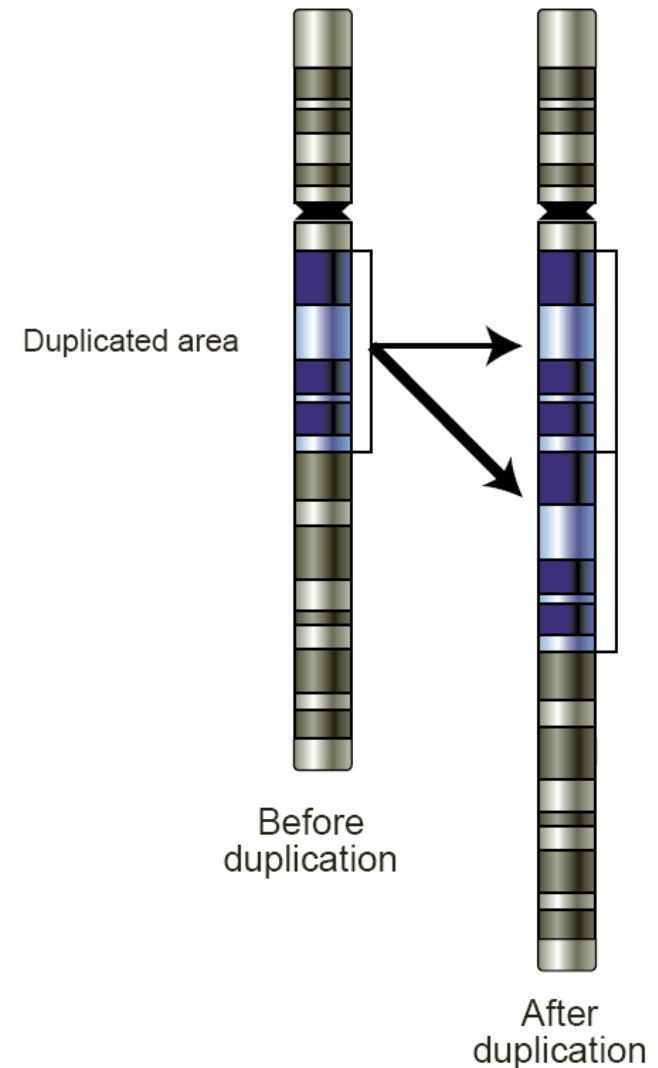


NGSデータを用いたコピー数多型 (CNV) 解析

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

- ゲノムワイドのCNV解析には、従来からマイクロアレイが使用されているが、基本的にサイズの大きい領域の検出に制限される。
- SNPや小サイズのInsertion/Deletionのような、ゲノム上の変異の検出については、すでに次世代シーケンサーを使用した疾患遺伝子パネル解析、全エクソーム解析や全ゲノム解析が主流になっている。
- Golden Helix社VarSeq[®]のCNV Caller機能では、このような次世代シーケンサーの変異解析パイプラインのデータをそのまま用いて、1エクソンの小サイズから、1遺伝子あるいは染色体全体のような大サイズのCNVを検出することができる。



SNP & Variation Suite



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

VarSeq[®]

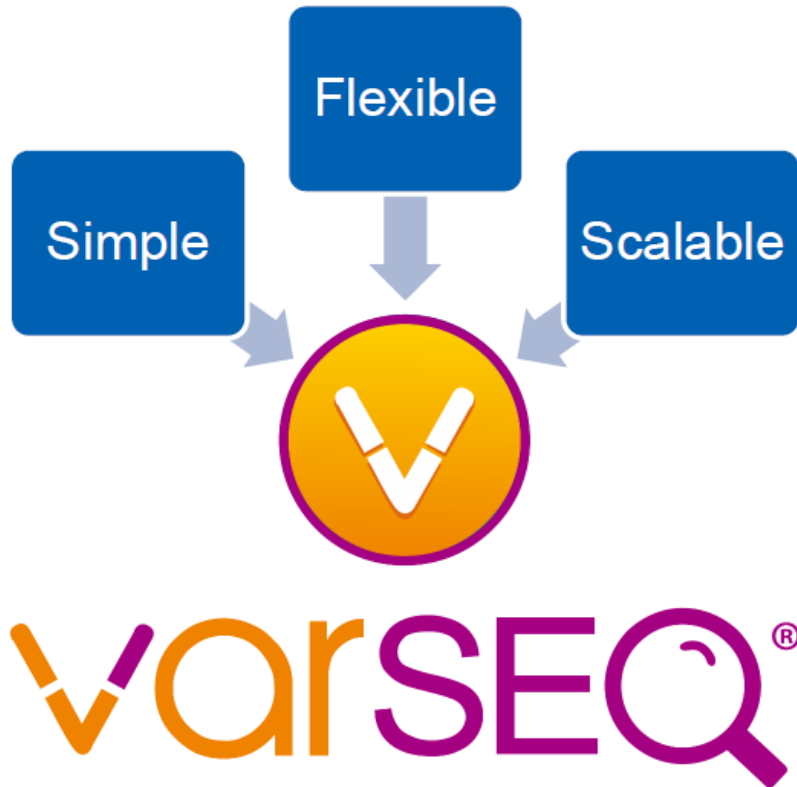


- Variant Interpretation
- Cancer Diagnostics
- CNV Calling
- Clinical Reporting
- High-throughput NGS Testing

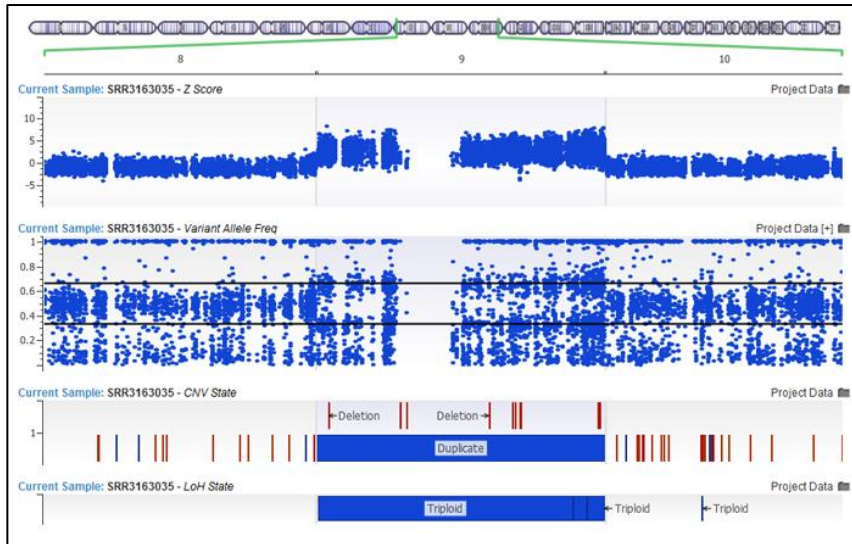
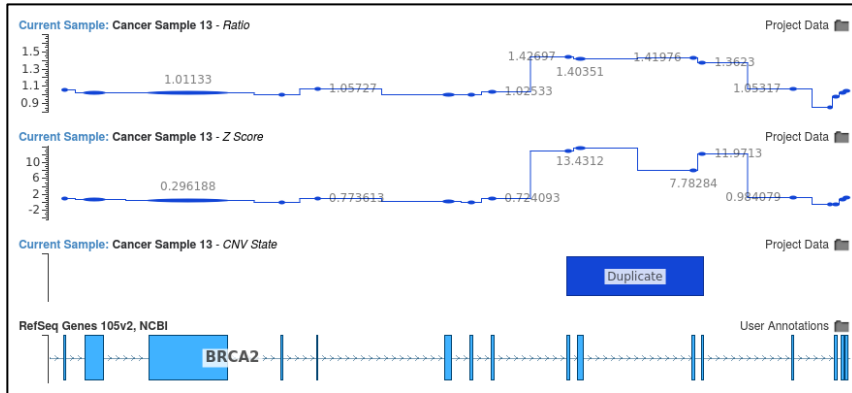
VSWarehouse



