Supporting Information

The Sweat Metabolome of Screen-Positive Cystic Fibrosis Infants: Revealing Mechanisms beyond Impaired Chloride Transport

Adriana N. Macedo,1 Stellena Mathiaparanam,1 Lauren Brick,2 Katherine Keenan,3 Tanja Gonska,3,4 Linda Pedder,2 Stephen Hill,5 and Philip Britz-McKibbin1,*

1 Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Ontario, L8S 4L8, Canada
2 Department of Pediatrics, McMaster University, Hamilton, ON, L8S 3Z5, Canada
3 Program in Translational Medicine, The Hospital for Sick Children, Toronto, ON, M5G 1X8, Canada
4 Department of Pediatrics, University of Toronto, Ontario, M5G 1E2, Canada
5 Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, L8S 3Z5, Canada

*Corresponding author: Philip Britz-McKibbin, britz@mcmaster.ca

Methods, Tables S1-S7, Figures S1-S6
METHODS

Chemicals and Reagents. Ultra-grade LC-MS solvents (water, methanol and acetonitrile) obtained from Caledon Laboratories Ltd. (Georgetown, ON, Canada) were used to prepare all buffer and sheath liquid solutions. All other chemicals were obtained from Sigma-Aldrich Inc. (St. Louis, MO, USA), except for malic acid, xanthine, uric acid, and citric acid, which were purchased from Alfa Aesar Inc. (Heysham, Lancashire, UK), and pilocarpic acid from Toronto Research Chemicals Inc. (Toronto, ON, Canada).

Study Design and Sweat Collection for CF Screen-Positive Infants. In this study, residual sweat samples were obtained from first-time/screen-positive CF infants following standardized pilocapine-stimulated sweat chloride testing, excluding samples with intermediate chloride levels (30-59 mmol/L). In most cases, a 5 µL aliquot was typically used for sweat metabolomic studies if available. All sweat specimens were de-identified and relabeled with a research ID at McMaster Children’s Hospital and The Hospital for Sick Children (Sick Kids) in order to protect patient privacy, as approved by the Hamilton Integrated Research Ethics Board and The Hospital for Sick Children’s Research Ethics Board. Patient medical records from Newborn Screening Ontario (NSO) and CF clinics were linked directly to the research ID, maintaining confidentiality of personal health information. Sweat collection was performed using a Webster Model 3700 Macroduct Sweat Capillary Collection System (Wescor Inc., Logan, UT, USA). Electrodes and agar gel discs containing 0.5% w/w pilocarpine nitrate were placed in the infant’s forearm for sweat stimulation after carefully cleaning the region with ethanol and distilled water. A low current (1.5 mA) was applied for 5 minutes to deliver pilocarpine to the sweat gland via iontophoresis. Sweat was then collected for 25 minutes using a coiled microbore tube containing a blue dye to facilitate visualization of the sample volume. Samples were transferred to a 0.5 mL centrifuge tube and stored in a fridge (+4°C) after taking an aliquot for chloride analysis by chloridometer (Wescor Sweat-Check Model 3120). Residual sweat samples were subsequently stored at -80°C prior to thawing when performing metabolomic studies. Overall, sixty-eight sweat specimens from screen-positive infants (≤ 3 months old) using a two-tiered algorithm based on elevated IRT/genetic mutation panel at NSO were analyzed in this study, including 50 CF unaffected infants (31 from McMaster, 19 from Sick Kids) and 18 CF affected (13 from McMaster, 5 from Sick Kids).
**Instrumentation and Sweat Analysis by MSI-CE-MS.**  Nontargeted metabolite profiling was performed using an Agilent 7100 capillary electrophoresis (CE) instrument and an Agilent 6550 quadrupole time-of-flight (QTOF) mass spectrometer equipped with a dual jet stream electrospray ion source (dual AJS ESI) with ion funnel technology (Agilent Technologies, Mississauga, ON, USA). Sweat samples were analyzed by MSI-CE-MS in positive (ESI+) and negative (ESI-) ion modes to cover cationic and anionic metabolites after thawing specimens slowly to room temperature on ice. System control and data acquisition were performed using the Mass Hunter Workstation Software (Data Acquisition, version B.07.01, Agilent Technologies, 2014). The CE separation was carried out using uncoated fused-silica capillaries (Polymicro Technologies, Phoenix, AZ, USA) with 50 µm inner diameter and 120 cm total length. The background electrolyte (BGE) used for cationic metabolites was composed of 1 M formic acid and 15% v/v acetonitrile, pH 1.8 [1], whereas the BGE for anionic species consisted of 50 mmol/L ammonium bicarbonate, pH 8.5 (adjusted with 10% v/v ammonium hydroxide) [2]. New capillaries were conditioned by flushing with methanol, water, 1 M sodium hydroxide, water, and formic acid as BGE (15 minutes each). For ESI-, this conditioning was followed by 15 minutes flush with water and 30 minutes flush with ammonium bicarbonate as BGE. The capillary was flushed with BGE for 15 minutes before each separation, followed by the MSI injection sequence in which seven samples were injected hydrodynamically (5 seconds, 50 mbar) and interspaced with BGE (40 seconds, 50 mbar), ending with BGE (5 seconds, 50 mbar). Separations were performed using normal polarity with an applied voltage of 30 kV and capillary maintained at 25°C. Pressure-assisted separation was used to speed up the analysis of anionic metabolites under high pH and normal polarity by applying a pressure gradient of 1 mbar/min from 10 to 20 minutes and 3 mbar/min from 20 to 38 minutes. A standard mixture, and a QC run with blank (i.e., internal standard solution in water) were analyzed at the beginning of each day to equilibrate the CE-MS system and ensure high data quality (i.e., sample carry-over effects, good reproducibility, ion spray and current stability etc.) prior to the randomized analysis of pairs of sweat samples with QC (i.e., the same pooled sweat was used consistently throughout study with a frozen aliquot thawed each day) when using MSI-CE-MS. For overnight storage, capillaries were flushed with water for 10 minutes and air dried for 10 minutes. As preventative maintenance, the CE electrode was cleaned daily with 50% v/v isopropanol-water and methanol.
The Q-TOF system was calibrated every morning before analysis using an Agilent tune mixture over a *m/z* 50-1700 range. An Agilent 1260 Infinity series isocratic pump equipped a 100:1 splitter was used to deliver sheath liquid to the CE-MS interface at a flow rate of 10 µL/min. The sheath liquid for ESI+ consisted of 60% *v/v* methanol in water with 0.1% *v/v* formic acid, whereas for ESI- it was 50% *v/v* methanol in water. Purine and HP-0921 (API-TOF Reference Mass Solution Kit, Agilent Technologies) were added to the sheath liquid as reference masses for real-time automatic mass calibration. The nebulizer gas kept at 8 psi, while the flow rate for the drying gas was maintained at 16 L/min (300°C), and the sheath gas flow was kept at 3.5 L/min (200°C). The capillary voltage, nozzle voltage, and fragmentor voltage were 2000 V, 2000 V and 380 V, respectively. The instrument was kept in extended dynamic range (EDR, 2 GHz) for ESI+ experiments to prevent saturation of highly abundant cationic metabolites, whereas ESI- experiments were performed in high resolution (HiRes 4 GHz) to improve detection of anionic metabolites that in general have lower sensitivity. As preventative maintenance, the ion source was wiped daily with lint-free cloth and 50% *v/v* isopropanol-water.

