New Investigator Tools for Finding Unique and Common Components across Multiple Samples with Comprehensive Two-Dimensional Chromatography

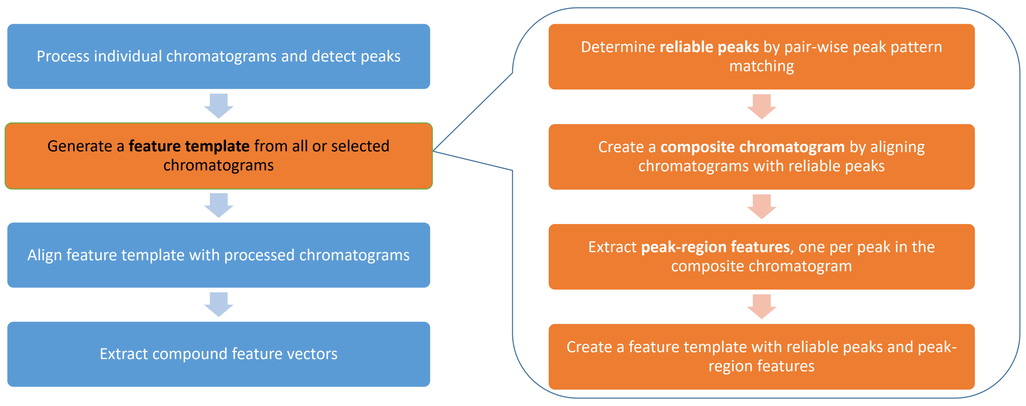
Qingping Tao;[1] Stephen E. Reichenbach;[1,2] Chase Heble;[1] Zhanpin Wu;[3]
[1] GC Image, LLC, Lincoln NE, USA; [2] University of Nebraska, Lincoln NE, USA; [3] Zoex Corporation, Houston TX, USA

INTRODUCTION

Comprehensive two-dimensional chromatography is a powerful technique for highly effective chemical separations of complex mixtures and is increasingly used for cross-sample analyses such as sample classification and biomarker discovery. These technologies, such as GCxGC and LCxLC, produce large data sets that are rich with information, but are highly complex and require automated processing with robust methods.

Our Investigator framework analyzes data from multiple samples to extract a feature template that comprehensively captures the pattern of peaks detected in the retention-times plane **[1]**. Then, for each sample chromatogram, the template is transformed to align with the detected peak pattern and generate a set of feature measurements for cross-sample analyses. The approach avoids the intractable problem of comprehensive peak matching, but can generate feature templates with thousands of features.

Automated workflow supported by GC Image's Investigator™ software



CHALLENGES

An important analysis step is to select markers that can be used effectively for clustering and classifying multiple samples. Visualization is useful to verify their effectiveness.

Challenge 1: Detect unexpected compounds that appear in some samples but not others:

- Unique Compounds that appear in one sample class but not others;
- Common Compounds that appear in all samples.

