

# Deep Learning-based Pseudo-Mass Spectrometry Imaging Analysis for Precision Medicine

Xiaotao Shen PhD

Stanford University School of Medicine  
Department of Genetics  
Dr. Snyder lab

October 9<sup>th</sup>, 2022 @ CASMS



shenxt@stanford.edu



shenxt.info



@xiaotaoshen1990



github.com/jaspershen



# About Me

Name: Xiaotao Shen/申小涛

Affiliation: Stanford University School of Medicine

Title: Postdoctoral Research Fellow

Email: shenxt@stanford.edu

Personal website: shenxt.info

GitHub: github.com/jaspershen

Twitter: @xiaotaoshen1990



shenxt.info

## </> Research:

Metabolomics & Bioinformatics

Microbiome

Multi-omics

Systems biology

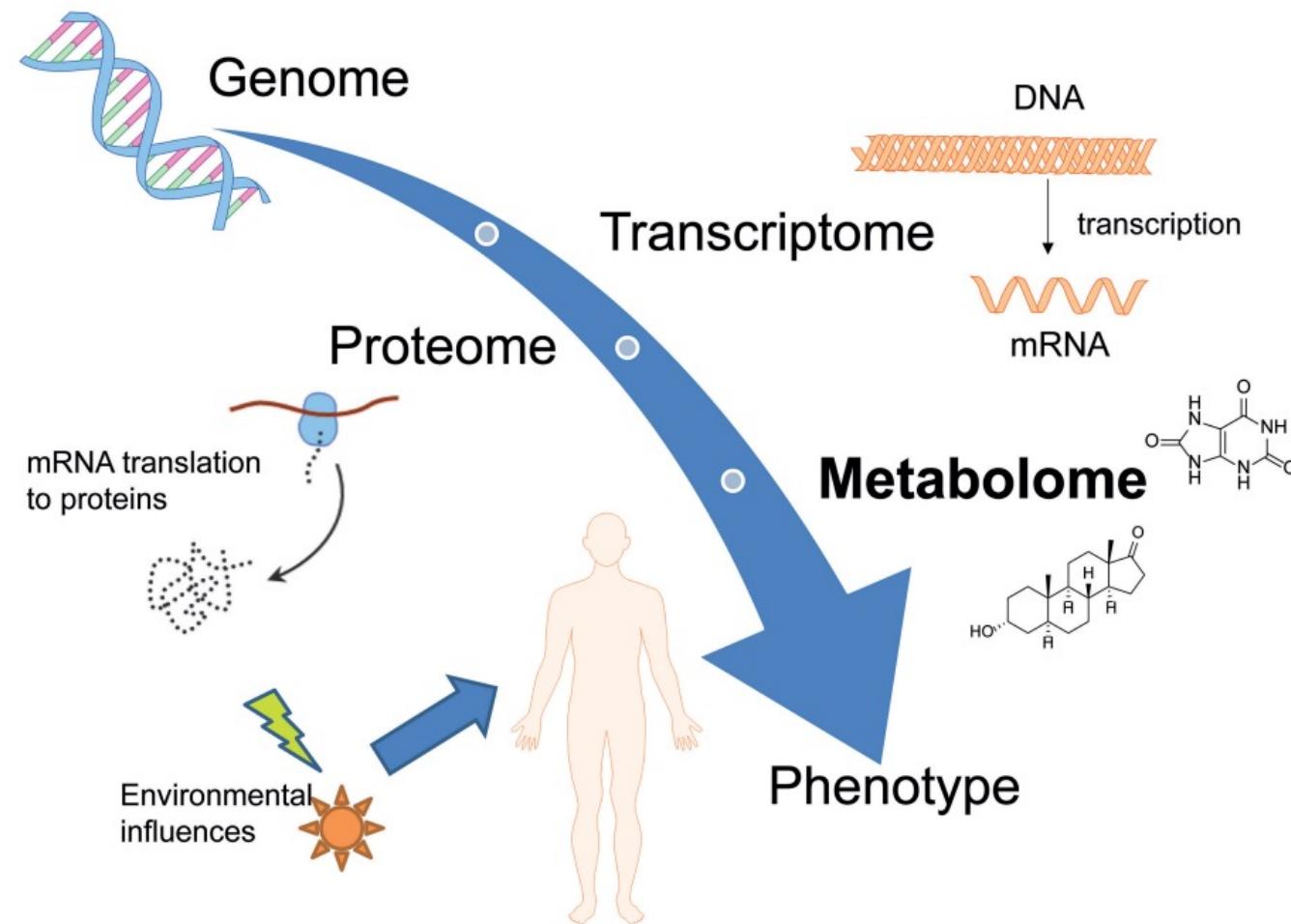
Aging

Precision medicine

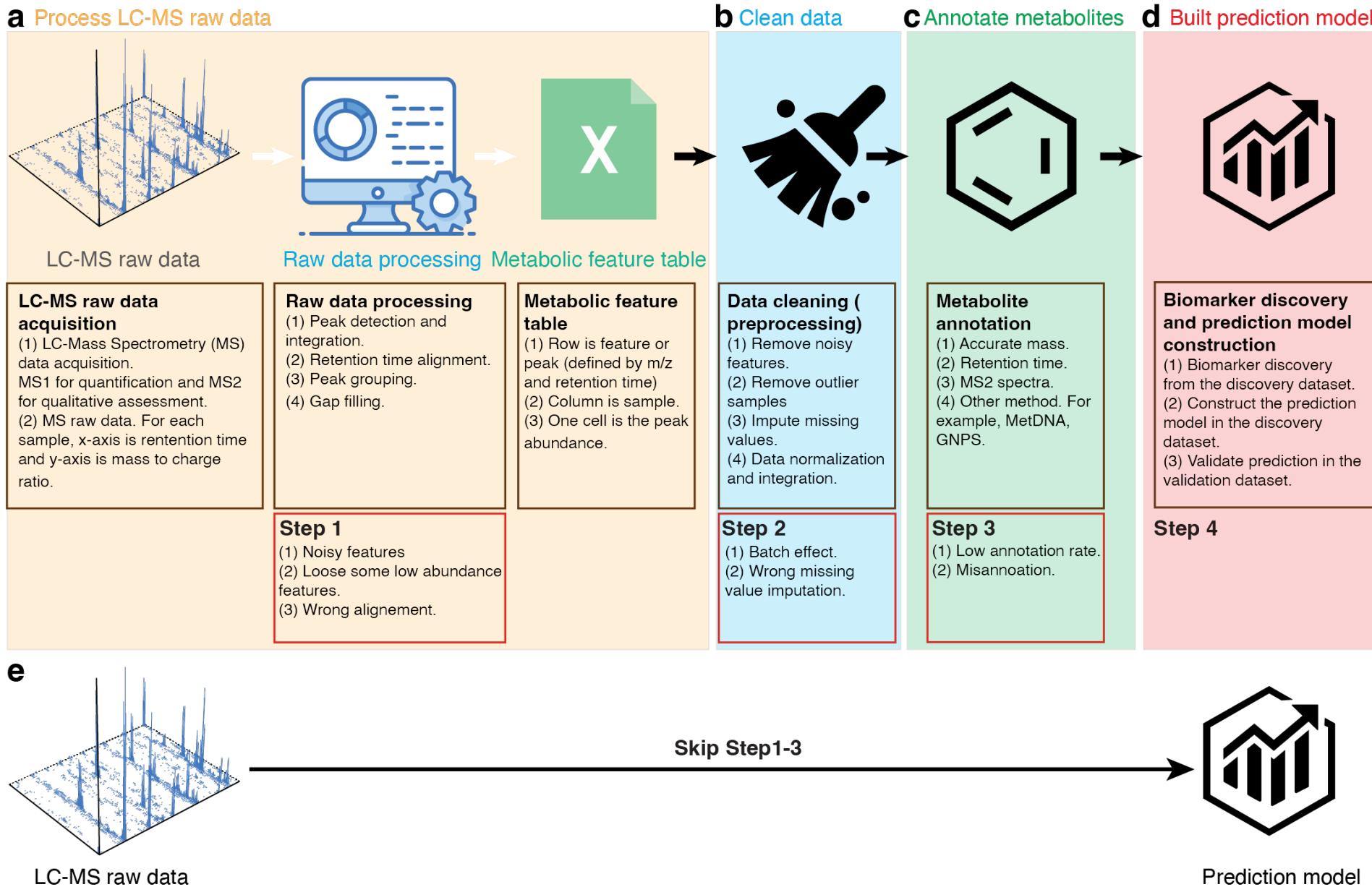
- </> **Background**
- </> **DeepPseudoMSI project: Deep Learning-based Pseudo-Mass Spectrometry Imaging**
- </> **Applications of deepPseudoMSI**
- </> **Summary**

# Background

# ➤ Metabolomics/lipidomics



# ► Conventional approaches for disease diagnosis using LC-MS



# ➤ Disadvantages of conventional methods

## **Identification of all metabolites is extremely difficult**

1. Most of the peaks have no MS<sup>2</sup> spectra (Only 30%-40% peaks have MS<sup>2</sup> spectra).
2. Some MS<sup>2</sup> spectra quality are not high.
3. We don't have enough MS<sup>2</sup> spectra.
4. We can not avoid the false identification.

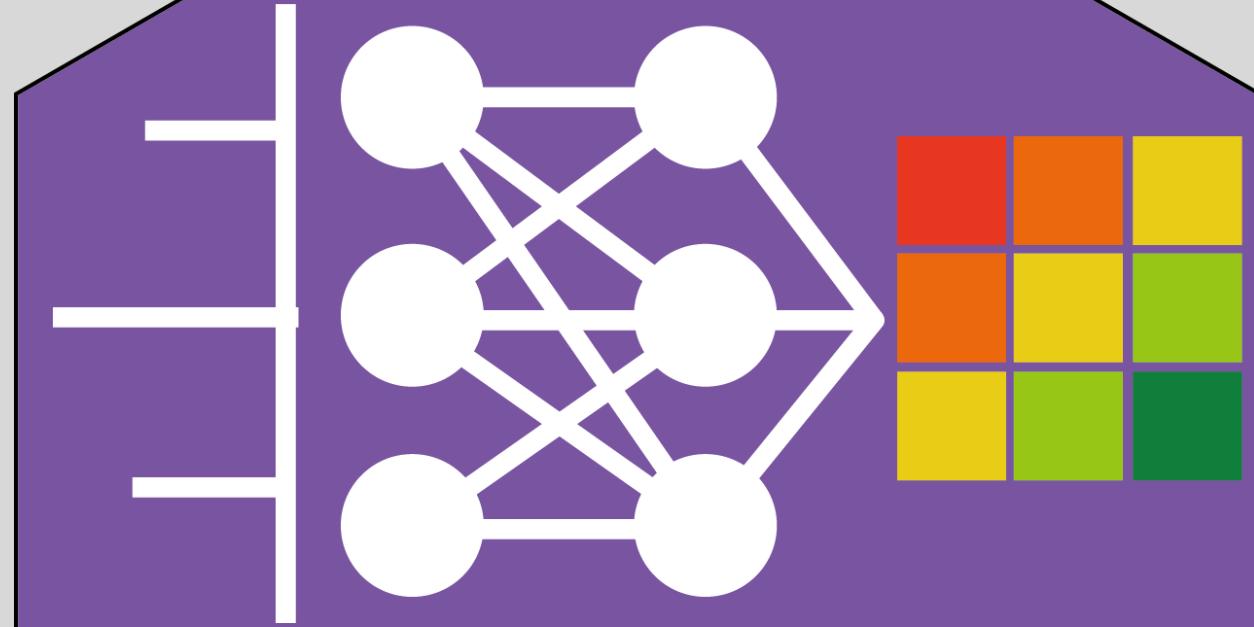
## **Prediction model with metabolites/lipids**

1. A lot of peaks with good prediction power will be lost.
2. It is extremely time-costing to identify metabolites (level 1, standard confirmation).

