



---

**Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio**

*Presented at the 2020 Plant and Animal Genome Conference*

*Author: Shawn Quinn (CTO, Curio Genomics)*

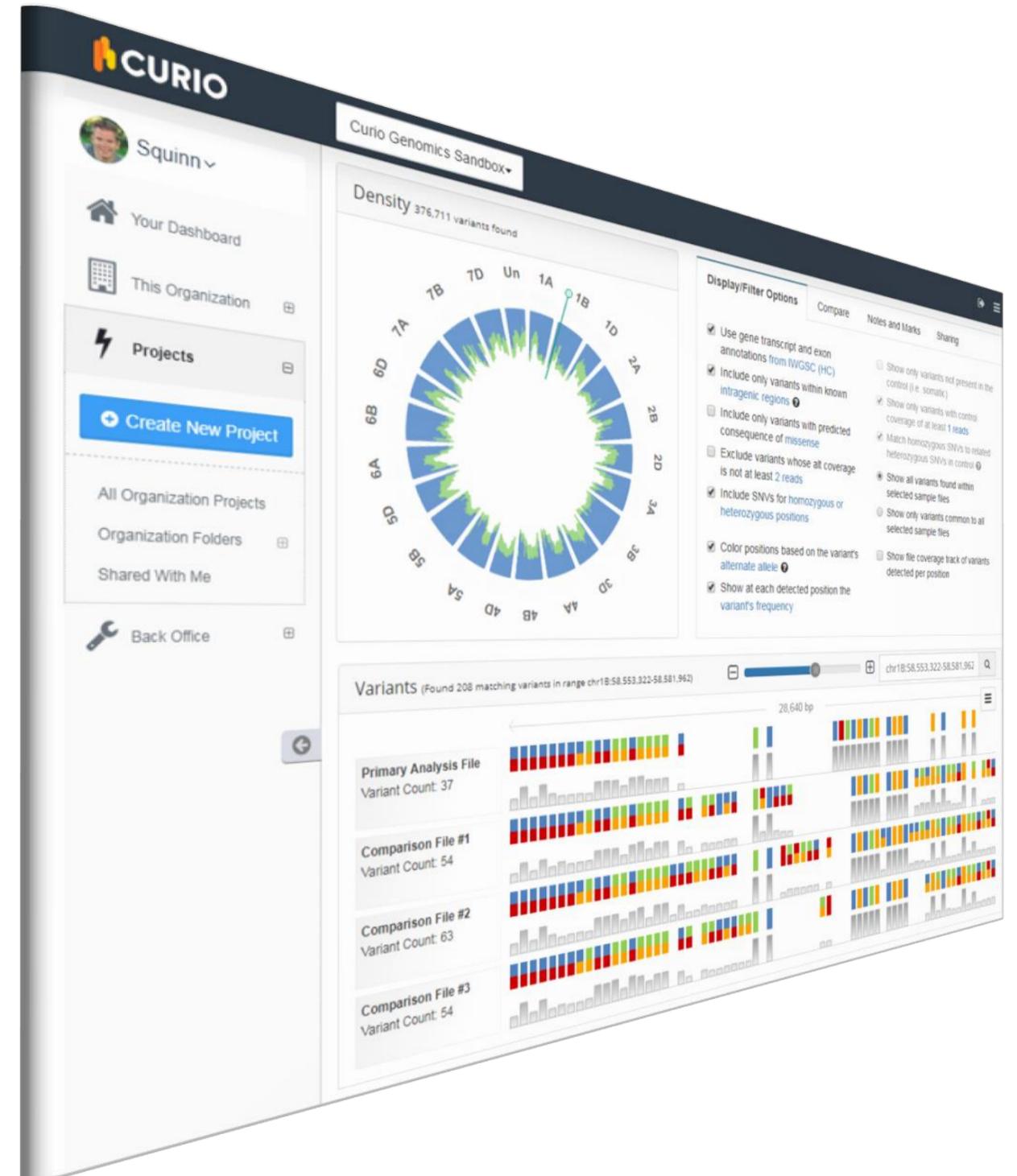
# Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio

## Presentation Overview

- What is Curio
- Sample Analysis with Chinese Spring Wheat
  - Read Alignment and Visualization
  - Handling Large Chromosomes
  - Mapping Algorithms & Coverage
  - Variant 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



# Bread Wheat DNA-Seq: Read Mapping and Navigation

- Incorporates Chinese Spring Wheat (*Triticum aestivum*) reference assembly from the IWGSC

Dashboard / All Projects / Project Details

Project Details

General Information

Project Name:

Species:

General Notes:

All Files **Raw Sequences**

[▶ Start An Alignment](#)

<input type="checkbox"/>	File Name			
<input type="checkbox"/>	SampleA-R2.fastq.gz			<a href="#">▶ Start an Alignment</a>
<input type="checkbox"/>	SampleA-R1.fastq.gz	2 Completed,	2019-12-06 17:23	<a href="#">▶ Start an Alignment</a>
<input checked="" type="checkbox"/>	SampleA Paired	16 Completed,	2019-12-06 17:00	<a href="#">▶ Start an Alignment</a>

Viewing 1-3

Align Paired-End Sequence Files ✕

Select Type:

Assembly: IWGSC WGA 1.0 (Chinese Spring Wheat)

---

- Alignment
- Pre-Processing
- Scoring
- Paired FASTQ

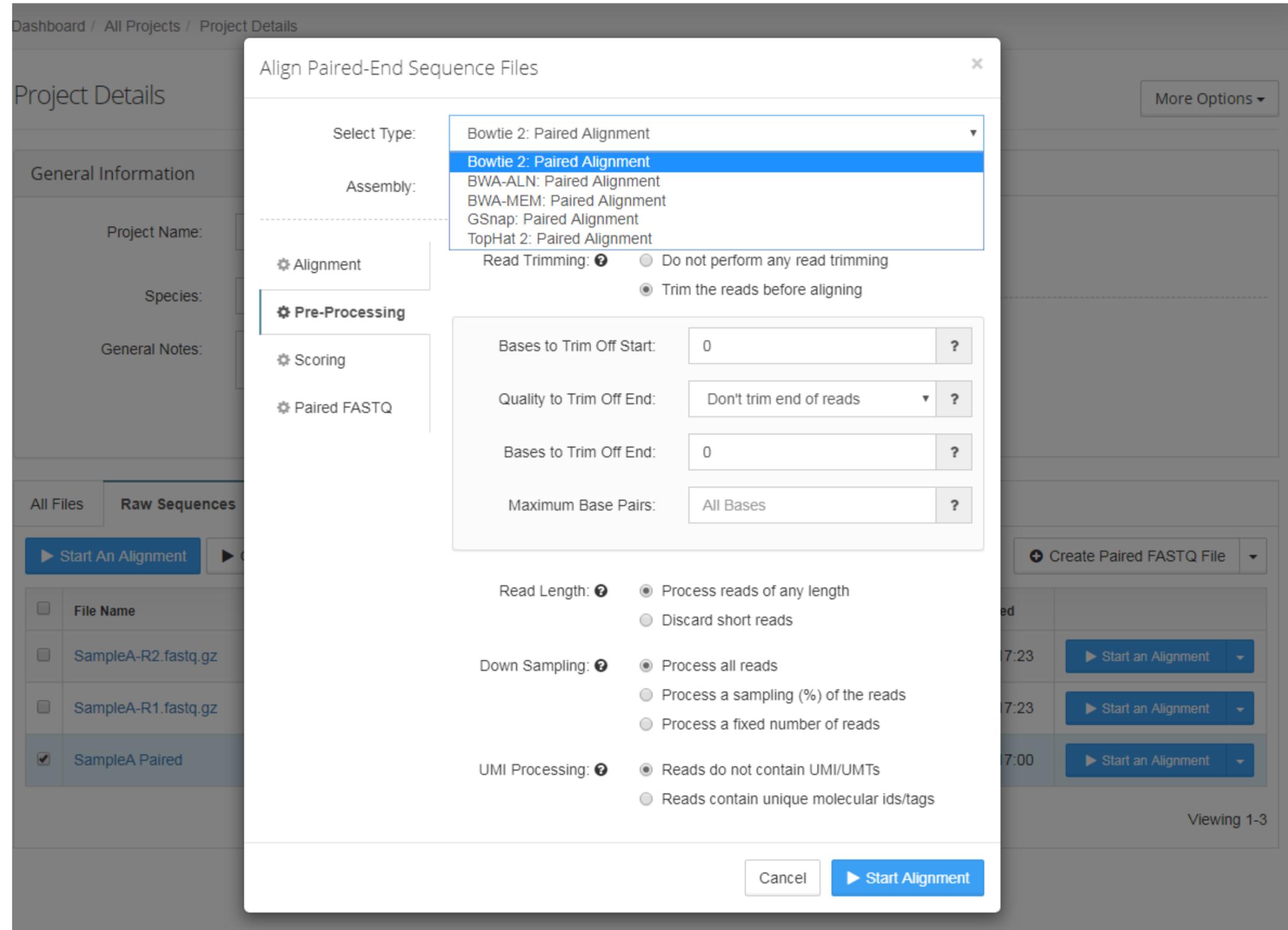
Read Trimming: ?  Do not perform any read trimming  
 Trim the reads before aligning

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

- Incorporates Chinese Spring Wheat (*Triticum aestivum*) reference assembly from the IWGSC
- 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



The screenshot displays the CURIO web interface with a modal dialog titled "Align Paired-End Sequence Files". The background shows a "Project Details" page with sections for "General Information" and "Raw Sequences".

**Align Paired-End Sequence Files**

Select Type: **Bowtie 2: Paired Alignment**

Assembly: [Empty]

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

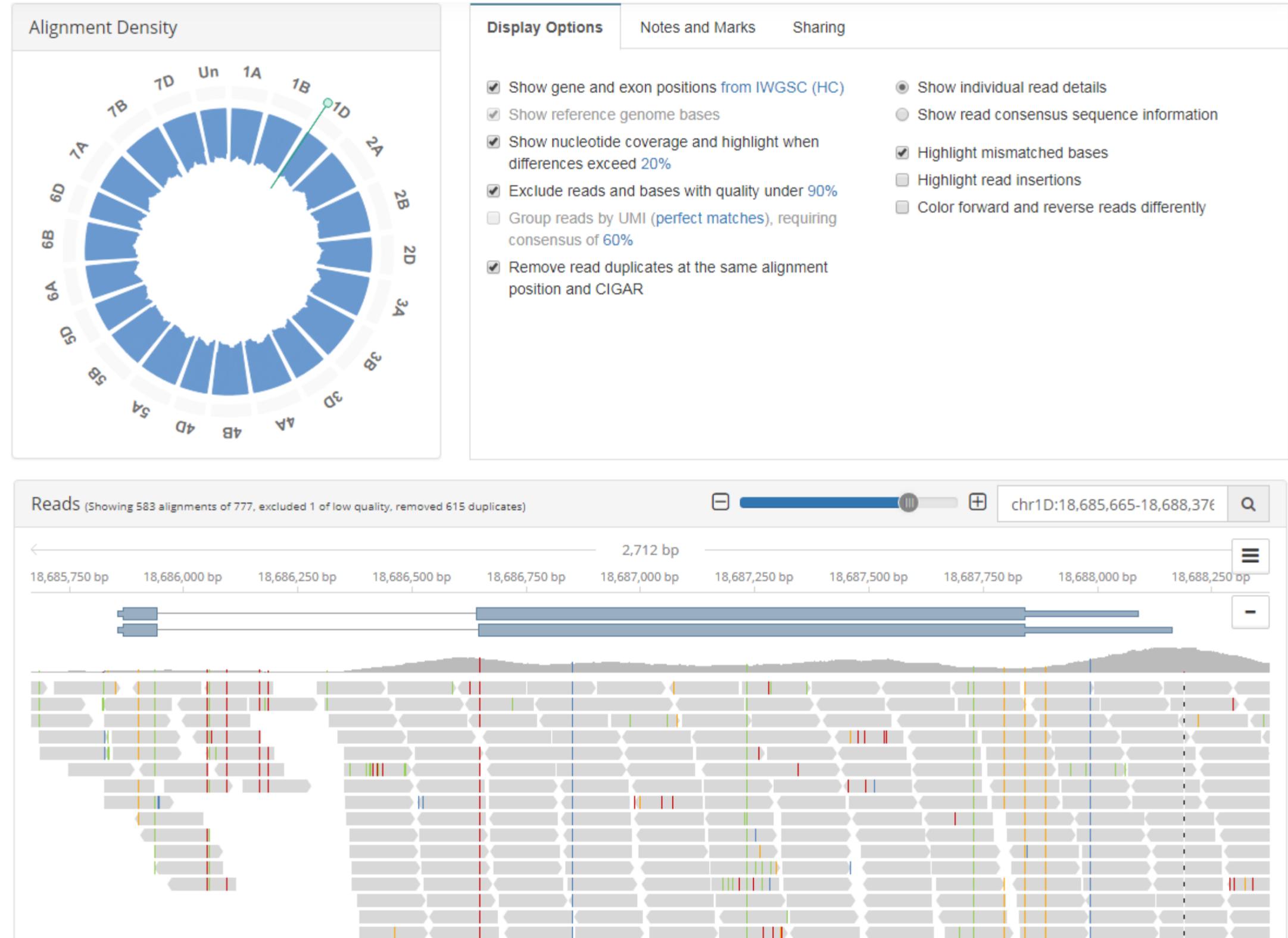
UMI Processing:  Reads do not contain UMI/UMTs  Reads contain unique molecular ids/tags

Cancel Start Alignment

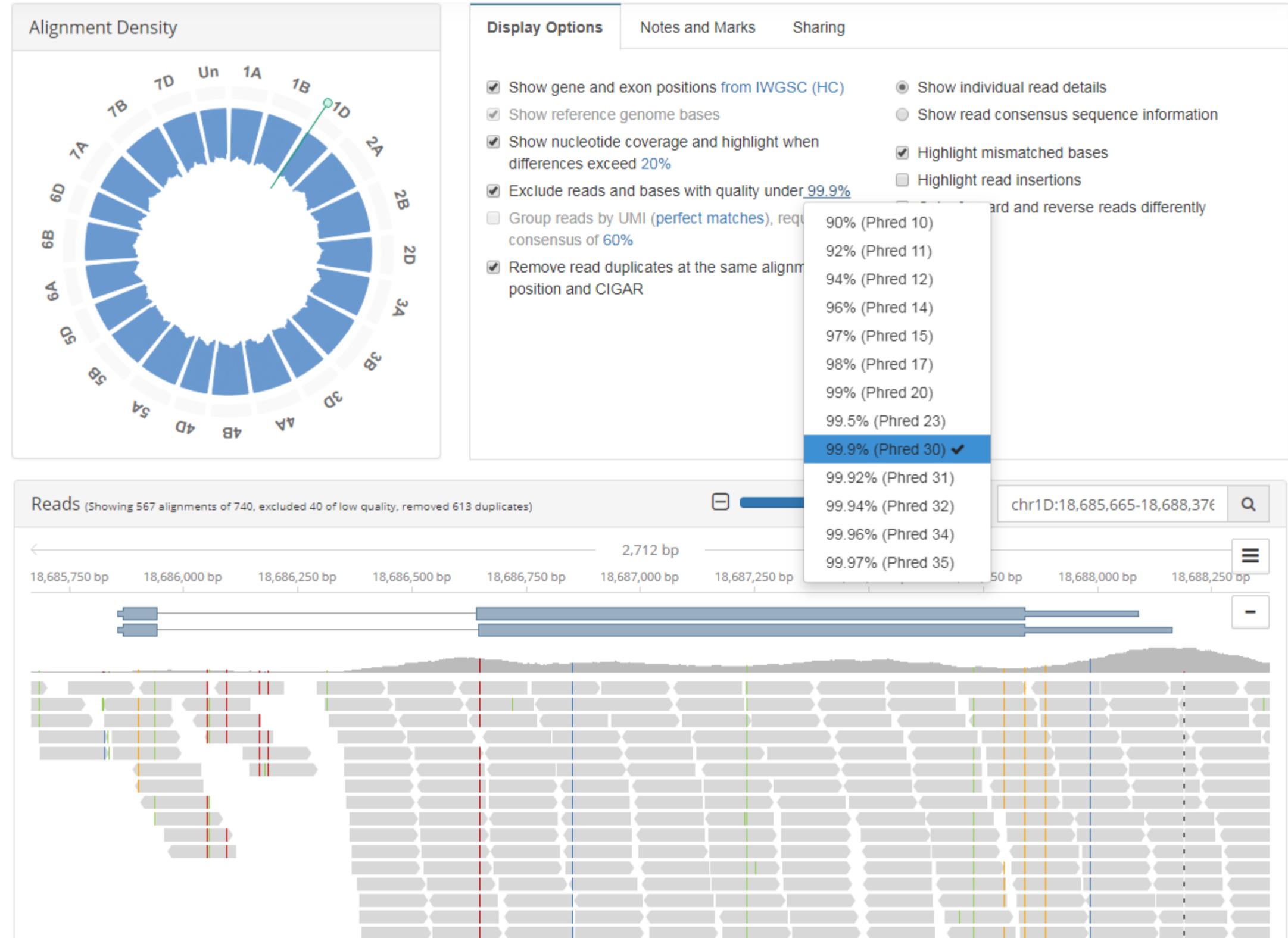
Background interface details:

