bioRxiv
THE PREPRINT SERVER FOR BIOLOGY
Cold Spring Harbor Laboratory
NEW RESULTS
BIOINFORMATICS

MEBOCOST: Metabolic Cell-Cell Communication Modeling by Single Cell Transcriptome

Rongbin Zheng [1 2], Yang Zhang [3 4], Tadataka Tsuji [3 4], Lili Zhang [1 2], Yu-Hua Tseng [3 4] and Kaifu Chen [1 2 *]

[1] *Basic and Translational Research Division, Department of Cardiology, Boston Children's Hospital, Boston, MA 02115, USA*

[2] *Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA*

[3] *Section on Integrative Physiology and Metabolism, Joslin Diabetes Center, Harvard Medical School, Boston, MA 02115, USA*

[4] *Department of Medicine, Harvard Medical School, Boston, MA 02115, USA*

[*] correspondence should be addressed to Kaifu Chen (Kaifu.Chen{at}childrens.harvard.edu)

bioRxiv preprint DOI: https://doi.org/10.1101/2022.05.30.494067

Posted: May 31, 2022, Version 1

Abstract

We developed MEBOCOST, a computational algorithm for quantitatively inferring metabolite-based intercellular communications using single cell RNA-seq data. MEBOCOST predicted cell-cell communication events for which metabolites, such as lipids, are secreted by one cell (sender cells) and traveled to interact with sensor proteins of another cell (receiver cells). The sensor protein on receiver cell might be cell surface receptor, cell surface transporter, and nuclear receptor. MEBOCOST relies on a curated database of metabolite-sensor partners, which we collected from the literatures and other public sources. Based on scRNA-seq data, MEBOCOST identifies cell-cell metabolite-sensor communications between cell groups, in which metabolite enzymes and sensors were highly expressed in sender and receiver cells, respectively. Applying MEBOCOST on brown adipose tissue (BAT) showed the robustness of predicting known and novel metabolite-based autocrine and paracrine communications. Additionally, MEBOCOST identified a set of intercellular metabolite-sensor communications that was regulated by cold exposure in BAT. Those predicted communicating metabolites and sensors may play important roles in thermogenesis regulation. We believe that MEBOCOST will be useful to numerous researchers to investigate metabolite-based cell-cell communications in many biological and disease models, thus will be useful to remove critical barriers impeding the development of new therapies to target these communications. MEBOCOST is freely available at https://github.com/zhengrongbin/MEBOCOST.

Introduction

Communication between cells, or cell-cell communication, is an integral part of cellular function in a human tissue. It is a critical process that maintains the functions and hemostasis of cells, organs, and intact systems[1]. Abnormal cell-cell communications are key contributors to many health conditions such as obesity[2], diabetes[3], heart disease[4], and cancer[5]. Communications between cells can be mediated by various types of molecules, e.g., proteins and metabolites. Protein-mediated cell-cell communications, e.g., those mediated by protein ligand-receptor pairs, have been the subject of many recent investigations based on single-cell RNA sequencing (scRNA-seq) and many robust algorithms[6]. Cell-cell metabolic reaction between cells is also frequently analyzed by inferring a metabolite generated by an enzyme in one cell and consumed as a substrate of another enzyme in a different cell[7]. For instance, FFA generated by lipase in adipocyte were reported to feed breast cancer cell, in which the FFA becomes substrate of acyl-CoA synthetases and is converted into Fatty Acel-CoA[8]. Several algorithms were recently reported to detect the generation and consumption of metabolites based on scRNA-seq data, thus indirectly enabled single-cell analysis of cell-cell metabolic reaction, e.g., COMPASS[9], scFEA[10], scFBA[11]. However, little computational resource is available to investigate metabolite-sensor communications.

In a cell-cell metabolite-sensor communication, a metabolite generated by one cell travels to another cell, which has a sensor protein that binds the metabolite to trigger a signaling pathway[12, 13]. For instance, polyamine produced and secreted by EC was reported to be sensed by β-adrenergic receptor on the surface of white adipocyte to regulate adiposity[14]. In contrast to enzymes that produce or consume the metabolite in a cell-cell metabolic reaction, sensor proteins often do not consume the metabolite. Instead, sensor proteins often bind and release the metabolites to trigger and end cell signaling, respectively. Due to this mechanistic difference in the underlying biology, existing methods for analyzing cell-cell metabolic reaction are not applicable to analysis of metabolite-sensor communications. Current algorithms to analyze ligand-receptor communications were focused on protein or peptide ligands and thus designed in ways that do not support analysis of metabolite-sensor communications. Two major constraints on investigation of cell-cell metabolite-sensor communications include the lack of a curated catalogue of reported metabolite-sensor pairs and a paucity of robust methods to detect active metabolite-sensor communications in a sample.

In this study, we addressed these most pressing constraints by developing a novel bioinformatics technology, called MEBOCOST, which enables researchers to detect cell-cell metabolite-sensor communications by analyzing transcriptomes of single cells. We applied MEBOCOST to study cell-cell metabolite-sensor communications during thermogenesis, which

is the process of heat production in brown adipose tissue in response to cold temperatures. MEBOCOST successfully identified known and novel metabolite-sensor communications in brown adipose tissue of mouse housed at cold temperature for 2 days. It further identified cold-sensitive communication events of which the communication was reprogrammed between mice housed under acute cold, chronic cold, room temperature, and thermal neutral conditions. By delivering a novel technology to enable systematic analysis of cell-cell metabolite-based communications, MEBOCOST will pave a new avenue for investigating the molecular basis of development and diseases.

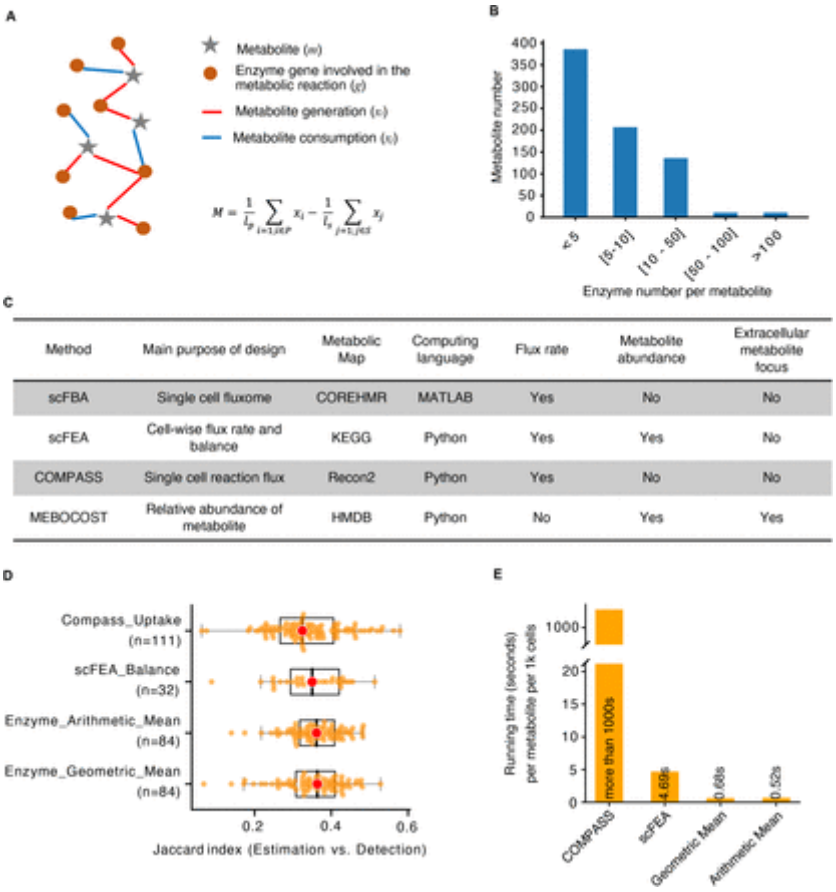
Results

The algorithm in MEBOCOST for detection of cell-cell metabolite-sensor communications

In metabolically active cells, expressed metabolic enzymes catalyze metabolic reactions to produce many metabolites. Some metabolites can diffuse into extracellular space and function as signaling molecules. Some extracellular metabolites can bind sensor proteins of spatially nearby cells. We termed the cells that secret the metabolites as sender cells and the cells that express the sensor proteins as receiver cells. Therefore, interactions between these metabolites and sensor proteins could mediate communications between sender and receiver cells in a paracrine manner. They may also mediate autocrine when the sender cell and receiver cell are the same cell. To analyze metabolite-sensor communications between sender and receiver cells, we developed a computational algorithm, called MEBOCOST, which takes processed scRNA-seq data as input.

MEBOCOST integrates single cell RNA expression data and prior knowledge of extracellular metabolites, metabolic enzymes, and metabolite-sensor partners to detect metabolite-sensor communications (**Figure 1**). MEBOCOST predicts communication events of which the enzymes and sensors of a metabolite were highly expressed in the sender and receiver cells, respectively. To this end, MEBOCOST first extracted gene expression of the enzymes from scRNA-seq data and inferred the presence of metabolite in a cell based on the enzyme RNA expression (**Supplementary Figure 1A, 1B**, details in Methods and Materials). The enzyme genes of extracellular metabolites were collected from the Human Metabolome Database (HMDB)[15-18] which provides comprehensive annotation for 220,945 metabolites. Considering that the relation between metabolite abundance and RNA expression level of metabolic enzymes are often nonlinear, we do not expect to quantitatively calculate the abundance of metabolite abundance. We instead aimed at inferring the presence of metabolite based on scRNA-Seq data and interpret it as a qualitative result. Several algorithms have been reported to accomplish such inference. Therefore, we designed MEBOCOST to take the output of these algorithms as its input data. These algorithms include the scFBA[1411], scFEA[1310], and