## **Prediction model with peaks**

1. Peaks are not robust.
2. It is difficult to use peaks in clinical application.

## **m/z, retention time and intensity (inter- intra-batch effect) shift**

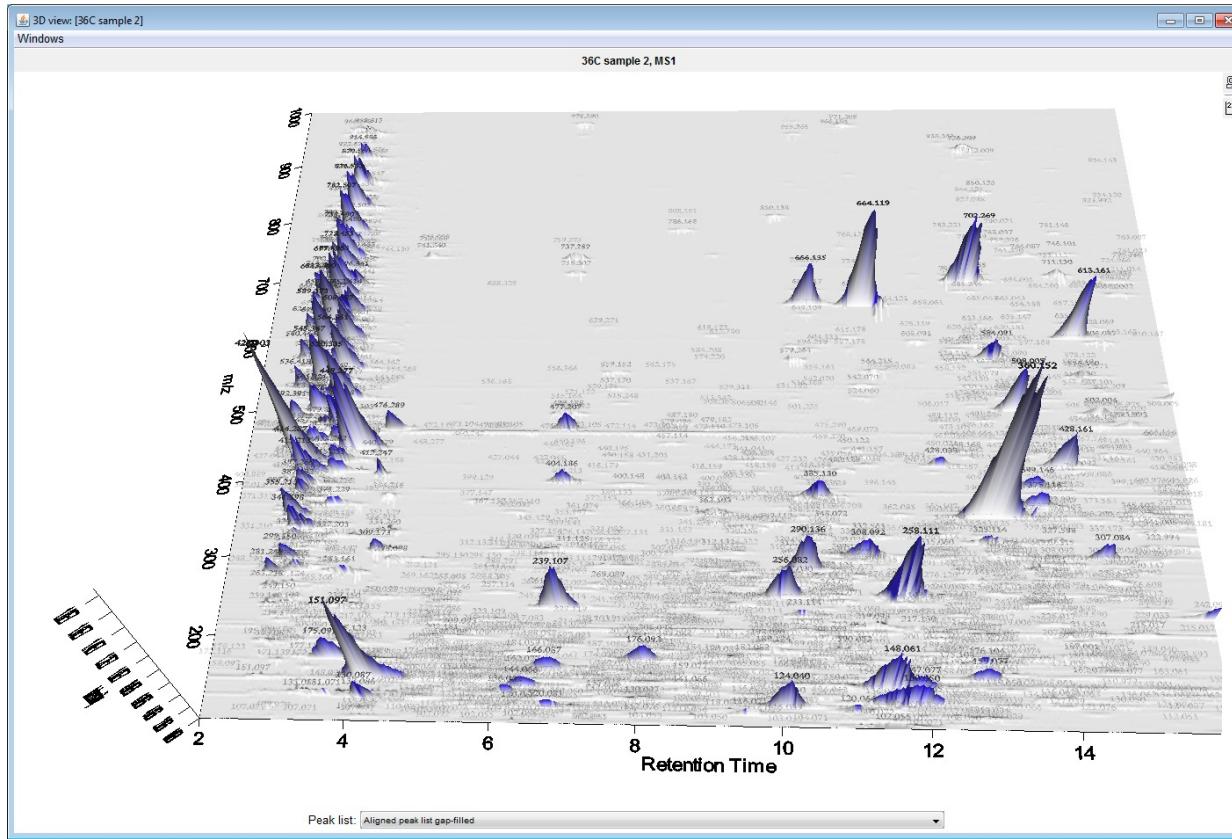


*deepPseudoMSI*

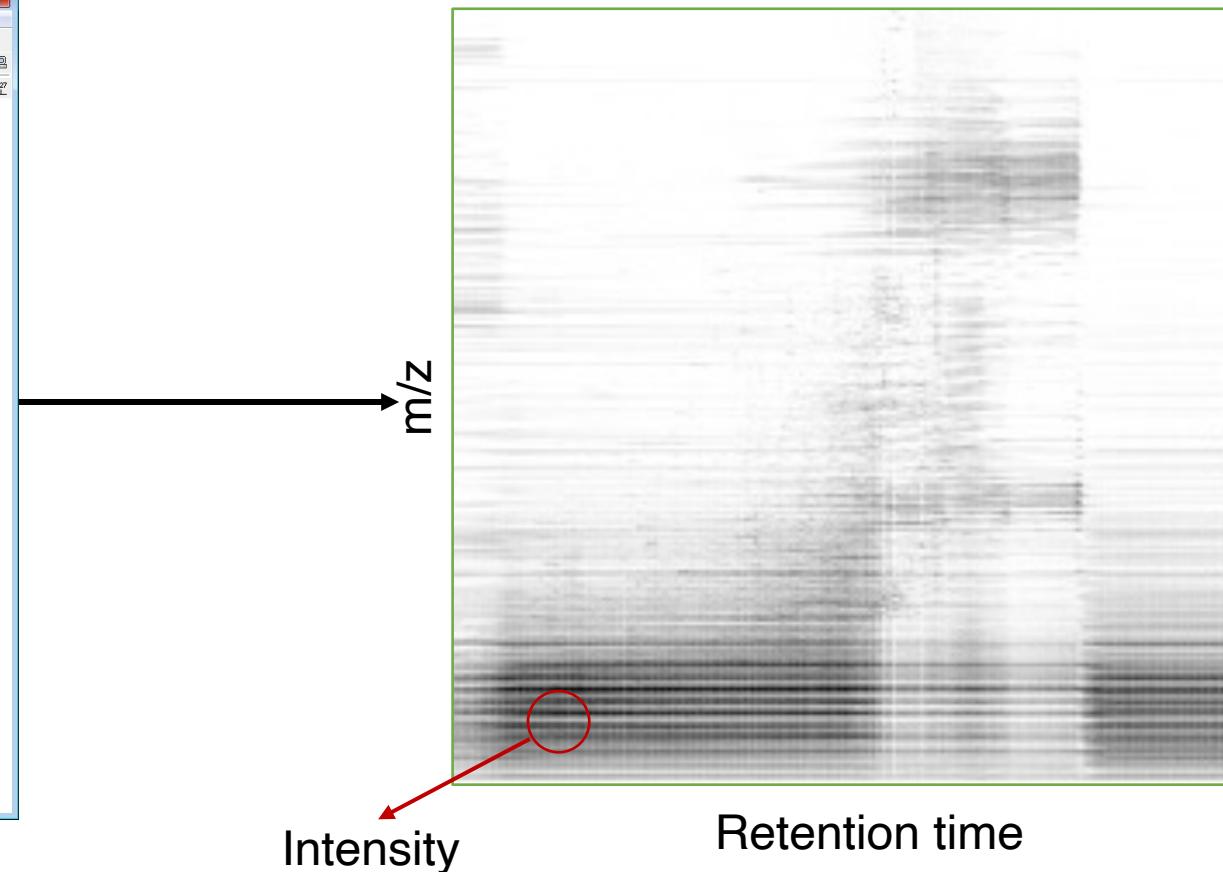
# DeepPseudo MSI project

# › LC-MS raw data: Image (pseudo Mass Spectrometry Image, pseudoMSI)

Raw MS data (Millions data points)



Metabolic image (pseudoMSI)



