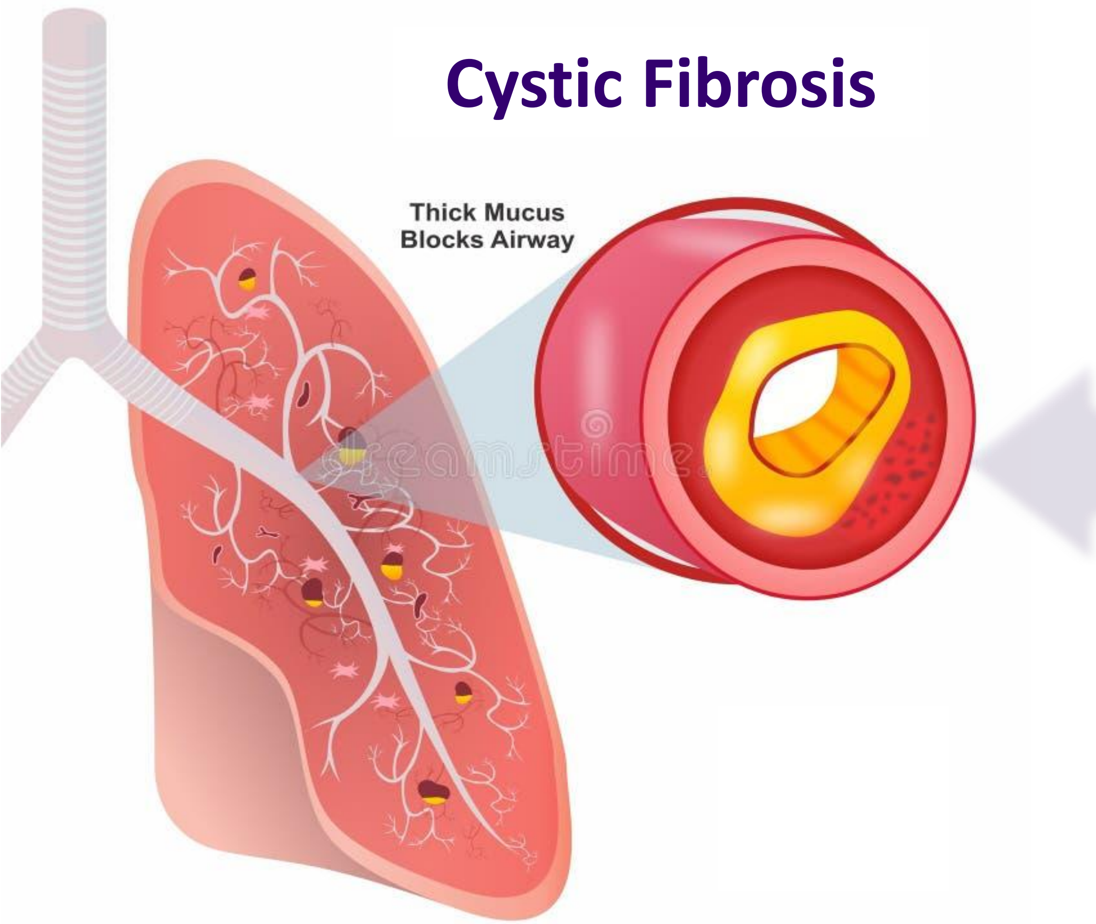
Tile-based Analysis of GC×GC-TOFMS Data of SPME Sampled VOCs Produced from *Pseudomonas aeruginosa* and *Aspergillus fumigatus*

Wenjing Ma, **Ting Wang, Jamison Whitten, Lauren J. Rajakovich, Ashleigh B. Theberge and Robert E. Synovec**

Department of Chemistry, University of Washington, Seattle, WA, USA

17th Multidimensional Chromatography Workshop

Project Motivation



- *Pseudomonas aeruginosa* and *Aspergillus fumigatus* are major pathogens found in the lungs of patients with Cystic fibrosis.
- Their coexistence worsens lung function and leads to poor clinical outcomes.
- The project aims to investigate their metabolic interactions.

Aim and Methodology

- To discover analytes that are **statistically different** in the headspace across all sample classes (Media, PA monoculture, AF monoculture, and their coculture) using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) and chemometrics.
- To explore appropriate experimental design approaches for generating high quality data and enabling high throughput experimentation.

Media

Aspergillus fumigatus (AF)

Pseudomonas aeruginosa (PA)

Coculture of PA and AF

Feature Selection Methods
Tile-based Analysis

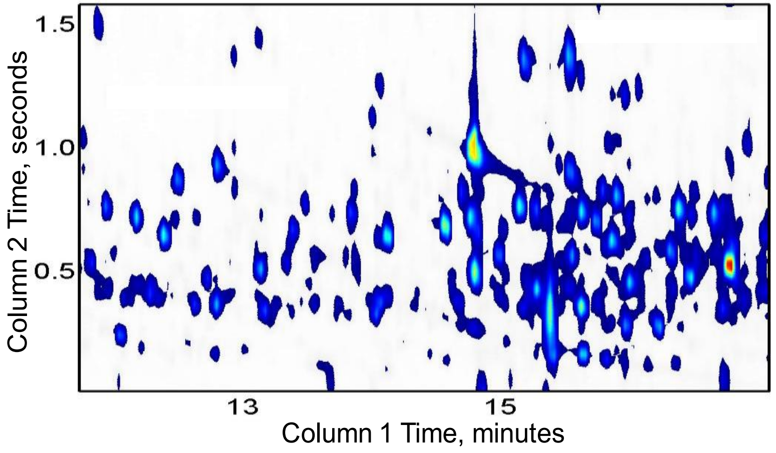
ChromaTOF Tile

GC×GC-TOFMS : Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectral Detection

- Mass spectra with electron impact ionization and unit mass resolution
  ➔ peak identification

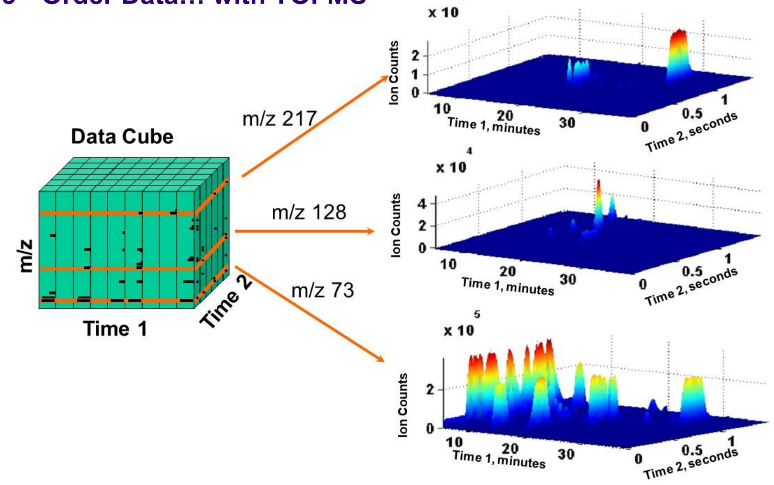
- Fast ➔ 500 spectra / second
  Column Two peak widths:
  ~ 50 ms to ~ 300 ms

- Adds another selective dimension
  ➔ 3rd - order data
  ➔ Provides informative data analysis