COMPASS[129] that each performed flux balance analysis in a unique way (**Supplementary Figure 1C**). Whereas flux balance analysis is popular for conventional metabolic analysis, the drawback is that it relies on some assumptions that might not be valid in many applications. Different assumptions will lead to different results. Also, flux balance analysis is computation-intensive and takes days or even weeks to do a genome-wide analysis of many single cells. In contrast, the mean RNA expression of ligand protein and receptor protein was frequently used to infer ligand-receptor communications, and is simple to calculate, biologically easy to interpret, and computationally efficient. A gaussian mixture model was reported to infer metabolite distribution based on expression of related enzymes[4319]. It is yet unclear which methods might perform better when applied to infer presence of metabolite based on RNA-seq data. Therefore, we performed a comparison between several methods, including scFEA, COMPASS, and the geometric or arithmetic mean of expression levels of related enzymes in GEMs from the HMDB database. The scFBA was excluded in this preliminary work because we developed MEBOCOST based on a free open access policy, whereas scFBA is implemented based on the commercial platform MATLAB. The gaussian mixture model is excluded because no software or code was released to use this model. We used the four algorithms to analyze the CCLE dataset[20, 21], which includes metabolomics data of 225 metabolites and genome-scale transcriptomics data from each of 928 cell lines. To measure the performance of each algorithm, we calculated Jaccard index for the overlap between metabolites inferred based on RNA expression and detected by metabolomics. Unexpectedly, despite being much simpler, we found that the arithmetic and geometric mean algorithms perform better (greater Jaccard index value) than the COMPASS algorithms and as good as the scFEA algorithm (**Supplementary Figure 1D**). Further, the arithmetic and geometric mean algorithms are about 10 folds faster than scFEA and 1000 folds faster than COMPASS (**Supplementary Figure 1E**). Arithmetic mean performs slightly better than geometric mean. Therefore, we designed MEBOCOST to be compatible with all these four algorithms, but the arithmetic mean algorithm will be the default option in MEBOCOST to infer metabolite presence.



Supplementary Figure 1.

Estimation of metabolite presence based on the enzymes of metabolites.
A, The schematic plot showed for metabolite presence estimation. The grey star donated metabolite; the orange dots donated enzyme genes for metabolic reactions. The line donated the relationship between enzyme genes and metabolite. The red lines indicated that the metabolites are products of the enzymes. The blue lines indicated that the metabolites are substrates of the enzymes. **B**, The number of enzymes used for metabolite inference was shown in the bar-plot. The x-axis was the interval of enzyme gene number per metabolite, y-axis was he number of metabolites. **C**, The summary of flux-balance analysis tools for single-cell RNA-seq data together with MEBOCOST. **D**, The accuracy of metabolite presence inference was shown for different methods. The RNA-seq data from CCLE datasets were used for estimation; the matched metabolomics profiles of same samples from CCLE dataset were used as a ground truth (detection) in this analysis. Metabolite balance result given by scFEA and uptake reaction prediction given by COMPASS were included in this analysis, together with two straightforward methods in which arithmetic mean or geometric mean of enzyme gene expression were used. The x-axis was the Jaccard index calculated based on the overlapping presence of metabolites between estimation and detection. The higher Jaccard index indicates greater similarity between estimation and detection. **E**, The per-1000 cells running time of four methods was shown. The x-axis were the four methods for estimation metabolite presence using scRNA-seq data. The y-axis was the running time in seconds of the four methods.

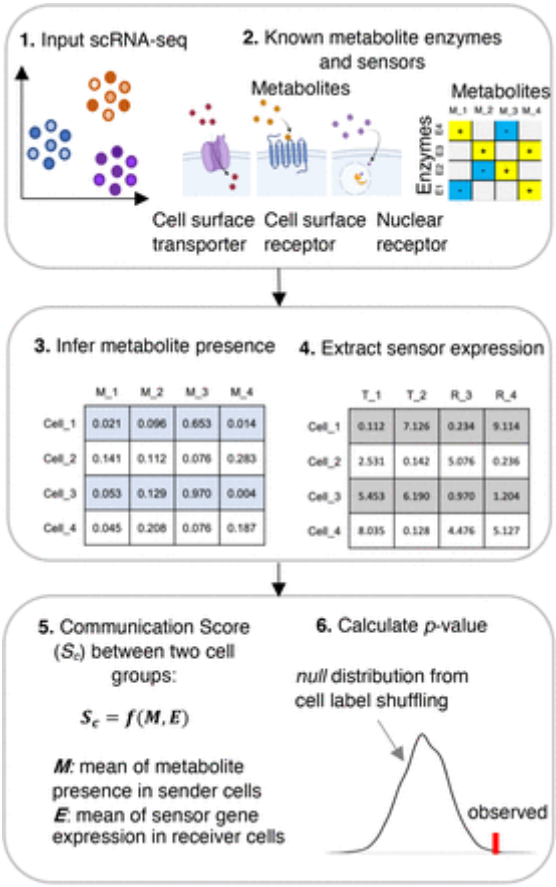


Figure 1.

The algorithm in MEBOCOST to preform cell-cell metabolite-sensor communications using scRNA-seq data. The prediction of cell-cell metabolite-sensor communications in MEBOCOST includes 5 steps. (1) The expression data and cell type annotation from processed scRNA-seq were taken as the input data. (2) The prior knowledge of metabolite-sensor partners and enzymes was incorporated into MEBOCOST. (3) The gene expression of metabolite enzymes was extracted from the scRNA-seq data. The presence of metabolite was infered by taking the mean of enzyme gene expression. (4) The gene expression of metabolite sensors was extracted from scRNA-seq data. (5) Calculation of communication score by taking the product of mean metabolite presence in sender cell population and mean sensor gene expression in receiver cell population. (6) Shuffling single cell labels to generate a statistical a *null* distribution to calculate *p* value for given communication score.

Meanwhile, expression values of sensor genes were extracted from the scRNA-seq data. Next, the score calculated by scFEA or COMPASS as estimated likelihood of metabolite presence, and sensor gene expression value, were averaged per cell groups. For a metabolite from a sender cell group and its sensor protein from a receiver cell group, the metabolite-sensor communication score will be calculated by taking the product of the average score of metabolite presence in the sender cell group and the average expression of the sensor protein in the receiver cell group. Such calculations were performed for each metabolite-sensor partner and each pair of cell types. Therefore, each pair of cell types will be assigned with a set of communication scores calculated for all the metabolite-sensor partners. A higher communication score represents a higher likelihood of a metabolite-sensor communication between the associated two cell types. Next, the significance of a communication score was evaluated. For each metabolite-sensor communication, MEBOCOST generated a *null* statistical distribution for the communication score by shuffling the cell labels of all cells in the scRNA-seq data 1000 times. Permutation test was performed to calculate the *p* value for the

communication score. The *p* values were further corrected to calculate false discovery rate (FDR) using the Benjamini-Hochberg's method[22].

A curated knowledge repository of enzyme-metabolite-sensor partners

To collect the extracellular metabolites and their enzymes (**Figure 2A**), we started from the 157 and 2,597 unique metabolites covered by scFEA and COMPASS, respectively. To focus on a subset of well-annotated metabolites, we retrieved the metabolites that could be mapped to those in the HMDB and further removed redundant metabolites that represent different names of the same metabolite. This resulted in 1,240 metabolites with unique accession number from the HMDB. As our interest is in cell-cell communications, we further narrowed down to 910 metabolites annotated as could be in extracellular space, blood, or cerebrospinal fluid. We defined these metabolites as extracellular metabolites in this study. To infer metabolite presence based on enzyme RNA expression, we next selected a subset of 441 metabolites that have annotated upstream enzymes in metabolic reactions.

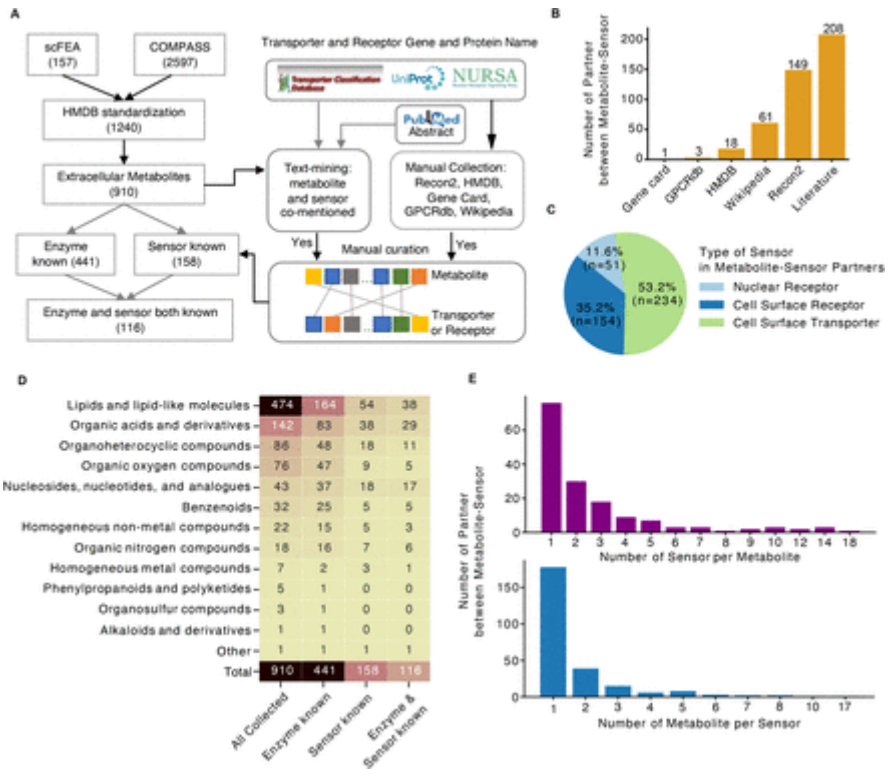


Figure 2.

The collection of extracellular metabolites, enzymes, and sensor proteins.
A, The pipeline of collecting extracellular metabolites and their enzymes as well as sensors. The extracellular metabolites and related enzymes were collected based on the annotation at Human Metabolome Database where the biological location was extracellular space, blood, and cerebrospinal fluid. The combination of text-mining and manual collection was performed to collect metabolite-sensor partners. **B**, The source of metabolite-sensor partners in our collection. X-axis represented the source; y-axis represented the number of metabolite-sensor partners. **C**, A pie plot to show the summary of sensor types of metabolite-sensor partners in our collection. The percentage of each category and the number in each category were shown. **D**, A summary of the collected extracellular metabolites. The columns correspond to each step of our focus in the collection. The rows were different classes of metabolites. The number labeled in the plot showed the number of metabolites in that class. **E**, The association between metabolites and sensors were summarized. Y-axis was the number of metabolite-sensor partners. The x-axis in the upper panel was the number of sensors per metabolite, and x-axis in the lower panel was the number of metabolites per sensor.

