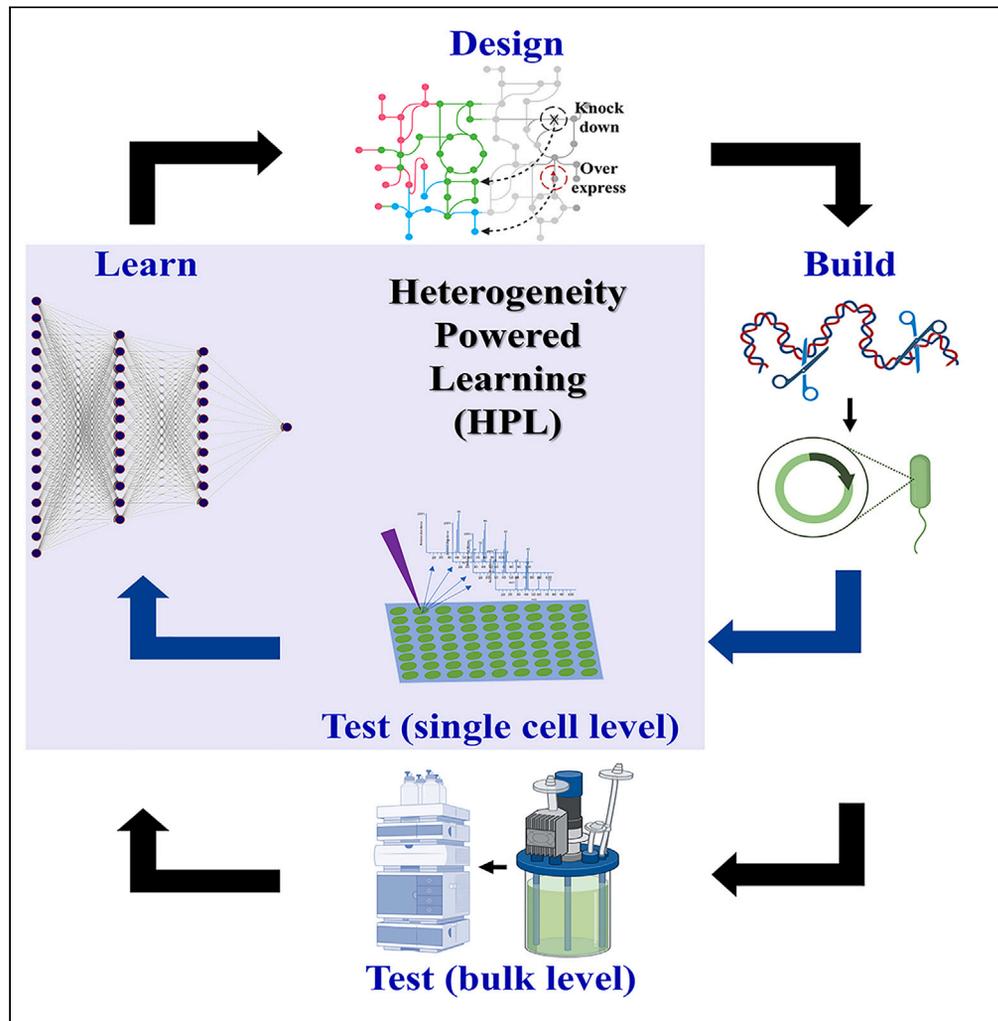**Article**

RespectM revealed metabolic heterogeneity powers deep learning for reshaping the DBTL cycle

Xuanlin Meng,
Ping Xu, Fei Tao

pingxu@sjtu.edu.cn (P.X.)
taofei@sjtu.edu.cn (F.T.)

Highlights

A strategy "RespectM" is proposed for microbial single-cell level metabolomics (MSCLM)

Over 4,321 MSCLM data are acquired by using the RespectM strategy

Heterogeneity-powered learning (HPL) is established with MSCLM data

The HPL-based model could reshape traditional DBTL cycle of synthetic biology

Article

RespectM revealed metabolic heterogeneity powers deep learning for reshaping the DBTL cycle

Xuanlin Meng,[1] Ping Xu,[1,*] and Fei Tao[1,2,*]

SUMMARY

Synthetic biology, relying on Design-Build-Test-Learn (DBTL) cycle, aims to solve medicine, manufacturing, and agriculture problems. However, the DBTL cycle's Learn (L) step lacks predictive power for the behavior of biological systems, resulting from the incompatibility between sparse testing data and chaotic metabolic networks. Herein, we develop a method, "RespectM," based on mass spectrometry imaging, which is able to detect metabolites at a rate of 500 cells per hour with high efficiency. In this study, 4,321 single cell level metabolomics data were acquired, representing metabolic heterogeneity. An optimizable deep neural network was applied to learn from metabolic heterogeneity and a "heterogeneity-powered learning (HPL)" based model was trained as well. By testing the HPL based model, we suggest minimal operations to achieve high triglyceride production for engineering. The HPL strategy could revolutionize rational design and reshape the DBTL cycle.

INTRODUCTION

Synthetic biology, relying on the Design-Build-Test-Learn (DBTL) cycle, provides engineering principles that allow the design and build of biological systems with new or enhanced functions.[1,2] Accelerating the DBTL cycle can achieve rapid and facile bioengineering of organisms for the production.[3] Therefore, researchers endeavor to accelerate DBTL through the integration of robotics and informatics.[4,5] The DBTL cycle's Learn (L) step leverages the data previously generated to inform the next design step and is fundamentally important for design section.[6] However, the L phase of the DBTL cycle has traditionally been the most weakly supported and developed, despite its critical importance to accelerate the full cycle.[7] Arguably, the main factors of the lack of emphasis on the L phase are: the lack of predictive power for biological system behavior,[8] resulted by the extreme asymmetry between the sparse learning data and the chaotic metabolic network.

Heterogeneity widely exists in natural cells.[9] This phenomenon can be found in multiple aspects of genomes,[10,11] transcriptomes,[12,13] proteomes,[14] and metabolomes.[15,16] In microbial studies, heterogeneity is generally considered to be a crucial factor in the evolution and stability of cell populations.[17,18] Recent studies have extended the appreciation of heterogeneity effects in several dimensions, beyond including bioprocess robustness,[19] and casted light on its roles in drug resistance,[20] quorum sensing,[21] symbiosis,[17] and adaptive evolution.[22] Besides, heterogeneity, specifying to the fluctuation of nucleic acids, proteins, and metabolites at single-cell levels,[23,24] can be used to generate big data at the status with or without external stimulation. Microbial heterogeneity encompasses massive metabolic associations among billions of cells, which implies that big data would be directly generated from heterogeneity exploring. These big data can be utilized to power the learning step of DBTL cycle and make deep learning algorithms applicable. Thus, there is an urgent demand for practicable single-cell tools to investigate cellular heterogeneity.

Single-cell analysis has emerged as a promising discovery direction and research hotspot in recent years. This cutting-edge technology enables the measurement of metabolic activity in individual cells, revealing heterogeneity and complexity of cellular metabolism.[14] With the ability to identify cellular subpopulations, track metabolic changes, and explore metabolic pathways, single-cell metabolism analysis holds great potential for advancing our understanding of cellular physiology and disease. To date, there are few single-cell microbial research methods unlike those for mammals, which restrains the intuitive access to the underlying mechanism of microbial heterogeneity. The development of single-cell methods for proteins

[1]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, P. R. China

[2]Lead contact

*Correspondence: pingxu@sjtu.edu.cn (P.X.), taofei@sjtu.edu.cn (F.T.)

https://doi.org/10.1016/j.isci. 2023.107069

and metabolites may enhance the understanding of direct protein-protein interaction (PPI) networks and metabolite reaction networks (MRN) to counteract the cognition of cellular heterogeneity. In the past decade, several studies have been conducted applying Matrix Associated Laser Desorption Ionization (MALDI) mass spectrometry to characterize metabolites at single-cell levels.[15,16,25–27] However, few studies have been performed to address the defects of poor reproducibility among chemical matrix backgrounds and inadequacy of analytical solution for microbial single-cell data analysis, which are the primary obstacles for establishing reliable microbial single-cell metabolomics methods.

In this study, we develop a new method, namely "RespectM," to perform single-cell metabolomics for microorganisms. In this method, the collection of single-cell data is precisely achieved by discontinuous mass spectrometry imaging (MSI). A sublimation method was applied for the MSI experiment.[28] Standard software SCiLS Lab (Bruker, USA) and open access R packages sclmpute, MetNormalizer, and Stream were integrated into the analytical system to form the pipeline.[29–31] These software packages aided in standardization of the method and readability for the data processing. To ensure the accuracy of data acquisition, we used microscopy-guided cell ablation coupled with *in situ* micron-level precise positioning.

Researchers revealed that cells can predictably respond to energy limitations in an artificially constructed microbial population, which likely contributes to the stability and robustness of microbial life.[32] Inspired by this work, we conducted a single-cell metabolomics analysis of a *Chlamydomonas reinhardtii* cell population to explore microbial metabolic heterogeneity. We established different *C. reinhardtii* cell groups in three aspects: temperature, cell wall-deficiency, and photosynthesis diversity within the total cell populations. RespectM provided a dataset of 4,321 cells with more than 600 metabolites. All metabolite identifiers conform to RespectM standards, which consist of lipids and metabolites with Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations. The identified series of features includes neutral glycerolipids (DG, TG), protective lipids (PE), signal transduction lipids (PIP), nucleotides (GMP), pigments (porphyrin), and metabolites belonging to the central metabolic pathway (such as oxaloacetate). Finally, 36 features (metabolites) were selected from dysregulated pathways for downstream single-cell analysis. The principle of choosing 36 features is described in the supporting information.

Our results show that RespectM can distinguish single microbial cells from the blank matrix with an accuracy between 95.3% and 98.3% using the Area Under the ROC (AUC) calculation, proving its capability to avoid matrix interference in the single-cell data acquisition process. We also applied the PLS-DA and support vector machine (SVM) method to classify two allelic *C. reinhardtii* mutants at a single-cell level.[30,33] In allelic strains, we achieved 91.9% classification accuracy. Then we applied a uniform manifold approximation and projection (UMAP) for dimension reduction chart to display the basal metabolic status of cells and introduced pseudo-time analysis to illustrate the cell trajectory.[34–36] Ultimately, we captured the cell trajectory and the dysregulated metabolite accumulation-induced cell grouping (DMACG) phenomena by the stream algorithm.

Based on RespectM, we obtained the information about accumulation of metabolites in total 4,321 cells. We learned single-cell metabolomics data, representing metabolic heterogeneity, by deep neural network (DNN). The heterogeneity-powered learning (HPL) based model was established by DNN with high accuracy (Training MSE: 0.0009546, Test MSE: 0.0009198). In the optimal model, it is discussed that synthesis-related genes of diglycerides (DG), phosphatidylserines (PS) and cobyrinate should be overexpressed first beside the variation of 17 metabolites to ensure the high triglycerides (TG) yield(s). High TG values corresponded pattern that can be achieved with minimal operation were suggested as the optimal pattern to realize high TG production.

