Investigating the metabolic consequences of protein misfolding in the yeast *Saccharomyces cerevisiae*

by

Tyler Louisa Howell-Bray

Canterbury Christ Church University

2019

Acknowledgements

I would like to express the sincerest gratitude to my supervisor Dr Lee Byrne, his patience, kindness, and unshakable faith in me has kept me motivated at the low points of my studies and kept me elevated during the highs. I am eternally grateful for his guidance, his knowledge, and our lunches. He is exactly the type of supportive and gentle supervisor, I needed, and I am very grateful to have had him on my side throughout this time.

I would also like to extend my thanks to Professor Simon Harvey and Professor Peter Vujakovic, for providing me with thought provoking and inspiring comments, affording me the opportunity to widen my understanding.

In addition, I would like to thank Canterbury Christ Church University, Dr Lee Byrne and Professor Simon Harvey for giving this school dropout, working class Londoner a chance to change her destiny.

My sincere thanks also go to Biomolecular Science Facility Manager, Kevin Howland, for granting me access to his laboratory and mass spectrometers and to Dr Kristopher Leslie for his assistance and patience, granting me server access and helping me to acquire the skills to navigate Linux. Without their support this research would not have been possible.

I would like to give thanks to my lab mates and colleagues here at Canterbury Christ Church University, almost everything I know is as a direct result of your hours of putting up with me. I have been truly blessed to have been part of such a knowledgeable and caring team, your smiles, and encouragement have played a huge role in my completion. Notably, my fellow instructors, the stability and support that we have provided for each other has been unthinkable and without most of you climbing this often seemingly unattainable mountain would have most days appeared insurmountable.

My heartfelt appreciation also goes to my lab volunteers Alex Neale (Monkey) and Angela Lucas, although our time in labs has been brief your assistance and our laughs at some of the toughest moments have been invaluable.

To all the students I have met and taught along the way, your wide-eyed stares at the thought that someone like you could achieve something like this have kept me going, that and your constant questioning about if I am finished yet. Honestly, I hope I have demonstrated that anything is possible if you work hard enough, dream big guys, always.

I would also like to thank the likely hundreds of scientists that have worked to create both excellent and freely available bioinformatic software for use within metabolomics, without you none of this would have been possible.

Last but not least, I would like to give thanks to my family: my husband, my mum, my dad, my sisters, my aunties, my cousins and my in-laws for their never-ending faith in what I am capable of and for always keeping me on my toes.

For those lost and those gained

For my Nanny, Joyce Louisa Bray

The clever little girl that won a scholarship to go to grammar school only to be denied it due to WWII, your encouragement, support, and general nagging has been instrumental to the scientist that I have become. I am deeply saddened that you are not here for me to read this down the phone to or to see me submit, and yet I know that you would be incredibly proud.

For my Nephew and Niece, Issac and Darcey

Welcoming you into the world is without question one the greatest honours of my life, you've kept things exciting and smashed the odd key whilst I have been writing. I hope that this inspires you both to go on to do whatever fills you with happiness, safe in the knowledge that you are capable and that you are loved.

Abstract

Protein misfolding disease can be loosely separated into two groups: the non-transmissible amyloid diseases (e.g. Alzheimer's and Huntington's) and the transmissible amyloid diseases (e.g. Scrapie and Kuru). The amyloid and prion diseases present the same misfolding mechanism and thus prions can be used to recapitulate the biochemical hallmarks of amyloid disease. The budding yeast *Saccharomyces cerevisiae* is a popular and valuable model for the study of prions. Whilst much progress has been made over the last century, our understanding of the specific mechanisms which underpin amyloid and prion disease is still incomplete.

The aim of this study was to enhance current understanding about the metabolic consequence of misfolded proteins. Yeast strains carrying different conformational variants of the known prion forming Rnq1 protein were used to obtain metabolic profiles and identify key perturbations. This approach began by determining a metabolomic method suitable for use in *S. cerevisiae*. Ultra-High-Performance Liquid Chromatography Mass Spectrometry (UHPLC-MS) was used to establish a sample preparation method that accurately revealed the metabolic state of *S. cerevisiae*. It was found that culturing yeast cells in a liquid medium and extracting metabolites with an acetonitrile: water (50:50) mix most accurately reported on the biological conditions imposed. The endogenous cellular role of the Rnq1 protein was studied, using a strain of *S. cerevisiae* in which the *RNQ1* gene had been deleted (Δ*rnq1*). A robust data analysis methodology was established and applied to the data obtained, utilising cross comparison of two widely used metabolomics analysis programs. Then, biomarkers and metabolic pathways associated with the presence of the Rnq1 protein were investigated, comparing [*RNQ*$^{+}$] and [*rnq*$^{-}$] cells. The toxicity of Rnq1 protein overexpression in a [*RNQ*$^{+}$] background was explored, via the expression of the *RNQ1* gene.

These studies reveal that the presence of the Rnq1 protein downregulates the ubiquinone biosynthesis pathways within cells, suggesting that the Rnq1 protein may play a lipid/mevalonate-based cytoprotective role as a regulator of ubiquinone production. Distinct perturbations in sphingolipid metabolism were observed in [*RNQ*$^{+}$] cells, with significant downregulation in metabolites within these pathways, providing new evidence of metabolic similarities between yeast and mammalian cells as a consequence of prion presence. Metabolic perturbations relating to general and specific stress responses caused by oxidative stress in the presence and absence of prions were also obtained. This work establishes the application of metabolomics as a tool to investigate prion-based phenomena.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACN	Acetonitrile
AD	Alzheimer's Disease
ALS	Amyotrophic Lateral Sclerosis
BE	Boiling Ethanol
BSE	Bovine Spongiform Encephalopathy
CJD	Creutzfeldt-Jakob Disease
CM	Chloroform: Methanol
COQ	Coenzyme Q
COSHH	Control of Substances Hazardous to Health
Cpd	Compound
CPEB3	Cytoplasmic Polyadenylation Element-Binding Protein 3
CRM	Charged Residue Module
CWD	Chronic Wasting Disease
DIA	Data Independent Acquisition
DNA	Deoxyribose Nucleic Acid
DPPS Score	Pathway Perturbation Within Multiple Data Sets
ECL	Enhanced Chemiluminescence
EIC	Extracted Ion Chromatogram
ER	Endoplasmic Reticulum
ESI	Electrospray Ionisation
FDR	False Discovery Rate
FFI	Fatal Familial Insomnia
GAL	Galactose
GC-MS	Gas Chromatography Mass Spectrometry
GdnHCl	Guanidine Hydrochloride
GMO	Genetically Modified Organism
GSS	Gerstmann-Straussler-Scheinker Syndrome
HILIC	Hydrophobic Interaction Liquid Chromatography
HIS	Histone
HMDB	Human Metabolome Database

HPLC	High Performance Liquid Chromatography
HRP	Horse Radish Peroxidase
Hsp	Heat Shock Protein
HSQC	Heteronuclear Single Quantum Coherence
HSR	Heat Shock Response
HX	Hydrogen Exchange
ID	Identification
IDP	Intrinsically Disordered Protein
IDR	Intrinsically Disordered Region
IEM	Ion Evaporation Mechanism
LC-MS	Liquid Chromatography Mass Spectrometry
m/z	Mass to Charge
mRNA	Messenger Ribose Nucleic Acid
MS	Mass Spectrometry
ND	Neurodegenerative Diseases
NIM	Negative Ionisation Mode
NMR	Nuclear Magnetic Resonance
NO	Nitric Oxide
NPC	Normal Phase Chromatography
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
OPR	Oligopeptide Repeat Region
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PD	Parkinson’s Disease
PFD	Prion Forming Domain
PIM	Positive Ionisation Mode
PLS-DA	Partial Least Squares Discriminant Analysis
PM	Pure Methanol
PPP	Pentose Phosphate Pathway
PRD	Prion Domain
Prion	Proteinaceous Infectious Particle

Q/N Rich Region	Asparagine and Glutamine Rich Region
QC	Quality Control
QToF	Quadrupole Time of Flight
RF	Radio Frequency
RNA	Ribose Nucleic Acid
ROC	Receiver Operating Characteristic
ROS	Reactive Oxygen Species
RPC	Reversed Phase Chromatography
SC	Synthetic Complete
SD	Synthetic Defined
SNP	Single Nucleotide Polymorphism
ssDNA	Sheared Salmon Sperm DNA
TIC	Total Ion Chromatogram
TMS	Tetramethylsilane
ToF	Time of Flight
TSE	Transmissible Spongiform Encephalopathy
UHPLC	Ultra-High-Performance Liquid Chromatography
UPR	Unfolded Protein Response
UPS	Ubiquitin-Proteasome System
USA	Ureidosuccinate
YMDB	Yeast Metabolome Database

Chapter One - Introduction

1.1. Prions

1.1.1. Protein folding

The term ‘protein folding’ refers to the process whereby a protein acquires its native conformation from a completely unfolded state. Efficient protein folding is critical to an organism’s health, however the complexities of how cells orchestrate this folding has long remained a mystery.

Genes, via translation, dictate the size and order of the 20 commonly occurring amino acids available with which to build a simple, linear, polypeptide. However, in reality these polypeptides are anything but simple given the incomprehensible magnitude of the variety of amino acid combinations and chain lengths that exist. Synthesis of polypeptides is not enough to create functional proteins and despite their already abundant complexity, these polypeptides fold into three-dimensional conformations aided by a large number of catalysts and molecular chaperones (Anfinsen *et al.* 1961). Even after this process, the native structure of the protein remains undetermined. Quite how proteins achieve their native conformations from the relatively simplistic genetic codes underlying this process remains one of the most difficult questions of modern science (Dill and MacCallum, 2012; Vendruscolo *et al.* 2003; Dobson 2003).

1.1.2. Understanding and modelling protein folding

It was once thought that proteins fold through a distinct pathway via distinct intermediate states (Figure 1.1. A). This classical or Levinthal view was drawn on the basis that undirected folding would result in randomised searching for proteins to take their native form through an extraordinary number of structural possibilities, given that this would take an extraordinarily long time, it was concluded that proteins must fold via predetermined pathways (Levinthal, 1968). However, upon experimentation, common intermediate structures could not be identified to support this theory and so further development resulted in the ‘new view’ hypothesis that moved the model from the two-dimensional pathway to a three-dimensional funnel-shaped energy landscape (Dill and Chan, 1997; Baldwin, 1994; Anfinsen *et al.* 1961). This hypothesis was based on the concept that the native form of a protein corresponds to the most stable form available given its structure and cellular conditions. In essence, the relative position and charge-charge interactions between amino

acid residues carry with them an inherent energy cost; therefore, those interactions that result in the lowest energy structure shape the folding/energy landscape and hence result in the native conformation (Figure 1.1. B) (Makhatadze, 2017). Although *in vitro*, this mechanism of folding occurs almost instantaneously, it is understood to be a fluid process requiring many stages including the formation of a folding core nucleus made up of a very small number of key residues within the polypeptide. This core then forces the surrounding amino acids into a native like form, coalescing the remaining protein around the nucleus and hence acquiring the native structure (Fersht, 2000). An important consideration within this process are the thermostable forces that shape the protein folding landscape which are varied and complex. Often there is only a modest gain in free energy associated with the acquisition of the native form compared with the still relatively low energy of a variety of misfolded states (Lindquist and Kelly, 2011).

Experimental evidence for such folding behaviour has been extremely difficult to obtain, as the rate of decay of most intermediates is less than 1 second and well beyond the capabilities of most structural biology methods such as Nuclear Magnetic Resonance (NMR) and crystallography. However, thanks to advancing computational power, NMR, Mass spectrometry (MS) technology and Hydrogen Exchange (HX) pulse labelling have been employed to monitor hydrogen exchange, making it possible to explore these theories macroscopically (Englander and Mayne, 2011). Interestingly, this research suggests that proteins are composed of separately cooperative ‘foldon’ building blocks which repeatedly fold and unfold throughout their journey to their native form (Englander and Mayne, 2011). Englander and Mayne, (2011) states that these experiments show ‘foldons’ providing structural guidance and a free energy bias which ultimately leads to native formation, supporting the ‘new view’ hypothesis. However, the formation of ‘foldons’ in their small cooperative units would not fulfil the energy bias required by such a model, suggesting that protein folding theories may necessitate one another and ultimately that protein folding may occur via a combination of both theories.

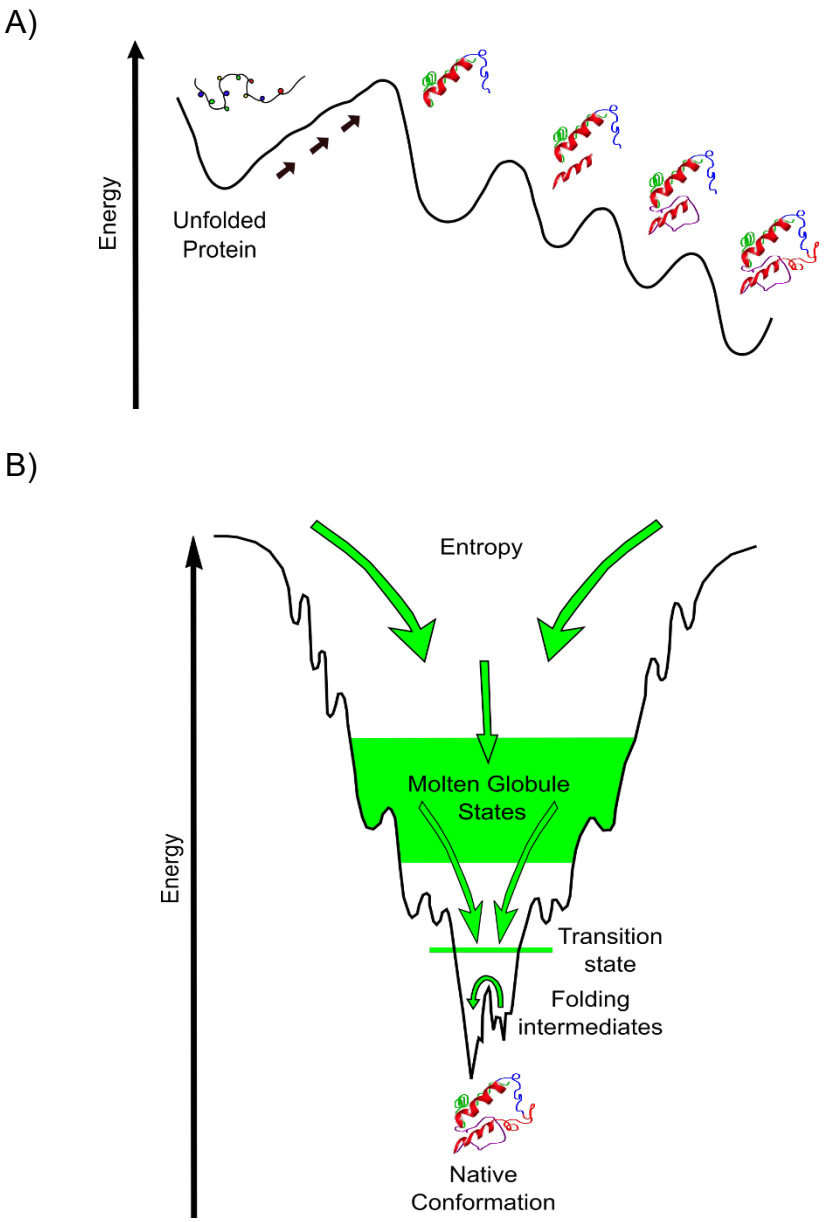


Figure 1.1. The classical model of protein folding alongside the Anfinsen thermodynamic model of protein folding within the context of energy requirement. A) Shows the classical view or model of defined protein folding B) Shows the Anfinsen thermodynamic model of protein folding (adapted from Englander and Mayne (2011))

It is important when considering the native conformation of a protein that it may operate a suite of functions within cells and thus will need to be flexible and dynamic. Thus, the 'correct' shape for a protein to carry out its role within the cell may be inherently disordered. Whilst some structured proteins require a well-defined three-dimensional shape to fulfil their cellular roles, it is possible for proteins to adopt a far more disordered structure and still function within cells (Tompa, 2012). Proteins range in shape from structured proteins, which adopt well-formed tertiary structures, to intrinsically disordered proteins (IDPs), which lack a tertiary structure of any kind (Figure 1.2.) (van der Lee *et al.* 2014).

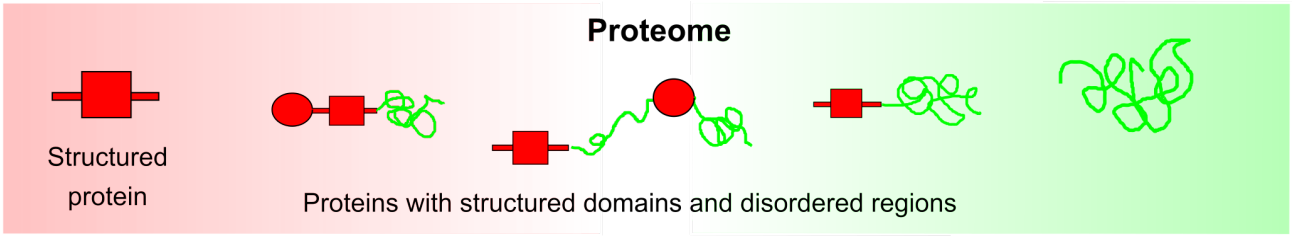


Figure 1.2. Five key areas in the spectrum of possible protein structures ranging from structured to disordered. From left to right the variation in protein structure from well-defined three-dimensional conformations to intrinsically disordered proteins (adapted from van der Lee *et al.* (2014))

IDPs and their intermediates: intrinsically disordered regions (IDRs), are not capable of forming tertiary structures due to the presence of non-biochemically compatible amino acids occurring within close proximity of each other (van der Lee *et al.* 2014). Whether IDPs or IDRs form generally depends on the size of this region within the protein. IDP formation is normally found when the protein lacks a region of bulky hydrophobic amino acids and hence is unable to form a folding core nucleus (Romero, 2001). Once considered to be solely associated with disease phenotypes, investigation has showed IDPs and IDRs to be present in large numbers within protein-encoded sequences of the human genome. In fact, 44% of open reading frames contain disordered regions of greater than 30 amino acids (Oates *et al.* 2013). The prevalence of these disordered regions presents challenges to our understanding of protein function and the classical ideas about protein structure. It is apparent that whilst knowledge about protein folding continues to increase, much remains unknown, uncharacterised, and misunderstood about this fundamental process (van der Lee *et al.* 2014; Oates *et al.* 2013; Babu *et al.* 2012).

1.1.3. Protein folding complications

It is evident that proteins face many trials throughout their journey to their native conformation. For example, *in vivo*, protein folding is unaided by the relative 'pollution' of the crowded cellular environment. The requirement for this complex process to be completed all the while being bombarded by macromolecules such as proteins, polysaccharides and lipids unsurprisingly causes complications (White *et al.* 2010; Ellis and Minton, 2006). These obstacles mean that many proteins fail to reach their native conformations or misfold into stable non-native ones.

Furthermore, protein folding occurs in a variety of cellular locations, some of which are highly specialised. These different cellular environments, such as the ER (endoplasmic reticulum) and the mitochondria, are incredibly variable in their biochemical nature and the cell must tackle these changing environments with each presenting a different set of problems for protein folding (Valastyan and Lindquist, 2014; White *et al.* 2010; Ellis and Minton, 2006).

1.1.4. Protein folding solutions

Many cellular defence mechanisms exist against misfolded proteins. One of these defence strategies is the constitutive expression of molecular chaperones throughout the cell. Playing a key role in correct protein folding and aiding misfolded proteins to regain their correct conformation, molecular chaperones respond to perturbations in protein homeostasis and are dynamically upregulated in response to the accumulation of misfolded or unfolded proteins (Kim *et al.* 2013). This response occurs across the cell but the most well characterised of these 'regulation events' occurs in the ER and the nuclear and cytosolic compartment. Known as the unfolded protein response (UPR) and heat shock response (HSR) respectively, it was thought that the upregulation of molecular chaperones was an emergency-based response which occurred as a result of environmental stress. Consequently, it is now understood that this is an ever-present dynamic process, constantly monitoring and responding to small changes in protein homeostasis (Hartl *et al.* 2011).

When the role of the molecular chaperone becomes obsolete and the misfolded protein cannot be refolded into its native form, secondary systems such as autophagy and the ubiquitin-proteasome system (UPS) are deployed to break down and recycle these misfolded proteins. This is not performed lightly; proteins take considerable energy to build, hence their breakdown is a costly but necessary expenditure. Primarily the UPS acts as the protein degradation system for short-lived proteins, fine tuning the background levels of regulatory proteins and maintaining the availability of amino acids during times of high stress (Nedelsky *et al.* 2008). Autophagy has historically been thought to play a role in the degradation of large proteins mainly during chronic stress but recently it has been found that autophagy plays a role in the defence against diseases caused by protein misfolding, equal to or greater than the UPR. Further work to understand these complex mechanisms is ongoing (Dasuri *et al.* 2013; Nedelsky *et al.* 2008).

1.1.5. Protein misfolding and aggregation

Given the inherent complexity and breadth of the process of protein folding, there is ample opportunity for mistakes within this process to occur. Importantly these mistakes may have the opportunity to lead to a disease phenotype (Valastyan and Lindquist, 2014; Dobson, 2003). The energy landscape which sculpts protein folding is a sensitive and dynamic process and perturbations can have major effects resulting in protein misfolding events. Such misfolding events are fundamentally rooted in pathological conditions (Figure 1.3.) (Valastyan and Lindquist, 2014; Kim *et al.* 2013).

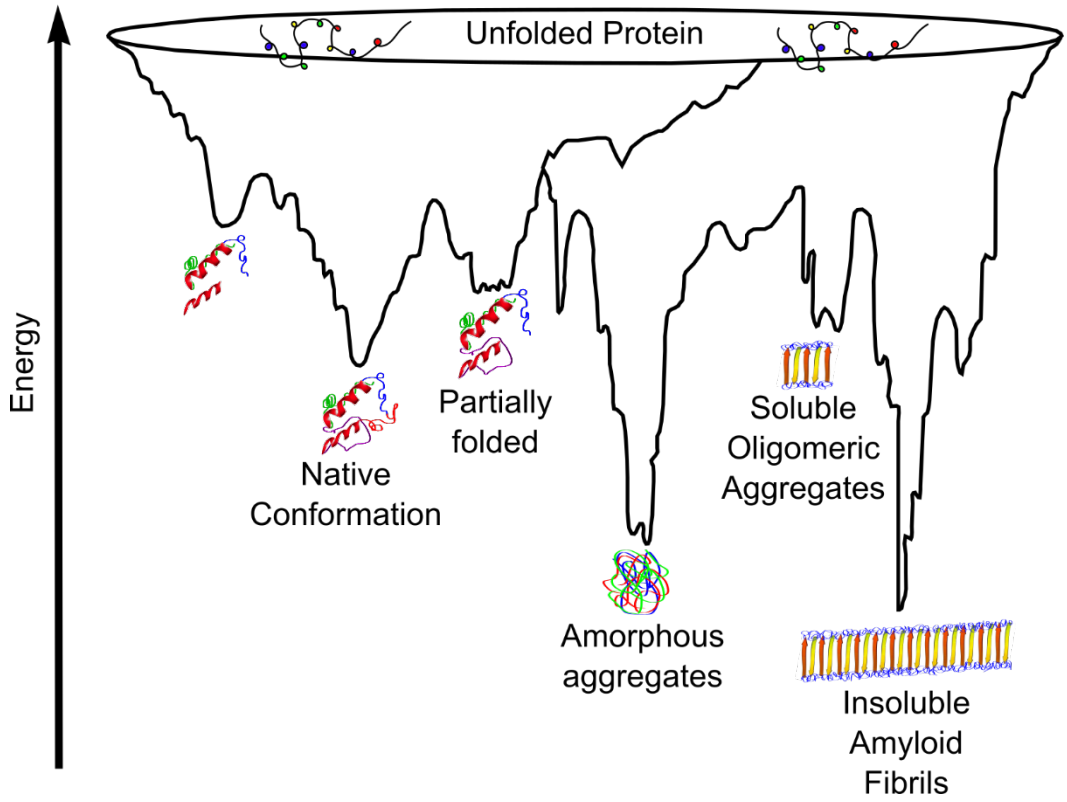


Figure 1.3. Six of the main protein structures ranging from unfolded to insoluble amyloid fibrils in the context of the cellular energy landscape. Depicts the energy landscape of protein folding and aggregation. Native state formation is shown on the left and 'other' non-native state formation is shown on the right. (adapted from Raskatov and Teplow (2017))

Destabilising mutations within the protein sequence (sometimes appearing as single nucleotide polymorphisms or as large repeat areas of genome) are a known cause of protein misfolding. Additionally, the IDRs which are now considered a 'normal' part of protein folding, can sometimes be associated with a disease phenotype (Valastyan and Lindquist, 2014;

van der Lee *et al.* 2014; Tompa, 2012). In general, it has been found that protein misfolding occurs when the native structure of the protein is perturbed by underlying mutations in the corresponding genetic code, either introducing or increasing the number of non-biochemically compatible amino acids occurring within close proximity to one another.

Non-biochemical compatibility is defined here as a region within a protein were there are many polar amino acids positionally near many hydrophobic amino acids, creating instability within a protein's structure. Many of these regions are known to be rich in the polar amino acid's asparagine, glutamine, and serine. It is known that the introduction of these amino acids (especially in high quantity, through large repeat sections of genome) reduces the solubility of the protein in question and increases the number of hydrogen and salt bridges throughout the protein, affecting secondary structure and making it much harder for the cell to degrade (Trevino *et al.* 2007; Perutz *et al.* 2002).

Table 1.1. Nature, name and associated three and single letter codes of the 20 most common amino acids.

Nature	Name	Three letter code	Single letter code
Charged	Arginine	Arg	R
	Lysine	Lys	K
	Aspartic Acid	Asp	D
	Glutamic Acid	Glu	E
Polar	Glutamine	Gln	Q
	Asparagine	Asn	N
	Histidine	His	H
	Serine	Ser	S
	Threonine	Thr	T
	Tyrosine	Tyr	Y
	Cysteine	Cys	C
	Methionine	Met	M
	Tryptophan	Trp	W
Hydrophobic	Alanine	Ala	A
	Isoleucine	Ile	I
	Leucine	Leu	L
	Phenylalanine	Phe	F
	Valine	Val	V
	Proline	Pro	P
	Glycine	Gly	G

The 20 most common amino acids, their names, three letter code and single letter code listed by their biological nature

Often with IDRs that cause mammalian disease, the normal native form of the protein will have been faithfully made by cells for years and then, via causes unknown, the IDRs suddenly start to have an effect on the overall structure of the protein causing it to misfold into an abnormal conformation. The presence of these misfolded proteins not only causes the protein to lose its functional role within the cell, but it initiates a cascading reconfiguration of the remaining native protein, causing misfolding and acquisition of the abnormal conformation. These misfolded proteins then aggregate in cells in many oligomeric stages, eventually leading to the formation of large amyloid fibres (Figure 1.4.).

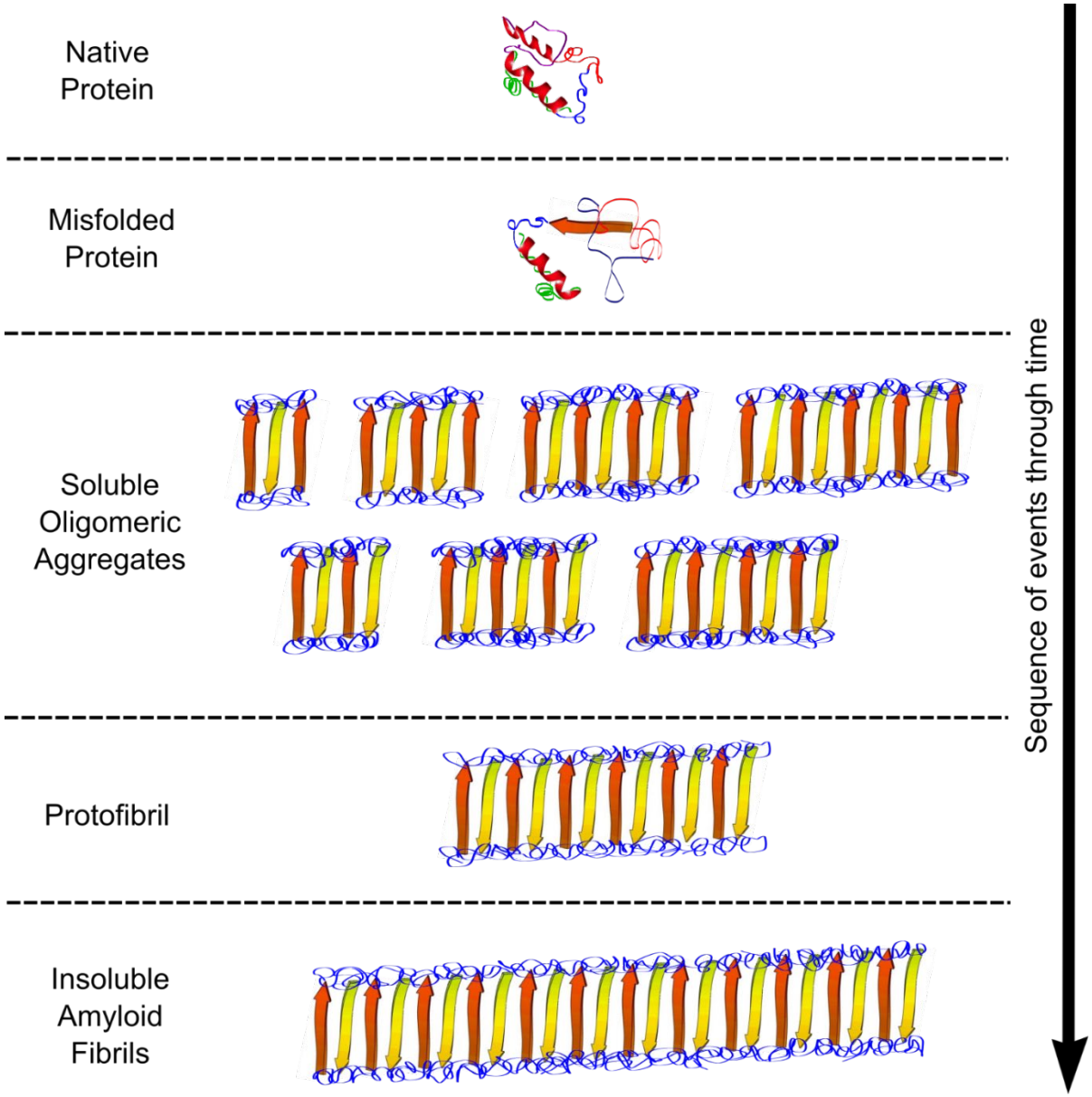


Figure 1.4. The sequence of events through time that occur when a protein misfolds. This figure depicts the current understanding of the sequence or mechanism of protein misfolding events through time which ultimately lead to amyloid formation.

These so called ‘amyloids’ (meaning starch-like) appear (as visualised by iodine staining under light microscopy) as fibril or thread-like (Hardy and Selkoe, 2002; Caughey and Lansbury, 2003). Closer examination using cryo-electron microscopy and solid-state NMR has revealed that amyloid fibrils are a few nanometres in diameter but can be micrometres in length. They are highly ordered in a cross-β pattern and all amyloids are surprisingly generic irrespective of the native protein which has misfolded to create them (Fitzpatrick *et al.* 2013; Dobson, 2003; Petkova *et al.* 2002). This finding suggests that the primary amino acid structure has little bearing on the overall architecture of the protein, causing much questioning to arise regarding our understanding of protein structure and indeed even the central dogma.

As amyloids have a large number of intermolecular contact and hydrogen bonds, they have very high levels of kinetic and thermodynamic stability and it has been suggested that the amyloid rather than being a rare disease-related phenomenon is simply an alternate state which may be acquired by any protein (Cremades *et al.* 2012). Experimentation *in vitro* has found that this is indeed the case and that the generic amyloid cross-β lateral structure can be adopted by any protein (Auer *et al.* 2008). It appears that this structure is favoured when high concentrations of the protein in question are present. This leads to the surprising discovery that the ‘critical’ concentration required for protein X to form amyloids *in vitro* is less than the concentration of protein X present *in vivo*, and yet in cells, the aggregates do not form. The explanation for this observation is that whilst the native state of a protein may not represent the global free energy minimum that the amyloid does, it may represent the local free energy minimum *in vivo*, or alternatively/additionally, that kinetic barriers to the amyloid state prevent formation (Baldwin *et al.* 2011).

1.1.6. Aggregation theories

Initially a lack of molecular knowledge underpinning how amyloids formed led to the conclusion known as ‘the amyloid hypothesis’. This stated that these large, protease resistant fibres are the cause of the toxicity associated with disease. However, definitive links between amyloid fibres, toxicity and disease phenotype have never been conclusively established, causing this theory to dwindle (Hardy and Selkoe, 2002).

