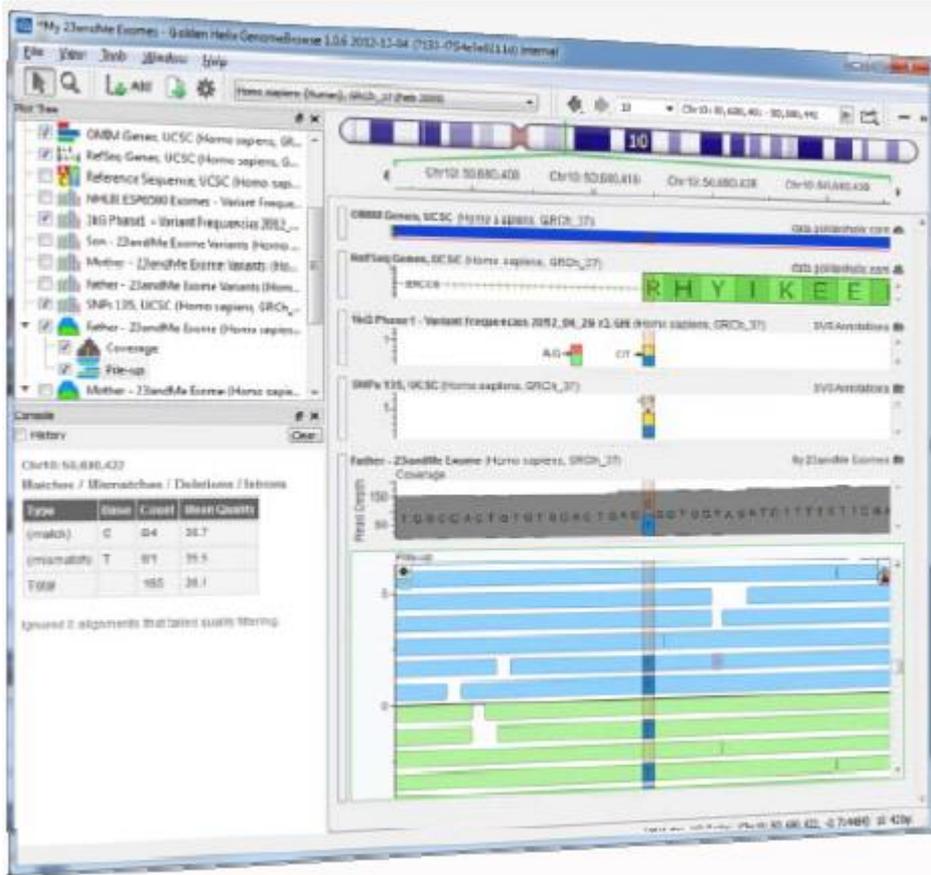
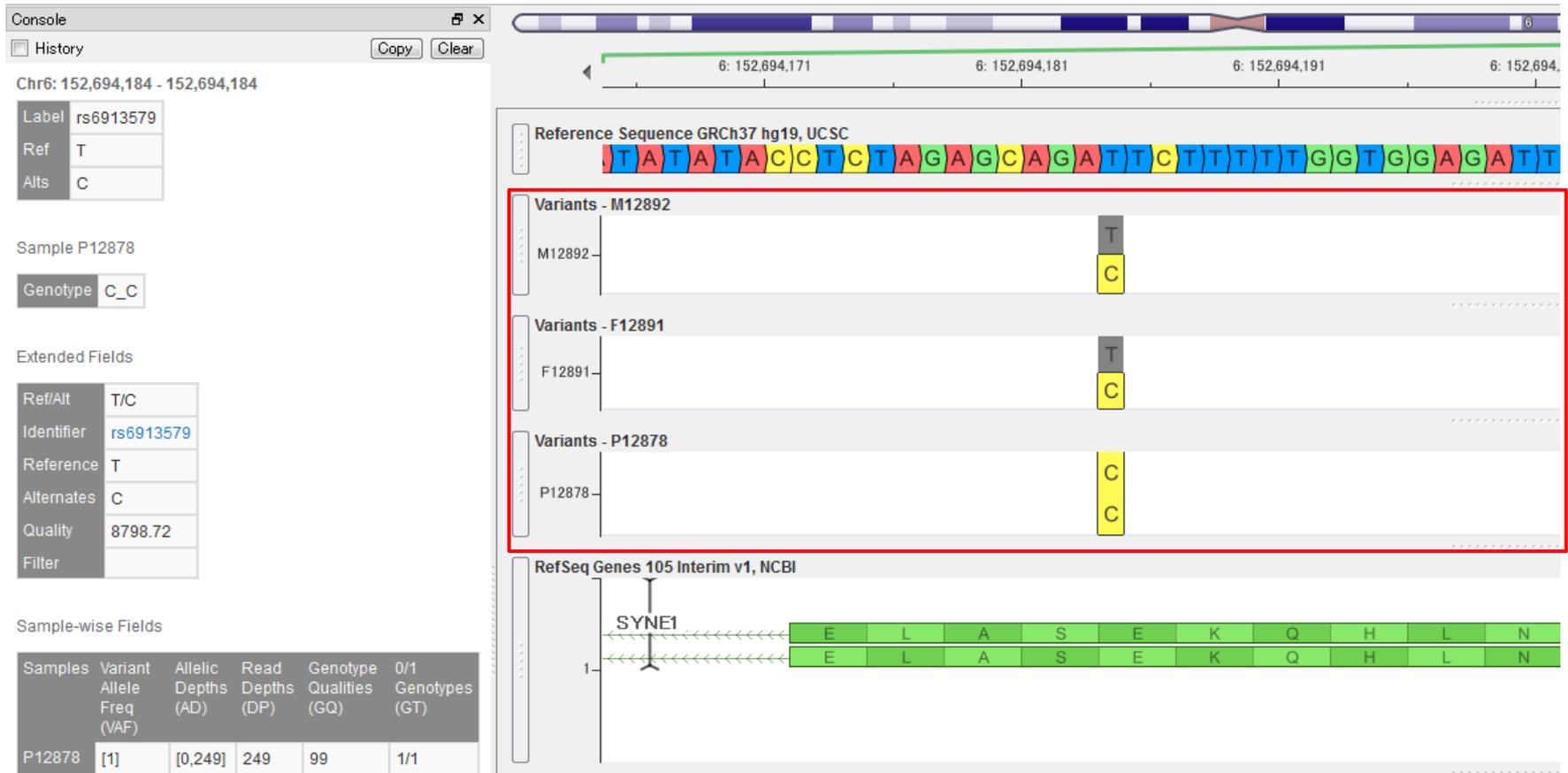