RESULTS

The RespectM method

We established a microbial metabolomics method at the single-cell level, named "RespectM". Including sample preprocessing, data acquisition, and data processing (Figures 1A–1C and S1).

The main challenges of MALDI-based single-cell metabolomics include precise single-cell data acquisition, and enough peak generation for identification. We applied cell dilution strategy, combined with laser etching guided droplet microarray (LEM), matrix sublimation,[37] and sparse data matrix generation to meet these challenges.
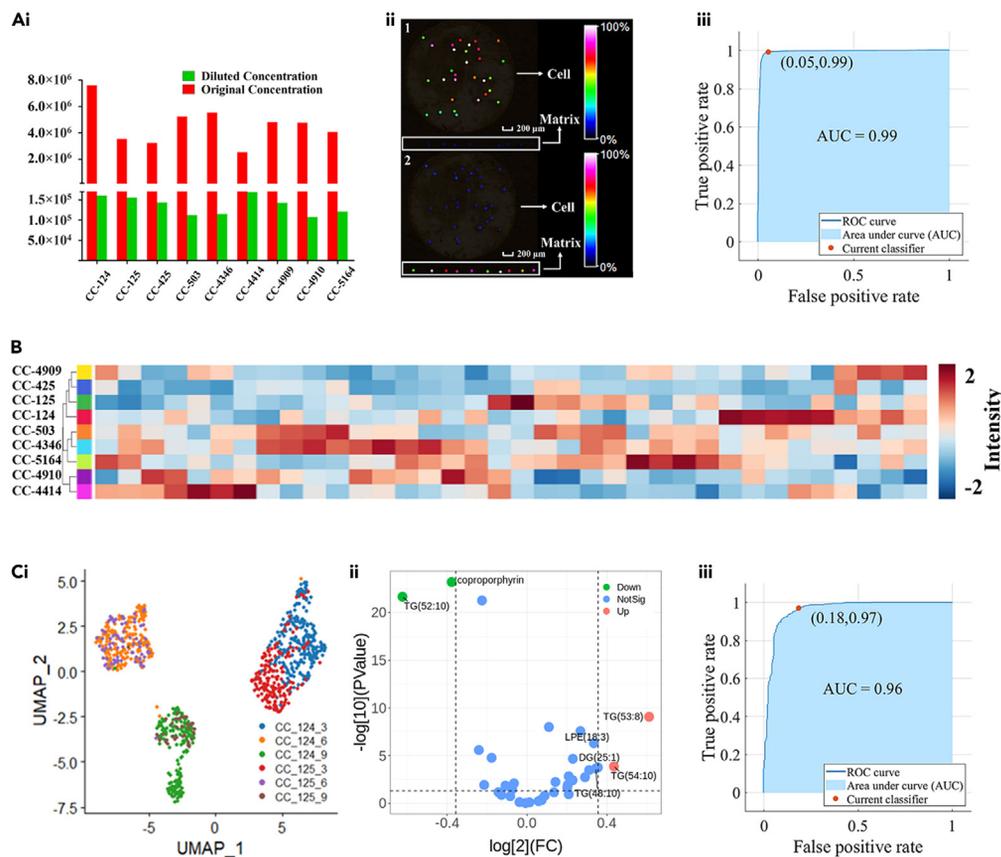
Figure 1. Workflow and validation of the RespectM method

(A) The workflow construction of RespectM. (i) Cell dilution from original density to 10^5 cells/mL. Corresponding pictures calculated by Cellometer Auto T4 can be found in Figure S3. (ii) Feature selection for distinguishing cell from the blank matrix. 1. Feature intensity in cells > Feature intensity in matrix and 2. Feature intensity in cells < Feature intensity in matrix. The latter is removed in feature selection. Scale bar: 200 μm. Bar color: From red to blue, intensity reduces. (iii) Matrix versus sample classification test based on Quadric SVM and reached 98.3% accuracy. A classifier with a larger AUC works better.

(B) Heatmap of dysregulated features based on LC-MS/MS identification, from left to right, we numbered 1–36.

(C) Validation of RespectM method among allelic strains. (i) UMAP visualization of strains CC-124 and CC-125 at three-time points. (ii) Dysregulate feature filtration among allelic strains by volcano plot. Fold change >1.2, p < 0.05. (iii) Allelic strain classification test based on Quadric SVM and reached 91.9% accuracy.

The most vital part of the experiment is to figure out an appropriate cell density to do single-cell level metabolomics data acquisition. To validate the cell size of nine algae, Cellometer Auto T4 (Nexcelom, USA) is applied to calculate the diameter of nine algae cell populations. The statistical analysis summarized that the majority scale of nine algae cells is around 10 μm as in the previous report (Figure S2). Before experiment, we diluted the original cell density of 10^6 to about 10^5 cells/mL level according to the Cellometer calculation (Figures 1A(i) and S3). Under the diluted cell density, after MALDI data acquisition, the spacing between cells is larger than MALDI laser ablation diameters (About 50 μm). Therefore, a single-cell level data acquisition could be achieved. The flow chart of MALDI mass spectrometry data acquisition is shown as Figure S4.

A low MALDI laser raster co-sampling rate is the prerequisite of single-cell level data acquisition. Generally, a higher cell density should improve the co-sampling rate. However, a lower cell density should reduce data acquisition efficiency. Under the diluted cell density (about 100 cells/μL), the co-sampling rate was manually calculated to 9.1% (Figure S5, 44 cells/0.5μL). It is unlikely that co-sampling would happen while the cell density is less than 90 cells/μL (Figure S5, 19 cells/0.5μL and 11 cells/0.5μL). To balance the efficiency and co-sampling rate, we chose around 100 cells/μL to perform 4,321 single cell level data acquisition.

To prove this method has single cell level sensitivity, we further tested the mass spectrum intensity among one, two, and three cells. It was revealed that the mass spectrum signal intensity was increased from one to three cells. The result shows that our method has enough sensitivity to detect metabolites from a single cell (Figures S6 and S7).

Through laser etching on the ITO slide, we printed multiple cross patterns arranged on the surface of the ITO glass slides. Therefore, to ensure that different droplets would not contaminate each other. We named it laser etching guided droplet microarray (LEM). After the cells were dropped at the designated location, a fine needle was used to draw three cross marks around the droplet for instrument positioning (Figures S4 and S8A). Finally, ten blank matrix points were collected as quality control (QC) samples at the end of each dripped spot sequence to reduce the batch effect (Figure 1A(ii)). This attempt is the first time that a QC-based correction strategy was integrated with MALDI single-cell level metabolomics has been proposed. We applied the Tissue Profiling model (flexImaging, Bruker) to perform data acquisition, a kind of discontinuous mass spectrometry imaging model that can be easily achieved by traditional three-point positioning (Figures S4 and S8B).[38] Based on the Tissue Profiling model, it is easy and free to choose the cell spot on the microscope picture (Figure S4). Finally, RespectM could provide a single-cell level data matrix corresponding to the *in situ* MSI acquisition sequence through SciLS Lab software (Bruker, USA). The reads included the data acquisition sequence, the cells *in situ* position in the corresponding series, and the peak annotation results (Figure S6, Table S1).

RespectM distinguishes allelic strains in single cell levels

In MSI studies, it is necessary for chemical matrix to associate the desorption of biological samples, but it is also complicated because of the background in the MSI data. Many studies have neglected the signal interference of the chemical matrix on biological samples. In this study, we propose a new solution to retain the feature (*m/z*) intensity in cells > feature (*m/z*) intensity in matrix. The feature signal response in the sample is higher than the feature in the blank chemical matrix (Figure 1A(ii)).

The sample data need to be distinguished from the blank matrix to ensure the accuracy of downstream single-cell analysis (Figure 1A(iii)). Dysregulated features were chosen to correspond to LC-MS/MS metabolomics and lipidomics for classification (Figures 1B, S8C and S9). On this basis, 29 classification strategies were applied to distinguish samples from the blank matrix. The results showed that these methods could achieve as high as 98.3% classification accuracy in the receiver operating characteristic curve (ROC curve) (Figures 1A(iii) and S8C).

Furthermore, we conducted Seurat (R package, open access) to perform analysis on strains CC-124 and CC-125 (Figure 1C(i)). We observed some different between CC-124, CC-125 (Figure 1C(ii)) and tested the ability of our method to distinguish allelic strains CC-124 from CC-125. Through the support vector machine (SVM) strategy, 91.9% classification accuracy of the two strains was achieved (Figure 1C(iii)). Because the reproduction time of *C. reinhardtii* is 24 h, the mixed cell state at the three timepoints needs to undergo significant changes. According to the instruction of R package, Seurat, the principal component analysis (PCA) reduction should be advanced before uniform manifold approximation and projection for dimension reduction (UMAP).[39] The UMAP chart showed that the distance of single-cell data between three timepoints was significant. We visualized the marker features used for automatic grouping and applied the Wilcoxson algorithm to calculate all 12 markers in auto-clusters (Figure S10). Results showed that the abundance of these markers were dysregulated in the auto-clusters. We found that the relative intensity of marker features was consistent under the premise of wild-type *C. reinhardtii* without external stimulation. In parallel, some parts of cells clustered at each timepoint, indicating the variability of single-cell metabolic adaptation.

RespectM is applied to visualize the heterogeneity in cell populations

The metabolic heterogeneity of microorganisms lacks an effective visualization method. This situation also limits the characterization of the metabolic status of different microbial populations at single-cell levels. For example, the metabolism of *C. reinhardtii* cells within the population is highly variable, which complicates the single-cell metabolic status collection. In addition, the cells sampled at different time points are in mixed states. However, traditional analysis methods are limited to find more information among single-cell level. Therefore, we conducted multiple strategies including UMAP and pseudo-time analysis to reveal the complex cell metabolism (Figures S10–S29). In addition, to analysis the single-cell dataset, we visualized total lipid accumulation among three time-points (Figure S29A). We found that lipids were dysregulated at different time-points in PCA analysis (Figure S29C).

First, we visualize the cell population based on UMAP to cope with this obstacle (Figure 2A). We visualized the data among time points (Figures S11–S13). Our results showed that there was no apparent clustering of cells observed at Days 3 and 9 (Figures 2B, S11, and S13). However, two obvious clusters occurred in the cell population at Day 6 (Figures 2B and S12). Combined with 3 group analysis (Figure 2C), it was found that the three different groups of *C. reinhardtii* had significant metabolic changes at Day 6, which is correspond to the algae morphology (Figure 2B).[40] In the temperature group (strains CC-124, CC-4414, and CC-5164), there was also no apparent clustering of cells at Days 3 and 9. Two clusters occurred in Day 6 (Figure S26). Here we focused on the joint clusters of strains CC-4414 and CC-5164, where the accumulation of metabolites helped explain the reasons for clustering. Therefore, we visualized the markers used to distinguish different clusters of temperature cell groups in the entire cell population (Figure S14).