- 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
- Create Paired FASTQ File
- 7:23 Start an Alignment
- 7:23 Start an Alignment
- 7:00 Start an Alignment
- Viewing 1-3

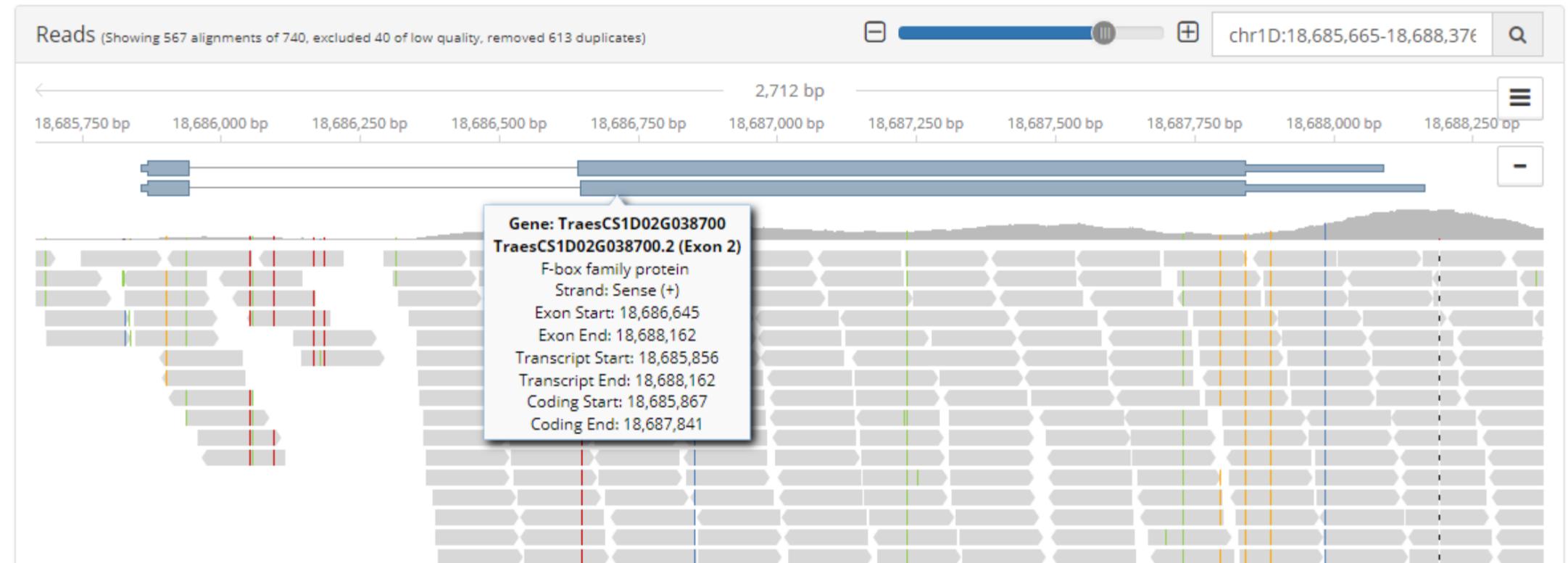
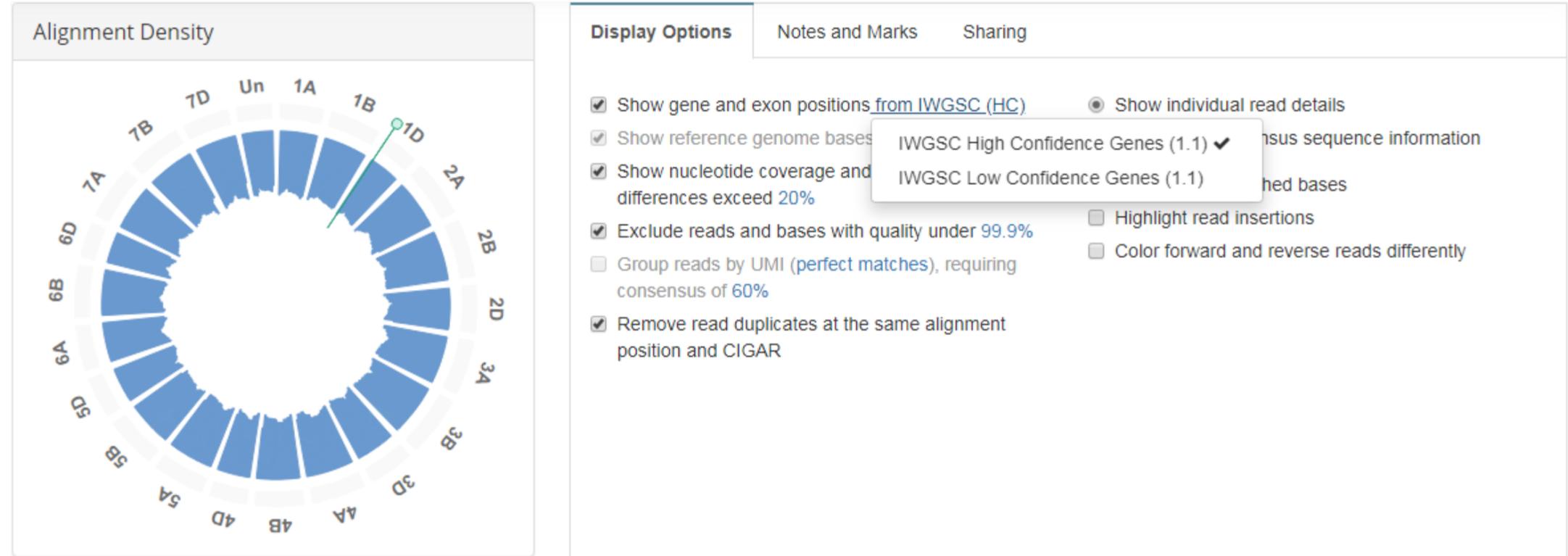
- Quickly browse and visualize reads from samples of any size, anywhere in the genome



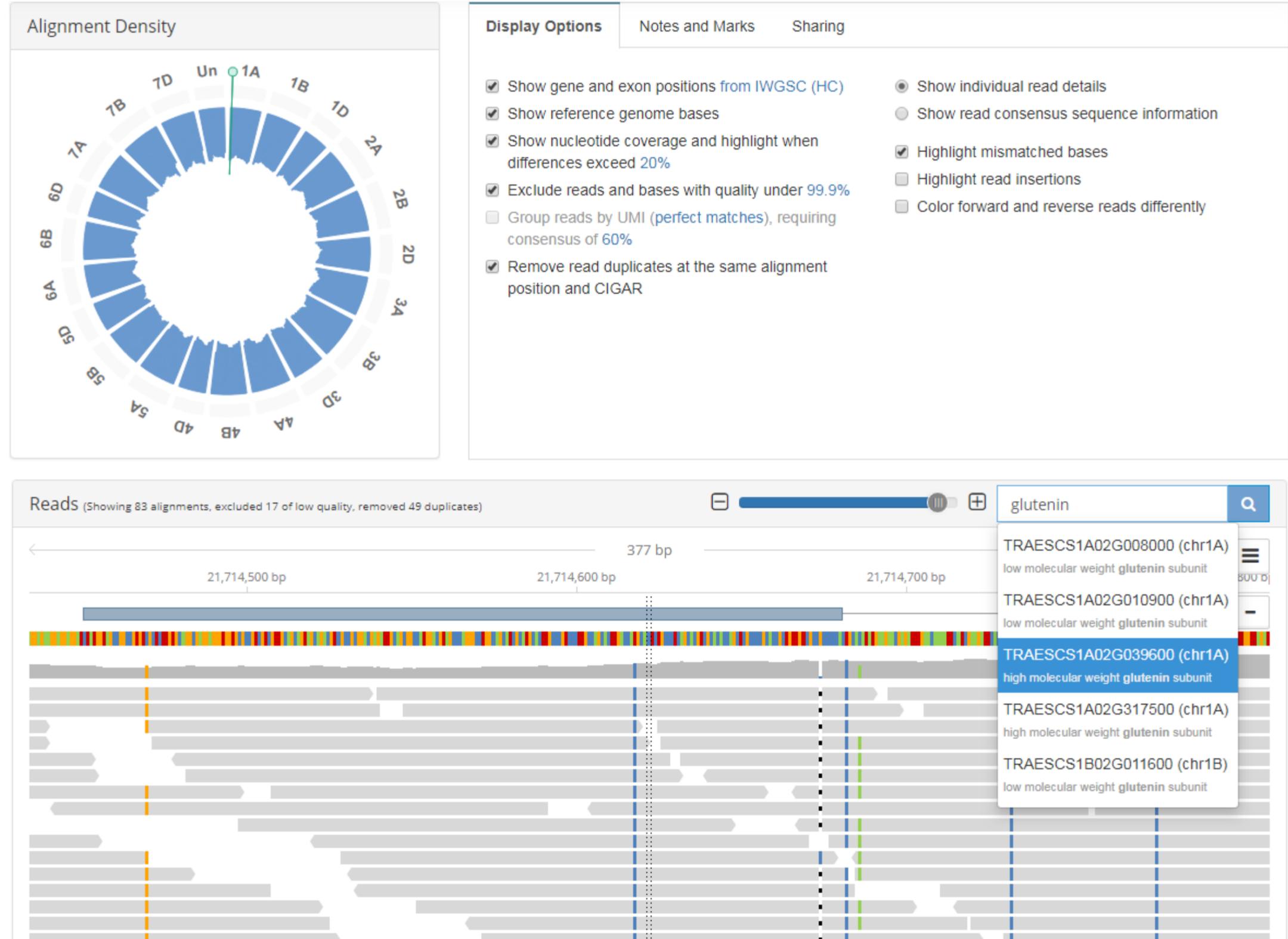
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly



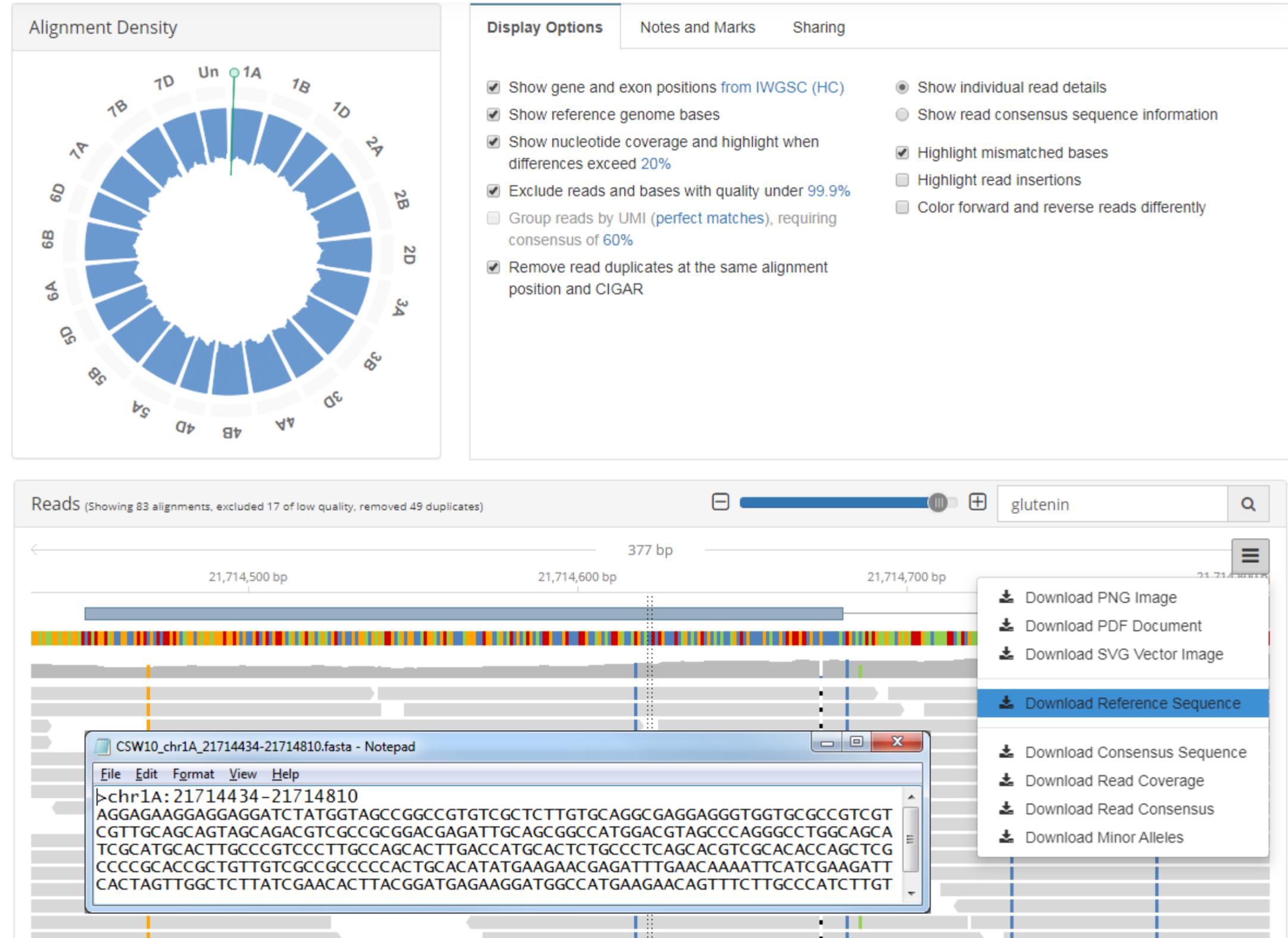
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline



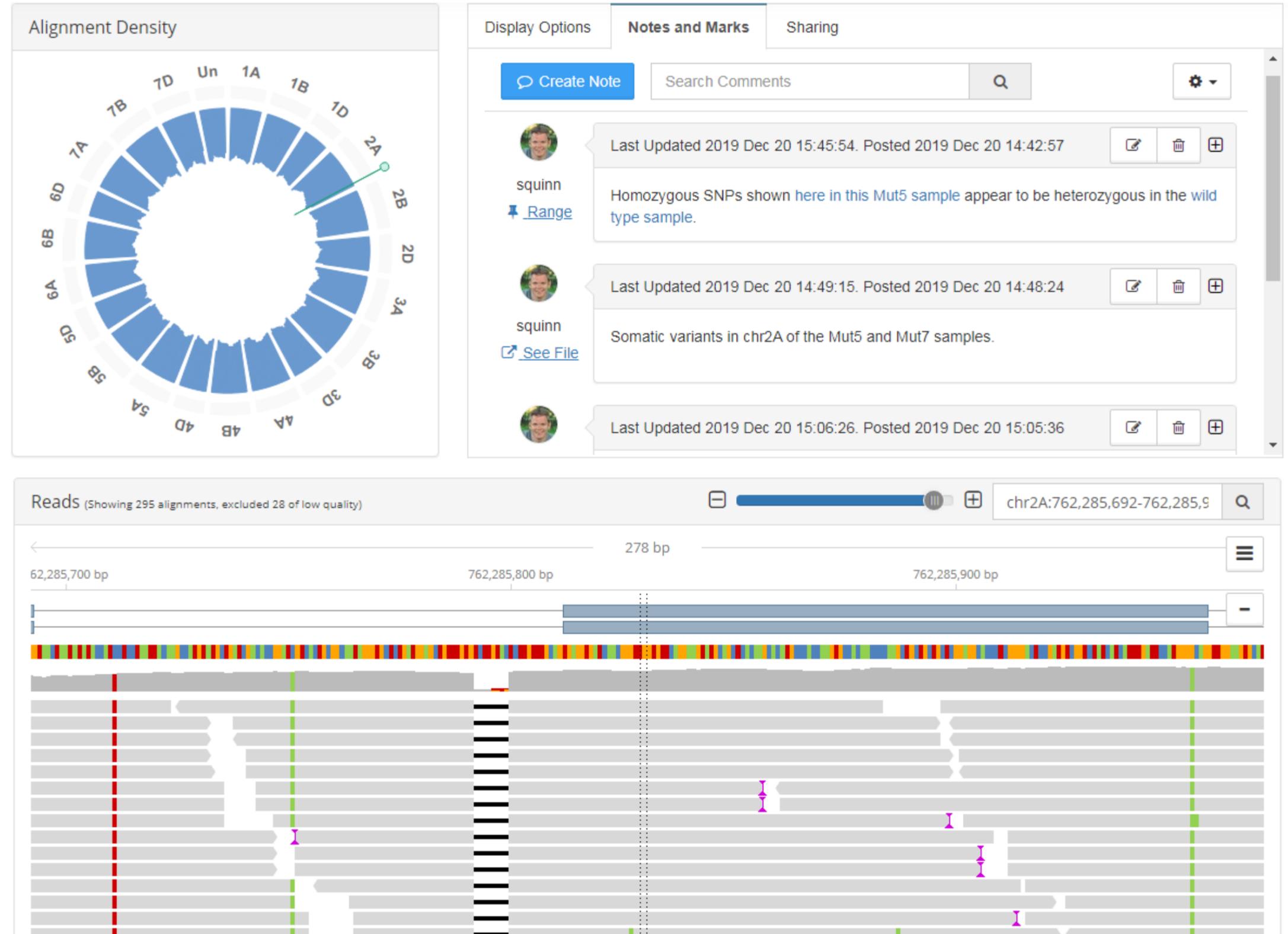
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline
- Quickly search and navigate the genome using the IWGSC annotation data as well