無料ゲノムブラウザーを用いた 遺伝子変異解析

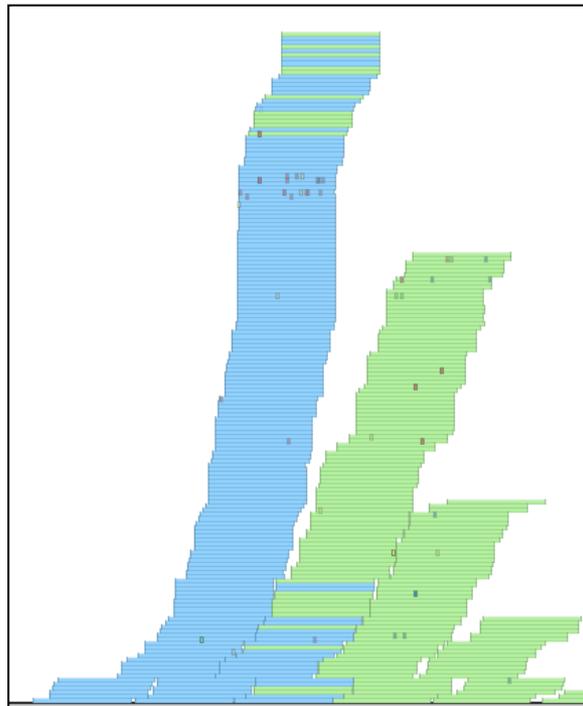
フィルジェン株式会社 バイオインフォマティクス部
(biosupport@filgen.jp)



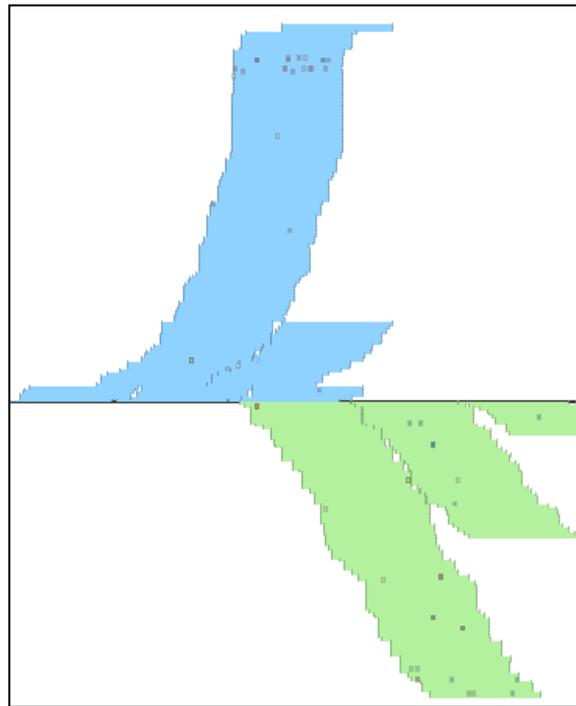
- 次世代シーケンサーより得られたシーケンスデータのクオリティチェックやビジュアライゼーションには、ゲノムブラウザーが用いられる
- Golden Helix社がアカデミックフリーで配布している高機能なゲノムブラウザー GenomeBrowse[®]を使用することにより、シーケンスデータや遺伝子変異データのクオリティチェックやビジュアライゼーション、さらに変異の各種データベースへの照合を容易に実行することができる



