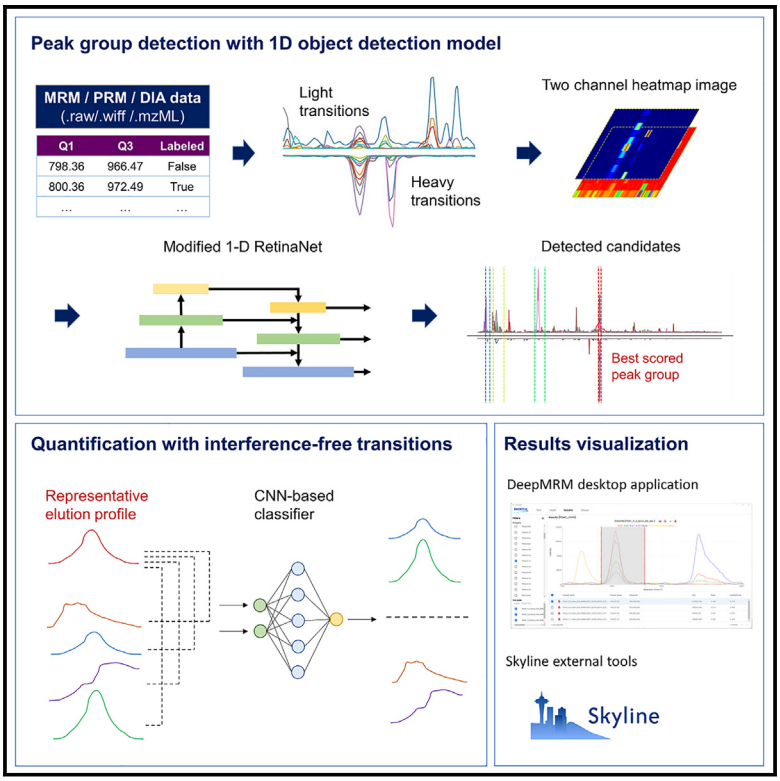
Cell Reports
Methods

Report

Targeted proteomics data interpretation with DeepMRM

Graphical abstract



Highlights

- DeepMRM utilizes AI for object detection in targeted proteomics data interpretation
- DeepMRM outperforms Skyline in quantification accuracy
- DeepMRM shows robust performance across MRM, PRM, and DIA data
- DeepMRM is available as a Windows desktop application and a Skyline external tool

Authors

Jungkap Park, Christopher Wilkins, Dmitry Avtonomov, ..., Brendan X. MacLean, Sang-Won Lee, Sangtae Kim

Correspondence

sangtae.kim@bertis.com

In brief

Park et al. present DeepMRM, a software package leveraging deep learning for object detection to minimize manual intervention in targeted proteomics data analysis. DeepMRM promotes high-throughput analysis and enhances the reproducibility and scalability of targeted proteomics in clinical settings.

Park et al., 2023, Cell Reports Methods 3, 100521
July 24, 2023
https://doi.org/10.1016/j.crmeth.2023.100521

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Report

Targeted proteomics data interpretation with DeepMRM

Jungkap Park,[1] Christopher Wilkins,[2] Dmitry Avtonomov,[2] Jiwon Hong,[3] Seunghoon Back,[3] Hokeun Kim,[3] Nicholas Shulman,[4] Brendan X. MacLean,[4] Sang-Won Lee,[3] and Sangtae Kim[2,5,*]

[1]Bertis, Inc., Seoul 06108, Republic of Korea
[2]Bertis Bioscience, Inc., San Diego, CA 92121, USA
[3]Department of Chemistry, Center for Proteogenomic Research, Korea University, Seoul 02841, Republic of Korea
[4]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[5]Lead contact
*Correspondence: sangtae.kim@bertis.com
https://doi.org/10.1016/j.crmeth.2023.100521

MOTIVATION In clinical proteomics, targeted proteomics approaches like multiple-reaction monitoring (MRM) or parallel-reaction monitoring (PRM) are widely used, but their application is often hampered by labor-intensive and error-prone manual data interpretation. Existing computational methods, while helpful, often demand a substantial degree of manual intervention or decoy-transition approaches, constraining the throughput and efficiency of clinical proteomics assays. To address these challenges, we developed DeepMRM, a targeted proteomics data interpretation package that facilitates high-throughput analysis and enhances the reproducibility and scalability of targeted proteomics in clinical settings.

SUMMARY

Targeted proteomics is widely utilized in clinical proteomics; however, researchers often devote substantial time to manual data interpretation, which hinders the transferability, reproducibility, and scalability of this approach. We introduce DeepMRM, a software package based on deep learning algorithms for object detection developed to minimize manual intervention in targeted proteomics data analysis. DeepMRM was evaluated on internal and public datasets, demonstrating superior accuracy compared with the community standard tool Skyline. To promote widespread adoption, we have incorporated a stand-alone graphical user interface for DeepMRM and integrated its algorithm into the Skyline software package as an external tool.

INTRODUCTION

In clinical proteomics laboratories, targeted proteomics approaches such as multiple-reaction monitoring (MRM; also known as selected-reaction monitoring [SRM]) or parallel-reaction monitoring (PRM) are widely employed due to their high sensitivity and reproducibility.[1–3] However, researchers often devote substantial time to manual peak selection, interference identification, and peak area adjustments to interpret targeted proteomics data.[4,5] This step is prone to human error and may result in inconsistent outcomes due to subjective evaluations. The reliance on manual inspection is a major barrier to the transferability, reproducibility, and scalability of targeted proteomics in clinical applications.

Numerous computational methods have been developed to address these challenges. Skyline, the most widely used open-source software for targeted quantitative proteomics, offers a graphical user interface (GUI) with a suite of tools for peak picking, peak integration, and quantitative analysis of the results.[5,6] For automated absolute quantification, an external calibration curve method was developed.[7] For quality assessment and control, a decoy-transition approach and machine learning models have been implemented.[4,8,9] Deep learning algorithms have recently been applied to the peak detection problem and have demonstrated a substantial improvement in accuracy compared with conventional methods.[10–12] However, existing approaches still necessitate substantial manual intervention or reliance on a decoy-transition approach,[4] which ultimately constrains the assay's throughput.

