Analysis of fecal biomarkers to impact clinical care and public health

by

Mariana Guadalupe Matus García

BSc, Universidad Nacional Autónoma de México (2009)
MSc, Wageningen University (2012)

Submitted to the Computational and Systems Biology Graduate Program
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 2018

© 2018 Massachusetts Institute of Technology 2018. All rights reserved.

Author

Signature redacted

Certified by

Signature redacted

Eric J. Alm
Professor, Biological Engineering
Thesis Supervisor

Accepted by

Signature redacted

Christopher B. Burge
Director, Computational and Systems Biology Graduate Program
Analysis of fecal biomarkers to impact clinical care and public health

by

Mariana Guadalupe Matus Garcia

Submitted to the Computational and Systems Biology Program
on June 6, 2018 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

DNA sequencing and metabolomics technologies have accelerated the discovery of novel biomarkers in clinical samples. In this thesis, I explore the potential of fecal biomarkers to impact clinical and public health practice through non-invasive assessments. First, I highlight the potential of the gut microbiome to provide novel diagnostic and therapeutic targets. By analyzing the gut microbiome and metabolome of mice exposed to a high salt diet, we identified *Lactobacillus* as a potential probiotic to counteract salt-sensitive conditions such as high blood pressure. Next, I present preliminary validation of wipe samples as a patient-friendly alternative to standard stool collection methods, in particular for the clinical management of Inflammatory Bowel Disease patients. By comparing paired stool and wipe samples, I show that wipe samples capture the same gut microbiome profiles as standard stool samples, and can also be used to quantify fecal calprotectin. Finally, I present the first ever analysis of the microbiome and metabolome of wastewater collected from a residential neighborhood. By testing samples collected hourly over one day, we identified thousands of bacteria and metabolites derived from human activity. Glucuronide compounds that directly reflect consumption of pharmaceutical products and drugs were identified for the first time in a wastewater epidemiology study. Our results highlight the potential of testing wastewater in geo-localized residential areas to produce high-quality data to inform public health practice. Together, these results show the potential of leveraging high-throughput technologies to create seamless readouts of human and population health.

Thesis Supervisor: Eric J. Alm
Title: Professor, Biological Engineering
Acknowledgements

My time at MIT has been life changing. Beyond being a great place to learn and contribute to science, the Alm lab and the MIT community have given me the opportunity to explore many interests and find a motivating path forward.

First and foremost, I would like to thank my advisor Prof. Eric Alm for giving me the opportunity to pursue new lines of research in his lab, for his creative feedback to make my projects always better, for his support of my entrepreneurial path, and for his encouragement during the low points. I am very happy and proud to have been part of his research group, and I am looking forward to continuing our partnership post-graduation!

Thank you to the past and current thesis committee members Prof. Jon Runstadler, Prof. Elizabeth Kujawinski, Prof. David Kong, Prof. Otto Cordero for their unconditional support and valuable feedback to my scientific work. Thank you to Dr. Luis Perez-Breva for contributing to my committee meetings from his entrepreneurial perspective. I am also grateful to my former advisor Mark van Passel for his mentorship and friendship during the early stages of my scientific career, and for encouraging me to apply to MIT just like his previous student Chris Smilie had! I would have never applied to MIT if it wasn’t for Mark.

Thank you to my many collaborators and mentors in the Alm lab, Senseable City Lab, Kujawinski lab, Runstadler lab, Kuwait Institute for Scientific Research, Max-Delbruk Center for Molecular Medicine in Berlin, Universidad de Guanajuato and Instituto Mexicano de Medicina Genomica. I would like to especially acknowledge Prof. Ilana Brito, Prof. Jon Runstadler, Prof. Elizabeth Kujawinski, Prof. Carlo Ratti, Claire Duvallet, and Newsha Ghaeli and Yaniv Turgeman for their contribution to the Underworlds Project. The best outcome of having started this project has been the opportunity to work with a super talented team. Projects like Underworlds give a glimpse of the potential of interdisciplinary science.
I was also fortunate to be part of many communities outside the Alm lab. Thank you to the Computational and Systems Biology Program, to our Director Chris Burge and to Jacquie Carota, as well as my fellow CSBs. In particular, thank you to the 2012 class for being awesome friends. Thank you to the Sidney Pacific Graduate Residence, housemasters Julie and Neel Shah, SPEC 2015, and all the students I got to live and work with to make everyone feel at home. Thank you to the MIT Mexican Students Association, Griselda Gomez –Director of MIT-Mexico, and the Mexican community in the Boston area including Consul Emilio Rabasa, Leonardo Lavalle and Carmen Alanis, for helping me stay close to my culture and enjoy the company of many talented Mexicans.

And thank you to my cofounder Newsha Ghaeli, for being the best partner in my entrepreneurship journey, and to the MIT entrepreneurship ecosystem for their support through our transition out of academia. Thank you to Dean Anantha Chandrakasan, Jinane Abounadi and the Sandbox program; Bill Aulet, Trish Cotter, Nick Meyer and everyone at the Martin Trust Center for MIT Entrepreneurship; Gilad Rosenzweig, Dennis Frenchman and the DesignX team; Keely Swan and the IDEAS Global Challenge program. Thank you for believing in us!

And thank you to my family for being my source of strength and inspiration. My husband Nezar, whom I met in my first semester at MIT and whom I married in the MIT Chapel! It’s like the geekiest fairy tale. My sister Monica and niece Sofia, my mother Mariana and father Fernando, my grandmother Carmen, and my family in-law Vicki, Alex and Camil. I know the future is bright because I have you all in my life.
Contents

1 Introduction .......................................................................................................................... 6

2 Effect of high salt diet on the gut microbiome and metabolome of mice .......................................................... 9

3 Microbiome sequencing from wipe samples ........................................................................ 28

4 Measurement of population health and activity through wastewater testing .......................................................... 40

5 Conclusion ....................................................................................................................... 58

A Experimental protocols developed for Chapter 3 ......................................................... 60

B Experimental protocols developed for Chapter 4 .......................................................... 74
Chapter 1

Introduction

The human gastrointestinal tract digests food, absorbs nutrients, and plays important roles in maintaining metabolic homeostasis. The microbial community residing in our gut harvests energy from the food we eat, trains our immune system, breaks down xenobiotics, and releases metabolites and hormones important for regulating our physiology. Disruption of this system is associated with disease and today can be assessed to some extent through the analysis of stool samples.

Today, fecal biomarkers are routinely used in the clinic to diagnose conditions affecting the gastrointestinal tract, including: poor nutrient absorption, inflammation, colorectal cancer, and infection with parasites or bacteria like Clostridium difficile. Two emerging high throughput technologies, DNA sequencing and metabolomics, are revolutionizing the way we analyze stool and are accelerating the discovery of novel fecal biomarkers that can be leveraged for clinical care and public health practice.

Next generation sequencing of the 16S rRNA gene is widely used to characterize the composition of complex microbial communities. The 16S rRNA gene is present in all bacteria and archaea and can be selectively amplified and sequenced to obtain the relative abundance of operational taxonomic units (OTUs) in a sample. Continuous improvements and decreased costs have enabled widespread adoption of the technology, which has spurred the discovery of gut bacteria associated with disease in hundreds of clinical studies.

However, 16S characterization is limited by a lack of quantitative functional annotation. The fecal metabolome provides a functional readout of microbial activity and is therefore complementary to sequencing-based approaches. Metabolomics is the study of low-molecular weight metabolites (<1500 Da) within a biological system, typically profiled and quantitated using mass spectrometry. While untargeted metabolomics aims to detect and characterize the full metabolome without bias, targeted
metabolomics focuses on quantifying a specific set of target compounds through the use of authentic standards.

In this dissertation, I present three projects that demonstrate the potential of applying high-throughput technologies such as DNA sequencing and metabolomics to discover and leverage fecal biomarkers.

The first project describes an analysis of the gut microbiome and metabolome in mice subjected to a high salt diet. High salt content in the Western diet has been implicated in several health disorders including high blood pressure. However, not much is known about the mechanistic effect of high salt on cardiovascular health. By analyzing the gut microbiome and metabolome of mice exposed to a high salt diet, we demonstrated that the gut microbiome links diet, the immune response and cardiovascular health. Furthermore, Lactobacillus was identified as a potential therapeutic target to counteract salt-sensitive conditions.

The second project aims to make stool collection more patient-friendly to facilitate continuous monitoring of fecal biomarkers, which is particularly relevant for the clinical care of Inflammatory Bowel Disease (IBD) patients. IBD is a chronic disease characterized by inflammation flares that affect the gastrointestinal tract and require increased medication and even surgery. Currently, the standard practice to assess disease severity is endoscopy which is expensive, invasive and prone to complications. Here we present preliminary validation of wipe samples as an alternative stool collection method to facilitate continuous monitoring of inflammation in IBD patients.

The third extends assessment of fecal biomarkers to the community level revealing the potential of wastewater testing to assess the health of communities in cities. To date, wastewater testing has been largely developed as a tool to collect data on illicit drug consumption. The current standard practice is to collect samples from wastewater treatment plants (WWTPs) and apply targeted metabolomics. However, the lack of geographic specificity of data from WWTPs limits its value in informing public health initiatives, especially within urban environments. Furthermore, high-throughput sequencing and untargeted metabolomics can expand the list of detectable human health biomarkers in wastewater beyond illicit drugs. Here we present the first unbiased assessment of bacteria and metabolites detectable in wastewater from a residential community, and also show how many biomarkers of
human activity are lost as wastewater travels to the treatment plant. Our results highlight the potential of sampling in localized residential areas to accelerate the discovery and validation of a breadth of biomarkers of population health.

These projects exemplify that human stool is not just waste, but rather a rich source of information about our health and behavior, at the individual and population level. Advances on the gut microbiome field are accelerating the discovery of microbial, metabolite or protein biomarkers associated with health and disease. Clinical care and public health practice are becoming more data-driven, and stool analysis presents an excellent opportunity to create non-invasive and versatile monitoring platforms.
Chapter 2

Effect of high salt diet on the gut microbiome and metabolome of mice


2.1 Abstract

High salt content in the Western diet has been implicated in several health disorders including cardiovascular disease. Based on epidemiological observations, medical guidelines recommend reducing salt intake to improve cardiovascular health. However, not much is known about the mechanistic effect of high salt on cardiovascular health. This is the first study to examine the effect of a high salt diet on the gut microbiome and metabolome of mice. We show that while a high salt did not affect the overall composition of the gut microbiome, the commensal Lactobacillus murinus was consistently depleted. In contrast, the gut metabolome was largely affected with most metabolites being decreased by high salt diet. Subsequent experiments showed that L. murinus prevented experimental salt-sensitive hypertension by modulating T_{h}17 cells. Our results highlight the gut microbiome as a link between diet, the immune response and cardiovascular health, and therefore its potential as therapeutic target to counteract salt-sensitive conditions.
2.2 Introduction

High salt content in the Western diet has been implicated in numerous disorders (1), particularly in cardiovascular health (2). Guidelines (3, 4) and public initiatives recommend reducing salt intake, but an improved mechanistic understanding is needed. The deleterious effect of a high salt diet (HSD) on cardiovascular health is driven by arterial hypertension and associated with increased morbidity and mortality (2, 5). So far, most studies have focused on the role of the kidneys, the sympathetic nervous system and direct effects on the vasculature (6). However, some studies have implicated the immune system in these processes (7), linking pro-inflammatory T cells to the development of hypertension (8). In particular, interleukin-17A (IL-17A)-producing CD4\(^+\) T\(_{17}\) cells may promote hypertension (9, 10). Studies have recently demonstrated that the generation of pathogenic T\(_{17}\) cells could be promoted by a high salt environment.

The interaction between HSD and the gut microbiome has not been thoroughly investigated. The intestine is exposed to varying salt loads of ingested foods. Gut microbes are known to respond to fluctuations in dietary composition (11) leading to transient or persistent alterations in the gut microbiome (12). Diet-induced shifts in microbiome composition may have profound effects on the host, especially on T cells (13). T\(_{17}\) cells are particularly affected by the abundance of specific commensal bacteria (14). Here we examine the effect of high salt challenges on the gut microbiome, metabolome and immune system of mice, and determined the implications of HSD for hypertension.
2.3 Materials and Methods

Animal ethics.

All animal experiments were conducted in accordance with institutional, state and federal guidelines and with permission of the local animal ethics committees (Landesamt für Gesundheit und Soziales Berlin, Germany; Regierung Unterfranken, Würzburg, Germany; Ethical Committee for Animal Experiments, Hasselt University, Belgium). Male mice were maintained on a 12:12 h day:night cycle with constant access to food and water.

Mouse high salt feeding and feces collection.

All normal salt (NSD, E15430-047) and high salt (HSD, E15431-34) purified diets used for mouse experiments were purchased from Ssniff (Soest, Germany). Diets were gamma-irradiated (25 kGy) and identical in composition except for NaCl content (NSD: 0.5% NaCl, HSD: 4% NaCl). Drinking water for HSD animals was supplemented with 1% NaCl.

For fecal microbiome analyses, male FVB/N mice aged 12 weeks were purchased from Charles River and accustomed to NSD. Control animals remained on the NSD (n = 8), others were switched to HSD (n = 12) for 14 days. A subgroup was switched back to NSD for another 14 days (n = 8). Body weight and food intake were monitored. To avoid cage effects, mice were housed individually. Fresh fecal pellets were collected directly from the anal orifices, immediately flash-frozen in liquid nitrogen and stored at −80 °C for later analyses.

