



**Extending the Curio Genomics Platform for the Wheat Research Community
DNA Sequences to Analysis Results**

Presented April 30th, 2020 to the IWGSC Community

Author: Shawn Quinn (CTO, Curio Genomics)

Extending the Curio Genomics Platform for the
Wheat Research Community:
DNA Sequences to Analysis Results

Presentation Overview

- What is Curio
- Demonstration of Several Crop/Wheat Research Examples
 - DNA-Seq and RNA-Seq
 - Read Mapping Analysis
 - Variant Analysis
 - Coverage Analysis
 - Expression Analysis
- Looking Ahead
- Acknowledgements

What is Curio?

- Modern big data management and genomic analysis platform, fully web-based, collaboration ready
- Supports both bioinformatic processing and scientific interpretive analysis
- Provides scalable data processing and interactive data visualizations using real-time databases and clustering technologies
- Designed for extensibility to continuously support new analysis methods, data types, etc.
- Includes complex crop research solutions, including tetraploid and hexaploid wheat DNA-Seq and RNA-Seq analysis



DNA-Seq Hexaploid Wheat: Read Mapping and Visualization

- Incorporates Chinese Spring Wheat (*Triticum aestivum*) reference assembly from the IWGSC (i.e. the International Wheat Genome Sequencing Consortium)

Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

▶ Start An Alignment ▶

▶ Create Paired FASTQ File ▶

File Name	Progress	Date	Action
<input type="checkbox"/> SampleA-R2.fastq.gz	None	2019-12-06 17:23	▶ Start an Alignment ▶
<input type="checkbox"/> SampleA-R1.fastq.gz	2 Completed,	2019-12-06 17:23	▶ Start an Alignment ▶
<input checked="" type="checkbox"/> SampleA Paired	16 Completed,	2019-12-06 17:00	▶ Start an Alignment ▶

Viewing 1-3

Align Paired-End Sequence Files

Select Type: Bowtie 2: Paired Alignment

Assembly: IWGSC WGA 1.0 (Chinese Spring Wheat)

Alignment

Pre-Processing

Scoring

Paired FASTQ

Read Trimming: ? Do not perform any read trimming
 Trim the reads before aligning

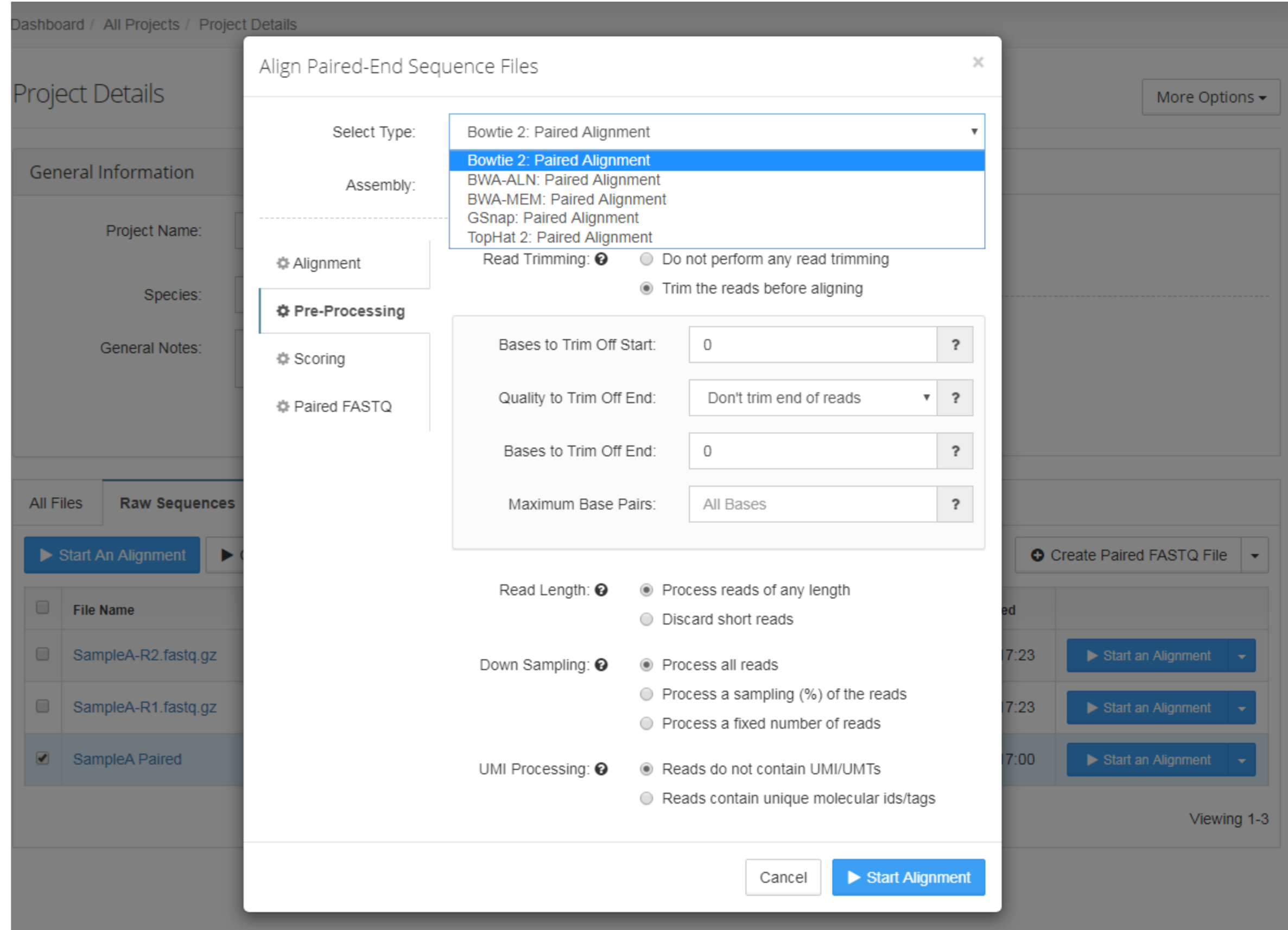
Read Length: ? Process reads of any length
 Discard short reads

Down Sampling: ? Process all reads
 Process a sampling (%) of the reads
 Process a fixed number of reads

UMI Processing: ? Reads do not contain UMI/UMTs
 Reads contain unique molecular ids/tags

Cancel ▶ Start Alignment

- Incorporates Chinese Spring Wheat (*Triticum aestivum*) reference assembly from the IWGSC (i.e. the International Wheat Genome Sequencing Consortium)
- Multiple read mapping algorithms with pre-built indexes that are deployed and ready on a computational cluster
- Experiment with various alignment and read processing options without requiring any pipeline configuration



Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Start An Alignment

File Name

SampleA-R2.fastq.gz

SampleA-R1.fastq.gz

SampleA Paired

Align Paired-End Sequence Files

Select Type: Bowtie 2: Paired Alignment

Assembly:

Alignment

Pre-Processing

Scoring

Paired FASTQ

Read Trimming: Do not perform any read trimming Trim the reads before aligning

Bases to Trim Off Start: 0 ?

Quality to Trim Off End: Don't trim end of reads ?

Bases to Trim Off End: 0 ?

Maximum Base Pairs: All Bases ?

Read Length: Process reads of any length Discard short reads

Down Sampling: Process all reads Process a sampling (%) of the reads Process a fixed number of reads

UMI Processing: Reads do not contain UMI/UMTs Reads contain unique molecular ids/tags

Cancel Start Alignment

More Options

Create Paired FASTQ File

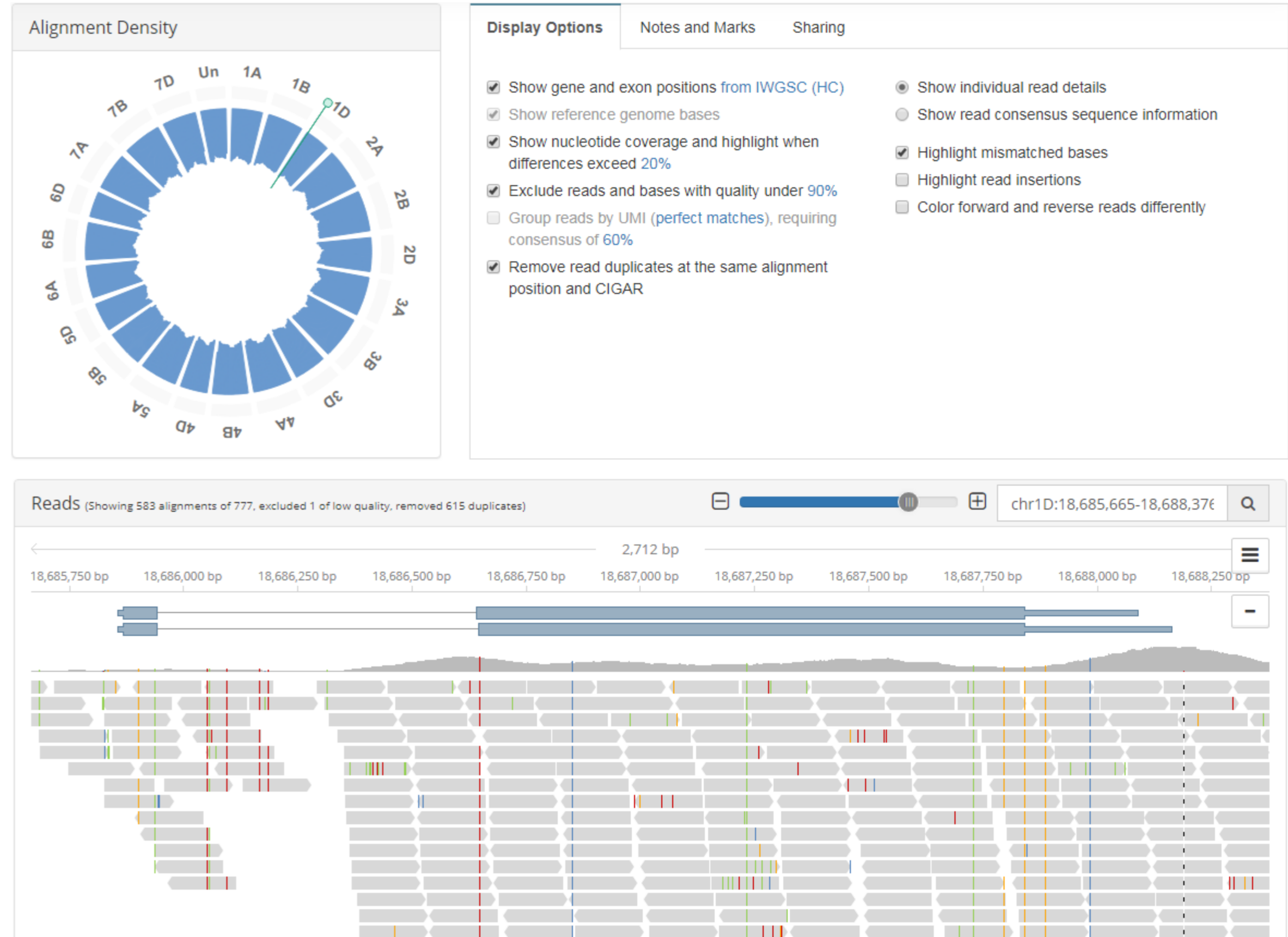
Start an Alignment

Start an Alignment

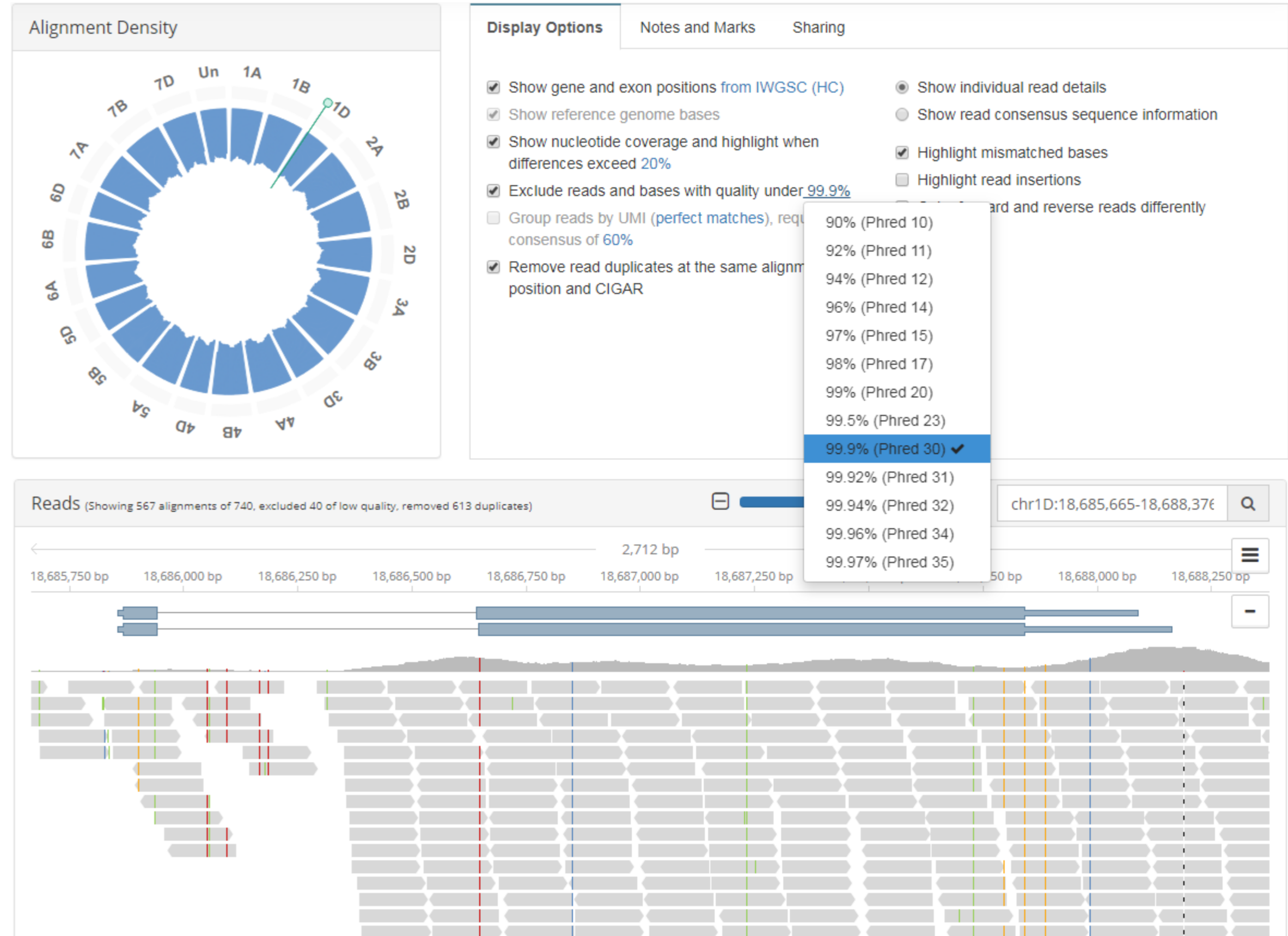
Start an Alignment

Viewing 1-3

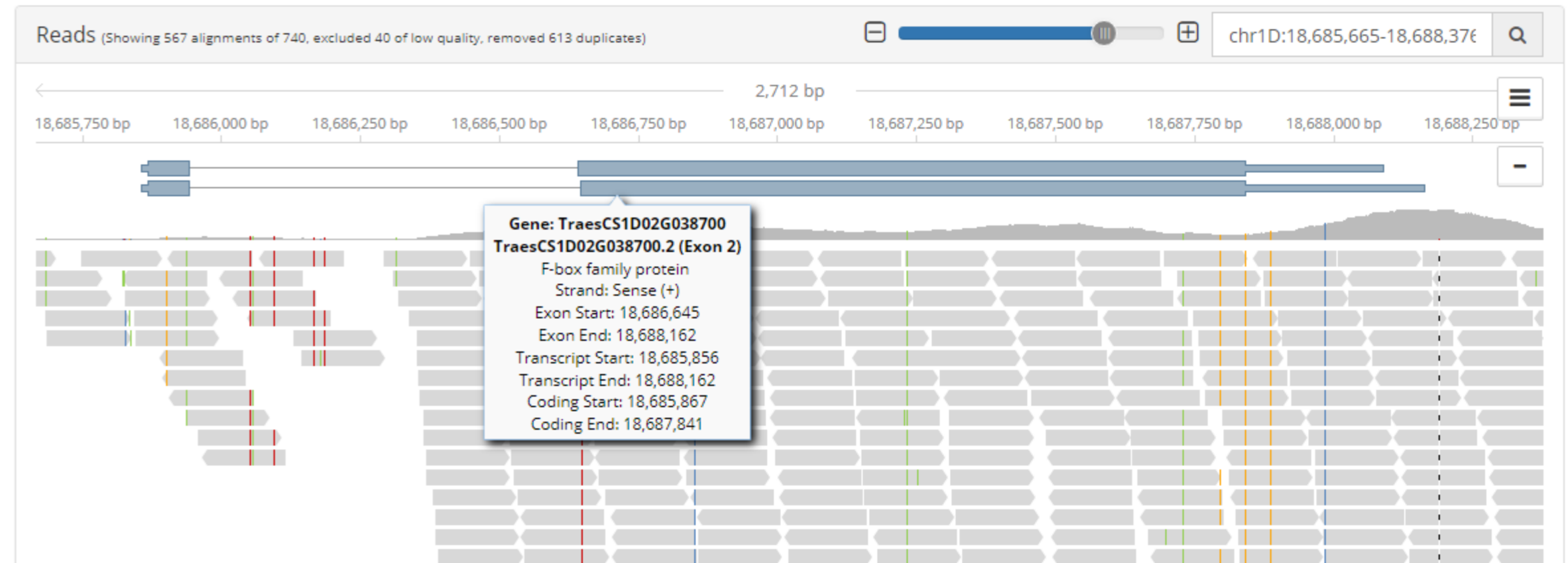
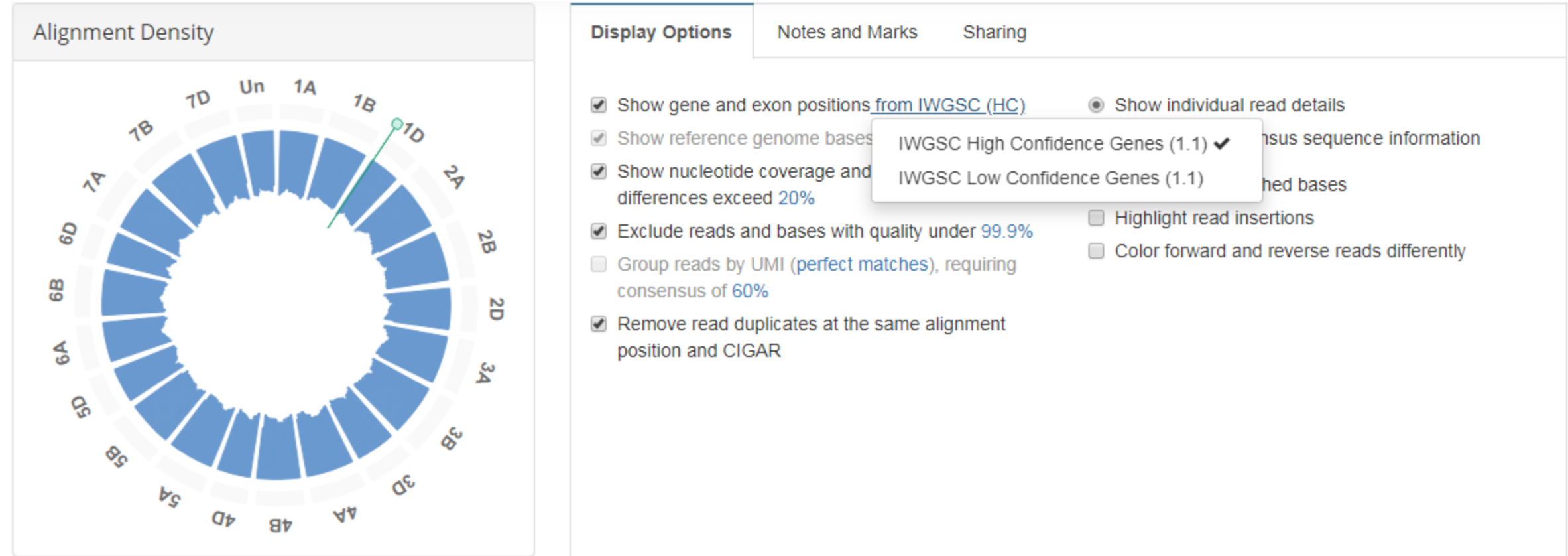
- Quickly browse and visualize reads from samples of any size, anywhere in the genome



