M²ara: unraveling metabolomic drug responses in whole-cell MALDI mass spectrometry bioassays

T. Enzlein[1], A. Geisel[1], C. Hopf[1, 2, 3, *], S. Schmidt[1, *]

[1]Center for Mass Spectrometry and Optical Spectroscopy (CeMOS), Mannheim University of Applied Sciences, Paul-Wittsack-Str. 10, 68165 Mannheim

[2]Medical Faculty, Heidelberg University, Im Neuenheimer Feld 280, 69117 Heidelberg, Germany

[3]Mannheim Center for Translational Neuroscience (MCTN), Medical Faculty Mannheim, Heidelberg University, Theodor Kutzer-Ufer 1-3, 68167 Mannheim, Germany

*To whom correspondence should be addressed.

Abstract

Summary: Fast computational evaluation and classification of concentration responses for hundreds of metabolites represented by their mass-to-charge (*m/z*) ratios is indispensable for unraveling complex metabolomic drug actions in label-free, whole-cell Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI MS) bioassays. In particular, the identification of novel pharmacodynamic biomarkers to determine target engagement, potency and potential polypharmacology of drug-like compounds in high-throughput applications requires robust data interpretation pipelines. Given the large number of mass features in cell-based MALDI MS bioassays, reliable identification of true biological response patterns and their differentiation from potentially present measurement artefacts is critical. To facilitate the exploration of metabolomic responses in complex MALDI MS datasets, we present a novel software tool, *M²ara*. Implemented as a user-friendly R-based shiny application, it enables rapid evaluation of Molecular High Content Screening (MHCS) assay data. Furthermore, we introduce the concept of Curve Response Score (CRS) and CRS fingerprints to enable rapid visual inspection and ranking of mass features. In addition, these CRS fingerprints allow direct comparison of cellular effects among different compounds. Beyond cellular assays, our computational framework can also be applied to MALDI MS-based (cell-free) biochemical assays in general.

Availability and implementation: The software tool, code and examples are available at https://github.com/CeMOS-Mannheim/M2ara and https://dx.doi.org/10.6084/m9.figshare.25736541.

Contact: c.hopf@hs-mannheim.de, s.schmidt@hs-mannheim.de

INTRODUCTION

Whole-cell MALDI MS assays are a versatile Molecular High Content Screening (MHCS) technology that unravels the metabolomic complexity of drug responses in a biological system(Weigt *et al.* 2018, 2019; Unger *et al.* 2020, 2021; Dueñas *et al.* 2023). In contrast to conventional HCS technologies that focus primarily on morphometric or functional measures such as cell size, neurite length, dendrite branching or calcium signaling(Yeyeodu *et al.* 2010; Wu *et al.* 2019), MHCS refers to the fact that MS cell assays can monitor hundreds of metabolites simultaneously, providing a snapshot of the metabolome. However, to date, the interpretation and classification of hyperspectral MS assay data remains a computational challenge in bioanalytics, pharmacology as well as chemical biology and life science research. In particular, in phenotypic assays where the mode of action of drugs can be investigated in a target-agnostic manner, sophisticated data interpretation pipelines are essential to transform individual cellular *metabolomic fingerprints* into a meaningful and simple form that can be classified, sorted and visualized based on well-defined scores. Hereby, mass-specific concentration response curves are deduced, followed by their classification and characterization using given assay metrics. Understanding metabolomic responses can enable the identification of useful pharmacodynamic biomarkers and thus allows the determination of potencies and effect sizes of test compounds in early drug discovery including early drug safety testing. Nevertheless, there is no common open-source or proprietary software tool capable of performing concentration-response regression analysis in an easy-to-use fashion(Di Veroli *et al.* 2015; Ritz *et al.* 2015; Wang *et al.* 2020; Unger *et al.* 2021) of complex MALDI MS data. Moreover, implementations designed for the classification and scoring of cellular metabolomic fingerprints across compound libraries or cell lines are missing. Recently, a statistical method has been introduced to address statistical significance into concentration-response curve analysis, allowing concentration-response relationships for large proteomics datasets to be visualized, evaluated and understood(Bayer *et al.* 2023).