By visualizing the highly variable features, the Seurat package was conducted to do marker feature selection (Figures 2D(i–iii) and 2E(i–iii)). In the temperature diversity group (CC-124, CC-4414, CC-5164), we found two lipids TG (53:8), DGDG (34:7) gradually increased with time (Figures S15 and S26), whereas MGDG (36:7) decreased. In the photosynthetic group (strains CC-124, CC-4909, CC-4910, and CC-4346), we found 12 marker features with high variables for analysis and visualization (Figures 2E(i) and S16). We focused on the two cell groups generated at Day 6. One group consisted of strains CC-4909 and CC-124, and the other group had strains CC-4910 and CC-4346. The two groups at Day 6 were independent (Figures S16 and S26). With respect to the distribution of metabolite abundance at Day 6, we found that TG (51:7) had high abundance in strains CC-4346 and CC-4910 of the cell groups (Figures 2C, S16 and S26). Over time, the abundance of PS (35:3), MGDG (36:7) and Cobyrinate decreased. In contrast, the abundance of TG (53:8) increased (Figures S17 and S26). In the final cell wall defect population, the 6-day cell population did not generate an obvious cluster (Figures S18, S19, and S26). Nevertheless, in both of temperature and photosynthetic groups, the abundance of MGDG (36:7) accumulated with time is contrary to that of TG (53:8) (Figures S15 and S17). In this section, we visualized the metabolic heterogeneity of *C. reinhardtii* at single-cell levels. However, UMAP analysis could not reveal the detailed microbial cell trajectory.

Identification of *C. reinhardtii* DMACG by RespectM

RespectM may provide a more intuitive representation of the heterogeneity of single-cell metabolism. The above results mainly showed the heterogeneity of *C. reinhardtii* in the two dimensions of timepoint and category. However, because of the complex cell mixing state at each timepoint, more heterogeneity information was hidden in the cell population (Figure 3A). Owing to the rapid reproduction of microorganisms, the metabolic state is unstable compared to higher organisms. This phenomenon complicated the capture of the metabolic state. Therefore, we applied a matured pseudo-time method Stream (http://stream.pinellolab.org/) to explore further microbial heterogeneity in the cell trajectory (Figures 3B–3D).[41]

Strains CC-124, CC-4414, and CC-5164 at the S3 node have different dysregulated metabolite accumulation-induced cell grouping (DMACG) phenomena in the temperature group (Figures 3B and S20). Strain CC-124 independently generated an S3-S5 branch, whereas strains CC-4414 and CC-5164 formed an S3-S4 branch (Figure S20). The results show that the intensity of MGTS, PG, and DGTS and PIP in the S3-S4 branch was higher than that of S3-S5, and the abundance of DG and DGDG in the S3-S5 branch was higher than that of S3-S4 in parallel (Figure S21). This phenomenon shows that strains CC-4414 and CC-5164 produce more PE under normal temperature conditions to adapt to the environment (Figure S21). Meanwhile, the high abundance of phytic acid and PIP lipids confirmed a difference in the phosphatidylinositol metabolism pathway in strains CC-4414, and CC-5164 at single-cell levels (Figure S21).

In the photosynthetic group, the inheritance of strains CC-124, CC-4909, and CC-4910 was continuous. This relationship was reflected in the S0-S1 and S0-S2 branches (Figures 3C and S22). Strains CC-4909 and CC-124 share the same branch, which is consistent with CC-124 being the background strain of CC-4909. This figure was also in line with strain CC-4910 and CC-4346 having a dysregulated metabolic capacity in lipid synthesis (Figure S22 and S23). In parallel, strain CC-4346 shares the same branch with the other three *Chlamydomonas* at S3-S4 (Figure S22). However, S3-S4 shows a higher lipid abundance than S0-S2 branches (Figure S23). This phenomenon implies that the CAO deficiency demands CC-4346 (an insertional mutation in the CAO gene encoding a chlorophyll(ide) oxygenase) to enhance metabolism to compensate for its photosynthetic defects.[42]
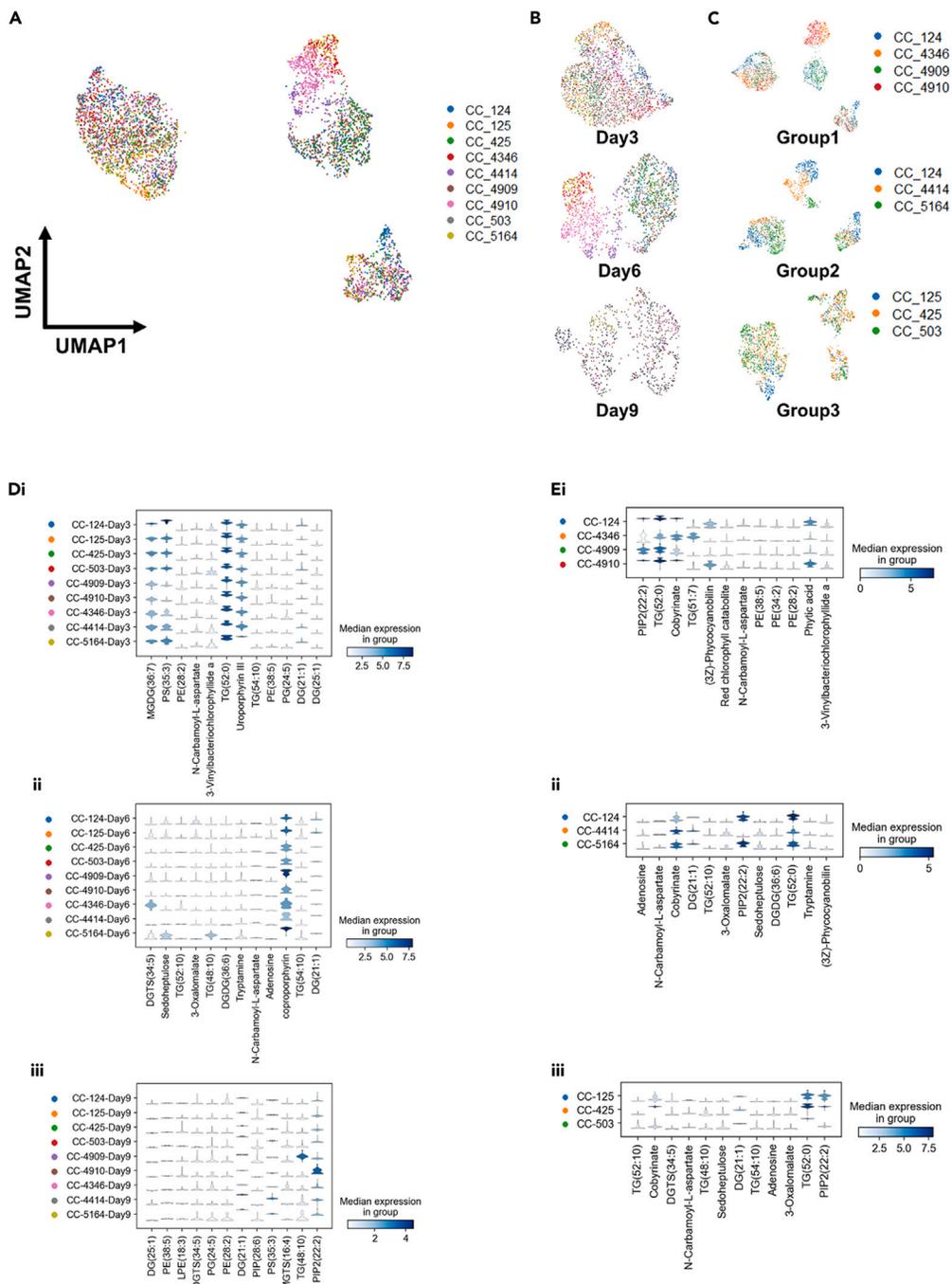
Figure 2. Visualization of microbial single-cell metabolomics data

(A) UMAP visualization of single-cell metabolic profiles of 4,321 singe cells.

(B) UMAP visualization of cells from Days 3, 6, 9.

(C) UMAP visualization of cells from groups 1, 2, 3 (Group 1: four *Chlamydomonas reinhardtii* strains with differences in photosynthetic capacity: strains CC-124, CC-4909, CC-4910, and CC-4346; Group 2: three *Chlamydomonas reinhardtii* strains with differences in temperature adaptability: strains CC-124, CC-4414, and CC-5164; Group 3: three *Chlamydomonas reinhardtii* strains with differences in cell structure: strains CC-125, CC-425, and CC-503).

(D) Violin Plot is used to display the distribution state and probability density of multiple time-point datasets. (i) Abundance of Day 3 marker features. (ii) Abundance of Day 6 marker features. (ii) Abundance of Day 9 marker features.

(E) Violin Plot displays multiple cell group datasets' distribution states and probability density. (i) Abundance of group1 marker features. (ii) Abundance of group 2 marker features. (ii) Abundance of group 3 marker features.
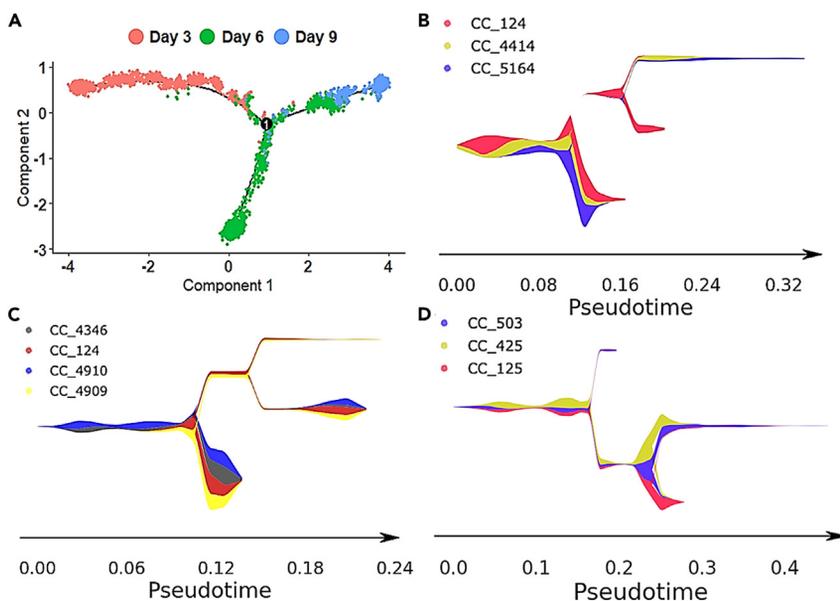
Figure 3. Pseudo-time analysis of microbial single-cell metabolomics data

(A) The Pseudotime trajectory among 4,321 cells of *Chlamydomonas reinhardtii* based on dysregulated metabolites.