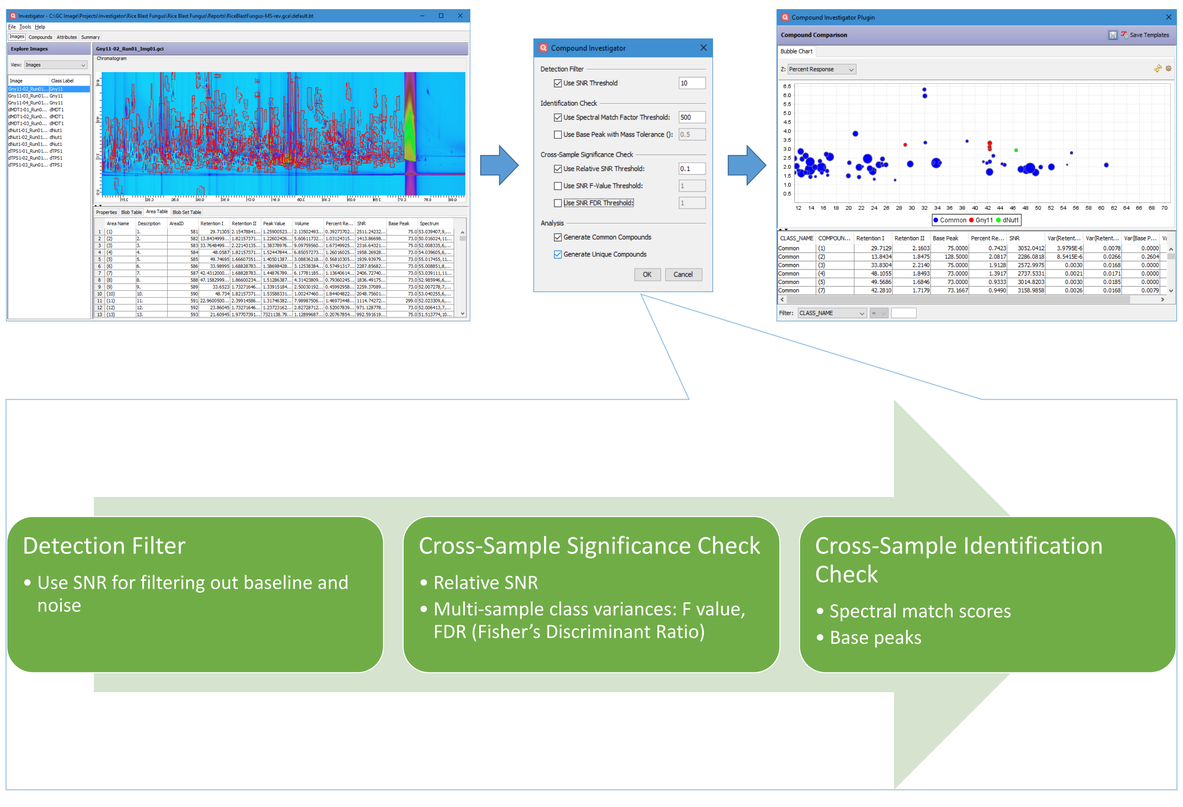
Challenge 2: Accommodate any type of chromatogram data analysis:

- **Multi-sample classes**: Multiple chromatogram runs are available for each sample class. This is the most common scenario for traditional data classification analysis. Classic variance-based statistics such as F value work well, but **do not provide clear predictions of commonality and uniqueness**.
- **One-sample classes**: Only one chromatogram run is available for each sample class. This is the most cost effective for sample collection and data acquisition, but **variance-based statistics cannot be computed.**

OUR APPROACH

We developed a new workflow and associated tools that extend the Investigator framework with specialized detection and identification constraints based on chromatographic and mass spectral information to distinguish targeted compounds. This new workflow allows analysts to detect common and unique compounds across many samples for either multi-sample classes or one-sample classes.

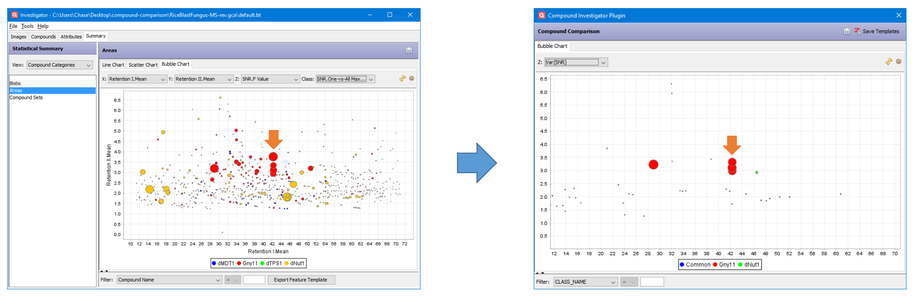
1. Extract compound features with
   - Chromatographic information: Retention Times, SNR, and Percent Response
   - Spectral information: Spectrum and Base Peak
2. Build a Compound-Sample-Class hierarchical association index (HAI)
3. Prune the HAI by applying specified criteria that can give analytically useful information on compound deviations between samples
4. Visualize the result table with a labeled bubble plot, which provides not only metric values, but also instructive predictions as to which features are effective for distinguishing samples.