To facilitate the exploration of cellular drug effects in comprehensive MALDI MS metabolomic datasets and to link measured effects to multiple drug-targets in high-throughput screens, we introduce a novel software tool, *M²ara*, for the interpretation of whole-cell MALDI MS bioassays, which includes a scoring system to filter for significant mass features. It is implemented as an open-source R-based *shiny* application or executable, including several modifications and additions compared to our previously published work(Unger *et al*. 2021). Specifically, we added a *shiny* based graphical user interface, including insightful plots to visualize spectra, dose-response curves, assay metrics and experimental parameters, introduced metrics for assay quality, and added functionalities for clustering of curves with similar shape as well as the option to perform sparse principal component analysis (PCA) as a multivariate alternative to univariate dose-response curves.

METHODS

The computational pipeline of *M²ara* (**Fig. 1a, Fig. S1**) is designed to decipher the molecular complexity of drug responses and to provide response characteristics in cellular systems. To this end, raw MS data is first imported using the *MALDIquantForeign* R-package(Gibb and Strimmer 2012) followed by (user-defined) data pre-processing such as baseline removal using the TopHat method, Savitzky-Golay smoothing of the spectra, square-root transformation to stabilize variance of the data and filtering for monoisotopic peaks (**Fig. 1a, i**). Next, *M²ara* offers recalibration of the data to a selected peak, for example using a *m/z* value of an internal standard, normalization (total ion current, internal standard, etc.) and alignment of the spectra. Since the MALDI target plate layout commonly includes several technical replicates per condition, potential outliers are highlighted using Chauvenet's Criterion(Chauvenet 1863) as well as information on the total ion current or signal intensity for a given feature per spot. After pre-processing, the mean spectra for all concentrations are calculated and a four-parametric logistic regression curve is fitted for each *m/z* feature (**Fig. 1a, ii**). Subsequently, performance measures (pEC50, log2-fold-change, SSMD, modified Z' factor, modified V' factor(Zhang, Chung and Oldenburg 1999; Bray, Carpenter, and Imaging Platform, Broad Institute of MIT and Harvard 2004)) are calculated (**Fig. 1a, iii, Suppl. Information**). In addition, we implemented a Curve Response Score (CRS) to yield a dedicated feature reduction and thus a focused selection and ranking of significant mass features (**Fig. 1a, iv**). The corresponding CRS fingerprint provides a unique visualization of regulated mass features unraveling metabolic responses essential for exploring cellular metabolic alterations for a given compound or even across different compounds. To this end, we combine three known assay quality metrics (**see Suppl. information**) to account for the effect size of the response curve (*log*2-fold-change), their variance (modified *Z*' factor) and the goodness of the fit (modified *V*' factor). Based on the given shape and results of the used metric, response curves are classified as either up- or down-regulated or even non-regulated. Results are presented to the user in a convenient table format and plots, including response curves as well as MS spectra. All performance measures can be exported as a .csv file for further analysis.

RESULTS

We introduce and benchmark *M²ara* by presenting previously published data, ranging from a cell-based MALDI MS fatty acid synthase (FASN) inhibitor assay(Weigt *et al*. 2019), a MALDI MS cellular screening assay for compounds that block the uptake of the substrate estrone-3-sulfate via the OATP2B1 transporter(Unger *et al*. 2020) and a phenotypic, i.e. target-agnostic cell-based assay for tyrosine-kinase inhibitors(Weigt *et al.* 2018) to a complement dependent cytotoxicity (CDC) assay for potency assignment of therapeutic antibodies(Schmidt *et al*. 2024). Beyond cellular assays, our computational framework can be used for MALDI MS-based biochemical assays in general. Examples of a previously published γ-secretase in vitro inhibition assay(Koch *et al*. 2023) (**Fig. S2**) and cell-free caspase-6 inhibitor assay (**Fig. S3**, https://www.ppscreeningcentre.com/label-free-mass-spectrometry-ultra-highthroughput-screening/) are provided.