(B) Three *Chlamydomonas reinhardtii* strains with differences in temperature adaptability: strains CC-124, CC-4414, and CC-5164. Stream plot shows a Pseudotime trajectory from homeostatic, intermediate, and metabolic variation among strains CC-124, CC-4414, and CC-5164.

(C) Four *Chlamydomonas reinhardtii* strains with differences in photosynthetic capacity: strains CC-124, CC-4909, CC-4910, and CC-4346. Pseudo-time analysis reveals the trajectory of three lines that have a coherent genetic relationship.

(D) Three *Chlamydomonas reinhardtii* strains with differences in cell structure: strains CC-125, CC-425, and CC-503. Stream plot shows a Pseudotime trajectory among wild type CC-125, and cell wall-deficiency strains CC-425 and CC-503.

Strains CC-425 and CC-503 are wall-deficient types belonging to the S0-S1 branch (Figures 3D and S24). The cells of the wild-type of strain CC-124 (with cell walls) were mainly distributed in the S0-S2 branch (Figure S24). The abundance of *N*-carbamoyl-L-aspartate in the wall-deficient type of strains CC-425 and CC-503 was higher than that of the wild type of strain CC-124, which implies that the wall-deficient type needs stronger primary metabolism to resist the defect of the wall. In the S0-S1 branch, MGDG, DGDG,[43] LPE, and PS abundances were higher than S0-S2 (Figure S25). It is also worth noting that when the nine strains of *C. reinhardtii* were analyzed in a pseudo-time series without comparison, the basal status of *C. reinhardtii* still produced the DMACG phenomenon in the cell trajectory. This result suggests that the RespectM method could be used to reveal basal metabolic heterogeneity by reordering cells (Figure S27).

Deep learning of big data generated by heterogeneity resolving

The information of metabolic heterogeneity could be revealed by analyzing *C. reinhardtii* single-cell metabolomics data. Generally, heterogeneity is mainly specified to the fluctuation of nucleic acids, proteins, and metabolites at single-cell levels,[23,24] which could generate massive data at the status with or without external stimulation. To learn from the information provided by metabolic heterogeneity, we established a metabolic model of triglycerides (TG) target metabolites by deep learning (Figures 4A–4F).

The data used for training are experimental measured single-cell level metabolomics data. After normalizing the original data to the 0–1 interval, we proceeded to downstream large-scale data analysis. The output TG data used for training are continuous actual measurement values rather than qualitative labels. An optimizable neural network (MATLAB, 2021b, Regression Learner) was applied to establish the relations between input and output nodes. All lipid data were processed by summation. Taking TG as an example, the TG related fatty acids are summing to form the total TG. After the summing process, the 36 features were reformed to 23 features. A correlation test was conducted to test the correlation among 23 features. We found that all features were correlated with TG (Figure 4F). Therefore, 22 features were chosen as predictors, and TG was selected to be the response. Before training
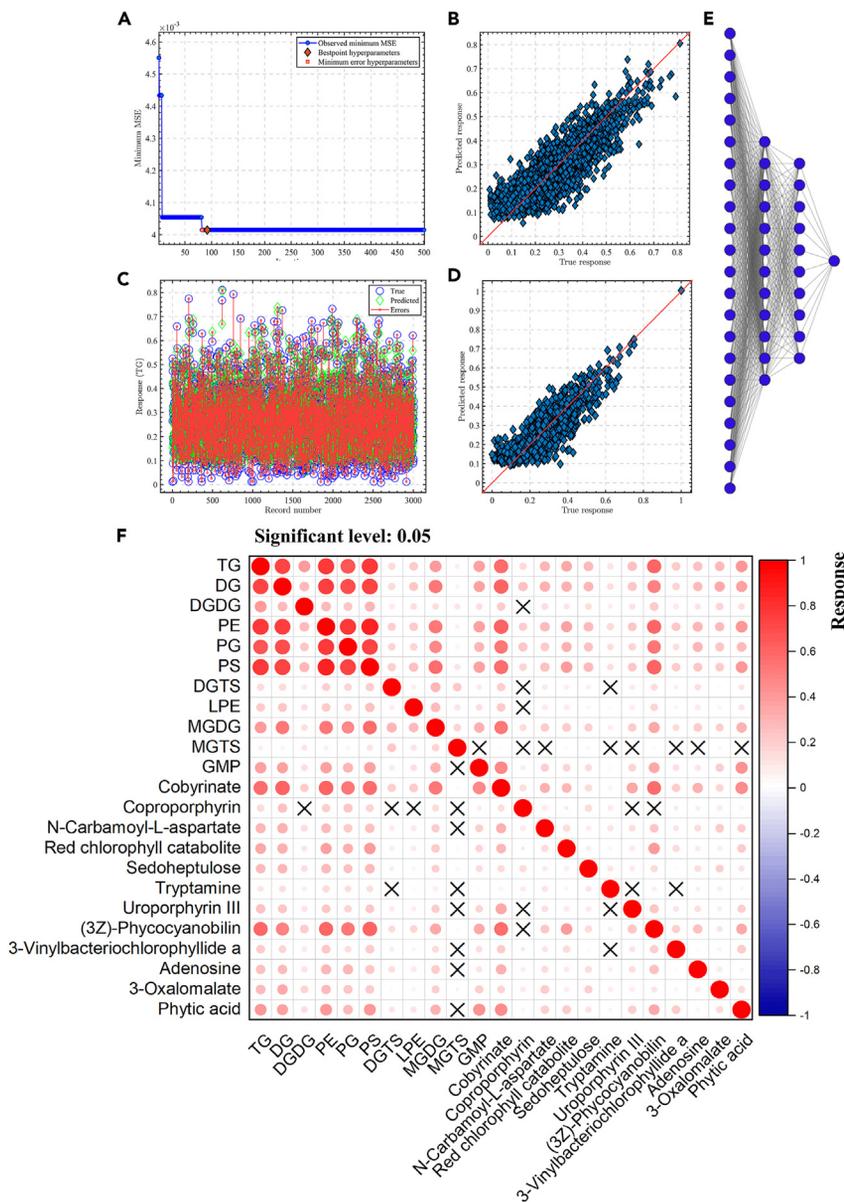
Figure 4. The construction of metabolic model at single-cell level

(A) Hyperparameter selection of optimizable neural network.

(B) The scatterplot of the predicted versus the validation data.

(C) Error between actual data and predicted data.

(D) The scatterplot of the predicted versus the test data.

(E) The structure of the optimized neural network. Includes 22 input nodes, 1 output node, and 2 hidden layers: 12 nodes in first layer, 10 nodes in second layer.

(F) Correlation result among metabolites at the single-cell levels. The false symbol represents insignificant. Red represents positive correlation; Blue represents negative correlation. The shape of the circle represents the correlation level. The larger circle represents a stronger correlation.

operation, five folder cross-validation was conducted to protect against overfitting by partitioning the dataset into folds and estimating accuracy on each fold. We used random search method to find the best parameters. Before dividing the two dataset of training and testing, we performed the shuffle operation on 4,321 cells. An optimized deep neural network (DNN) was established by training single-cell metabolomics data (The nodes of first hidden layer: 12; The nodes of the second layer: 10; R^2: 0.8312, Training MSE: 0.0009267; Test MSE: 0.0009198) (Figures 4A–4E).

Furthermore, to verify the model's generalization ability, 200 maximum and 200 minimum TG responses were extracted to be testing data. It was found that the model was well adapted to the fitting of the generalized data (R = 0.96575) (Figure S28). In summary, we established a TG metabolic model based on microbial heterogeneity. Because the model was learned from the 4,321 single-cell metabolomics data, representing the metabolic heterogeneity, we named it as "Heterogeneity-powered learning (HPL)" model. The training result and preliminary testing of the HPL model demonstrate that single-cell metabolic data could be well learned.

Rational design based on the HPL metabolic model

Triglycerides (TG) occur widely in animals, plants, and microbial organisms and have been used as raw materials for biofuel production.[44] To achieve the high yield(s) of TG in *C. reinhardtii*, we intend to find the high TG values corresponding metabolic pattern by designing and verifying a model based on HPL. In the macroscopic dimension, metabolic patterns are generally presented by the average of all cells. Therefore, we set the original state of the 22 features as their average values. To test the model, first, the value of each feature from 0 to twice of its own maximum value (2 max) is increased, respectively, whereas maintaining the average values of the rest features. We hope to find the feature whose value increase has the most significant effect on TG. When the values of nine lipids (diglycerides (DG), digalactosyl diglyceride (DGDG), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylserine (PS), and lyso-phosphatidylethanolamine (LPE)) are increased, the value of TG increases as well (Figure 5A). In parallel, when the values of eight metabolites (GMP, cobyrinate, *N*-carbamoyl-L-aspartate, Red chlorophyll catabolite, phytic acid, sedoheptulose, uroporphyrin III, 3-oxalamate, 3-vinybactcrio-chlorophylide a) are increased, TG's values increase as well (Figure 5B). On the contrary, when the values of three lipids diglycero-3-phospho-*O*-serine (DGTS), monoglycero-3-phospho-*O*-serine (MGTS), Monogalactosyldiacylglycerol (MGDG) are increased, the value of TG decreases (Figure 5C). Also, when the values of five metabolites (coproporphyrin, tryptamine, (3Z)-phycocyanobilin, adenosine) are increased, the value of TG decreases (Figure 5D).

Second, 100,000 patterns of 22 features (Figure 5E) were randomly generated. The principles of data generation are (1) each feature independently generates data in the range of 0–2 max; (2) the randomly generated data are in average distribution rather than the normal distribution; (3) total data randomly consist of 23 features generated data. Then all generated data were predicted by the trained HPL based model. To display the high TG value corresponding metabolic pattern, a polar chart was conducted to visualize the pattern of the highest 1,000 TG values. We found that the glycerophospholipid and glycerolipid metabolism ensembled most of the highest TG value patterns (Figure 5E). Furthermore, when the values of DG, PS and cobyrinate are increased, the TG value increases the most. This result indicates that DG, PS, and cobyrinate could significantly affect TG accumulation. To validate the result, we conducted genetic algorithm and simulate annealing algorithm to approximate the global optimum (Figure S29D).[45,46] By SHAP explainer analysis,[47] we found that DG, PS are the most important lipids contribute to the TG output (Figure S29D); also, the metabolite is the most important metabolite contribute to the TG output. To validate our biological finding that DG, PS and cobyrinate increase should promote the TG, we also predicted our generated data using genetic algorithm and simulate annealing algorithm. Result shows that DG, PS and cobyrinate increase still improved TG. We also have plans to find the global optimum in theory by optimize the current model. We can use brute-force algorithms to exhaustively enumerate all combinations of inputs and outputs so that each value can be trained, and the global optimal solution of low-dimensional data can be approximated.[48] Also, we can use convex optimization methods to accurately find the global optimum of two-layer ReLU neural networks according to the previous reports and quantitative the output.[49,50] Plans can be found in Github repository (dawnmengsjtu/RespectM).