- 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



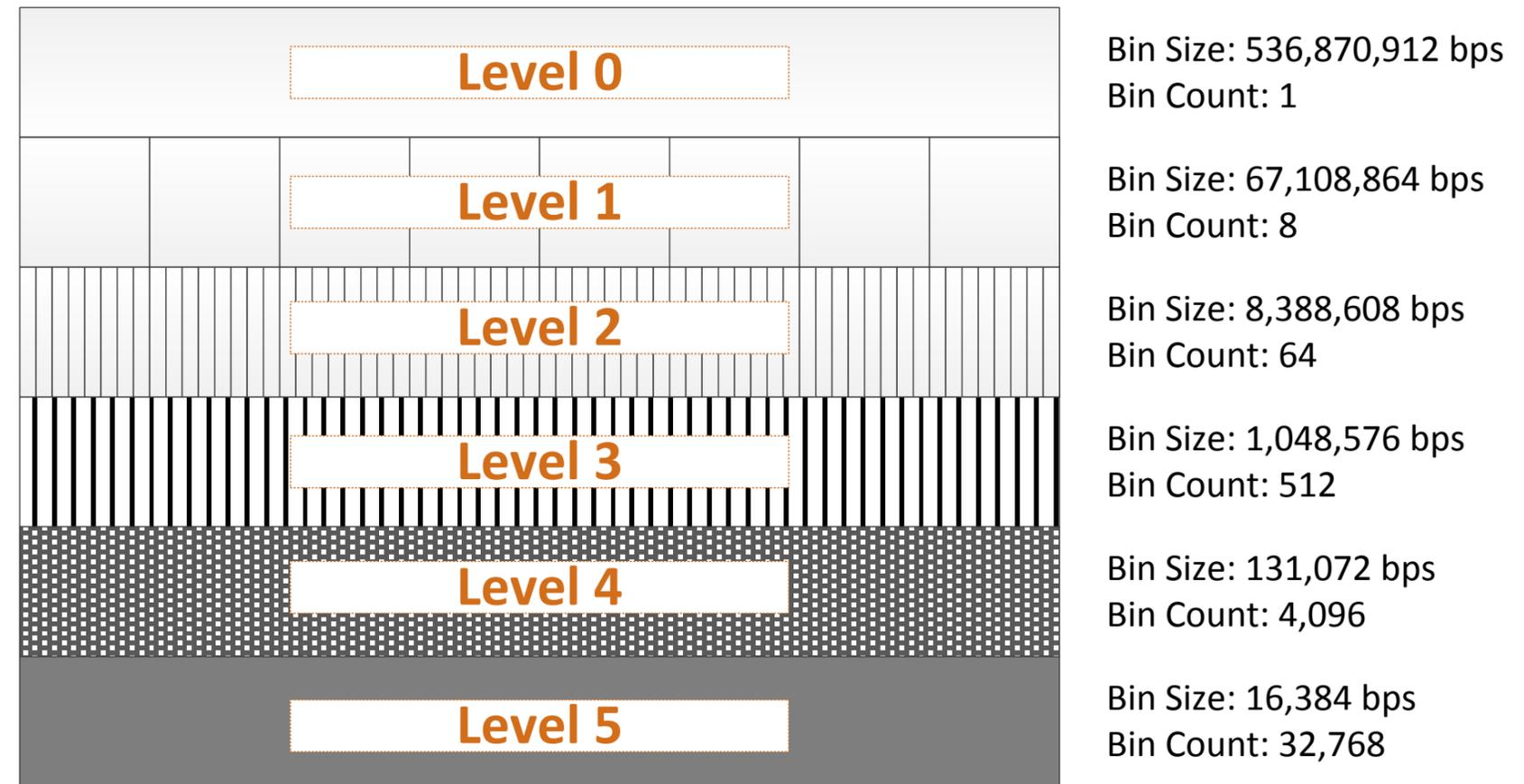
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline
- Quickly search and navigate the genome using the IWGSC annotation data as well
- Conveniently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



# Handling Large Chromosomes

# Standard Bin Strategy

- Binning strategy used in BAM index files (i.e. a “BAI” file)
- 6 levels deep and maximum bin size of: **536 million bases**
- Largest human chromosome (chr1): **249 million bases**
- Largest CSW chromosome (chr3B): **837 million bases**
- Forces a split chromosome approach or causes various tool compatibility issues



hg38: chr1

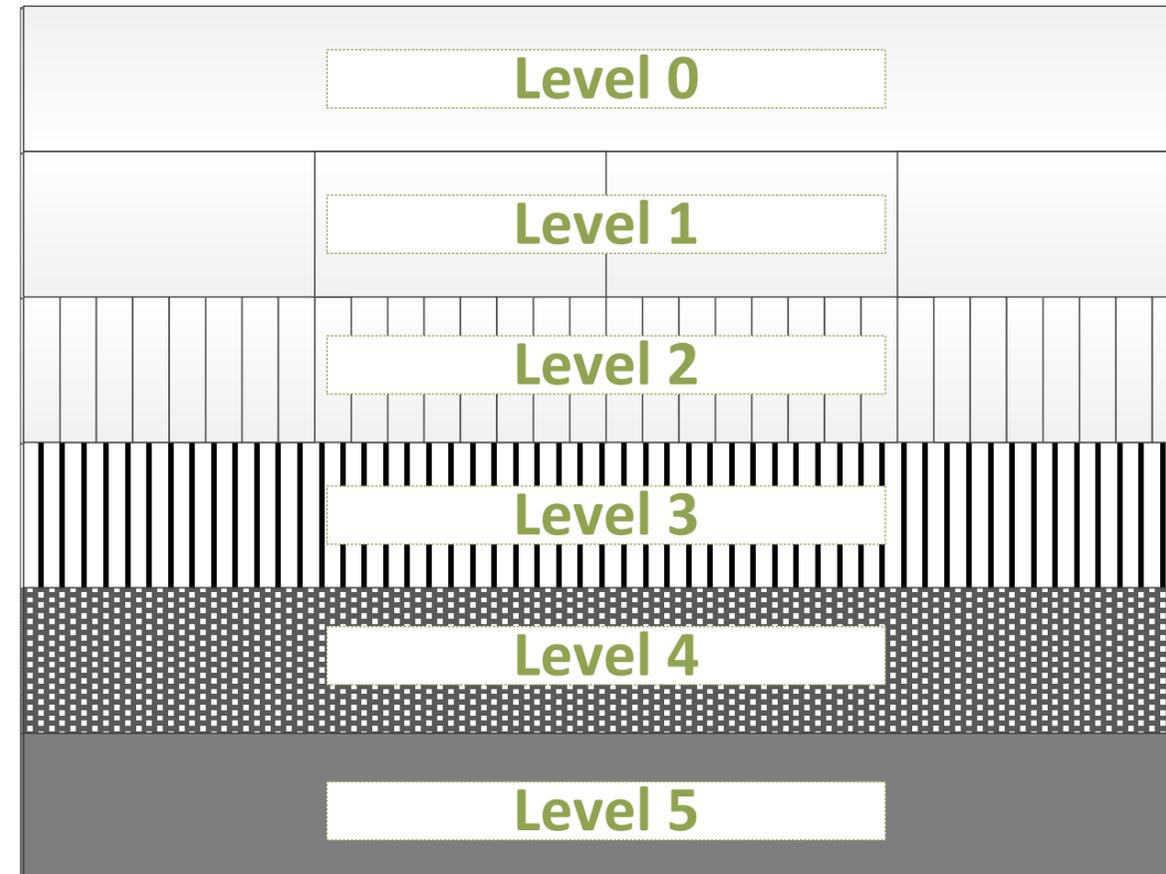
**56% too large**



Chinese Spring Wheat: chr3B

## A Better Approach

- Dynamic binning strategy based on species & assembly
- For wheat utilize a strategy based on the coordinate-sorted index specification with a minimum bit shift of 16
- Maintain 6 levels of depth & the number of bins per level, but use the first half of each level
- Max bin size: **1 billion bases**
- Largest CSW chromosome (chr3B): **837 million bases**



Bin Size: 1,073,741,824 bps  
Bin Count: 1

Bin Size: 268,435,456 bps  
Bin Count: 4 of 8

Bin Size: 33,554,432 bps  
Bin Count: 32 of 64

Bin Size: 4,194,304 bps  
Bin Count: 256 of 512

Bin Size: 524,288 bps  
Bin Count: 2,048 of 4,096

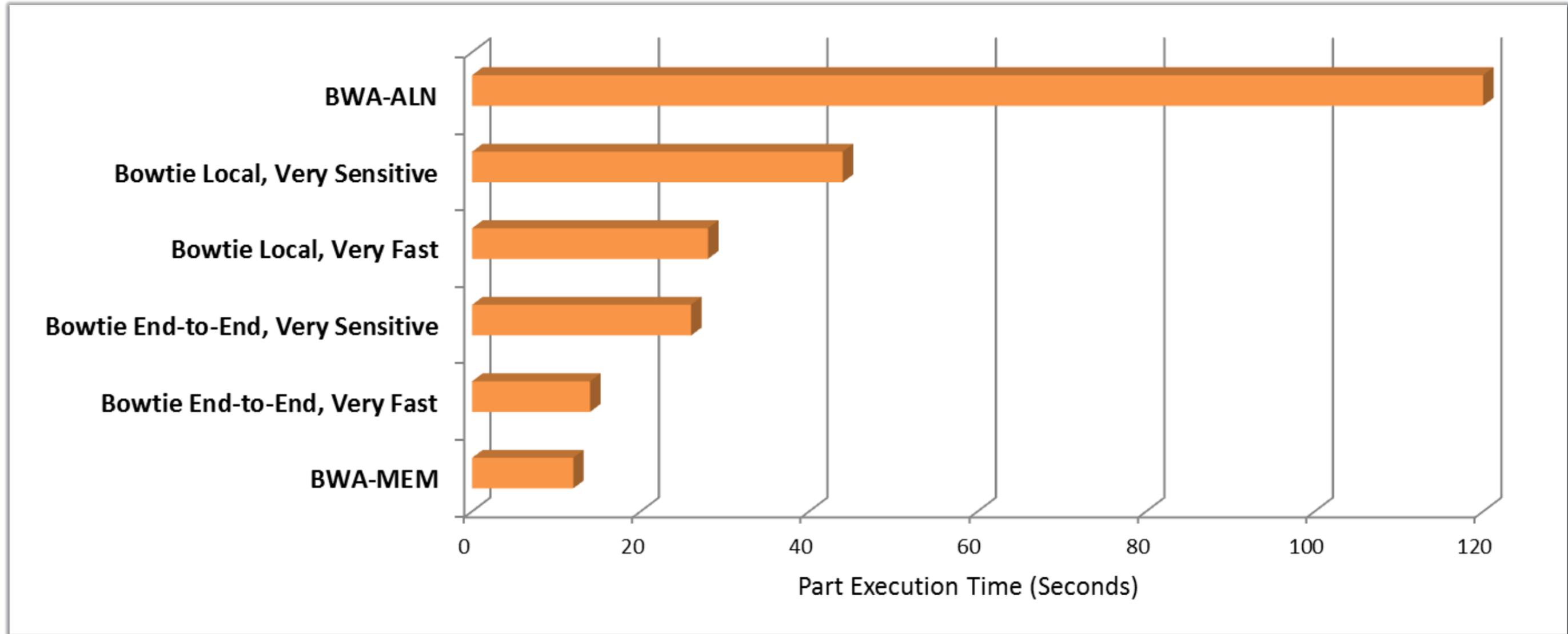
Bin Size: 65,536 bps  
Bin Count: 16,384 of 32,768



Chinese Spring Wheat: chr3B

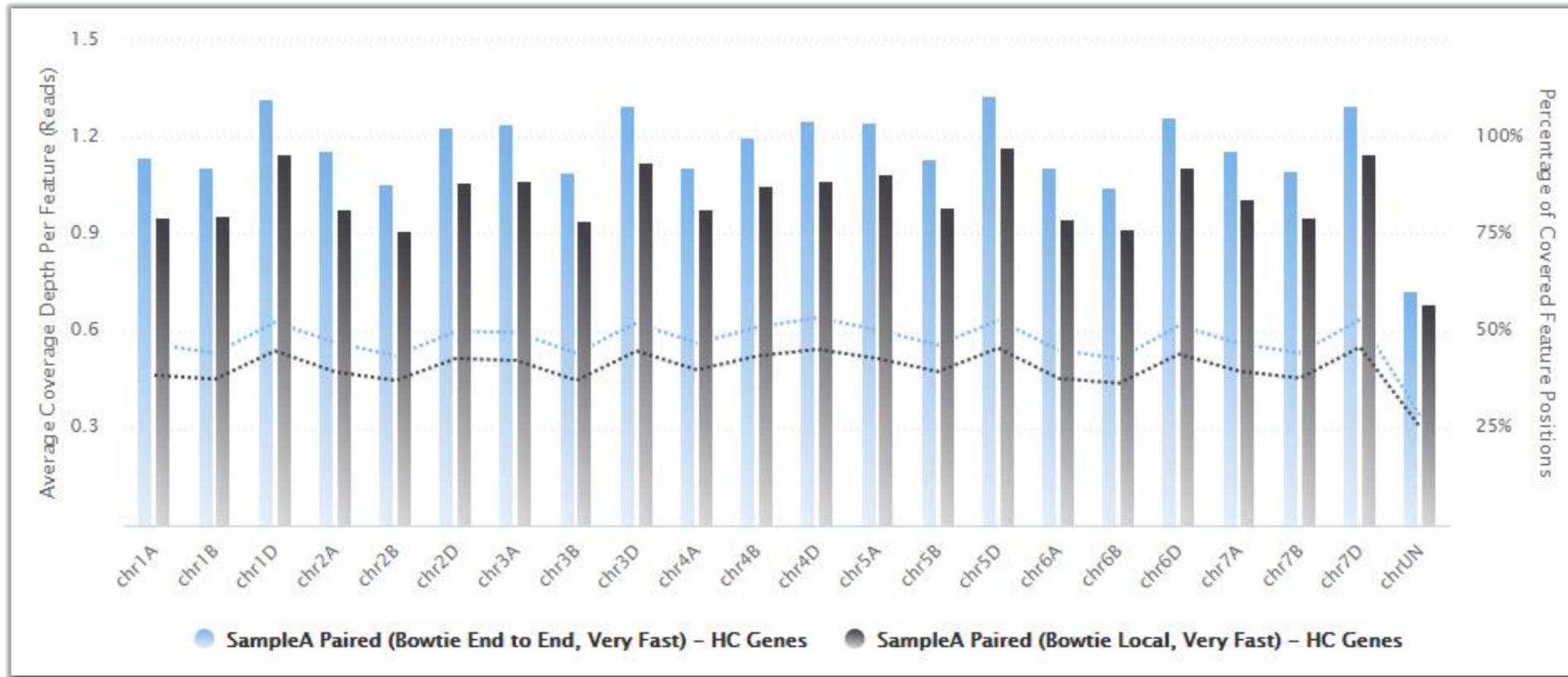
# DNA-Seq Bread Wheat: Alignment Algorithm Coverage Impacts

