
# 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)*

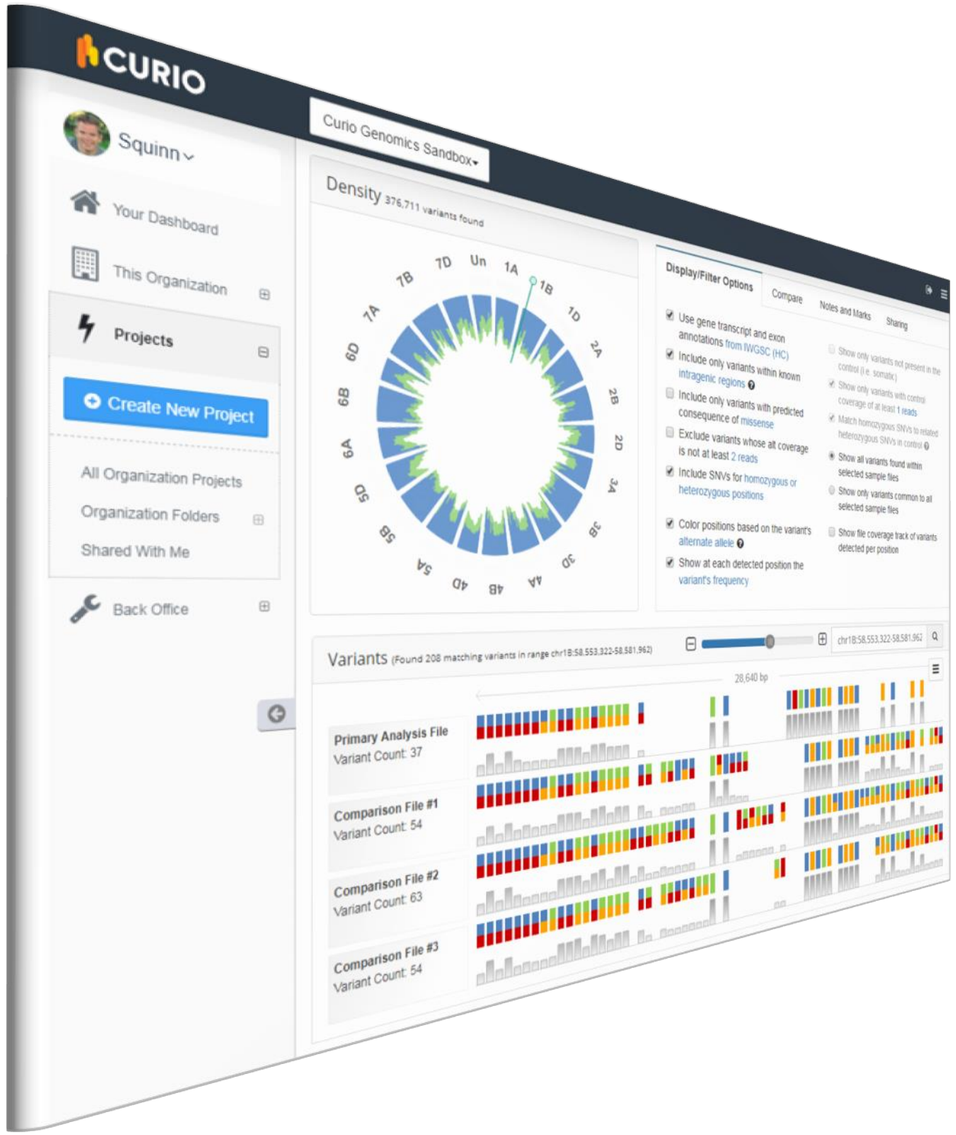
# Extending the Curio Genomics Platform for the Wheat Research Community: *DNA Sequences to Analysis Results*

# 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?

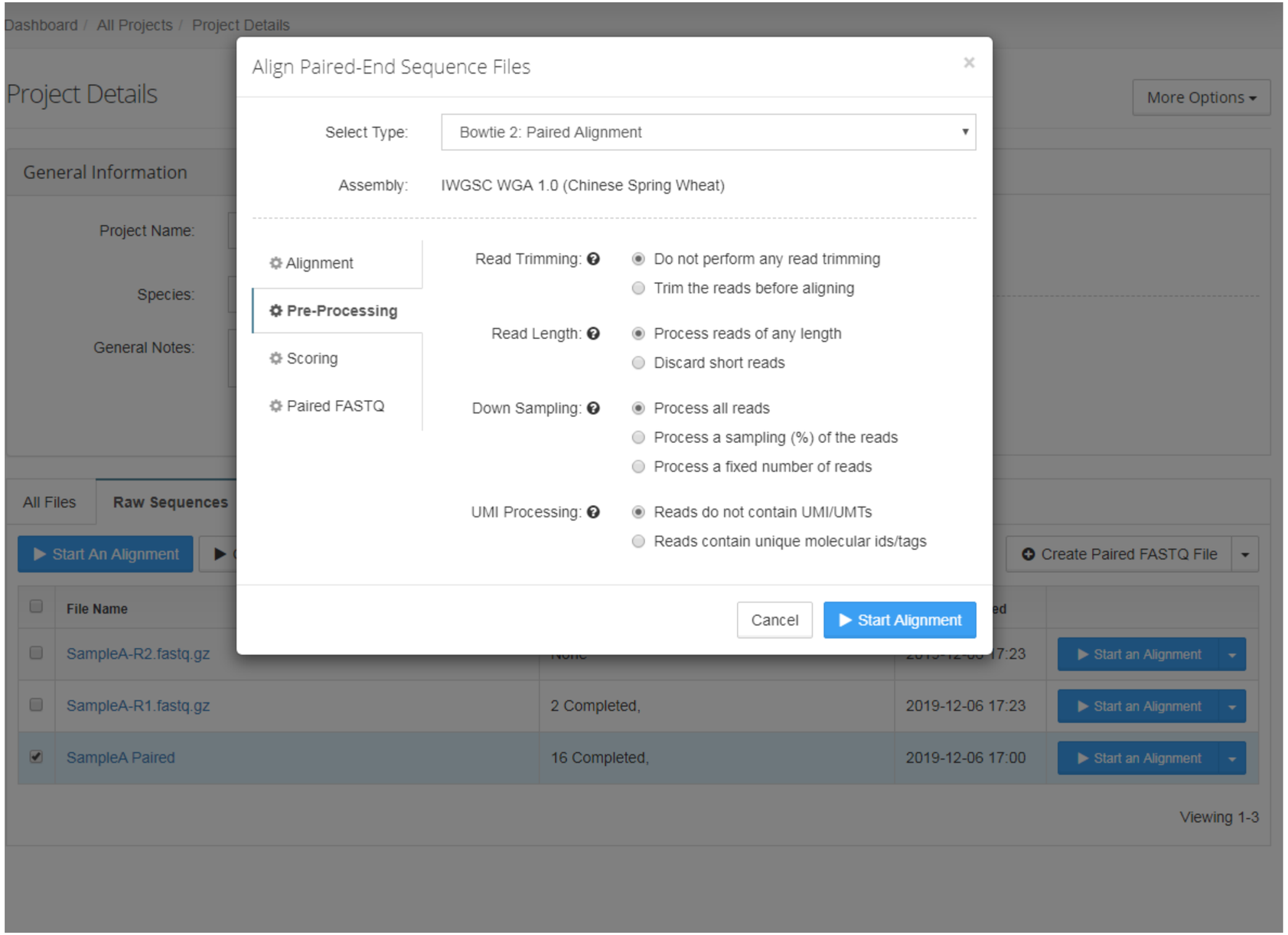
- Modern big data management and genomic analysis platform, fully web-based, collaboration ready
- Supports both bioinformatic processing and scientific interpretive analysis
- Provides scalable data processing and interactive data visualizations using real-time databases and clustering technologies
- 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

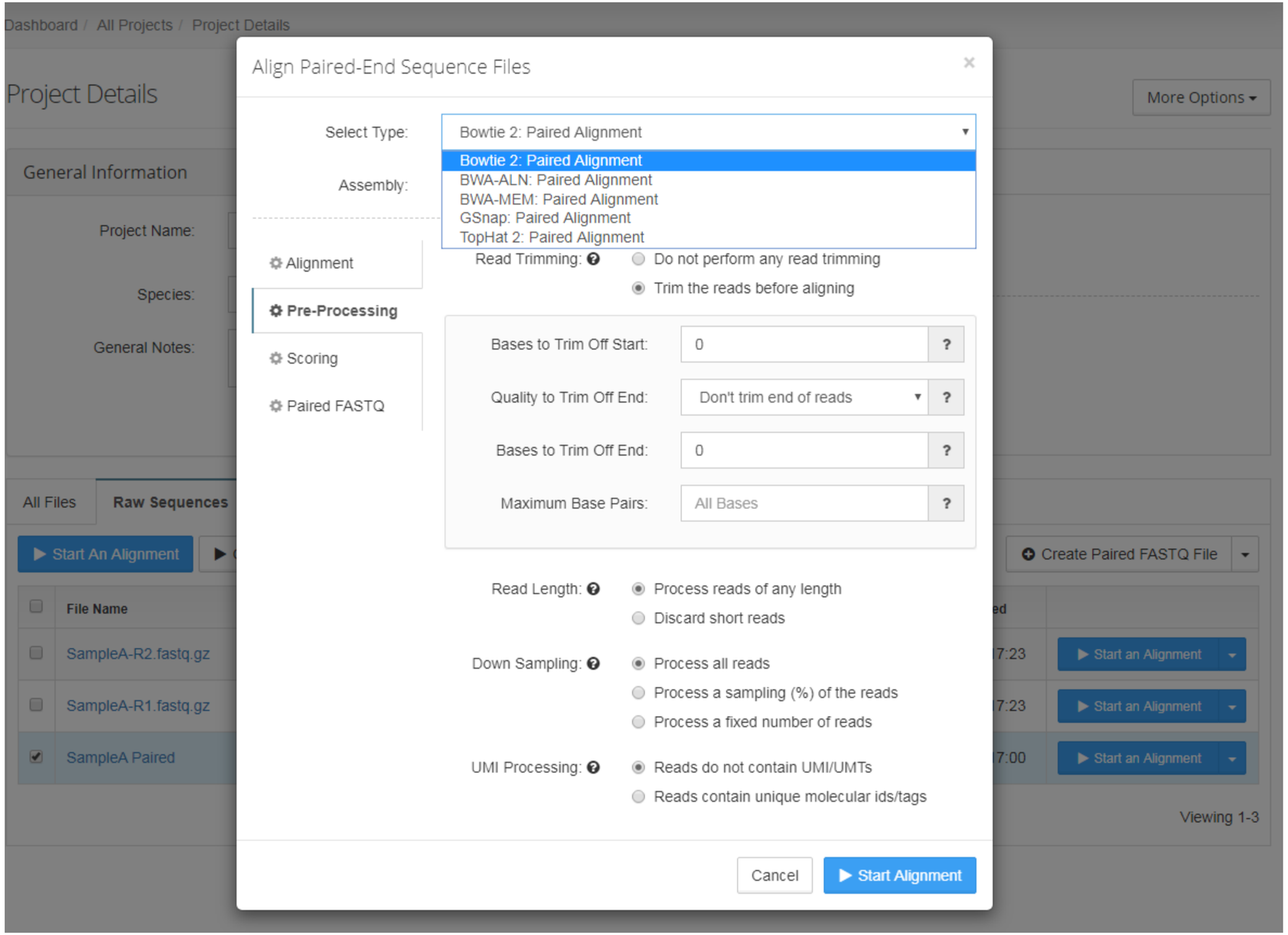
# Hexaploid Wheat: Read Mapping

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



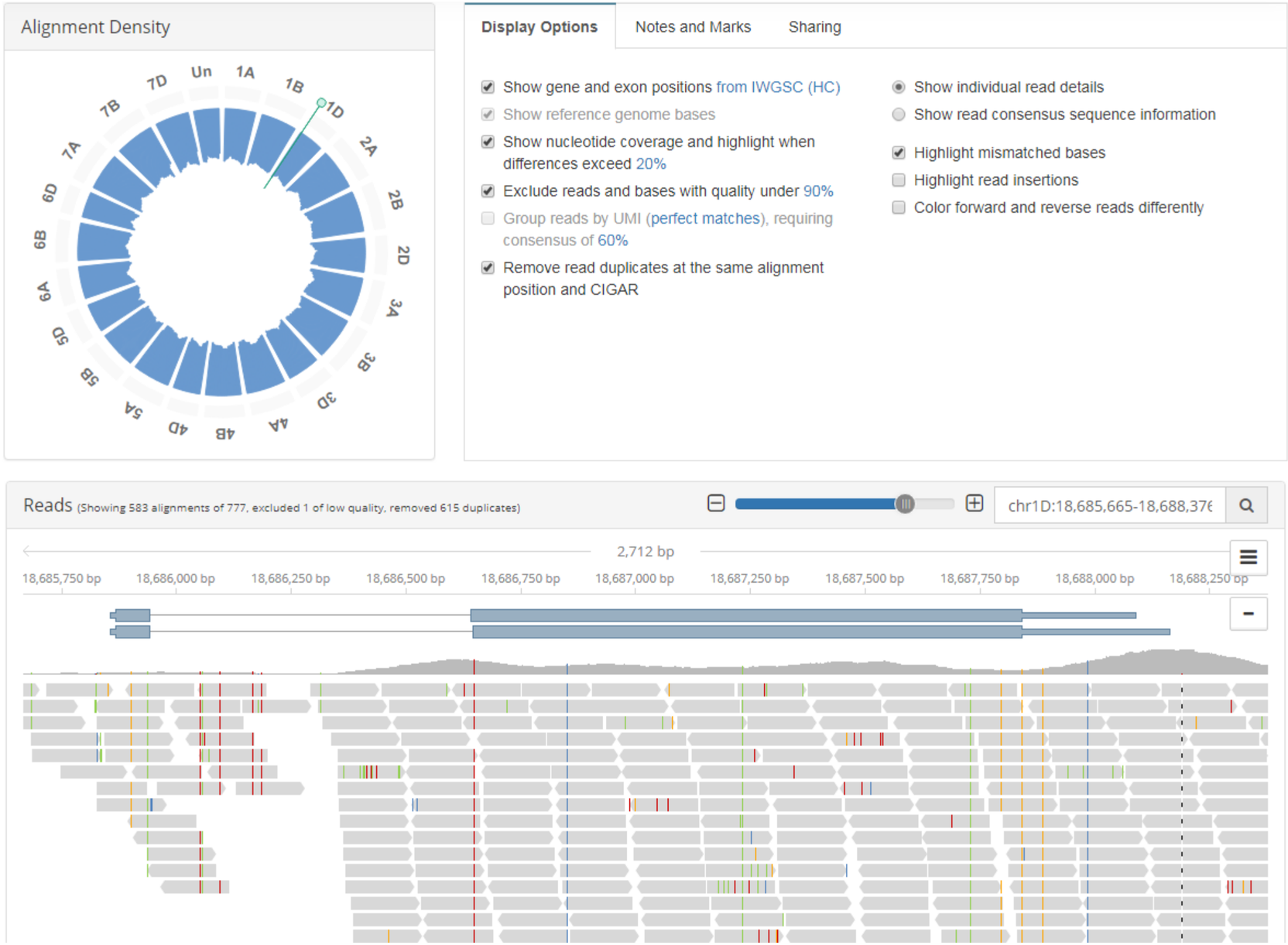
# Hexaploid Wheat: Read Mapping

- 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



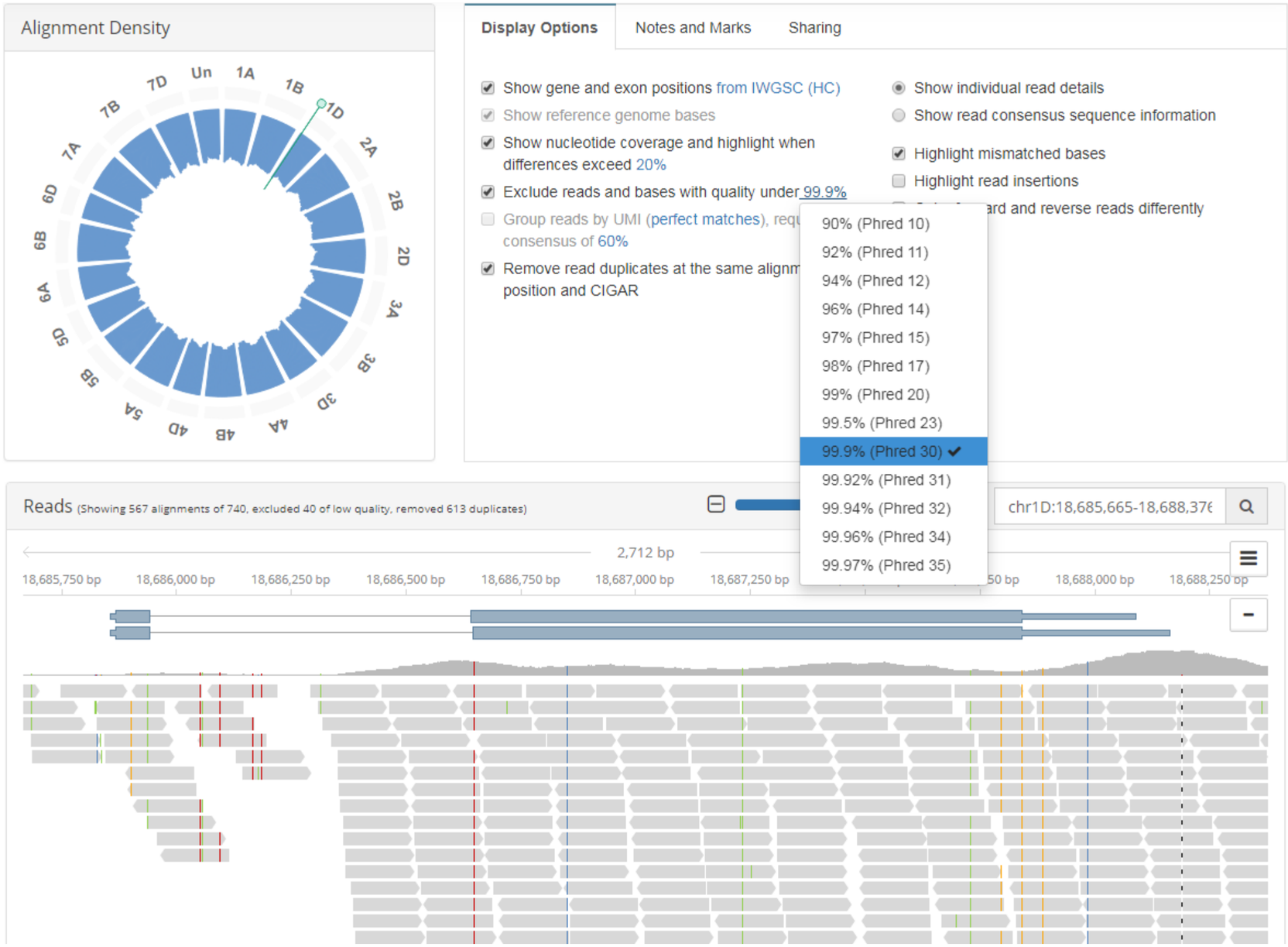
# Hexaploid Wheat: Alignment Visualization