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

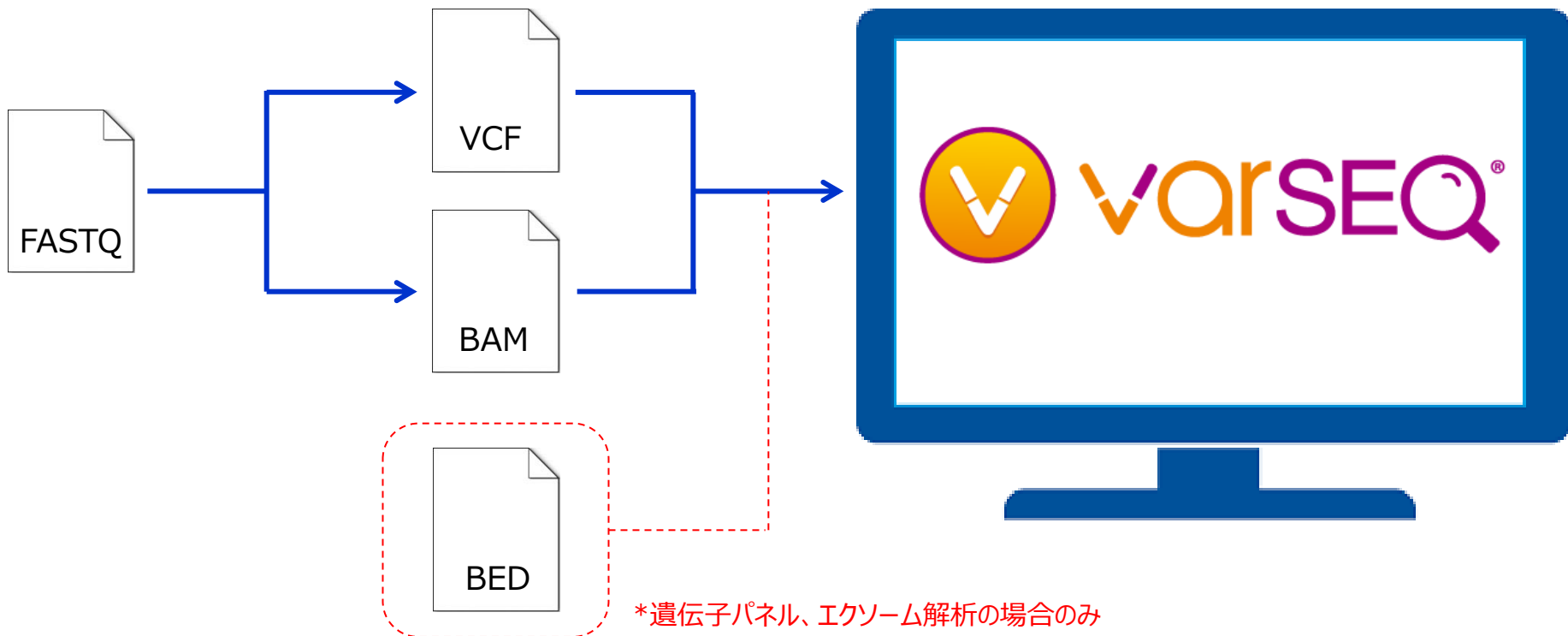


- キュレーションされた様々なデータリソースを使用し、変異データへアノテーション付けを実行
 - dbSNP
 - RefSeq Genes
 - COSMIC
 - 1000 Genome
 - NHLBI 6500 Exomes
 - ExAC Variant
 - gnomAD Exomes
 - SIFT and PolyPhen
 - dbNSFP Functional Predictions
 - ClinVar
 - CIVic
 - ICGC Simple Somatic Mutation
 - 各種遺伝子パネルのターゲットデータ ...など
- VCFファイルに含まれる変異データから、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
- HGVDなどの独自定義ファイルもアノテーションデータとして利用可能
- カバレッジ計算やトリオ解析、表現型情報に基づく遺伝子ランキングなどの解析アルゴリズムを搭載
- 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
- 無償提供のビューワーソフトウェアが利用でき、解析結果を容易にシェアすることが可能



- VarSeq[®]専用の、機能追加用の有償アドオン
- 遺伝子パネルや全エクソーム、全ゲノムシーケンズデータを使用した、CNV領域の検出
- 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
- LoH (Loss of Heterozygosity) 検出も可能

- 次世代シーケンサーの変異解析用パイプラインで出力された、変異データ（VCFファイル）とリード配列のアライメントデータ（BAMファイル）を使用
- 遺伝子パネルやエクソーム解析の場合は、ゲノム上のターゲットキャプチャー領域データ（BEDファイル）が必要