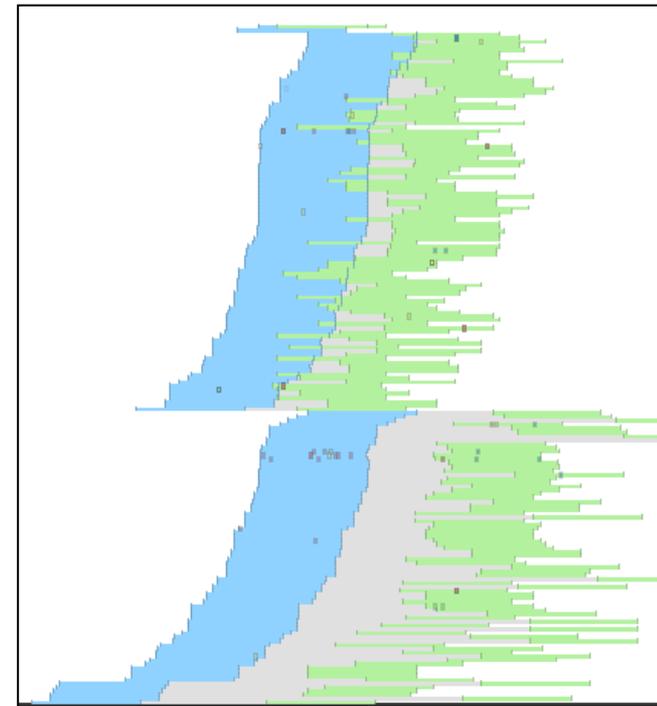
- VCFファイルに含まれるアليل情報をプロットし、またコンソールにてアليل頻度やリード深度などの詳細情報を確認
- 複数のVCFファイルを並べてプロットし、サンプル間の比較を行うことも可能



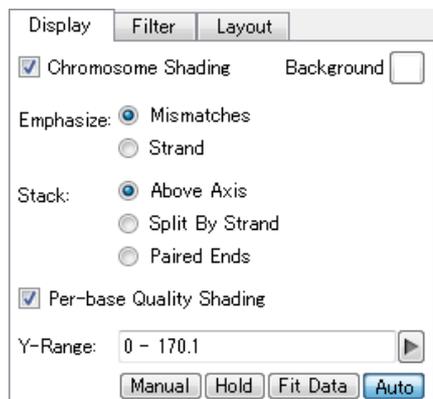
通常プロット



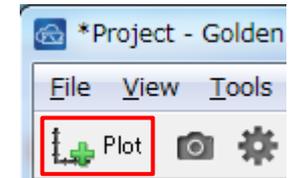
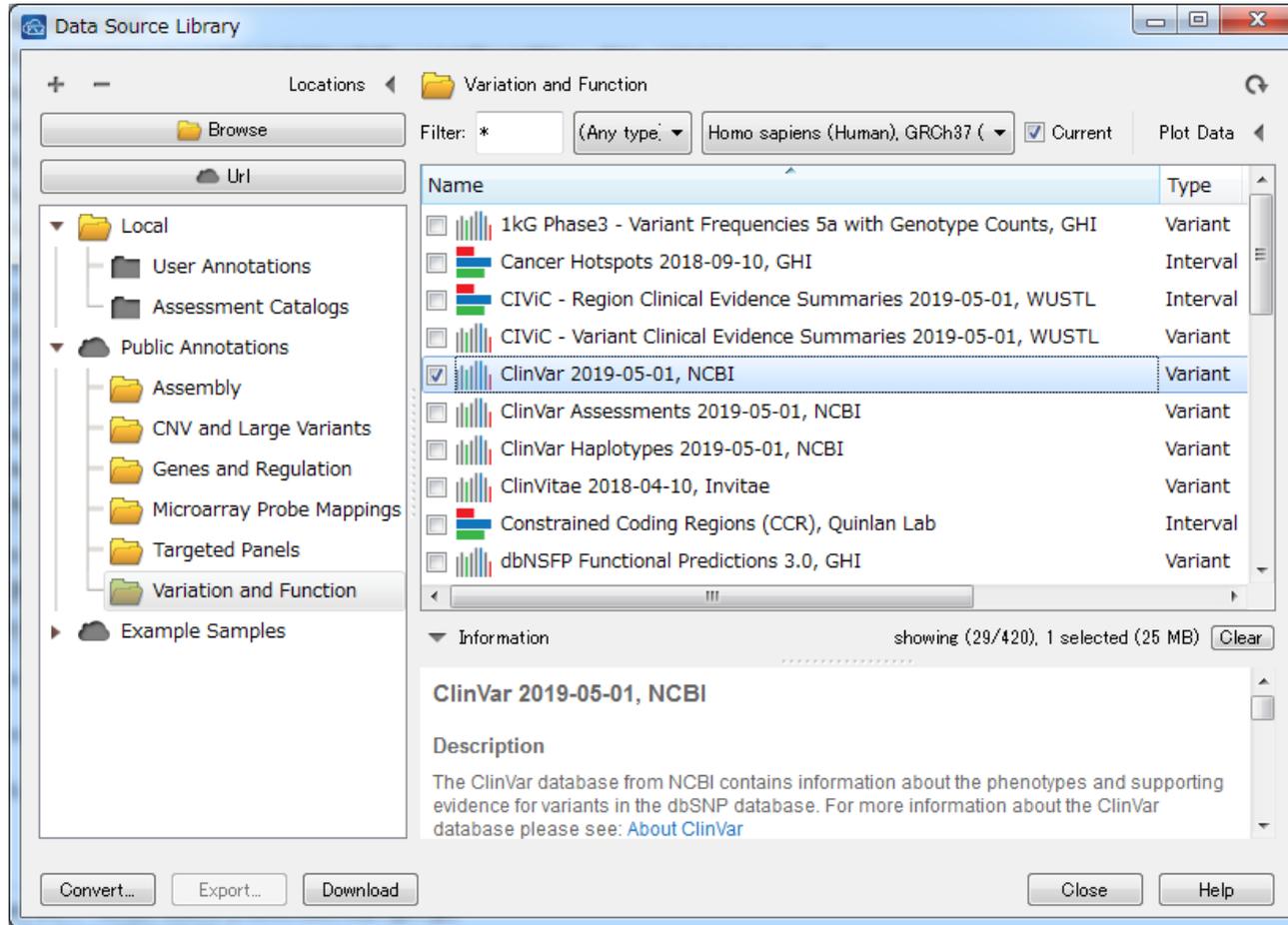
Split by Strandプロット