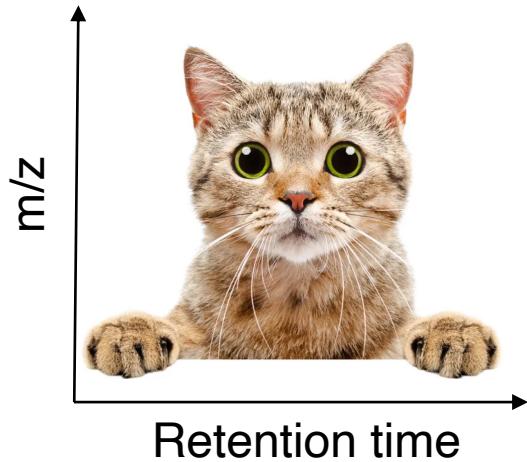
## ➤ Advantages of pseudoMS image

1. All the information are in the image



# ➤ Advantages of pseudoMS image

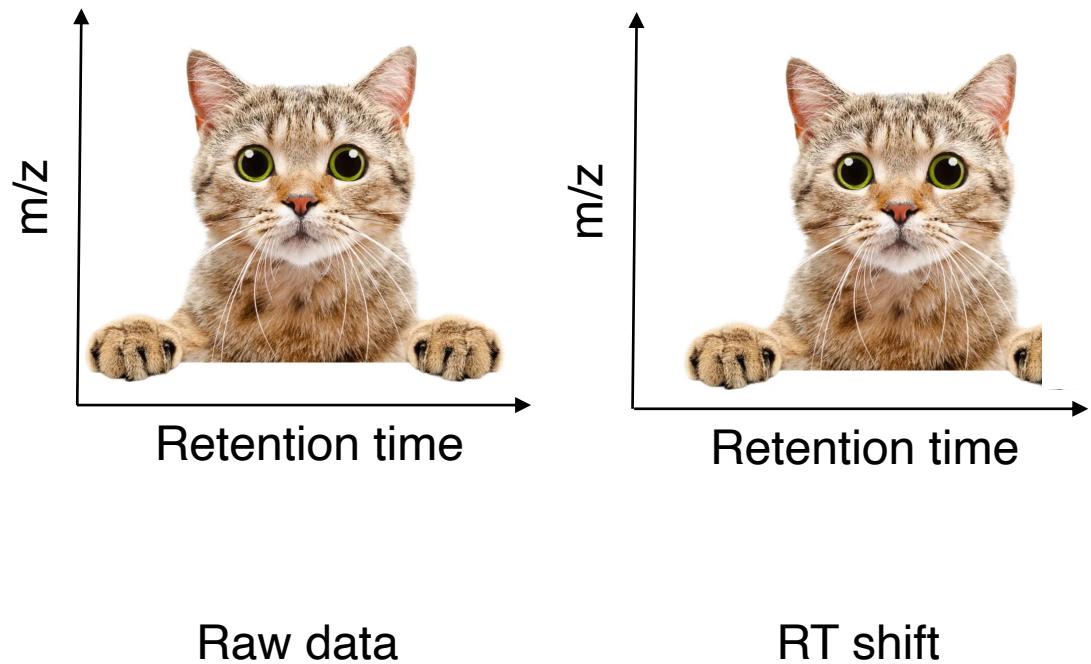
## 2. Handle batch effect



Raw data

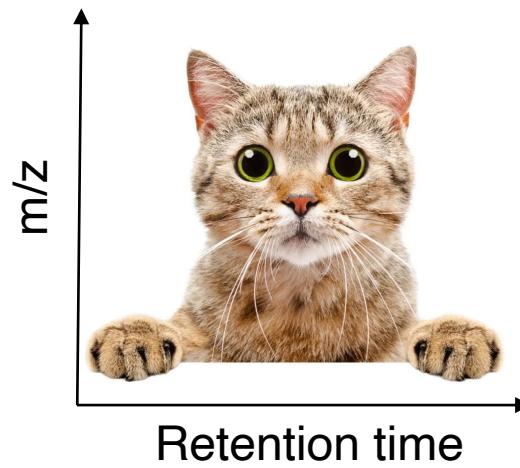
# ➤ Advantages of pseudoMS image

## 2. Handle batch effect

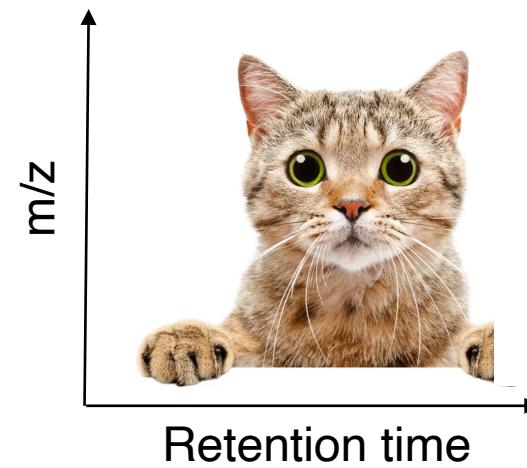


# ➤ Advantages of pseudoMS image

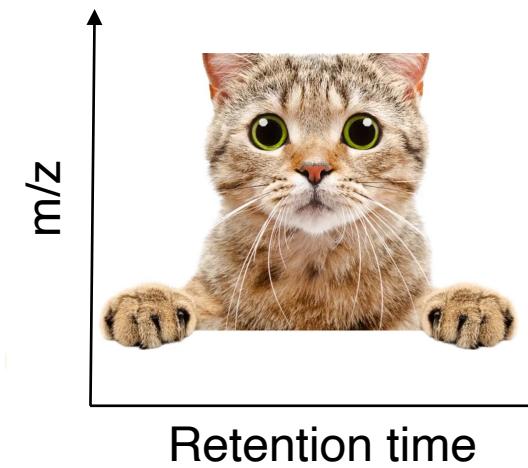
## 2. Handle batch effect



Raw data



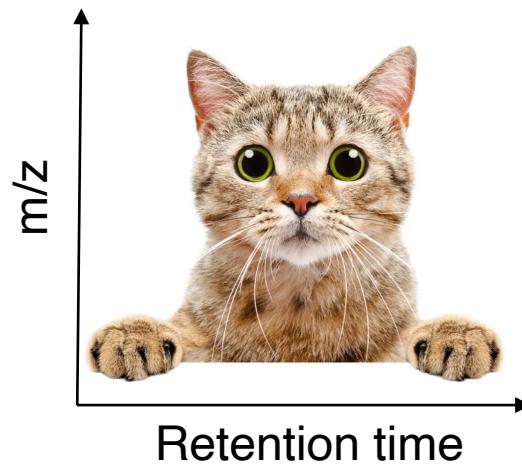
RT shift



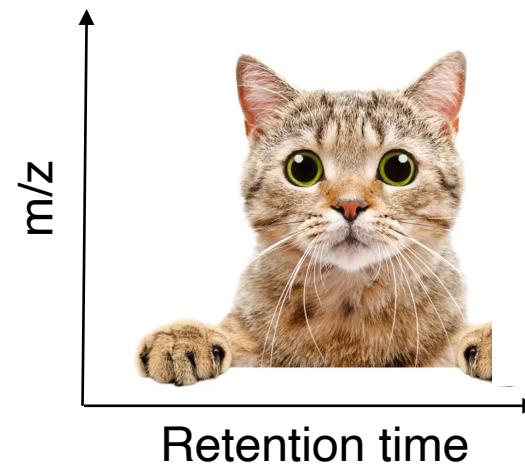
$m/z$  shift

# Advantages of pseudoMS image

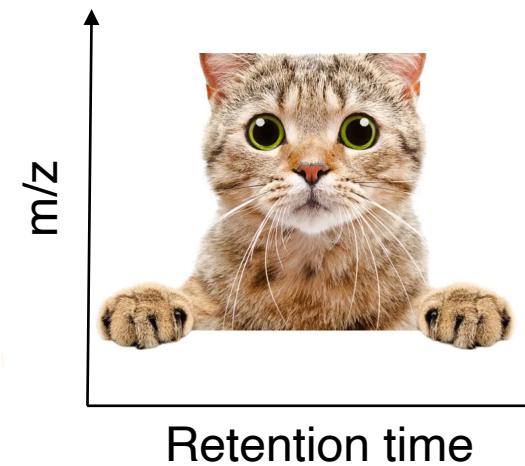
## 2. Handle batch effect



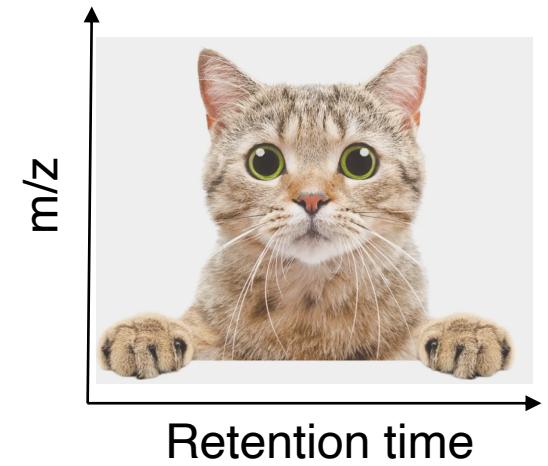
Raw data



RT shift



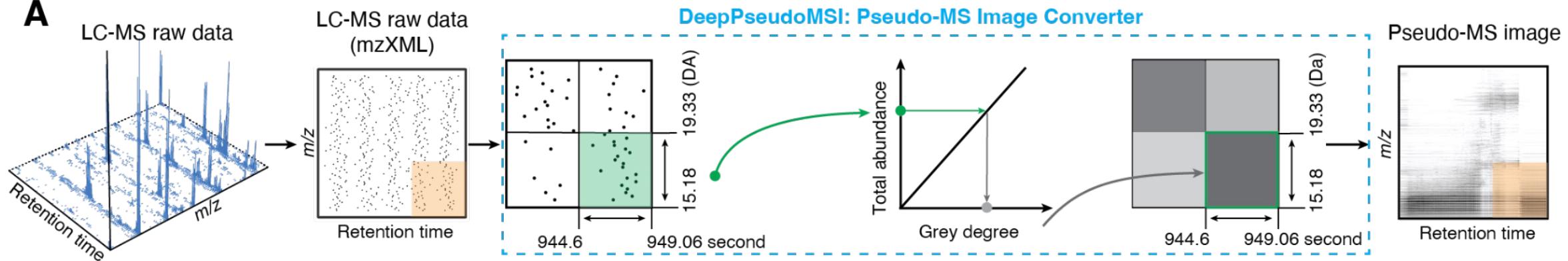
$m/z$  shift



Intensity shift

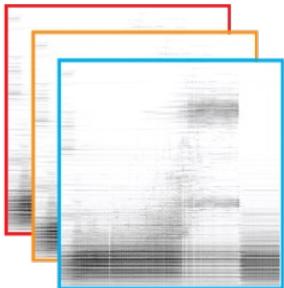
# ► Workflow of deepPseudoMSI

**A**



**B**

Training pseudo-MS images



DeepPseudoMSI: Image predictor

VGG16 Feature Extraction Network

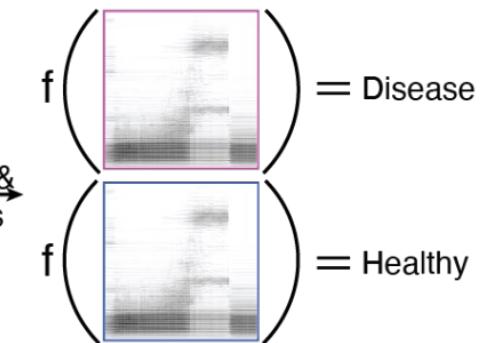
Global Average Pooling

Flatten

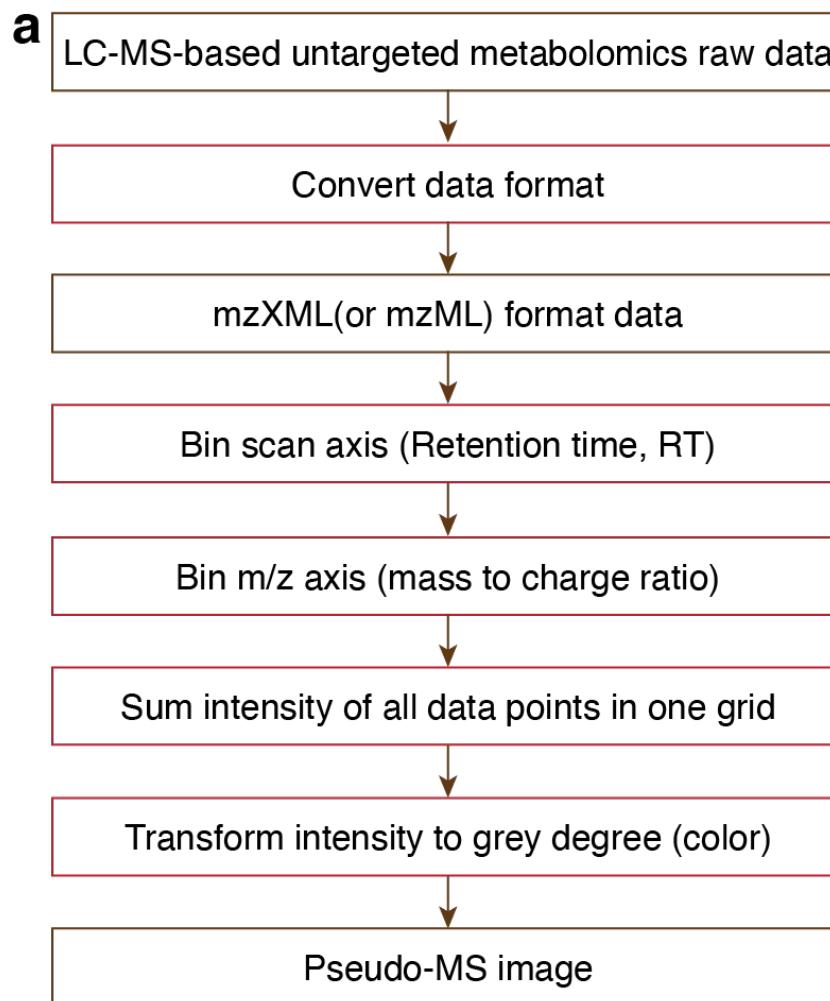
Dense Regression Network

Prediction & Diagnosis

New pseudo-MS images

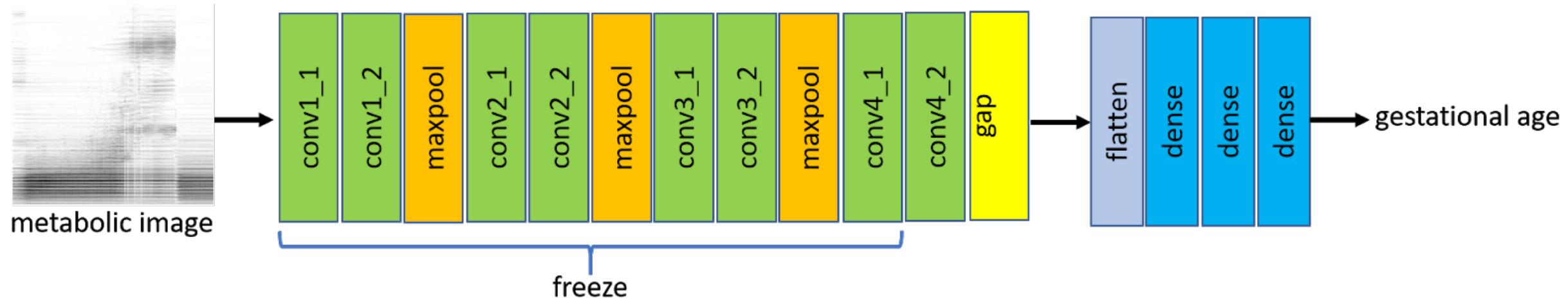


# ► PseudoMS Image converter

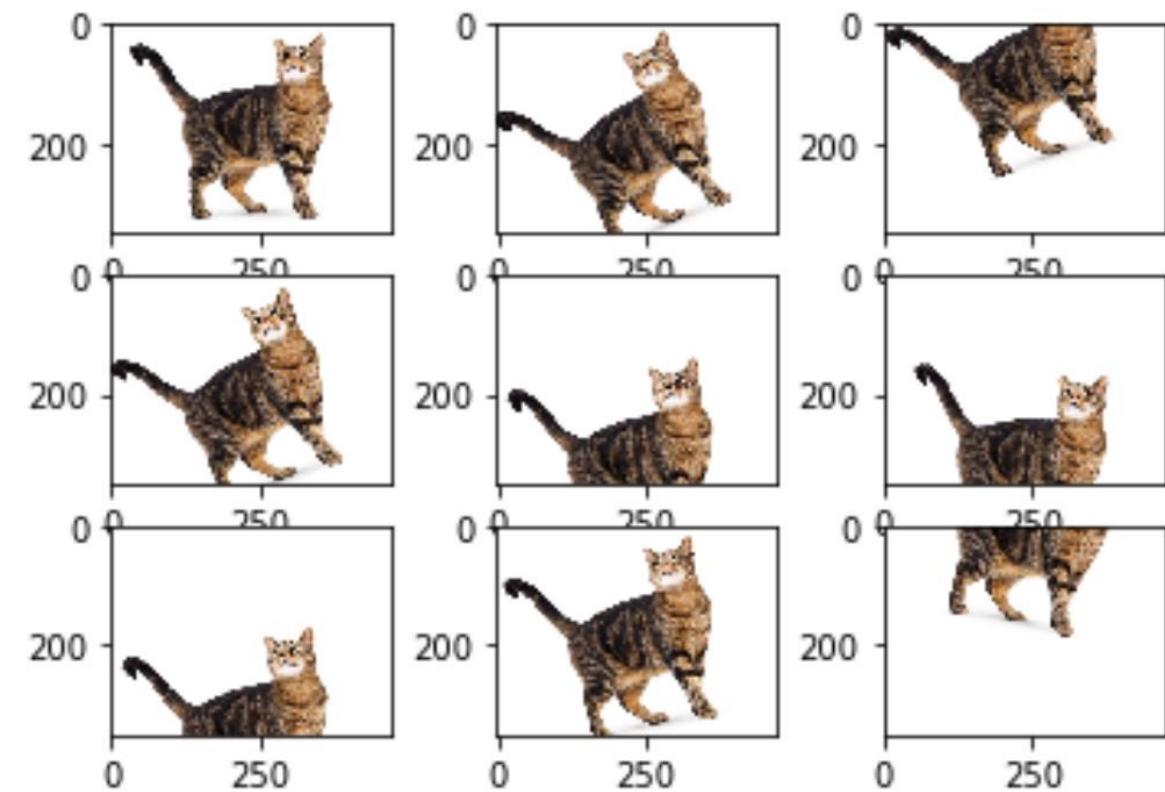


- a**
- Vendor format: Thermo raw, Waters raw, Sciex wiff, Agilent .D.