# Chinese Spring Wheat: Aligner Performance

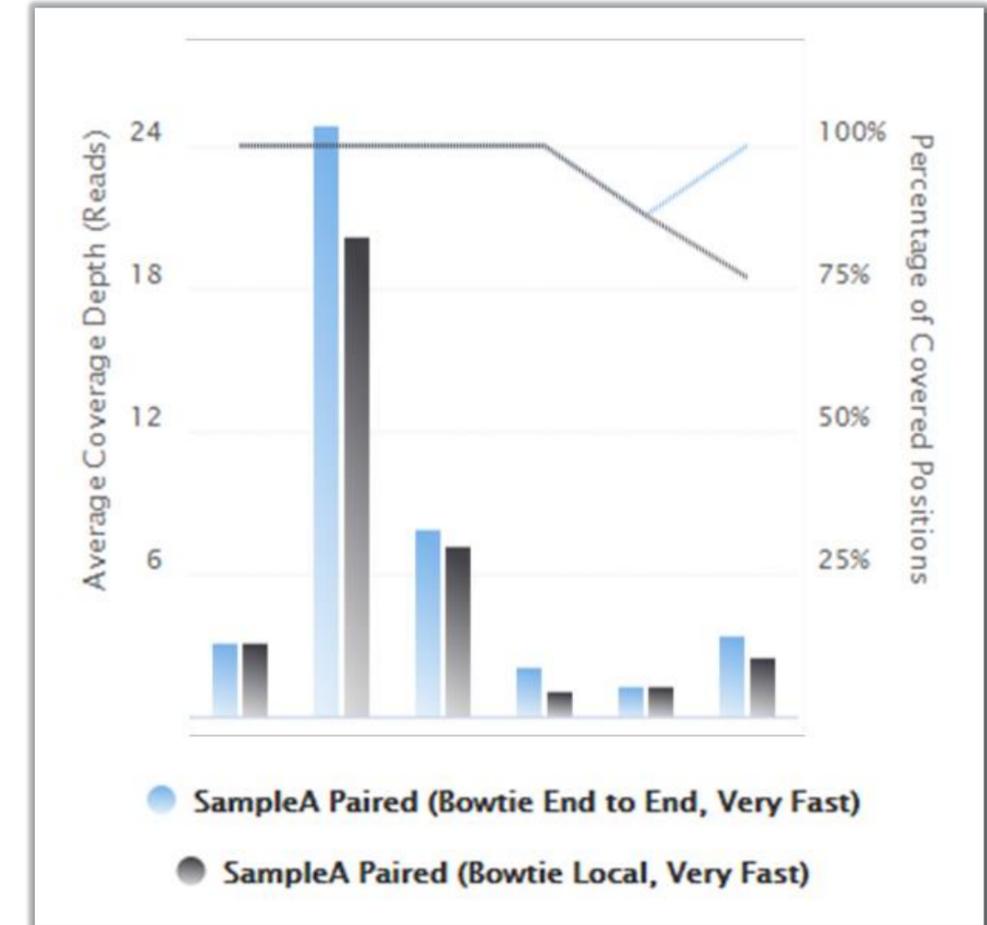


Note: the time needed for both the “bwa aln” and “bwa sampe” commands are included in the BWA-ALN metric shown above.

# Bowtie2 “End to End” vs “Local” Algorithm, Exome Coverage



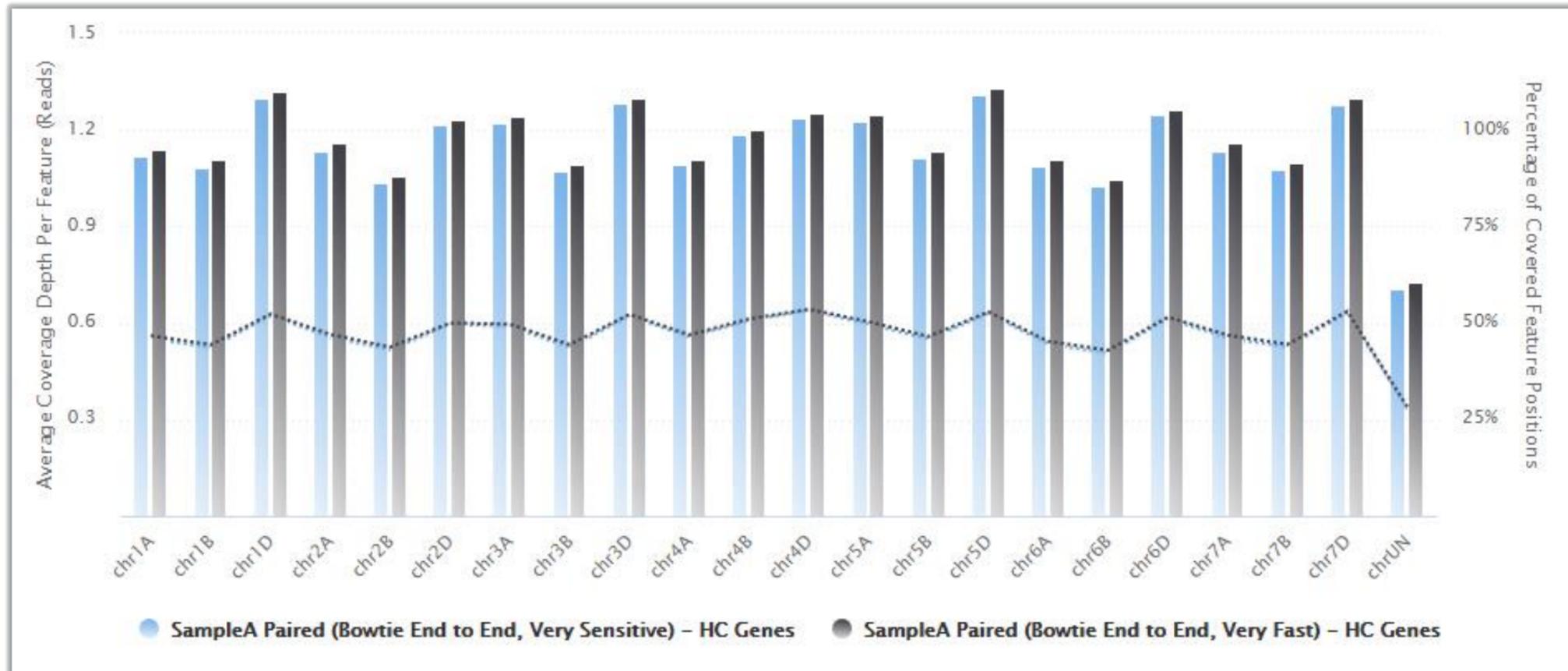
Average Exome Coverage Across all Chromosomes



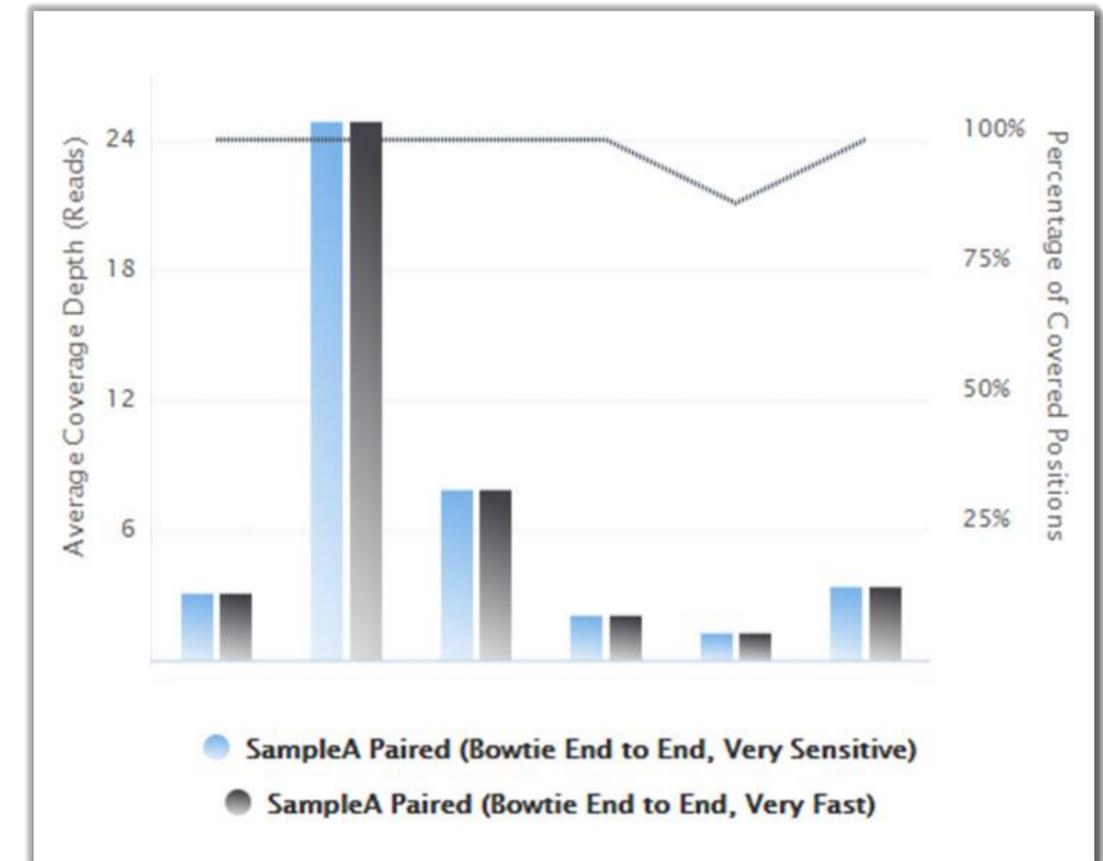
Exons of Single Gene

Note: L-lactate dehydrogenase gene (TRAESCS1A02G238700) shown here

# Bowtie2 “Very Sensitive” vs “Very Fast” Presets, Exome Coverage



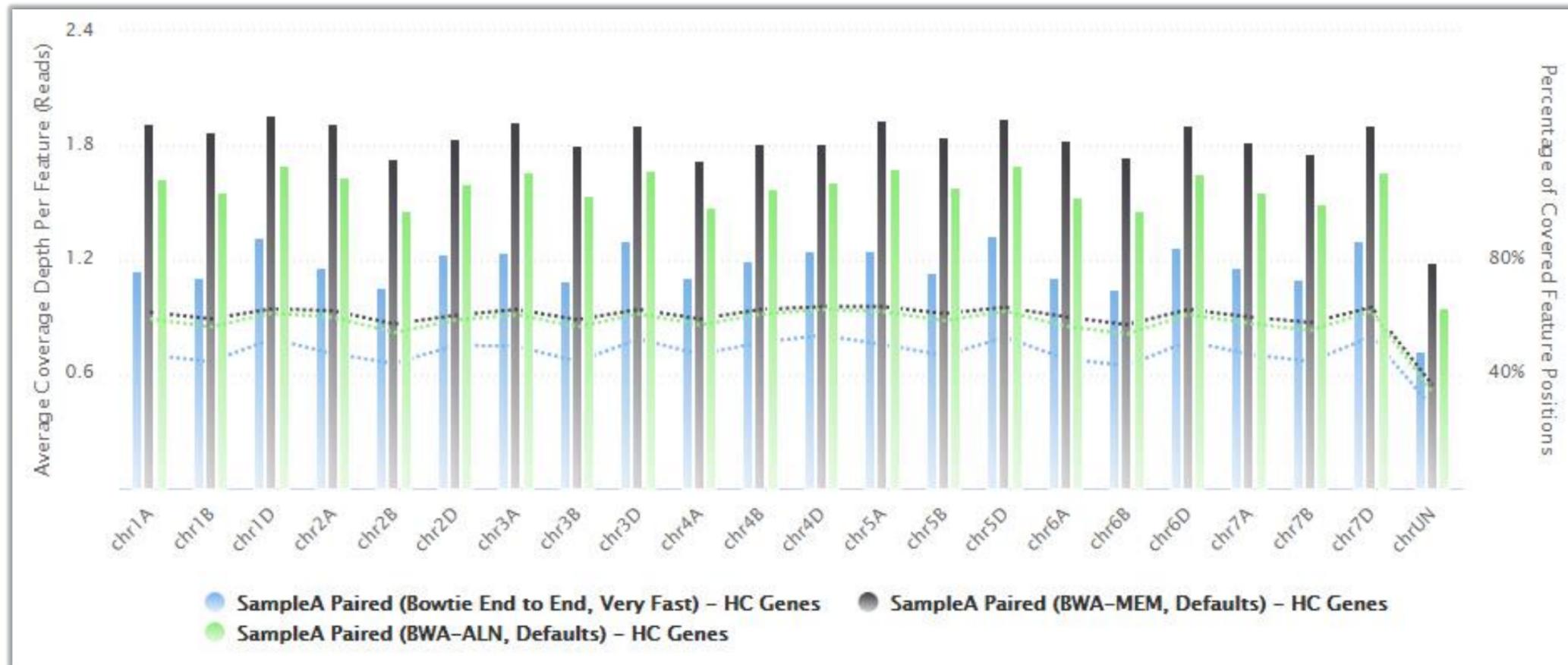
Average Exome Coverage Across all Chromosomes



