

# TIMS DIA-NN 3.0: Improved Algorithms to enhance peptide detection with confidence and accuracy

Robin Park<sup>1</sup>, Yani Zhao<sup>2</sup>, Dennis Trede<sup>2</sup>, Tharan Srikumar<sup>3</sup>, Tomasz Siwarga<sup>2</sup> and Jonathan Krieger<sup>3</sup>,

<sup>1</sup>Bruker Scientific LLC, Billerica MS, USA; <sup>2</sup>Bruker Daltonics GmbH & Co. KG, Bremen, Germany; <sup>3</sup>Bruker Ltd., Milton, ON; <sup>4</sup>Bruker Switzerland AG, Faellanden, Switzerland

## Introduction:

Due to the rising popularity of data-independent acquisition (DIA), we previously developed and launched an automated spectral library generation tool and the CCS-aware TIMS DIA-NN 2.0 for DIA analysis, for the purpose of analyzing high-throughput mass spectrometry data. Subsequent to this initial launch, we have continuously developed TIMS DIA-NN 3.0 by redesigning and implementing new algorithms. The primary focus in this iteration has been on substantially improving the confidence, accuracy, and reliability of both identification and quantification processes.

In processing raw files, the TIMS DIA-NN effectively identifies local maxima by clustering and merging peaks, employing a 4D-Proteomics approach. This method integrates key dimensions - collision cross-section (CCS), retention time, mass-to-charge ratio (m/z), and MS/MS fingerprinting - to optimize peak detection. Enhancements have been made with the introduction of multiple new algorithms, including new mass drift and mass tolerance calibration algorithms, which improve the detection of peptide candidates and the precision of chromatogram construction. Additionally, we have integrated novel algorithms, including a Gaussian correlation score for peak detection in neural network analysis, peak boundary detection in chromatograms, and fragment ion selection for MBR to improve reproducibility. These improvements collectively contribute to the improvement of the consistency of the coefficient of variation (CV) across replicates, thereby elevating the robustness and accuracy of the data analysis.

## Methods:

To evaluate the improvements, we first re-processed a multi-species datasets with known ratios between two samples measured in quintuplet in three released versions of the algorithm. This dataset allows the evaluation of the accuracy and precision of the algorithm.

Next, we applied the updated algorithm to a cross-site study. 200ng of K562 lysate (Promega) was analyzed on four timsTOF HT systems in four labs spanning 2 continents. 5min active gradients were used for peptide separation on 5cm Aurora columns (IonOpticks) in conjunction with nanoElute LC systems. Each site was required to acquire at least 10 injections. Raw data were uploaded to common server for processing.