All samples were prepared by a simple dilution step (typically 4-fold in in ultra-grade LC-MS water) using a volume of 5 µL (unless otherwise stated) of residual infant sweat specimen from screen-positive CF infants containing 10 µM of the internal standards, 3-chloro-L-tyrosine (Cl-Tyr) and sodium 2-naphthalenesulfonate (NMS), followed by mixing using a vortex for 30 seconds. A pooled quality control (QC) sample (*n*=10) was prepared by combining equal aliquots of sweat samples collected from 10 children, including five confirmed CF cases and five CF unaffected individuals. Those samples were not individually analyzed in this study for exceeding the age criteria (≤ 3 months old) or for lacking clinical information from NBS (e.g., siblings of recently diagnosed CF infant, CF affected children who were born before implementation of NBS for CF). In this case, the QC was not prepared from the actual study samples due to limited volumes remaining in some infant’s sweat specimens after chloridometer analysis.

Metabolomic studies were initially performed using a dilution trend filter in MSI-CE-MS [1, 3] with a QC sample for rigorous peak selection/data filtering of cationic and anionic metabolites derived from infant sweat in ESI+ and ESI-, respectively (Figure S1). This process allows for unambiguous identification of authentic yet reproducible molecular features (*i.e.*, [M+H]+ and
[M-H]-) from background or spurious signals based on their distinctive dilution signal pattern that has adequate precision (RSD < 40%, n=3) with no measurable signal in the blank. Using this approach, a total of 52 spurious/background signals were removed from over 206 features originally obtained from Mass Hunter Molecular Feature Extractor (MFE). A list of authentic features defined by their characteristic mass-to-charge ratio and relative migration time (m/z:RMT) was then further filtered to remove redundant signals from the same compound which have the same RMT as the ionized molecule, including adducts, dimers, isotopes and in-source fragments. This data filtering procedure was confirmed with authentic standards whenever available as summarized in Table S1 based on confidence levels for metabolite identification (i.e., level 1 indicative of confirmation of unknown ion with authentic standard) based on co-migration and MS/MS match of product ion spectra at different collisional energies as recommended by the Metabolomics Standards Initiative [4]. MFE was also performed in sweat samples from four authentic CF and two unaffected CF infants in order to identify unique molecular features (m/z:RMT) that were not detected in the original pooled QC.