Paired Endsプロット



- Pile-Upグラフは、リード配列の表示形式を3タイプに切り替えることが可能
- 変異などの参照ゲノム配列とのミスマッチや、リード配列のクオリティーに基づいた表示を行うことで、変異やシーケンス結果のクオリティーチェックに活用できる



ダウンロードしたデータベースを自由に選択し、ゲノムブラウザにプロット

- Golden Helix社によってデータの精査・メンテナンスが行われている各種アノテーションリソースを、Data Source Libraryより自由にダウンロードし、ゲノムブラウザにプロット
- 日本人変異データベースHGVDなどのカスタムアノテーションデータをインポートし、アノテーションリソースとして使用が可能

Assemblies

- GRCh 37 (hg19)
- GRCh 38 (hg38)

Genes

- RefSeq
- Ensembl

Clinical Annotations

- ClinVar
- GWAS Catalog

Cancer Annotations

- CIViC
- Cancer Hotspots
- ICGC Simple Somatic Mutations
- TCGA Variants
- MSK Impact

Functional Annotations

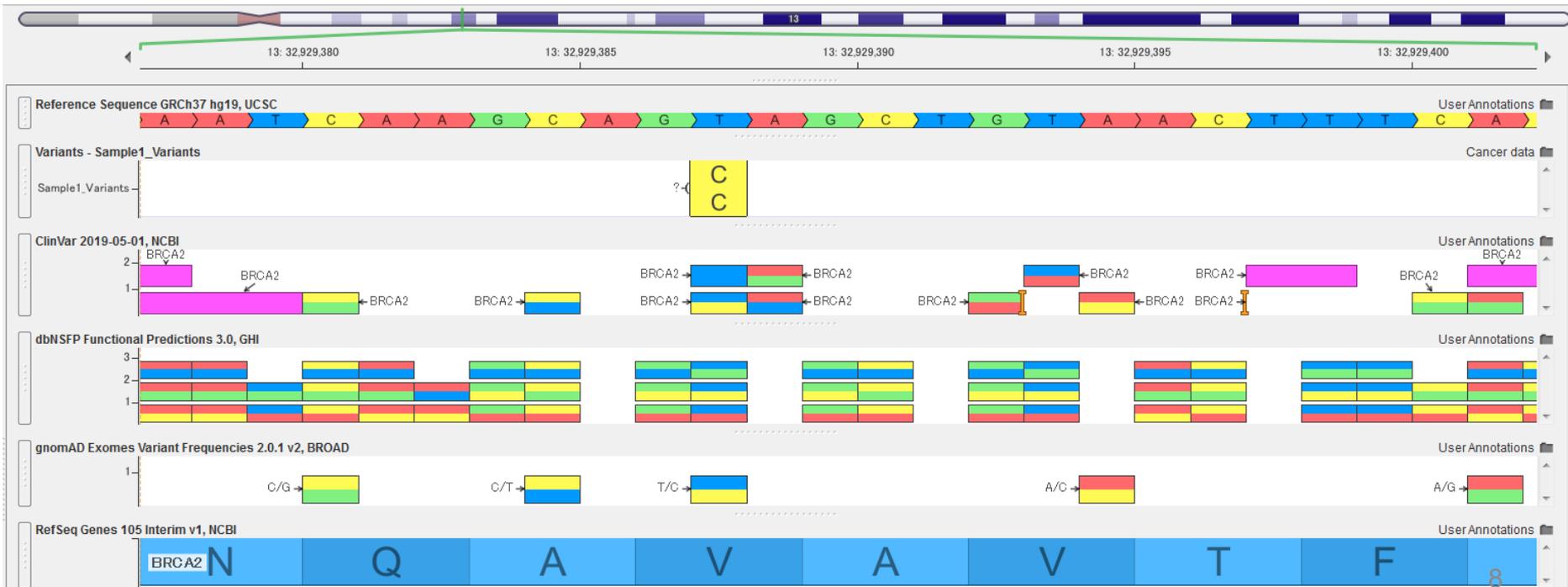
- dbNSFP (SIFT, PolyPhen, MutationTaster...)

