A Powerful Tool For Mapping Of Short Reads

INTRODUCING NOVOALIGN

Precise and Seamless Mapping

novoAlign is a formidable solution for researchers seeking to map short reads onto a reference genome from Illumina and MGI Tech sequencing platforms. It is designed to streamline and optimize this critical process, ensuring precise alignments and facilitating a deeper understanding of your genetic data. With unsurpassed features and robust performance, it's time to elevate your research with seamless short read mapping. Unlock the potential of your data and accelerate your discoveries with our cutting-edge technology.

Why novoAlign?

01

High concordance rate

Excels in high concordance rates for variant calling, with seamless GATK4 integration providing consistent and reliable genetic analyses.

02

Precise alignment

Designed for precise alignment, novoAlign consistently achieves top-tier precision to minimize false callings in your alignment, and ensure accurate alignment.

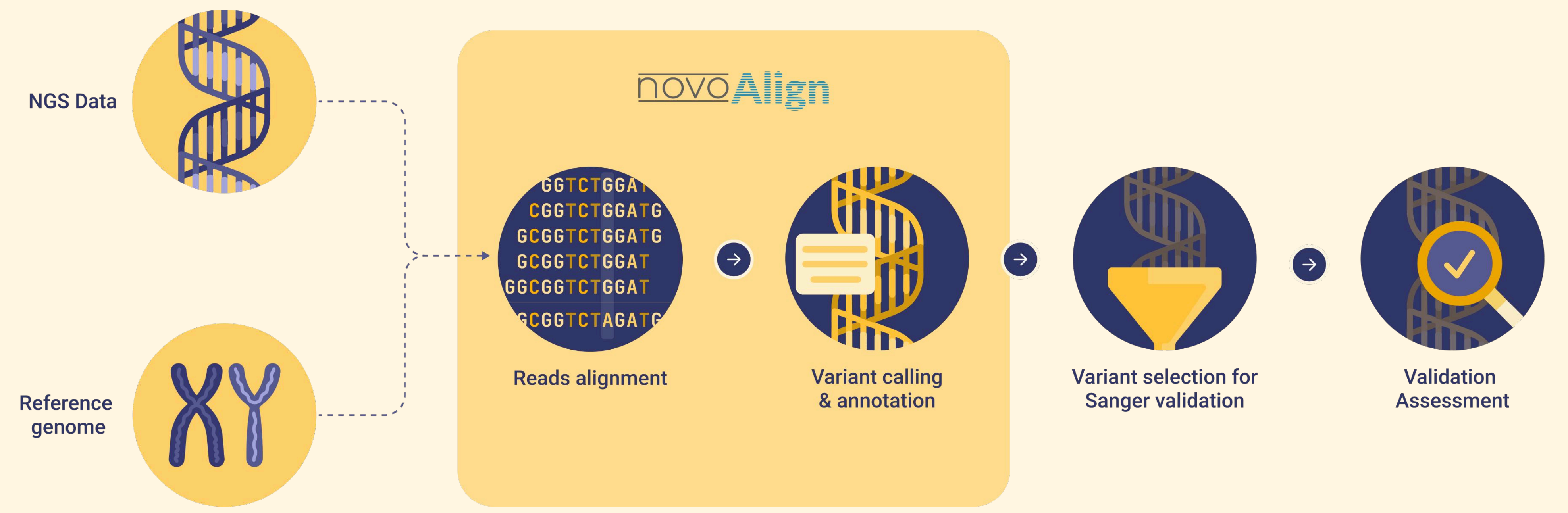
03

High sensitivity

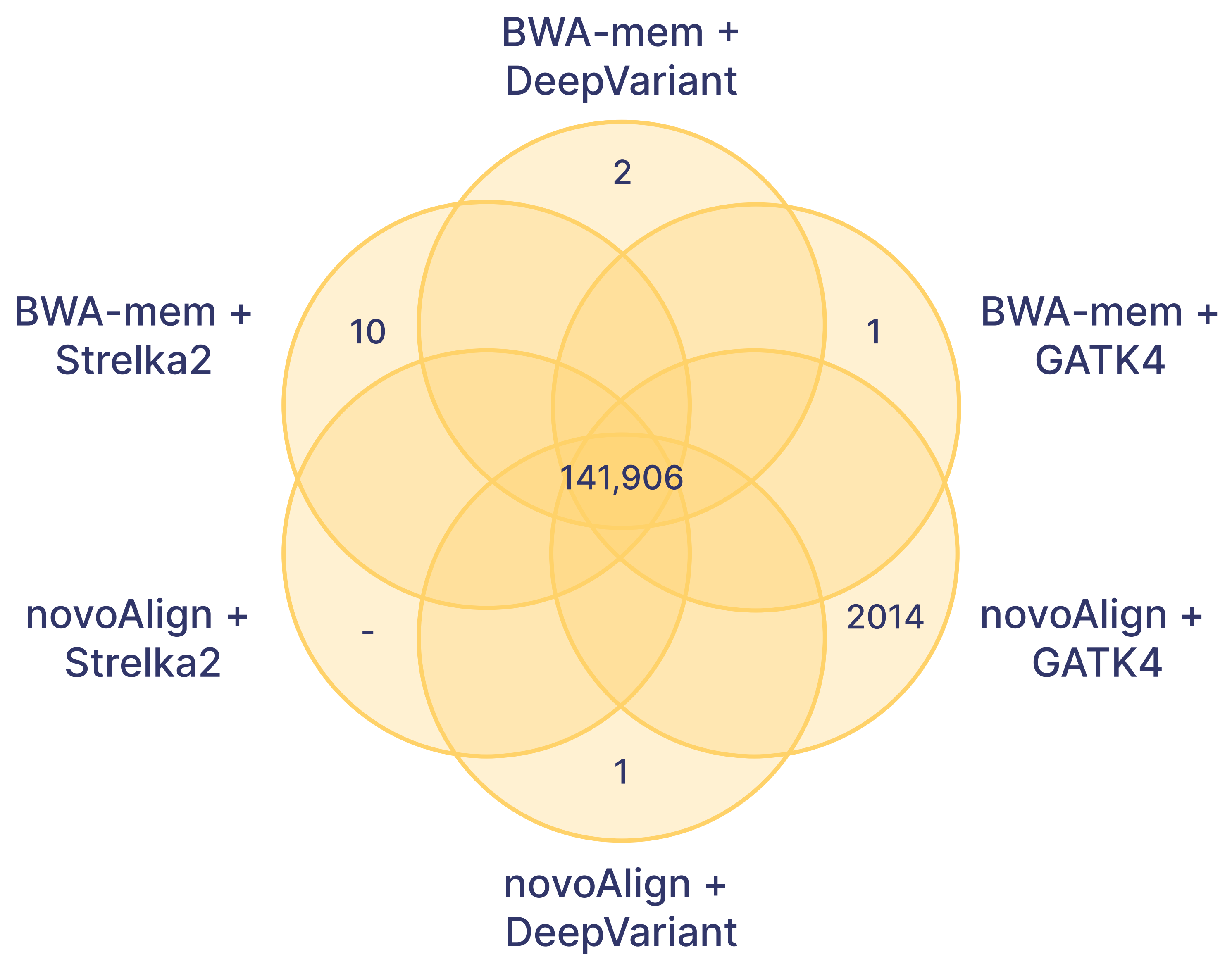
Achieves superior sensitivity, mapping more reads than other aligners to capture all critical genetic variations and provide a comprehensive view of genomic data.

> **“novoAlign showed the highest accuracy in all alignments. It is also reached the best mapping score for uniquely aligned coding sequences in WGS analysis”**
>
> Neural Computing and Applications, 32(22), 15669-15692 (2021)

How novoAlign Works?



