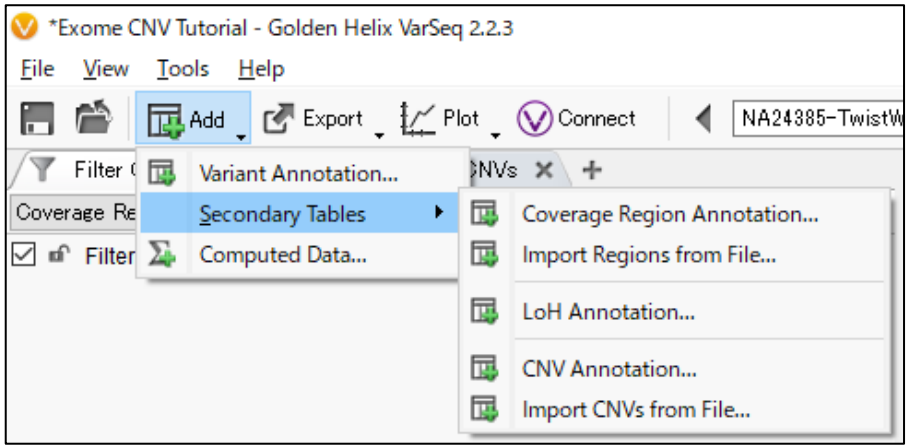
+

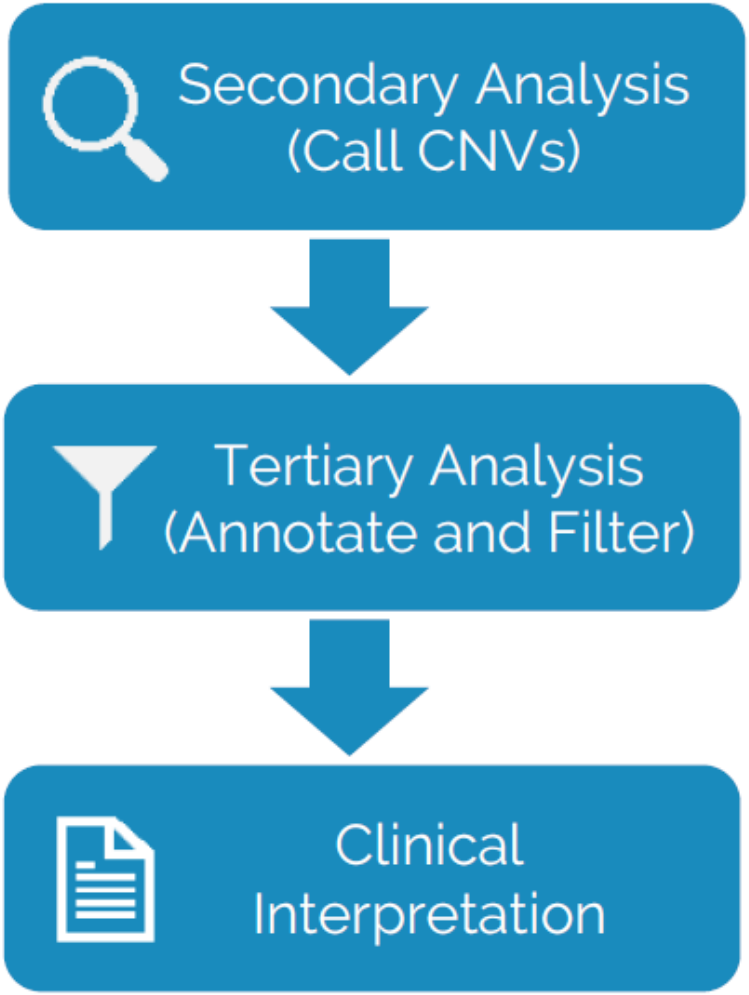
CNV Info				Male 9					
Region	Type	# Targets	Span	CNV State	Avg Target Mean Depth	Avg Z Score	Avg Ratio	Karyotype	p-value
13:20000001-115000000	Loss	95	95000000	Het Deletion	0.0125498	-2.96658	0.471608	45,XX,-13	1.23984244499035e-135