We introduce DeepMRM, a targeted proteomics data interpretation package leveraging deep learning algorithms for object detection specifically designed to substantially reduce the manual inspection burden, even for noisy and complex data (Figure 1). By reframing the challenge of finding peaks for targeted peptides as an object detection task, DeepMRM detects instances of peak groups within a set of 1D chromatograms, akin to how traditional object detection models identify object instances of a specific class within a 2D image. DeepMRM accepts transition

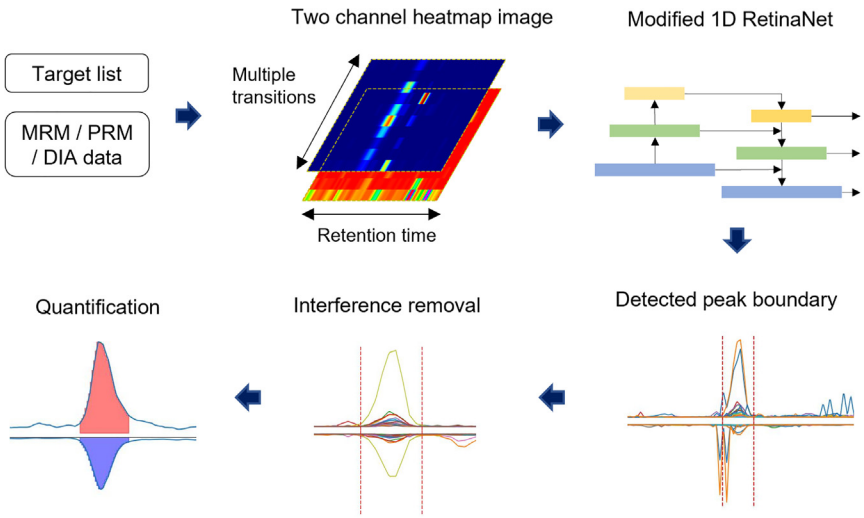


Figure 1. DeepMRM workflow for detecting peak groups of targeted peptides

Given the target list and MRM/PRM/DIA data as input, transition chromatograms of both heavy and light peptides are fed into the model as 2-channel heatmap images. Multiscale 1D feature maps are extracted and processed by two subnetworks: a classifier for determining whether candidate peak groups are present and a regressor for detecting the boundary of candidate peak groups. Detected peak groups are then examined by a CNN-based transition classification model to select transitions unaffected by interference or noise. The abundances of target peptides are estimated based on peak areas of heavy and light peptides.

chromatograms of varying lengths and numbers as input, generating chromatographic peaks of targeted peptides, along with their abundances and confidence scores. Detection results and transition chromatograms are visualized in a GUI, enabling users to review and adjust the results as needed rapidly. The performance of DeepMRM was benchmarked against Skyline and validated using internal and public datasets generated by different acquisition methods, such as MRM, PRM, or data-independent acquisition (DIA) experiments. DeepMRM facilitates high-throughput analysis of targeted proteomics by automated quantification and quality control without requiring decoy transitions.

RESULTS

Overview of DeepMRM

DeepMRM is a robust and highly accurate peak detection model designed for interpreting targeted proteomics data, complemented with a user interface for visualizing detection results. DeepMRM takes MRM, PRM, or DIA data and a target list as input, generating peaks corresponding to targeted peptides, their abundances, and confidence scores. The workflow consists of four steps (Figure 1; see STAR Methods). First, for each targeted peptide, transition chromatograms of both heavy (stable isotope labeled [SIL]) and light (endogenous) peptides are extracted and converted into 2-channel heatmap images. Second, a deep learning-based peak detection model extracts 1D feature maps at various scales and identifies peak group candidates. Third, another convolutional neural network (CNN)-based transition classification model examines individual pairs of heavy and light transitions in the detected peak groups. Transitions unaffected by interference or noise are selected for calculating the abundances (peak areas) of target peptides. Lastly, quantification results are visualized alongside transition chromatograms in a GUI, allowing users to swiftly examine candidate peak groups and transitions selected for quantification and to make adjustments to their boundaries if necessary (Figure S1).

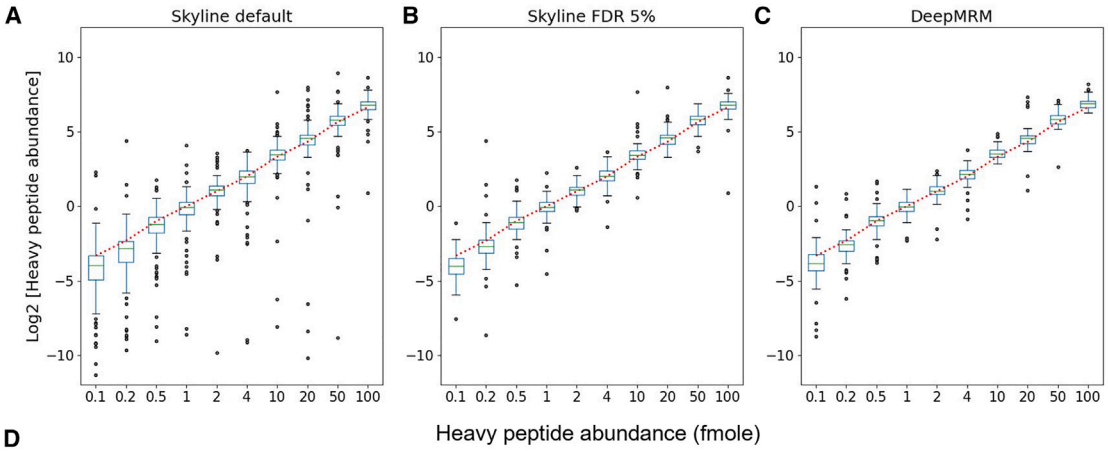
DeepMRM comprises two neural network models: a peak detection model for detecting peak groups and a transition classification model for selecting interference- and noise-free transitions. The peak detection model features a modified architecture of RetinaNet,[13] a popular neural network model for object detection. DeepMRM can accept transition chromatograms of any length and number as input as a fully convolutional network. For the backbone convolutional architecture, we modified ResNet-18 to better learn relationships between transitions of the same peptide, as well as heavy and light peptides (see STAR Methods). Additionally, to improve model generalization across different target peptides and experimental settings unexplored during training, data augmentation schemes such as resizing, cropping, retention time shifting, and intensity jittering were incorporated during training. The transition classification model employs the same backbone architecture with different kernel sizes.

To train and evaluate DeepMRM, we used liquid chromatography (LC)-MRM-mass spectrometry (MS) data derived from tumor tissues of 66 patients with pancreatic ductal adenocarcinoma (PDAC)[14] (see STAR Methods). The MRM assay targeted 153 endogenous peptides and their heavy SIL peptides with three transitions each. Human experts examined the data using Skyline[6] and manually adjusted the targeted peptides' peak boundaries. Additionally, peak quality was assessed using evaluation criteria such as equal retention time between heavy and light peptides, peak shape, intensity ratio consistency across transitions, and removal of transitions with interferences. A total of 30,294 peak groups were annotated, with 27,432 peak groups identified with peak boundaries and 19,230 peak groups deemed quantifiable.