Exemplarily, the CRS fingerprint of the FASN bioassay (**Fig. 1b**) features a CRS for the up-regulated metabolite malonyl-CoA (*m/z* 854.1, $[M+H]^+$) of the fatty acid synthesis pathway of 83.9 %, whose corresponding response curve is depicted in **Fig. 1b**. Further, a non-regulated mass feature (*m/z* 760.6, likely PC34:1$[M+H]^+$) with a CRS of 0% is presented. A similar simple CRS fingerprint was computed for the OATP2B1 transporter inhibition assay (**Fig. 1c**). As expected, the *m/z* value for the down-regulated substrate estrone-3-sulfate (*m/z* 349.1, $[M-H]^-$) yielded a high CRS of 65.1%. More complex patterns were found for the phenotypic assays presented by(Weigt *et al.* 2018) (**Fig. S4**) and (Schmidt *et al*. 2024), with e.g. glutathione as the most significant response marker identified in the latter case study (**Fig. S5**).

Expanding the evaluation to multiple drugs allows compound classification based on either a single mass feature or the entire CRS fingerprint (**Fig. 1a, v**), exemplified for a small number of compounds using the FASN inhibitor data set (**Fig. S6a, c**). **Fig. S6a** shows the pEC50 as well as the CRS for different FASN inhibitors using malonyl-CoA (*m/z* 854.1, $[M+H]^+$) enrichment as the read-out. **Fig. S6b** on the other hand shows all *m/z* features with a non-zero CRS vs. their respective pEC50 values for different drugs and is an example of an untargeted approach where the response fingerprint is of interest. This can be useful to identify off-target-effects of known compounds.

Besides univariate analysis, *M²ara* offers a set of multivariate analysis tools. For example, the global response can be visualized with a PCA (**Fig. S6c**). This approach can serve as a feasibility check for experiments at an early stage of assay development. Moreover, by leveraging a sparse PCA that uses an L1-norm to represent the large feature space with just a small subset of *m/z* features, it is also possible to identify multivariate responses, where all curves examined alone might not cross the threshold to be identified as valid responses, but taken together provides insight into complex mechanisms.

Finally, to assess concentration response similarities, *M²ara* features the classification of response curves based on their sigmoidal shape (**Fig. S2c**). This serves as an additional selection or similarity feature for deciphering metabolic responses. In particular, if response markers with a positive CRS are present, it may help to identify other *m/z* features with the same (but weak-er/noisier) response curve characteristics. An example would be the identification of additional dysregulated Aβ-species that, although below the CRS quality threshold, still follow the same trend as the Aβ-species detected by CRS

alone (**Fig. S2c and d**). Additionally, this technique might also be used to identify markers with a reverse response.

DISCUSSION

In the last decade, mass spectrometry based assay technology has emerged as a promising high-throughput technology in drug discovery(Unger *et al*. 2021; Dueñas *et al.* 2023) including MALDI MS or its variants like IR-MALDESI MS or acoustic mist ionization/acoustic ejection MS(Belov *et al*. 2020; Simon *et al*. 2020, 2021; Pu *et al*. 2023). Standardized protocols exist for the establishment of both, biochemical or cellular assays(De Cesare *et al*. 2020; Unger *et al*. 2020). The latter provide a unique framework for early drug discovery as compounds are directly tested in cells, offering the potential for identification of more specific target compounds(Unger *et al*. 2020; Pu *et al*. 2023). Beyond that, for phenotypic cellular experiments, in which the metabolic response to a pharmaceutical reagent is usually unknown, computational data evaluation pipelines implemented in a single platform are indispensable to handle data load, data complexity and to provide a fast and user-friendly way for data visualization and data interpretation. With *M²ara* it becomes straightforward to perform response curve analysis for the identification of e.g. potential response markers as well as mode-of-action studies. *M²ara* facilitates the evaluation and classification of mass-specific response curves enabling fast decision making in assay development. For compound library screens, *M²ara* provides statistical tools for the visualization and analysis of screening results including similarity tests and potency assignments of compounds.

M²ara allows to import data in the Bruker Flex (*.fid) format as well as the open mzML standard. In future trapped ion mobility could be included(Willems *et al*. 2021; Luu *et al*. 2022) for next level analytical assay analysis. Furthermore, advanced machine learning tools might also help to further improve statistical evaluation and classification of the significance of response curves, feature classification and pattern recognition.