Create Sample Project



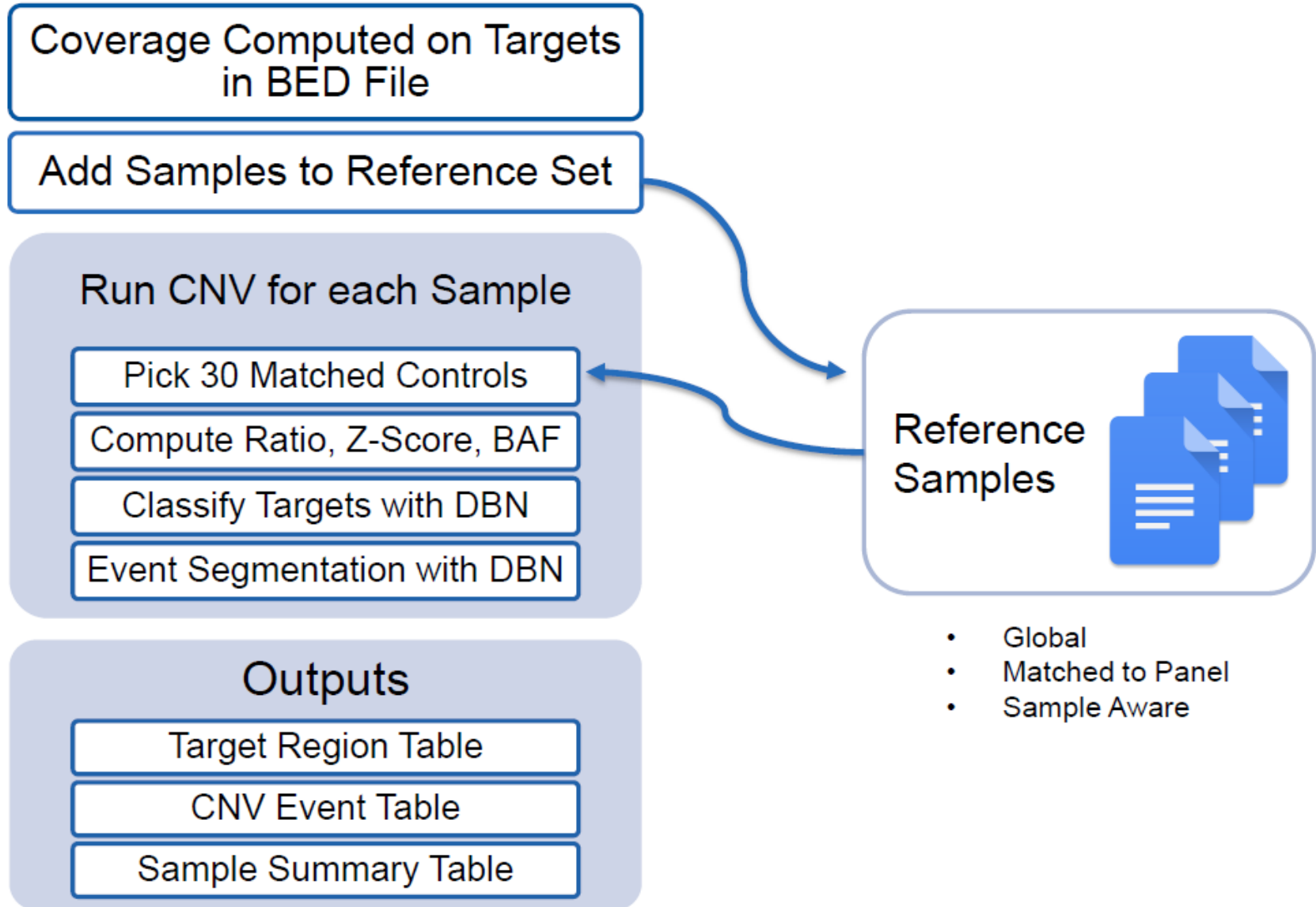
Compute Sample Coverage



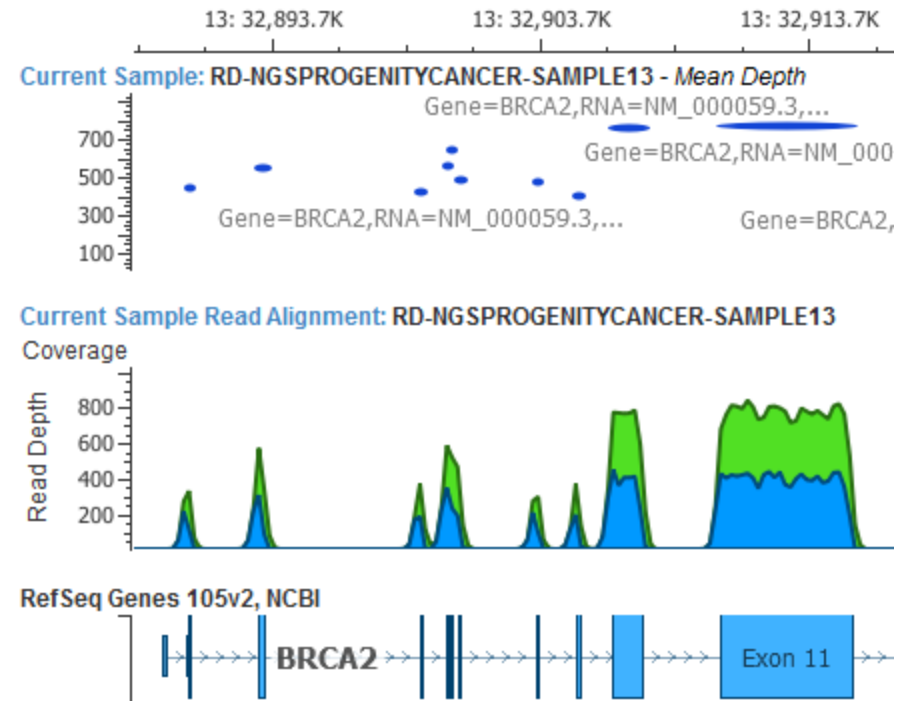
Call CNVs



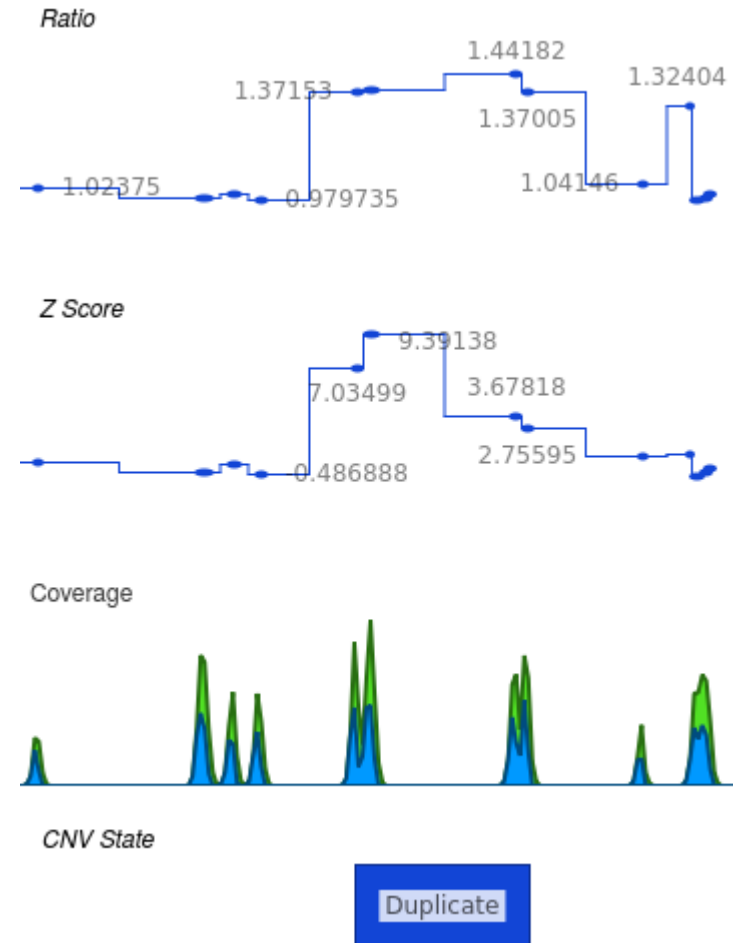
Plot on Genome Browser



- リードアライメントデータとターゲット領域データを使用し、各ターゲット領域ごとのカバレッジを計算
- 全ゲノムシーケンスの場合は、ゲノム全体を一定の間隔で分割し、分割された各領域ごとのカバレッジを計算する



- カバレッジデータのサンプル間の正規化を行った後、各サンプルごとに、リファレンスサンプルセットとカバレッジを比較して、各ターゲット領域ごとのRatioとZ-scoreを計算する
 - Z-Score: サンプルカバレッジが、リファレンスサンプルの平均カバレッジから、標準偏差の何倍離れているか
 - Ratio: サンプルカバレッジと、リファレンスサンプルの平均カバレッジとの比率
- 同時にDeletionやDuplicationなどの、ゲノム上でコピー数異常が検出された領域を示すデータ (CNV State) が出力される
- 計算されたRatioやZ score、CNV Stateなどのデータは、そのままゲノムブラウザにプロットすることができる

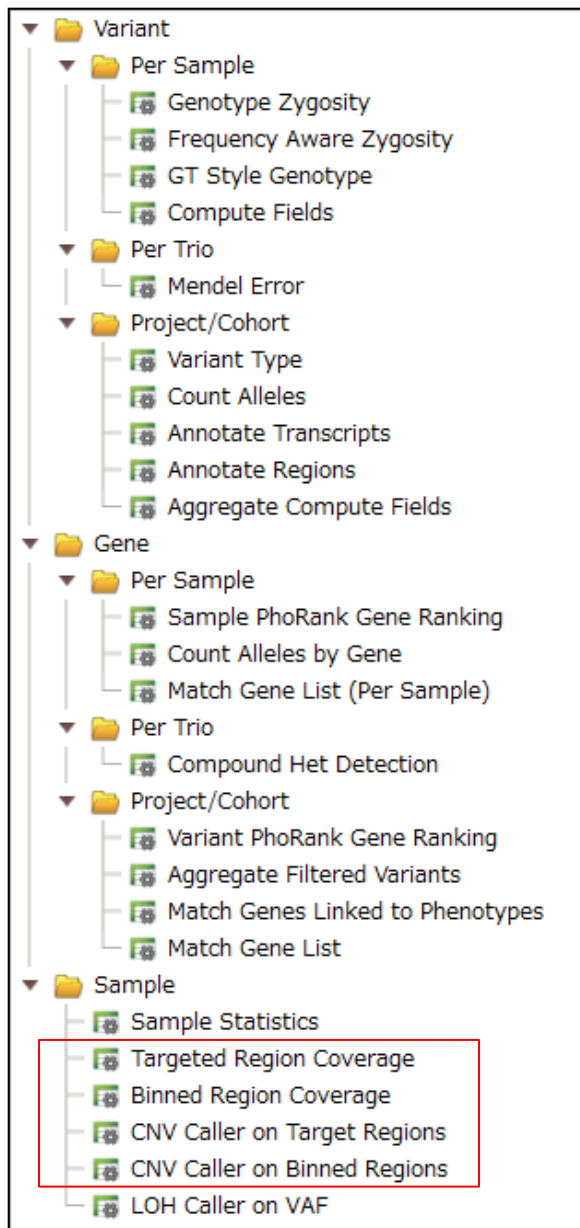


● カバレッジ

- 遺伝子パネル、エクソーム解析の場合は、100 x が必要

● サンプル

- リファレンスサンプルセットに、最低10サンプル、できれば30サンプル以上が望ましい
- すべて同一の解析プラットフォーム（パネルの種類、サンプル調整の条件、シーケンスデータ量など）である
- 性染色体の解析を行う場合は、サンプルの性別も合わせる必要がある