A now widely accepted alternate theory, ‘the oligomeric hypothesis’ (Cremades *et al.* 2012), postulates that amyloid formation is a mechanism employed to aid cell survival, suggesting

that during the assembly process of amyloids, toxic intermediates are generated. These toxic oligomeric intermediates are considered to be the most damaging to cells (Cremades *et al.* 2012; Campoioni *et al.* 2010; Haass and Selkoe, 2007). However, given the variety of size and structure of oligomers throughout the intermediate stages of amyloid formation, experimental evidence for the causative toxic protein species that may introduce disease and be responsible for transmission remains unattainable (Karran *et al.* 2011). It is possible that almost any of these misfolded species are likely to expose groups of amino acids that are not conducive of cellular biochemical activity, hence generating a source of toxicity and causing cellular dysfunction. Of these misfolded species, the larger rudimentary amyloid core oligomers, with their large hydrophobic surfaces, are considered to have the most devastating interactions with cellular components (Cremades *et al.* 2012; Campoioni *et al.* 2010). However once proteo- and amyloid fibres are produced, the percentage of this hydrophobic surface which is exposed is greatly reduced, thus the toxic effect is ameliorated or reduced (Auer *et al.* 2008; Cheon *et al.* 2007).

1.1.7. Protein misfolding disease

Protein misfolding disease can be loosely separated into two groups: the amyloid diseases and the prion diseases. Pathologically, a well understood and frequent characteristic of amyloid disease and prion disease is the disruption of protein homeostasis leading to the accumulation and aggregation of misfolded proteins (Knowles *et al.* 2014). Amyloid disease and prions differ in that prions are considered transmissible and amyloid diseases have not. However, recent evidence of iatrogenic transmission in humans and experimental transmission to mouse models of amyloid disease have led to a reconsideration of the definition between the two (Jaunmuktane *et al.* 2015). In reality, prion proteins can be capable of performing normal physiological functions, although they are also associated with a plethora of disease states. Recent discoveries of prion formation by the Parkinson's-causing protein α-synuclein, commonly considered a hallmark of amyloid disease, has again provided evidence for a revision of the definition between amyloid and prion (Pruisner *et al.* 2015).

Throughout the 20th Century, economic growth and improved knowledge has allowed medical science to extend the average human lifespan significantly. The average global lifespan is estimated to have been 33 years in 1900 rising to 67 years in 2000. Remarkably this trend is clear, regardless of region of the world; all seem to be subject to an approximate

doubling of average life expectancy within this century (Roser, 2017). Whilst this triumph of medical science is, and should, be celebrated, it is not without its challenges. The relatively small-time span of human evolution has not shaped our cellular biology to be durable enough to cope with the consequences of greatly extended longevity and so now the community faces new challenges (Knowles *et al.* 2014; Dobson, 2003). Today approximately 50 human disorders relating to protein misfolding and aggregation are known to exist. These are predicted to stop the great increases in life expectancy mentioned and promises to place great economic and emotional burden on the people, communities, and governments of the world. As such, the need for greater understanding of the underlying disease biology becomes increasingly urgent (Knowles *et al.* 2014; Valastyan and Lindquist, 2014).

Poised to be the 'new great plague' it is estimated that there will be 84 million new cases of Alzheimer's disease in the next 30 years (World Alzheimer report, 2016; World Alzheimer report, 2010). Some amyloid diseases are already abundant within populations with type II diabetes affecting 300 million people worldwide (Olshansky *et al.* 2005). Whilst intensive research has been directed towards deciphering the mechanisms of diseases such as Alzheimer's and Parkinson's, current understanding remains incomplete (Knowles *et al.* 2014; Valastyan and Lindquist, 2014). For instance, despite genetic studies identifying key mutations in genes that clearly cause Alzheimer's and Parkinson's, exactly how (or why) these mutations lead to the death of neurons remains a mystery (Knowles *et al.* 2014; Golde and Petrucelli, 2009). Exploration of the relationships between the quantity of amyloid present and disease progression show that in systemic amyloidosis, such as type II diabetes, a clear relationship between these two factors is observed (Eisenberg and Jucker, 2012; Haass and Selkoe, 2007). In contrast, the neurodegenerative diseases (ND) such as Alzheimer's and Parkinson's show no correlation between detectable protein level and disease progression or cellular damage (Eisenberg and Jucker, 2012; Haass and Selkoe, 2007). This, combined with a lack of reliable biomarkers for ND, means that accurate diagnosis of these disorders remains largely qualitative in nature, often requiring high levels of disease progression to be present for diagnosis. What has become clear is that a plethora of factors including intracellular mechanisms, local tissue environment, systemic environment and mechanisms related to neurodevelopment and ageing, all play a role in the development of these complex diseases (Ramanan and Saykin, 2013).

Table 1.2. The most common amyloid and prion diseases, including the disease name, mode of transmission, causal protein, and the host species.

A)

The Amyloid Diseases			
Disease	Mode of transmission	Protein	Host Species
Alzheimer's	Sporadic (95%) or inherited (5%)	Amyloid β and tau	Human
Parkinson's	Mostly Sporadic or inherited (10%)	α-Synuclein	Human
Huntington's	Inherited (autosomal dominant)	Huntingtin (exon 1, htt)	Human
Amyotrophic lateral sclerosis (ALS)	Sporadic (90%) or inherited (10%)	Superoxide dismutase (SOD1)	Human

B)

The Prion Diseases			
Disease	Mode of Transmission	Protein	Host Species
Scrapie	Sporadic (90%), inherited (8%) or infectious (2%)	PrP^C	Sheep, Goats
Bovine Spongiform encephalopathy (BSE)			Cattle
Chronic Wasting Disease			Pigs, Deer, Elk, Mule, Moose
Feline Spongiform encephalopathy			Feline
Creutzfeldt-Jakob Disease (CJD)			Human
Gerstmann-Straussler-Scheinker syndrome (GSS)			Human
Fatal familial insomnia (FFI)			Human
Kuru			Human

a) Examples of common mammalian amyloid diseases including their mode of transmission, the name of the protein which mis-folds thereby resulting in the disease phenotype and the species effected. b) Examples of common mammalian prion diseases including their mode of transmission, the name of the protein which mis-folds thereby resulting in the disease phenotype and the species effected.

1.1.8. Prion Disease

Many landmark discoveries have contributed to our current understanding of prion disease and the prion hypothesis; that misfolded proteins are responsible for the maintenance and transmission of the proteinaceous infectious particles known as prions (Griffith, 1967; Prusiner, 1982).

Clear descriptions from across Europe of the prion disease scrapie appear in the literature as early as 1750 (Leopoldt, 1750; Comber, 1772). However, despite being a prevalent problem for many years, it was not until 1936 that the first evidence of natural disease transmission was provided via intraocular inoculation of infected spinal cord tissue (Cullie and Chelle, 1936). This established the infectious nature of the disease as one that can be passed from one individual to another. Confirmation of the infectious nature of scrapie occurred in 1936 by William Gordon, when scrapie was accidentally transmitted during inoculation against a common virus: unknowingly, the inoculation contained a formalin-extract derived from an infected animal (Gordon, 1966; Bradley *et al.* 2002). Further reports confirming the transmissible nature of scrapie and its long incubation period (estimated to be 2 years) were later published for sheep, goat, and mice (Cullie and Chelle, 1939; Chandler, 1961).

Kuru was the first of the human prion diseases to be described, reported in 1950 as a ‘new disease’ by anthropologists visiting Papua New Guinea (Linsley, 1951). It took some time to convince the medical profession that the disease was not a genetically determined disorder; a reasonable, albeit inaccurate, assumption, occurring only in some cannibalistic tribes in Papua New Guinea (Bennett *et al.* 1959; Lindenbaum, 2008). Klatzo’s (1959) observations of the histopathology of Kuru infected brains showed a remarkable similarity to the lesion-ridden brains described by Jakob and Creutzfeldt. In 1966 it was shown that Kuru infected human tissue and was capable of transmitting the disease to monkeys. This was followed by similar experimentation involving Creutzfeldt-Jakob disease and Gerstmann-Straussler-syndrome (Gajdusek, 1977; Gibbs *et al.* 1968; Masters *et al.* 1981).

Until the 1960’s it had been assumed that the pathogen responsible for all of the diseases mentioned was microbial based. However, through a ground-breaking series of experiments, Alper and her team demonstrated that the infectious agent of scrapie was ‘of an unusual nature’ (Alper *et al.* 1967; Alper *et al.* 1966; Alper, 1993).

Using procedures which were known to destroy nucleic acids, such as high levels of ionizing radiation and ultraviolet light, Alper showed that the infectious material remained intact (Alper *et al.* 1967). The team also estimated the minimum molecular weight required to maintain infectivity to be $2x10^5$ Da, too small to possibly be a virus or any other known pathogen (Alper *et al.* 1966).

In 1967 Pattison and Jones' extraction experiments led them to the conclusion that the transmissible agent of scrapie may be a small basic protein. That same year Griffith (1967) based on these findings proposed three methods by which a protein could act as an infectious agent: -

1. The protein could switch on a damaging reaction in the host that is normally off
2. An aberrant form of the protein that has spontaneously produced, could serve as a template to induce production of the more aberrant forms
3. A protein may take on a diseased form when it passes from animal to animal

During the 1980's Stanley Prusiner tested these theories, coining the phrase prion or 'proteinaceous infectious particle' (Prusiner, 1982). By isolating the PrP 27-30 protease K-resistant C-terminal core of the full-length pathological prion protein (PrP^{Sc}) from infectious material and positively identifying it as a protein, Prusiner's team had finally provided experimental validation for the protein-only theory (Bolton *et al.* 1982). They went on to show that PrP^{Sc} was protease resistant, that infectivity was proportional to protein concentration, and invaluably, that infectivity decreased in the presence of both chemicals which destroy protein structure and anti-PrP antibodies (Gabizon *et al.* 1988). Subsequently in 1997 Prusiner was awarded the Nobel Prize in Medicine for "his discovery of Prions - a new biological principle of infection" (NobelPrize, 2018).

The identification of the single host gene encoding PrP (*PRNP*) and the corresponding mRNA in mammals, found no significant differences between genotype or expression in healthy and infected animals (Oesch *et al.* 1985; Chesebro *et al.* 1985). This led to the conclusion that the prion protein could exist in two distinct conformations: a normal native non-pathogenic conformation termed PrP^{C} and a misfolded pathogenic isoform, PrP^{Sc} (Balser *et al.* 1986; Prusiner, 1998). Continuation of genetic analysis discovered the first *PRNP* mutation linked to familial prion disease; there are now several mutations known to

either cause, or increase the likelihood of, developing prion disease. As previously mentioned, these range from SNPs (single nucleotide polymorphisms) changing single amino acids to large genomic repeats (Finckh *et al.* 2006; Mastrianni, 2003; Goldfarb *et al.* 1992; Hsiao *et al.* 1989). Once *PRNP* familial mutations were discovered, the opportunity to study PrP^{Sc} *in vitro* arose. Transgenic mice were created and overexpression of their mutated *PRNP* genes resulted in a transmissible neurodegenerative disease similar to prion disease (Hsiao *et al.* 1990; Jackson *et al.* 2009; Sigurdson *et al.* 2009; Telling *et al.* 1996a; Telling *et al.* 1996b; Nazor *et al.* 2005). Further experimentation demonstrated that the absence of the *PRNP* gene in knock out mice left hosts unable to contract infection or develop any symptoms (Bueler *et al.* 1993).

One rather particular feature of prions is the existence of distinct strains. Although distinct strains of pathogens can be explained by differences in their genome, it has long been unclear how a protein could be responsible for multiple distinct phenotypes (Collinge and Clarke, 2007). It has been observed experimentally, through proteolytic fragment size and glycoform ratios following proteinase K digestion, that different fragment sizes of PrP^{Sc} and thereby alternate conformations exist. Indeed, these changes in fragment size of PrP^{Sc} are associated with different clinical phenotypes of prion disease, including incubation time, brain pathology and protease sensitivity (Safar *et al.* 1998; Hsaio *et al.* 1990; Gibbs *et al.* 1968). The variance in proteinase K digestion kinetics, denaturation curves and the faithful replication of strain-associated biochemical characteristics when amplified provide compelling evidence that prion strains are associated with a conformation change in PrP (Sigurdson *et al.* 2009; Castilla *et al.* 2005; Saborio *et al.* 2001).

Today it is understood that PrP^{Sc} is responsible for the prion diseases in mammals; the transmissible spongiform encephalopathies (TSE's) are currently untreatable and therefore fatal neurodegenerative disorders. With similar symptoms to the Amyloid diseases, TSE's are characterised by spongiform degradation, neuronal loss, brain vacuolation, astrogliosis and the formation of large amyloid plaques in the brain (Collinge, 2005; Unterberger *et al.* 2005). It is now well understood that these symptoms are caused by a conformational change in a normal cell-surface glycoprotein (PrP^{C}) creating an altered protease-resistant isoform (PrP^{Sc}). This process can be facilitated by enzymes or chaperones or can occur spontaneously. The presence of this degradation resistant PrP^{Sc} modifies the synthesis of its normal cellular counterpart causing it to become highly infectious and accumulate in the brain in large numbers (Balser *et al.* 1986; Prusiner, 1998).

This templating action, seen in all amyloid or prion disease, is due to monomers adding to the end of fibrils. As they do so, they adopt a cross-β conformation to match the peptides in the aggregates, thus provide a template for newly formed monomers. As is the case with many prion proteins, the normal cellular function of PrP^C is unclear, although it is posited that it plays a cytoprotective role against internal and external stresses (Linden, 2017). Several other theories exist including roles in neuronal development, synaptic plasticity, iron uptake and neuronal myelin sheath maintenance (Westergard *et al.* 2007; Mani *et al.* 2003; Taylor *et al.* 2009; Steele *et al.* 2006). Whilst much progress has been made over the last century, our understanding of the specific mechanisms which underpin amyloid and prion disease is still incomplete. Interestingly, for the pathology of Alzheimer's disease, recent findings in mouse models suggest a relationship between PrP^c signalling disruption and continuing cognitive decline, with late depletion of PrP^c capable of restoring cognition in mice (Lima-Filho and Oliveira, 2018; Salazar and Strittmatter, 2017).

Although prions were traditionally associated with disease, understanding of their complex nature has expanded to encompass a wide range of functional prion-like proteins. Stephan *et al.* (2015) state that whilst the switch to the prion state is spontaneous and *de novo* in nature for pathological prions, functional prion switching appears to be tightly regulated by cells involving specific stimuli. Expression of mammalian cytoplasmic polyadenylation element-binding protein 3 (CPEB3) in yeast resulted in the formation of heritable aggregates (Stephen *et al.* 2015). This, alongside studies in mouse models, confirms the role of CPEB3 in long-term memory maintenance as an actin/CPEB3 feedback mechanism for synaptic plasticity induced by serotonin (Stephen *et al.* 2015; Si *et al.* 2010). Similar prion-based determinants or strong prion-like characteristics have been found in antiviral immune defence signalling and the immune response as well as within luminidependens, understood to be a key mechanism of cold tolerance and seasonal variation within flowering plants (Chakrabortee *et al.* 2016; Cai *et al.* 2014). Despite this expansion in understanding of the 'prion concept', prions are still considered an excellent model for the study of the underlying pathology of amyloid disease.

For some years, this study has been hindered by the lack of an appropriate model organism from which to study prion and amyloid formation within a reasonable timescale. However, in 1994, Reed Wickner correctly hypothesised that the unusual heritable traits observed in the yeast *Saccharomyces cerevisiae* were, in fact, prion forms of normal cellular proteins (Wickner, 1994). These unusual heritable trains were first reported by Brian Cox who

searched for the underlying genetic factors responsible for the [*PSI*$^{+}$] trait but had been unable to identify a mutation responsible (Cox *et al.* 1988). On discovery of yet another unusual pattern of inheritance, [*URE3*], Wickner drew the conclusion that prion forms of Sup35 and Ure2 respectively were responsible for the associated traits, establishing the presence of prions in the Fungal kingdom and a new model organism for prion experimentation (Wickner, 1994). Whilst these proteins are no relation and share no sequence similarity to their lethal mammalian counterpart, PrP, they exhibit the hallmark templating mechanism of prions, are composed of protein fibrils, propagate by seeding and exist in both soluble and insoluble amyloid isoforms (Caughey and Lansbury, 2003). A suitable model system had been found.

1.1.9. Prions in Yeast

The budding yeast *Saccharomyces cerevisiae* is known to contain over 20 prion forming IDRs or PrDs (prion domains), ten of which that have been well-characterised (Wickner *et al.* 2015; Liebman and Chernoff, 2012). Among eukaryotic model organisms, *S. cerevisiae* offers many advantages for the study of the mechanism and cellular influence of prion formation. This is demonstrated by *S. cerevisiae* 's ability to recapitulate many cellular and molecular features of amyloid-based disease, including the existence of two stable protein states: a relatively unstructured but soluble form and a self-perpetuating amyloid. Unlike more complex eukaryotes, *S. cerevisiae* can be grown on defined media within timescales similar to bacterial growth (Feldmann 2012). This, combined with the availability of powerful genetic and molecular techniques, makes it a popular and valuable model organism for the study of prions (Botstein and Fink 2011; Sherman, 2002; Botstein, 1997).

The growing number of known yeast prions exhibit a surprising amount of diversity, including proteins from many different functional classes involved in a host of cellular functions including gene transcription, mRNA translation and metabolic regulation. Offering such an array of potentially short lived or Lamarckian diversity, it has long been argued that the epigenetic influence of prions gives them the potential to influence survival outcomes in changing environments (Chakravarty *et al.* 2020; Jarosz *et al.* 2014a; Halfmann *et al.* 2010; Sondheimer and Lindquist, 2000). It is important to note, however, that yeast prions can also be damaging to cells; this can be caused by their presence having a negative impact on cell survivability or by introducing cellular toxicity via amyloid formation (Wickner et al. 2020; Tuite, 2015; Kelly et al. 2012a; Nakayashiki *et al.* 2005).

Table 1.3. Table of the most well characterised yeast prions with a summary of information regarding their native function, prion phenotype, and other prion relevant summary information.

Prion	[*PSI*$^+$]	[*RNQ*$^+$]/[*PIN*$^+$]	[*URE3*]	[*SWI*$^+$]	[*OCT*$^+$]	[*MOT3*]	[*ISP*$^+$]
Protein determinant	Sup35	Rnq1	Ure2	Swi1	Cyc8	Mot3	Sfp1
Native function	Translation termination	Unknown	Nitrogen regulation	Transcriptional regulation	Transcriptional regulation	Transcriptional regulation	Transcriptional regulation
Prion Phenotype	Increased nonsense suppression	Heterologous prion appearance	Use of poor nitrogen source	Loss of function	Loss of function	Loss of function	Opposite of loss of function
Infectivity of fibres	Yes	Yes	Yes	Yes	ND	Yes	ND
Amyloid	Yes	Yes	Yes	Yes	ND	Yes	ND
QN-rich domain	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variants Isolated	Yes	Yes	Yes	ND	ND	ND	ND
Overproduction induces	Yes	ND	Yes	ND	Yes	Yes	Yes
Cured by hsp104Δ	Yes	Yes	Yes	Yes	Yes	Yes	No
Cured by GdnHCl	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Found in Wild	No	Yes	No	ND	ND	ND	ND

Not determined is signified by the letters ND (adapted from Liebman and Chernoff (2012) and Chernova *et al.* (2014)).

1.1.10. Sup35p and the [*PSI*$^+$] Prion

The *S. cerevisiae* prion, [*PSI*$^+$] is the most extensively studied of yeast prions, understood to be the prion form of the translation termination factor Sup35p. Encoded for by the 2058 base pair sequence of the *SUP35* gene on chromosome IV, Sup35p consists of 685 amino acids. The *N*-terminal PrD of Sup35p appears between amino acid 1-114, the highly charged middle domain is present between amino acid 124-253 and the functional *C*-domain responsible for its role in translation termination appears at amino acid 254-685 (Derkatch *et al.* 2004; Lindquist *et al.* 2001). Shown in Figure 1.5.

Essential for prion formation and propagation, the *N*-terminal PrD consists of two distinct regions: an atypical asparagine and glutamine (Q/N) rich region (between residues 1-40) and a series of five near identical oligopeptide repeats (amino acids 41-97) known as the oligopeptide repeat region (OPR). It is thought that the Q/N rich region provides critical stability for amyloid aggregation, aiding in the formation of the amyloid core via the nucleation and polymerization of Sup35p (Derkatch *et al.* 2004; Chien and Weissman, 2001; DePace *et al.* 1998). The OPR exhibits the same repeat pattern that is seen in mammalian PrP and indeed some sequence similarity exists between the two. Important for the propagation of the [*PSI*$^+$] prion state, Sup35p OPR is implicated in mediating chaperone access to regions of the protein and/or in stabilizing Sup35p-Sup35p interactions. The *N*-terminal region is not required for prion maintenance or the normal function of the Sup35 protein, oddly the '*M*' domain of Sup35p seems to play no role in either the normal or prion function and is therefore entirely dispensable for both, where, by contrast, the *C*-terminal domain is essential for cell viability (Shukundina *et al.* 2006). As is the case with all amyloids, the conversion from normal Sup35p to the abnormal prion conformation is associated with the formation of thermostable, protease resistant, *β*-sheet rich amyloid fibres and whilst the *M* and *C*-domains of the protein may retain their normal folds, the cross-*β* interactions previously mentioned connect the prion domains and allow templating to occur (Toombs *et al.* 2010; Chernoff, 2007; Ross *et al.* 2005; Derkatch *et al.* 2004)

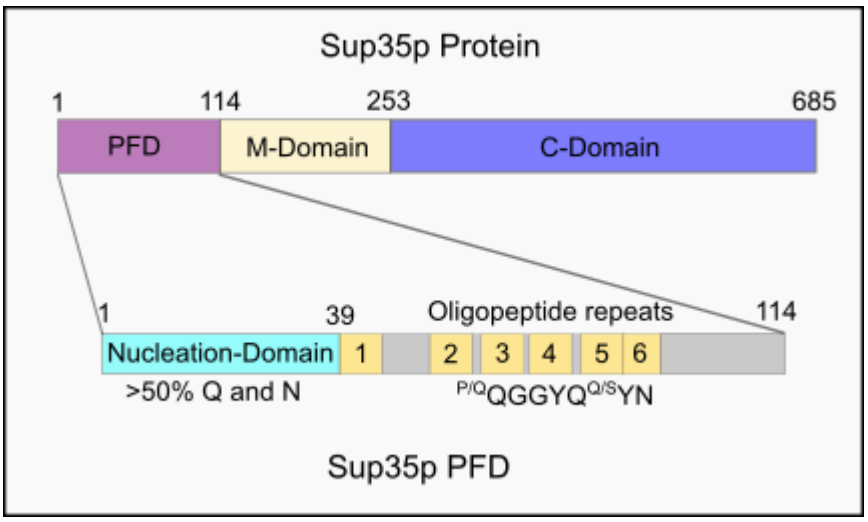


Figure 1.5. Schematic diagram of Sup35 protein domains and prion domains, highlighting the Q/N rich nucleation domain of Sup35p and the oligopeptide repeats of the Prion forming domain. At the top of the Figure is the entire Sup35p protein, showing the prion forming domain (PFD) in purple between amino acid 1-114, the middle (M) domain in cream between amino acid 114-253 and the Core (C) domain in purple between amino acid 253 and 685. At the bottom of the Figure the PFD is magnified to reveal, the Glutamine (Q) and Asparagine (N) content of the nucleation domain between amino acids 1 and 39 and between amino acids 39 and 114 oligopeptide repeats are shown in yellow with the single letter amino acid code of this repeat detailed (adapted from Toombs *et al.* (2011) and Osherovich *et al.* (2004))

Known to alter cellular fitness, the presence of the [*PSI*$^{+}$] prion interferes with translation fidelity by affecting reading frame selection and decreasing cellular translation termination efficiency. Fortuitously, this role in translation termination results in an easily identifiable phenotypic assay in all yeast strains carrying the *ade1-14* or *ade2-1* mutation (Byrne *et al.* 2009; Cox *et al.* 2003; Ugolini and Bruschi, 1996). Mutation in the adenine biosynthetic pathway leads to the accumulation of a red pigment in yeast colonies which contain the normal conformation of the Sup35p protein, known as [*psi*$^{-}$] (Ugolini and Bruschi, 1996). However, in [*PSI*$^{+}$] strains, the prion form of Sup35, now known as Sup35p, causes a read through of the stop codon within this biosynthetic pathway, resulting in no accumulation of red pigment and therefore white colonies (Byrne *et al.* 2009; Chernoff, 2007; Cox *et al.* 2003). As well as being useful for identification, this allows colony selection based on the bias of adenine selection, given that [*psi*$^{-}$] cells cannot produce adenine due to the nonsense mutation and therefore cannot grow on adenine-free defined media (Figure 1.6). It is worth noting that experimentally the use of synthetically defined drop-out media within the study of *S. cerevisiae* is a very powerful tool, troubled by the increasing resistance and general tolerance of fungal organisms to antibiotics, drop out media presents a more reliable and easy to use alternative and is therefore widely used for selection.

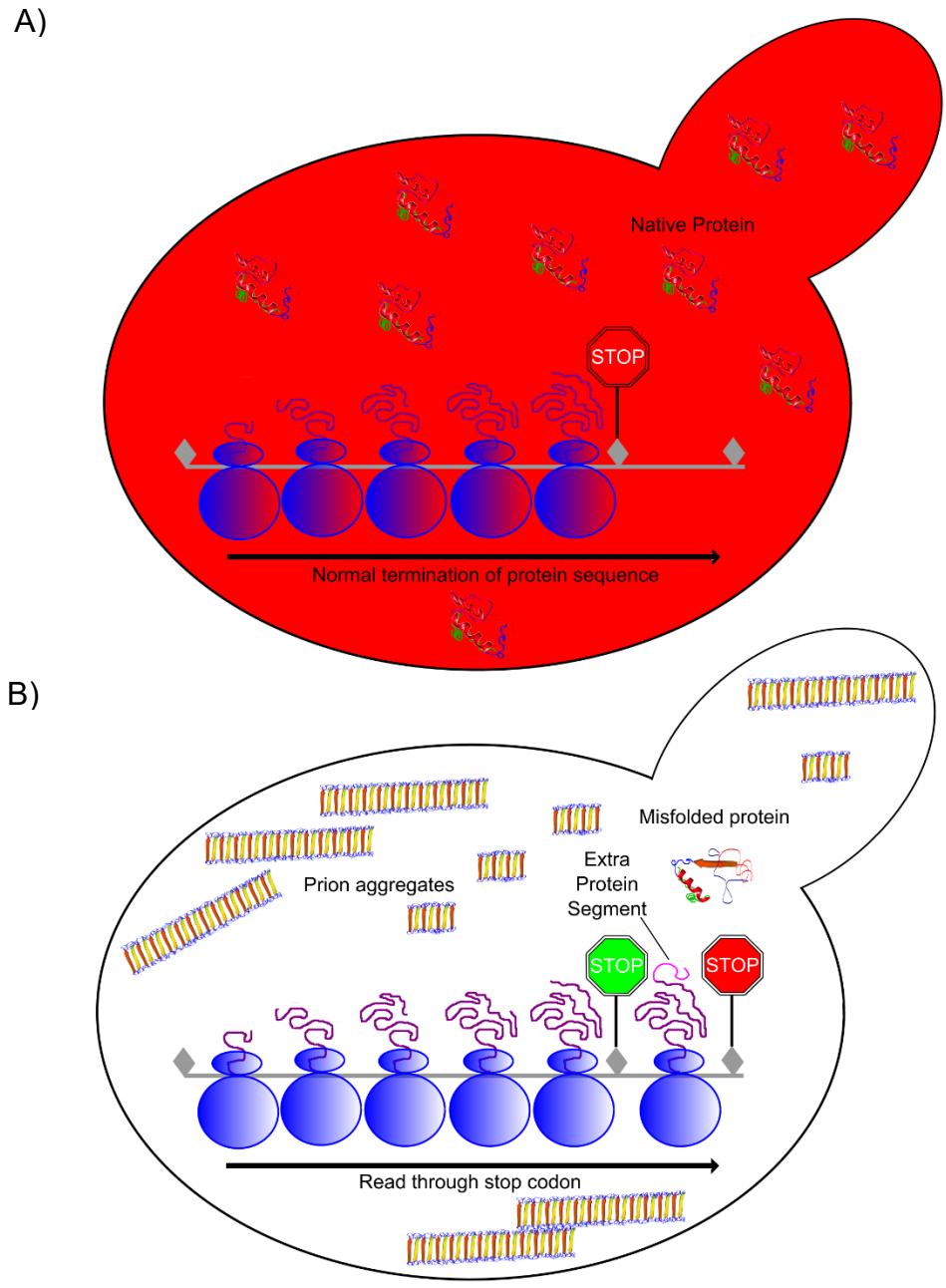


Figure 1.6. Assaying the presence of the [*PSI*$^{+}$] prion by suppression of the *ade1-14* nonsense mutation. Blue motifs are representative of ribosomes, effective stop codons are shown in red and ineffective stop codons are shown in green. The proteins within the cell are reflective of the prion status of the cell, either containing only native conformation [*prion*$^{-}$] or prion aggregates [*prion*$^{+}$]. The colour of the yeast cells is indicative of the colour change seen as a result of the a nonsense adenine mutation presence in cells hence allowing for the build-up of a red colony pigment in (A) [*psi*$^{-}$] due to the reading of the stop codon allowing for adenine biosynthesis or a white colony pigment (B) [*PSI*$^{+}$] due to the readthrough and loss of function of the subsequent biosynthetic pathway.

1.1.11. The molecular chaperone Hsp104p

Molecular chaperones interact, stabilise and aid proteins to maintain cellular protein homeostasis. To fulfil this role, they must recognise and bind to unfolded or misfolded proteins and mechanically facilitate their folding into the native conformation (Hartl *et al.* 2011). Hsp104p is one such molecular chaperone, belonging to the family of Hsp's (heat shock proteins). Hsp104p is a homohexameric AAA ATPase and (*in vivo*) it is responsive to stresses including heat, ethanol, and sodium arsenite and is required for the propagation of yeast prions (with the exception of [*ISP*$^{+}$]) (Liebman and Chernoff, 2012). Inhibition of the ATPase activity of Hsp104p by guanidine hydrochloride (GdnHCl) leaves cells unable to efficiently propagate [*PSI*$^{+}$] or any other prion (Grimminger *et al*. 2004; Jung *et al*. 2002; Ferreira *et al*. 2001). Small concentrations, between 3-5mM, of GdnHCl are enough to achieve this; a process commonly referred to as 'curing' (Byrne *et al.* 2007; Grimminger *et al.* 2004; Tuite *et al.* 1981). This Hsp104p mediated 'curing' of GdnHCl is known to block the ability of Hsp104p to cleave the [*PSI*$^{+}$] amyloid fibres into smaller, heritable oligomers, known as 'propagons' (prion seeds) (Park *et al.* 2012; Byrne *et al.* 2009; Byrne *et al.* 2007). In doing so the number of 'propagons' required to successfully transmit the prion from mother to daughter are depleted and efficient prion seeding is not achieved for all daughter cells (Byrne *et al.* 2009; Byrne *et al.* 2007).

Deletion of Hsp104p eliminates the presence of the [*PSI*$^{+}$] prion and, surprisingly, the same is true if Hsp104p is overexpressed (Chernoff *et al*. 1995). Paushkin *et al.* (1996) reasoned that this curing via over expression of Hsp104p may be as a result of increased cleavage or fragmentation, breaking down amyloids at such a rate that *'propagons'* are refolded by the UPS into the correct conformation without the chance for templating to occur. However, recently Ness *et al.* (2017) have provided evidence that no such breakdown of Sup35p aggregates occur when Hsp104p is overexpressed. Rather it is the distribution of 'propagons' between mother and daughter cells that is unequally partitioned via a mechanism not currently known. It was reported that mother cells retain all 'propagons' in approximately 10% of divisions per generation, creating daughter cells completely devoid of prion and thus [*psi*$^{-}$] (Ness *et al.* 2017).

1.1.12. Rnq1p and the [*RNQ*$^{+}$] Prion

The [*RNQ*$^{+}$] or [*PIN*$^{+}$] prion is known to facilitate the *de novo* appearance of prions or other PrD-containing proteins, including [*PSI*$^{+}$] (Derkatch *et al*. 2001) and exon 1 from the human

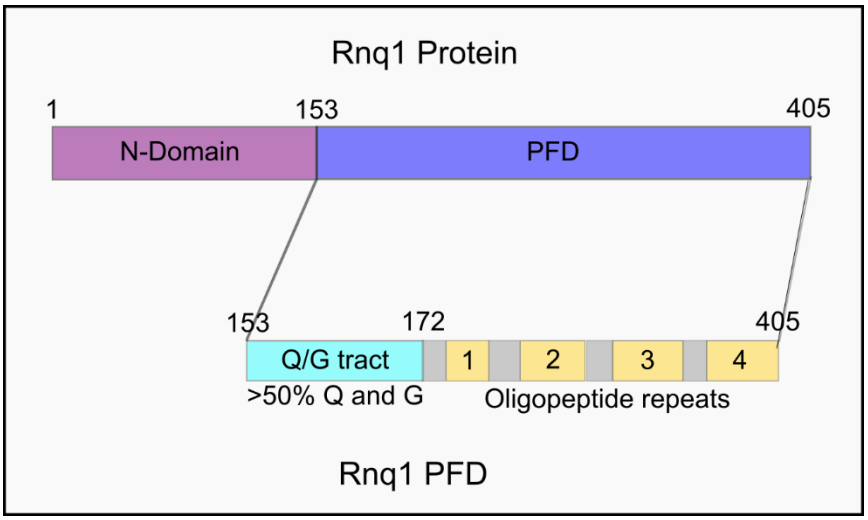
huntingtin protein (Derkatch *et al.* 2004), allowing them to convert from their soluble state to their amyloid states. Clear experimental evidence supports this concept as Rnq1 proteins in their prion conformation (known as [*RNQ*$^{+}$]) have been found within Sup35p polymers, asserting the role that [*RNQ*$^{+}$] has in the seeding of Sup35p (Liebman and Chernoff, 2012). Initially named [*PIN*$^{+}$] as a result of this function (meaning [*PSI*$^{+}$] inducing), it was considered that the [*RNQ*$^{+}$] trait was an intermediate conformation of the non-amyloid form of Sup35 (Serio *et al.* 2000). However, subsequent experimentation has revealed that this role in heterologous prion appearance facilitates the *de novo* formation of many prions and so [*RNQ*$^{+}$] (meaning rich in asparagine (N) and glutamine (Q)) is the more common nomenclature (Sondheimer and Lindquist, 2000).