  - ProteoWizard. massConverter, fileConverter
  - mzXML is a XML (eXtensible Markup Language) based common file format for metabolomics/proteomics mass spectrometric data.
  - mzXML can be read using the R package MSnbase. And for the scan axis, we combine scans who are in the same range.
  - For the m/z axis, we combine m/z who are in the same range.
  - All the data points in the raw data are grouped into different grids. The median intensity is used to represent the intensity of one grid.
  - The color of one grid is represented by the grey degree. The grey degree is linearly transformed from the intensity of one grid.
  - Pseudo-MS image is a black and white graph with defined resolution. For example, 224 \* 224.

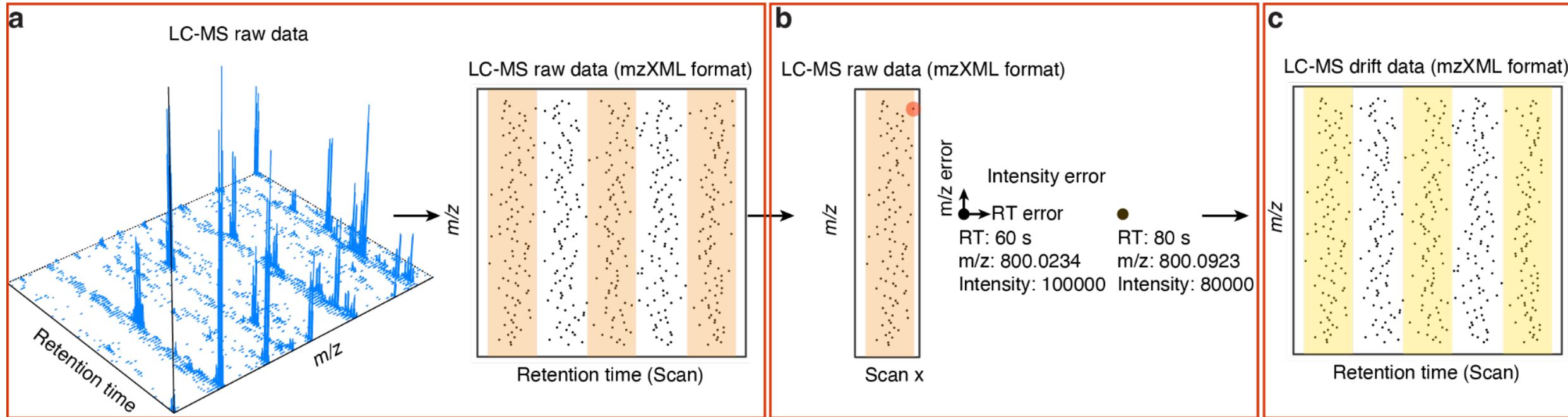
# › PseudoMS Image predictor



## ► Workflow of pseudo-MS image augmentation



# ► Workflow of pseudo-MS image augmentation

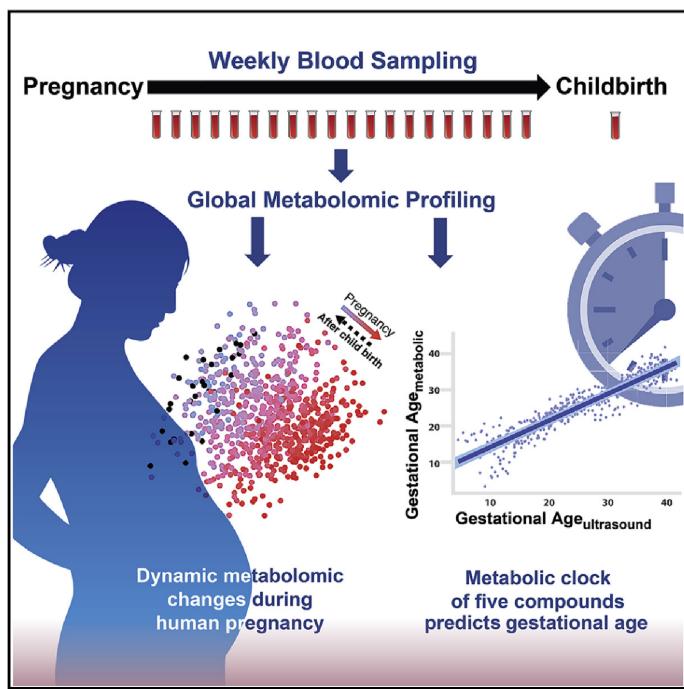


# ➤ Applications of deepPseudoMSI

## Cell

### Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women

#### Graphical Abstract



#### Resource

#### Authors

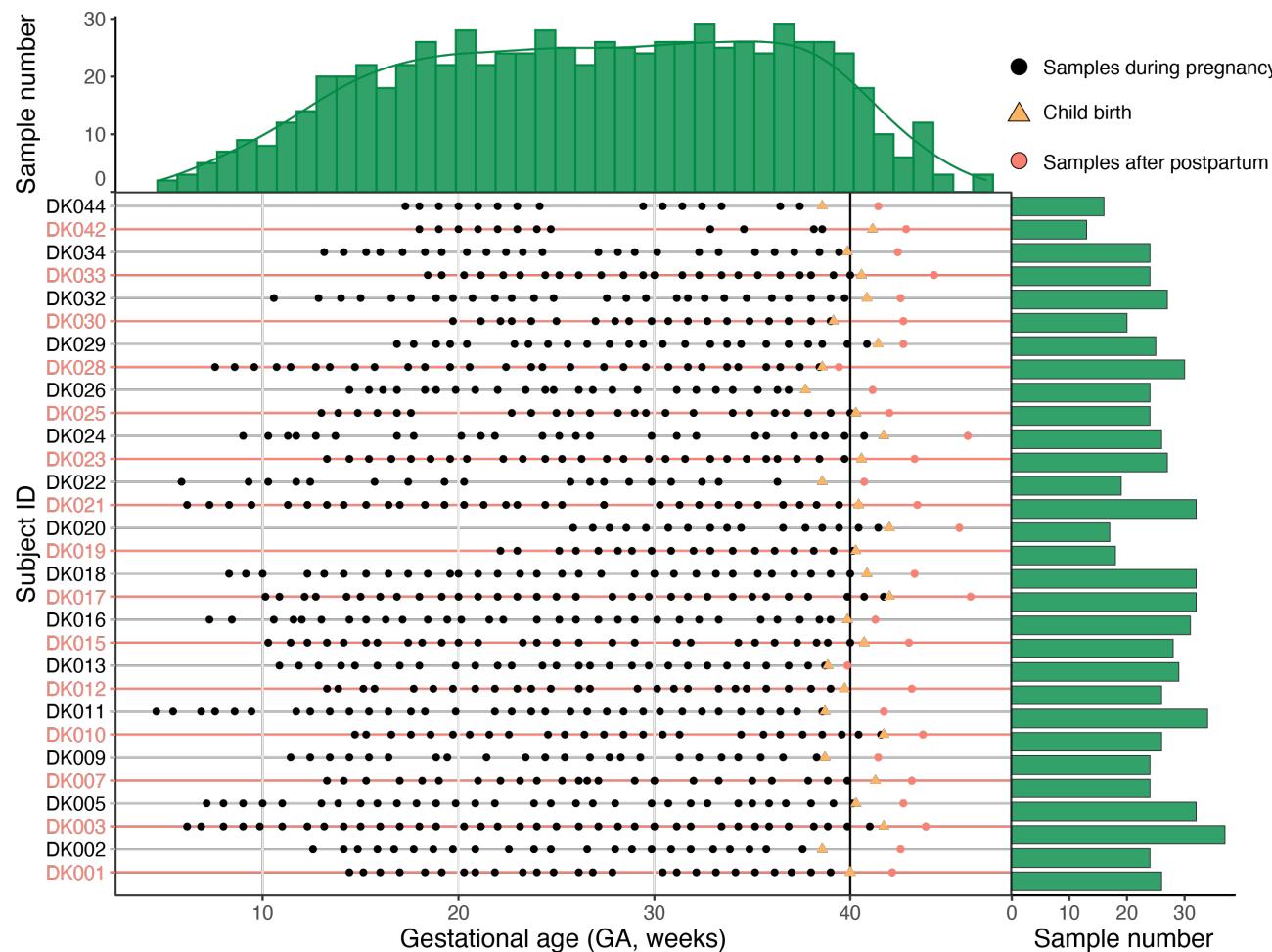
Liang Liang,  
Marie-Louise Hee Rasmussen,  
Brian Piening, ..., Hanyah Zackriah,  
Michael Snyder, Mads Melbye

#### Correspondence

mpsnyder@stanford.edu (M.S.),  
mmelbye@stanford.edu (M.M.)