To find the closest path corresponding to the minimal operations to achieve a high TG yield(s) in engineering (Figure 6A), we first define a "high TG value" as follows: the TG value predicted after 22 random feature variation/average TG value in the 4,321 cells >1.6 (Figure 6B). Therefore, the highest 100 TG values corresponding patterns were chosen to be tested by three distance calculation algorithms (Euclidean, Manhattan, and Harmonic means). We found that pattern 94 is the closest pattern to the original 22 feature pattern (Figure 6C). More precisely, pattern 94 is closest to basal status among the three algorithms and is the most uncomplicated engineering strategy to achieve a high TG value. In this pattern, 3 metabolites should be knocked down, 17 metabolites should be overexpressed (Figure 6D).
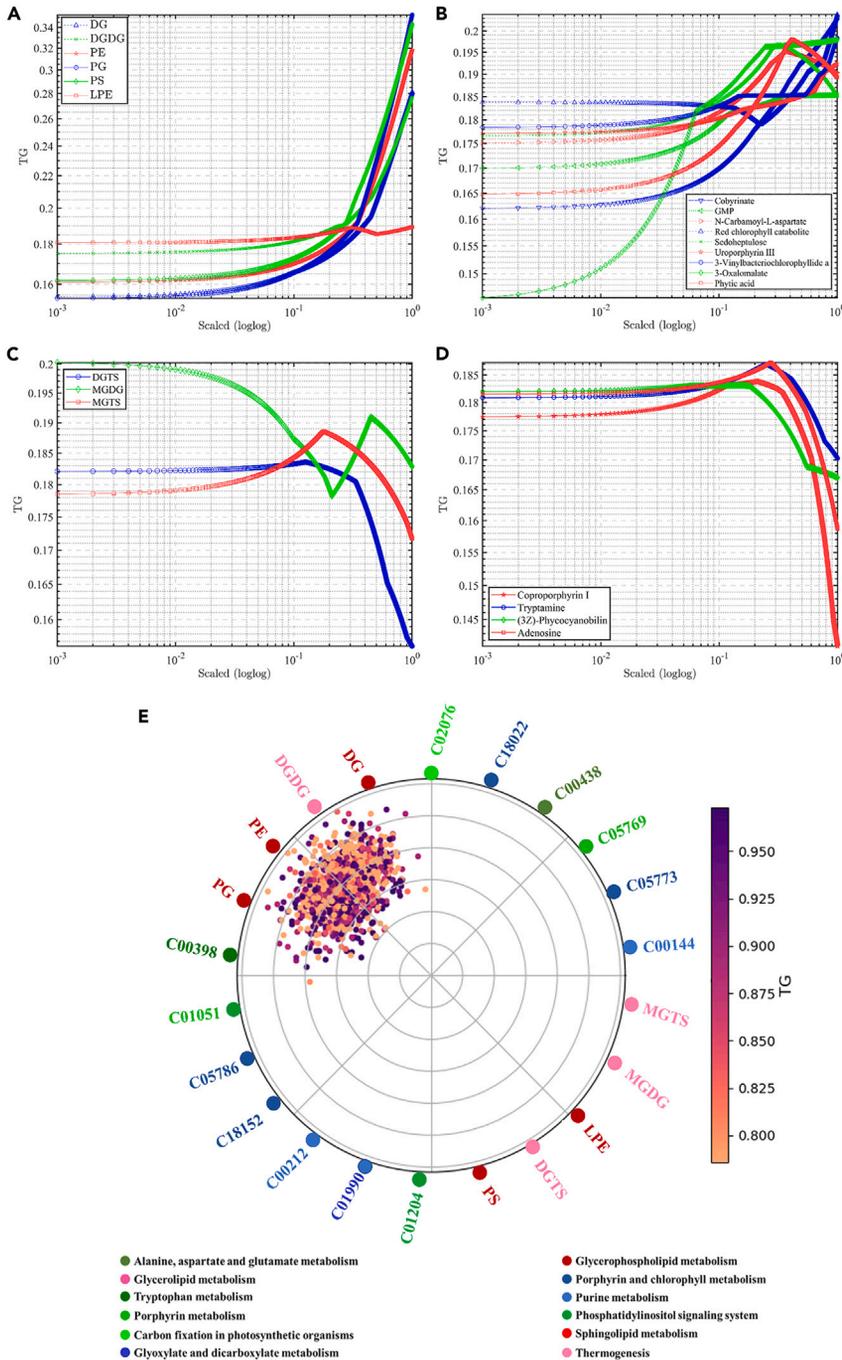
Figure 5. Visualization of metabolic results by using model prediction

The original state of the 22 features is set as their average value. To test the model, first, the value of each feature from 0 to twice of its own maximum value (2 max) is increased, respectively, while maintaining the average value of the rest features. (A–B) represents the increase of TG values during the increase of lipids and metabolites values, respectively; (C–D) represents the decrease of TG values during the increase of lipids and metabolites values, respectively.

(E) 100,000 patterns of 22 features are randomly generated, each feature independently generates data in the range of 0–2 max 100,000 data were generated randomly. The highest 1,000 TG intensity data distributes on the polar chart. Color bar: Dark colors represent the high TG intensities; light colors represent the low TG intensities.
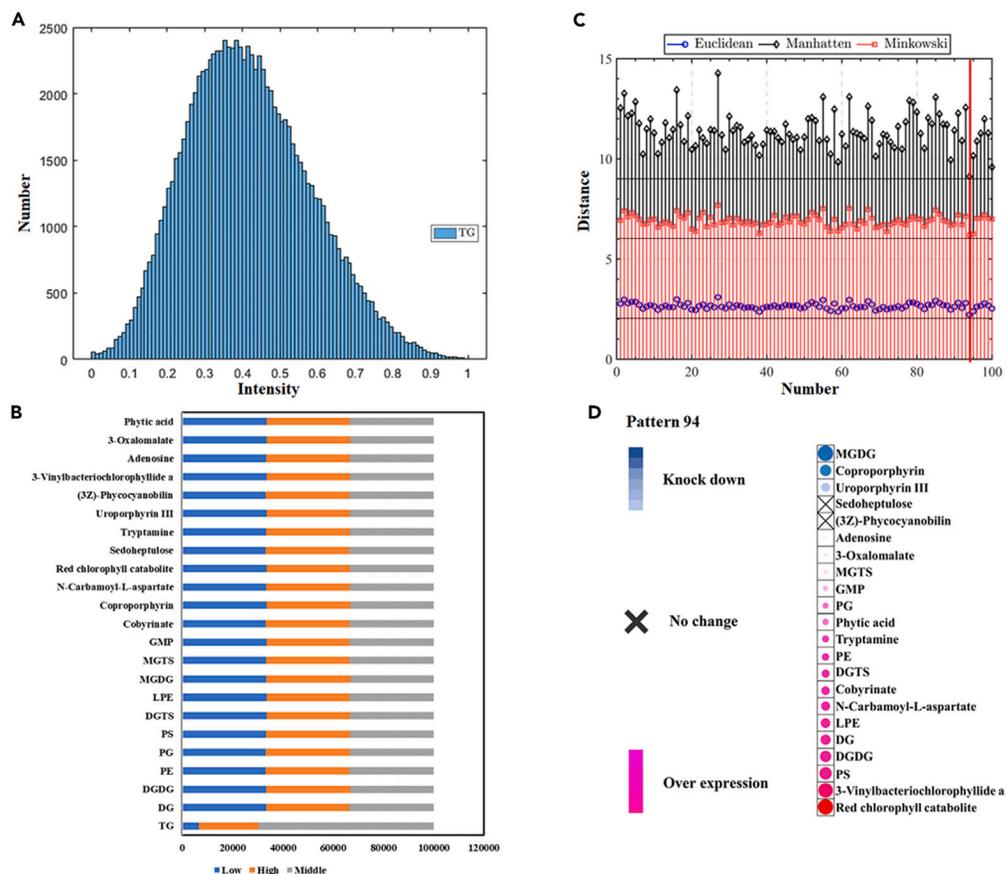
Figure 6. Optimal model selection corresponds to high TG values

(A) The value distribution of 100,000 random generated data prediction results.

(B) Data distribution of all features includes TG. Low: 0–0.333; Medium: 0.333–0.666; High: 0.666–1. All values were normalized in columns.

(C) Visualization of the distance between highest 100 TG values belonging model and original 4,321 cells trained model. Pattern 94 is the closest pattern to the original in 3 algorithm calculations.

(D) The visualization of the optimal model–pattern 94. Blue bar represents that these metabolites need to be knocked down; Red bar represents that these metabolites need to be overexpressed; X symbol represents that these metabolites are no need to be changed.

DISCUSSION

Single-cell metabolic analysis is at the forefront of current technological advancements and holds significant importance for microbial research.[15,16] With the ability to analyze metabolic activity at the individual cell level, researchers can uncover metabolic heterogeneity and identify novel metabolic pathways in microbial populations. Single-cell metabolic analysis also enables researchers to investigate the impact of environmental factors on microbial metabolism, providing insights into microbial behavior and their ecological roles.[15,16] This technology has the potential to advance our understanding of microbial physiology and inform the development of new treatments for microbial diseases.

Because previous DBTL testing data offer to learn is few compared to the complex metabolic network, the development of rational design is limited. Especially the yields for biological chemical production in microorganisms approach their traditional theoretical maximum; therefore, metabolic engineering and synthetic biology require new tools and approaches for improvements beyond what traditional strategies could achieve.[51] Herein, we propose the "heterogeneity-powered learning (HPL)" strategy to promote the development process of rational design. To explore the heterogeneity of cells, researchers have collected many data in mammals.[52–54] However, fewer microbial single-cell metabolic fingerprints constrain the visualization of microbial metabolic heterogeneity. Our study aims to improve the research on microbial single-cell

metabolomics by the developed RespectM method. The advantages of RespectM include several dimensions. First, the LEM cell chip strategy simplified the preprocessing steps before MSI data collection and optimized the single cell *in situ* positioning accuracy during MSI data collection. Second, a confidence grading system including the MS and MS/MS identification strategies was developed for single-cell metabolic identification. Third, the QC-based MSI calibration method was developed in MALDI-based single-cell metabolomics (Figure S1). Altogether, these factors underscore the unique advantages of RespectM in convenience, accuracy, and throughput. By establishing the RespectM methodology, we detect more than 600 metabolites at a rate of 500 cells per hour.