Moreover, gene deletion of the Rnq1p determinant is sufficient to convert a strain from [*RNQ*$^{+}$] to [*rnq*$^{-}$], as is the atypical prion treatment of GdnHCl curing (Derkatch *et al.* 2001; Osherovich and Weissman 2001; Sondheimer and Lindquist, 2000). As expected of prion behaviour, the presence or absence of the prion form of Rnq1p correlates with the appearance and disappearance of aggregate formation and intermediate non-aggregated Rnq1p induces the appearance of [*RNQ*$^{+}$] when transformed into [*rnq*$^{-}$] background (Patel and Liebman, 2007; Derkatch *et al.* 2001). Although required for the *de novo* appearance of prions, once a prion form of a protein has been established cellularly the [*RNQ*$^{+}$] prion is no longer needed and hence plays no role in prion propagation (Liebman and Chernoff, 2012; Derkatch *et al.* 2000; Derkatch *et al.* 1997). This role as a facilitator or seeding mechanism indicates that there is a hierarchy present within yeast prions and whilst other amyloid forms of other Q/N-rich regions can perform a similar seeding function, this is uncommon and always in a reduced capacity, leaving [*RNQ*$^{+}$] firmly placed at the top of this hierarchy (Liebman and Chernoff, 2012; Derkatch *et al.* 2000; Derkatch *et al.* 1997).

The presence of distinct variants has also been observed within [*RNQ*$^{+}$] cells, via the manipulation of the [*PSI*$^{+}$] reporter system detailed previously. Fusing the prion forming domain of the Rnq1 protein to the translation termination domain of the Sup35 protein allows the indicative phenotypic red to white assay to inform as to the ‘strength’ of the [*RNQ*$^{+}$] strain (Bardill and True, 2010). Interestingly, this work has revealed a relationship between temperature and the distribution of variant strength, with fibres formed at lower temperatures exhibiting ‘strong’ phenotype with morphologically distinct, short, and curly amyloid fibres and higher temperatures favouring a ‘weak’ phenotype with large bundled fibres. Stein and True (2011) showed that the ‘weak’ variants lost their prions at a much higher rate (~40%)

when compared with the strong strains (<1%). These observations support the morphological data, as seeding is much more likely to be unsuccessful with the large insoluble aggregates seen in 'weak' strains and vice versa. Curiously, despite these differences, both the 'weak' and 'strong' strains' ability to induce [*PSI*$^{+}$] was reported as being unhindered with only medium variants showing low levels of [*PSI*$^{+}$] induction. An explanation for this observation was provided with there being high levels of self-templating but a less efficient interaction with Sup35p within 'strong' strains (and vice versa with 'weak' variants), hence leaving medium variants lacking in both functions (Kalastavadi and True, 2010).

a)



b)

MDTDKLISEAESHFSQGNHAEAVAKLTSAAQSNPNDEQMSTIESLIQKIAGYVMDNR
SGGSDASQDRAAGGGSSFMNTLMADSKGSSQTQLGKLALLATVMTHSSNKGSSNRGF
DVGTVMSMLSGSGGGSQSMGASGLAALASQFFKSGNNSQGQGQGQGQGQGQGQGQ
GSFTALASLASSFMNSNNNNQQGQNQSSGGSSFGALASMASSFMHSNNNQNSNNSQQ
GYNQSYQNGNQNSQGYNNQQYQGGNGGYQQQQGQSGGAFSSLASMAQSYLGGGQTQS
NQQQYNQQGQNNQQQYQQQGQNYQHQQQGQQQQQGHSSSFSALASMASSYLGNNSNS
NSSYGGQQQANEYGRPQQNGQQQSNEYGRPQYGGNQNSNGQHESFNFSGNFSQQNNN
GNQNRY

Figure 1.7. Rnq1p detailed protein domain and amino acid information a) Schematic diagram of Rnq1p domains and prion domains, highlighting the Q/G rich tract and the oligopeptide repeats of the *C*-terminal or PFD of Rnq1. At the top of the Figure is entire Rnq1p, showing the nitrogen (N) domain in purple between amino acid 1-153, the protein forming domain (PFD) in purple between amino acid 153-405. At the bottom of the Figure the PFD is magnified to reveal, the glutamine (Q) and glycine (G) content of the nucleation domain between amino acids 153 and 172 and between amino acids 172 and 405 oligopeptide repeats are shown in yellow. b) Amino acid code of Rnq1p. All hydrophobic amino acids are shown in red, polar amino acids in green and charged amino acid in black.

Coded for by the 1218bp gene *RNQ1* on chromosome III of the *S. cerevisiae* genome, the Rnq1p protein is known to contain 405 amino acids. Compared to the well-studied Sup35p, little is known about the protein domains of Rnq1p. It is known that the *N*-terminal plays no part in prion formation, instead the PrD appears at the *C*-terminal of the protein between residue 153-402 and contains four discrete Q/N rich residues (shown in Figure 1.7.) (Stein and True, 2011). Experimentation on these regions has shown that whilst one specific region of these four is required for prion formation, prions can be maintained if any two are present (Kadnar *et al.* 2010). The native form of Rnq1pis soluble and relatively unstructured in its conformation; atypically the PrD region is known to form β-sheet rich amyloids in its prion formation (Liebman and Chernoff, 2012).

Despite the efforts of several labs, the native role of the non-aggregated conformation of Rnq1p has remained elusive. Known to be a non-essential gene, its only known role is as an epigenetic modifier of prion formation (Sondheimer and Lindquist, 2000). The cellular presence of the [*RNQ*$^{+}$] prion does not result in a measurable growth defect and is not toxic *per se*, but overexpression of Rnq1p or a polyQ-containing protein in a [*RNQ*$^{+}$] background causes cell toxicity (Douglas *et al.* 2008). Interestingly, it has been shown that this toxicity of overexpression is not caused by generalised proteomic stress but instead is as a result of highly specific mitotic arrest (Treusch and Lindquist, 2012). Past experimentation has revealed that this toxic effect can be elevated via the overexpression of molecular chaperones involved in the propagation of amyloid formation and accentuated by their suppression (Douglas *et al.* 2008). This highlights the role that intermediates to fibrillation play in the toxic effects seen, supporting the oligomeric hypothesis, and bringing into question their interactions with normal cellular functioning.

Whilst other prions such as [*PSI*$^{+}$] and [*URE3*] are not found in natural, industrial and clinical isolates, [*RNQ*$^{+}$] has been found in approximately 16% of 70 wild populations tested by Kelly *et al.* (2012a), similar to the previous findings of Nakayashiki *et al.* (2005). However, a larger study by Halfmann *et al.* (2012) of 690 natural *S. cerevisiae* isolates detected [*RNQ*$^{+}$] in approximately 6% of samples. This study also detected weak [*PSI*$^{+}$] in approximately 1%; however, all isolates in which [*PSI*$^{+}$] was detected were from wineries and therefore arguably their status as 'wild' is debatable (Kelly et al. 2012; Halfmann *et al.* 2012).

Contrasting ideas about the nature of prions in the wild exist (Debets *et al.* 2012). Some argue that the ability to aggregate and/or interact with other Q/N rich proteins is key to

Rnq1p's biological role (Kadnar *et al.* 2010). With experimental data showing that [*RNQ*$^{+}$] is predominantly benign and between 25% and 40% of the time, beneficial (Halfmann *et al*. 2012; Halfmann *et al*. 2010). These groups argue that the prion determinant of Rnq1p is maintained within wild populations in small numbers as a sophisticated evolutionary "bet-hedging" tool (Halfmann *et al*. 2012; Halfmann *et al*. 2010). Capable of coping with unexpected stress, [*RNQ*$^{+}$] significantly rises in number (up to 60-fold) within a population during stressed conditions acting as a valuable mechanism of phenotypic variability. In laboratory conditions, once the stressor is removed, the favoured [*prion*$^{-}$] state is quickly restored (Treusch and Lindquist, 2012; Halfmann *et al*. 2012; Halfmann *et al*. 2010). In contrast, it could be argued that if any of the yeast prions were indeed beneficial then they would have been brought to fixation within wild populations (Wickner et al. 2020; Kelly et al. 2012a; Nakayashiki *et al.* 2005). Kelly *et al.* (2012b) show that maintenance of [*RNQ*$^{+}$] only appears to be maintained in wild populations with highly polymorphic *RNQ1* loci's, suggesting that these sequence differences are capable of blocking or ameliorating prion propagation or its pathogenetic effects.

Arguments regarding the role of yeast prions as a disease, based on the fixation of prions within wild populations, are largely built on the functional [*Het-s*] prion of *Podospora anserine*. The [*Het-s*] prion is reported in 92% of *het-s* isolates with evidence for this prion's role in heterokaryon incompatibility corresponding to a cell death reaction; an event which only occurs upon the fusion of genetically distinct variants (Debets *et al.* 2012). However, studies of other yeast prions appear to support the alternative epigenetic or Lamarckian inheritance model (Treusch and Lindquist, 2012; Halfmann *et al*. 2012; Halfmann *et al*. 2010). The [*GAR*$^{+}$] prion, for instance, causes the spontaneous switching of the cell's metabolic requirements from glucose to alternative carbon sources (Jarosz et al. 2014a). This prion manifests via a curious relationship between bacteria and yeast when they occupy the same environment. [*GAR*$^{+}$] is induced by a factor secreted by cohabiting bacteria in correlation with changes in the environmental availability of glucose (Jarosz *et al.* 2014b). A not dissimilar prion, [*SMAUG*$^{+}$], regulates the breeding preference and survival strategies of yeast cells, allowing for pre-adaptation to nutrient repletion (Chakravarty *et al.* 2020; Itakura *et al.* 2020). Thus, Jarosz *et al.* (2014a) and Chakravarty *et al.* (2020) argue that these prions are conserved as part of an adaptive strategy, allowing continued population growth and maintenance in diverse and changeable environments.

1.2 Metabolomics

1.2.1. Metabolites, the metabolome, and metabolomics

The term 'metabolites' refers to the small biological molecules present both intra- and extracellularly, with the 'Metabolome' being the nomenclature for the entire suite of metabolites (Patti *et al.* 2012; Nielsen and Oliver, 2005). Metabolites are usually of low molecular weight (<1500Da), have a varied biochemical make up and are often present in small quantities all of which makes their detection a complex endeavour. Considered indicative of the pathway, chemical intermediates and intended modification of the parental chemical compound, metabolites represent the biochemical molecular phenotype of cells. Metabolites, produced as direct result of the cellular processes the 'omics' field fixates on; DNA (deoxyribose nucleic acid) is transcribed (genomics and transcriptomics), this is in turn translated into protein sequence and proteins are formed (proteomics), enzymatic reactions within the cell alongside a host of biochemical compound import and export cause metabolic reactions to take place, and so metabolism occurs with a set of chemical conversions leading to metabolites being produced or consumed (metabolomics) (Figure 1.8.) (Roberts *et al.* 2013; Patti *et al.* 2012; Duportet *et al.* 2011). Chemical compounds considered as energy sources and building blocks of the cell are used to perform and maintain all of the cellular functions, achieved via the process of oxidation or conversion to more stable compounds allowing for a release of energy available for the cell to store or to use for other anabolism. Whilst engaged within this chemical process of metabolism (Greek '*metabole'* meaning change or transformation) these chemical compounds are referred to as metabolites (Pinu *et al.* 2017; Baidoo *et al.* 2012; Fiehn, 2002).

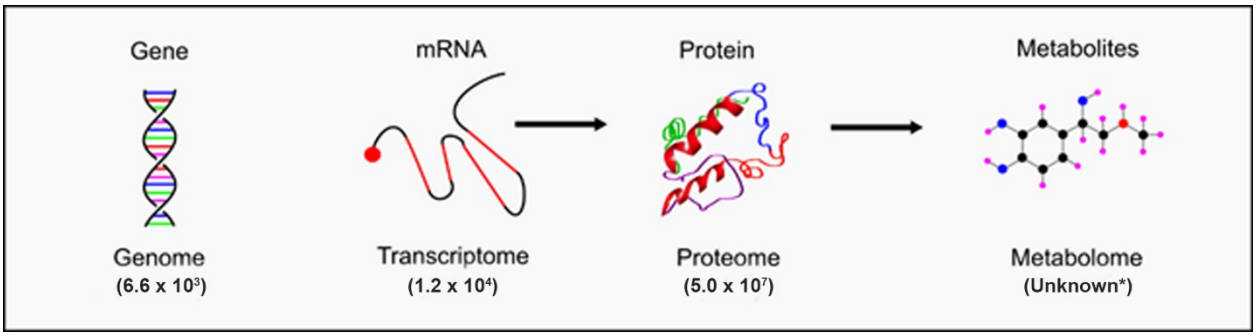


Figure 1.8. The 'omics' cascade, showing the relationship between the genome, proteome and metabolome within *S. cerevisiae*. The possible number of variables at each stage within *S. cerevisiae.* *The size of *S. cerevisiae's* metabolome is unknown, however there are currently 1.6 x 10^4 identified metabolites, involving approximately 900 genes, the size of the metabolome itself is estimated to be upwards of 1.0 x 10^5 (adapted from Patti *et al.* 2012 and Gerszten and Wang, 2008 using Ramirez-Gaona *et al.* 2017; BNID 100237, BNID 102988, BNID 106198, Milo *et al.* 2010).

The wide-scale cellular implications of metabolism and the constant requirement of cells for energy means that any given metabolite does not exist for long or in isolation. The formation of metabolites depends entirely on the nature of the enzyme that acts upon a parent compound and the composition of the parent compound's structure. Within primary metabolism, metabolites have a fast turnover in cells which reflects the fast-paced nature of biological life and illustrates why primary metabolism is often highly conserved across species. The need for cells to generate energy and synthesise biological components means that metabolites are not generally permitted by cells to accumulate (Pinu *et al.* 2017; Villas-Boas *et al.* 2007). Since metabolites are the intermediate and by-products of metabolic reactions, these processes are dynamic and continuingly fluctuating and so metabolites within primary metabolism are likely to be in low abundance (Pinu *et al.* 2017; Villas-Boas *et al.* 2007). More diverse across species, secondary metabolism however is typically a slower process and linked to survivability or fecundity. Second metabolism consists of a smaller number of reactions and hence turnover rate is slower and therefore the relative levels of compounds are often higher (Villas-Boas *et al.* 2007). This change in turnover rate is often determined as metabolic flux and understanding of this interconversion rate has in recent times been a strong focus within the field (Anotoniewicz, 2015; Villas-Boas *et al.* 2007).

The structural and thus chemical and physical diversity of metabolites is enormous and provides additional challenges not faced by other 'omics' disciplines (Villas-Boas *et al.* 2007). It is currently thought that upwards of 110,000 metabolites are present within human cells and the number is even higher in plants at upwards of 1,000,000 (HMDB, 2019; Chae *et al.* 2014). Metabolites range in molecular weight, molecular size, and polarity, encompassing many different chemical classes including the nucleotides, sugars, amino acids, alcohols, phenolics, steroids and lipids among others (Villas-Boas *et al.* 2007; Nielsen and Oliver, 2005; Dunn and Ellis, 2005).

Unlike genes or proteins that are subject to post translation modifications, metabolites with their complexity and specificity, are considered to be a functional end point within cells. Metabolomics, said to provide "a functional read out of the physiological state" of an organism (Mathew and Padmanaban 2013), is the study of the entire metabolome of an organism, aiming to measure, identify and quantify metabolites in cells, tissues and biological fluids at a given time point using sophisticated analytical technologies (Roessner and Bowne, 2018; Mathew and Padmanaban 2013; Weckwerth 2010; Fiehn 2002; Oliver *et*

al. 1998). A combined effort over the last two decades has seen rapid improvements in the 'omics' fields. This is certainly apparent within metabolomics, with the prevalence of metabolomics publications rising exponentially since the start of the century. A search of Medline for the term metabolomics, reveal 2 publications in 2000 rising to 4800 in 2018 (Web resource at URL:http://dan.corlan.net/medline-trend.html). This is, in part, due to a rise in awareness of the potential applications of the field but can largely be attributed to an increased availability of the analytical technologies, method development and the increasing commonality of computational power. Metabolomics is increasingly used to study a wide range of topics including disease mechanisms, biomarker discovery, cellular networking, environmental response, and evolutionary studies within a variety of systems ranging from microbial to mammalian (Zhang *et al.* 2013). Metabolomics is poised to be a key player in the development of fundamental knowledge, forming links in our understanding between genomics, transcriptomics, proteomics and metabolomics and, by doing so, better understanding how cellular biology interacts with its own architecture and its environment (Ramanan and Saykin, 2013; Patti *et al.* 2012; Gieger *et al.* 2008). However, both the transient nature of many metabolites and their complex diversity, present many challenges to the metabolomics community in their pursuit of this understanding.

1.2.2. Metabolomics workflow

When considering metabolomics experimentation, it is first necessary to determine the number of metabolites that require measurement. In doing so, researchers are able to define their experimental workflow in accordance with the recommendations for either 'targeted' or 'untargeted' metabolomics (Figure 1.9.).

So called 'targeted' metabolomics aims to accurately identify, measure, and quantify a predefined list of metabolites (a subset of the metabolome), normally associated with one or more specific pathways of interest (Bingol, 2018). Using authenticated standards researchers apply this method in the pursuit of specific biochemical questions and hypotheses regarding particular pathways (Patti *et al.* 2012). This method is commonly used during the development of metabolomic methodologies, enabling the measurement of a range of intracellular metabolites and allowing for the assessment of any given protocol by way of comparison to others. This has allowed for the development of robust and highly sensitive methods which enable reliable quantification of relatively low-level metabolites within cells (Roberts *et al.* 2013). In addition, this method is well used to study the effects of

pharmaceutical and therapeutic interventions via drug metabolism (Patti *et al.* 2012). Limited to hundreds of metabolites, targeted metabolomics data analysis is simpler and although a strength, it is also compromised in that its analysis is limited to only what is currently known (or contained within a given metabolite library) reducing the scope for discovery and introducing bias (Bingol, 2018; Paglia *et al.* 2018; Patti *et al.* 2012).

Untargeted metabolomics by contrast is technically more demanding with data pre-processing and analysis being both time and bioinformatically challenging. It does however introduce no bias to the study by way of metabolite selection, aiming to sample the entire complement of metabolites produced by a cell at a given time point (Paglia *et al.* 2018; Patti *et al.* 2012). Untargeted metabolomics allows for the measurement of thousands of features and the identification of novel and unexpected metabolites but is considered to be one of the most widely accepted bottlenecks for untargeted metabolomics (Bingol, 2018). Online algorithms and databases are capable of defining and annotating compound specific features allowing for neutral masses to be calculated and putative identifications given. There are however limitations to this approach. Whilst lists improve daily (and the addition and development of MS/MS spectral databases have been of great assistance with identification), metabolite databases are not comprehensive, are often method-dependent and have inherent biases towards synthesised compounds (due to high levels of fragmentation detail being available) (Bowen and Northern, 2010). Other methods of identification such as extensive structural study via fragmentation spectra or experimental chemical shift patterns are possible but these are often cost and time intensive, requiring large scale isolation of compounds which can often prove impractical at best and impossible at worse (Bingol, 2018; Bowen and Northern, 2010).

Understanding the biological implications of this type of global information can be challenging, as frequently data sets are very large and difficult to comprehend. To tackle this, untargeted metabolomic experiments often focus on metabolic profiling, aiming to sort and compare different biological samples and then ultimately identify the metabolic changes or differences between groups. This does not require a pre-existing knowledge of the changes that may occur (that would be pre-requisite of a targeted metabolic study) and does allow for the generation of hypotheses regarding the variation between groups or perturbations (Patti *et al.* 2012). Although identification of these differences may be achieved, this method does not easily allow for the quantification of metabolites and despite considerable work in this area involving quantified spikes and stable isotope labelling, these

methods are often expensive and much better suited to confirmatory or follow up targeted experimentation (Paglia *et al.* 2018; Dumas and Davidoc, 2013; Patti *et al.* 2012). Another possible approach is to use metabolic fingerprinting thereby removing the need to identify large numbers of metabolites, although this limits the scope of the study to a mere classification tool.

Table 1.4. The differences between untargeted and targeted metabolomics experiments.

Untargeted Metabolomics Experiments	Targeted Metabolomics Experiments
Global detection of a wide range of metabolites.	Quantification of small number of specific and related metabolites.
Hypothesis generating experiment, measuring unexpected or unknown changes in known and unknown metabolites.	Hypothesis testing experiment, measuring the expected changes of known metabolites.
>1000s of metabolites measured with a small number identified	Approximately 20 metabolites measured, all being identified with confidence.
Data acquisition without prior knowledge of interesting metabolites, gunshot.	Specific metabolites targeted, involving extensive specialisation of sample preparation and analytical collection technique.
Explorative experimentation.	Deterministic experimentation.
Semi-quantitative/relative quantification via the addition of standards or isotopic internal standards.	Absolute quantification via the addition of standards or isotopic internal standards.
No chemical standards required.	Chemical standards are required.
Technically challenging. The amount of accessible information generated is based on appropriate experimental design and data analysis methods.	Technically simple. Although experimental design and data analysis methods are still important, results are more robust and less subject to small changes in experimental design.
Unbiased (relatively) as there is no preselection of metabolites.	Limited scope due to preselection of metabolites.

(Naz *et al.* 2014; Patti *et al.* 2012; Dunn *et al.* 2011).

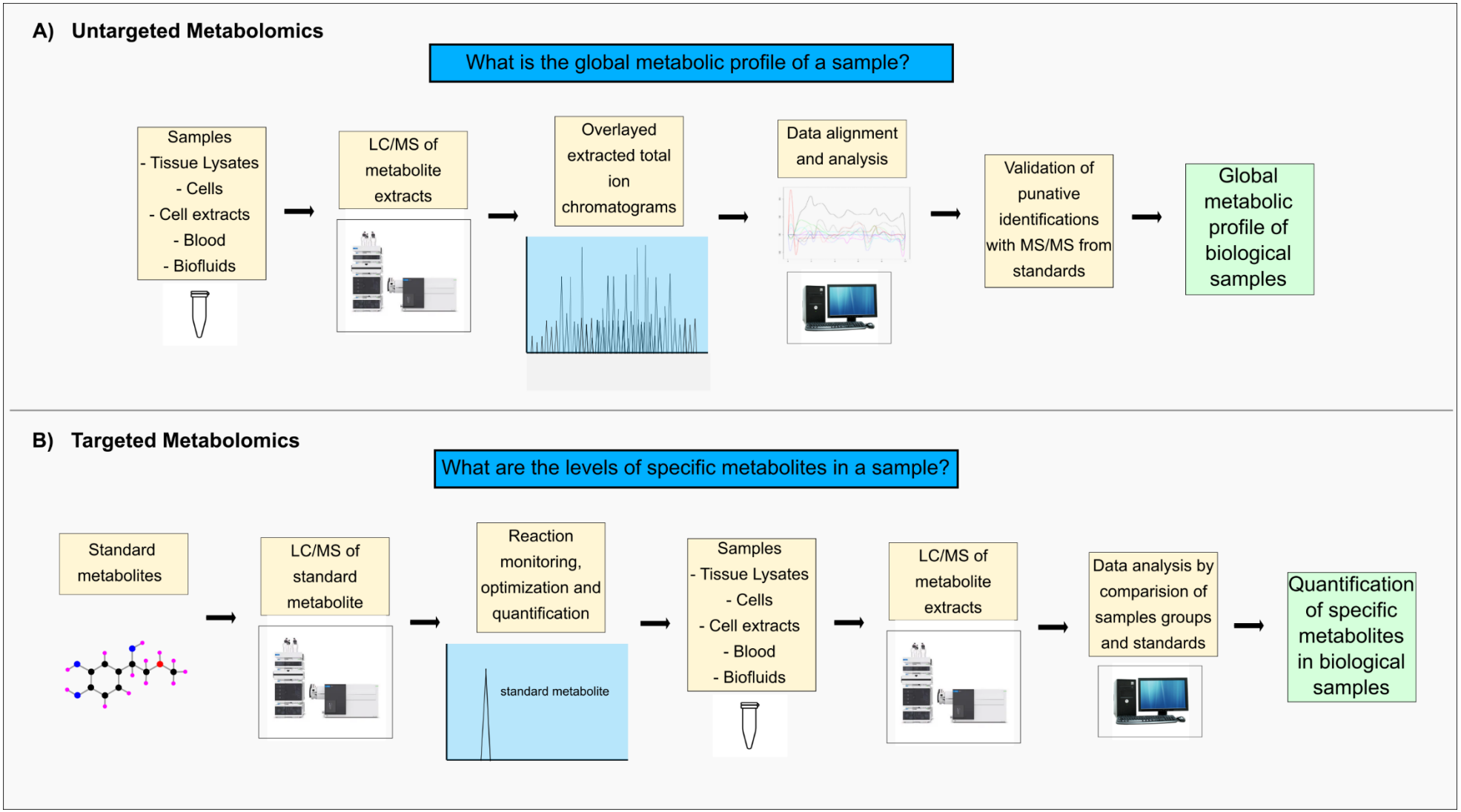


Figure 1.9. LC/MS analysis sequence in untargeted and targeted metabolomics. A) Shows the aims and typical LC/MS analysis sequence of untargeted metabolomics B) Shows the aims and typical LC/MS analysis sequence of targeted metabolomics (adapted from Patti *et al.* 2012).

1.2.3. Measuring the metabolome (methods)

Once a method of analysis has been established (targeted or untargeted), sample preparation and experimentation can be considered. Many unique and specific methods for the various steps involved in this process are available, even when considering a single model organism or cell culture type. As such, great emphasis has been put on reviewing the methods currently available (Riekeberg and Powers, 2017; Faijes *et al.* 2007; Rabinowitz, 2007; Villas-Bôas *et al.* 2005). Sample preparation is understood to be a limiting factor within metabolomics, and acts as a large source of variability within metabolomic analysis. Sample preparation here refers to the release and collection of metabolites from cells at a known time point (Villas-Boas *et al.* 2006). Researchers aim to establish a sample preparation workflow that does not impact upon the number of metabolites obtained or affect the internal metabolite signature of a sample (Kapoore *et al.* 2017). Sample preparation can be broken down into four specific areas outlined in Figure 1.10.

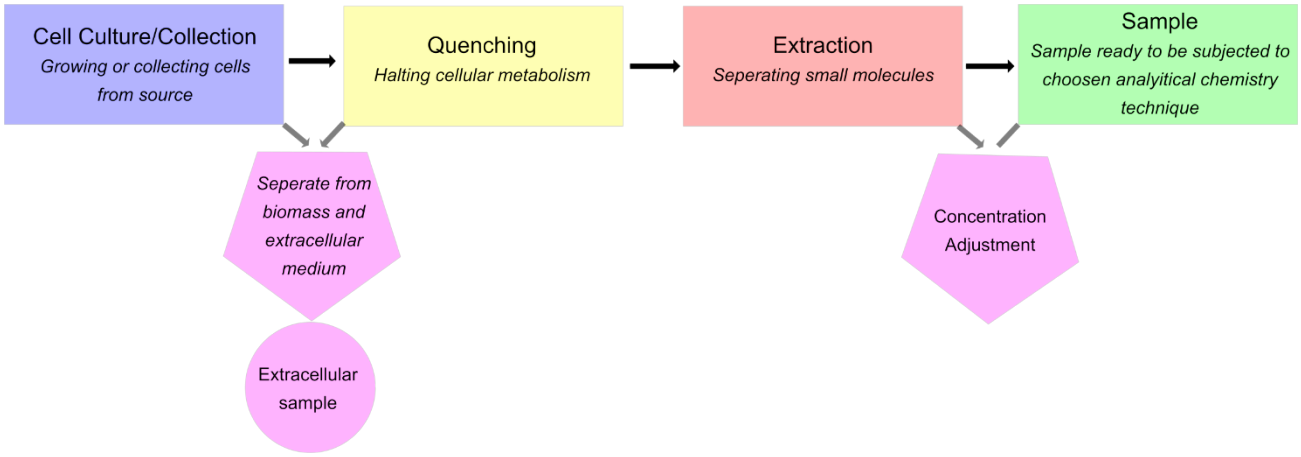


Figure 1.10. General steps involved in metabolite sample preparation. Solid black arrows represent the main sequence of events, light grey arrows indicate alternative/optional steps (adapted from Villas-Boas *et al.* 2006).

Cell growth is normally considered routine with few researchers placing much emphasis on this and instead following whatever common practise is in place for their model or cell line of choice. However, recent investigation concerning mammalian cell line culture shows the choice of culture media can significantly impact on the ability to detect metabolite levels (Daskalaki *et al.* 2018; Huang *et al.* 2015). Within the microbial community, limited work has been done to assess growth conditions effects on subsequent metabolite analyses, although some groups have developed novel methods of cell growth. (Tang, 2011;

Rabinowitz, 2007; Brauer *et al.* 2006). Commonly, extracellular metabolites are measured alongside intracellular metabolites which arguably reduces the overall impact that the choice of cell growth may have to the overall detection of any given metabolite, although mis-location may occur (Pinu and Villas-Boas, 2017). This methodology does not account though for the complexity caused by cell growth media creating additional hurdles for analysis, adding noise and influencing metabolite utilisation by cells (Creek *et al.* 2013). All of these factors cause complications to metabolite detection and hence more research area is required in this area.

Quenching aims to halt the metabolism, thereby providing a snapshot of the *in-vivo* metabolic state of a cell or organism at a given point in time. Pertaining to the dynamic and fast turnover rate of metabolites, especially those within primary metabolism mentioned previously, sample quenching works to minimize the formation or degradation of metabolites by inactivation of cellular metabolism and enzymatic activity (Kapoore *et al.* 2017; Sasidharan *et al.* 2012; Canelas *et al.* 2010; Winder *et al.* 2008; Faijes *et al.* 2007; Rabinowitz, 2007; Villas-Boas *et al.* 2007). This process is usually conducted by placing cells in contact with a very cold (<-40°C) or a relatively hot (>80°C) solutions. Different cells require different treatments, with considerations regarding their cell wall make up crucial to ensuring this vital stage is conducted without negative consequence (Pinu *et al.* 2017). Without such consideration it is possible that leakage of metabolites will occur due to damage to cell wall structures resulting in intracellular metabolites (normally the target for collection) leaking into the extracellular medium (generally removed via centrifugation) leading to a loss of metabolites and thus effecting the estimation and detection of metabolite abundances (Pinu *et al.* 2017; Kim *et al.* 2013; Canelas *et al.* 2009). Often several cellular washing steps are employed during quenching in an effort to remove all extracellular metabolites and media that may contaminate readings. However, this process often adds considerable time to the intended rapid process of cellular quenching and is often subject of suspicion when discussing metabolite leakage. In mammalian studies Ser *et al.* (2015) suggest that no washing step should be used, stating that they appear to play a big role in interfering with metabolite signals detected. Kapoore *et al.* (2017) however reported that employing a single washing step was preferable, resulting in the minimum of intracellular leakage of the methodologies tested. In microbial cells, washing steps are now considered arbitrary, adding undesirable complexity, and lengthening the time of the quenching procedure thereby harbouring the maintenance of very low experimental temperatures (Canelas *et al.* 2010; Winder *et al.* 2008; Rabinowitz, 2007).

Regardless of which quenching methodology is used, metabolites are still required to be extracted from cells. Within sample preparation, extraction methods have disproportionally been the focus of many of the reviews within the field, and so a great number of methods are available. This can also be attributed to the large variety of metabolites present in cells, as many argue that no single extraction technique is capable of retrieving all types of metabolites simultaneously, whilst others argue that the use of multiple methods should be employed to ensure metabolome wide extraction (Riekeberg and Powers, 2017; Canelas *et al.* 2009; Oldiges and Takors 2005; Mashego *et al.* 2003). Extraction can be approached in either a chemical or mechanical fashion, or occasionally using a mixture of the two. The scope of extraction methods is so large it would be unrealistic to discuss the entirety. They range from solvent based solutions such as boiling ethanol and cold methanol to mechanical methods like ultra-sonification and microwave extractions. Some groups choose to extract after storage, however Fiehn (2002) argued that freeze drying samples can also lead to metabolites becoming irreversibly bound to cell walls and membranes and thus decrease the efficiency of subsequent extraction. To address this, extraction is now commonly carried out prior to storage removing or separating undesirable cellular components that may have negative consequence on experimentation (Tredwell *et al.* 2011).

The last step in sample preparation is storage, two alternatives exist for this: freezing at very low temperatures (-80°C) or freeze drying. Both methods aim to limit or stop any cellular activity or change that may occur in between this stage and running samples on the analytical method of choice. Freezing is both fast and accessible for most laboratories, given that -80°C freezers are commonplace. This process maintains sample integrity as cellular metabolism and enzymatic activity cannot continue at these temperatures. Efforts to avoid partial thawing should be made and dependent on the size of the experiment may be limited by the availability of physical storage space (Villas-Boas *et al.* 2007). Recent findings also suggest that -80°C reduces the impact of the extraction method on the metabolites within samples, reducing the number of artefacts present (Sauerschnig *et al.* 2018). Freeze drying by contrast takes considerably longer and although low temperatures can be employed, they are not comparable to -80°C of freezing. Freeze drying can cause degradation of metabolites, as this process takes time due to the presence of water and degradative processes may still continue (Villas-Boas *et al.* 2007). However, freeze drying is extensively used in the field and is often a required process of many of the analytical techniques in use within metabolomics, owing to the necessity of removing water because of its ability to interfere with instrumentation.