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



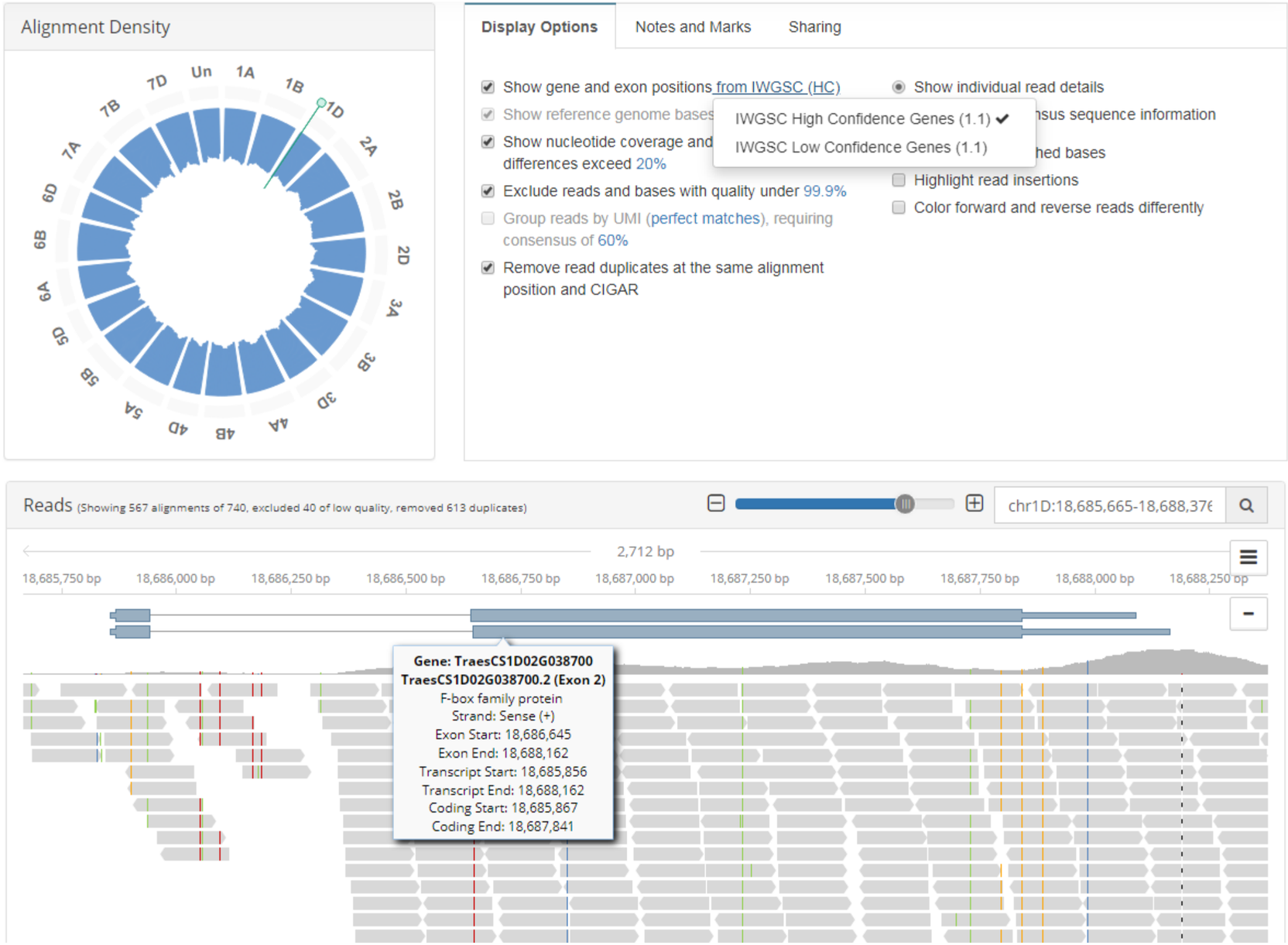
# Hexaploid Wheat: Alignment Visualization

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



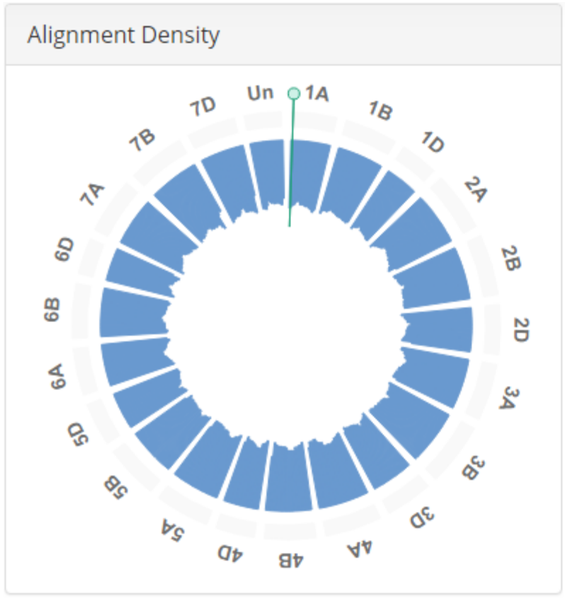
# Hexaploid Wheat: Alignment Visualization

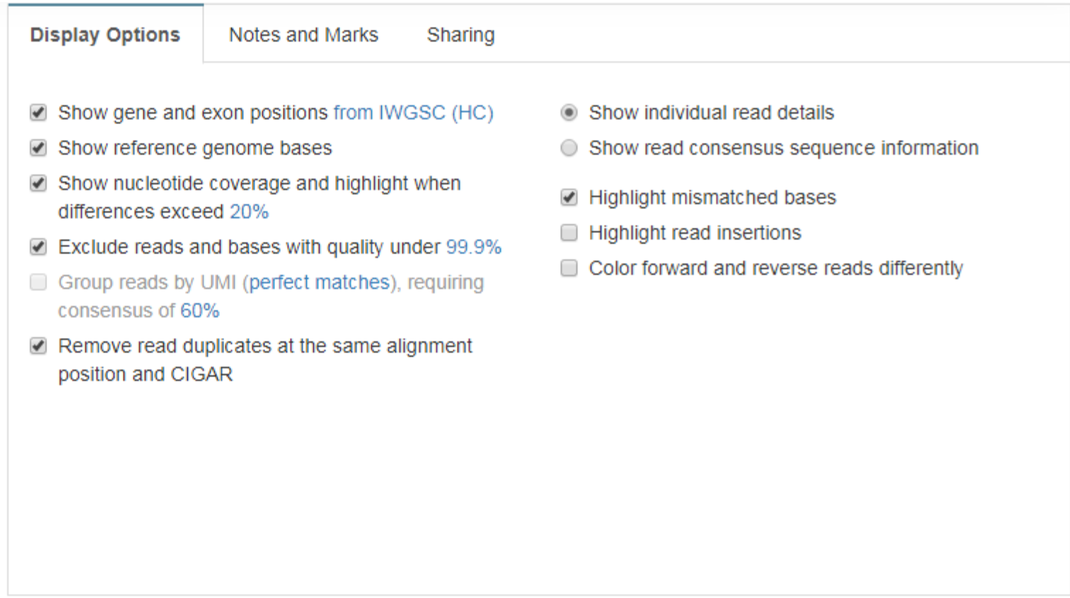
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline

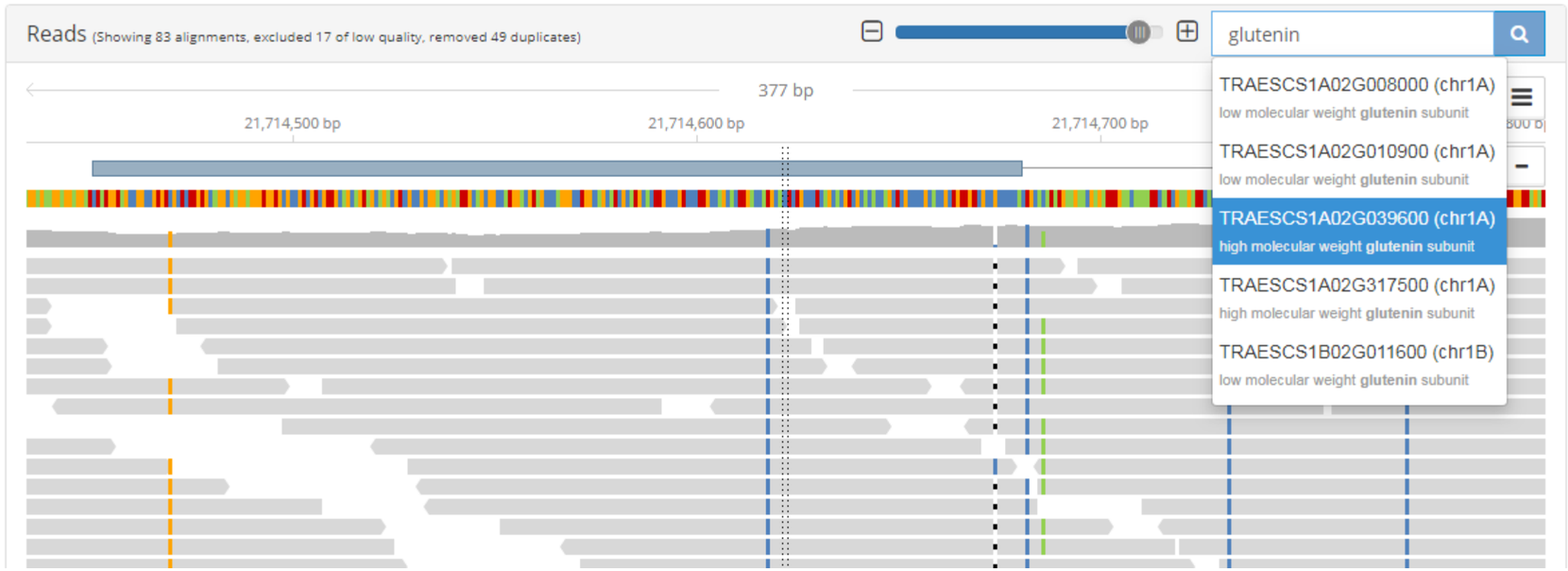


# Hexaploid Wheat: Alignment Visualization

- 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

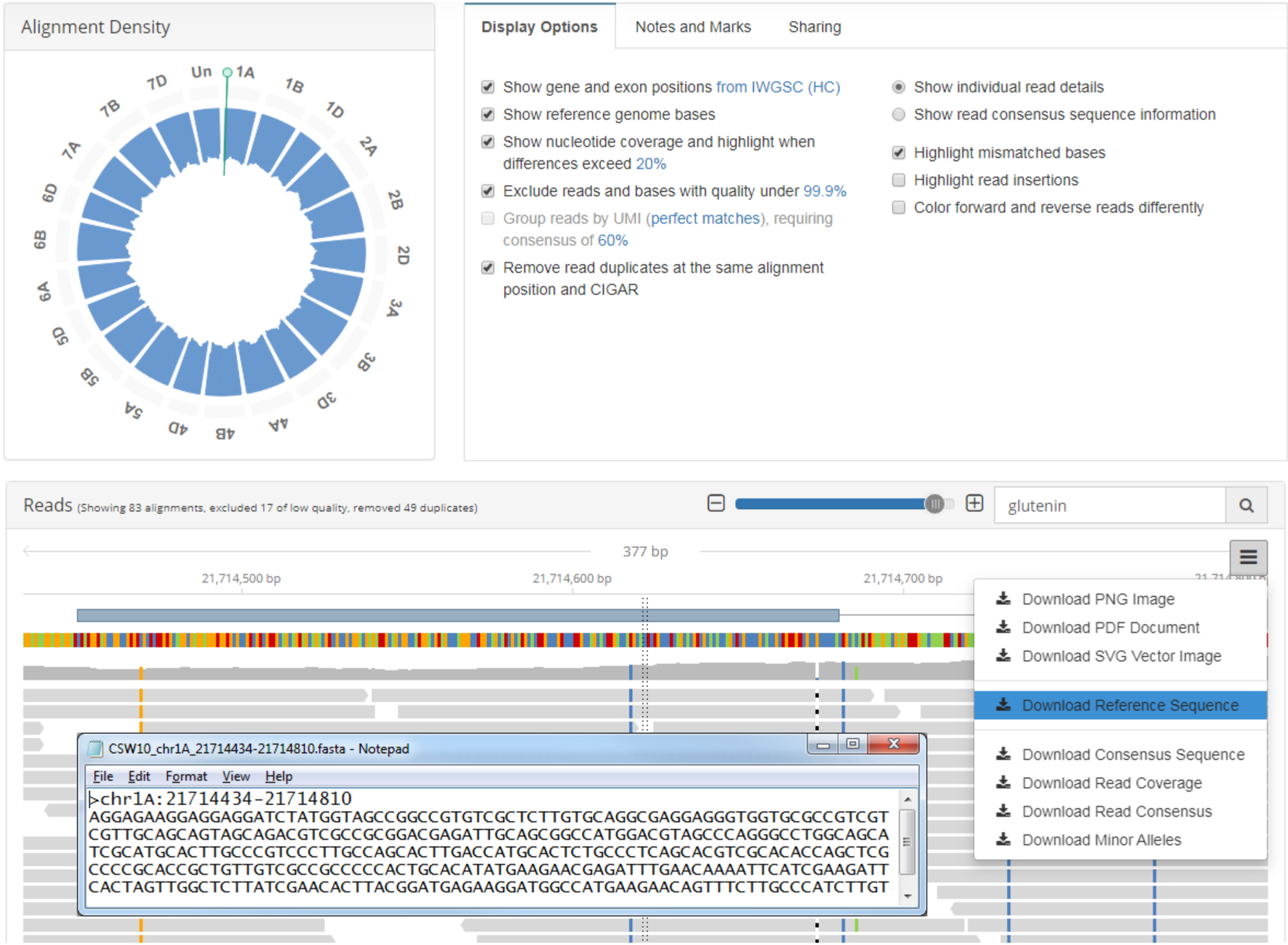


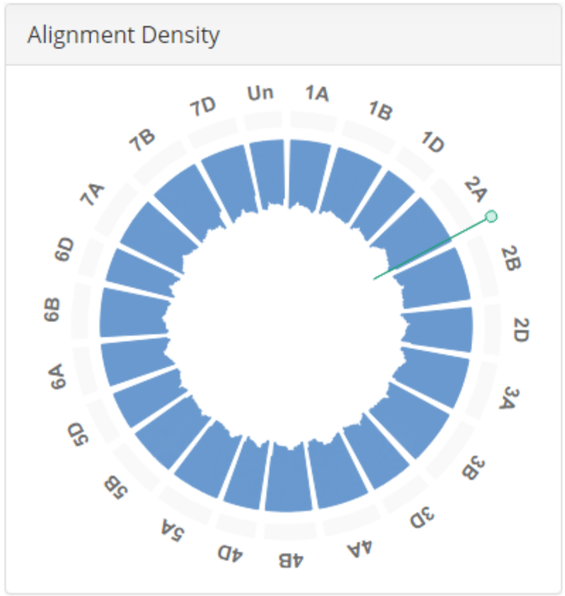
# Hexaploid Wheat: Alignment Visualization

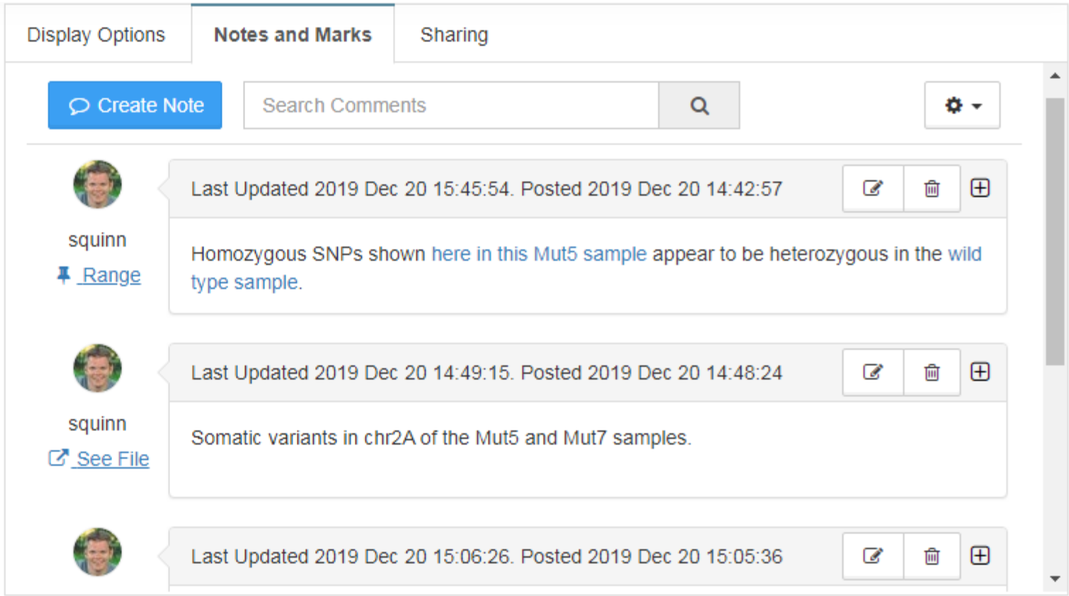
- Quickly browse and visualize reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and functional annotations from the IWGSC inline
- Quickly search and navigate the genome using the IWGSC annotation data as well
- Conveniently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