Although metabolite-sensor partners have been reported by many individual studies, few effort were taken to thoroughly collect and curate the catalogue of reported metabolite-sensor partners. To systematically collect reported metabolite-sensor partners, we developed a computational pipeline for comprehensive literature mining (**Figure 2A**). We first retrieved the 233 transporters from the Transporter Classification Database[23-27], the 1,522 cell surface receptor proteins from the UniProt database[28], and the 48 nuclear receptors from the Nuclear Receptor Signaling Atlas (NURSA) database[29]. Next, a text-mining approach was used to parse the metabolite-sensor partners from the abstracts of publications in the PubMed database. Briefly, an abstract will be selected if any pair of metabolite and sensor were co-mentioned in a sentence from the abstract. The obtained metabolite-sensor partners together with the corresponding abstracts were then subjected to three rounds of manual curation. At least three curators read each abstract independently and filter out the metabolite-sensor partners for which the evidence in the abstracts was insufficient. We also searched for annotated metabolite-sensor partners in five annotation databases: Recon2[30], HMDB[15-18], GeneCards[31], GPCRdb[32, 33], and the Nuclear Receptor page of the Wikipedia site. Altogether, these procedures result in 440 metabolite-sensor partners in total, of which 208, 149, 61, and 22 partners were collected from literature, Recon2, Wikipedia, and other sources, respectively (**Figure 2B**). Among these partners, 53.2% was metabolite-transporter partners, 35.2% was metabolite-cell surface receptor partners, and 11.6% was metabolite-nuclear receptor partners (**Figure 2C**). There are 116 metabolites in these partners (**Figure 2A**). Most of the metabolites (33%) in these partners are lipids and lipid-like molecules (**Figure 2D**). Organic acids and derivatives also constitute a major proportion (25%) of the metabolites. Other major categories include Nucleosides, nucleotides, and analogues (15%), Organoheterocyclic compounds (9%), along with 10 other minor types. Interestingly, among the collected metabolite-sensor partners, most metabolites each have one sensor, while some metabolites each have multiple sensors (**Figure 2E**). Also, most sensors each have one metabolite, while some sensors each have multiple metabolites. This database of metabolite-sensor partners represents a rich research resource for the community to perform systematic analysis of metabolite-sensor communications.

Scalable Computing Resource Usage and Robust Stability of MEBOCOST

To evaluate the usage of computing resource, we focused on a scRNA-seq dataset from brown adipose tissue (BAT) of mouse housed at cold temperature (4 ˚C) for 2 days (Cold2), which includes 33,470 cells in total. We applied MEBOCOST on the Cold2 data as well as a series of down-sampled data to test the scalability. We found that the time and computer memory taken by MEBOCOST had a linear correlation with the data size. The time to run MEBOCOST ranged between 9 and 7 minutes when the number of cells was sampled down from 33,470 to 11,157 (**Figure 3A**), where the peak memory usage ranged from 2.5 to 1.5Gb (**Figure 3B**). We also

evaluated the effect of sequencing depth on the performance of MEBOCOST. The result indicated that the number of detected communications showed little decrease even after the total cell numbers were sampled down to keep only 30% or the original data (**Figure 3C**). The Jaccard index of overlapped communications between original data and down sampled data were further calculated to evaluate the stability of the prediction. It was observed that the Jaccard index was stably greater than 0.9 even when the number of total cells was sampled down to 30% of the original data (**Figure 3D**). Taken together, MEBOCOST is computationally efficient and showed great resilience to low sequencing depth.

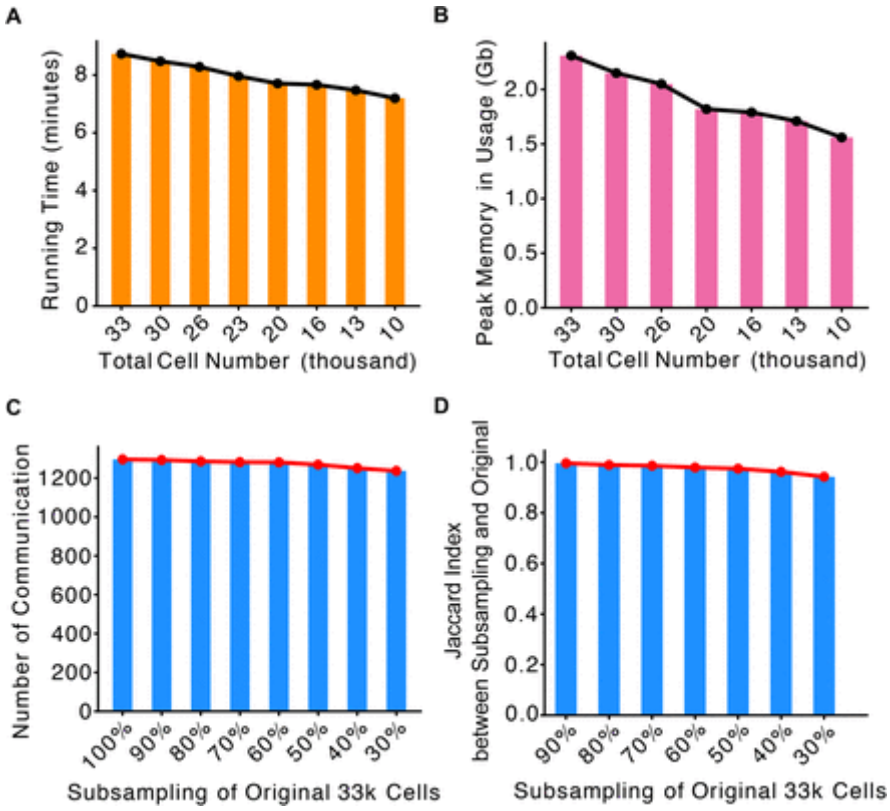


Figure 3.

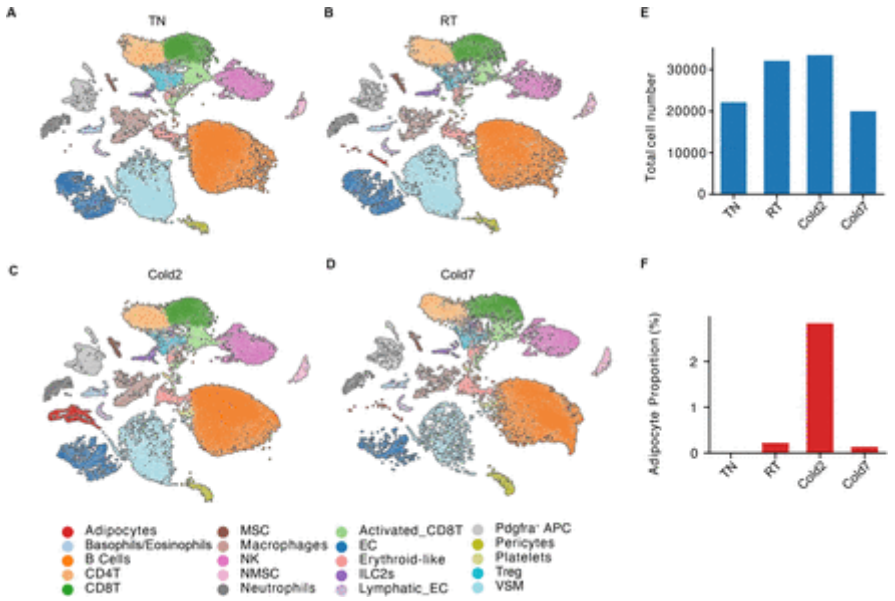
Robust performance and great scalability of MEBOCOST.

A, The running time in minutes was shown in bar-plot across a series number of total cells. The x-axis was the total number of cells in the test. The y-axis was the running time taken by MEBOCOST. **B**, Peak memory usage was summarized and shown by bar-plot. The x-axis was the total number cells included in the test. The y-axis was the peak memory in Gb. **C**. The number of communications detected by MEBOCOST using scRNA-seq data with different number of total cells. The x-axis was the subsampling percentage of total cells from original Cold2 BAT scRNA-seq data. The y-axis was the number of significant communications. A significant communication was defined if FDR < 0.05 and the presence of metabolite and sensor was at least in 25% of the cells in the associated population. **C**, The Jaccard index of overlapping communications between subsampling data and the original data was shown. The x-axis represented subsampling percentage of total cells from the original Cold2 BAT scRNA-seq data, and the y-axis represented the Jaccard index score.

Metabolic communication in mouse brown adipose tissue

Brown adipose tissue (BAT) is a metabolically active tissue and is specialized to dissipate chemical energy in the form of heat in thermogenesis. Cold exposure has been known as a strong activator for BAT[34-36]. It has been well accepted that cold exposure promotes not only thermogenesis but also the process of adiposity or differentiation of brown adipocytes[37]. The process of brown adipocyte differentiation can be regulated by metabolites in BAT[38, 39]. We

hypothesized that different cell types in BAT can communicate through metabolites, and some of the metabolites can target to adipocytes in order to regulate the function and differentiation process of adipocytes. Therefore, we performed scRNA-seq data for mouse BAT stromal vascular fraction. The data is consisted of 107,679 high-quality cells (**Supplementary Figure 2E**) and 468 million reads from BAT of mice housed at different environmental temperatures (**Supplementary Figure 2A-D**), including thermoneutral (TN) temperature (30 ˚C for a week), room temperature (RT, 22 ˚C), or cold temperature (4 ˚C for 2 days and 7days, Cold2 and Cold7). 20 cell types were identified. These include adipocytes, schwann cells, vascular cells, immune cells, etc. Mature adipocytes are enriched with lipid droplets and tend to be excluded from stromal vascular fraction, thus tend to be not captured by our scRNA-seq experiment. Therefore, most of adipocytes in this scRNA-seq dataset were differentiating adipocytes.