#### In Brief

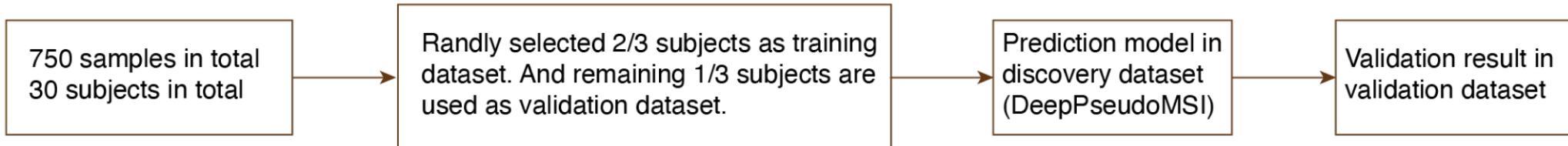
Identification of blood metabolites in pregnant women that can accurately predict gestational age and provide insights into pregnancy variations undetected by ultrasound.



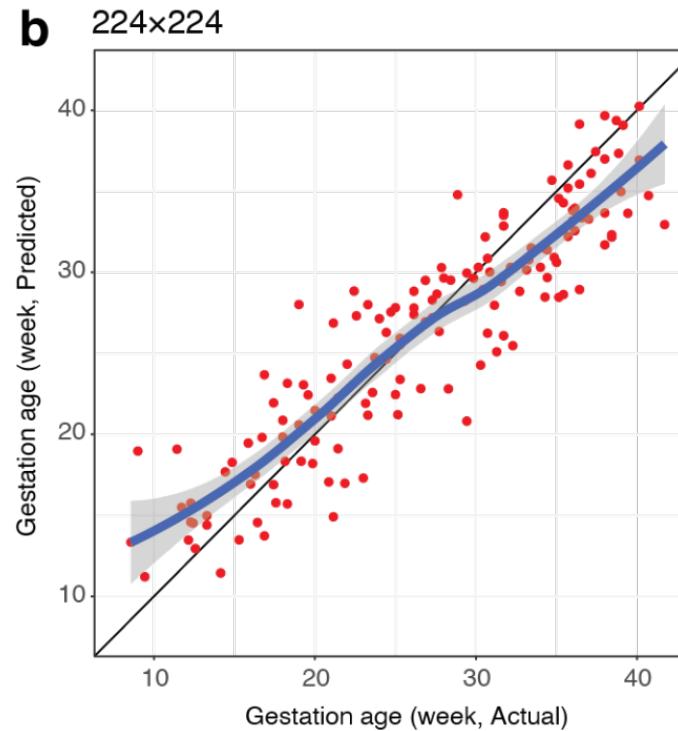
Data from Liang et al, Cell, 2020.

