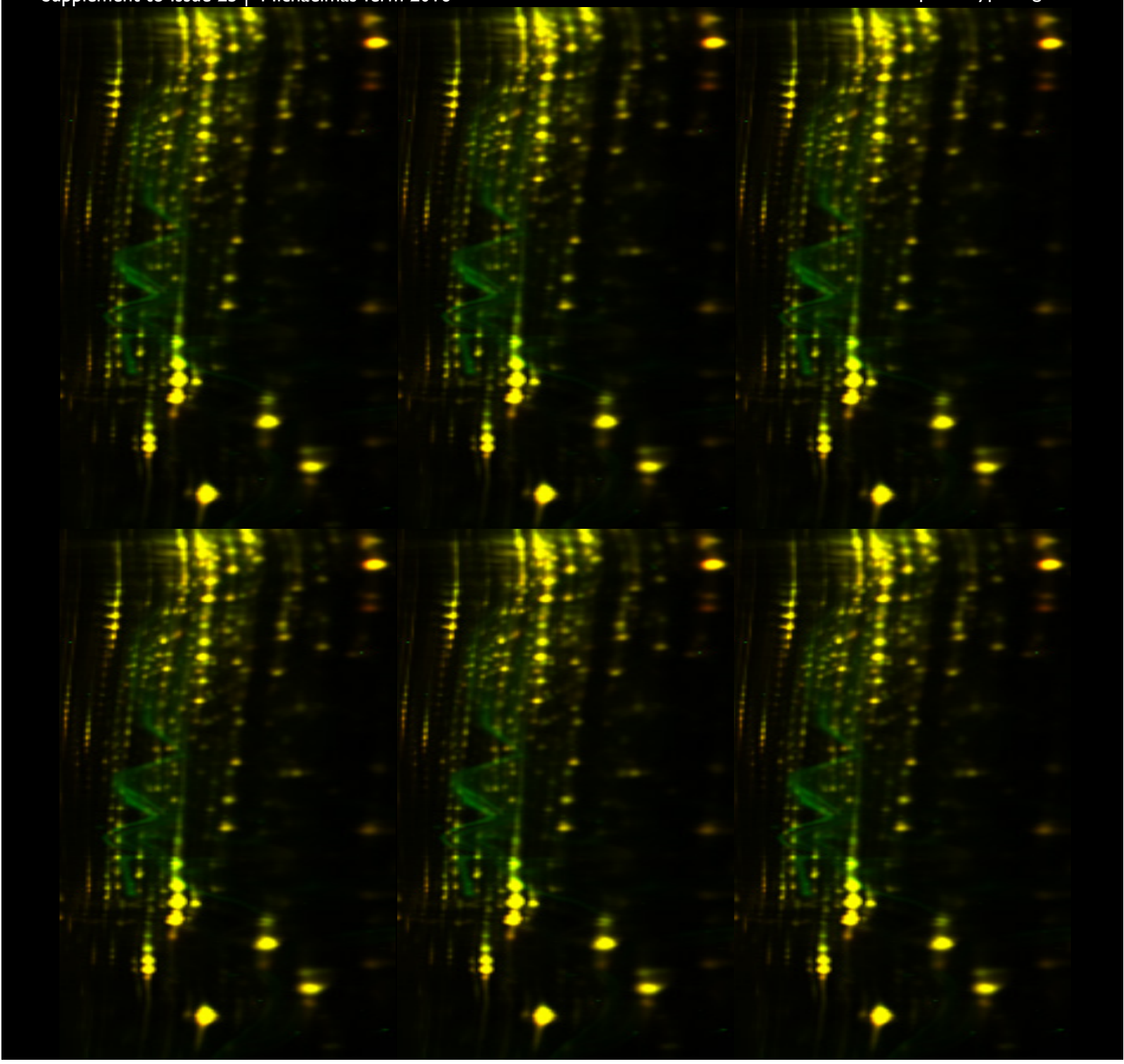
PHENOTYPE

'Omics

Supplement to Issue 25 | Michaelmas Term 2016

www.phenotype.org.uk



Supporting proteomics use through training

Transcriptomics and iPSCs

Metabolomics: a snapshot

OXFORD UNIVERSITY BIOCHEMICAL SOCIETY

OUBS

CONTENTS

'Omics

OXFORD UNIVERSITY BIOCHEMICAL SOCIETY

DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

MEDICAL SCIENCES Skills Training

OXFORD UNIVERSITY PRESS

WILEY-BLACKWELL

EDITORIAL TEAM

Cover Image: Tiled image of a 2D DIGE gel, Dr Jillian Simon

Supporting access and use of proteomics through training

by Dr Jillian Simon

From interrogating protein complexes to identifying potential disease biomarkers, mass spectroscopy-based proteomics plays a significant role in modern day biomedical research and more and more investigators are depending on 'omics approaches to address their research questions. Yet, at a place like Oxford – where facilities are stocked with cutting-edge equipment and our scientists are at the forefront of methodologic development – accessibility, training and financial feasibility are still a rate-limiting step for applying proteomics techniques to any research programme. The Proteomics Training Scheme, supported by Oxford's British Heart Foundation Centre for Research Excellence and the Target Discovery Institute, aims to tackle these issues by offering an array of training opportunities, easier access to state-of-the-art technology and financial incentives for those interested in playing a more active role in their proteomics studies.

Proteomics has become an integral part of biomedical research in the last several decades and mass spectroscopy has played a crucial role in the field, allowing scientists to evolve our understanding of biological systems. As scientists continue to push the technical limits of mass spectroscopy, biotechnology engineers have kept pace by designing cutting edge instrumentation capable of churning out large amounts of data from single experiments. Yet, we find ourselves unequipped to efficiently manage, analyse and interpret this data, leading to a bottleneck in the translation to impactful discoveries. In an effort to bridge this gap, initiatives aimed at educating and training biomedical researchers in proteomic workflows and data analysis have increased world-wide. Here at Oxford, the British Heart Foundation's Centre for Research Excellence (BHF CRE) and the Target Discovery Institute (TDI) have partnered to develop the Proteomics Training initiative, launched in October 2015. The initiative, led by Jillian Simon (Radcliffe Department of Medicine, RDM), Mark Crabtree (RDM) and Roman Fischer (TDI), is a coordinated effort to increase awareness, accessibility and training for proteomics research while reducing the financial burden through a number of training opportunities and financial incentives.