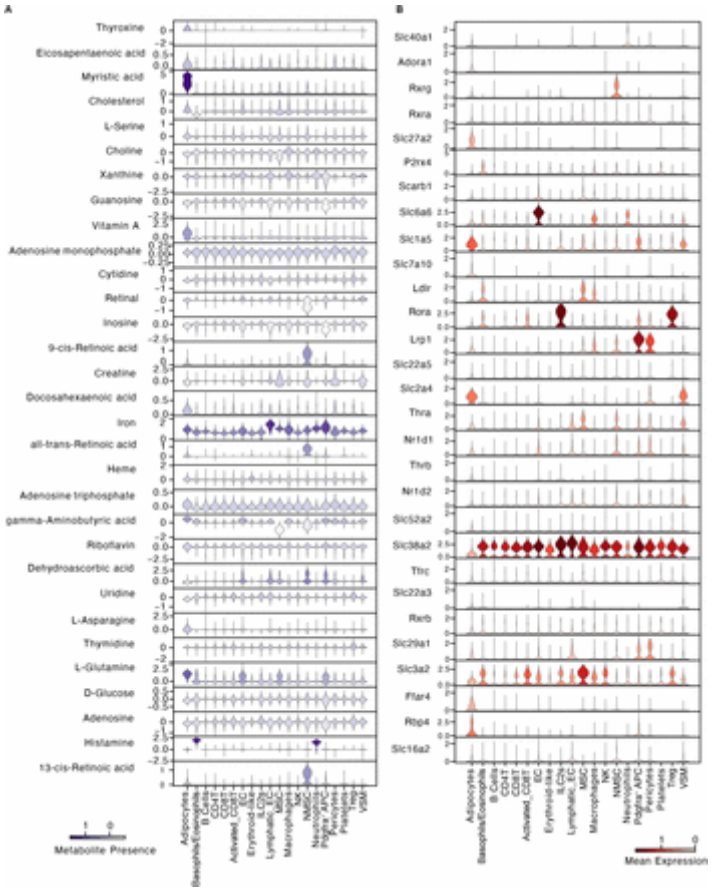
Supplementary Figure 2.

Single cell RNA-seq data of mouse adipose tissue

A-D, The UMAP visualization of scRNA-seq data from four conditions, including TN (30 °C for a week), RT (room temperature), Cold2(4 °C for 2 days), and Cold7 (4 °C for 7 days). **E**, The total number of cells in the scRNA-seq data from four conditions. The x-axis was the conditions. The y-axis was the total number of high-quality cells in the scRNA-seq. **F**, The proportion of differentiating adipocytes across four conditions. The x-axis was conditions, and y-axis was the proportion of differentiating adipocytes in the total cells of the condition.

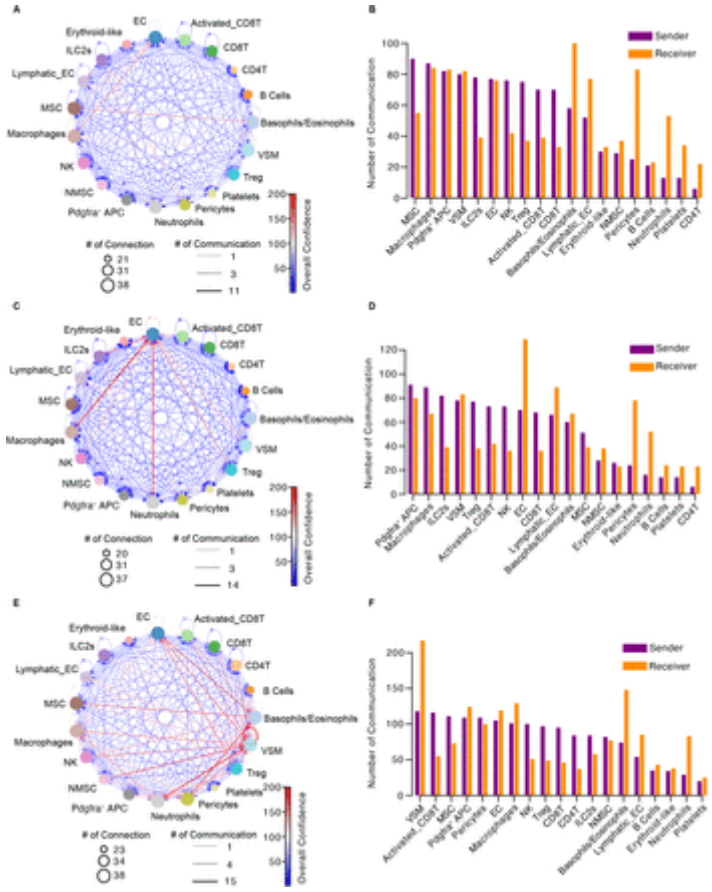
We first analyzed the BAT scRNA-seq data of Cold2 condition, as it contains more adipocytes when compared to the other three conditions (**Supplementary Figure 2E-F**). In total, 1,296 metabolite-sensor communications were detected in Cold2 BAT (FDR < 0.05, **Figure 4A**). In addition to paracrine, autocrine communications were also observed in many cell types such as Adipocytes, VSM, EC, Lymphatic EC, NK, NMSC, MSC, Pdgfra^{+} APC, etc. Among all communications, adipocytes showed the largest number of communications compared to other cell types (**Figure 4A, B**). Many of the communications sent to adipocytes showed higher overall confidence score compared to other communications (**Figure 4A**). Notably, adipocytes can be both senders and receivers in the metabolite-sensor communications in BAT. However,

much greater number of communication events were observed for adipocytes as receiver cells (256 events) than as sender cells (157 events) (**Figure 4B**). In addition, of all the adipocytes related communications, autocrine of adipocytes happened much more frequently than paracrine (**Figure 4C**). Such metabolite-sensor partners include Myristic acid ~ Slc27a2, L-Glutamine ~ Slc1a5, L-Glutamine ~ Slc38a2, L-Glutamine ~ Slc3a2, Vitamin A ~ Rbp4, Eicosapentaenoic acid ~ Ffar4, and Docosahexaenoic acid ~ Ffar4, and so on (**Figure 4D-E**). Some of the metabolites, sensors, and metabolite-sensor partners have been reported to be involved in regulating the function of BAT. For example, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), a type of omega-3 fatty acid, were reported to play an important role in brown adipocyte differentiation and thermogenesis[40]. Particularly, Jiyoung Kim and colleagues have reported that EPA can potentiate brown adipocyte thermogenesis dependent on Ffar4[41]. The critical role of vitamin A transport has also been demonstrated in adipose tissue browning and thermogenesis in cold exposure[42, 43]. Such importance was proven by knocking out Rbp, which is a gene encoding the vitamin A transporter[43]. Glutamine was reported as a major source of *de novo* fatty acid synthesis in a brown adipocyte cell line[44], and the utilization of glutamine was reported to be enhanced in brown adipose tissue by acute cold exposure[45]. Meanwhile, some metabolite-sensor communications were newly identified by MEBOCOST in this study. Among those, myristic acid and Slc27a2 were particularly noticed, since it had a highest communication score in adipocyte autocrine communications (**Figure 4D, 4E**). Interestingly, myristic acid was specifically and highly enriched in adipocytes when compared to other cell types in Cold2 BAT (**Supplementary Figure 3A**). Similarly, the mRNA expression of Scl27a2, a transporter for myristic acid, was also expressed specifically and higher in the adipocytes than in other cell types (**Supplementary Figure 3B**). However, the communication mediated by myristic acid and Scl27a2 in BAT were rarely studied. These results showed that MEBOCOST not only recaptured known but also discovered new metabolite-sensor communications for adipocyte autocrine.



Supplementary Figure 3.

The presence of significant communication-associated metabolites and sensors in cell types in brown adipose tissue. **A**, Violin plots showing the estimated metabolite presence for metabolites in significant communications predicted from Cold2 BAT scRNA-seq data. The x-axis represented the cell types, and y-axis represented the metabolites. The color of violin showed the averaged metabolite presence of cells in the cell type. **B**, Violin plot showing the gene expression level of metabolite sensors that was used in significant communications predicted from Cold2 BAT scRNA-seq data. The x-axis represented the cell types, and y-axis represented the sensor genes. The color of violin showed the averaged sensor gene expression of cells in the cell type.



Supplementary Figure 4.

The metabolite-sensor communications predicted by MEBOCOST in TN, RT, and Cold7 BAT.
A, C, E. Circle plot showing the cell-to-cell metabolite-sensor communications in TN, RT, and Cold7 BAT, respectively. Each dot was a cell type. The size of dots represented the number of communications with other cell types. The directional line represented the communication from sender cell type to receiver cell type. The line width indicated the number of metabolite-sensor communications between a sender and a receiver. The color of lines showed the overall confidence which was calculated by the sum of -log10(FDR) of all metabolite-sensor communications between a sender and a receiver. **B, D, F**, The number of communications for senders and receivers were shown in bar-plots for TN, RT, and Cold7 BAT, respectively. The x-axis was the cell types in BAT scRNA-seq data. The y-axis was the number of communications. The orange bars and purple bars were the number of communications for sender and receiver cells, respectively.

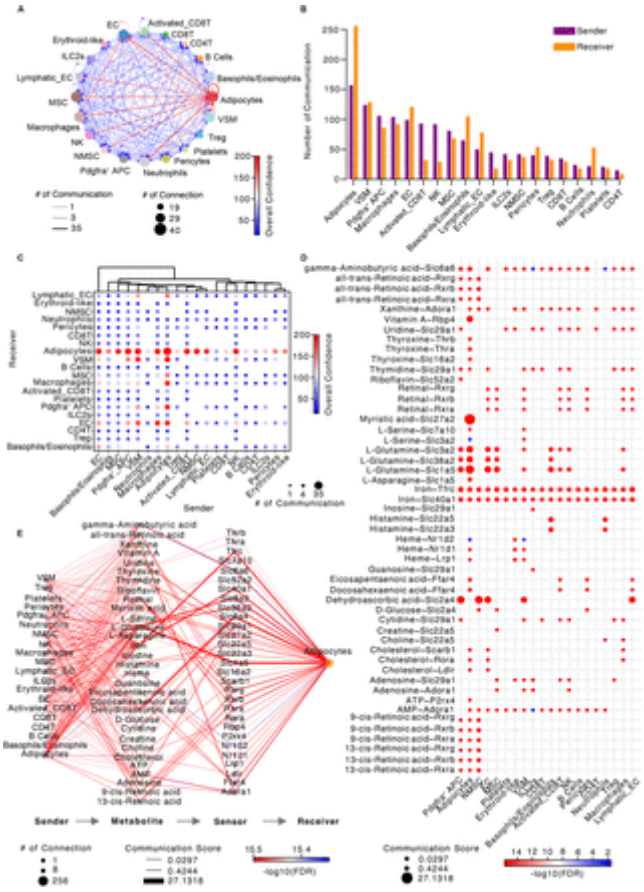


Figure 4.