Individual sweat samples were analyzed in duplicate using temporal signal pattern recognition in MSI-CE-MS for confident assignment of peaks that also confirms their sample of origin even when a compound is not detected in one or more samples in the serial injection sequence [5]. As shown in Figure 1, three sweat samples were prepared in duplicate using different dilution factors, which results in a pattern of ion signals as a function of time that allows for simple visual peak assignment. Furthermore, this configuration allows the introduction of a pooled QC within each run, which is critical for evaluating long-term system variability/technical variation during data acquisition. Sweat samples were analyzed in randomly assigned order and injection positions by MSI-CE-MS with a QC randomly inserted in positions 1, 3, 5 or 7. Sample dilution was optimized for each ionization mode in order to prevent saturation and signal suppression due to certain high abundant sweat components. In this case, sweat samples were 10- and 20-fold diluted for cation analysis and 4 and 8-fold diluted for anion analysis. Sweat samples were analyzed in two separate batches of runs performed about 10 months apart due to delays for acquisition of sweat specimens from an adequate number of screen-positive CF infants given the incidence rate of CF in the population (1:3600) as compared to total carriers (1:30), however only 39 CFTR mutations/3 variants are included within screening panel for CF in Ontario [6].
**Data Processing and Statistical Analysis.** Raw data was extracted using Mass Hunter Workstation Software (Qualitative Analysis, version B.06.00, Agilent Technologies, 2012). Untargeted analysis was performed using MFE and Molecular Formula Generator tools followed by a personal compound database search. Features were extracted in centroid using 10 ppm mass window and labeled according to their m/z, RMT and ionization mode (p = ESI+, n = ESI-). Peaks were integrated after smoothing (quadratic/cubic Savitzky-Golay, 15 points) and peak areas and migration times were transferred to Excel (Microsoft Office) for calculation of relative integrated peak area (RPA), relative migration time (RMT), relative standard deviation (RSD) and average fold-changes (FC), as well as conversion into a data matrix format. Only molecular features detected in at least 75% of the samples were included in the data matrix for subsequent data transformation and statistical analysis. Due to between-subject variations in sweat rate/volume among infants, a probabilistic quotient normalization (PQN) method was evaluated for normalizing sweat metabolomic data to correct for dilution/hydration status [7]. Briefly, the most probable dilution factor determined by PQN was derived from the ratio of the median response for ions measured in a sample relative to pooled sweat sample (QC) used as a reference. Additionally, batch effect correction was performed via an empirical Bayes method (ComBat) on MetaboAnalyst 3.0 [8]. Other batch correction algorithms were also evaluated based on QC s or study samples [9]. Although the top-ranked metabolites remained consistent when using different batch adjustment methods, ComBat was adopted for providing better overlap between batches, as indicated by a principal component analysis (PCA) 2D scores plots (data not presented).

Supervised multivariate data analysis was also performed on MetaboAnalyst 3.0 using partial least-squares-discriminate analysis (PLS-DA) in order to identify top-ranked sweat metabolites associated with CF disease status in affected infants based on variable importance in projection (VIP) scores on PC1. The PLS-DA model was validated using leave-out-one-at-at-time cross-validation in conjunction with permutation testing. Normality tests, Spearman rank correlation and nonparametric statistical analysis (Mann-Whitney U test) were performed using the Statistical Package for the Social Science (SPSS, version 18), and MedCalc (MedCalc Inc.) was used for ROC curves and boxplots.

**Structural Elucidation of Unknown Sweat Metabolites by CE-MS/MS.** MS/MS experiments were performed using a single long injection in CE-MS/MS to enhance signal response when
detecting fragments for low abundance sweat metabolites, with 90 seconds injection (50 mbar) for cations prepared in 200 mmol/L ammonium acetate buffer, pH 5.0, and 50 seconds injection (50 mbar) for anions prepared in water. Data was acquired using the targeted MS/MS function on Mass Hunter, using the EDR 2 GHz mode and collision-induced dissociation experiments with cycling collision energies of 10, 20 and 40 V. MS/MS spectra were extracted using the Find by Targeted MS/MS function, generating an individual spectrum per collision energy, except in the case of very low abundance compounds (not detected by Find by Targeted MS/MS), which were manually extracted as an average fragment spectrum for 10, 20 and 40 V. Agilent Mass Hunter Molecular Structure Correlator (MSC) was used for assignment of molecular formula candidates, with subsequent identification of candidate molecular structures from ChemSpider, Human Metabolome Database and an internal Agilent Metlin Personal Metabolite Database/Library. MS and MS/MS search were also performed using open-source databases (HMDB and Metlin) in combination with *in silico* fragmentation using CFM-ID [10] for putative candidates with consistent electromigration behavior under CE separation conditions. Additionally, manual MS/MS assignment of unknown metabolites based on common neutral losses was also performed to identify putative metabolite classes. Compound identities were confirmed by comparing RMT and MS/MS spectra with authentic chemical standards whenever available. Unknown yet low abundance metabolites that prevented acquisition of MS/MS spectra and did not have eligible structures within public metabolomic databases were annotated based on their accurate mass (*m/z*), ionization mode, most likely molecular formula and RMT. Mass accuracy for most detectable sweat metabolites were under 5.0 ppm (average of 2 ppm) unless otherwise noted as summarized in Table S1.