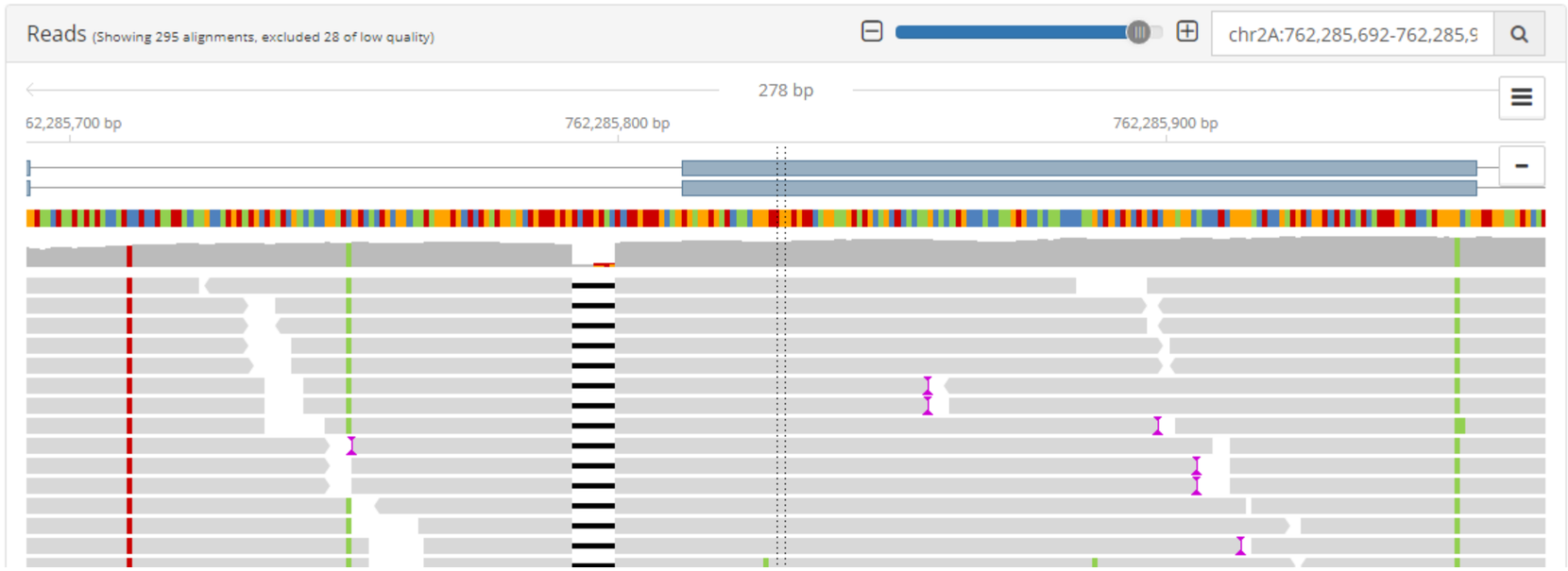


# Hexaploid Wheat: Alignment Visualization

- 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



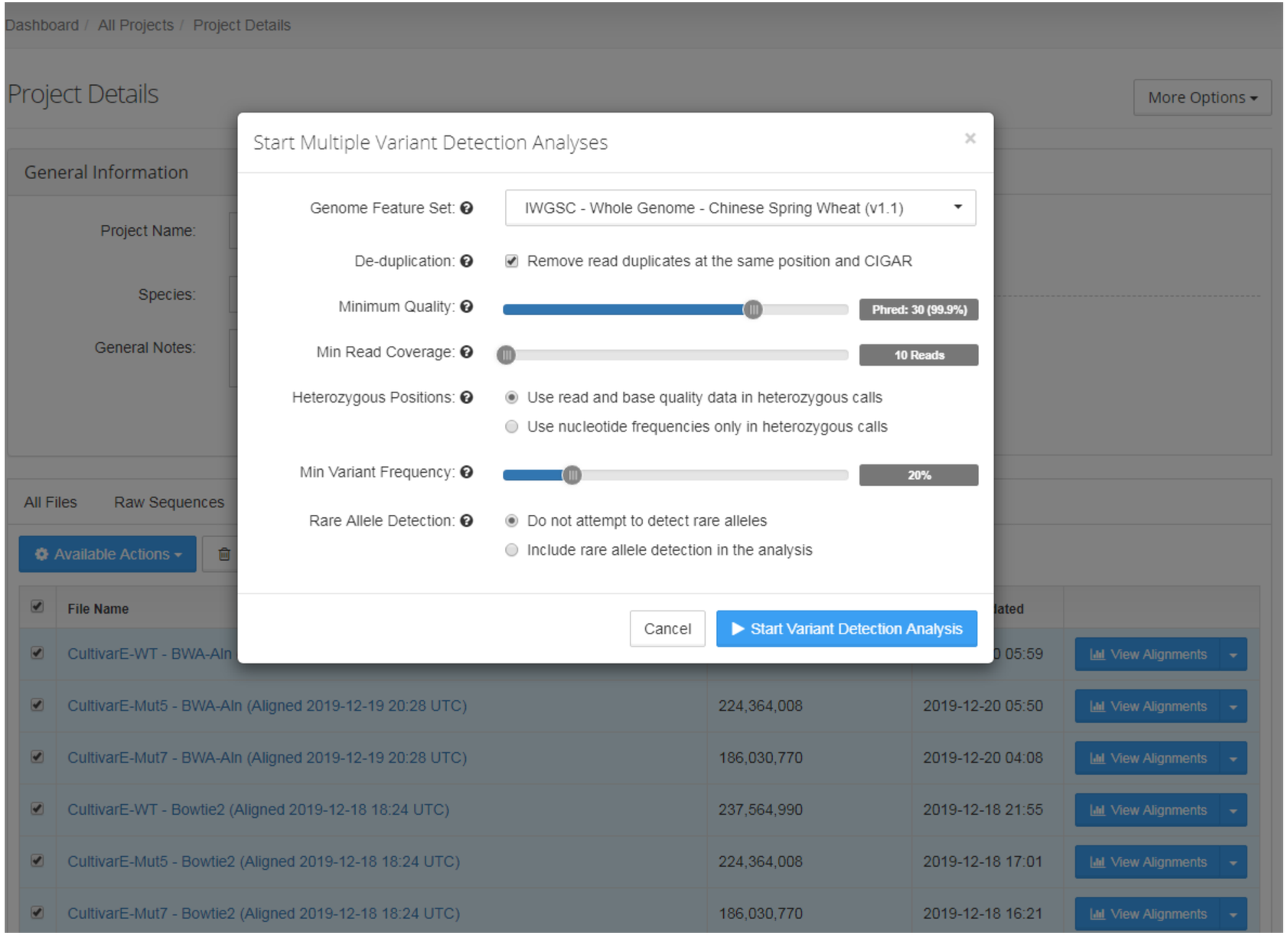



# DNA-Seq Hexaploid Wheat: Variant Analysis

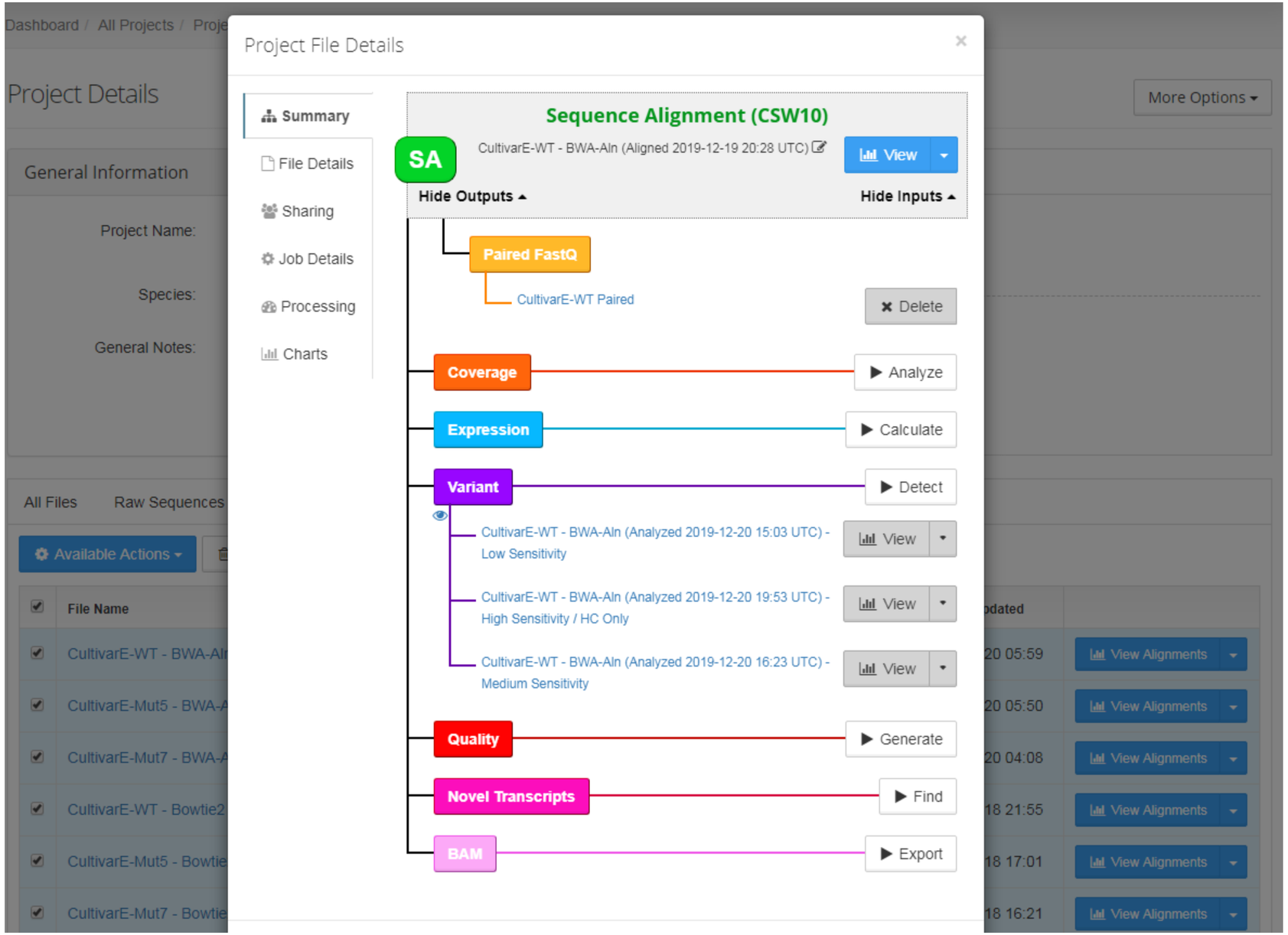
# Hexaploid Wheat: Variant Analysis

- 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



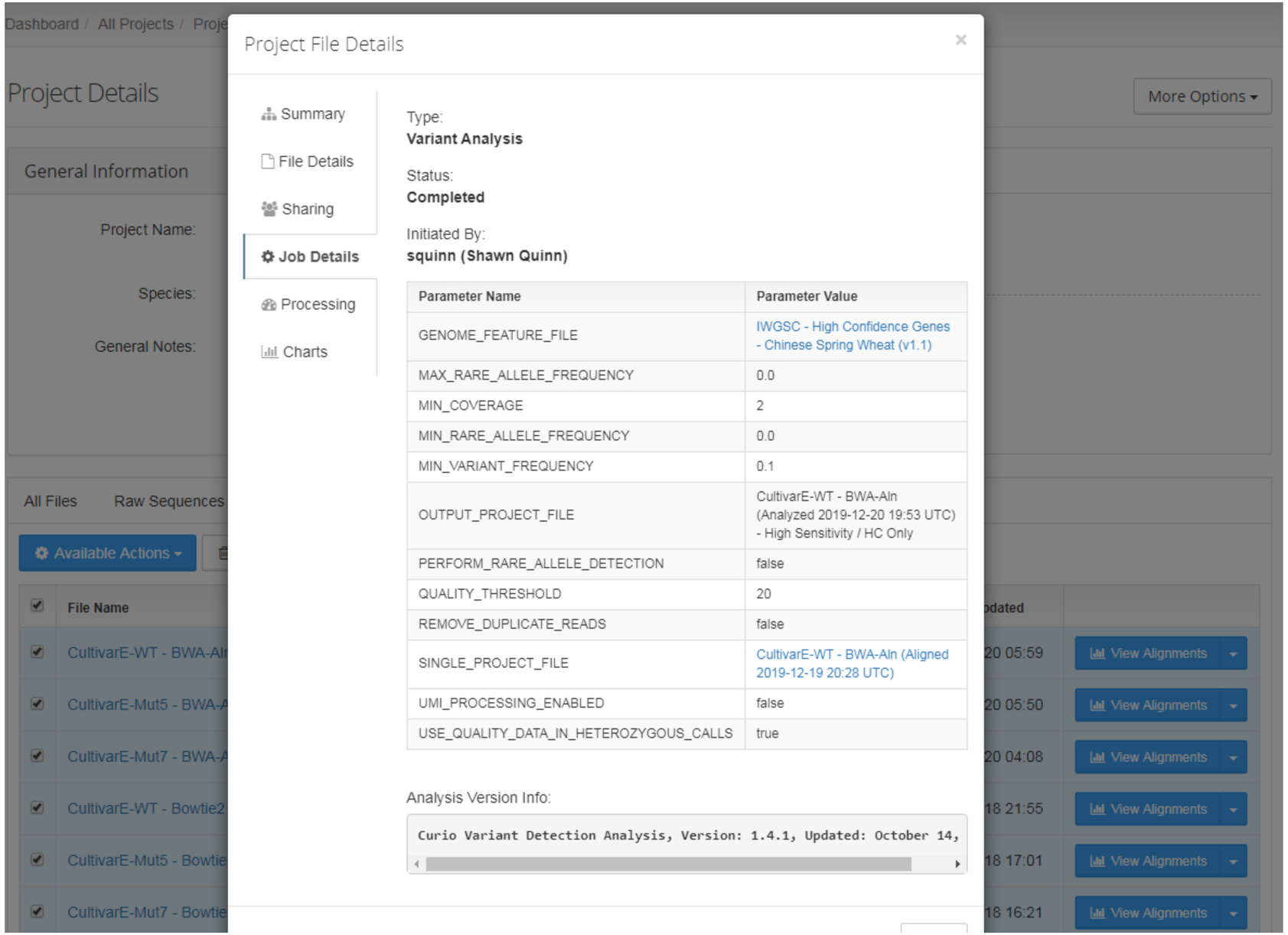
# Hexaploid Wheat: Variant Analysis

- 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
- Automatically track the provenance of every file…



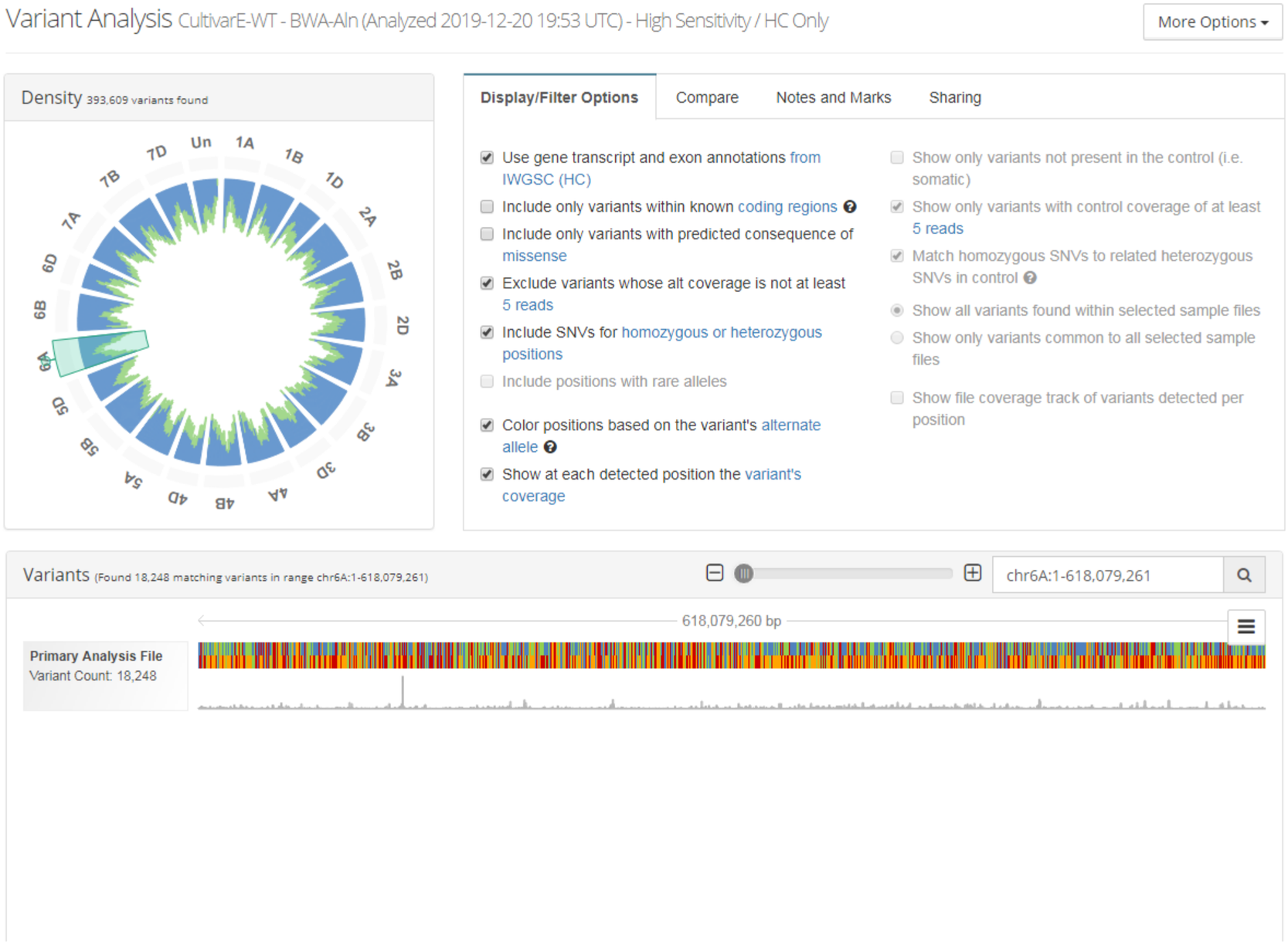
# Hexaploid Wheat: Variant Analysis

- 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
- Automatically track the provenance of every file…
- …including the specific options that went into any analysis