Autocrine and paracrine of metabolite-sensor communications in brown adipose tissue from mouse exposed to cold for 2-days.

A, Circle plot showing cell-to-cell metabolite-sensor communications in Cold2 BAT predicted by MEBOCOST. Each dot was a cell type. The size of dots represented the number of communications with other cell types. The directional line represented the communication from sender cell type to receiver cell type. The line width indicated the number of metabolite-sensor communications between a sender and a receiver. The color of lines showed the overall confidence which was calculated by the sum of -log10(FDR) of all metabolite-sensor communications between a sender and a receiver. **B**, The number of communications for senders and receivers were shown in a bar-plot. The x-axis was the cell types in BAT scRNA-seq data. The y-axis was the number of communications. The orange bars and purple bars were the number of communications for sender and receiver cells, respectively. **C**, A dot-plot showing the communications between cell types in pairwise, where x-axis was the sender cell types, the y-axis was the receiver cell types. The dot size indicated the number of metabolite-sensor communications between a sender and a receiver. The color of the dot represented the overall confidence of communications between the sender and receiver. The calculation was the same with the overall confidence showed in panel A. **D**, A dot-plot showing metabolite-sensor communications to adipocytes from adipocytes themselves and other cell types. The rows were the metabolite-sensor partners, the columns are the sender cell types for adipocytes. The dot size indicated the communication score calculated by MEBOCOST. The color of the dot was the - log10(FDR) for the metabolite-sensor communication. **E**, A flow diagram showing the information flow of metabolite-sensor communications from sender cell type to receiver cell type through metabolites and sensors. The size of dots represented the number of connections in the diagram, indicating the frequency of usage among all the communications in the figure. The lines connect the sender, metabolite, sensor, and receiver. The color of the line indicated the -log10(FDR) for the metabolite-sensor communication. The width of lines represented the communication score. Figure 5

In addition to autocrine, adipocytes received many paracrine-style metabolite-sensor communications from other cell types in the brown adipose tissue. The most frequent sender cell types for adipocytes were Pdgfra^{+} APC, VSM, EC, MSC, NMSC, Activated CD8T, NK, and Lymphatic EC (**Figure 4C**). Among these sender cells, the communications between vascular system and adipocytes were well known[46]. For instance, Jennifer H Hammel and Evangelia Bellas reported that EC-adipocyte crosstalk can improve adipocyte browning, as proven based on an EC-adipocyte co-culture system[47]. Cholesterol-Scarb1 communication was reported to

regulate EC functions[48-51] as well as adipocyte function[52, 53], and were detected by MEBOCOST for both cell types. Meanwhile, some senders were newly detected by MEBOCOST. For instance, activated CD8 T cells, a subtype of CD8 T cells expressing cytotoxic markers, were much less reported regarding metabolite-sensor communication with adipocytes in BAT. Interestingly, 16 communications between activated CD8 T cells and adipocytes were detected in Cold2 BAT, while only 8 communications were detected for naïve CD8 T cells (**Figure 4C, 4D, 4E**). This observation indicates that the role of activated CD8 T cells in the process of thermogenesis is overlooked in the past. Strikingly, among all metabolite-sensor partners, the communication score of dehydroascorbic acid and Slc2a4 were much higher than other metabolite-sensor partners (**Figure 4D-E**). Such communications mediated by dehydroascorbic acid and Slc2a4 were detected between Pdgfra$^+$ APC and Adipocytes, NMSC and Adipocytes, EC and Adipocytes, VSM and Adipocytes, as well as Lymphatic EC and Adipocytes. Moreover, the metabolite presence inference showed that dehydroascorbic acid was highly enriched in those sender cells including Pdgfra$^+$ APC, NMSC, EC, and VSM when compared to other cell types. Meanwhile, we found that Slc2a4 was also highly and specifically expressed in brown adipocytes, indicating that those communications discussed above may function importantly for adipocytes. In summary, the analysis of MEBOCOST based on Cold2 BAT scRNA-seq data showed that the brown adipose resident cells frequently communicate with brown adipocytes through metabolite-sensor patners.

Identification of cold temperature-sensitive metabolite-sensor communications in BAT

We next performed comparative analysis of metabolite-sensor communications in BAT among the four conditions including TN, RT, Cold2, and Cold7. We constructed the landscape of metabolite-sensor communications in BAT under the different conditions. BAT scRNA-seq data from four conditions was analyzed by MEBOCOST separately. The total number of communications increased during thermogenesis in response to cold exposure, especially by the chronic cold exposure for 7 days (**Figure 5A**). In total, 1,032, 1,014, 1,296, and 1,557 communication events were detected in BAT under the TN, RT, Cold2, and Cold7 conditions, respectively (**Figure 5A**). Many metabolites and sensors appeared to be cell type- and condition-specific (**Figure 5B, 5C**). Furthermore, the composition of communications between cell types was also dramatically changed by temperature. For instance, the Basophils/Eosinophils received the largest number of communications from other cell types in TN, while EC, Adipocytes, and VSM were the cell types receiving the most communications in RT, Cold2, and Cold7, respectively. Similar changes were observed for senders. The MSC was the most frequently used sender cell type in TN, while Pdgfra$^+$ APC, Adipocytes, and VSM were identified to send the greatest number of metabolites to other cell types in RT, Cold2, and Cold7, respectively. These results suggested that the metabolite-sensor communications in mouse BAT were regulated by environment temperatures.

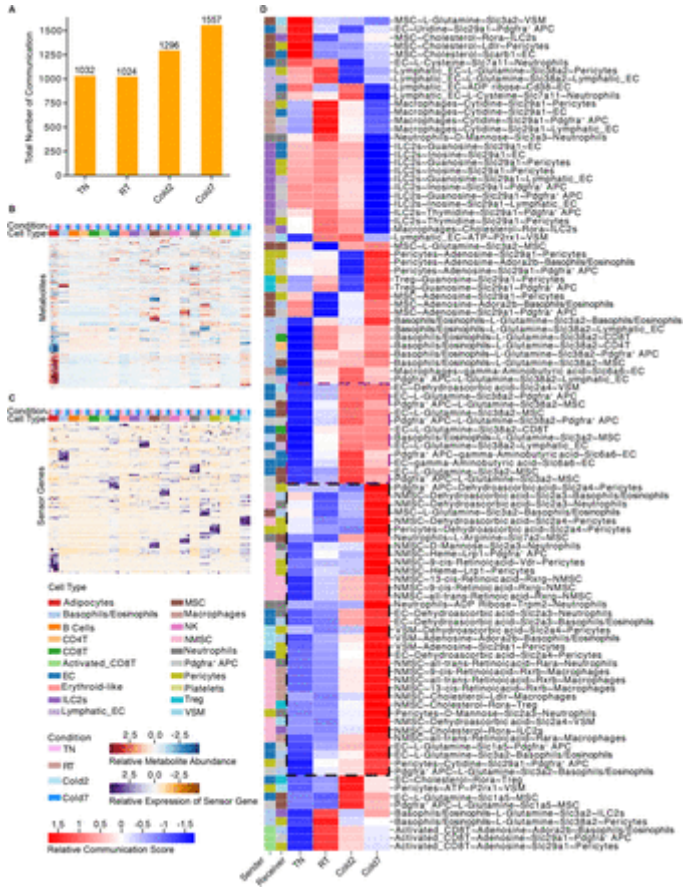


Figure 5.

Cold temperature-sensitive metabolite-sensor communications in brown adipose tissue.
A, A bar-plot showing the total number of communications detected by MEBOCOST in BAT scRNA-seq data. The x-axis was the four conditions of BAT scRNA-seq data, including TN (30 °C for a week), RT (room temperature), Cold2(4 °C for 2 days), and Cold7 (4 °C for 7 days). The y-axis was the total number of significant communications. A significant communication was defined if FDR < 0.05 and the presence of metabolite and sensor in a cell type was at least in 25% of the cell population. **B**, A heatmap to show the estimated metabolite presence. Columns were cell types in four conditions of BAT scRNA-seq data, rows were metabolites. The conditions and cell types were colored on the top of the heatmap. **C**, A heatmap to show the gene expression level of metabolite sensors. Columns were cell types in four conditions of BAT scRNA-seq data, rows were sensors. The conditions and cell types were colored on the top of the heatmap. **D**, Metabolite-sensor communications were shown due to their high sensitivity to cold exposure. Each row represented one cell-cell metabolite-sensor communication. The communication was defined as sender-metabolite-sensor-receiver. The columns in the main heatmap were four conditions of scRNA-seq data. The sender and receiver cell types were colored at the left side of the main heatmap. The black box with dashed lines labeled the Cold7 specifically enhanced communications, while the purple dashed box labeled enhanced communications for both Cold2 and Cold7.

Next, to identify the temperature-sensitive metabolite-sensor communications, all cells from the four conditions were pooled for MEBOCOST analysis and then grouped by cell types and conditions. In such strategy, the communication score and *p* value of same metabolite-sensor partners across conditions were comparable, because same population of cells were used for the *null* distribution estimation. With the detected communications from the four conditions, index of dispersion (IOD) of the communication scores was calculated to characterize the sensitivity of communications to cold exposure. 100 temperature-sensitive communications were identified as with communication scores highly variable across conditions (**Figure 5D**). Interestingly, we observed that some of the communications showed increased communication score under any of the two types of cold exposure including both Cold2 and Cold7, while some other communications were strongly and specifically increased by chronic cold exposure

in Cold7. Among the Cold7-specific increased communications, NMSC was a major sender compared to other cell types, indicating that chronic cold exposure may increase the secretion of signaling metabolites from NMSC and increase its effects on other adipose-resident cells. Altogether, MEBOCOST enabled detection of condition-specific cell-cell metabolite-sensor communications in brown adipose tissue.