DNA extraction from mouse feces and 16S sequencing.

DNA was extracted from a single fecal pellet from each mouse using the Power Soil kit (MO BIO Laboratories). The protocol was modified from the manufacturer’s instructions to include proteinase K treatment to further lyse the cells. After addition of proteinase K (final concentration 5 mg ml⁻¹) samples were incubated at 65 °C for 10 min and a further 10 min at 95 °C. Plates were inverted to mix during both incubations. The V4 region of the 16S rRNA gene was amplified with 515F and 806R primers (15) using a two-step PCR library preparation as previously described (16). An Illumina MiSeq was run for 250 cycles to produce paired-end reads.
16S rDNA data processing.

The raw sequences were de-multiplexed, allowing at most 2 mismatches in the barcode before discarding a sequence. Primers sequences were removed, allowing at most 2 mismatches in the primer sequence before discarding a sequence. Forward and reverse reads were merged by comparing alignments with lengths of 253 ± 5 nucleotides. The alignment with the fewest mismatches was used unless the number of mismatches was greater than 2, in which case the read pair was discarded. Merged reads were filtered for quality by removing reads with more than 2.0 expected errors (17). Each unique sequence was assigned a taxonomy using the ribosomal database project (18), truncating the taxonomy to the highest taxonomic level with at least 80% support. Sequences that were assigned the same taxonomy were then placed in the same OTU. De novo OTUs were also called using usearch (19).

16S rDNA data analysis.

For most analyses, three samples were excluded because their read counts were low (<1,000 counts). The MDS ordination and PERMANOVA test were computed using the vegan package (20) in R. The AdaBoost classifier (21) was run with 10⁷ estimators using the scikit-learn module (22) in Python. The random forest classifier (23) was run with 10⁶ estimators also using the scikit-learn module in Python.

Fecal metabolite analysis.

An extraction mixture of methanol:chloroform:water (MCW; 5:2:1 v:v:v) (methanol LC–MS-grade, chloroform, Reagent Plus 99,8% Sigma-Aldrich) with cinnamic acid (2 μg ml⁻¹, Sigma-Aldrich) as internal standard was added to the sample. Samples were dissolved in MCW (1 ml per 60 mg of sample) using the tissue lyser (Precelllys 24 lysis and homogenization, Bertin Technologies), samples were cooled on ice between the shaking cycles. Samples were shaken at 1,000 r.p.m. and 4 °C for 60 min. After addition of ice-cold water (half of the MCW volume), samples were shaken at 1,000 r.p.m. and 4 °C for 10 min. Samples were centrifuged for 10 min at 14,000 r.p.m. to separate the polar (top), lipid (bottom) and interface (tissue debris) layers. The polar phase containing metabolites was dried under vacuum for 12 h. Samples were derivatized as follows: the dried extracts were dissolved in 20 μl of methoxyamine hydrochloride solution (Sigma-Aldrich; 40 mg ml⁻¹ in pyridine (Roth)) and incubated for 90 min at 30 °C shaken at 1,000 r.p.m. followed by the addition of 40 μl of N-methyl-N-[trimethylsilyl]trifluoroacetamide (MSTFA; Machery-Nagel) and incubation at 37 °C for 45 min agitated at 1,000 r.p.m. The extracts were centrifuged for 10 min at 14,000 r.p.m., and aliquots of 30 μl
were transferred into glass vials (Chromacol) for gas chromatography–mass spectrometry (GC–MS) measurements. Metabolite analyses were performed with a Pegasus IV mass spectrometer (LECO) as described previously (24). The GC–MS chromatograms were pre-processed with the ChromaTOF software (LECO). Calculation of retention index, mass spectra identification and metabolite quantification were performed using the in-house Maui-SILVIA software tool (25).

Measured values from 66 metabolites were obtained. Because a paired analysis (metabolites at NSD baseline versus metabolites after HSD) was performed, absence of a given value made the exclusion of the corresponding second value necessary. This was in only 33 out of 1,056 cases (3.1%). A small pseudocount value (0.001) was added to all metabolite values and data were log_{10} transformed. Data from each metabolite were normalized by subtracting the minimum and dividing by the maximum value across all eight mice. A PCA was performed using the PCA module in the scikit-learn package of Python. The heat map was prepared with the clustermap function of the Seaborn package.

**Isolation of L. murinus.**

Fecal samples from healthy male NSD-fed FVB/N mice were dissolved and diluted at a 1:10 dilution in anaerobic phosphate-buffered saline (PBS) (pH 7.6) containing l-cysteine HCl at 0.1% in a Coy Anaerobic Chamber (5% H2, 20% CO2, 75% N2). Samples were diluted tenfold and each dilution spread on LAMVAB agar (26). Plates were incubated at 37 °C under anaerobic conditions and examined for growth at 24 h. Individual colonies growing at the highest dilution were picked into LAMVAB medium and grown for an additional 16 h. Liquid cultures were stored in 15% DMSO. For identification of isolates, DNA was extracted by adding 5 μl liquid culture to 20 μl sterile distilled water and storing at 4 °C overnight; 2 μl of this extract was amplified with Phusion HF polymerase in a 20-μl reaction using universal 16S primers 27F (5' -AGAGTTTGATCMTGGCTCAG-3') and 1492R (5' -TACGGYTACCTTGTTACGACTT-3'). PCR products were purified using Agencourt AMPure XP and submitted with the 27F primer for Sanger sequencing. An isolate for which the full-length 16S sequence shared 100% identity with the V4–V5 region of the *Lactobacillus* species identified in the 16S library was selected for further study. Frozen stocks of *L. murinus* (in PBS with 25% glycerol) were prepared, stored at −80 °C and used for gavage of salt-sensitive mice.

**Salt-sensitive hypertension in mice.**

To induce salt-sensitive hypertension, the l-NAME/salt mouse model was used as described previously (27). This non-surgical intervention closely recapitulates salt-sensitive hypertension
common in humans (27). In brief, NSD-fed male FVB/N mice, aged 10–12 weeks, received
pretreatment with L-NAME (0.5 mg ml\(^{-1}\), Sigma-Aldrich) in drinking water for three weeks, followed
by a one week washout period with NSD and normal drinking water. Then, mice were switched to
either HSD with oral administration of \(L\). \textit{murinus} (daily gavage of 200 \(\mu\)l 10\(^7\) CFU ml\(^{-1}\) \(L\). \textit{murinus}
 suspension), HSD with oral administration of control solution (daily gavage of 200 \(\mu\)l PBS:glycerol)
or NSD for two weeks. For blood pressure measurements, mice were implanted with miniature
subcutaneous radiotelemetry devices under anaesthesia (Data Sciences International) before the L-
NAME/salt protocol. Thereby, systolic and diastolic blood pressures were recorded continuously at
5-min intervals in freely moving mice. Following hypertension induction with HSD for three weeks,
HSD was continued and mice were concomitantly gavaged with 200 \(\mu\)l 10\(^7\) CFU ml\(^{-1}\) \(L\). \textit{murinus}
in PBS and glycerol daily. Mice were euthanized under anaesthesia and spleens and intestines were
collected. Single-cell suspensions of the small intestinal (si) and colonic (c) lamina propria lymphocytes
(LPL) were obtained by enzymatic and mechanical dissociation using the Mouse Lamina Propria
Dissociation kit (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer’s
protocol. Cell debris was removed using a Percoll (GE Healthcare) density gradient centrifugation as
described previously (28). Splenocyte single-cell suspensions were obtained using 70-\(\mu\)m strainers,
followed by erythrocyte lysis and subsequent filtering using a 40-\(\mu\)m mesh. Cells were counted by
trypan blue exclusion and labelled for flow cytometry analysis. Isolated immune cells were either
directly stained for surface markers using the respective fluorochrome-conjugated antibodies (30 min
in PBS supplemented with EDTA and BSA) or restimulated with 50 ng ml\(^{-1}\) phorbol-12-myristate-
13-acetate (PMA, Sigma-Aldrich), 750 ng ml\(^{-1}\) ionomycin (Sigma-Aldrich) and 0.75 \(\mu\)l ml\(^{-1}\) GolgiStop
(BD Bioscience) for 4 h at 37 \(^{\circ}\)C and 5% CO\(_{2}\) in RPMI 1640 medium (Sigma-Aldrich) supplemented
with 10% FBS, 1% penicillin–streptomycin. For all measurements, dead cell exclusion was performed
using a fixable viability dye for 405 nm (Thermo Fisher). For intracellular staining, cells were
permeabilized and fixed using the FoxP3 Staining Buffer kit (eBioscience) and labelled using the
respective antibodies. Antibodies used are listed below. Cells were analyzed with the BD FACSCanto
II flow cytometer and BD FACSDiva software (BD Bioscience). Data analysis was performed with
FlowJo v.10 (FlowJo LLC).

Statistics.
Power calculation is a prerequisite for any animal experiment according to the local animal law and
was performed using G* Power software v.3.1.9.2. Effect sizes were calculated from previously
published experiments. Animals were randomly assigned to the respective body weight-matched
groups, probiotic and control treatment were administered without knowledge of the treatment
groups. Unless otherwise specified, box plots show median and IQR with whiskers showing minimum
and maximum values, bar graphs show mean ± s.e.m. Outliers identified by Grubbs’ test were
excluded. Normality was assessed using a Kolmogorov–Smirnov test. To compare independent
measurements, we used a t-test and Mann–Whitney U-test, as appropriate. To compare dependent
measurements, we used a paired t-test or Wilcoxon signed-rank test, as appropriate. To compare more
than two groups, we used one-way ANOVA followed by Tukey’s post hoc test or Kruskal–Wallis test
followed by Dunn’s post hoc test, as appropriate. Statistical analysis was performed using GraphPad
Prism v.6. To analyze mouse blood pressure telemetry data, we conducted repeated measurement
analysis by using linear mixed models. We tested a random intercept versus a random intercept–slope
model and selected the best-fit model. Data analysis was performed with R (v.3.1.1 R Foundation,
Vienna, Austria) using the packages lme4 and nlme. A P value < 0.05 was considered statistically
significant.

**Code availability.**

Code used for the 16S rDNA data analysis has been uploaded to a github repository
(https://github.com/almlab/analysis-salt-responsive). Software was obtained from publicly available
sources; papers describing the software are cited in the text.

**Data availability.**

Raw files of the bacterial V4–V5 16S rRNA data and the L. murinus genome have been uploaded to
the NCBI Sequence Read Archive as Bioproject PRJNA400793.
2.4 Results

**High salt diet does not affect the overall composition of the gut microbiome in mice.**

To determine the effect of a HSD on the composition of the gut microbiome, we fed eight FVB/N mice a normal salt diet (NSD) or HSD for 14 days and analyzed their gut microbiome by 16S ribosomal DNA (rDNA) sequencing. Another eight mice remained under NSD through the treatment period as control. Both diets were equally well-tolerated, indicated by similar body weight and food intake. HSD-fed mice had a significantly higher fluid and salt intake than NSD-fed mice, but similar intestinal transit. There was high level of variation in the gut microbiome across all mice, and within the same animal (Figure 1A, B). This high level of variation may be at least partially due to housing mice individually. The overall microbial composition (based on operational taxonomic units, OTUs, assigned using the Ribosomal Database Project) showed no obvious pattern shifts between HSD- and NSD-fed mice (Figure 2).

**High salt diet alters specific OTUs of the fecal microbiome and decreases *Lactobacillus* in mice.**

The absence of large-scale taxonomic differences was unexpected but consistent with the fact that the two diets are identical in energy content and only differ in salt content. To identify the specific bacterial OTUs that did change across diet, we used a sensitive machine learning approach. An AdaBoost classifier trained to distinguish NSD from HSD samples on day 14 of the treatment identified 8 OTUs with nontrivial feature importance (Figure 3A) with 92% accuracy (Figure 3B). These OTUs varied in maximum relative abundance (from 0.04% to 19.5%) and responded differently to the HSD (Figure 3A, C). OTUs identified as most important were consistent across different cross-validation runs and across different algorithms. The most important OTU (25% feature importance) was a member of the genus *Lactobacillus* and was depleted after HSD (Figure 3D). Other features included OTUs from *Prevotellaceae*, *Pseudoflavonifractor*, *Clostridia*, *Parasutterella*, *Akkermansia*, *Bacteroidetes* and *Alistipes* (Figure 3A, C and Figure 4). *Lactobacillus* depletion showed a quick onset that was detectable one day after the initiation of the HSD, remaining at low levels during the HSD with the lowest abundance on day 14. When the mice were returned to a NSD, the *Lactobacillus* OTU abundance returned to baseline levels (Figure 3D).
Because the *Lactobacillus* OTU was the bacterial group most strongly associated with high salt, we isolated a *Lactobacillus* strain from mouse feces. The 16S rDNA sequence of the isolate shared 100% identity with the V4–V5 16S region of the OTU described above, and was identified as *L. murinus*. We confirmed the decrease in abundance of this strain after a HSD using qPCR (Figure 3E, F).

**High salt diet strongly alters the fecal metabolite profile.**

Notably, analysis of fecal metabolites from central carbon and nitrogen metabolic pathways by gas-chromatography mass spectrometry showed clear differences between the two groups (Figure 5A, B). Principal component analysis and hierarchical clustering showed grouping of NSD and HSD fecal metabolites. HSD led to an overall reduction in total metabolite peak intensities in fecal samples (Figure 5D). The fecal levels of the nucleoside adenosine were similar in both diets suggesting that the change in metabolites is not due to a decrease in overall bacterial biomass (Figure 5C). Together, these results suggest that bacterial metabolism may be inhibited by high salt.