MULTI-SAMPLE CLASS EXAMPLE: RICE BLAST FUNGUS

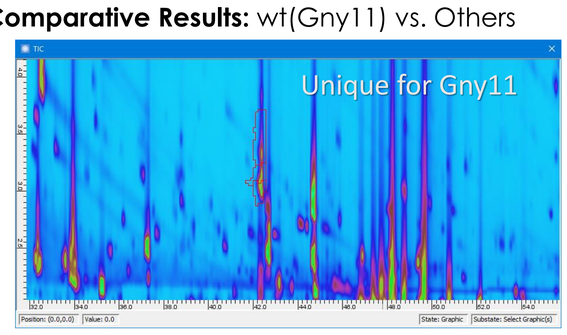
❖ Samples: Non-targeted analysis of 4 types of rice blast fungus (Magnaporthe oryzae) including wt(Gny11), Δnut1, Δtps1, and Δmdt1 **[2]**. Each type has 3 samples.
❖ Instrument: Agilent 7890B/ZOEX ZX2 thermal modulation system coupled with Agilent 7200 Q-TOF

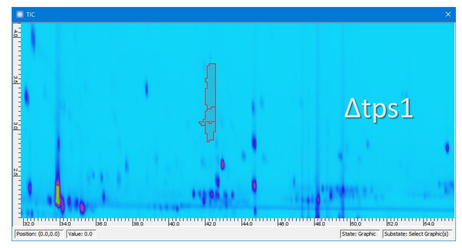
Criteria: SNR > 10, Spectral Match Factor Threshold = 500, SNR F Value Threshold = 5
Results: 35 common features and 5 unique features from two types are detected from total 572 features.

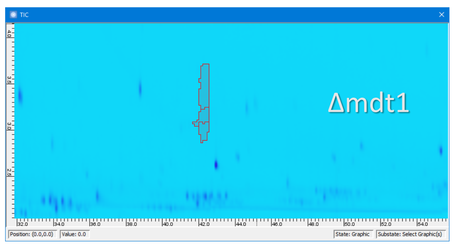


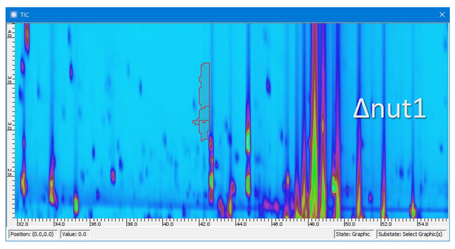
- Left: A bubble plot of all features with F value as bubble sizes and class labels determined by FDR.
- Right: A bubble plot of common and unique features with F value as bubble sizes and class labels determined by above criteria.
- **Not all significant compounds are also unique makers for a specific class.**