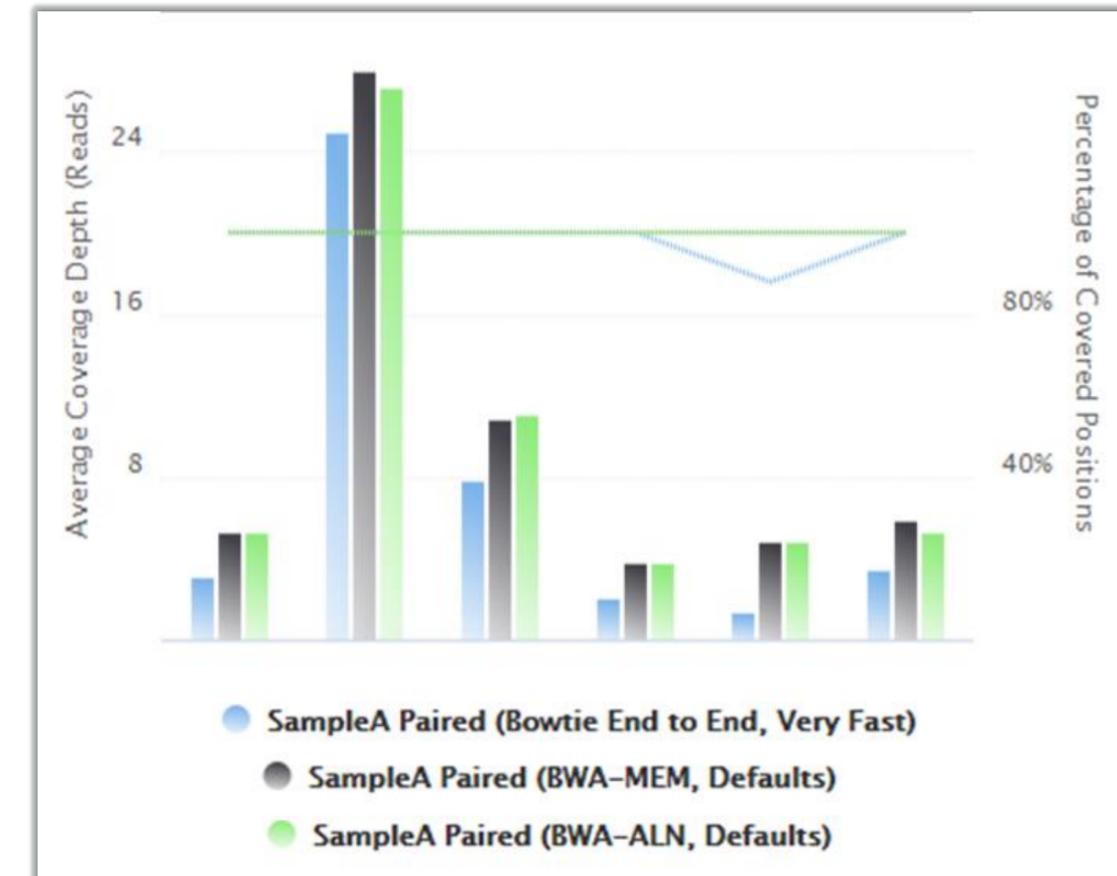
Exons of Single Gene

Note: L-lactate dehydrogenase gene (TRAESCS1A02G238700) shown here

# Aligner Algorithms, IWGSC “HC” Genes, Default Quality Filtering



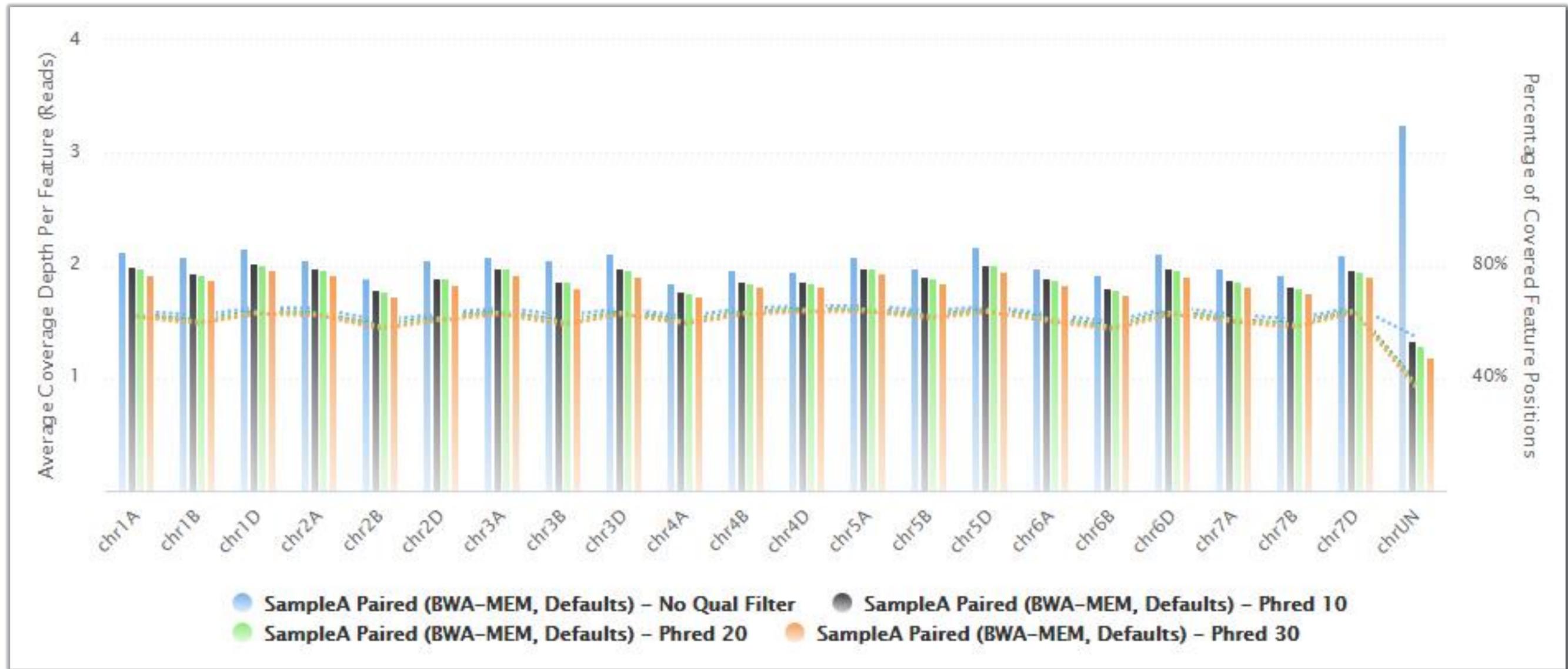
Average Exome Coverage Across all Chromosomes



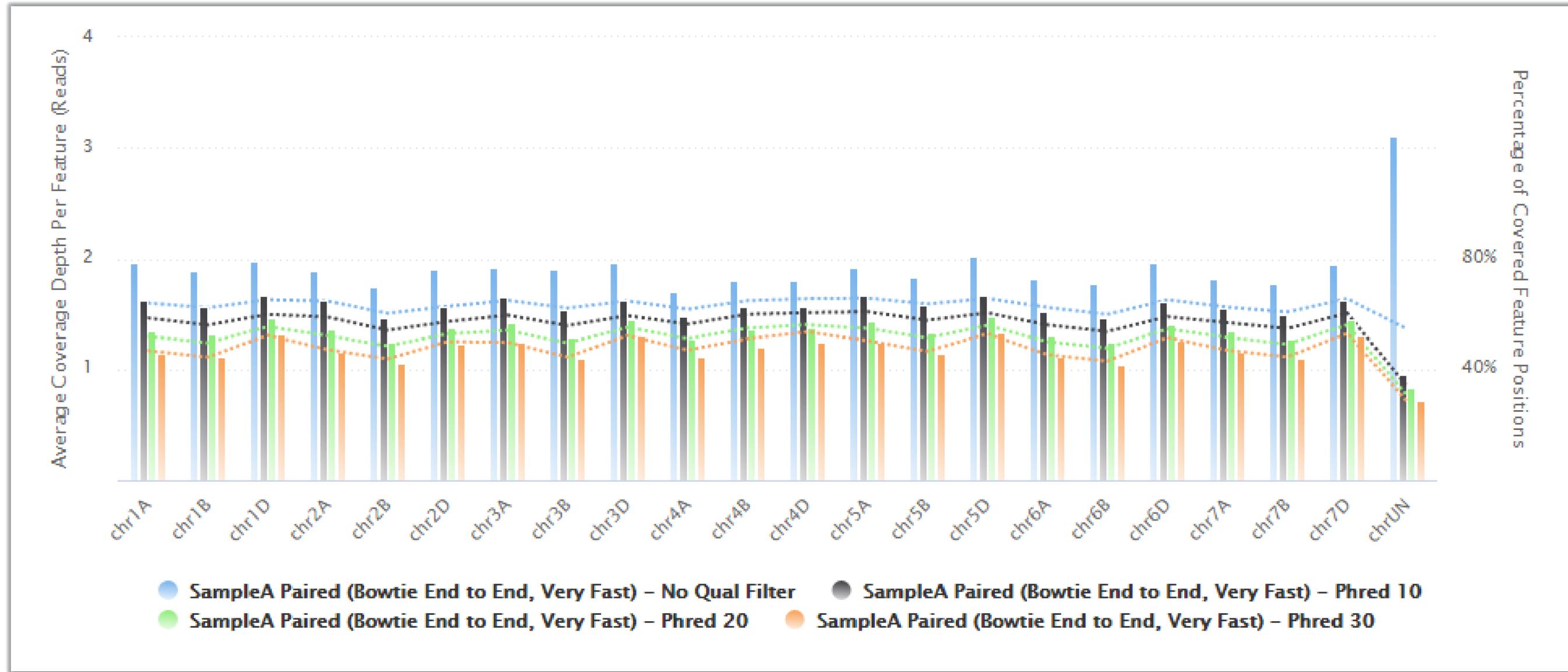
Exons of Single Gene

Note: L-lactate dehydrogenase gene (TRAESCS1A02G238700) shown here

# BWA-MEM Read Mapping Quality Impact on Coverage

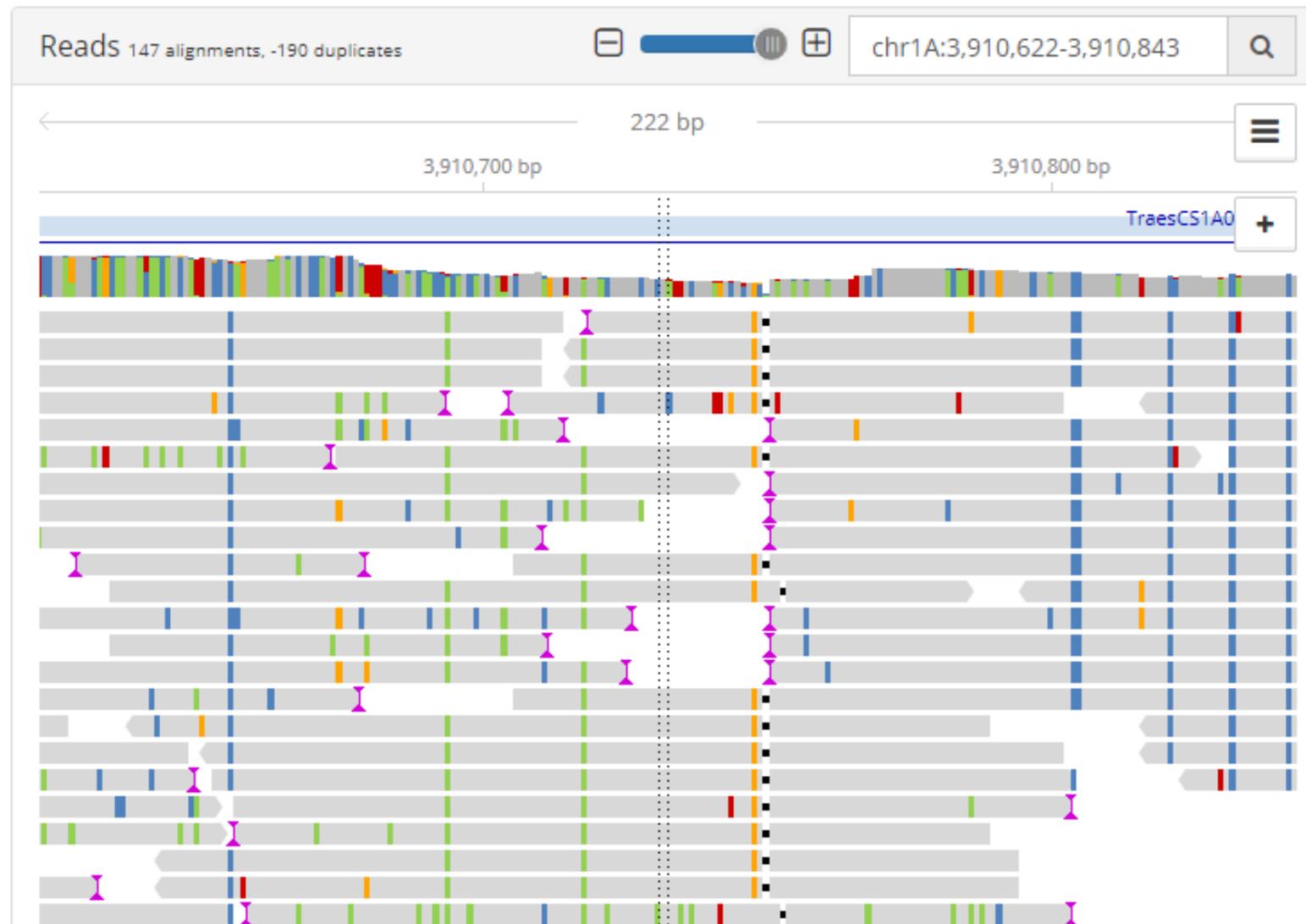


# Bowtie2 Read Mapping Quality Impact on Coverage



- BWA-MEM and Bowtie2 with no quality filter applied
- Visualizing low quality base calls within aligned reads

BWA-MEM / No Quality Filter

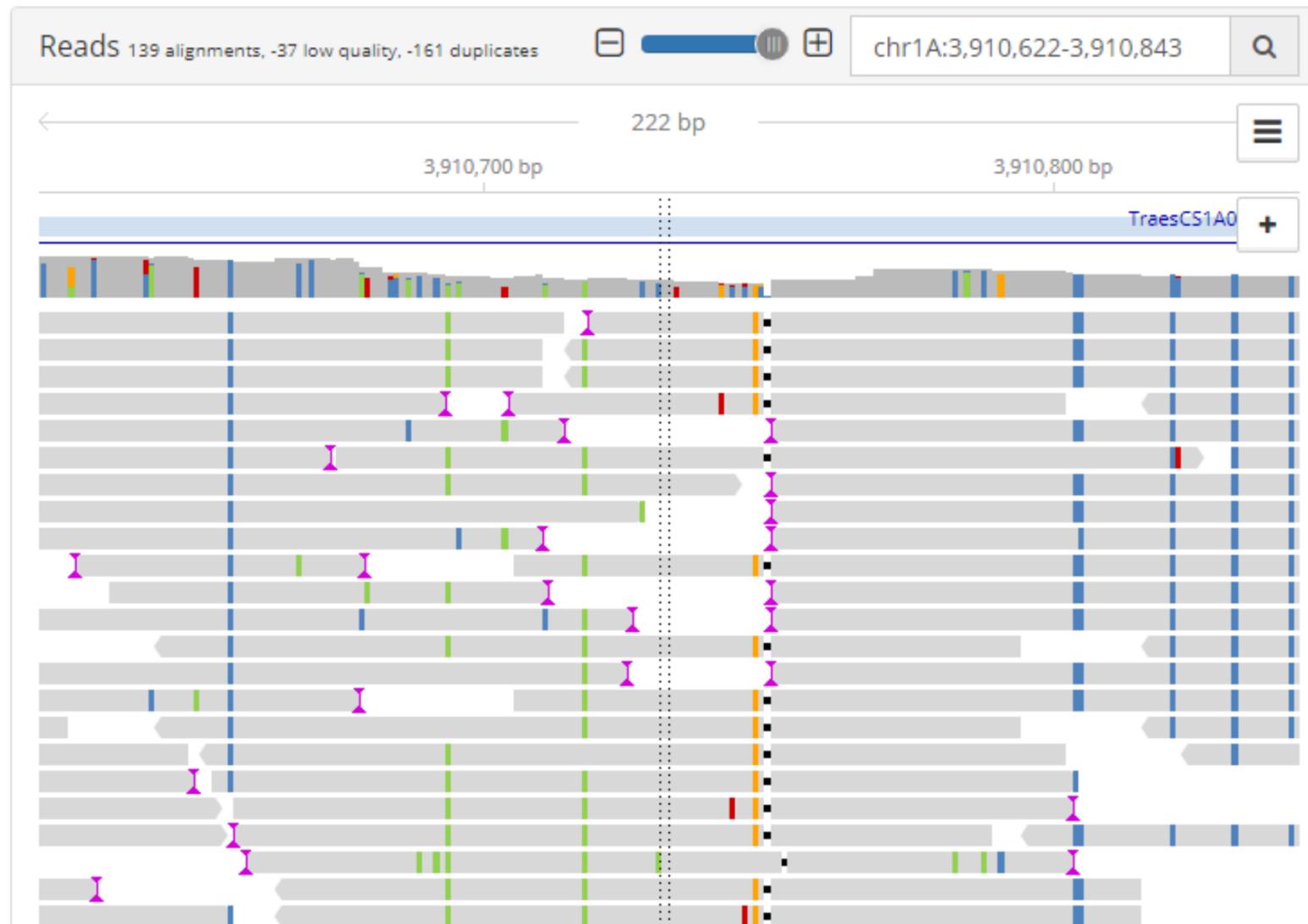


Bowtie2 / No Quality Filter

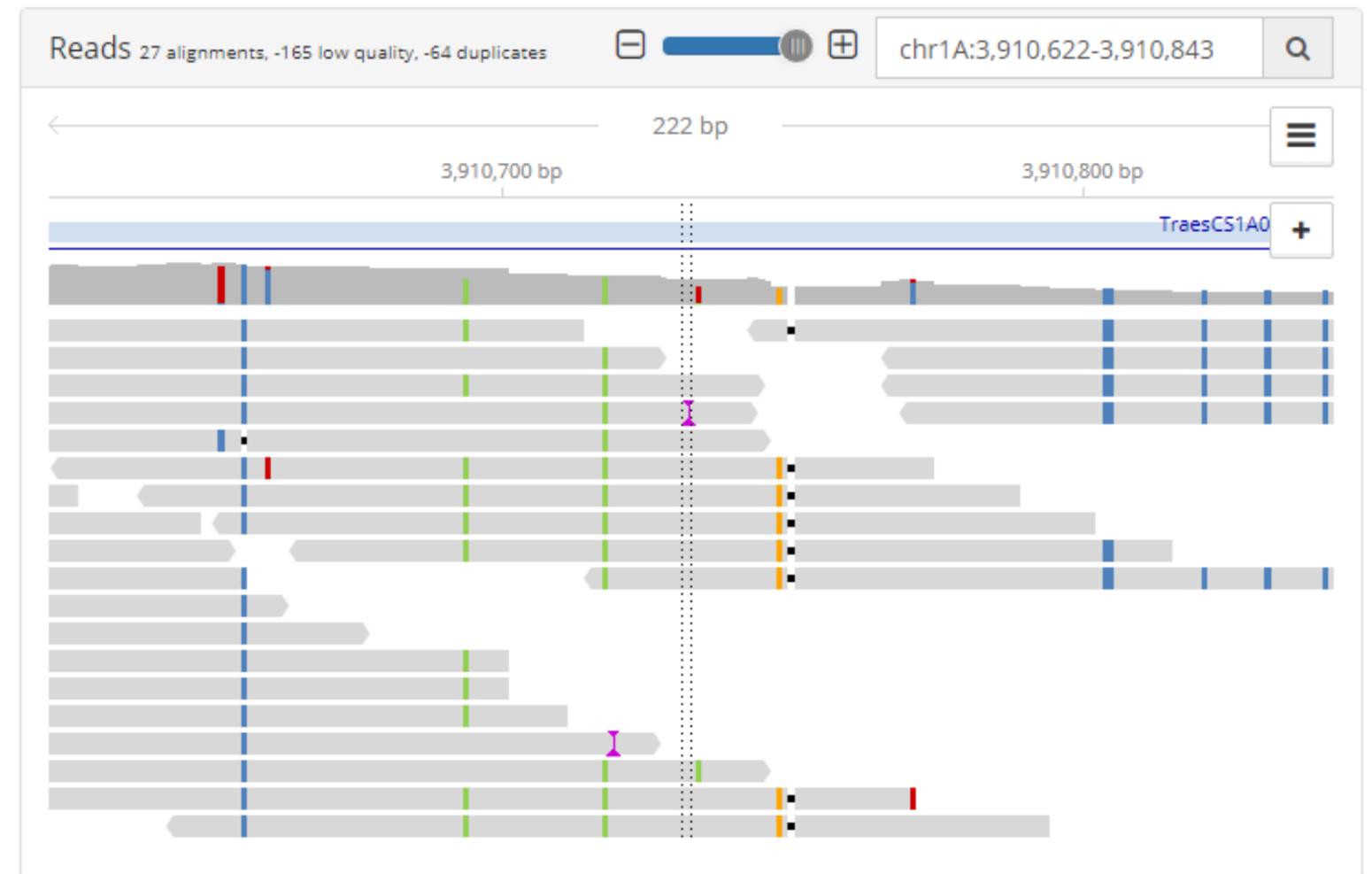


- BWA-MEM and Bowtie2 with a quality filter set to 99% (i.e. Phred 30)
- Visualizing impacts of different mapping algorithm approaches to alignment quality

