

# Partek Flow DNA-Seq Analysis

Eric

# Import Data

- Import data by **Automatically create samples from files**
- Browse to **/home/flow/FlowData/DNA-seq**
- Select the available **.bam** file
- Select data type as **DNA-Seq**

Import data

Automatically create samples from files

Import samples from data repository

Create a new blank sample

Import single cell RNA-seq count matrix

Import single cell ATAC matrix

Import single cell V(D)J data

Import bcl files

Select files from [Paretek Flow Server](#) [My Computer](#) [URL](#) [GEO / ENA project](#)

Select files

Current directory [/home/flow/FlowData/NGSTrainingFiles/DNA-seq\\_Data](#) [Goto](#)

Server Computer

- home
  - flow
    - ATAC-seq Data
    - FlowData
    - library\_files
    - NGSTrainingFiles
      - Active
        - CHIP-seq\_Data
        - DNA-seq\_Data
      - Gene Lists
      - NGSTrainingFiles
      - RNA-seq-2Factors
      - RNA-seq\_SAZA
      - scRNA-PBMC3k
      - scRNA\_normalized
      - Single\_Cell\_PBMC3k
      - STAR 2.5.3a index
      - tutorial\_data
    - Output
      - RNA-seq\_SAZA
      - scRNA-PBMC3k

1 files selected	
Name	Size
<input checked="" type="checkbox"/> Exome-chr22.bam	7.63 MB

Valid files are: CEL, bam, bcf, bgx, bgz, bpm, fa, fa.gz, fasta, fasta.gz, fastq, fastq.gz, fna, fna.gz, fq, fq.gz, idat, probe\_tab, raw, sam, sff, sra, tar, tar.gz, txt, vcf, vcf.gz or zip

Select data type

DNA-Seq

[Back](#) [Create sample](#)

# Coverage Report

Once the download completes, the sample table will appear in the Data tab.

- Click **Analyses** tab
- Click **DNA-seq** data node
- Click **Coverage report** in the QA/QC section of the task menu



# Coverage Report

- Set the Assembly to **Homo sapiens (human) – hg19\_chr22** and the Annotation model to **RefSeq**
- Click **Finish** to run

Select Annotation file

Assembly

Annotation model

Add minimum coverage levels

Coverage level *i* 1x  20x  100x

x

Advanced options

Strand specificity *i*

Generate target enrichment graph *i*

Use multithreading *i*

# Coverage Report

- Double click the **Coverage Report** node
- Click a sample name for a sample-level table

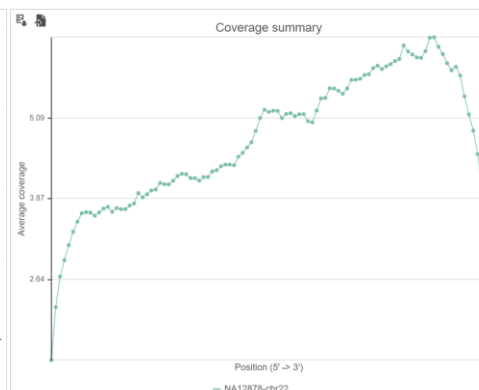
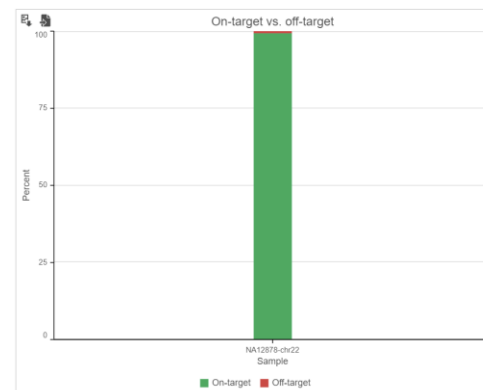
Annotation model RefSeq  
Strand-specificity No

Optional columns

Sample name	1x	20x	100x	Average coverage	Average quality	On-target alignments	Off-target alignments
NA12878-chr22	30.76%	3.74%	0%	3.46	24.02	215925	1488