Discussion

Cell-cell communication is a fundamental mechanism that coordinates cellular activities in development and disease[6]. In addition to protein ligands, metabolites are another major type of signaling molecules to mediate cell-cell communications, which are studied extensively based on experimental approaches. However, few computational tools were appliable to study metabolite-based intercellular communications based on single cell RNA-seq data. Motivated by the fundamental importance of metabolic communications in development and diseases, we developed MEBOCOST was to fill up the technology gap to infer the intercellular metabolite-sensor communications using scRNA-seq. To this end, we first constructed a metabolite-sensor partner database, which covered sensor types including cell surface receptor, cell surface transporter, and nuclear receptor. We developed the first computational tool that perform thorough literature mining to collect reported metabolite-sensor partners from literature. The collected metabolite-sensor partners provided a new foundation of knowledge for investigating cell-cell metabolite-sensor communications. Having those partners, communication scores were calculated based on the expression of metabolic enzymes and sensors for each pair of cell types and each pair of metabolite-sensor to characterize the communication likelihood. By applying to a BAT scRNA-seq datasets, MEBOCOST successfully recaptured known metabolite-sensor communications and further uncovered new communications. We demonstrated that MEBOCOST is a versatile and easy-to-use Python-based package to perform cell-cell metabolic communication analysis. MEBOCOST enables researchers to analyze cell-cell metabolite-sensor communications by scRNA-seq data in numerous human and mouse samples.

As far as we known, MEBOCOST is the first computational algorithm for genome-scale systematic detection of cell-cell metabolite-sensor communications. Although the current version has already enabled the prediction of cell-cell communications for 440 metabolite-sensor partners, a further extension of metabolite-sensor partners will increase the capability of MEBOCOST. We plan to enlarge the number of metabolite-sensor partners through two aspects in the future. First, we will increase the number of extracellular metabolites by integrating more metabolite resources such as HMDB, Recon2, Metabolic Atlas, etc. Second, we plan to optimize the text-mining pipeline to improve the recognition of metabolite-sensor partners. For example, we will enable the pipeline to use synonymous of metabolites and

sensors to parse the metabolite-sensor partners from PubMed abstract. Also, due to the complicated biological mechanism of cellular communication, taking more information into consideration in communication prediction will increase the performance of the algorithm. Cell surface transporter or receptor for metabolites might function by a protein complex. For example, SLC7A11 is a cell surface transporter for Cysteine curated in our database, and SLC3A2 is an extra need for forming the SLC7A11 mediated metabolite-sensor communications[54]. Therefore, it might be useful to improve MEBOCOST by incorporating the protein complex information of sensors into the calculation of communication score. Finally, MEBOCOST infers potential metabolite-sensor communications using single cell transcriptomics data without considering the spatial proximity of the cells. Although a robust performance has been observed in the current MEBOCOST, we believe that it will have the potential to provide a more comprehensive view of metabolite-sensor communications by combining spatial distribution of the cells into consideration.

Methods and Materials

Software and package version

Cellranger 6.1.2, scFEA 1.1.2, COMPASS (https://github.com/YosefLab/Compass), Python 3.8, and python packages including pandas 1.4.1, scipy 1.8.0, scanpy 1.8.2, matplotlib 3.5.1, seaborn 0.11.2, adjustText 0.7.3, network 2.7, jupyter 1.0.0.

Collection of extracellular metabolites and related enzyme genes

We were motivated to focus on metabolites that are predictable from gene expression data. scFEA and COMPASS are two software for metabolic flux-balance analysis in scRNA-seq data. The metabolites involved in the two software were focused on. Specifically, the metabolite names and their annotation information in scFEA were downloaded from https://github.com/changwn/scFEA/blob/master/data/Human_M168_information.symbols.csv .

The metabolite name and annotation in COMPASS were downloaded from https://github.com/YosefLab/Compass/blob/master/compass/Resources/Recon2_export/met_md.csv. We noticed that HMDB at https://hmdb.ca/ is a comprehensive database that include detailed annotation for 220,945 metabolites. Therefore, we decided to standardize the collected metabolite names from both scFEA and COMPASS into the annotation provided by HMDB. To do this, we focused on the metabolites for which the HMDB accession number was known. HMDB accessions of metabolites can be directly accessed in COMPASS, while only KEGG accessions were provided for scFEA metabolites. To map those scFEA metabolites into HMDB annotation, a parser script was developed to covert KEGG compound accession number

into HMDB accession number. Taking Acetyl-CoA as an example, C00024 is the KEGG compound accession number, and the KEGG page of C00024 was derived by link of https://www.genome.jp/entry/C00024. The related annotation in HMDB for Acetyl-CoA can be further derived by hyperlinks at “All links” session. In this specific case, the HMDB accession number for Acetyl-CoA can be collected from https://www.genome.jp/dbget-bin/get_linkdb?-t+hmdb+cpd:C00024. We applied the strategy for all the metabolites from scFEA to find their the HMDB accessions. For those metabolites that were successfully mapped to HMDB, biological location annotation, including cellular locations and biospecimen locations, were extracted. To focus on potential intercellular signaling metabolites, the metabolites in “extracellular space”, “blood”, or “cerebrospinal fluid” were retained and named as “extracellular metabolites”. Meanwhile, other basic annotation information for those metabolites were also collected, such as synonyms, metabolite class, and protein associations.

To collect metabolite enzymes, we focused on the extracellular metabolites collected by the procedures mentioned in above. A parser script was designed specifically for HMDB to collect the reaction and related enzymes of extracellular metabolites. Briefly, the webpage of metabolite in HMDB was visited based on the given HMDB accession number, and the annotation in “Enzymes” sessions was collected. Taking D-Lactic acid as an example, the annotation page of D-Lactic acid can be retrieved by https://hmdb.ca/metabolites/HMDB0001311 where HMDB0000171 is the HDMB accession number. In the “Enzymes” tab, the reactions such as “S-Lactoylglutathione + Water → Glutathione + D-Lactic acid” and the related gene name (e.g. HAGH) for the reactions were collected. Only complete reactions containing subtract metabolite names and product metabolite names were included in the collection. Such procedures produce a list of reactions as well as the corresponding enzyme gene names and metabolite names. The collection of reaction and related enzymes were used for the inference of metabolite presence.

Collection of metabolite-sensor partners

The detection of metabolite-based intercellular communications relied on the prior knowledge of metabolite-sensor partners. The sensor proteins in this study mainly include three types, namely, cell surface transporter, cell surface receptor, and nuclear receptors. To collect known pairs of metabolite-sensor partners, a workflow that combines computational text-mining and manual collection was designed. In the text-mining part, a list of metabolites and a list of sensor gene names were needed. Publicly reported pairs of metabolite and sensor will be recognized and curated at the manual collection step of the workflow. To initiate the text-mining, extracellular metabolites were focused on. Additionally, sensor gene names were collected by the following steps. For cell surface receptor, we obtained the list of receptor names mainly from two sources. Firstly, we downloaded a list of receptors from NicheNet[55], a

protein ligand-receptor communication analysis R package, by taking the “to” genes in file of https://zenodo.org/record/3260758/files/lr_network.rds. Secondly, we collected receptor genes from UniProt database[28] by searching for the keyword “receptor” through https://www.uniprot.org/uniprot/?query=name%3Areceptor+reviewed%3Ayes+organism%3A%22Homo+sapiens+%28Human%29+%5B9606%5D%22&sort=score. The receptors from the two sources were united and further filtered by focusing on cell membrane proteins based on the cellular location annotation in UniProt database. For cell surface transporters, we noticed that Transporter Classification Database (TCDB)[23-27] contains a comprehensive list of transporter proteins; thus, transporter gene names from TCDB were collected. For nuclear receptors, Nuclear Receptor Signaling Atlas (NURSA)[29] in dkNET project at https://dknet.org/data/source/nif-0000-03208-6/search was used for collecting known names of nuclear receptor names.

Having the names of metabolites and sensors, text-mining based on PubMed abstracts were performed. Firstly, 711,272 PubMed abstracts with unique PMID were obtained from https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=%28metabolism%5BTitle%2FAbstract%5D%29&retmax=15000000. We started from those abstracts as they were related to metabolism. Such strategy helped us to narrow down to a subset of abstracts, thus reduced the time of text-mining compared to going through all the abstracts in the PubMed. Secondly, the context of publication titles, MeSH words, and abstracts was downloaded for those PMID. Thirdly, potential combinations of metabolite and sensor names was checked in each of the collected context of publication titles, MeSH words, and abstracts. The PMID, metabolite names, and sensor names were recorded if the pair of metabolite name and sensor name were co-mentioned in the same sentence in the context of publication titles, MeSH words, and abstracts. Next, the collected pairs of metabolite-sensor partners together with the PMID and evidence in the context of publication title, MeSH words, or abstracts were subjected to a manual curation.

Besides the text-mining, we also manually collected metabolite-sensor partners from other well-known databases such as HMDB[15-18], Recon2[30], GPCRdb[32, 56], Wikipedia, and GeneCards[31]. For HMDB, we collected metabolite and transporter pairs if the metabolite associated protein were annotated as a transporter and the protein belongs to a type of cell membrane protein. For Recon2, since metabolites were further annotated by cellular localization such as [e] for extracellular and [c] for cytosol, so we specifically collected the reaction associated genes if the reaction happened by transporting the same metabolite from extracellular [e] into cytosol [c]. However, cell surface receptors and nuclear receptors were not included in both HMDB[15-18] and Recon2[30]. Therefore, pairs of metabolite and cell surface transporter partners were mainly focused on those from these two databases. We additionally collected metabolite and cell surface receptor partners from GPCRdb[32, 33] which is a database

for G protein-coupled receptors (GPCR) basic annotation and their ligands. The metabolite ligand of the GPCR protein were collected from the webpage at https://gpcrdb.org/ligand/statistics if the metabolite is an extracellular metabolite. Additionally, one pair of metabolite-sensor partner, pyruvic acid and SLC16A11, was collected from the Gene Card database by reading the description of *SLC1611* at https://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC16A11. Next, we manually collected metabolite and nuclear receptor partners from a Wikipedia page at https://en.wikipedia.org/wiki/Nuclear_receptor#Ligands. To this end, the preliminary automatic and manual collection of metabolite-sensor partners were completed. All the metabolite-sensor partners were then further manually curated by not less than three curators separately. Notably, the species and cell type were not restricted during the process of text-mining and manually collection, although the metabolite enzymes and sensor names were obtained mostly from databases for human. To generate a collection for mouse, the collected metabolite enzyme genes and metabolite sensor genes were match to mouse by homology gene pairs between human and mouse.