**Pilocarpine Gel Extracts and Blanks for Sweat Collection Device.** A small piece of one new and two used gel discs (≈ 0.2 g) was transferred to a 1.5 mL centrifuge tube, sonicated for 10 min with 1 mL of water and centrifuged for 10 min at 14,000 g to separate gel debris. Two aliquots of the supernatant were collected and diluted 20 and 40-fold in water containing the internal standards Cl-Tyr and NMS before analysis by MSI-CE-MS. A standard solution of pilocarpine was also extracted as a control, indicating a satisfactory mean recovery of 90% for pilocarpine. A blank for the sweat collection device was prepared to evaluate the presence background components in sweat specimens collected using this system. First, a 55 µL aliquot of
water was inserted in the central opening of the coiled microbore tube and collected in the opposite side of the tube in order to capture compounds potentially originated from the plastic material. Then, another 55 µL aliquot of water was placed in the back of the collection device, over the blue dye spot located next to the tube opening. In this case, the objective was to evaluate if any compounds identified in sweat samples were coming from the blue dye. The blanks for the collection tube and blue dye were diluted 4, 8, and 16-fold in water containing Cl-Tyr and NMS and analyzed by MSI-CE-MS in ESI+ and ESI-, respectively.

Supporting References

[1] Kuehnbaum, N. L.; Kormendi, A.; Britz-McKibbin, P. Multisegment Injection-Capillary Electrophoresis-Mass Spectrometry: A High-Throughput Platform for Metabolomics with High Data Fidelity. *Anal. Chem.* 2013, 85, 10664–10669.
[2] Yamamoto, M.; Ly, R.; Gill, B.; Zhu, Y.; Moran-Mirabal, J.; Britz-McKibbin, P. Robust and High Throughput Method for Anionic Metabolite Profiling: Preventing Polyimide Aminolysis and Capillary Breakages under Alkaline Conditions in Capillary Electrophoresis-Mass Spectrometry. *Anal. Chem.* 2016, 88, 10710-10719.
[3] Kuehnbaum, N. L.; Gillen J.B.; Kormendi, A.; Lam, K.P; DiBattista, A.; Gibala, M.J.; Britz-McKibbin, P. Multiplexed Separations for Biomarker Discovery in Metabolomics: Elucidating Adaptive Responses to Exercise Training. *Electrophoresis* 2015, 36, 2226-2236.
[4] Salek, R.M; Steinbeck, C.; Viant, M.R.; Goodacre, R.; Dunn, W.B. The Role of Reporting Standards for Metabolite Annotation and Identification in Metabolomic Studies. *Gigascience* 2013, 2, 13.
[5] DiBattista, A.; McIntosh, N.; Lamoureux, M.; Al-Dirbashi, O.Y.; Chakraborty, P.; Britz-McKibbin, P. Temporal Signal Pattern Recognition in Mass Spectrometry: A Method for Rapid Identification and Accurate Quantification of Biomarkers for Inborn Errors of Metabolism with Quality Assurance. *Anal. Chem.* 2017, DOI: 10.1021/acs.analchem.7b01727.
[6] https://www.newbornscreening.on.ca/en/disease/cystic-fibrosis-cf
[7] Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. *Anal. Chem.* 2006, 78, 4281-4290.
[8] Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0 – Making Metabolomics More Meaningful. *Nucleic Acids Res.* 2015, 43, W251-W257.
[9] Wehrens, R.; Hageman, J. A.; van Eeuwijk, F.; Kooke, R.; Flood, P. J.; Wijnker, E.; Keurentjes, J. J. B.; Lommen, A.; van Eekelen, H. D. L. M.; Hall, R. D.; Mumm, R.; de Vos, R. C. H. Improved Batch Correction in Untargeted MS-based Metabolomics. *Metabolomics* 2016, 12, 88.
[10] Allen, F; Greiner, R.; Wishart, D. Competitive Fragmentation Modeling of ESI-MS/MS Spectra for Putative Metabolite Identification. *Metabolomics*, 2015, 11, 98.
Table S1. Summary of 64 metabolites detected in sweat from screen-positive CF infants. Metabolites described in terms of m/z, RMT, ionization mode (p = ESI+, n = ESI-), molecular formula, mass error, compound ID or tentative ID, confidence level for identification, and chemical classification.