1.2.4. Analytical variation in metabolomics studies

The inherent features that make metabolites a desirable area of study are likewise the very same features that make metabolites difficult to obtain and measure. Sample preparation must be well designed in order to reliably determine metabolites of interest, but careful consideration should also be given to the choice of analytical method. Currently the most commonly used techniques are that of Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS) (Caudy *et al.* 2017).

1.2.5. Nuclear Magnetic Resonance (NMR)

NMR is widely used within metabolism research, especially for clinical based application due largely to its rapid turnover, minimal sample preparation, excellent structural elucidation, and non-destructive nature. Allowing for the measurement of *in vivo* metabolites, NMR enables clinical studies to measure a cell's response to specific stresses. Often this involves the measurement of biofluids dosed with pharmaceutical drugs, allowing for the metabolic turnover and cellular disruption to be estimated in real time (Caudy *et al.* 2017; Dunn *et al.* 2005; Dunn and Ellis, 2005).

There are a variety of NMR experimental techniques ranging from simple ^{1}H or proton studies, to the more complex two-dimensional experiments such as HSQCs (Heteronuclear Single Quantum Coherence). All, however, rely on the use of strong magnetic fields and radio frequency pulses to the nuclei of isotopic elements, namely ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P. The surrounding magnetic field introduces nuclear spin to these isotopes and the absorption of radio frequency pulses then allows nuclei to be promoted from low energy to high energy spin states, inducing what is known as a magnetic moment. This alternation of low and high energy spin states emits measurable radiation, or resonance, during the relaxation process (from high to low energy). The precise resonance of this relaxation is dependent on the effective magnetic field experienced by the nucleus, thus the chemical environment and the electrons shielding the nucleus can affect this resonance, among other factors. This influence of shielding electrons on resonance is known as chemical shift, in order to produce readable spectra, information about the environment and shielding effect must be gathered in order to adjust for this. Common references are customary depending on the type of NMR being conducted, for instance in a ^{1}H or proton studies the resonance of an observed proton is compared to that of tetramethyl silane (TMS). This information via Fourier transform is

then converted into readable spectra for molecular interpretation (Dunn and Ellis, 2005; Mitchell and Costisella, 2004).

Two-dimensional experimentation aims to focus on the interactions between two of the observable features, so for instance HSQCs are used to determine the proton-carbon single bonds. Many other types of two-dimensional experimentations are possible all providing information about the bonds and atoms surrounding those isotopes which have been measured, this allows for considerable structural information to be obtained (Mitchell and Costisella, 2004).

There are some notable disadvantages of NMR for use in metabolomics, the greatest of which being the limited sensitivity of NMR (typically μM to nM) when compared to MS (typically fM to aM), thus reporting only on high concentration metabolites (Tredwell *et al.* 2011; Dunn *et al.* 2005; Dunn and Ellis, 2005). Also worthy of consideration is the relatively large amounts of sample that are needed to conduct NMR experimentations and the extensive cost of introducing isotopes to cells if the natural abundance of them does not suffice for experimentation (Dunn and Ellis, 2005; Griffin *et al.* 2002). Although it is possible to improve the sensitivity of experiments via the use of cryogenic probes and the application of higher magnetic fields (Keun *et al.* 2002).

1.2.6. Mass spectrometry (MS)

In contrast to the requirements of NMR spectroscopy, the wide variety of MS instruments available offer very high sensitivity, relatively small sample requirements and a variety of ionization approaches with which to streamline experimentation.

GC-MS (Gas chromatography mass spectrometry), the standard for substance identification, is a commonly used and well evidenced method in metabolomics and often coupled with NMR (Tredwell *et al.* 2011). GC-MS offers many benefits: high chromatographic resolution, high sensitivity, fast turnover rate, small injection volumes and is capable of analysing highly volatile, thermo- and energetically stable low molecular weight compounds in the gas state. This makes GC-MS an excellent tool for the chemical analysis of pheromones, plant volatiles and molecules in breath; chemical compounds that would be incredibly difficult to resolve (Oleander *et al.* 2019; Deng *et al.* 2004; Perera *et al.* 2002). However, GC-MS has its limitations. Many polar molecules are poorly volatile and therefore

require extensive two-stage derivatization to ionize. This process often requires the use of high temperatures (a concern regarding sample stability) and thus reduces the reliability of experimentation (Caudy *et al.* 2017; Dunn and Ellis, 2005). LC-MS (Liquid Chromatography mass spectrometry) however, allows for analysis at lower temperatures in a liquid state, removing the need for extensive sample preparation. This permits LC-MS to separate a broad range of metabolites and is the most widely used method within metabolomics (Caudy *et al.* 2017; Dunn and Ellis, 2005).

UHPLC-QTOF-MS (Ultra High-Performance Liquid Chromatography – Quadrupole Time of Flight – Mass Spectroscopy) utilising electrospray ionisation (ESI) was the platform selected for this research. In the initial phases of this hyphenated technique, chromatography is employed to separate molecules within complex mixtures into single chromatographic peaks. This separation is achieved using a two-phase system, involving interactions between solute molecules (known as the mobile phase) and specific ligands bound to chromatography matrixes (known as the stationary phase) (Aguilar, 2004).

Reversed phase chromatography (RPC) was the method employed throughout this research. RPC is capable of separating molecules that possess hydrophobic properties with excellent recovery and resolution. In addition, the use of ion pairing modifiers in the mobile phase can also allow for the separation of some charged solutes. In RPC the mobile phase is hydrophilic in nature and the stationary phase is hydrophobic (Moldoveanu and David, 2013). To achieve a hydrophobic stationary phase, hydrophobic molecules are added to the hydrophilic silica particles of the column. One of the most common hydrophobic molecules used in this process is stearic acid (C18 group). These hydrophobic C18 groups coat the silica particles as shown in Figure 1.11 (Waters, 2021a).

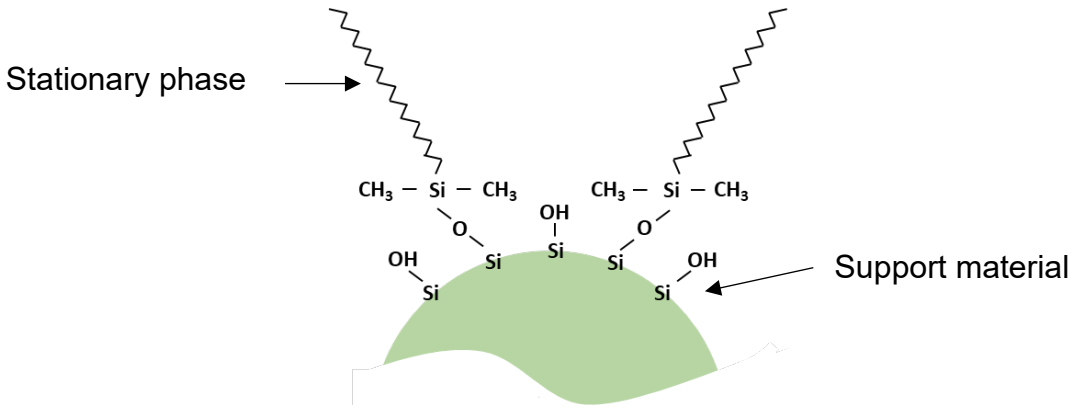


Figure 1.11. The attachment of hydrophobic molecules to the silica particles of the column.

The retention of molecules in RPC is reliant upon the degree of interaction between the analytes in the hydrophilic mobile phase with the hydrophobic molecules of the stationary phase (Moldoveanu and David, 2013). As molecules flow through the chromatography column they partition between the stationary and mobile phases depending upon their degree of hydrophobicity (Figure 1.12.). The more hydrophobic a molecule is, the more interactions it will have with the stationary phase and the longer it will be retained on the column giving it a higher retention factor (k). Thus, less hydrophobic molecules will elute from the column first. (Figure 1.13.) (Stanbury *et al.* 2017; Jandera, 2005).

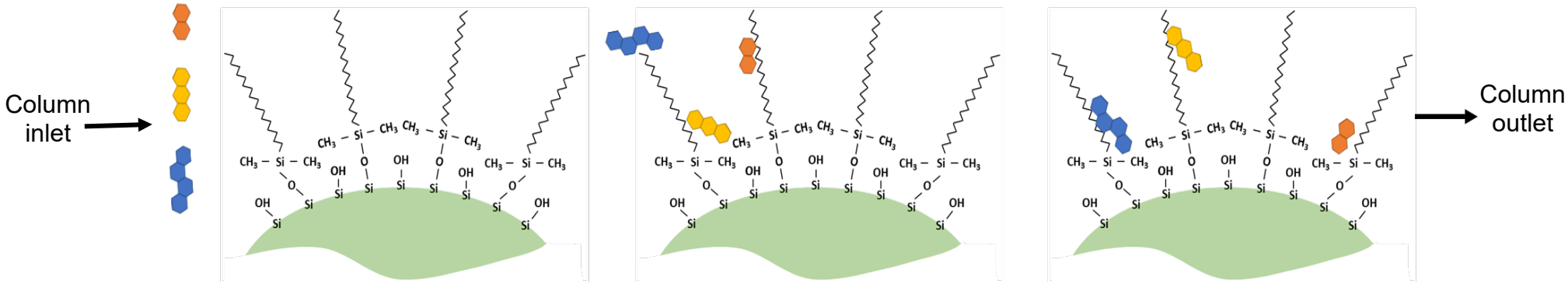


Figure 1.12. RPC Column interactions of molecules varying in hydrophobicity. Entering from the left is the analyte solution containing three different molecules ranging in hydrophobicity. The orange molecule represents low hydrophobicity, the yellow molecule medium hydrophobicity and the blue molecule high hydrophobicity. Their time in transit and interactions with the hydrophobic attachment on the silica particles are shown, with each still representative of a single point in time. The column outlet, to the mass spectrometer is shown on the right, with the image showing which molecules will reach this end first.

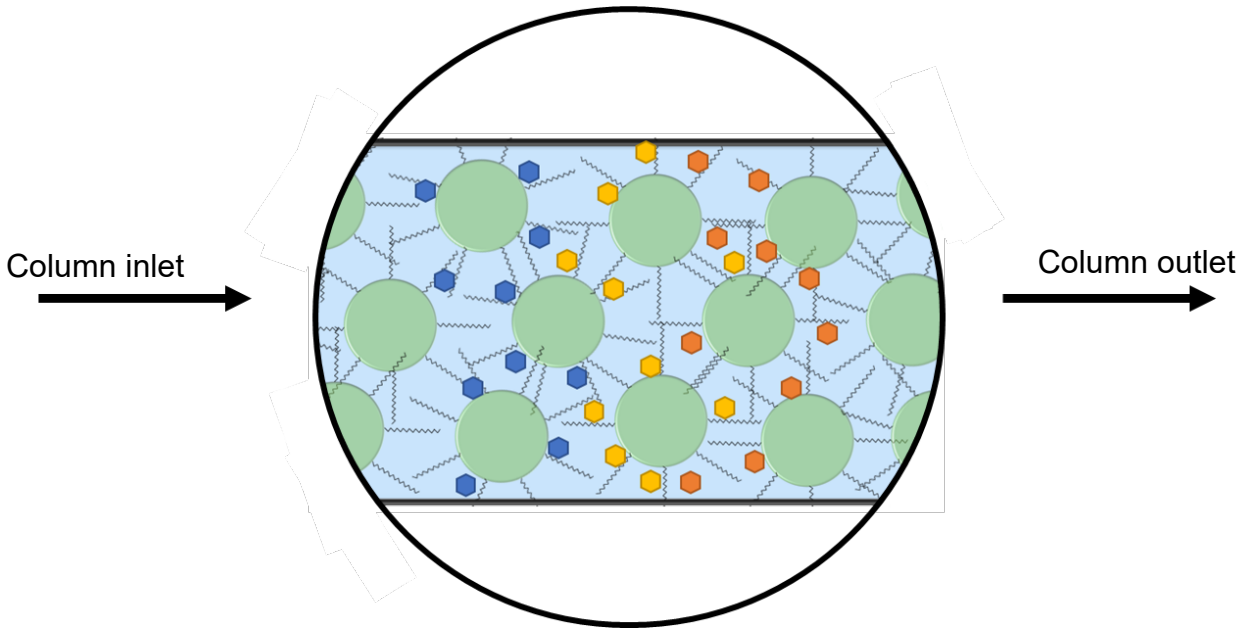


Figure 1.13. Cut through of a column operating in RPC. The attachment of hydrophobic molecules to the silica particles of the column. Entering from the left is the analyte solution containing three different molecules ranging in hydrophobicity. The orange molecule represents low hydrophobicity, the yellow molecule medium hydrophobicity and the blue molecule high hydrophobicity. Their time in transit and interactions with the hydrophobic attachment on the silica particles are shown. The column outlet, to the mass spectrometer is shown on the right, with the image showing that the orange molecules will reach this end first, then the yellow then the blue.

In GC, the mobile phase plays no part in separation; however, in HPLC, the ability to change the mobile phase is considered one of its most powerful attributes. By adding organic modifiers (such as methanol or acetonitrile) to the mobile phase, the hydrophobicity of the mobile phase can be controlled. If the percentage of organic modifier in the mobile phase is decreased, the hydrophobic molecules in the solute will spend more time interacting with the stationary phase, increasing their retention time (Moldoveanu and David, 2013).

The resolution of a molecule is determined by the amount of time it spends interacting with the stationary phase. If the concentration of the organic modifier present in the mobile phase is in such high concentrations that it fulfils the same hydrophobic interaction as the stationary phase, then all molecules will elute at the same time leading to poor resolution (Figure 1.14.) (Moldoveanu and David, 2013; Jandera, 2005). However, as the amount of organic modifier within the mobile phase is reduced, the hydrophobic interactions of the molecule with the stationary phase increase leading to an increase in interaction time and subsequent elution time. This allows for greater resolving power (Figure 1.14.) (Moldoveanu and David, 2013; Jandera, 2005).

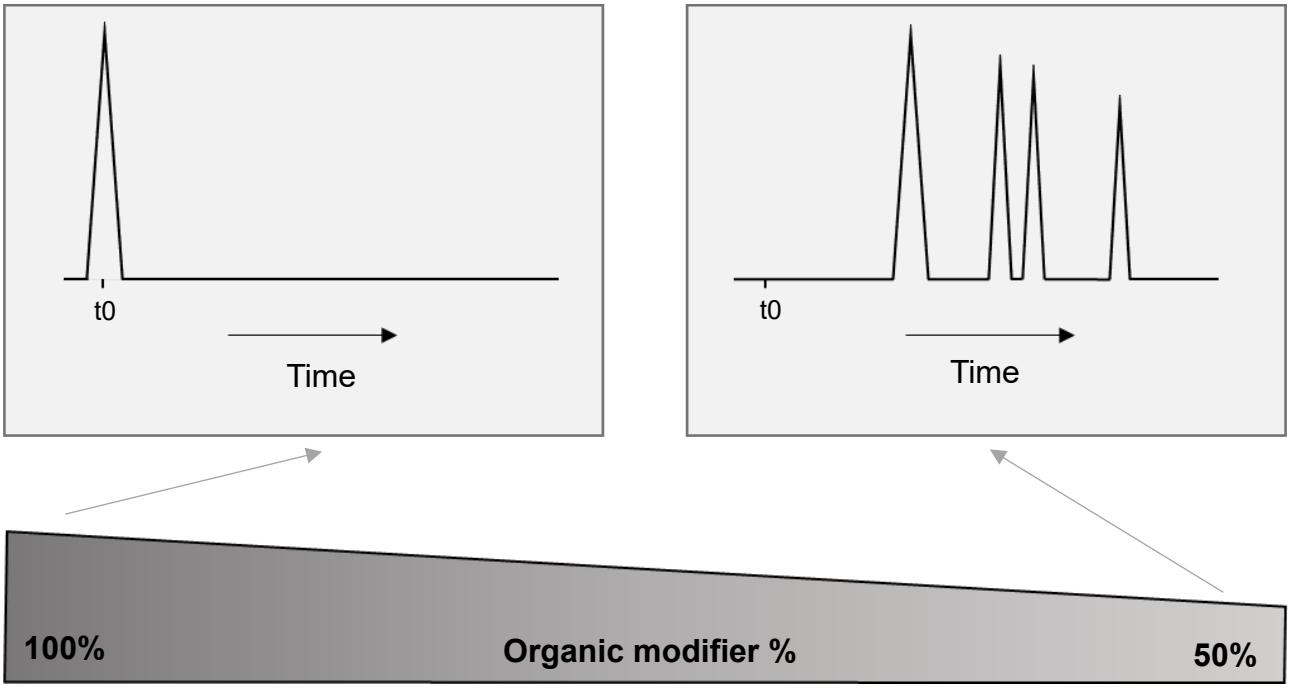


Figure 1.14. The effect of organic modifier percentage on the peak resolution in RPC. The percentage of organic modifier is shown on the gradient bar at the bottom, with arrows indicating which percentage is being to produce the above chromatograms.

The type of organic modifier used will significantly impact the hydrophobicity of the mobile phase due to the number of carbon atoms present. As the mobile phase become less hydrophobic, interactions with the stationary phase increase and therefore the retention time of molecules increase. In addition, other factors such as UV absorbance, viscosity, toxicity, and boiling point can affect the choice or modifier (Waters, 2021a; Nikitas *et al.* 2002; Valko *et al.* 1993).

Organic modifier percentages can be used in two ways: either isocratic elution or gradient elution. Isocratic elution involves the continued use of the same percentage of water to organic modifier (known as a static ratio) throughout the duration of the experimentation. However, isocratic elution can lead to poor resolution of closely related molecules and lengthy analysis times. It is often therefore used alongside size exclusion chromatography (Waters, 2021a; Nikitas *et al.* 2002; Valko *et al.* 1993). More commonly used is gradient elution, whereby the percentage or ratio of the organic modifier is varied throughout the analysis. Generally increasing linearly overtime, the rate of this elution can be modified with a balance between compound retention and experimental run time being the desired outcome. Gradient elution greatly improves the run time of experimentation and allows for

greater resolution of closely related species (Nikitas *et al.* 2002). This increase in resolution is caused by changes in retention time due to the changing hydrophobicity of the mobile phase having varying effects on the interactions between then stationary phase and molecule. It will depend on the molecular nature of the analyte, and to a greater extent, because of the physical difference between the injector end and the detector end of the column. The molecules at the injector end interact with the higher organic modification percentage before those at the detector end. This causes the injector end molecules to be travelling at a slightly higher velocity as they partition from mobile to stationary phase, creating sharper peaks and a higher resolution / sensitivity (Nikitas *et al.* 2002).

In contrast to RPC, Normal Phase Chromatography (NPC) and Hydrophilic Interaction Liquid Chromatography (HILIC) are capable of separating molecules that possess hydrophilic properties. Silica bound columns (without additions) create a hydrophilic stationary phase with hydrophobic molecules quickly eluting from the column. Hydrophilic molecules, however, have a range of interactions with the stationary phase and hence variable retention times, thus allowing for their resolution (Stanbury *et al.* 2017; Thermo Fisher Scientific, 2014; Jandera, 2005; Plumb *et al.* 2004). In NPC the mobile phase used is 100% organic; hydrophilic analytes interact strongly with the stationary phase and may not elute from the column. HILIC, often described as ‘reverse RPC’, employs a hydrophilic stationary phase (typically silica) alongside a miscible organic solvent (e.g. acetonitrile) with the aqueous portion varying in percentage, or used as a modifier (Thermo Fisher Scientific, 2014; Jandera, 2005). This exerts an opposite action to that of the organic modifier in RPC and makes HILIC’s selectivity complementary. Thus, HILIC and RPC are often run in tandem with each other (Caudy *et al.* 2017; Thermo Fisher Scientific, 2014; Dunn and Ellis, 2005).

Column efficiency is measured using theoretical plates, *N*, and is normalised using the length of the column, *L*, to give the height equivalent theoretical plate, *HETP* or *H* (Figure 1.15a).

A)

$$HETP = \frac{L}{N}$$

B)

$$HETP = a(d_p) + \frac{b}{u} + c(d_p)^2 u$$

Figure 1.15. Equations to determine *HETP*. A) Simplified equation to determine *HETP*. [L] is column length, [*N*] is plate count and [*HETP*] is height equivalent to a theoretical plate. B) van Deemter equation (The Van Deemter equation, 2021; Waters, 2021b).

The van Deemter equation (Figure 1.15b) describes the various factors which can influence *HETP*, and is comprised of three terms:

- The A term (eddy diffusion) is related to the particle size (d_p) of the packing material, as well as the nature and quality of the packing itself. Eddy diffusion is the result of multiple flow paths through a packed column. As such, the uniformity or non-uniformity of the column is also an important consideration. Factors influencing this term include particle size (d_p), particle shape, pore structure, quality of column packing, column wall material and column diameter (The Van Deemter equation, 2021; Waters, 2021b; Edge, 2003).

- The B term (longitudinal diffusion) concerns the diffusion of the analytes in the mobile phase and on the stationary phase. This term decreases with increased linear velocity of the mobile phase. Factors influencing the B term include the linear velocity of the mobile phase (u), diffusion coefficient of the analyte, mobile phase viscosity, temperature and the molecular mass of the analyte (The Van Deemter equation, 2021; Edge, 2003).

- The C term (mass transfer) is related to both the linear velocity of the mobile phase and the particle size squared. Mass transfer is the interaction between analyte molecules with the stationary phase and the diffusion distance in and out of packing material pores. Factors influencing this term include particle size (d_p), linear velocity of the mobile phase (u), diffusion coefficient of the analyte, porosity of the packing particles, mobile phase viscosity, temperature and retention factor (The Van Deemter equation, 2021; Edge, 2003).

The lower the *HETP* value the greater the efficiency of the column. By selecting smaller particle sizes such as those present in UHPLC (<2 μm), a smaller H (larger *N*) can be achieved due to impacts on the A and C term of the van Deemter equation (Waters, 2021b; Henry, 2014). This is turn impacts the optimal operating range with respect to linear velocity which impacts the B term. Larger particles (ie.10 μm) have very narrow optimal operating ranges. If the speed of the mobile phase is too slow or too fast, an increase in *HETP* is seen (Waters, 2021b; Henry, 2014). Smaller particles have a larger linear velocity range, with much lower *HETP* at higher mobile phase velocities. For these reasons UHPLC offers advantages to the more traditional HPLC (High performance liquid Chromatography) via the use of columns packed with smaller particles that allow for higher mobile phase velocities (Waters, 2021b). This results in higher resolution chromatography that operates within a faster run time (Plumb *et al.* 2004).

In addition, UHPLC have shorter column lengths than HPLC which has positive *HETP* effects on the equation shown in Figure 1.15a, as well as reducing band spreading. Specialised pumps within UHPLC are capable of dealing with higher pressure, again having positive effects on mobile phase velocity as well as having faster run time (without compromising performance) and considerably less solvent use than traditional HPLC. The narrower chromatographic peaks produced in UHPLC make the process of successfully analysing complex mixtures much more attainable, a very useful addition for metabolomics (Waters, 2021b; de Hoffmann and Stroobant, 2007).

Following on from chromatographic separation fractionated eluates are then introduced into the ionisation chamber, where ionisation takes place. ESI is known as a ‘soft ionisation’ technique, owing to the production of intact ions related to analyte molecules and fewer fragment ions than vacuum ionisation techniques such as electron ionisation (Banerjee and

Mazumdar, 2011). In an ESI ion source, an analyte sample (often a chromatography eluent) is injected through a small metal needle or capillary (~0.1mm) at a low flow rate (1-20uL/min). The tip the needle or capillary is held at very high voltage (2-6kV) relative to a counter electrode at the MS interface (located 1-3cm from the spray needle) (Banerjee and Mazumdar, 2011). Accompanying this needle is a flow of inert gas (usually nitrogen) at atmospheric pressure. The gas flow is used to more efficiently nebulise or break up the liquid stream into tiny droplets and helping to direct the stream toward the MS (Figure 1.16A). One of the benefits of this system is that ESI allows for both negative and positive ionisation mode permitting the acquisition of a greater range of ionised species. The sign of the potential difference between the capillary needle and the MS interface will determine which ion type is produced (Figure 1.16A) (Villas-Boas *et al.* 2007).

This strong electric field (created due to the potential difference between the high voltage of the needle tip and the relatively low voltage of the MS interface) has a dispersion effect on the sample in solution and creating an aerosol of highly charged electrospray droplets. Ion formation in ESI must occur in solution prior to entering the gas phase (Konermann *et al.* 2013; Banerjee and Mazumdar, 2011). It is thought that molecules undergo electrochemical redox reactions at the liquid metal interface of the capillary needle or acid-base reactions in solution. The exact mechanism by which ions desolve in the gas phase is still a matter of debate and research; however, it appears the nature of the analyte itself informs the choice of mechanism (Konermann *et al.* 2013). There are currently two mechanisms considered for this process, which include:

- The Ion Evaporation mechanism (IEM), whereby the electric field at the surface of the highly charged droplets becomes so highly charged that larger droplets break down into smaller ones (Konermann *et al.* 2013). This happens sequentially until just the desolved ion remains (Figure 1.16B.). It is thought this is the favoured method for smaller ions (Konermann *et al.* 2013).

- The Charged Residue Module (CRM), whereby ions eventually desolve as solvent molecules sequentially leave the surface of droplets until just desolved ions are left (Figure 1.16C.) (Konermann *et al.* 2013). This method is believed to be favoured by larger species with multiple charges (Konermann *et al.* 2013).

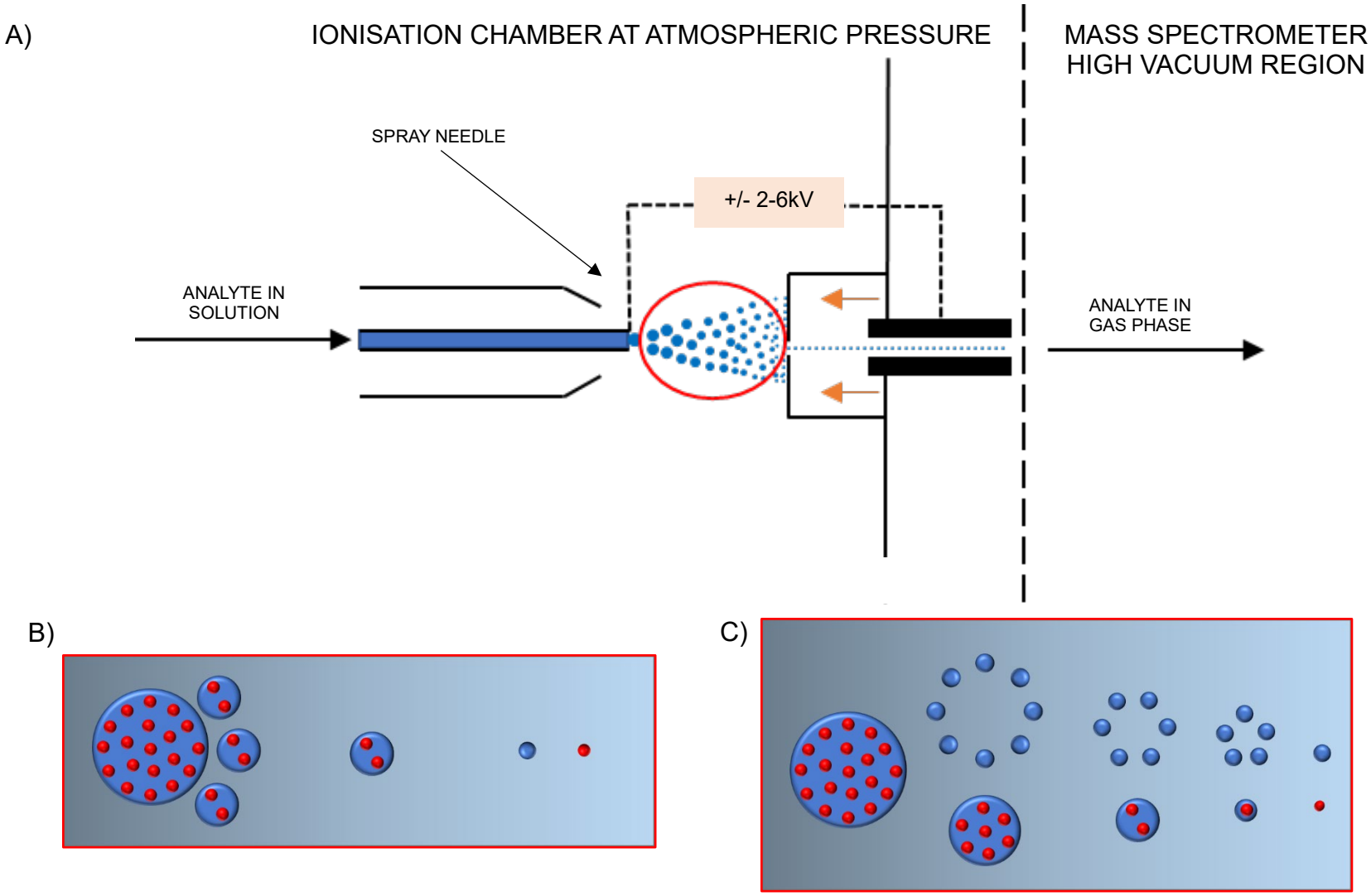


Figure 1.16. Schematic illustration of Ion desolvation. A) Schematic of ESI, with red circle indicating the area which has been enhanced in part b and c B) Ion Evaporation mechanism of desolvation, blue represents solvent and red represents ions C) Charged Residue Module of desolvation, blue represents solvent and red represents ions.

The Waters SYNAPT G2-Si (Figure 1.17.) operates in two modes: HDMS^{TM} mode (using Triwave ion mobility separation) and ToF mode. The operation mode used throughout this research was ToF mode (Waters, 2015). In ToF mode once the sample has eluted from UHPLC, it is injected into the ion source where the molecules are converted into ions. There are many types of ion source used within metabolomics, with ESI (the technique used throughout this research) being the most common (Dettmer *et al.* 2007).

Desolved ions are then attracted into the mass analyser and are accelerated into the quadrupole. The quadrupole module consists of four quadrupoles (metal rods), arranged in parallel and connected to a radio frequency (RF) voltage supply where an alternating electrical field is created between the quadrupole rods. Once molecules reach the quadrupole they start to spin within an imaginary cylinder, the size of which is dependent on the m/z and the RF voltage (Villas-Boas *et al.* 2007). In MS and MS/MS mode the first quadrupole (Q1) can act as a mass filter for the selection of a specific range of ions based on their m/z. Alternatively it can act in RF-only mode allowing all ions to be transmitted (Allen and McWhinney, 2019). In MS mode the second quadrupole (Q2) acts in RF-only mode and so not inducing fragmentation and allowing accurate determination of the mass of unfragmented precursor ions to be detected. In MS/MS mode, Q2 acts as a collision cell. Ions are bombarded by a neutral gas such as nitrogen or argon, inducing fragmentation of the ions and allowing the acquisition of product or fragment ions (Allen and McWhinney, 2019). The ToF then applies high voltage pulses to accelerate the ions into a high vacuum flight tube. Within the spectrometer (Figure 1.17.), a dual stage reflector is employed to separate out the ions based on m/z and reflect them back to the detector. The m/z is relative to the ToF with smaller m/z's reaching the detector first (Allen and McWhinney, 2019; Waters, 2015). Therefore, UHPLC-QTOF-MS's high-resolution power, excellent sensitivity and high selectivity makes the use of a great candidate for untargeted metabolomics (Allen and McWhinney, 2019).