# Hexaploid Wheat: Variant Analysis

- Browse and visualize variants across the entire CSW genome

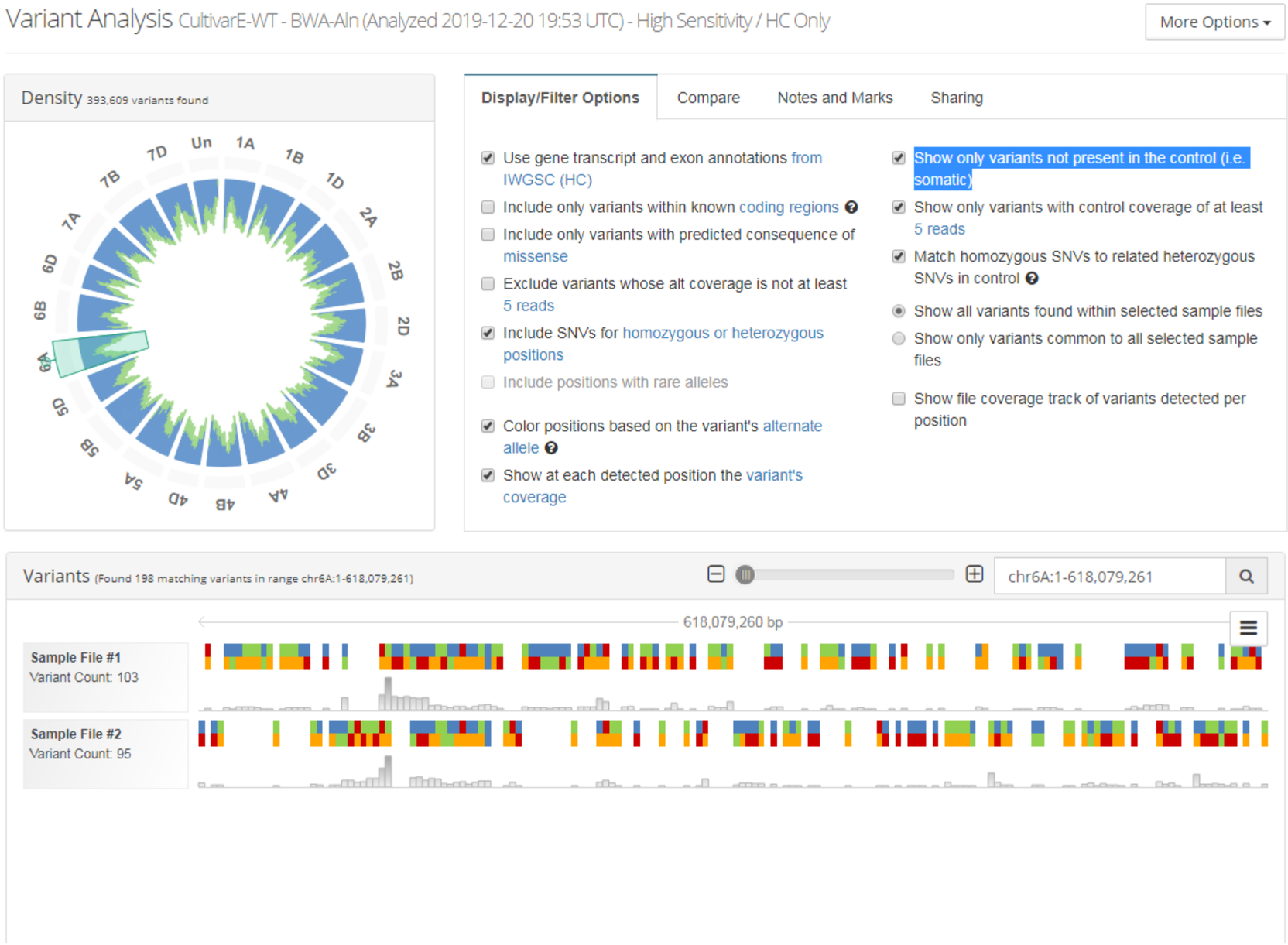


# Hexaploid Wheat: Variant Analysis

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

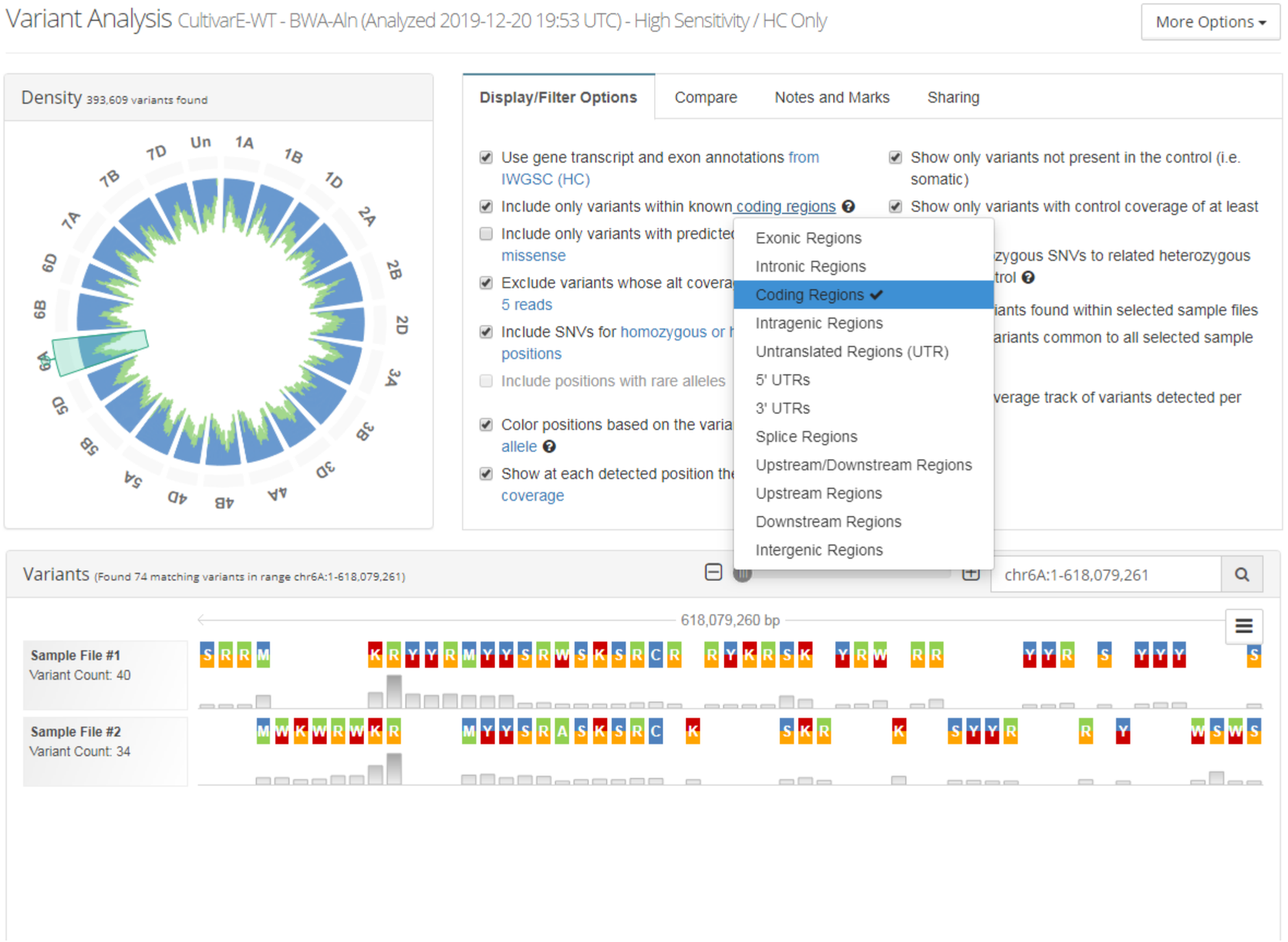
# Hexaploid Wheat: Variant Analysis

- 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



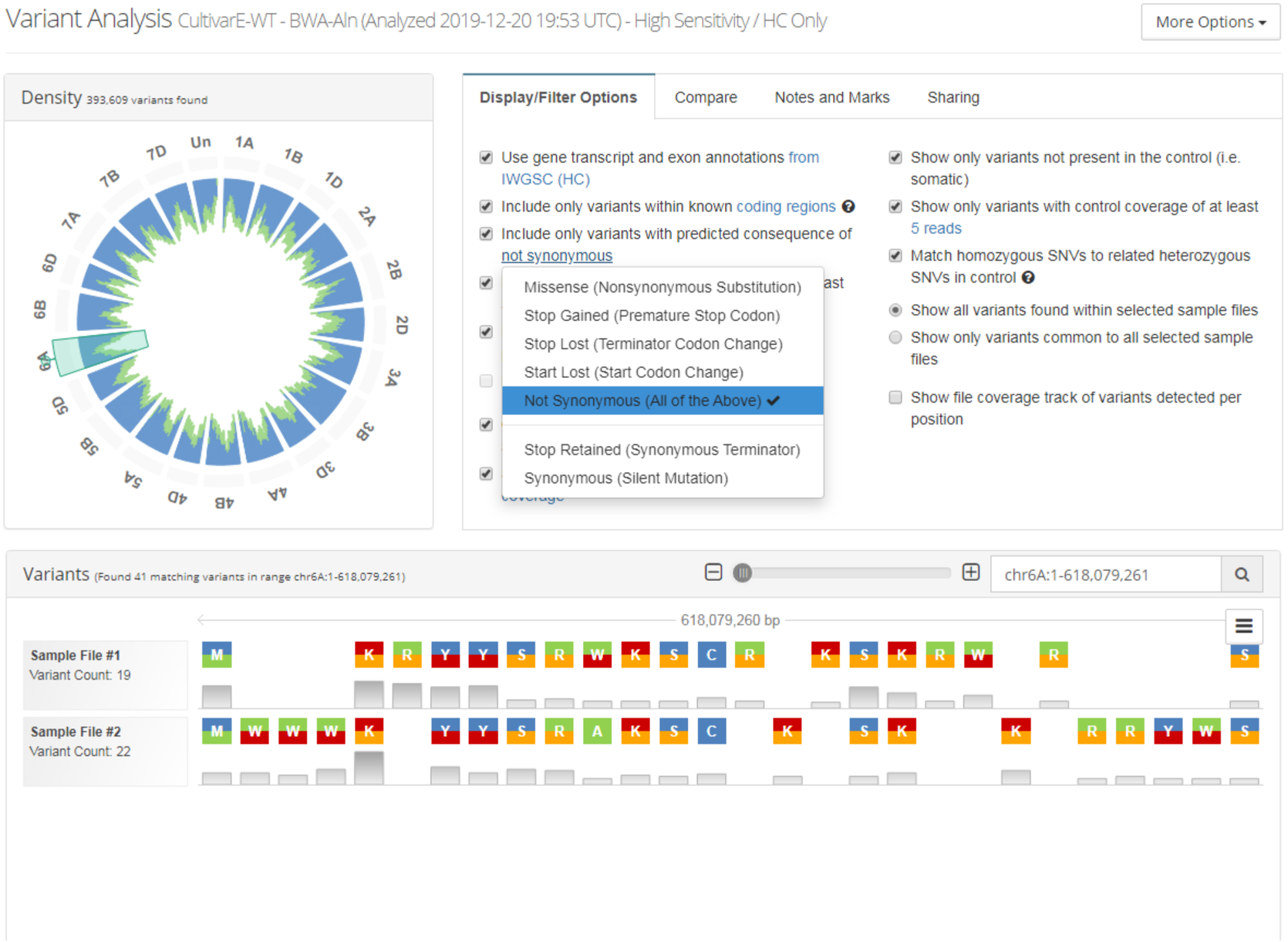
# Hexaploid Wheat: Variant Analysis

- 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



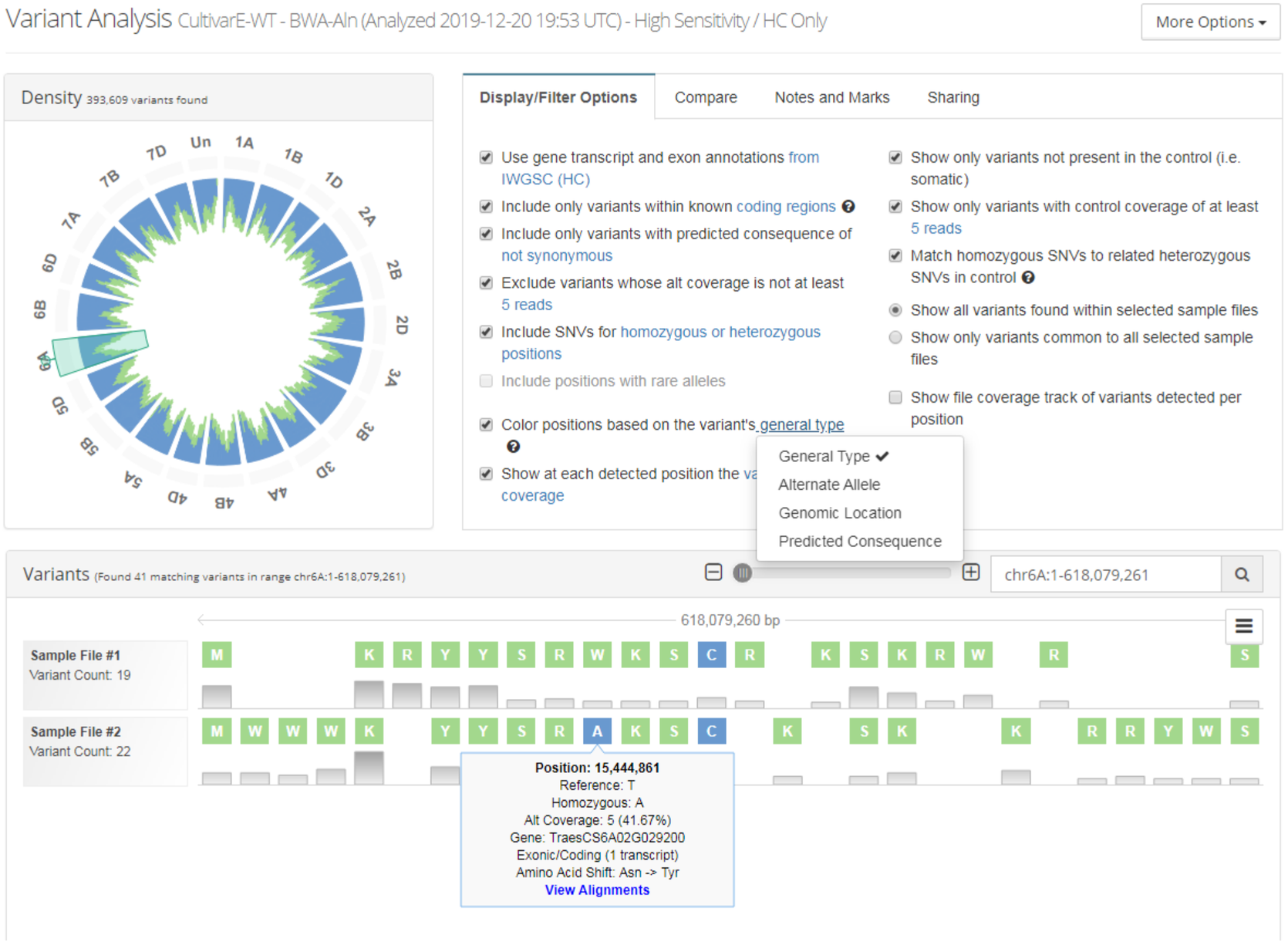
# Hexaploid Wheat: Variant Analysis

- 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



# Hexaploid Wheat: Variant Analysis

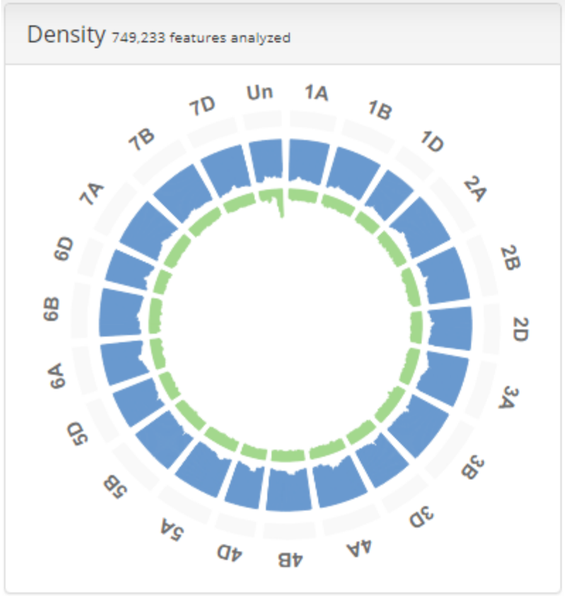
- Browse and visualize variants across the entire CSW genome
- Simultaneously compare and visualize multiple samples
- Use one or more samples as a control to exclude variants present in the wild type
- Focus on genomic regions of interest based on IWGSC gene transcript annotations
- Access predicted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



# DNA-Seq Hexaploid Wheat: Coverage Analysis

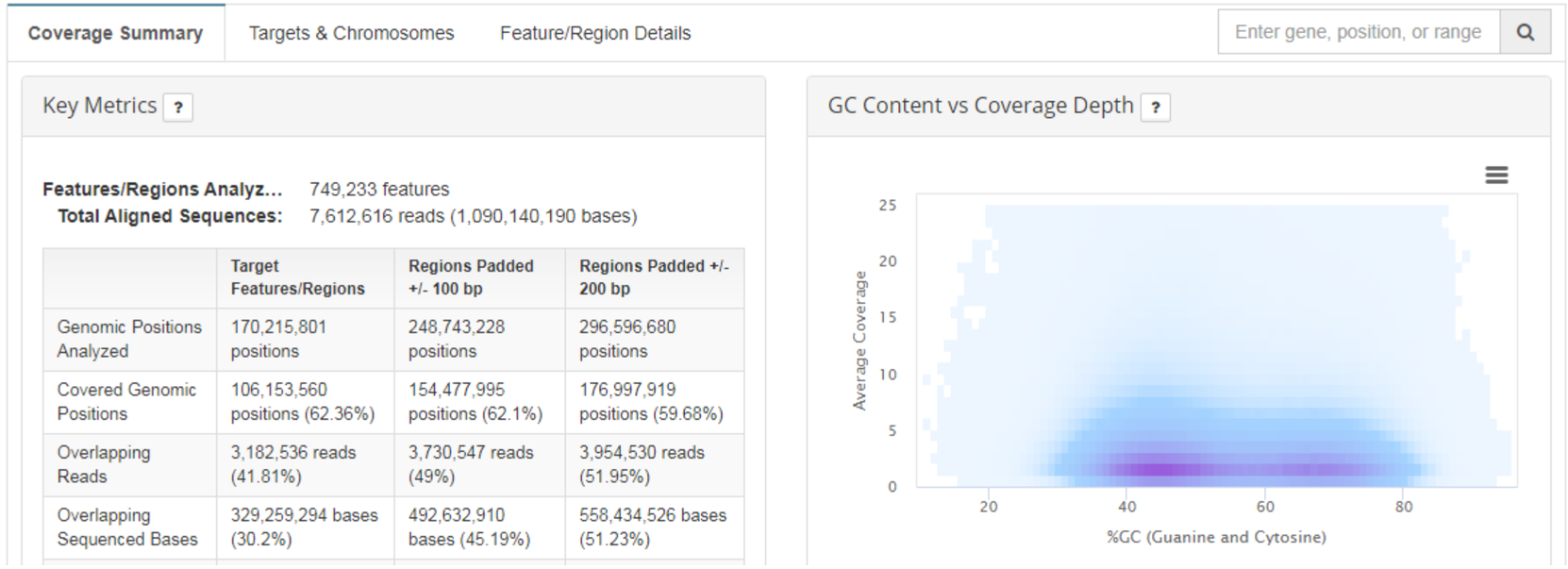
# Hexaploid Wheat: Coverage Analysis

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





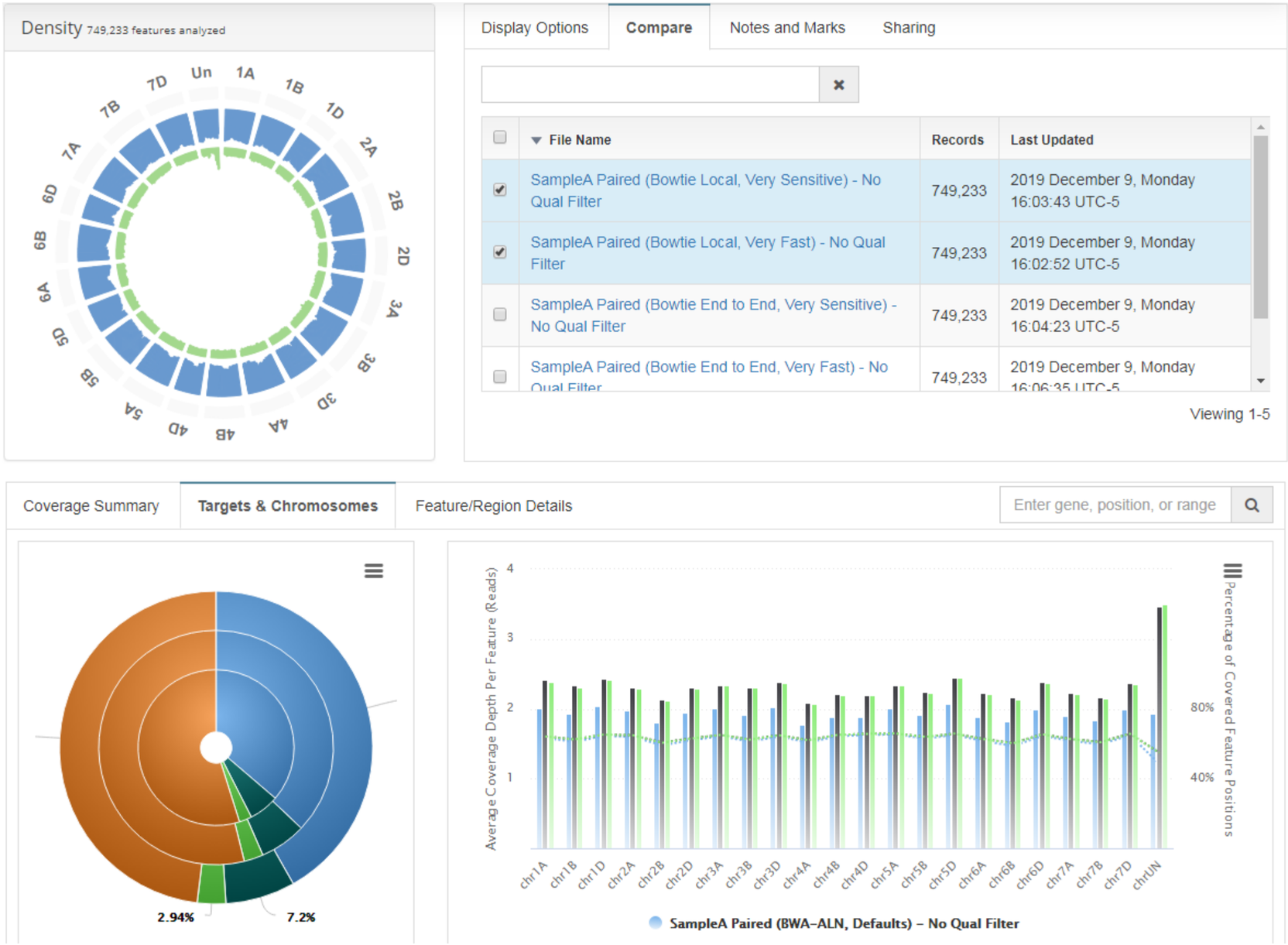





| | Target Features/Regions | Regions Padded +/- 100 bp | Regions Padded +/- 200 bp |
|---|---|---|---|
| Genomic Positions Analyzed | 170,215,801 positions | 248,743,228 positions | 296,596,680 positions |
| Covered Genomic Positions | 106,153,560 positions (62.36%) | 154,477,995 positions (62.1%) | 176,997,919 positions (59.68%) |
| Overlapping Reads | 3,182,536 reads (41.81%) | 3,730,547 reads (49%) | 3,954,530 reads (51.95%) |
| Overlapping Sequenced Bases | 329,259,294 bases (30.2%) | 492,632,910 bases (45.19%) | 558,434,526 bases (51.23%) |