| m/z:RMT:mode  | Formula | Mass error (ppm) | Compound ID | Confidence | Classification |
|---------------|---------|------------------|-------------|------------|----------------|
| 76.0393:0.757:p | C₂H₆N₂O₂ | 17 | Glycine | 1 | Amino acid |
| 90.0550:0.805:p | C₂H₆N₂O₂ | 8.9 | Alanine | 1 | Amino acid |
| 104.1075:0.645:p | C₃H₁₁NO | 1.9 | Choline | 1 | Quaternary ammonium salt |
| 106.0499:0.874:p | C₂H₇NO₃ | 3.8 | Serine | 1 | Amino acid |
| 114.0662:0.687:p | C₄H₉NO₃ | 1.8 | Creatinine | 1 | Heterocyclic compound derivative |
| 116.0706:0.920:p | C₂H₉NO₂ | 0.9 | Proline | 1 | Amino acid |
| 118.0863:0.868:p | C₃H₁₁NO₂ | 1.7 | Valine | 1 | Amino acid |
| 120.0655:0.909:p | C₂H₇NO₃ | 2.5 | Threonine | 1 | Amino acid |
| 132.0655:1.009:p | C₃H₇NO₃ | 0.8 | 3-Hydroxyproline | 1 | Amino acid derivative |
| 132.0766:0.798:p | C₂H₁₀N₂O₂ | 2.3 | Creatine | 1 | Amino acid |
| 132.1019:0.879:p | C₃H₁₃NO₂ | 1.5 | Isoleucine | 1 | Amino acid |
| 132.1019:0.888:p | C₃H₁₃NO₂ | 1.5 | Leucine | 1 | Amino acid |
| 133.0608:0.773:p | C₂H₇N₂O₃ | 3.0 | Glycylglycine | 2 | Dipeptide |
| 133.0608:0.912:p | C₂H₇N₂O₃ | 3.0 | Asparagine | 1 | Amino acid |
| 133.0972:0.645:p | C₂H₇N₂O₂ | 2.3 | Ornithine | 1 | Amino acid |
| 134.0448:0.973:p | C₂H₇N₄O₄ | 2.2 | Aspartic acid | 1 | Amino acid |
| 139.0500:0.772:p | C₃H₉N₂O₂ | 3.6 | Urocanic acid | 1 | Amino acid derivative |
| 147.0764:0.930:p | C₃H₁₀N₂O₃ | 2.7 | Glutamine | 1 | Amino acid |
| 147.1128:0.648:p | C₄H₁₄N₂O₂ | 0.7 | Lysine | 1 | Amino acid |
| 148.0604:0.940:p | C₂H₇N₂O₄ | 0.7 | Glutamic acid | 1 | Amino acid |
| 150.0583:0.916:p | C₃H₁₁NO₂S | 5.3 | Methionine | 1 | Amino acid |
| 150.1119:0.763:p | C₃H₁₃NO₃ | 4.0 | Triethanolamine | 2 | Amine |
| 156.0768:0.686:p | C₂H₇N₂O₂ | 0.6 | Histidine | 1 | Amino acid derivative |
| 162.1125:0.782:p | C₃H₁₃NO₃ | 1.9 | Carnitine | 1 | Amino acid derivative |
| 163.0719:0.827:p | C₅H₁₀N₂O₄ | 8.6 | Glycylserine | 2 | Dipeptide |
| 166.0863:0.940:p | C₄H₁₁NO₂ | 0.6 | Phenylalanine | 1 | Amino acid |
| 168.0770:0.733:p | C₂H₇N₃O₂ | 1.8 | Amino acid derivative | 3 | Amino acid derivative |
| 175.1190:0.677:p | C₆H₁₄N₂O₂ | 1.1 | Arginine | 1 | Amino acid |
| 176.1030:0.952:p | C₆H₁₄N₂O₃ | 1.1 | Citrulline | 1 | Amino acid |
| 182.0812:0.967:p | C₆H₁₁NO₃ | 1.6 | Tyrosine | 1 | Amino acid |
| 194.1380:0.802:p | C₇H₁₉NO₄ | 3.6 | Unknown | 4 | Unknown |
| 205.0972:0.944:p | C₁₁H₂₂N₂O₂ | 2.4 | Tryptophan | 1 | Amino acid |
| 209.1290:0.795:p | C₁₁H₁₆N₂O₂ | 1.4 | Pilocarpine | 1 | Exogenous |
| 213.0990:0.635:p | C₇H₁₆N₂O₃ | 2.3 | Glycylhistidine | 1 | Dipeptide |
| 235.1199:0.841:p | C₁₁H₁₄N₂O₂ | 3.8 | Unknown | 4 | Unknown |
| 89.0244:1.146:n | C₃H₆O₃ | 0.0 | Lactic acid | 1 | Hydroxy-acid |
| 96.9696:1.686:n | H₃PO₄ | 2.1 | Phosphate | 1 | Oxoanionic compound |
| Mass        | Molecular Formula | Mass/Charge | Molecular Weight | Description                           | Source          |
|-------------|-------------------|-------------|------------------|---------------------------------------|-----------------|
| 128.0352:1.020 | C₅H₇NO₃       | 1.6         | 128.0352         | Oxoproline                           | Amino acid      |
| 133.0149:1.857 | C₆H₆O₅         | 0.0         | 133.0149         | Malic acid                           | Hydroxy diacid  |
| 135.0299:0.994 | C₆H₇O₅         | 2.2         | 135.0299         | Threonic acid                        | Sugar acid      |
| 143.1078:0.899 | C₈H₁₆O₂        | 1.4         | 143.1078         | Caprylic acid                        | Medium-chain    |
| 151.0262:0.938 | C₅H₉N₂O₂       | 0.7         | 151.0262         | Xanthine                             | Purine derivative |
| 151.0402:0.755 | C₆H₈O₃         | 0.7         | 151.0402         | Methylparaben                        | Exogenous       |
| 160.0615:0.904 | C₄H₆NO₆        | 1.2         | 160.0615         | α-Aminoacidic acid                   | Amino acid      |
| 167.0211:0.962 | C₅H₉N₂O₃       | 1.8         | 167.0211         | Uric acid                            | Purine derivative |
| 171.1391:0.856 | C₁₀H₂₀O₂       | 1.8         | 171.1391         | Capric acid                          | Medium-chain fatty acid |
| 179.0714:0.725 | C₁₀H₁₂O₃       | 1.1         | 179.0714         | Propylparaben                        | Exogenous       |
| 188.0929:0.860 | C₈H₁₃NO₄      | 0.0         | 188.0929         | Unknown                              | Unknown         |
| 191.0197:2.064 | C₈H₆O₇         | 1.0         | 191.0197         | Citric acid                          | Tricarboxylic acid |
| 195.0510:0.880 | C₆H₁₂O₇        | 1.0         | 195.0510         | Gluconic acid                        | Sugar acid      |
| 199.0725:0.868 | C₈H₁₃N₂O₄     | 0.5         | 199.0725         | Unknown                              | Unknown         |
| 199.1704:0.824 | C₁₂H₂₅O₂       | 1.0         | 199.1704         | Lauric acid                          | Medium-chain fatty acid |
| 201.0769:0.819 | C₆H₁₄O₅        | 0.5         | 201.0769         | Unknown                              | Unknown         |
| 215.0673:0.866 | C₈H₁₂N₂O₅     | 0.0         | 215.0673         | Unknown                              | Unknown         |
| 225.1245:0.777 | C₁₁H₁₈N₂O₅    | 1.3         | 225.1245         | Pilocarpic acid                      | Exogenous       |
| 227.1041:0.834 | C₁₀H₁₆N₂O₄    | 0.4         | 227.1041         | Prolyl-4-hydroxyproline              | Dipeptide       |
| 243.0624:1.276 | C₉H₁₄N₂O₆     | 0.4         | 243.0624         | Unknown                              | Unknown         |
| 257.1146:0.813 | C₁₁H₁₈N₂O₅    | 1.2         | 257.1146         | Unknown                              | Unknown         |
| 265.0940:0.817 | C₁₁H₁₄N₂O₄    | 0.0         | 265.0940         | Oxoprolylhistidine                   | Dipeptide       |
| 265.1478:0.837 | C₁₂H₂₆O₂S     | 0.4         | 265.1478         | Lauryl sulfate                       | Exogenous       |
| 277.1445:0.794 | C₁₆H₂₂O₄      | 2.5         | 277.1445         | MEHP                                  | Exogenous       |
| 320.1900:0.779 | C₁₅H₃₁NO₄S    | 0.3         | 320.1900         | N-Lauroyl-N-methyltaurine            | Exogenous       |
| 329.2333:0.739 | C₁₈H₃₄O₅      | 2.1         | 329.2333         | TriHOME                               | Hydroxy fatty acid |
| 373.0720:0.874 | C₃₇H₃₇N₂O₆S₃ | 0.8         | 373.0720         | FD&C blue dye no. 1                  | Exogenous       |