Infering metabolite presence in a cell based on gene expression of enzymes

Estimation of metabolite presence is a key step in MEBOCOST algorithm. Although several software, such as scFBA, scFEA, and COMPASS have been reported to predict metabolic fluxes by scRNA-seq data, none of them were specifically designed for estimating the metabolite distribution (**Supplementary Figure 1C**). Although the balance result in scFEA and uptake reaction result in COMPASS can be treated as metabolite abundance as a by-product of the prediction, those tools still have limitations to be fully integrated into MEBOCOST. For example, scFEA predicts only less than 100 metabolites, and COMPASS algorithm was computing intensive so that takes more than 1,000 seconds to run the prediction for a metabolite in 1,000 cells (**Supplementary Figure 1E**). Therefore, we hope to find a method that can predict the metabolite abundance in a cell by scRNA-seq data with less running time but without losing the accuracy compared to the existing tools.

Although we did not incorporate the running of scFEA and COMPASS in MEBOCOST running, the development of such tools showed the power of gene expression of metabolic enzymes on metabolism modeling. We reasoned that the level of a given metabolite in a cell should be dynamically influenced by two types of metabolism reactions. Some metabolism reactions take the metabolites as products; thus, the expression of the enzymes should correlate with the accumulation of the metabolites. Some other metabolism reactions may take the metabolites as substrates. The happening of such reactions will convert the given metabolites into other ones, so the enzyme gene expression should correlate with the depletion of the metabolites of interest. Therefore, the relative presence of a given metabolite in a cell can be estimated by

the average expression of enzymes in reactions that take the metabolite as a product after subtracting the average expression of enzymes in reactions that take the metabolite as a substrate. The estimation of metabolite presence based on enzyme gene expression was formulated as:

where *M*is the estimated presence of a given metabolite in a cell. *P* is a set of enzyme genes that associated with the accumulation of the metabolite. *i* is the i^{th} enzyme gene in the *P.* I_p is the total number of enzyme genes in the *P*. In addition, *S* is a set of enzyme genes that associated with the depletion of the metabolite. *j* is the j^{th} enzyme gene in the *S.* I_s is the total number of enzyme genes in the *S.* x_i and x_j are the expression level of enzyme *i*and *j* in a cell. MEBOCOST applied this formula to calculate the relative presence of each metabolite in each cell by scRNA-seq data.

To test the proposed method together with scFEA and COMPASS, we downloaded matched metabolomics and RNA-seq data of 928 cancer cell lines from CCLE project[20, 21]. The metabolomics data was downloaded by https://depmap.org/portal/download/api/download?file_name=ccle%2Fccle_2019%2FCCLE_metabolomics_20190502.csv&bucket=depmap-external-downloads. The RNA-seq data was downloaded from DepMap data portal at https://depmap.org/portal/download/ (file named CCLE_expression.csv). Next, we applied scFEA, COMPASS, and our method on the expression matrix of the RNA-seq data to estimate the metabolite level in each cell line. In addition to taking the average (arithmetic mean) of the enzyme expression, we also calculated the geometric mean of the enzyme gene expression in the estimation as a comparison. To evaluate the similarity between the estimation and the real detection of metabolomics, the Jaccard index between the results of estimation and detection was calculated for each metabolite. To calculate the Jaccard index, metabolite abundant cell lines were obtained for each estimation method as well as the real metabolomics data. Among the 910 cell lines, a metabolite abundant cell line was counted if the value of a given metabolite above the mean value of all the cell lines. For each metabolite, the metabolite abundant cell lines from estimation data can be donated as A, and the metabolite abundant cell lines from metabolomics was donated as B. The Jaccard index score *J* was calculated by

Interestingly, the Jaccard index value of methods by calculating mean expression of enzyme genes showed quiet compatible result with scFEA and slightly better result than COMPASS (**Supplementary Figure 1D**). However, the method of taking arithmetic mean of enzyme gene expression saved many time for the computing (**Supplementary Figure 1E**).

Calculation of metabolic communication scores

Having the estimation of metabolite presence from gene expression data and the prior knowledge of metabolite-sensor partners, MEBOCOST computed the communication score for each pair of cell types and for each pair of metabolite-sensor partners. Given a pair of cell types *i* and *j*, and given a pair of metabolite-sensor partners which metabolite was donated as *m*, and sensor was donated as *s*. We donated as the mean metabolite abundance in cell type *i*, and as the mean expression level of sensor gene in cell type *j*. Then, the communication score S_c was computed as the following:

The communication score was computed for cell types in pairwise and for each pair of metabolite-sensor partners, so that the MEBOCOST give equal change for each cell type as a sender or as a receiver in all calculations. To evaluate the statistical significance, we performed a shuffling on cell labels of the scRNA-seq data by default 1,000 times. The users are allowed to change the time of cell label shuffling. For each cell type pair and each metabolite-sensor partner, the same method was applied to calculate the communication score for the shuffled scRNA-seq data. This procedure generated 1,000 communication scores as a statistical null distribution for each cell type pair and each metabolite-sensor partner. Based on the *null* distribution, *p* value was computed by permutation testing[57]. All the *p* values in the datasets were subjected to a false discovery rate (FDR) correction by Benjamini-Hochberg procedure[22]. We reasoned that the baseline level of metabolites and sensors could be vary a lot. To reduce the bias of communication scores from the baseline, the real communication score was further normalized by the mean of 1,000 communication scores calculated from shuffled data. Furthermore, to increase the statistical power of the prediction, we focused on the highly abundant metabolites and highly expressed sensors in the cell type. By default, the metabolite and sensor were considered as informative if they were both active in at least 25% of the cell population of the cell type, else the *p* value and FDR will be converted to 1.

Single cell RNA-seq data processing

A single cell RNA-seq data of mouse brown adipose tissue was used in the current study. The dataset was generated by 10X Genomics platform. Raw data was deposited to NCBI GEO database under accession number GSE160585. Cellranger count was applied to map the raw sequence to mouse reference genome (mm10) and obtain read count over each gene, and r1-length parameter was set to 26, while other parameters were set by default. Next, the data processing, including data normalization, dimension reduction, clustering, and visualization of gene expression were performed using Scanpy[58] in Python. Cells were filtered to have at least 800 UMIs and 400 detected genes. Genes were filtered to be at least detected in 10 cells. To reduce the doublet effect, cells were removed if the total UMIs greater than 50,000 or the number of detected genes greater than 7500. The number of nearest neighbors was set to 10

in Scanpy find neighbor function. The top 40 principal components were included in clustering and UMAP analysis[59]. The visualization of the clusters was performed by the UMAP method. Cell annotation was done based on the cell type marker genes collected from PanglaoDB[60] at https://panglaodb.se/markers.html.

Evaluation of running time and memory usage in peak

We tested MEBOCOST on the BAT scRNA-seq data from mouse housed at cold temperature for 2 days (Cold2)[61]. The dataset contains 33,470 cells and 20 cell types including adipocytes. To evaluate the usage of computing resource of our algorithm on different cell numbers, 90%, 80%, 70%, 60%, 50%, 40%, 30% of the total number of cells were subsampled. Next, MEBOCOST was applied on the original Cold2 dataset as well as the subsampling dataset. All the jobs were run at same computing server with 8 cores. The running time were recorded by python “time” module, and the peak memory usage were also recorded by the python “tracemalloc” module. The running time in minutes and the peak memory usage in Gb were showed by bar plot using Python “matplotlib” package.

Stability evaluation

The prediction result of cell-cell metabolic communications in Cold2 original dataset and subsampling datasets were further used to evaluate the stability of MEBOCOST. We reasoned that a good algorithm should be less influenced by the sequencing depth and the total cell number of scRNA-seq. Therefore, we compared the prediction results of subsampling datasets with the result of original dataset in two aspects. First, we compared the total number of significant communication events between subsampling datasets and original datasets. The communications were deemed significant at a False Discovery Rate (FDR) of 0.05 and 25% of metabolite and sensor active cells in the cell type. Second, the similarity score between the prediction result of subsampling datasets and the original dataset was calculated. The similarity score S was computed based on the overlap of communications via the Jaccard similarity defined by:

where the V and V' are two sets of intercellular communication events predicted from original dataset and subsampling dataset, respectively.

Identification of most cold-sensitive communication events

All the cells of BAT scRNA-seq data from TN, RT, Cold2, and Cold7 were pooled and grouped by cell types and conditions when running MEBOCOST. Communications were deemed as significant if FDR less than 0.05 and the fraction of metabolite and sensor active cells greater than

0.25. Index of dispersion (IOD) was calculated using communication score across four conditions. Given one sender-metabolite-sensor-receiver communication, the IOD was defined by:

where σ^2 is the variance of the four communication score, μ is the mean of the four communication score. Sender-metabolite-sensor-receiver communications included in this cold-sensitive communication analysis were predicted as significant at least in one of the four conditions. The IOD scores of communications were ranked from high to low, and 100 top-ranked were selected as cold-sensitive communications.

Code availability and data availability

All the analysis including MEBOCOST were implemented by Python. The collected enzyme and sensor genes of extracellular metabolites, source code of MEBOCOST, and the detailed instruction of usage were available at https://github.com/zhengrongbin/MEBOCOST. The singe cell RNA-seq data of mouse brown adipose tissue was deposited to NCBI GEO under accession GSE160585.

Author Contributions

K. C. and Y. H. T. conceived the project. R. Z. developed the algorithm, curated the metabolite-sensor partners, performed the computational analysis. Y. Z. and T. T. curated the metabolite-sensor partners. K.C., Y.H.T., R.Z., and L.Z analyzed and interpreted the results. R. Z. and K. C. drafted and edited the manuscript.

