ELRIG
Drug Discovery 2019

High-throughput metabolomics for liver safety and mechanisms of toxicity

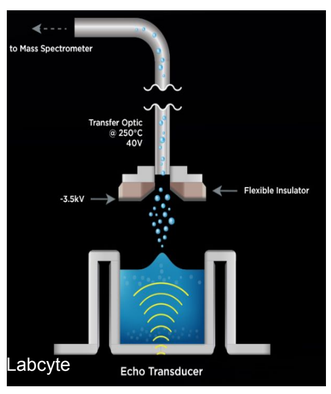
Delyan Ivanov[1], Maria Luisa Guerriero[2], Natalie Kurbatova[2], Ian Sinclair[3], Jonathan Wingfield[1], Liz Roberts[1], Ian Barrett[2]
[1] Mechanistic Biology & Profiling, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK
[2] Data Sciences and Quantitative Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK
[3] Sample Management, Discovery Sciences, R&D, AstraZeneca, Alderley Park, UK

Aim: Improve sensitivity for cell-based liver toxicity predictions and gain mechanistic insight by measuring intracellular metabolites

I. The challenge of predicting Liver Injury

- Current cell-based Liver Toxicity assays mostly measure cell death
- Functional liver damage occurs much earlier than death
- Cytotoxicity endpoints peak at ~60% Sensitivity and 90% Specificity
- Measuring metabolites better reflects cell functioning and health
- Multiple metabolites can be linked in pathways to reveal mechanisms

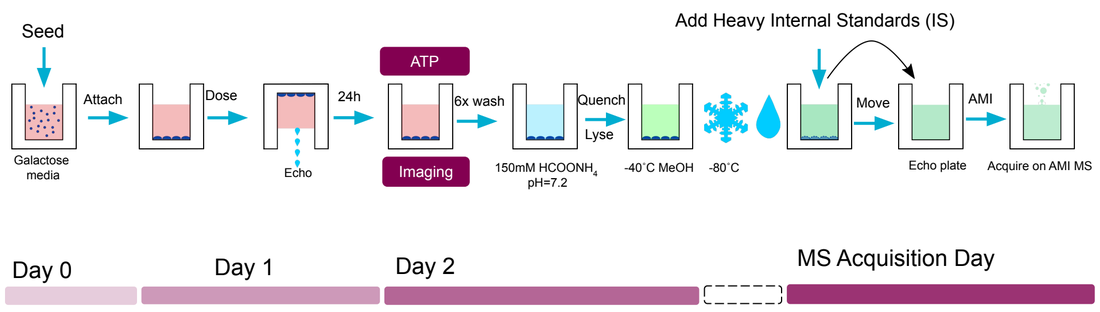
II. The opportunity of AMI MS



Acoustic Mist Ionisation Mass Spectrometry

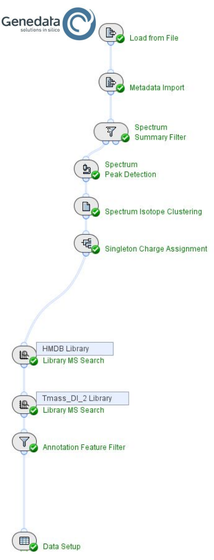
- Prototype technology conceived in DS
- Acoustic transducer ejects femtolitre droplets
- Tiny volumes - 150nL sample used
- High speed - 2 sec/sample; 20mins / 384 wells
- No chromatography selection
- ToF Detector for wide metabolite coverage

III. Experimental workflow



- 150 DILI (Drug Induced Liver Injury)-relevant compounds
- 9-point, half-log concentration-response
- 316µM top concentration

IV. Data processing



- Pre-processing performed in GeneData Expressionist
- Isotope cluster patterns used to detect metabolites
- Human Metabolite Database (HMDB) queried for best match
- Metabolite hits confirmed with manually curated library
- Detected >3000 peaks; 1000 have annotation
- Expressionist greatly reduced pre-processing times
- Before Expressionist – days to pre-process; now minutes

V. The data analysis challenge

- Highly multidimensional data with thousands of features spanning 6 orders of magnitude
- High proportion of raw features are noise
- Many features have high proportion of missing values
- Batch effects – within-plate, plate-to-plate and day-to-day variability
- Need to find “real” signals in variable background
- Non-monotonic concentration-responses – hormetic and biphasic

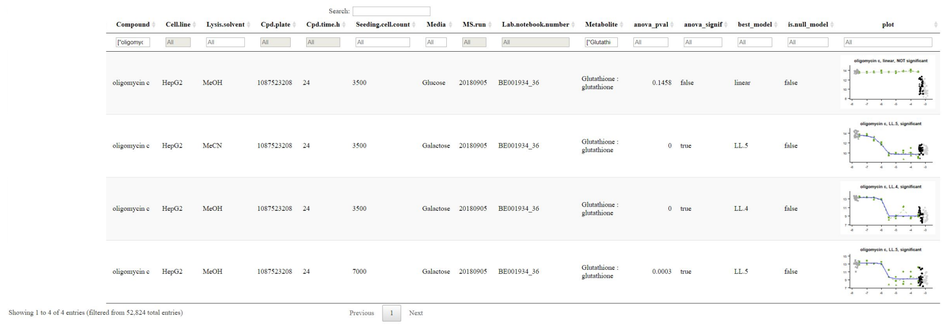
VI. The analysis pipeline

log-transformation | Missing values imputation | Plate-to-plate batch correction | Plate effect correction | Sample normalisation

- Custom pipeline in R
- Log-transformation to counter skewness
- k-NN (k-nearest neighbours) imputation of missing values
- Batch correction to reduce plate-to-plate and day-to-day variability
- Spatial plate effects normalisation to reduce within-plate variability
- Internal standard (IS) dependent and IS independent normalisation

VII. Automated concentration-response fitting

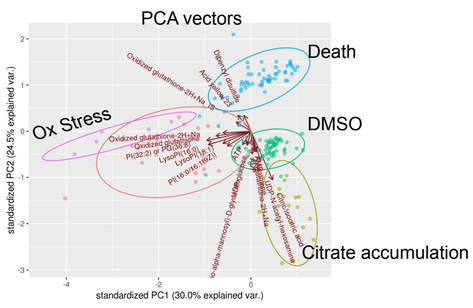
- GD Screener struggles to fit increasing and non-monotonic responses
- We built an automatic workflow to fit multiple dose-response models
- Fit straight line; sigmoidal; hormesis; multiphasic and bell-shaped models
- Used Bayesian information criterion (BIC) to penalise complex models
- Automated fit >60,000 metabolite-compound curves per experiment
- Interactive, searchable HTML table to quickly filter and examine curve fits



Left: Screenshot of the interactive report. **Right**: Examples of model fits to the response of two different metabolites to chlorpromazine treatment: LL4 (above) and gaussian (below).

VIII. Future application – towards a predictive assay

- Developed a pipeline to analyse challenging multiparametric dataset
- Working towards launching a metabolomics assay with improved sensitivity



Clustering of compounds to reveal mechanistic insight

We thank Martin Bachman, Muireann Coen, Vasudev Kantae, Mark Viant, Ralf Weber and Matthew Smith for sharing mass spectrometry data processing expertise and metabolomics insight.

AstraZeneca