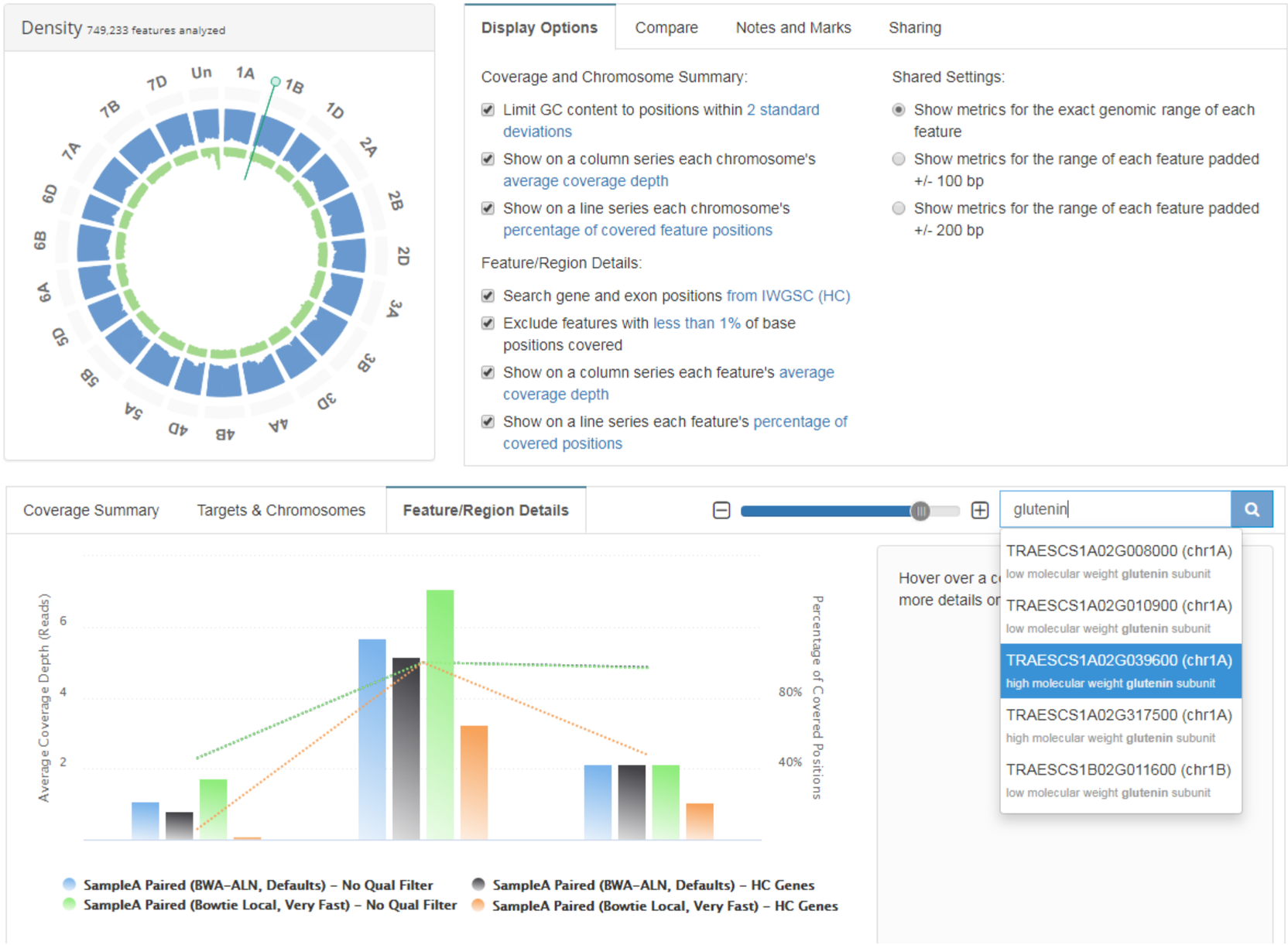
# Hexaploid Wheat: Coverage Analysis

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



# Hexaploid Wheat: Coverage Analysis

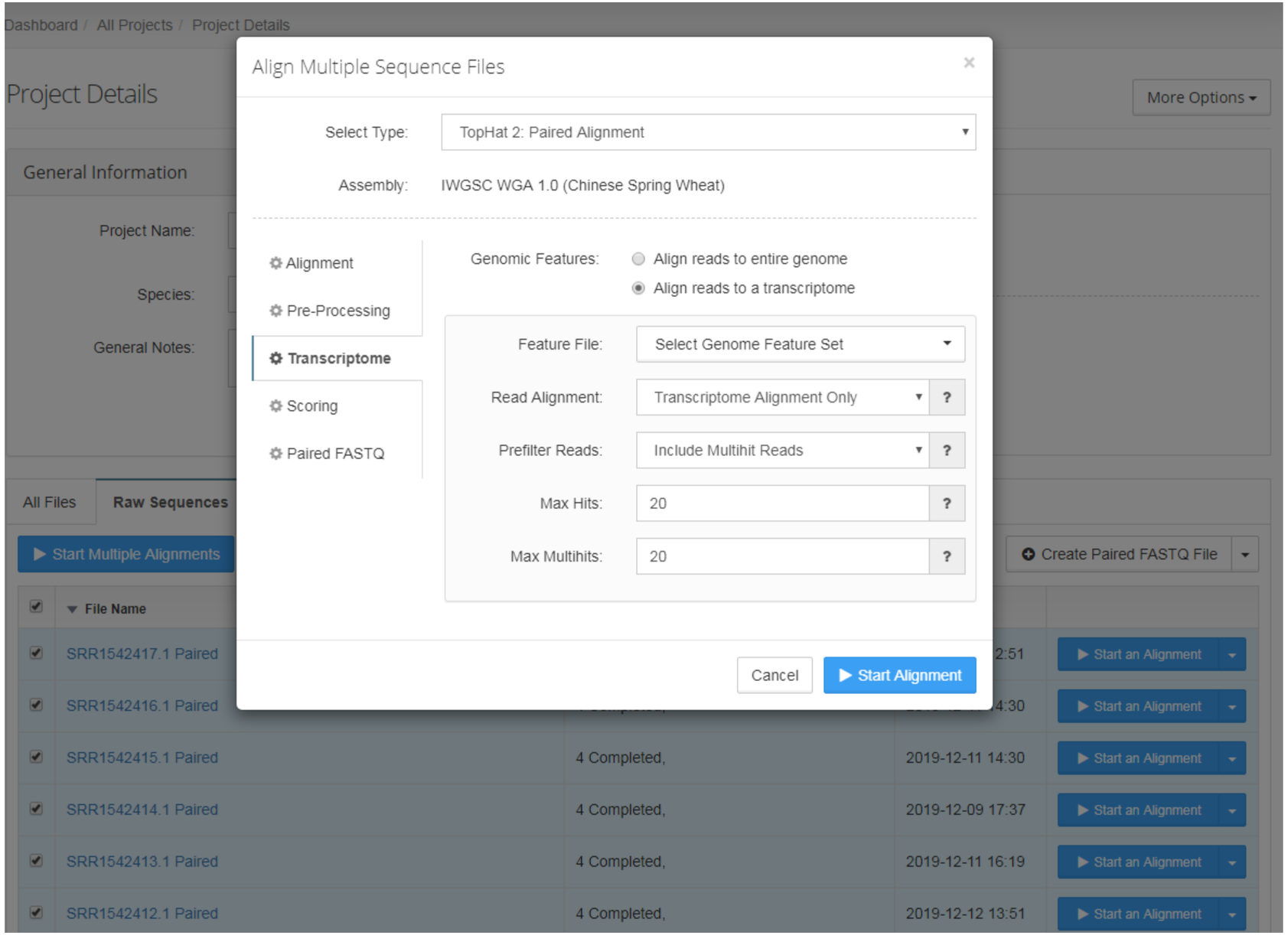
- Measure library / kit exome or custom target area key coverage metrics
- Compare and visualize multiple samples simultaneously
- Access individual gene or transcript coverage metrics leveraging IWGSC annotation sets



# RNA-Seq Hexaploid Wheat : Read Mapping and Visualization

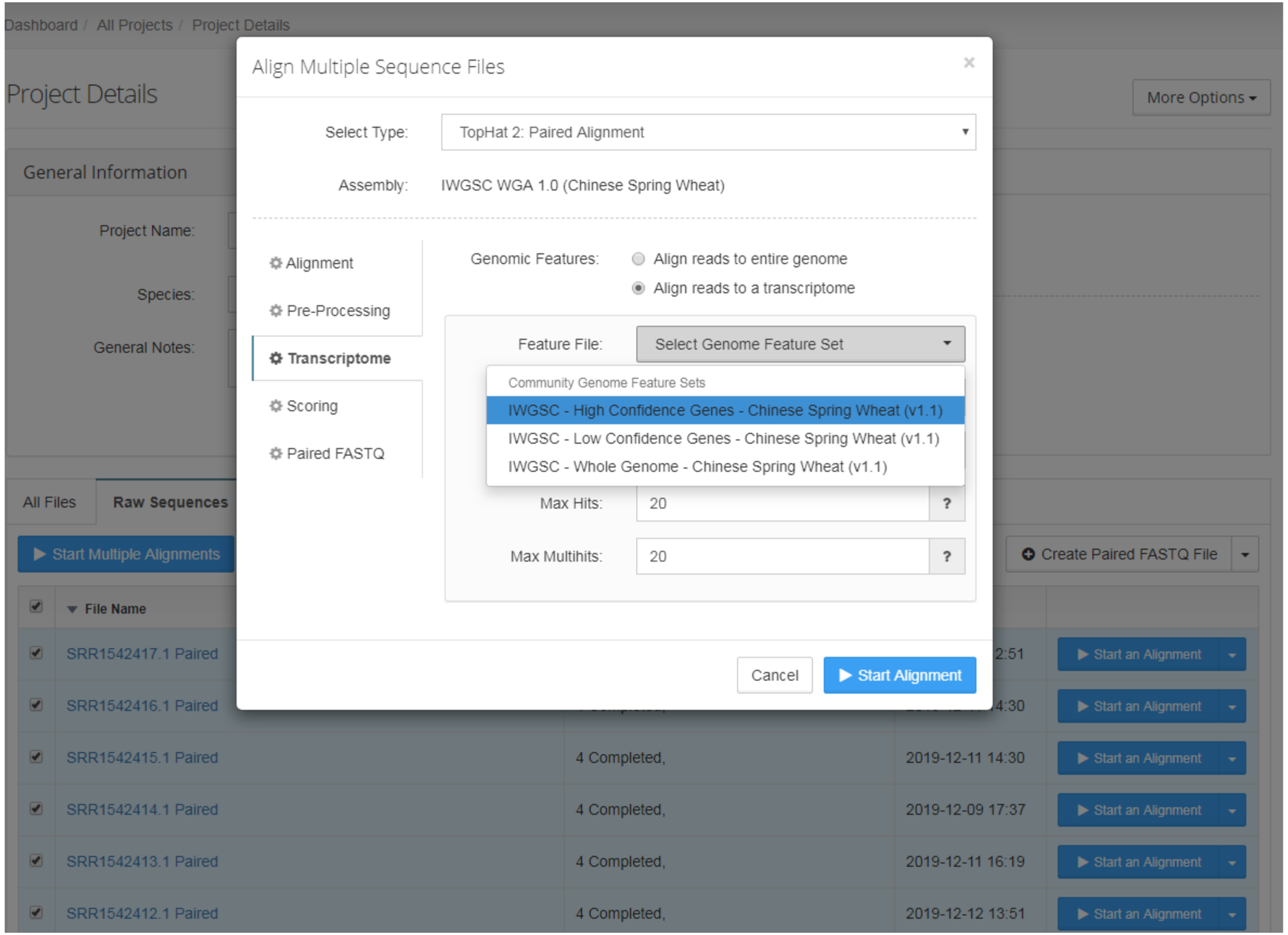
# Bread Wheat: RNA-Seq Read Mapping

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



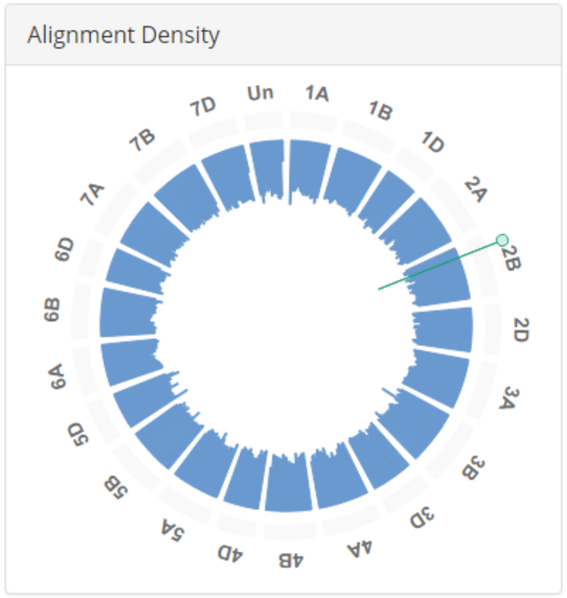
# Bread Wheat: RNA-Seq Read Mapping

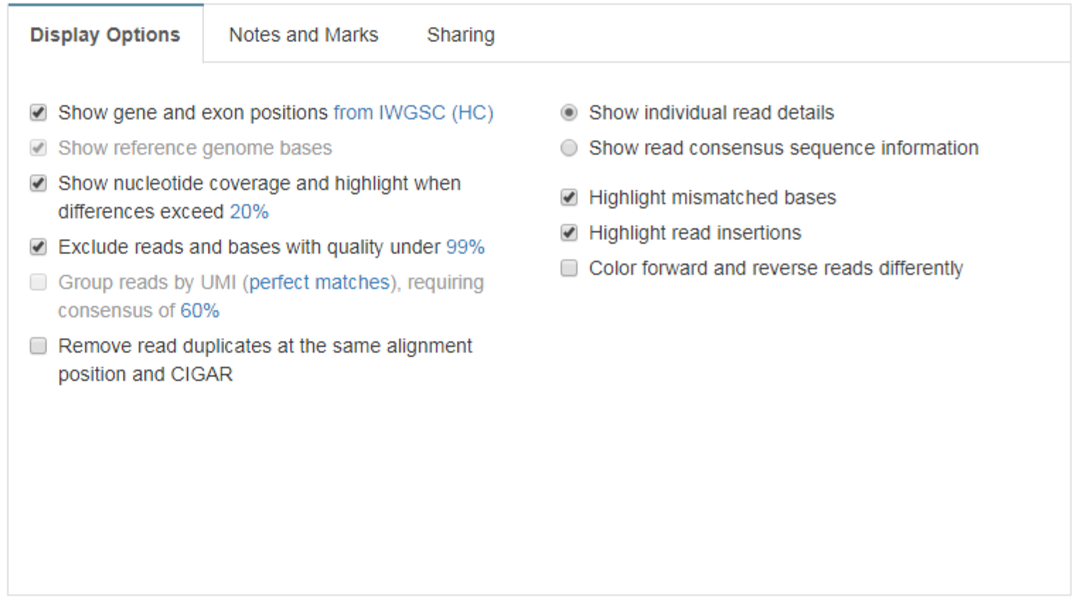
- Conveniently select to align to either the full Chinese Spring Wheat genome or a transcriptome
- Transcriptome is dynamically assembled on demand using selected annotation set from the IWGSC