Rows per page 25 (1 of 1) Download

Region coverage summary  
Download all samples region reports



Region name	Chromosome	Start	Stop	Strand	Total exon length	Reads	% GC	% N	1x	20x	100x	Average coverage	Average quality	
NR_122113>DUXAP8	22	16,150,529	16,193,010	-	2,107	375	44.98%	0.31%	97.06%	14.19%	0%	10.08	25.23	all ↕
NM_001136213>POTEH	22	16,256,332	16,287,938	-	2,037	473	47.79%	0.17%	94.31%	1.52%	0%	5.46	24.74	all ↕
NR_046571>POTEH-AS1	22	16,274,609	16,277,578	+	1,492	285	45.17%	0.15%	99.73%	0%	0%	9.67	25.85	all ↕
NM_001005239>OR11H1	22	16,448,824	16,449,805	-	981	197	43.56%	0.37%	100.00%	2.14%	0%	10.18	25.65	all ↕
NM_014406>CCT8L2	22	17,071,648	17,073,701	-	2,053	541	55.70%	0.15%	100.00%	8.57%	0%	13.25	24.12	all ↕
NR_001591>TPTEP1	22	17,082,801	17,129,721	+	1,430	271	52.25%	0.23%	96.43%	17.34%	0%	9.81	24.35	all ↕
NR_040115>ANKRD62P1-PARP4P3	22	17,134,599	17,156,431	-	1,538	218	38.39%	0.23%	91.55%	0.98%	0%	7.04	26.01	all ↕
NM_001318251>XKR3	22	17,264,306	17,302,590	-	1,689	248	38.64%	0.18%	97.63%	0%	0%	6.83	26.11	all ↕
NM_175878>XKR3	22	17,264,306	17,302,590	-	1,686	248	38.60%	0.18%	97.63%	0%	0%	6.83	26.11	all ↕
NR_003607>HSFY1P1	22	17,308,364	17,310,226	+	1,381	357	37.36%	0.22%	98.19%	4.63%	0%	11.94	26.05	all ↕
NM_001037814>GAB4	22	17,442,827	17,489,113	-	2,629	426	56.81%	0.34%	93.65%	2.09%	0%	7.94	23.71	all ↕
NR_015352>CECR7	22	17,517,460	17,540,961	+	4,004	678	56.53%	0.17%	67.96%	5.24%	0%	8.72	23.74	all ↕
NR_152825>CECR7	22	17,517,460	17,540,961	+	3,942	678	56.44%	0.17%	67.45%	5.33%	0%	8.66	23.73	all ↕
NR_152826>CECR7	22	17,517,460	17,540,961	+	3,560	678	58.38%	0.20%	63.96%	4.52%	0%	7.85	23.26	all ↕
NM_001289905>IL17RA	22	17,565,849	17,596,585	+	8,506	7	0%	0%	0.01%	0%	0%	0	6.00	all ↕
NM_014339>IL17RA	22	17,565,849	17,596,585	+	8,608	7	59.94%	0%	1.07%	0%	0%	0.04	19.04	all ↕
NM_001163079>TMEM121B	22	17,597,189	17,602,258	-	3,961	118	69.84%	1.27%	9.67%	0%	0%	0.28	20.06	all ↕
NM_031890>TMEM121B	22	17,597,189	17,602,214	-	5,025	118	72.93%	0.30%	28.80%	0%	0%	1.13	19.94	all ↕
NR_103793>LINC01664	22	17,602,485	17,612,995	+	1,175	206	54.75%	0.39%	99.49%	0%	0%	8.96	23.29	all ↕
NM_017829>HDHD5	22	17,618,410	17,646,178	-	1,726	86	48.18%	0%	4.58%	0%	0%	0.22	27.57	all ↕
NM_033070>HDHD5	22	17,618,410	17,640,170	-	1,801	3	81.60%	0%	6.22%	0%	0%	0.09	17.48	all ↕
NR_024482>HDHD5-AS1	22	17,640,279	17,646,336	+	340	83	71.83%	1.19%	42.06%	0%	0%	1.97	21.05	all ↕
NR_024483>HDHD5-AS1	22	17,640,279	17,646,336	+	626	83	65.04%	0.35%	68.53%	0%	0%	5.94	22.14	all ↕
NM_001282225>ADA2	22	17,659,680	17,700,476	-	4,505	53	56.37%	0%	3.88%	0%	0%	0.18	24.77	all ↕
NM_001282226>ADA2	22	17,659,680	17,700,476	-	4,472	53	56.69%	0%	3.18%	0%	0%	0.13	23.90	all ↕