## Results:

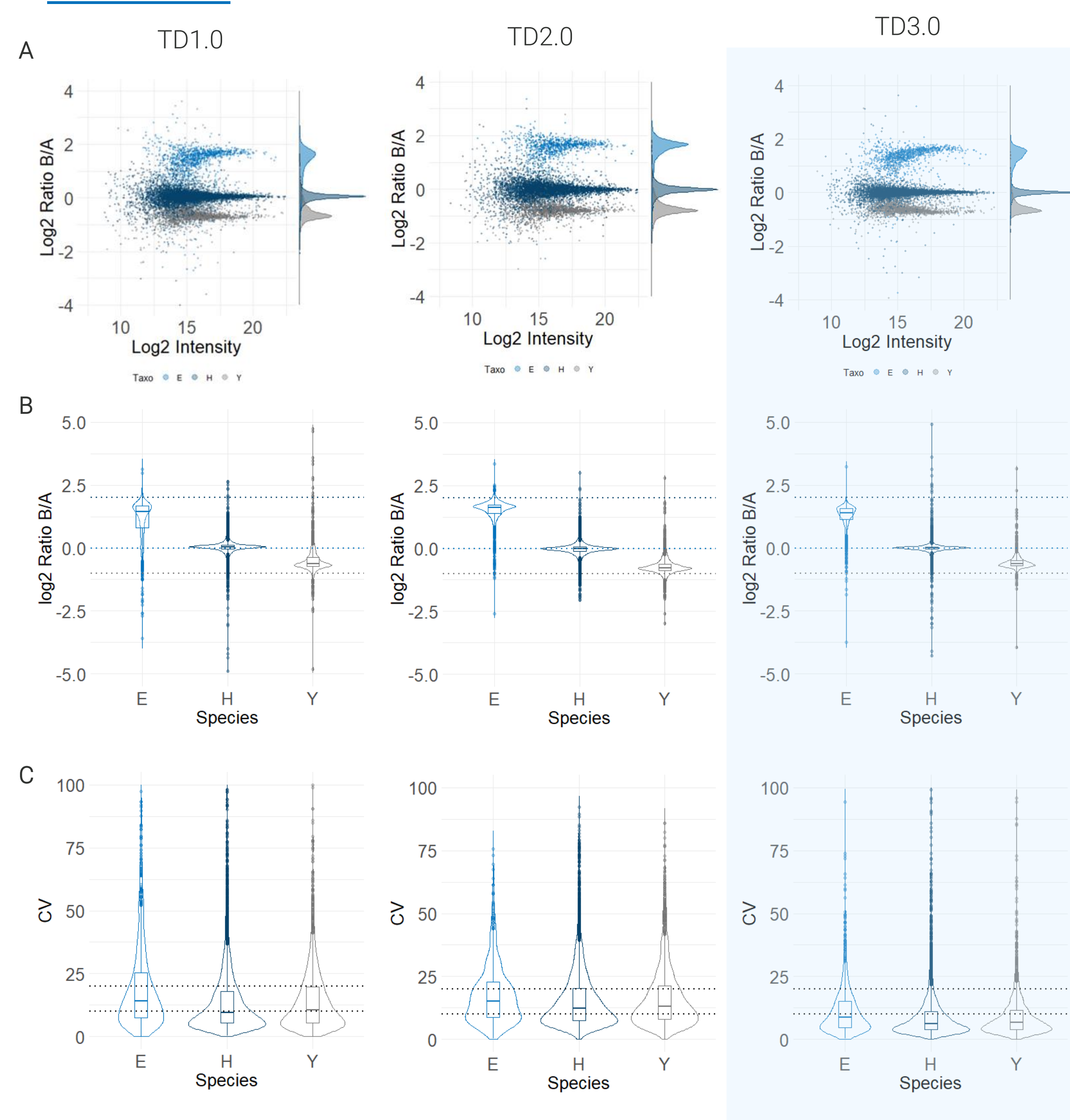
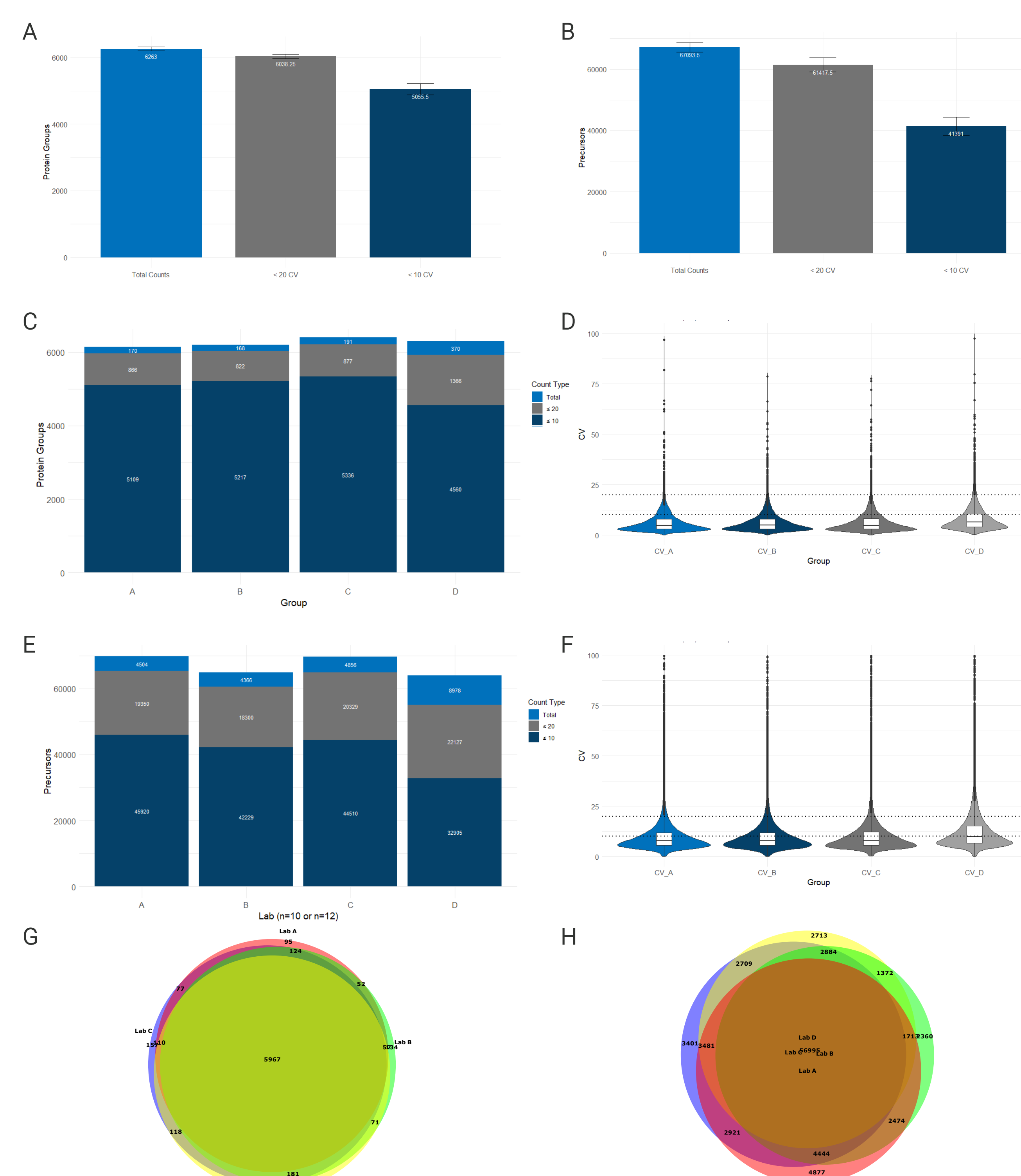