The programme is designed as a 3-tier training system (Figure 1) which offers opportunities for researchers to engage in a number of ways to increase their awareness of accessibility to proteomics applications. The first tier, which is applicable to the majority of researchers, includes seminars covering a broad range of 'omics applications for biomedical research, such as the one held on 20th October 2015 to kick-start the training initiative. This one-day seminar served as a general overview of proteomics topics, including an introduction to mass spectroscopy-based proteomic workflows, case studies on how proteomics has been applied clinically here at Oxford, as well as mass spectroscopy-based approaches for metabolomic and lipidomic analysis. Future seminars are likely to include application of proteomics to large patient cohorts, overviews of the latest proteomic methods available and how to harvest information from biomedical 'omics data.

The second and third tiers of training, whose purpose is two-fold, are offered to those looking to use proteomics more regularly in their research programme. Through termly hands-on training, small groups of researchers (up to 8) learn what types of workflows can be applied to address their specific research questions. They'll also be taught, through practical modules, how to process, run and appropriately analyse samples using the mass spectrometer. Once properly trained, researchers can then benefit from substantial price reductions by independently carrying out many of the steps that would otherwise be performed, and charged for, by the core facility. Advanced training and access is also offered for a select number of individuals within the CRE who use mass spectroscopy-based proteomics routinely as part of their research programme. This also includes DPhil students who wish to rotate in the Advanced Proteomics Facility as part of their DPhil training.

Apart from training, the initiative has also included a large financial contribution from the CRE, along with several other institutions, to purchase the most state-of-the-art mass spectrometer on the market, the Orbitrap Fusion™ Lumos™. This novel instrumentation offers many advantages over previous models, including increased sensitivity and resolution required to perform robust quantitation of proteins using multiplexing, as well as the ability to carry out deep mining of post-translational modifications. Now Oxford researchers will be able to perform proteomic analysis with greater detail and from new angles. In addition, the new Fusion™ Lumos™ will improve the ability to analyse complex samples, particularly those coming from the clinic which suffer from large heterogeneity and high dynamic ranges of protein expression, and make high-throughput analysis more feasible. These capabilities are particularly important given the emergence of personalized medicine and 'Big Data', which rely heavily on 'omics approaches to help with patient stratification and understanding disease mechanisms (1).