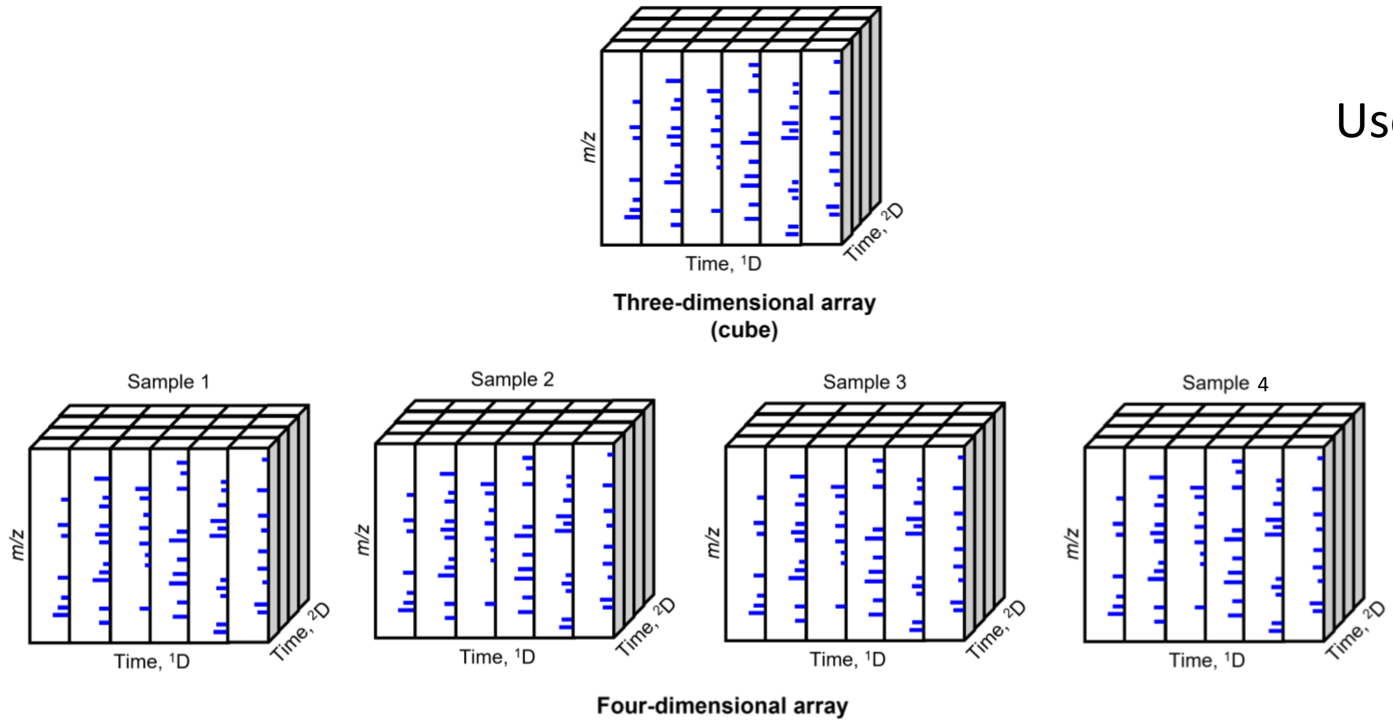


3rd Order Data… with TOFMS

Extracted Ion Chromatograms



Methods of Handling GC×GC-TOFMS Data



Use of a TOFMS with a high collection frequency can create a large amount of data

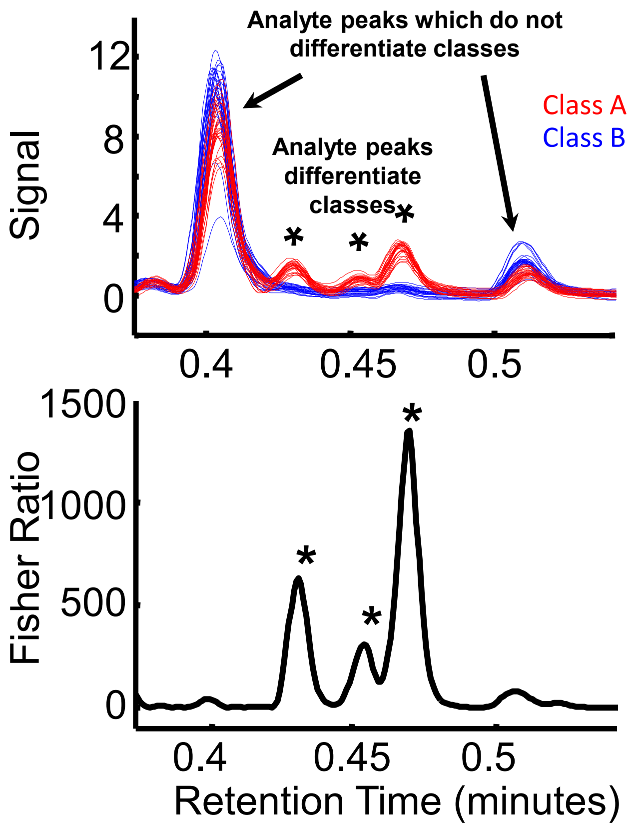
> 10^6 data points per chromatogram!

Data reduction strategies

- ***Assemble peak tables***
- ***Average the chromatographic signal (i.e., binning)***
- ***Discovery-based Feature selection***

Tile-based Fisher Ratio (F-ratio) Analysis

Shown for 1D data….



…. applied to **Class A** samples versus **Class B** samples

$$Standard\ F-ratio = \frac{Between\ Class\ Variance}{\sum(Within\ Class\ Variance)}$$

- K.J. Johnson, R.E. Synovec, J. Chemom. Intell. Lab. Syst., 2002, 60, 225-237.
- L.C. Marney, W.C. Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Talanta 2013, 115, 887–895.
- B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Anal. Chem. 2015, 87, 3812–3819.
- N.E. Watson, B.A. Parsons, R.E. Synovec, J. Chromatogr. A, 2016, 1459, 101-111.
- B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, J. Chromatogr. A, 2016, 1440, 179-190.
- B.C. Reaser, B.W. Wright, R.E. Synovec, Anal. Chem., 2017, 89, 3606-3612.

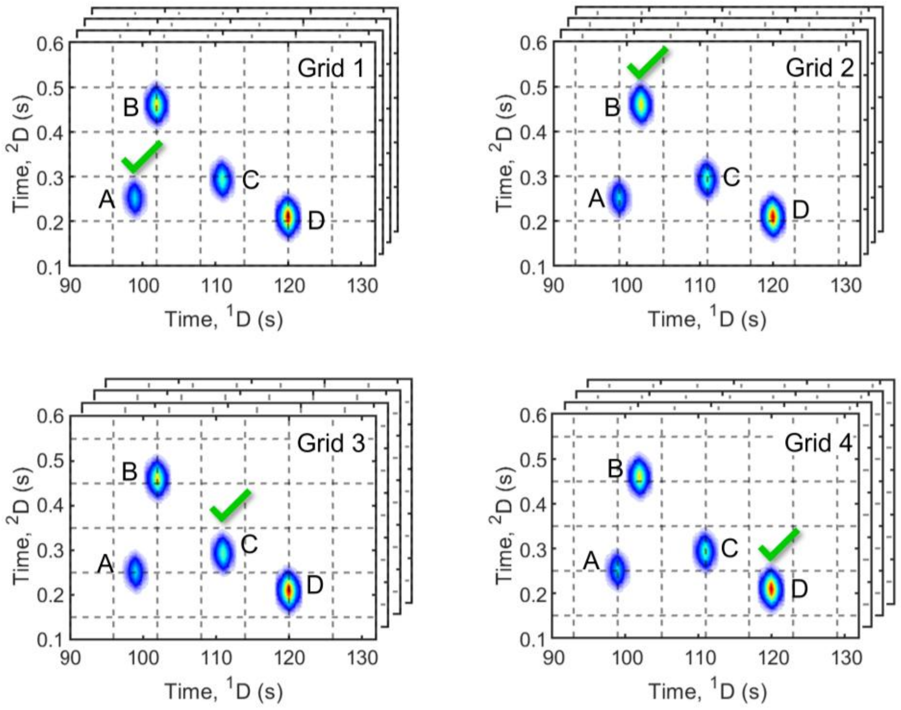
F-ratio Analysis provides a ranked hitlist of analytes that are likely to be statistically different in concentration (p-value < 0.05) between sample classes.

Hit List

Hit #	Fisher Ratio	Analyte
1	High	AAA
↓	↓	↓
N	Low	ZZZ

Tile-based Fisher Ratio (F-ratio) Analysis

Illustration of how 4-grid tile-based FRA works.