Optimising variant calling with novoAlign + GATK4



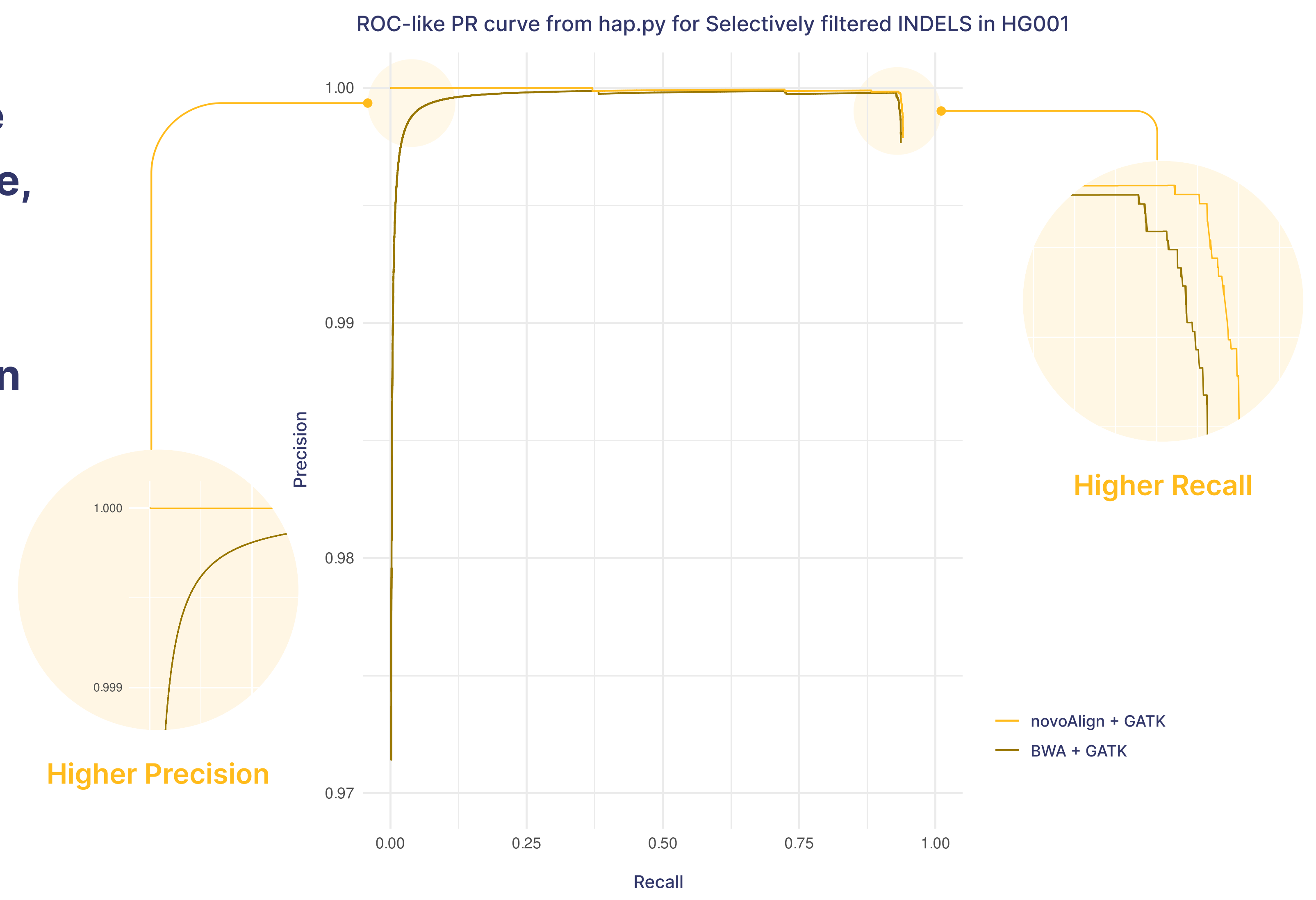
99.87%

ALIGNMENT ACCURACY

novoAlign + GATK detects more variants than the other. In a previous study, it showed that novoAlign + GATK pipeline detects more variants than the other. Specifically, out of the 2014 variants identified by novoAlign + GATK , 1560 come from chromosome 6. Among these, 1564 variants are within the major histocompatibility complex (MHC) region. Impressively, this region closely matches the microarray chip and WGS genotypes with a 99.87% accuracy, with just 2 mismatches detected.

- Park et al., Genes & Genomics (2023)

novoAlign+GATK pipeline scores a higher AUC value, higher recall, and higher precision than the BWA+GATK pipeline when compared to GIAB benchmark.