*a* Most probable formula is presented in the case of unknowns or tentatively identified compounds  
*b* Stereochemistry (D/L) not confirmed (presumably mainly L-amino acids)  
*c* Confidence levels: (1) confidently identified compound; (2) putatively identified compound; (3) putatively annotated compound class; (4) unknown compound  
*d* Compounds not included in the data matrix for being non-detectable in more than 25% of the samples  
*e* Tentatively identified as one of the isomers of trihydroxyoctadecenoic acid
**Table S2.** Overall variability within sweat samples from CF affected and non-affected infants, as well as pooled sweat QC samples. Comparisons based on original integrated relative peak area (RPA) data, as well as after batch correction or PQN normalization of samples analyzed in batch 1 (non-CF, \(n=30\); CF, \(n=8\)) and batch 2 (non-CF, \(n=20\); CF, \(n=8\)). Values represent the median RSD for 54 sweat metabolites that were present in at least 75% of the samples.

| Response considered / batch                  | RSD within groups (%) |
|---------------------------------------------|-----------------------|
|                                             | CF        | Non-CF    | QC     |
| RPA / batch 1                               | 53        | 58        | 8      |
| RPA / batch 2                               | 88        | 95        | 20     |
| RPA / batches 1 and 2                       | 76        | 82        | 25     |
| Batch-corrected RPA / batches 1 and 2       | 76        | 68        | 24     |
| PQN-normalized RPA / batches 1 and 2        | 67        | 67        | -      |
Table S3. Top-ranked sweat metabolites when comparing non-corrected relative peak areas (RPA) for CF affected and unaffected (carriers) infants.

| m/z:RMT:mode | Compound ID      | p-value\(^b\) | Effect size | Fold-change\(^c\) | q-value   |
|--------------|------------------|----------------|-------------|-------------------|-----------|
| 225.1245:0.777:n | Pilocarpic acid | 9.33E-07**     | 0.55        | 0.36              | 5.04E-05* |
| 133.0608:0.912:p | Asparagine      | 5.34E-04**     | 0.41        | 3.63              | 1.44E-02* |
| 188.0929:0.860:n | Unknown         | 4.01E-03       | 0.34        | 0.33              | 6.39E-02  |
| 277.1445:0.794:n | MEHP            | 4.73E-03       | 0.34        | 0.58              | 6.39E-02  |
| 147.0764:0.930:p | Glutamine       | 1.70E-02       | 0.29        | 1.34              | 1.72E-01  |
| 151.0402:0.755:n | Methylparaben   | 1.91E-02       | 0.28        | 0.64              | 1.72E-01  |
| 213.0990:0.635:p | Glycylhistidine | 2.37E-02       | 0.27        | 1.47              | 1.83E-01  |
| 134.0448:0.973:p | Aspartic acid   | 3.32E-02       | 0.26        | 1.39              | 2.24E-01  |
| 168.0770:0.733:p | Amino acid derivative\(^a\) | 4.22E-02 | 0.25 | 0.83 | 2.53E-01 |