**Lactobacillus murinus reduces salt-sensitive hypertension.**

Accumulating evidence suggests that T\(_{h17}\) cells have a role in the genesis of hypertension (7). Moreover, a recent meta-analysis provided preliminary support that *Lactobacillus*-rich probiotics might affect blood pressure in hypertensive subjects (29). Therefore, we tested whether *L. murinus* treatment decreased experimental salt-sensitive hypertension. Blood pressure increased over three weeks of HSD (Figure 6A and B). Concomitant daily treatment with *L. murinus* led to a significant reduction in systolic blood pressure and normalization of diastolic blood pressure (Figure 6A and B). We next asked whether *L. murinus* treatment affects T\(_{h17}\) cells in experimental salt-sensitive hypertension and analyzed intestinal and splenic lymphocytes by flow cytometry. Compared to NSD, a HSD led to a significant increase in CD4\(^+\)ROR\(_{yt}\)^+ T\(_{h17}\) cell frequencies in siLPL, which was significantly reduced by *L. murinus* treatment (Figure 6C). In addition, flow cytometry analysis of siLPL, cLPL and splenic lymphocytes revealed a significant reduction in T\(_{h17}\) cell frequencies by *L. murinus* treatment compared to HSD feeding alone (Figure 6D–F). Thus, *L. murinus* prevents HSD-induced generation of T\(_{h17}\) cells and consequently ameliorates salt-sensitive hypertension.
2.5 Discussion

Here we show the effect of increased salt consumption on intestinal bacteria in mice and expand the existing knowledge on the effects of this nutrient. Contrary to other dietary interventions, high salt diet did not alter the overall composition of the gut microbiome. Several intestinal bacteria were affected by high salt; particularly, *Lactobacillus* spp. were suppressed. Interestingly, analysis of fecal metabolites shows a drastic decrease in the levels of most intestinal metabolites. Future studies could investigate if high salt content decreases metabolic activity in the gut microbiome.

Changes in the gut microbiome and metabolome of mice may contribute to salt-induced TH17 cell responses and salt-sensitive conditions. The development of microbiota-targeted therapies is an intriguing new avenue for many diseases. Nevertheless, changes in microbiome composition or function must first be carefully shown to contribute to any disease. Our experimental data in mice suggest that the gut microbiota might serve as a potential target to counteract salt-sensitive conditions. The identification of *Lactobacillus* as a ‘natural inhibitor’ of high salt-induced TH17 cells in mice could serve as a basis for the development of novel prevention and treatment strategies. It is up to randomized controlled trials in humans with diseases to test this hypothesis. Moreover, any future dietary salt intervention trial should thus consider monitoring the microbiome to expand on our observations.
2.6 References


2.7 Figures

A

NSD → NSD

Animal

Relative abundance

-1 14 -1 14 -1 14 -1 14 -1 14 -1 14

OTU

Day

Root: Bacteria
Bacteroidetes; Bacteroidia; Bacteroidales
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides
Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae
Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae, Barnesellia
Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae, Parabacteroides
Bacteroidetes; Bacteroidia; Bacteroidales; Rikenellaceae, Alitipes
Candidatus _Saccharibacteria
Deferrribacteres; Deferrribacteres; Deferrribacteres; Deferrribacteres; Mucispirillum
Firmicutes
Firmicutes; Bacill; Bacillales, Alicyclobacillaceae, Tunebacillus
Firmicutes; Bacill; Lactobacillales, Lactobacillaceae; Lactobacillus
Firmicutes; Clostridia, Clostridiales
Firmicutes; Clostridia, Clostridiales, Lachnospiraceae
Firmicutes; Clostridia, Clostridiales, Lachnospiraceae, Clostridium XIVa
Firmicutes; Clostridia, Clostridiales, Lachnospiraceae, Clostridium XIVb
Firmicutes; Clostridia, Clostridiales, Ruminococcaceae
Firmicutes; Clostridia, Clostridiales, Ruminococcaceae, Clostridium IV
Firmicutes; Clostridia, Clostridiales, Ruminococcaceae, Flavonifractor
Firmicutes; Clostridia, Clostridiales, Ruminococcaceae, Oscillibacter
Firmicutes; Clostridia, Clostridiales, Ruminococcaceae, Pseudoflavonifractor
Firmicutes, Erysipelotrichia; Erysipelotrichales, Erysipelotrichaceae
Proteobacteria, Betaproteobacteria, Burkholderiales, Sutterellaceae, Parasutterella
Proteobacteria, Gammaproteobacteria, Alteromonadales, Shewanellaceae, Shewanella
Proteobacteria, Gammaproteobacteria, Enterobacteriales, Enterobacteriaceae, Escherichia/Shigella
Proteobacteria, Gammaproteobacteria, Oceanospirillales, Halomonadaceae, Halomonas
Figure 1. Fecal microbiome profiles of mice kept on a NSD or HSD over time. Taxonomic bar charts showing relative abundance of ribosomal database project-based OTUs on indicated days.

A. Mice remaining on a NSD for 14 days served as NSD controls. Baseline NSD day -1 and NSD day 14 are shown.

B. Separate mice were switched from NSD (days -2 and -1) to HSD for 14 days, and were finally re-exposed to NSD for another 14 days (recovery). For time course analyses, fecal samples from baseline NSD days (-1 and -2), early (days 1–3) and late (day 14) HSD days and NSD recovery days (days 15–17, 19, 22, 28) are shown. \( n = 8 \) mice per group.
Figure 2. High salt diet does not affect the overall gut microbiome profile in mice. Mouse 16S rDNA fecal microbiome samples do not separate by diet in a MDS ordination (white, NSD samples; black, HSD samples; grey, recovery on NSD).
Figure 3. High salt diet alters specific OTUs of the fecal microbiome and decreases *Lactobacillus* in mice.

A. AdaBoost identified eight 16S rDNA OTUs distinguishing samples of NSD-fed mice from HSD-fed mice.

B. AdaBoost classifier accuracy per mouse and diet.

C. Relative OTU abundances on HSD day 14 (n = 12 mice; n = 8 NSD control mice).

D. *Lactobacillus* abundance over time. Samples with >1% are not shown.

E, F, *L. murinus* qPCR (n = 8 mice).
Figure 4. Time series for seven OTUs, other than *Lactobacillus*, identified by the classifier. NSD and HSD phases are indicated by white and grey backgrounds. Mice (*n* = 12) were switched from NSD to HSD and back to NSD (subgroup of *n* = 8). Other control mice (*n* = 8) that remained on NSD are shown in white. Box plots show median, IQR with whiskers of 1.5× IQR, circles represent samples from independent mice.
Figure 5. High salt diet strongly alters the fecal metabolite profile.
Male FVB/N mice (n = 8) were fed a NSD and then switched to HSD. Metabolites were extracted from fecal pellets taken with NSD (day –3) and HSD (day 13), and analyzed by GC–MS.
A. HSD samples are clearly distinguishable from NSD samples in a principal component analysis for fecal metabolites.
B. Fecal metabolites clearly cluster by treatment. The majority of fecal metabolites are reduced by HSD. Hierarchically clustered heat map, metabolites shown in alphabetical order. Metabolites were normalized by subtracting the minimum and dividing by the maximum value across all mice.
C. Fecal levels of the nucleoside adenosine were similar in both diets, suggesting that the change in metabolites is not due to a decrease in overall bacterial biomass.
D. HSD leads to a reduction in total metabolite peak intensities in fecal samples. **P < 0.01; two-tailed paired Student’s t-test.
Figure 6. *L. murinus* ameliorates salt-sensitive hypertension and reduces the number of Th17 cells.

Continuous blood pressure recordings in n=7 FVB/N mice.

A. Mean systolic pressure over time.

B. Box plots of systolic and diastolic blood pressures. Box plots show median and IQR, whiskers are 1.5x IQR. ###P-value < 0.001 HSD versus NSD, ***P-value < 0.001 HSD with *L. murinus* versus HSD; linear mixed model.

C. CD4⁺RORγ⁺ siLPL in mice fed a NSD (n = 7), HSD (n = 8) or HSD with *L. murinus* (n = 9).

D-F. CD4⁺IL-17A⁺IFNγ⁻ siLPL (D), cLPL (E) or splenocytes (F) out of the total number of CD4⁺ cells in mice fed a NSD (n=5), HSD (n=6; siLPL, n=7) and HSD with *L. murinus* (n=6; siLPL, n=5). Representative plots per group are shown, the quantification shows the mean±s.e.m., circles represent individual mice. *P < 0.05; **P < 0.01; one-way ANOVA and post hoc Tukey’s test (C, E, F), Kruskal–Wallis and post hoc Dunn’s test (D).
Chapter 3

Microbiome sequencing from wipe samples

3.1 Abstract

Inflammatory bowel disease (IBD) is a chronic disease characterized by relapsing inflammation of the gastrointestinal tract. The standard practice to assess inflammation is endoscopy, which is expensive, invasive and uncomfortable for patients. Fecal markers, like the gut microbiome or fecal calprotectin, hold the potential to largely replace endoscopy to inform clinical care of IBD patients. However, current methods to collect stool are cumbersome and uncomfortable for patients. Here we present a preliminary validation that wipe samples are suitable for continuous monitoring of IBD patients. Wipe samples capture the same 16S rDNA microbiome composition as standard stool samples. Wipe samples could also be used to quantify fecal calprotectin, a clinical marker of gut inflammation, in six IBD patients. An at-home wipe collection kit was developed. Patients can drop this kit in any USPS mailbox instead of having to schedule courier pick-ups. The standard DNA extraction protocol for stool samples was adapted for wipe samples and can be completed by laboratory technicians in the same amount of time. Based on this preliminary study, we recommend that wipe samples are validated in larger cohorts of healthy subjects and IBD patients, with specific focus on the stability of biomarkers during transportation.
3.2 Introduction

Inflammatory bowel disease (IBD), including its two subtypes ulcerative colitis (UC) and Crohn's disease (CD), is a chronic disease characterized by recurring inflammation of the gastrointestinal tract. In the U.S. alone, one million patients suffer UC. The standard practice to assess disease severity is endoscopy with biopsies. This method is expensive, invasive, uncomfortable for patients, and prone to complications. Lack of tools to continuously monitor patients' disease severity and tailor treatment leads to poor disease management and expensive care that requires increased medication and even surgery. These, in turn, account for most of the $13.6B/year burden.

Fecal biomarkers hold the potential to largely replace endoscopy to inform clinical care of IBD patients. For example, fecal calprotectin has been proposed as a non-invasive marker of mucosal inflammation. Calprotectin was first described in 1980. It is a 36.5 kDa calcium-binding hetero-trimer present in the cytoplasm of neutrophils where it comprises about 60% of soluble proteins. It is also secreted extracellularly from stimulated neutrophils and monocytes. Calprotectin has been detected in plasma, urine, cerebrospinal fluid, feces, saliva, synovial fluid and colonic biopsies. Fecal calprotectin is clinically used to diagnose IBD from other gastrointestinal disorders, monitor disease severity and even predict relapse (1, 2). An at-home lateral flow-device to quantify fecal calprotectin has been developed (3) and is making its way into clinical care in European countries.

The gut microbiome has more recently been investigated as a source of non-invasive markers to diagnose IBD and tailor treatment (4). UC and CD share many epidemiologic, immunologic, therapeutic and clinical features, and therefore can be hard to discriminate. Microbial biomarkers to distinguish UC and CD were recently described (5). The microbial signature combined with calprotectin data could help decision-making when the diagnosis is uncertain among CD, UC and IBS.

Beyond diagnosis, continuous monitoring of fecal biomarkers has the potential to assess the effectiveness of treatment to decrease inflammation, and ultimately prevent IBD flares and disease complications. This would help patients stay out of the hospital, and overall contribute to the understanding of the disease. However, current methods to collect stool require patients to produce samples directly into a clean container and to manually sub-sample into a sterile vial. This method is
cumbersome, and therefore not optimal for continuous monitoring of IBD patients. Soiled toilet paper has been investigated as an alternative source of gut microbiome 16S rDNA profiles, but it was found to produce slightly different compositions from control stool samples (6). Here we present preliminary validation that commercial moistened wipes produce the same 16S rDNA microbiome composition as paired stool samples, and potentially the same level of fecal calprotectin too. Further testing in larger healthy and IBD cohorts is needed to confirm these preliminary results.
3.3 Materials and Methods

Collection of paired stool and wipe samples.
Nine paired stool and wipe samples were obtained from three healthy individuals on three separate days. RNAlater was added immediately after collection. Samples were stored at -80°C within five hours of collection. Six paired stool and wipe samples were obtained from three Ulcerative Colitis patients and three Crohn’s disease patients (Table 1). Samples were frozen at -80°C immediately after collection. Collection and testing protocols were approved by Institutional Review Boards at MIT and the Massachusetts General Hospital.

DNA extraction and amplicon based Illumina sequencing of 16S rRNA genes.
RNAlater was removed and samples were washed with phosphate-buffered saline (PBS) buffer twice. DNA was extracted with the Power Soil kit (MO BIO Laboratories Inc.) according to manufacturer’s instructions. Paired-end Illumina sequencing libraries were constructed using a two-step PCR approach targeting 16S rRNA genes previously described (7). All paired-end libraries were multiplexed into one lane and sequenced with paired end 300 bases on each end on the Illumina MiSeq platform.