Pipeline for SNPs	AUC
🟨 novoAlign + GATK	0.9402
🟫 BWA + GATK	0.9360

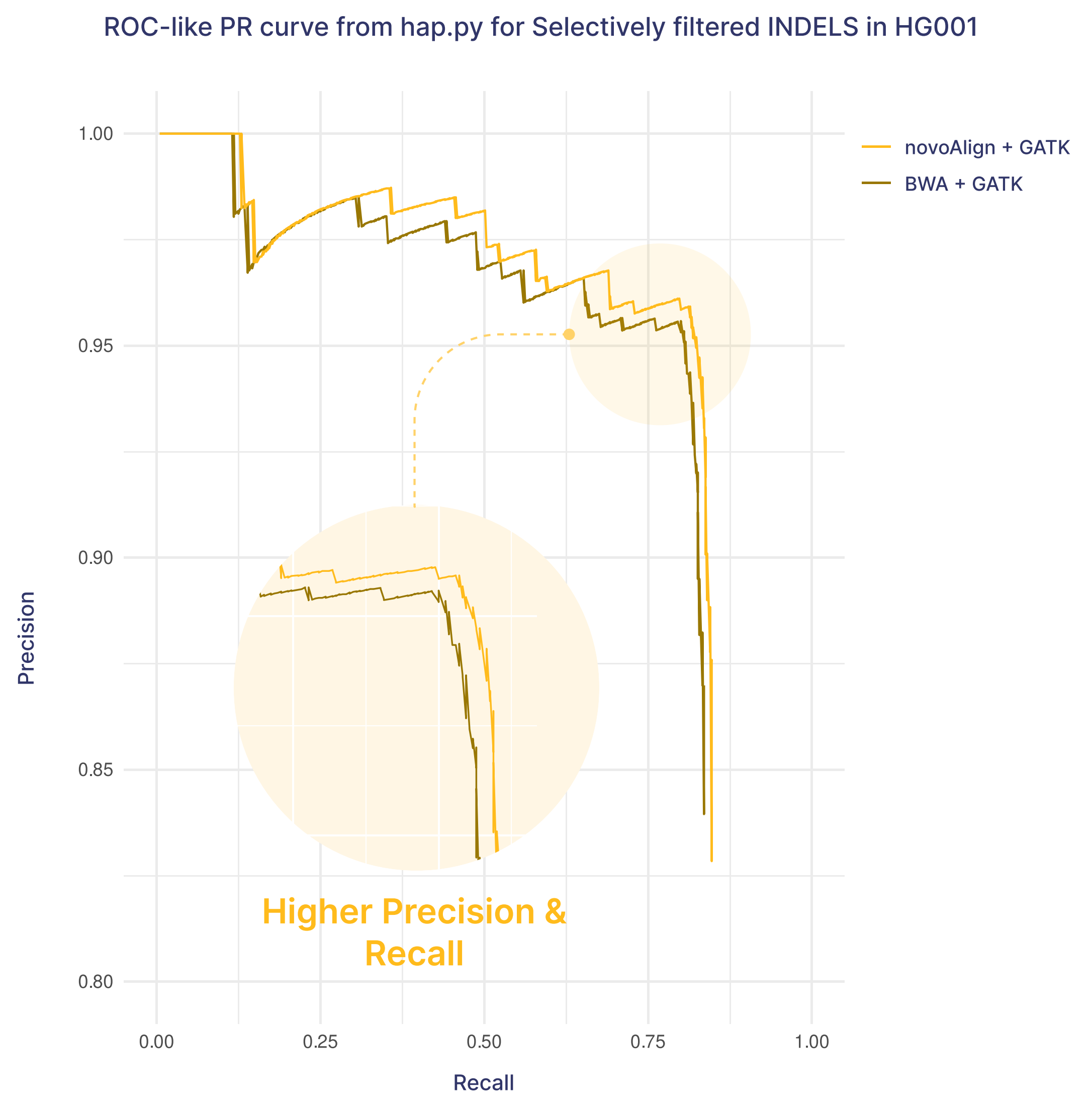
ROC-like PR curves from hap.py results comparing GIAB benchmark **SNPs** with Novoalign+GATK and BWA+GATK pipeline SNPs. The graph focuses on high-quality SNPs, and the AUC value is calculated using the "bayestestR" library in R.

0.9402

AUC VALUE FOR SNPS

0.8233

AUC VALUE FOR INDELS



ROC-like PR curves from hap.py results comparing GIAB benchmark **Indels** with Novoalign+GATK and BWA+GATK pipeline Indels. The graph focuses on high-quality Indels, and the AUC value is calculated using the "bayestestR" library in R.