In summary, we present a novel, open-source software tool for the evaluation and interpretation of MHCS MALDI MS data sets. The data analysis pipeline facilitates robust feature identification in minutes by introducing a Curve Response Score that is based on common assay metrics. Those drug-dependent metabolic fingerprints are reduced in features space to allow for fast and significant similarity tests across compounds. Novel response *m/z* biomarkers can be identified rapidly by ranking of significant features including potency assignment. Predominantly, *M²ara* serves as a computational tool in assay development for biomarker identification and the investigation of alterations in cell metabolism upon drug treatment.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

brief description (file type, i.e., PDF)

AUTHOR INFORMATION

Corresponding Author

* Stefan Schmidt, s.schmidt@hs-mannheim.de, Carsten Hopf, c.hopf@hs-mannheim.de, Center for Mass Spectrometry and Optical Spectroscopy (CeMOS), Mannheim University of Applied Sciences, Paul-Wittsack-Str. 10, 68165 Mannheim

Author Contributions

This work was conceived by T.E., S.S. and C.H. T.E. developed and implemented the software, wrote code and performed statistical data evaluation. S.S. coordinated this work, provided input on statistical metrics, evaluated all data and tested the software. T.E. and S.S. wrote the manuscript. A.G. evaluated the software and provided input on statistical features. C.H. supervised the overall work, evaluated the results, provided infrastructure and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENT

We acknowledge the work of B. Munteanu, S. Schulz, M. Unger, and D. Weigt for their work in the development of cellular MALDI-TOF MS assays. Further, we thank B. Fröhlich for fruitful discussions on assay technology and his contributions in laboratory automation processes. In addition, we thank S. Honarnejad, F. Kleinpenning and N. Ottenheym from Pivot Park Screening Centre for sharing the cas-6 data set. In addition, we thank Novartis Technical Research and Development for sharing the complement dependent cytotoxicity (CDC) bioassay data. This work was supported by the BMBF (German Federal Ministry of Research) as part of the Innovation Partnership "Multimodal Analytics and Intelligent Sensorics for the Health Industry" (M²Aind), projects "Drugs4Future" (grant 12FH8I05IA) and "DrugsData" (grant 13FH8I09IA) and by the Ministerium für Wissenschaft, Forschung and Kunst (MWK) Baden-Württemberg by the Mittelbauprogramm. In addition this work was supported by DFG grant INST874/9-1.

REFERENCES

Bayer FP, Gander M, Kuster B *et al*. CurveCurator: a recalibrated F-statistic to assess, classify, and explore significance of dose-response curves. *Nat Commun* 2023;**14**:7902.

Belov AM, Kozole J, Bean MF *et al*. Acoustic Mist Ionization-Mass Spectrometry: A Comparison to Conventional High-Throughput Screening and Compound Profiling Platforms. *Anal Chem* 2020;**92**:13847–54.

Bray M-A, Carpenter A, Imaging Platform, Broad Institute of MIT and Harvard. Advanced Assay Development Guidelines for Image-Based High Content Screening and Analysis. In: Markossian S, Grossman A, Arkin M, et al. (eds.). *Assay Guidance Manual*. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004.

Chauvenet W. *A Manual of Spherical and Practical Astronomy: Theory and Use of Astronomical Instruments, Method of Least Squares*. J. B. Lippincott & Company, 1863.

De Cesare V, Moran J, Traynor R *et al*. High-throughput matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry-based deubiquitylating enzyme assay for drug discovery. *Nat Protoc* 2020;**15**:4034–57.

Di Veroli GY, Fornari C, Goldlust I *et al*. An automated fitting procedure and software for dose-response curves with multiphasic features. *Sci Rep* 2015;**5**:14701.

Dueñas ME, Peltier-Heap RE, Leveridge M *et al.* Advances in high-throughput mass spectrometry in drug discovery. *EMBO Mol Med* 2023;**15**:e14850.

Gibb S, Strimmer K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* 2012;**28**:2270–1.

Koch M, Enzlein T, Chen S-Y *et al.* APP substrate ectodomain defines amyloid-β peptide length by restraining γ-secretase processivity and facilitating product release. *The EMBO Journal* 2023;**n/a**:e114372.