臨床的意義の確認

- VarSeqでは、VS-CNVによるCNVコールだけではなく、検出された各CNVへのアノテーション付けによる、臨床的意義の確認も可能
  - CNVデータベース
  - 表現型オントロジー
  - ACMG & ClinGenガイドライン（VSClinical）





CNVデータベース

■ ソフトウェア搭載のデータソースライブラリーよりダウンロードが可能

■ **アレル頻度データベース**

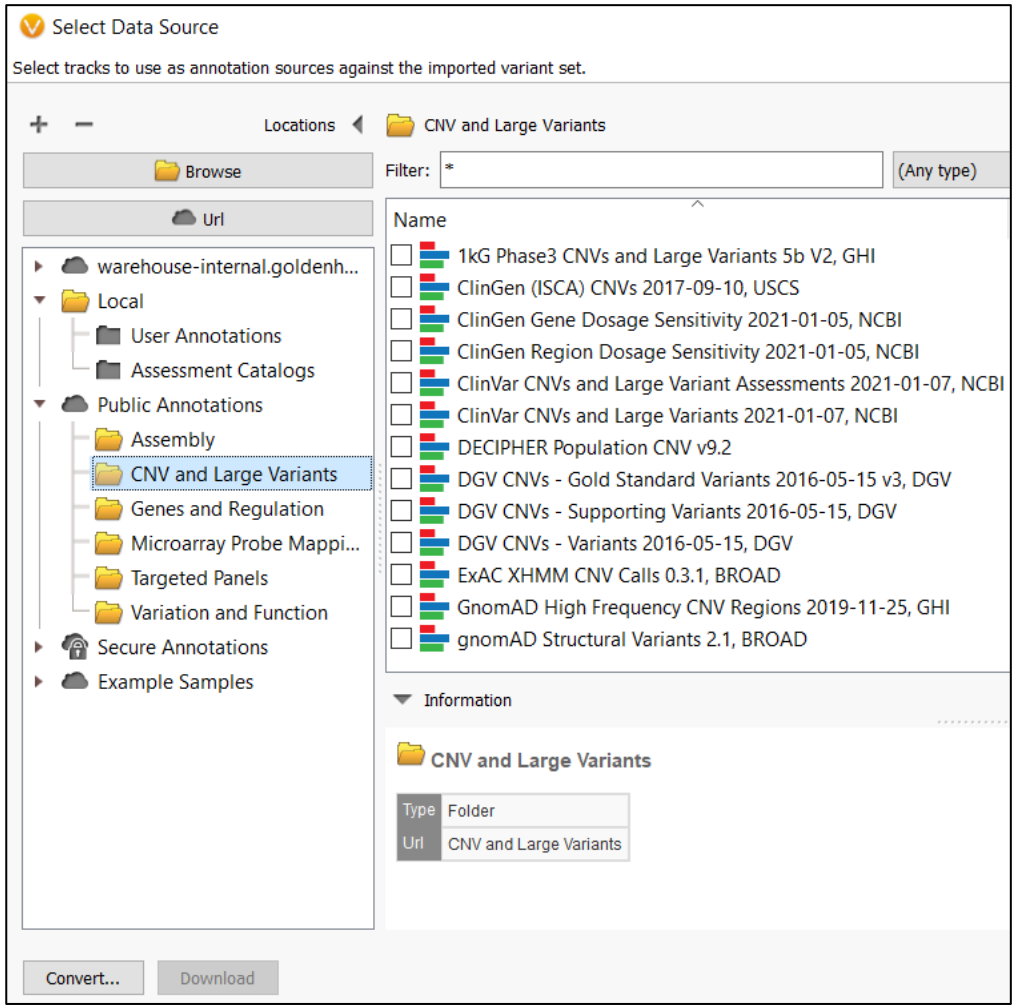
- 1000 Genomes Phase 3 / ExAC
- DGV / gnomAD

