



# Enhanced usability in the MSstats family of statistical analysis software for quantitative mass spectrometry(MS)-based proteomics



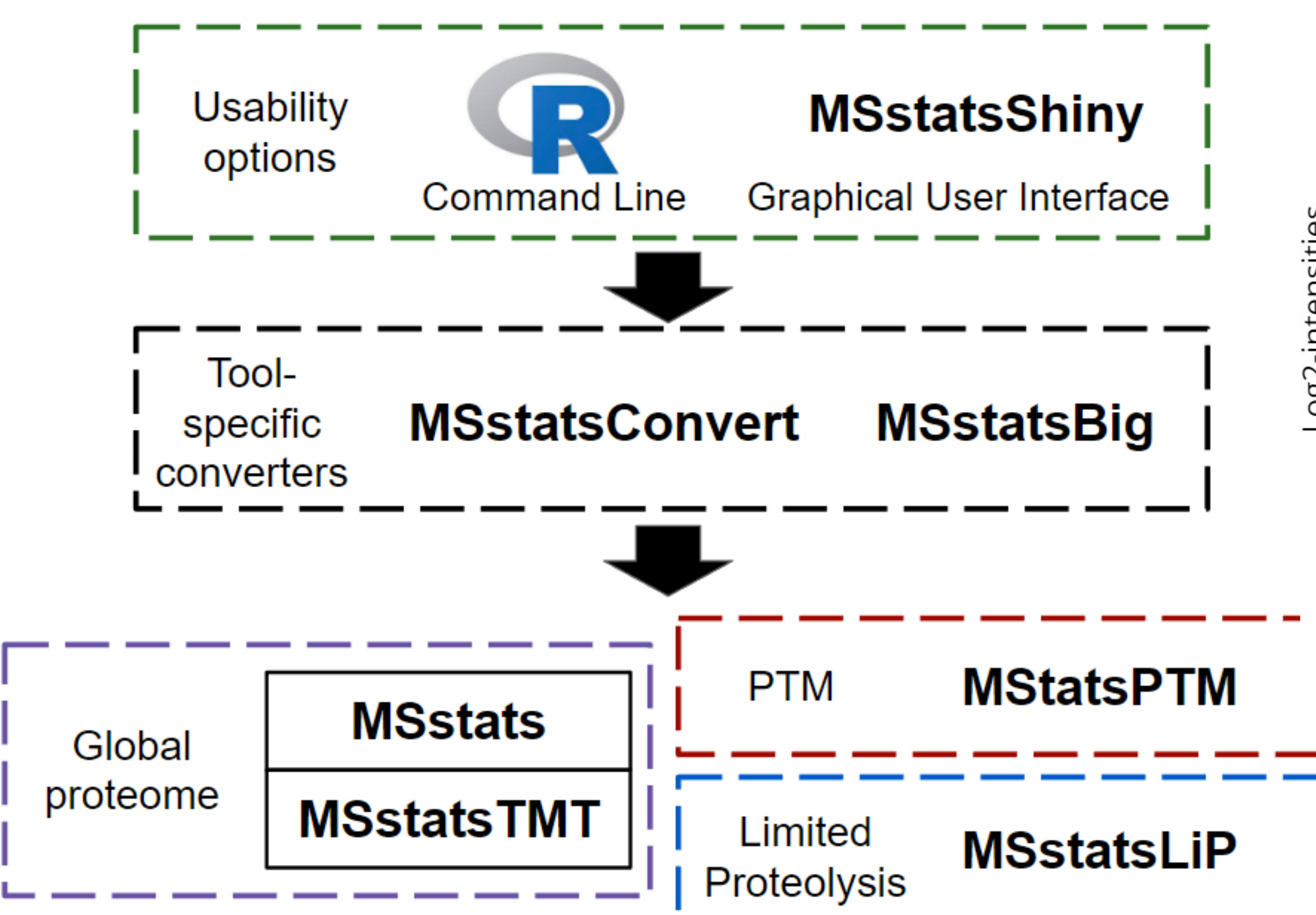
Devon Kohler<sup>1,2</sup>, Mateusz Staniak<sup>3</sup>, Deril Raju<sup>1</sup>, Sarah Szvetecz<sup>1,2</sup>, Anthony Wu<sup>1,2</sup>, Olga Vitek<sup>1,2</sup>

<sup>1</sup>Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA

<sup>2</sup>Barnett Institute for Chemical and Biological Analysis, Northeastern University, Boston, MA, USA