Fig. 1: Iterative improvements in accuracy and precision. A multi-species mixed dataset with known ratio between two samples measured in quintuplet was analyzed with different versions of TIMS DIA-NN. (A) Scatterplot of ratio vs intensity colored by species. (B)  $\log_2$  protein ratio between the two samples for each species and (C) the coefficient of variation in the quintuplet measurements.



TIMS DIA-NN has been iteratively improved since its initial release. We investigated the improvements in accuracy and precision following the integration of a Gaussian correlation score for peak detection in neural network analysis, peak boundary detection in chromatograms, and fragment ion selection for MBR to improve reproducibility. We show that average CV in TD3.0 is <10% (see Fig. 1) and has similar accuracy to other algorithms (data not shown).

We next analyzed a cross-site study employing 5 min active LC gradients and with each site producing a minimum of 10 replicates. On average, >6200 protein groups were identified in each injection, using a library consisting of >500,000 precursors mapping to >13,000 protein groups, without employing MBR. Of the identified protein groups >96% (6038 protein groups) could be quantified with a  $CV \leq 20\%$ , while 80% (5055 protein groups) could be quantified with a  $CV \leq 10\%$ . Similarly, >92% of precursors could be quantified with  $CV \leq 20\%$ . Additionally, the overlap of protein groups and peptides was excellent (see Fig. 2 G&H).

Taken together, timsTOF HT system together with 4D-Proteomics algorithms (such as TIMS DIA-NN, as well as Spectronaut, DIA-NN, Fragpipe, MaxQuant and others) are capable of generating data that leads the field in quantitative performance allowing users to gain true meaningful biological information from their samples.

Fig. 2: Cross-site study on variation. Raw data from four labs were processed with TIMS DIA-NN 3.0 using a human spectral library and without MBR. The average number of protein groups (A) and precursors (B) identified and quantified at  $CV \leq 20\%$  and  $CV \leq 10\%$  across all sites. The number of protein groups (C) and precursors (E) identified and quantified at  $CV \leq 20\%$  and  $CV \leq 10\%$  at each site. The CV for all protein groups (D) and precursors (F) at each site. Overlap of protein groups (G) and precursors (H) was very high.

## Conclusion:

- TIMS DIA-NN 3.0 improves accuracy and precision over previous versions, allowing for reliable quantification of data-independent acquisition (DIA) data.
- TIMS DIA-NN 3.0 was able to quantify >96% of protein groups with a  $CV \leq 20\%$  from a 5min LC gradient.
- TIMS DIA-NN 3.0 has been integrated into the ProteoScope platform.

Technology