■ **臨床情報データベース**

- ClinVar

■ **用量感受性遺伝子データベース**

- ClinGen



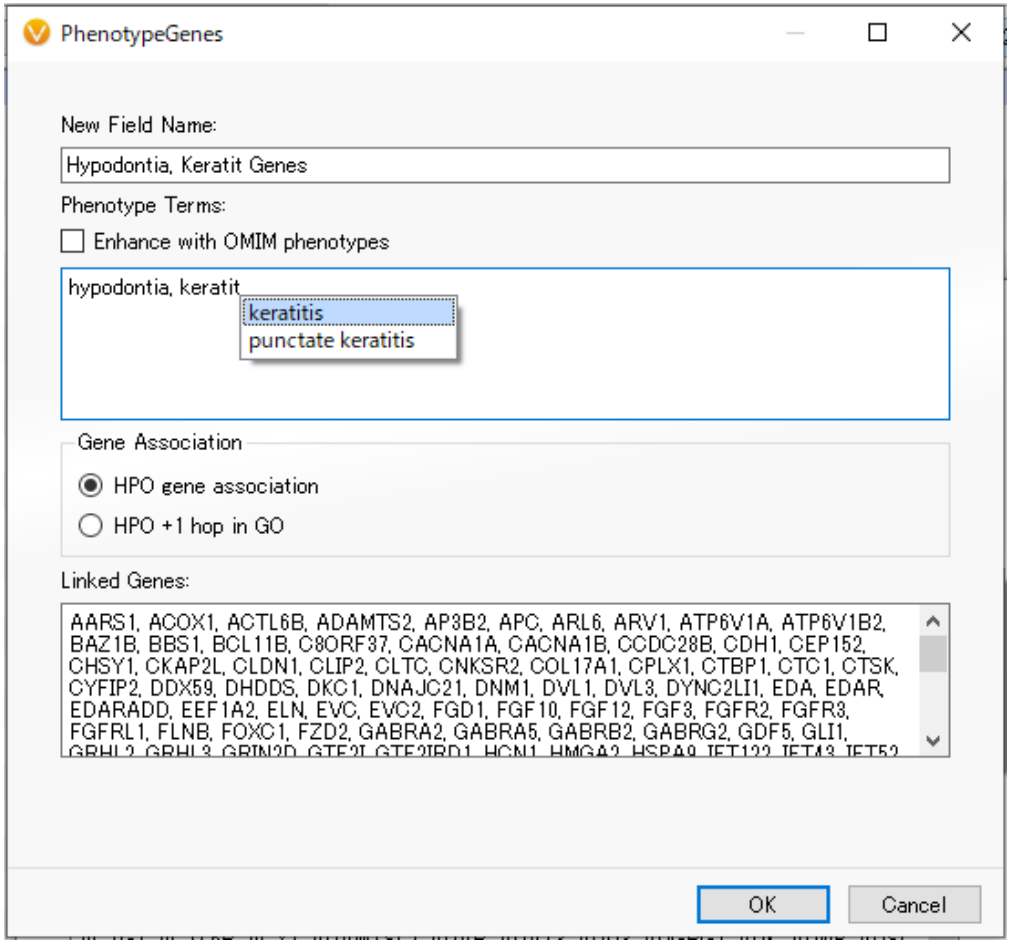
CNVデータベース

Overlapping CNVs ClinVar CNVs and Large Variants 2021-08-05, NCBI						Overlapping CNVs GnomAD High Frequency CNV Regions 2019-11-25, GHI		
Region	dbVar ID	Classification	Review Status	Gene Names	Conditions	Overlap Type	Average Alt Allele Counts (AC)	Average Alt Allele Freq (AF)
16:89814557-8981818...	nssv1415435,ns...	Uncertain Significance,Un...	(0 Stars) No Assertion Crite...	FANCA,LOC...	See cases,See cases	?	?	?
17:7668436-7676594	?	Uncertain Significance	(1 Stars) Criteria Provided, ...	TP53	Li-Fraumeni syndrome	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	Within Region	8834	0.407232
17:19020809-1919673...	?,?,nssv581965,...	Benign,Benign,Benign,Lik...	(0 Stars) No Assertion Crite...	GRAP,GRAP...	not provided,not provid...	Partial Overlap	3673	0.169337
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	Within Region	12253	0.564826
17:46061148-4666196...	nssv578359,nsv...	Benign,Likely Benign,Beni...	(1 Stars) Criteria Provided, ...	NSF,LRRC37...	See cases,See cases,See ...	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
19:54222913-5424045...	nssv13654532,n...	Benign,Benign	(0 Stars) No Assertion Crite...	LILRB3,LILR...	See cases,See cases	?	?	?
?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?
20:1570841-1600450	?	Benign	(0 Stars) No Assertion Crite...	SIRPB1	not provided	Within Region	17864	0.823469

■ データベースの情報を付加すると、検出結果のCNVテーブルに、各アノテーションのフィールドが追加される

表現型オントロジー

- ユーザー指定の表現型オントロジー (HPO) に基づき、表現型と関連する遺伝子を検索するツールを搭載
- 検索結果の遺伝子に対して、単純にアノテーション付けを行う場合と、関連の強さに応じたランク付けを行う2種類の手法を選択可能
- **Match Genes Linked to Phenotype**
  - ユーザー入力の表現型オントロジーと関連する、すべての遺伝子にアノテーション付けを行う
- **CNV PhoRank Gene Ranking**
  - ユーザー入力の表現型オントロジーとの関連の強さに応じて、遺伝子ごとのランク付けを行う



表現型オントロジー

Filgen® biosciences & nanosciences

■ Match Genes Linked to Phenotype

Match Genes Linked to Phenotypes	
Global Developmental Delay Genes	Match
True	NPM1
True	RAC1
False	?
False	?
False	?
True	FANCG
False	?
False	?
True	SET
True	PTEN
True	SUFU
False	?
False	?