<sup>3</sup>University of Wrocław, Wrocław, Poland

## MSstats: Versatile software for differential analysis of MS-based proteomics experiments



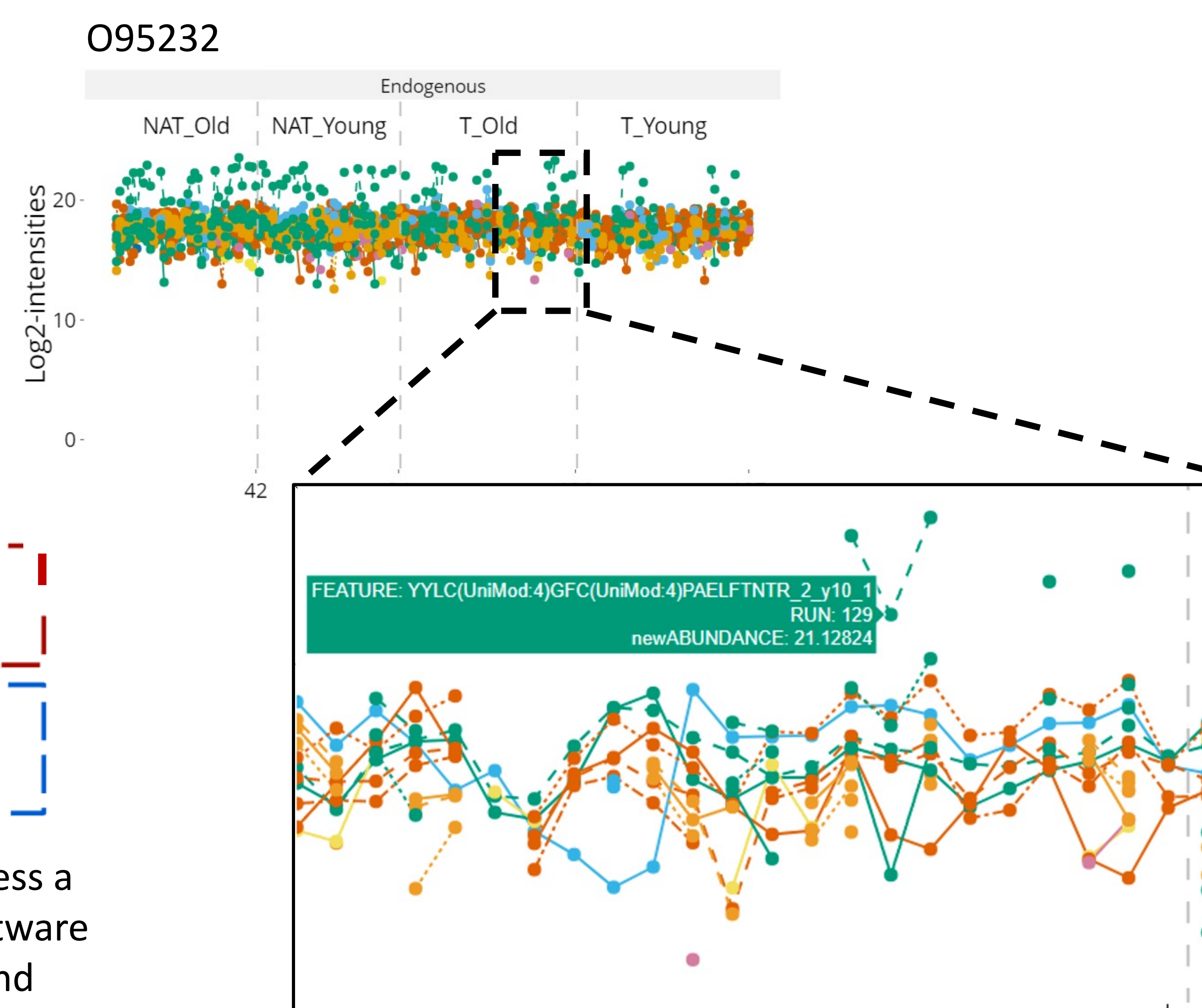
MSstats is a family of statistical analysis software which can address a variety of biological questions and experimental designs. The software integrates with spectral processing tools used for identification and quantification and can be run via the command line or a GUI.