● カバレッジ計算

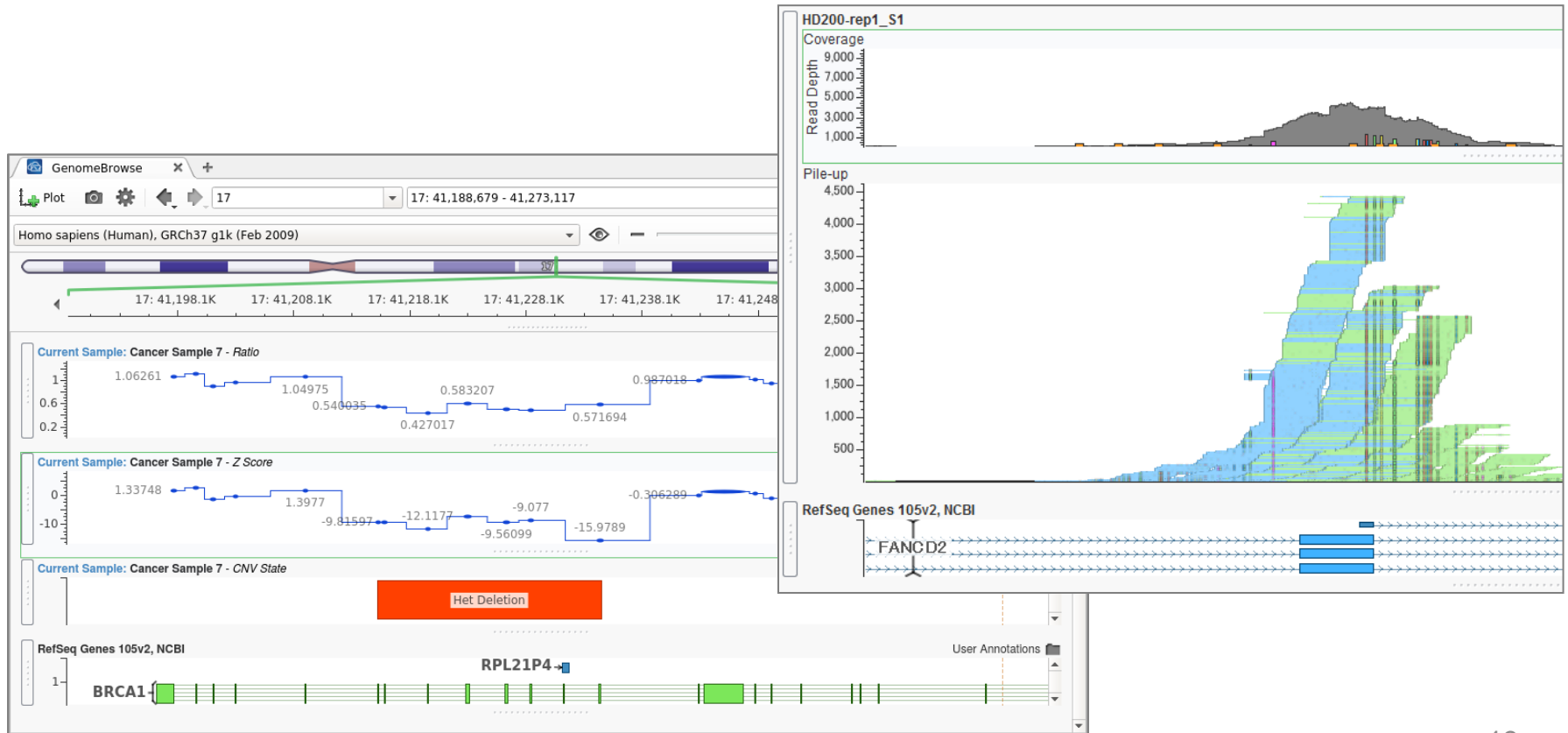
- **Targeted Region Coverage** (遺伝子パネル、エクソーム解析用)
- Bedファイルなどで指定されたターゲット領域ごとにカバレッジを計算
- **Binned Region Coverage** (全ゲノム解析用)
- ゲノム全体を任意の間隔で分割した領域ごとにカバレッジを計算

● CNV検出

- **CNV Caller on Target Regions** (遺伝子パネル、エクソーム解析用)
- Targeted Region Coverageで計算されたカバレッジデータを使用したCNV検出
- **CNV Caller on Binned Regions** (全ゲノム解析用)
- Binned Region Coverageで計算されたカバレッジデータを使用したCNV検出

● ゲノムブラウザー

- 検出されたCNV State、さらに各領域ごとのZ-ScoreやRatioをプロットすることで、ゲノム上のコピー数異常領域をビジュアル表示できる
- リードアライメントデータや、各種データソースの情報も、プロットが可能

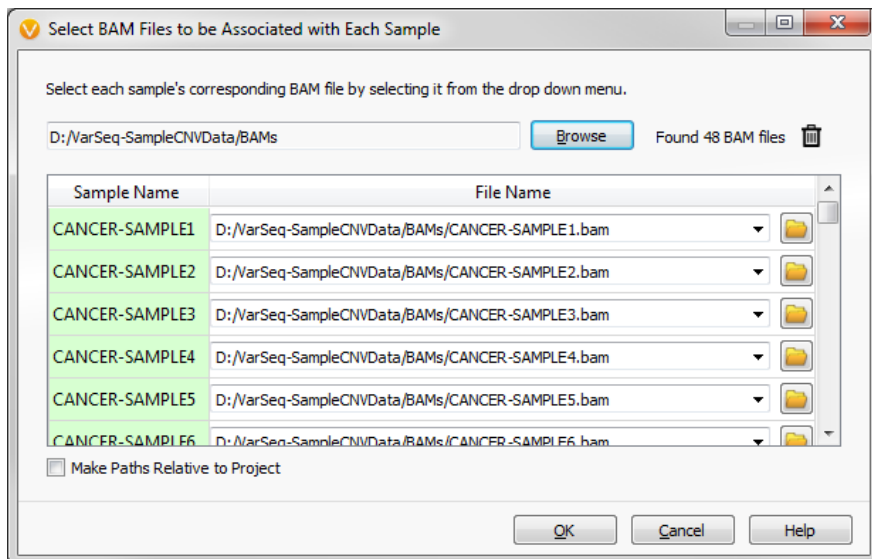
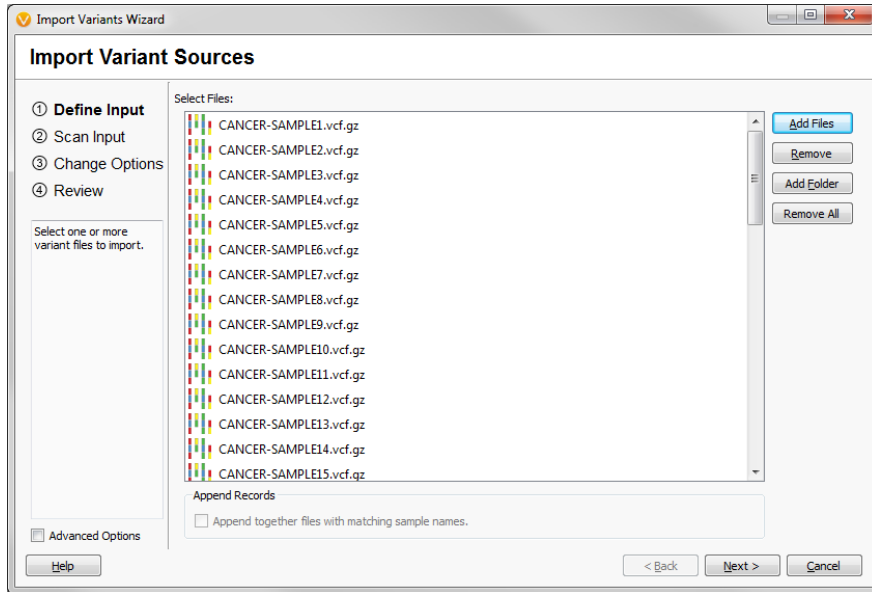


遺伝子パネル：**Illumina TruSight Cancer panel**

解析サンプル：**腫瘍細胞（48サンプル）**

48サンプルの腫瘍細胞を、Illumina社のがん遺伝子パネルとシーケンサーを用いてターゲットシーケンスを行い、得られた48サンプル分のFASTQファイルから、アライメントデータのBAMファイルと、変異コールデータのVCFファイルを、それぞれのサンプルに対して取得した。