\(^a\) Compounds tentatively identified
\(^b\) Two-tailed exact p-values from Mann-Whitney U test
\(^c\) Fold-change calculated from median RPAs for CF/non-CF
** Compounds significantly different after Bonferroni correction (p < 9.26E-04)
* Compounds significantly different when using FDR (q < 0.05)

Table S4. Top-ranked sweat metabolites when comparing probabilistic quotient normalization (PQN) normalized relative peak areas (RPA) for CF affected and unaffected (carriers) infants.

| m/z:RMT:mode | Compound ID      | p-value\(^b\) | Effect size | Fold-change\(^c\) | q-value   |
|--------------|------------------|----------------|-------------|-------------------|-----------|
| 225.1245:0.777:n | Pilocarpic acid | 1.92E-06*      | 0.54        | 0.24              | 1.04E-04** |
| 133.0608:0.912:p | Asparagine      | 1.07E-03       | 0.39        | 4.67              | 2.88E-02** |
| 277.1445:0.794:n | MEHP            | 1.91E-03       | 0.37        | 0.37              | 3.43E-02** |
| 151.0402:0.755:n | Methylparaben   | 3.48E-03       | 0.35        | 0.68              | 4.70E-02** |
| 188.0929:0.860:n | Unknown         | 5.52E-03       | 0.33        | 0.22              | 5.87E-02  |
| 151.0262:0.938:n | Xanthine        | 6.97E-03       | 0.32        | 0.53              | 5.87E-02  |
| 168.0770:0.733:p | Amino acid derivative\(^a\) | 7.61E-03 | 0.32 | 0.63 | 5.87E-02 |
| 147.0764:0.930:p | Glutamine       | 1.03E-02       | 0.31        | 1.48              | 6.95E-02  |
| 329.2333:0.739:n | TriHOME\(^a\)   | 3.09E-02       | 0.26        | 0.81              | 1.85E-01  |

\(^a\) Compounds tentatively identified
\(^b\) Two-tailed exact p-values from Mann-Whitney U test
\(^c\) Fold-change calculated from median PQN normalized RPAs for CF/non-CF
* Compounds significantly different after Bonferroni correction (p < 9.26E-04)
** Compounds significantly different when using FDR (q < 0.05)
Table S5. Comparison of top significant metabolites in screen-positive CF infants stratified by sex, sweat collection site, age, gestational age and birth weight. Metabolite levels (batch corrected RPAs) in the stratified categories were compared using Mann-Whitney U test for sex, sweat collection site and birth weight, whereas Kruskal-Wallis test was used for age and gestational age.

| Variables                  | Pilocarpic acid | MEHP | Asparagine | Glutamine |
|----------------------------|-----------------|------|------------|-----------|
| Sex                        |                 |      |            |           |
| Female                     | 0.492           |      | 0.329      | 0.632     |
| Male                       | 0.364           |      | 0.366      | 0.241     |
| Collection site            |                 |      |            |           |
| McMaster                   | 0.505           |      | 0.050      | 0.061     |
| Sick Kids                  | 0.644           |      | 0.636      | 0.869     |
| Age                        |                 |      |            |           |
| 0-30 days                  | 0.214           |      | 0.428      | 0.869     |
| 31-60 days                 | 0.091           |      | 0.866      | 0.428     |
| 61-95 days                 | 0.428           |      | 0.428      | 0.869     |
| Gestational age            |                 |      |            |           |
| 37.0-38.5 weeks            | 0.312           |      | 0.999      | 0.717     |
| 38.6-40.0 weeks            | 0.562           |      | 0.999      | 0.717     |
| 40.1-41.5 weeks            | 0.999           |      | 0.999      | 0.717     |
| Birth weight               |                 |      |            |           |
| 2230-3780 g                | 0.821           |      | 0.598      | 0.58      |
| 3781-5330 g                | 0.366           |      | 0.598      | 0.58      |

Table S6. Correlation matrix for pilocarpic acid, MEHP, asparagine and glutamine in comparison to sweat chloride concentrations. Data matrix shows Spearman’s rank correlation coefficients with p-values (in brackets).