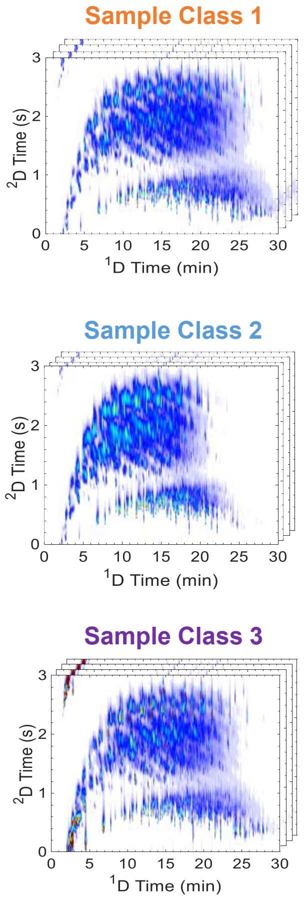


Hit List

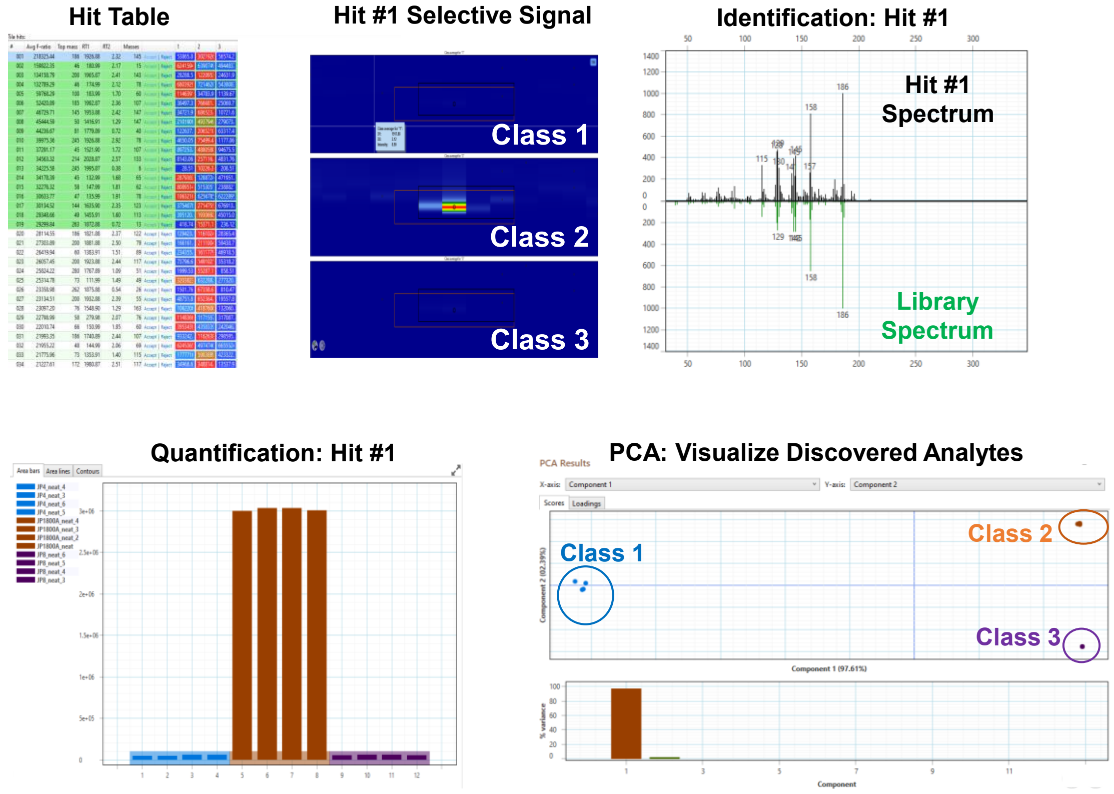
Hit #	F-ratio	(^1t_R, ^2t_R)
1	High	(x,y)
N	Low	(x_N,y_N)

Software Development at UW

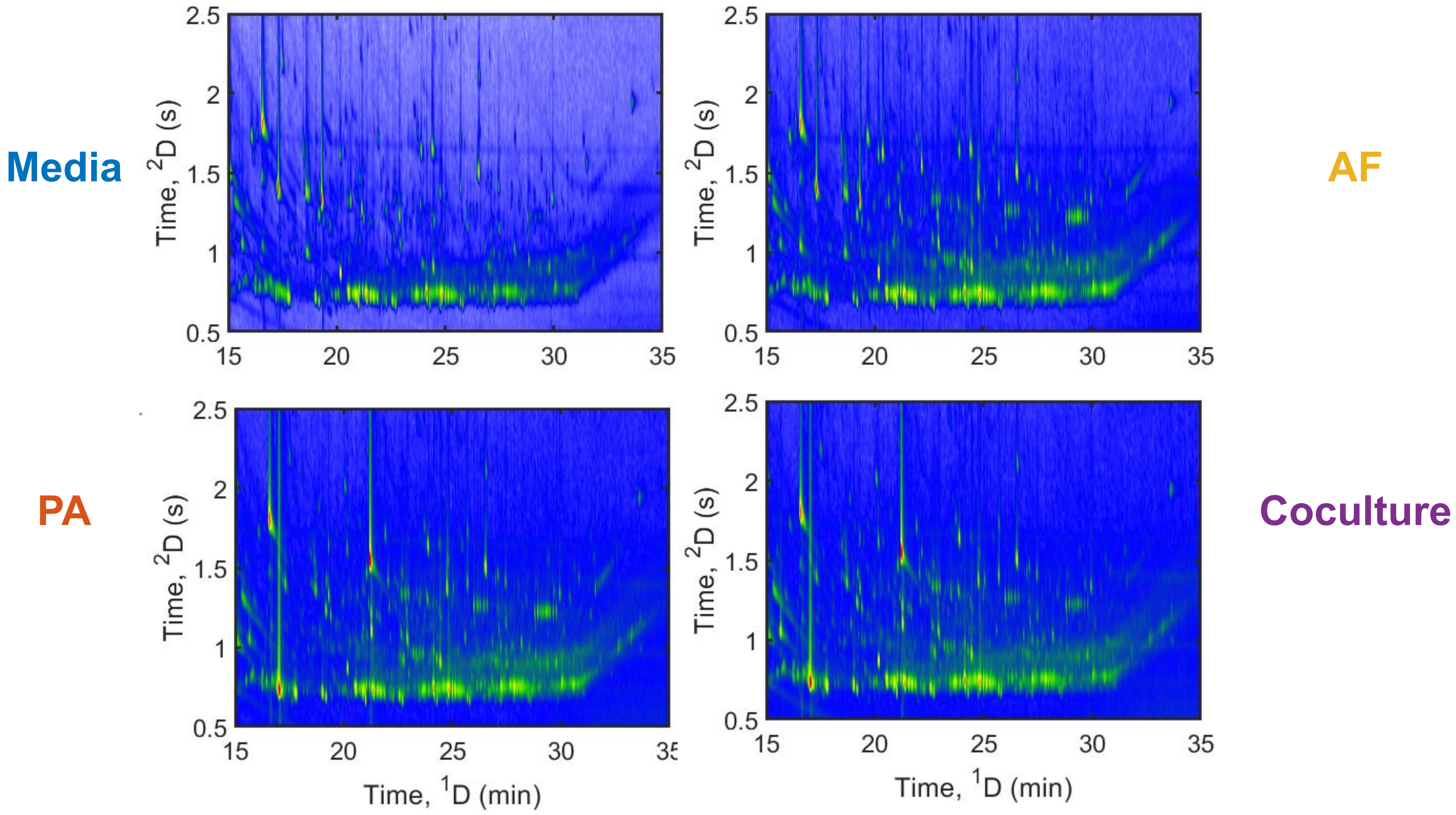
Tile-based Fisher Ratio Analysis



Apply ChromaTOF Tile



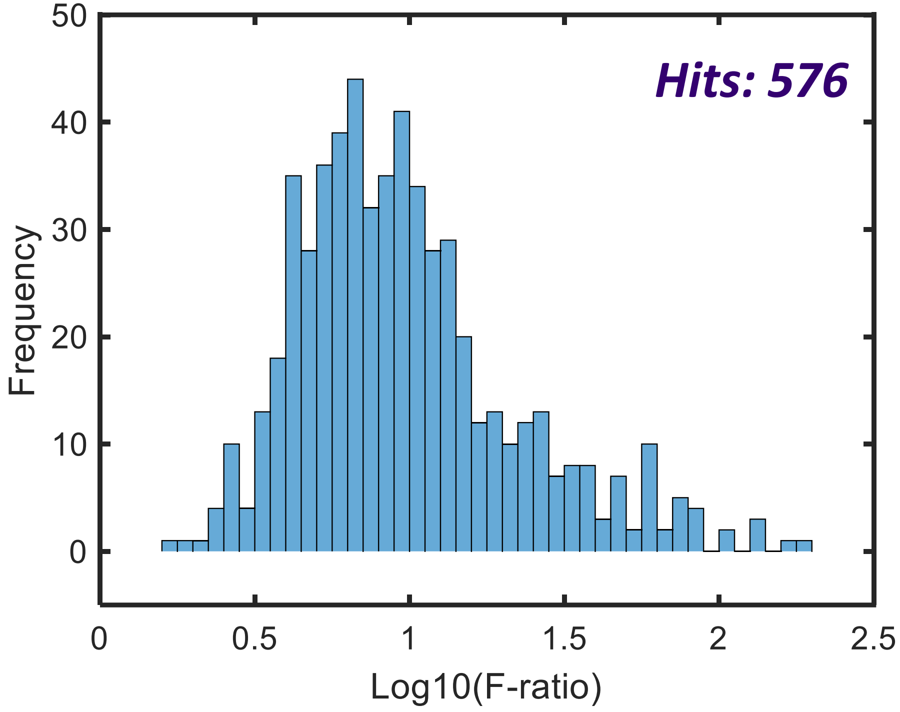
Comparative Analysis of GC×GC-TOFMS Data