In single-cell MSI data acquisition, the cleanliness of the matrix background and accurate cell positioning is the basis. Currently, spray coating and sublimation are the standard matrix application methods.[37,55] On the one hand, the optimized matrix auto-sprayer could achieve a high ion intensity reproducibility across samples.[37] On the other hand, it is reported that the matrix application of the sublimation strategy could retain the *in situ* metabolic information with a high spatial resolution and the lowest de-localization effect.[28] Sublimation also has good reproducibility and eliminates the potential for spreading analytes from solvent deposition during matrix application.[56] Furthermore, the DHB sublimation performs well on MSI while using Solarix FT-ICR instrument.[57] Therefore, we chose the matrix sublimation method for RespectM.

Several factors currently restrict the application of MSI in microbial single-cell studies. For instance, the size of a single microbial cell is only 1–10 µm, whereas the raster ablation size is usually around 50 µm. Therefore, optical figure splicing operation is inappropriate for microbial single-cell data acquisition. Taking ultrafleXtreme (Bruker, USA) as an example, the physical movement accuracy of the mechanical stage is 5 µm; optical image stitching will also generate micron-level errors. This phenomenon indicates that positioning errors also increase when MSI *in situ* range is expanded. Rappez et al. proposed a set of machine learning-based localization strategies and achieved sub-micron accuracy. Our strategy is still based on the traditional microscopic guided three-point localization method.[38] However, the convenience of this method is still remarkable in our experiments.

To ascertain the cleanliness of single-cell data, we first compared RespectM with methods requiring cell staining and fixation.[55] First, RespectM has a shorter preprocessing time, and the cell sample is obtained through regular cultivations. Both cell-bulk metabolomics and single-cell metabolomics are directly sampled from culturing cells simultaneously. Second, we applied the discontinuous MSI method (tissue profiling) in RespectM to prevent cross-contamination between adjacent data points in data collection. The accuracy distinguishing cells from blank matrix reached 98.3%, which yields strong confidence in microbial single-cell acquisition. Thirdly, there is no method for batch correction between MSI internal data points. Therefore, MALDI-based single-cell data collection needs to cope with a new batch correction strategy. We introduced matrix background points as QC to correct batch effects between multiple data acquisitions. This strategy differs from the batch correction method of single-cell transcription, and mature metabolomics algorithms MetNormalizer aids in more confidence to RespectM.[30]

In exploring microbial cell heterogeneity, only a few studies on single-cell transcriptomics were attempted.[58,59] In addition to the omics level, fluorescence-based flow cytometry and mass spectrometry flow cytometry can detect a few features at single-cell levels for heterogeneity analysis.[60,61] However, heterogeneity analysis has not been established before. In this study, RespectM can complete an omics analysis of 4,000+ cells within a few hours and reach the number of metabolic identifiers to the omics level. Furthermore, we analyzed microbial single-cell metabolomics data by the UMAP algorithm. Based on the metabolic heterogeneity visualization through UMAP, in the timepoint aspect, a clear grouping was revealed at Day 6 (Figure 2B); in the species aspect, more cell groups were generated (Figure 2C). The cognition of heterogeneity will likely change depending on the observation angle.

All cells of *C. reinhardtii* were reordered through pseudo-time analysis. The microbial metabolic pseudo-time analysis is more diverse than the mammalian metabolic pseudo-time analysis. Our results show a re-ordered dysregulated metabolite accumulation-induced cell grouping (DMACG) phenomenon consistent with the biochemical and genetic relationships among the three cell populations. Without external stimulation, the metabolic heterogeneity of *C. reinhardtii* produced DMACG in different dimensions. The dysregulated-accumulated metabolites in various cells allow us to analyze the joint clustering of cells from different species, combined with the "cell age" from the pseudo-time analysis.

After obtaining the 4,321 single-cell metabolomics data, we conducted a deep learning network (DNN) to learn the single-cell metabolomics data and then screened the metabolic model corresponding to a high TG value. Generally, the traditional metabolic model also contains the relation among metabolites; however, it cannot predict a pattern with a high yield of the target product. Through the HPL strategy, we built an HPL-based model and predicted a high TG yield(s) metabolic pattern with minimal engineering operations, resulting in the deep learning of 4,321 single microbial metabolic statuses. Combined with the results concluded by the loglog chart and polar chart above (Figure 5), the accumulation DG, PS, and cobyrinate should be necessary to achieve high TG values. Phospholipid diacylglycerol acyltransferases (PDATs) synthesize TG by transferring a phospholipid donor's fatty acid to DG.[62] As a result, the increase of DG value significantly improves TG.

Meanwhile, the phosphorylation of lipid metabolic enzymes by protein kinase C requires the coordination of phosphatidylserines and diglycerides, implying that the rationale of increasing PS value improves TG as well.[63] Surprisingly, cobyrinate could significantly affect the TG accumulation, belonging to porphyrin metabolism. Porphyrin, a macrocyclic tetrapyrrole, and metal-containing derivatives play essential roles in many biological systems, especially chlorophyll (Mg, Zn), for photosynthesis in microbes.[64] Therefore, improving porphyrin metabolism should promote the photosynthesis ability of *C. reinhardtii* and realize a high TG accumulation under the premise of normal conditions.

To accurately understand the complex relationships between triacylglycerols (TG) and other metabolites, it is essential to find and approximate the global optimal solution, as the current HPL model only provides a local optimal solution to increase TG output. In this study, we used multiple deep learning algorithms to approximate a global optimum, developed plans to optimize current models, and quantitatively predict the TG output. We applied the genetic algorithm and simulated annealing algorithm to constrain the existing neural network models.[45,46] It is suggested that the results of both algorithms are consistent with the results generated by the previous HPL model, indicating that we approximated a global optimal solution. We also presented plans for using the brute force algorithm and convex optimization methods to optimize the HPL model.[48–50] Therefore, to better obtain the global optimal solution, we propose combining the HPL strategy with well-established quantitative methods such as metabolic flux analysis or metabolic kinetic analysis.[65] These methods can help researchers determine the quantitative effects of metabolic changes, optimize the efficiency and stability of the modeling process, and gain a better understanding of the complex relationships between TG and other metabolites, leading to insights into optimizing TG output.

In summary, the HPL strategy should promote studies on microbiology. The feasibility of this method is demonstrated by characterizing the DMACG phenomenon in *C. reinhardtii*. The established protocol and high compatibility combined with microbiology and bioinformatics may enhance further studies in single-cell microbial metabolomics, greatly contributing to synthetic biology.

All abbreviations in this manuscript are summarized in Table S2 as a glossary.

Limitations of the study

RespectM method has limitations. The 50 µm MALDI laser raster inevitably causes a 9.1% co-sampling rate in the cell density of 100–200 cells/µL, while acquiring microbial single cell level metabolomics data. Although the RespectM method is easy to operate, it's hard to acquire single cell level data in a high microbial cell density. In the future, the applications of "MicroMS" and "SpaceM" methods could be employed in the pre- and post-MALDI session to do automated cell filtering under a high cell density.[55,66]

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 SUBJECT DETAILS
  - Strains, and growth conditions

- METHOD DETAILS
  - Sample preprocessing and microarray preparation
  - Matrix application and cell localization
  - Data acquisition of single cell level metabolomics by MALDI-MRMS
  - Sample preparation for acquisition of LC-MS/MS data
  - LC-MS/MS methods for metabolomics
  - LC-MS/MS methods for lipidomics
  - RespectM pipeline of single cell metabolite-lipid identification
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107069.

AUTHOR CONTRIBUTIONS

X.M. and F.T. performed the experiments and analyzed the data. X.M., F.T., and P.X. designed the research. X.M. wrote the manuscript. F.T. and P.X. reviewed the manuscript and revised it. All authors approved the submitted version.

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

REFERENCES

1. Dahabieh, M.S., Thevelein, J.M., and Gibson, B. (2020). Multimodal microorganism development: integrating top-down biological engineering with bottom-up rational design. Trends Biotechnol. 38, 241–253.

2. Wang, J., Nielsen, J., and Liu, Z. (2021). Synthetic biology advanced natural product discovery. Metabolites 11, 785.

3. Ando, D., and Garcia Martin, H. (2018). Two-scale 13 C metabolic flux analysis for metabolic engineering. In Synthetic Metabolic Pathways (Springer), pp. 333–352.

4. Aurand, E., Keasling, J., Friedman, D., and Salis, H. (2019). Engineering Biology: A Research Roadmap for the Next-Generation Bioeconomy (Engineering Biology Research Corporation).

5. Zhang, J., Chen, Y., Fu, L., Guo, E., Wang, B., Dai, L., and Si, T. (2021). Accelerating strain engineering in biofuel research via build and test automation of synthetic biology. Curr. Opin. Biotechnol. 67, 88–98.

6. Radivojević, T., Costello, Z., Workman, K., and Garcia Martin, H. (2020). A machine learning Automated Recommendation Tool for synthetic biology. Nat. Commun. 11, 1–14.

7. Nielsen, J., and Keasling, J.D. (2016). Engineering cellular metabolism. Cell 164, 1185–1197.

8. Gardner, T.S. (2013). Synthetic biology: from hype to impact. Trends Biotechnol. 31, 123–125.

9. Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? Cell 141, 559–563.

10. Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338, 1622–1626.

11. Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A.R., Zhu, P., Hu, X., Xu, L., Yan, L., et al. (2012). Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science 338, 1627–1630.

12. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1214.

13. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201.

14. Spitzer, M.H., and Nolan, G.P. (2016). Mass cytometry: single cells, many features. Cell 165, 780–791.

15. Krismer, J., Sobek, J., Steinhoff, R.F., Fagerer, S.R., Pabst, M., and Zenobi, R. (2015). Screening of Chlamydomonas reinhardtii populations with single-cell resolution by using a high-throughput microscale sample preparation for matrix-assisted laser desorption ionization mass spectrometry. Appl. Environ. Microbiol. 81, 5546–5551.

16. Krismer, J., Tamminen, M., Fontana, S., Zenobi, R., and Narwani, A. (2017). Single-cell mass spectrometry reveals the importance of genetic diversity and plasticity for phenotypic variation in nitrogen-limited Chlamydomonas. ISME J. 11, 988–998.

17. Hillesland, K.L., and Stahl, D.A. (2010). Rapid evolution of stability and productivity at the origin of a microbial mutualism. Proc. Natl. Acad. Sci. USA 107, 2124–2129.

18. Martins, B.M.C., and Locke, J.C.W. (2015). Microbial individuality: how single-cell heterogeneity enables population level strategies. Curr. Opin. Microbiol. 24, 104–112.

19. Delvigne, F., and Goffin, P. (2014). Microbial heterogeneity affects bioprocess robustness: dynamic single-cell analysis contributes to understanding of microbial populations. Biotechnol. J. *9*, 61–72.

20. Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., Nyerges, Á., Szamecz, B., Fekete, G., Papp, B., et al. (2017). Phenotypic heterogeneity promotes adaptive evolution. PLoS Biol. *15*, e2000644.

21. Samal, B., and Chatterjee, S. (2019). New insight into bacterial social communication in natural host: evidence for interplay of heterogeneous and unison quorum response. PLoS Genet. *15*, e1008395.