# ➤ Resolution optimization

a

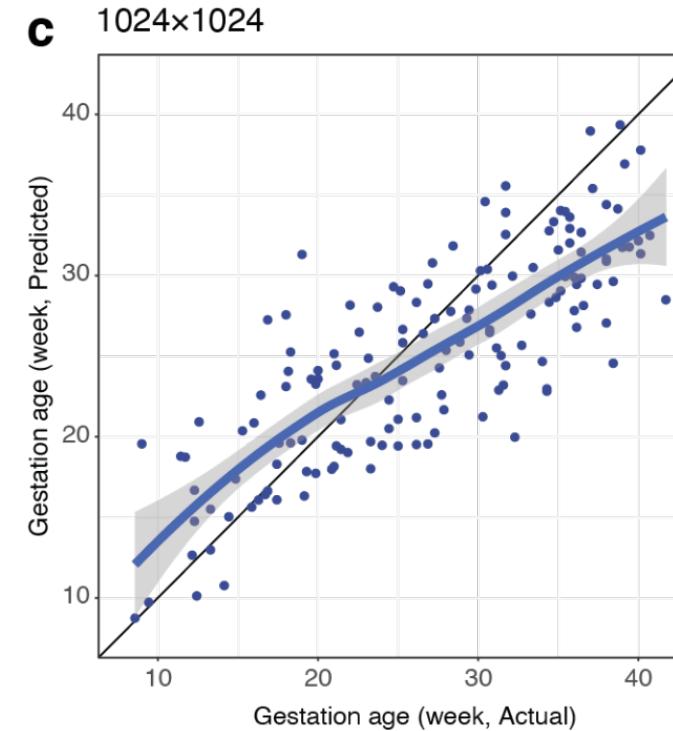


b

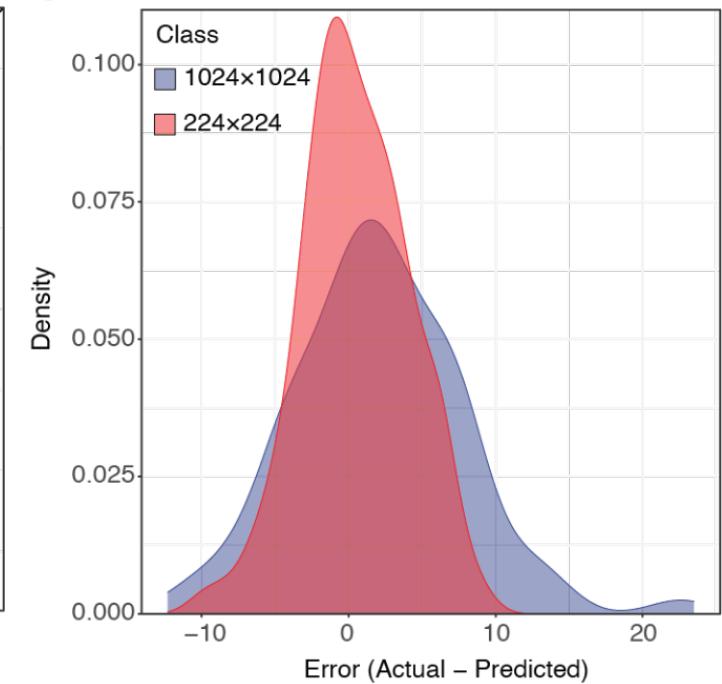


c

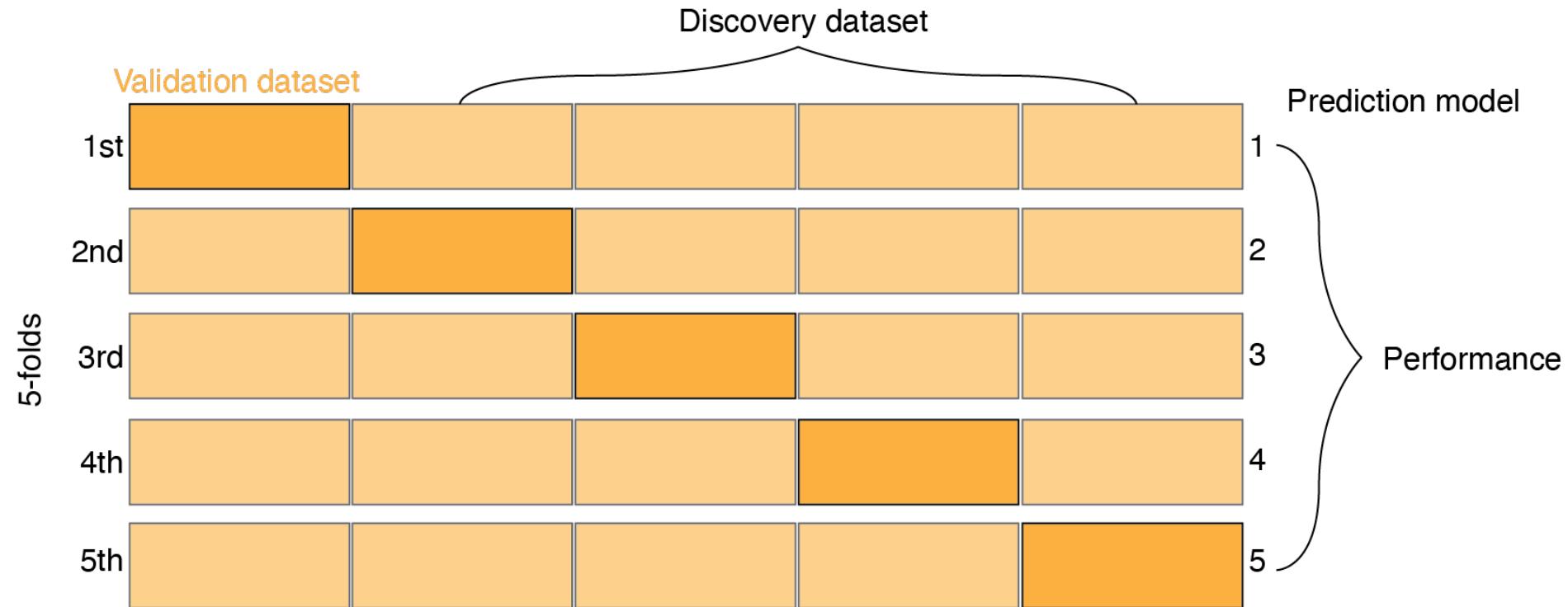
1024×1024



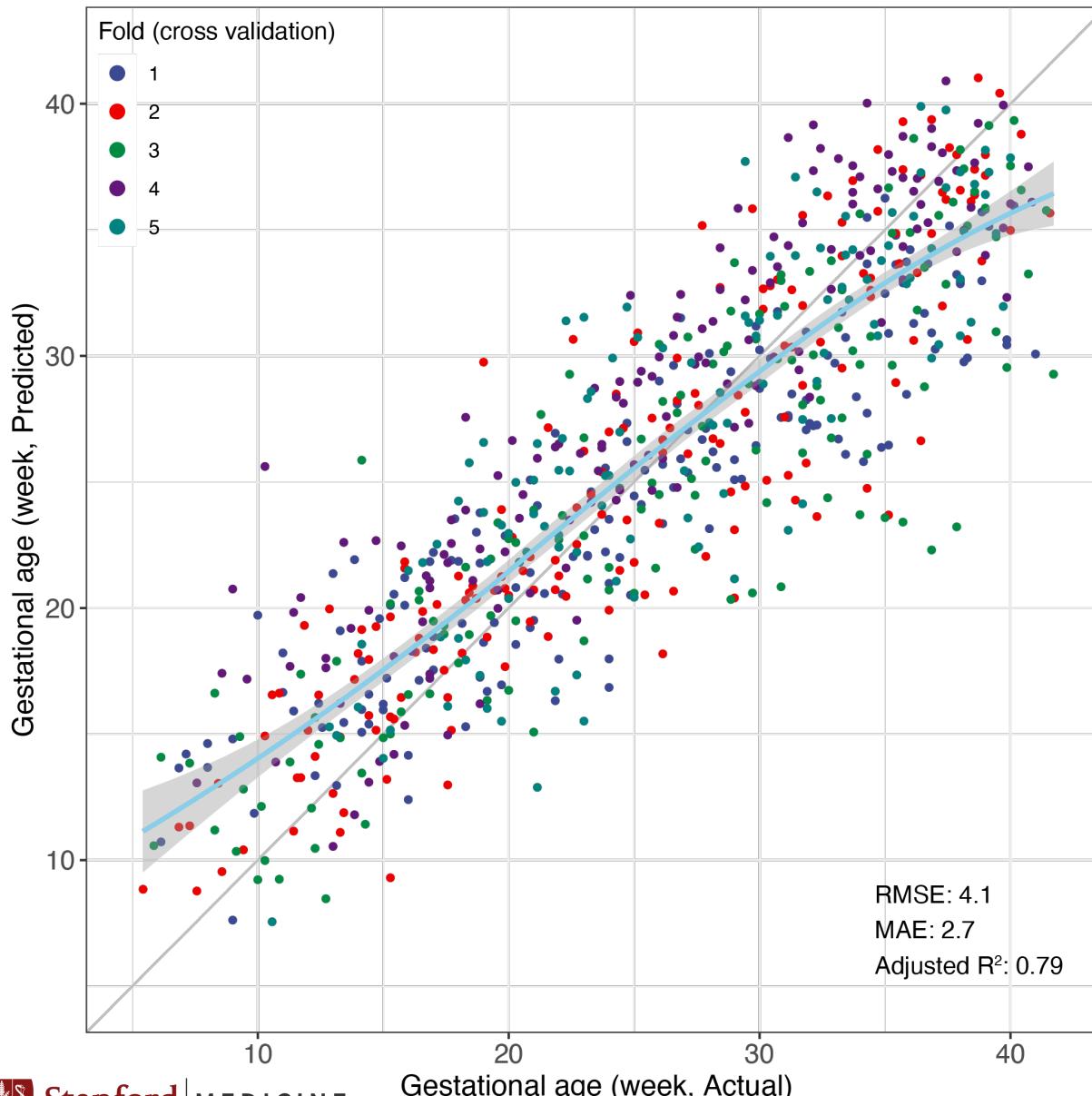
d



# DeepPseudoMSI predicts gestation age



# DeepPseudoMSI predicts gestation age

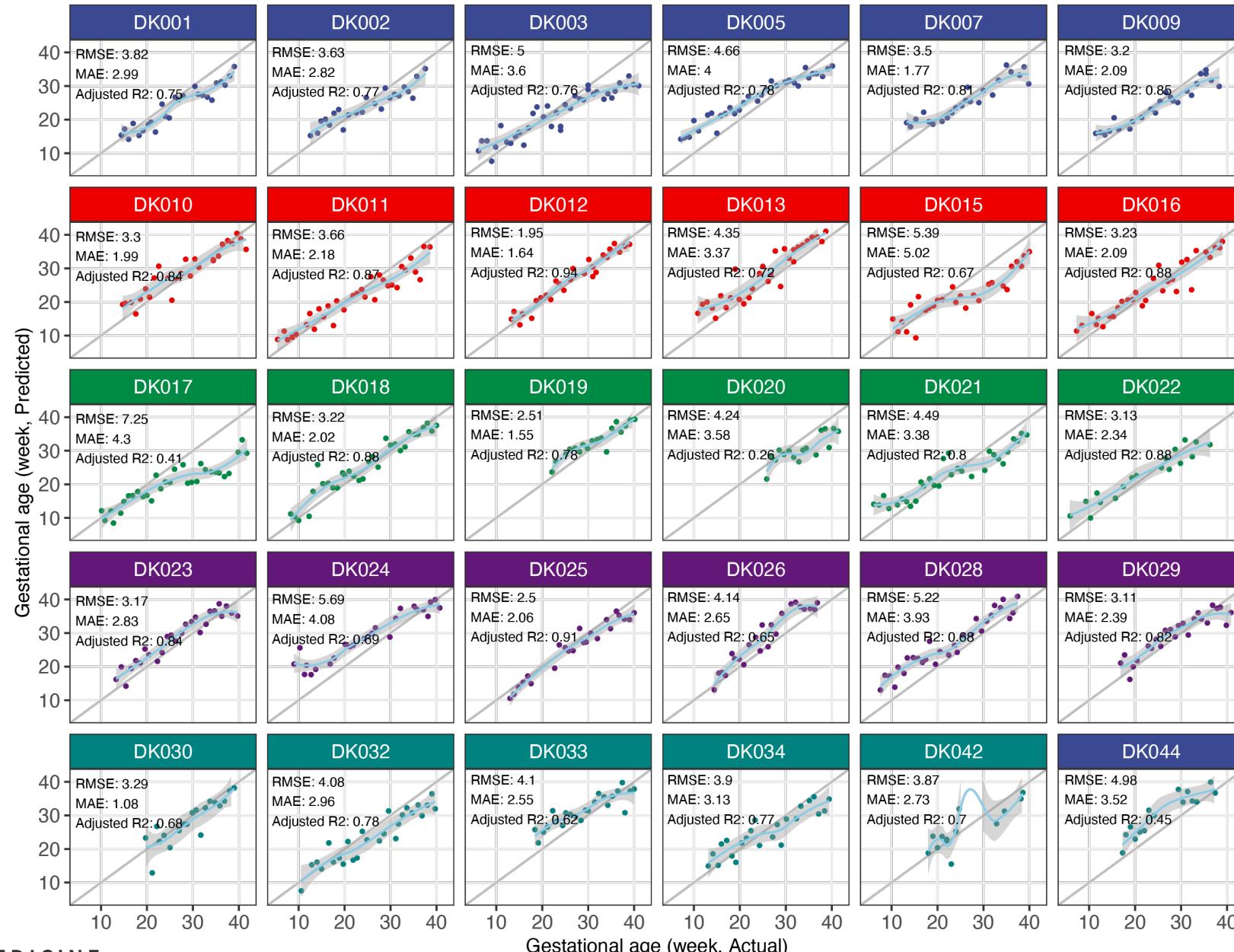


RMSE (Root Mean Squared Error): 4.1 weeks

MAE (Mean Average Error): 2.7 weeks

Adjusted R<sup>2</sup>: 0.79

# DeepPseudoMSI predicts gestation age (Individual)



Fold 1

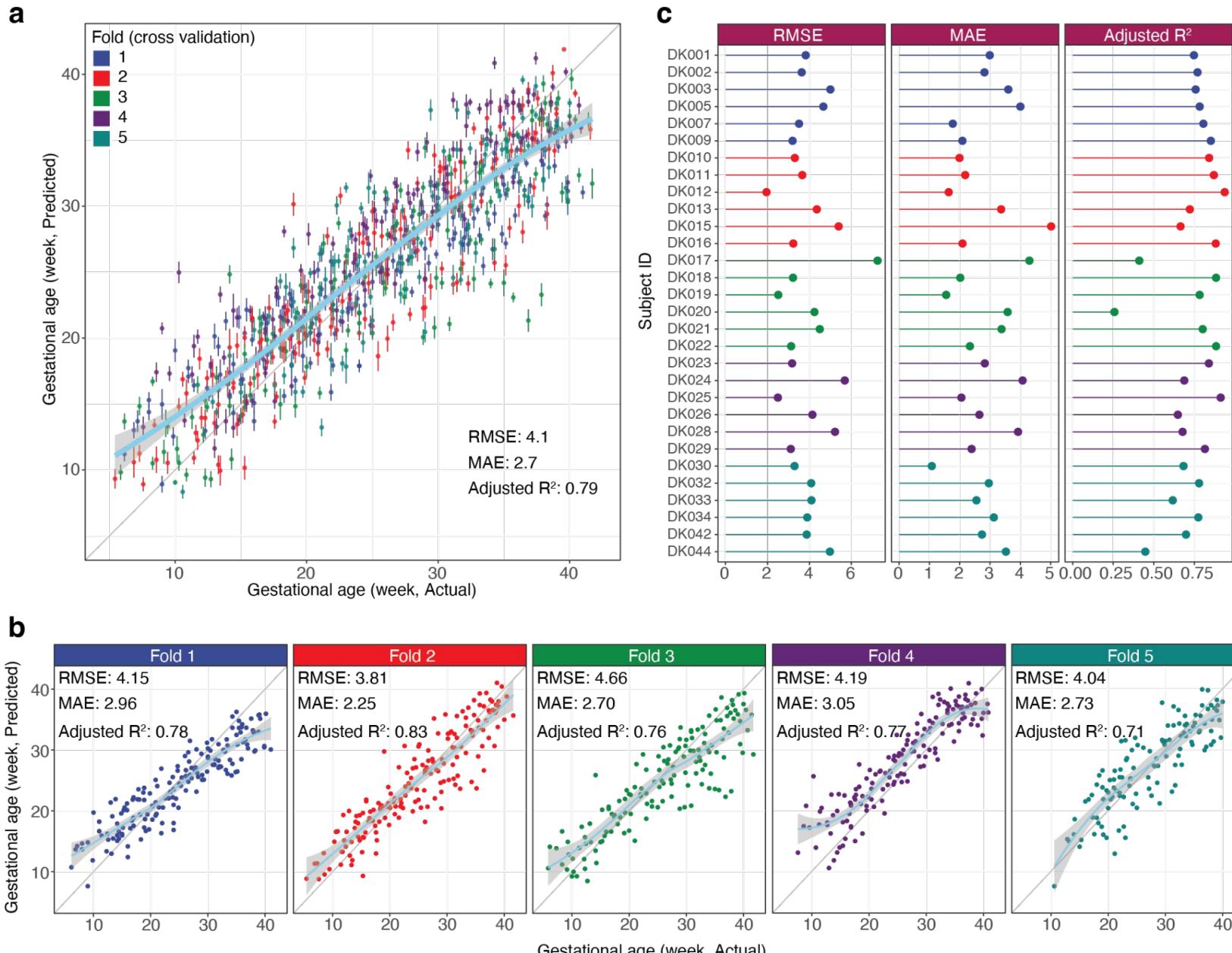
Fold 2

Fold 3

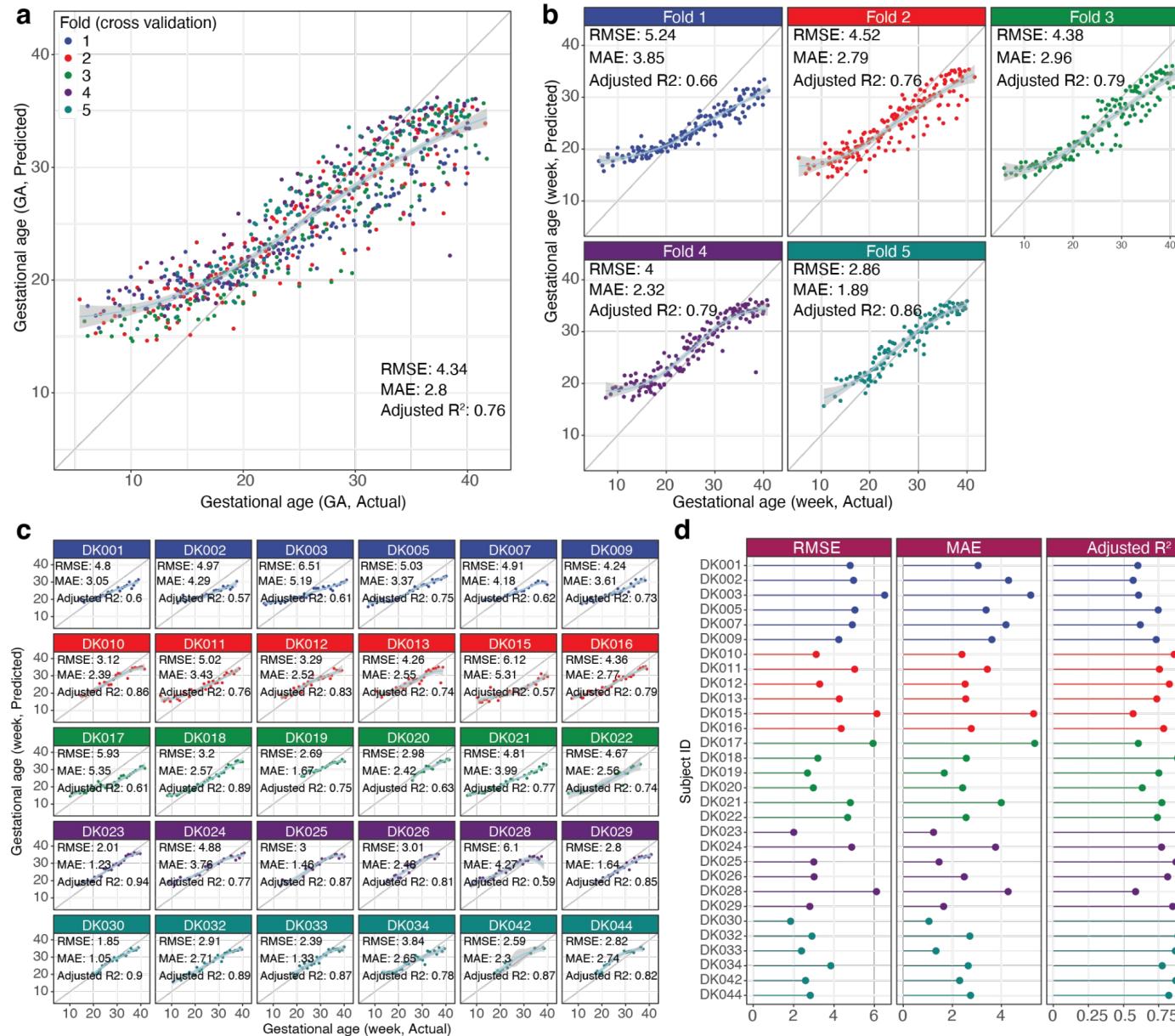
Fold 4

Fold 5

# DeepPseudoMSI predicts gestation age



# ➤ Conventional method



# ➤ Other public databases

Received: 4 October 2021 | Revised: 9 January 2022 | Accepted: 13 January 2022

DOI: 10.1002/ijc.33943

## TUMOR MARKERS AND SIGNATURES



### A serum lipidomics study for the identification of specific biomarkers for endometrial polyps to distinguish them from endometrial cancer or hyperplasia

Xingxu Yan<sup>1</sup> | Wen Zhao<sup>2</sup> | Jinxia Wei<sup>1</sup> | Yaqi Yao<sup>1</sup> | Guijiang Sun<sup>3</sup> |  
Lei Wang<sup>4</sup> | Wenqing Zhang<sup>1</sup> | Siyu Chen<sup>1</sup> | Wenjie Zhou<sup>1</sup> | Huan Zhao<sup>1</sup> |  
Xiaomeng Li<sup>1</sup> | Yu Xiao<sup>5</sup> | Yubo Li<sup>1</sup>

## scientific reports

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open Access](#) | Published: 17 March 2020

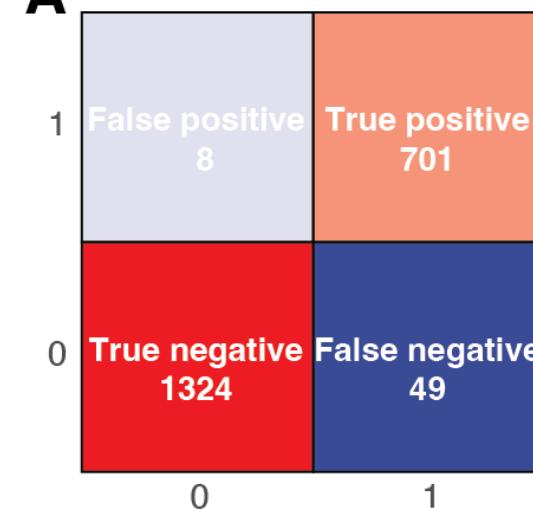
### Sex Differences in Colon Cancer Metabolism Reveal A Novel Subphenotype

[Yuping Cai](#), [Nicholas J. W. Rattray](#), [Qian Zhang](#), [Varvara Mironova](#), [Alvaro Santos-Neto](#), [Kuo-Shun Hsu](#),  