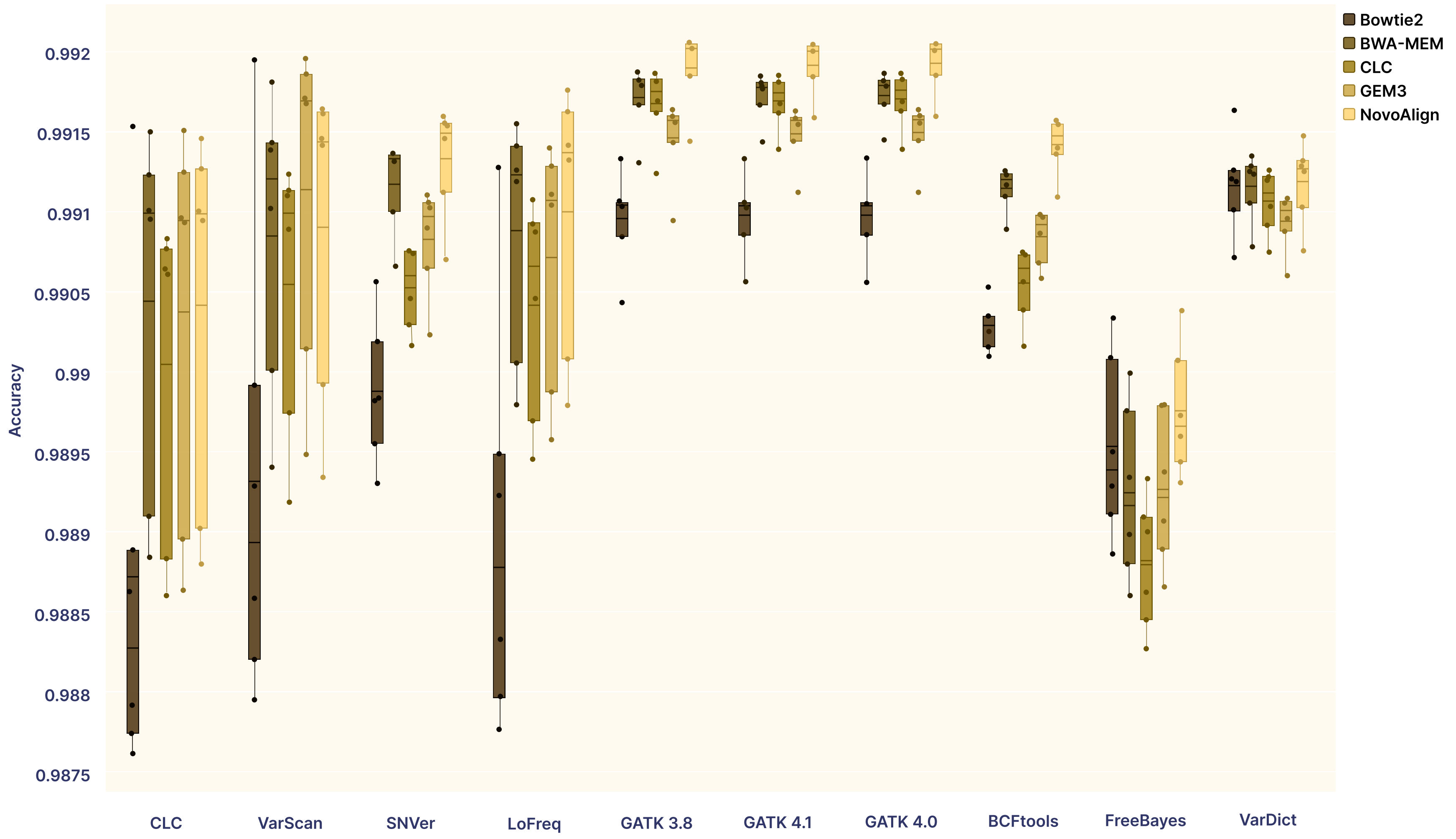
Pipeline for INDELS	AUC
🟨 novoAlign + GATK	0.8233
🟫 BWA + GATK	0.8087

The Benchmark for High-Complexity Agri-Genomics

Plant genomes present a unique challenge in bioinformatics due to massive sizes, dense repeat content, and polyploidy, which often cause standard aligners to fail or produce ‘multi-mapping’ errors. These errors create bioinformatic noise that directly compromises the integrity of SNP calling and structural variant analysis.

novoAlign overcomes these hurdles by replacing simple heuristics with a sophisticated probabilistic mapping algorithm designed to identify the true origin of every read, even within highly similar sub-genomes. By prioritizing accuracy and sensitivity over the speed-focused shortcuts of general mappers, novoAlign provides the essential, high-precision foundation required for reliable agri-genomic research.