Introducing new methods and applications for proteomics, increasing access to cutting edge technology, and training our researchers to conduct, interpret and analyse proteomics datasets is essential to create a foundation on which we can continue to build a strong proteomics programme here at Oxford. In a 'Big Data' era, where much emphasis will be placed on 'omics tools to drive biomedical research, this foundation will be vital to helping researchers interface disciplines and work together effectively to tackle their scientific questions. It is our hope that this training scheme will aid in laying that foundation and will encourage more researchers to play an active role in their proteomics studies.

For those interested in finding out more about what training is currently being offered, how to access the financial benefits, or for general inquires you can visit the scheme's website at: http://www.cardioscience.ox.ac.uk/mass-spectroscopy-based-omics-training-and-facilities.

References:

1.Alyass A, *et al* (2015) From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med Genomics* 8: 33.

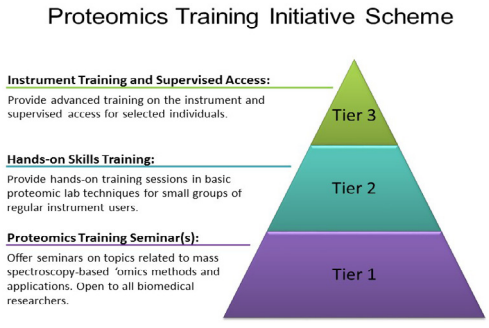


Figure 1: Overview of 3-tier training system offered as part of the Proteomics Training Initiative.

Dr Jillian Simon is a Postdoctoral Researcher in the Casadei group, Department of Cardiovascular Medicine

Induced pluripotent stem cell-derived neuronal cells:

by
Dr Cynthia Sandor

Despite significant research into neurodegenerative disorders including Alzheimer's (AD) and Parkinson's disease (PD), our understanding of the pathogenic mechanisms leading to neuronal cell death remains poor. This slow progress is mainly due to the inaccessibility of the human brain. Advances in induced pluripotent stem cell (iPSC) technology enable the generation of clinically relevant neuronal cells *in vitro* through the reprogramming of human somatic cells into pluripotent stem cells, which can then be re-differentiated into disease-specific cell types of interest.

In the last four years, multiple studies on iPSC-derived neurons generated from patients affected by monogenic forms of AD or PD have demonstrated that this cellular model can be useful for identifying altered cellular phenotypes. For example, Reinhardt *et al* found that the correction of *LRRK2-G2019S* genetic variant, the most common known genetic cause of PD, in iPSC-derived dopaminergic neurons generated from PD patients resulted in the rescue of aberrant cellular phenotypes (1). However, all studies published so far present two weaknesses: (i) no systematic evaluation of the cellular identity of the iPSC-derived neurons that were generated and (ii), in most cases, the cellular disease-associated phenotypes identified were already known or predicted, and no unbiased systematic investigation of the underlying molecular perturbation(s) was performed. Given the genetic predisposition being studied in these models, a valuable hypothesis-free tool to elucidate cellular molecular perturbations is to generate genome-wide transcriptomic profiles. Here, I will briefly explore some lessons learned from transcriptomic studies performed in iPSC-derived neuronal cell populations within Oxford University's (2) and StemBANCC's research programmes (3).

Lesson 1: iPSC-derived neurons are immature

The comparison of global transcriptional profiles of iPSC-derived neurons with those of post-mortem human brain tissue at different developmental stages shows that the iPSC-derived cultures are dominated by immature neuronal populations. This observation raises the question of how relevant this *in vitro* model is for neurodegenerative disorders, for which age-related changes in the brain are key to the pathogenic process. However, different iPSC-derived neuronal types appear to achieve different maturities *in vitro*, with dopaminergic neurons significantly more mature than cortical neurons. Indeed, a recent developmental atlas of gene expression in the Macaque found that the neocortex matured relatively very late (4), suggesting that driving the maturity of iPSC-derived cortical neurons will be a challenge.