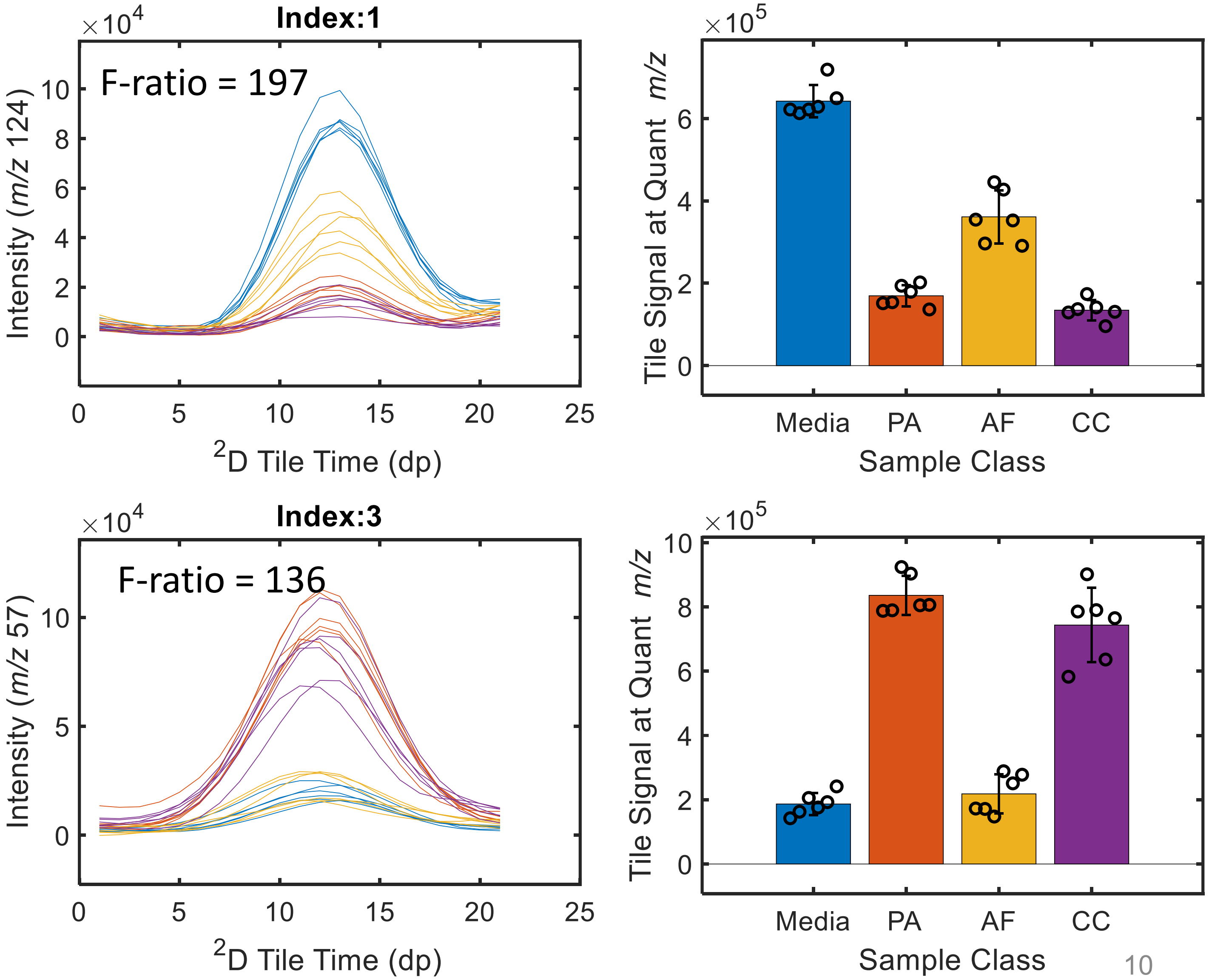


~ 500 peaks detected in the TIC chromatogram

The media sample appears to have less overall signal

Tile-based Fisher Ratio (F-ratio) Analysis

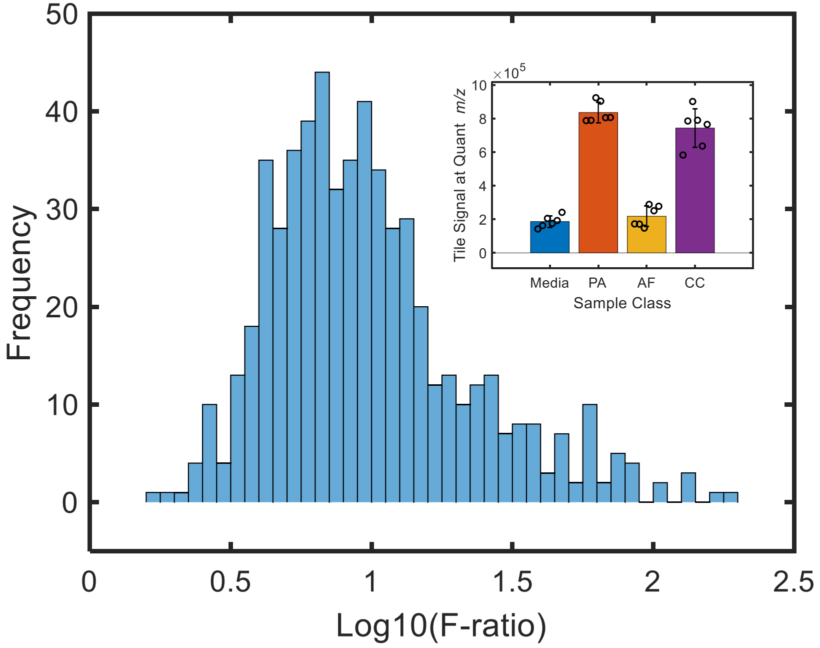




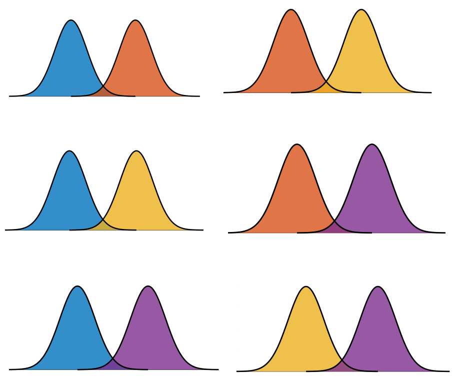
Followed Up: Pairwise t-Test (95% CL)

- F-ratio analysis discovers analytes that are statistically different across all sample classes.
- Pairwise t-tests were formed to find out which classes differ.
- Significance set at $p < 0.05$ (95% confidence level).
- For 4 classes, this results in 6 pairwise comparisons.

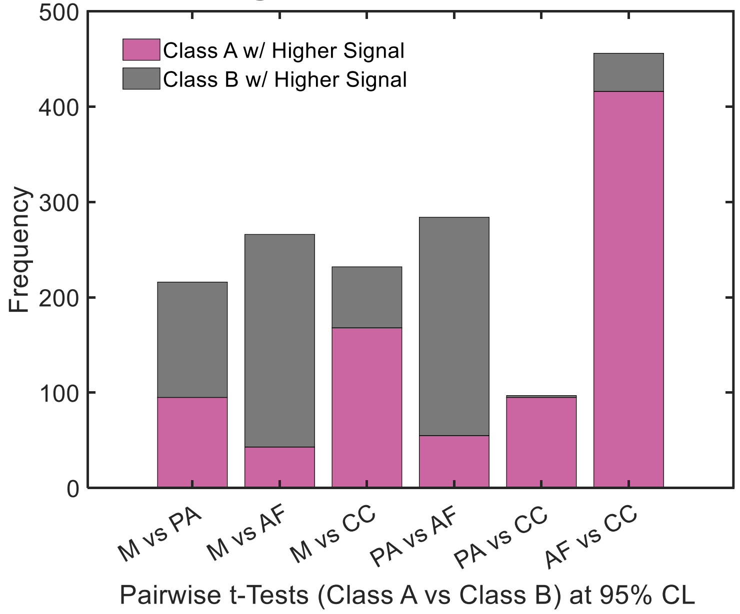
F-ratio analysis



Pairwise t-Tests



Significant Results



Aim and Methodology

- To discover analytes that are **statistically different** in the headspace across all sample classes using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) and chemometrics
- To explore appropriate experimental design approaches for generating high quality data and enabling high throughput experimentation

Feature Selection Methods
Tile-based Analysis

ChromaTOF Tile

Fisher Ratio

Coefficient of Variation

Fold Change

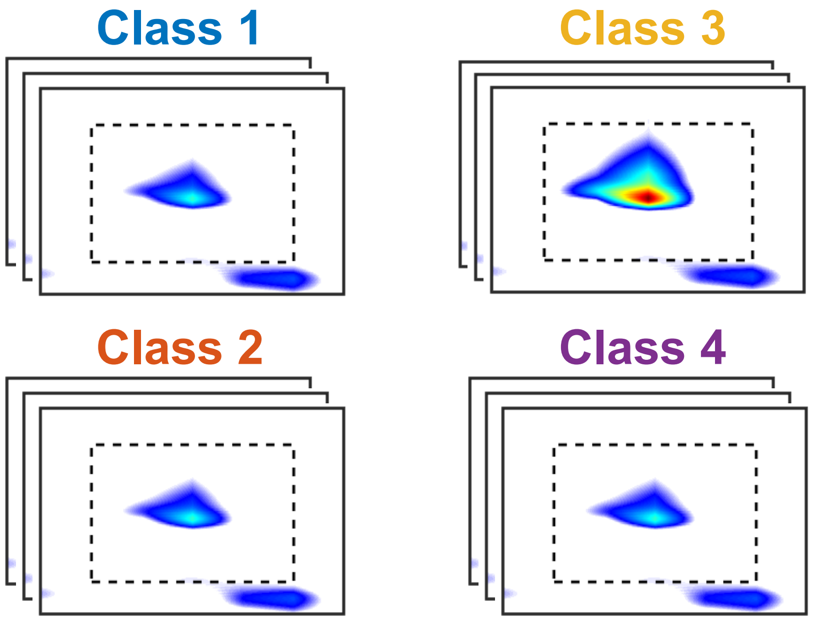
Experimental Design Considerations Under Practical Constraints

- Replicates may not always be available due to sample, time, and/or cost limitations
- Replicates may not be reproducible: consistency across replicates is required
- Highly precision required during sample preparation for device fabrication
- Time sensitivity of sample preparation and data collection
    - SPME-GC×GC-TOFMS dataset collection required 24-hour spacing between the completion of sample preparation and the start of volatile absorption.