Acknowledgements

This project is supported in part by R01GM125632 (Kaifu Chen), 1R01HL148338 (Kaifu Chen), R01DK132469 (Yu-Hua Tseng), and R01HL155632 (Lili Zhang).

Abbreviations

EC

Endothelial cells

CD8T

$CD8^+$ T lymphocytes

Activated CD8T

Activated CD8 T lymphocytes

CD4T

$CD4^+$ T lymphocytes

Erythroid-like

Erythroid-like and erythroid precursor cells

NK

Natural killer cells

MSC

Myelinating Schwann cells

NMSC

Non-myelinating Schwann cells

Treg

Regulatory T cells

VSM

Vascular smooth muscle cells

ILC2s

Type 2 innate lymphoid cells

BAT

Brown adipose tissue

TN

Thermoneutral

RT

Room temperature

Cold2

Cold temperature for 2 days

Cold7

Cold temperature for 2 days

References

1.Song, D., Yang, D., Powell, C.A. & Wang, X. Cell-cell communication: old mystery and new opportunity. *Cell Biol Toxicol* **35**, 89–93 (2019).

2.Lafontan, M. Fat cells: afferent and efferent messages define new approaches to treat obesity. *Annu Rev Pharmacol Toxicol* **45**, 119–146 (2005).

3.Roy, S., Kim, D. & Lim, R. Cell-cell communication in diabetic retinopathy. *Vision Res* **139**, 115–122 (2017).

4.Tirziu, D., Giordano, F.J. & Simons, M. Cell communications in the heart. *Circulation* **122**, 928–937 (2010).

5.AlMusawi, S., Ahmed, M. & Nateri, A.S. Understanding cell-cell communication and signaling in the colorectal cancer microenvironment. *Clin Transl Med* **11**, e308 (2021).

6.Armingol, E., Officer, A., Harismendy, O. & Lewis, N.E. Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* **22**, 71–88 (2021).

7.Martins Conde Pdo, R., Sauter, T. & Pfau, T. Constraint Based Modeling Going Multicellular. *Front Mol Biosci* **3**, 3 (2016).

8.Dias, A.S., Almeida, C.R., Helguero, L.A. & Duarte, I.F. Metabolic crosstalk in the breast cancer microenvironment. *Eur J Cancer* **121**, 154–171 (2019).

9.Wagner, A. et al. Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. *Cell* **184**, 4168–4185 e4121 (2021).

10.Alghamdi, N. et al. A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. *Genome Res* **31**, 1867–1884 (2021).

11.Damiani, C. et al. Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. *PLoS Comput Biol* **15**, e1006733 (2019).

12.Wang, Y.P. & Lei, Q.Y. Metabolite sensing and signaling in cell metabolism. *Signal Transduct Target Ther* **3**, 30 (2018).

13.Liu, S., Alexander, R.K. & Lee, C.H. Lipid metabolites as metabolic messengers in inter-organ communication. *Trends Endocrinol Metab* **25**, 356–363 (2014).

14.Monelli, E. et al. Angiocrine polyamine production regulates adiposity. *Nat Metab* (2022).

15.Wishart, D.S. et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* **46**, D608–D617 (2018).

16.Wishart, D.S. et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res* **50**, D622–D631 (2022).

17.Wishart, D.S. et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **41**, D801–807 (2013).

18.Wishart, D.S. et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res* **35**, D521–526 (2007).

19.Tonn, M.K., Thomas, P., Barahona, M. & Oyarzun, D.A. Computation of Single-Cell Metabolite Distributions Using Mixture Models. *Front Cell Dev Biol* **8**, 614832 (2020).

20.Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

21.Li, H. et al. The landscape of cancer cell line metabolism. *Nat Med* **25**, 850–860 (2019).

22.Hochberg, Y.B.a.Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc* **57**, 289–300 (1995).

23.Saier, M.H. et al. The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res* **49**, D461–D467 (2021).

24.Saier, M.H., Jr.., Reddy, V.S., Tamang, D.G. & Vastermark, A. The transporter classification database. *Nucleic Acids Res* **42**, D251–258 (2014).

25.Saier, M.H., Jr.. et al. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* **44**, D372–379 (2016).

26.Saier, M.H., Jr.., Tran, C.V. & Barabote, R.D. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**, D181–186 (2006).

27.Saier, M.H., Jr.., Yen, M.R., Noto, K., Tamang, D.G. & Elkan, C. The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**, D274–278 (2009).

28.UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).

29.Becnel, L.B. et al. Nuclear Receptor Signaling Atlas: Opening Access to the Biology of Nuclear Receptor Signaling Pathways. *PLoS One* **10**, e0135615 (2015).

30.Thiele, I. et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol* **31**, 419–425 (2013).

31.Safran, M. et al. GeneCards Version 3: the human gene integrator. *Database (Oxford)* **2010**, baq020 (2010).

32.Kooistra, A.J. et al. GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res* **49**, D335–D343 (2021).

33.Pandy-Szekeres, G. et al. The G protein database, GproteinDb. *Nucleic Acids Res* **50**, D518–D525 (2022).

34.Lee, P. et al. Temperature-acclimated brown adipose tissue modulates insulin sensitivity in humans. *Diabetes* **63**, 3686–3698 (2014).

35.Peres Valgas da Silva, C., Hernandez-Saavedra, D., White, J.D. & Stanford, K.I. Cold and Exercise: Therapeutic Tools to Activate Brown Adipose Tissue and Combat Obesity. *Biology (Basel)* **8** (2019).

36.van Marken Lichtenbelt, W.D. et al. Cold-activated brown adipose tissue in healthy men. *N Engl J Med* **360**, 1500–1508 (2009).

37.McMillan, A.C. & White, M.D. Induction of thermogenesis in brown and beige adipose tissues: molecular markers, mild cold exposure and novel therapies. *Curr Opin Endocrinol Diabetes Obes* **22**, 347–352 (2015).

38.Whitehead, A. et al. Brown and beige adipose tissue regulate systemic metabolism through a metabolite interorgan signaling axis. *Nat Commun* **12**, 1905 (2021).

39.Singh, R., Braga, M. & Pervin, S. Regulation of brown adipocyte metabolism by myostatin/follistatin signaling. *Front Cell Dev Biol* **2**, 60 (2014).

40.Leiria, L.O. et al. 12-Lipoxygenase Regulates Cold Adaptation and Glucose Metabolism by Producing the Omega-3 Lipid 12-HEPE from Brown Fat. *Cell Metab* **30**, 768–783 e767 (2019).

41.Kim, J. et al. Eicosapentaenoic Acid Potentiates Brown Thermogenesis through FFAR4-dependent Up-regulation of miR-30b and miR-378. *J Biol Chem* **291**, 20551–20562 (2016).

42.Herz, C.T. & Kiefer, F.W. The Transcriptional Role of Vitamin A and the Retinoid Axis in Brown Fat Function. *Front Endocrinol (Lausanne)* **11**, 608 (2020).

43.Fenzl, A. et al. Intact vitamin A transport is critical for cold-mediated adipose tissue browning and thermogenesis. *Mol Metab* **42**, 101088 (2020).

44.Yoo, H., Antoniewicz, M.R., Stephanopoulos, G. & Kelleher, J.K. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. *J Biol Chem* **283**, 20621–20627 (2008).

45.Okamatsu-Ogura, Y. et al. UCP1-dependent and UCP1-independent metabolic changes induced by acute cold exposure in brown adipose tissue of mice. *Metabolism* **113**, 154396 (2020).

46.Shamsi, F., Wang, C.H. & Tseng, Y.H. The evolving view of thermogenic adipocytes - ontogeny, niche and function. *Nat Rev Endocrinol* **17**, 726–744 (2021).

47.Hammel, J.H. & Bellas, E. Endothelial cell crosstalk improves browning but hinders white adipocyte maturation in 3D engineered adipose tissue. *Integr Biol (Camb)* **12**, 81–89 (2020).

48.Huang, L. et al. SR-B1 drives endothelial cell LDL transcytosis via DOCK4 to promote atherosclerosis. *Nature* **569**, 565–569 (2019).

49.Zhang, Y. et al. Shear stress regulates endothelial cell function through SRB1-eNOS signaling pathway. *Cardiovasc Ther* **34**, 308–313 (2016).

50.Vaisman, B.L. et al. Endothelial Expression of Scavenger Receptor Class B, Type I Protects against Development of Atherosclerosis in Mice. *Biomed Res Int* **2015**, 607120 (2015).

51.Yeh, Y.C., Hwang, G.Y., Liu, I.P. & Yang, V.C. Identification and expression of scavenger receptor SR-BI in endothelial cells and smooth muscle cells of rat aorta in vitro and in vivo. *Atherosclerosis* **161**, 95–103 (2002).

52.Bartelt, A. et al. Thermogenic adipocytes promote HDL turnover and reverse cholesterol transport. *Nat Commun* **8**, 15010 (2017).

53.Yvan-Charvet, L. et al. In vivo evidence for a role of adipose tissue SR-BI in the nutritional and hormonal regulation of adiposity and cholesterol homeostasis. *Arterioscler Thromb Vasc Biol* **27**, 1340–1345 (2007).

54.Koppula, P., Zhuang, L. & Gan, B. Cystine transporter SLC7A11/xCT in cancer: ferroptosis, nutrient dependency, and cancer therapy. *Protein Cell* **12**, 599–620 (2021).

55.Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**, 159–162 (2020).

56.Husted, A.S., Trauelsen, M., Rudenko, O., Hjorth, S.A. & Schwartz, T.W. GPCR-Mediated Signaling of Metabolites. *Cell Metab* **25**, 777–796 (2017).

57.Moore, J.H. Bootstrapping, permutation testing and the method of surrogate data. *Phys Med Biol* **44**, L11–12 (1999).

58.Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).

59.Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* (2018).

60.Franzen, O., Gan, L.M. & Bjorkegren, J.L.M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019** (2019).

61.Shamsi, F. et al. Vascular smooth muscle-derived Trpv1(+) progenitors are a source of cold-induced thermogenic adipocytes. *Nat Metab* **3**, 485–495 (2021).