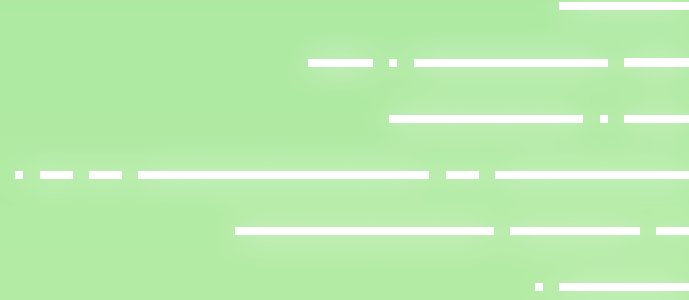


Comparative Benchmarking of MiXCR and other VDJ analysis tools



Key findings

- MiXCR offers the widest range of functionalities from a single tool, including a built-in reference for a large number of species and single cell analysis capabilities.
- MiXCR has the fastest processing time – up to 6x faster in some cases.
- MiXCR is the most accurate tool on all tested datasets.

Introduction

This document benchmarks MiXCR, Immcantation and TRUST4, well known tools for preprocessing VDJ sequencing data. VDJ sequencing facilitates the investigation of diverse antigen receptor repertoires within T and B cells, providing insights into the adaptive immune response. Due to the complexity and variability of VDJ data, selecting the processing software that is both efficient and accurate is essential for effective data interpretation and the success of immunological studies. This work demonstrates the predominance of MiXCR in terms of functionality, speed and accuracy, particularly under challenging conditions.

Functionalities comparison

In this section, we explore the specific features and capabilities of each software in our benchmarking study. The table below offers a snapshot of how MiXCR compares to other VDJ data processing tools.

	MiXCR	Immcantation	TRUST4	IgBlast	IMGT
Built-in reference library	✓	✗	✗	✗	✓
Targeted TCR/BCR data	✓	✓	✓	✓	✓
Bulk RNA-seq/Exome-seq data	✓	✗	✓	✗	✗
Single-cell data	✓	✗	✓ no noise filters	✗	✗
Incomplete rearrangements	✓	✗	✗	✗	✗
PCR/Sequencing error correction	✓	✓	✗	✗	✗
UMI based error correction	✓	✓	✗	✗	✗
Fine tuning for low quality data	✓	✓	✗	✗	✗
QC text reports and plots	✓	✓	✗	✗	✗
Downstream analysis	✓	✓	✗	✗	✗
Lineage trees reconstruction	✓	✓	✗	✗	✗

MiXCR offers the most comprehensive set of functionalities in a convenient single tool, including single cell and RNA-seq analysis capabilities. MiXCR also supports a broad range of species with its unique built-in reference, which is continuously updated. Additionally, MiXCR is compatible with a wide range of technologies. It includes optimized presets for numerous public protocols and commercially available bulk TCR/BCR-seq kits. Regarding single-cell sequencing, it can be used with broadly used technologies such as 10x Genomics, Parse Biosciences, BD Rhapsody.

Run time comparison

Preprocessing and downstream analysis of raw VDJ data can be time consuming, especially when large datasets are involved. In order to assess the run time of the tools we are comparing datasets of different sizes—from small to large—were identically processed with MiXCR, Immcantation and TRUST4. Immcantation toolset was run using the Airrflow pipeline. IgBlast and IMGT are excluded from the subsequent analysis as they lack the downstream features necessary for comparison with the other tools.

In all tested datasets, MiXCR was the fastest performer by far, even up to six times faster in some cases (Figure 1).