Lesson 2: iPSC-derived neurons populations are heterogeneous

Analysis of transcriptomic profiles of technical replicates of iPSC-derived neurons shows that they vary in maturity, demonstrating significant cellular heterogeneity. This cellular heterogeneity can be a potential confounder in the comparison of gene expression profiles between iPSC-derived cell lines. To address this problem, there are currently two solutions. The first is to purify iPSC-derived neurons by using specific markers to the neuronal identify. For example, iPSC-derived human dopaminergic neurons can be FACS sorted using the intracellular marker tyrosine hydroxylase (5). However, this requires prior knowledge of neuronal markers expressed and that the cells are fixed prior to sorting, and thus killed, although this is perhaps not the case for an extracellular marker. A second approach is to employ single-cell sequencing approaches to examine the transcriptomes of individual cells from cultures and then to identify equivalent populations for comparative studies (6) (Figure 1A).

Lesson 3: Differential expression gene analyses allows the identification of molecular and phenotypic perturbations

Currently, a major issue for biologists is that the functions of most genes are poorly understood and it can be difficult to formulate hypotheses as to the perturbed molecular and cellular

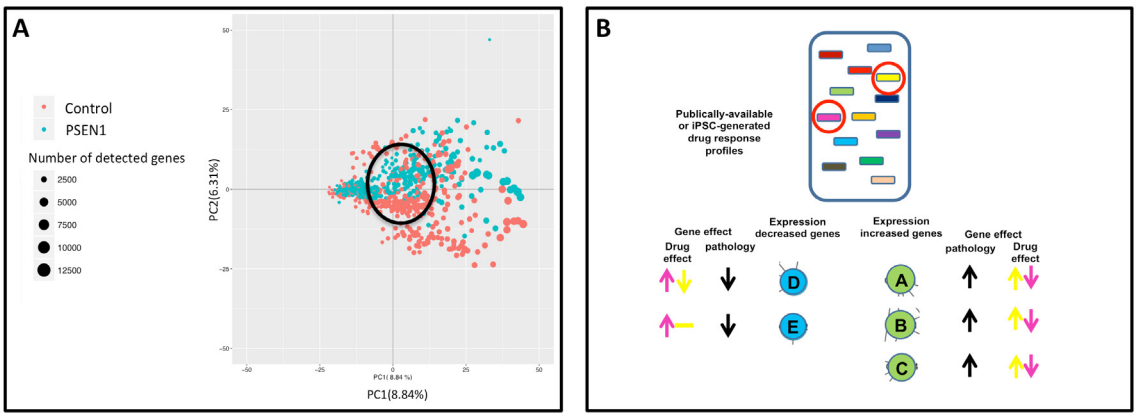


Figure 1: Transcriptomic analyses in iPSC derived neurons.
(A) Transcriptomic profile of individual iPSC derived cortical neurons of two cell lines. This graph illustrates the possibilities offered by this type of technology to identify differentially expressed genes between two iPSC lines (green and red) in a selected homogenous cell population (black circle) (figure from Volpato *et al.* (8)
(B) Given a disease transcriptomic profile, with some genes showing increased expression and other showing decreased expression, we can search publicly available drug response profiles for a drug that is known to have the opposite effect on cells. Conversely, some drugs have a gene expression profile similar to disease transcriptomic profiles (yellow drug) and can induce the disease.

process from a list of differentially expressed genes. To identify them systemically, it is possible to use gene network-based methods, which assume that disease-associated genes operate through a limited number of shared molecular processes and seek to detect patterns of functional convergence across a given set of genes. For example, we found that genes that were differentially expressed between *PD LRRK2-G2019S* converge functionally. We showed also that the abnormal motor capabilities/coordination/movement mouse phenotype annotation, coinciding with clinical manifestations of PD, is overrepresented in the annotations of this gene set (5).