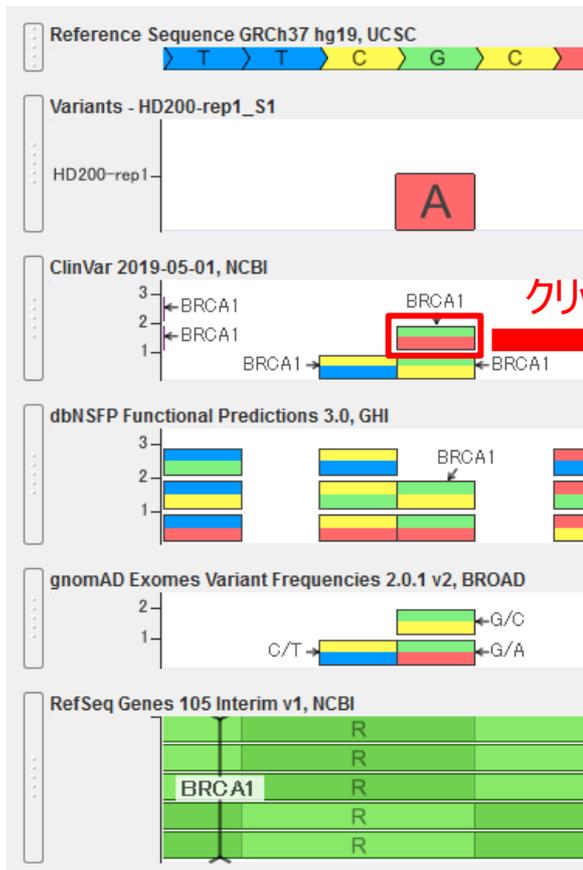
Population Catalogs

- dbSNP
- 1000 Genome
- NHLBI 6500 Exomes
- ExAC Variant
- gnomAD Exomes
- TOPMed

Targeted Panels

- TruSight
- Ion AmpliSeq



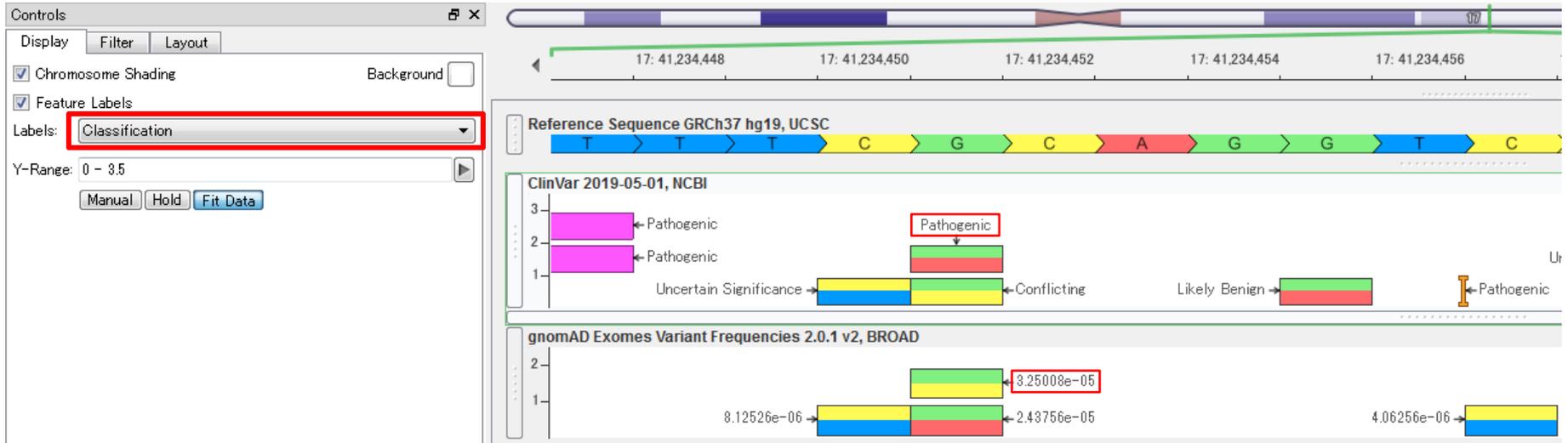


Chr17: 41,234,451 - 41,234,451
[NCBI](#), [UCSC](#), [Ensembl](#)
 Normalized Alleles

Extended Fields

Ref/Alt	G/A
Variant ID	17675
Classification	Pathogenic
Clinical Significance	Pathogenic
Aggregate of Interpretations from Submissions	Pathogenic (32)
Review Status	(3 Stars) Reviewed By Expert Panel
RSID	rs41293455
Gene Names	BRCA1
Gene IDs	672
HGVS g. Name	NC_000017.10:g.41234451G>A
HGVS g. Name (GRCh38)	NC_000017.11:g.43082434G>A
HGVS c. Name	NM_007300.3:c.4327C>T
HGVS p. Name	NP_009231.2:p.Arg1443Ter
Per Condition Record Accessions	RCV000019244, RCV000048523, RCV000131880, RCV000159989, RCV000235131, RCV000239083, RCV000515235, RCV000735445, RCV000763000
Conditions	Breast-ovarian cancer, familial 1, Hereditary breast and ovarian cancer syndrome, Hereditary cancer-predisposing syndrome, Familial cancer of breast, not provided, not specified, Breast-ovarian cancer, familial 1; Familial cancer of breast; Pancreatic cancer 4, Breast and/or ovarian cancer, Breast-ovarian cancer, familial 1; FANCONI ANEMIA, COMPLEMENTATION GROUP S; Familial cancer of breast; Pancreatic cancer 4

- ゲノムブラウザ上でプロットをクリックすると、コンソールに各プロットの詳細が表示される



- 各リソースのプロットラベルは、コンソールより表示フィールドの選択が可能
- その他、プロットするデータの選択フィルターや、レイアウトの編集など、ゲノムブラウザ上での表示設定を、任意に変更することが可能