DeepMRM showed better quantification accuracy compared with Skyline

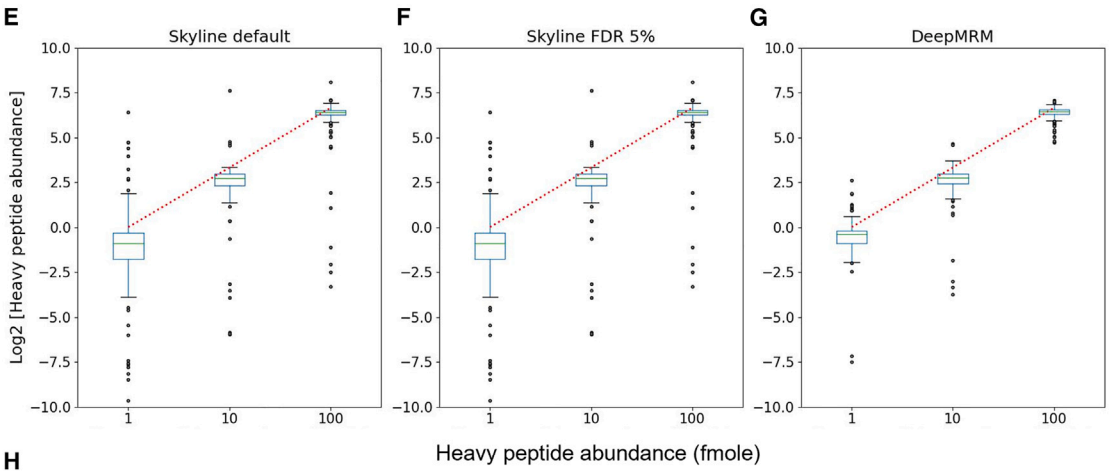
We first evaluated DeepMRM's quantification performance using a public MRM dataset consisting of two dilution series[7] (Table S1; see STAR Methods) and compared it with Skyline. In this dataset, heavy peptide abundances ranged from 0.1 to 100 fmol, while light peptide abundances remained constant. Absolute quantification was accomplished using the external calibration curve method as described in the original paper.[7] We evaluated the linear relationship between known and measured abundances of heavy peptides using Pearson's

CellPress
OPEN ACCESS



D

	Peak groups	PCC	SPC	MAAPE	Number of points in quantile box									
					0.1	0.2	0.5	1	2	4	10	20	50	100
Skyline default	1287	0.9284	0.9413	0.3713	129	129	129	130	128	130	129	128	129	126
Skyline FDR 5%	971	0.9482	0.9723	0.3011	70	90	89	94	101	100	103	107	106	111
DeepMRM	1183	**0.9736**	**0.9855**	**0.2988**	40	60	94	99	100	110	108	109	112	115



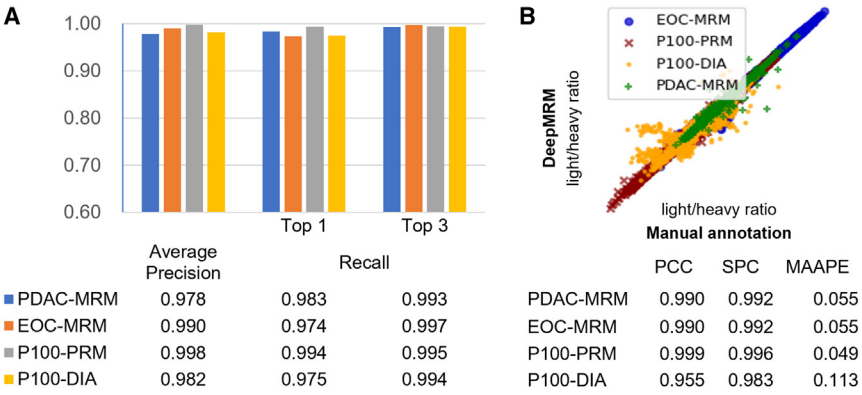
H

	Peak groups	PCC	SPC	MAAPE	Number of points in quantile box		
					1	10	100
Skyline default	386	0.9683	0.9237	0.3897	129	128	129
Skyline FDR 5%	386	**0.9683**	**0.9237**	0.3897	129	128	129
DeepMRM	323	0.9605	0.9159	**0.3025**	103	120	126

Figure 2. Comparison of DeepMRM and Skyline with or without mProphet

Relative quantification and distribution of the abundance of heavy peptides for (A–D) noisy dataset and (E–H) complex background dataset. There is no difference between Skyline default and Skyline at FDR 5% in the complex background dataset because mProphet did not filter any results. Red dotted lines represent the theoretical values according to the heavy peptide abundance. In the boxplots, the centerline, edges, and whiskers represent the median, the first and third quartile, and 1.5× interquartile range, respectively. Outlier points outside of the whiskers are indicated by dot symbols. The table below the boxplot shows averaged correlation coefficients, arctangent absolute percentage error over 43 targeted peptides, and the number of peak groups in each quantile box (PCC, Pearson's correlation; SPC, Spearman's rank correlation; MAAPE, mean arctangent absolute percentage error).

See also Figure S3.



	Average Precision	Recall Top 1	Recall Top 3
PDAC-MRM	0.978	0.983	0.993
EOC-MRM	0.990	0.974	0.997
P100-PRM	0.998	0.994	0.995
P100-DIA	0.982	0.975	0.994

	PCC	SPC	MAAPE
PDAC-MRM	0.990	0.992	0.055
EOC-MRM	0.990	0.992	0.055
P100-PRM	0.999	0.996	0.049
P100-DIA	0.955	0.983	0.113

Figure 3. DeepMRM performances on MRM, PRM, and DIA datasets

(A and B) (A) Average precision (AP) and recall (RC) scores for MRM, PRM, and DIA datasets and (B) scatterplot comparing the results of light/heavy ratios calculated by peaks manually annotated and those by peaks detected by DeepMRM. The scale of axes is log2. The table below the scatterplot shows correlation coefficients (PCC, Pearson's correlation; SPC, Spearman's rank correlation) and mean arctangent absolute percentage error (MAAPE).