Peak GATK Accuracy in *Arabidopsis* study



99.2%

ALIGNMENT ACCURACY

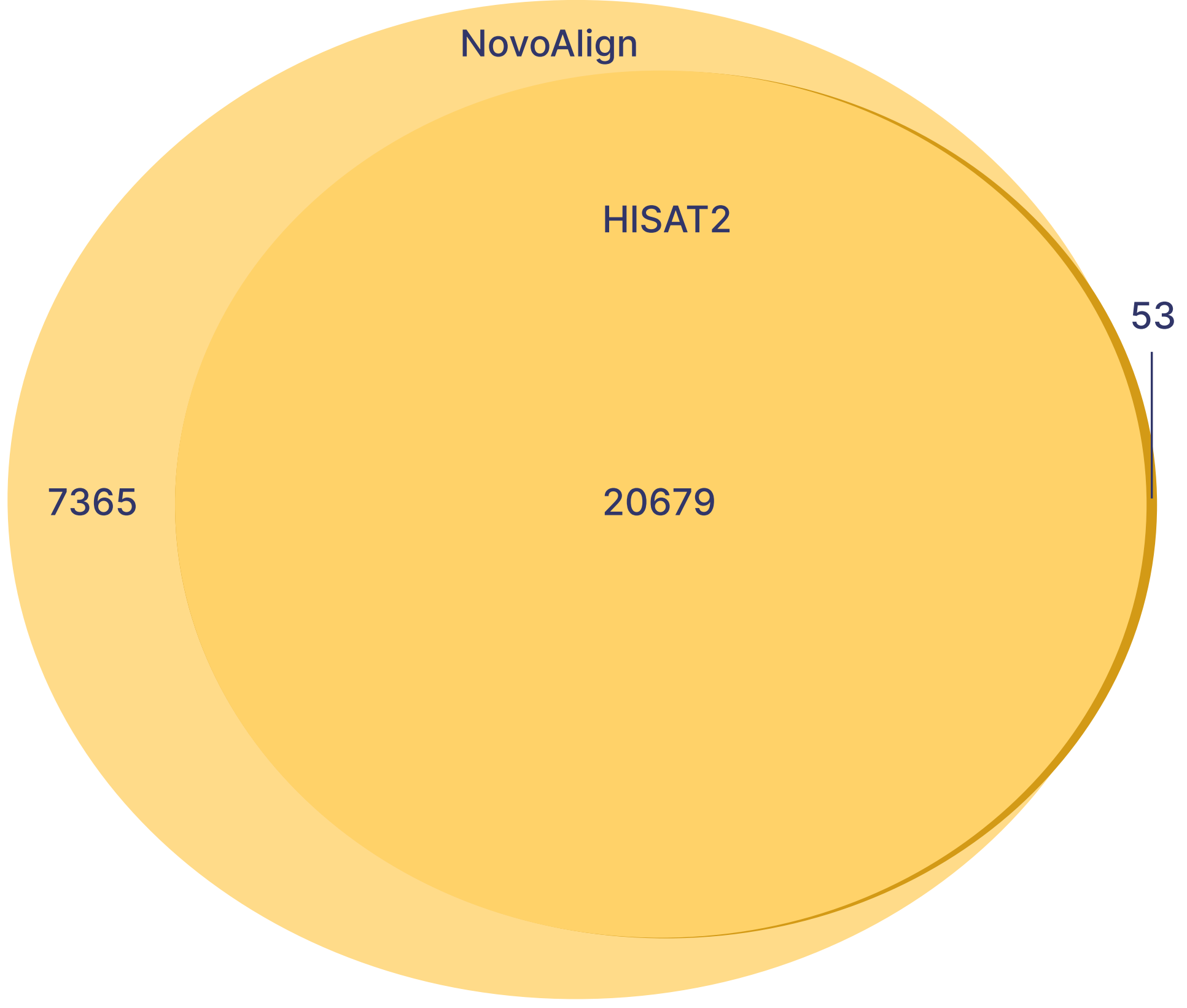
0.604

F-VALUE METRICS

A comparative study of the ***Arabidopsis thaliana*** NGS data was done. It shows that GATK worked best on alignments produced by novoAlign. It reached the best (median) results with respect to accuracy (0.922) compared to other aligners, and had the best mean performance with respect to the F1 score.