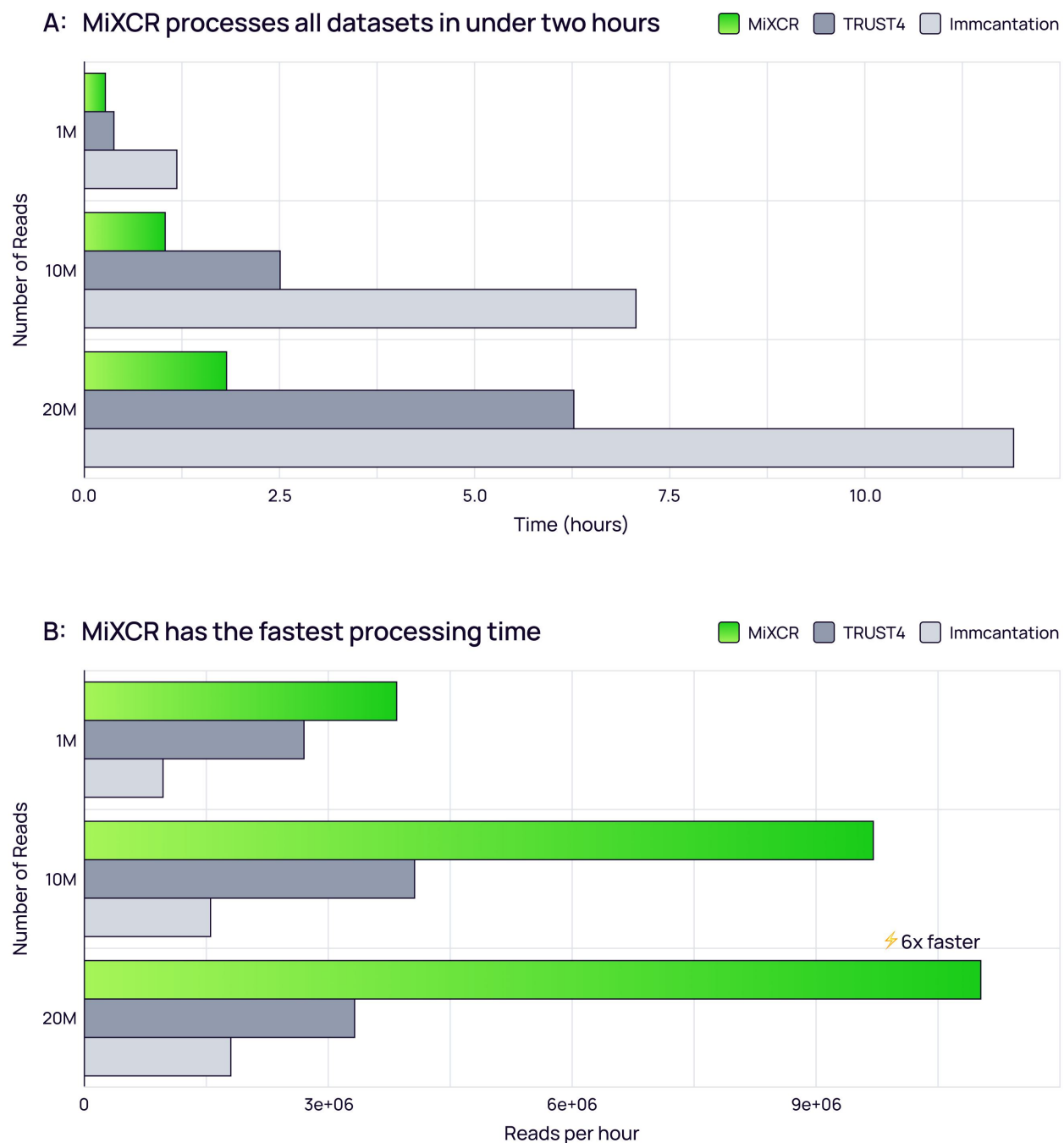


Figure 1: The barplots illustrate A) the execution times and B) the processing speed of MiXCR, Immcantation (implemented with Airrflow), and TRUST4 when analyzing standard bulk TCR sequencing data with UMIs. The tools were tested on datasets of varying sizes, containing 1 million, 10 million, and 20 million reads. Data obtained from computations were performed on a server with 24 CPU cores and 128 GB of RAM.

Importantly, the speed difference between the compared tools is considerable, especially in the largest dataset containing 20 million reads. This highlights the amount of time that can be saved using MiXCR.

Performance comparison

To assess the accuracy of the benchmarked tools, we evaluated the VDJ sequences that were identified in three simulated UMI containing datasets with varying clonal abundances and increasing frequencies of sequencing errors.