## New converters integrate directly with spectral processing tools, reducing cumbersome manual work

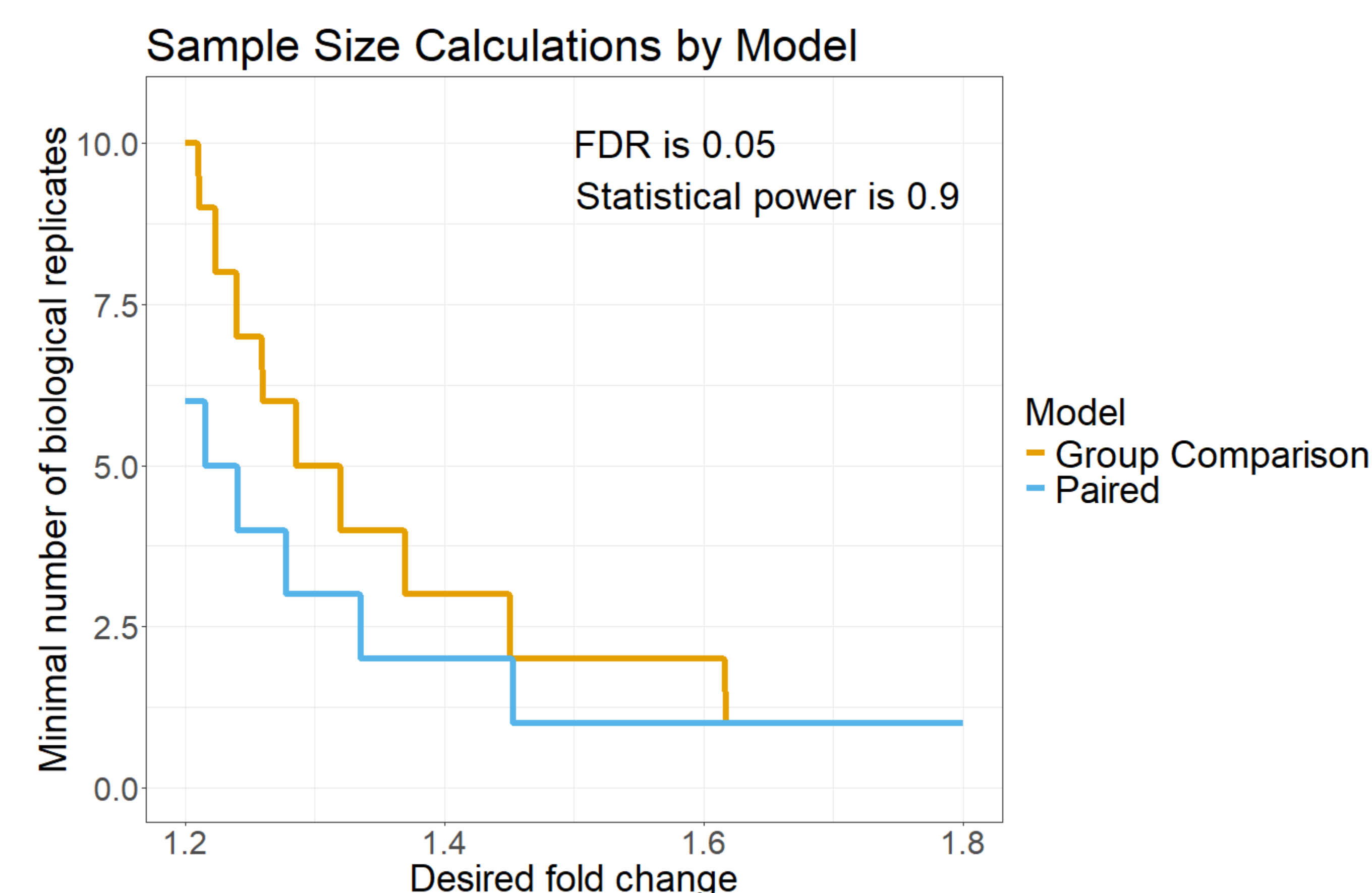
Dedicated Converter	LF	TMT	PTM	LiP
FragPipe	✓*	✓	✓	
Skyline	✓		✓	✓
Spectronaut	✓*		✓	✓
MaxQuant	✓	✓	✓	
Proteome Discoverer	✓	✓	✓	
DIA-NN	✓		✓	
OpenMS	✓	✓		
MetaMorpheus	✓		✓	
Progenesis	✓		✓	
SpectroMine		✓		
OpenSWATH	✓			
PEAKS Studio			✓	
DIA-Umpire	✓			