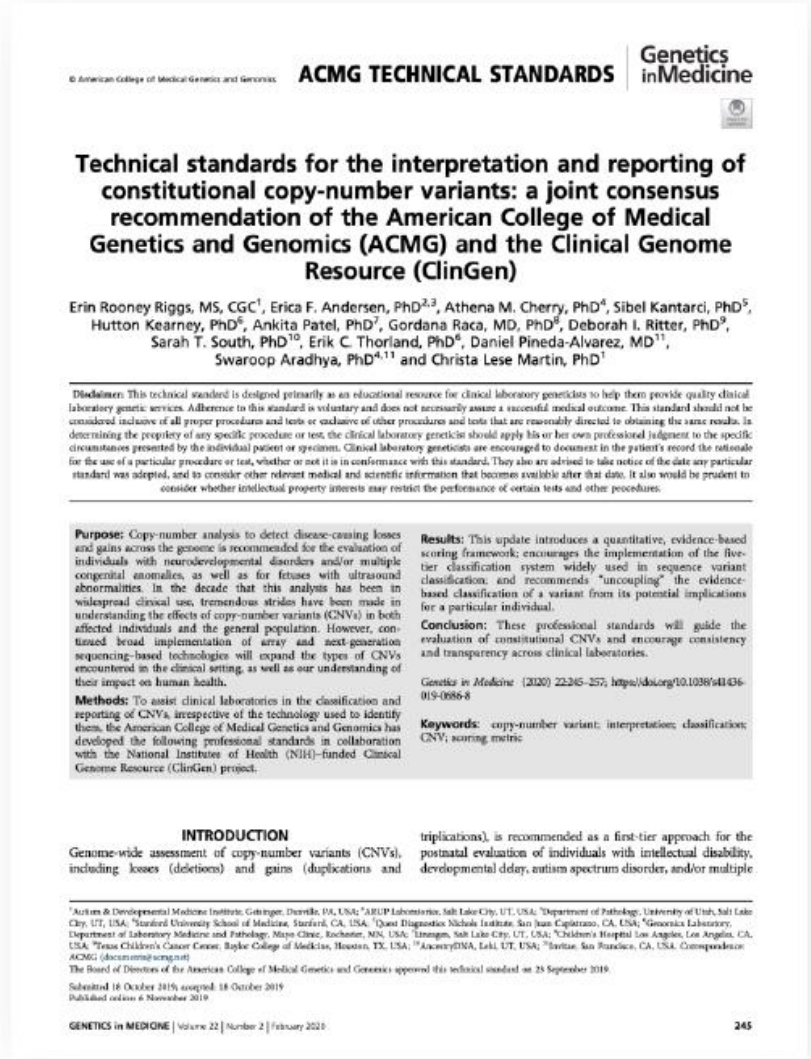
■ CNV PhoRank Gene Ranking

Global Developmental Delay PhoRank					
Sum of Scores	Max Score	Gene Name	Gene Rank	Gene Score	Paths
1	1	NPM1	1	0.0512553	NPM1 / HP:0...
1	1	RAC1	1	0.0512553	RAC1 / HP:0...
0.632099	0.632099	ADAM32	0.632099	0.000453124	ADAM32 / G...
0.501235	0.501235	LRRC69	0.501235	0.000135725	LRRC69 / GO...
0.639506	0.639506	PTK2	0.639506	0.000474464	PTK2 / GO:00...
1	1	FANCG	1	0.0512553	FANCG / HP:...
1.2642	0.632099	CCDC107,SIT1	0.632099,0.63...	0.000453124,...	CCDC107 / G...
0.62716	0.62716	CBWD5	0.62716	0.000405753	CBWD5 / GO...
1.84691	1	SET,DYNC2I2	1,0.846914	0.0512553,0.0...	SET / HP:000...
1	1	PTEN	1	0.0512553	PTEN / HP:0...
1	1	SUFU	1	0.0512553	SUFU / HP:0...
0.590123	0.590123	SPTY2D1OS	0.590123	0.000252965	SPTY2D1OS /...
0.622222	0.622222	SELENOH	0.622222	0.000378687	SELENOH / ...

- Match Genes Linked to Phenotypeでは、CNVとオーバーラップする遺伝子に対して、入力した表現型オントロジーとの関連を、TRUEとFALSEの2値で判別したアノテーションを、CNVテーブルに付加
- CNV PhoRank Gene Rankingでは、遺伝子ごとの表現型との関連の強さをスコア化し、ランク付けしたデータをCNVテーブルに付加

ACMG & ClinGenガイドライン

- ACMGとClinGenのガイドラインに基づいた、生殖細胞系列CNVの評価用ツール
- CNVごとの病原性評価（Pathogenic, Benignなど）を自動で実行し、評価結果をCNVテーブルにアノテーション付け
- ダッシュボード画面で、評価の手動での調整や、レポート作成も可能
- 使用するには、別途VSClinicalライセンスが必要



ACMG TECHNICAL STANDARDS | Genetics inMedicine

Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen)

Erin Rooney Riggs, MS, CGC[1], Erica F. Andersen, PhD[2,3], Athena M. Cherry, PhD[4], Sibel Kantarci, PhD[5], Hutton Kearney, PhD[6], Ankita Patel, PhD[7], Gordana Raca, MD, PhD[8], Deborah I. Ritter, PhD[9], Sarah T. South, PhD[10], Erik C. Thorland, PhD[6], Daniel Pineda-Alvarez, MD[11], Swaroop Aradhya, PhD[4,11] and Christa Lese Martin, PhD[1]

Purpose: Copy-number analysis to detect disease-causing losses and gains across the genome is recommended for the evaluation of individuals with neurodevelopmental disorders and/or multiple congenital anomalies, as well as for fetuses with ultrasound abnormalities. In the decade that this analysis has been in widespread clinical use, tremendous strides have been made in understanding the effects of copy-number variants (CNVs) in both affected individuals and the general population. However, continued broad implementation of array and next-generation sequencing–based technologies will expand the types of CNVs encountered in the clinical setting, as well as our understanding of their impact on human health.