Results of this comparison are shown in Figure 2. From the initial data point in all three plots, where the dataset contains no errors, MiXCR demonstrates greater sensitivity than Immcantation and TRUST4 under baseline conditions. Importantly, this trend persists as errors are introduced into the data, with MiXCR consistently outperforming the other two tools.

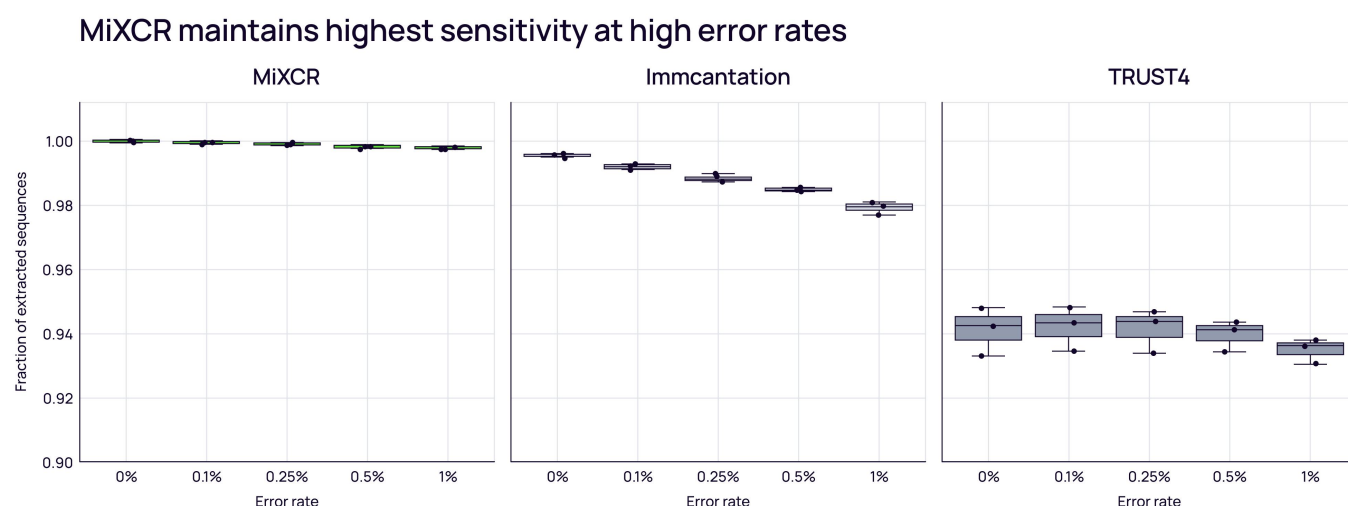


Figure 2: The boxplots illustrate the sensitivity of MiXCR, Immcantation, and TRUST4 on simulated data with UMIs. Sensitivity was determined by calculating exact VDJ sequence matches to the true repertoires.

The ability of MiXCR to effectively handle sequencing errors is crucial when working with biological data, where errors are commonly encountered. The ability to manage these errors can significantly influence the outcomes of a study, affecting both the identification of crucial clones and the accurate assessment of clonal diversity.

To further explore performance differences, we conducted an analysis of hybridoma datasets. In this analysis, we expect to detect only a small number of clones, indicative of the monoclonal nature of hybridoma cell lines, with each line originating from a single B cell clone. While generally stable, these original clones may undergo limited somatic hypermutations, potentially introducing slight increases in diversity.