解析サンプル：腫瘍細胞と正常細胞のペアサンプル

解析に用いるデータファイル：各サンプルのVCFファイル、BAMファイル

解析の目的：がん遺伝子における体細胞変異の抽出と、臨床アノテーションの確認

胃腺がん患者の腫瘍細胞サンプルと正常細胞サンプルのゲノムシーケンスより取得したVCFファイルとBAMファイルを使用し、特定のがん遺伝子の体細胞変異の抽出とクオリティチェックを行い、さらにデータベースに照合することで、病原性の有無や抗がん剤情報といった、臨床アノテーションの確認を行う。

手順1 : VCF, BAMファイルの取り込み

The screenshot displays the Golden Helix GenomeBrowse 3.0.0 software interface. The main window shows a genomic track for Homo sapiens (Human), GRCh37 (hg19) (2 2009). The track is currently empty, with a 'RefSeq Genes 105 Interim v1, NCBI' track visible below it. A large blue arrow labeled 'ドラッグ & ドロップ°' (Drag & Drop) points from a file selection box to the main track area. The file selection box contains the following files:

- N990005.bam
- N990005.vcf.gz
- T990005.bam
- T990005.vcf.gz

The interface also includes a 'Controls' panel on the left, a 'Console' panel at the bottom left, and a 'Navigation' panel at the bottom right. The console panel displays the text: 'Need to filter your data?' and 'Try VarSeq'. The navigation panel shows the current coordinates: '(3: 116,584,959, 1,92) 1 - MT 3.1 Gbp'.

手順2：ターゲットとする遺伝子名の検索

The screenshot displays the Golden Helix GenomeBrowse 3.0.0 interface. The search results for 'PIK3CA' are shown in a dropdown menu, with the following entries:

PIK3CA (Gene Name)	3: 178,866,311 - 178,957,881
PIK3CA (Transcript Name)	15: 71,173,681 - 71,184,772

The main view shows genomic tracks for 'Somatic_cancer' and 'RefSeq Genes 105 Interim v1, NCBI'. The tracks include:

- Variants - T990005**: rs139255255, rs2699896, rs121913273
- Variants - N990005**: rs139255255, rs2699896
- T990005 Coverage**: Read Depth plot (0-500)
- Pile-up**: Read alignment plot (0-800)
- N990005 Coverage**: Read Depth plot (0-300)
- Pile-up**: Read alignment plot (0-600)

The console on the left shows the selected gene: **RefSeq Genes 105 Interim v1, NCBI**. The description states: "This track contains RefSeq Gene transcripts annotated by the NCBI Homo sapiens Annotation Release 105 Interim v1. This annotation contains features projected from the current RefSeq transcripts and curated genomic sequences (with accession prefixes NM_ or NR_ and NG_ respectively) placed on either the GRCh37.p13 or GRCh38.p10 assembly. The current RefSeqs include transcript variants that are new or have been updated since the last full annotation (Annotation Release 105 for GRCh37.p13 released in December 2013 or Annotation Release 108 for GRCh38.p7 released in June 2016). The GRCh37.p13 annotation does not include any Gnomon gene or transcript predictions, whereas the GRCh38.p10 annotation does include Gnomon models".

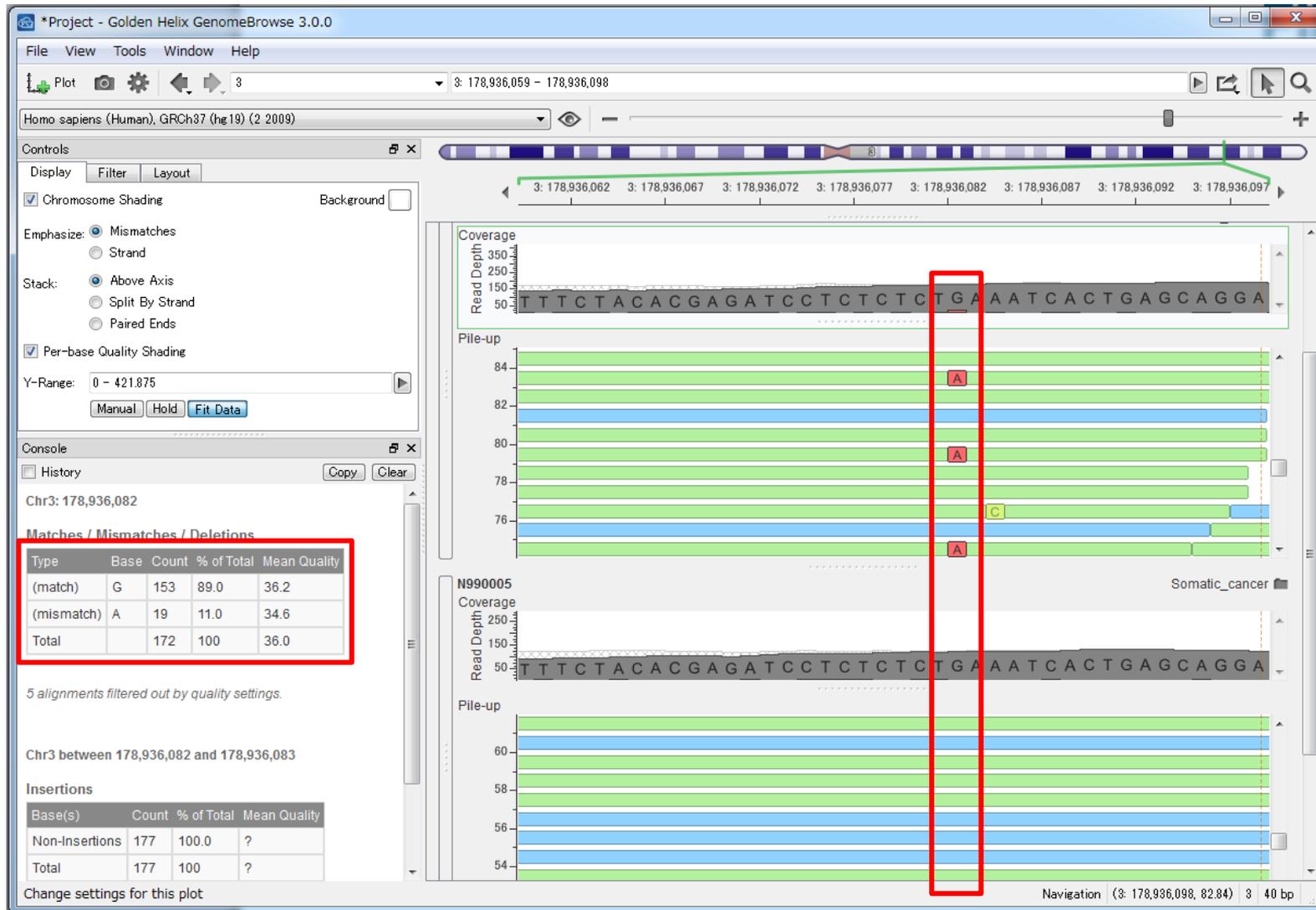
Navigation: (3: 81,214,263, 0.625) 3 91.6 Kbp