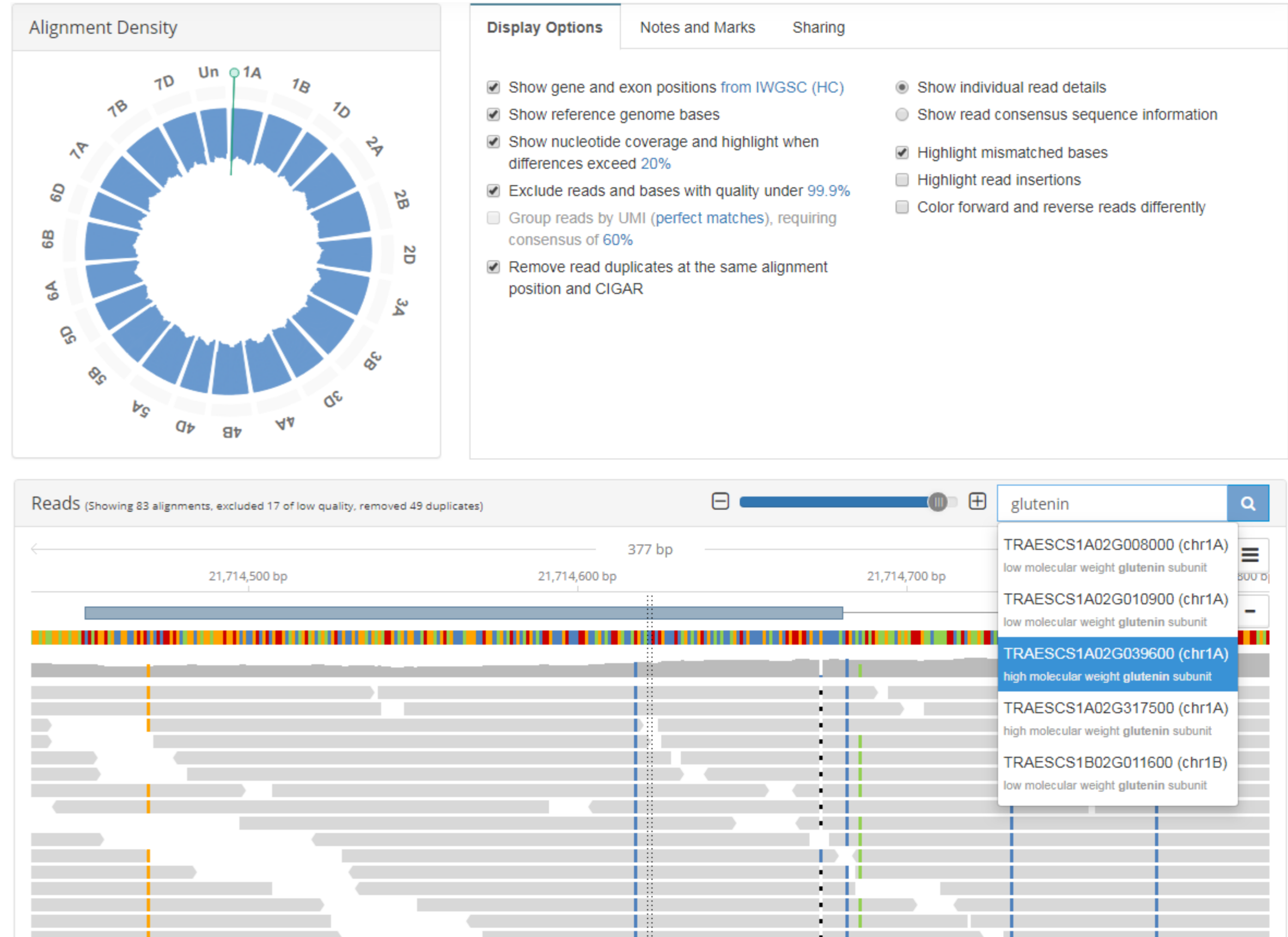
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly



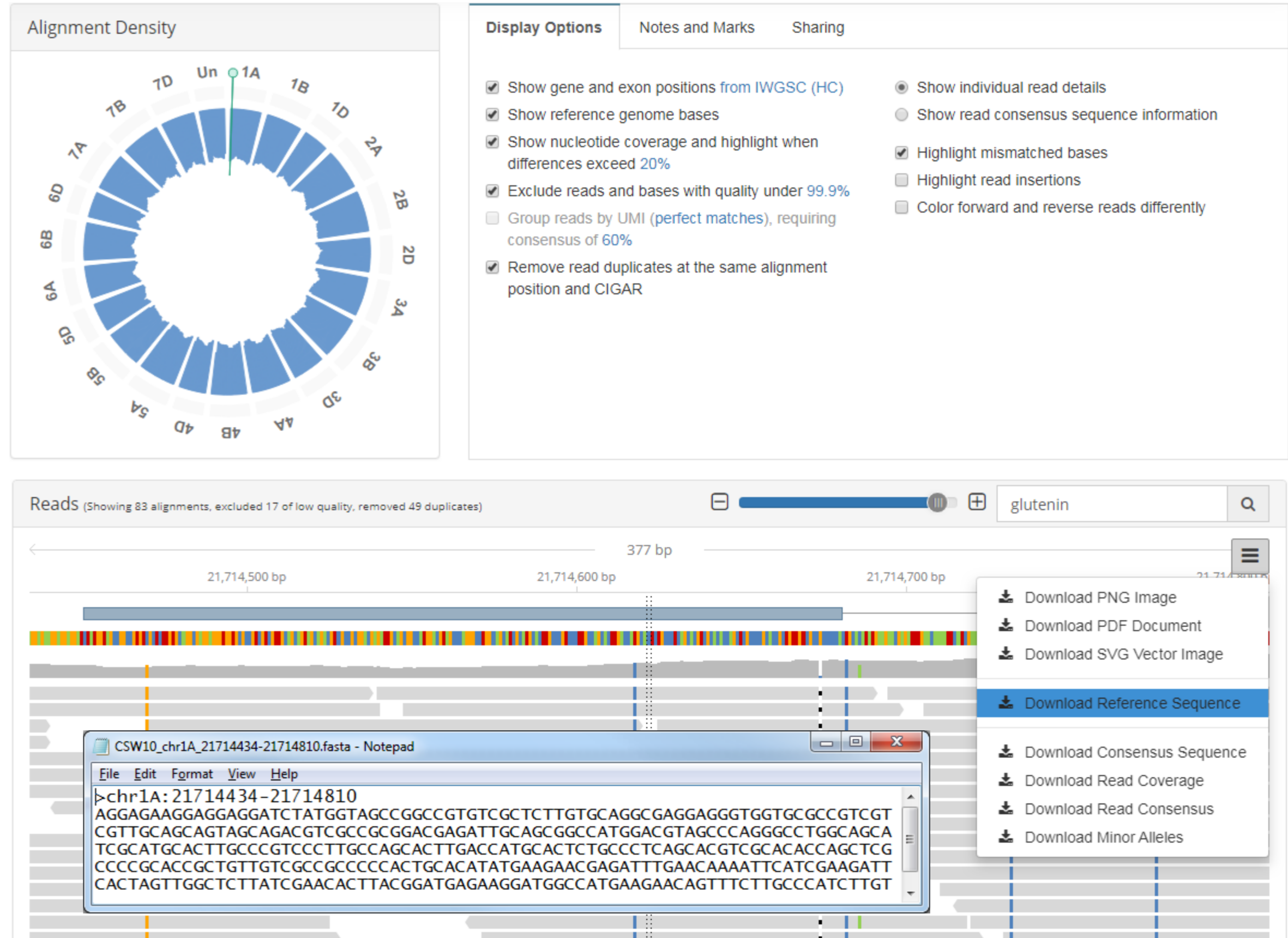
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline



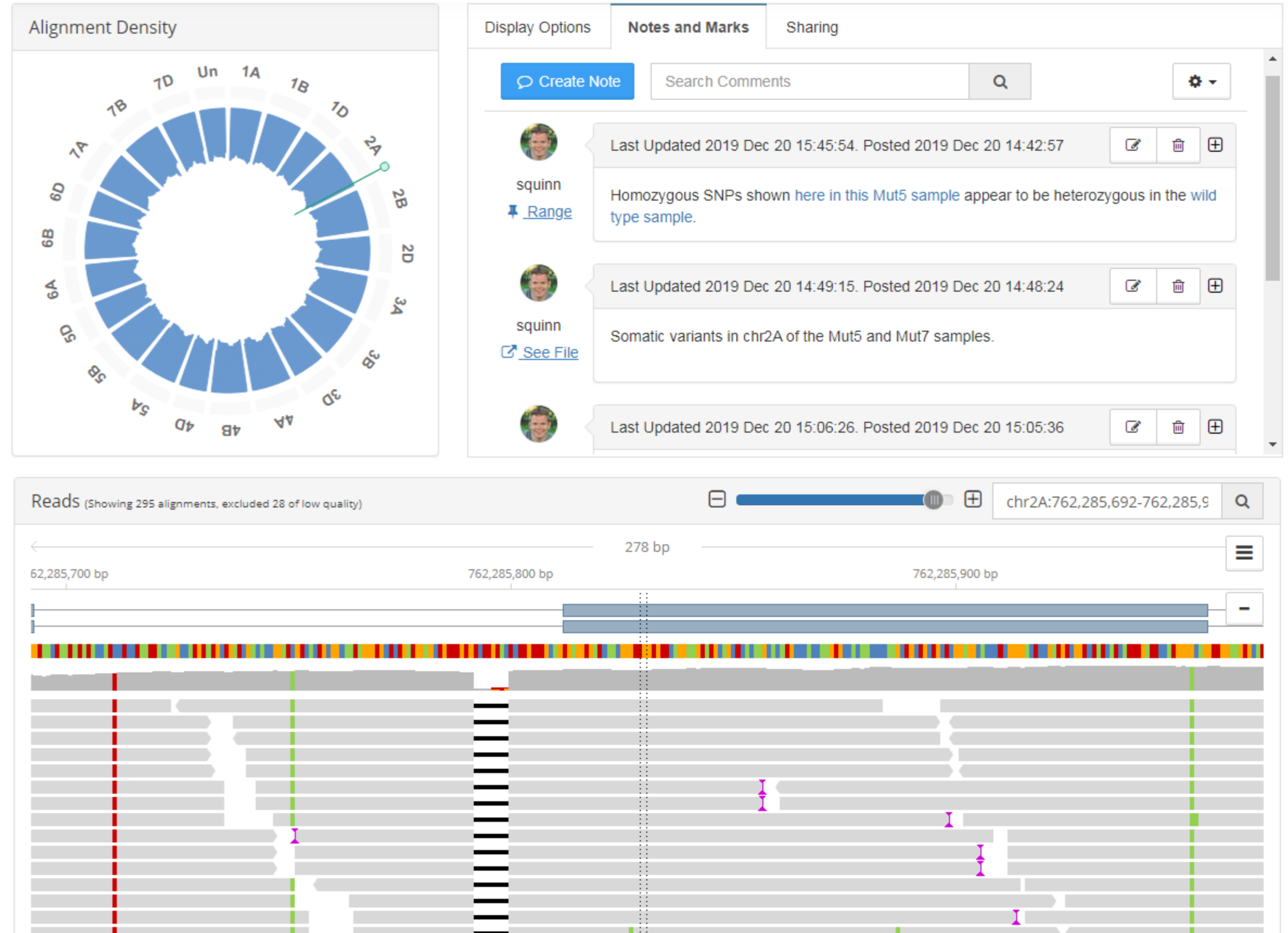
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline
- Quickly search and navigate the genome using the IWGSC annotation data as well



- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline
- Quickly search and navigate the genome using the IWGSC annotation data as well
- Conveniently access reference or consensus sequences, and read coverage details



- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline
- Quickly search and navigate the genome using the IWGSC annotation data as well
- Conveniently access reference or consensus sequences, and read coverage details
- Navigate and collaborate through annotated comments



DNA-Seq Hexaploid Wheat: Variant Analysis

- Call variants leveraging the Chinese Spring Wheat reference assembly and related annotations from the IWGSC
- Operate on one or multiple samples of any read depth simultaneously
- Conveniently adjust for sensitivity and specificity

Dashboard / All Projects / Project Details

Project Details More Options ▾

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Available Actions ▾

File Name	Size	Created	Actions
<input checked="" type="checkbox"/> CultivarE-WT - BWA-Align		2019-12-20 05:59	View Alignments ▾
<input checked="" type="checkbox"/> CultivarE-Mut5 - BWA-Align (Aligned 2019-12-19 20:28 UTC)	224,364,008	2019-12-20 05:50	View Alignments ▾
<input checked="" type="checkbox"/> CultivarE-Mut7 - BWA-Align (Aligned 2019-12-19 20:28 UTC)	186,030,770	2019-12-20 04:08	View Alignments ▾
<input checked="" type="checkbox"/> CultivarE-WT - Bowtie2 (Aligned 2019-12-18 18:24 UTC)	237,564,990	2019-12-18 21:55	View Alignments ▾
<input checked="" type="checkbox"/> CultivarE-Mut5 - Bowtie2 (Aligned 2019-12-18 18:24 UTC)	224,364,008	2019-12-18 17:01	View Alignments ▾
<input checked="" type="checkbox"/> CultivarE-Mut7 - Bowtie2 (Aligned 2019-12-18 18:24 UTC)	186,030,770	2019-12-18 16:21	View Alignments ▾

Start Multiple Variant Detection Analyses ✕

Genome Feature Set:

De-duplication: Remove read duplicates at the same position and CIGAR

Minimum Quality: Phred: 30 (99.9%)

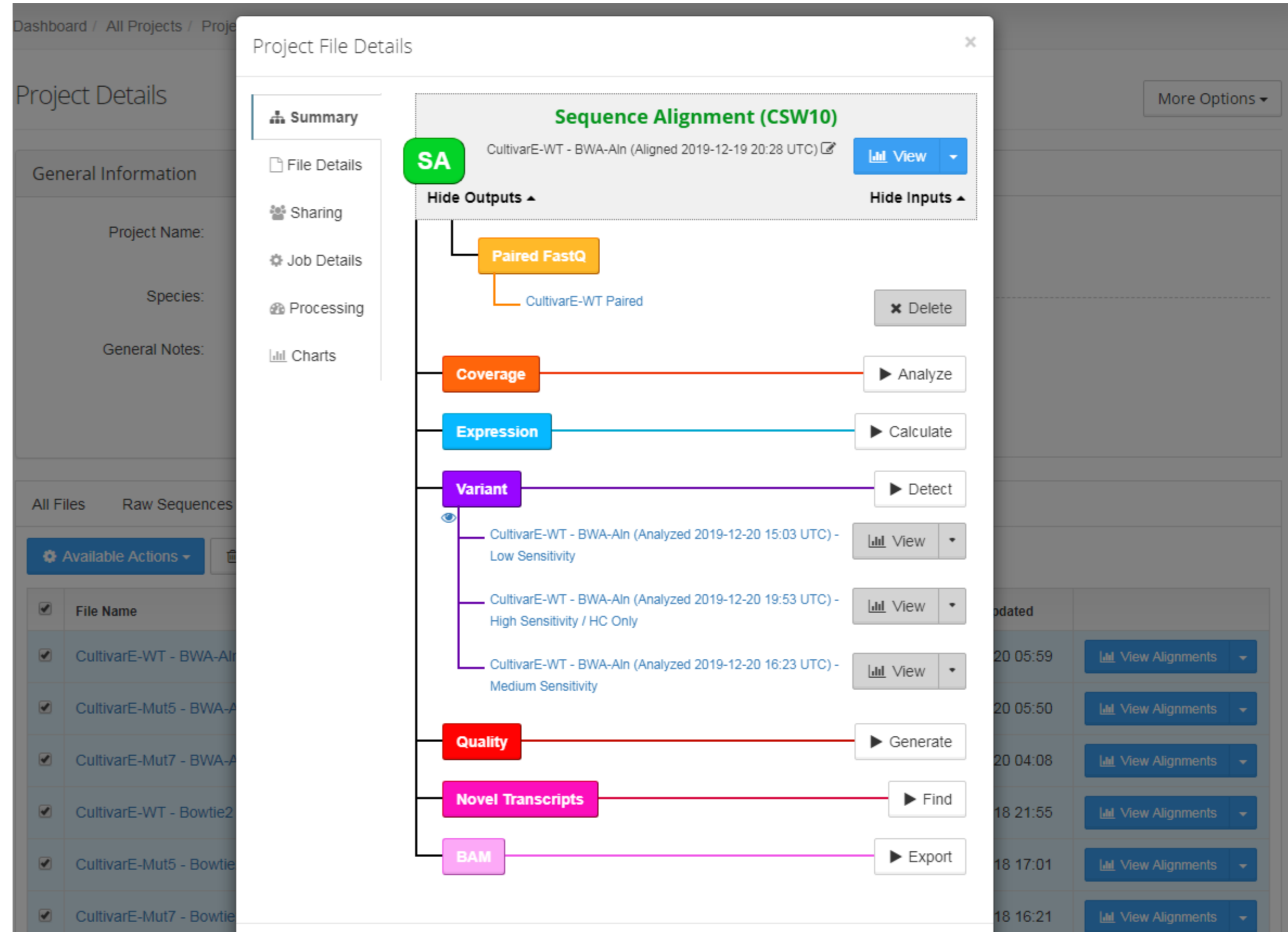
Min Read Coverage: 10 Reads

Heterozygous Positions: Use read and base quality data in heterozygous calls
 Use nucleotide frequencies only in heterozygous calls

Min Variant Frequency: 20%

Rare Allele Detection: Do not attempt to detect rare alleles
 Include rare allele detection in the analysis

- Call variants leveraging the Chinese Spring Wheat reference assembly and related annotations from the IWGSC
- Operate on one or multiple samples of any read depth simultaneously
- Conveniently adjust for sensitivity and specificity
- Automatically track the provenance of every file...



Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Available Actions

File Name

✓ CultivarE-WT - BWA-Aln

✓ CultivarE-Mut5 - BWA-A

✓ CultivarE-Mut7 - BWA-A

✓ CultivarE-WT - Bowtie2

✓ CultivarE-Mut5 - Bowtie

✓ CultivarE-Mut7 - Bowtie

Project File Details

Summary

File Details

Sharing

Job Details

Processing

Charts

Sequence Alignment (CSW10)

CultivarE-WT - BWA-Aln (Aligned 2019-12-19 20:28 UTC)

View

Hide Outputs

Hide Inputs

Paired FastQ

CultivarE-WT Paired

Delete

Coverage

Analyze

Expression

Calculate

Variant

Detect

CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 15:03 UTC) - Low Sensitivity

View

CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

View

CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 16:23 UTC) - Medium Sensitivity

View

Quality

Generate

Novel Transcripts

Find

BAM

Export

More Options

Updated

20 05:59

View Alignments

20 05:50

View Alignments

20 04:08

View Alignments

18 21:55

View Alignments

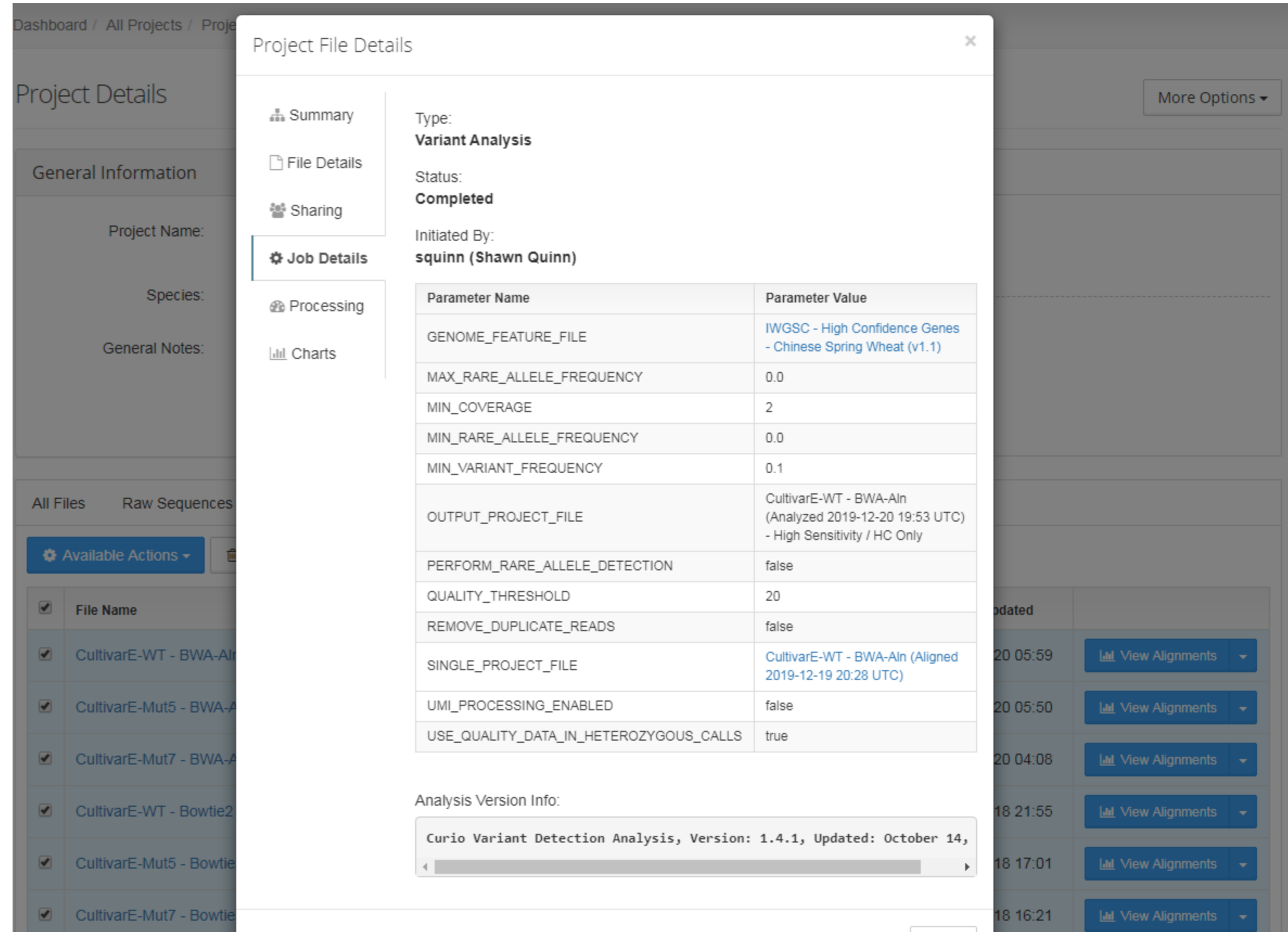
18 17:01

View Alignments

18 16:21

View Alignments

- Call variants leveraging the Chinese Spring Wheat reference assembly and related annotations from the IWGSC
- Operate on one or multiple samples of any read depth simultaneously
- Conveniently adjust for sensitivity and specificity
- Automatically track the provenance of every file...
- ...including the specific options that went into any analysis



The screenshot displays the CURIO web interface. A modal window titled "Project File Details" is open, showing the following information:

- Type:** Variant Analysis
- Status:** Completed
- Initiated By:** squinn (Shawn Quinn)

The modal also contains a table of parameters used in the analysis:

Parameter Name	Parameter Value
GENOME_FEATURE_FILE	IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)
MAX_RARE_ALLELE_FREQUENCY	0.0
MIN_COVERAGE	2
MIN_RARE_ALLELE_FREQUENCY	0.0
MIN_VARIANT_FREQUENCY	0.1
OUTPUT_PROJECT_FILE	CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only
PERFORM_RARE_ALLELE_DETECTION	false
QUALITY_THRESHOLD	20
REMOVE_DUPLICATE_READS	false
SINGLE_PROJECT_FILE	CultivarE-WT - BWA-Aln (Aligned 2019-12-19 20:28 UTC)
UMI_PROCESSING_ENABLED	false
USE_QUALITY_DATA_IN_HETEROZYGOUS_CALLS	true

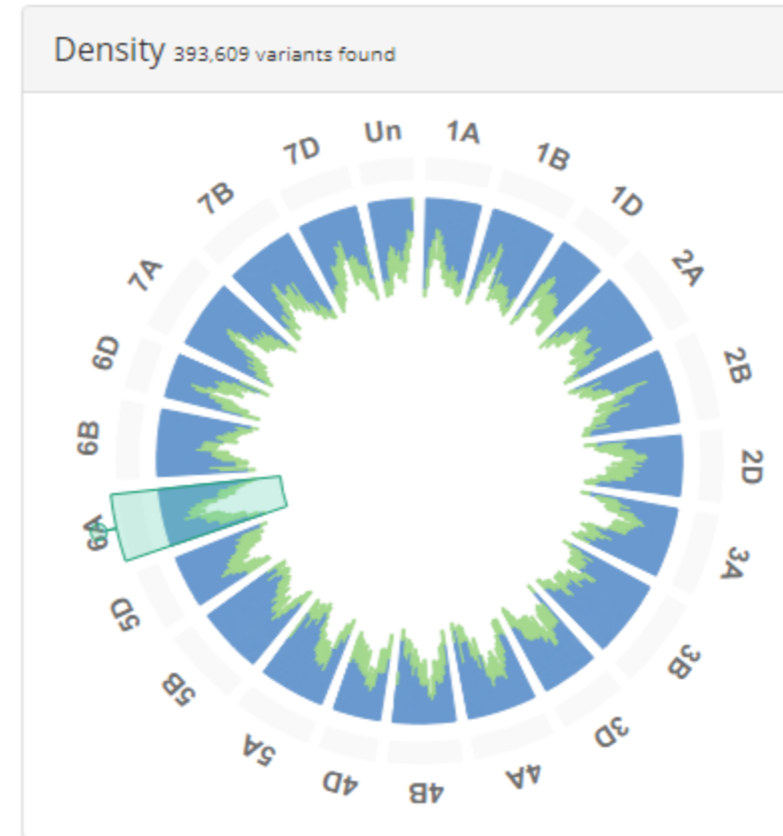
Below the table, the "Analysis Version Info" is displayed: "Curio Variant Detection Analysis, Version: 1.4.1, Updated: October 14,"

The background interface shows a "Project Details" page with a sidebar menu (Summary, File Details, Sharing, Job Details, Processing, Charts) and a list of files under "All Files" and "Raw Sequences".

- Browse and visualize variants across the entire CSW genome

Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

More Options ▾



Display/Filter Options

Compare

Notes and Marks

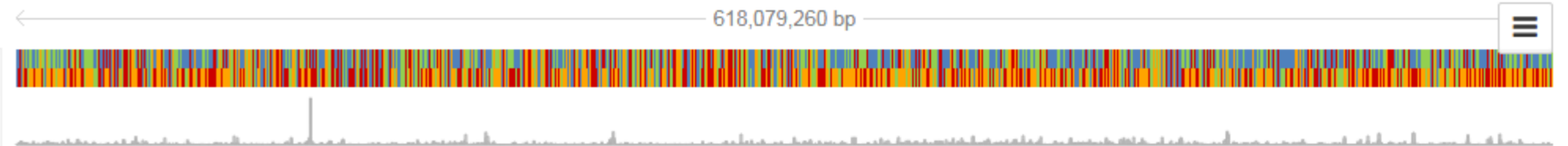
Sharing

- Use gene transcript and exon annotations from [IWGSC \(HC\)](#)
- Include only variants within known [coding regions](#) ⓘ
- Include only variants with predicted consequence of [missense](#)
- Exclude variants whose alt coverage is not at least [5 reads](#)
- Include SNVs for [homozygous or heterozygous positions](#)
- Include positions with rare alleles
- Color positions based on the variant's [alternate allele](#) ⓘ
- Show at each detected position the [variant's coverage](#)
- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least [5 reads](#)
- Match homozygous SNVs to related heterozygous SNVs in control ⓘ
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position

Variants (Found 18,248 matching variants in range chr6A:1-618,079,261)



chr6A:1-618,079,261

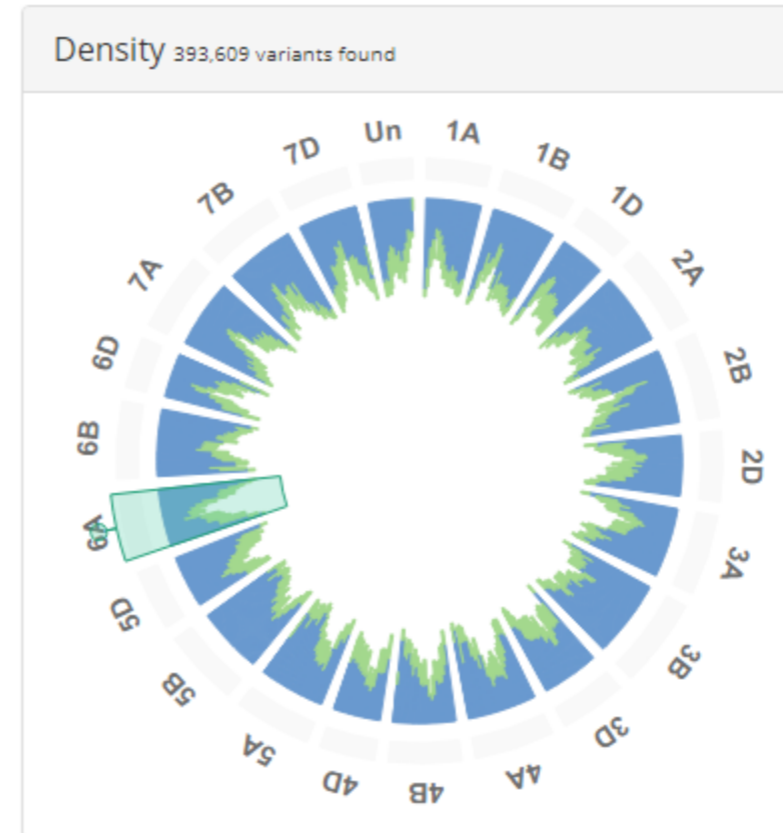


Primary Analysis File
Variant Count: 18,248

- Browse and visualize variants across the entire CSW genome
- Simultaneously compare and visualize multiple samples

Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

More Options ▾



Display/Filter Options **Compare** Notes and Marks Sharing

Type here

<input type="checkbox"/>	Type	File Name	Records	Last Updated
<input checked="" type="checkbox"/>	Sample ▾	CultivarE-Mut5 - BWA-Aln (Analyzed 2019-12-20 15:03 UTC)	376,711	2019 December 20, Friday 10:56:25 UTC-5
<input checked="" type="checkbox"/>	Sample ▾	CultivarE-Mut7 - BWA-Aln (Analyzed 2019-12-20 15:03 UTC)	306,472	2019 December 20, Friday 10:55:12 UTC-5
<input type="checkbox"/>	Sample ▾	CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 16:23 UTC) - Medium Sensitivity	5,859,463	2019 December 20, Friday 11:41:27 UTC-5

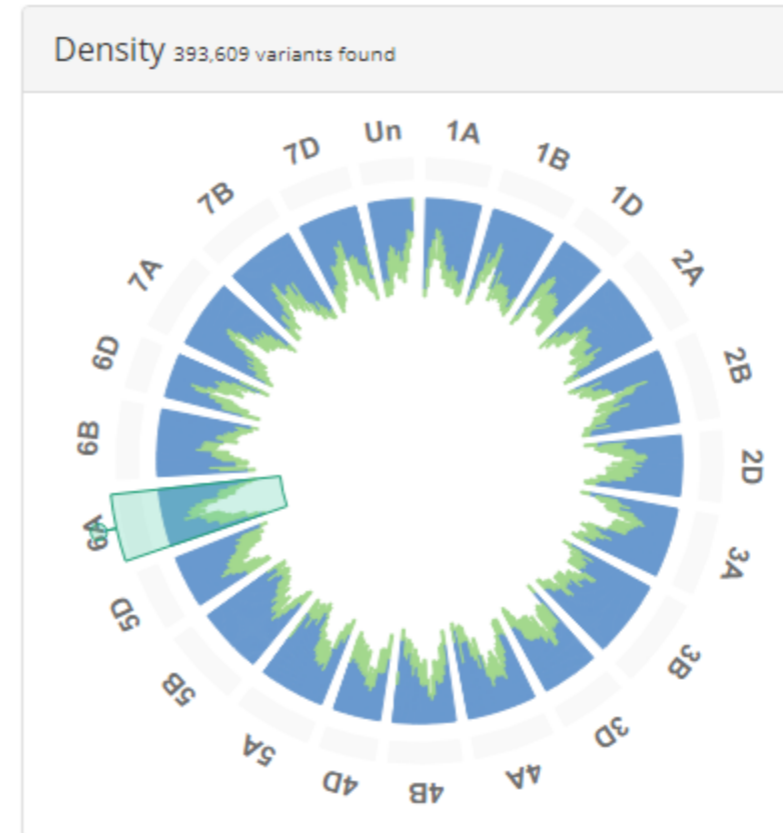
Viewing 1-8 of 8



- Browse and visualize variants across the entire CSW genome
- Simultaneously compare and visualize multiple samples
- Use one or more samples as a control to exclude variants present in the wild type

Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

More Options ▾



Display/Filter Options

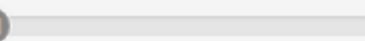
Compare

Notes and Marks

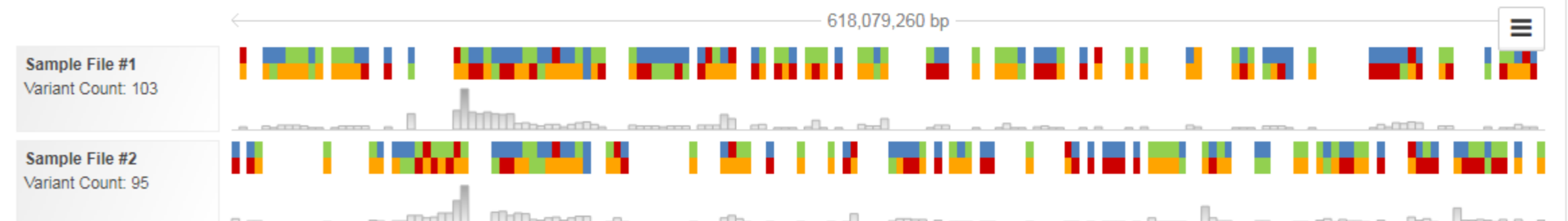
Sharing

- Use gene transcript and exon annotations from IWGSC (HC)
- Include only variants within known coding regions ?
- Include only variants with predicted consequence of missense
- Exclude variants whose alt coverage is not at least 5 reads
- Include SNVs for homozygous or heterozygous positions
- Include positions with rare alleles
- Color positions based on the variant's alternate allele ?
- Show at each detected position the variant's coverage
- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least 5 reads
- Match homozygous SNVs to related heterozygous SNVs in control ?
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position

Variants (Found 198 matching variants in range chr6A:1-618,079,261)



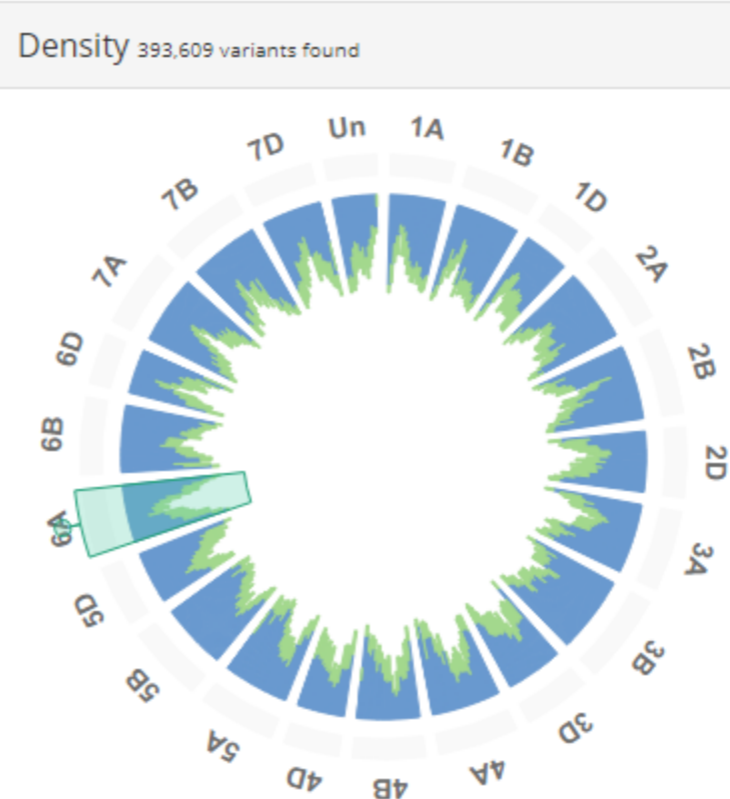
chr6A:1-618,079,261



- Browse and visualize variants across the entire CSW genome
- Simultaneously compare and visualize multiple samples
- Use one or more samples as a control to exclude variants present in the wild type
- Focus on genomic regions of interest based on IWGSC gene transcript annotations

Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only More Options ▾

Density 393,609 variants found



Display/Filter Options Compare Notes and Marks Sharing

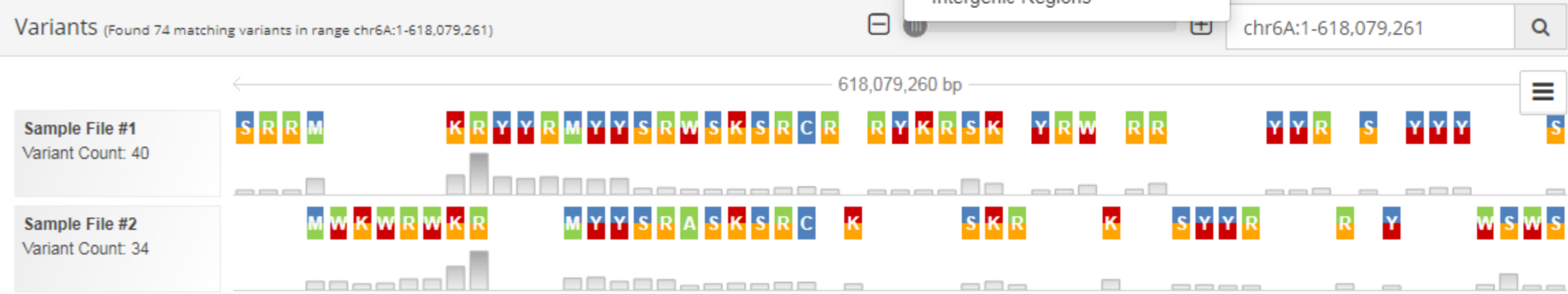
- Use gene transcript and exon annotations from [IWGSC \(HC\)](#) Show only variants not present in the control (i.e. somatic)
- Include only variants within known [coding regions](#) Show only variants with control coverage of at least
- Include only variants with predicted [missense](#)
- Exclude variants whose alt coverage is [5 reads](#)
- Include SNVs for [homozygous or heterozygous](#) positions
- Include positions with rare alleles
- Color positions based on the variant [allele](#)
- Show at each detected position the [coverage](#)

- Exonic Regions
- Intronic Regions
- Coding Regions ✓**
- Intragenic Regions
- Untranslated Regions (UTR)
- 5' UTRs
- 3' UTRs
- Splice Regions
- Upstream/Downstream Regions
- Upstream Regions
- Downstream Regions
- Intergenic Regions

Variants (Found 74 matching variants in range chr6A:1-618,079,261) chr6A:1-618,079,261 🔍

Sample File #1
Variant Count: 40

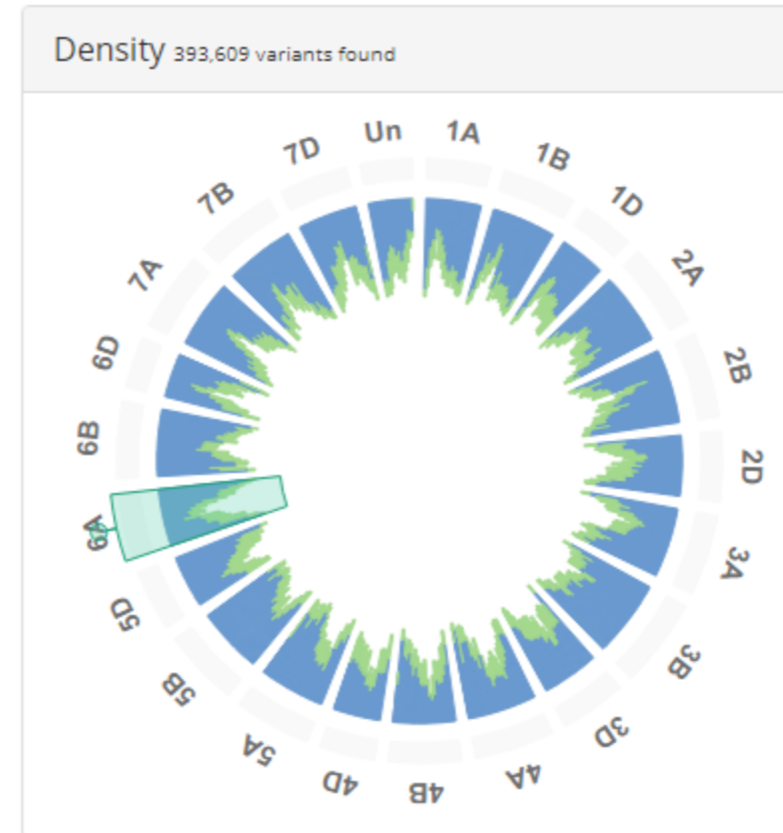
Sample File #2
Variant Count: 34



- Browse and visualize variants across the entire CSW genome
- Simultaneously compare and visualize multiple samples
- Use one or more samples as a control to exclude variants present in the wild type
- Focus on genomic regions of interest based on IWGSC gene transcript annotations
- Access predicted biological consequences based on calculated amino acid shifts

Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

More Options ▾



Display/Filter Options | Compare | Notes and Marks | Sharing

- Use gene transcript and exon annotations from [IWGSC \(HC\)](#)
- Include only variants within known [coding regions](#) ⓘ
- Include only variants with predicted consequence of [not synonymous](#)
 - Missense (Nonsynonymous Substitution) last
 - Stop Gained (Premature Stop Codon)
 - Stop Lost (Terminator Codon Change)
 - Start Lost (Start Codon Change)
 - Not Synonymous (All of the Above) ✓
 - Stop Retained (Synonymous Terminator)
 - Synonymous (Silent Mutation)

- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least [5 reads](#)
- Match homozygous SNVs to related heterozygous SNVs in control ⓘ
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position

Variants (Found 41 matching variants in range chr6A:1-618,079,261) chr6A:1-618,079,261

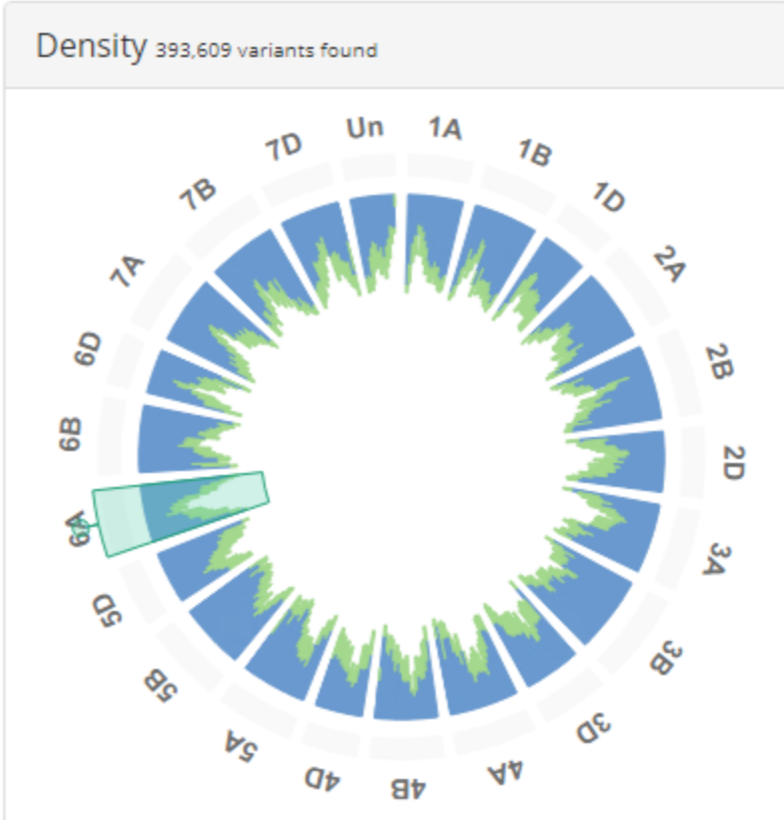
618,079,260 bp

Sample File #1 Variant Count: 19	M	K	R	Y	Y	S	R	W	K	S	C	R	K	S	K	R	W	R	S			
Sample File #2 Variant Count: 22	M	W	W	W	K	Y	Y	S	R	A	K	S	C	K	S	K	K	R	R	Y	W	S

- Browse and visualize variants across the entire CSW genome
- Simultaneously compare and visualize multiple samples
- Use one or more samples as a control to exclude variants present in the wild type
- Focus on genomic regions of interest based on IWGSC gene transcript annotations
- Access predicted biological consequences based on calculated amino acid shifts
- Interactive visualizations provide deeper access to the information behind each call

Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only More Options ▾

Density 393,609 variants found



Display/Filter Options | Compare | Notes and Marks | Sharing

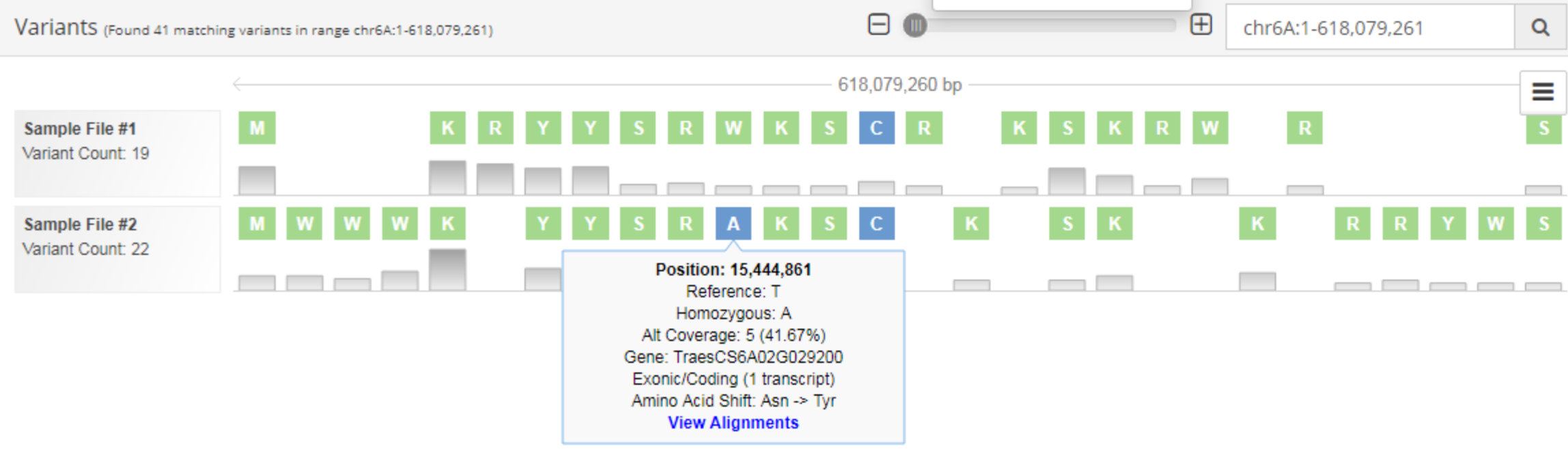
- Use gene transcript and exon annotations from [IWGSC \(HC\)](#)
- Include only variants within known [coding regions](#) ⓘ
- Include only variants with predicted consequence of [not synonymous](#)
- Exclude variants whose alt coverage is not at least 5 reads
- Include SNVs for [homozygous or heterozygous positions](#)
- Include positions with rare alleles
- Color positions based on the variant's [general type](#) ⓘ
- Show at each detected position the variant's [coverage](#)

- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least 5 reads
- Match homozygous SNVs to related heterozygous SNVs in control ⓘ
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position

Variants (Found 41 matching variants in range chr6A:1-618,079,261) chr6A:1-618,079,261 🔍