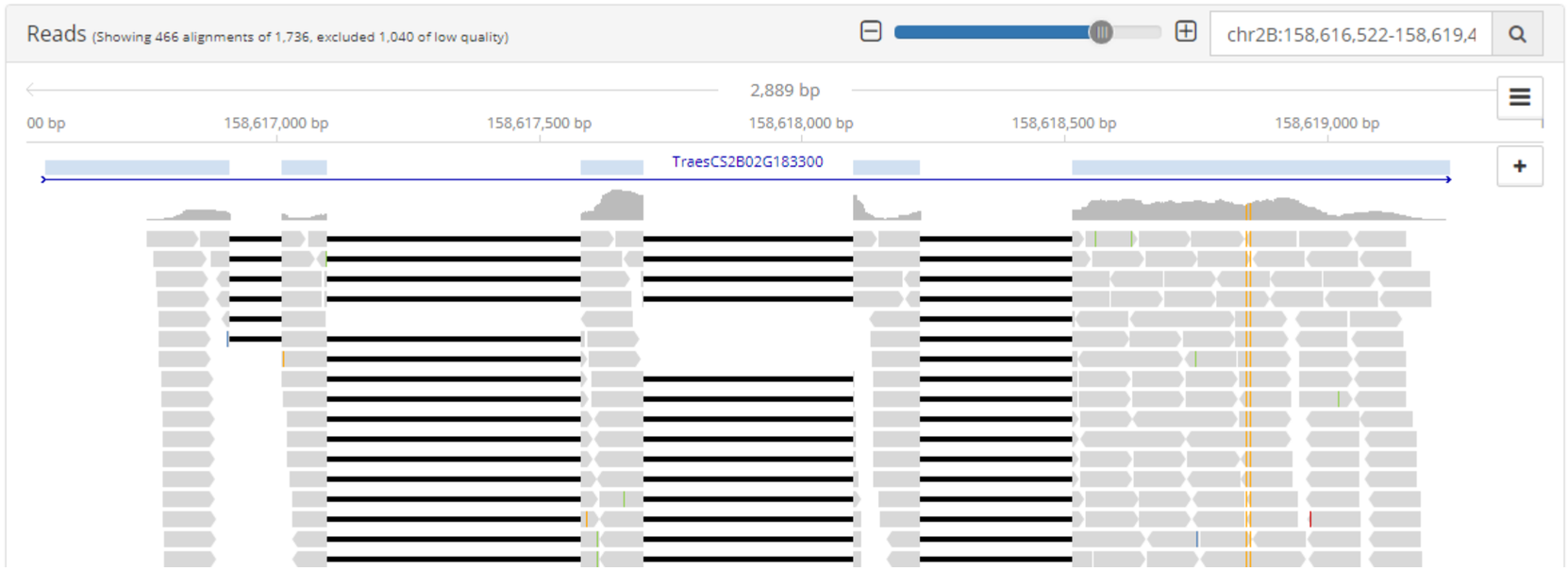


# RNA-Seq Read Mapping Visualization

- 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

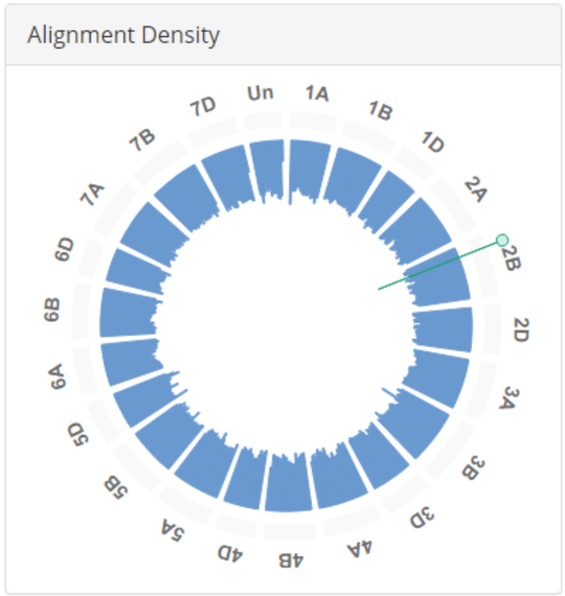


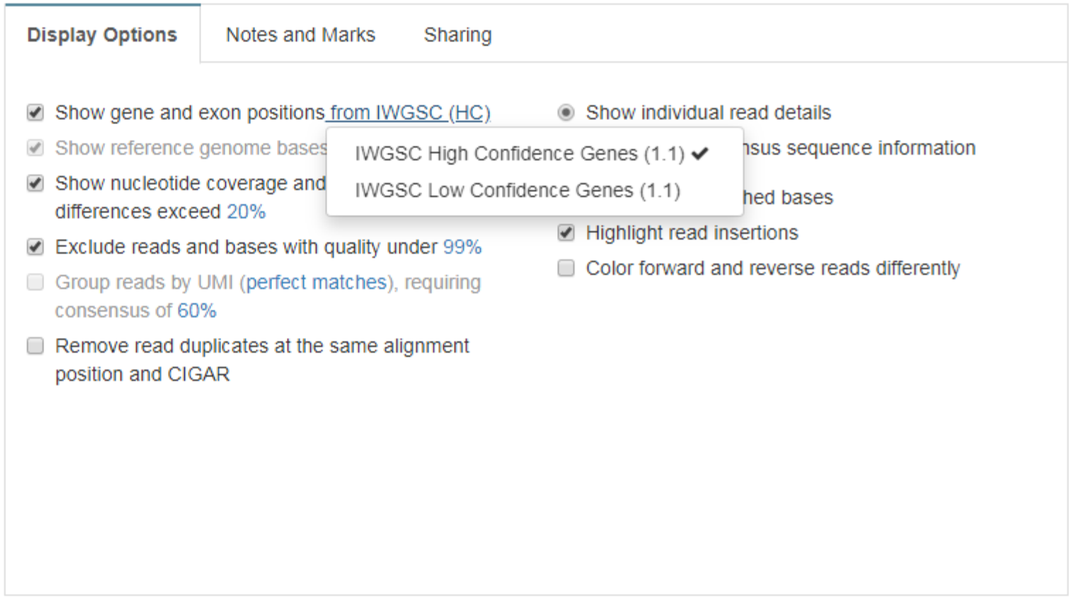



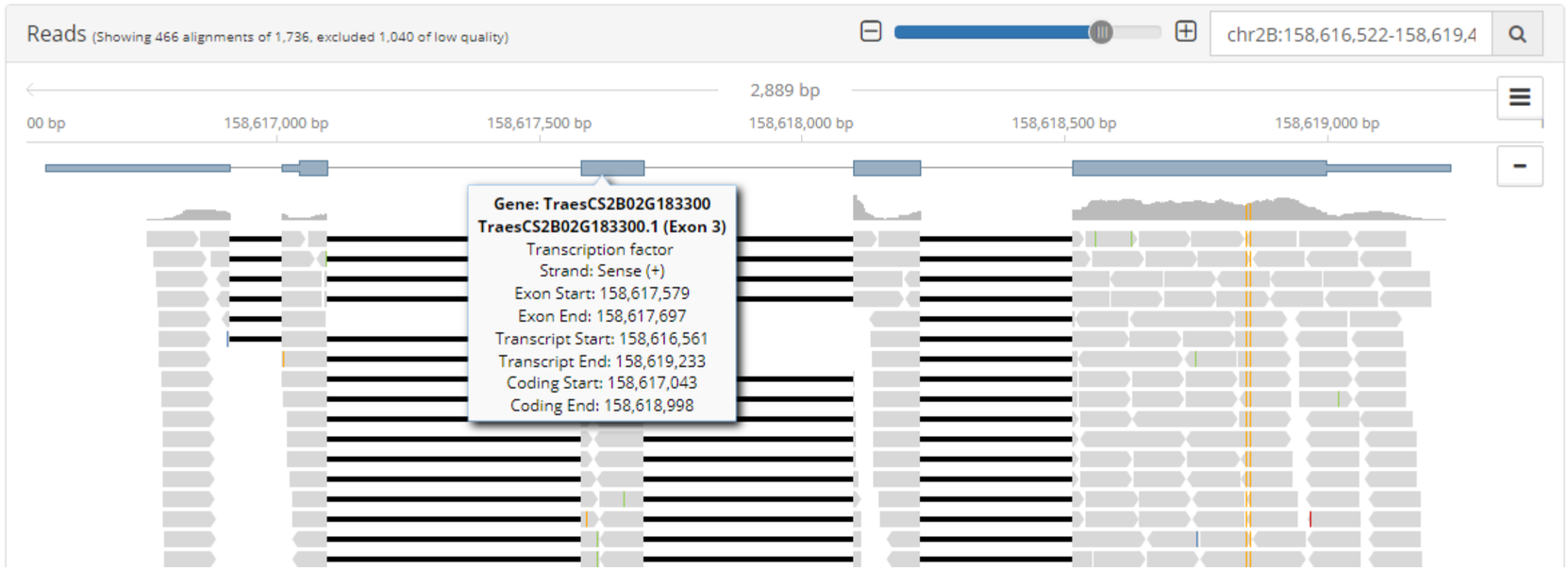



# RNA-Seq Read Mapping Visualization

- 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



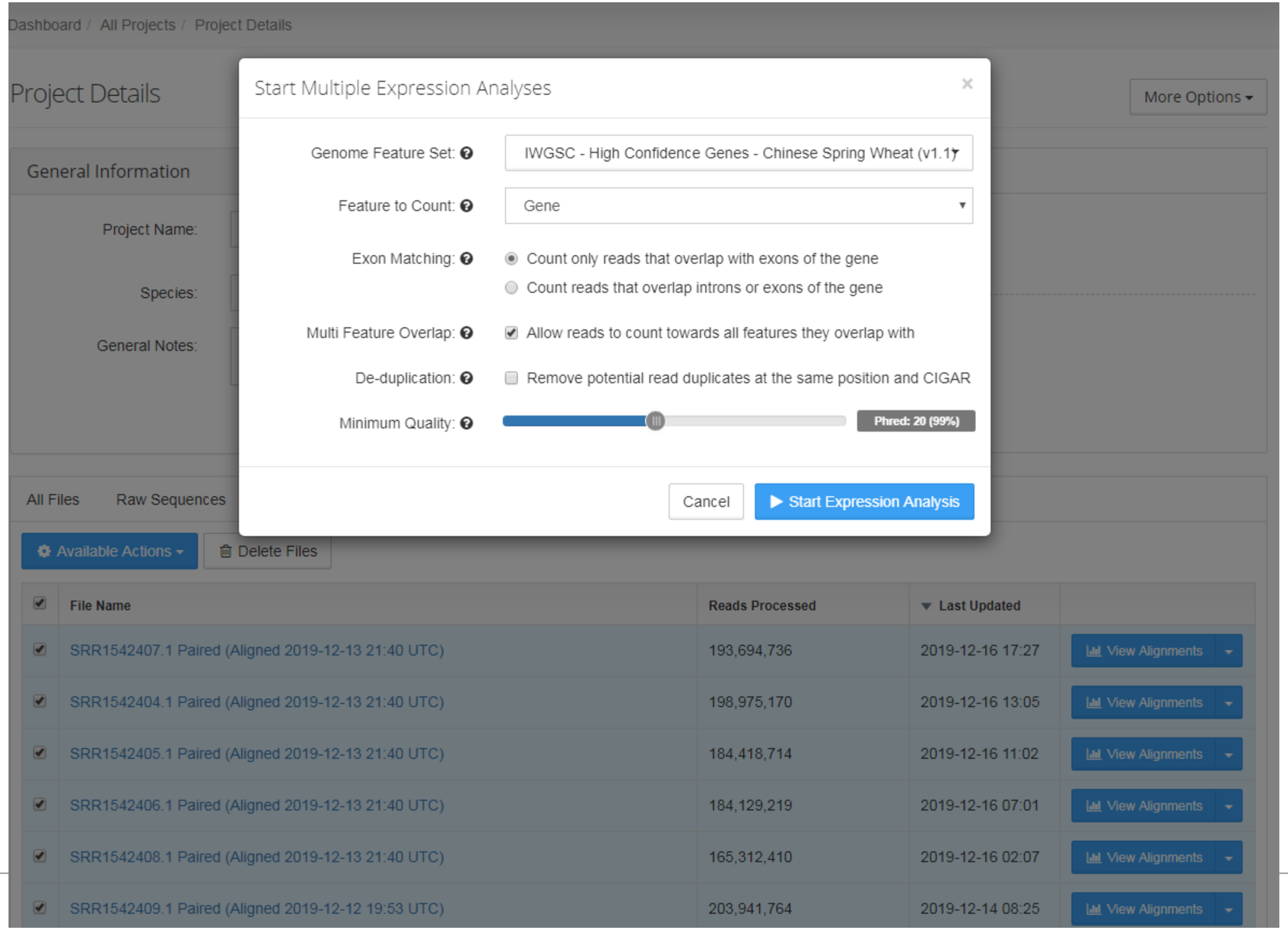



# RNA-Seq Hexaploid Wheat: Expression Analysis

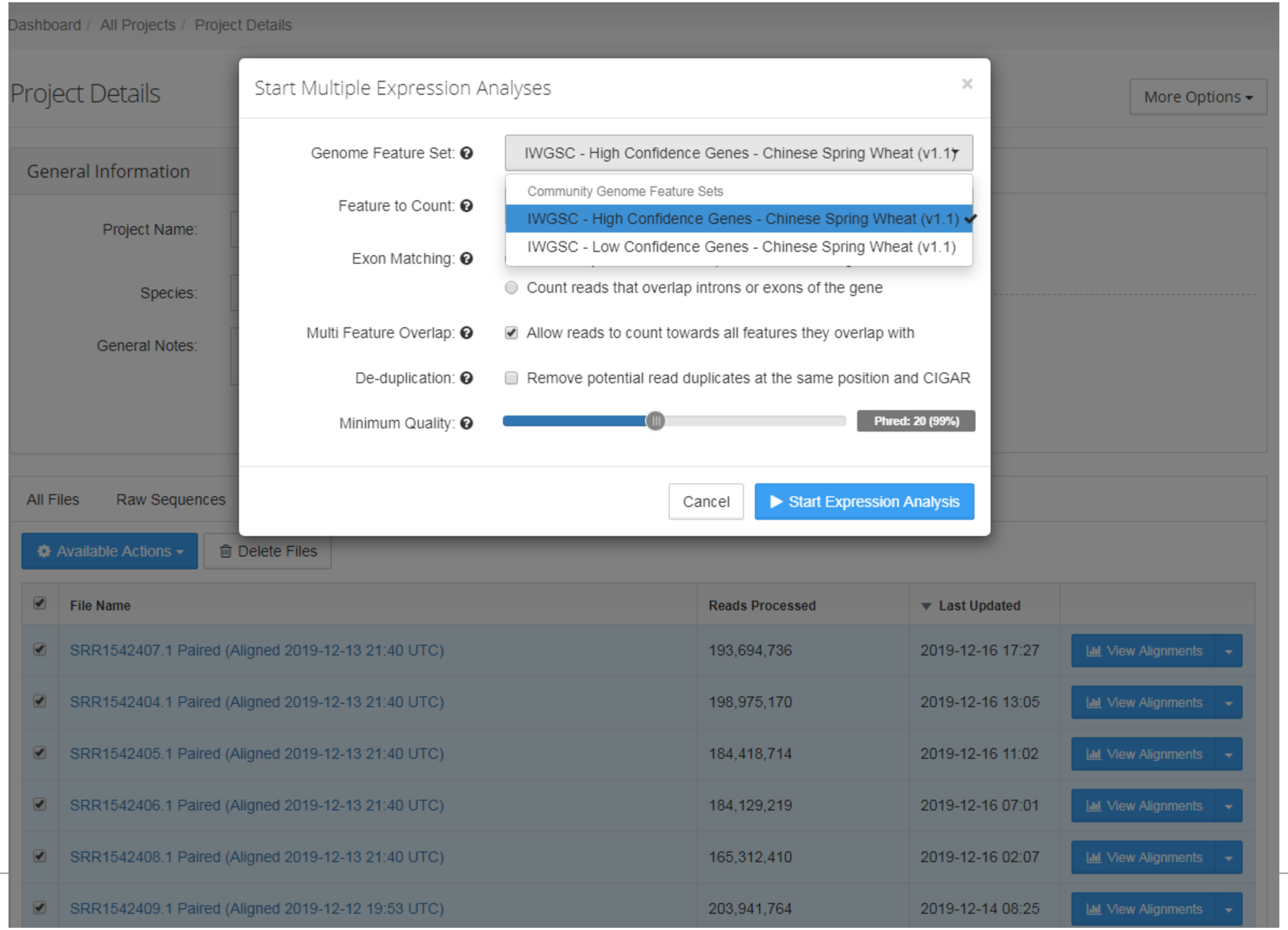
# Bread Wheat: Expression Analysis

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly



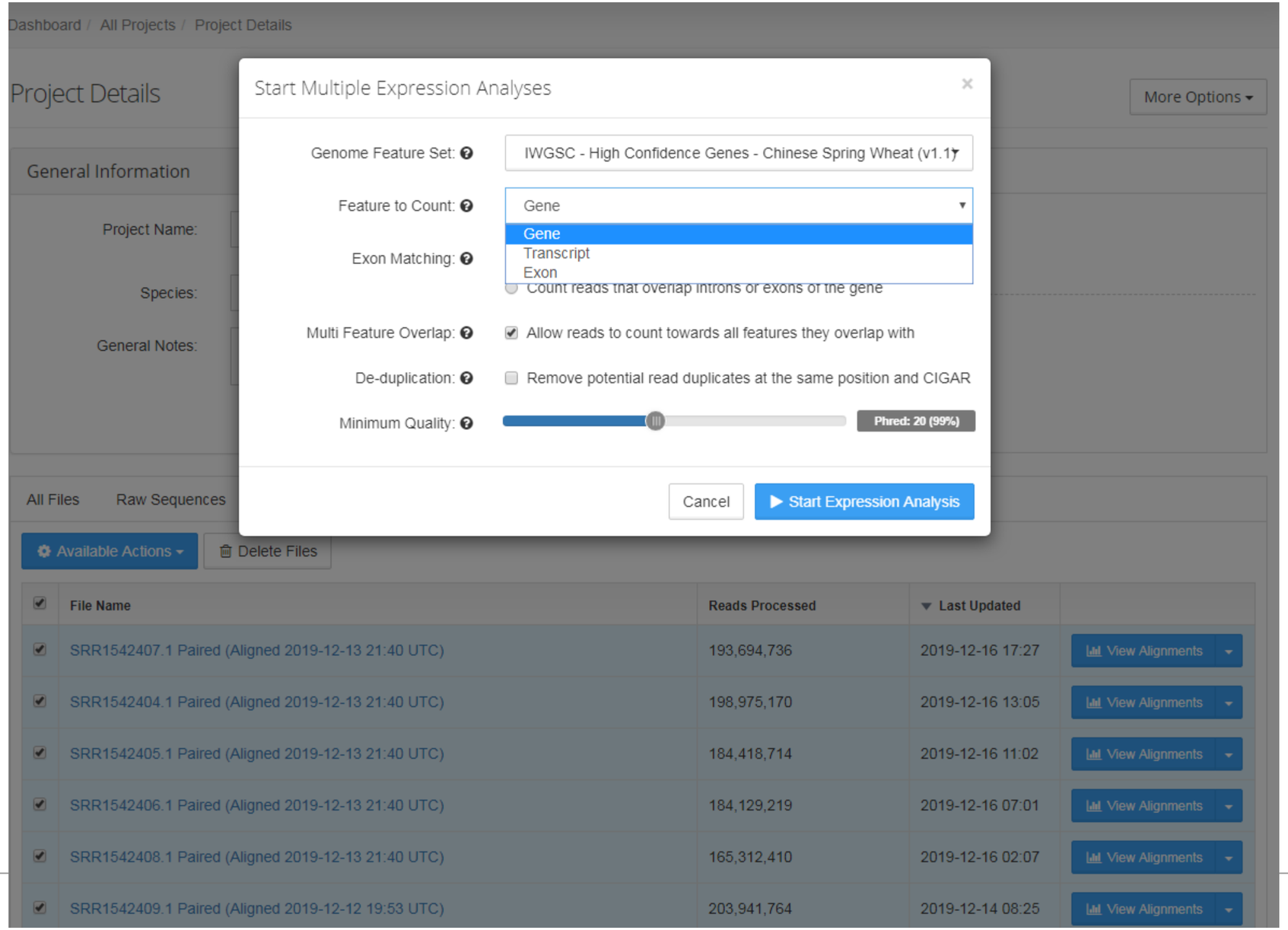
# Bread Wheat: Expression Analysis

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



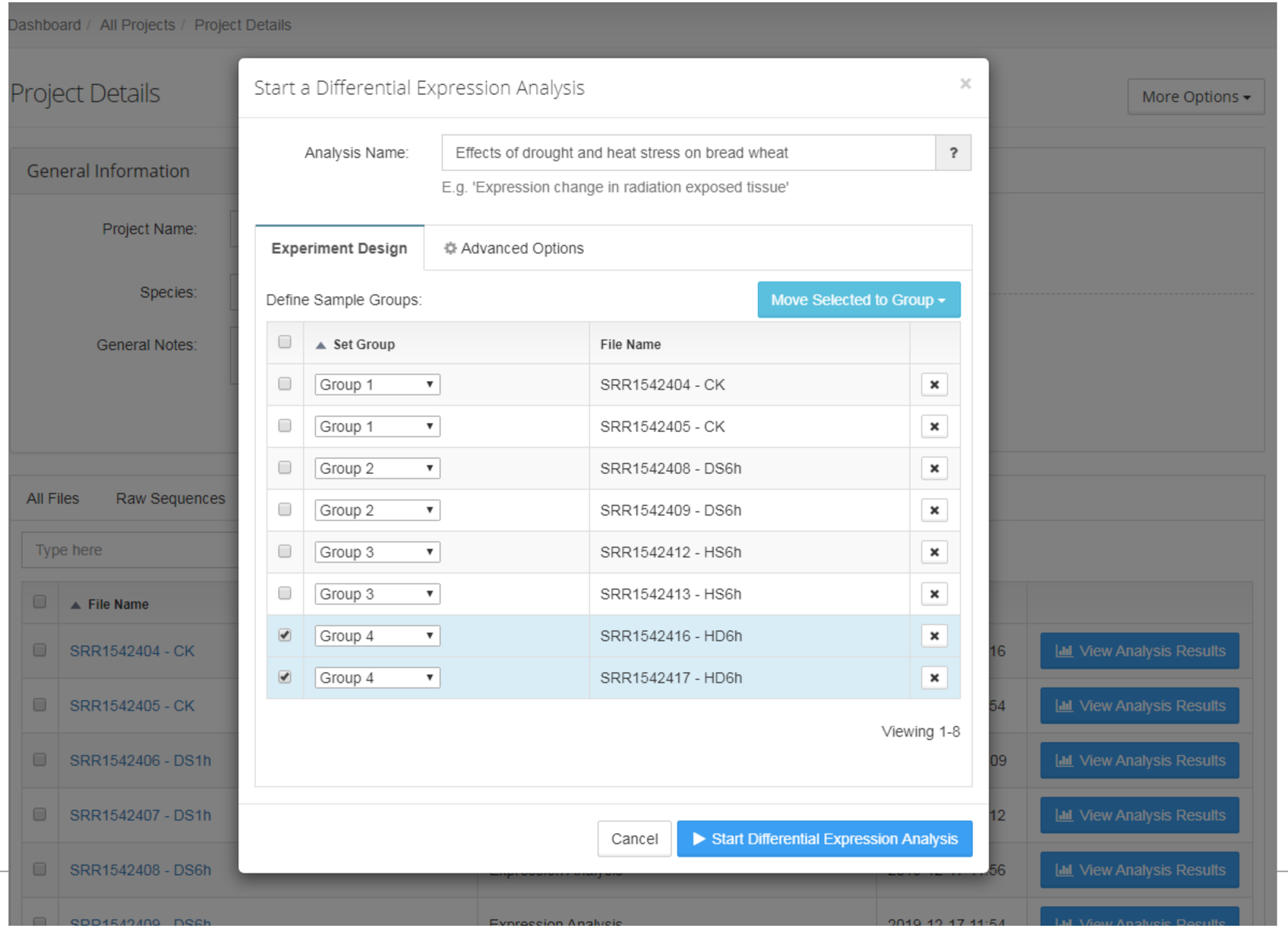
# Bread Wheat: Expression Analysis

- 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



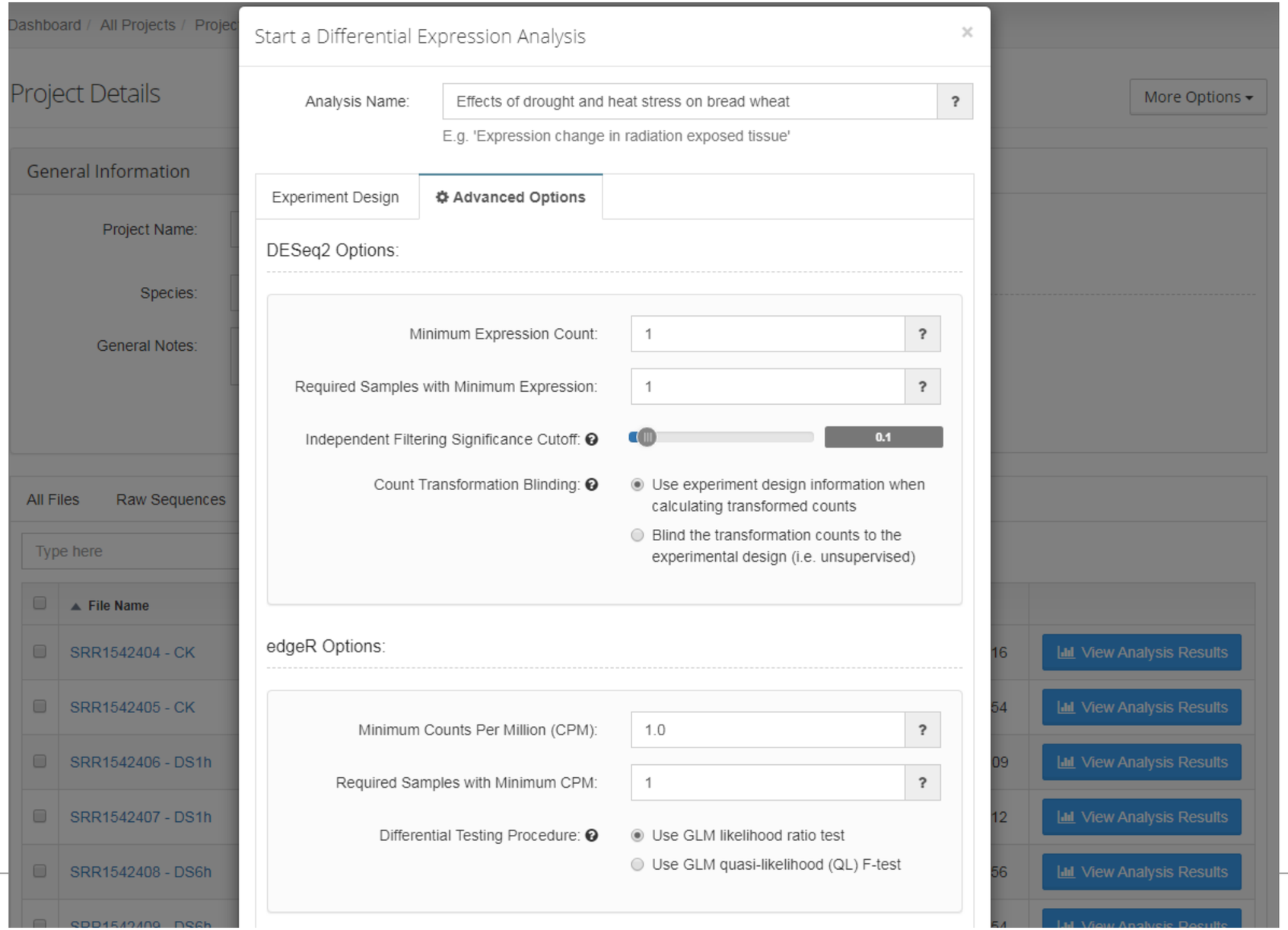
# Bread Wheat: Differential Expression

- 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…



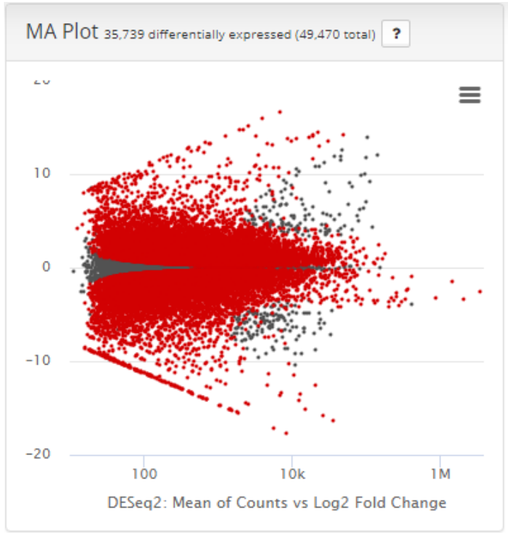
# Bread Wheat: Differential Expression

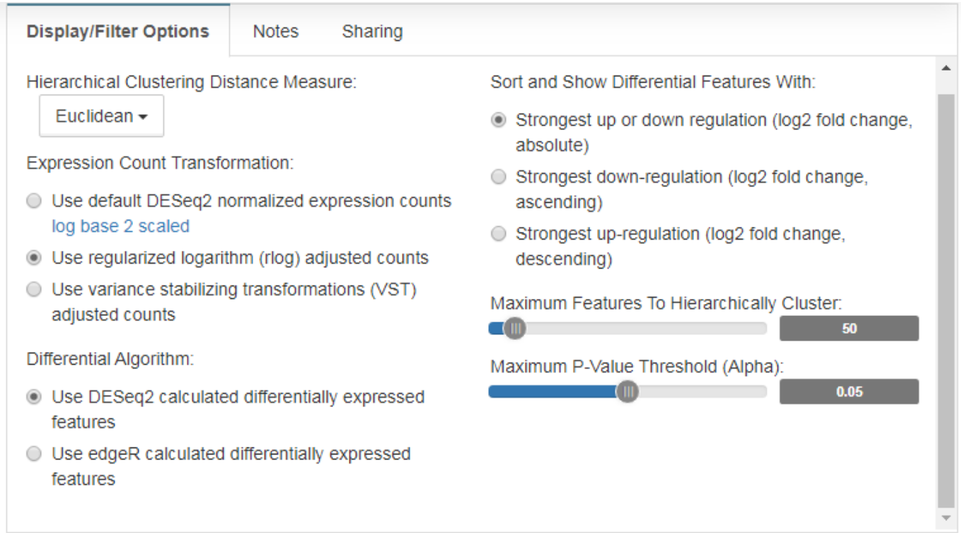
- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC “High Confidence” or “Low Confidence” annotation sets
- Measure expression levels of genes, transcripts, or individual exons
- Group samples for a differential expression analysis…
- …and conveniently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