BWA-MEM / Phred 30 or Higher



Bowtie2 / Phred 30 or Higher



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

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%

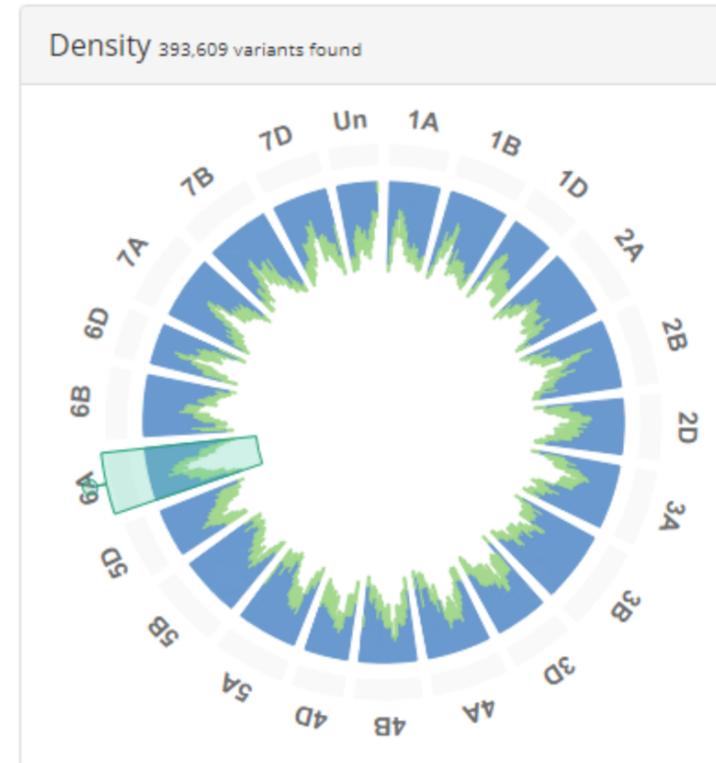
Rare Allele Detection:  Do not attempt to detect rare alleles  
 Include rare allele detection in the analysis

Cancel ▶ Start Variant Detection Analysis

- 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

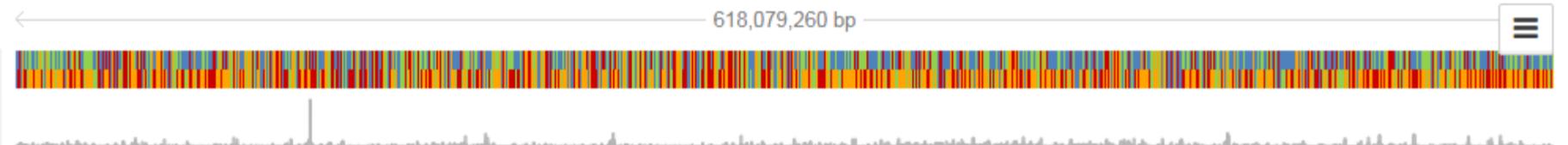
Variants (Found 18,248 matching variants in range chr6A:1-618,079,261)



chr6A:1-618,079,261



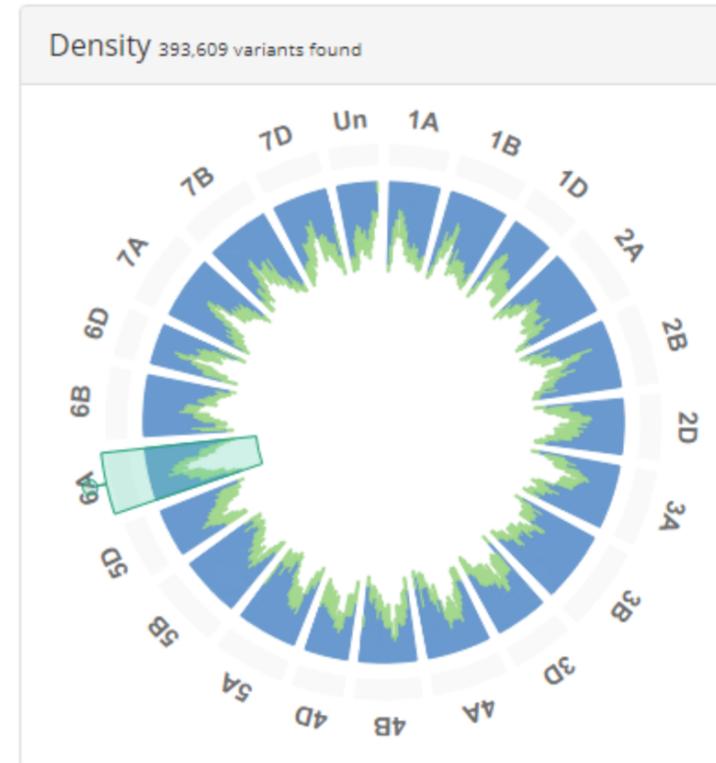
Primary Analysis File  
Variant Count: 18,248



- 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

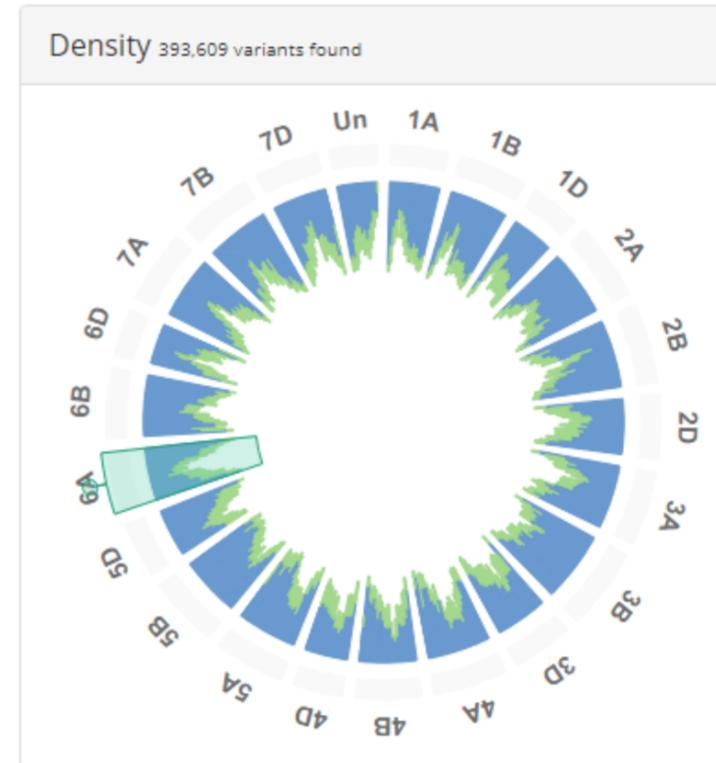
Viewing 1-8 of 8



- 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

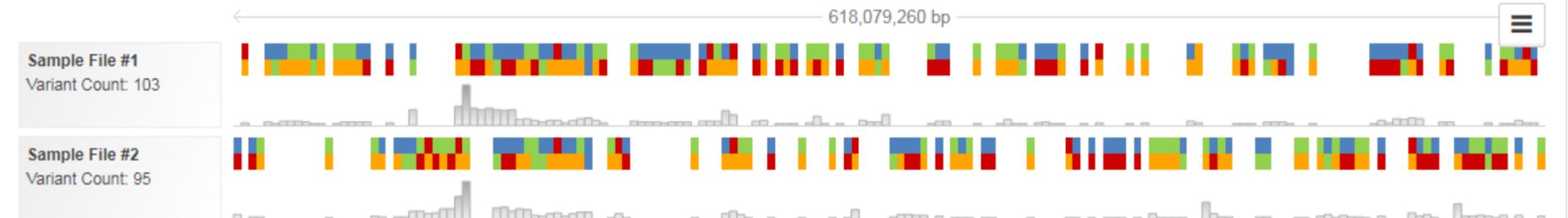
Sharing

- 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)



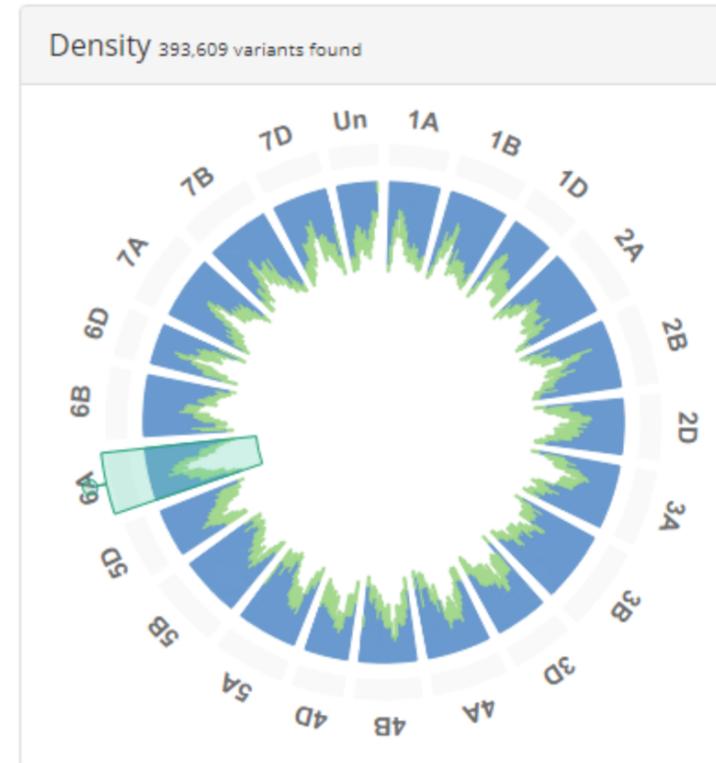
chr6A:1-618,079,261



- 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 ▾



### Display/Filter Options

Compare

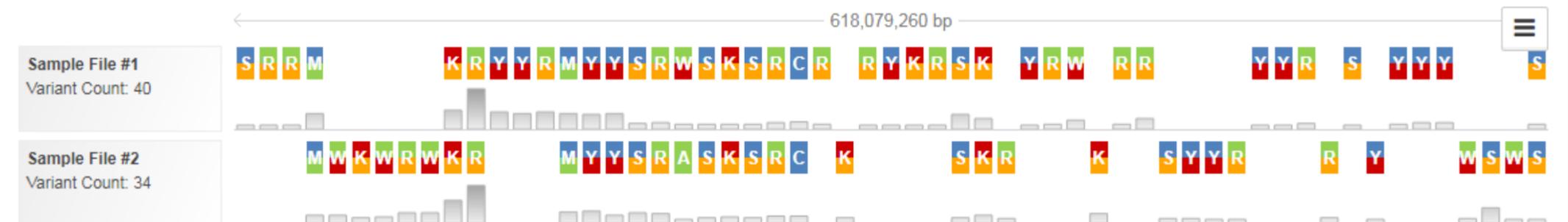
Notes and Marks

Sharing

- Use gene transcript and exon annotations from [IWGSC \(HC\)](#)
  - Include only variants within known [coding regions](#)
  - Include only variants with predicted [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](#)
  - Show only variants not present in the control (i.e. somatic)
  - Show only variants with control coverage of at least
- 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)

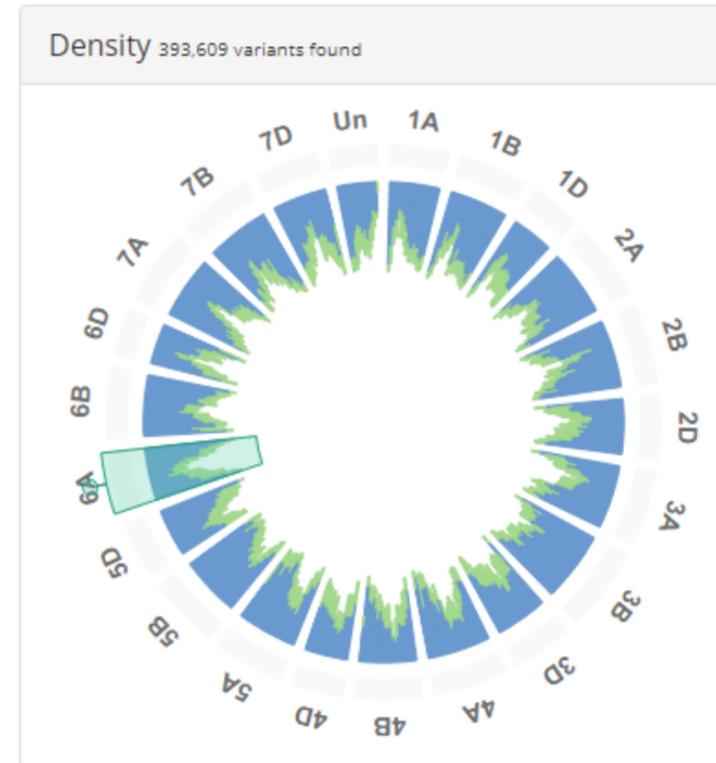
chr6A:1-618,079,261



- 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

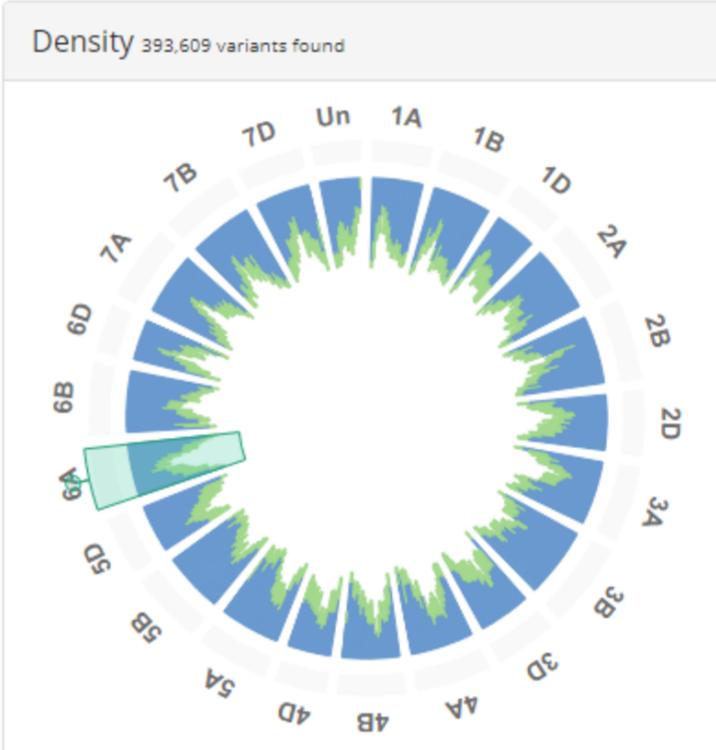
618,079,260 bp

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

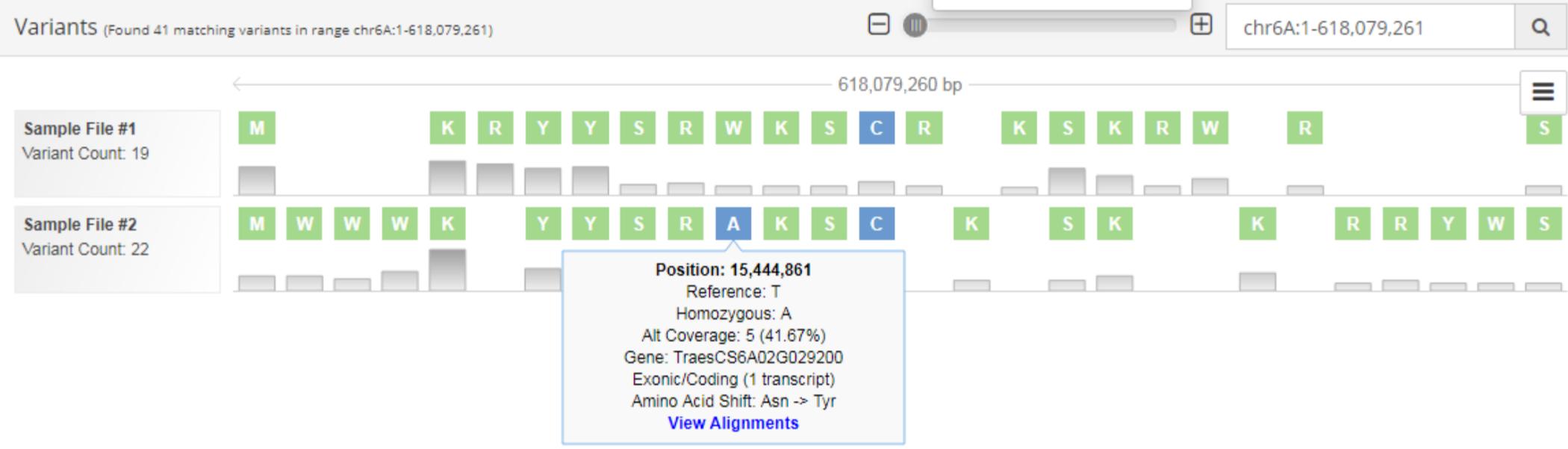
- Use gene transcript and exon annotations from [IWGSC \(HC\)](#)
- Include only variants within known [coding regions](#) ⓘ
- Include only variants with predicted consequence of [not synonymous](#)
- Exclude variants whose alt coverage is not at least 5 reads
- Include SNVs for [homozygous or heterozygous 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](#)