Sample File #1
Variant Count: 19

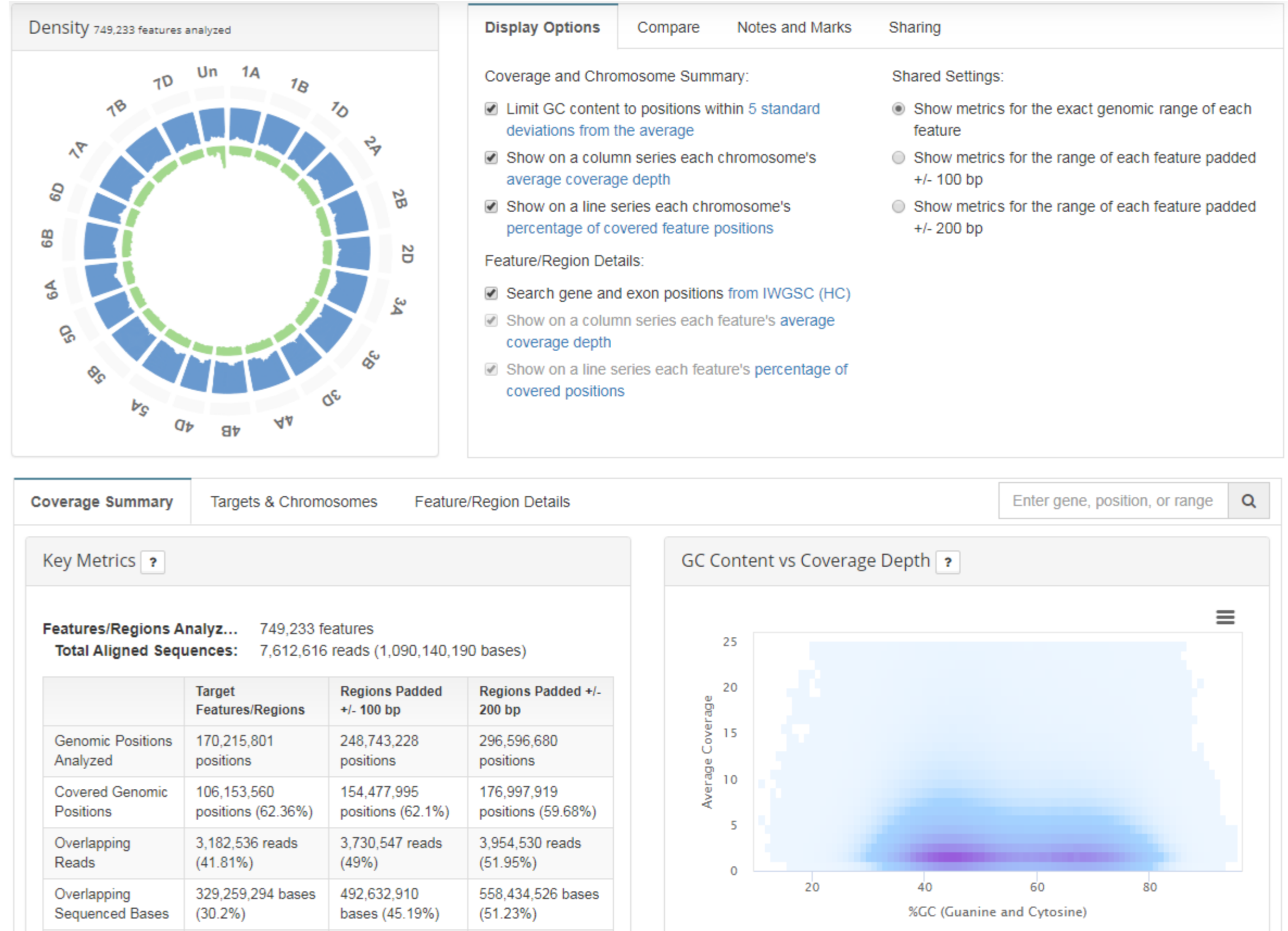
Sample File #2
Variant Count: 22



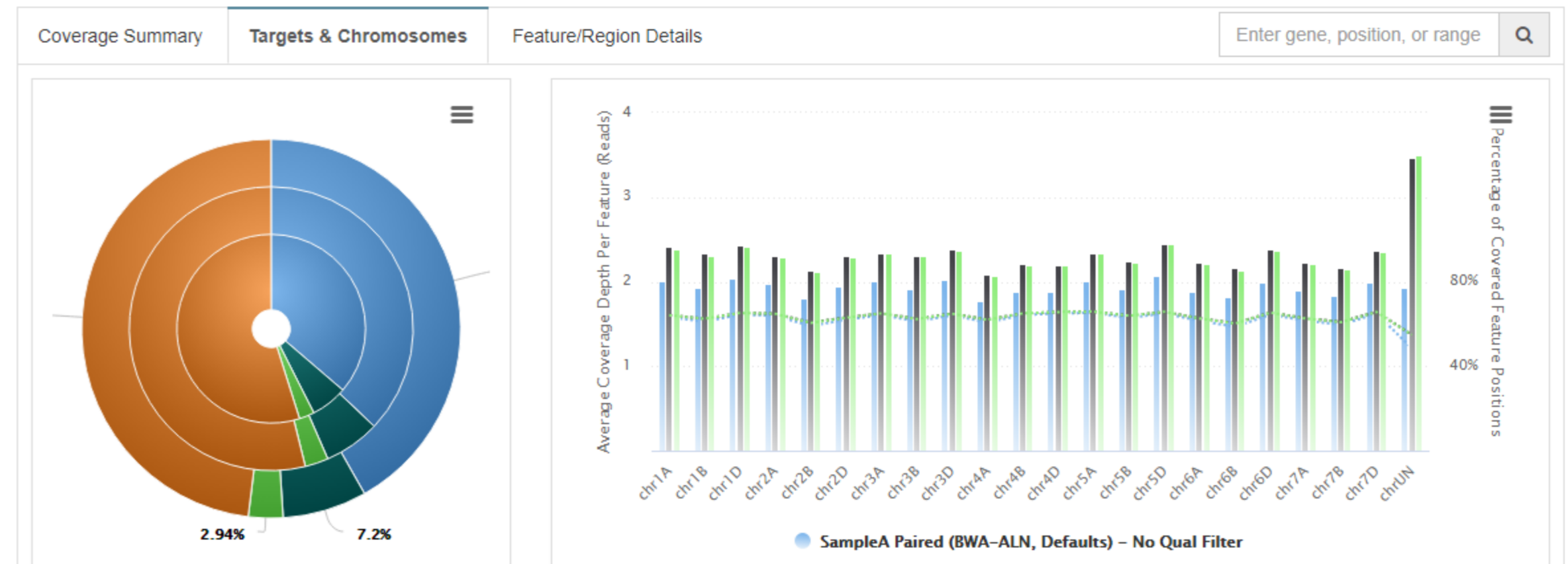
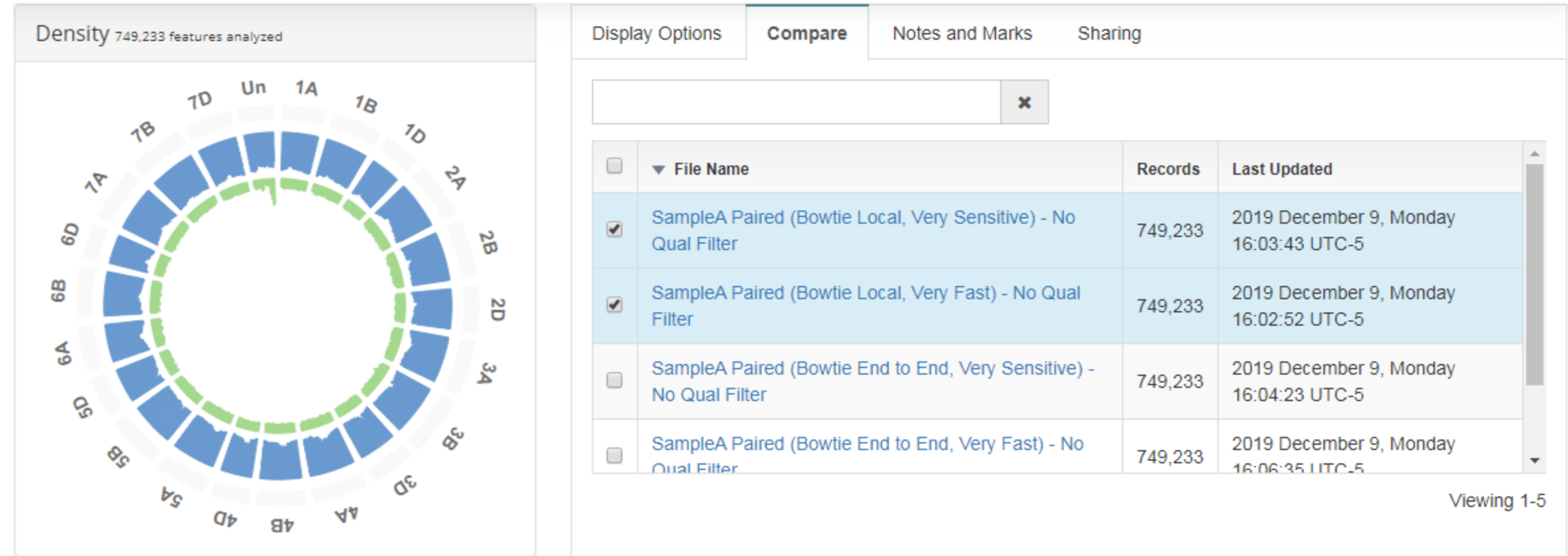
Position: 15,444,861
 Reference: T
 Homozygous: A
 Alt Coverage: 5 (41.67%)
 Gene: TraesCS6A02G029200
 Exonic/Coding (1 transcript)
 Amino Acid Shift: Asn -> Tyr
[View Alignments](#)

DNA-Seq Hexaploid Wheat: Coverage Analysis

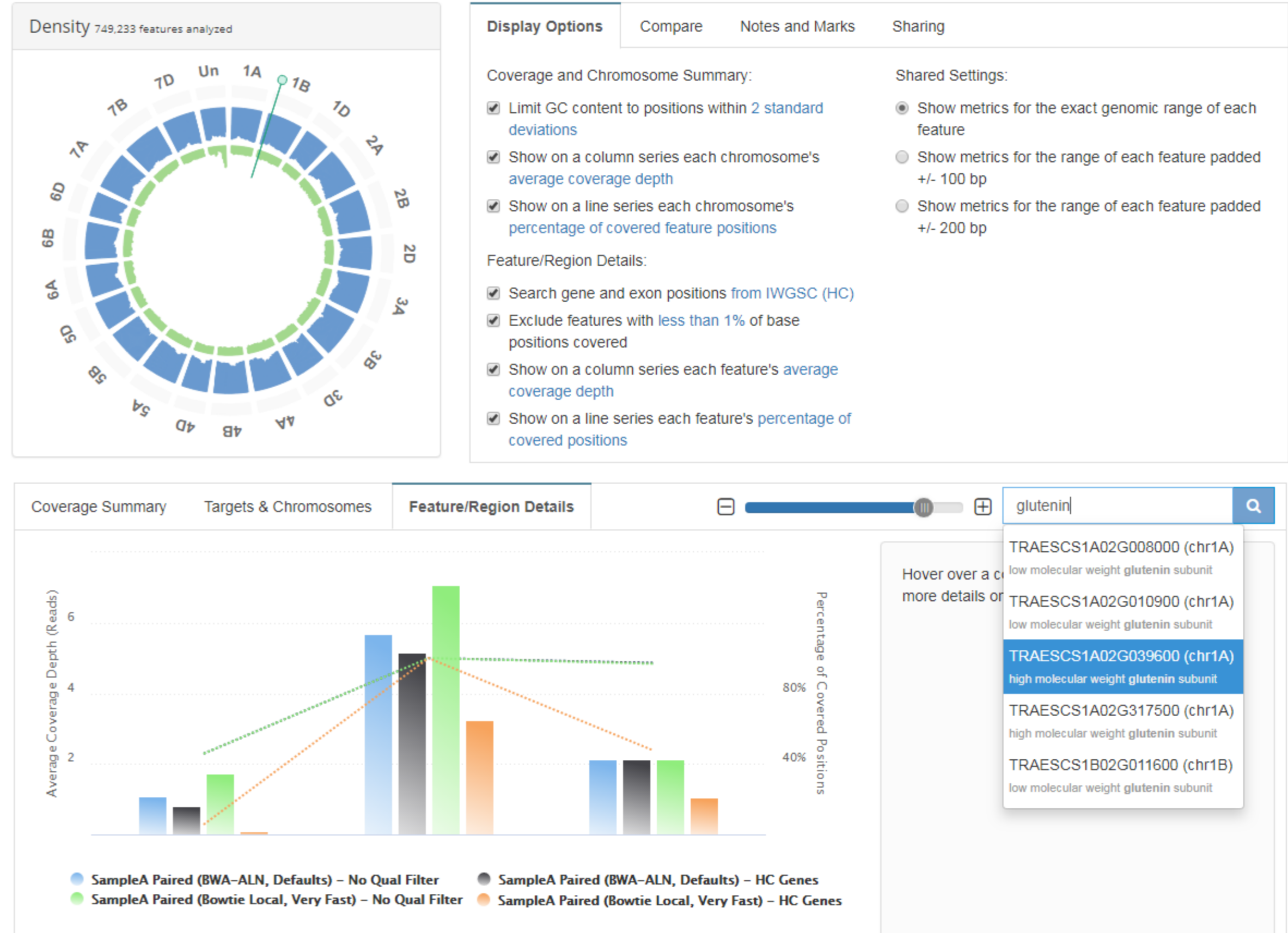
- Measure library / kit exome or custom target area key coverage metrics



- Measure library / kit exome or custom target area key coverage metrics
- Compare and visualize multiple samples simultaneously

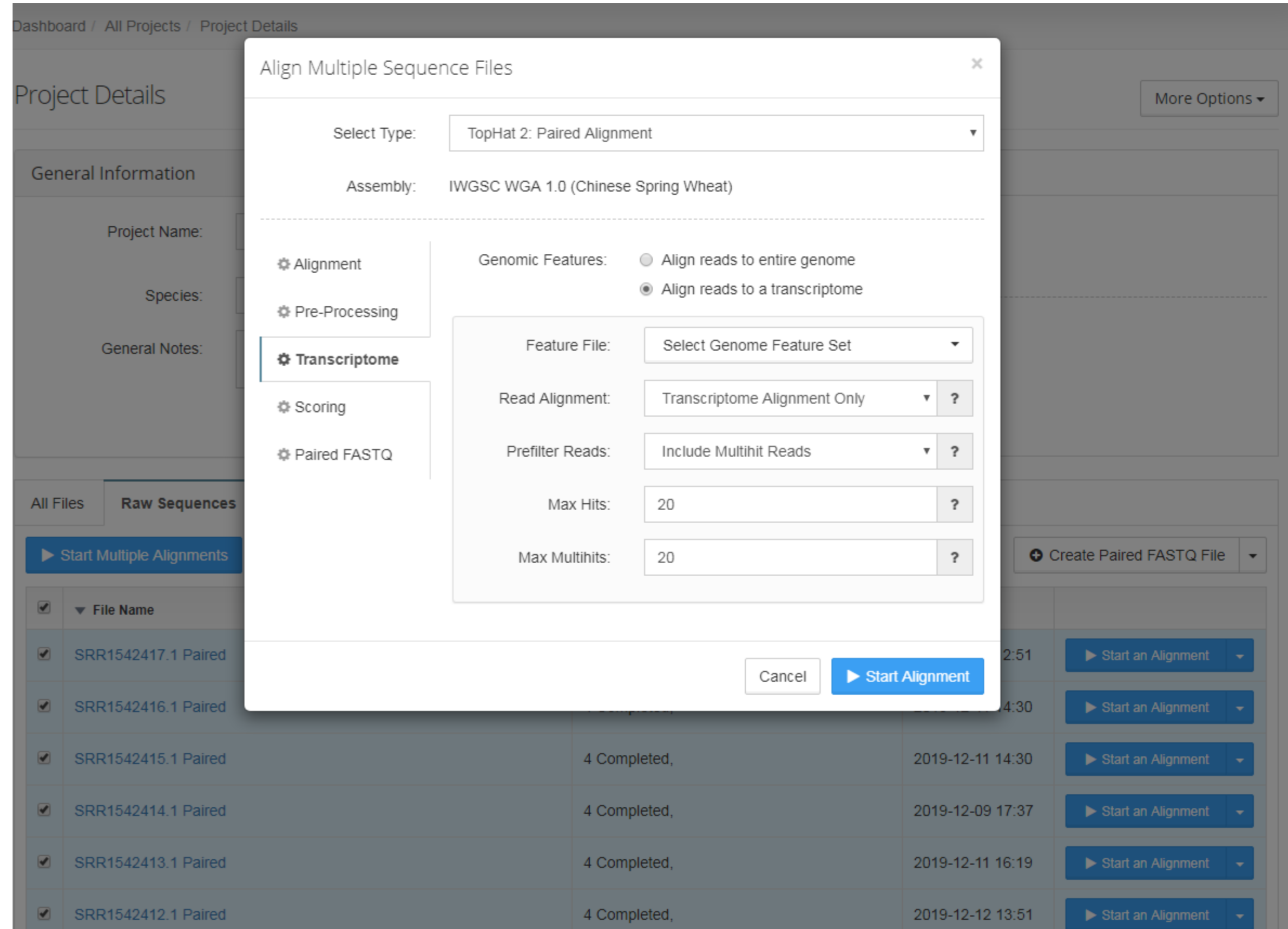


- Measure library / kit exome or custom target area key coverage metrics
- Compare and visualize multiple samples simultaneously
- Access individual gene or transcript coverage metrics leveraging IWGSC annotation sets



RNA-Seq Hexaploid Wheat : Read Mapping and Visualization

- Conveniently select to align to either the full Chinese Spring Wheat genome or a transcriptome



Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Start Multiple Alignments

File Name

<input checked="" type="checkbox"/>	SRR1542417.1 Paired			2:51	Start an Alignment
<input checked="" type="checkbox"/>	SRR1542416.1 Paired			4:30	Start an Alignment
<input checked="" type="checkbox"/>	SRR1542415.1 Paired	4 Completed,		2019-12-11 14:30	Start an Alignment
<input checked="" type="checkbox"/>	SRR1542414.1 Paired	4 Completed,		2019-12-09 17:37	Start an Alignment
<input checked="" type="checkbox"/>	SRR1542413.1 Paired	4 Completed,		2019-12-11 16:19	Start an Alignment
<input checked="" type="checkbox"/>	SRR1542412.1 Paired	4 Completed,		2019-12-12 13:51	Start an Alignment

Align Multiple Sequence Files

Select Type: TopHat 2: Paired Alignment

Assembly: IWGSC WGA 1.0 (Chinese Spring Wheat)

Alignment

Pre-Processing

Transcriptome

Scoring

Paired FASTQ

Genomic Features:

Align reads to entire genome

Align reads to a transcriptome

Feature File: Select Genome Feature Set

Read Alignment: Transcriptome Alignment Only ?

Prefilter Reads: Include Multihit Reads ?

Max Hits: 20 ?

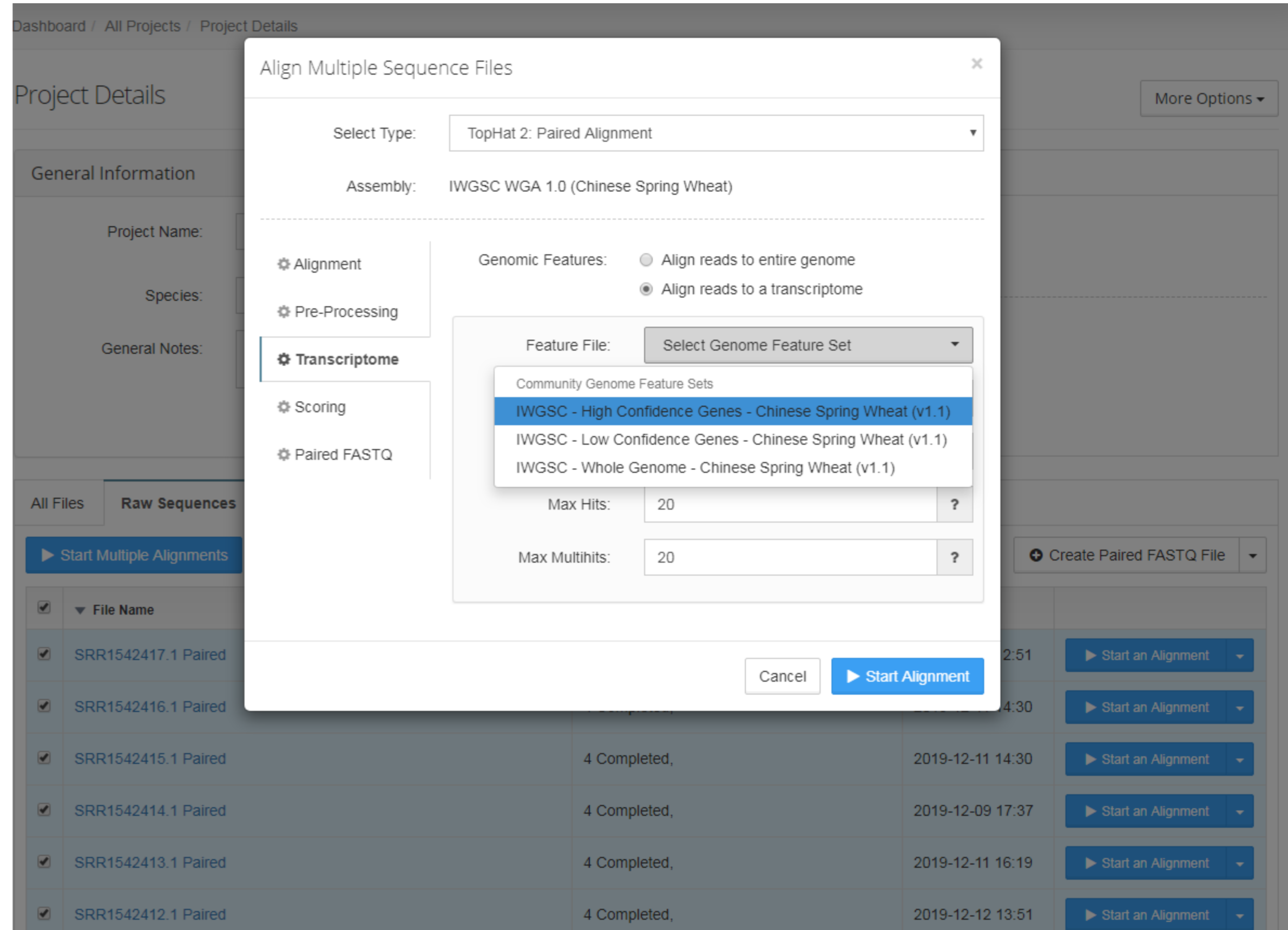
Max Multihits: 20 ?