Comparative Results: wt(Gny11) vs. Others





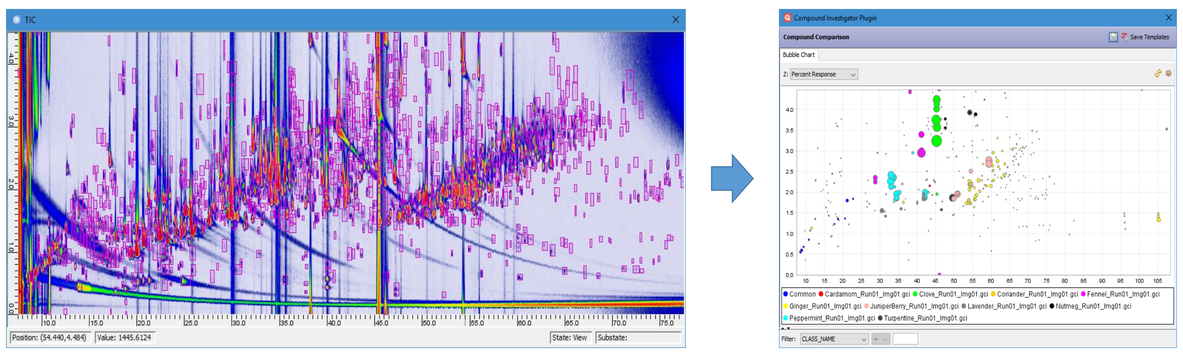




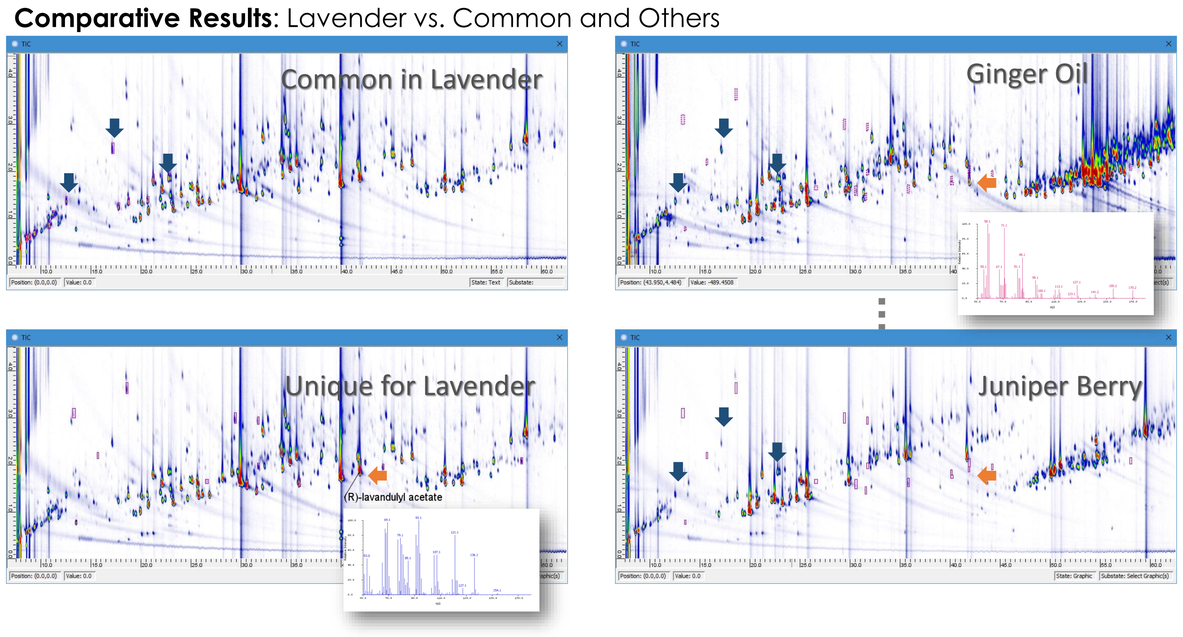
ONE-SAMPLE CLASS EXAMPLE: ESSENTIAL OILS

❖ Samples: Non-targeted analysis of 10 essential oils including Cardamom, Clove Bud, Coriander, Fennel, Ginger Oil, Juniper Berry, Lavender, Nutmeg, Peppermint, and Turpentine **[3]**. Each type has one sample.
❖ Instrument: Agilent 7890A/ZOEX ZX2 thermal modulation system coupled with Agilent 7200 Q-TOF

Criteria: SNR > 10, Spectral Match Factor Threshold = 500, Relative SNR Threshold = 0.1
Results: 12 common features and 319 unique features are detected from 1352 features.



Comparative Results: Lavender vs. Common and Others



References

1. S. Reichenbach, X. Tian, C. Cordero, Q. Tao. "Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography." Journal of Chromatography A, 1226:140-148, 2012.
2. S. Aronova, W.C. Ledford, M. Marroquin-Guzman, R.A. Wilson, Q. Tao, S.E. Reichenbach, Z. Wu, E.B. Ledford, J. Gushue, and H. Prest. "Untargeted Metabolomics Study of the Plant-Pathogenic Fungus Magnaporthe Oryzae by GCxGCxQTOFMS." GCxGC Symposium, Riva del Garda, IT, May 2014.
3. E. Ledford, Z. Wu, S. Nieto, S. Reichenbach, and Q. Tao. "GCxGCxQ-TOF-MS Survey of Essential Oils." ASMS Conference on Mass Spectrometry and Allied Topics, San Antonia TX, June 2016.
4. GC Image R2.7a GCxGC-HRMS, a custom alpha version used for the data processing (Visit www.gcimage.com for current R2.7 public releases).

GC IMAGE
Software for Multidimensional Chromatography

UNIVERSITY OF Nebraska Lincoln

ZOEX | CORPORATION