See also Figure S2 and Table S2.

correlation coefficient (PCC) and Spearman's rank correlation coefficient (SPC). Additionally, we calculated the mean arctangent absolute percentage error (MAAPE).[15] Quantification results of Skyline reported in the previous benchmark test[7] were analyzed in two scenarios: original results (denoted as Skyline default) and results filtered by the mProphet[4] scoring model with a false discovery rate of 5% (denoted as Skyline FDR 5%) based on decoy transitions. Each technical replicate was treated as an independent sample in this test, and the quantification values for the same peptide in technical replicates were considered separate observations.

DeepMRM demonstrated higher quantification accuracy in detecting quantifiable peak groups compared with both Skyline default and Skyline 5% FDR (Figure 2). Although Skyline default reported the highest number of peak groups, it lacked quality control, resulting in the lowest correlation coefficients and highest MAAPE values. Skyline 5% FDR with the mProphet model improved quantification accuracy by filtering out 25% of results in the noisy dataset but did not filter any results in the complex background dataset, where a relatively small number of peak groups were considered. DeepMRM detected 85%–90% of peak groups for targeted peptides in both datasets and exhibited higher correlation coefficients and lower MAAPE values in quantification results compared with Skyline default and Skyline 5% FDR (Figures 2D and 2H). Upon visual inspection of the quantification results, we observed that DeepMRM outperformed other methods in accurately determining peak boundaries in noisy examples and effectively filtering out non-quantifiable transitions or peak groups from quantification (Figure S3).

DeepMRM showed robust performance across MRM, PRM, and DIA data

We evaluated DeepMRM on a hold-out test set of MRM data from PDAC samples (PDAC-MRM) and three external datasets: an MRM dataset used in an epithelial ovarian cancer study (EOC-MRM)[16] and PRM and DIA datasets employed to profile phosphosignaling responses (P100-PRM and P100-DIA)[17,18] (Table S1; see STAR Methods). In this test, we calculated recall (RC) and average precision (AP), which are conventional metrics used in the object detection task.[19,20] A detected peak group is considered a true positive only if it has an Intersection over Union (IoU) greater than a certain threshold with the manually annotated peak group (i.e., ground truth). Since chromatogram peaks often have long tails, which may result in large variations in determining peak endpoints, and these variations do not substantially impact the peak area, we opted for a less stringent IoU threshold of 0.3 rather than 0.5, a typical threshold in object detection studies. RC was calculated for the top 1 and 3 candidates per heatmap image (RC_1 and RC_3). Also, to validate the quality of true positive peak groups, we evaluated PCC, SPC, and MAAPE between the light/heavy ratios obtained by manual annotation and those obtained by DeepMRM.

DeepMRM exhibited an AP of 98%–99% and an RC of 97%–99% (Figures 3A and S2; Table S2). The light/heavy ratios estimated by peaks detected by DeepMRM displayed correlation coefficients of 0.96–0.99 when compared with manually annotated peaks (Figure 3B). MAAPE was in the range of 5%–11%. The relatively high MAAPE observed in P110-DIA data was mainly due to many peak groups with light/heavy ratios close to zero. In particular, more transitions were considered in DIA data compared with MRM data, which led to relatively larger variations in quantification results depending on transition selection. If these low-ratio peak groups are excluded (e.g., <0.05), the MAAPE decreased to 0.07.

To confirm the effectiveness of the data augmentation schemes on model performance, we trained the DeepMRM model without data augmentation. From the comparisons with and without augmentation, we concluded that the accuracy and robustness were improved by data augmentation (Table S3). Furthermore, we examined DeepMRM's performance in selecting good

Table 1. Comparison between quantification results obtained using transitions selected based on their quality and those obtained using all available transitions

		All transitions	Selected transitions
PDAC-MRM	PCC	0.993	0.990
	SPC	0.991	0.992
	MAAPE	0.061	0.055
P100-DIA	PCC	0.925	0.955
	SPC	0.983	0.983
	MAAPE	0.161	0.113

transitions used for quantification. Our analysis revealed that quantification results using transitions selected by DeepMRM were better aligned with the reference results obtained from manually selected transitions compared with those obtained using all available transitions with a decrease of up to 4.8% in MAAPE (Table 1). These findings demonstrate the effectiveness of the transition classification model in selecting transitions free of interference and noise.

Windows desktop application and integration into Skyline

DeepMRM is packaged as a Windows desktop application, streamlining the detection task and visualization of results. The GUI in the desktop application enables users to examine detection results and input transition chromatograms swiftly. Additionally, it allows for loading multiple samples together, simplifying the comparison of results for a target peptide. The desktop application supports the community standard format for MS data (mzML[21]) and proprietary formats like Thermo RAW and Sciex WIFF. In addition to the standalone desktop software, we have integrated the DeepMRM algorithm into the Skyline software package as an external tool.

The running time of DeepMRM (excluding time for chromatogram extraction) is a few seconds (<10 s) per sample on a Windows computer with a 3.8 GHz CPU (AMD Ryzen) and 32 GB memory (Table S4). Although using a GPU can further decrease the running time for data with a high number of transitions and longer chromatogram lengths (e.g., PRM or DIA data), the difference is not substantial in practice.

DISCUSSION

DeepMRM is a robust and highly accurate peak detection model for interpreting targeted proteomics data featuring a user interface for visualizing detection results. Designed to minimize human intervention in data interpretation, DeepMRM accepts targeted MS data and a target list as input and provides quantification results and confidence scores. Also, the GUI enables users to swiftly examine the results and make necessary corrections. We demonstrated that DeepMRM outperformed the community standard tool Skyline in quantification accuracy and exhibited robust accuracy across MRM, PRM, and DIA datasets. Importantly, DeepMRM does not require decoy transitions, which can limit assay throughput. With automated quantification and quality control, we anticipate that DeepMRM will facilitate high-throughput analysis of targeted proteomics.

Limitations of study

To enhance the applicability and accuracy of DeepMRM, we have planned several improvements. First, DeepMRM currently requires stable isotope-labeled peptides to quantify target peptides. We aim to expand the model to enable label-free quantification without heavy peptides. Second, isotopic peak cluster patterns in high-resolution PRM/DIA data are not currently utilized. We anticipate that incorporating the learning and utilization of these patterns will improve detection accuracy. Finally, DeepMRM has been designed and trained for targeted proteomics data interpretation. We are working on adapting DeepMRM to identify peptides in untargeted DIA data without prior transition information on targeted peptides. These developments will enhance the versatility of DeepMRM for various proteomics studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANTS DETAILS
- METHOD DETAILS
  - Cryopulverization, protein extraction and digestion of PDAC tissue samples
  - Synthesis and purification of stable isotope labeled peptides
  - Optimization of MRM conditions, LC-MRM-MS experiments, and analysis
  - Benchmark datasets
  - DeepMRM model
  - Benchmark test
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2023.100521.