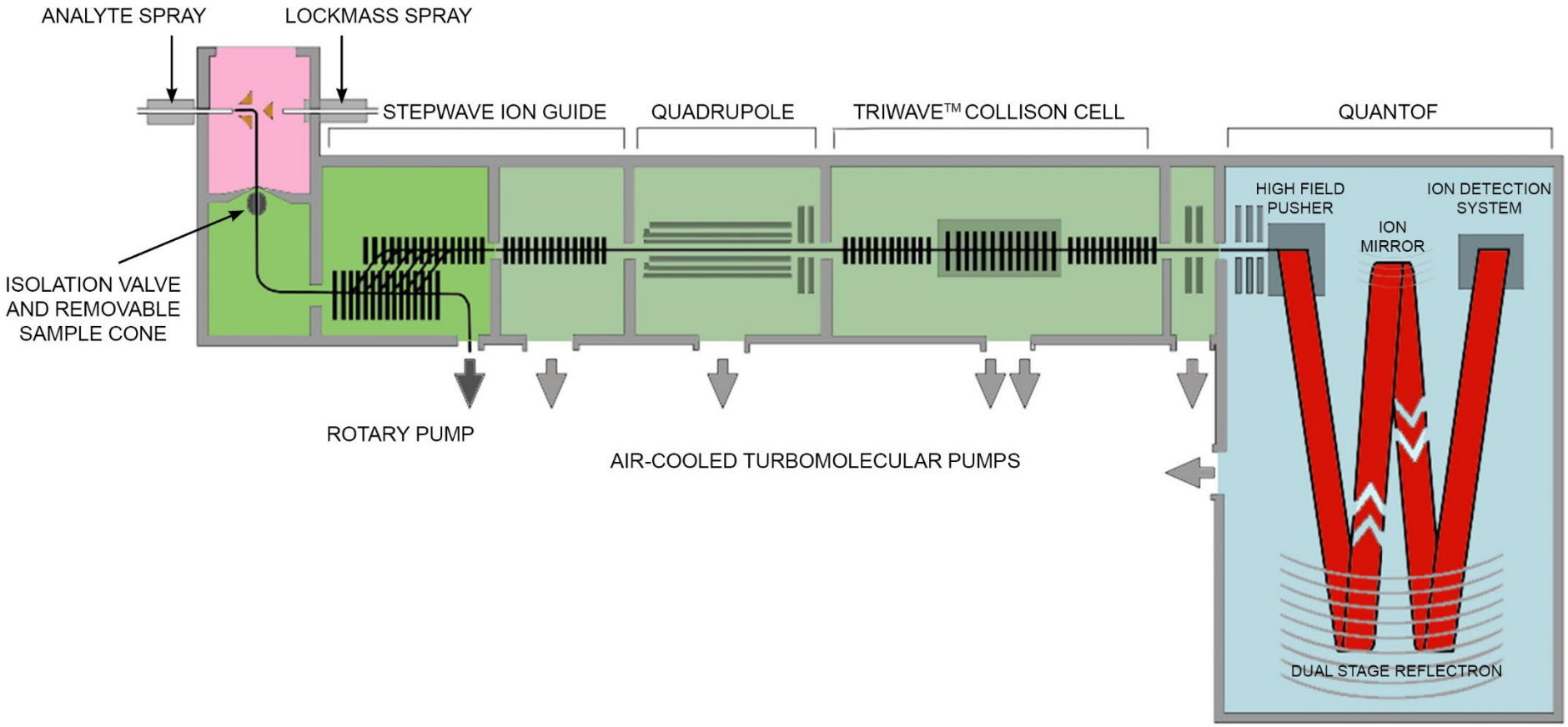


Figure 1.17. Schematic illustration of a Waters SYNAPT G2-Si. (adapted from Waters, 2005).

1.2.7. Data analysis

Having acquired data by NMR or MS, converting the raw instrumental data into data that can be processed by the considerable number of statistical and metabolite identification tools is a challenging endeavour. Within LC-MS, pre-processing such as deconvolution, baseline drift, normalisation, noise, and peak detection needs to be carried out prior to data being analysed in a traditional discovery-based fashion (Hendriks *et al.* 2011). Fortunately, there are many freely available data processing programs for use with LC-MS, such as MzMine and XCMS (Ni *et al.* 2016; Benton *et al.* 2014; Gowda *et al.* 2014; Zhu *et al.* 2013; Tautenhahn *et al.* 2012a; Tautenhahn *et al.* 2012b; Pluskal *et al.* 2010). Within NMR and other methodologies, the same pre-processing steps must be performed however these are normally internally governed by the types of machines used to acquire the data, with programs and databases provided to users at a cost (Hendriks *et al.* 2011).

Once pre-processing has been achieved, multivariate statistical analysis and metabolite identification are readily employed. Again, within LC-MS thanks to the advancement of bioinformatic tools, metabolite identification has become a largely automated process, with several programs offering peak picking, retention time alignment and statistical analysis (Chong *et al.* 2019a; Chong *et al.* 2019b; Ni *et al.* 2016; Benton *et al.* 2014; Gowda *et al.* 2014; Zhu *et al.* 2013; Tautenhahn *et al.* 2012a; Tautenhahn *et al.* 2012b; Pluskal *et al.* 2010). Metabolite identification is still often one of the main bottlenecks of metabolomics experimentation. In recent years, great effort has been made to increase the breadth of online compound databases, but many metabolites are left unidentified. Without confirmatory runs of expensive known standard compounds, any database matches that are achieved, even using retention time and MS/MS data, can only be considered punitive identifications.

The most common type of statistical analysis within metabolomics is multivariate analysis, where the power of computational models is used to assess and compare a multitude of groups and establish the factors and strength of the factors on which those models are built. These include a PCA (principal component analysis) which is a separation method that uses linear transformation to preserve as much variation within the data as possible without taking account of the predefined groups. PCA provides an unbiased separation, only revealing group structure if enough variation is present within groups (Hendriks *et al.* 2011; Worley and Powers, 2016). By contrast PLS-DA (partial least squares discriminant analysis) (for

use in multiple group comparisons) and OPLS-DA (orthogonal partial least squares discriminant analysis) (for use in pairwise comparisons) force separation between groups, building models around the strongest sources of variation between groups (Worley and Powers, 2016).

It is important to understand that these two modelling types are different, whilst they look similar and both aim to report on general trends within a data set. OPLS-DA and PLS-DA force separation between two pre-defined groups at the expense of model reliability, but PCA deals with data as if it exists as part of a single group and then subsequently labels groups. This forced separation of OPLS-DA/PLS-DA has led to this modelling method being commonly used, especially within metabolomics community and hence consideration of OPLS-DAs without subsequent PCA analysis is misguided (Worley and Powers, 2016). It must be conceded, therefore that OPLS-DA/PLS-DA models must be built using weaker sources of variation than their PCA counterparts (Worley and Powers, 2016). This being the case, visible separation in OPLS-DA/PLS-DA is expected, and so CV-ANOVA scores act as more reliable source of information with regards to the significance of the separation seen in OPLS-DAs/PLS-DAs.

Once metabolites have been identified and subject to statistical analysis, significant features can then be mapped via pathway analysis. Once again there are many online programs capable of assisting with pathway analysis, the choice of the pathway analysis program used is often based on user preference and the organism which has been the subject of their studies as so pathway analysis tools offer larger databases for certain organisms than others (Chong *et al.* 2019a; Chong *et al.* 2019b; Karp *et al.* 2017; Hendriks *et al.* 2011; Caspi *et al.* 2007; Kanehisa and Goto, 2000).

1.2.8. The application of metabolomics for disease research

Given that proper metabolic functioning is vital for the health of an organism, the pivotal role that metabolomics is set to take in the future of medicine cannot be overstated. Studies are available that focus on almost every human disease, including diabetes, cancer, neurodegeneration, non-fatty liver disease, obesity, and cardiovascular health to name a few (Pinu *et al.* 2019; Zhang *et al.* 2017; He *et al.* 2015; Lewis *et al.* 2008).

Neurodegeneration has received some attention from the metabolomics community over the last decade, with the pathogenesis of the amyloid diseases (namely Alzheimer's and Parkinson's) chief among these studies (Bourgognon *et al.* 2018; Oresic *et al.* 2011). Metabolomics and neurodegeneration have risen from being virtually non-existent in the literature (0-2 papers per year from 2008 till 2012), to 33 per year in 2018 (Web resource at URL:http://dan.corlan.net/medline-trend.html). Utilising the power of metabolomics within neurodegeneration has the potential to aid progress toward the development of effective treatments, currently halted by a lack of understanding of the early pathology of disease (Wilkins and Trushina, 2018). In addition, it can provide reliable and disease specific biomarkers from minimally invasive samples (such as blood, saliva, or urine) capable of early or even pre-symptomatic diagnosis (Bamji-Stocke *et al.* 2018; Dong *et al.* 2018; Liang *et al.* 2016). This would end the qualitative assessments currently in use, many of which are subject to user manipulation, late stage diagnosis or at worst misdiagnosis. To date, definitive diagnosis of Alzheimer's disease (AD) can only be achieved post-mortem via brain tissue examination (Galasko and Golde, 2013).

Current findings from cerebrospinal fluid metabolomics suggest that the metabolic perturbations resulting from these conditions is significant and widescale, even in patients with mild cognitive impairment (Wilkins and Trushina, 2018). Lewitt *et al.* (2013) reports the formation of the 3-hydroxykynurenine and reductions of antioxidant glutathione signalling as disease relevant markers for Parkinson's and Alzheimer's. More commonly reported are changes in the biosynthesis and metabolism of lipids, cortisone, and amino acids, as well as disruption to energy, urea, and bile acid metabolism (Varma *et al.* 2018; Wilkins and Trushina, 2018; Laurens *et al.* 2015; Han *et al.* 2011; Oresic *et al.* 2011). Wilkins and Trushina (2018) reason that reduced glucose utilization is understandably problematic for the highly glucose-dependent brain, suggesting that compensatory mechanisms switch to alternative energy sources, including the catabolism of lipids, to maintain homeostasis within this vital organ. Bourgognon *et al.* (2018) state that the most relevant disease alterations in neuronal metabolism for use as biomarkers are changes in glycolysis, pentose phosphate pathways, polyamines, ceramides, and prostaglandins. They claim that the perturbations to these pathways increase with increasing redox stress and advancing disease progression (Bourgognon *et al.* 2018). While metabolomic findings have provided significant insight into the mechanisms of protein misfolding, there is still an inconsistency and confliction in its results (Wilkins and Trushina, 2018). This is unaided by the common use of mouse models carrying known AD disease genes (such as *APOE4*), as although these models are helpful,

the vast majority of human cases of AD are sporadic with no direct genetic cause and so the amount of information that can be inferred is limiting (Eckert *et al.* 2009). However, the most promising and consistent hallmarks of AD studies appear within lipid metabolism (Varma *et al.* 2018; Toledo *et al.* 2017; Laurens *et al.* 2015; Han *et al.* 2011; Oresic *et al.* 2011). In one of the most significant and largest blood-based metabolomic studies Toledo *et al.* (2017) identified preclinical biomarkers of alterations in sphingomyelins and ether-containing phosphatidylcholines, associated and correlating with AD stage. This is supported by the findings that alterations in the metabolism of branched chain amino acids and bile acids impact directly on lipid metabolism (Pan *et al.* 2017).

Prion focussed studies are rarer by comparison to amyloid studies, however their value for disease pathology and progression as well as increasing and validating biomarkers should not be overlooked. Bourgognon *et al.* (2018) states that until their experimentation, the metabolic profiles of prion-containing neuronal cells had not been examined. Despite large bottlenecks in the amounts of compounds they were able to identify, they conclusively found that mice with prion-infected brains displayed altered metabolic profiles to disease-free mice. As seen within the amyloid diseases, these changes included reduced glycolysis and energy production, changes to amino acid utilisation and enhanced metabolism of the sphingolipids in diseased cells (Bourgognon *et al.* 2018). Importantly, this study found these results prior to the establishment of the traditional symptoms of prion disease, indicating that key metabolic differences are apparent and measurable long before the disease appears to show any atypical symptoms. This is a vital piece of information for the translation of this into biomarker detection (Bourgognon *et al.* 2018). Using serum samples, Pushie *et al.* (2011) found several metabolites associated with chronic wasting disease (CWD), a prion disease effecting cervids (typically white-tail deer and elk). Pushie *et al's.* (2011) NMR studies correlate well with those previously reported for BSE and Scrapie infected sheep, with disruption in degradation pathways and fatty acid synthesis (Pushie *et al.* 2011; Allison *et al.* 2008; Allison *et al.* 2007).

The parallels between amyloid and prion metabolomics findings provide an encouraging level of similarity, suggesting that universal effects of as a result of protein misfolding disease are present within mammals. However, the distinction between these diseases may therefore be difficult to obtain, or otherwise require a new approach to the atypical, disease free, controls used within metabolomics. For instance, many diseases result in increasing redox stress, including disruption to Ca^{2+} signalling, nitric oxide (NO) signalling and general

neuroinflammation (Varma *et al.* 2018; Wang *et al.* 2017; Nakamura and Lipton, 2017). It seems unreasonable to suggest that cells that are coping with any disease would not be presenting with signs of stress, perhaps this perceived overlap in biomarkers such as NO signalling are simply a demonstration of a diseased or stressed brain cell. For the findings to be specific to the disease in question it is arguably more appropriate to compare one disease state to another albeit different disease. Indeed, as studies and knowledge of metabolic perturbations of amyloid and prion disease increase, so too must comparison between findings, if this powerful and precise methodology is to deliver within the field of diagnosis, disease progression and therapeutic targets. Such effort and focus from the metabolomics community would be welcome news, as these are all areas that have made only limited progression over the last half century (Shao and Le, 2019; Paglia *et al.* 2016; Trushina and Mielke, 2014).

1.2.9. Metabolomics in yeast

An incredibly valuable and insightful model organism, *Saccharomyces cerevisiae,* has been extensively used to study of metabolism for many years. Currently the best eukaryotic model available, considerable time and effort has gone into developing and significantly expanding the YMDB (yeast metabolome database) which includes essential information about compounds, metabolic and signalling pathways and spectral data (Ramirez-Gaona *et al.* 2017; Caudy *et al.* 2017). Although much improvement is still needed in order for the database to be as effective or to cover as broad a range as its human based counterpart HMDB (human metabolome database). Its value as a workhorse within metabolomics for biotechnology, fundamental discovery and food production however cannot be overstated (Caudy *et al.* 2017).

1.3 Aims and Objectives

1.3.1. Overview of aims and objectives

Despite knowing for many years, the link between amyloid formation and disease, the specific mechanisms underlying these events remain a mystery. Although it is understood that the journey to the amyloid state is a complex, multi-layered phenomenon involving many oligomeric intermediates, it is important to uncover the specific mechanism by which aggregation occurs and thus reveal the way in which amyloid disease induces pathogenicity (Knowles *et al.* 2014; Valastyan and Lindquist, 2014).

Yeast prions share many of the molecular features of amyloid disease, including highly specific cellular responses to the toxicity of overexpression. This has illustrated how promiscuous prion protein interactions can elicit highly specific toxicities within the proteome (Liebman and Chernoff, 2012; Treusch and Lindquist, 2012). Here the use of the well-documented model organism, *S. cerevisiae,* which can recapitulate the biochemistry of amyloidosis, is suggested as it may provide insights that are not currently identifiable within the complex biochemistry of the brain.

Given that information regarding these promiscuous protein interactions has remained elusive and yet is key to the fundamental understanding of amyloidosis and the development of appropriate medical interventions, here a novel approach is proposed. Using metabolomics, the aim was to analyse the effect of the aberrant and promiscuous protein-protein interactions of yeast prions in order better understand, isolate, and influence the effect that these rogue oligomers and amyloid proteins have on metabolic pathways.

1.3.2. Specific aims and objectives

The first aim of this project was to determine which metabolomic method was most suitable for use in *S. cerevisiae.* As discussed previously, little consensus exists regarding which methodology is or should be favoured, even when considering only one step in the multi-step process that is metabolomics experimentation. Many research groups use a variety of methods for cell growth, quenching of the metabolism, metabolite extraction, detection, and computational analysis. This makes it difficult to draw parallels between the studies and their often-contradictory findings (Duportet *et al.* 2011; Canelas *et al.* 2009). Assessing the capability of any given methodology to reflect known metabolic changes should be key in determining its future use for biological interpretation. As such the aim was to use the simple,

predictable, and independent (of metabolomics findings) metabolic influencers of aerobic respiration, anaerobic respiration, and ammonium as sole nitrogen source to assess the suitability of the currently available sample preparation methodologies. These experiments involved exploration of suitable growth and extraction methodologies using Ultra High-Performance Liquid Chromatography Mass Spectrometry (UHPLC-MS), to establish a sample preparation method which accurately revealed the metabolic state of *S. cerevisiae.*

This justifiable and valid metabolomics methodology was then used to investigate the endogenous cellular role of the Rnq1 protein, using a strain of *S. cerevisiae* in which the *RNQ1* gene had been deleted (∆rnq1). Understanding the endogenous role of the Rnq1 protein (the facilitator of prions and one of the only wild prions known) may aid in our understanding of the mechanisms controlling this phenomenon. Halfmann *et al.* (2012) suggest that this aspect of fungal prions overshadows the traditional associations with mammalian amyloidosis; the researchers state that prion-driven phenotypic diversity increasing under stress provides evidence that fungal prions can create protein-based molecular memories.

The previously established method was then used to measure the metabolic profiles of [*RNQ*$^{+}$] and [*rnq*$^{-}$] cells, comparing 'disease' vs 'non-disease' state in *S. cerevisiae*; biomarkers and metabolic pathways associated with the presence of Rnq1 protein protein and/or amyloidosis were identified. An atypical stress response in *S. cerevisiae* was used as a control; the aim was to identify biomarkers or pathways which are specific for the presence of Rnq1 protein protein and/or amyloidosis rather than simply being indicative of a general stress phenotype. Previous stress response studies in *S. cerevisiae* suggest that a near universal metabolic stress response is seen in heat shock, oxidative stress, and high pH (Kang *et al.* 2012). Given that the other stress responses tested appear to give a variety of results and the known biochemical associations between heat/oxidative stress and amyloidosis, the decision to use a mild oxidative stress in our control [*rnq*$^{-}$] was taken. The aim was to detract from atypical stress response influencers, ensuring that our results reflected the 'disease state' in question.

Using the same methodologies, the toxicity of Rnq1 overexpression in a [*RNQ*$^{+}$] background was also explored. Douglas et al. (2008) show that Rnq1 overexpression in [*RNQ*$^{+}$] cells causes approximately 25% culture death four hours after induction. The cellular changes

that occur over this time were monitored, with the aim to better understand why the formation of amyloids is often associated with cell death, without a known direct influence themselves.

1.3.3. Summary

Merging the fields of prion/amyloid biology and metabolomics promises to be an informative and enlightening endeavour. This project was subject specific providing information such as metabolomic methodology use in *S. cerevisiae* and the native role of Rnq1 protein. At the same time, the research shed light on fundamental questions regarding amyloid formation, addressing the current need for more detailed understanding of disease biology. The results revealed cellular perturbations caused by yeast prions, which can then be used to direct future studies and to identify homologous circuits in higher multi-cellular eukaryotic organisms.

Chapter Two - Materials and Methods

2.1. Chemicals and reagents

Appropriate risk assessments including COSHH and GMO where carried out for all protocols involving the use of hazardous chemicals or equipment, and suitable control measures employed. All laboratory work was carried out in an ACDP category two laboratory despite the quantity and nature of the organisms used only requiring level one. Prior to experimentation, any freshly sterilised glassware or media for use was left for 24 hours after preparation to check for contamination.

Table 2.1. Chemicals and reagents.

Materials	Content	Source
Chemicals and reagents	Ethanol, Chloroform, Methanol, Acetonitrile, Formic Acid, Tricine, Guanidine hydrochloride, Tris-HCL, EDTA, Lithium acetate, PEG 3350, glycerol, Potassium acetate, Calcium chloride dihydrate, Manganese (II) chloride tetrahydrate, Magnesium chloride hexahydrate	Sigma Aldrich, Fisher Scientific
Reaction kits	QIAprep Spin Miniprep Kit	QIAGEN
Media and amino acid drop-outs	Kaiser Complete SC mixture (SC); non drop-outs: SC Complete Supplement mixture (CSM); single drop-outs: SC -his, SC-Ura; double drop-outs: SC, -his, -ura	Formedium, Fisher Scientific

Chemicals, reagents, reaction kits and amino acid drop-outs are detailed, including their name and source.

2.2. Growth media

All components for buffers and media were weighed to three decimal places using an analytical balance. Buffers and media components were, where appropriate, dissolved in laboratory grade distilled water (Sartorious) and autoclaved at 121°C with a 15-minute cycle using a bench top autoclave (Astell). The autoclave was used to provide sterile media with which to work, eliminating the risk of contamination. Liquid media recipes are detailed in Table 2.2 and 2.3. Where solid media was required, granulated agar (Fisher Scientific) was added at a final concentration of 2% (w/v) prior to autoclaving. In the event an amino acid drop-out mix was used a single pellet of NaOH was also added to the media prior to autoclaving to aid pH balance and solidification of the agar. Solid media were prepared using standardised aseptic technique, with ≈ 20 mL of molten media in each petri dish.

2.2.1. Yeast media for the culture of *S. cerevisiae*

Table 2.2. Yeast growth media used in this study.

Media	Recipe
YEPD (Yeast extract, peptone, glucose) complete medium	2% (w/v) glucose, 1 % (w/v) yeast extract, 2% (w/v) bactopeptone
¼ YEPD (Yeast extract, peptone, glucose) complete medium	2% (w/v) glucose, 0.25 % (w/v) yeast extract, 2% (w/v) bactopeptone
SC (Synthetic complete) 2% glucose drop-out medium (used in chapter 3)	2 % (w/v) glucose, 0.89% Yeast Nitrogen Base (with amino acids, with ammonium sulphate)
SD (Synthetic defined) 2% glucose medium	2 % (w/v) glucose, 0.67 % Yeast Nitrogen Base (without amino acids, with ammonium sulphate), the appropriate concentration of amino acid supplementation as per the individual strain's requirements
SC (Synthetic complete) 2% glucose drop-out medium (used in chapters 4,5 and 6*)	2 % (w/v) glucose, 0.17 % Yeast Nitrogen Base (without amino acids, without ammonium sulphate), 0.5% ammonium sulphate, the appropriate concentration of yeast synthetic complete supplement mixture or synthetic complete drop-out media supplement
SC (Synthetic complete) 2% galactose drop-out medium	2 % (w/v) galactose, 0.17 % Yeast Nitrogen Base (without amino acids, without ammonium sulphate), 0.5% ammonium sulphate, the appropriate concentration of yeast synthetic complete supplement mixture or synthetic complete drop-out media supplement

Media name and recipe details given. *an alternative recipe for SC was employed for chapters 4, 5 and 6 to ensure that media components remained identical wherever possible

The use of YEPD and ¼ YEPD although employed to grow up strains was not used throughout experimentation due the additional metabolite 'noise' which can be created by more traditional yeast media and SD, and versions of SC were considered preferable.

Guanidine Hydrochloride (GdnHCl) was used as a curing agent in the media of *S. cerevisiae* cells that required a [*prion*$^{-}$] status. Filter-sterilised GdnHCl was added to all the yeast medias when appropriate at a final concentration of 3mM after autoclaving and cooling to ≈ 50°C.

2.2.2. Growth media for the culture of *Escherichia coli*

Table 2.3. *E. coli* growth media in this study.

Media	Recipe
LB (Luria Bertani) Medium	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) sodium chloride
SOB (Super Optimal Broth) medium	2% (w/v) tryptone, 0.5 % (w/v) yeast extract, 10mM sodium chloride, 20mM magnesium sulphate, 2mM potassium chloride

Media name and recipe details given.

Ampicillin was used to select as a selection tool for *E.coli* cells that had been transformed with plasmids containing the AmpR gene. Filter-sterilised ampicillin was added to the Lb medium at a final concentration of 100 µg/ml, from a 100 mg/ml stock, after autoclaving and cooling to ≈ 50°C.

2.3. Strains and plasmids

All *S. cerevisiae* strains (Table 2.4.) were maintained on solid media in 30°C incubator. All strains were kept as glycerol stocks and stored in the -80°C freezer and streaked at least three times before use (Sherman, 2002).

2.3.1. *S. cerevisiae* strains

Table 2.4. *S. cerevisiae* strains.

Strain	Notes	Genotype	References
74D-694	[*pin*$^-$]	*MATa ade1-14(UGA) trp1-289(UAG) ura3-52 his3-Δ200 leu2-3, 112*	Chernoff et al. 1993
74D-694	[*PIN*$^+$]	*MATa ade1-14(UGA) trp1-289(UAG) ura3-52 his3-Δ200 leu2-3, 112*	Chernoff et al. 1993
74D-694	*RNQ1* deleted	*MATa ade1-14(UGA) trp1-289(UAG) ura3-52 his3-Δ200 leu2-3, 112*	G. L. Staniforth

Strain, notes, genotype, and reference given (all strains were kindly supplied by Kent fungal group of the University of Kent).

2.3.2. Starter cultures

Prior to the commencement of each experiment within this thesis, each yeast strain was initially grown as a starter culture. 50 mL of the appropriate media was added to a bunged and autoclaved 250 mL conical flask, inoculated with the strain and grown overnight at 30 °C with shaking at 180 rpm. Next morning, the culture was then diluted to an OD600 of ≈ 0.05 in 50 mL of the appropriate minimal media added to a fresh bunged and autoclaved 250 mL conical flask and allowed to grow at 30°C with shaking at 180rpm until the desired OD600 was obtained.

2.3.3. Filter cultures for cell growth

When the diluted starter culture had reached early exponential phase (OD_{600} of ≈ 0.1), 2 mL of the culture was passed through a sterile nylon membrane filter (0.2 µm pore size, hydrophilic nylon membrane, 47 mm diameter, Sigma) using a Nalgene sterile filter unit (Fisher). This filter was then placed cell-side face up onto a corresponding synthetic media plate, placed in a 30°C incubator and grown to an OD_{600} of ≈ 0.6 (plate repeats allowed for OD_{600} assessment). (Yuan et al. 2008; Brauer et al. 2006; Rabinowitz 2007)

2.3.4. Broth cultures for cell growth

Typically, 50 mL liquid cultures were set up in 250 ml bunged and sterilised conical flasks, the appropriate volume of starter culture (as determined by OD assessment) was then added. Broths were then incubated at 30°C with shaking at 180rpm, until an OD_{600} of ≈ 0.6 equivalent to mid-log phase growth was reached

2.3.5. *E.coli* strain

This *E.coli* strain (Table 2.5) was maintained on solid LB medium (0.5 % (w/v) NaCl, 0.5 % (w/v) yeast extract, 1.0 % (w/v) peptone, 2 % (w/v) agar) in a 37°C incubator. This strain was kept as glycerol stocks (LB media, 40% (v/v) glycerol) and stored in the -80°C freezer. When required for experimentation, the strain was streaked on to solid LB media and grown overnight at 37^0 C before being transferred to liquid LB cultures. Liquid cultures were grown in sterilised 1.5 mL Eppendorf microfuge tubes, 50 mL Falcon centrifuge tubes or 250 mL conical flasks containing the appropriate volume of LB broth (0.5 % (w/v) NaCl, 0.5 % (w/v) yeast extract, 1.0 % (w/v) peptone) at 30°C with shaking at 180rpm or on solid media in 30°C incubator.

TOP10 strain given (kindly supplied by Kent fungal group of the University of Kent) with a genotype of F- **mcr**A Δ(**mrr-hsd**RMS-**mcr**BC) Φ80**lac**ZΔM15 Δ**lac**X74 **rec**A1 **ara**D139 Δ(**ara leu**) 7697 **gal**U **gal**K**rps**L (StrR) **end**A1 **nup**G was used.

2.3.6. Plasmid for the overexpression of Rnq1 protein

Plasmid pYES2, with *GAL1*P, *URA*3, 2μ, Amp^R characteristics and *RNQ1* insert, was kindly supplied by Kent fungal group of the University of Kent. Present with *GAL1*P: promoter of the *GAL1* gene; *URA*3: selective marker of uracil biosynthesis in *S. cerevisiae*.

2.4. Recombinant DNA methods

2.4.1. Preparation of competent *E.coli* cells

A single colony of *E.coli* T10 strain was picked from a SOB plate (2% (w/v) tryptone, 0.5 % (w/v) yeast extract, 10mM sodium chloride, 20mM magnesium sulphate, 2mM potassium chloride, 2% (w/v) agar) and transferred to 5 mL of SOB media in a 25 mL falcon tube and placed at 37°C with shaking at 180 rpm overnight (≈ 12 -16 hours). 600 µl of glycerol was then added to the culture (15 % (v/v) final concentration), and 1.0 mL of these seed stocks were placed into 1.5 ml Eppendorf microfuge tubes and put in -80°C freezer. One of these 1.0 mL seed stock aliquots was then added to 250 mL of SOB media in a 1 litre bunged and autoclaved conical flask. This culture was grown at 30° C with shaking at 180 rpm, until an OD_{600} of 0.3 was reached. Cells were then harvested by centrifugation in a benchtop centrifuge at 3000rpm and 4°C for 10 minutes. The spent-broth supernatant was discarded, and any residual liquid removed using a sterile Pasteur pipette. The cell pellet was then gently resuspended (swirled rather than shaking or pipetting) in 80 mL of ice-cold CCMB80 buffer (10 mM KOAc, 80 mM $CaCl_2.2H_2O$, 20mM $MnCl_2.2H_2O$, 10mM $MgCl_2.6H_2O$, 10 % (v/v) glycerol) and incubated on ice for 20 minutes. Cells were then subject to further centrifugation in a benchtop centrifuge at 3000 rpm and 4° C for 10 minutes and the supernatant discarded as before. Cells where then resuspended in a further 10 mL of ice cold CCMB80. 50 µl of this resuspension was added to 200 µl of SOB and the OD_{600} measured until 1.0 – 1.5 was reached (via further dilution of the resuspension with CCMB80 if required). Once the desired OD_{600} was reached, 100 µl aliquots of cells were pipetted into ice-chilled 1.5 ml Eppendorf microfuge tubes and placed immediately into the -80°C freezer until required.

2.4.2. Transformation of plasmid DNA into competent *E.coli* cells

In separate sterile 1.5 mL microfuge tubes, 5ng of miniprep-grade plasmid DNA of the desired plasmid and 5ng of miniprep-grade plasmid DNA of the control plasmid (usually pUC 19) were added to each respective tube and placed on ice. Three 100 µl aliquots of TOP10 competent *E.coli* were allowed to thaw on ice (≈ 10 minutes). Once thawed, the plasmid DNA was added to the aliquoted cells. Nothing was added to the third aliquot of cells. Each tube containing cells was incubated on ice for 20-30 minutes. Cells were then subject to heat shock by quickly transferring the tubes to a heat block set to 42^0 C for between 30-60 seconds. Cells were then immediately returned to ice for two minutes. After this time, 1.0 mL of SOB medium was added to each tube and incubated at 37°C for 30-60 minutes with gentle shaking. 100 µl of cells were then plated out onto appropriate antibiotic-based selective and non-selective media and incubated inverted at 37°C overnight.

2.4.3. Transformation of plasmid DNA into *S. cerevisiae*

A single yeast colony was used to inoculate 5 mL of YEPD media and grown over night at 30º C with shaking at 180rpm. In the morning, 10 mL of fresh YEPD (pre-warmed at 30º C) was inoculated with 0.5 mL of stationary phase cells from the overnight culture and incubated at 30º C with shaking for ≈ 4.5 hours. When an OD_{600} of 0.5-0.6 was reached, the cells were harvested by transferring to a sterile 50 mL microfuge tube and centrifuged in a benchtop centrifuge at 3000 rpm, room temperature for 3 minutes. The spend broth supernatant was then discarded, and the cell pellet resuspended in 5 mL of LiAC solution (10 % (v/v) TE (10x pH 8.0), 10% (v/v) Lithium acetate (1M)). The resuspended cells were then re-centrifuged under the same conditions to wash them. The supernatant was discarded, and the cell pellet resuspended in 500 µl LiAC solution and transferred to a 1.5mL Eppendorf microfuge tube. Sheared salmon sperm DNA (ssDNA) was prepared by dissolving ssDNA in water at a concentration of 5 mg/ml, passing the solution through a narrow-gauge needle multiple times (≈ 50) and then boiling at 100^0 C for 20-30 minutes before cooling on ice. 100 µl of the carrier ssDNA was then mixed with the competent cell solution, and 60 µl of the cell/ssDNA mix aliquoted into ten 1.5 mL Eppendorf microfuge tubes. 2 µl of miniprep plasmid DNA containing 10ng of DNA and 500 µl of PEG solution (10 % (v/v) TE (10x pH 8.0), 10% (v/v) lithium acetate (1 M) and 80 % (v/v) PEG 3350 solution (50 % w/v), filter sterilised) was then added to each tube. Tubes were incubated at 30º C with shaking at 180 rpm for 45 minutes, followed by heat shock at 42º C with shaking at 180 rpm for 15 minutes. The cell transformation mix was then aliquoted in 100 µl - 200 µl

amounts onto appropriate selective and non-selective solid media and incubated upside down at 30º C for 2-4 days.

2.5. Growth conditions and analysis

Throughout experimentation, every effort was made to ensure that cells were grown, extracted, and analysed together. When this was not possible due to the number of samples, all experimental classes were split across (up to) three weeks, with each week containing no less than two class replicates.

2.5.1. Determination of cellular density in liquid or broth culture

To measure the cell density of a culture rapidity, a UV- spectrophotometer was used to obtain an OD_{600} reading of a sample the culture. Routinely, 1.0 mL of the cell culture was aseptically removed from the culture and transferred to a 1.0 mL plastic cuvette. A separate cuvette was also prepared that contained 1.0 mL of fresh growth medium to act as a 'blank' or reference for the spectrometer. The spectrophotometer was set to a wavelength of 600 nm and the reading used as to estimate the cell density of liquid broths. It was assumed that an OD_{600} = 1.0 is approximately 1.0 x 10^7 cells/mL (Sherman, 2002). If the initial OD_{600} reading obtained was above 1.0, then the sample was diluted before reading to ensure the measurement was within the linear range of the spectrophotometer.

2.5.2. Determination of cellular density on solid or filter culture

This method is used to determine the cell density of cells growing on solid filters using a modified protocol of that stated in section 2.5.1. Filters that have had cells grown on their surface were placed inside 50 mL falcon tubes and 2 mL of fresh liquid media added. The Falcon tubes containing the filters and media were then vortexed for 2 minutes to wash the cells from the filter surface into the media. 1.0 ml of media now containing the cells was pipetted into a 1.0 ml plastic cuvette and the OD_{600} measured against a 1.0 mL fresh media reference. For filter cultures, additional plates were prepared to enable this estimation of OD_{600} and so determine the rate of growth of the cultures within an experiment.