Luu GT, Freitas MA, Lizama-Chamu I *et al.* TIMSCONVERT: a workflow to convert trapped ion mobility data to open data formats. *Bioinformatics* 2022;**38**:4046–7.

Pu F, Radosevich AJ, Bruckner BG *et al.* New Platform for Label-Free, Proximal Cellular Pharmacodynamic Assays: Identification of Glutaminase Inhibitors Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization Mass Spectrometry. *ACS Chem Biol* 2023;**18**:942–8.

Ritz C, Baty F, Streibig JC *et al.* Dose-Response Analysis Using R. *PLoS One* 2015;**10**:e0146021.

Schmidt S, Geisel A, Enzlein T *et al.* Label-free Assessment of Complement-Dependent Cytotoxicity of Therapeutic Antibodies via a Whole-Cell MALDI Mass Spectrometry Bioassay. 2024:2024.05.03.592336.

Simon RP, Häbe TT, Ries R *et al.* Acoustic Ejection Mass Spectrometry: A Fully Automatable Technology for High-Throughput Screening in Drug Discovery. *SLAS Discov* 2021;**26**:961–73.

Simon RP, Winter M, Kleiner C *et al.* MALDI-TOF Mass Spectrometry-Based High-Throughput Screening for Inhibitors of the Cytosolic DNA Sensor cGAS. *SLAS Discov* 2020;**25**:372–83.

Unger MS, Blank M, Enzlein T *et al.* Label-free cell assays to determine compound uptake or drug action using MALDI-TOF mass spectrometry. *Nat Protoc* 2021;**16**:5533–58.

Unger MS, Schumacher L, Enzlein T *et al.* Direct Automated MALDI Mass Spectrometry Analysis of Cellular Transporter Function: Inhibition of OATP2B1 Uptake by 294 Drugs. *Anal Chem* 2020;**92**:11851–9.

Wang D, Hensman J, Kutkaite G *et al.* A statistical framework for assessing pharmacological responses and biomarkers using uncertainty estimates. *Elife* 2020;**9**:e60352.

Weigt D, Parrish CA, Krueger JA *et al.* Mechanistic MALDI-TOF Cell-Based Assay for the Discovery of Potent and Specific Fatty Acid Synthase Inhibitors. *Cell Chem Biol* 2019;**26**:1322-1331.e4.

Weigt D, Sammour DA, Ulrich T *et al.* Automated analysis of lipid drug-response markers by combined fast and high-resolution whole cell MALDI mass spectrometry biotyping. *Sci Rep* 2018;**8**:11260.

Willems S, Voytik E, Skowronek P *et al.* AlphaTims: Indexing Trapped Ion Mobility Spectrometry–TOF Data for Fast and Easy Accession and Visualization. *Molecular & Cellular Proteomics* 2021;**20**:100149.

Wu N, Nishioka WK, Derecki NC *et al.* High-throughput-compatible assays using a genetically-encoded calcium indicator. *Sci Rep* 2019;**9**:12692.

Yeyeodu ST, Witherspoon SM, Gilyazova N *et al.* A Rapid, Inexpensive High Throughput Screen Method for Neurite Outgrowth. *Curr Chem Genomics* 2010;**4**:74–83.

Zhang JH, Chung TD, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* 1999;**4**:67–73.