ACKNOWLEDGMENTS

This work was supported by Bertis, Inc., and a grant from the National Research Foundation (NRF-2022M3H9A2086450) funded by the Korean Ministry of Science and ICT (MSIT). The biological specimens used in this study were provided by the Biobank at Seoul National University Hospital (SNUH), a member of the Korean Biobank Network (KBN4_A03), and the SNUH Cancer Tissue Bank.

AUTHOR CONTRIBUTIONS

J.P. developed and implemented DeepMRM. J.P., B.X.M., and S.K. designed and performed benchmarking tests. C.W. and D.A. implemented desktop applications. C.W., N.S., and B.X.M. performed a proof-of-concept integration with Skyline. J.H., S.B., H.K., and S.-W.L. conducted LC-MRM-MS experiments and analyzed the data. J.P. and S.K. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

J.P. is an employee of Bertis, Inc., and C.W., D.A., and S.K. are employees of Bertis Bioscience, Inc., both of which are companies developing proteomics-based diagnostics solutions.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors employed the ChatGPT service to refine English sentences during the preparation of this manuscript. After using the service, the authors

reviewed and edited the content as necessary and take full responsibility for the publication's content.

Received: January 5, 2023
Revised: April 18, 2023
Accepted: June 15, 2023
Published: July 11, 2023

REFERENCES

1. Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. Nat. Methods *9*, 555–566. https://doi.org/10.1038/nmeth.2015.
2. Doerr, A. (2013). Mass spectrometry–based targeted proteomics. Nat. Methods *10*, 23. https://doi.org/10.1038/nmeth.2286.
3. Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G., and Aebersold, R. (2017). Quantitative proteomics: challenges and opportunities in basic and applied research. Nat. Protoc., 1289–1294. https://doi.org/10.1038/nprot.2017.040.
4. Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.Y., Hengartner, M.O., and Aebersold, R. (2011). mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat. Methods *8*, 430–435. https://doi.org/10.1038/nmeth.1584.
5. Pino, L.K., Searle, B.C., Bollinger, J.G., Nunn, B., MacLean, B., and MacCoss, M.J. (2020). The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. Mass Spectrom. Rev. *39*, 229–244. https://doi.org/10.1002/MAS.21540.
6. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968. https://doi.org/10.1093/BIOINFORMATICS/BTQ054.
7. Nasso, S., Goetze, S., and Martens, L. (2015). Ariadne's thread: A robust software solution leading to automated absolute and relative quantification of SRM data. J. Proteome Res. *14*, 3779–3792. https://doi.org/10.1021/PR500996S/SUPPL_FILE/PR500996S_SI_001.ZIP.
8. Toghi Eshghi, S., Auger, P., and Mathews, W.R. (2018). Quality assessment and interference detection in targeted mass spectrometry data using machine learning 03 Chemical Sciences 0301 Analytical Chemistry. Clin Proteomics *15*, 1–13. https://doi.org/10.1186/S12014-018-9209-X/FIGURES/4.
9. Abbatiello, S.E., Mani, D.R., Keshishian, H., and Carr, S.A. (2010). Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. Clin. Chem. *56*, 291–305. https://doi.org/10.1373/CLINCHEM.2009.138420.
10. Wu, Z., Serie, D., Xu, G., and Zou, J. (2020). PB-Net: Automatic peak integration by sequential deep learning for multiple reaction monitoring. J. Proteonomics *223*, 103820. https://doi.org/10.1016/J.JPROT.2020.103820.
11. Melnikov, A.D., Tsentalovich, Y.P., and Yanshole, V.v. (2020). Deep Learning for the Precise Peak Detection in High-Resolution LC-MS Data. Anal. Chem. *92*, 588–592. https://doi.org/10.1021/ACS.ANALCHEM.9B04811/ASSET/IMAGES/LARGE/AC9B04811_0002.JPEG.
12. Xu, L.L., and Röst, H.L. (2020). Peak Detection On Data Independent Acquisition Mass Spectrometry Data With Semisupervised Convolutional Transformers. Preprint at arxiv. https://doi.org/10.48550/arxiv.2010.13841.
13. Lin, T.Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2020). Focal Loss for Dense Object Detection. IEEE Trans. Pattern Anal. Mach. Intell. *42*, 318–327. https://doi.org/10.48550/arxiv.1708.02002.
14. Hyeon, D.Y., Nam, D., Han, Y., Kim, D.K., Kim, G., Kim, D., Bae, J., Back, S., Mun, D.G., Madar, I.H., et al. (2022). Proteogenomic landscape of human pancreatic ductal adenocarcinoma in an Asian population reveals tumor cell-enriched and immune-rich subtypes. Nat Cancer *4*, 290–307. https://doi.org/10.1038/s43018-022-00479-7.
15. Kim, S.W., Cho, I.H., Park, J.N., Seo, S.M., and Paek, S.H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. Int. J. Forecast. *16*, 669–679. https://doi.org/10.1016/J.IJFORECAST.2015.12.003.
16. Hüttenhain, R., Choi, M., Martin de la Fuente, L., Oehl, K., Chang, C.Y., Zimmermann, A.K., Malander, S., Olsson, H., Surinova, S., Clough, T., et al. (2019). A Targeted Mass Spectrometry Strategy for Developing Proteomic Biomarkers: A Case Study of Epithelial Ovarian Cancer. Mol. Cell. Proteomics *18*, 1836–1850. https://doi.org/10.1074/MCP.RA118.001221.
17. Abelin, J.G., Patel, J., Lu, X., Feeney, C.M., Fagbami, L., Creech, A.L., Hu, R., Lam, D., Davison, D., Pino, L., et al. (2016). Reduced-representation phosphosignatures measured by quantitative targeted MS capture cellular states and enable large-scale comparison of drug-induced phenotypes. Mol. Cell. Proteomics *15*, 1622–1641. https://doi.org/10.1074/MCP.M116.058354/ATTACHMENT/23207A23-E223-4AE4-907C-5A30D82B48DE/MMC1.ZIP.
18. Vaca Jacome, A.S., Peckner, R., Shulman, N., Krug, K., DeRuff, K.C., Officer, A., Christianson, K.E., MacLean, B., MacCoss, M.J., Carr, S.A., and Jaffe, J.D. (2020). Avant-garde: an automated data-driven DIA data curation tool. Nat. Methods, 1237–1244. https://doi.org/10.1038/s41592-020-00986-4.
19. Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., and Dollár, P. (2015). Microsoft COCO: Common Objects in Context. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn., 3686–3693.
20. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., Everingham, M., Van Gool Leuven, L.K., CKI Williams, B., Winn, J., and Zisserman, A. (2009). The Pascal Visual Object Classes (VOC) Challenge. Int. J. Comput. Vis. *88*, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
21. Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W.H., Römpp, A., Neumann, S., Pizarro, A.D., et al. (2011). mzML–a community standard for mass spectrometry data. Mol. Cell. Proteomics *10*, R110.000133. https://doi.org/10.1074/MCP.R110.000133.
22. Mun, D.G., Bhin, J., Kim, S., Kim, H., Jung, J.H., Jung, Y., Jang, Y.E., Park, J.M., Kim, H., Jung, Y., et al. (2019). Proteogenomic Characterization of Human Early-Onset Gastric Cancer. Cancer Cell *35*, 111–124.e10. https://doi.org/10.1016/J.CCELL.2018.12.003.
23. Louwagie, M., Kieffer-Jaquinod, S., Dupierris, V., Couté, Y., Bruley, C., Garin, J., Dupuis, A., Jaquinod, M., and Brun, V. (2012). Introducing AAA-MS, a rapid and sensitive method for amino acid analysis using isotope dilution and high-resolution mass spectrometry. J. Proteome Res. *11*, 3929–3936. https://doi.org/10.1021/PR3003326/SUPPL_FILE/PR3003326_SI_001.PDF.
24. Lee, H., Mun, D.G., Bae, J., Kim, H., Oh, S.Y., Park, Y.S., Lee, J.H., and Lee, S.W. (2015). A simple dual online ultra-high pressure liquid chromatography system (sDO-UHPLC) for high throughput proteome analysis. Analyst *140*, 5700–5706. https://doi.org/10.1039/C5AN00639B.
25. Wang, M., Wang, J., Carver, J., Pullman, B.S., Cha, S.W., and Bandeira, N. (2018). Assembling the Community-Scale Discoverable Human Proteome. Cell Syst. *7*, 412–421.e5. https://doi.org/10.1016/j.cels.2018.08.004.
26. Chambers, M.C., MacLean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. *30*, 918–920. https://doi.org/10.1038/nbt.2377.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Tumor samples from human pancreatic ductal adenocarcinoma patients	Seoul National University Hospital (SNUH) and Supplemental table of Hyeon et al.[14]	https://www.nature.com/articles/s43018-022-00479-7#Sec34
Deposited data		
PDAC-MRM	This paper	MassIVE: MSV000089914; K-BDS: PRJKA2103083
P100-PRM	Abelin et al.[17]	MassIVE:MSV000079524
P100-DIA	Vaca Jacome et al.[18]	MassIVE:MSV000085540
Two-dilution series	Nasso et al.[7]	PeptideAtlas:PASS00456
Software and algorithms		
DeepMRM	This paper	Zenodo: https://doi.org/10.5281/zenodo.7964403; GitHub: https://github.com/bertis-informatics/deep-mrm