Processing of 16S rRNA gene sequencing data.
Raw data was processed with an in-house 16S processing pipeline (7). To assign OTUs, we clustered OTUs at 97% similarity using USEARCH (8) and assigned taxonomy to the resulting OTUs with the RDP classifier (9) and a confidence cutoff of 0.5. For each dataset, we removed samples with fewer than 100 reads and OTUs with fewer than 10 reads. We calculated the relative abundance of each OTU by dividing its value by the total reads per sample. We then collapsed OTUs by their RDP assignment up to genus level by summing their respective relative abundances.

Fecal calprotectin measurement with immunoassay.
Fecal calprotectin was measured with an ELISA immunoassay (Buhlmann Labs) on 10 mg of stool samples and small pieces of wipe samples. Stool was dosed with the CALEX cap (Buhlmann Labs). A standard curve was performed and fit to a four-parameter logistic (4 PL) regression curve, as recommended by the vendor. The detection range of the kit is 30 to 1800 μg/g of fecal calprotectin.
Estimation of stool concentration in wipe samples.
Stercobilin in fecal extracts from wipe samples was quantified by measuring absorbance at 489nm, and comparing it to a standard curve.
3.4 Results

**Comparison of 16S rDNA microbiome profiles in paired stool and wipe samples from healthy individuals.**

The microbiome composition of paired stool and wipe samples was compared at the OTU level. We computed the Jensen Shannon Distance (JSD) of paired stool and wipe samples, and contrasted it to the JSD of the same individual on different days, as well as between subjects (Figure 1A). Although the sample size is small, we find that the differences between paired wipe and stool samples do not appear to be any greater than differences between days within an individual (either method) with indications of an even tighter spread. This is contrasted to the differences observed between subjects. When the top ten most abundant bacterial phyla are ranked by order of abundance from an arbitrary sample in each of these subjects, the rank is mostly preserved (Figure 1B).

**Fecal calprotectin measurement and normalization in wipe samples from IBD patients.**

Fecal calprotectin (µg/g) was measured in paired stool and wipe samples from six IBD patients (Figure 2A, left panel). As expected, paired samples did not produce the same result ($R^2 = 0.39$, Figure 2A, right panel). The ELISA fecal calprotectin assay requires a pre-weighted amount of stool (10 mg) to be diluted in the kit’s buffer to a final stool concentration of 2mg/ml. A fecal extract can be obtained from wipe samples by cutting out a piece of soiled wipe and adding kit’s buffer, but the stool concentration is unknown. To produce quantitative fecal calprotectin measurements from wipes, it is necessary to estimate the concentration of the fecal extract. Stercobilin is a tetrapyrrolic bile pigment product of heme metabolism that gives stool its brown color. This molecule can be measured spectrophotometrically at 489 nm (10). A standard curve to match absorbance at 489nm with stool concentration was prepared with fecal samples from healthy individuals (Figure 2C). A linear relation between absorbance at 489nm and stool concentration was observed in the range 0.2 to 20 mg/ml ($R^2 = 0.99$). The 489nm absorbance measured in fecal extracts from wipe samples were all included in the range of the standard curve. Fecal calprotectin concentrations were corrected based on the standard curve (Figure 2B, left panel). After correction, paired stool and wipe samples produced the same fecal calprotectin concentration ($R^2 = 0.97$, Figure 2B, right panel).
Development of at-home wipe collection kit

With the support of the Environmental Health Services (EHS) Department at MIT, an at-home wipe collection kit was developed. To ensure the stability of DNA for 16S rDNA microbiome sequencing, the kit includes RNAlater buffer that is added to the wipe sample immediately after collection. To comply with shipping regulations for biological materials, the kit contains wipe samples within three layers. And to make collection patient-friendly, the kit can be dropped in any USPS mailbox and reach MIT within 2 days from any location in the United States. The latter feature drastically reduces the cost of sample collection, since it replaces expensive courier services that pick up stool samples from patients’ homes. Appendix A includes the sampling protocol developed for patients.

Development of protocol for high-throughput DNA extraction from wipe samples

Another consideration for substituting standard stool samples for microbiome studies is the time spent on extracting DNA. With the support of the MOC Unit at the Broad Institute, a high-throughput protocol was developed and preliminarily validated. The protocol takes the same amount of time as standard stool samples and is included in Appendix B of this thesis.
3.5 Discussion

Fecal biomarkers have the potential to make IBD patient treatment more personalized to improve quality of life and reduce healthcare costs. Fecal calprotectin has been extensively studied as an alternative to endoscopy to assess gut inflammation, and it is slowly making its way into clinical practice. Gut microbiome markers have been described more recently to improve diagnosis. Combining fecal calprotectin with microbial biomarker could enhance decision-making.

To enable adoption of fecal biomarker testing by patients, further technology development is needed. Here I focused on validating wipe samples as a patient-friendly alternative to standard stool samples. Preliminary results suggest that wipe samples can provide the same 16S rDNA microbial composition and the same fecal calprotectin levels as standard stool samples. Logistically, wipe samples also present the advantage of allowing patients to drop off their samples at any USPS mailbox instead of having to schedule courier pick-ups. From the laboratory perspective, wipe samples take as much time to process as regular stool samples.

Further validation is required in larger cohorts of healthy individuals and IBD patients. In particular, the effect of RNAlater buffer on the stability and detection of fecal calprotectin needs to be properly tested. RNAlater is a strong base that preserves DNA integrity by denaturing proteins and enzymes. Therefore, addition of RNAlater may protect the DNA material but disrupt fecal calprotectin, or other biomarkers. Immunoassays that recognize unfolded calprotectin may be required to account for RNAlater addition.

Another interesting direction is to expand the list of biomarkers that are compared in paired stool and wipe samples. For example, there may be biomarkers that allow for a better estimation of stool concentration in wipes than the approach implemented here. It is unclear as well what the effect of RNAlater would be on the stability and detection of metabolites.

Based on this preliminary work, we recommend that wipe samples continue being developed as a platform to enhance research and clinical care of IBD patients.
3.6 References


3.7 Tables

Table 1. Characteristics of IBD patients that provided paired and wipe stool samples.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Study ID</th>
<th>Diagnosis</th>
<th>Age</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UC-20B</td>
<td>Ulcerative Colitis</td>
<td>29</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>UC-21B</td>
<td>Ulcerative Colitis</td>
<td>50</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>UC-22B</td>
<td>Ulcerative Colitis</td>
<td>26</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>10037577B</td>
<td>Crohn’s Disease</td>
<td>49</td>
<td>F</td>
</tr>
<tr>
<td>5</td>
<td>10037552B</td>
<td>Crohn’s Disease</td>
<td>19</td>
<td>M</td>
</tr>
<tr>
<td>6</td>
<td>10037639C</td>
<td>Crohn’s Disease</td>
<td>18</td>
<td>F</td>
</tr>
</tbody>
</table>
Figure 1. Comparison of 16S rDNA microbiome profiles in paired stool and wipe samples.
A. Paired stool and wipe samples differ as much as the day-to-day variability in each individual (seven paired samples from 3 individuals), and are more similar than samples from different individuals.
B. Abundance of top 10 bacterial families in paired stool (blue bars) and wipe samples (green bars).
Figure 2. Estimation of stool concentration in wipe samples enables more accurate quantification of fecal calprotectin (µg/g) in wipe samples.

A. Comparison of fecal calprotectin (µg/g) in paired stool and wipe samples (left). Correlation between fecal calprotectin measured in stool and wipe samples, $R^2=0.39$ (right).

B. Comparison of fecal calprotectin (µg/g) in paired stool and normalized wipe samples. Measurement from wipe samples was corrected to the amount of fecal matter in each wipe (left). Correlation between fecal calprotectin in stool and normalized wipe samples, $R^2=0.98$ (right).

C. Stool concentration (mg/ml) and absorbance at 489nm correlate linearly in the range of interest.
Chapter 4

Measurement of population health and activity through wastewater testing

The contents of this chapter have been prepared for submission to PNAS: Matus et al., 2018.

4.1 Abstract

Urbanization is a globally increasing phenomenon. Wastewater has been proposed as a source of data on population health and behavior to inform public health efforts, and has been largely developed as a tool to collect data on illicit drug consumption. Currently, the standard practice is to collect samples from wastewater treatment plants (WWTPs) and apply targeted mass spectrometry methods. Data from WWTPs is not optimal for city officials since it lacks geographic specificity. Furthermore, high-throughput sequencing and metabolomics methods can be leveraged to expand the list of human health biomarkers in wastewater beyond illicit drugs. In this study, we propose sampling from small geographic units representing residential communities in a city to produce geo-localized data relevant to municipal public health departments. Through untargeted microbiome and metabolomics analyses, we show that wastewater from residential catchments is mostly composed of biomarkers derived of human activity including gut microbiome bacteria, urinary and fecal metabolites, and glucuronide compounds of hormones, bile acids and pharmaceutical products. Glucuronide compounds and some gut bacterial families were decreased or absent in wastewater collected at a downstream WWTP suggesting that biomarkers in sewers can be lost faster than they are replenished by additional sources. Together, our results show that residential catchments are ideal geographic units to produce high-quality biomarkers of human activity to build out metrics for municipal public health departments.
4.2 Introduction

Urbanization is a globally increasing phenomenon. By 2050, 70% of the world's population will live in cities where 80% of the global GDP will be generated. Cities are thus prime locations to deliver services that improve the wellbeing of residents. However, cities tend to operate under financial constraints and with limited data to establish priorities and measure the impact of their programs. Wastewater has been proposed as a source of data on population health and behavior that can fill this gap for public health officials (1). For example, methods to measure drug metabolites in wastewater have been developed and standardized over the last decade to provide continuous data on illicit drug consumption at the national- and regional-level in Australia (2) and over 60 European cities and towns (3, 4).

To produce estimates of illicit drug consumption, the standard practice is to collect 24-hour composite samples from wastewater treatment plants (WWTPs) and apply targeted mass spectrometry methods (5–7). While practical and cost-effective to generate national estimates, data from WWTPs is not optimal for city officials since it lacks geographic specificity. And beyond illicit drugs, public health officials are interested in a broad range of human health biomarkers but it is not straightforward to know what markers are stable and quantifiable in wastewater collected from different regions (8).

Here we propose that wastewater testing can be adapted for municipal public health departments by collecting wastewater from curated residential areas in a city. While previous studies have looked at individual buildings (e.g. a college campus (9) or prison (10)), our approach is to select residential catchments that represent areas with >5,000 inhabitants to guarantee anonymity, >80% residential land use to capture mostly resident rather than transient populations, and <60 minutes wastewater travel time to ensure the integrity of biomarkers. Based on these criteria, we selected one residential catchment in the city of Cambridge, Massachusetts and generated untargeted metabolomics and 16S rDNA microbiome profiles from hourly samples collected through one day. A correlation analysis of the wastewater microbiome and metabolome uncovered biomarkers correlated to the diurnal flow rate pattern. Tentatively identified as bacterial and chemical biomarkers derived from human activity, these biomarkers made up most of the wastewater microbiome (up to 80% of Operational Taxonomic Units (OTUs)) and metabolome (up to 90% of metabolites). Several glucuronide compounds of hormones,
bile acids and pharmaceutical products were detected. For comparison, wastewater collected at a downstream WWTP was also tested. 40% of metabolites had differential concentration, with glucuronide compounds being largely reduced or absent in wastewater collected at the WWTP. As glucuronide groups can be enzymatically-cleaved by many gut bacteria and utilized as a carbon source, these results suggest that biomarkers in sewers can be lost to microbial activity faster than they are replenished by additional sources. Therefore, sampling at residential catchments enables the collection of a wider range of biomarkers of human activity including glucuronide compounds which are main byproducts of human metabolism. Together, our results show that residential catchments are ideal geographic units to produce high-quality biomarkers of human activity that can be used to build out metrics for municipal public health departments.
4.3 Materials and Methods

**Selection of Residential Catchment sampling site.**

The sampling site was identified by cross-referencing Cambridge’s wastewater network maps, which contained the layout of pipes and manholes, as well as flow direction, with demographic and land use data. We selected a location where the upstream catchment fulfilled three major requirements: 1) a land use of over 80% residential so that we avoid transient populations and can characterize the demographics of the catchment population accurately, 2) a total catchment population >5,000 people to provide an anonymous reading of the community, and 3) the wastewater travel time from the furthest point in the catchment is <60 minutes to preserve the integrity of the sample.

**Wastewater collection and processing.**

Grab wastewater samples (500 mL) were taken every hour from 10 AM on April 8th 2015 to 10 AM on April 9th 2015 (25 samples total). Samples were collected from the selected manhole with a commercial peristaltic pump (Global Water) sampling at a rate of 100 mL/min. 200 ml of collected sewage were filtered through 10-μm PTFE membrane filters. 150 ml of the outflow were filtered through 0.2-μm PTFE membrane filters. 100 ml of the final outflow were acidified with concentrated HCl (Optima grade, Fisher Scientific) to pH 3.0 and frozen at -80 degrees Celsius until metabolomics analysis. PTFE membrane filters were kept in RNAlater at -80 degrees Celsius until DNA extraction. The lab filtration system consisted of a Masterflex peristaltic pump (Pall), Masterflex PharMed BPT Tubing (Cole-Palmer), 47 mm PFA filter holders (Cole-Palmer) and 47mm PTFE Omnipore filter membranes (Millipore). Both the tubing and filter holders were previously cleaned with HCl (10% v/v) and ultrapure deionized water.