Lesson 4: Therapeutic molecules that can reverse disease-associated molecular perturbations can be identified

One key aim following the identification of a molecular perturbation associated with disease is to identify a therapeutic molecule that might reverse that perturbation. The transcriptomic response of cells following exposure to a given drug is known for over 1000 compounds, and there are significant efforts underway to expand both the list of compounds and the cell types exposed. Given a gene expression signature corresponding to a molecular perturbation, the aim is to identify compounds that provoke an anti-correlated transcriptomic response (Figure 1B) reasoning that the drug-induced profile will cancel out the disease-induced profile. For example, in our work, the *PD LRRK2-G2019S* disease-associated profile was opposed by that induced by clioquinol, a drug known to rescue dopamine neuron loss and Parkinsonian behavioural phenotypes in mouse models (7).

These lessons taken together, despite significant challenges remaining, show that RNA sequencing profiles obtained from iPSC-derived cell types have the translational potential to identify both the perturbed molecular pathways and the therapeutic molecules that may act to ameliorate the effects, but care must be taken to consider both the maturity and cellular heterogeneity of iPSC-derived populations.

References:

1. Reinhardt P, *et al* (2013) Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell* 12(3): 354–367.
2. ODPC (2010) The Oxford Parkinson's Disease Centre. Available at http://www.everychildmatters.gov.uk/health/healthyschools/ [Accessed 10th August 2016]
3. StemBANCC (2012) Stem Cells for Drugs Discovery Available at http://www.stembancc.org/ [Accessed 10th August 2016]
4. Bakken T, *et al* (2016) A comprehensive transcriptional map of primate brain development. *Nature* 13;535(7612):367-375.
5. Sandor C, *et al* (2016) Transcriptomic profiling of purified patient-derived dopamine neurons identifies convergent perturbations and therapeutics for Parkinson's disease. *Hum Mol Genet* (in revision)
6. Stegle O, *et al* (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 16 (3):133–45.
7. Lei P, *et al* (2015) Clioquinol rescues Parkinsonism and dementia phenotypes of the tau knockout mouse. *Neurobiol Dis* 81:168-175.
8. Volpato V, *et al* (2016) Multi-omic experiments (in preparation)

Cynthia Sandor is a Postdoctoral Researcher in Dr Caleb Webber's group in the Department of Physiology, Anatomy and Genetics, Oxford

Metabolomics: a snapshot

by Professor James McCullagh

The understanding of biological systems has seen significant contributions from the application of 'omics' technologies in recent years. One of the youngest of these, 'metabolomics' aims to provide a window on global metabolic changes and is rapidly becoming an indispensable tool for discovering how small molecule chemistry integrates with cell biology. In this context, metabolomics has a tremendous potential for enabling a better understanding of disease processes and contributing to the development of medicines that are tailored to an individual's personalised cellular response (1).

The McCullagh lab, based in the Mass Spectrometry Research Facility in the Department of Chemistry, is focussed on the development and application of analytical methods for understanding small molecule chemistry at the interface with biology and medicine, with a special interest in metabolomics. We are currently developing analytical methodologies and performing metabolomics experiments with a number of collaborators across the University and beyond. This short article provides a brief snapshot of metabolomics in the context of an ongoing project to look at the metabolic impact of a single gene mutation.

On the screen in front of me is a database of over 18,000 animal and plant metabolites containing molecular structures, chemical formulae, isotope abundances and predictive fragmentation patterns. I am in the process of comparing information in the database with experimental data from liquid chromatography coupled to high mass accuracy tandem mass spectrometry (LC/MS/MS) in order to identify metabolites. In our sample extracts over 7000 compounds have been measured.

We are conducting a mass spectrometry-based metabolomics experiment comparing an immortalised glioblastoma cell line (LN18) with a mutated form in which the gene for isocitrate dehydrogenase (IDH1) has an amino acid change in the active site. This mutation is found in over 70% of low grade gliomas and secondary glioblastomas that have a heterozygous somatic mutation (R132H) in IDH1. Figure 1 shows a 3-D principal component analysis (PCA) of the 7000 compounds measured across all 18 samples. Samples clustering into 'wild type' and 'mutant' groups suggests compositional differences in compound abundance and composition. One of the beautiful aspects of metabolomics, and omics experiments generally, is that they can capture very large amounts of data in a relatively unbiased way; allowing hypotheses to be generated as well as tested. I am currently in the process of mapping the changes observed onto metabolic pathways in order to determine how expression of the mutant enzyme affects metabolism. Before discussing the results further I will provide a brief overview of metabolomics, some of its current challenges and the scope for future developments and applications.