Methods: To assist clinical laboratories in the classification and reporting of CNVs, irrespective of the technology used to identify them, the American College of Medical Genetics and Genomics has developed the following professional standards in collaboration with the National Institutes of Health (NIH)–funded Clinical Genome Resource (ClinGen) project.

Results: This update introduces a quantitative, evidence-based scoring framework; encourages the implementation of the five-tier classification system widely used in sequence variant classification; and recommends "uncoupling" the evidence-based classification of a variant from its potential implications for a particular individual.

Conclusion: These professional standards will guide the evaluation of constitutional CNVs and encourage consistency and transparency across clinical laboratories.

Genetics in Medicine (2020) 22:245–257; https://doi.org/10.1038/s41436-019-0686-8

Keywords: copy-number variant; interpretation; classification; CNV; scoring metric

INTRODUCTION

Genome-wide assessment of copy-number variants (CNVs), including losses (deletions) and gains (duplications and triplications), is recommended as a first-tier approach for the postnatal evaluation of individuals with intellectual disability, developmental delay, autism spectrum disorder, and/or multiple

[1]Autism & Developmental Medicine Institute, Geisinger, Danville, PA, USA; [2]ARUP Laboratories, Salt Lake City, UT, USA; [3]Department of Pathology, University of Utah, Salt Lake City, UT, USA; [4]Stanford University School of Medicine, Stanford, CA, USA; [5]Quest Diagnostics Nichols Institute, San Juan Capistrano, CA, USA; [6]Genomics Laboratory, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; [7]Invitae, Salt Lake City, UT, USA; [8]Children's Hospital Los Angeles, Los Angeles, CA, USA; [9]Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX, USA; [10]AncestryDNA, Lehi, UT, USA; [11]Invitae, San Francisco, CA, USA. Correspondence: ACMG (documents@acmg.net)

The Board of Directors of the American College of Medical Genetics and Genomics approved this technical standard on 25 September 2019.

Submitted 18 October 2019; accepted: 18 October 2019
Published online 6 November 2019

Riggs ER, Andersen EF, Cherry AM, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med*. 2020;22(2):245-257. doi:10.1038/s41436-019-0686-8
https://clinicalgenome.org/