**DNA extraction and amplicon based Illumina sequencing of 16S rRNA genes.**

0.2-μm filter membranes were thawed in ice. RNAlater was removed and filters were washed with phosphate-buffered saline (PBS) buffer twice. DNA was extracted from each filter with Power Water extraction kit (MO BIO Laboratories Inc.) according to manufacturer’s instructions. The only modification was that DNA from up to three filters from the same sample was pooled into the same DNA-binding column. Paired-end Illumina sequencing libraries were constructed using a two-step PCR approach targeting 16S rRNA genes previously described (11). All paired-end libraries were
multiplexed into one lane and sequenced with paired end 300 bases on each end on the Illumina MiSeq platform at the MIT Biomicro Center.

**Processing of 16S rRNA gene sequencing data.**

Raw data was processed with an in-house 16S processing pipeline (11). To assign OTUs, we clustered OTUs at 99% similarity using USEARCH (12) and assigned taxonomy to the resulting OTUs with the RDP classifier (13) and a confidence cutoff of 0.5. For each dataset, we removed samples with fewer than 100 reads and OTUs with fewer than 10 reads. We calculated the relative abundance of each OTU by dividing its value by the total reads per sample. We then collapsed OTUs by their RDP assignment up to genus level by summing their respective relative abundances.

**Untargeted metabolomics analytical methods.**

Acetonitrile and deuterated biotin were added to the 0.2 um-filtrate (acetonitrile final concentration was 5%, deuterated biotin was 0.05 µg ml\(^{-1}\)). The resulting solution was analyzed using liquid chromatography (LC) coupled via electrospray ionization (negative ion mode) to a linear ion trap-7T Fourier-transform ion cyclotron resonance hybrid mass spectrometer (Thermo Scientific, FT-ICR MS; LTQ FT Ultra). Chromatographic separation was performed using a Synergi Fusion C18 reversed phase column (2.1 x 150mm, 4um, Phenomenex). Samples were eluted at 0.25 mL/min with the following gradient: an initial hold at 95% A (0.1% formic acid in water) : 5% B (0.1% formic acid in acetonitrile) for 2 minutes, ramp to 65% B from 2 to 20 minutes, ramp to 100% B from 20 to 25 min, and hold at 100% B until 32.5 minutes. The column was re-equilibrated for 7 min between samples with solvent A. The injected sample volume was 20 uL. Full MS data were collected in the FT-ICR cell from m/z 100-1000 at 100,000 resolving power (defined at 400 m/z). In parallel to the FT acquisition, MS/MS scans of the four most intense ions were collected at nominal mass resolution in the ion trap (LTQ). Samples were analyzed in random order with a pooled sampled run every six samples in order to assess instrument variability.

**Untargeted metabolomics raw data processing.**

Raw XCaliber files were converted to mzML files with MSConvert (threshold = 1000 for the negative ion mode files). Peaks were picked using the function xcmsSet from the R package xcms, with the following parameters (negative mode): method = 'centWave', ppm = 2, sntresh = 10, prefilter = (5, 1000), mzCenterFun = 'wMean', integrate = 2, peakwidth = (20, 60), noise = 1000, and mzdiff = -0.005. To align retention times, we used the retcor.obiwarp function with the following parameters:
plottype='deviation', profStep=0.1, distFunc='cor', gapInit=0.3, and gapExtend=0.4. To group peaks from different samples we used the group.density function with the following parameters: minfrac=0, minsamp=1, bw=30, and mzwid=0.001. To integrate areas of missing peaks, we used the fillPeaks function with method='chrom'.

**Mass matching.**

The first step in identification was comparison of exact mass values with masses of metabolites present in METLIN (14). We searched METLIN for 397 of the most abundant features from the data and putatively identified 126 of these based on m/z matches (ppm error <= 2). We also programmatically mapped all metabolite m/z values to the Human Metabolome Database (HMDB) database (15). Untargeted m/z values were first converted to their expected neutral mass, assuming H+ or Cl- ions, and then scanned against the neutral masses of all HMDB compounds, with an error tolerance of 1 ppm. All HMDB compounds within the error tolerance were returned. An m/z with a single hit in HMDB (“hit_type” column = hmdb_name) shows the chemical name of the HMDB entry and its HMDB ID. When an m/z had multiple HMDB hits, the feature was named with the common chemical taxonomic classification (“hit_type” column). The list of HMDB ID hits is shown too. The HMDB database was downloaded in May 2016. The code to reproduce this analysis is available at https://github.com/cduvallet/blast_hmdb.

**Glucuronides annotation through MS2 match**

We attempted to confirm all putative glucuronides (from m/z hits through METLIN and HMDB searching) via matching of MS/MS data. MS/MS spectra were extracted from individual mzML files (via MZMine (16)) and matched to predicted in silico fragmentation patterns from MetFrag (17). We considered a glucuronide to be confirmed in this way if the expected glucuronide was among the top MetFrag hits and the observed fragments included peaks corresponding to the un-glucuronidated compound, diagnostic glucuronide derivatives (m/z = 113, 157, 175), or other diagnostic peaks (parent compound minus a CO2 and/or OH).

**Metabolite standards confirmation**

We were able to confirm the putative identifications of several of the metabolites in Table 2 (‘standard’) at the most confident level using authentic standards analyzed with the LC-MS method described here. Each standard was analyzed individually in pure solvent as well as spiked into the 1 pm and 2 am samples. The expected m/z and retention time for each standard was determined from
the standard-only runs and confirmed by identifying the corresponding m/z and retention time in the spike-in samples.
4.4 Results

**Selection of Residential Catchment through analysis of wastewater network, census and land use GIS datasets.**

We selected a sampling site to represent a community with at least 80% residential land use, a population size of over 5,000 people, and a sewage travel time of less than 60 minutes. These criteria can be used to curate Residential Catchments that represent residential areas in a city (Figure 1A). Besides geographical specificity, Residential Catchments mitigate confounding factors imposed by wastewater infrastructure. The proposed selection criteria naturally produce catchments with approximately the same population size (5,000-10,000 inhabitants) instead of the large range observed (1,000 – 2,000,000 inhabitants) in WWTPs (18). To some extent, this mitigates the need to correct for unequal population size when comparing different geographic locations. Data from Residential Catchments also presents a more even representation of all households in the basin compared to WWTPs since sewage travel time from every source is short (in minutes) and narrowly-distributed (Figure 1A). From a usability perspective, data from Residential Catchments is more actionable by municipal public health departments because it mostly captures resident rather than transient populations; it reflects the city's spatial heterogeneity; and it enables the design of public health interventions within their city.

**Correlation analysis of wastewater microbiome and metabolome sampled hourly over 24 hours identifies bacteria and metabolites derived from human activity.**

To characterize the type of biomarkers available at this geographic resolution, we set to identify bacteria and metabolites sourced from human activity. We collected hourly samples over 24 hours from the Residential Catchment and produced 16S rDNA sequencing and untargeted metabolomics profiles (Figure 1B). Metabolic features (n=3,672) were co-clustered with microbiome OTUs (n=254) based on Spearman correlation coefficients (Figure 2). Metabolic features grouped into two main clusters (M1 and M2), while bacterial OTUs produced three groups (O1 and O2 are highlighted). The M1 metabolite cluster (n=2,815) had strong positive correlation with the O1 bacterial cluster (n=41) and negative correlation to the O2 bacterial cluster (n=84). Metabolites in cluster M1 were tentatively identified as being derived from human activity (Figure 3A). These metabolites are the most abundant in the metabolome during the daytime (90% of total ion count) but drastically decrease at night, following the same diurnal flow rate pattern. The cluster M1 was statistically enriched in matches to
human metabolites, compared to the M2 cluster (Table 1) and includes known urinary and fecal metabolites. Twenty-two metabolic features in M1 were identified as glucuronide compounds of hormones, bile acids and pharmaceuticals. This was an unexpected finding because previous studies have not been able to measure them in wastewater (19). Cluster M2 had metabolites with constant abundance through day and nighttime and therefore is likely to represent chemicals derived from the environment or sewer biochemistry.

Bacteria in the cluster O1 were putatively identified as coming from human gut microbiomes. Cluster O1 included 41 bacterial OTUs primarily from the Firmicutes and Bacteroidetes phyla that composed 40-80% of the wastewater microbiome. Similar to what was observed in the metabolites, these bacteria had a dip in relative abundance during the nighttime (Figure 3B). Cluster O2 included 84 bacterial OTUs, mainly from the Proteobacteria phylum, which made up to 30% of the wastewater microbiome and increased in relative abundance at night. Furthermore, the top 10 most abundant bacterial families in human stool (Human Microbiome Project) make 86% of the community in the O1 group and only 0.05% of the O2 group (Figure 3C).

**Comparison of wastewater metabolome and microbiome sampled at a Residential Catchment and a downstream WWTP.**

The metabolome and microbiome analyses showed that most of the metabolites and bacteria in residential wastewater are derived from human activity. Unexpected were the detection of several glucuronide compounds, which have been hard to detect in previous studies (19), as well as a high proportion of human gut bacteria. Newton et al. (18) sequenced the wastewater microbiome of samples collected from over 70 WWTPs across the U.S. They found that on average 15% of sequences in sewage have fecal origin, while we found 40-80% depending on the time of day. Therefore, we proceeded to compare the wastewater metabolome and microbiome of samples collected at the selected Residential Catchment and a downstream WWTP (Figure 4A and Table 2). Three grab samples were collected on the same day at noon and processed as before. Untargeted metabolomics data showed that 40% of detected metabolites (n=284/710) had differential concentration in both sample types (Figure 4B). Of those, 80% (n=203) of metabolites decreased or disappeared in wastewater collected at the WWTP, including many metabolites from human activity (Figure 4B, yellow points).
The glucuronide compounds were significantly reduced or altogether absent in wastewater collected at the WWTP. Alpha-phenylacetyl glutamine (APG), the most abundant urinary metabolite detected in this study, was also significantly affected. In contrast, several bile acids, urobilinogen and sucralose had approximately the same concentration (Figure 4C). Many gut bacteria carry glucuronidase enzymes that can release the glucuronide group to use as carbon source (20). APG can also be enzymatically cleaved to produce carbon and nitrogen. These results indicate that ~60% of metabolites have balanced input/loss to the sewer system and 40% have faster loss than replenishment. Metabolites that are absent in wastewater from WWTPs are likely metabolized by bacteria in sewers. The wastewater microbiome composition also changed (Figure 4D). There were overall less bacteria of fecal origin, with the Ruminococcaceae, Lachnospiraceae, and Veillonellaceae families being most affected. Together, these results indicate that wastewater from residential catchments which has been traveling in sewers for less than one hour still contains many human biomarkers that are absent in wastewater collected at WWTPs. Therefore, sampling at residential catchments extends the possible biomarkers that can be monitored through wastewater epidemiology.
4.5 Discussion

This is the first study that adapts wastewater-based epidemiology methods to be relevant to city-level officials, and the first one to leverage untargeted metabolomics and 16S rDNA microbiome sequencing to examine the breadth of human biomarkers available in residential wastewater and how it changes with travel time through sewers.

WWTPs are easily accessible and cost-effective, but impose barriers to data interpretation. Depending on the treatment plant, the wastewater can be a combination of waste from residential, commercial, and industrial buildings, and can represent one or multiple municipalities. If the interest is to compare several cities, collection from different wastewater treatment plants or other downstream sites may introduce confounding factors, such as: representation of different population sizes (21), large variation in wastewater travel time from different sources, different rates of in-sewer degradation of chemical (22) and bacterial biomarkers, and collection with ad-hoc sampling equipment that prevents researchers from optimizing collection parameters to obtain representative samples (23, 24).

Data collected from curated residential catchments has geographic specificity, mitigates confounding factors imposed by the wastewater infrastructure, and is mostly composed of human-derived bacterial and chemical biomarkers, including glucuronide compounds and gut bacteria that are rapidly lost in sewers. Our results suggest that compounds that can be enzymatically cleaved, such as glucuronides and APG, are metabolized by the bacterial community in sewers faster than they are replenished by downstream sources. Similarly, there are some gut bacterial families that decay quickly and end up underrepresented in the microbiome of wastewater collected at WWTPs.

Given increased sampling costs and logistics, residential catchments could be optimized for covering as much of the population with the least number of sites, or for geographic specificity if neighborhood-level resolution is desirable. We propose that sampling campaigns with sub-city resolution should obtain buy-in from local public health officials and community leaders to guarantee that only data that can benefit the community is produced.

To extend this work, further technological development is needed including software to facilitate selection of sampling sites within a city, and hardware that enables wastewater collection from manholes in a cost-effective way.
4.6 References

2. Australian Criminal Intelligence Commission National Wastewater Drug Monitoring Program.
3. SCORE: Sewage Analysis CORe group Europe.


4.7 Tables

**Table 1.** Metabolites in cluster M1 are enriched in metabolites derived from human activity (Fisher's exact test p-value = 2.4e-11).