Why metabolomics?

Metabolites in a biological system are collectively known as the metabolome. The idea of taking a snapshot of the entire metabolome, to learn about changes in metabolism in response to experimental changes, encapsulates the process of metabolomics (2). In the context of disease for example, the metabolome can reflect both up stream genetic changes and influences from the external environment. The genome, transcriptome and proteome dictate what metabolism is possible but small molecules, which may have an endogenous or exogenous source, can also modulate gene activity and protein expression. The metabolome therefore is closest to representing a cell's molecular phenotype, the product of the interaction between genes and the environment (3). This makes metabolic changes a useful indicator of how a cell responds to disease processes and the potential to find vulnerable pathways suitable for therapeutic intervention.

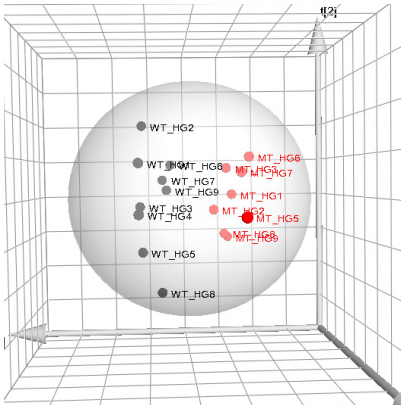


Figure 1. 3D Principal component analysis (PCA) showing analysis of over 7000 compounds across 9 wild type (WT) and 9 mutant (MT) samples. These data suggests the WT and MT groups are significantly different in chemical composition.

The experimental workflow is conceptually straightforward, analogous to that of proteomics and critical to successful results (Figure 2). Extraction of metabolites can be particularly challenging due to the significant chemical differences between some metabolites and the lack of common monomeric units like those found in proteins and DNA. Additionally, large differences in polarity can lead to challenges in choosing suitable extraction solvents which can bias the experimentally accessible metabolome (4).

Analytical developments

Although the proteome and genome are larger entities, the human metabolome, with an estimated size of 3000-5000 metabolites, is particularly challenging to capture analytically. This is mainly due to the chemical diversity and wide range in compound abundances represented in biological systems, however the sensitivity and speed of LC/MS/MS makes it an increasingly important analytical approach for metabolomics. The McCullagh lab has developed the use of ion-chromatography interfaced directly to mass spectrometry via an electrolytic ion suppressor which overcomes many of the problems traditionally associated with coverage or highly polar and ionic compounds. This enables us to focus on pathways of central metabolism, which amongst others include glycolysis, the citric acid cycle, the pentose phosphate pathway, purine and pyramidine metabolism and multiple pathways associated with amino acid metabolism. We are also working on complimentary analytical approaches for lipid and fatty acid analysis.

Identification of metabolites

The identification of large numbers of metabolites in a single sample requires robust, multiple, independent measurements for each compound. We use accurate mass measurement, fragmentation patterns, isotope ratios and chromatographic retention times for each compound. Such measurements are made for every compound in experimental samples and compared to a database of values obtained from the analysis of authentic standards. Recently we received funding to create a database containing 500 metabolites. A list of those currently in our database can be found on the Mass Spectrometry Facility website (5).