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sangtae Kim (sangtae.kim@bertis.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- PDAC-MRM dataset has been deposited to the MassIVE proteomics repository and K-BDS repository, and it is publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table. The source code also can be accessed at https://github.com/bertis-informatics/deep-mrm, and the DeepMRM desktop application is available for download at https://github.com/bertis-informatics/deep-mrm/releases. The DeepMRM external tool for Skyline can be downloaded from Skyline Tool Store at https://skyline.ms/skyts/home/software/Skyline/tools/begin.view.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANTS DETAILS

PDAC tissue samples were collected from patients who underwent surgery at Seoul National University Hospital (SNUH) from March 2010 to December 2016, following the procedures as previously described.[14] The patients included in the study were diagnosed with various stages of pancreatic cancer including IA, IB, IIA, IIB, or III. The age of patients ranged from 35 to 87, with a male-to-female ratio of 57:43. Ethical approval was obtained from the Institutional Review Board at Seoul National University Hospital (SNUH 1705-031-852). In brief, tissues were collected following a standardized operation procedure in the operating room for minimized cold ischemic time then immerged in liquid nitrogen and stored at −80°C until sampling.

METHOD DETAILS

Cryopulverization, protein extraction and digestion of PDAC tissue samples

66 PDAC patient tissue samples were cryopulverized, lysed and tryptic digested as previously described.[22] Briefly, cancer tissues were washed in PBS buffer to eliminate blood contamination then cryopulverized with a Cryoprep device (CP02, Covaris). Lysis buffer (4% SDS in 0.1 M Tris-HCl (pH 7.6) with phosphatase inhibitor (04906837001, Roche)) was added to the obtained tissue powder to

perform lysis using a probe sonicator (Q55 Sonicator, Qsonica). After removal of debris from centrifugation at 16,000 g for 10 min, the supernatant lysate protein concentration was measured with Pierce BCA assay kit (Thermo Scientific) according to the manufacturer's instructions for protein quantification. 500 μg of the measured protein was digested through a modified filter-associated sample preparation (FASP) method. In brief, reduction was performed on the proteins with SDT buffer (4% SDS in 0.1 M Tris-HCl, pH 7.6, and 0.1 M DTT) at 37°C for 45 min at 300 rpm then sequentially boiled for 10 min at 95°C. Bath sonication was performed for 10 min, then the reduced proteins were centrifuged at 16,000 g for 5 min where the supernatants were transferred to a Microcon device YM-30 filter (MRCF0R030, Millipore Corporation). For detergent removal, the filters were added 200 μL of 8 M urea in 0.1 M Tris-HCl (pH 8.5) then centrifuged at 14,000 g for 30 min. This step was repeated three times. Alkylation was performed by adding 100 μL of 50 mM iodoacetamide in 8M urea to each filter with incubated for 25 min at room temperature in dark then centrifuged at 14,000 g for 30 min 200 μL of 8 M urea was added followed by centrifugation at 14,000 g for 30 min for wash, which was repeated four times. 100 μL of 50 mM NH_4HCO_3 wash was then performed twice for buffer exchange. Collection tubes were exchanged, then trypsin (V5111, Promega) was added with an enzyme-to-protein ratio of 1:50 (w/w) to each filter for overnight incubation at 37°C. Second digestion was sequentially performed for 6 h with additional trypsin (enzyme-to-protein ratio of 1:100). Peptides were eluted with centrifugation at 14,000 g for 20 min, then added 60 μL of 50 mM NH_4HCO_3 twice to rinse. Peptide concentration was measured using Pierce BCA assay kit then vacuum-dried (concentrator plus, Eppendorf).