サンプルデータのインポート

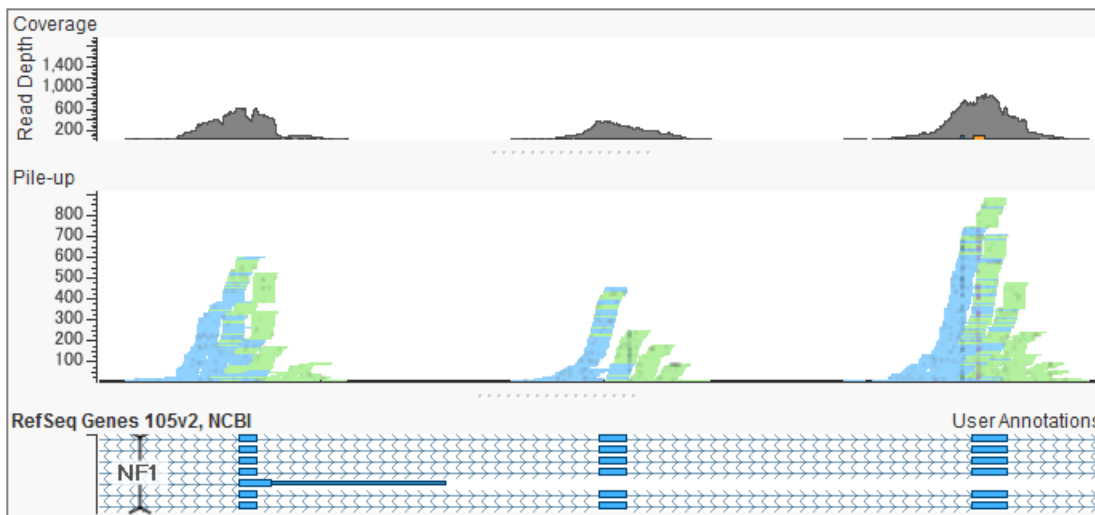


- 各サンプルのVCFファイルと同時に、BAMファイルもインポートを行う

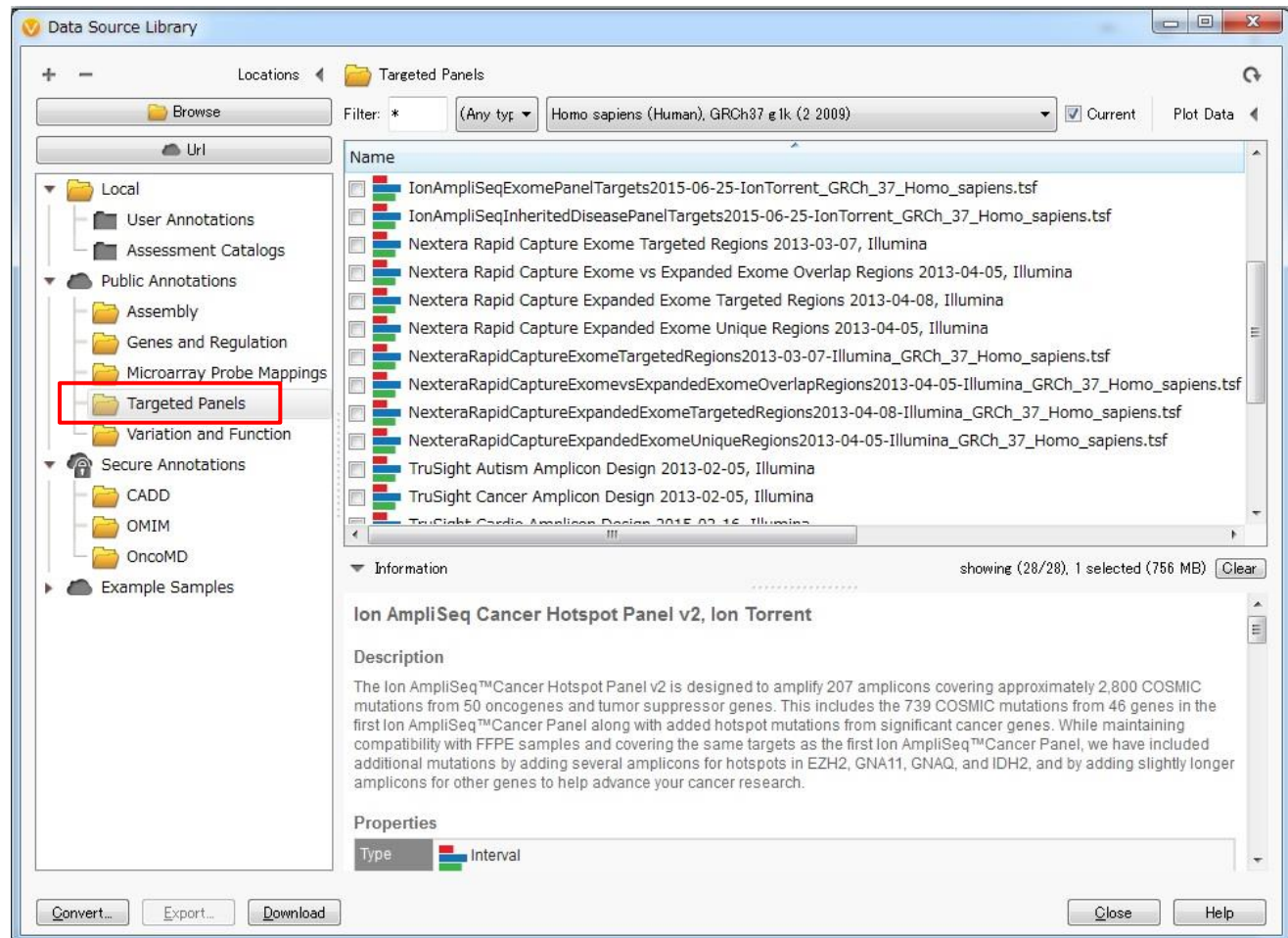
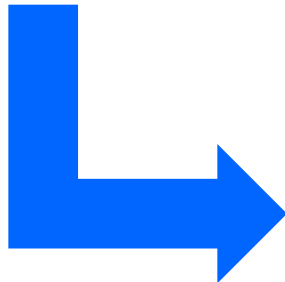
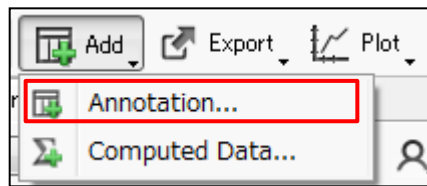
Variant Info			Cancer Sample 7		
Chr:Pos	Ref/Alt	Identifier	Read Depth (DP)	Variant Allele Freq	Allelic Depths (AD)
17:41244000	T/C	rs16942	193	0.458333	104,88
17:41244301	T/C	?	?	?	?,?
17:41244435	T/C	rs16941	187	0.526882	88,98
17:41244524	C/T	rs1800704	?	?	?,?
17:41244815	T/C	rs1800740	?	?	?,?
17:41244936	G/A	rs799917	211	0.438095	118,92
17:41245090	T/C	rs56082113	?	?	?,?
17:41245237	A/G	rs16940	242	0.39834	145,96
17:41245466	G/A	rs1799949	227	0.537445	105,122
17:41245471	C/T	rs4986850	229	0.532751	107,122
17:41246481	T/C	rs1799950	?	?	?,?
17:41246567	T/C	rs1800063	?	?	?,?
17:56772341	A/G	rs45511291	?	?	?,?
17:56780540	G/T	rs193023469	?	?	?,?
17:56787304	G/A	rs147241704	?	?	?,?
17:56798128	A/G	rs28363317	?	?	?,?
17:59760996	A/G	rs4986763	188	0.484043	97,91
17:59763347	A/G	rs4986764	195	0.471795	103,92
17:59763465	T/C	rs4986765	152	0.536913	69,80