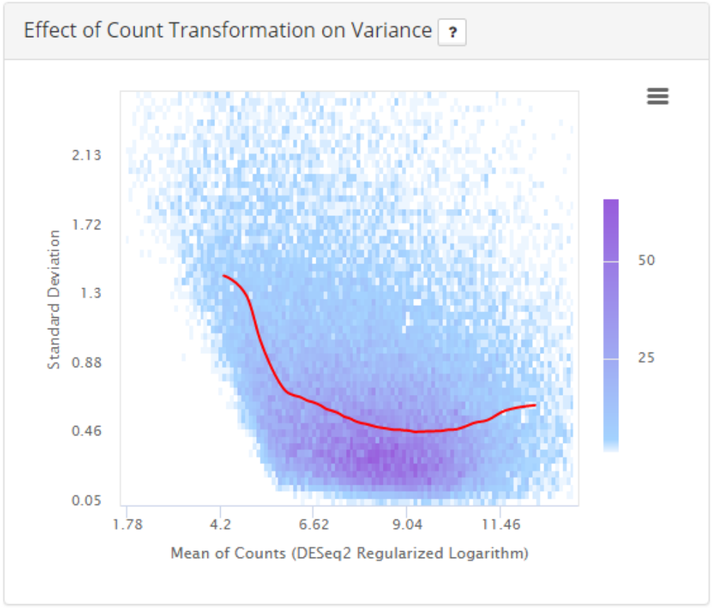
# Bread Wheat: Differential Expression

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



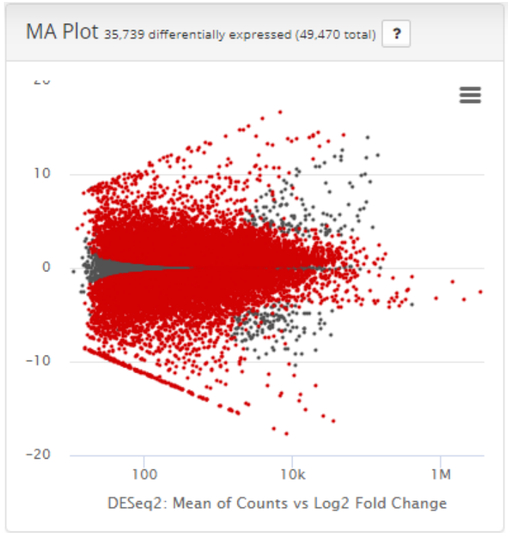



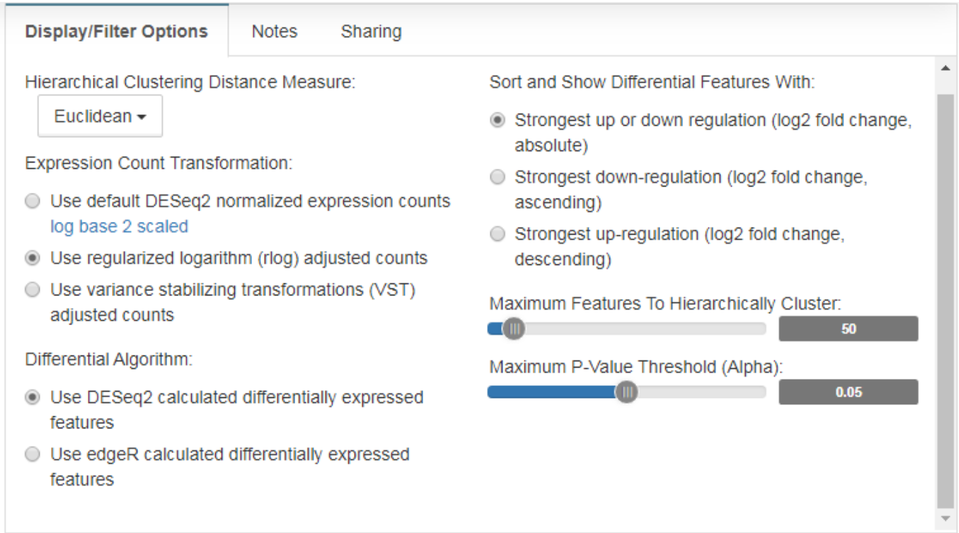


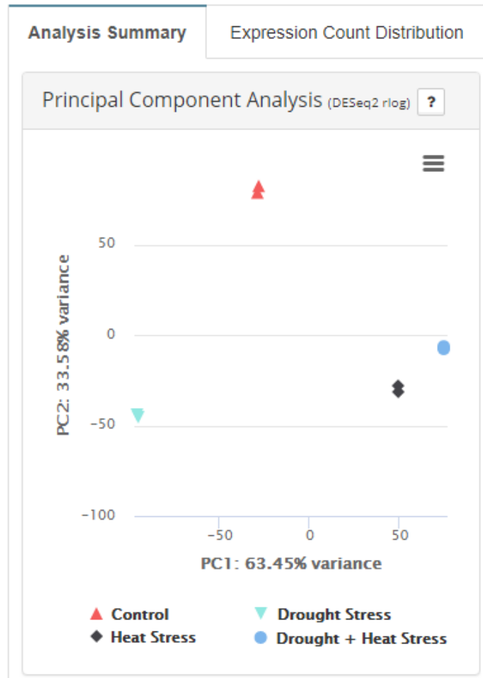


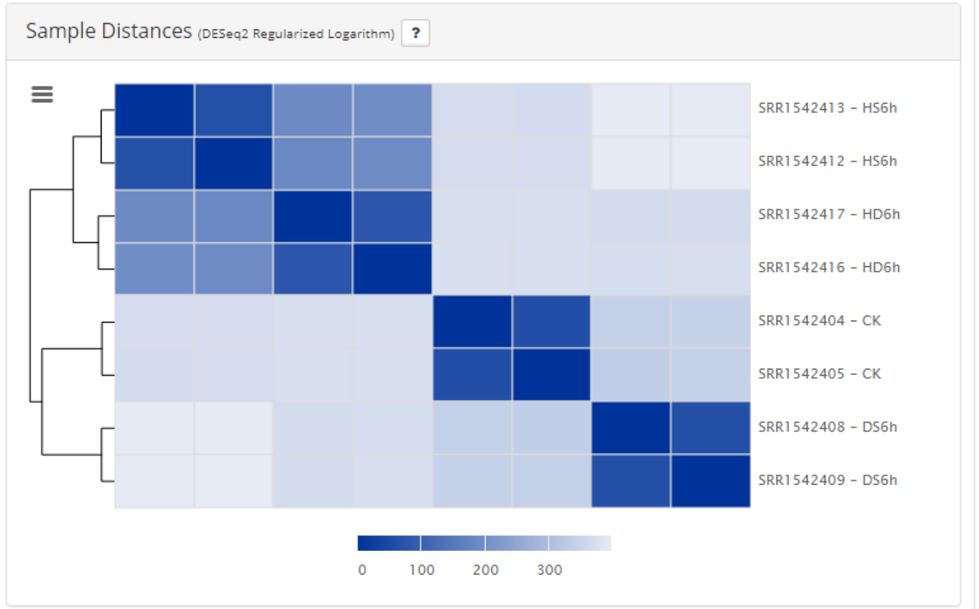
# Bread Wheat: Differential Expression

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



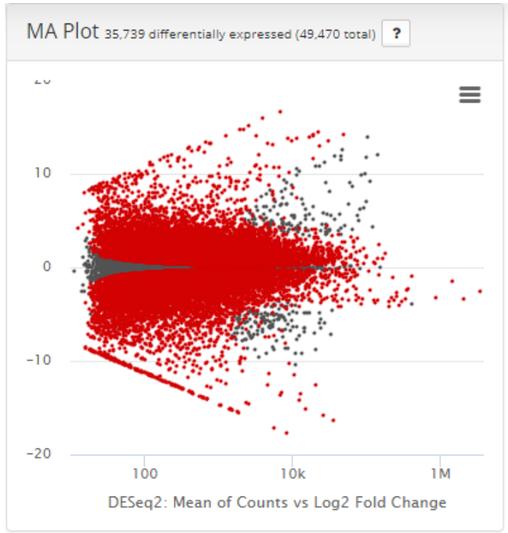



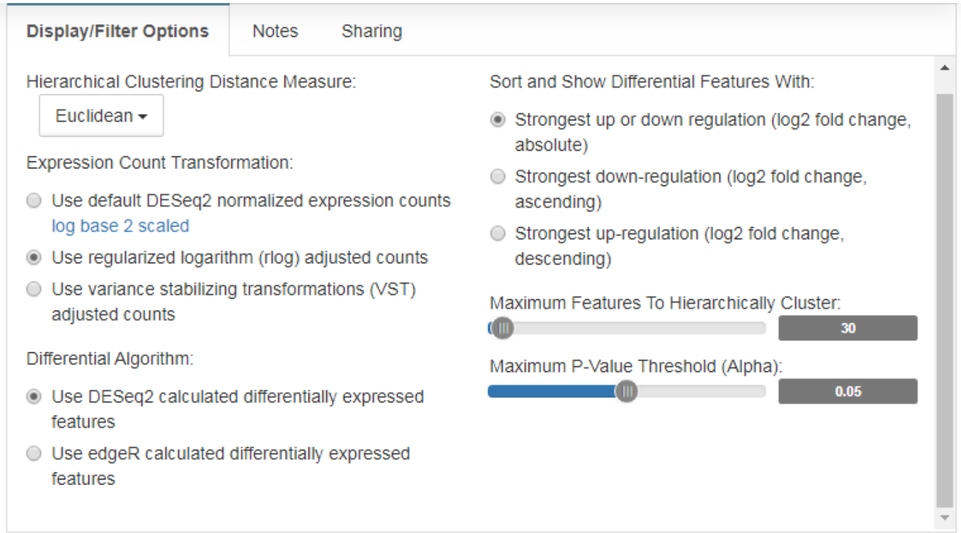



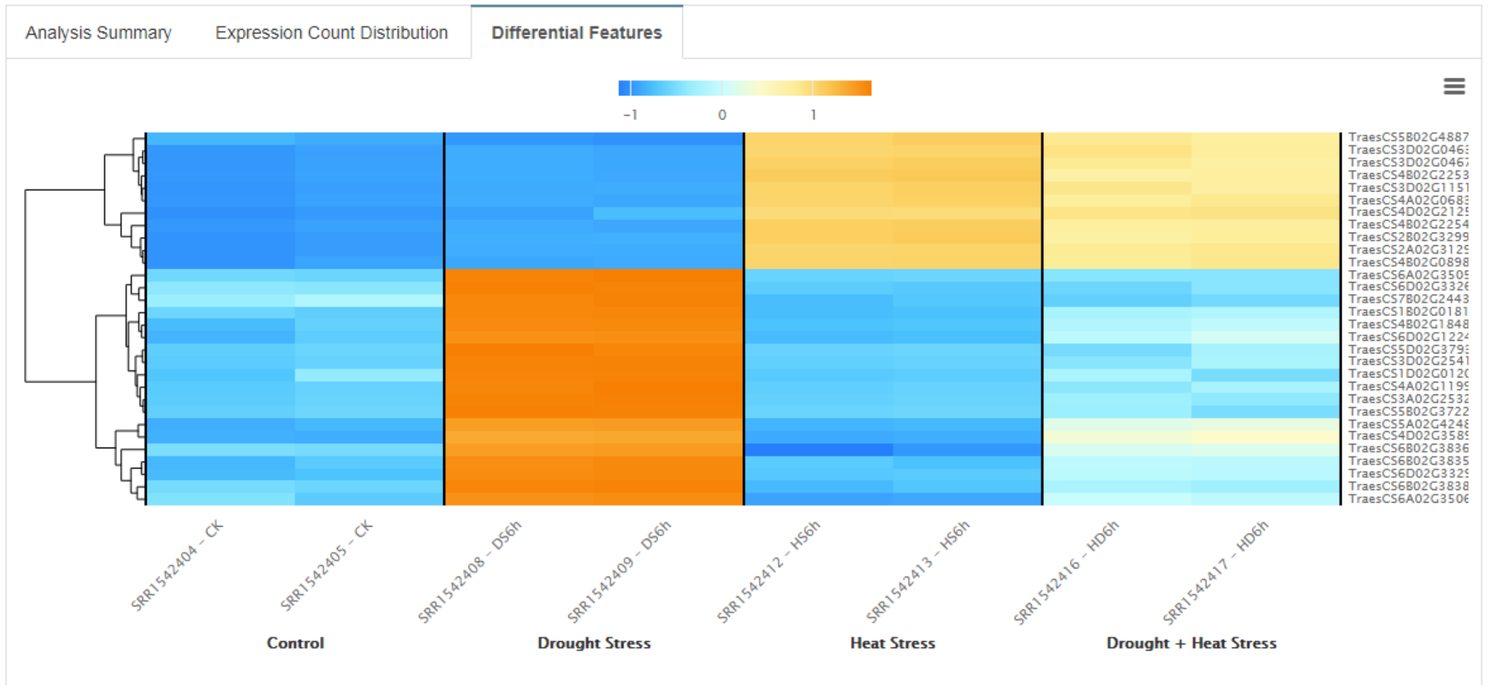



# Bread Wheat: Differential Expression

- 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

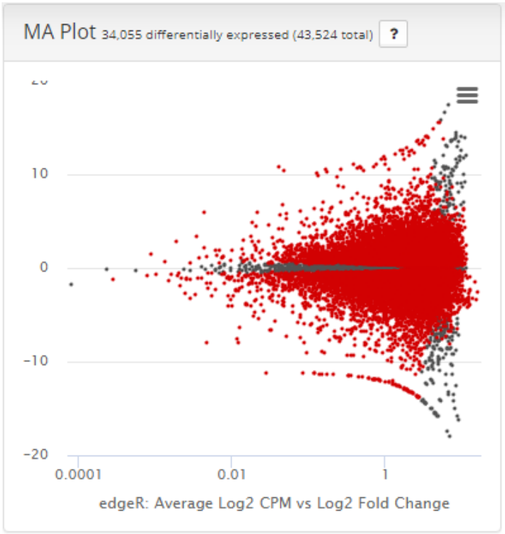


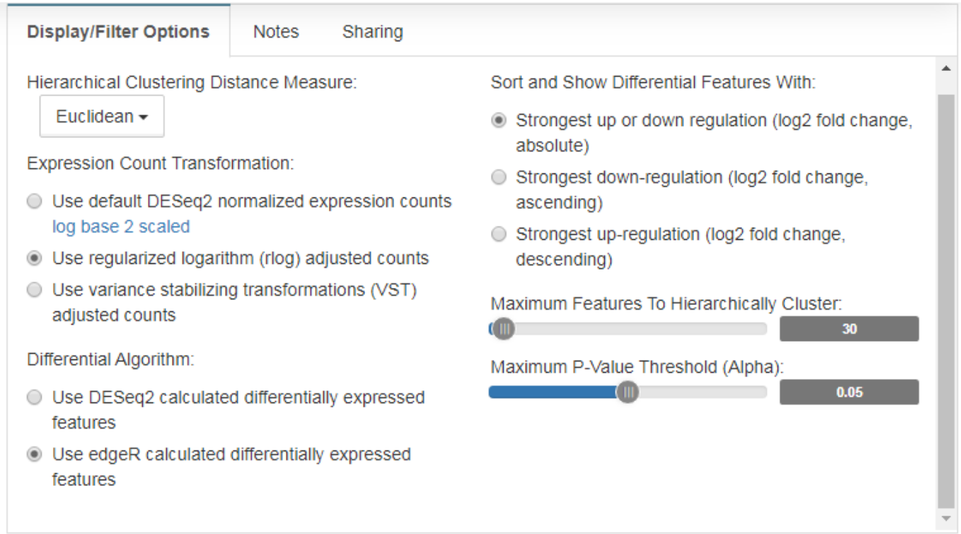



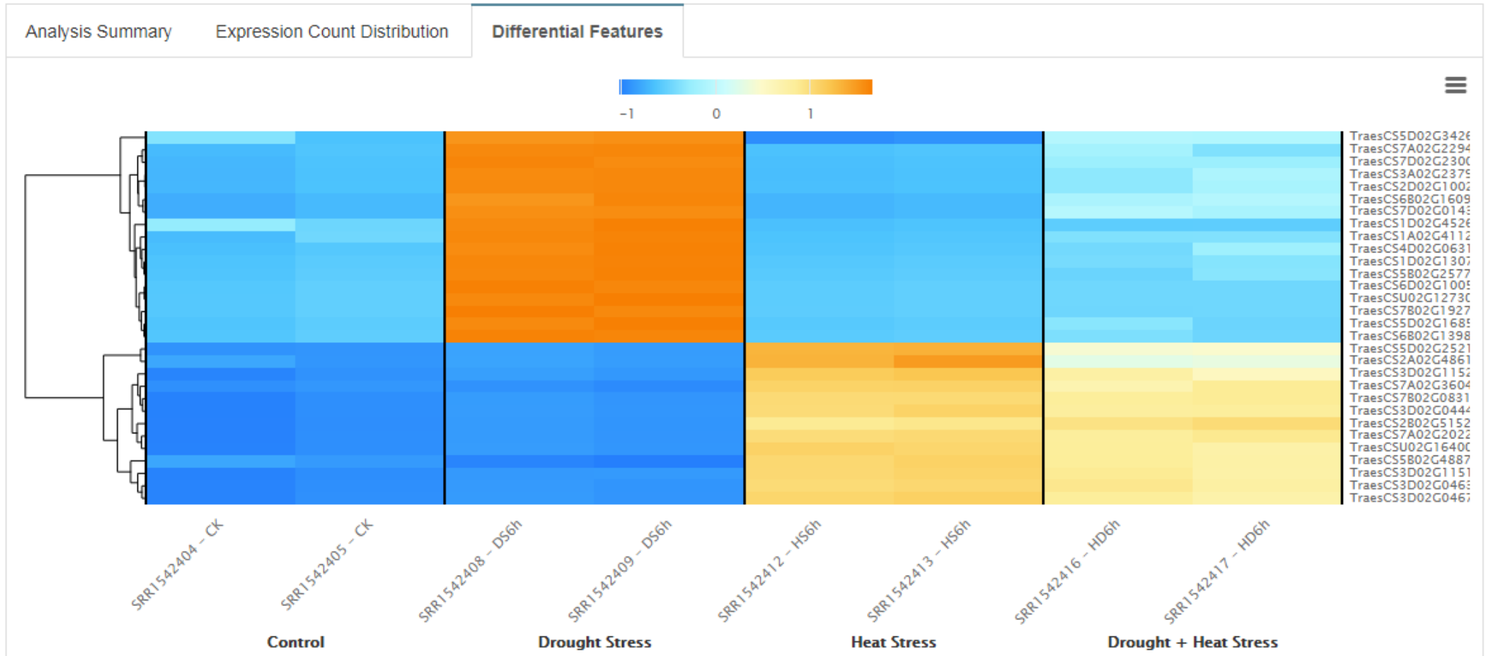



# Bread Wheat: Differential Expression

- 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



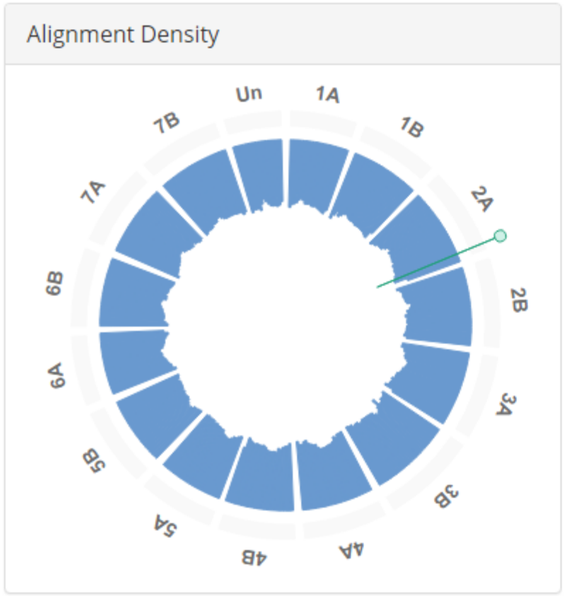



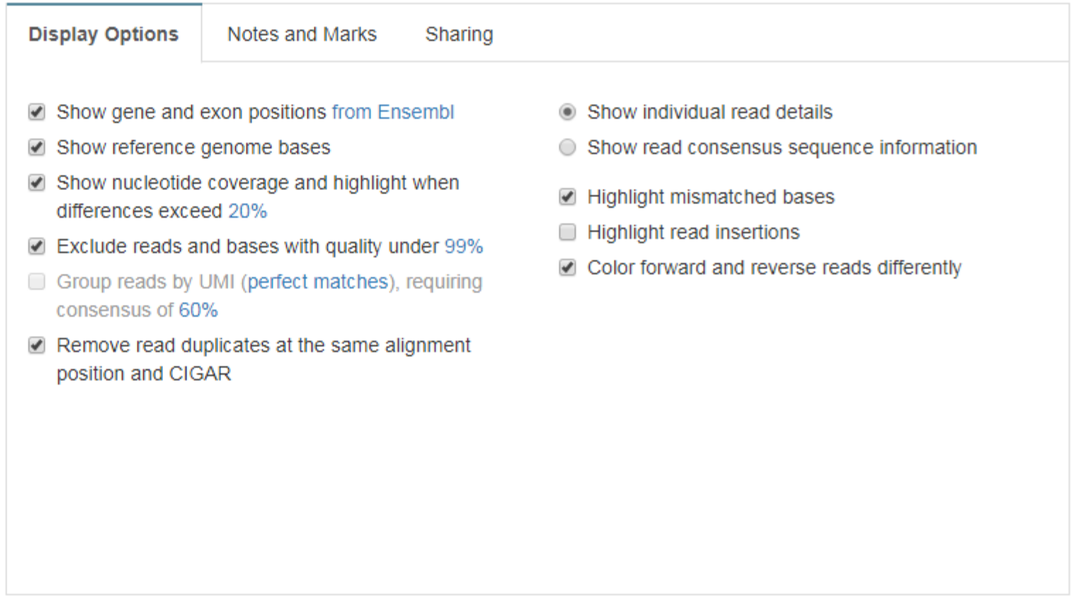