手順3：腫瘍サンプルと正常サンプルの比較による、体細胞変異の探索

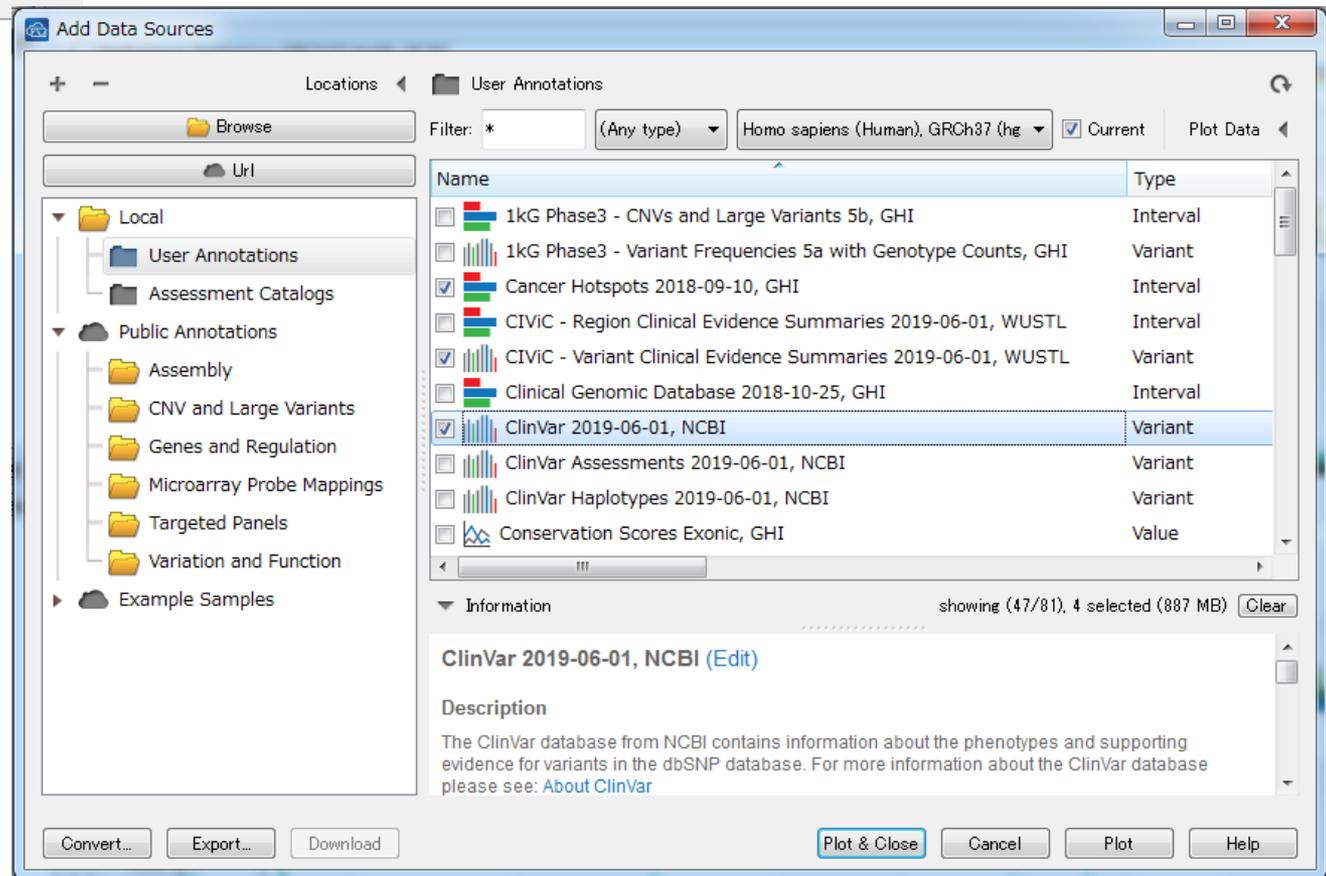
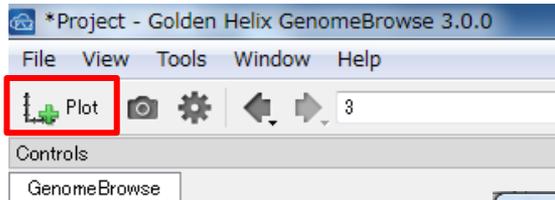
The screenshot displays the Golden Helix GenomeBrowse 3.0.0 interface. The main window shows a genomic track for chromosome 3, with a zoomed-in view of a region from 178,936,072 to 178,936,091. The track includes a reference sequence (GRCh37 hg19, UCSC) and a variants track. A somatic variant, rs121913273, is highlighted with a red box. The variant is located at position 178,936,082 on chromosome 3, with a reference allele of G and an alternate allele of A. The variant is associated with the identifier rs121913273 and has a quality score of 136.76. The track also shows coverage and read depth for the variant, with a pile-up view at the bottom.