- インポートした変異データ（VCFファイル）と、リードアライメントデータ（BAMファイル）を表示

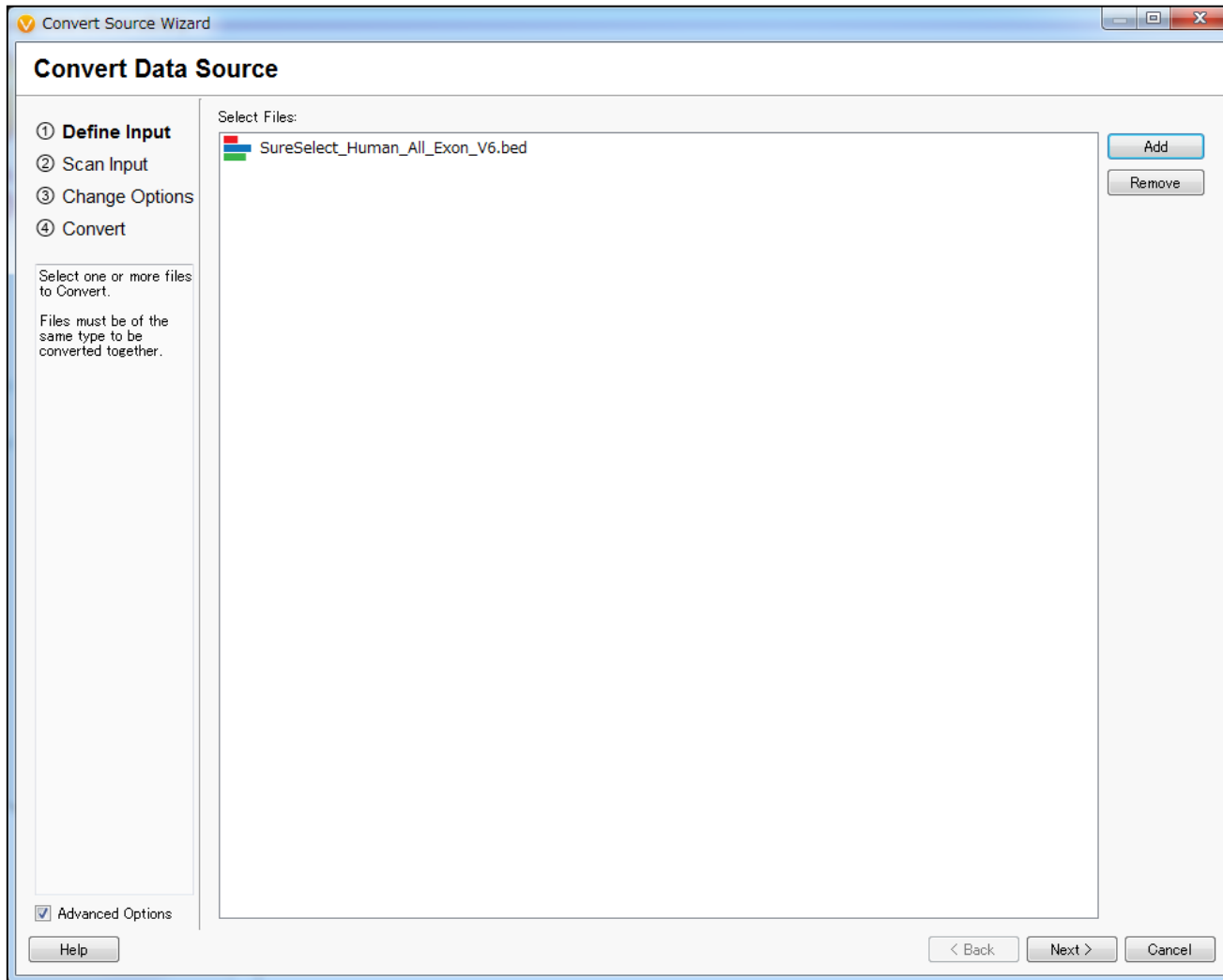
✓ 変異データ



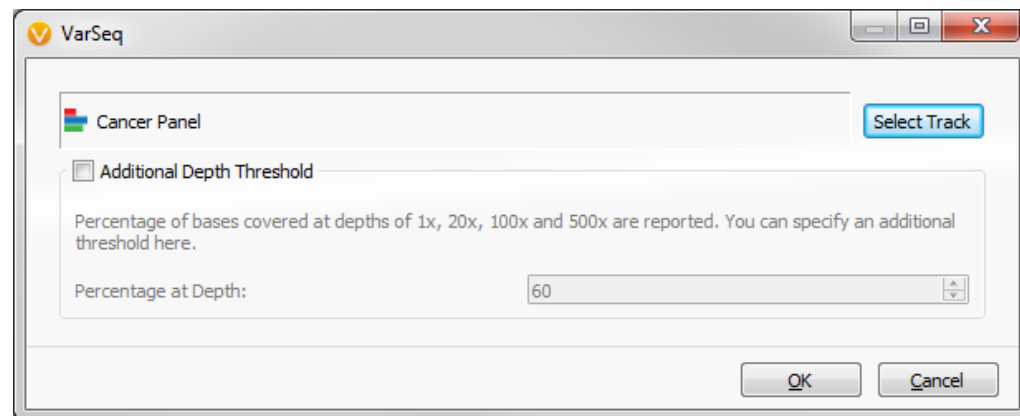
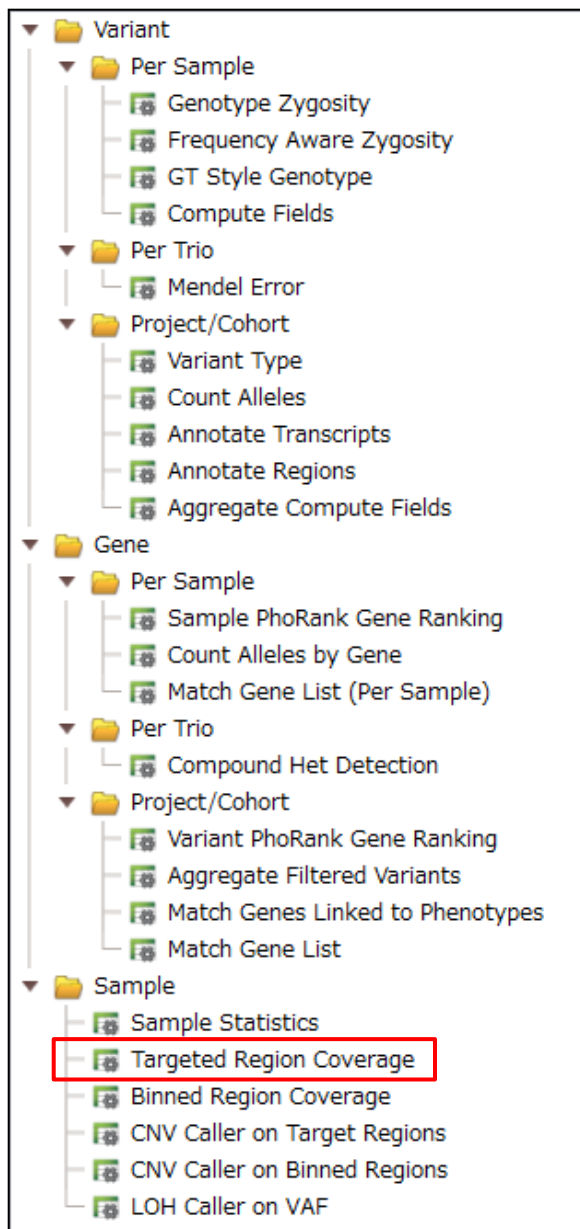
✓ アライメントデータ



- Illumina社、Thermo Fisher Scientific社の遺伝子パネル、エクソーム解析用のターゲット領域データは、VarSeq[®]のダウンロードツールより、ダウンロードが可能



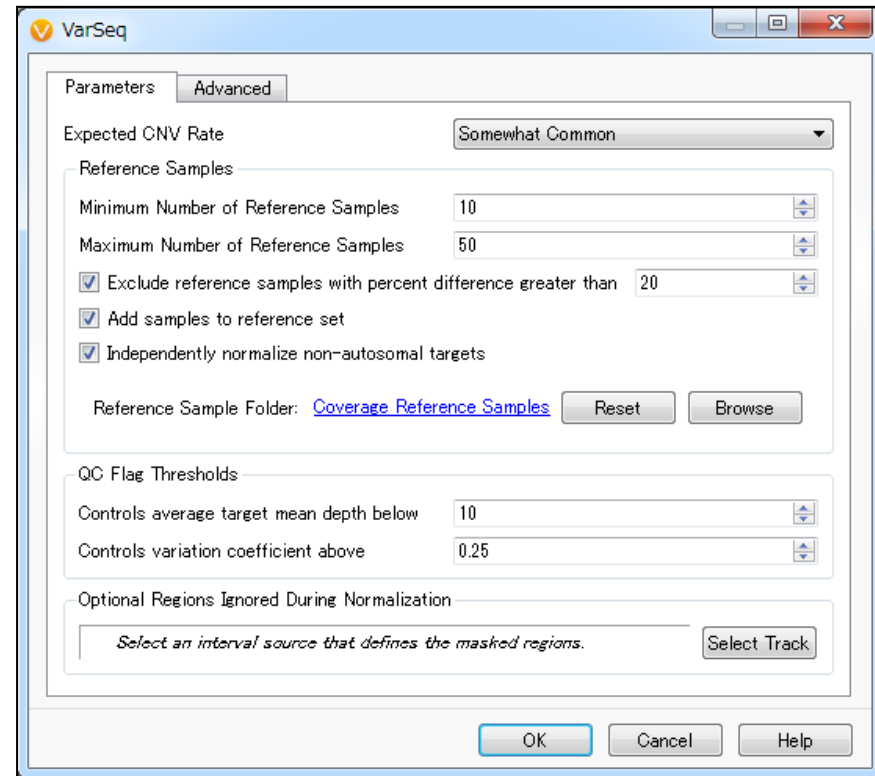
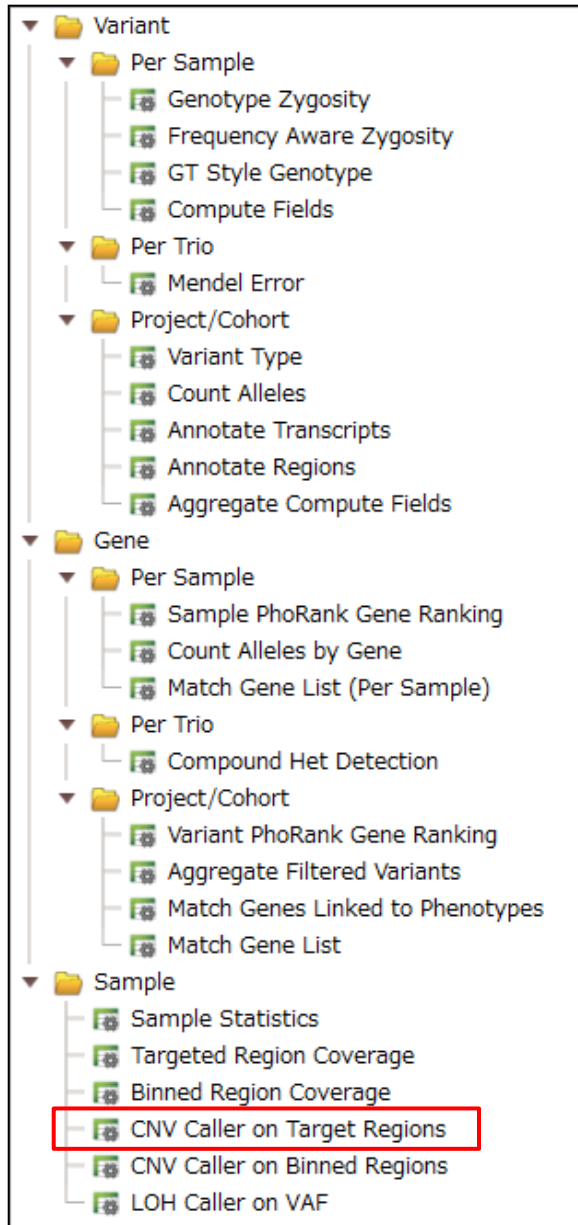
- ダウンロードツールにないターゲット領域データを使用する場合は、BEDファイルをインポートして使用する