Experimental Design Approaches for High-Quality Data via High Throughput Experimentation

Tile-Based Feature Selection Metrics

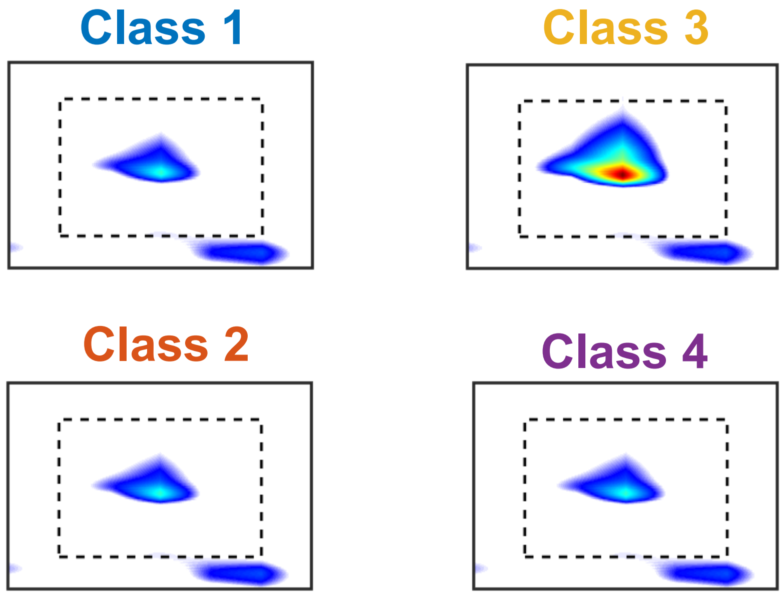
Tile-Based F-ratio Analysis



$$F - ratio = \frac{Between\ Class\ Variance}{\sum(Within\ Class\ Variance)}$$

Replicates may not always be available due to sample, time, and/or expense limitations

Tile-Based Coefficient of Variation Analysis



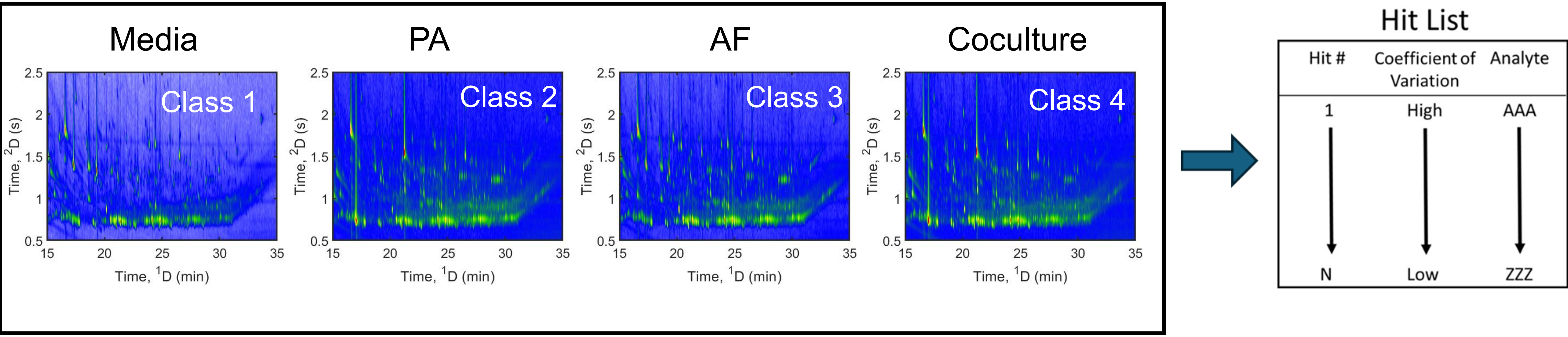
$$CV = \frac{STD}{Mean}$$

Overcomes issues associated with pixel-based subtraction plots

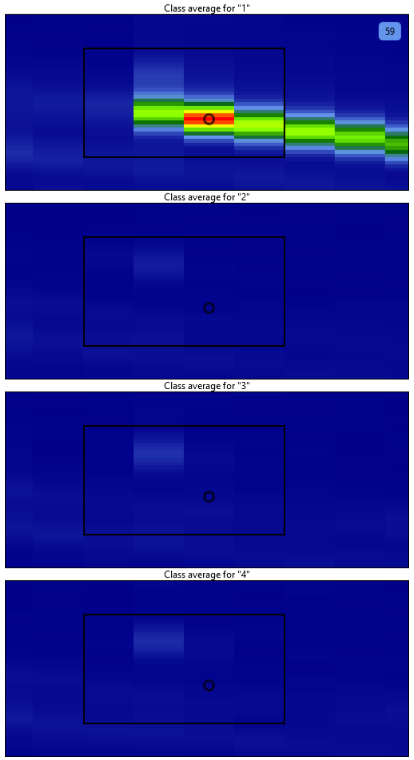
Tile-Based Coefficient of Variation Analysis

We applied ChromaTOF Tile (Coefficient of Variation method) to simultaneously to obtain a single hitlist that relates all analytes in all samples in terms of up and down regulations of their concentrations.

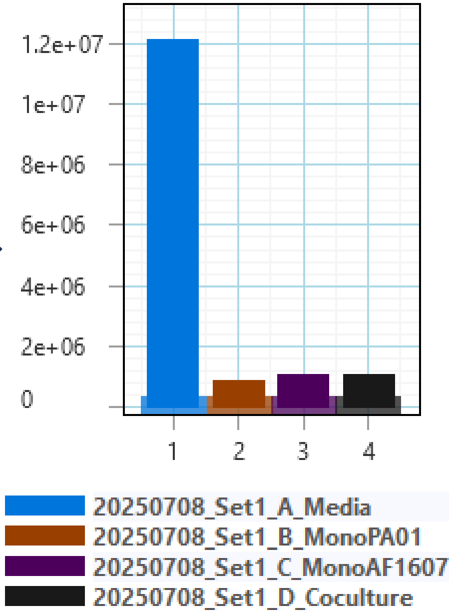
This required only one replicate from each sample to obtain this comprehensive hitlist of relative concentration for each analyte.