- Schilbert et al., Plants (2020)

Maximized Unique RNA-Seq Reads in Sugarcane study



82.0%

UNIQUE MAPPING

In the complex, high-polyploidy genome of *Saccharum*, novoAlign significantly outperformed HISAT by uniquely mapping 82% of genes. This ensures higher data retention and more reliable expression profiles in difficult-to-map grass species.

- Zhu et al., Euphytica (2018)

Doubled SNP Discovery Yield in *Sorghum* study

Pipeline	Total candidates	Recall (Sensitivity)	Confirmation rate
novoAlign + SAMtools	283,000	91%	82%
SOAP2 + simple filter	155,000	77%	79%

A comparative study demonstrated that Novoalign nearly doubled the yield of actionable SNPs (283K vs. 155K) compared to SOAP2-based pipelines, while maintaining a superior 91% recall rate for true biological variants.

- Nelson et al., BMC Genomics (2011)

Automated Precision via Hard-Clipping in 17GB Wheat study

Stringency (Min. Reads)	Novoalign (Total EMS)	BWA-MEM (Total EMS)	Novoalign Advantage
3 Reads (Standard)	2,669	2,223	+20% more SNPs
5 Reads (Moderate)	995	896	+11% more SNPs
8 Reads (Strict)	619	587	+5% more SNPs

Navigating hexaploid complexity, Novoalign delivered 20% more high-confidence SNPs than BWA-MEM. By integrating automated hard-clipping of poor-quality bases and adapter sequences, Novoalign resolves sub-genomic ambiguities that standard mappers often miss, eliminating the need for manual pre-processing.

- King et al., PLoS One (2015)

novoAlign in Plant & Agri-Genomics

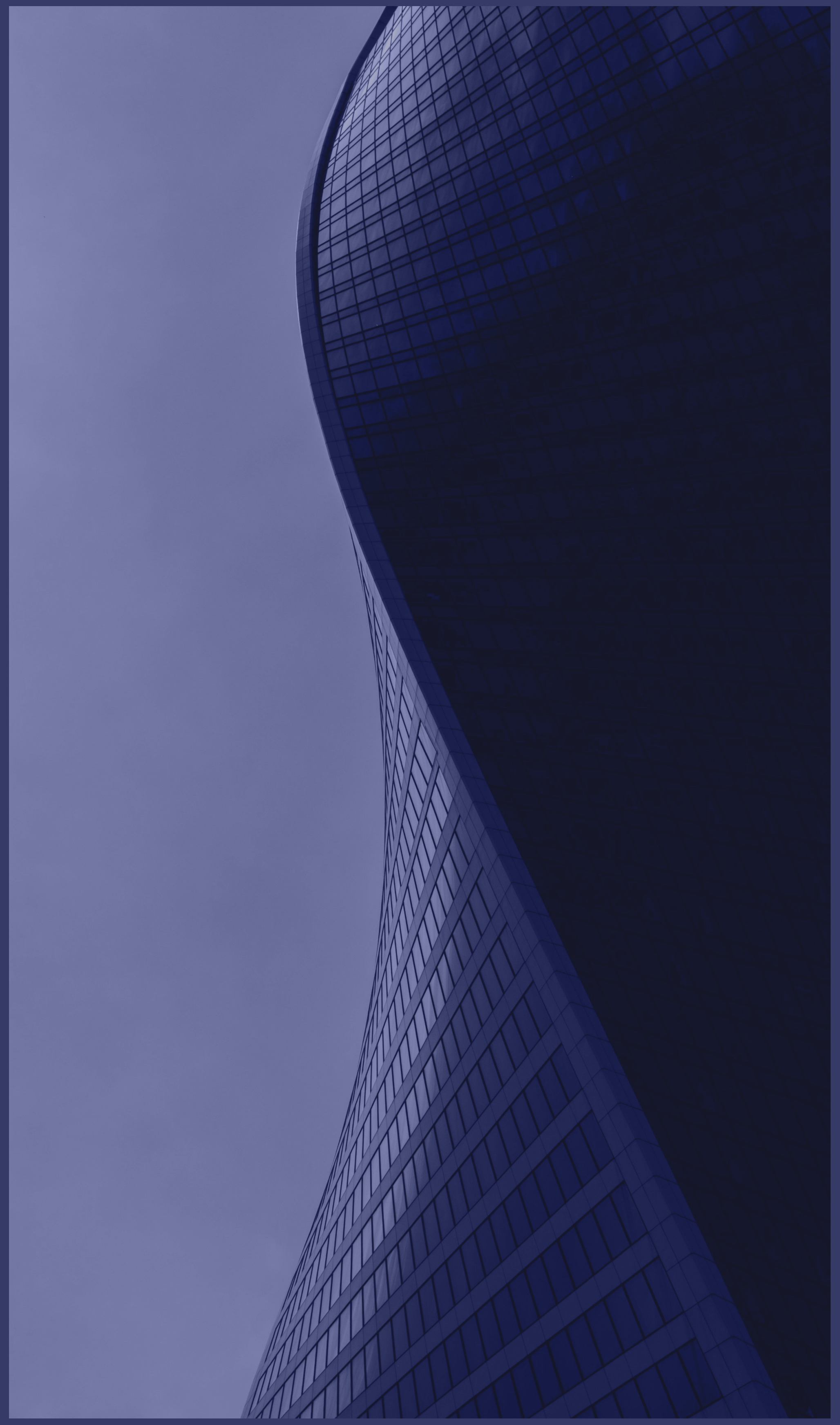
Novoalign has been applied across diverse plant species and data types, from chloroplast and nuclear genome assembly to RNA-seq, exon capture, and variant discovery in major crops.