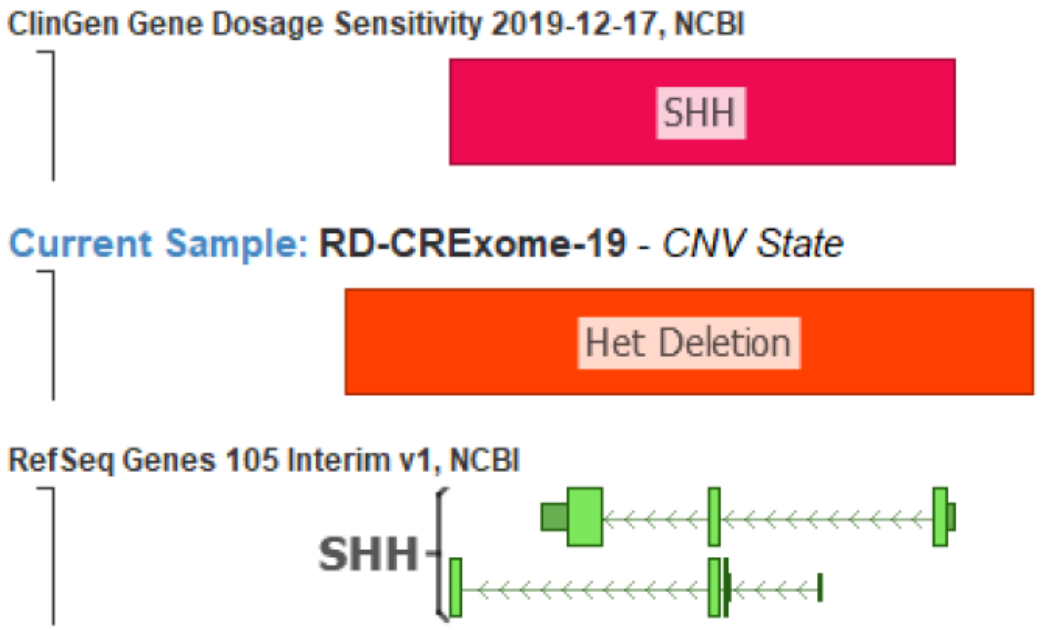
ACMG & ClinGenガイドライン

■ 評価基準

- ✓ CNVがオーバーラップしている遺伝子上の領域
- ✓ オーバーラップしている遺伝子数
- ✓ オーバーラップしている遺伝子の用量感受性（ハプロ不全など）
- ✓ 既知の疾患と遺伝子の関連情報、CNVの遺伝様式

…など

CNVごとに病原性を自動で評価



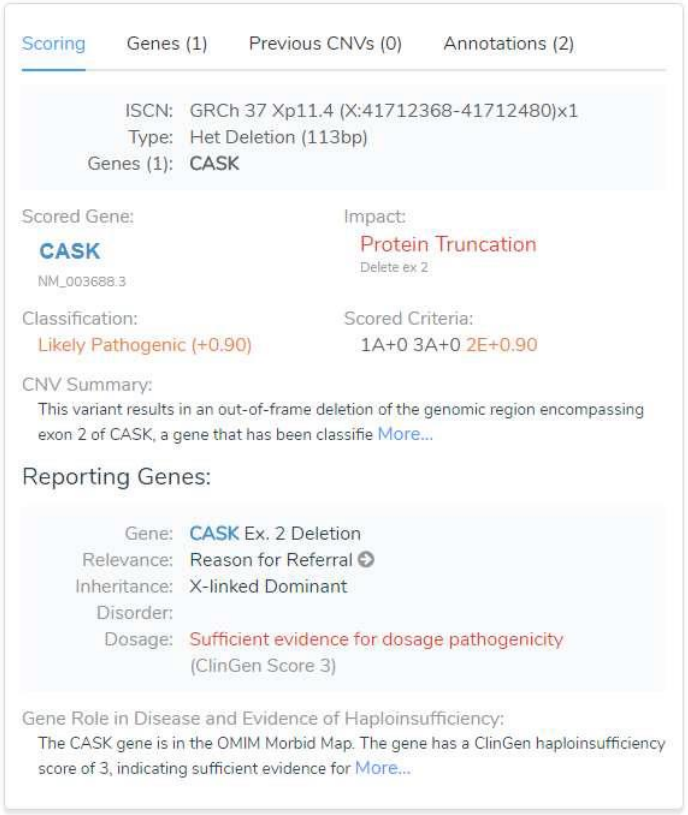
ACMG & ClinGenガイドライン

■ **CNVテーブルにアノテーション付けを行った場合**

ACMG Sample CNV Classifier for NA24385-TwistWE_v1-200826_A00726_0168_AHNKJCDRXX

Gene List	Critical Gene List	Scored Gene	Scored Gene Transcript	Scored Gene Impact Score	Additional Score	Total Score	Classification	Criteria
IKBKG	IKBKG	IKBKG	NM_003639.4	0.9	0	0.9	Likely Pathogenic	2E:0.9 Both b...
F8	F8	F8	NM_000132.4	-0.6	0	-0.6	VUS	2B:0 The cnv ...
F8,H2AB1	F8	F8	NM_000132.4	0	0	0	VUS	2B:0 The cnv ...
F8	F8	F8	NM_000132.4	-0.6	0	-0.6	VUS	2B:0 The cnv ...
TMLHE	?	?	?	?	-1	-1	Benign	1A:0 The CN...