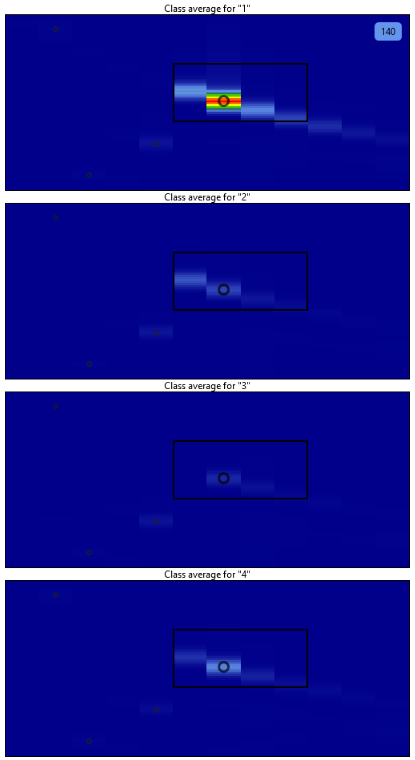


Example Hits with Highest Signals in Class 1 – Media

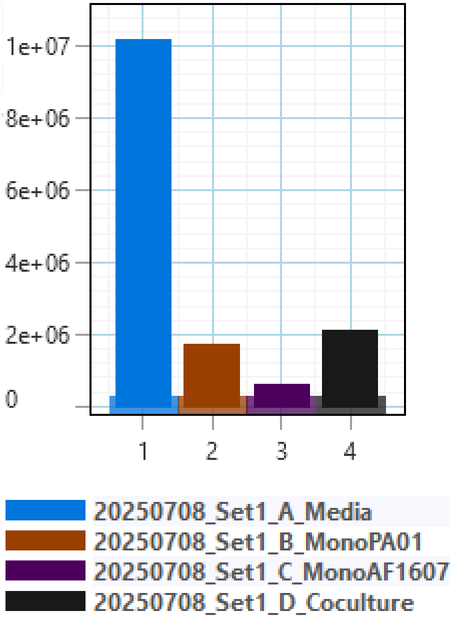


'2-Cyclobutyl-2-propanol'
MV=741 , CV = 1.48

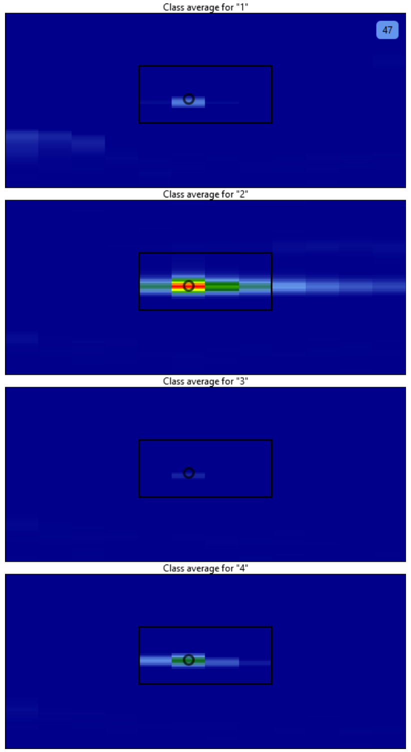




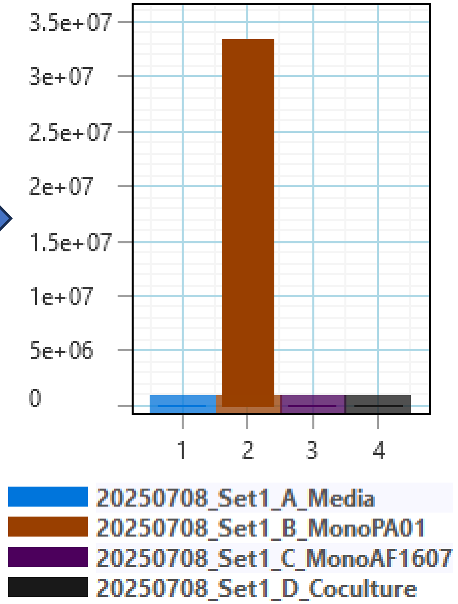
'4-Piperidinone, 2,2,6,6-tetramethyl-'
MV = 821 , CV = 1.2