Cancel Start Alignment

More Options

Create Paired FASTQ File

- Conveniently select to align to either the full Chinese Spring Wheat genome or a transcriptome
- Transcriptome is dynamically assembled on demand using selected annotation set from the IWGSC



Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files **Raw Sequences**

Start Multiple Alignments

File Name	Progress	Time	Action
<input checked="" type="checkbox"/> SRR1542417.1 Paired		2:51	Start an Alignment
<input checked="" type="checkbox"/> SRR1542416.1 Paired		4:30	Start an Alignment
<input checked="" type="checkbox"/> SRR1542415.1 Paired	4 Completed,	2019-12-11 14:30	Start an Alignment
<input checked="" type="checkbox"/> SRR1542414.1 Paired	4 Completed,	2019-12-09 17:37	Start an Alignment
<input checked="" type="checkbox"/> SRR1542413.1 Paired	4 Completed,	2019-12-11 16:19	Start an Alignment
<input checked="" type="checkbox"/> SRR1542412.1 Paired	4 Completed,	2019-12-12 13:51	Start an Alignment

Align Multiple Sequence Files

Select Type: TopHat 2: Paired Alignment

Assembly: IWGSC WGA 1.0 (Chinese Spring Wheat)

Alignment

Pre-Processing

Transcriptome

Scoring

Paired FASTQ

Genomic Features: Align reads to entire genome Align reads to a transcriptome

Feature File: Select Genome Feature Set

Community Genome Feature Sets

- IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)**
- IWGSC - Low Confidence Genes - Chinese Spring Wheat (v1.1)
- IWGSC - Whole Genome - Chinese Spring Wheat (v1.1)

Max Hits: 20 ?

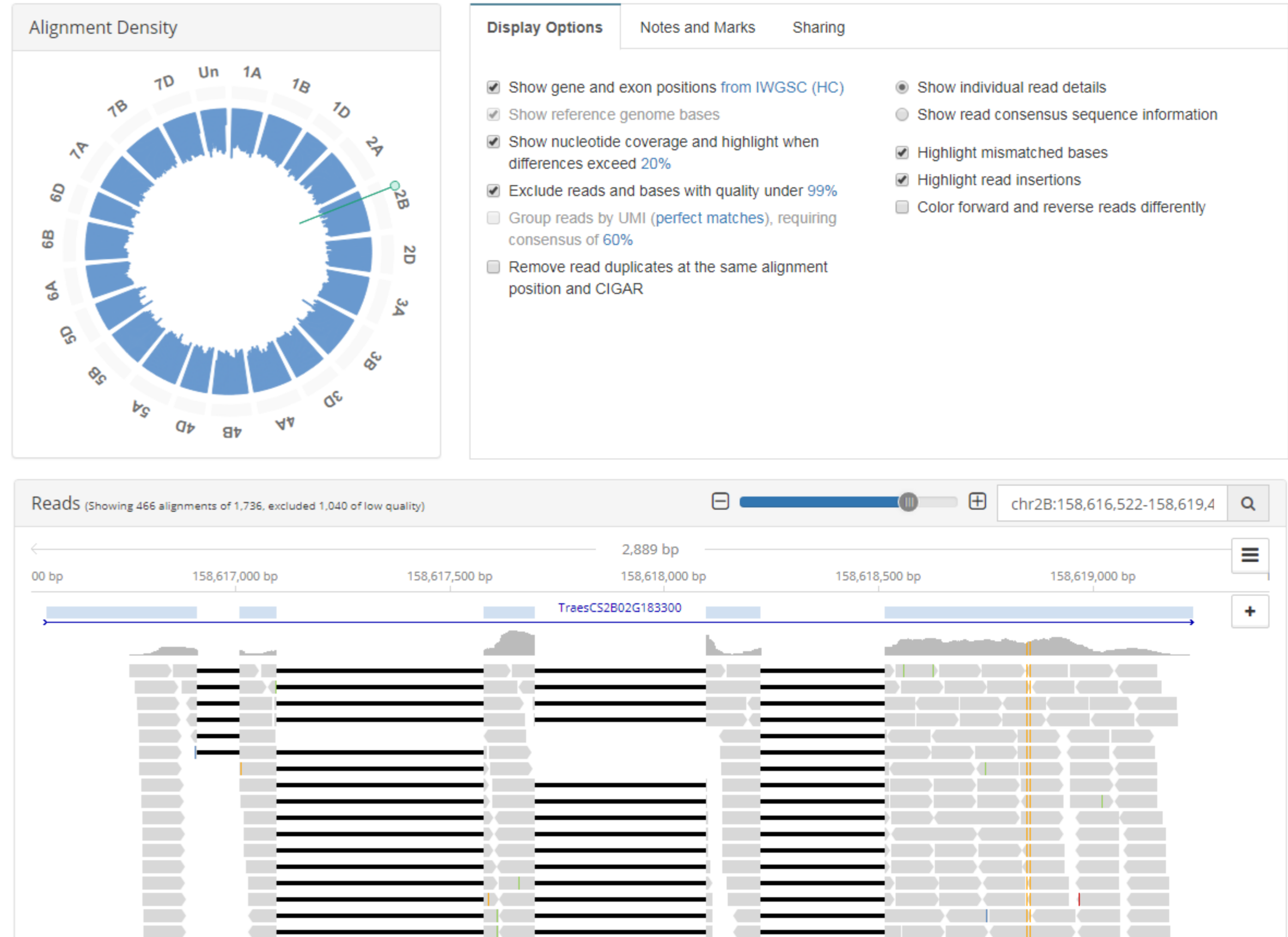
Max Multihits: 20 ?

Cancel Start Alignment

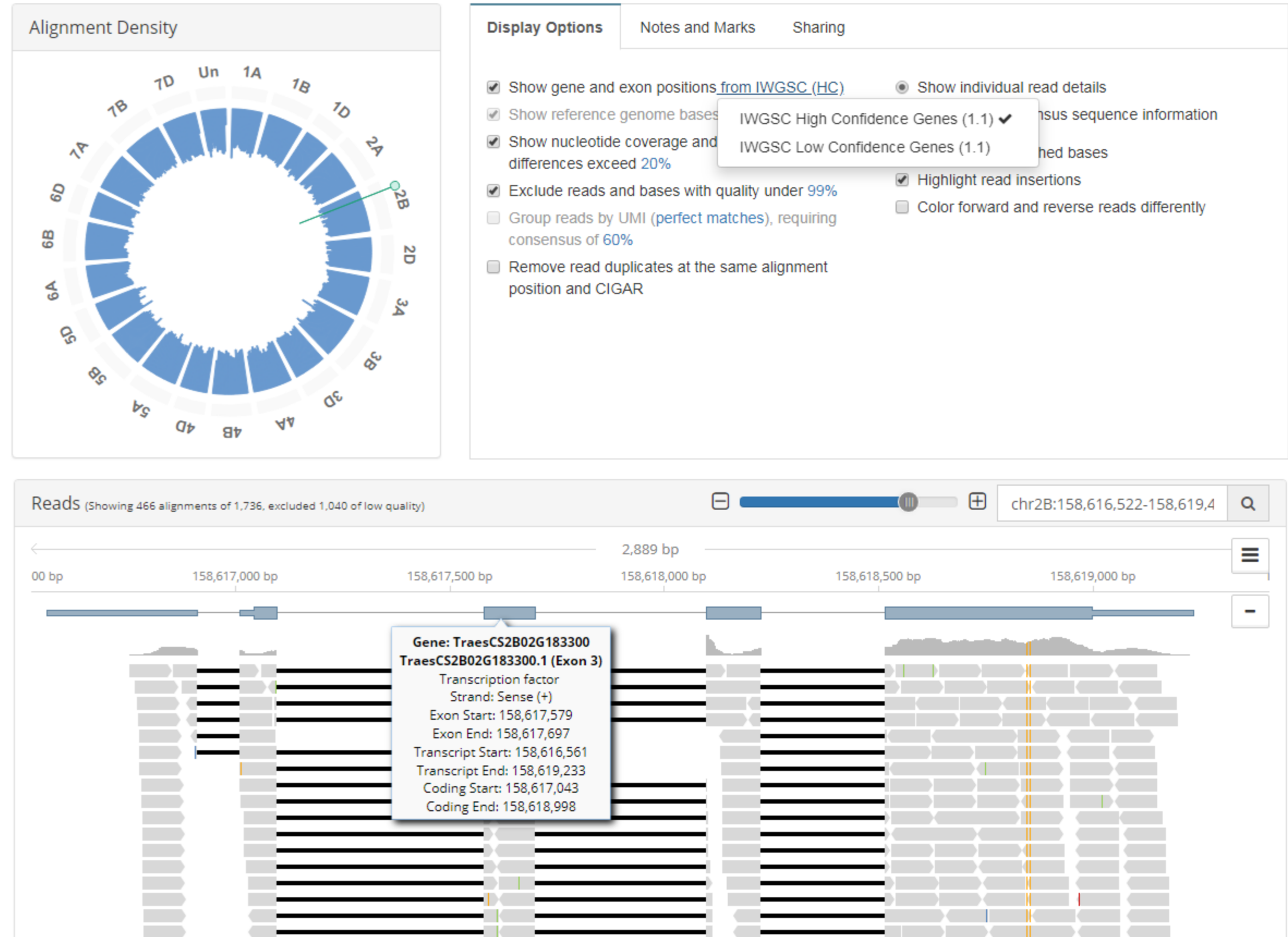
More Options

Create Paired FASTQ File

- Conveniently select to align to either the full Chinese Spring Wheat genome or a transcriptome
- Transcriptome is dynamically assembled on demand using selected annotation set from the IWGSC
- Quickly visualize and navigate RNA-Seq data from samples of any size



- Conveniently select to align to either the full Chinese Spring Wheat genome or a transcriptome
- Transcriptome is dynamically assembled on demand using selected annotation set from the IWGSC
- Quickly visualize and navigate RNA-Seq data from samples of any size
- Reference IWGSC transcript and functional annotations on the fly



RNA-Seq Hexaploid Wheat: Expression Analysis

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly

Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Available Actions Delete Files

Start Multiple Expression Analyses

Genome Feature Set: IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)

Feature to Count: Gene

Exon Matching:

- Count only reads that overlap with exons of the gene
- Count reads that overlap introns or exons of the gene

Multi Feature Overlap: Allow reads to count towards all features they overlap with

De-duplication: Remove potential read duplicates at the same position and CIGAR

Minimum Quality: Phred: 20 (99%)

Cancel Start Expression Analysis

<input checked="" type="checkbox"/>	File Name	Reads Processed	Last Updated	
<input checked="" type="checkbox"/>	SRR1542407.1 Paired (Aligned 2019-12-13 21:40 UTC)	193,694,736	2019-12-16 17:27	View Alignments
<input checked="" type="checkbox"/>	SRR1542404.1 Paired (Aligned 2019-12-13 21:40 UTC)	198,975,170	2019-12-16 13:05	View Alignments
<input checked="" type="checkbox"/>	SRR1542405.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,418,714	2019-12-16 11:02	View Alignments
<input checked="" type="checkbox"/>	SRR1542406.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,129,219	2019-12-16 07:01	View Alignments
<input checked="" type="checkbox"/>	SRR1542408.1 Paired (Aligned 2019-12-13 21:40 UTC)	165,312,410	2019-12-16 02:07	View Alignments
<input checked="" type="checkbox"/>	SRR1542409.1 Paired (Aligned 2019-12-12 19:53 UTC)	203,941,764	2019-12-14 08:25	View Alignments

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC “High Confidence” or “Low Confidence” annotation sets

Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Available Actions Delete Files

Start Multiple Expression Analyses

Genome Feature Set: IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)

Feature to Count: IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)

Exon Matching: Count reads that overlap introns or exons of the gene

Multi Feature Overlap: Allow reads to count towards all features they overlap with

De-duplication: Remove potential read duplicates at the same position and CIGAR

Minimum Quality: Phred: 20 (99%)

Cancel Start Expression Analysis

File Name	Reads Processed	Last Updated	
SRR1542407.1 Paired (Aligned 2019-12-13 21:40 UTC)	193,694,736	2019-12-16 17:27	View Alignments
SRR1542404.1 Paired (Aligned 2019-12-13 21:40 UTC)	198,975,170	2019-12-16 13:05	View Alignments
SRR1542405.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,418,714	2019-12-16 11:02	View Alignments
SRR1542406.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,129,219	2019-12-16 07:01	View Alignments
SRR1542408.1 Paired (Aligned 2019-12-13 21:40 UTC)	165,312,410	2019-12-16 02:07	View Alignments
SRR1542409.1 Paired (Aligned 2019-12-12 19:53 UTC)	203,941,764	2019-12-14 08:25	View Alignments

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC “High Confidence” or “Low Confidence” annotation sets
- Measure expression levels of genes, transcripts, or individual exons

Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Available Actions Delete Files

Start Multiple Expression Analyses

Genome Feature Set: IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)

Feature to Count: Gene

Exon Matching: Count reads that overlap introns or exons of the gene

Multi Feature Overlap: Allow reads to count towards all features they overlap with

De-duplication: Remove potential read duplicates at the same position and CIGAR

Minimum Quality: Phred: 20 (99%)

Cancel Start Expression Analysis

File Name	Reads Processed	Last Updated	
<input checked="" type="checkbox"/> SRR1542407.1 Paired (Aligned 2019-12-13 21:40 UTC)	193,694,736	2019-12-16 17:27	View Alignments
<input checked="" type="checkbox"/> SRR1542404.1 Paired (Aligned 2019-12-13 21:40 UTC)	198,975,170	2019-12-16 13:05	View Alignments
<input checked="" type="checkbox"/> SRR1542405.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,418,714	2019-12-16 11:02	View Alignments
<input checked="" type="checkbox"/> SRR1542406.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,129,219	2019-12-16 07:01	View Alignments
<input checked="" type="checkbox"/> SRR1542408.1 Paired (Aligned 2019-12-13 21:40 UTC)	165,312,410	2019-12-16 02:07	View Alignments
<input checked="" type="checkbox"/> SRR1542409.1 Paired (Aligned 2019-12-12 19:53 UTC)	203,941,764	2019-12-14 08:25	View Alignments

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC “High Confidence” or “Low Confidence” annotation sets
- Measure expression levels of genes, transcripts, or individual exons
- Group samples for a differential expression analysis...

Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Type here

<input type="checkbox"/>	▲ File Name
<input type="checkbox"/>	SRR1542404 - CK
<input type="checkbox"/>	SRR1542405 - CK
<input type="checkbox"/>	SRR1542406 - DS1h
<input type="checkbox"/>	SRR1542407 - DS1h
<input type="checkbox"/>	SRR1542408 - DS6h
<input type="checkbox"/>	SRR1542409 - DS6h

Start a Differential Expression Analysis ✕

Analysis Name: ?
 E.g. 'Expression change in radiation exposed tissue'

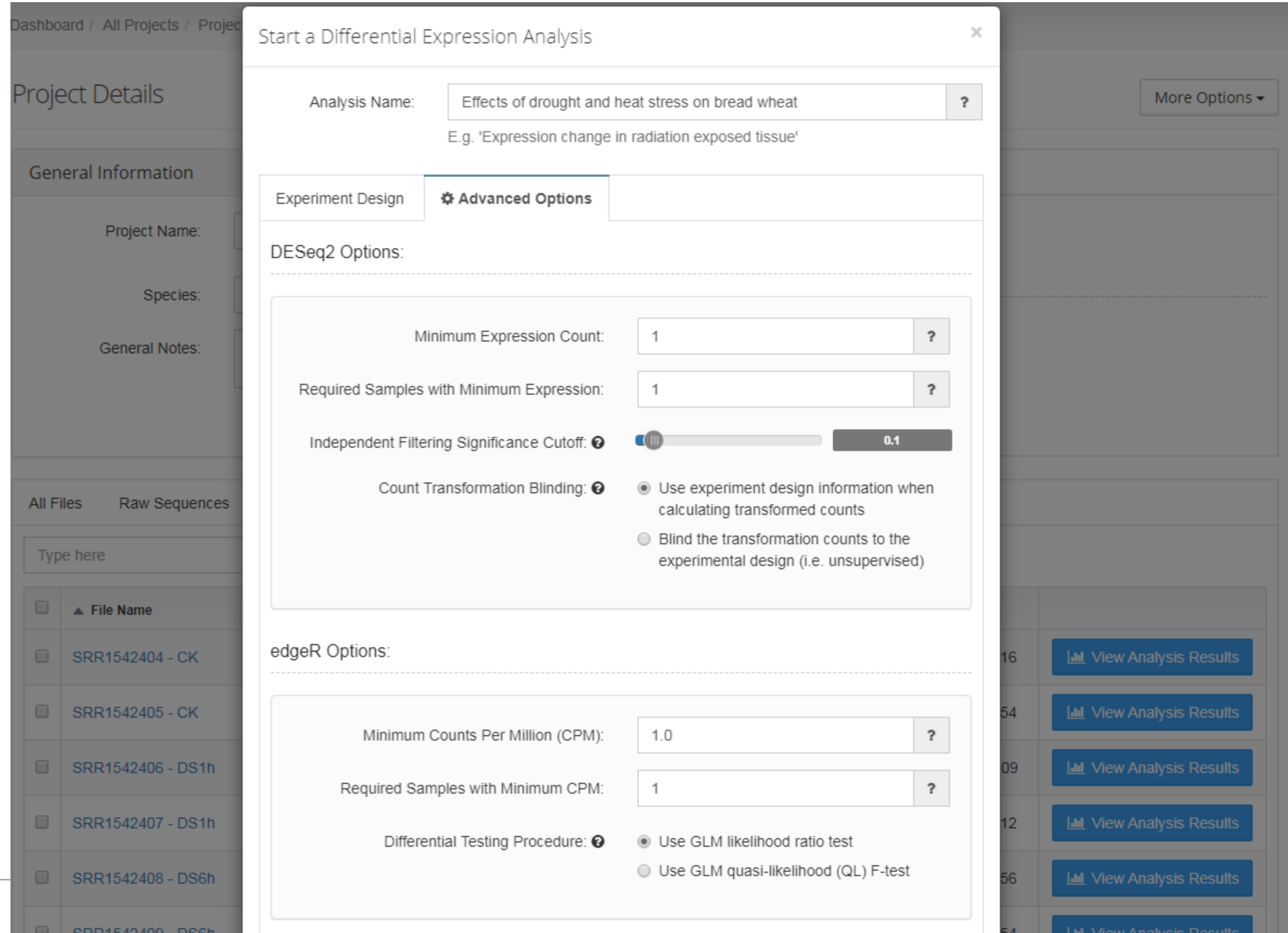
Experiment Design ⚙️ Advanced Options

Define Sample Groups: Move Selected to Group ▾

<input type="checkbox"/>	▲ Set Group	File Name	
<input type="checkbox"/>	Group 1 ▾	SRR1542404 - CK	✕
<input type="checkbox"/>	Group 1 ▾	SRR1542405 - CK	✕
<input type="checkbox"/>	Group 2 ▾	SRR1542408 - DS6h	✕
<input type="checkbox"/>	Group 2 ▾	SRR1542409 - DS6h	✕
<input type="checkbox"/>	Group 3 ▾	SRR1542412 - HS6h	✕
<input type="checkbox"/>	Group 3 ▾	SRR1542413 - HS6h	✕
<input checked="" type="checkbox"/>	Group 4 ▾	SRR1542416 - HD6h	✕
<input checked="" type="checkbox"/>	Group 4 ▾	SRR1542417 - HD6h	✕

Viewing 1-8

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC “High Confidence” or “Low Confidence” annotation sets
- Measure expression levels of genes, transcripts, or individual exons
- Group samples for a differential expression analysis...
- ...and conveniently tune industry standard ‘R’ based algorithms



Dashboard / All Projects / Project

Project Details

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Type here

<input type="checkbox"/>	▲ File Name
<input type="checkbox"/>	SRR1542404 - CK
<input type="checkbox"/>	SRR1542405 - CK
<input type="checkbox"/>	SRR1542406 - DS1h
<input type="checkbox"/>	SRR1542407 - DS1h
<input type="checkbox"/>	SRR1542408 - DS6h
<input type="checkbox"/>	SRR1542409 - DS6h

Start a Differential Expression Analysis

Analysis Name: Effects of drought and heat stress on bread wheat ?
E.g. 'Expression change in radiation exposed tissue'

Experiment Design **Advanced Options**

DESeq2 Options:

Minimum Expression Count: 1 ?

Required Samples with Minimum Expression: 1 ?

Independent Filtering Significance Cutoff: 0.1

Count Transformation Blinding: Use experiment design information when calculating transformed counts
 Blind the transformation counts to the experimental design (i.e. unsupervised)

edgeR Options:

Minimum Counts Per Million (CPM): 1.0 ?

Required Samples with Minimum CPM: 1 ?

Differential Testing Procedure: Use GLM likelihood ratio test
 Use GLM quasi-likelihood (QL) F-test

More Options ▾

16 [View Analysis Results](#)

54 [View Analysis Results](#)

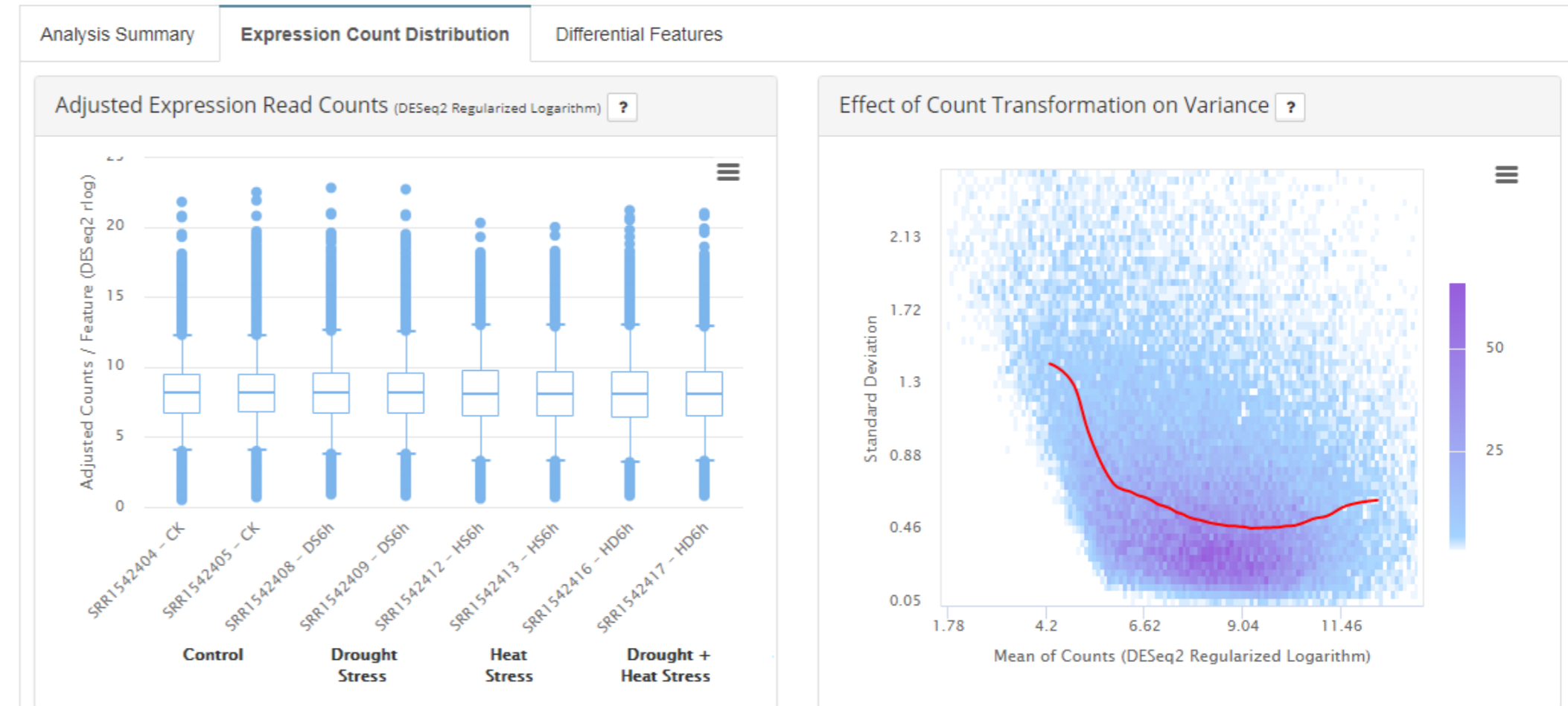
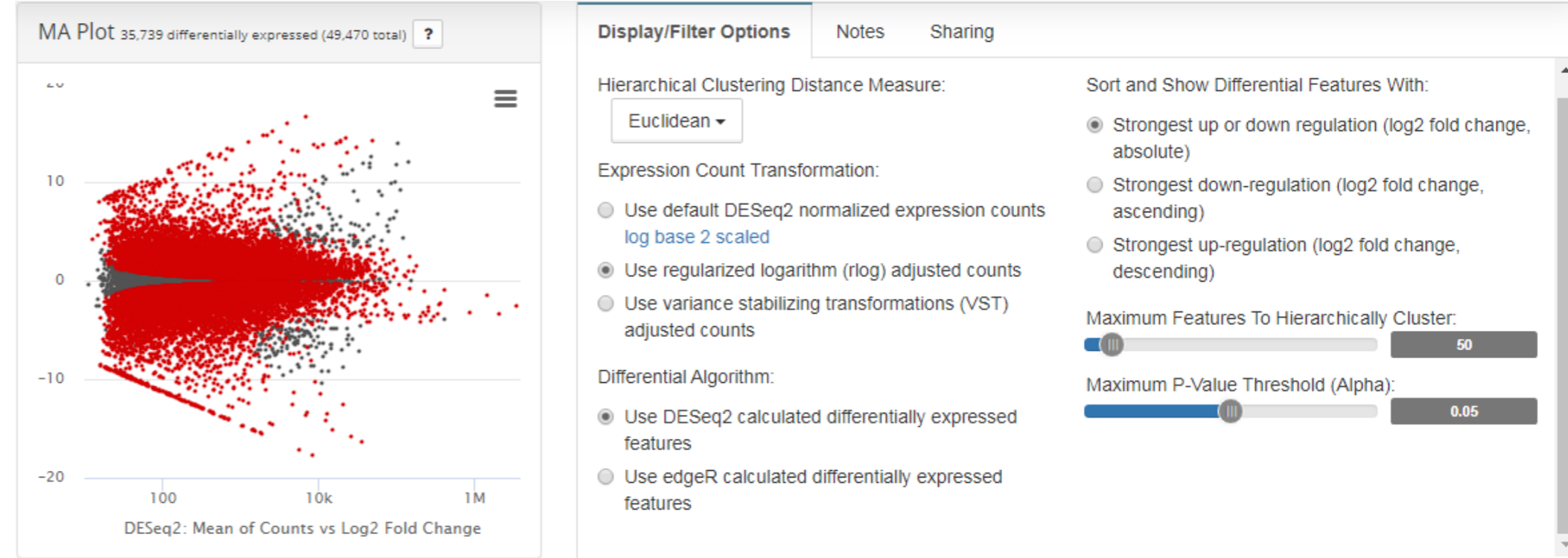
09 [View Analysis Results](#)

12 [View Analysis Results](#)

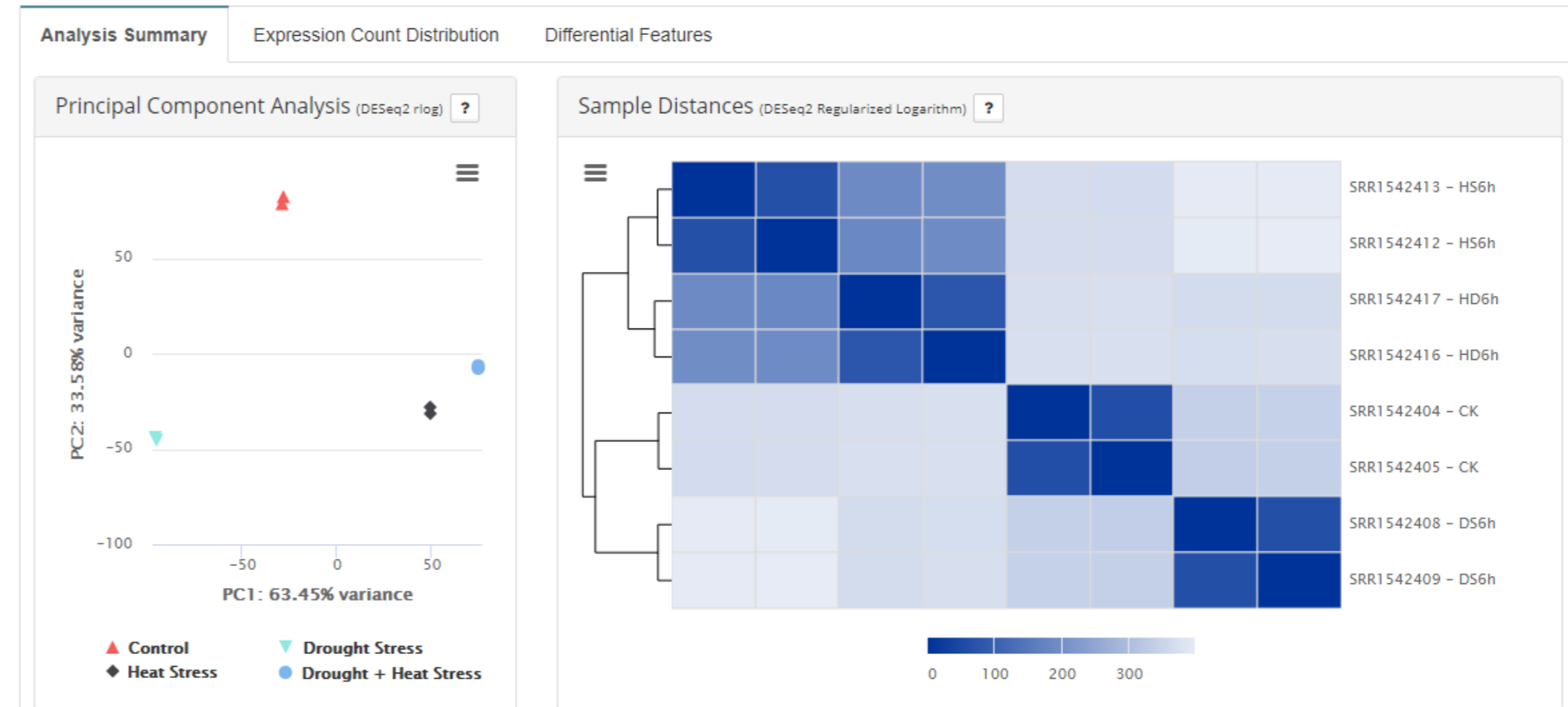
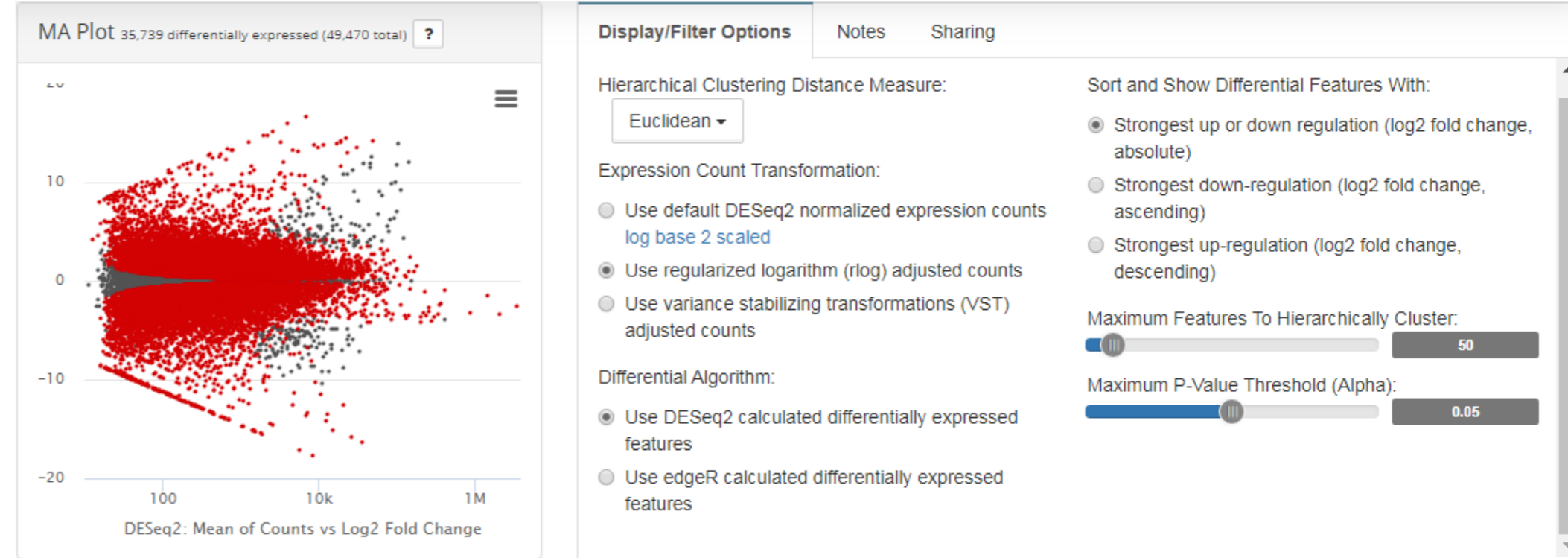
56 [View Analysis Results](#)

54 [View Analysis Results](#)

- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets



- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets
- Validate sample distances through principal component analysis



- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets
- Validate sample distances through principal component analysis
- Highlight up or down regulated genes through interactive heat maps