Rows per page 25 (1 of 49) Download

# Filter Alignments

- Click the project name to return to the Analyses tab
- Select the **DNA-Seq** node
- Click **Filter alignments** in the Post-alignment tools

The screenshot displays a software interface with a top navigation bar containing tabs: Analyses, Data, Log, Project settings, Notebook, Data viewer, and Attachments. A purple 'List generator' button is located in the top right corner. The main workspace shows a 'DNA-Seq' node, represented by a blue circular icon with a white bar chart, which is highlighted with a red square. An arrow points from this node to a 'Coverage report' icon. On the right side, a dropdown menu is open, showing a list of options under the 'DNA-Seq' header. The 'Post-alignment tools' section is expanded, and the 'Filter alignments' option is highlighted with a red rectangle. Other options in the menu include 'Partek development', 'QA/QC', 'Downscale alignments', 'Convert alignments to unaligned reads', 'Combine alignments', 'Deduplicate UMIs', 'AGeNT LocatIt', and 'Quantification'.

# Filter Alignments

- To get rid of the PCR duplicates check the **Filter duplicates** option
- Choose **Consider each read end separately** and duplicate as **Same start and same sequence** option
- Select **Filter low mapping quality**
- Push **Finish**

The screenshot shows the 'Filter Alignments' interface with the following settings:

- Filter duplicates** (highlighted in red)
- Paired-end reads:  Consider both ends together,  **Consider each read end separately** (highlighted in red)
- Keep duplicates up to: 1
- Treat the alignment as duplicate if:  Same start position,  **Same start and same sequence** (highlighted in red)
- Keep the alignment with:  Highest mapping score,  Randomly selected
- Filter low mapping quality** (highlighted in red)
- Min mapping quality: 20
- Filter alignments with mismatches
- Max mismatched bases: 2
- Filter by sequence names
- Filter by sequence names: include, chromosome, in, Terms, OR, AND
- Filter by genomic locations
- Include region overlapping alignments:
- Exclude region overlapping alignments:
- Additional filters:
  - Remove singletons:
  - Remove unaligned reads:
  - Remove not primary reads:
  - Remove failed platform/vendor quality reads:
  - Remove pcr/optical duplicates:


Buttons: Back, Finish

# Detect Variants

- Select the **Filtered reads** data node
- Choose **Samtools** in Variant Callers from the task menu
- Choose Assembly to **Homo sapiens (human) – hg19\_chr22** as reference
- Click **Finish**

Variant detection method

Select Reference sequence

Assembly 

Advanced options

Option set  [Configure](#)

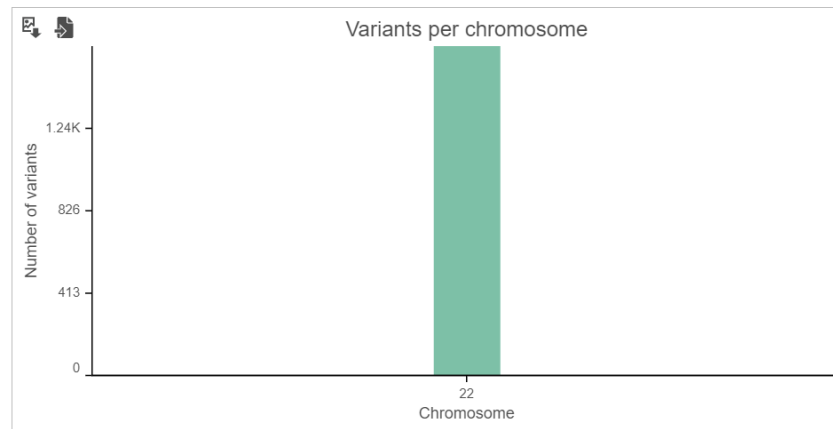


# Inspect Variants

- Select the **Variants** data node
- Choose **View variants** in Variant analysis from the task menu
- Use default settings and click **Finish**
- Double click the **View variants** task node for results