| Chloride | Chloride | Asn  | Gln  | PA      | MEHP     |
|----------|----------|------|------|---------|----------|
| Chloride | 1.000    |      |      | -0.327 (6.55E-03) | -0.241 (5.09E-02) |
| Asn      | 0.366 (4.35E-03) | 1.000 |      | -0.341 (8.15E-03) | -0.141 (2.97E-01) |
| Gln      | 0.194 (1.24E-01) | 0.277 (3.69E-02) | 1.000 | -0.184 (1.45E-01) | -0.290 (2.24E-02) |
| PA       | -0.327 (6.55E-03) | -0.341 (8.15E-03) | -0.184 (1.45E-01) | 1.000     | 0.444 (1.90E-04) |
| MEHP     | -0.241 (5.09E-02) | -0.141 (2.97E-01) | -0.290 (2.24E-02) | 0.444 (1.90E-04) | 1.000     |
Table S7. Statistical analysis for pilocarpic acid, MEHP, asparagine and glutamine in the first and second batches separately in samples of screen-positive CF infants.

| Compound ID | Parameters | Batch 1 | Batch 2 |
|-------------|------------|---------|---------|
| Pilocarpic acid | $p$-value$^a$ | 5.03E-05** | 7.25E-03 |
| | Effect size | 0.59 | 0.50 |
| | Fold-change$^b$ | 0.33 | 0.36 |
| | $q$-value | 2.77E-03* | 4.43E-02* |
| MEHP | $p$-value$^a$ | 1.62E-03 | 2.31E-01 |
| | Effect size | 0.48 | 0.23 |
| | Fold-change$^b$ | 0.43 | 0.87 |
| | $q$-value | 3.66E-02* | 3.43E-01 |
| Asparagine | $p$-value$^a$ | 1.99E-03 | 7.55E-02 |
| | Effect size | 0.47 | 0.34 |
| | Fold-change$^b$ | 3.46 | 6.88 |
| | $q$-value | 3.66E-02* | 1.66E-01 |
| Glutamine | $p$-value$^a$ | 9.99E-02 | 3.46E-03 |
| | Effect size | 0.26 | 0.54 |
| | Fold-change$^b$ | 1.30 | 6.63 |
| | $q$-value | 3.89E-01 | 3.36E-02* |

$^a$Two-tailed exact $p$-values from Mann-Whitney U test
$^b$Fold-change calculated from median RPA for CF/non-CF
**Compounds significantly different after Bonferroni correction ($p < 9.26E-04$)
*Compounds significantly different when using FDR ($q < 0.05$)
Figure S1. Dilution trend filter for nontargeted metabolite profiling of the sweat metabolome based on temporal signal pattern recognition when using MSI-CE-MS. (A) Injection configuration for the dilution trend filter using a pooled sweat QC (n = 10) serially diluted by factors of 1, 2, 5, and 10-fold, including a triplicate for the least diluted sample and a blank. (B) Example extracted ion electropherogram (EIE) of an authentic feature (citrulline, m/z 176.1030, ESI+), which follows the dilution trend ($R^2 = 0.989$), can be reliably measured with good precision (RSD = 4.2%, n=3), and shows no background signal in the blank (i.e., signal is derived from sweat). (C) Example of a spurious signal (m/z 178.1588, ESI+), which does not follow the expected dilution trend and can be confidently excluded from the mass list.
Figure S2. Comparison of MS/MS spectra for unambiguous identification of pilocarpic acid and MEHP as lead exogenous metabolites that differentiate affected CF infants from non-affected carriers. Mirror plots comparing (A) pilocarpic acid and (B) MEHP in sweat samples relative to authentic standards, with matching scores of 98.25 and 93.33%, respectively.
**Figure S3.** Overview of batch effects and application of a batch-correction algorithm on representative metabolites in original data as summarized in a 2D scores plot using PCA. QC plots for some example compounds comparing changes in RPAs (A) and batch-corrected RPAs (B). The step-change in RPAs observed for glutamine was corrected by batch effect adjustment. On the other hand, the correction algorithm had a negative impact on some compounds not initially affected by batch effect (e.g., asparagine and pilocarpic acid). PCA plots show overall variability in the responses for anions (C) and cations (D) before and after batch effect adjustment, respectively.
Figure S4. Boxplots comparing age, birth weight, sweat chloride and dried blood spot IRT levels in CF and non-CF screen-positive CF infants. Infants in the two groups had no significant difference with respect to testing age (A) and birth weight (B), whereas sweat chloride (C) and IRT (D) were significantly elevated in screen-positive affected CF cases.
Figure S5. Percentage of pilocarpine and pilocarpic acid in gel disc water extracts. Samples of new and used gel discs used in pilocarpine-stimulated iontophoresis for sweat collection were extracted in water and compared to a control (i.e., pilocarpine solution in water). A low background fraction of pilocarpic acid (0.1-1.2%) was detected in the extracts, which provides evidence that the much higher pilocarpic acid concentrations measured in sweat were generated in vivo from screen-positive infants.
Figure S6. Supporting evidence of the origin of the weakly basic plasticizer metabolite, MEHP (277.144 m/z as [M-H]). Pooled infant sweat samples were analyzed as a QC run (A), and a dilution trend filter (B) for feature selection was used in MSI-CE-MS under alkaline conditions and negative ion mode detection. In both cases, MEHP is reproducibly detected in infant sweat samples with good precision having a distinct linear dilution trend (C) that does not originate from the background during spray formation.