[Zahra Rattray](#), [Justin R. Cross](#), [Yawei Zhang](#), [Philip B. Paty](#), [Sajid A. Khan](#) & [Caroline H. Johnson](#)

[Scientific Reports](#) **10**, Article number: 4905 (2020) | [Cite this article](#)

4498 Accesses | 19 Citations | 40 Altmetric | [Metrics](#)

**A** MTBLS3444

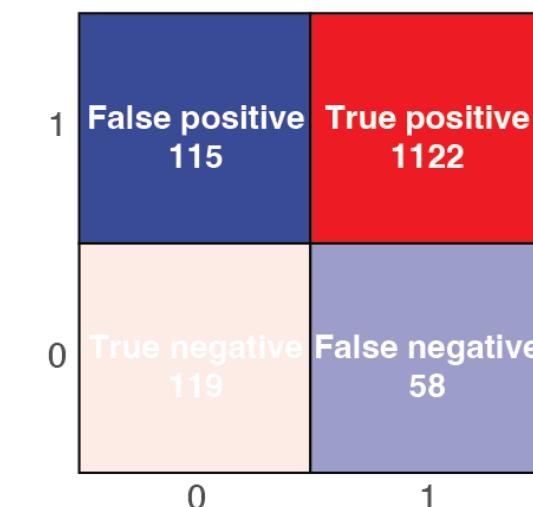


$$\text{Accuracy} = \frac{701 + 1324}{701 + 1324 + 8 + 49} = 97.3\%$$

$$\text{Specificity} = \frac{1324}{1324 + 49} = 96.4\%$$

$$\text{Sensitivity} = \frac{701}{701 + 8} = 98.9\%$$

**B** MTBLS1129

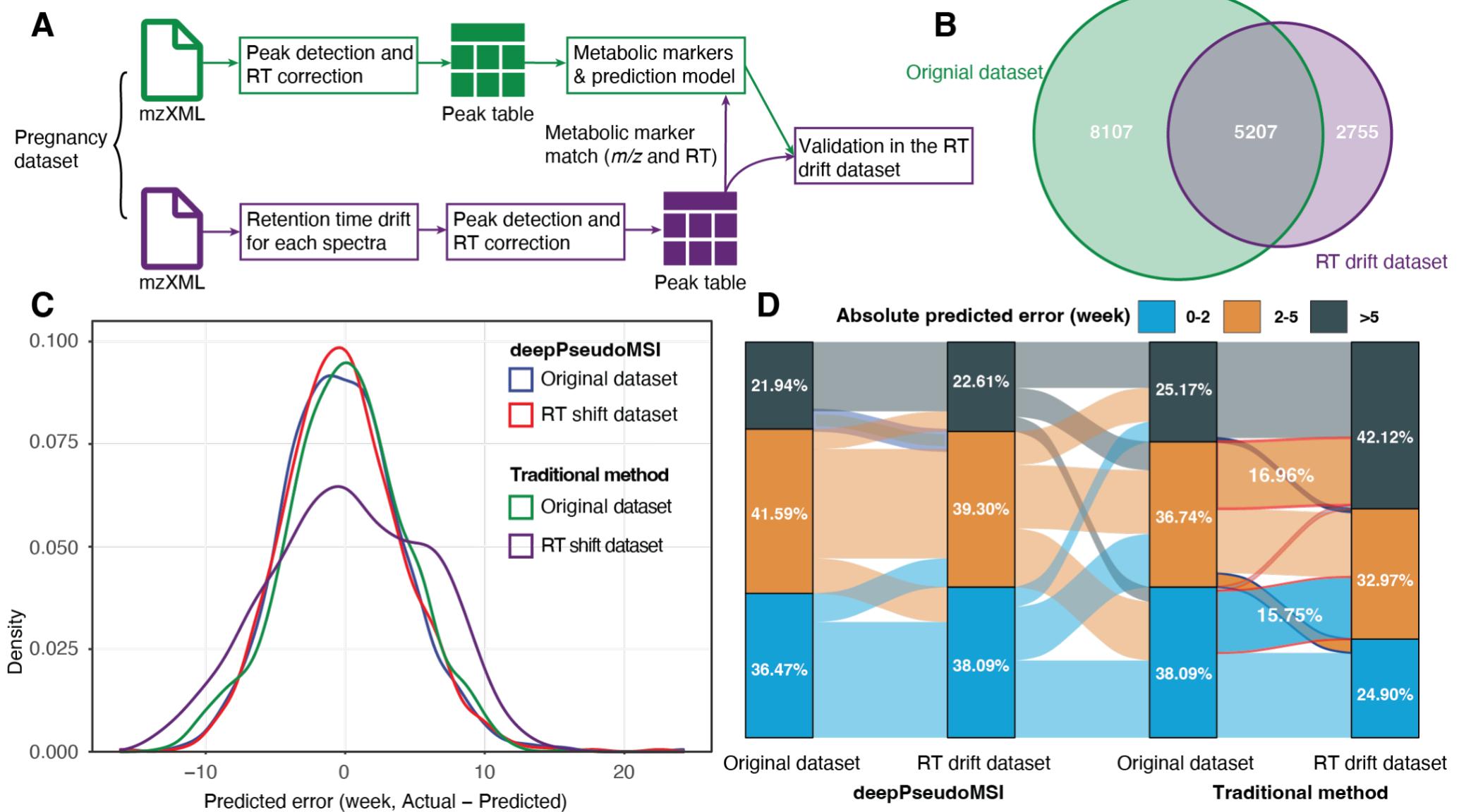


$$\text{Accuracy} = \frac{1122 + 119}{1122 + 119 + 115 + 58} = 87.8\%$$

$$\text{Specificity} = \frac{119}{119 + 58} = 67.2\%$$

$$\text{Sensitivity} = \frac{1122}{1122 + 115} = 90.7\%$$

# ➤ Evaluation of deepPseudoMSI



## ➤ Future works

The screenshot shows a web browser window with the title "DeepPseudoMSI". The top navigation bar includes links for "About", "Case study", "Tutorial", "Team", "Tools", and "Contact". Below the title, there is a large heading "DeepPseudoMSI project" followed by a descriptive paragraph about the project's purpose. At the bottom left are buttons for "Source code </>" and "Ask a question >". On the right side, there is a diagram illustrating the workflow: "LC-MS raw data" (3D plot of m/z vs. retention time) is converted into a "Pseudo-MS image" (2D grayscale plot of m/z vs. retention time), which is then processed by a "Deep learning" model (represented by a neural network diagram). A detailed inset titled "DeepPseudoMSI: Pseudo-MS Image Converter" shows how a specific peak at 944.6 Da is mapped onto a grid of pixels, with dimensions of 15.18 seconds in retention time and 19.33 Da in m/z.