Sample name ↕	Average variant quality ↕	Variants ↕	SNVs ↕	Indels ↕	Ti/Tv ratio ↕	Het/Hom ratio ↕
NA12878-chr22	60.64	1,652	1,552	100	2.12274 (1055/497)	3.39 (1276/376)

Rows per page: 25 | (1 of 1) | [Download](#)



# Annotate Variants

- Click the project name to return to the Analyses tab
- Select the **Variants** data node
- Choose **Annotate variants** in Variant analysis
- Select **Annotate with genomic features** and point to **RefSeq**
- Select **Annotate with known variants** and point to **dbSNP**
- Click **Finish** to start

Assembly Homo sapiens (human) - hg19\_chr22

Annotate with genomic features

Annotation model RefSeq (Administrator) v

Promoter upstream limit *i* 5000

Promoter downstream limit *i* 5000

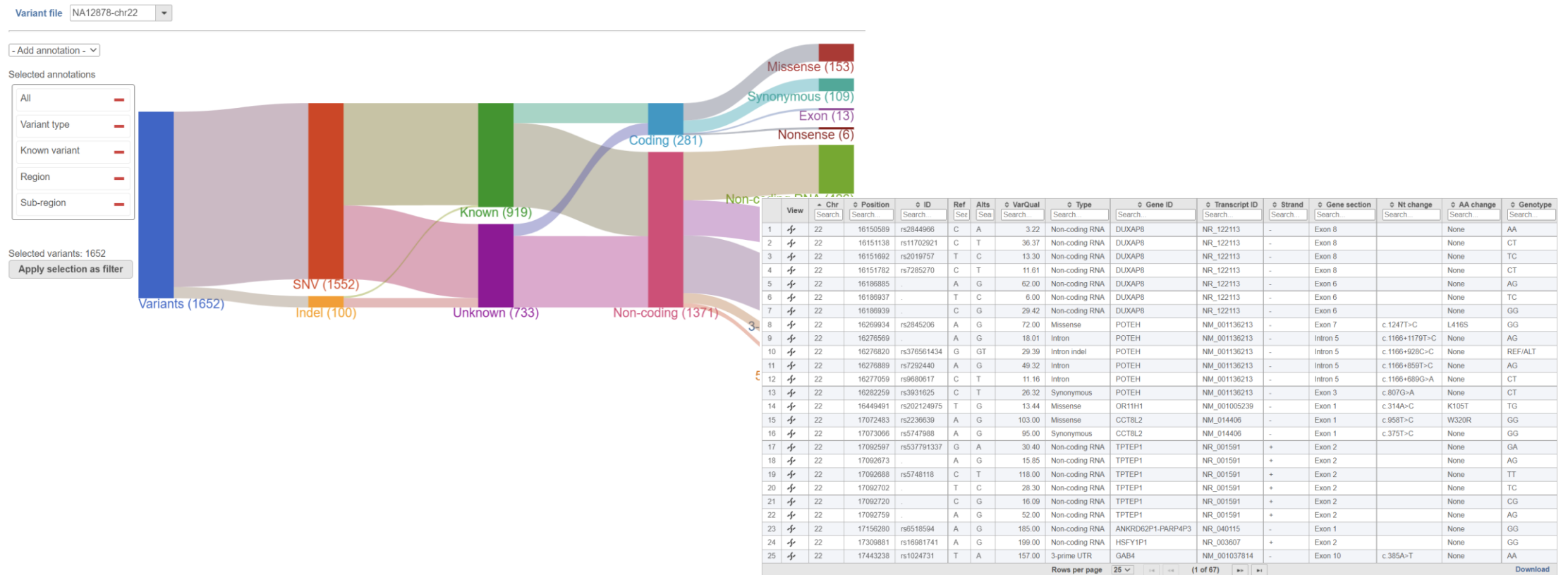
Annotate with known variants

Variant annotation dbSNP (Simit Patel) v

Back Finish

# Variants Report

- Double click the **Annotated variants** task node for reports



# Filter Variants

- Click the project name to return to the Analyses tab
- Select the **Annotated variants** data node
- Choose **Filter variants** in Variant analysis
- Set Minimum read depth to **20**
  - The filter shows variants covered at least 20x
- Set Minimum alternate calls to **10**
  - The filter shows variants with at least 10 supporting reads
- Push **Finish**