\*Includes MSstatsBig Functionality

## GUI data visualizations enhance the analysis of large experiments via increased interactivity

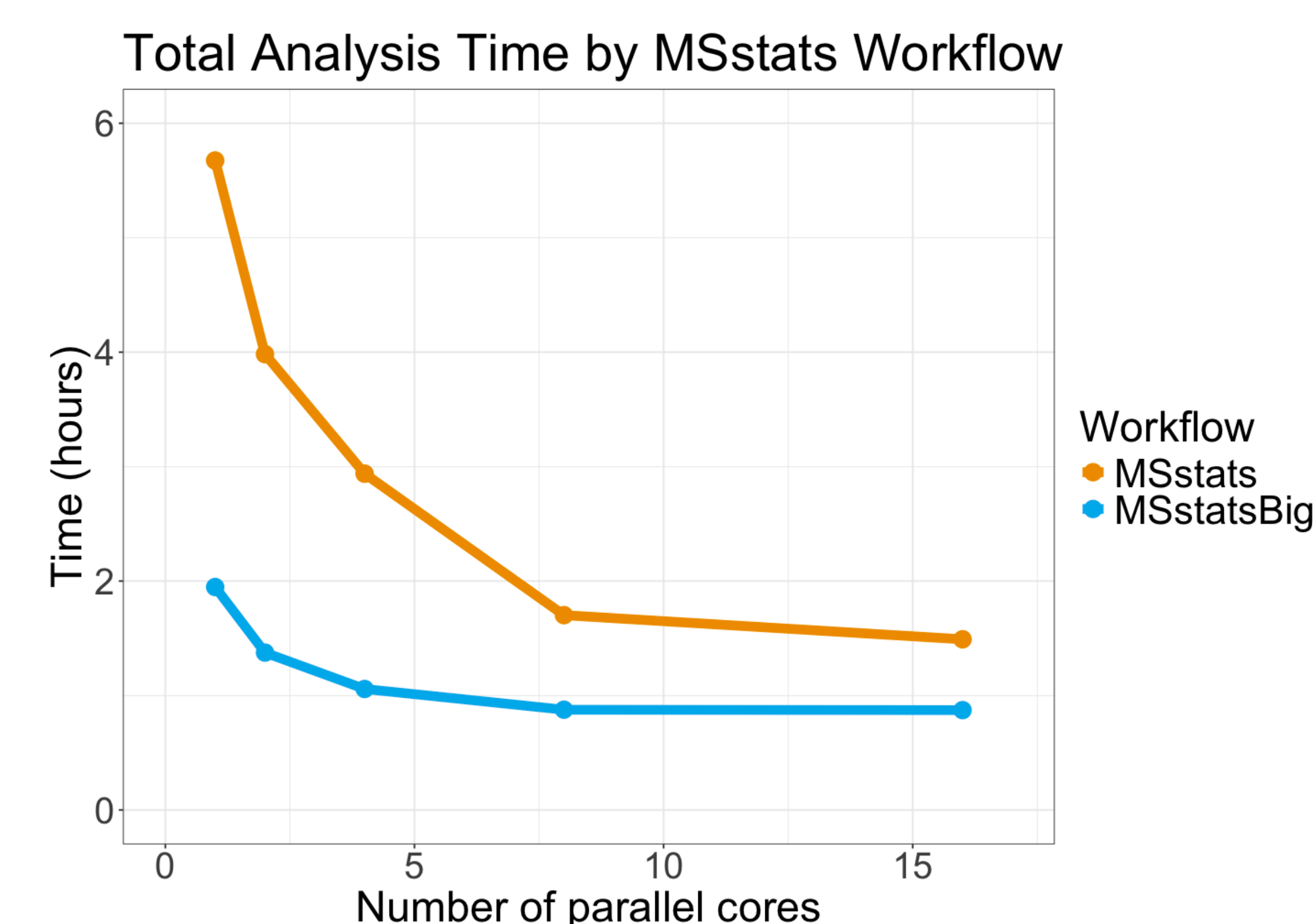


## New TMT sample size calculations provide increased insight into experimental power<sup>1</sup>



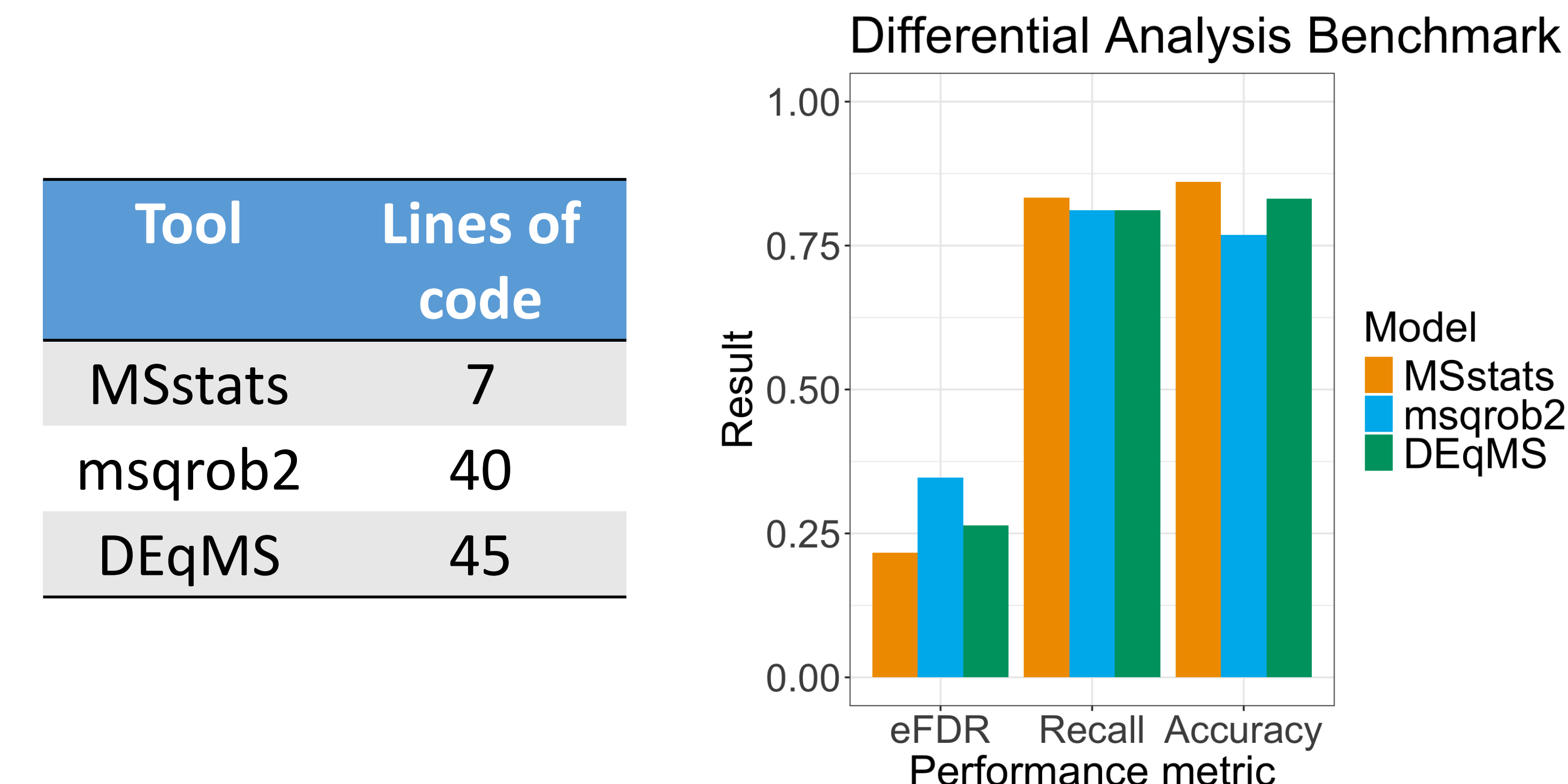
$$J = \frac{2(\sigma + \sigma_S + \sigma_T)}{\Delta^2} ; \alpha = (1 - \beta) \frac{q}{1 + (1 - q)m_0/m_1}$$
$$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(z_{1-\alpha/2} + z_{1-\beta})^2}$$

## MSstatsBig and parallel processing greatly speed up workflows on large multi-gigabyte PSM files



All benchmarks were run using a large (187 sample) DIA experiment by Clark et al.<sup>2</sup> Leveraging MSstatsBig and multiple cores reduced the analysis time by 83% over the standard MSstats workflow.

## MSstats analysis pipeline enables high quality analysis without major implementation burden



Comparisons were run using a DDA LFQ benchmark dataset from Van Puyvelde et al.<sup>3</sup> The experiment included two mixtures with differing levels of human, yeast, and E.coli samples.

## References

- Bojkova et al. Nature 583, 469-472 (2020).
- Clark et al. Cell 179, 964-983 (2019).
- Van Puyvelde et al. Sci Data 9, 126 (2022).



The authors declare no competing financial interest