Chr	Start	Stop	Ref/Alt	Identifier	Reference	Alternates	Quality	Filter
3	178866327	178866327	C/T	rs139255255	C	T	40.63	PASS
3	178922274	178922274	C/A	rs2699896	C	A	2314.94	PASS
3	178936082	178936082	G/A	rs121913273	G	A	136.76	PASS

手順4 : BAMファイルプロットによる、検出変異のクオリティー確認



手順5 : アノテーションリソースのプロット



手順6 : アノテーションリソースのプロットラベルを変更

The screenshot shows the Golden Helix GenomeBrowse 3.0.0 interface. The 'Controls' panel on the left has the 'Feature Labels' checkbox checked, and the 'Labels' dropdown menu is set to 'Residue', which is highlighted with a red box. The main plot area displays various genomic tracks for the region Chr3: 178,936,072 - 178,936,091. The tracks include:

- Cancer Hotspots 2018-09-10, GHI:** Shows blue boxes for residues P539, E542, and E545.
- CIVIC - Variant Clinical Evidence Summaries 2019-06-01, WUSTL:** A bar chart showing evidence for variants across different tissues like Thyroid, Bowel, Breast, and Head and Neck.
- ClinVar 2019-06-01, NCBI:** Shows variant annotations with labels like 'Likely Pathogenic'.
- gnomAD Exomes Variant Frequencies 2.0.1 v2, BROAD:** Shows variant frequencies with p-values like 4.08257e-06 and 4.08533e-06.
- Variants - T990005:** Shows variant rs121913273.
- Variants - N990005:** Shows another variant.
- RefSeq Genes 105 Interim v1, NCBI:** Shows the protein sequence PIK3CA P L S E I T E.

The 'Console' panel on the left shows the current gene (PIK3CA) and residue (E542) information, including amino acid counts and tumor types.

手順7：体細胞変異の臨床アノテーションを確認

The screenshot shows the Golden Helix GenomeBrowse 3.0.0 interface. The variant E542K is selected, and its clinical annotations are displayed in the left sidebar (highlighted with a red box) and the main panel.

Variant Information (Left Sidebar):

- Variant Name: E542K
- CIVIC Variant ID: 103
- Variant Origin: Somatic Mutation
- Variant Summary: PIK3CA E545K/E542K are the second most recurrent PIK3CA mutations in breast cancer, and are highly recurrent mutations in many other cancer types. E545K, and possibly the other mutations in the E545 region, may present patients with a poorer prognosis than patients with other PIK3CA variants or wild-type PIK3CA. There is also data to suggest that E545/542 mutations may confer resistance to EGFR inhibitors like cetuximab. While very prevalent, targeted therapies for variants in PIK3CA are still in early clinical trial phases.
- Phenotypes: ?
- Disease: Colorectal Cancer
- Disease Ontology ID: 9256
- Drugs: Regorafenib
- Drug Interaction Type: ?

Main Panel Annotations:

- Cancer Hotspots 2018-09-10, GHI:** Shows hotspots P539, E542, and E545.
- CIVIC - Variant Clinical Evidence Summaries 2019-06-01, WUSTL:** A bar chart showing evidence across various cancer types: Bowel, Breast, Head and Neck, Lung, and Skin.
- ClinVar 2019-06-01, NCBI:** Shows ClinVar entries for E542K, all classified as "Likely Pathogenic".
- gnomAD Exomes Variant Frequencies 2.0.1 v2, BROAD:** Shows variant frequencies across populations, with a frequency of 4.08257e-06.
- Variants - T990005:** Shows variant rs121913273.
- Variants - N990005:** Shows variant N990005.
- RefSeq Genes 105 Interim v1, NCBI:** Shows the PIK3CA gene structure with the variant E542K highlighted.

Navigation: (3: 178,936,091, 35) 3 20 bp

GenomeBrowse®ダウンロード :

<http://goldenhelix.com/products/GenomeBrowse/index.html#download>

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