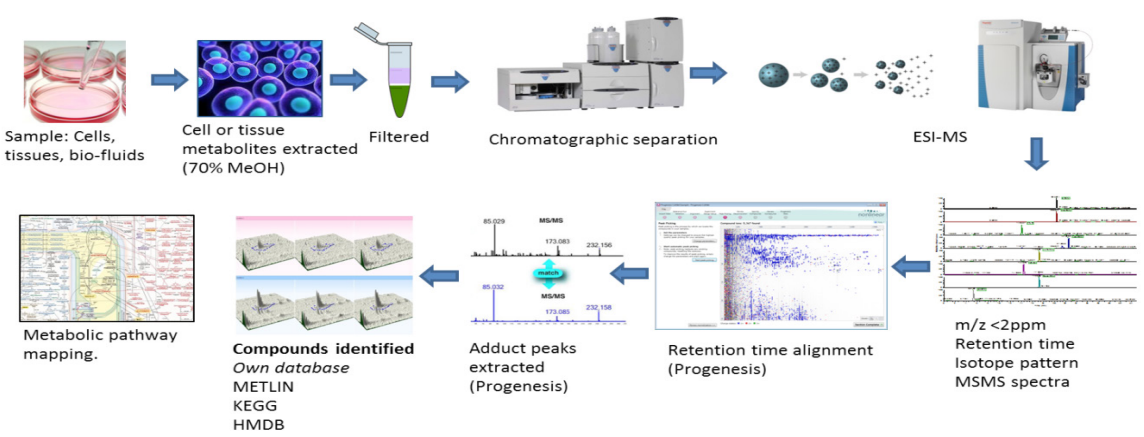


Figure 2. A metabolomics workflow which can be divided into four parts: 1) Sample preparation, 2) Sample analysis, 3) data processing, and 4) data interpretation.

Interpreting Results

The general metabolomic experiment is usually characterised by a large number of variables (metabolites) and a relatively small number of samples. Therefore multivariate analysis is required to ensure that any differences between experimental groups are statistically significant. Because ion abundances are recorded for all compounds this can be performed on identified and unidentified compounds and usually involves both unsupervised (PCA) and supervised discriminant analysis (PLS-DA or OPLS-DA). PCA analyses show trends and outliers within the dataset (as in Figure 1), while supervised approaches provide a model that ranks the compounds according to their contribution to the differences observed between the experimental groups considered (6,7).

There are usually two main aims for the analysis of metabolomics data. The first is the identification of statistically significant biomarker compounds. The second, which is really an extension of the first, is to identify and represent metabolic changes in a biologically meaningful context. This often involves identifying which metabolic pathways have been affected by comparing changes in the abundance of their constituent metabolites (8).

Results of the LN18 mutant and wild type experiment

The experiment introduced at the start was conducted in order to investigate the metabolic impact of a specific IDH mutation. Untargeted differences in metabolite profile were observed between the two experimental groups and displayed in a PCA plot. From our database 220 compounds were identified. Multivariate statistical analysis using OPLS-DA was conducted to create a model of the differences observed and a list of statistically significant potential metabolite biomarkers associated with the mutation (unpublished, data not shown).

One compound in particular, 2-hydroxyglutarate (2-HG), showed a 432 fold change between the wild type and mutant cell line. This confirmed previous findings, from both in vitro and *in vivo* studies, that showed the IDH1 enzyme, which normally converts isocitrate to α-ketoglutarate in the TCA cycle, has a change of function associated with the mutant form (9,10). Additional LC/MS analyses showed that 2-HG formation was enantiomerically selective, with the D (and not L) enantiomer being significantly increased in mutant cells (Figure 3). Although there are dramatic increases in the concentration of 2-HG in mutant cells and tumours, its tumorigenic function remains unclear. One of the aims of our research is to use targeted metabolomics to identify additional metabolic changes associated with the mutation.

Conclusions

Many complex diseases with a significant impact on modern society, including cancer, diabetes and dementia, have complex genetic mutations which ultimately manifest in metabolic effects that remain relatively little explored at a systems level (11). Understanding these genetic changes helps show where potential alterations in cellular function may occur but it is at the metabolic level where the effects can be interpreted in the context of cellular function. It is still early in its development but metabolomics has exciting potential to play an important role in better understanding disease processes and their treatment in the future.

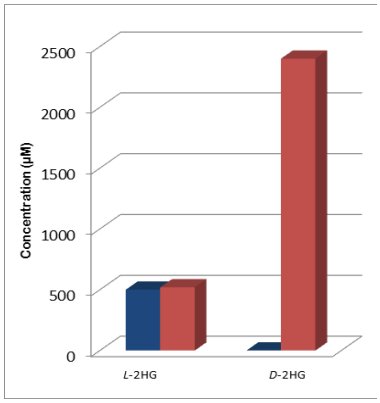


Figure 3. Comparison of the abundance of both 2-hydroxyglutarate enantiomers found in the wild type and mutant cells. The mutant cells show an over 400 fold increase in D-2-HG compared to the L form.