■ **バリアントダッシュボードで評価を行った場合**

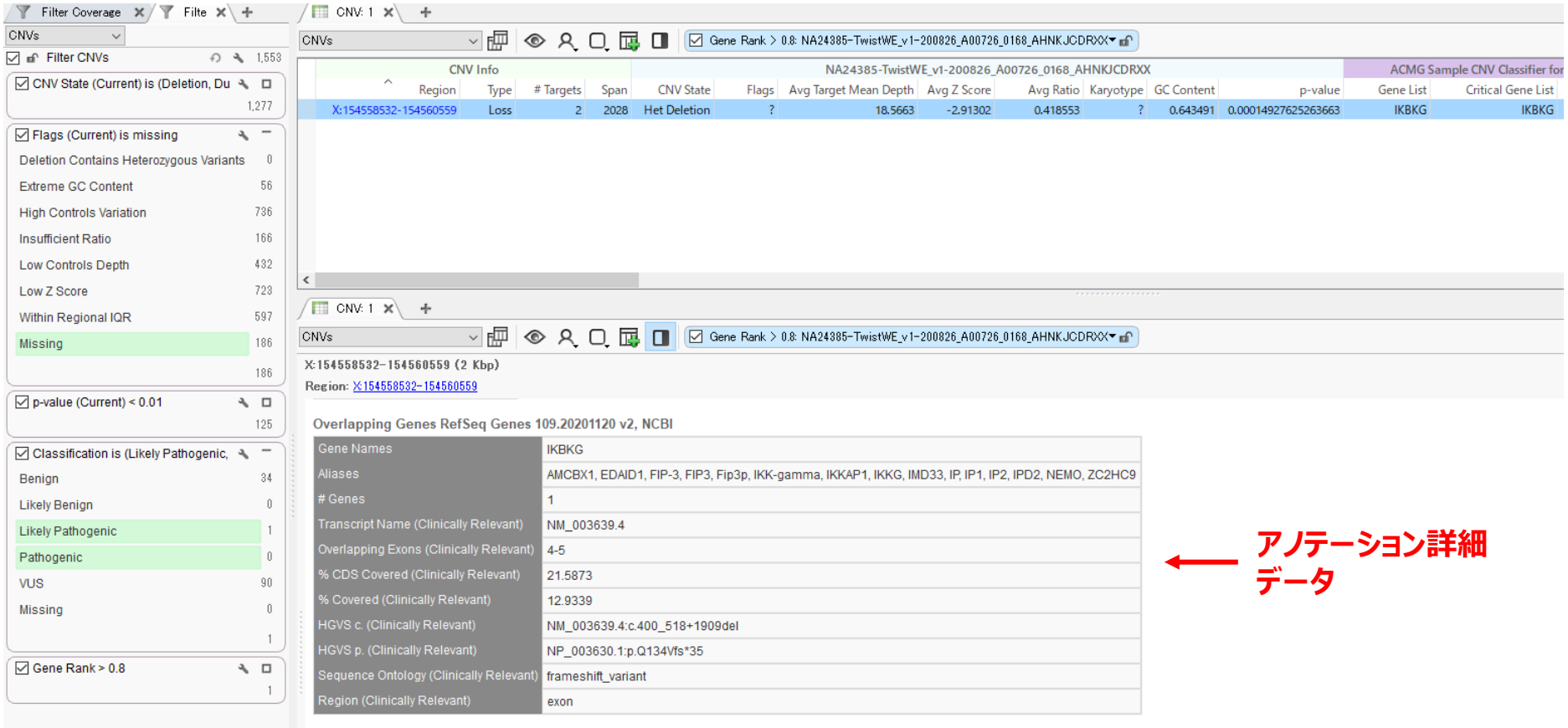


- CNVテーブル上の全CNVには、評価結果の各種データがアノテーションとして追加される
- バリアントダッシュボードでは、評価結果を手動で調整することや、評価に関する説明や根拠など、詳細情報を確認することが可能

フィルタリングとアノテーションの確認

フィルタリングワークフロー

CNVテーブル



アノテーション詳細データ

- CNVテーブルにて、各データ（Flag, P-value, アノテーションなど）に基づきフィルタリングを実行
- 各CNVに付加されたアノテーションは、詳細画面から全アノテーションの詳細データを確認することも可能

お問い合わせ先：フィルジェン株式会社

TEL: 052-624-4388 (9:00～18：00)

FAX: 052-624-4389

E-mail: biosupport@filgen.jp