# Bread Wheat RNA-Seq: 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 ?

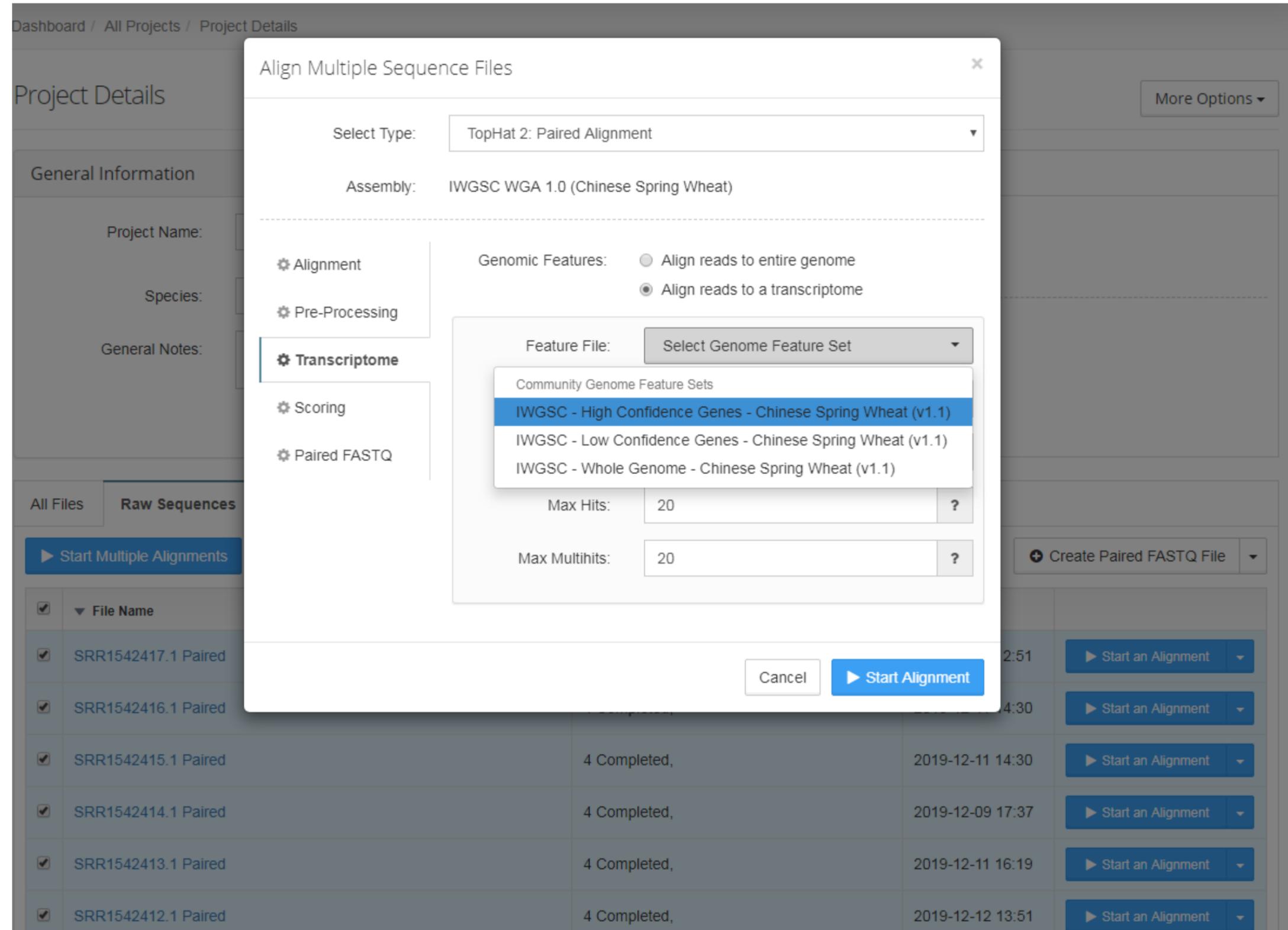
Max Multihits: 20 ?

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 ?

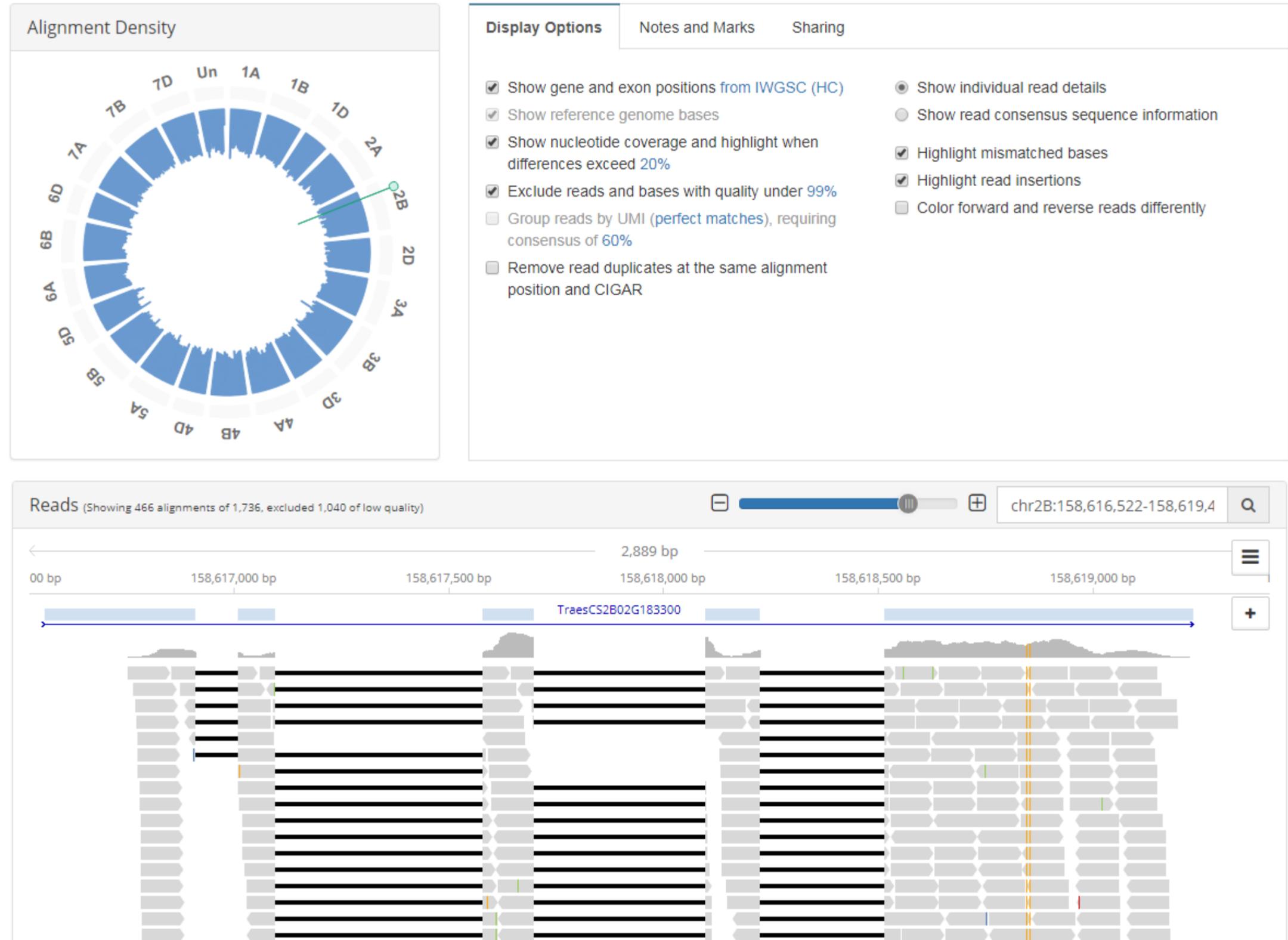
Max Multihits: 20 ?

Cancel Start Alignment

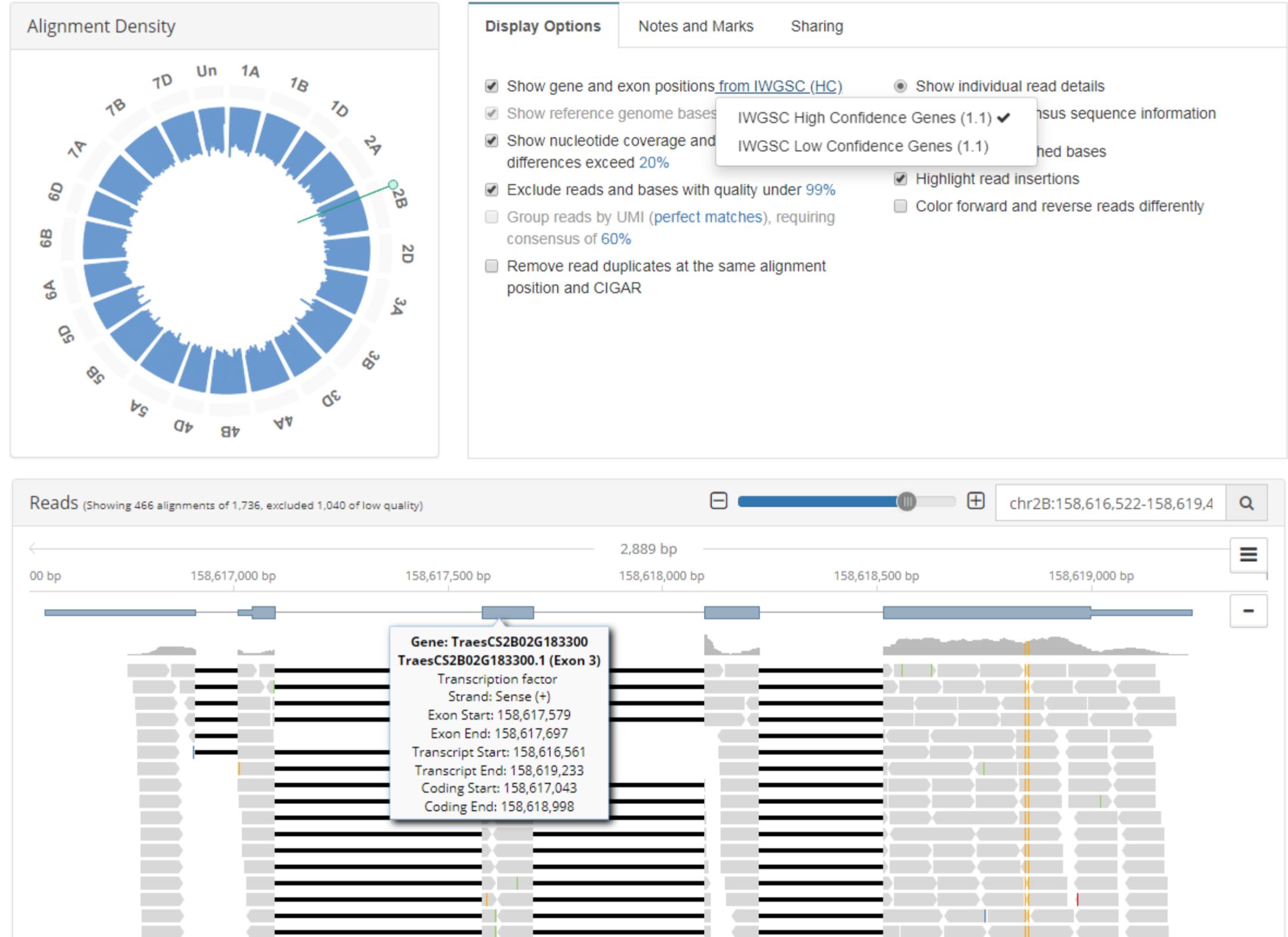
More Options

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



# Bread Wheat RNA-Seq: Expression Analysis

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly

Dashboard / All Projects / Project Details

Project Details More Options ▾

General Information

Project Name:

Species:

General Notes:

All Files Raw Sequences

Available Actions ▾ Delete Files

Start Multiple Expression Analyses ✕

Genome Feature Set:

Feature to Count:

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%)

<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	<a href="#">View Alignments</a> ▾
<input checked="" type="checkbox"/>	SRR1542404.1 Paired (Aligned 2019-12-13 21:40 UTC)	198,975,170	2019-12-16 13:05	<a href="#">View Alignments</a> ▾
<input checked="" type="checkbox"/>	SRR1542405.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,418,714	2019-12-16 11:02	<a href="#">View Alignments</a> ▾
<input checked="" type="checkbox"/>	SRR1542406.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,129,219	2019-12-16 07:01	<a href="#">View Alignments</a> ▾
<input checked="" type="checkbox"/>	SRR1542408.1 Paired (Aligned 2019-12-13 21:40 UTC)	165,312,410	2019-12-16 02:07	<a href="#">View Alignments</a> ▾
<input checked="" type="checkbox"/>	SRR1542409.1 Paired (Aligned 2019-12-12 19:53 UTC)	203,941,764	2019-12-14 08:25	<a href="#">View Alignments</a> ▾

- 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	<a href="#">View Alignments</a>
SRR1542404.1 Paired (Aligned 2019-12-13 21:40 UTC)	198,975,170	2019-12-16 13:05	<a href="#">View Alignments</a>
SRR1542405.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,418,714	2019-12-16 11:02	<a href="#">View Alignments</a>
SRR1542406.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,129,219	2019-12-16 07:01	<a href="#">View Alignments</a>
SRR1542408.1 Paired (Aligned 2019-12-13 21:40 UTC)	165,312,410	2019-12-16 02:07	<a href="#">View Alignments</a>
SRR1542409.1 Paired (Aligned 2019-12-12 19:53 UTC)	203,941,764	2019-12-14 08:25	<a href="#">View Alignments</a>

- 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

Start Multiple Expression Analyses

Genome Feature Set:

Feature to Count:

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

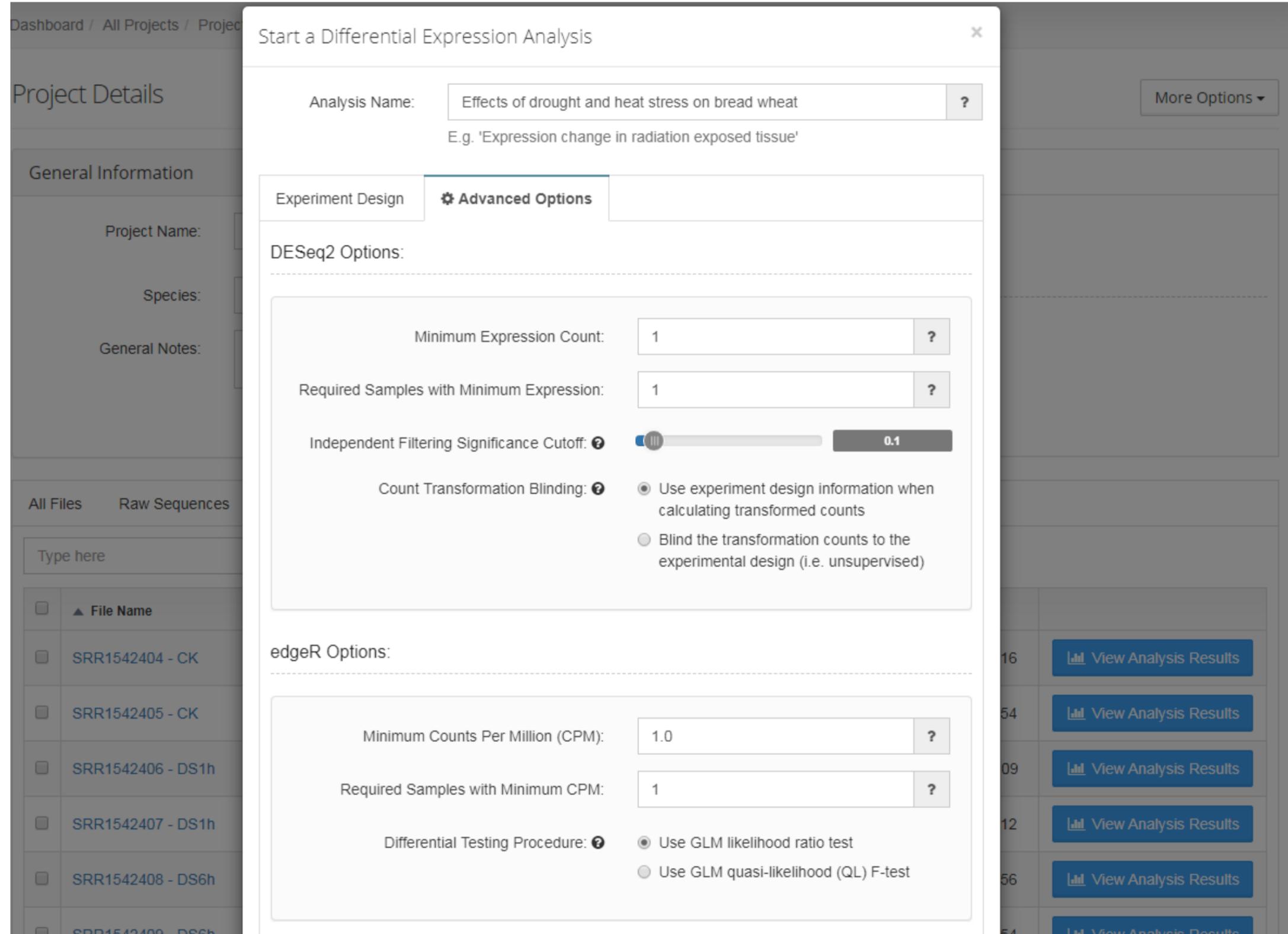
De-duplication:  Remove potential read duplicates at the same position and CIGAR

Minimum Quality:  Phred: 20 (99%)

<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	<a href="#">View Alignments</a>
<input checked="" type="checkbox"/>	SRR1542404.1 Paired (Aligned 2019-12-13 21:40 UTC)	198,975,170	2019-12-16 13:05	<a href="#">View Alignments</a>
<input checked="" type="checkbox"/>	SRR1542405.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,418,714	2019-12-16 11:02	<a href="#">View Alignments</a>
<input checked="" type="checkbox"/>	SRR1542406.1 Paired (Aligned 2019-12-13 21:40 UTC)	184,129,219	2019-12-16 07:01	<a href="#">View Alignments</a>
<input checked="" type="checkbox"/>	SRR1542408.1 Paired (Aligned 2019-12-13 21:40 UTC)	165,312,410	2019-12-16 02:07	<a href="#">View Alignments</a>
<input checked="" type="checkbox"/>	SRR1542409.1 Paired (Aligned 2019-12-12 19:53 UTC)	203,941,764	2019-12-14 08:25	<a href="#">View Alignments</a>



- 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](#)

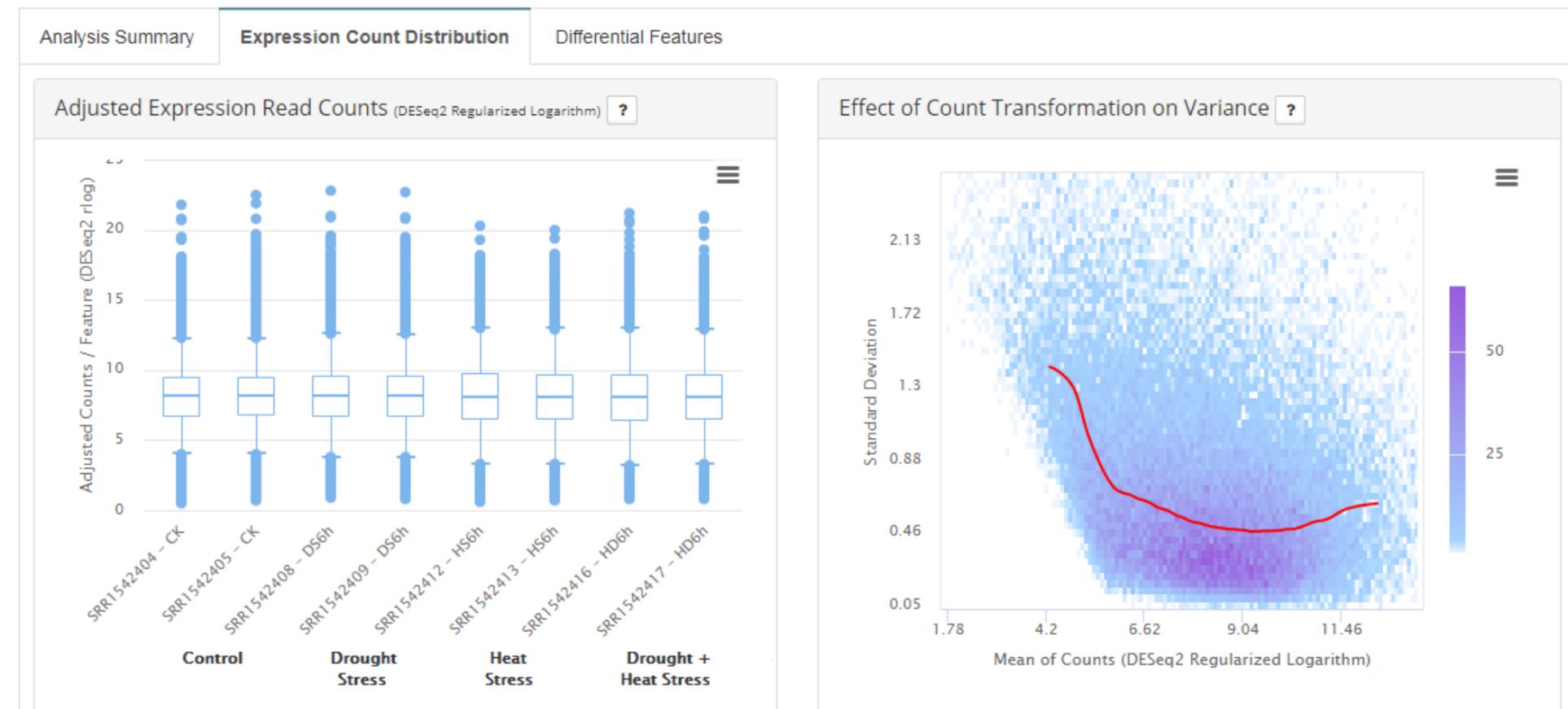
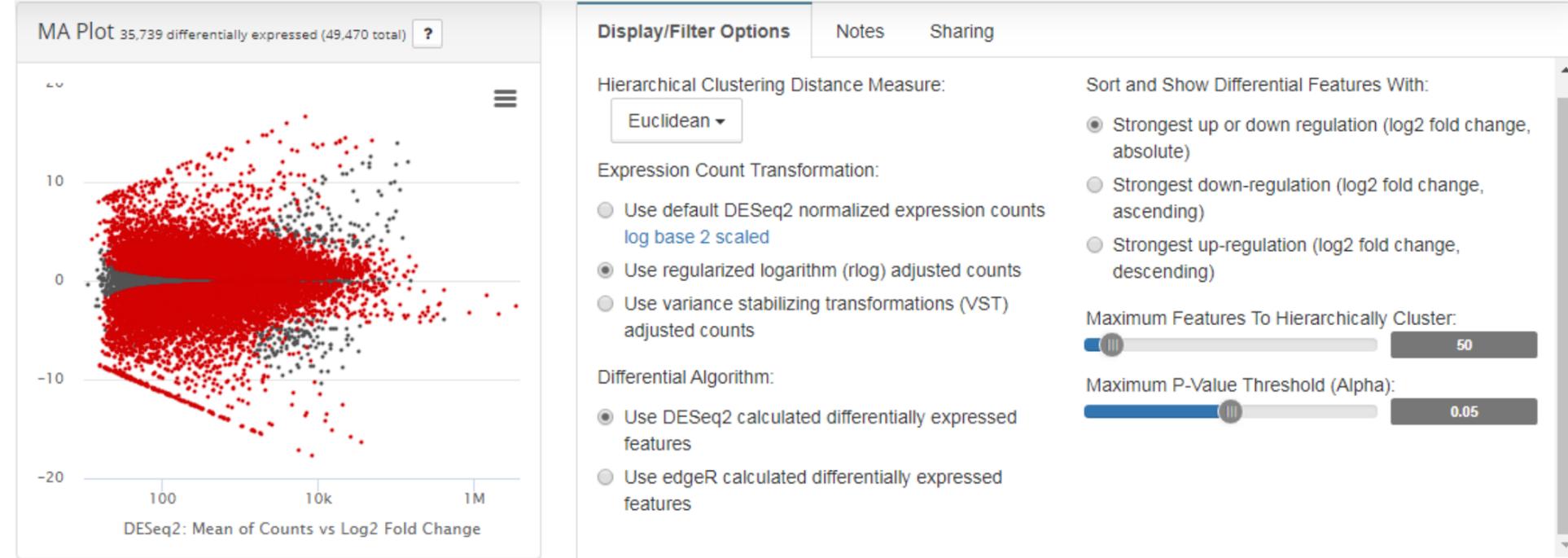
09 [View Analysis Results](#)

12 [View Analysis Results](#)

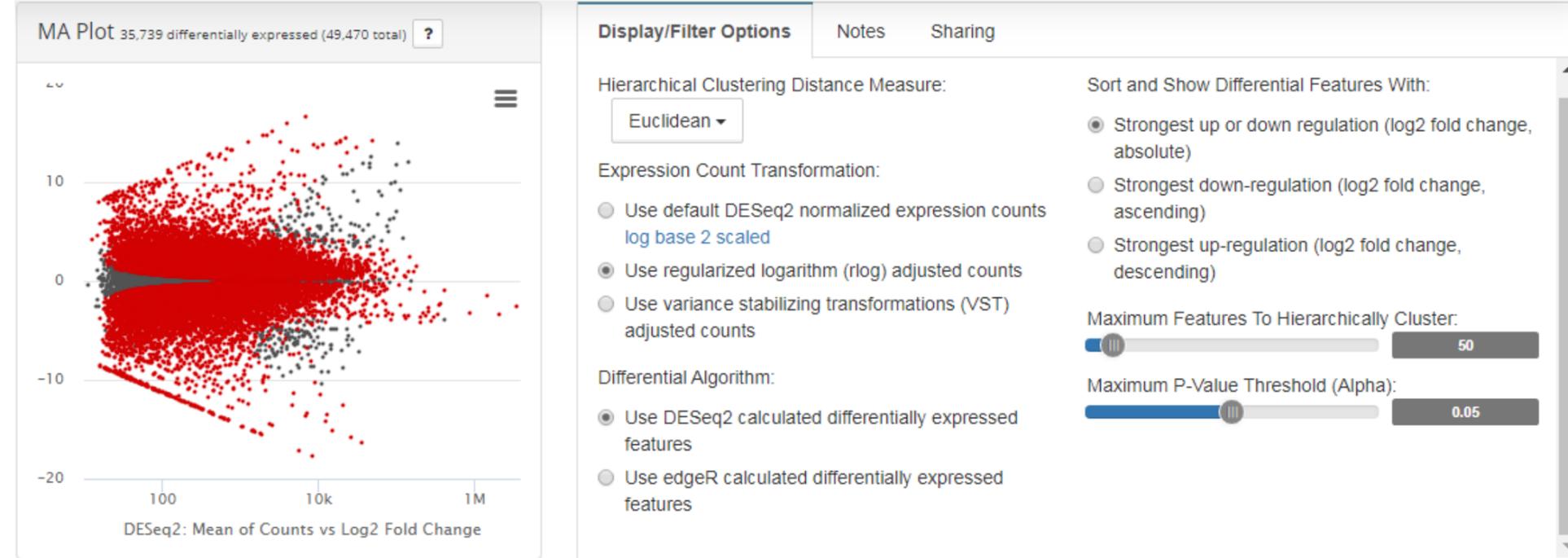
56 [View Analysis Results](#)

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



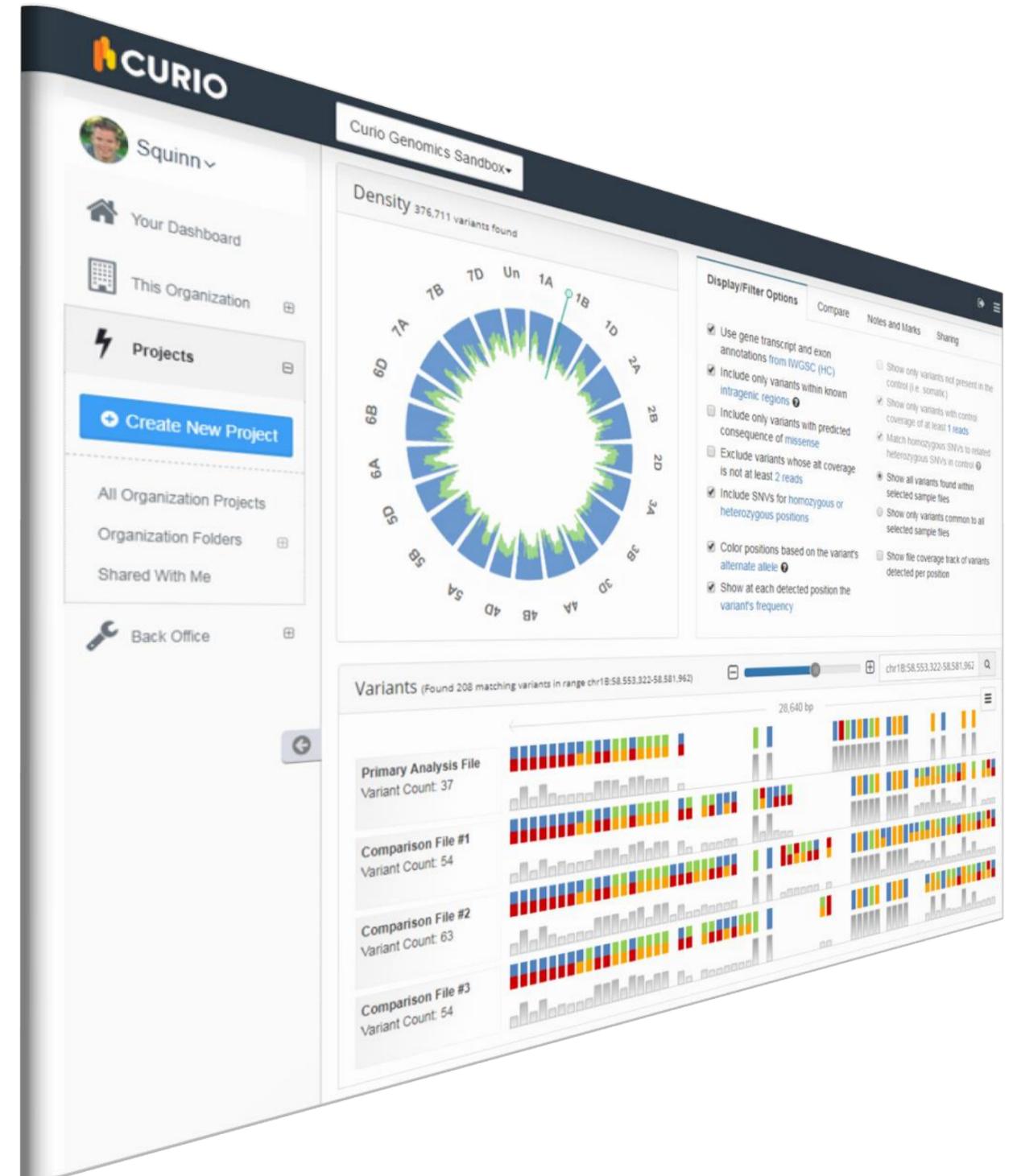
- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets
- Validate sample distances through principal component analysis
- Highlight up or down regulated genes through interactive heat maps
- Compare and contrast different industry standard algorithms in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



# 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 standard hexaploid and tetraploid wheat assemblies and 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)

## 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