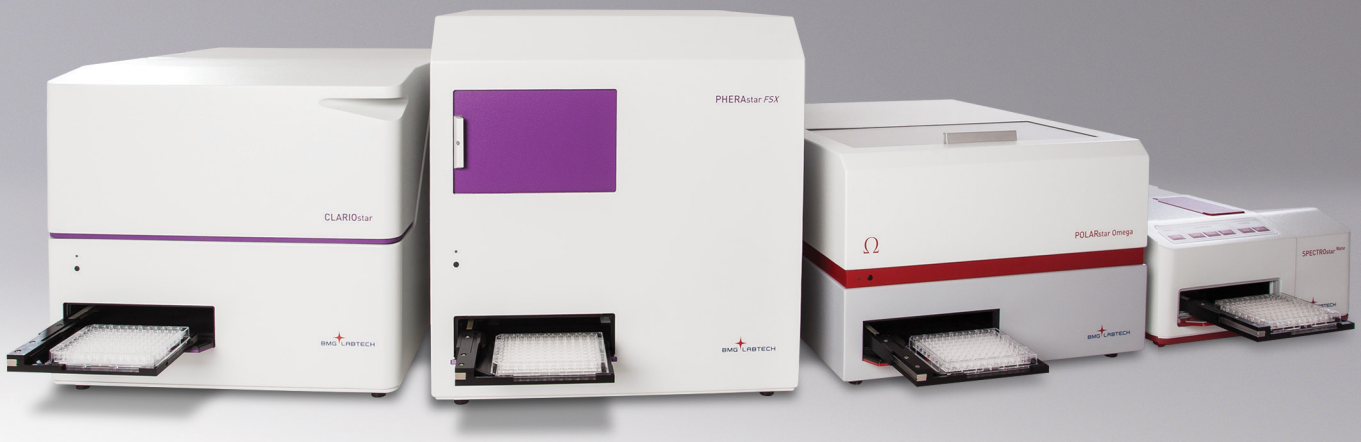
References

1 Wishart DS. (2016) Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov*15: 473-484.
2.Patti GJ, *et al* (2012) Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Bio.* 13:263-266.
3. Ryan D and Robards K. (2006) Metabolomics: The Greatest Omics of Them All? *Anal. Chem.* 78: 7954-7958.
4.Want EJ, *et al.* (2013) Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc.* 8(1) 17-32.
5. University of Oxford (2008) Mass Spectrometry Research Facility. Available at http://www.chem.ox.ac.uk/spectroscopy/mass-spec/index.html [Accessed 4th October 2016]
6. Johnson C, *et al.* (2016) Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 17:451-459.
7. Sugimoto M, *et al.* (2012) Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. *Current Bioinformatics.* 7 (1): 97-108.
8. Pirhaji L, *et al.* (2016) Revealing disease-associated pathways by network integration of untargeted metabolomics. *Nat Methods.* 13(9): 770-776.
9. Dang L, *et al.* (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature.* 462: 739-745.
10. Cairns RA and Mak TW. (2013) Oncogenic Isocitrate Dehydrogenase Mutations: Mechanisms, Models and Clinical Opportunities. *Cancer Discovery.* 7: 730-740.
11. Dang CV. (2015) Links between metabolism and cancer. *Genes Dev.* 26:877-890.

James McCullagh is an Associate Professor and Head of the Mass Spectrometry Facility in the Department of Chemistry, University of Oxford

The BMG LABTECH All Stars

Innovative, high-performance microplate readers for all assay needs



CLARIOstar®

The most sensitive monochromator-based microplate reader. Equipped with revolutionary LVF monochromators™, it is the perfect reader for flexible assay development.

PHERAstar® *FSX*

The new gold standard for High Throughput Screening. The PHERAstar *FSX* is the new reference multi-mode plate reader, combining highest sensitivity with the fastest read times.

Omega series

Upgradeable single to multi-mode filter-based microplate readers. The Omega series offers a combination of flexibility and performance for any life science application.

SPECTROstar® *Nano*

Microplate reader with ultra-fast detection of UV/Vis absorbance. Spectra from 220 - 1000 nm are detected in less than one second per sample in microplates and cuvettes.

Find all our microplate readers on www.bmglabtech.com

BMG LABTECH

The Microplate Reader Company