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster M1</th>
<th>Cluster M2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine metabolites (Level 1*)</td>
<td>300</td>
<td>32</td>
<td>332</td>
</tr>
<tr>
<td>Bile acids (Level 1*)</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sucralose (Level 1*)</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Glucuronides (Level 1 or 2*)</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Human Metabolome DB (Level 3*)</td>
<td>22</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>2,515</td>
<td>825</td>
<td>3,340</td>
</tr>
<tr>
<td>Total</td>
<td>2,815</td>
<td>857</td>
<td>3,672</td>
</tr>
</tbody>
</table>

* Metabolite Standards Association

**Table 2.** Comparison of infrastructure characteristics of the Residential Catchment and WWTP Catchment sampled in this study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Residential Catchment</th>
<th>WWTP Catchment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographic resolution</td>
<td>1 residential area</td>
<td>18 municipalities</td>
</tr>
<tr>
<td>Population served</td>
<td>approx. 5,300 people</td>
<td>approx. 642,000 people</td>
</tr>
<tr>
<td>Land use</td>
<td>84% residential</td>
<td>44% residential</td>
</tr>
<tr>
<td>Mean sewage travel time</td>
<td>25 minutes</td>
<td>4 hours</td>
</tr>
<tr>
<td>Max. sewage travel time</td>
<td>45 minutes</td>
<td>10 hours</td>
</tr>
<tr>
<td>Average flow rate</td>
<td>0.4 MGD (all days)</td>
<td>94.4 MGD (all days)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>81.2 MGD (dry days)</td>
</tr>
<tr>
<td>Sewer type</td>
<td>100% Gravity-fed.</td>
<td>Mostly gravity-fed. Several upstream subarea pump stations.</td>
</tr>
</tbody>
</table>
4.8 Figures

Figure 1. Method to select residential catchments within a city, and experimental design of 24-hour time series study.

A. Residential Catchments are selected by integrating GIS datasets to find basins that meet land use, population size, and wastewater travel time criteria. The map shows the residential catchment in the city of Cambridge, MA that was sampled in this study, as well as its land use and wastewater travel time distribution.

B. Experimental design of the 24-hour time series study. The sample processing steps (dotted line) were also applied to samples collected in the WWTP.
Figure 2. Correlation analysis of wastewater metabolome and microbiome sampled from one location over 24 hours.

Metabolites and bacterial OTUs were clustered based on Spearman correlation coefficients calculated from 20 samples. Metabolites form two main clusters (M1, yellow bar; M2, gray bar). Metabolites in the M1 cluster correlate with bacterial OTUs in the O1 cluster (orange bar) and negatively-correlate with OTUs in the O2 cluster (dark gray bar).
Figure 3. Identification of human-derived metabolites and bacteria in wastewater through analysis of diurnal dynamics of the metabolome and microbiome.

A. Diurnal dynamics of metabolites in clusters M1 (yellow) and M2 (gray) identified in Figure 2 (top panel), and of known urinary (middle panel) and fecal metabolites included in M1 (bottom panel).

B. Diurnal dynamics of bacterial OTUs in clusters O1 (orange) and O2 (dark gray) identified in Figure 2 (top panel). Phylum-level plots of OTUs in the O1 cluster (middle panel) and the O2 cluster (bottom panel).

C. Relative abundance of top 10 bacterial families in human stool, the O1 and O2 clusters.
Figure 4. Comparison of wastewater metabolome and microbiome of samples collected at a residential and a WWTP catchment.

A. Sampling sites at the residential catchment and downstream WWTP catchment.

B. Comparison of metabolite mean abundance from three replicates in wastewater collected in the residential catchment (sewer) and the WWTP (left). Volcano plot showing the fold change versus p-value (right). Metabolites from human activity that are significantly different are shown in yellow. The fold-change and p-value boundaries for significance are shown (blue dotted lines).

C. Total ion counts measured in wastewater collected in the residential (red) and WWTP catchments (green). Several glucuronide compounds and urinary alpha-phenylacetyl glutamine significantly decreased in the WWTP samples (**p-value < 0.05). Bile acids, urobilinogen and sucralose did not change.

D. The proportion of bacteria of fecal origin decreased in wastewater samples collected at the WWTP compared to the residential catchment.
Chapter 5

Conclusion

In this thesis, I sought to apply DNA sequencing and metabolomics to fecal samples to further our understanding of human and population health.

In the first project, I analyzed the gut microbiome and metabolome of mice subjected to high salt diet. High salt diet did not affect the overall composition of the gut microbiome, but did affect specific intestinal bacteria. The gut metabolome was drastically altered by high salt diet, with most metabolites being decreased. *Lactobacillus murinus* was consistently decreased by high salt diet. This bacterium was isolated from mice fecal pellets and fed to mice with experimentally-induced salt-sensitive hypertension. Our results show that *L. murinus* ameliorated salt-sensitive hypertension by modulating T,17 cells. This highlights the gut microbiome as a link between diet, the immune responded and cardiovascular health, and poses *Lactobacillus* as a potential therapeutic target. However, future work needs to assess the effect of high salt diet on the human microbiome as there may be an analogous mechanism but it is unclear if it could be driven by *Lactobacillus* or other bacteria.

In the second chapter, I analyzed fecal calprotectin and the gut microbiome composition of paired stool and wipe samples. Wipe samples stored in RNAlater had the same 16S microbiome composition as regular stool samples from healthy individuals. Fecal calprotectin could be quantified in wipe samples to produce the same results as stool samples, only after correcting for the amount of stool present in fecal extracts from wipes. Wipes are compact and can be included in at-home collection kits that can be dropped off in any USPS mailbox. Finally, extracting DNA from wipe samples to do 16S rDNA sequencing takes the same amount of time as regular stool samples. Collecting stool with wipes preserved with RNAlater has the disadvantage of denaturing proteins which may complicate simultaneous measurement of the gut microbiome and fecal calprotectin. It also prevents researchers from isolating live cultures from stool samples. Future work should further validate wipe samples in larger cohorts of healthy individuals and IBD patients; assess the stability and detection of fecal
calprotectin from wipes shipped in RNAlater; and extend the list of fecal biomarkers that can be reliably in wipe samples.

In the third chapter, I analyzed the wastewater microbiome and metabolome of samples collected hourly over one day from a residential community. Bacteria and metabolites derived from human activity were identified by doing a correlation analysis of the microbiome and metabolome. Our results show that human-derived bacteria and metabolites dominated the composition of wastewater collected from a residential community, and included glucuronide compounds. Glucuronidation is a major metabolic strategy to clear out xenobiotics from the human body and therefore is a direct readout of consumption. This was the first study to find an abundance of glucuronides in wastewater, and we showed that those compounds indeed disappear if wastewater is collected after traveling to wastewater treatment plants. These results suggest that testing wastewater from geo-localized residential areas has the potential to accelerate the discovery and validation of human biomarkers of population health. The caveats of this approach are increased logistical and analytical cost, the need for better sampling equipment that researchers can use at manholes, and the need to partner with local municipalities to gain access into the infrastructure and buy-in from the community. Future studies could develop software and hardware tools to facilitate sampling from manholes, as well as extend the approach to varied locations to ensure the replicability of the system.

This thesis highlights the versatility of stool analysis to provide an understanding of human and population health, the impact of high-throughput technologies on extending our analytical capabilities, and the complexity of making fecal biomarkers accessible to clinical care and public health practice. This research integrates into the broader fields of the human gut microbiome, metabolomics and wastewater epidemiology. Further technological advancement in those fields and integration have the potential to create seamless monitoring systems for clinicians managing patients, as well as local public health officials creating policies to increase community wellbeing.
Appendix A

Experimental protocols developed for Chapter 3

Instructions for patients to collect wipe samples at-home

Important note about sample collection and shipping: It is important that you mail your sample within 24 hours of collecting it. To ensure your sample is received quickly, please mail your sample on a MONDAY, TUESDAY, OR WEDNESDAY.

Please contact the research coordinator if you have any questions about when or how to collect your sample, or how to mail it back.

Part 1 - Check your kit

Your kit should contain:
1 Ziploc bag
1 package of pre-moistened wipes
1 pair of gloves (optional to wear)
1 tube of clear liquid (RNAlater liquid buffer)
1 biohazard bag with a neon sticker, white tissue inside, and packing list outside
1 mailing box
1 piece of tape

Part 2 - Get Ready

1. Wash your hands with soap and water, then rinse. Dry your hands. If you want, put on the gloves.
2. Open the Ziploc bag at the top, and place it in a convenient, easy to reach location.
3. Open the pack of wipes, and have a wipe ready to grab.
Part 3 - Collect your sample
1. Poop into the toilet as usual.
2. When you’re done, wipe your bottom using one of the wipes. Do not put the wipe down. Place the wipe directly into the Ziploc bag. This is your sample.
3. Finish wiping with more wipes or with toilet paper these wipes and paper can be thrown away. Do not put them into the Ziploc bag.
4. Wash your hands with soap and water, then rinse. Dry your hands.

Part 4 – Pack up your sample
1. Pick up the Ziploc bag containing the wipe. Add ALL of the contents from the tube containing RNAlater liquid buffer into the Ziploc bag with the wipe. Throw away the empty tube.
2. Close the Ziploc bag tightly, pushing out most of the air.

Part 5 – Ship your sample
Remember, it’s important to ship your sample within 24 hours of collection.
1. Find the biohazard bag with the neon sticker. On the neon sticker, use a pen to write the date (MM/DD/YYYY) you collected the sample.
2. Place the Ziploc bag with your wipe into the biohazard bag. Do not remove the absorbent material (white tissues) from inside the bag or the packing list outside the bag.
3. Seal the biohazard bag by pulling off the white strip to expose the sticky strip. Fold the top of the bag over, so that the sticky strip sticks to the bag. Press firmly to seal the bag.
4. Put the sealed biohazard plastic bag into the cardboard box and close the box. Use the tape provided to secure the box lid.
5. Drop off the package at any USPS mailbox within 24 hours of sample collection

Thank you for participating in this study!
DNA EXTRACTION FROM WIPE SAMPLES

Developed by MM, TG, TP and the MOC

Day 1. Remove RNAlater and cut out piece of wipe.

Materials
- Biosafety cabinet level 2
- Personal protective equipment
- Dry heat sterilizer
- Flat surface
- Scissors
- 2 pairs of tweezers
- 2 96-well plates (2 ml capacity). Power Soil 96-well Bead Plate if doing HT format.
- Freshly prepared bleach 20%
- Flask to collect RNAlater waste

Procedure
1. Thaw wipes on the bench.
2. With sterile scissors, cut a lower corner of the Ziploc bag. Squeeze out RNAlater and pour into waste container.
3. Carefully take out wipe from bag with sterile tweezers and lay on sterile flat surface. Discard Ziploc bag.
4. With sterile tweezers, fold wipe such that soiled area is positioned in a corner.
5. With sterile scissors, cut out 1x1 inch square from wipe by positioning scissors on the corner containing soiled area. Cut square into half. Discard leftover wipe.
6. With sterile tweezers, transfer each piece of wipe to a separate 96-well plate (2ml capacity). Use tweezers to roll the wipe samples so they enter nicely into the plate wells. If you are using the Power Soil high-throughput format, transfer one piece of wipe to the Power Soil beads plate. Keep the rubber plate cover as you’ll need it for the DNA extraction protocol.
7. Wipe tweezers and scissors with 20% bleach. Dry with paper towel (to avoid stickiness to glass beads) and insert into dry heat sterilizer. Let sit for 30 seconds to 1 minute.
8. Wipe flat surface with 20% bleach. Dry with paper towel. Lay down scissors and tweezers. You are ready for the next sample.
9. Store one 96-well plate at -80°C for back-up. Store the other plate overnight at 4°C and continue with the protocol the next day.

10. Discard RNAlater in the sink. RNAlater cannot be mixed with bleach.

**Day 2. Wash wipe pieces with PBS and extract DNA.**

*Materials*
- Biosafety cabinet level 2
- Personal protective equipment
- PBS buffer
- Multichannel pipette and tips (1-ml capacity)
- Plastic reservoirs

*Procedure*
1. Get 96-well plate with samples out of fridge.
2. Add PBS 1X to plastic reservoir.
3. With multi-channel pipette, add 1 mL PBS per well. Before dispensing the liquid, press the wipe sample to the bottom of the well to avoid overflowing. Carefully pipette up and down to mix, then remove all liquid by pressing with the pipette tip on the wipe sample at the bottom of the well. Discard PBS into a plastic waste reservoir.
4. Repeat wash (step 3) once more.
5. You are ready to extract DNA with the Power Soil kit.

**Power Soil DNA extraction**

If using the Tube format:
- With sterile tweezers (use bleach 20% and dry heat sterilizer), transfer wipe squares to the Power soil bead tubes, and continue the protocol according to the Alm lab’s instructions.

If using the High-throughput (96-well) format:
- Wipe samples are already in the Beads Plate. Continue the protocol according to the Alm lab’s instructions.
- *Note:* Use 2-ml collection plates for all steps (not 1 ml) otherwise the volume does not fit.

Note: QIAGEN has recently acquired MO-BIO. The new vendor may have different instructions.

## Power Soil DNA extraction protocol (tube format)

### Materials
- MOBIO Power Soil DNA extraction kit
- Biosafety cabinet level 2
- Microcentrifuge
- Bench top vortex
- MOBIO Vortex adapter
- Two water baths or heat blocks (one at 65°C and one at 95°C)
- Micropipettes and pipette tips
- Proteinase K solution
- Ice

### Procedure *(Alm lab modifications in blue font)*

1. All wipe samples must already be contained in the Power Soil bead tubes.

2. Check **Solution C1**. If Solution C1 is precipitated, heat it to 60°C until dissolved before use.

3. Add 60 μL of **Solution C1** and invert several times or vortex briefly.

4. **Add 20 μL of Proteinase K** solution and invert several times or vortex briefly.

5. Incubate samples at 65°C for 10 minutes. Invert the tubes to mix a few times during the incubation.

6. Secure tubes horizontally using the MOBIO Vortex Adapter tube holder for the vortex.

7. Vortex at maximum speed (3000 rpm) for 10 minutes.
8. Incubate samples at 95°C for 10 minutes. Invert the tubes to mix a few times during the incubation.