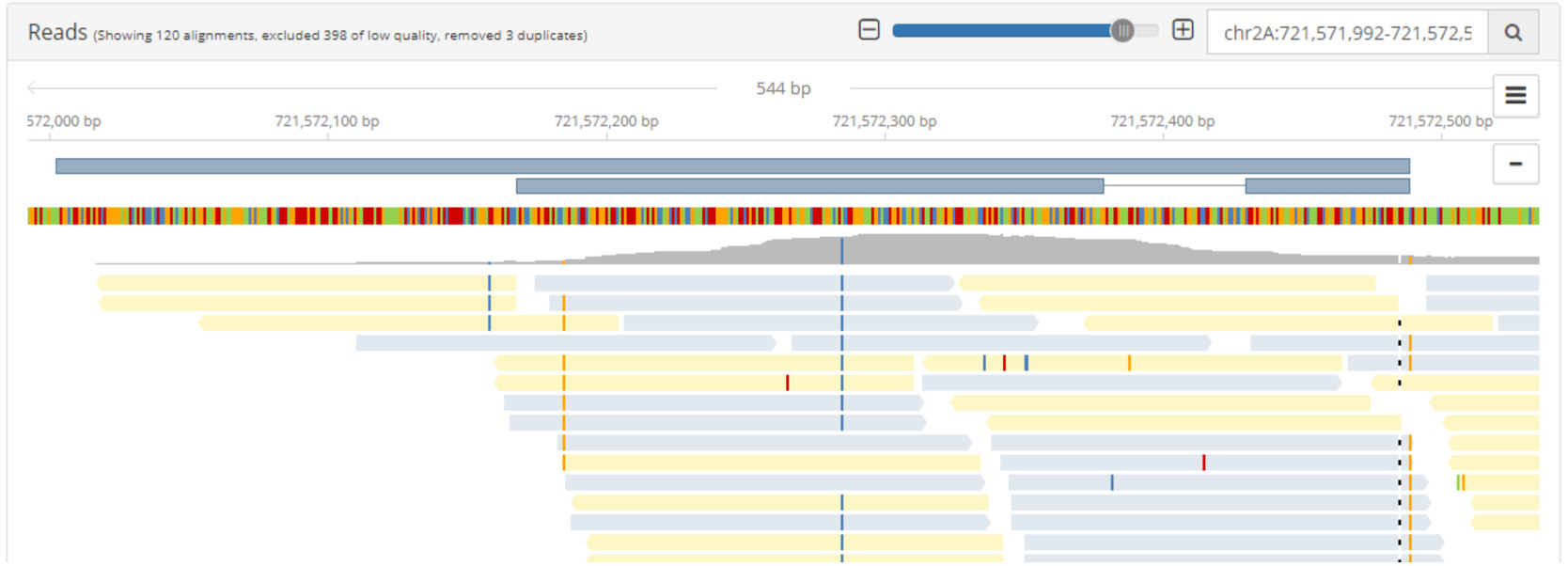
# DNA-Seq: Tetraploid Wheat

# Tetraploid Wheat: Read Mapping

- 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

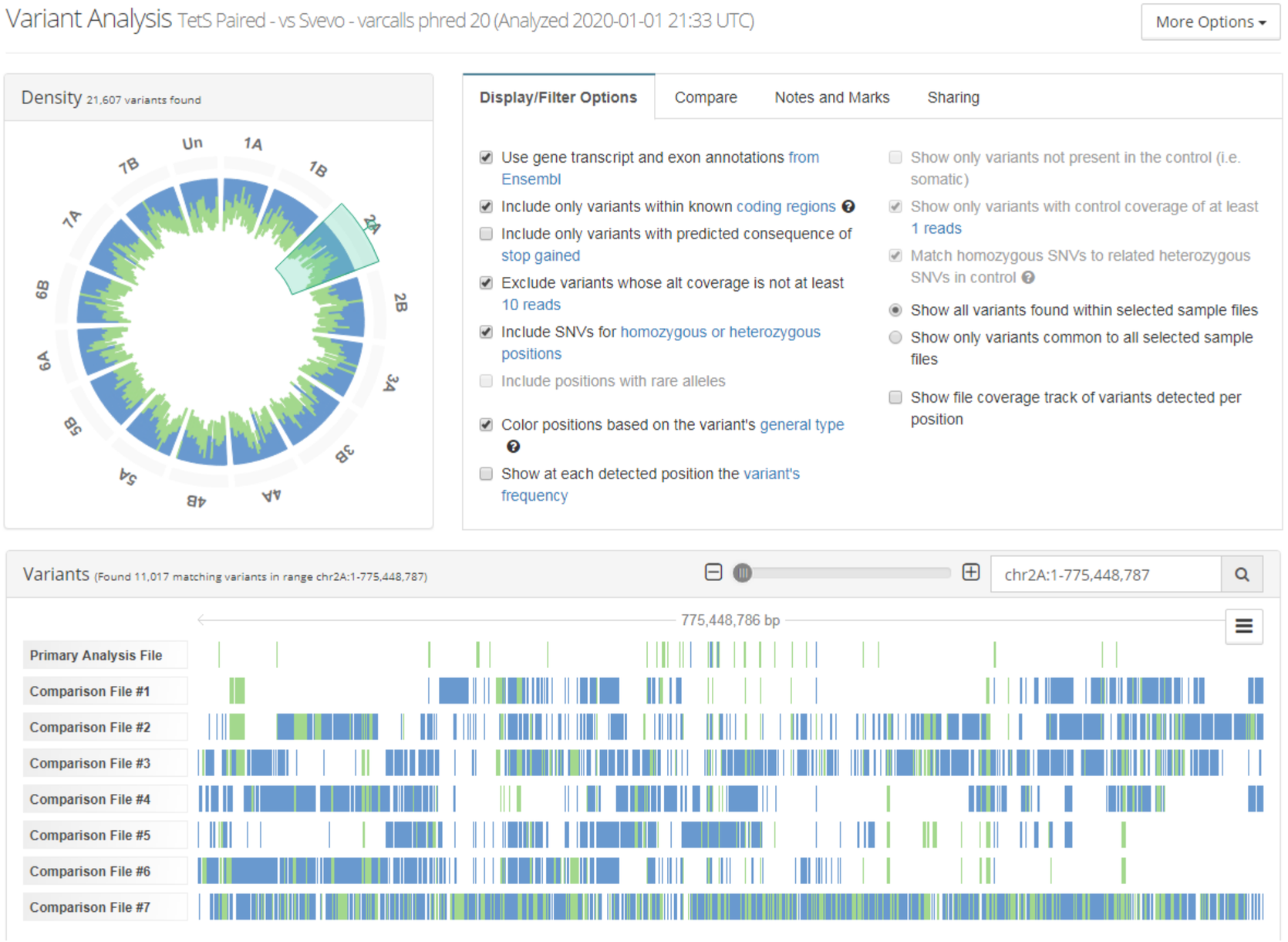


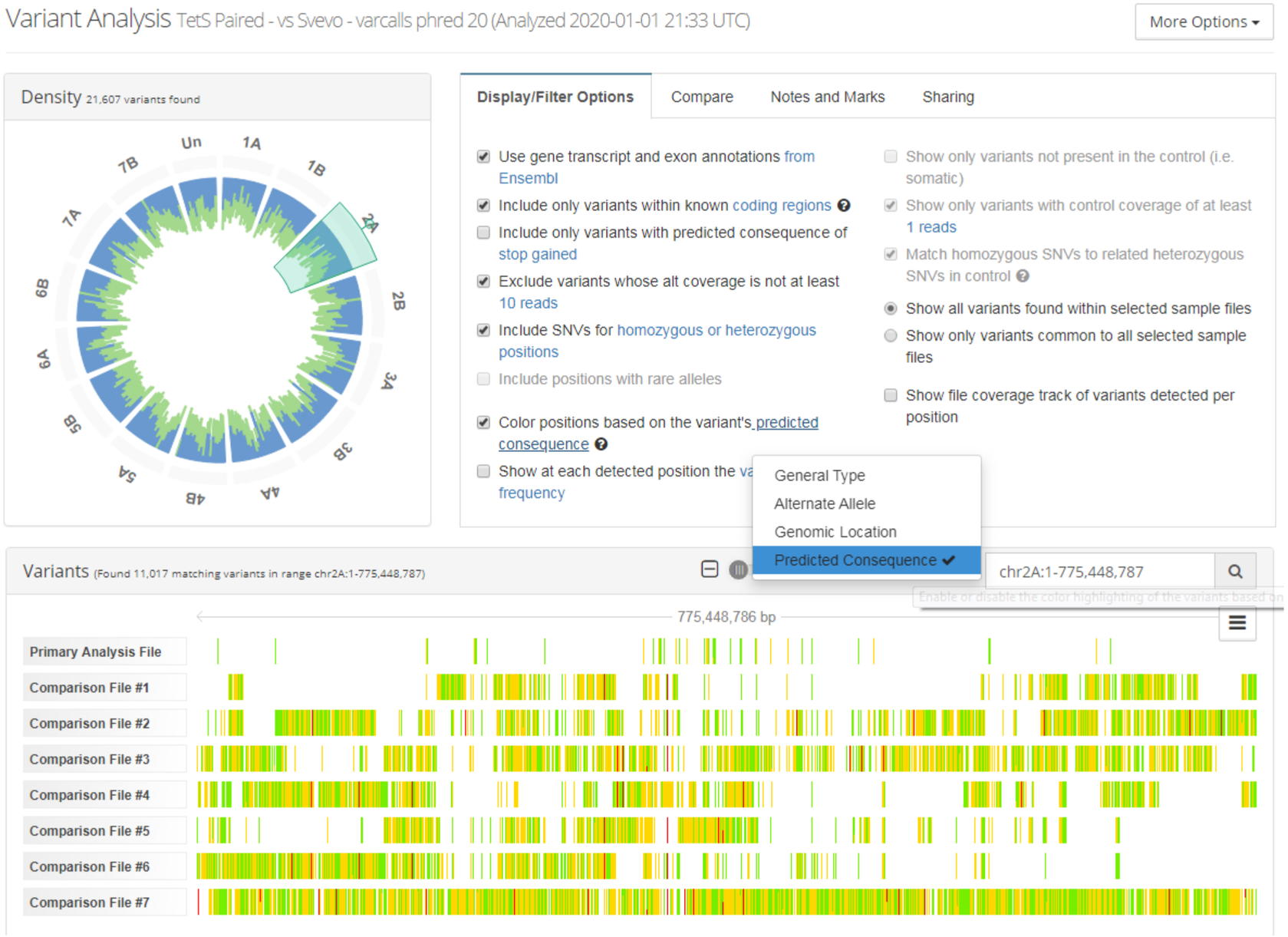
# Tetraploid Wheat: Variant Analysis

- 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



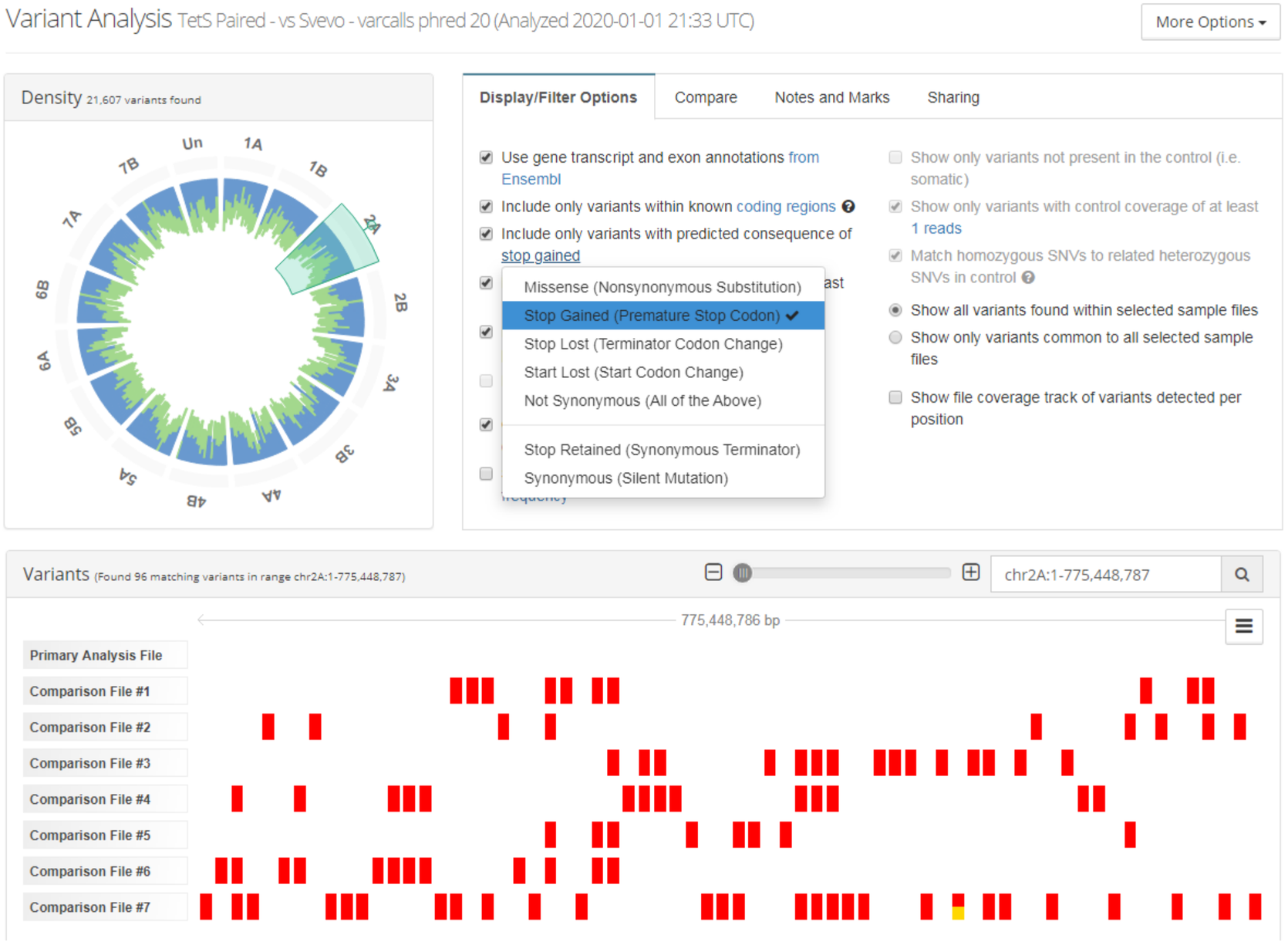
# Tetraploid Wheat: Variant Analysis

- 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



# Tetraploid Wheat: Variant Analysis

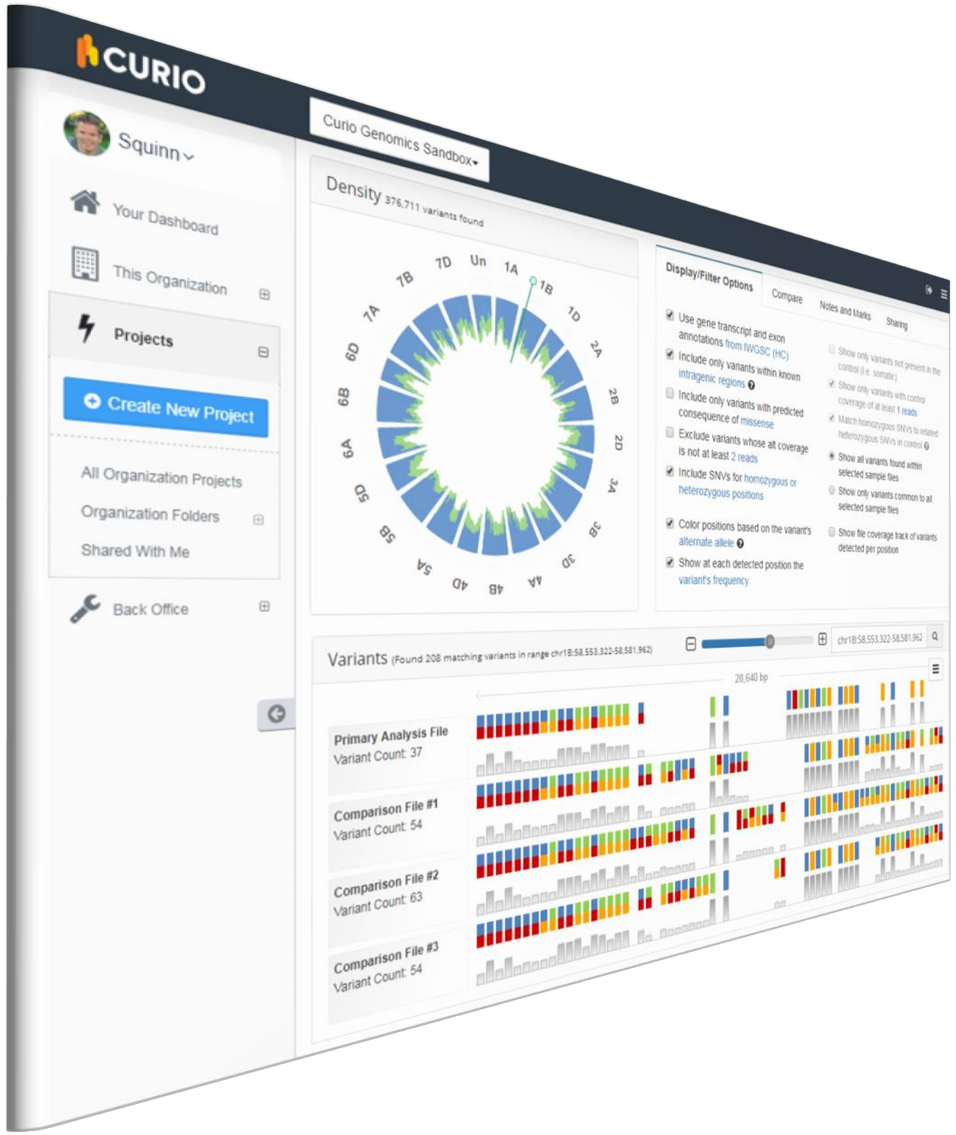
- Call and analyze variants leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and visualize variant density across multiple samples
- Dynamically filter to different genomic region types based on annotations published by the IDWGSC via Ensembl
- Access predicted biological consequences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



# 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