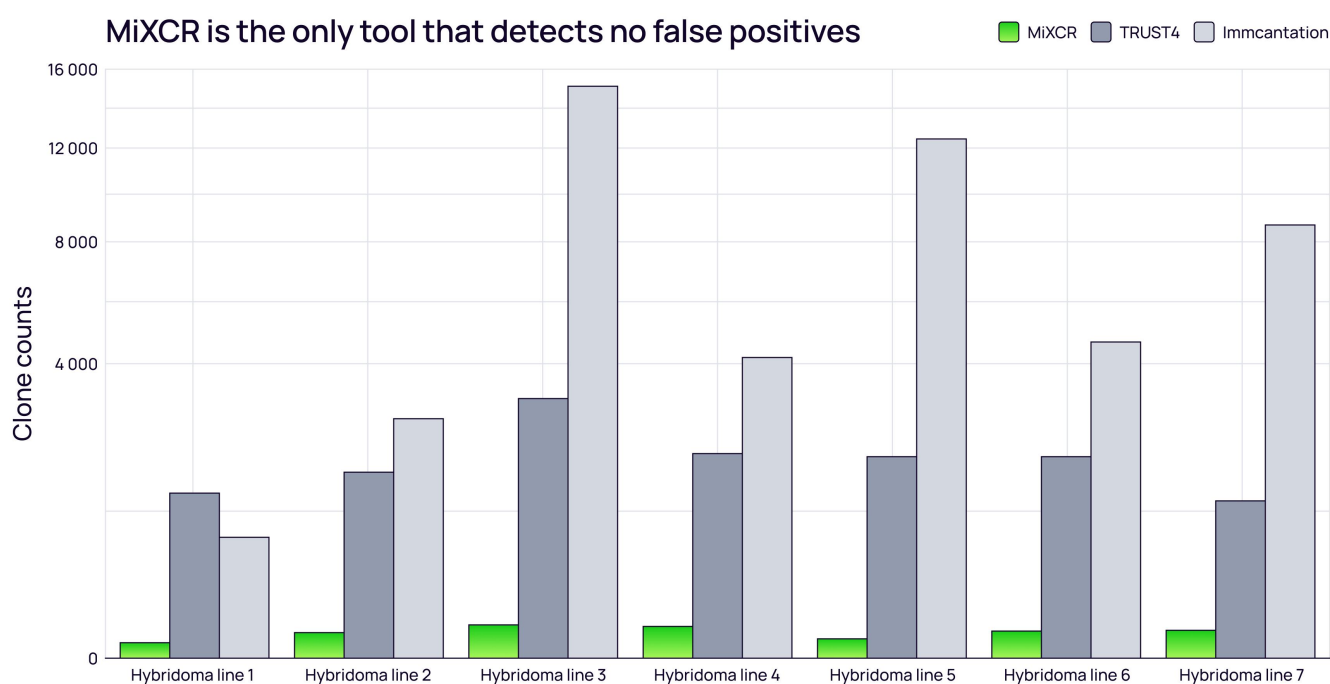


Figure 3: The bar plot illustrates the number of clones identified from sequencing data across seven different cell lines using MiXCR, TRUST4, and Immcantation tools. Please note the scale of the y-axis is in square root to ensure visibility of the MiXCR bar.

As expected given the monoclonal nature of the datasets, MiXCR detected a small number of clones (Figure 3). In contrast, TRUST4 identified approximately 20 times more clones than MiXCR, while Immcantation detected between 100 and 200 times more clones. This disparity illustrates how reduced accuracy can result in a significant number of false positives, thereby affecting the biological conclusions drawn from the data.

Conclusion

MiXCR has demonstrated unparalleled performance in the preprocessing of VDJ sequencing data, distinguishing itself as the best performing platform. MiXCR offers the most comprehensive set of functionalities, supports a broad range of species, and is compatible with various sequencing technologies. This flexibility allows researchers to streamline single cell and RNA-seq analyses within a single tool, eliminating the challenges of using multiple software platforms and simplifying their workflows.

The processing speed of MiXCR is unmatched, consistently surpassing that of Immcantation and TRUST4 across all dataset sizes. Its ability to process data at 6x the speed in some cases allows for enhanced productivity, particularly in large-scale projects where time is a critical factor.

Furthermore, MiXCR's precision to accurately identify clones, even in datasets with introduced errors, underscores its reliability. This accuracy is crucial in biological studies where the correct assessment of clones directly influences the validity of the scientific conclusions. In scenarios like the analysis of hybridoma datasets, where the expected clone diversity is low, MiXCR's precise detection of a small number of clones prevents the generation of false positives—an issue evident with the other tools tested, which reported excessively higher numbers of clones.

In conclusion, MiXCR not only excels in its technical and functional capacities but also enhances the overall efficiency and reliability of immunological research. Its robust performance across multiple facets—functionality, speed, and accuracy—confirms that MiXCR is an essential tool for researchers dedicated to advancing the field of immunology.