Paper	Organism/data	Application
Redwan et al. 2015, BMC Plant Biol 15:196	*Ananas comosus* (MD-2 pineapple)	Chloroplast genome assembly
Redwan et al. 2016, DNA Res. 23(5):427–439	*Ananas comosus* (MD-2 pineapple)	Draft nuclear genome assembly
Sass et al. 2016, PeerJ 4:e1584	*Zingiberales* (8 plant families) exon-capture phylogenomics	Phylogenomics / evolutionary history
Souza et al. 2022, Chem. Biol. Technol. Agric. 9(1):29	*Zea mays* (maize) root RNA-seq	Transcriptomics of plant hormone crosstalk (humic acids)
Wu et al. 2018, Nat. Commun. 9:4734	*Solanum lycopersicum* (tomato) resequencing (Yellow Pear, Gold Ball Livingston)	Crop trait discovery (fruit morphology)
Rodriguez et al. 2018, Front. Plant Sci. 9:522	*Zea mays* (maize) stem under *Sesamia nonagrioides* attack (RNA-seq)	Biotic stress transcriptomics (plant–insect interaction)
Xin et al. 2013, Tree Genet. Genomes 9(5):1343–1349	*Vitis vinifera* (grapevine) resequencing	Fruit quality / sugar metabolism genetics
Qi et al. 2013, Nat. Genet. 45(12):1510–1515	*Cucumis sativus* (cucumber) resequencing	Domestication & diversity genomics

Other unsurpassed features

- Bisulphite alignment mode for analysis of methylation status
- Automatic base quality calibration
- Handles single end and paired end reads up to 950bp/read
- In built adapter trimming and base quality trimming
- Option for amplicon primer trimming
- Mapping with base quality values
- Alignment quality scores using posterior alignment probability
- Paired end alignment
- Mismatches and gaps of up to 50% of read length
- Use of ambiguous codes in reference sequences can be used to reduce allelic bias

- Bisulphite alignment mode for analysis of methylation status



For diverse applications

- Whole genome sequence
- Whole exome sequence
- Human Leukocyte Antigen (HLA) Typing
- Amplicon
- RNA
- miRNA

For diverse research areas

- Genomics
- Transcriptomics
- Metagenomics
- Pharmacology
- Cancer research
- Human genetics
- Plant genomics
- Microbiology

Market’s leading aligner, currently licensed by over 100 institutions, research centers, universities and private corporations.

NOVOCRAFT
TECHNOLOGIES SDN BHD

BIOTECHCORP
BIONEXUS
STATUS COMPANY

For more information

NOVOCRAFT TECHNOLOGIES SDN. BHD.
C-23A-05, 3 Two Square, Section 19, 46300, Petaling Jaya, Selangor, Malaysia.

✉ office@novocraft.com ☎ +603 7960 0541 🌐 www.novocraft.com