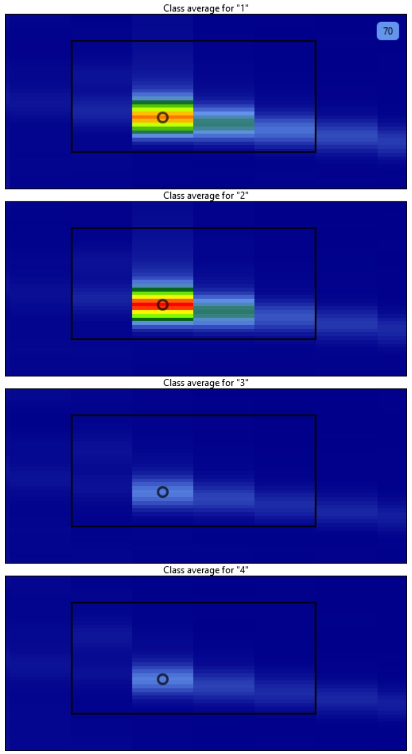


Example Hits with Highest Signals in Class 2 – PA monoculture

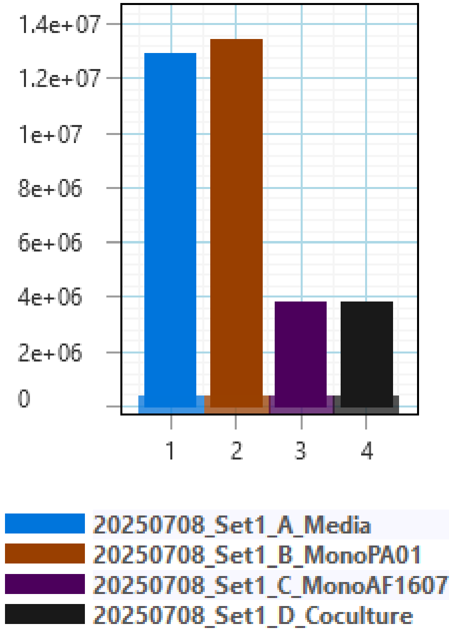


'Pentaborane(11)'
MV= 733 , CV = 2

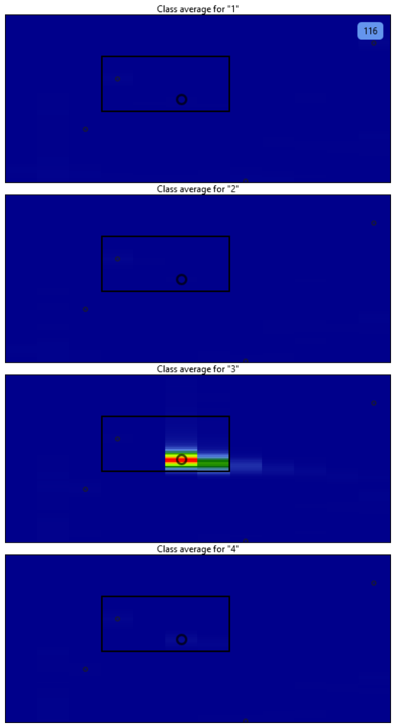




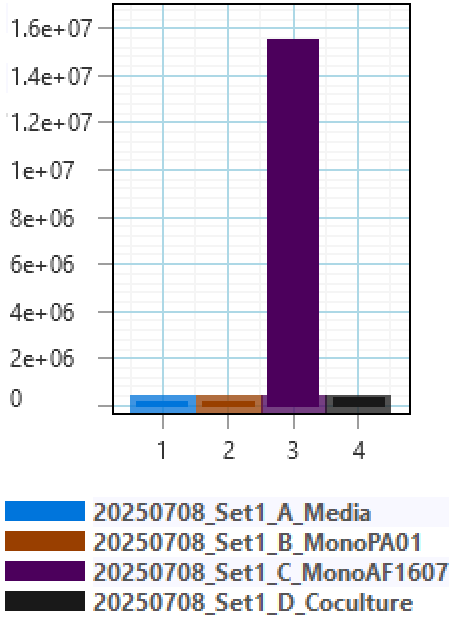
'Heptanal'
MV= 881 , CV = 0.64

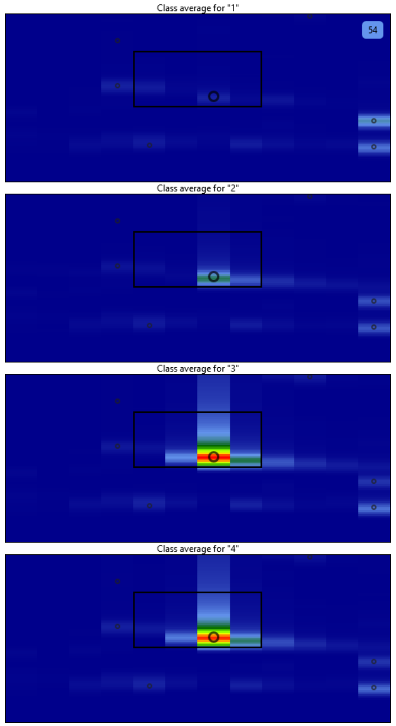


Example Hits with Highest Signals in Class 3 – AF monoculture

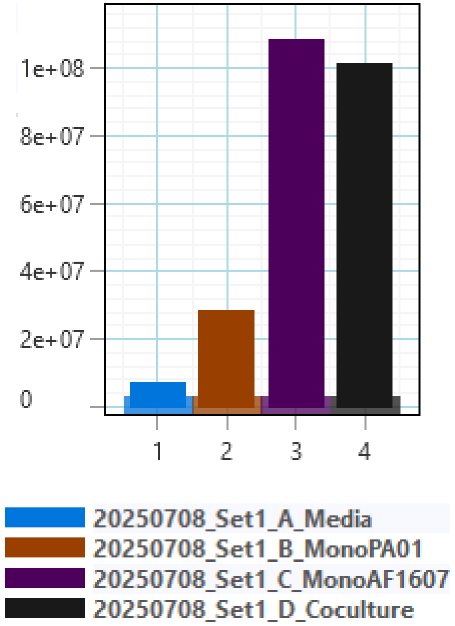


'3(2H)-Thiophenone, dihydro-2-methyl-'
MV = 818, CV = 1.9

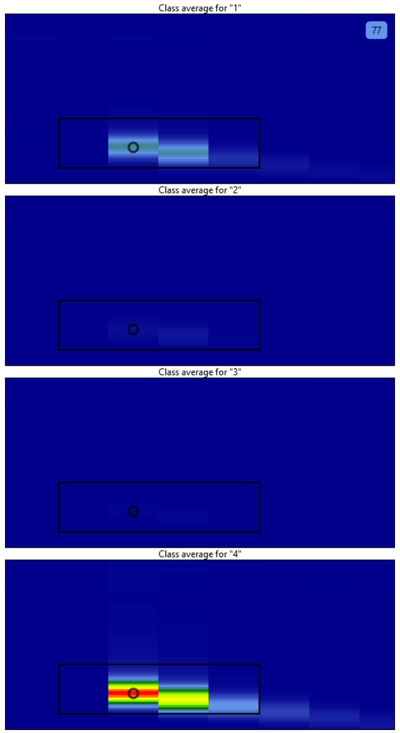




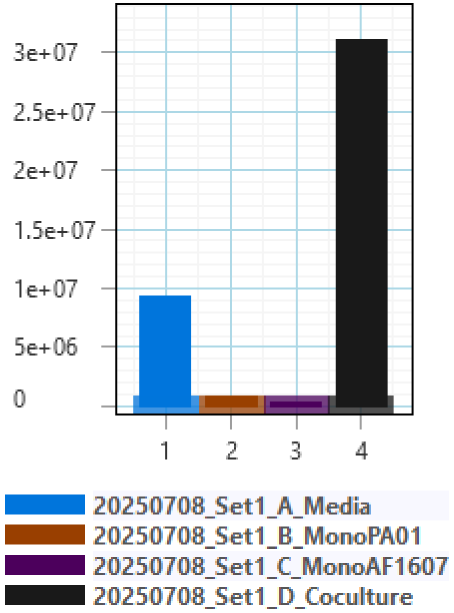
'1-Octen-3-ol'
MV = 837, CV = 0.83

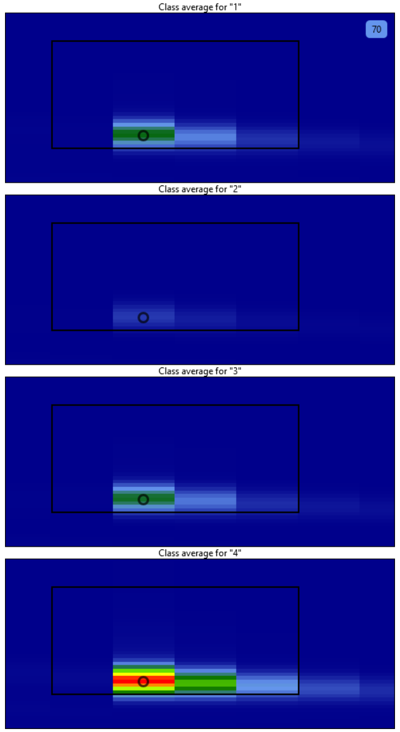


Example Hits with Highest Signals in Class 4 - Coculture

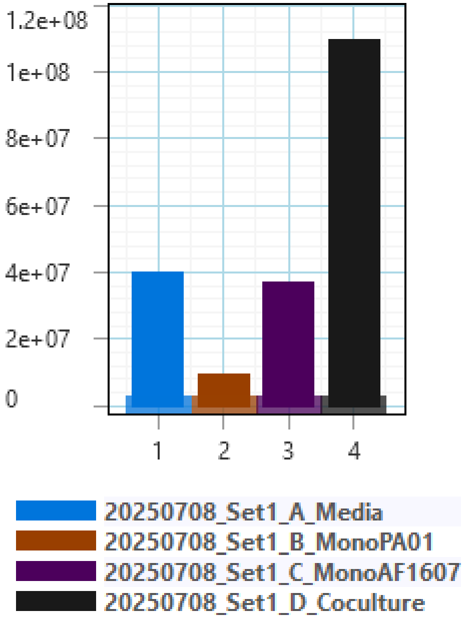


‘1H,1H,2H-Perfluoro-1-octene’
MV = 798, CV = 1.39





'2,4-Dimethyl-1-heptene’
MV = 915, CV = 0.87



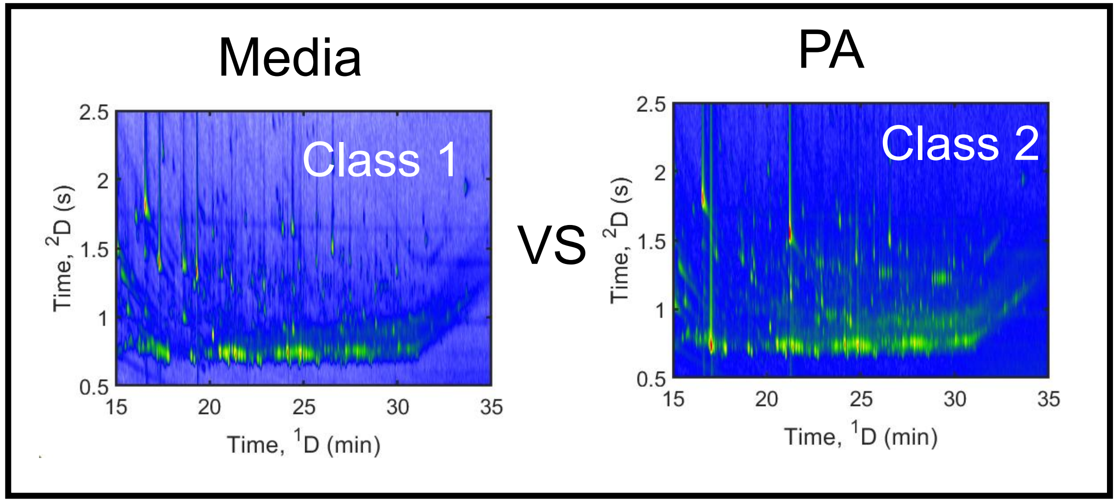
Tile-based Pairwise Analysis

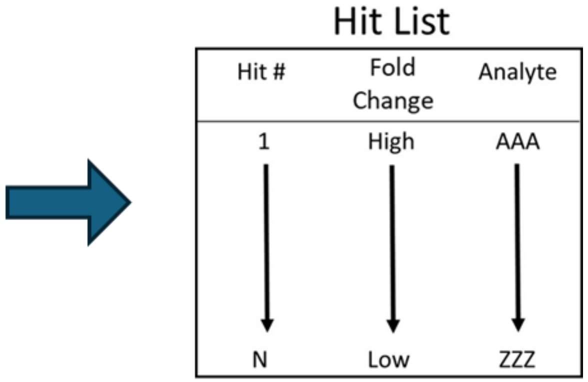
Non-Targeted Discovery Analysis Comparing Two Chromatograms

Utilize a tile-based relative change method that discovers analytes that are different in concentration between two samples with only one replicate.

$$Fold\ Change = \left(\frac{Class\ 1\ Tile\ Signal\ Sum}{Class\ 2\ Tile\ Signal\ Sum}\right)$$

Discover any hits with specified fold change (FC) threshold.





Conclusions and Next Steps

- Tile-based F-ratio analysis discovered 576 sample class distinguishing analytes.
- Additional pairwise t-tests can further determine which sample classes differ.
- Under constrained experimental conditions, the tile-based coefficient of variation (CV) method is more suitable for achieving high-quality data in high-throughput experimentation.
- Appropriate feature selection methods (F-ratio, CV, and FC) should be selected based on the specific experimental design.
- Further analyte identification can be performed using the chemometric tool PARAFAC.

Acknowledgements

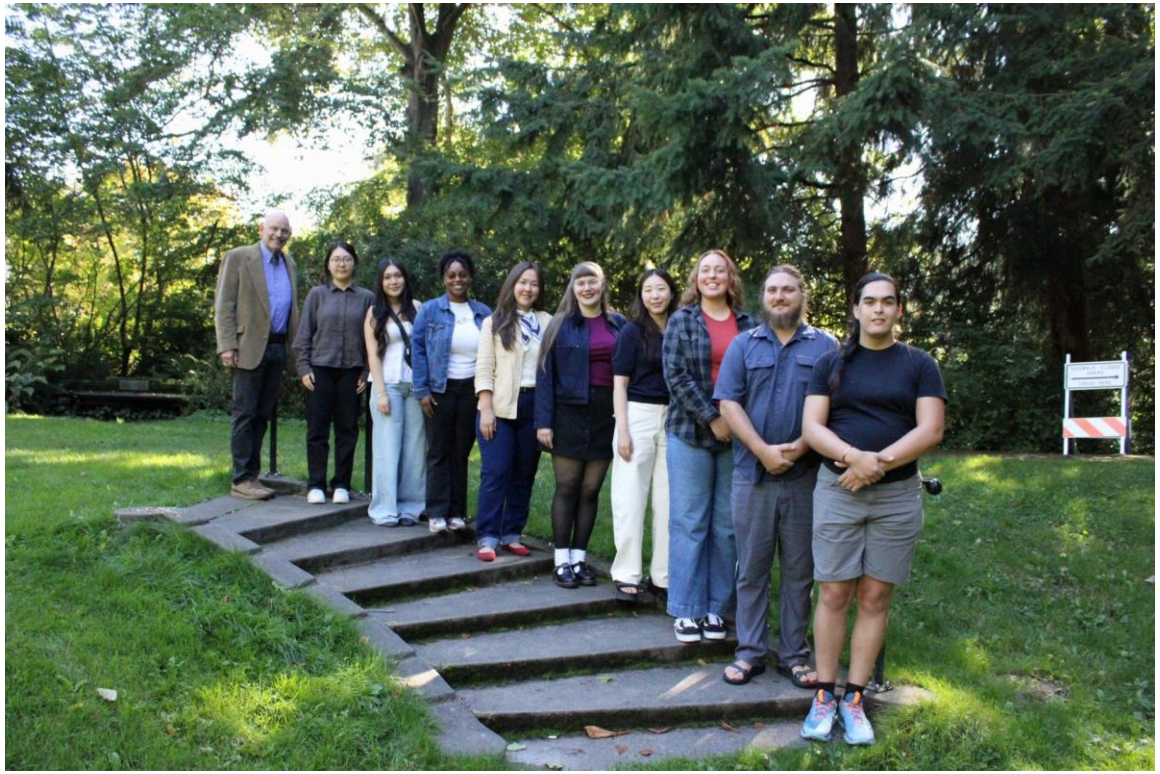
Principal Investigator
- Dr. Robert Synovec