- Targeted Region Coverageを起動し、ダウンロードしておいたターゲット領域データを指定する

Coverage Region Info		Coverage Statistics Cancer Panel for Cancer Sample 1					
Region	Mean Depth	Min Depth	Max Depth	% 1x	% 20x	% 100x	% 500x
1:45794959-45795129	468.094	331	522	100	100	100	28.655
1:45796169-45796249	441.012	397	474	100	100	100	0
1:45796835-45797026	722.182	621	769	100	100	100	100
1:45797073-45797248	690.602	597	761	100	100	100	100
1:45797314-45797541	659.588	566	711	100	100	100	100
1:45797676-45797778	710.757	633	809	100	100	100	100
1:45797819-45798002	862.989	811	916	100	100	100	100
1:45798044-45798180	747.438	687	829	100	100	100	100
1:45798227-45798379	703.353	653	746	100	100	100	100
1:45798416-45798526	662.928	635	697	100	100	100	100
1:45798571-45798651	701.309	677	727	100	100	100	100
1:45798750-45798862	683.195	651	716	100	100	100	100
1:45798938-45799016	756.43	712	790	100	100	100	100
1:45799066-45799295	722.583	563	791	100	100	100	100
1:45800044-45800203	388.581	300	455	100	100	100	0
1:45805872-45805946	556.973	521	588	100	100	100	100
2:47596626-47596740	208.47	157	234	100	100	100	0
2:47600583-47600729	456.068	372	508	100	100	100	6.80272
2:47600928-47601207	550.189	379	635	100	100	100	82.8571
2:47602354-47602458	455.657	396	517	100	100	100	16.1905
2:47604134-47604236	453.165	382	510	100	100	100	13.5922
2:47606073-47606213	410.319	327	472	100	100	100	0
2:47606889-47607128	514.317	286	637	100	100	100	62.5
2:47612286-47612369	513.036	464	549	100	100	100	67.8571
2:47613692-47613772	551.58	433	674	100	100	100	72.8395

- パネルに含まれている、各ターゲット領域ごとのカバレッジデータがテーブル形式で出力される



- CNV Caller on Target Regionsを起動し、リファレンスサンプルやQCチェックのパラメータなどを指定する

● Sample Table

- 各サンプルごとのQCチェックデータ（平均カバレッジや低カバレッジフラグなど）

Sample Info		Coverage Statistics Cancer Panel							Copy Number Variants		
Samples	Sample Mean Depth	SampleMeanForwardDepth	Sample Mean Reverse Depth	Sample % 1x	Sample % 20x	Sample % 100x	Sample % 500x	Sample Flags	Inferred Gender	# CNV Events	
Cancer Sample 25	144.426	74.6347	69.791	100	99.869	81.5915	0	?	?	0	
Cancer Sample 26	681.601	353.186	328.415	100	100	99.7713	79.765	?	?	0	
Cancer Sample 27	556.331	288.364	267.966	100	100	99.8437	64.2316	?	?	0	
Cancer Sample 28	586.246	302.465	283.781	100	100	99.7656	68.7547	?	?	2	
Cancer Sample 29	183.889	95.2023	88.6863	100	100	97.2301	0	?	?	1	
Cancer Sample 30	147.068	75.6568	71.4111	100	99.9471	95.23	0	?	?	2	
Cancer Sample 31	552.113	281.229	270.884	100	100	99.7501	62.769	?	?	0	
Cancer Sample 32	570.24	290.126	280.115	100	100	99.6581	65.1278	?	?	0	
Cancer Sample 33	401.821	206.623	195.198	100	99.8983	99.3529	23.0273	?	?	0	
Cancer Sample 34	19.4118	9.82568	9.58581	99.9837	45.5645	0	0	High IQR,Low Sample Mean Depth	?	25	
Cancer Sample 35	172.921	89.1312	83.7899	100	100	98.0066	0	?	?	1	
Cancer Sample 36	682.728	350.094	332.634	100	100	99.8046	82.7505	?	?	0	
Cancer Sample 37	771.953	395.482	376.471	100	100	99.7949	89.547	?	?	1	
Cancer Sample 38	296.208	150.766	145.442	100	99.9821	99.3513	0	?	?	0	
Cancer Sample 39	617.615	316.354	301.261	100	100	99.6614	73.9583	?	?	1	
Cancer Sample 40	475.563	241.922	233.641	100	99.8763	99.1877	42.1924	?	?	2	
Cancer Sample 41	28.3141	14.3683	13.9457	99.9943	76.2877	0	0	High IQR,Low Sample Mean Depth	?	14	
Cancer Sample 42	412.22	212.505	199.715	100	100	99.5124	24.4785	?	?	1	
Cancer Sample 43	619.86	321.093	298.767	100	100	99.8185	73.4291	?	?	1	
Cancer Sample 44	393.09	203.393	189.697	100	100	99.7306	16.6892	?	?	2	
Cancer Sample 45	328.852	169.296	159.556	100	100	99.5393	0.175813	?	?	2	

● Coverage Region Table

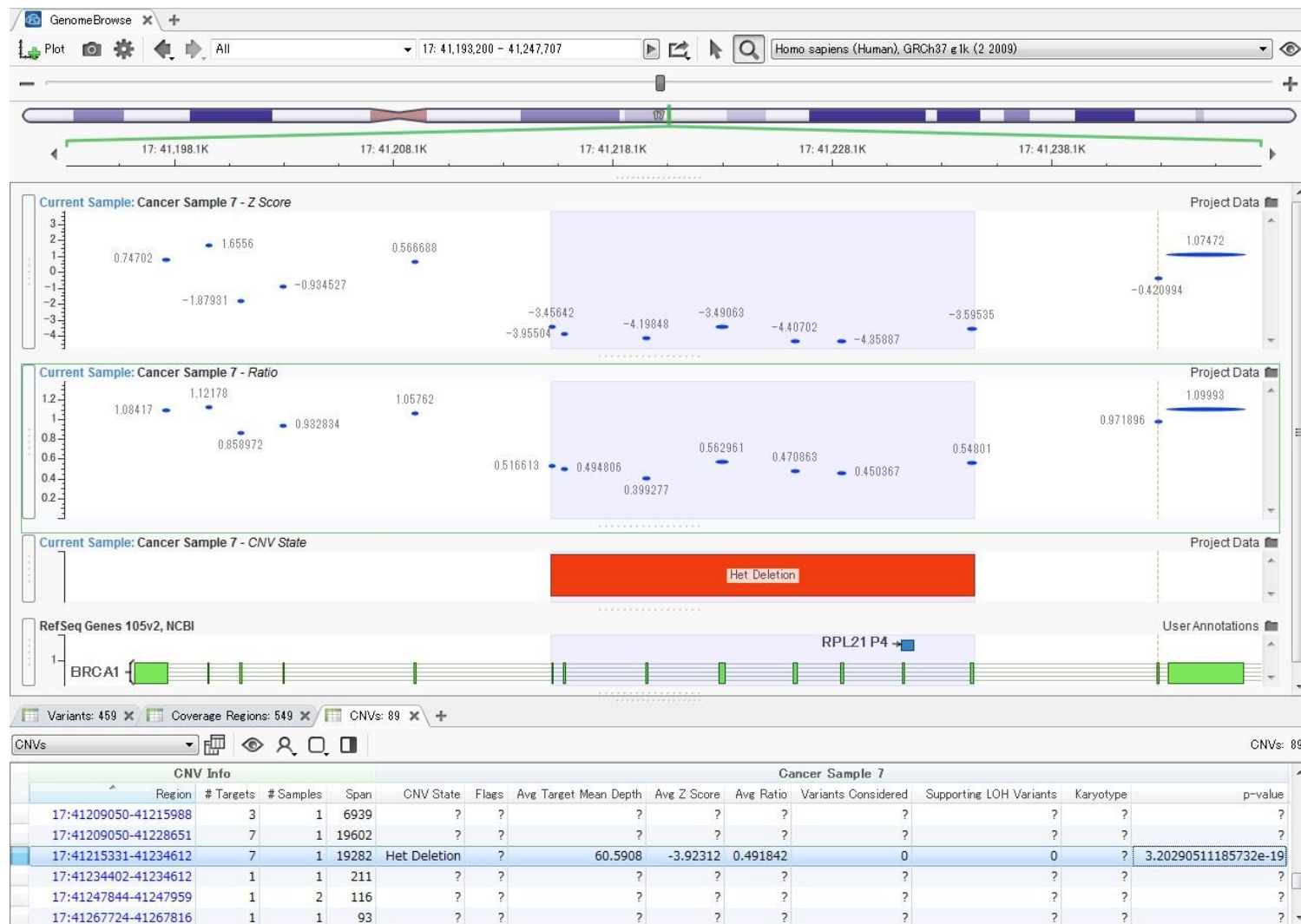
- 各ターゲット領域ごとのコピー数状態（Diploid、Deletion、Duplicationなど）と Ratio、Z-scoreなど