22. Mould, D.L., and Hogan, D.A. (2021). Intraspecies heterogeneity in microbial interactions. Curr. Opin. Microbiol. *62*, 14–20.

23. Takhaveev, V., and Heinemann, M. (2018). Metabolic heterogeneity in clonal microbial populations. Curr. Opin. Microbiol. *45*, 30–38.

24. Gasol, J.M., Zweifel, U.L., Peters, F., Fuhrman, J.A., and Hagström, A. (1999). Significance of size and nucleic acid content heterogeneity as measured by flow cytometry in natural planktonic bacteria. Appl. Environ. Microbiol. *65*, 4475–4483.

25. Ibáñez, A.J., Fagerer, S.R., Schmidt, A.M., Urban, P.L., Jefimovs, K., Geiger, P., Dechant, R., Heinemann, M., and Zenobi, R. (2013). Mass spectrometry-based metabolomics of single yeast cells. Proc. Natl. Acad. Sci. USA *110*, 8790–8794.

26. Amantonico, A., Urban, P.L., Fagerer, S.R., Balabin, R.M., and Zenobi, R. (2010). Single-cell MALDI-MS as an analytical tool for studying intrapopulation metabolic heterogeneity of unicellular organisms. Anal. Chem. *82*, 7394–7400.

27. Amantonico, A., Oh, J.Y., Sobek, J., Heinemann, M., and Zenobi, R. (2008). Mass spectrometric method for analyzing metabolites in yeast with single cell sensitivity. Angew. Chem. Int. Ed. *47*, 5382–5385.

28. Bien, T., Bessler, S., Dreisewerd, K., and Soltwisch, J. (2021). Transmission-mode MALDI mass spectrometry imaging of single cells: optimizing sample preparation protocols. Anal. Chem. *93*, 4513–4520.

29. Li, W.V., and Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat. Commun. *9*, 997–999.

30. Shen, X., Gong, X., Cai, Y., Guo, Y., Tu, J., Li, H., Zhang, T., Wang, J., Xue, F., and Zhu, Z.-J. (2016). Normalization and integration of large-scale metabolomics data using support vector regression. Metabolomics *12*, 1–12.

31. Trede, D., Schiffler, S., Becker, M., Wirtz, S., Strehlow, J., Kobarg, J.H., Steinhorst, K., Aichler, M., Oetjen, J., and Dyatlov, A. (2012). O5. scils lab: software for analysis and interpretation of large maldiims datasets. OurCon *2012*, 50.

32. Shoemaker, W.R., Jones, S.E., Muscarella, M.E., Behringer, M.G., Lehmkuhl, B.K., and Lennon, J.T. (2021). Microbial population dynamics and evolutionary outcomes under extreme energy limitation. Proc. Natl. Acad. Sci. USA *118*, e2101691118.

33. Ballabio, D., and Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. Anal. Methods-UK. *5*, 3790–3798.

34. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.

35. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

36. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

37. Gemperline, E., Rawson, S., and Li, L. (2014). Optimization and comparison of multiple MALDI matrix application methods for small molecule mass spectrometric imaging. Anal. Chem. *86*, 10030–10035.

38. Patterson, N.H., Tuck, M., Van de Plas, R., and Caprioli, R.M. (2018). Advanced registration and analysis of MALDI imaging mass spectrometry measurements through autofluorescence microscopy. Anal. Chem. *90*, 12395–12403.

39. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.

40. Nowicka, B., Zyzik, M., Kapsiak, M., Ogrodzińska, W., and Kruk, J. (2021). Oxidative stress limits growth of Chlamydomonas reinhardtii (Chlorophyta, Chlamydomonadales) exposed to copper ions at the early stage of culture growth. Phycologia *60*, 1–11.

41. Chen, H., Albergante, L., Hsu, J.Y., Lareau, C.A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A.N., Bauer, D.E., Aryee, M.J., et al. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat. Commun. *10*, 1903.

42. Tanaka, A., Ito, H., Tanaka, R., Tanaka, N.K., Yoshida, K., and Okada, K. (1998). Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. Proc. Natl. Acad. Sci. USA *95*, 12719–12723.

43. Li-Beisson, Y., Beisson, F., and Riekhof, W. (2015). Metabolism of acyl-lipids in Chlamydomonas reinhardtii. Plant J. *82*, 504–522.

44. Long, F., Liu, W., Jiang, X., Zhai, Q., Cao, X., Jiang, J., and Xu, J. (2021). State-of-the-art technologies for biofuel production from triglycerides: a review. Renew. Sustain. Energy Rev. *148*, 111269.

45. Lambora, A., Gupta, K., and Chopra, K. (2019). Genetic algorithm-A literature review. In 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon), IEEE), pp. 380–384.

46. Guilmeau, T., Chouzenoux, E., and Elvira, V. (2021). Simulated annealing: a review and a new scheme. In 2021 IEEE Statistical Signal Processing Workshop (SSP) (IEEE), pp. 101–105.

47. Van den Broeck, G., Lykov, A., Schleich, M., and Suciu, D. (2022). On the tractability of SHAP explanations. J. Artif. Intell. Res. *74*, 851–886.

48. El Emam, K., Dankar, F.K., Issa, R., Jonker, E., Amyot, D., Cogo, E., Corriveau, J.P., Walker, M., Chowdhury, S., Vaillancourt, R., et al. (2009). A globally optimal k-anonymity method for the de-identification of health data. J. Am. Med. Inf. Assoc. *16*, 670–682.

49. Pilanci, M., and Ergen, T. (2020). Neural Networks Are Convex Regularizers: Exact Polynomial-Time Convex Optimization Formulations for Two-Layer Networks (ICML), pp. 7695–7705.

50. Wang, Y., Lacotte, J., and Pilanci, M. (2021). The Hidden Convex Optimization Landscape of Regularized Two-Layer Relu Networks: An Exact Characterization of Optimal Solutions (ICLR).

51. Schmitz, A.C., Hartline, C.J., and Zhang, F. (2017). Engineering microbial metabolite dynamics and heterogeneity. Biotechnol. J. *12*, 1700422.

52. Lau, A.N., Li, Z., Danai, L.V., Westermark, A.M., Darnell, A.M., Ferreira, R., Gocheva, V., Sivanand, S., Lien, E.C., Sapp, K.M., et al. (2020). Dissecting cell-type-specific metabolism in pancreatic ductal adenocarcinoma. Elife *9*, e56782.

53. Li, H., Ning, S., Ghandi, M., Kryukov, G.V., Gopal, S., Deik, A., Souza, A., Pierce, K., Keskula, P., Hernandez, D., et al. (2019). The landscape of cancer cell line metabolism. Nat. Med. *25*, 850–860.

54. Robinson, J.L., Kocabaş, P., Wang, H., Cholley, P.-E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., et al. (2020). An atlas of human metabolism. Sci. Signal. *13*, eaaz1482.

55. Rappez, L., Stadler, M., Triana, S., Gathungu, R.M., Ovchinnikova, K., Phapale, P., Heikenwalder, M., and Alexandrov, T. (2021). SpaceM reveals metabolic states of single cells. Nat. Methods *18*, 799–805.

56. Hankin, J.A., Barkley, R.M., and Murphy, R.C. (2007). Sublimation as a method of matrix application for mass spectrometric imaging. J. Am. Soc. Mass Spectrom. *18*, 1646–1652.

57. Huizing, L.R.S., Ellis, S.R., Beulen, B.W.A.M.M., Barré, F.P.Y., Kwant, P.B., Vreeken, R.J., and Heeren, R.M.A. (2019). Development and evaluation of matrix application techniques for high throughput mass spectrometry imaging of tissues in the clinic. Clin. Mass Spectrom. *12*, 7–15.

58. Jackson, C.A., Castro, D.M., Saldi, G.-A., Bonneau, R., and Gresham, D. (2020). Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. Elife 9, e51254.

59. Jariani, A., Vermeersch, L., Cerulus, B., Perez-Samper, G., Voordeckers, K., Van Brussel, T., Thienpont, B., Lambrechts, D., and Verstrepen, K.J. (2020). A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. Elife 9, e55320.

60. Holmes, D., Pettigrew, D., Reccius, C.H., Gwyer, J.D., van Berkel, C., Holloway, J., Davies, D.E., and Morgan, H. (2009). Leukocyte analysis and differentiation using high speed microfluidic single cell impedance cytometry. Lab Chip 9, 2881–2889.

61. Di Palma, S., and Bodenmiller, B. (2015). Unraveling cell populations in tumors by single-cell mass cytometry. Curr. Opin. Biotechnol. 31, 122–129.

62. Dahlqvist, A., Ståhl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., and Stymne, S. (2000). Phospholipid: diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc. Natl. Acad. Sci. USA 97, 6487–6492.

63. Dey, P., Su, W.-M., Han, G.-S., and Carman, G.M. (2017). Phosphorylation of lipid metabolic enzymes by yeast protein kinase C requires phosphatidylserine and diacylglycerol. J. Lipid Res. 58, 742–751.

64. Yamagata, H., Matoba, R., Fujii, T., and Yukawa, H. (1998). Application of photosynthetic bacteria for porphyrin production. In Studies in Surface Science and Catalysis (Elsevier), pp. 475–478.

65. Kim, G.B., Kim, W.J., Kim, H.U., and Lee, S.Y. (2020). Machine learning applications in systems metabolic engineering. Curr. Opin. Biotechnol. 64, 1–9.

66. Comi, T.J., Neumann, E.K., Do, T.D., and Sweedler, J.V. (2017). microMS: a Python platform for image-guided mass spectrometry profiling. J. Am. Soc. Mass Spectrom. 28, 1919–1928.

67. Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., Yang, C., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. Cell 172, 249–261.e12.

68. Meng, X., Zhao, X., Ding, X., Li, Y., Cao, G., Chu, Z., Su, X., Liu, Y., Chen, X., Guo, J., et al. (2020). Integrated functional omics analysis of flavonoid-related metabolism in atmyb12 transcript factor overexpressed tomato. J. Agric. Food Chem. 68, 6776–6787.