Synovec Lab Members
- Rachel Halvorsen
- Austin Dobrecevich
- Wenjing Ma
- Haylee Meissner
- Cassandra Padilla
- Peri Abdigali
- Valencia Parker
- Jungho Ahn
- Stephanie Nguyen

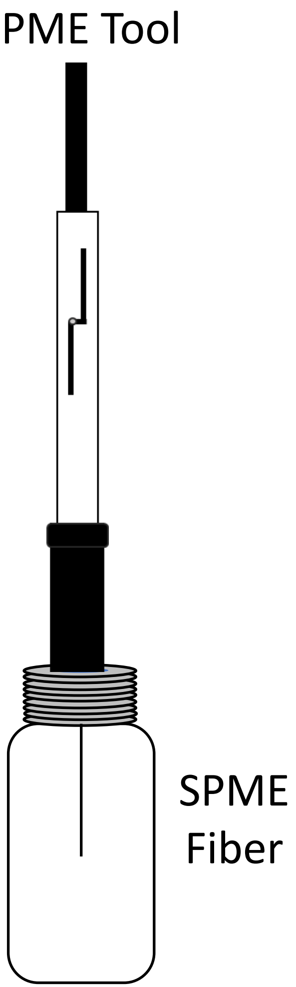
Synovec Lab

Gas Chromatography, Liquid Chromatography, and Mass Spectrometry, with Multi-Dimensional Data Analysis

Seattle's Best Chromatography

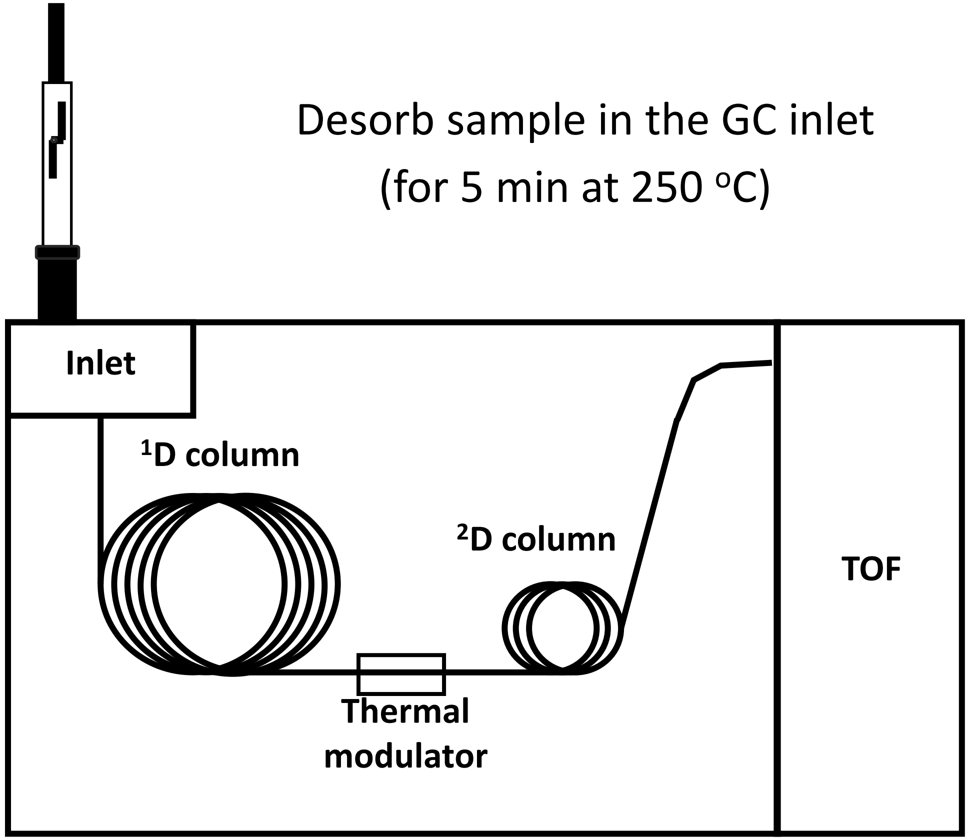


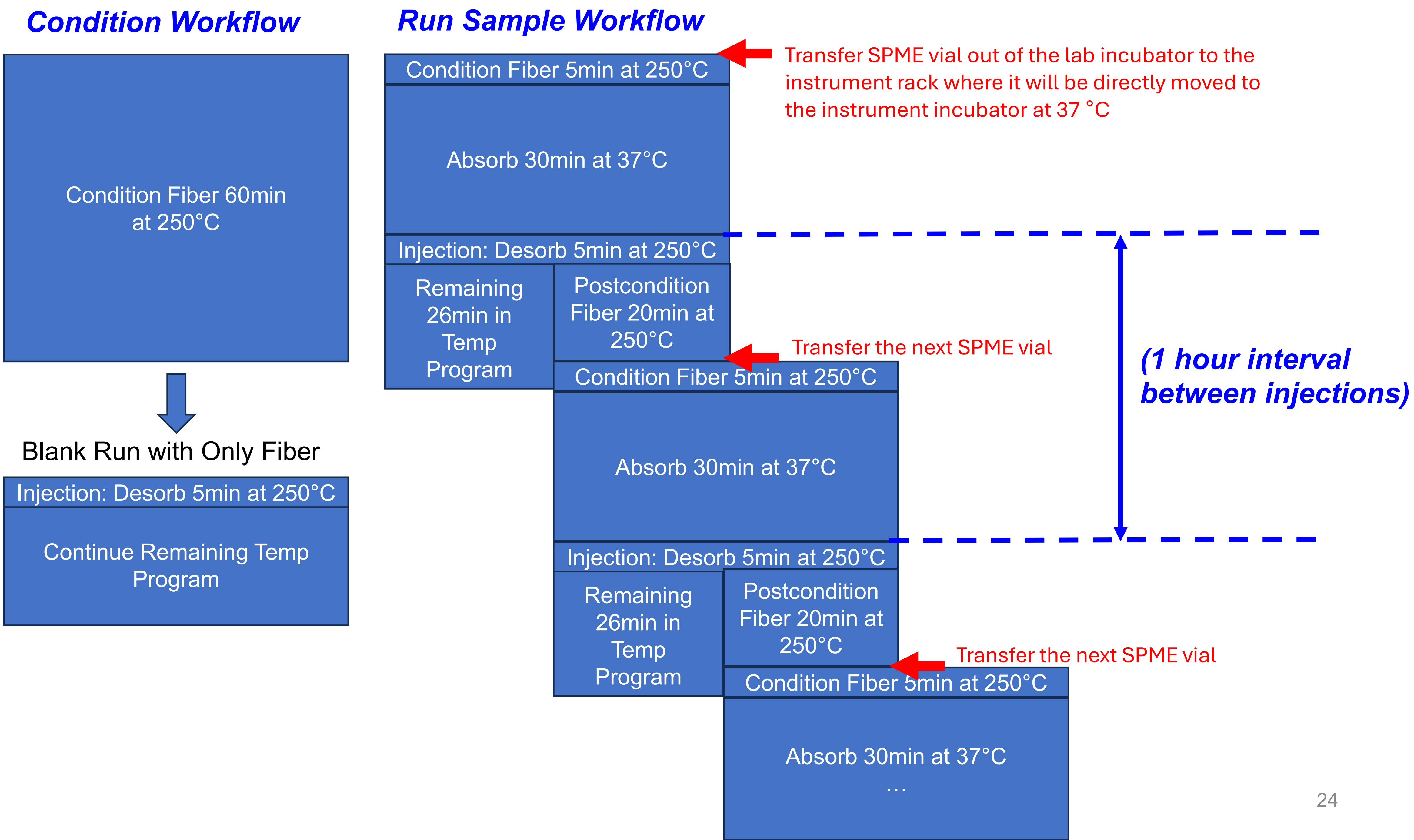
Thank you for listening! Any questions?



Expose SPME to sample
(for 30 min at 37 ^{o}C)

Desorb sample in the GC inlet
(for 5 min at 250 ^{o}C)




Condition Workflow
Condition Fiber 60min at 250°C
Blank Run with Only Fiber
Injection: Desorb 5min at 250°C
Continue Remaining Temp Program
Run Sample Workflow
Condition Fiber 5min at 250°C
Transfer SPME vial out of the lab incubator to the instrument rack where it will be directly moved to the instrument incubator at 37 °C
Absorb 30min at 37°C
Injection: Desorb 5min at 250°C
Remaining 26min in Temp Program
Postcondition Fiber 20min at 250°C
Transfer the next SPME vial
Condition Fiber 5min at 250°C
Absorb 30min at 37°C
(1 hour interval between injections)
Injection: Desorb 5min at 250°C
Remaining 26min in Temp Program
Postcondition Fiber 20min at 250°C
Transfer the next SPME vial
Condition Fiber 5min at 250°C
Absorb 30min at 37°C
…