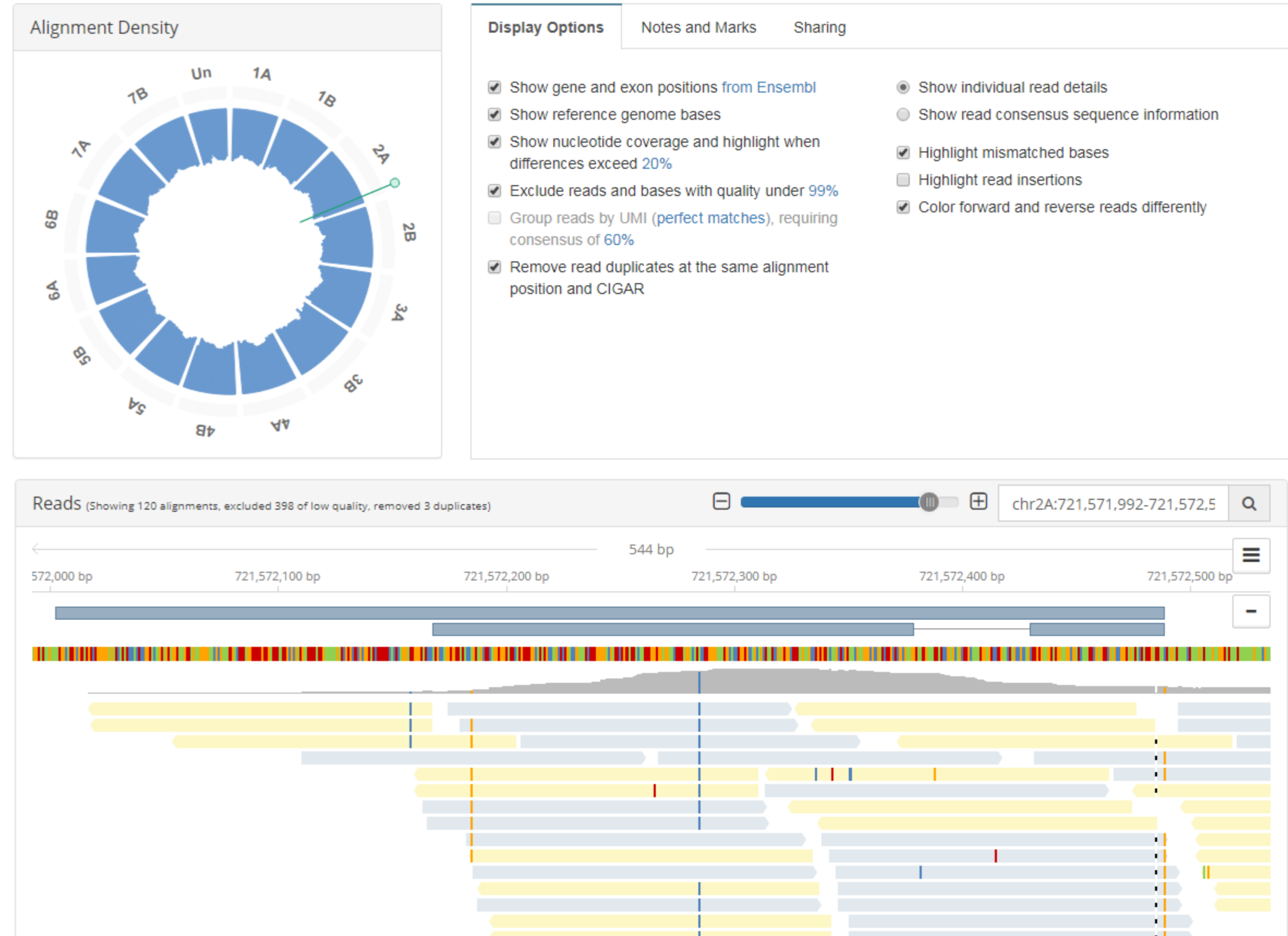


- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets
- Validate sample distances through principal component analysis
- Highlight up or down regulated genes through interactive heat maps
- Compare and contrast different industry standard algorithms in real-time



DNA-Seq: Tetraploid Wheat

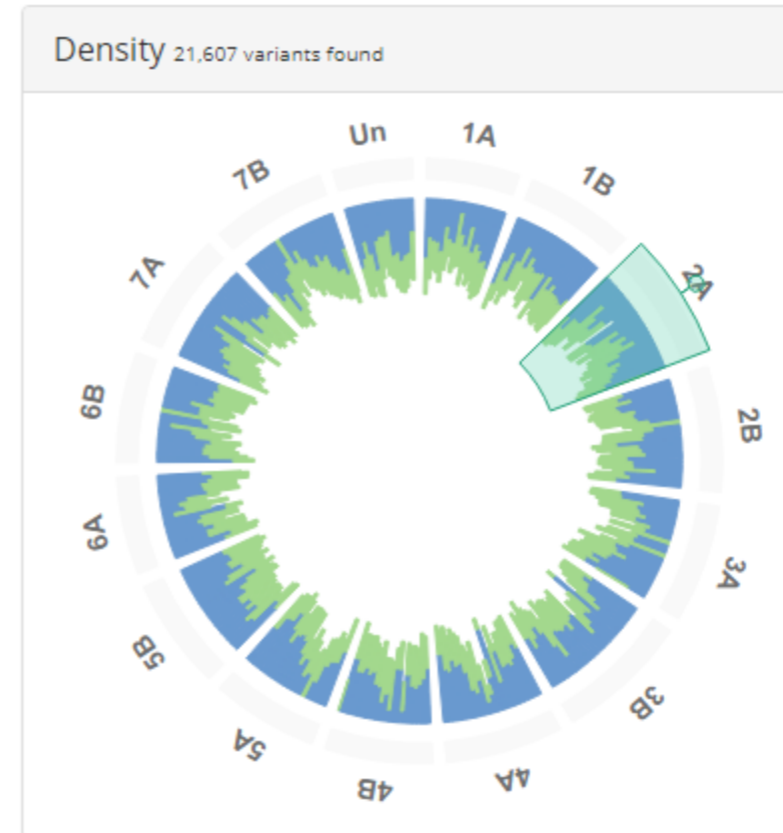
- Also incorporates the Svevo (*Triticum turgidum*) reference assembly from the IDWGSC (i.e. the International Durum Wheat Genome Sequencing Consortium)
- Instantly browse any portion of the IDWGSC reference sequence and visualize read alignments of samples of any size
- Dynamically leverage gene/transcript/exon annotations for Svevo from Ensembl



- Call and analyze variants leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and visualize variant density across multiple samples
- Dynamically filter to different genomic region types based on annotations published by the IDWGSC via Ensembl

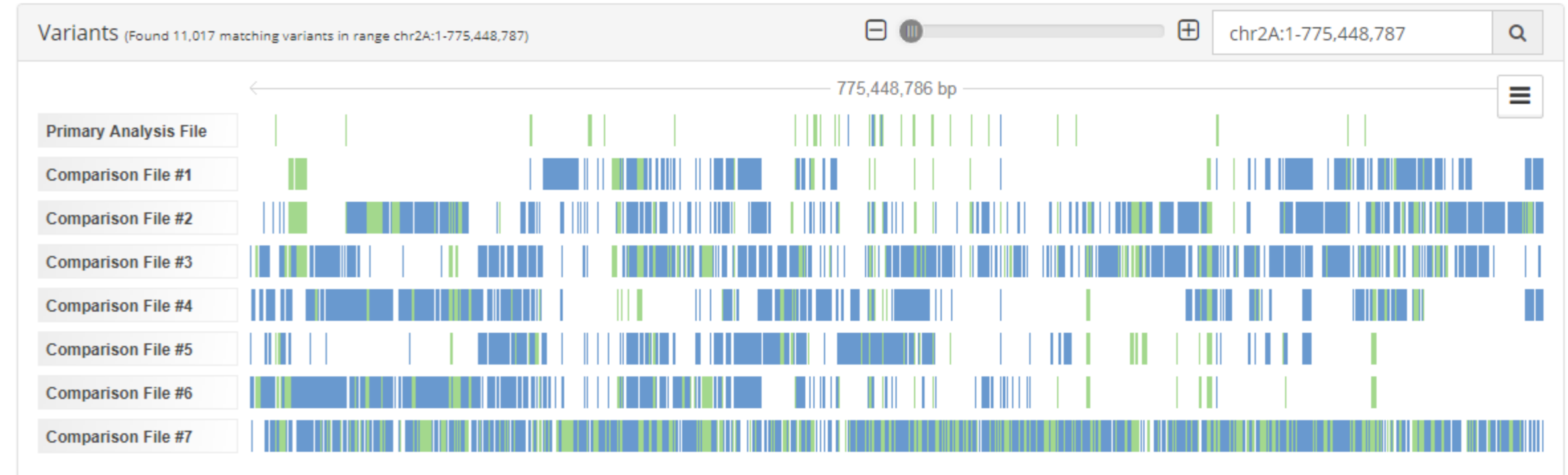
Variant Analysis TetS Paired - vs Svevo - varcalls phred 20 (Analyzed 2020-01-01 21:33 UTC)

More Options ▾



Display/Filter Options Compare Notes and Marks Sharing

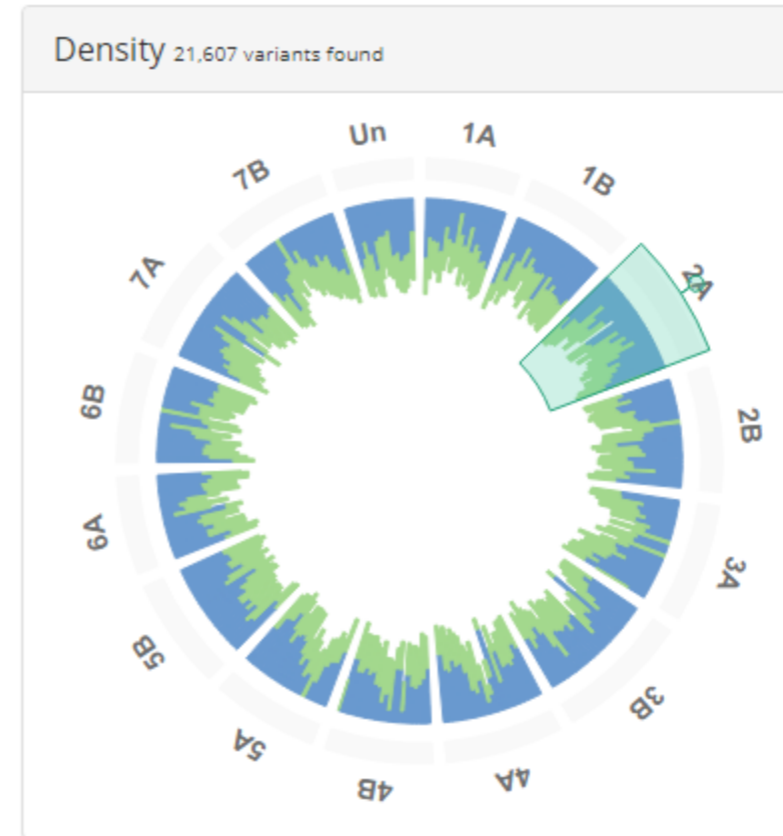
- Use gene transcript and exon annotations from [Ensembl](#)
- Include only variants within known [coding regions](#) ⓘ
- Include only variants with predicted consequence of [stop gained](#)
- Exclude variants whose alt coverage is not at least **10 reads**
- Include SNVs for [homozygous or heterozygous positions](#)
- Include positions with rare alleles
- Color positions based on the variant's [general type](#) ⓘ
- Show at each detected position the [variant's frequency](#)
- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least **1 reads**
- Match homozygous SNVs to related heterozygous SNVs in control ⓘ
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position



- Call and analyze variants leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and visualize variant density across multiple samples
- Dynamically filter to different genomic region types based on annotations published by the IDWGSC via Ensembl
- Access predicted biological consequences based on calculated amino acid shifts

Variant Analysis TetS Paired - vs Svevo - varcalls phred 20 (Analyzed 2020-01-01 21:33 UTC)

More Options ▾



Display/Filter Options

Compare

Notes and Marks

Sharing

- Use gene transcript and exon annotations from [Ensembl](#)
- Include only variants within known [coding regions](#)
- Include only variants with predicted consequence of [stop gained](#)
- Exclude variants whose alt coverage is not at least **10 reads**
- Include SNVs for [homozygous or heterozygous positions](#)
- Include positions with rare alleles
- Color positions based on the variant's [predicted consequence](#)
- Show at each detected position the variant's [frequency](#)
- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least **1 reads**
- Match homozygous SNVs to related heterozygous SNVs in control
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position

- General Type
- Alternate Allele
- Genomic Location
- Predicted Consequence**

Variants (Found 11,017 matching variants in range chr2A:1-775,448,787)

775,448,786 bp

chr2A:1-775,448,787

Q

Enable or disable the color highlighting of the variants based on

Primary Analysis File

Comparison File #1

Comparison File #2

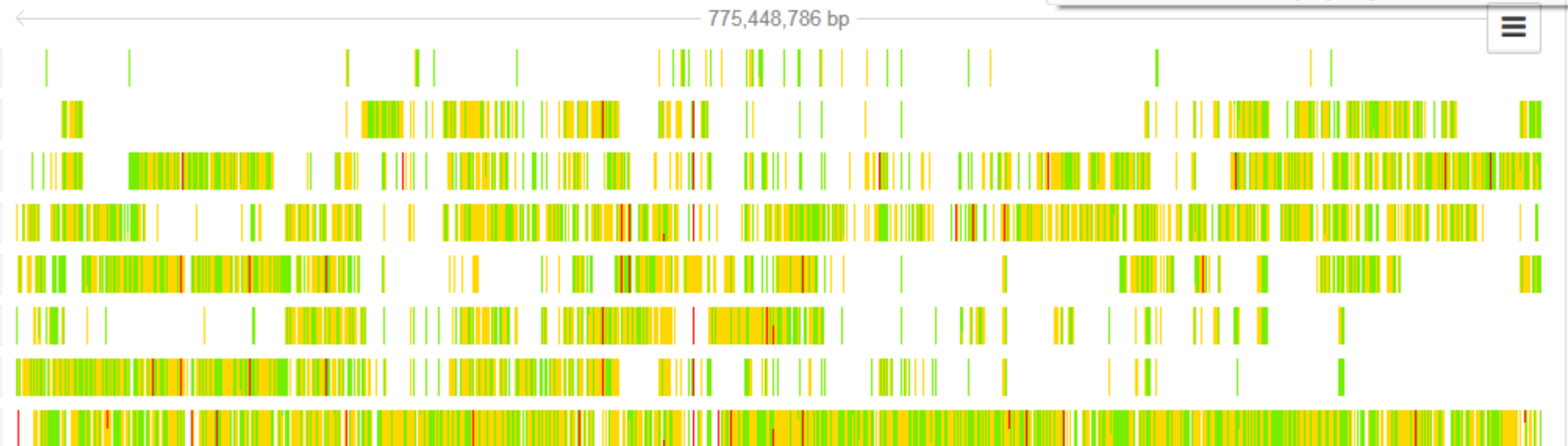
Comparison File #3

Comparison File #4

Comparison File #5

Comparison File #6

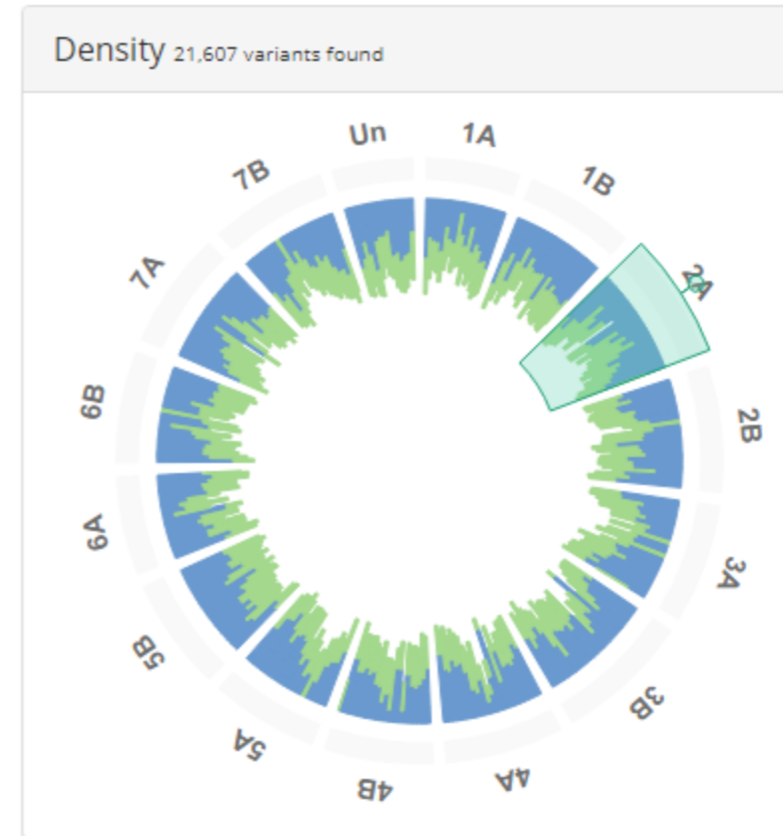
Comparison File #7



- Call and analyze variants leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and visualize variant density across multiple samples
- Dynamically filter to different genomic region types based on annotations published by the IDWGSC via Ensembl
- Access predicted biological consequences based on calculated amino acid shifts
- Filter down to consequences of interest

Variant Analysis TetS Paired - vs Svevo - varcalls phred 20 (Analyzed 2020-01-01 21:33 UTC)

More Options ▾



Display/Filter Options | Compare | Notes and Marks | Sharing

- Use gene transcript and exon annotations from [Ensembl](#)
- Include only variants within known [coding regions](#) ⓘ
- Include only variants with predicted consequence of [stop_gained](#)
- Missense (Nonsynonymous Substitution) ast
- Stop Gained (Premature Stop Codon) ✓**
- Stop Lost (Terminator Codon Change)
- Start Lost (Start Codon Change)
- Not Synonymous (All of the Above)
- Stop Retained (Synonymous Terminator)
- Synonymous (Silent Mutation)

- Show only variants not present in the control (i.e. somatic)
- Show only variants with control coverage of at least **1 reads**
- Match homozygous SNVs to related heterozygous SNVs in control ⓘ
- Show all variants found within selected sample files
- Show only variants common to all selected sample files
- Show file coverage track of variants detected per position



In Conclusion

Looking Ahead

- Incorporation of the recently released IWGSC 2.0 reference assembly
- Addition of the soon-to-be released IWGSC 2.0 transcript and functional annotations
- Consideration of other wheat and related species, along with related annotation sets
- Support for custom reference assemblies
- Additional crop-research-specific analysis types and interpretive visualizations



Acknowledgements

Special thanks to collaborators at:

- Arbor Biosciences
- John Innes Centre (JIC)
- French National Institute for Agricultural Research (INRA)
- University of Adelaide, Plant Genomics Centre

And the collective efforts of the:

- International Wheat Genome Sequencing Consortium (IWGSC)
- International Durum Wheat Genome Sequencing Consortium (IDWGSC)

Differential expression analysis utilized FASTQ files published as part of the following study:

- Liu Z, et al. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.)
BMC Plant Biol. 2015;15:152. doi: 10.1186/s12870-015-0511-8.

Q&A