Annotation

Include region overlapping variants

Exclude region overlapping variants

Samples

Filter by samples

Variant Type

Indels  Indels only  Non-Indels  All

Zygosity  Heterozygous only  Homozygous only  All

Mutation type  Synonymous  Missense  Nonsense  Exonic indels

Feature section  Splice-5  Splice-3  Non-coding RNA  5-prime UTR  3-prime UTR  Intron  Promoter  Intergenic

Variant Novelty

Known variants  Known only  Novel only  All

Quality

Minimum read depth  20

Minimum high-quality alternate calls  10

Minimum mean-square mapping quality  10.0

Minimum variant quality  30.0

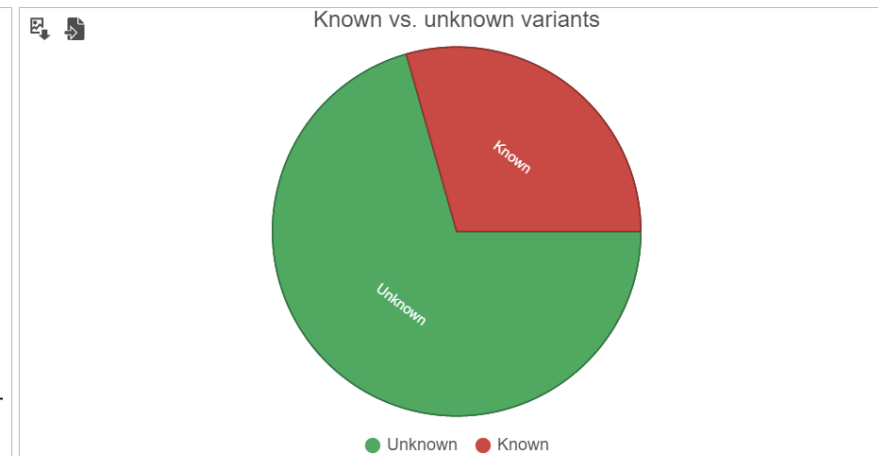
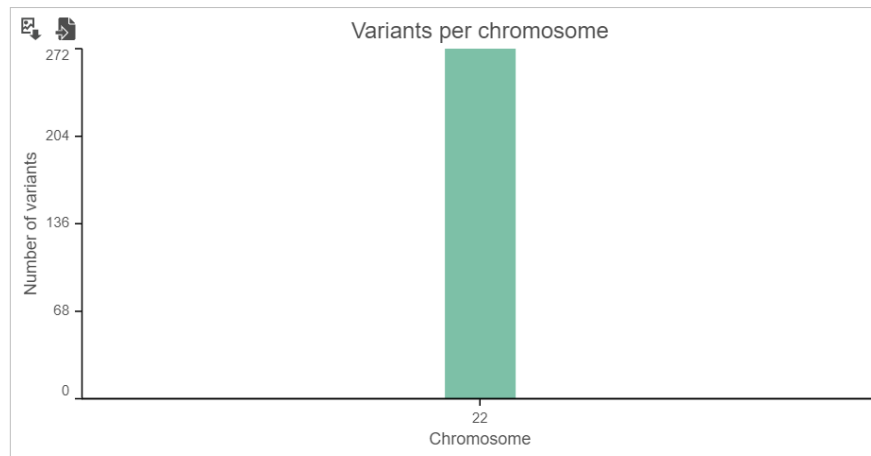
Back Finish

# Inspect Variants

- Select the **Variants** data node
- Choose **View variants** in Variant analysis from the task menu
- Use default settings and click **Finish**
- Double click the **View variants** task node for results

Sample name	Average variant quality	Variants	SNVs	Indels	Known variants	% known variants	Ti/Tv ratio	Het/Hom ratio
NA12878-chr22	83.20	272	263	9	80	29.41%	2.05814 (177/86)	23.73 (261/11)

Rows per page: 25 (1 of 1) [Download](#)



# Filter Results Table




- Click a sample name for a sample-level table
- To select missense variants only, type **Missense** under Type