69. Wang, R., and Miao, X. (2022). Lipid turnover and SQUAMOSA promoter-binding proteins mediate variation in fatty acid desaturation under early nitrogen deprivation revealed by lipidomic and transcriptomic analyses in Chlorella pyrenoidosa. Front. Plant Sci. 13, 987354.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Chlamydomonas reinhardtii	All strains used in this study were listed in Table S3	CC124, CC-125, CC-425, CC-503, CC-4346, CC-4414, CC-4909, CC-4910, CC-5164
Chemicals, peptides, and recombinant proteins		
Methanol	Sigma-Aldrich	Cat# 900688-1L
Acetonitrile	Thermo Fisher	Cat#A955-4
Dichloromethane	Macklin	Cat#D960496-10L
Chloroform	Sigma-Aldrich	Cat#366919-1L
Isopropanol	Aladdin	Cat#I573164-2.5L
Ammonium formate	Sigma-Aldrich	Cat#V900127-500G
Formic acid	Sigma-Aldrich	Cat#695076-100ML
2,5-Dihydroxybenzoic acid (DHB)	Aladdin	Cat#D119198-5g
Deposited data and code		
Metabolomics data	This study, MetaboLights	MTBLS3830
Source code	This study, Github	https://github.com/dawnmengsjtu/RespectM
Software and algorithms		
Matlab (v2021b)	Matlab software	https://www.mathworks.com/products/matlab.html
Seurat R package (v3.1.5)	cran.r-project.org	https://github.com/satijalab/seurat
scImpute R package (v0.0.8)	cran.r-project.org	https://github.com/Vivianstats/scImpute
MetNormalizer R package (v1.3.2)	cran.r-project.org	https://github.com/jaspershen/MetNormalizer
Stream Bioconda package (v1.1)	pinellolab.partners.org	https://github.com/pinellolab/STREAM

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Fei Tao (taofei@sjtu.edu.cn).

Materials availability

All the requests for the generated plasmids and strains should be directed to the lead contact and will be made available on request after completion of a Materials Transfer Agreement.

Data and code availability

- Metabolomics data is publicly available as of the date after publication. Accession numbers are listed in the key resources table.

- All original code and any additional information required to reanalyze the data reported in this paper are available from the lead contact upon reasonable request.

- All data generated during this study are available from the lead contact upon reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains, and growth conditions

Chlamydomonas reinhardtii strains were obtained from Chlamydomonas Resource Center University of Minnesota (http://www.chlamycollection.org) and Freshwater Algae Culture Collection at the Institute of

Hydrobiology (FACHB) (http://algae.ihb.ac.cn/) (Table S3). All species of algae were grown up on a solid TAP medium. Single clones were then transferred to a liquid TAP medium for three replicates, respectively. Cells were cultured in a shaker under ~100 μEinstein m^{-2} s^{-2} at 100 rpm.

METHOD DETAILS

Sample preprocessing and microarray preparation

For microarray preparation, the preprocessing part of this method is quite simple. We designed patterns on the surface of the ITO slide and used lasers for pattern-based etching. Since the endogenous single-cell status is different at each timepoint, we choose Days 3, 6, 9 for cell harvesting. When sampling for LC-MS/MS metabolomics and lipidomics, a small amount of cell fluid was taken to centrifuge and washed with deionized water. 10 mL *Chlamydomonas reinhardtii* cell culture was taken from Days 3, 6, 9. Samples were centrifuged at 2,000 × *g* for 5 minutes and resuspended two times in deionized water to reduce the remaining TAP medium among cells. Cells were then dilute to an appropriate density to balance the low co-sampling rate and acquisition efficiency. After dripping around 0.5ul cell suspension to the designated area of laser etching guided microarray (LEM), the chips were placed on the clean bench and waited 15–20 minutes for the blow-dry. Before matrix sublimation, it is necessary to take bright field optical graphs by stereomicroscope (Olympus & MVX10, Japan) of each cell droplet unit on the prefabricated cell chip. The purpose of this step is to recognize cells in optical images.

Matrix application and cell localization

After preparing microarrays, we used DHB as a matrix because it is suitable to analyze small molecules. Matrix was applied by sublimation on LEM. The parameter of sublimation was 181°C, 12 minutes (iMLayer, SHIMADZU, Japan). Since high-purity DHB is unstable in the air, we completed the downstream preparations within 20 minutes (https://www.aladdin-e.com/zh_cn/d119198.html).

To localize cells on microarrays, we used a stereomicroscope to take optical images. Needles were applied to mark the original laser-etched pattern on the LEM surface. Hence the micron level positioning was achieved. Meanwhile, the calculation of laser spot offset is the prerequisite to collect the data of single cell level metabolomics automatically. Generally, the laser does not bombard at the absolute central in flexcontrol (Bruker, USA), demanding that the relative offset from the experiment strike position was calculated to the absolute central position. Based on the calculation, the cells were marked on the adjusted positions in tissue profiling mode, fleximaging (Bruker, USA). Optical images were taken after matrix application for MSI *in situ* data acquisition. Cells were recognized combined with original optical images.

Data acquisition of single cell level metabolomics by MALDI-MRMS

Throughout the experiments, MALDI-TOF (UltrafleXtreme, Bruker) was applied to do pretest and MALDI-MRMS (SolariX 7.0T, Bruker) was used for high-resolution MSI acquisition. The temperature in the mass spectrometer room was set to 18°C, and the humidity is maintained between 50% and 60%. The instrument parameter settings are as follows. Funnel RF amplitude: 150 V; time of flight = 0.9 ms; transfer optic freq = 4 MHz; Q1 mass: 300 m/z; RF amplitude = 450 Vpp; funnel RF amplitude = 150 Vpp. The laser offset calculation should be performed before the experiments. The acquisition point is then marked in the correct position on the fleximaging software. After data acquisition, optical images were taken to compare before laser ablation to ensure the accuracy of laser bombardment position. We collected ten blank matrix spots in each sequence for the quality control. In parallel, 20 standards were selected due to their biological function and m/z. The m/z values of A, B, C, D, and E were 100, 300, 500, 700, and 900, respectively. The selection of 5 m/z can ensure the confidence of calibration between 60–1,500 m/z. After the acquisition, data is processed by Root Mean Square (RMS) algorithm in SCiLS Lab software (Bruker, USA).

Sample preparation for acquisition of LC-MS/MS data

For the extraction of metabolites, we used 80% methanol as the solvent, sonicate and centrifuge at 3,000 rpm for 15 minutes. The supernatant is then taken out for centrifugation at 12,000 rpm for 30 minutes. The extracted metabolites were lyophilized, and then reconstituted in 80% methanol.; Folch reagent (chloroform: methanol = 2:1) was used for lipid extraction, then freeze-dried, and re-dissolved in a solvent (chloroform: methanol = 1:1). The extracted lipid were lyophilized, and then reconstituted in methanol: dichloromethane = 1:1.

LC-MS/MS methods for metabolomics

This study applied Thermo UPLC Q-Extractive plus (QE) coupled with ESI ion source for metabolomics data acquisition. Acquity UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) was chosen for positive and negative mode analysis. In the mobile phase selection, we used 0.1% formic acid in diluted water as aqueous phase A and 0.1% formic acid in pure acetonitrile as organic phase B.[67,68] The LC gradient elution program was as follows: t = 0.0 min, 99% A; t = 5.0 min, 99% A; t = 5.5 min, 70% A; t = 9 min, 100% B; t = 11 min, 100% B; and t = 12.1 min, 99% A, respectively.

The MS parameters of C18–ESIMS in the analysis of positive and negative ionization mode were as follows: the mass range was set from m/z 80 to 1,000, with a spectra collection rate of 2.0 Hz and capillary voltage of 4,500 V; the gas flow rate of the nebulizer was 1.6 bar; the velocity of dry gas at 220°C was 6.0 L/min; funnel 1 and 2 radio frequencies (RFs) were set as 150 Vpp and 200 Vpp; collision-induced dissociation (CID) energy was set as 0 eV; quadrupole ion energy was 5 eV; collision cell energy was 7.0 eV; pre-pulse storage was 5.0 μs; collision RF ramp was from 400 Vpp to 800 Vpp; transfer time ramp was from 50 μs to 100 μs.

QE plus with ESI ion source was also applied for metabolomics data acquisition. Acquity UPLC BEH C18 column (1.7 μm, 2.1 × mm) was also chosen for positive and negative mode analysis. The data acquisitions of metabolomics and lipidomics were performed on Xcalibur (ThermoFisher, USA). The following parameters were set to correspond to Meng et al.[68]

LC-MS/MS methods for lipidomics

This study also applied Thermo UPLC Q-Extractive plus (QE) coupled with ESI ion source for lipidomic data acquisition. Acquity UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) was chosen for positive and negative mode analysis. In the mobile phase selection, we used Acetonitrile/water (60:40) with 10 mM ammonium formate and 0.1% formic acid as phase A, and Isopropanol/acetonitrile (90:10) with 10 mM ammonium formate and 0.1% formic acid as phase B. The LC gradient elution program was as follows: t = 0.0 min, 99% A; t = 5.0 min, 99% A; t = 5.5 min, 70% A; t = 9 min, 100% B; t = 11 min, 100% B; and t = 12.1 min, 99% A, respectively.

We chose the scan mode as follow: DDA mode, 1 full scan followed by 10 MS/MS scans. Collision energy is NEC 20, 45 to fragment the ions. Nitrogen (99.999%) was used as collision-induced dissociation gas. Full scan range: 150 to 2000 amu; Full scan resolution: 70000, AGC: 1e6, IT: 100ms; dd-MS/MS resolution: 17500, AGC:5e5, IT: 50ms; spray voltage 3.8 kV (positive mode); 3.0 kV (negative mode); capillary temperature: 320°C; s-lens RF level: 50 V.

RespectM pipeline of single cell metabolite-lipid identification

Compound Discover, commercial software was applied to identify metabolites. Identification results can be found in supporting information. In brief, we have classified the identification results of MALDI single-cell metabolomics with different confidence levels. The metabolites and lipids that bulk omics can directly annotate were classified into Level 1 and 2, and the MS1 peaks identified by the metID algorithm were classified as Level 3 according to adduct annotation. MS spectra of metabolites have been provided for validation.[68]

Commercial softwares LipidSearch and TraceFinder were applied to identify lipids. To determine the if the lipids belong to *Chlamydomonas reinhardtii* species or not, we consulted the expert in algae research. We set the parameter to identify lipids as follow: m/z delta = ±5 ppm; PQ (the "highest quality" parameter calculated by LipidSearch software) > 0.85; CV (the standard deviation of the peak area for three replicates of a biological sample divided by the mean peak area) < 0.3; R (the linear correlation of a three-point dilution series of a biological sample) > 0.9; m-score (the degree of matching between identified substance and its fragmentation pattern) > 2.0; Rej (the "Reject" parameter calculated by LipidSearch) = 0. According to the advice and references, we kept 48-54 carbon chain TG for single cell analysis and modeling. We kept the most lipid identified by LipidSearch with the grade of A and B, finally 8 class of lipids remains. After the parameters set above, we checked the lipidion according to the annotation of LipidSearch and retained the identified lipids as the principle above. The DGTS and MGTS is not included in the LipidSearch software,[69] therefore, we applied the TraceFinder to do identification. Based on the calculation of the isotope internal standard, we obtained an average recovery rate of 94.01%.

QUANTIFICATION AND STATISTICAL ANALYSIS

At least three biological replicates were performed for each experiment. Data are given as the means $\pm$ standard deviation (SD). All comparisons to determine differences were performed by applying Student's t test. R statistical software (R Core Team, Vienna, Austria) was used for data processing and analysis. Mean values were regarded as significantly different at $p < 0.05$.