Coverage Region Info		Coverage Statistics Cancer Panel for Cancer Sample 1							Target Copy Number State for Cancer Sample 1				
Region	Mean Depth	Min Depth	Max Depth	% 1x	% 20x	% 100x	% 500x	CNV State	Flags	Z Score	Ratio	Variants Considered	
1:45798416-45798526	662.928	635	697	100	100	100	100	Diploid	?	-0.0328397	0.99772	0	
1:45798571-45798651	701.309	677	727	100	100	100	100	Diploid	?	0.712468	1.05367	0	
1:45798750-45798862	683.195	651	716	100	100	100	100	Diploid	?	1.16314	1.10108	0	
1:45798938-45799016	756.43	712	790	100	100	100	100	Diploid	?	1.06012	1.10258	0	
1:45799066-45799295	722.583	563	791	100	100	100	100	Diploid	?	0.609163	1.05001	0	
1:45800044-45800203	388.581	300	455	100	100	100	0	Diploid	?	-0.529925	0.957499	0	
1:45805872-45805946	556.973	521	588	100	100	100	100	Diploid	?	0.392859	1.04146	0	
2:47596626-47596740	208.47	157	234	100	100	100	0	Het Deletion	?	-2.88048	0.528975	0	
2:47600583-47600729	456.068	372	508	100	100	100	6.80272	Diploid	?	0.238906	1.03	0	
2:47600928-47601207	550.189	379	635	100	100	100	82.8571	Diploid	?	0.500308	1.04318	1	
2:47602354-47602458	455.657	396	517	100	100	100	16.1905	Diploid	?	-0.389887	0.957959	0	
2:47604134-47604236	453.165	382	510	100	100	100	13.5922	Diploid	?	-0.343476	0.970031	0	
2:47606073-47606213	410.319	327	472	100	100	100	0	Diploid	?	-0.658619	0.926245	0	
2:47606889-47607128	514.317	286	637	100	100	100	62.5	Diploid	?	0.614464	1.05725	0	
2:47612286-47612369	513.036	464	549	100	100	100	67.8571	Diploid	?	-0.10762	0.988171	0	
2:47613692-47613772	551.58	433	674	100	100	100	72.8395	Diploid	?	0.212828	1.02333	0	

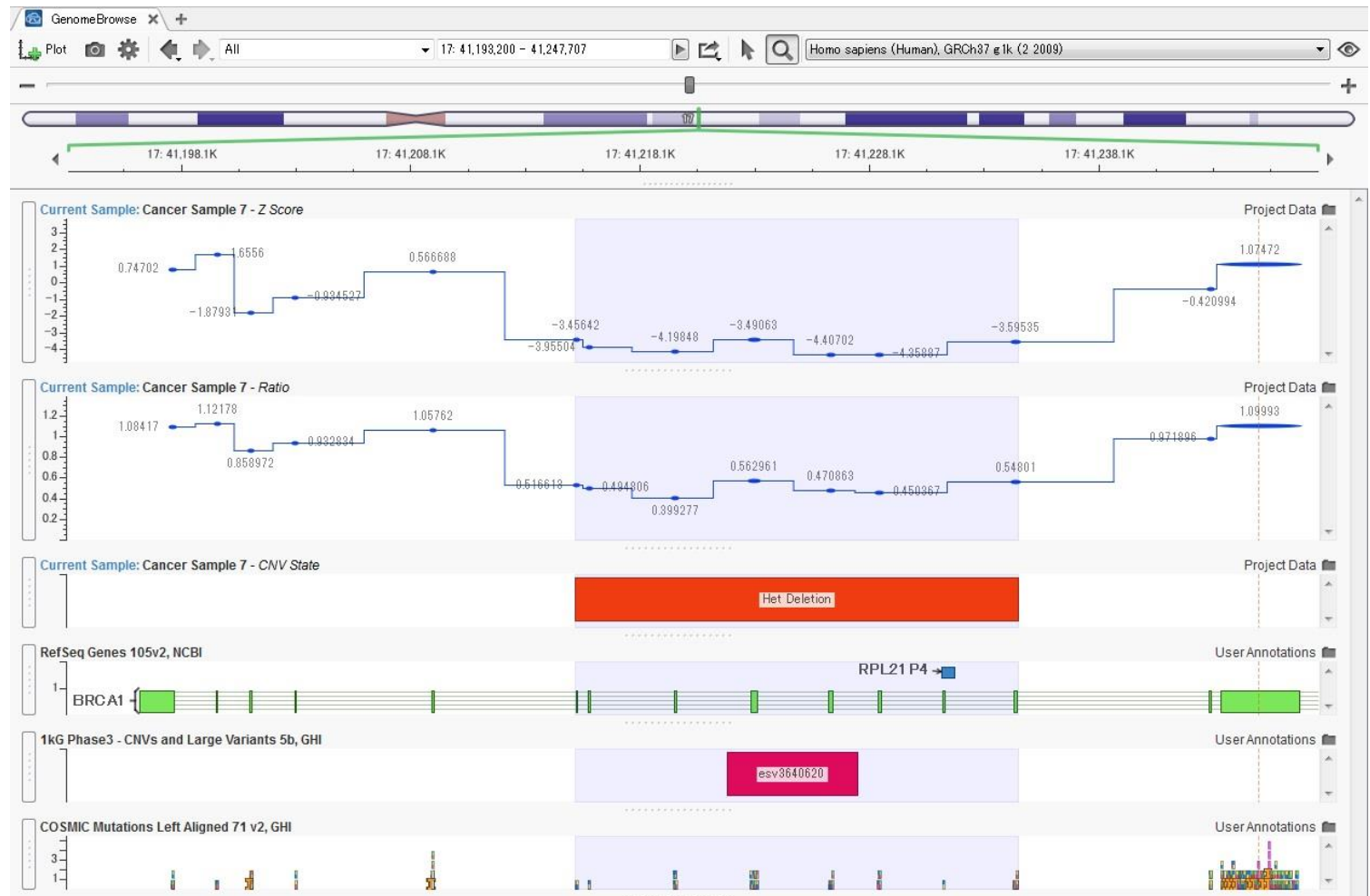
● CNVs Table

- ゲノム全体のコピー数異常領域とその状態（Diploid、Deletion、Duplicationなど）、さらにP-valueなど

CNV Info			Cancer Sample 3							
Region	# Targets	Span	CNV State	Flags	Avg Target Mean Depth	Avg Z Score	Avg Ratio	Supporting LOH Variants	p-value	
7:6013011-6018347	3	5337	Het Deletion	High Controls Variation,Low Z Score	52.8089	-1.62329	0.380012	0	0.0329238548874855	
10:43606636-43609143	4	2508	Duplicate	Within Regional IQR	153.515	2.57413	1.23655	0	0.000513269216753542	
11:108213930-108236255	7	22326	Het Deletion	?	75.8131	-3.97302	0.504004	0	8.05861182905793e-14	
17:7579293-7579932	3	640	Duplicate	Within Regional IQR	153.536	2.51492	1.2535	0	0.00327462935820222	
17:29422309-29422407	1	99	Duplicate	?	124.061	3.13198	1.40498	0	0.0324721410870552	
17:29557841-29559227	2	1387	Duplicate	?	152.233	2.94003	1.42884	0	0.00469905091449618	



- 出カテーブルの各データを、ゲノムブラウザーにプロットする



- Ratioなどの数値データを、変化が分かりやすいように線でつなげたり、各データソースのSNVやCNVデータを、プロットに加えることができる

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