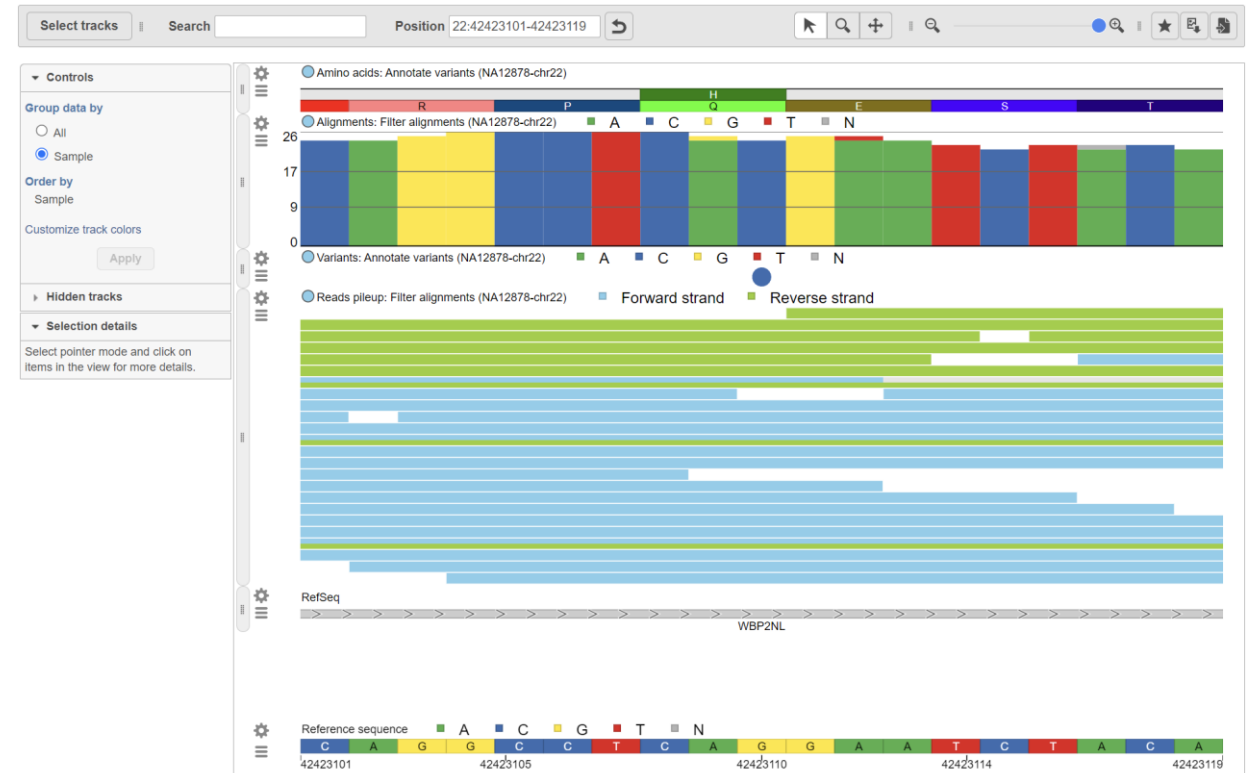
Sample NA12878-chr22 Variants 272  Show one overlap per variant  Apply precedence rules  Show all overlaps [i](#)

Optional columns

	View	▲ Chr	↕ Position	↕ ID	Ref	Alts	↕ VarQual	↕ Type	↕ Gene ID	↕ Transcript ID	↕ Strand	↕ Gene section	↕ Nt change	↕ AA change	↕ Genotype
		Search...	Search...	Search...	See	See	Search...	Missense	Search...	Search...	Search...	Search...	Search...	Search...	Search...
1		22	42423110	rs2301521	G	C	216.00	Missense	WBP2NL	NM_152613	+	Exon 6	c.855G>C	Q285H	CC
Rows per page 25 <input type="button" value="←"/> <input type="button" value="&lt;&lt;"/> (1 of 1) <input type="button" value="&gt;&gt;"/> <input type="button" value="→"/>															
<a href="#">Download</a>															

# Exploring Browser View

- Click on  to browse in chromosome view
- Click on **Select tracks** to add or remove tracks
- On each track, click on  to configure the track
- Click on  to hide or pin track
- Drag and drop to change the order of the tracks



# DNA-Seq Pipeline