Synthesis and purification of stable isotope labeled peptides

Stable isotope labeled (SIL) peptides were synthesized and purified for spiking to conduct LC-MRM experiments of 66 PDAC patient sample tissues. 153 peptides selected based from mass spectrometry data of a previous PDAC proteogenomic study.[14] For peptide selection, adequacy for LC-MRM experiments were considered such as sequence length to be 7–25 amino acids, having no missed cleavages and exclusion of methionine along with its importance as a PDAC marker. Selected peptides were stable isotope labeled on their lysine or arginine at C-terminus with $^{13}C_6$ lysine or $^{13}C_6$ arginine analogs, respectively. The purity of the SIL peptides was higher than 95% which then the purified peptides were amino acid analyzed by AAA-MS method[23] for absolute quantification.

Optimization of MRM conditions, LC-MRM-MS experiments, and analysis

MRM conditions were optimized for peptide amount, collision energy for either of the 2+ and 3+ precursor ions and retention time of the 153 target peptides. If a peptide had similar intensity for 2+ and 3+ precursor ions, both transitions were included in the target list yet only one precursor with higher intensity and less interference was selected for quantitation afterward.

For each of 66 PDAC patient tissues, three replicates of the corresponding tissue peptide samples, each spiked with the SIL peptide mixture, were analyzed by LC-MRM-MS experiments. A home-built dual online nano-flow LC system[24] (Ultimate 3000 NCP-3200RS, Thermo Fisher Scientific) coupled with Agilent 6495C triple quadruple mass spectrometer platform at the Center for Proteogenome Research was used for LC-MRM-MS experiments. The injection amount of PDAC tissue peptide was 5 μg. Dynamic MRM was performed with a time window 3–5 min for the three best y-ion transitions of the 153 targets excluding y1 and y2 ions. The spray voltage of 2500 V, drying gas flow of 5 L/min, and gas temperature of 225°C were used. Both Q1 and Q3 resolutions were set to Unit, collision energies were set at 10–40 according to their previously optimized values, and a cycle time of 800 ms was used. Columns were manufactured in-house with Jupiter C18, 3 μm, 300 Å particles to 75 μm × 50 cm (Phenomenex), and the column temperature was kept at 60°C. A 60 min gradient (10%–40% solvent B over 47 min, 40%–80% over 5 min, 80% for 6 min, and 10% for 2 min, 400 nL min^{-1}) was used for each experiment. Solvent A was 0.1% formic acid in water, and solvent B was 0.1% FA in acetonitrile (ACN). The resultant 198 LC-MRM-MS data were analyzed with Skyline version 21.1.0.146,[6] and the transitions were manually examined with evaluation criteria such as equal retention time between heavy and light peptides, peak shape, intensity ratio consistency across transitions, and removal of transitions with peak interference. For quantifiable peptides, the ratios of light/heavy peptides were determined based on peak areas. All transition lists are uploaded to MassIVE (identifier: MSV000089914) along with the raw data.

Benchmark datasets

We obtained four datasets comprising LC-MRM/PRM/DIA-MS experiments, which were utilized for evaluation (Table S1).

EOC-MRM, P100-PRM, and P100-DIA

Three external datasets were retrieved from the MassIVE repository,[25] comprising of an MRM dataset for biomarkers in epithelial ovarian cancer (EOC-MRM) samples[16] (MassIVE identifier: MSV000084048), a PRM validation dataset of approximately 100 phosphopeptides (P100-PRM) samples, produced for the Library of Integrated Network-based Cellular Signatures (LINCS) project[17] (MassIVE identifier: MSV000079524), and a DIA data of the same phosphoproteomics samples (P100-DIA) that was utilized to assess the accuracy of the Avan-Garde (AvG) tool against expert manual curation[18] (MassIVE identifier: MSV000085540). For all these datasets, Skyline files generated during the quantification analysis were obtained.

Dilution series datasets (noisy and complex background)

MRM datasets, previously generated to benchmark quantification algorithms in a study,[7] were obtained from the PeptideAtlas repository (Dataset identifier: PASS00456). These datasets were produced from two dilution series in which 43 SIL peptides were spiked under varying sample and acquisition conditions. Specifically, the complex background dataset was created with a complex background, while the noisy dataset was generated under suboptimal conditions without background. In these datasets, the

abundances of heavy peptides ranged from 0.1 to 100 fmol, while the abundance of light peptides remained constant. The quantification analysis results from the benchmark study were also acquired from https://github.com/saranasso/Ariadne.

DeepMRM model

Data preprocessing

Given the input data, which includes the LC-MS data and the target list, DeepMRM begins by extracting transition chromatograms for targeted precursor ions and subsequently transforms them into heatmap images with two channels. The entire transition chromatograms for targeted peptides are extracted unless reference retention times with time windows are provided. Linear interpolation is applied to all transition chromatograms, ensuring they share the same length and a scanning interval of 0.5 s. All the chromatograms are concatenated, with the horizontal and vertical axes representing the transition index and retention time, respectively. Consequently, two concatenated matrices (i.e., a two-channel heatmap image) are generated: one channel for light transition chromatograms and another for heavy transition chromatograms. The dimensions of heatmap images correspond to [2, number of transitions, length of chromatograms]. Lastly, as part of the preprocessing step, each transition pair is scaled from 0 to 1.

Model architecture

DeepMRM comprises two neural network models: a peak detection model for identifying peak groups and a transition classification model for selecting interference- and noise-free transitions.

Peak detection model. The architecture of our peak detection model is based on RetinaNet, a widely-used neural network model for object detection.[13] The network comprises a backbone network for extracting features and two small sub-networks. The first sub-network classifies the presence of peak groups on the extracted features; the second sub-network performs peak boundary regression.

We chose ResNet18 as the backbone network and modified it to accommodate the peak detection problem. First, the kernel size of conv1 is changed from 7x7 to 1x7, and the stride and padding sizes are adjusted accordingly. In the first convolutional layer, grouped convolutions are applied so that the heavy and light channels are convolved separately. Before the first residual block, an adaptive average pooling layer is inserted such that the height of the feature map becomes one. The kernel and padding sizes in the subsequent residual blocks are adjusted to 1x3 and 0x1, respectively. As a result, all the feature maps generated by the backbone are one-dimensional vectors.