2.5.3. Determination of aerobic growth condition

Yeast cell cultures, whether filter or liquid broth grown, were provided with fresh synthetic complete media and incubated at 30º C in the as detailed in section 2.2. (250 mL foam

bunged conical flasks for broth and unsealed lidded petri dishes in large incubator for filters) in the presence of air.

2.5.4. Determination of anaerobic growth condition

Yeast cell cultures, whether filter or liquid broth grown, were provided with fresh synthetic complete media, and incubated at 30º C. Liquid cultures were grown in closed 50 mL Falcon centrifuge tubes, sealed with parafilm and wrapped in parafilm. Filter grown cells were grown as detailed in section 2.3.2. but with the petri dish lids sealed with parafilm and placed inside an air-tight box. All cultures were incubated at 30º C.

2.5.5. Determination of ammonium as a sole nitrogen source growth condition

Yeast cultures whether as a filter format or liquid broth culture were provided with synthetic defined (SD) media that was appropriate to the genotype of the yeast strain being investigated. The synthetic defined media contained only the amino acids required by the strain used for survival (auxotrophic), all other amino acid would need to be synthesised (autotrophic). Cultures were incubated at 30ºC in the as stated previously (Section 2.3.) (250 mL foam bunged conical flasks for broth and unsealed lidded petri dishes in large incubator for filters) in air.

2.5.6. Determination of mild oxidative stress condition

To introduce a mild oxidative stress to the culture, this was achieved by the addition of H_2O_2 (final concentration 0.2 mM) to the culture medium.

2.5.7. Induction of the *GAL1* promoter in pYES2 plasmid

Cells on nylon membrane filters (see section 2.3.) were allowed to grow to an OD_{600} of 0.5 on SC (synthetic complete) 2% glucose drop-out medium (used in chapters 4,5 and 6*) (section 2.2.1.), at which point nylon membrane filters were transferred onto solid media (in line with filter growth methodology) in which glucose was replaced, in equal proportion (20%), with Galactose (section 2.2.1.). Throughout the experiment and irrespective of media cells were incubated at 30º C.

2.5.8. Determination of timed sample collection

The transfer of the cultures to galactose media was considered to be the initial starting point of induction (T0). Cultures were subsequently incubated for various times from this point as shown in the Table 2.7 below.

Table 2.5. Details of timed groups post galactose induction.

Group name	Experimental description
T1	Samples acquired one hour after being placed on galactose media
T2	Samples acquired two hours after being placed on galactose media
T3	Samples acquired three hours after being placed on galactose media
T4	Samples acquired four hours after being placed on galactose media

Describes and names the group breakdown, with reference to their experimental handling regarding the time of sample collection.

2.5.9. Prion status, plasmids, growth conditions and media used within each experimental chapter

In line with the objectives of this thesis, outlined in section 1.3., a variety of prion status, plasmids, growth conditions and medias were used throughout this research. As a point of reference, Table 2.8. provides an overview of each of the particular conditions and strains used within the following experimental chapters.

Table 2.6. Details the prion status, plasmids, growth condition and media used in each experimental chapter.

Chapter	Prion Status	Growth Condition	Media
3	[*rnq*$^-$]	Aerobic	SC 2% glucose medium
	[*rnq*$^-$]	Anaerobic	SC 2% glucose medium
	[*rnq*$^-$]	Ammonium only	SD 2% glucose medium
4	Δrnq	-	SC 2% glucose medium -his
	[*rnq*$^-$]	-	SC 2% glucose medium with GdnHCl
5	[*rnq*$^-$]	-	SC 2% glucose medium with GdnHCl
	[*rnq*$^-$]	H_2O_2	SC 2% glucose medium with GdnHCl
	[*RNQ*$^+$]	-	SC 2% glucose medium
	[*RNQ*$^+$]	H_2O_2	SC 2% glucose medium
6	[*rnq*$^-$]	-	SC 2% glucose medium with GdnHCl followed by SC 2% galactose medium with GdnHCl
	[*rnq*$^-$] with overexpression of Rnq1	-	SC 2% glucose medium -ura with GdnHCl followed by SC 2% galactose medium -ura with GdnHCl
	[*RNQ*$^+$]	-	SC 2% glucose medium followed by SC 2% galactose medium
	[*RNQ*$^+$] with overexpression of Rnq1	-	SC 2% glucose medium -ura followed by SC 2% galactose medium -ura

Chapter number, prion status, plasmid details, growth condition and media used are shown, further details pertaining to these are available in sections 2.2, 2.3, 2.5. and 2.10.

2.6. Metabolite quenching

2.6.1. Cold methanol quenching protocol

Using a 50 mL falcon tube, 2 mL samples from a liquid batch culture or 1 membrane filter containing cells grown on its surface was plunged into 5 mL of pure methanol at -60° C. The mixture was quickly vortexed and placed back at -60°C (filters removed). Extracellular medium was separated by centrifugation (5000 x g, 5 minutes, -20° C pre-cooled). The supernatants were discarded, and the cell pellets were then subjected to the appropriate extraction methodology (Canelas *et al.* 2008).

2.7. Metabolite extraction method

2.7.1. Boiling ethanol (BE)

Tubes containing 5 mL of 75% (v/v) ethanol, 0.1 M tricine (pH 7) were pre-heated in a water bath at 80° C for 5 minutes. Cell pellets/filters were removed from cold storage at -80^0 C and the boiling ethanol quickly poured over the cell pellet and re-placed into the 80° C water bath for 3 minutes. The solution was then cooled on ice for 3 minutes, followed by centrifugation (5000xg, 5 minutes, -20°C, precooled). Extracts were then stored at -80°C until further use (Tredwell *et al.* 2011; Canelas *et al.* 2009; Gonzalez *et al.* 1999).

2.7.2. Chloroform: Methanol (CM)

Throughout the procedure, the temperature of samples and solutions was maintained as close as possible to -60° C. Each cell pellet/filter was suspended in 2.5 mL of pre-cooled 50 % (v/v) aqueous methanol, after which 2.5 mL of precooled chloroform was added. All samples were then placed inside a -80° C freezer, removed and vigorously shaken for every five minutes, the temperature was monitored throughout, never above -50° C or below -65° C). The samples were then centrifuged (5000 x g, 5 min, -20° C, precooled); the upper water/methanol phases were collected separately, and the lower layers were re-extracted with 2.5 mL precooled 50 % (v/v) methanol by vortexing for 30 s. After centrifugation, the upper phases were pooled with the first extracts, and the combined extracts were then stored a -80° C until further use (Canelas *et al.* 2009; Villas-Bôas *et al.* 2005; de Koning and van Dam 1992).

2.7.3. Pure methanol (PM)

Cell pellets/filters were quickly vortexed in the presence of 5 mL of pure methanol (<-50°C), the mixture was then placed at -80° C for 3 minutes. The frozen suspension was then thawed in an ice bath for 15 minutes and centrifuged (5000 x g, 5 min, -20° C, precooled). Extracts were then stored a -80° C until further use (Tredwell *et al.* 2011; Villas-Bôas *et al.* 2005; Prasad Maharjan and Ferenci 2003).

2.7.4. Acetonitrile: water (ACN:water)

Cell pellets/filters were quickly mixed with 5 mL of acetonitrile/water mixture (1:1 v/v) at <-50° C, the mixture was then then placed at -80° C for 3 minutes. The frozen suspension was then thawed in an ice bath for 15 minutes and centrifuged (5000g, 5 min, -20°C, pre-cooled). Extracts were then stored a -80° C until further use (Kim *et al.* 2013; Boer *et al.* 2010).

2.7. Mass Spectrometry

2.7.1. Overview of mass spectrometry methodology employed

Several techniques were employed to enable UHPLC-MS experimentation throughout this thesis (Table 2.9).

Table 2.7. Overview of mass spectrometry techniques employed in this study.

Chapter	No. of Biological classes	n of each class	Total samples	Sample Preparation	Ionisation mode	UHPLC-MS run details	Data analysis
3	24	6	144	Multiple alternative sample preparation was employed as shown in Figure 2.1.	Positive only	Two samples from each class were grown, extracted and analysed each week, over the course of three weeks to eliminate experimental and machine bias	All data were analysed, normalised, and pre-processed collectively in Mzmine2.
4	2	6	12	Cells were grown using filter growth (section X), quenched using cold methanol (section x) and metabolites were extracted using the BE (section x) in line with original Chapter 3 analysis.	Positive and Negative	All samples were grown, extracted, and analysed together in a single run to eliminate experimental and machine bias	All data were analysed, normalised, and pre-processed collectively in Mzmine2, XCMS and MZmine2 and XMCS comparatively.
5	4	6	24		Positive and Negative		All data were analysed, normalised, and pre-processed collectively in MZmine2 and XCMS comparatively in line with chapter 4 analysis.
6	16	6	96		Positive only		

Chapter number, number of biological classes, *n* number of each class, total number of samples, sample preparation, ionisation mode, UHPLC-MS run details and data analysis details are shown. Further information is available throughout section 2.7, 2.8 and 2.10.

2.7.2. UHPLC-MS sample preparation

2.0 mL samples were delicately concentrated using an SPD 111V Thermo Scientific Savant speed vacuum at 35° C for ≈ 3 hours (1.5 hours for chloroform: methanol extraction samples), samples were then resuspended in 500µL of LC/MS grade water and lyophilised overnight. Lyophilised samples were then resuspended in 200µL of 0.1M formic acid and vortexed to ensure they were well mixed prior to 100 µL being loaded into vials for LC/MS.

2.7.3. UHPLC-MS conditions

Each sample (10 µL) was injected into the UHPLC equipped with a 1.7 µm C_{18} BEH column (Waters Corporation, Wilmslow, UK), heated to 35° C and operated with a 10-minute gradient from 0 % to 50 % acetonitrile (0.1% formic acid) at a flow rate of at 500 nL/min. The eluate was analysed as determined by randomisation detailed in section 2.7.4. using an ACQUITY SYNAPT G2-Si Mass Spectrometer (Waters Corporation, Wilmslow, UK). The UHPLC was directly interfaced to a Waters ACQUITY SYNAPT G2-Si Mass Spectrometer operating in electrospray resolution mode. Both positive and negative ion data were collected; positive ion data was collected first followed by negative ion data. A capillary voltage of 2.5 kV and a cone voltage of 40 V were used. Nitrogen, the desolvation gas, was set to 800 L/hr at a temperature of 400 °C. MS^E continuum mode was used for data acquisition over the mass range of 50-1200 m/z, with a scan time of 1 second, which was programmed to step between low (10 eV) and elevated ramp collision energies (20-35 eV) with Argon was used as the collision gas. The data was collected in Data Independent Acquisition (DIA) MS/MS mode. Water blanks were also analysed every 50 samples to check for contamination.

2.7.4. Randomisation and quality controls

Microsoft Excel was used to list samples, random numbers were then generated and associated with each sample. Random numbers and hence samples were the sorted by ascending number, allocating them a numbered position for loading the LC/MS. For quality control were stated, 25 µL of each sample was pooled to create a quality control sample, these were run ≈ every 10 samples.

2.8. Data Analysis

2.8.1 Data acquisition and preparation

The Masslynx (Waters, Version 4.1, Waters, Milford, MA, USA) software collected centroided mass spectra in real time using leucine-enkephalin (556.2771 m/z) as a lock-spray standard injected every 10 seconds for 1 second to calibrate mass accuracy. The raw data (.raw) files were automatic peak detected via Masslynx (Waters, Version 4.1, Waters, Milford, MA, USA) and then converted to .mzML format using Proteowizards MSConvert, at which point high energy scans (function 002) were removed due to lack of subsequent programs capable of analysing them (Chambers *et al.* 2012). Lock Mass correction based on the mass of Leucine enkephalin (556.2771 m/z) was applied within Proteowizards MSConvert.

2.8.2. XCMS independent

XCMS Online was accessed via https://xcmsonline.scripps.edu, (various versions of 3rd edition), mzML. files were uploaded as biological classes and pairwise and multigroup jobs were created using the modified parameters, positive ionisation parameters as detailed in Table 2.10. and negative ionisation parameters as detailed in Tables 2.11. Data files were then downloaded locally (Benton *et al.* 2014; Gowda *et al.* 2014; Zhu *et al.* 2013; Tautenhahn *et al.* 2012a; Tautenhahn *et al.* 2012b).

2.8.3. MzMine2 independent

Various versions of MZmine2 ranging from 2.32 to 2.37, alongside R and Java were loaded onto Canterbury Christ Church Universities Linux server, and accessed remotely. mzML files were uploaded to the server as biological classes, and imported into MZmine2, in most cases this was achieved using all negative or positive ionisation data chapter by chapter, the only variance being for chapter 6, where data was separated by time stamp. All data was then analysed using MZmine2 batch modes with parameters as detailed in Table 2.12. Data was then exported as csv. files and stored locally (Pluskal *et al.* 2010; Ni *et al.* 2016).

2.8.4. XCMS and MzMine2 collaborative

VBA programming was used to compare the exported m/z. values between resultant XCMS-diffreport-multiclass .xlsx files and MZmine2 .csv files. Allowing for variation between the two m/z. values of 0.0001. Overlapping values were then considered for subsequent analysis.

Table 2.8. Positive ionisation data parameter settings used in XCMS.

General	
Polarity	Positive
Retention time format	Minutes
Feature Detection	
Method	centwave
ppm	15
peak width	2 - 30 seconds
S/N threshold	10
mzdiff	0.01
Prefilter peaks	3
Prefilter Intensity	100
Noise Filter	0
EIC width	200
Retention time correction	
Method	obiwarp
profStep	0.5
Annotation and Putative Identification	
Biosource	*S. cerevisiae*
Camera Annotation Isotopes and Adducts	M+H, M+NH4, M+Na, M+K, M+2H, M+2Na
Identification	METLIN
Pathway mapping	Mummichog
Statistics	
Statistical Tests	Unpaired parametric t-test (Welch t-test)
Fold Change	1.5
p-value	0.01

Shows positive ion data analysis parameter settings for general, feature detection, retention time correction, annotation and putative identification and statistics, XP1, as used in XCMS online

Table 2.9. Negative ionisation data parameter settings used in XCMS.

General	
Polarity	Negative
Retention time format	Minutes
Feature Detection	
Method	centwave
Ppm	15
peak width	2 - 30 seconds
S/N threshold	10
Mzdiff	0.01
Prefilter peaks	3
Prefilter Intensity	100
Noise Filter	0
EIC width	200
Retention time correction	
Method	obiwarp
profStep	0.5
Annotation and Putative Identification	
Biosource	*S. cerevisiae*
Camera Annotation Isotopes and Adducts	M-H, M+Cl
Identification	METLIN
Pathway mapping	Mummichog
Statistics	
Statistical Tests	Unpaired parametric t-test (Welch t-test)
Fold Change	1.5
p-value	0.01

Shows negative ion data analysis parameter settings for general, feature detection, retention time correction, annotation and putative identification and statistics, XN1, as used in XCMS online.

Table 2.10. Parameter settings used in MZmine2.

Mass detection	
mass detector	Centroid
noise	0
Chromatogram builder ADAP	
min group size	0.01
group intensity	100
min height	100
ppm	15
Retention time normalization	
ppm	15
retention time tolerance	0.5
min standard intensity	100
Chromatogram deconvolution	
algorithm	Wavelets (ADAP)
s/n	10
min	10
ceo thres	110
peak duration	0-0.5
RT wavelet range	0-0.1
RANSAC Peak Alignment	
mz tolerance	15
RT tol	0.02
RT tol after	0.05
RANSAC iterations	10000
min no of points	30
threshold value	0.03
Filtering – Rows filter	
no. of rows	3
Gap filling - Peak finder	
mz tolerance	15
RT tolerance	0.5

Shows MZmine2 parameter settings for mass detection, chromatogram builder ADAP, retention time normalization, chromatogram deconvolution, RANSAC peak alignment, row filtering and gap filling used for both positive and negative ionisation data

2.9. Multivariate Statistical and Pathway Analysis

2.9.1. SIMCA analysis

After pre-processing and (when relevant) comparative analysis, data was then exported to SIMCA 14.1, a commercial multivariate statistical analysis software owned by Umetrics and commonly used within metabolomics studies. In SIMCA 14.1, the Omics package was employed using relative intensities labelled with their m/z. values, pareto scaling (which improves medium weight features without increasing unrequired noise, stated as the preferred method for use with metabolomics and mass spectrometry) was applied. Selected group comparisons were then processed producing unsupervised PCA models using a 95% confidence interval. PCAs were then observed and OPLS-DA (for pairwise comparisons) or PLS-DA (for groups comparisons) models were, if appropriate, then constructed to identify and highlight the differences between distinct sample groups.

The use of the Omics package allowed for models to be created by way of an autofitting routine, which helpfully avoids overfitting of the data. The models were assessed for quality and reliability via R^2 (representative of the variance within a data set) and Q^2 (representative of the variation within the training set) scores, with a good model considered to have a score of greater than 0.5 and an excellent model considered to have a score of greater than 0.9 for both parameters (ideally both values are relatively similar). The critical p-value used was <0.05 for CV-ANOVA when performed on OPLS-DA models. Whilst continued analysis within SIMCA is possible by way of S-plots and VIP scores, the lack of punitive identifications presented issues and so the decision was made to move to Metaboanalyst.

2.9.2. Metaboanalyst and Mummichogg analysis

Similar models were built via Metaboanalyst 4.0, another commonly used opensource metabolomics analysis program accessed via www.metaboanalyst.ca (Chong *et al.* 2019a; Chong *et al.* 2019b; Chong and Xia, 2018; Chong *et al.* 2018; Xia and Wishart, 2016; Xia *et al.* 2015; Xia *et al.* 2013; Xia *et al.* 2012; Xia *et al.* 2011; Xia and Wishart, 2011a; Xia and Wishart, 2011b; Xia *et al.* 2009; Xia and Wishart, 2010a; Xia and Wishart, 2010b; Xia *et al.* 2009)

Initially within Metaboanalyst 4.0, Statistical analysis was achieved using interquartile range filtering and normalization, using quantile normalization, log transformation and pareto scaling, the results of this normalization were viewed and visually assessed. Some of these models are shown, however the stringency of SIMCA autofitting and the overall modifiable nature and appearance of SIMCA models were the preferred multivariate models. Metaboanalyst 4.0 however provided valuable univariate tests with Bonferroni correction, such as T-tests (for pairwise comparisons), and multivariate tests such as ANOVAs (for group comparisons), available in a program transferable way, and so these T-tests or ANOVAs were then used to move into Metaboanalysts Mummichogg function MS Peaks to Pathways. MS peaks to pathways function now offers multiple versions of the algorithm and the algorithms selected are detailed in the appropriate chapter. Likewise, two *S. cerevisiae* pathway libraries are available, and the details of the pathway library used is detailed within the chapters. Resultant punitive compound and pathways identities were then downloaded as csv. files. In addition, it is possible via Metaboanalyst 4.0 statistical function to perform additional cluster analysis including heatmaps and dendrograms although I have used this function none of these Figures are presented in this thesis, simply due to personal preference.

2.9.3. Pathway analysis

Within Chapter 3 punitive compounds were then overlaid onto KEGG biosynthetic pathways, imaged and interpreted (Kanehisa and Goto, 2000). Within all other chapters BioCyc's omics dashboard was used to compare relative abundances of the most significantly altered metabolites between groups. BioCycs cellular overview was then used, with the same relative abundance data, to create a pathway Table and pathway collages of the most significantly altered metabolic pathways. These pathways were then overlaid with standardized data. This was achieved via Microsoft Excels STANDARDIZE function, using the raw relative abundance of the sample in question, the average mean of all data, the standard deviation of all data. This provided a normalized value known as a z-score or standard score, which represents the number of standard deviations a given data point is from the mean. Positive z-scores indicate a value greater than the mean and negative scores indicate a value less than the mean, this allowed for a visually accessible directionality that would have otherwise been unobtainable, given the large range of values that were present among the

relative abundances. These overlaid pathways were then explored via their connections and surrounding pathways via BioCycs online interactive framework (Karp *et al.* 2017; Caspi *et al.* 2007). Images were exported via png files and Tables were copied into excel.

2.10. Chapter specific experimental workflow

Detailed visualisations of the workflows used throughout subsequent chapters have been produced (Figure 2.1, 2.2, 2.3 and 2.4). To effectively communicate the complex testing of multiple biological classes. As well as the detail the programs used in incremental stages throughout analysis.

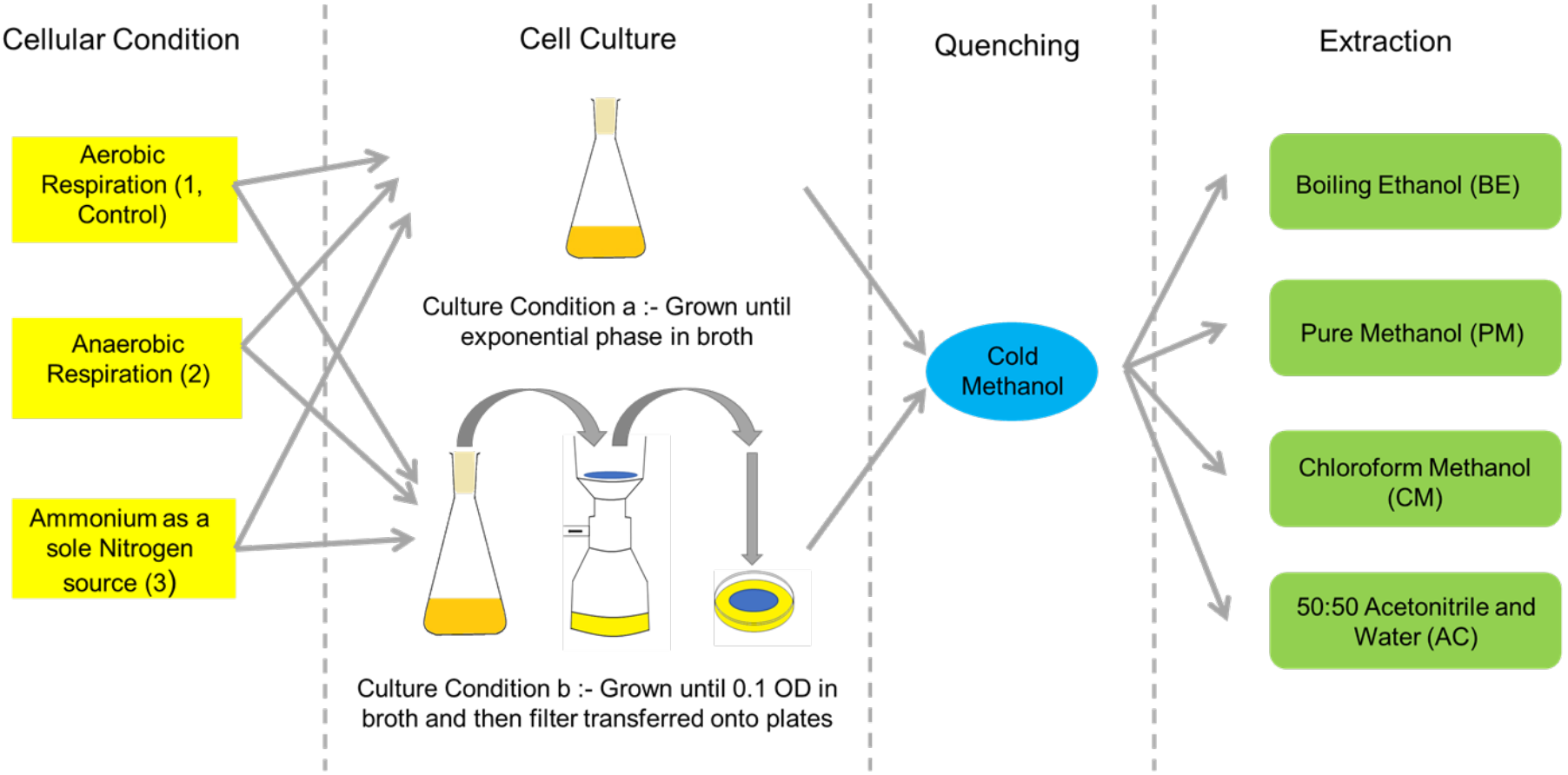


Figure 2.1. Visual representation of experimental workflow used in chapter three. Depicted is how the three cellular conditions, our control aerobic respiration, anaerobic respiration, and ammonium as a sole nitrogen source (shown in yellow) each feed into the two cell culture conditions, broth cell culture or filter cell culture. Experimental flow is indicated by the grey arrows. Both cell cultures then continue show the experimental fate of those groupings, via grey arrows, through to the single quenching method (shown in blue) and onto one of the four extraction methods tested, boiling ethanol, pure methanol, chloroform methanol and 50:50 acetonitrile and water (shown in green).

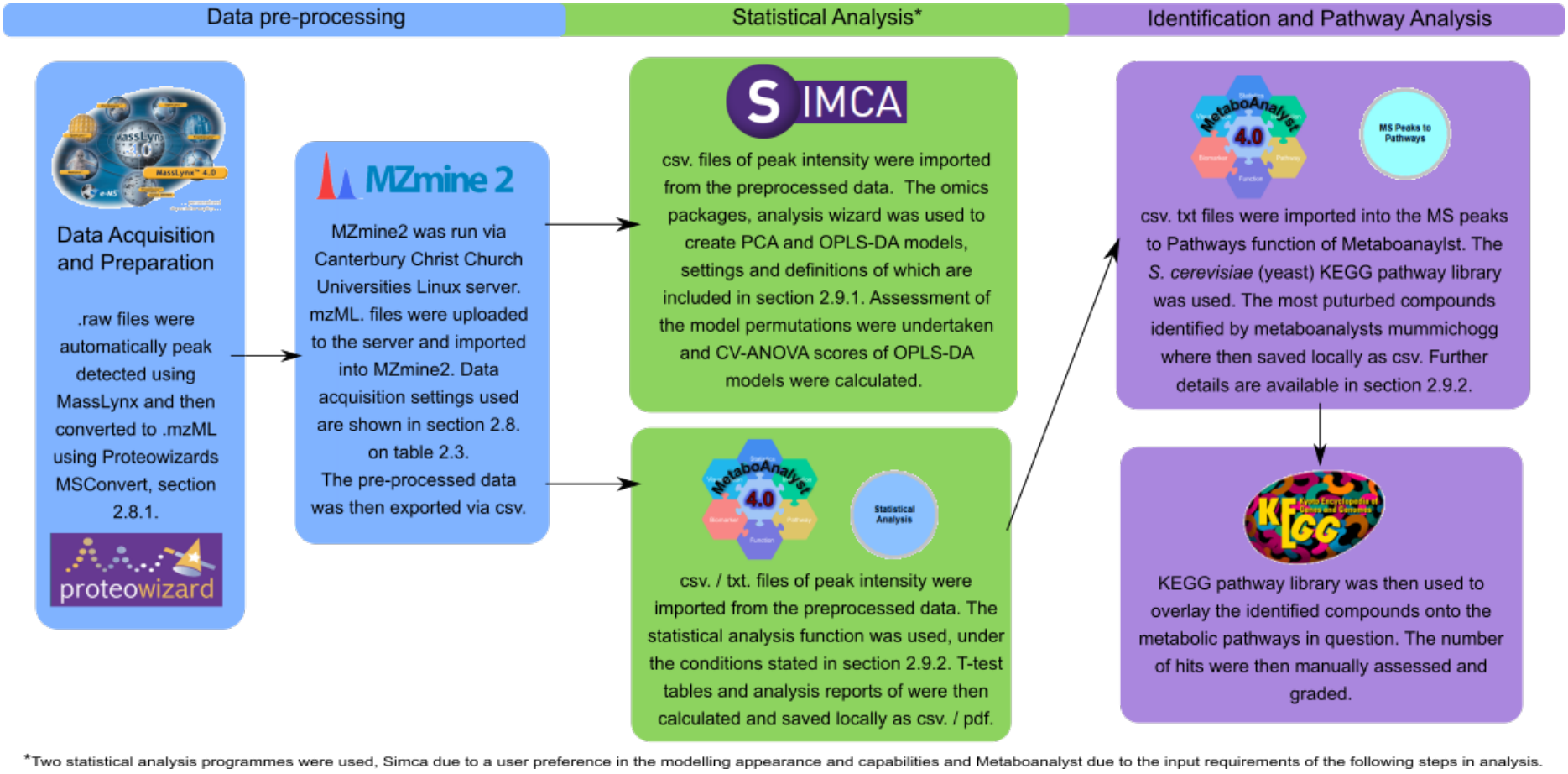


Figure 2.2. Visual representation of the workflow of data analysis in chapter three. From data acquisition within Masslynx, pre-processing in MZmine2 (shown in blue), statistical analysis in SIMCA and Metaboanalyst (shown in green) and Identification and Pathway analysis conducted in Metaboanalyst and KEGG pathway (shown in purple). Black arrows indicate the direction of workflow.

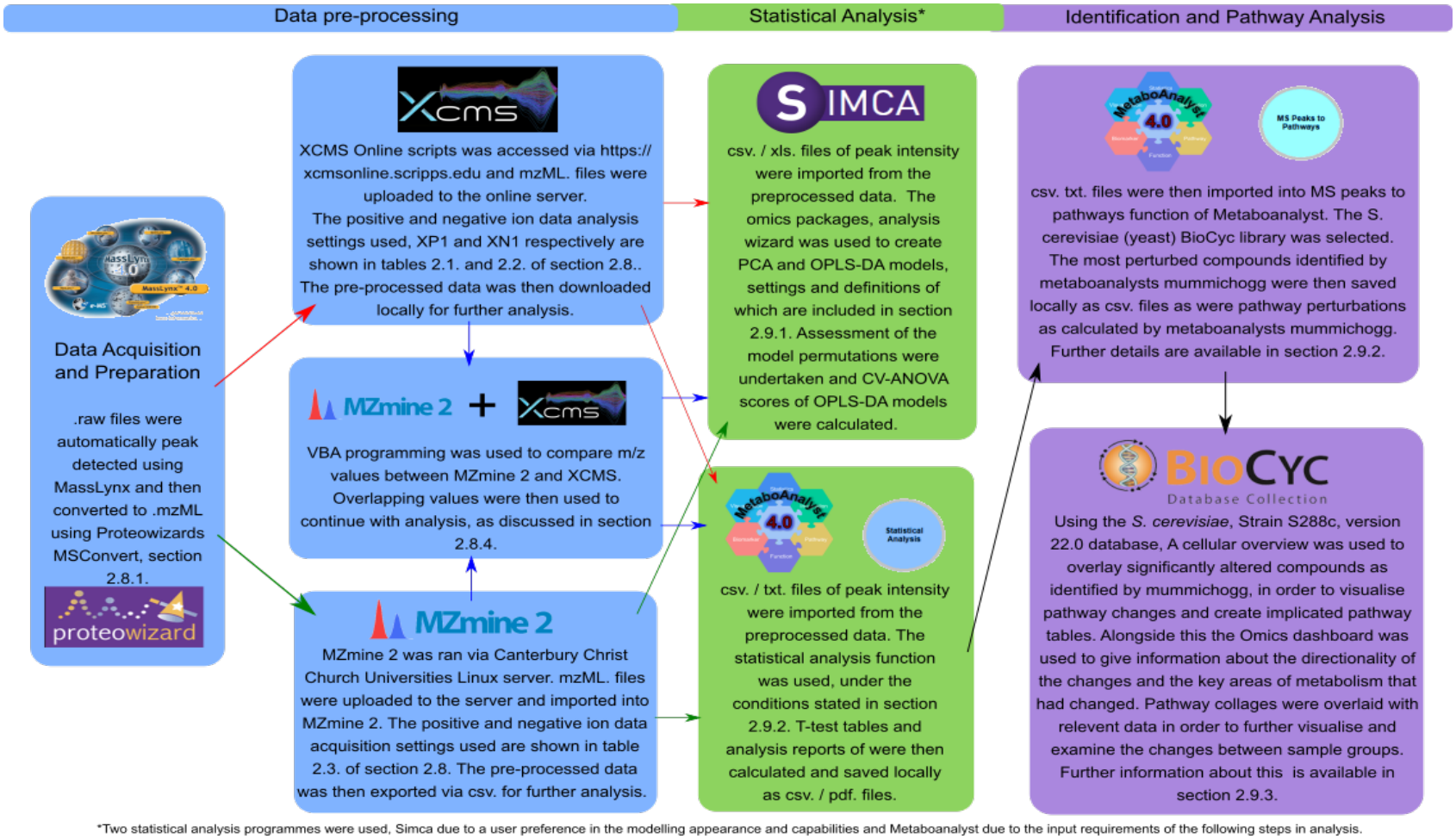


Figure 2.3. Visual representation of the workflow of data analysis in chapter four. From acquisition within Masslynx, pre-processing in MZmine2 and XCMS, and the comparative analysis (shown in blue), statistical analysis in SIMCA and Metaboanalyst (shown in green) and Identification and Pathway analysis conducted in Metaboanalyst and BioCyc (shown in purple). Red arrows depict the events of section 4.3., green arrows the events of 4.4. and blue arrows the events of 4.5., black arrows are used by all workflows.