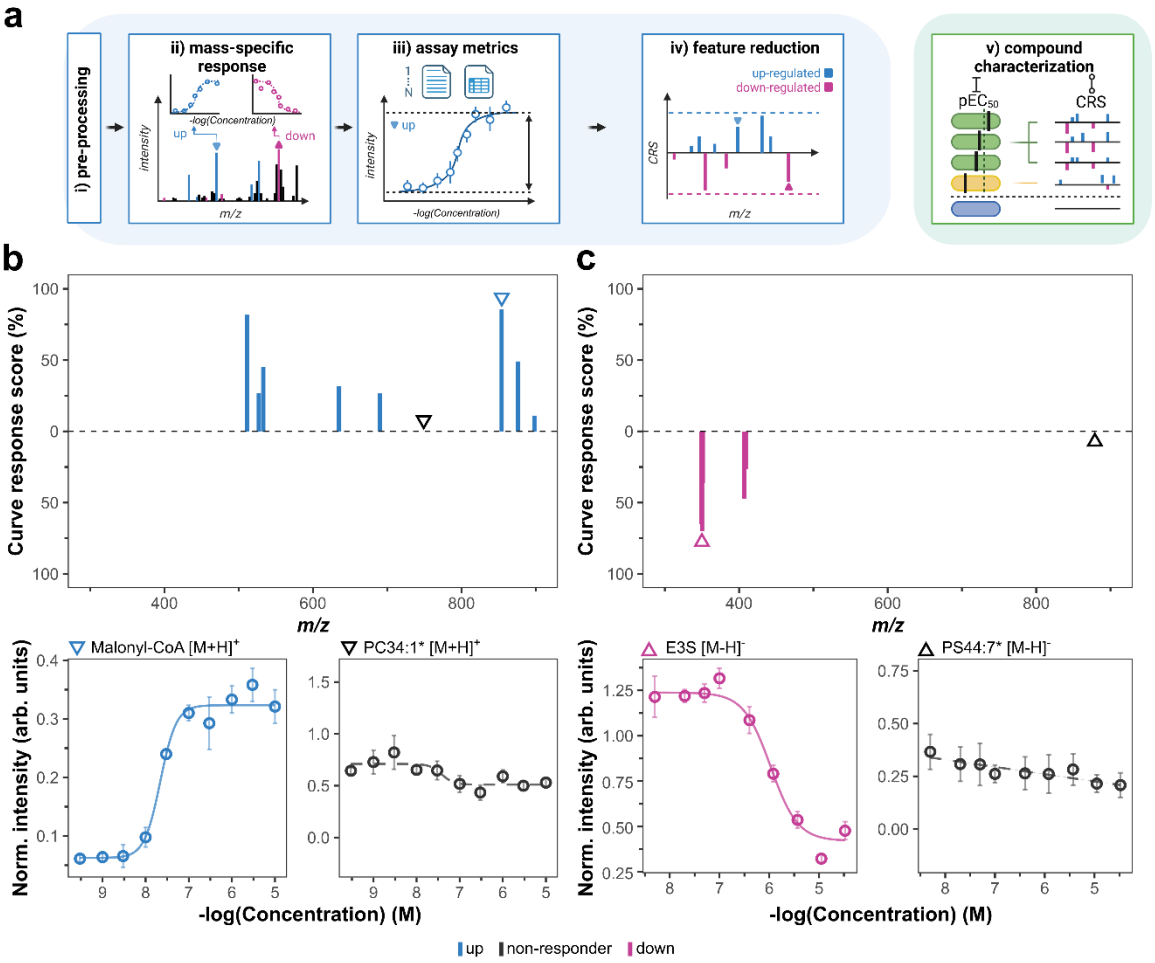

Figure 1: Schematic of the computational framework of *M²ara* for the interpretation and classification of metabolic drug responses in complex MALDI-TOF-MS assays. a, The data interpretation pipeline of *M²ara* includes i) data-preprocessing, ii) fitting of a sigmoidal response curve, iii) calculation of assay quality measures and iv) feature reduction based on an introduced Curve Response Score, a composite score taking into account different assay metrics. The drug action can be repeatedly calculated for multiple compounds, allowing for compound classification based on their potency or direct comparison of metabolic response patterns using CRS fingerprints. **b**, CRS fingerprint for the FASN inhibitor MALDI cell assay presented in Weigt et al., 2019, for A549 cells treated with compound GSK2194069. Lower panel shows characteristic drug-response curves of two selected mass features, malonyl-CoA (left) (*m/z* 854.1, $[M+H]^+$) as an up-regulated responder as well as a non-responder (*PC34:1, putative, *m/z* 760.6, right). **c**, CRS fingerprint for the OATP2B1 inhibitor erlotinib based on a cell assay presented in Unger et al., 2020. Lower panel shows characteristic drug-response curves of two selected mass features E3S (*m/z* 349.1, $[M-H]^-$) as a down-regulated responder, and *m/z* 888.6 (*PS44:7 $[M-H]^-$, putative) as a non-responder. Uncertainties are presented as standard deviation of four technical replicates.