Both sub-networks also employ convolutional layers with a kernel size of 1x3 and a padding size of 0x1. Since the regressor needs to predict only two points for a peak boundary rather than four points for a bounding box, the final output channel size is altered to twice the number of anchors. The anchors are generated with a single scale for feature level. The classifier subnetwork is configured for a binary classification problem: {background versus peak group}.

Transition classification model. A CNN-based classification model is developed to select transitions unaffected by interference or noise from the detected peak groups. This model takes a pair of transitions as input and classifies whether the pair is quantifiable by considering factors such as peak shape similarity, retention time match, light/heavy ratio similarity, and interference. The transition classification model is a binary classifier that is applied to the backbone architecture used in the peak detection model, with the only difference being that the kernel size of the first convolutional layer (conv1) is changed to 2x7.

Transition selection

Once peak groups are identified by the peak detection model, noise- and interference-free transitions are selected using the transition classification model. This model evaluates the quantitative feasibility of transition pairs. To assess the quantitative feasibility of each transition, a representative transition profile is generated as a reference chromatogram. All pairs of available transitions are evaluated by the transition classification model. Quantifiable transitions are then combined to create the representative transition profile. If no quantifiable transition pairs are found, the representative transition profile is created by summing all transitions. To reduce the computational cost, only up to six transitions with the highest peak heights are considered. As a last step, each transition is paired with the representative transition profile and assessed using the transition classification model for its quantitative feasibility. The classification model's confidence score is considered a transition quality score.

Abundance calculation

Upon selecting the quantifiable transitions, the target peptide's abundance (peak area) is calculated. If there are fewer than two quantifiable transitions, the peak area is computed using all transitions. The average of the transition quality scores of the selected transitions is reported as a quantification confidence score. The peak area is calculated using the trapezoidal numerical integration method.

Data augmentation

To enhance the robustness and applicability of DeepMRM, a data augmentation strategy is employed during model training. Augmentation methods include random resizing, cropping, intensity jittering, retention time shifting, and transition rescaling. The retention time shift destroys the alignment of the light and heavy transition peaks, and the transition rescaling makes the light/heavy ratio across transitions inconsistent. During the augmentation, the label data is also transformed accordingly. All these augmentation methods are applied prior to generating the heatmap image.

Training

The in-house dataset (PDAC-MRM) was split into training, validation, and test sets at a ratio of 8:1:1. The model was trained using the Adam optimizer with default parameters (lr = 1.e−3, $\beta 1 = 0.9$, $\beta 2 = 0.999$). The training was performed on NVIDIA GeForce 3090 GPUs

for 100 epochs with a batch size of 512. The learning rate was decreased by 0.5 every 10 epochs. The training was stopped when there was no improvement over 50 epochs.

Confidence score

DeepMRM comprises a peak detection model and a transition classification model, which compute and report two confidence scores: peak group confidence score and quantification confidence score. The peak group confidence score is calculated by a peak detection model using a sigmoid function to estimate the probability of a peak group being present within the candidate boundary region.

The quantification confidence score is determined by evaluating the quality of selected transitions for quantification. Each transition's quality is assessed by comparing it to a representative transition profile to determine if it contains interference or noise and if its retention time, peak shape, and heavy/light ratio are sufficiently similar for quantification. Details on creating the representative transition profile are described in the transition selection section. Then, the average quality score of the selected transitions for quantification is reported as the quantification confidence score.

Benchmark test

Data preparation

All mass spectrometry files were converted to mzML format using ProteoWizard's msConvert toolkit.[26] Then, pyOpenMS library was utilized to extract chromatograms from the mzML files. Before extracting chromatograms, all PRM and DIA spectra were centroided, and a 20ppm extraction window was set. In the case of the P100-DIA dataset, a retention time window of 20 min was used, with the reference retention time specified in the spectral library. For other MRM and PRM datasets, the complete transition chromatograms acquired for the targeted peptides were extracted and utilized as input for DeepMRM.

The boundary of peak groups, list of transitions, and annotation information about quantitative transitions (if available) were extracted from Skyline files for three external datasets. In the case of P100-PRM dataset, the quantification results were filtered and normalized using an in-house downstream analysis protocol by the original authors.[17] Since the filtered quantification results were not available, we employed Skyline's "dotp" score of 0.7 to exclude unreliable measurements. This filtering process removed 2,117 out of 13,629 peak groups. We manually inspected the excluded peak groups and observed that most lacked light or heavy peptide signals. For P100-DIA, we evaluated DeepMRM using the AvG open curation dataset.

Evaluation

We evaluated DeepMRM's performance in two major tasks. The first task is identifying peak groups relevant to the target peptide in the given chromatograms (i.e., object detection task). Since object detection involves searching for the target object from various combinations of location and scale, the precision-recall curve is typically used rather than the ROC curve commonly used for classification tasks.[20] A detected peak group is considered a true positive only if it has an Intersection over Union (IoU) greater than a certain threshold with the manually annotated peak group (i.e., ground truth). Here, we calculated and reported the average precision (AP) metric averaged across all recall levels ranging from 0 to 1. AP can summarize the precision-recall curve as a single value. We also calculated recall considering only the top 1 or 3 candidates per input image at a certain IoU threshold (RC_1 and RC_3). Recall can measure the detection model's ability to identify all peak groups (actual positive instances). Chromatogram peaks often have long tails, which may result in large variations in determining the end of peaks. Since this variation does not significantly affect the peak area, we used a less stringent IoU threshold of 0.3 rather than 0.5, a common threshold in object detection studies.

The second task involves selecting the transitions unaffected by noise and interference from the detected peak groups and estimating the abundance (peak area) of the targeted peptides. We evaluated how well the estimated abundance matched the manually calculated or ground truth abundances (if available). We calculated the linear relationship between the known and measured abundances of targeted peptides using Pearson's correlation coefficient (PCC) and Spearman's rank correlation coefficient (SPC). Additionally, we calculated the mean arctangent absolute percentage error (MAAPE).[15] For the two-dilution series dataset with available ground truth abundance, the abundance of the peak group with the highest score was considered. For other benchmark datasets, the abundance of true positive peak groups was taken into account.

Confidence score threshold

In the benchmark experiments, only peak groups with a peak group confidence score of 0.05 or higher were considered, and only quantification results with a quantification confidence score of 0.01 or higher were considered.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was performed in Python 3.9 with numpy and scipy libraries. More information on statistical tests used are outlined in the method details above, and also indicated in the figure legends.