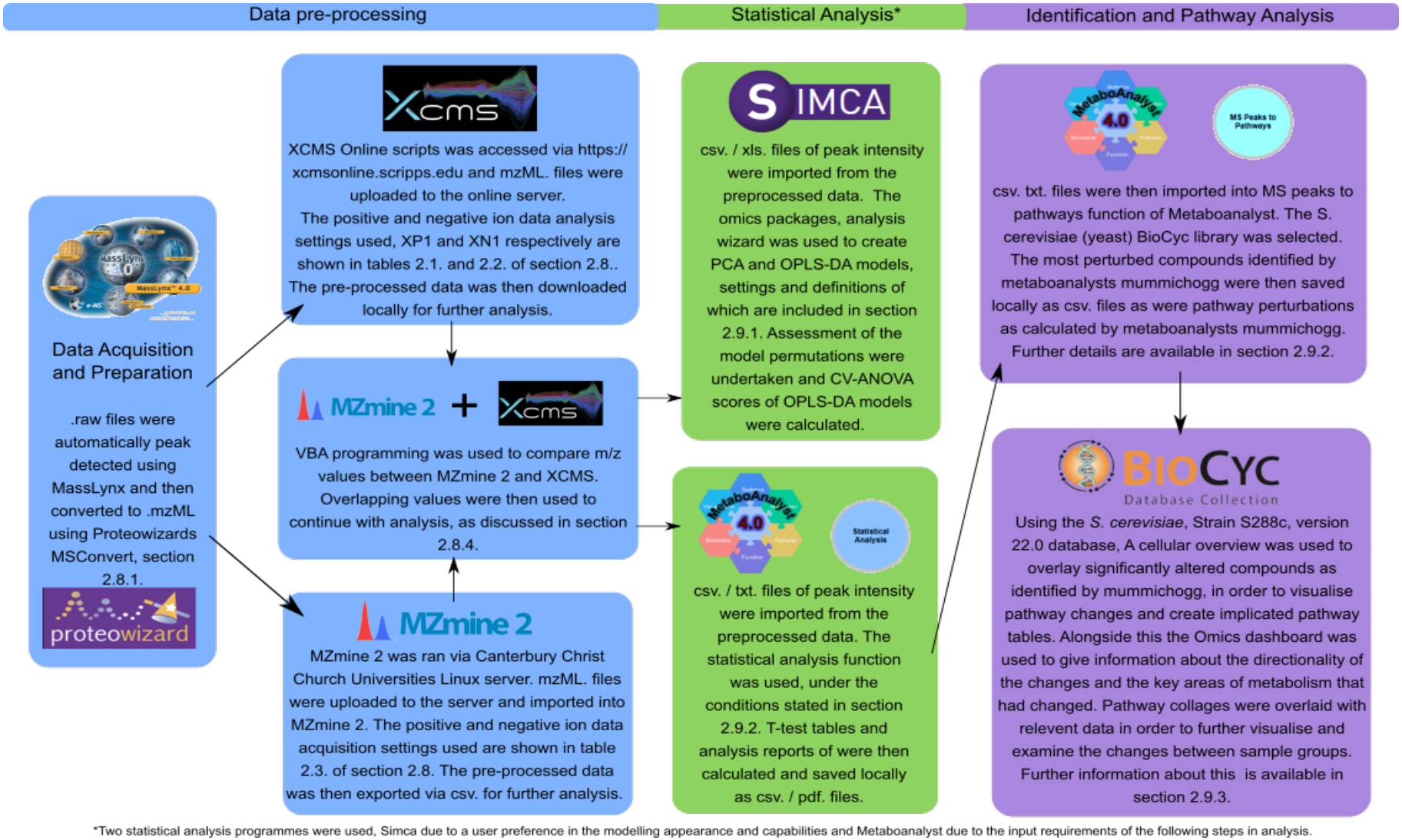


Figure 2.4. Visual representation of the workflow of data analysis in chapters five and six. (indicated via black arrows) within this chapter from acquisition within Masslynx, pre-processing in MZmine2 and XCMS, and the subsequent comparative analysis (shown in blue), statistical analysis in SIMCA and Metaboanalyst (shown in green) and Identification and Pathway analysis conducted in Metaboanalyst and BioCyc (shown in purple). Black arrows indicate workflow.

Chapter Three - Developing a reliable metabolomic methodology in *S. cerevisiae*

This chapter focuses on the testing and development of an appropriate metabolomic method for use in *S.cerevisiae*. Three cellular conditions (aerobic, anaerobic and ammonium as a sole nitrogen source), two types of cell culture (broth and filter), one quenching method and four extraction protocols (boiling ethanol, chloroform methanol, 50:50 acetonitrile and pure methanol) were studied. Detailed methods are described in chapter two.

3.1 Introduction

Given its many benefits, outlined in chapter one, the use of *S. cerevisiae* has led the way in the field of systems biology, especially within the field of metabolomics (Roessner and Bowne, 2018; Sanchez and Neilsen, 2015; Mathew and Padmanaban 2013; Kim *et al.* 2013). Botstein and Fink (2011) state that this model organism appears to be at the forefront of a new frontier, by expanding our understanding of the functions of genes and protein networks and how they partake in the maintenance of metabolism. Promising such insight and biological advancement, it is no surprise that metabolomics is becoming an ever more popular field. Within protein-based research, the prevalence of metabolomics as a method to approach protein-misfolding questions and gain insight into the underlying cellular disturbances has increased in recent times. However, Paglia *et al.* (2018) argue that among Alzheimer's disease studies, despite growing interest in proteomic analysis, insufficient dedication has been given to small molecule 'omics'.

It is known that small molecules (metabolites) play fundamental roles in brain chemistry and despite low-level interest within the field, studies of human disease have found significant perturbations (Paglia *et al.* 2018; Dumas and Davidoc, 2013), including changes in central metabolism occurring in the presence of known protein misfolding diseases, cancers and metabolic conditions (Dumas and Davidoc, 2013; Liesenfeld *et al.* 2013). By using metabolic profiles correlated with next generation sequencing methods researchers have been able to describe links between genotype

and phenotype, and by mapping metabolic findings, identify key biomarkers linked to perturbation and (it is hoped) diagnosis (Riekeberg and Powers, 2017; Sanchez and Neilsen, 2015; Dumas and Davidoc, 2013; Smolinska *et al.* 2012; Patti *et al.* 2012). It is, however, important to consider that metabolomics is a discipline in the early stages of development and thus its full potential has yet to be achieved (Paglia *et al.* 2018; Patti *et al.* 2012).

One of the key and most pertinent limiting factors regarding this technique's potential is the different experimental, laboratory and computational requirements for proper analysis (Riekeberg and Powers, 2017). This, combined with a lack of consensus within the metabolomic community regarding appropriate protocols, has questioned the biological implications of current methodologies and thereby the biological significance of the subsequent findings (Riekeberg and Powers, 2017; Duportet *et al.* 2011; Canelas *et al.* 2009; Castrillo *et al.* 2003).

When considering the overall aims of the research reported within this thesis, it was clear that the desire to monitor the global metabolic changes that occur in response to a variety of proteins (misfolded or otherwise), presented conditions that were suited to untargeted metabolomic investigation. By providing large scale analysis of the 'entire' complement of small molecules produced by a cell, untargeted metabolomic investigation allows for profile comparisons of cellular states via multivariate analysis (Paglia *et al.* 2018; Dumas and Davidoc, 2013; Patti *et al.* 2012). Paglia *et al.* (2018) state that this method provides information regarding profound biochemical alterations without the bias view of targeted metabolomics. However, it does not easily allow for the quantification of metabolites and although work regarding this is ongoing the experimental requirements are often expensive and much better suited to its counterpart, targeted metabolomics (Paglia *et al.* 2018; Dumas and Davidoc, 2013; Patti *et al.* 2012). Given that by way of comparison, untargeted metabolomics seeks to generate hypotheses for variation between groups or perturbations, this appeared to represent our aims well and so it was concluded that any methodologies tested should also use this type of investigation (Patti *et al.* 2012).

Experimental variations within metabolomics investigations can be loosely divided into two groups: the analytical method and sample preparation. The analytical methods

available for use within metabolomics have been outlined (Section 1.2.4. to 1.2.6). Based on this knowledge and our intended untargeted metabolic approach, the decision was taken to use mass spectrometry (MS), largely due to MS's high levels of sensitivity, the variety of ionization approaches, the relatively large online databases streamlined to MS data and the fact that this methodology is the most widely used within the field (Patti *et al.* 2012). By coupling this with complementary UHPLC, it is widely accepted that the largest separation of small molecules within a wide range of polarities can be achieved, thereby improving MS sensitivity and reproducibility and alleviating the effect of background noise and matrix interferences.

Regardless of the analytical methodology used, metabolites need to be released from cells prior to analysis (Faijes *et al.* 2007; Rabinowitz, 2007; Villas-Bôas *et al.* 2005). As a result, great emphasis has been placed on reviewing the effectiveness and reliability of sample preparation methodology (Riekeberg and Powers, 2017).

There are three main components of sample preparation:

- Cell growth
- Quenching of the metabolism (in an effort to halt metabolism and thereby provide a clear snapshot of the sampled time point)
- Extraction of metabolites

A detailed review of sample preparation methods present in the literature appeared to offer no clear and definitive methodology and little consensus exists regarding which methodology should be favoured; even when considering a single model organism, such as *S. cerevisiae* (Liesenfeld *et al.* 2013; Sasidharan *et al.* 2012; Villas-Bôas *et al.* 2005). Different research groups use different methods for cell growth, quenching of the metabolism, metabolite extraction, detection and assessment of extraction methods, making it difficult to draw parallels between the studies and their often-contradictory findings (Liesenfeld *et al.* 2013; Smolinska *et al.* 2012; Duportet *et al.* 2011; Canelas *et al.* 2009; Rabinowitz, 2007; Villas-Bôas *et al.* 2005). Some attribute this variance in results to a methodology's exclusion of particular metabolite classes

(Duportet *et al.* 2011). Some posit that the notion of an extraction method which could extract the full complement of compounds from cells is an impossible task, whilst others suggest that using two very different extraction techniques has the potential to achieve metabolome-wide extraction (Riekeberg and Powers, 2017; Canelas *et al.* 2009; Oldiges and Takors 2005; Goodacre *et al.* 2004; Mashego *et al.* 2003). Clearly, opinion is divided and, at the time of writing, appears to be no closer to reaching a reliable and satisfactory conclusion.

When selecting methods to be tested, the three primary experimental components that have been shown to affect metabolomics results, namely cell growth, quenching and metabolite extraction, were considered independently of each other. Selection of an appropriate quenching methodology appeared on initial inspection to be fraught with a multitude of very similar protocols, each critical of another. Hence, deciding which methodologies to use required an extensive literature review. Among the most popular methods employed when studying single celled organisms, including *S. cerevisiae,* are techniques involving ethanol, methanol, perchloric acid and/or chloroform, at a variety of temperatures (ranging from 4°C to -80°C), including or excluding a cell washing step (Liesenfeld *et al.* 2013; Duportet *et al.* 2011; Sasidharan *et al.* 2012; Canelas *et al.* 2009; Villas-Bôas *et al.* 2005).

Maintaining very low temperatures (<-40°C) is critical to the success of cellular quenching, allowing high efficiency and minimal sample decomposition (Smolinska *et al.* 2012; Sasidharan *et al.* 2012; Tredwell *et al.* 2011; Canelas *et al.* 2010; Winder *et al.* 2008; Faijes *et al.* 2007; Rabinowitz, 2007). The inclusion of washing steps, introduce an undesirable complexity to experimental procedures and can rapidly inhibit the necessity to maintain the low temperatures required for effective cellular quenching to take place (Tredwell *et al.* 2011; Canelas *et al.* 2010; Winder *et al.* 2008; Rabinowitz, 2007). Pinu *et al.* (2017) state that the 60% (v/v) cold methanol-based extraction protocol originally proposed by de Koning and van Dam (1992) is considered the “gold standard” despite controversial evidence suggesting intracellular metabolite leakage into the extracellular medium during its use. Canelas *et al.* (2009 and 2008) offered evidence that their 100 % cold methanol quenching procedure that employed rapid/sub-second sampling, avoided metabolite changes during harvesting. However, Kim *et al.* (2013) state that this method also resulted in large levels of

metabolite leakage from *S. cerevisiae* cells. Upon further inspection, this critique seemed unjust given that Kim *et al.* (2013) failed to follow Canelas *et al.* (2009) strict guidelines regarding temperature (between -20°C and -80°C), instead choosing to conduct their centrifugation and other experimentation at a much higher 4°C. Therefore, cold-methanol quenching was accepted as the quenching methodology to be employed in this study based upon Canelas *et al.* (2009 and 2008) rigorous analysis, the speed of use and its popularity, alongside methanol's miscible nature with water, lack of viscosity (allowing for easy centrifugation) and very low freezing point.

In contrast to the considerable experimental scrutiny of quenching methods, cellular growth conditions have received very little attention with most choosing to use the atypical liquid broth growth considered standard for *S. cerevisiae* (Kang *et al.* 2012; Canelas *et al.* 2009; Faijes *et al.* 2007; Villas-Bôas *et al.* 2005). A novel approach to cell growth, the filter method, first used by Brauer *et al.* (2006), makes use of filter-based culture to allow exponential growth, fast and effective transfer of cells from one media type to another and fast quenching of the metabolism. Rabinowitz (2007) argues that this method of cell growth in both *E. coli* and *S. cerevisiae* allows for a reliable and broad-spectrum overview of analytes, resulting in a net loss in the artefacts and contamination commonly associated with 'atypical' broth growth and subsequent washing steps (Yuan *et al.* 2008; Brauer *et al.* 2006). This method was included alongside traditional liquid media-based growth in an attempt to identify the most suitable cellular growth method protocol for obtaining *S. cerevisiae* metabolites.

When considering metabolite extraction methods, the large number of methods available created considerable choice (Riekeberg and Powers, 2017). To narrow the possible methodologies commonly cited methods in papers using *S. cerevisiae* were examined and their metabolite-based findings compared (Citation numbers between 330 and 477 each (Google Scholar)). This selection process included consideration of the practical restrictions of this research, including access to instrumentation and time scales of experiments. Canelas *et al.* (2009), Villas-Bôas *et al.* (2005) Gonzalez *et al.* (1999) and de Koning and van Dam (1992) were most often cited among this group, each proposing an alternative protocol as the most effective, with each considering the number of metabolites or a fixed recovery rate as their measurement of success (see Table 3.1.).

Table 3.1. Comparison of metabolite extraction methodologies.

Source	Measurement of success	Extraction method
Canelas *et al.* (2009)	C^{13} labelling on half of the samples to determine recovery rate of 19 amino acids, 19 sugars and 8 other metabolites.	Boiling Ethanol and Chloroform/Methanol
de Koning and van Dam (1992)	Spiking recovery rate of up to 13 metabolites	Chloroform/Methanol
Villas-Bôas *et al.* (2005)	Number of peaks and spiking recovery rate of up to 27 metabolites	Pure Methanol
Gonzalez *et al.* (1999)	Spiking recovery rate of 7 metabolites	Boiling Ethanol

The tables display an overview of findings from the selected metabolomics extraction methods employed by each author group, including how each group have determined the success of the methods tests and which extraction method their results indicate is the most effective.

Some of the most often cited methodologies (between 330 and 477 citations each (Google Scholar)), it seemed reasonable to use the methods utilised by these authors in our preliminary investigations to identify the methodology most suitable for use in our subsequent experiments. The extraction methods selected included the boiling ethanol method, the chloroform/methanol and the pure methanol method. However, given that all of these methods measured success of any given protocol in a way that appeared questionable, as outlined above, a method using a less selective technique to rate protocol success was needed so as to not narrow our field of enquiry.

Most method review papers have tested another commonly used extraction protocol, acetonitrile/water (1:1, v/v), and found it to be less effective. However, Kim *et al.* (2013) using peak intensity and hierarchical clustering of 110 identified metabolites found this method to be the most effective at capturing the largest proportion of metabolic change across a cell. Tambellini *et al.* (2013) used similar data interpretation techniques, arguing that this type of analysis of extraction protocol for any given

biological system is fundamentally important for experimental design and the continuation of a given protocols use, hence it was decided that acetonitrile/water (1:1, (v/v)), should be included in the preliminary investigation too.

Given that the purpose of this complex field of study is to gain insight into the biological implications of cellular conditions (both genetic and environmental) assessing the capability of any given methodology to reflect known metabolic changes should be key in determining its future use for biological interpretation. Despite this, protocols are often judged based on the maximum number of certain metabolites detected, with little regard as to the original source of these metabolites and whether they are an artefact or a true interpretation of cellular conditions (Canelas *et al.* 2009; Villas-Bôas *et al.* 2005; Gonzalez *et al.* 1999; de Koning and van Dam,1992).

To address the issues discussed throughout this introduction within *S. cerevisiae,* the aim was to use simple, predictable, and independent (of metabolomics findings) metabolic influencers to assess the suitability of the most widely used sample preparation techniques. During experimental design, priority was given to the simplicity of these metabolic influencers, in terms of the understanding of the metabolic pathways involved and the technical or lab-based requirements of the particular protocol. This led to the selection of three separate test conditions: yeast respiring aerobically with an appropriate supply of amino acids would be treated as a control group and our two test groups would consist of yeast respiring anaerobically and yeast with ammonium as their sole nitrogen source (Table 3.2.).

Table 3.2. Details of growth conditions that would result in clear, known and measurable metabolic changes for use within the preliminary investigation.

Condition one (1)	Condition two (2)	Condition three (3)
Aerobic respiration	**Anaerobic respiration**	**Ammonium as sole nitrogen source**
Grown until the point of quenching in the presence of oxygen on/in a synthetic complete media	Grown until the point of quenching in an anaerobic chamber on/in a synthetic complete media	Grown until the point of quenching in synthetic defined media (with the addition of adenine, tryptophan, histidine, leucine, and uracil, as per this strains requirements)

By causing fundamental changes to growth conditions that produce clear and well-understood changes to metabolism, the resulting changes in the metabolome obtained by each of the methodologies under investigation can be compared.

It was proposed that this approach would allow for an assessment of the respective, respiration or amino acid biosynthetic pathways to be made, providing a biochemical, metabolically relevant assessment of the protocols tested ensuring their validity for future use. For further details of the methodology used within this chapter please see Figure 2.1. and 2.2. in section 2.10.

3.2. Results

3.2.1. MZMine2 results

Data mining in MZmine2, following the parameters outlined in section 2.8., revealed over 22,000 aligned features in positive ionisation mode. It was determined that data analysis and subsequent results would need to be broken down into a few main objectives.

1) Are widescale statistical differences present between the two different culture conditions?
2) Are there widescale statistical differences between the two experimental growth conditions and the control growth condition?
3) If so, between individual groups (i.e. 1aBE and 2aBE) how well do the changes detected reflect the biochemical pathway known to have been disturbed?

3.2.2. Preliminary multivariate modelling in MZmine2

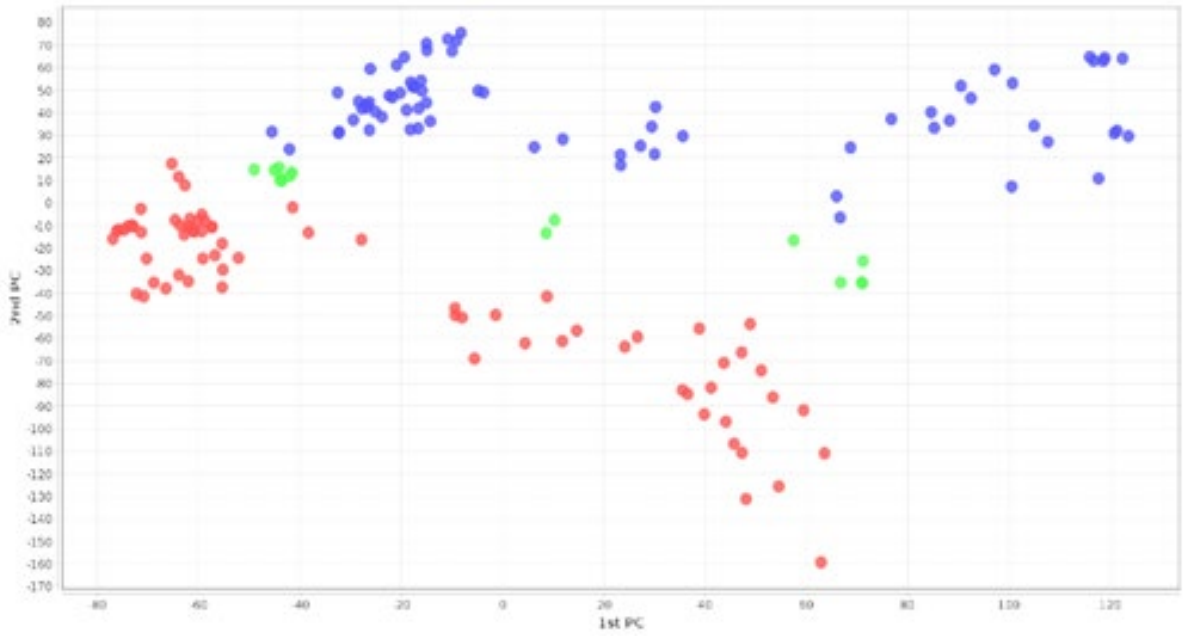


Figure 3.1. Preliminary PCA models of MZmine2 analysis of filter and broth cell culture. Preliminary PCA models via MZmine2 reveals separation between peak lists obtained via filter cell culture (shown in red) and peak lists obtained via broth cell culture (shown in blue). Quality controls (shown in green).

Preliminary MZmine2 PCA analysis revealed strong separation between culture conditions, irrespective of the growth or extraction method used (Figure 3.1.). Quality controls (shown in Figure 3.1.) show three distinct clusters, this is less than desirable and will need to be further examined via SIMCA analysis to establish data quality.

3.2.3. Quality control assessment via SIMCA

Upon further inspection, QCs clusters appear to be separated by the date on which the sample was run. Examining Figure 3.2. reveals that within sample dates little variance and machine drift exists for two of the three experimental runs (Figure 3.2. (B + D)). However, one of the runs shows two clusters as opposed to one within QCs (Figure 3.2. (C)). Laboratory notes elucidate that at about a third of the way through this run, one of the sample buffers ran out and so the experiment was restarted the following day on the remaining samples.

The MS in use was subject to routine maintenance shortly after the first and second run, contributing to the unavoidable variances that may exist within the data. Having examined the PCA and OPLS-DAs that are discussed within this chapter in some detail, these variances factor into the 1st component of analysis. However, confirmatory data analysis that considers each week as an individual data set for each of the factors tested was performed and showed no significant difference between the findings shown below. It was therefore determined that the observed changes or machine influence had had an equal effect on all findings, given that the *n* of 6 was comprised of 2 samples run each week, it was concluded that whilst this may have negatively affected the reported R^2 and Q^2 values of our models, the maintenance of our n of 6 was preferable to splitting up the data by weeks, simply to reach the same conclusions.

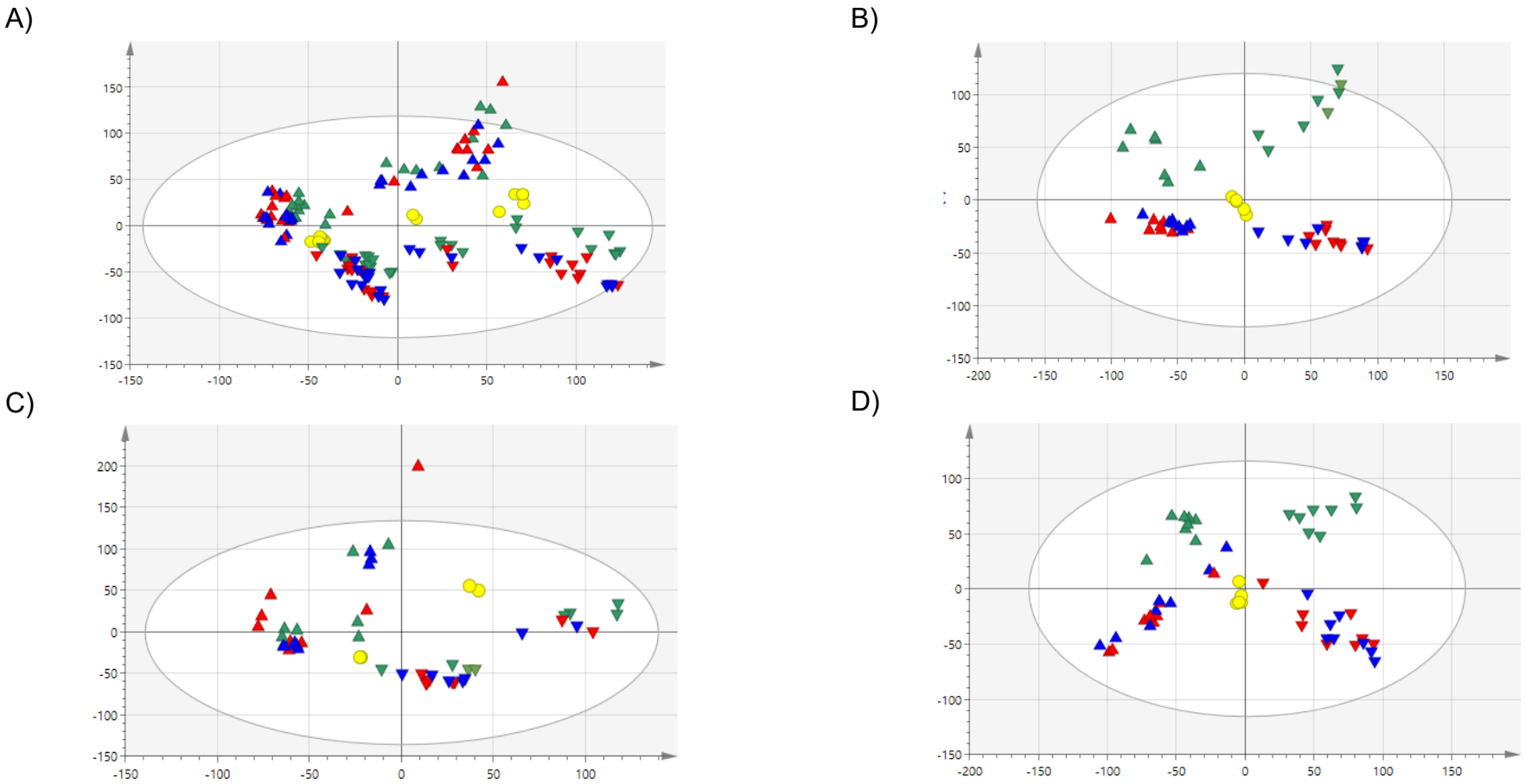


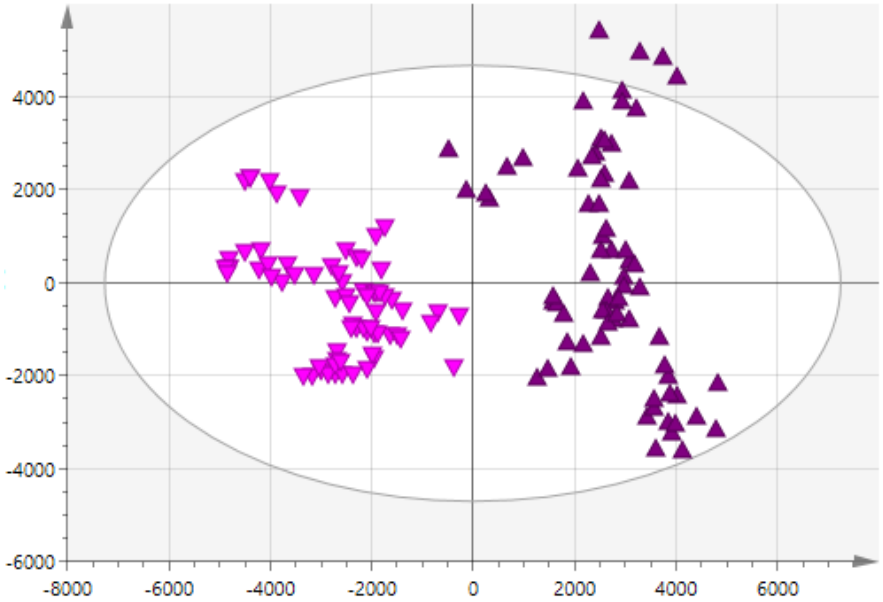
Figure 3.2. PCA models of MZmine2 analysis comparing growth conditions and QC validity. PCAs showing QC relationship between data sets and groupings, yellow circles represent quality controls, red represents growth condition one, blue represents growth condition two, green represents growth condition three, triangles represent samples cultured using the filter cell culture method and inverted triangle show samples cultured using the broth cell culture method (A) All data collected PCA scores plot a model with R^2 = 0.593 Q^2 = 0.397 values. (B) Data collected from MS run one PCA scores plot a model with R^2 = 0.0457 Q^2 = 0.258 values. (C) Data collected from MS run two PCA scores plot a model with R^2 = 0.493 Q^2 = 0.266 values. (D) Data collected from MS run three PCA scores plot a model with R^2 = 0.488 Q^2 = 0.252 value

3.2.4. Multivariate modelling differentiates between culture conditions

To address objective one (*are widescale statistical differences present between the two different culture conditions?),* comparison of the culture conditions of all 144 samples present was made possible by considering samples as belonging to one of two groups: filter cell culture or broth cell culture, irrespective of their extraction method or growth condition. PCA modelling in SIMCA revealed visible separation and clustering between the two culture conditions. The PCA score plot gave relatively good confidence levels of the model with the R^2 and Q^2 values falling within acceptable limits for biological relevance, $R^2 = 0.765$ and $Q^2 = 0.644$. A few scattered outliers were seen in this model, but these were considered acceptable given that the spread of the results may well be indicative of the other potentially influencing factors known to be present here, namely the extraction method and growth condition (Figure 3.3. (A)). An OPLS-DA model was built in a similar vein, including all 144 samples that focused on comparing filter cell culture conditions to broth cell culture conditions and revealed strong significant separation of culture conditions with $R^2(X) = 0.494$, $R^2(Y) = 0.975$, $Q^2 = 0.959$ and a CV-ANOVA p-value = 0.00×10^{-7} (Figure 3.3. (B)).

This information revealed little about which cell culture condition should be favoured, simply that they were in fact different from each other. To establish which one was preferable, more information about the growth conditions, (outlined in objective two) was needed. To achieve this, the metabolic profile of each growth condition was analysed in a similar manner. All 144 samples, irrespective of their extraction method or culture condition, were considered as belonging to either growth condition one (aerobic growth), growth condition two (anaerobic growth) or growth condition three (ammonium as a sole nitrogen source). Data from growth condition one was used as a standard control with which to compare the data from the other two growth conditions.

A)



B)

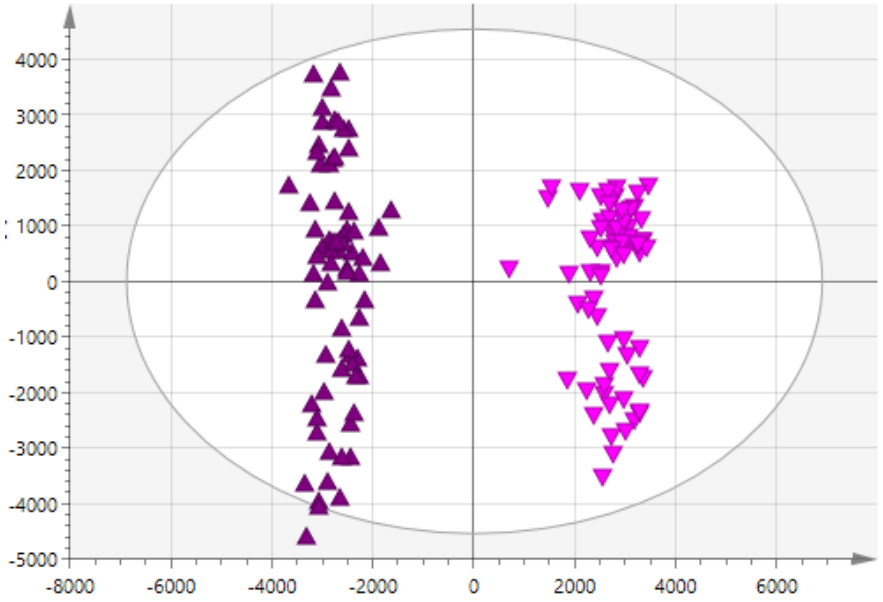


Figure 3.3. PCA (A) and OPLS-DA (B) models of MZmine2 analysis comparing filter and broth cell culture. Overview of the data shows up to 4 outlying samples within the 95% confidence range in each model. Purple triangles represent samples cultured using the filter cell culture method and Pink inverted triangles show samples cultured using the broth cell culture method. (A) PCA scores plot a model with $R^2 = 0.765$ $Q^2 = 0.644$ values. (B) OPLS-DA scores plot a model with $R^2(X) = 0.494$, $R^2(Y) = 0.975$, $Q^2 = 0.959$ and a CV-ANOVA *p-value* = 0.00×10^{-7}.

3.2.5. PCA model reveals no difference between growth condition one and growth condition two

When comparing data from growth condition two (anaerobic respiration) to data from control growth condition one (aerobic respiration), PCA modelling of the 96 samples shown in Figure 3.4. revealed that growth condition two, could not be separated from the control, growth condition one, irrespective of the culture or extraction method used. The PCA model gave values of $R^2 = 0.779$ $Q^2 = 0.635$ indicative of good confidence levels in the model, leading to the conclusion that a lack of distinguishing features between sample groups was responsible for the lack of separation seen. An OPLS-DA model was attempted to see if separation could be achieved, however a reliable model could not be obtained, validating our initial PCA model results. This experiment therefore provided relatively little information about which methods were most appropriate for use with *S. cerevisiae* and suggested that that the anaerobic conditions of growth condition two must not have been maintained successfully throughout the duration of the experiment.