9. Centrifuge tubes at 10,000 xg for 30 seconds at room temperature. Note: Be sure not to exceed 10,000 xg (10,000 rcf) or tubes may break.

10. Transfer 600 µL of supernatant to a clean 2 mL Collection tube (provided). Supernatant may still contain some particles.

11. Add 250 µL of Solution C2 and vortex for 5 seconds. Incubate at 4°C (ice) for 5 minutes.

12. Centrifuge tubes at 10,000 xg for 1 minute at room temperature.

13. Avoiding the pellet, transfer up to 600 µL of supernatant to a clean 2 mL Collection tube (provided).

14. Add 200 µL of Solution C3 and invert to mix. Incubate at 4°C (ice) for 5 minutes.

15. Centrifuge the tubes at 10,000 xg for 2 minutes at room temperature.

16. Avoiding the pellet, transfer up to 750 µL of supernatant to a clean 2 mL Collection tube (provided).

17. Shake Solution C4 to mix. Add 1200 µL (2 x 600 µL) of Solution C4 to the supernatant. Vortex briefly or invert to mix. Note: Solution C4 contains guanidine hydrochloride and cannot be mixed with bleach.

18. Load 650 µL onto a Spin Filter and centrifuge at 10,000 xg for 1 minute at room temperature. Discard the flow through.

19. Repeat step 18 two more times until all sample has been processed.

20. Add 500 µL of Solution C5. Centrifuge at 10,000 xg for 30 seconds at room temperature. Discard the flow through.

21. Centrifuge again at 10,000 xg for 1 minute at room temperature.

22. Carefully place Spin Filter in a clean and labeled 2 mL Collection Tube. Avoid splashing any Solution C5 onto the Spin Filter.
23. Let Spin Filters air dry for 5 minutes at room temperature.

24. Add 100 uL of Solution C6 to the center of the white filter membrane.

25. Let tubes sit for 5 minutes at room temperature.

26. Centrifuge at 10,000 xg for 30 seconds at room temperature.

27. Discard the Spin Filter. DNA is in the tube now ready for any downstream application.

**Power Soil DNA extraction protocol (plate format)**

**Materials**
- Power Soil HT kit, adapter plates, plate shaker
- Bench centrifuge that fits 96-well plates
- Micropipettes, multichannel pipettes, micropipette tips
- Plastic reservoirs
- Proteinase K solution
- Dry oven or shaker at 65°C and 95°C

**Procedure (Alm lab modifications in blue font)**

1. **BEFORE THE FIRST USE ONLY,** Solution C5-D must be prepared. Add an equal volume of 100% Ethanol to Solution C5-D (for the 4 prep kit = 120 ml, or for the 12 prep kit = 360 ml). Mix well. Put a check mark in the “ethanol added” box on the bottle cap label.

2. **All samples must be in the Bead Plate.** Take Bead Plate out of 4°C fridge.

3. Add 750 uL of Bead Solution to the wells of the Bead Plate. Before dispensing liquid, press wipe samples to the bottom of wells to avoid overflowing.

4. **Check Solution C1.** If Solution C1 has precipitated, heat the solution at 60°C until the precipitate has dissolved. Mix gently before using.

5. Add 60 uL of Solution C1.

6. **Add 20 uL of Proteinase K.** Secure the Square Well Mat tightly to the Bead Plate. A proper seal is critical to prevent loss of sample and leakage that might cause contamination or damage the shaker.
7. Incubate samples at 65°C for 10 minutes.

8. Place Bead Plate with mat securely fastened between 2 adapter plates (MO BIO Catalog# 11990) and place on the 96 Well Plate Shaker (MO BIO Catalog# 11996). Reference the protocol provided with the adapter plates for proper placement.

9. Shake at speed 20 for 10 minutes, remove plates and re-orient them so that the side closest to the machine body is now furthest from the machine body and shake again at speed 20 for 10 minutes.

10. Incubate samples at 95°C for 10 minutes.

11. Centrifuge at room temperature for 6 minutes at 4500 x g.

12. Remove the Bead Plate from the centrifuge. Carefully and without splashing remove and discard the Square Well Mat and transfer the supernatant to a clean 2 mL Collection
Plate. Note: Don’t use 1 mL Collection Plates because they overflow. The supernatant may still contain some particles.

13. Add 250 µl of **Solution C2** to each well and apply **Sealing Tape** to 2 ml Collection Plate. Vortex for 5 seconds and incubate at 4°C for 10 minutes. Centrifuge the 2 ml Collection Plate at room temperature for 6 minutes at 4500 x g. Remove and discard Sealing Tape.

14. Avoiding the pellet, transfer entire volume of supernatant to a new 2 mL **Collection Plate**.

15. Apply new **Sealing Tape** to the 2 ml Collection Plate and centrifuge at room temperature for 6 minutes at 4500 x g. Transfer entire volume of supernatant to a new 2 ml Collection Plate.

16. Add 200 µl of **Solution C3** and apply Sealing Tape to the 2 ml Collection Plate. Vortex for 5 seconds and incubate at 4°C for 10 minutes. Centrifuge at room temperature for 6 minutes at 4500 x g. Remove and discard Sealing Tape.

17. Avoiding the pellet, transfer entire volume of supernatant to a new 2 mL **Collection Plate**.

18. Apply new **Sealing Tape** to 2 ml Collection Plate. Centrifuge the 2 ml Collection Plate at room temperature for 6 minutes at 4500 x g.

19. Transfer no more than 650 µl of supernatant to a 2 mL **Collection Plate** avoiding any residual pellet.

20. Add 650 µl of **Solution C4** to each well of the 2 ml Collection Plate.

21. Add a second 650 µl of Solution C4 to each well of the 2 ml Collection Plate. Note: It is safe to stop the protocol at this step and store the samples covered with Sealing Tape (user supplied) at 4°C.

22. Pipet samples ‘up and down’ to mix.
23. Place **Spin Plate** onto a new 0.5 mL Collection Plate.

24. Load approximately 650 uL into each well of the **Spin Plate** and apply **Centrifuge Tape**.

25. Centrifuge at room temperature for 3 minutes at 4500 x g. Discard the flow through and place the Spin Plate back on the same 0.5 ml Collection Plate. Discard the Centrifuge Tape.

26. Repeat steps 20-21 until all the supernatant has been processed. Discard the final flow through.

27. Place the Spin Plate back on the same 0.5 ml Collection Plate.

28. **Confirm that ethanol has been added to Solution C5-D (see step 1).** Add 500 µl of **Solution C5-D** to each well of the Spin Plate. Apply Centrifuge Tape to the Spin Plate.

29. Centrifuge at room temperature for 3 minutes at 4500 x g. Discard the flow through and place the same 0.5 ml Collection Plate beneath the Spin Plate.

30. Centrifuge again at room temperature for 5 minutes at 4500 x g. Discard the flow through.

31. Carefully place the Spin Plate onto a **Microplate**. Remove Centrifuge Tape and discard.

32. Allow to air dry for 10 minutes at room temperature.

33. Add 100 µl of **Solution C6** to the center of each well of the Spin Plate. Apply **Centrifuge Tape**.

34. Centrifuge at room temperature for 3 minutes at 4500 x g. Remove Centrifuge Tape and discard.

35. Cover wells of Microplate with the **Elution Sealing Mat** provided. DNA is now ready for any downstream application. No further steps are required.
Buhlmann Fecal Calprotectin ELISA
Last updated by MM on 3/27/17

MATERIALS AND EQUIPMENT

- Buhlmann Fecal calprotectin ELISA 96-well format kit
- Biosafety cabinet
- 96-well plate reader (450 nm wavelength)
- Scale with mg resolution
- Small vortex inside biosafety cabinet
- Benchtop vortex with vortex adapter for 2 ml Eppendorf tubes
- Microcentrifuge for 2 ml Eppendorf tubes
- 15-ml conical tubes
- Disposable inoculation loops
- 2-ml Eppendorf tubes
- Benchtop plate rotator (450 rpm speed)
- 96-well plastic plate covers
- 96-well aluminum plate covers
- Multi-channel pipette and tips (100 μL – 300 μL range)
- Plastic reservoirs
- Paper towels

PREPARE SAMPLES AND REAGENTS

1. Thaw stool samples on ice for at least 2.5 hours. If stool samples are stored in RNAlater and will also be sequenced: Wash stool samples with PBS twice. Keep two 100 – 200 mg aliquots of stool for DNA sequencing in their respective tubes.

Note: Keep the number of freeze-thaw cycles consistent throughout a study. The more cycles you have, the more human cells are broken, and the higher fecal calprotectin levels you will measure.
2. Bring ELISA kit reagents to room temperature before starting, EXCEPT the 96-well plate and tubes with standards, high control and low control. Those stay at 4 degrees until you use them.

3. Draw plate map. Determine how many strips of the 96-well you will need.

   Note: Each sample and standard must be measured in duplicate. Also reserve three wells for the low control, high control and blank.


**STOOL SAMPLE EXTRACTION**

1. Label and weigh empty 15-ml conical tubes together with a disposable inoculation loop. Record weights in a paper sheet.

2. Take out 0.05 to 0.1 g of stool sample (50 to 100 mg) by means of the inoculation loop and place it into the pre-weighted tube.


4. Calculate grams of stool by subtracting weights for each tube pre- and post-stool.

5. Add **Extraction Buffer** according to the formula:

   \[ x \text{ g stool} \times 49,000 = y \text{ uL extraction buffer} \]

   For example: If you weigh 0.05 g stool, add 2450 uL of Extraction Buffer.

6. Swirl the inoculation loop to wash off any stool and discard it. Homogenize sample by vortexing for 30 seconds. Transfer 2 ml of sample to an Eppendorf tube.

   Note: This is the only modification to the manufacturer's protocol. We transfer 2 ml to an Eppendorf tube and then vortex for 30 minutes because we don't have a vortex-adapter for 15 ml conical tubes.
7. Secure tubes to vortex adapter and vortex at maximum speed for 30 minutes.
8. Centrifuge tubes at 3000xg for 5 minutes.
9. Transfer the supernatant into a fresh, labeled tube.

Note: Extracts are stable for 6 days at 2-8 degrees Celsius, or for 18 months at -20C. It could be a good idea to break here and continue with dilution and assay the next day. However, any storage condition must be kept consistent throughout a study.

SAMPLE DILUTION

HEALTHY samples:
1. Dilute stool extracts 1:50 with Incubation Buffer and mix well. For example:

   \[ 20 \, \text{uL extract} + 980 \, \text{uL incubation buffer} \]

2. Let the samples equilibrate for at least 5 minutes at room temperature.

IBD samples:
1. Dilute stool extracts 1:150 with Incubation Buffer and mix well. For example:

   \[ 10 \, \text{uL extract} + 1490 \, \text{uL incubation buffer (pipette 745 uL twice)} \]

2. Let the samples equilibrate for at least 5 minutes at room temperature.

ELISA ASSAY
1. Prepare a plate with sufficient strips to test the required number of standards, controls and diluted samples. Remove excess strips from the holder and re-seal them in the foil pouch together with the desiccant packs without delay. Store refrigerated.
Note: Each sample and standard must be measured in duplicate. Also reserve three wells for the low control, high control and blank.

2. Wash wells twice with 300 uL of Wash Buffer per well. Wait for 20 seconds. Empty wells at the sink and tap plate firmly onto paper towels.

3. Pipet 100 uL of Standards A-E, low control, high control and blank in their respective wells.

4. Vortex each diluted sample for 5 seconds and pipet 100 uL into their respective wells.

5. Cover the plate with a plate sealer and incubate for 30 minutes (+ 5 min max) on a plate rotator set at 450 rpm at room temperature.

6. Carefully remove and discard the plate sealer. Empty wells at the sink and tap plate firmly onto paper towels.

7. Wash three times with 300 uL of Wash Buffer per well. Wait for 20 seconds for each wash. Empty wells at the sink and tap plate firmly onto paper towels.

8. Pipet 100 uL of Enzyme Label to each well.

9. Cover the plate with a plate sealer and incubate for 30 (+ 5 min max) on a plate rotator set at 450 rpm at room temperature.

10. Carefully remove and discard the plate sealer. Empty wells at the sink and tap plate firmly onto paper towels.

11. Wash five times with 300 uL of Wash Buffer per well. Wait for 20 seconds for each wash. Empty wells at the sink and tap plate firmly onto paper towels.

12. Pipet 100 uL of the TMB Substrate Solution to all wells.

13. Cover the plate with an aluminum plate sealer. Protect the plate from direct light and incubate for 15 min (+ 2 min max) on a plate rotator set at 450 rpm at room temperature.

14. Pipet 100 uL of Stop Solution to all wells. Remove air bubbles with a pipette tip.

15. Read absorbance at 450 nm within 30 minutes. Export CVS file from Excel and save in a USB stick. Name file in the following format:

Date_Study_Initials.cvs

For example: 170327.FLAGSHIP.Longitudinal_TP.xls would correspond to a file from the longitudinal Flagship study run on March 27th, 2017 by Tsultrim Palden.
Appendix B

Experimental protocols developed for Chapter 4

Sewage processing for generating microbiome and metabolome data

Developed by Mariana Matus (Alm Lab), Claire Duvallet (Alm Lab) and Damon Baptista (EHS)

Time required per sample (3 replicates + 2 blanks): 2 hours

Updates: May 23, 2017
October 22, 2016
September 8, 2016
May 9, 2016

NOTES

Excuse yourself from the experiment if you feel unwell.
Dress code: pants, closed shoes and with your hair up.
Bandage any wounds on your hands, arms and face.
Otherwise stated, all work is done in a biosafety cabinet.