# DeepPseudoMSI

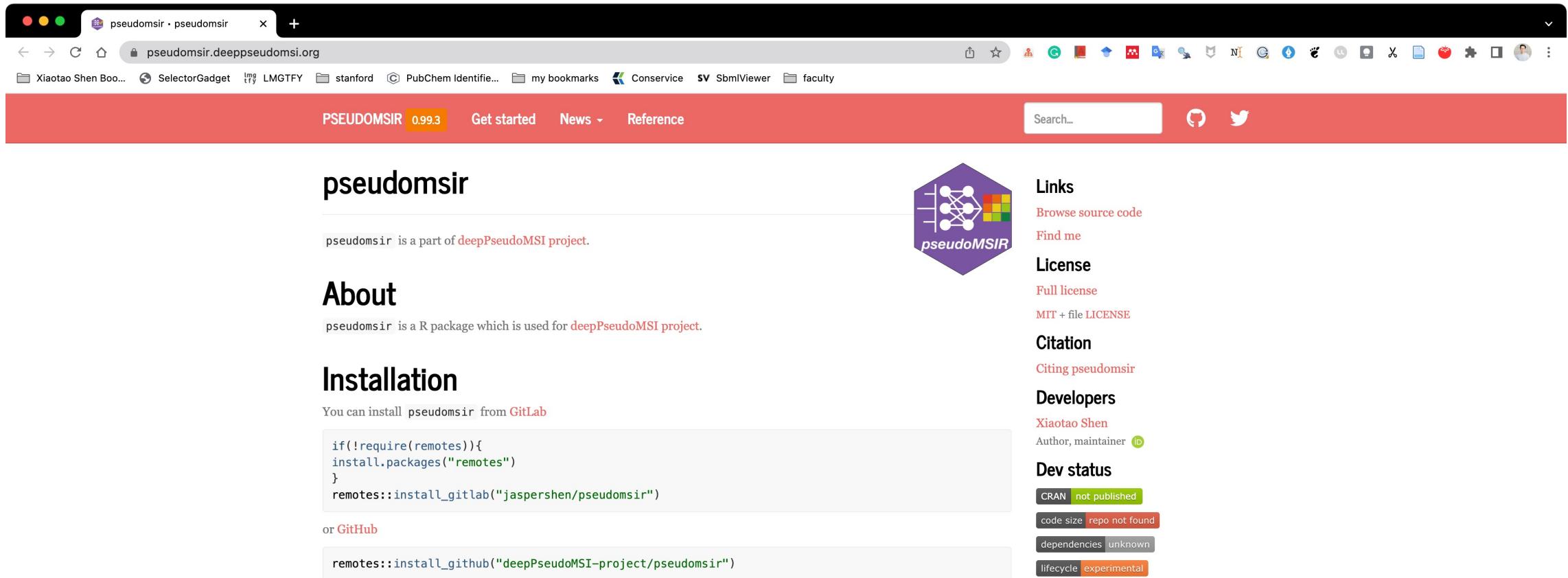
## Deep learning-based pseudo-mass spectrometry imaging analysis for precision medicine

The **deepPseudoMSI** project is the first method that convert LC-MS raw data to “images” and then process them using deep learning method for diagnosis. It contains two parts.

## 1. Pseudo-MS image converter

One LC-MS raw data usually contains millions of data points, so we need to divide it into different pixels (or grids) based on the

# ➤ Future works



The screenshot shows a web browser window with the URL [pseudomsir · pseudomsir](https://pseudomsir.deeppseudoMSI.org). The page has a red header bar with the text "PSEUDOMSIR 0.99.3" and navigation links for "Get started", "News", and "Reference". A search bar and social media icons for GitHub and Twitter are also present. The main content area features a purple hexagonal logo with a neural network icon and the text "pseudoMSIR". The page is divided into several sections: "About" (describing it as a part of the deepPseudoMSI project), "Installation" (with R code snippets for GitLab and GitHub), "Usage" (with a note to see the "Get started" page), "Need help?" (with contact information including email, Twitter, and address), and "Links", "License", "Citation", "Developers", and "Dev status" sections on the right.

## pseudomsir

pseudomsir is a part of [deepPseudoMSI](#) project.

## About

pseudomsir is a R package which is used for [deepPseudoMSI](#) project.

## Installation

You can install `pseudomsir` from [GitLab](#)

```
if(!require(remotes)){
  install.packages("remotes")
}
remotes::install_gitlab("jaspershen/pseudomsir")
```

or [GitHub](#)

```
remotes::install_github("deepPseudoMSI-project/pseudomsir")
```

## Usage

Please see the [Get started](#) page to get the instruction of `pseudomsir`.

## Need help?

If you have any questions about `pseudomsir`, please don't hesitate to email me ([shenxt@stanford.edu](mailto:shenxt@stanford.edu)).

➤ [shenzutao1990](#)  
✉ [shenxt@stanford.edu](mailto:shenxt@stanford.edu)  
🐦 [Twitter](#)  
📍 [M339, Alway building, Cooper Lane, Palo Alto, CA 94304](#)

### Links

[Browse source code](#)  
[Find me](#)

### License

[Full license](#)  
[MIT + file LICENSE](#)

### Citation

[Citing pseudomsir](#)

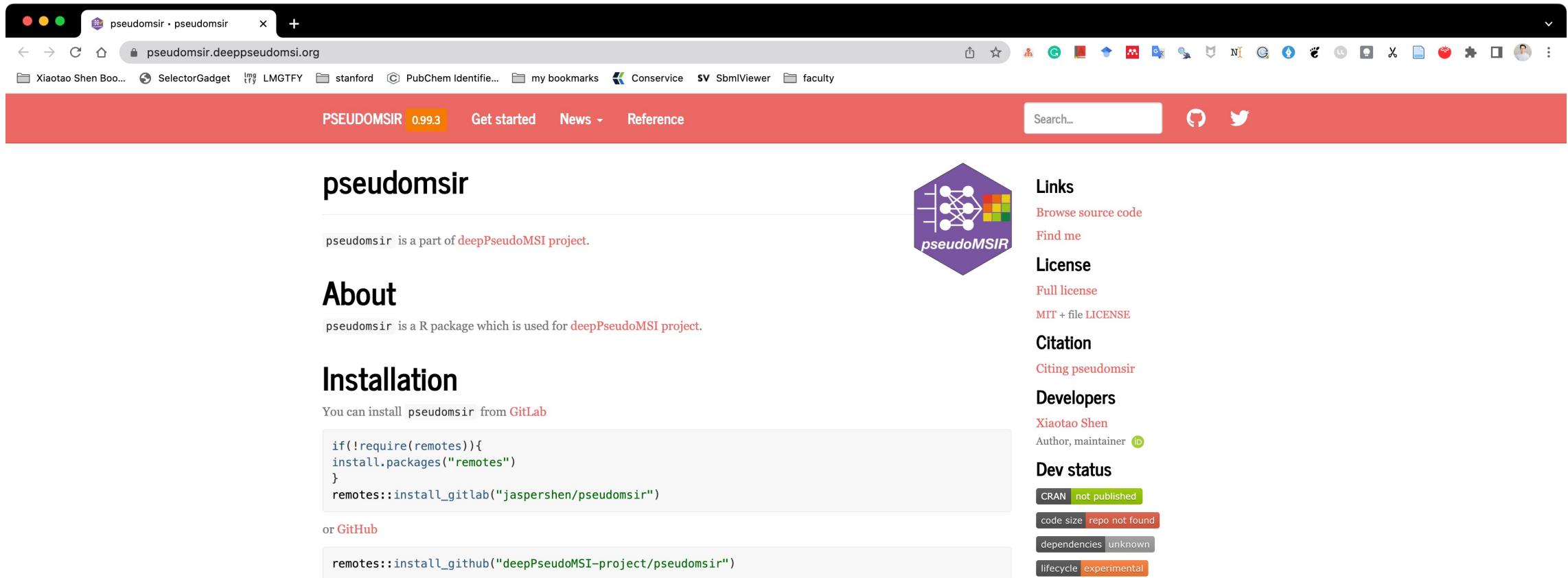
### Developers

Xiaotao Shen  
Author, maintainer [ID](#)

### Dev status

CRAN not published  
code size repo not found  
dependencies unknown  
lifecycle experimental

# ➤ Future works



The screenshot shows a Mac OS X browser window displaying the `pseudomsir` package page. The URL is `pseudomsir.deeppseudoMSI.org`. The page has a red header bar with the title "PSEUDOMSIR 0.99.3" and navigation links for "Get started", "News", and "Reference". A search bar and social media icons for GitHub and Twitter are also present. The main content area features a purple hexagonal logo with a neural network icon and the text "pseudoMSIR". Below the logo, there are sections for "Links", "License", "Citation", "Developers", and "Dev status". The "About" section includes a brief description of the package and its connection to the `deepPseudoMSI` project. The "Installation" section provides R code for installation from GitLab and GitHub. The "Usage" section directs users to the "Get started" page. The "Need help?" section provides contact information for the developer.

pseudomsir • pseudomsir

pseudomsir.deeppseudoMSI.org

Xiaotao Shen Boo... SelectorGadget LMGTFY stanford PubChem Identifier... my bookmarks Conservice SbmViewer faculty

PSEUDOMSIR 0.99.3 Get started News Reference Search... GitHub Twitter

# pseudomsir

pseudomsir is a part of [deepPseudoMSI](#) project.

## About

pseudomsir is a R package which is used for [deepPseudoMSI](#) project.

## Installation

You can install `pseudomsir` from [GitLab](#)

```
if(!require(remotes)){
  install.packages("remotes")
}
remotes::install_gitlab("jaspershen/pseudomsir")
```

or [GitHub](#)

```
remotes::install_github("deepPseudoMSI-project/pseudomsir")
```

## Usage

Please see the [Get started](#) page to get the instruction of `pseudomsir`.

## Need help?

If you have any questions about `pseudomsir`, please don't hesitate to email me ([shenxt@stanford.edu](mailto:shenxt@stanford.edu)).

➤ [shenzutao1990](#)

✉ [shenxt@stanford.edu](mailto:shenxt@stanford.edu)

🐦 [Twitter](#)

📍 [M339, Alway building, Cooper Lane, Palo Alto, CA 94304](#)

## Links

[Browse source code](#)

[Find me](#)

## License

[Full license](#)

[MIT + file LICENSE](#)

## Citation

[Citing pseudomsir](#)

## Developers

Xiaotao Shen

Author, maintainer [ID](#)

## Dev status

CRAN [not published](#)

code size [repo not found](#)

dependencies [unknown](#)

lifecycle [experimental](#)

## ► Future works

- ❖ How to combine datasets of different chromatography and Electrospray ionization (ESI) modes to increase the prediction accuracy.
- ❖ A large-scale dataset for cancer diagnosis.

## ➤ SUMMARY

- </> DeepPseudoMSI is the first systematic study that converts the LC–MS-based untargeted metabolomics data to pseudo-MS images and then takes advantage of the power of deep learning in image processing for precision medicine.
- </> We also demonstrate that the deepPseudoMSI can overcome the limitations of the traditional method for LC-MS data in precision medicine.

## ACKNOWLEDGEMENTS



**Prof. Wei Shao**  
(University of Florida)



**Dr. Chuchu Wang**  
(Stanford University)



**Prof. Mirabela Rus**  
(Stanford University)



**Prof. Michael Snyder**  
(Stanford University)

Stanford University  
Liang Liang PhD  
Songjie Chen PhD  
Sai Zhang PhD



# Thanks for your attention!

## Q&A

Xiaotao Shen PhD

Stanford University School of Medicine  
Department of Genetics



shenxt@stanford.edu



shenxt.info



@xiaotaoshen1990



github.com/jaspershen