MATERIALS AND SUPPLIES

- 20% Bleach in squirt bottle
- 70% Ethanol in spray bottle
- 1 x 500 ml 20% Bleach in bottle for disinfecting filtration line
- 1 x 500 ml 20% Bleach in beaker for disinfecting filter holder
- 1 x 250 ml glass bottle for collecting waste
PTFE 0.2 um 47 mm filter membranes
50 ml Falcon tubes
2 autoclaved tweezers
pH paper range 0-14
pH paper range 0-6
Concentrated HCl (37% v/v) in PTFE container
10 ul, 100 ul, 200 ul pipettes and pipette tips
Absorbent pads
Paper towels
Ice bucket and ice

Materials to be cleaned according to SOP “HOW TO CLEAN MATERIALS FOR MASS SPEC DATA COLLECTION”:

1 x 40 cm piece of Masterflex L/S #16 flexible tubing
1 PTFE filter holder 47 mm
30 ml Polycarbonate bottles (6 per sampling site)
50 ml Sterile diWater in polycarbonate or glass bottle (3 p/ sampling site)
600 ml Sterile diwater in polycarbonate or glass bottle (1 p/ sampling site)

EQUIPMENT

Biosafety cabinet, regularly logged
Chemical fume hood, regularly logged
Minimate peristaltic pump
Dry heat sterilizer
-80C freezer

FILTRATION SCHEDULE FOR TRIPlicate SAMPLES COLLeCTED FROM ONE LOCATION

1. Blank 1 (30 ml sterile diWater)
2. Sewage replicate 1 (30 ml)
Clean line with sterile diWater for 5 minutes (about 100 ml)

3. Sewage replicate 2 (30 ml)
   Clean line with sterile diWater for 5 minutes (about 100 ml)

4. Sewage replicate 3 (30 ml)
   Clean line with sterile diWater for 5 minutes (about 100 ml)

5. Blank 2 (30 ml sterile diWater)

---

**PROCEDURE**

**GEAR UP**

1. Put on personal protective equipment.
   - Lab coat
   - Safety goggles
   - Face shield
   - Two pairs of gloves

2. Disinfect biosafety cabinet surfaces with bleach 20% and ethanol 70%. Dry and start cabinet.

3. Turn on dry heat sterilizer and let it sit until it reaches 220 degrees Celsius. Lay tweezers on sterile surface (inside beaker or empty petri dish).

**0.2-MICRON FILTRATION**

1. Load 0.2-micron filter paper into filter-holder as described below:
   - Put new absorbent pad in biosafety cabinet.
   - Open filter-holder. Make the interior surfaces of the filter-holder face up.
   - Open carefully (and far from edge of biosafety cabinet) one box of 0.2-micron filters.
   - Disinfect tweezers by holding them 1 minute in the dry heat sterilizer.
   - With clean tweezers, grab one 0.2-micron filter, which is the **white paper**. You can discard any blue papers.
1. Put filter inside filter-holder. Make sure it is completely flat on the surface of the holder. Use tweezers to flatten it if necessary.

2. Put tweezers back in dry heat sterilizer.


2. Connect 0.2-micron filter-holder to filtration line in an **upright position**. Connect 'left-to-right' as described below:

   - Load flexible yellow tubing into peristaltic pump.
   - Screw filter-holder into yellow tubing. Make sure to include the small plastic fittings in between the plastic holder and its screws. This will make the line watertight.
   - Hold filter holder on top of collection (filtrate) bottle.

3. Put yellow tubing inlet inside one bottle with sewage (or distilled water for blanks).


5. Check filtration line for any obvious disconnection.

6. Set pump speed to 6 rpm and turn it on. Filter 30 mL of sewage (or distilled water for blanks) and stop pump.

7. Constantly check the line for leaks. If a leak starts, the filter has clogged. Stop the pump and disconnect the filter-holder from the yellow tubing.

8. Retrieve the clogged filter with sterile tweezers and store in a 50 ml falcon tube. Wipe tweezers with bleach 20%, dry with paper towel and put in dry heat sterilizer for 1 minute. Take them out and lay them down on a sterile surface.

9. Repeat steps 1 - 8 until you have filtered 30 ml of sewage (or distilled water for blanks).

10. Keep in ice until acidification and storage at -80 degrees Celsius.
CLEAN FILTRATION LINE IN BETWEEN SAMPLES

1. Connect (empty) filter-holder back into the line.

2. Put yellow tubing outlet in waste collection bottle.

3. Turn on pump to 20 rpm. Circulate sterile distilled water (in a clean polycarbonate or glass bottle) for 5 minutes.

4. Empty line and stop the pump.

5. The system is ready to filter the next sample.

ACIDIFY FILTERED SEWAGE FOR METABOLOMICS

1. Take the 30 mL bottle of filtered sewage to chemical fume hood.

2. Acidify to pH 2.0:
   - Take out a pH strip and lay it flat on an absorbent pad.
   - Pipette out 100 uL of sewage and drop it in the pH strip (Note: Don't dip pH strip into bottle). Record in metadata sheet initial pH.
   - Add 10 uL of concentrated HCl.
   - Close bottle and shake vigorously to mix.
   - Pipette out 100 uL of sewage and drop it in a pH strip (Note: Don't dip pH strip into bottle). Add more acid if necessary and repeat pH measurement until it reads between 2 and 3 (Note: Don't go below pH 2). Record pH and volume of HCl added in each step.
STORE PROCESSED SAMPLES

1. Store filtrate bottle, and the two falcon tubes with filters, at -80C.

2. Make sure you have labeled processed samples recorded all metadata.

DISINFECT MATERIALS

1. Connect (empty) filter-holder back into the line.

2. Put yellow tubing outlet in waste collection bottle.

3. Pass 20% bleach solution for 10 minutes.

4. Put filtration line components in bleach 20% bath. Let sit for 20 minutes.

5. Add bleach 20% to all bottles that contained sewage, blanks or waste. Let sit for 20 minutes.

6. Wash and autoclave all components.
How to clean materials to remove chemical contaminants

Developed by Mariana Matus, Claire Duvallet, Krista Longnecker, Liz Kujawinski

Motivation

Mass spectrometry is a very sensitive technique that will detect all ionizable chemicals in a sample. Plastics are overall very ionizable and therefore a common source of contamination. Chemically-inert materials that don’t bleed plasticizers into a sample include glass, polycarbonate and Teflon (PTFE). Additionally, all materials that make contact with sewage (glass, polycarbonate or Teflon (PTFE)) must be cleaned with soap and acid and autoclaved as described in this SOP. Failure to follow these procedures will result in mass spec data that is purely dominated by the plastics from the collection materials and therefore useless for research. Positive ion mode is especially affected by plastic contamination.

IMPORTANT NOTE 1: START CLEANING YOUR MATERIALS TWO WEEKS BEFORE SAMPLING DAY. THE PROTOCOL TAKES AT LEAST SIX DAYS.

IMPORTANT NOTE 2: ALWAYS WEAR GLOVES AND PPE TO PROTECT YOURSELF AND TO AVOID CONTAMINATING THE MATERIALS.

Supplies

- Concentrated HCl (37% v/v)
- Lab dish soap
- MiliQ water
- 1 x 100 ml glass measuring cylinder
- 4 x 250 ml glass beaker
- 2 x 500 ml glass beaker
- Personal Protective Equipment: lab coat, gloves, safety glasses
- Poly carboy container to safely dispose of HCl 10%. Concentrated HCl cannot be flushed down sink drains. Waste must be collected in a container which is the right material to withstand acid and provided by the Safety Officer.
- Labels resistant to freezing
Equipment
- Chemical fume hood
- Peristaltic pump
- Autoclave

Procedure to clean up bottles
1. Soak all bottles on soapy water overnight. Make sure all interior surfaces are making contact with the soapy water.
2. Discard soapy water in sink. Rinse each bottle with hot tap water three times.
   Short clip: https://goo.gl/photos/gPxn7s9U7ppFQ54D9
3. Let bottles dry on an absorbent pad on your bench.
4. In chemical fume hood and wearing full PPE: Prepare 500 ml of 10% HCl in distilled water (v/v) by using the glass measuring cylinder and glass beaker: IT IS VERY IMPORTANT THAT YOU FIRST POUR 50 ML OF HCL INTO THE BEAKER AND THEN YOU ADD 450 ML OF DISTILLED WATER, THAT IS, ADD WATER TO THE ACID. YOU CANNOT DO IT THE OTHER WAY AROUND (ADD ACID TO WATER) BECAUSE THAT REACTION IS HIGHLY EXOTHERMIC AND THEREFORE UNSAFE. You can store this 10% HCl solution after the cleanup in a glass container and reuse it up to five times. This reduces work and the amount of hazardous waste produced.
5. In chemical fume hood and wearing full PPE: Carefully fill ⅗ of each bottle’s capacity with 10% HCl and cap the bottle.
   Short clip: https://goo.gl/photos/aN2MYXCdJfQuqkBaA
6. Take the bottles back to your bench and lay them down on their side. Incubate for 24 hours so the acid cleans away all chemicals from the internal surfaces. Leave a note indicating: ‘10% HCl solution. CAREFUL. Don’t touch or move.” Also include your name, phone and date so people can contact you if they run into any problem.
   Short clip: https://goo.gl/photos/WQ8VFSa1Q8z929WYA
7. Rotate the bottles on their other side so the acid makes contact with all internal surfaces of the bottle. Incubate for 24 hours. Leave a note indicating: ‘10% HCl solution. CAREFUL. Don’t touch or move.” Also include your name, phone and date so people can contact you if they run into any problem.
8. In chemical fume hood and wearing full PPE: Carefully dump all 10% HCl into the poly
carboy disposal container. If you have a clean glass container, you can collect the 10% HCl
and reuse it for cleanup of other materials up to five times. This reduces work and the amount
of generated waste.

Short clip: https://goo.gl/photos/DokpKCv4zEoSC1VB9

9. Rinse each bottle with deionized water (miliQ) three times. Make sure you shake the bottle
vigorously to remove all acid from the internal surfaces of the bottles.

Short clip: https://goo.gl/photos/EVjYleEu3zSgMYc7

10. Let bottles dry on an absorbent pad on your bench. Make sure nothing touches the interior of
the bottle anymore.

Short clip: https://goo.gl/photos/YfjYVtsk6Qh1rFQz8

11. Proceed to next cleaning steps according to instructions in Table 1. As indicated, you need to
add diWater to some bottles before autoclaving. This water will be used as a blank or to clean
the filtration line in between samples.

**Procedure to clean up filter holder parts and tubing**

1. Put the filter holder parts and the piece of tubing inside a 500 ml glass beaker. Roll the piece
of tubing so it fits.
2. Fill beaker with soapy water, cover with aluminum foil and let it sit overnight on your bench.
3. Discard soapy water in sink. Rinse materials with hot tap water three times.
4. Take each material out of the beaker and let it dry on a clean absorbent pad on your bench.
   Remember to wear gloves so you don’t contaminate materials when you touch them.
5. Put the filter holder parts and piece of tubing back in the 500 ml glass beaker.
6. In chemical fume hood and wearing full PPE: Fill beaker with 10% HCl. Cover the beaker
with aluminum foil and leave it in a corner of the chemical fume hood for 24 hours. The
chemical fume hood is a public space so leave the beaker in a corner where it cannot be spilled
by accident and caution fellow lab workers with a note indicating: ‘10% HCl solution.
CAREFUL. Don’t touch or move.’ Also include your name, phone and date so people can
contact you if they run into any problem.
7. In chemical fume hood and wearing full PPE: Carefully dump all 10% HCl into the poly
carboy disposal container. If you have a clean glass container, you can collect the 10% HCl
and reuse it for cleanup of other materials up to five times. This reduces work and the amount
of generated waste.

8. Rinse materials with miliQ water **three times**.

9. Take each material out of the beaker and let it dry on a clean absorbent pad on your bench.
 Remember to **wear gloves** so you don’t contaminate materials when you touch them.

10. Put all the parts of the filter holder system into an autoclave baggie. Be careful not to lose the
 small round adapter that makes the connection watertight. Autoclave.

11. Now you need to clean the tubing on the inside.

**Procedure to clean up interior surface of tubing**

1. Prepare soapy water in a clean empty 250 ml glass beaker.

2. Setup the peristaltic pump on a bench. Set speed to 30.

3. Load the yellow tubing on the peristaltic pump. Put both ends of the tubing inside the beaker
 with soapy water.

4. Start the pump and circulate soapy water for 5 minutes. Stop the pump.

5. Remove both ends of the tubing from the beaker with soapy water. Empty tubing by starting
 the pump and collecting the outflow in a ‘waste’ beaker (a clean empty 250 ml beaker).

6. Start the pump and rinse the line with diWater for 5 minutes.

7. Put 100 ml of 10% HCl (from the solution you already prepared) in a clean empty 250 ml glass
 beaker.

8. Put both ends of the tubing inside the beaker with 10% HCl.

9. Start the pump and circulate 10% HCl for 5 minutes. Stop the pump.

10. Carefully remove both ends of the tubing from the beaker with 10% HCl. Empty tubing by
 starting the pump and collecting the outflow in the 10% HCl beaker (you don’t want to lose
 this HCl to the waste beaker because it can be reused).

11. Pass diWater through the tubing for 5 minutes. Collect the outflow in the ‘waste’ beaker.


13. Put tubing inside an autoclave baggie and autoclave.