This was not unexpected given the rudimentary way that growth condition two was maintained, which in principle, should have been sufficient to maintain anaerobic respiration, but evidently in practice was not. Some effort was made to measure ethanol content via specific gravity to check the maintenance of this condition but in practise the growth stage and ethanol content would have been too low to detect via this method, even if conditions had been maintained well. It may also have been possible, on reflection, that the differences between aerobic and anaerobic respiration are not biochemically different enough to be distinguishable by such test, given that this was an untargeted experiment and there are only a handful of small molecules that would be changed between the two conditions. Hence, it may be possible that such findings may be lost in within such a large number of small molecule data.

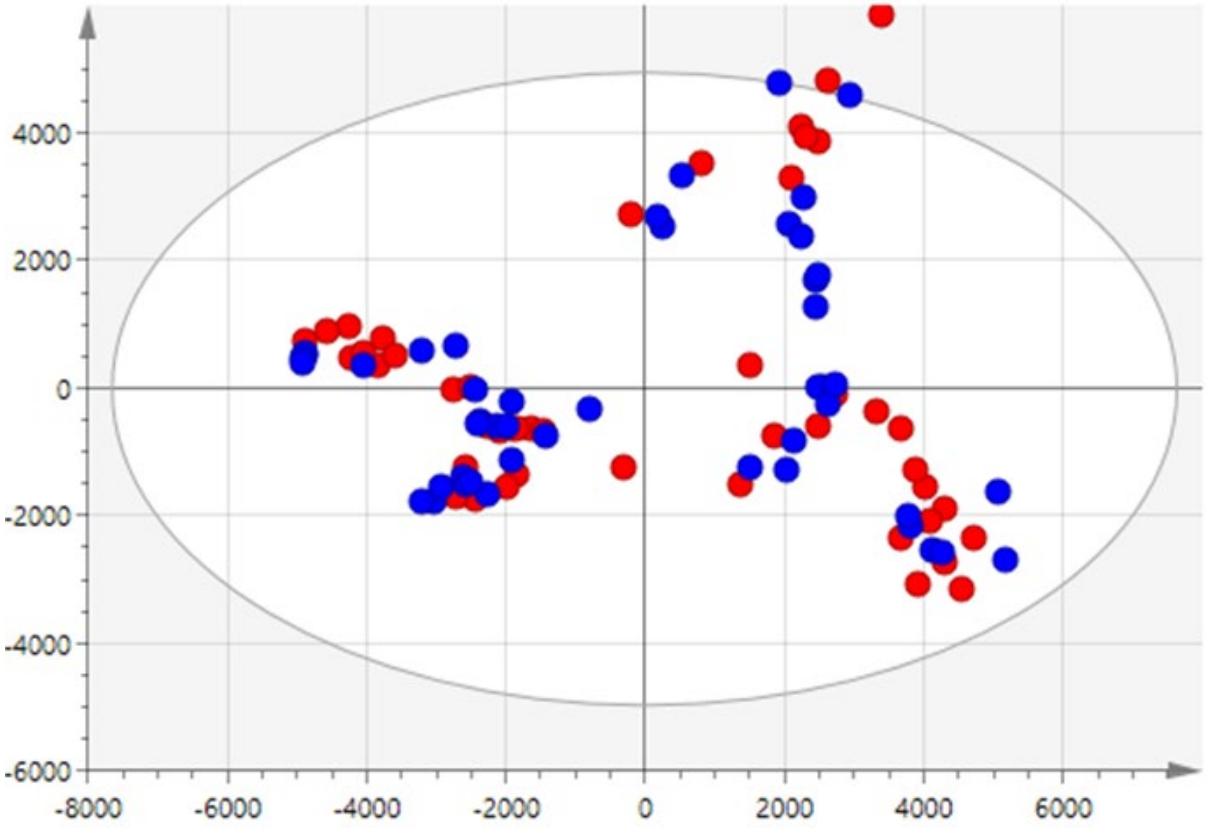


Figure 3.4. PCA model of MZmine2 analysis comparing aerobic and anaerobic growth conditions. Overview of the data shows only a 3 outlying samples within the 95% confidence range. PCA scores plot a model with R^2 = 0.779 Q^2 = 0.635 values. Red circles represent all samples subjected to aerobic growth condition one (control group) and Blue circles represent all samples subjected to anaerobic growth condition two.

3.2.6. Multivariate modelling of shows significant difference between growth condition one and growth condition three

PCA modelling of the 96 samples shown in Figure 3.5. revealed that growth condition three (ammonium as a sole nitrogen source), showed some separation of the data from growth condition one (aerobic respiration) and good confidence levels in the model were achieved R^2 = 0.729 Q^2 = 0.575. Although separation could be observed, it was not distinct; it was considered that this may have been due the variety of underlying variables, such as extraction method and culture conditions. Further OPLS-DA modelling of these two growth conditions were attempted to see if more clearly defined separation could be achieved. This model showed strong, significant separation of growth condition one and growth condition three with $R^2(X)$ = 0.595, $R^2(Y)$ = 0.982, Q^2 = 0.933, evidence of excellent confidence in the model. A CV-

ANOVA p-value = 0.00×10^{-7} showed a strong, significant difference between the groups. Together, this provided evidence for objective two (*Are there widescale statistical differences between the two experimental growth conditions and the control growth condition?*) and allowed us to conclude that experimental growth condition three (ammonium as a sole nitrogen source) results in a widescale statistical difference in metabolic profile between this condition and our control growth condition (aerobic respiration).

This work established that the use of alternative cell culture methods can produce statistically different results. The attempt at growing cultures under anaerobic conditions (growth condition two) was unsuccessful and this was revealed under statistical examination where results could not be separated from the control. This data was no longer considered relevant for further methodological examination. However, data obtained from cultures grown on an alternate nitrogen source (growth condition three) had been shown to be significantly different from our control and subjected to further testing, in the pursuit of answers for objective three (how well do the changes detected reflect the biochemical pathway known to have been disturbed?).

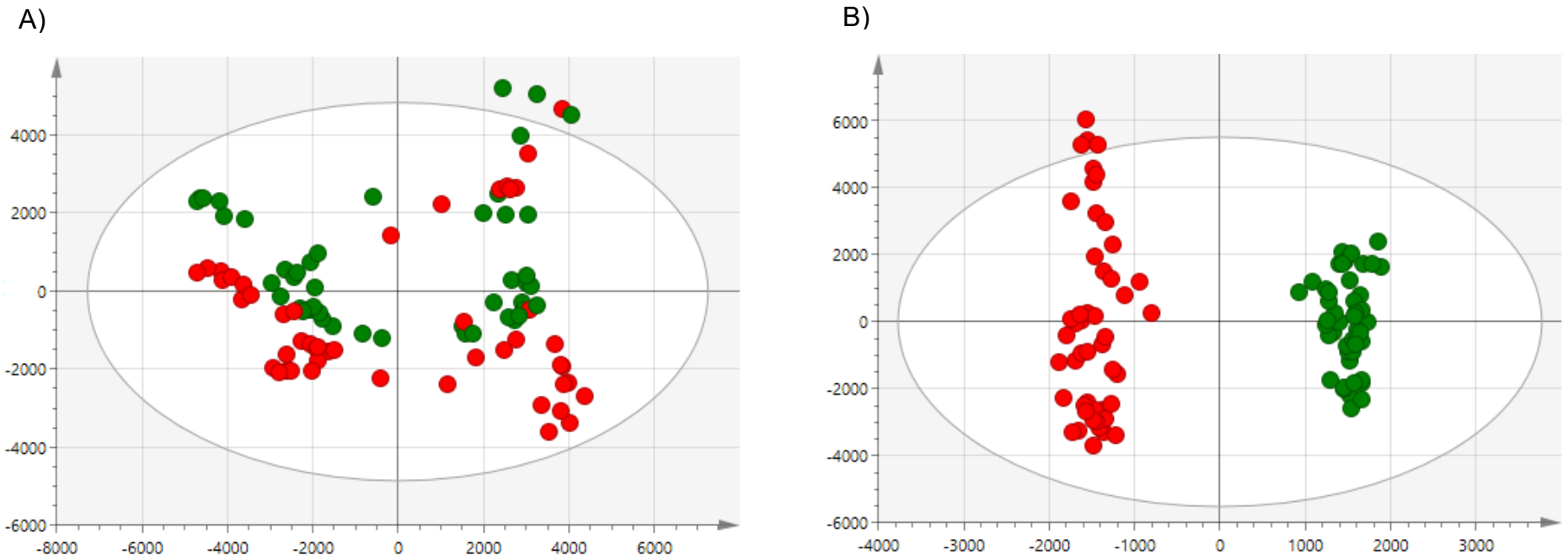


Figure 3.5. PCA (A) and OPLS-DA (B) models of MZmine2 analysis comparing aerobic and ammonium as a sole nitrogen source growth condition. Overview of the data shows only a 4 outlying samples within the 95% confidence range within each model. (A) PCA scores plot a model with $R^2 = 0.729$ $Q^2 = 0.575$ values. Red circles represent all samples subjected to aerobic growth condition one (control group) and green circles represent all samples subjected to ammonium as a sole nitrogen source, growth condition three. (B) OPLS-DA scores plot a model with $R^2(X) = 0.595$, $R^2(Y) = 0.982$, $Q^2 = 0.933$ and a CV-ANOVA *p-value* = 0.00×10^{-7}. Once again, red circles represent all samples subjected to aerobic growth condition one (control group) and green circles represent all samples subjected to ammonium as a sole nitrogen source, growth condition three.

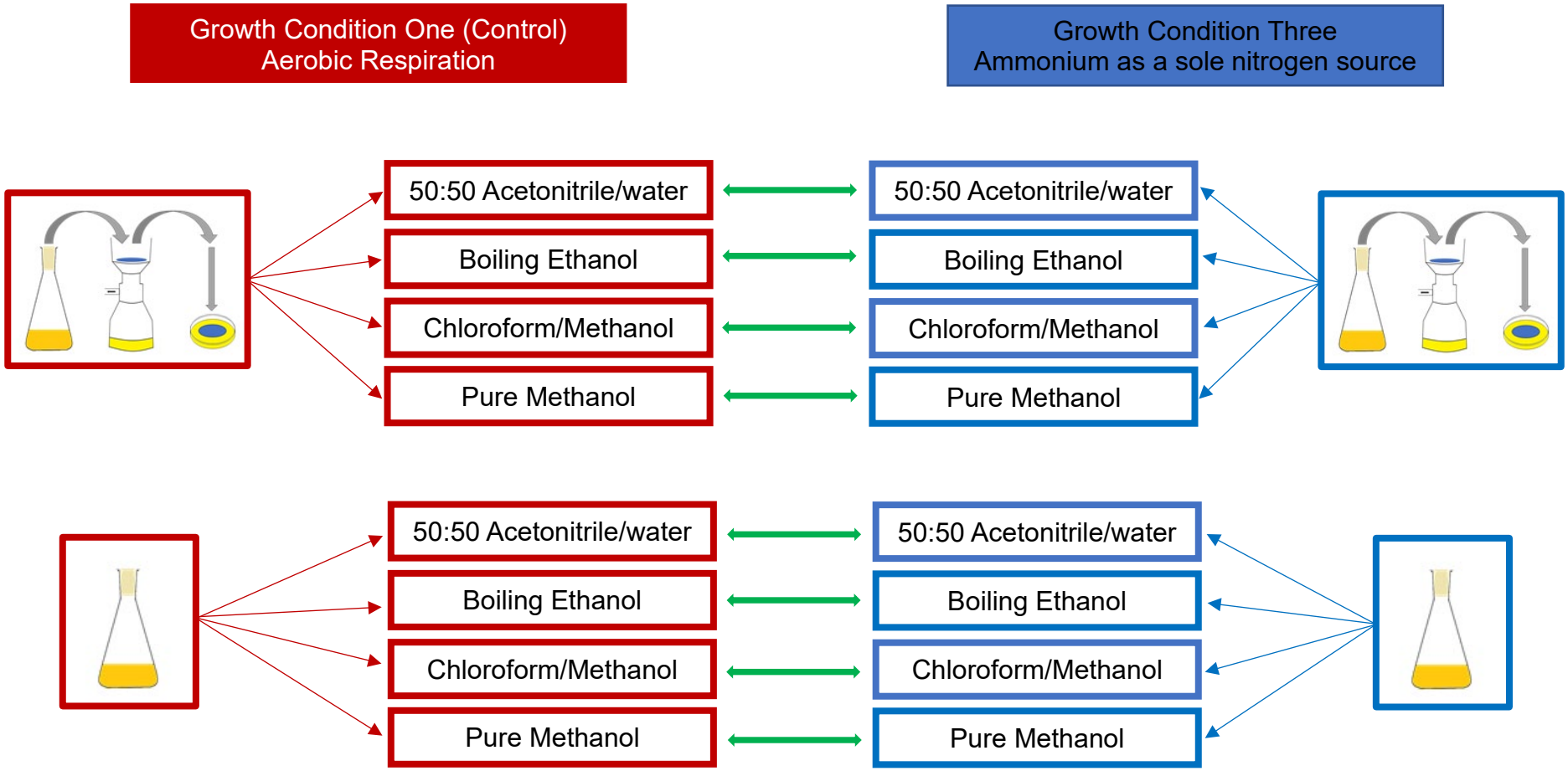


Figure 3.6. Visual representation of the pairwise comparisons made between the culture and extraction conditions of aerobic growth and ammonium as a sole nitrogen source. Depicts the pairwise comparisons made, green arrows show which groups were compared to each other, red boxes denote samples grown using growth condition one, red arrows indicate the flow of experimental methodology from cell culture to extraction method, likewise blue boxes denote samples grown using growth condition three, blue arrows indicate the flow of experimental methodology from cell culture to extraction method.

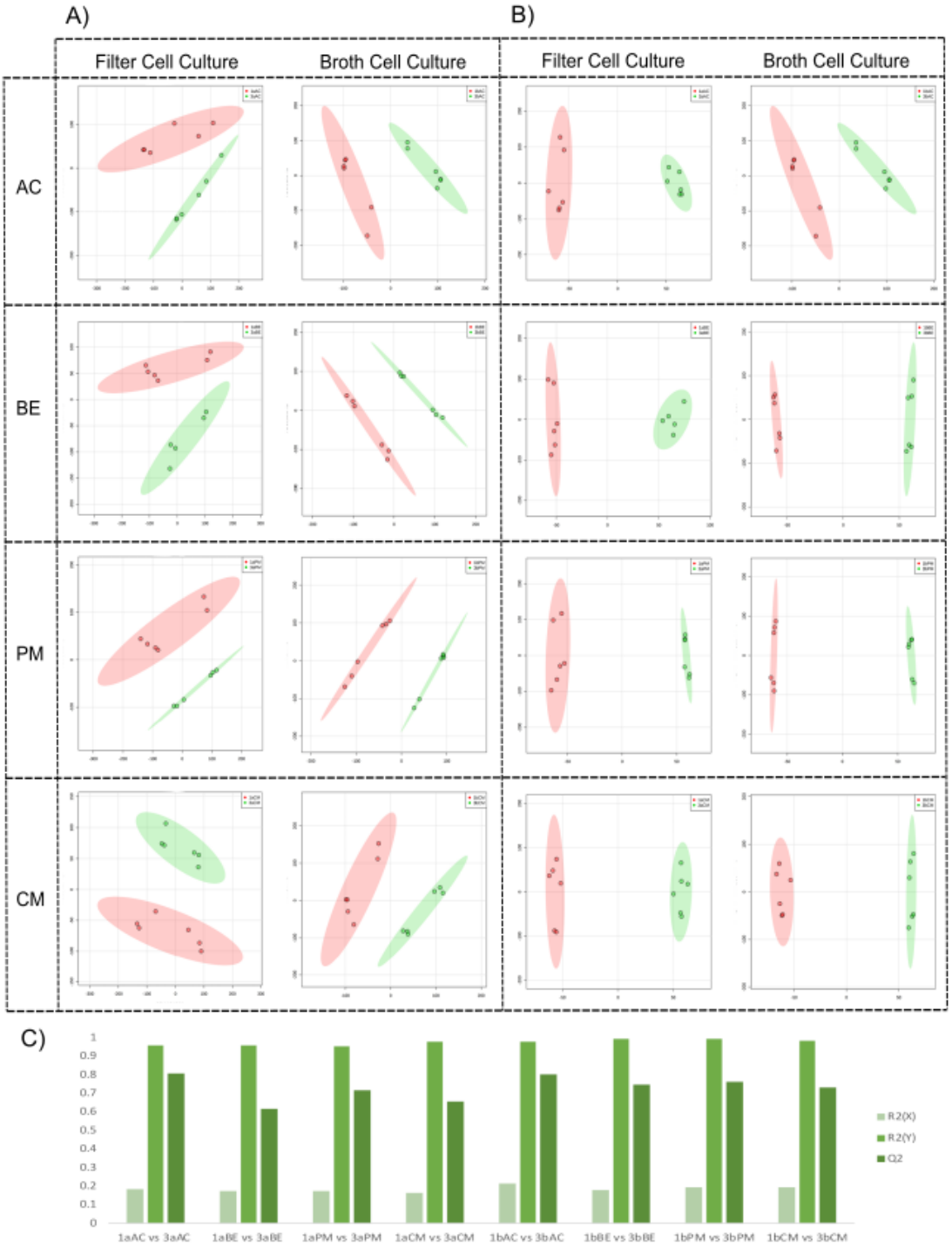


Figure 3.7. PCA and OPLSDA of pairwise comparisons made between the culture and extraction conditions of aerobic growth and ammonium as a sole nitrogen source. Metaboanalyst multivariate statistics of pairwise comparisons between growth condition one (red circles) and ammonium as a sole nitrogen source, growth condition three (green circles), extraction methods are separated by row and indicated on the left of the Figure, cell culture conditions are separated by column and labelled at the top of the Figure. All score plots show 95% confidence regions around each individual group with no overlap (A) All eight PCA scores plots (B) All eight OPLS-DA scores plots, (C) Shows the relative $R^2(X)$, $R^2(Y)$ and Q^2 values for OPLS-DAs.

3.2.7. Metaboanalyst finds significant differences between growth condition one and growth condition three irrespective of extraction or cell culture method

Group data from the 96 samples contained within growth condition one and growth condition three were initially divided into distinct cell culture and extraction method groups (shown in Figure 3.6.). Pairwise comparison was then conducted in Metaboanalyst, using settings described in section 2.9. Resultant PCA's and OPLS-DAs score plots respective $R^2(X)$, $R^2(Y)$ and Q^2 values were then assessed for validity, revealing strong confidence in all group models. Visual assessment of the comparisons showed significant clustering and separation with no overlap (shown in Figure 3.7.). This suggested that all of the cell culture and extraction methods used had permitted the ability to statistically distinguish between the two growth conditions: encouraging given the prevalence of these methodologies with the literature.

3.2.8. Pathway analysis of individual extraction and cell culture methods confirms their ability to detect perturbations

Continued statistical analysis on individual groups in the form of T-tests with the adjusted p-*values* of =>0.01 were then performed. To establish the identity of these increased and decreased features and run pathway analysis, these T-tests were then processed via Metaboanalyst's, Mummichog, *Peaks to Pathways* function, as indicated by Figure 2.1., a *p-value* of $1.0\ x\ 10^{-5}$ allowed for the acquisition of between 300 and 700 results (as required for algorithm functioning). KEGG's Fungi pathway library for *Saccharomyces cerevisiae* (yeast) was selected. This provided details of the most significant pathway hits and most significantly contributing compounds within said pathways. Encouragingly, the top pathway hits (as determined by Metaboanalyst's Mummichogg) were biochemical pathways involved in amino acid biosynthesis, which was expected given the nature of the difference between growth condition three (ammonium and a sole nitrogen source) and the control, growth condition one (aerobic respiration). These findings demonstrate that all methods tested here are capable of reflecting the biochemical pathways perturbed in this experiment.

3.2.9. KEGG pathway analysis of MZmine2 metabolomic data reveals most predicted change reported by culture condition b and extraction method AC

A key requirement of objective three (if so, between individual groups (i.e. 1aBE and 2aBE) how well do the changes detected reflect the biochemical pathway known to have been disturbed?), was to determine which method most accurately reported on the expected metabolic changes caused by a known disturbance. To achieve this, the resultant compound hits were then measured in relation to the amino acid biosynthetic pathways of *S. cerevisiae*, as described by KEGG pathways. Pathways were aligned to suit the genetic background of the strain in question and the most significant compounds, as determined by Metaboanalyst's Mummichog, were then overlaid onto these pathways. The number of metabolite hits within these pathways were then manually assessed and the number of hits declared (shown in Table 3.3.).

Given that the number of pathway hits within the relevant amino acid biosynthetic pathways were highest when cultures were grown using broth cell culture and metabolite extraction was performed using 50:50 acetonitrile/water, it was concluded that these were the methods that best reflected the expected metabolic changes from the perturbation present in condition three (ammonium as a sole nitrogen source) (Figure 3.8.).

Table 3.3. The number of relevant pathway hits for each of the eight identified candidate methodologies.

KEGG Pathway Name	1aAC vs 3aAC	1aBE vs 3aBE	1aPM vs 3aPM	1aCM vs 3aCM	1bAC vs 3bAC	1bBE vs 3bBE	1bPM vs 3bPM	1bCM vs 3bCM
Serine and threonine metabolism	1	1	1	1	1	1	2	1
Cysteine and methionine metabolism	1	1	2	1	4	3	3	3
Branched chain amino acid metabolism	2	2	1	2	2	1	2	1
Lysine metabolism	3	3	1	1	3	3	3	3
Arginine and proline metabolism	3	2	2	2	3	3	3	3
Aromatic amino acid metabolism	4	2	5	4	6	4	5	7
Biosynthesis of amino acids	29	25	25	24	42	32	37	38

Details the number of metabolite hits for each of the eight individual group comparisons, for six named metabolic KEGG pathways within the amino acid biosynthetic pathways, alongside the total number of hits with the super pathway as a whole (inclusive of hits within individual pathways named and those not named). The highest number of metabolite hits within the eight individual groups tested is indicated via the green column.

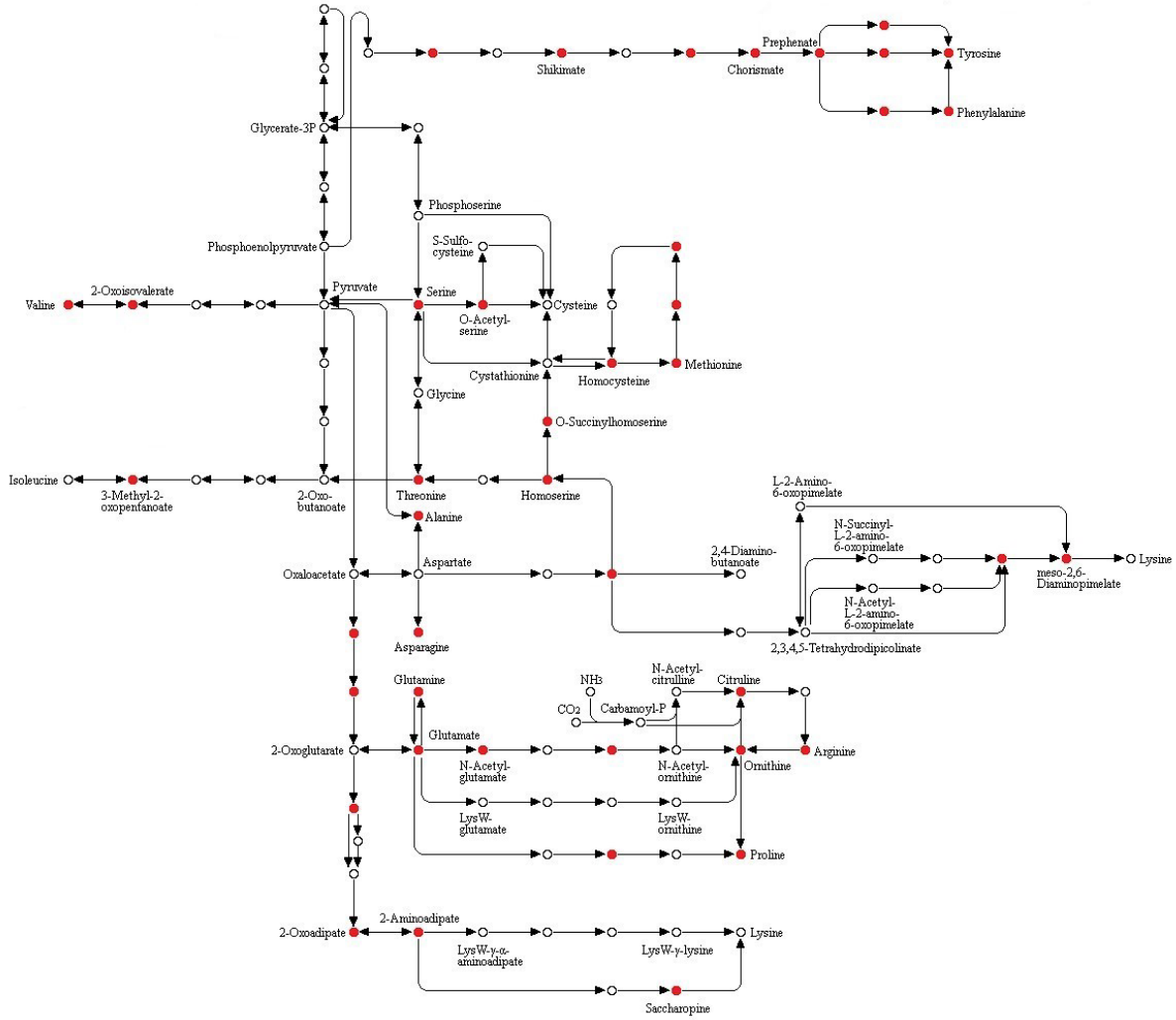
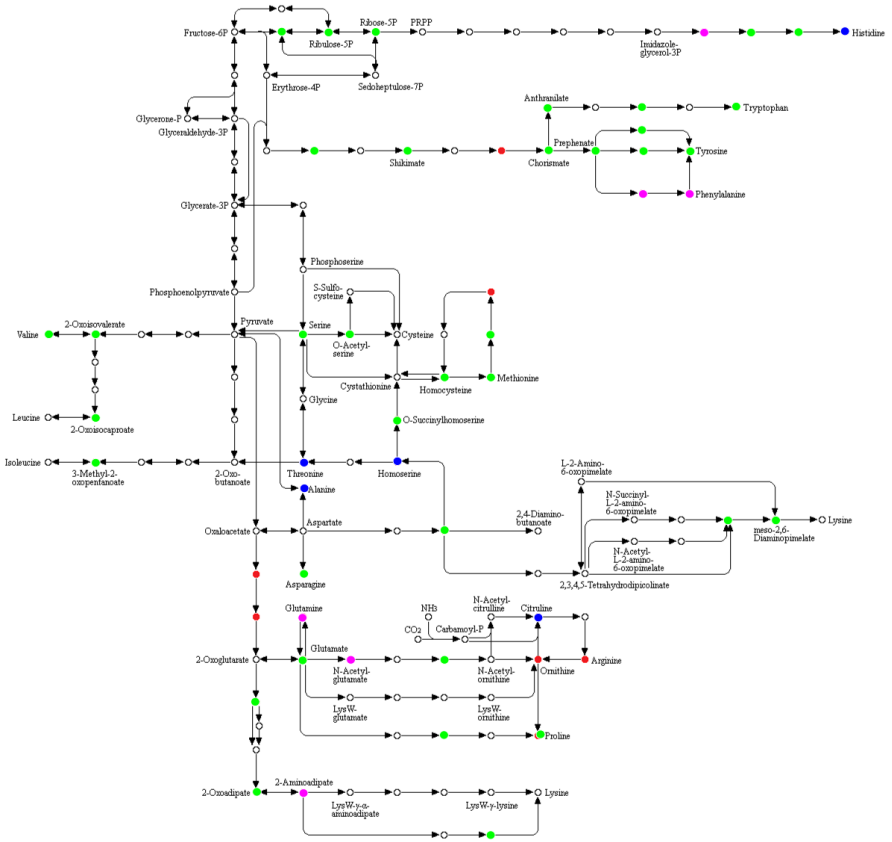


Figure 3.8. KEGG metabolic pathway of amino acid biosynthesis highlighting all of the features detected using broth cell culture and acetonitrile extraction. Red indicates metabolites detected within samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method and acetonitrile extraction protocol. A larger version of this Figure is available in Appendix F.

A)



B)

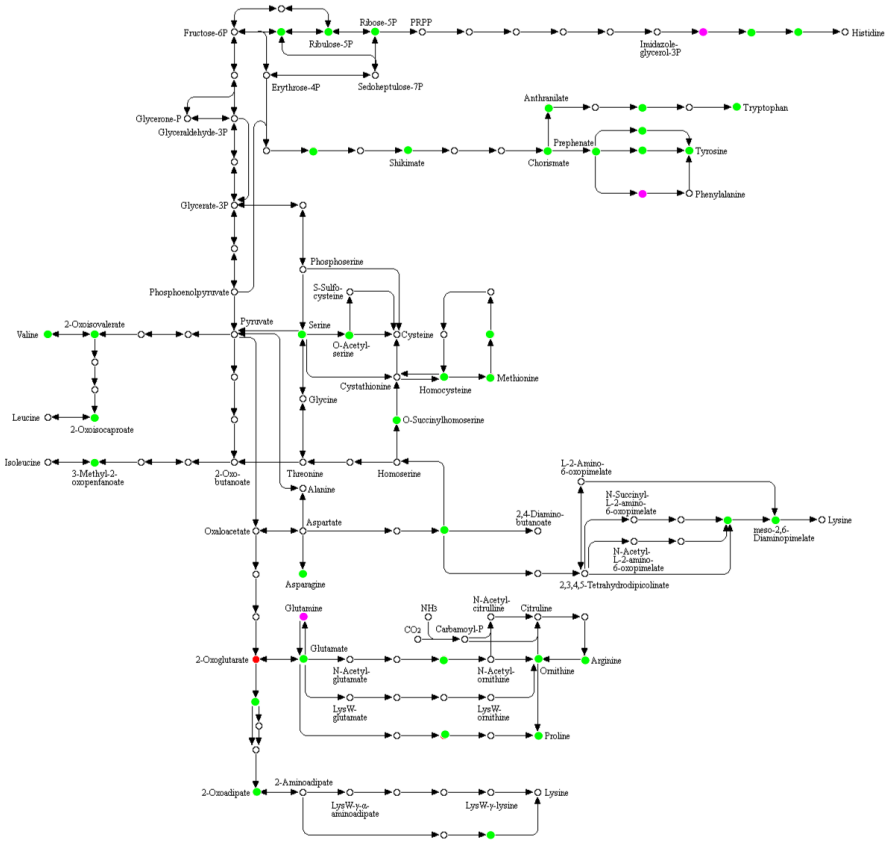
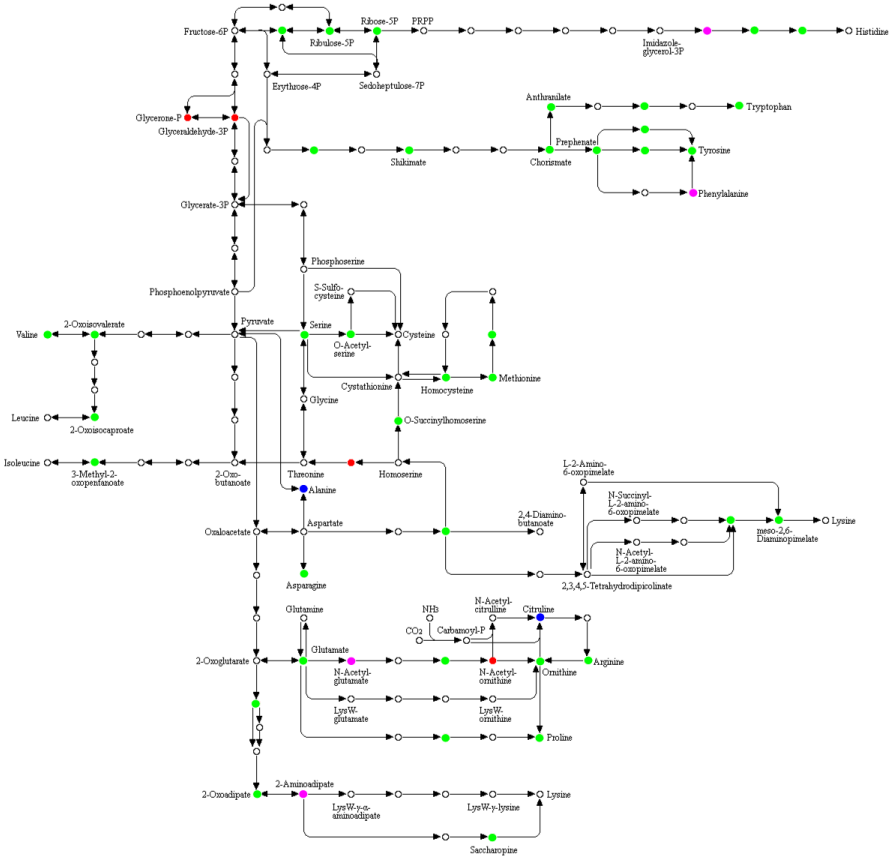


Figure 3.9. KEGG metabolic pathway of amino acid biosynthesis, overlaid with ranked metabolic hits for samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method. A) 50:50 Acetonitrile/water extraction B) Boiling ethanol extraction. Green circles indicate metabolites hit by all pathways, Fuchsia circles indicate metabolites hit by three methods, Blue circles indicate metabolites hit by two methods and Red circles indicate metabolites detected by only one method. A larger version of this Figure is available in Appendix F.

C)



D)

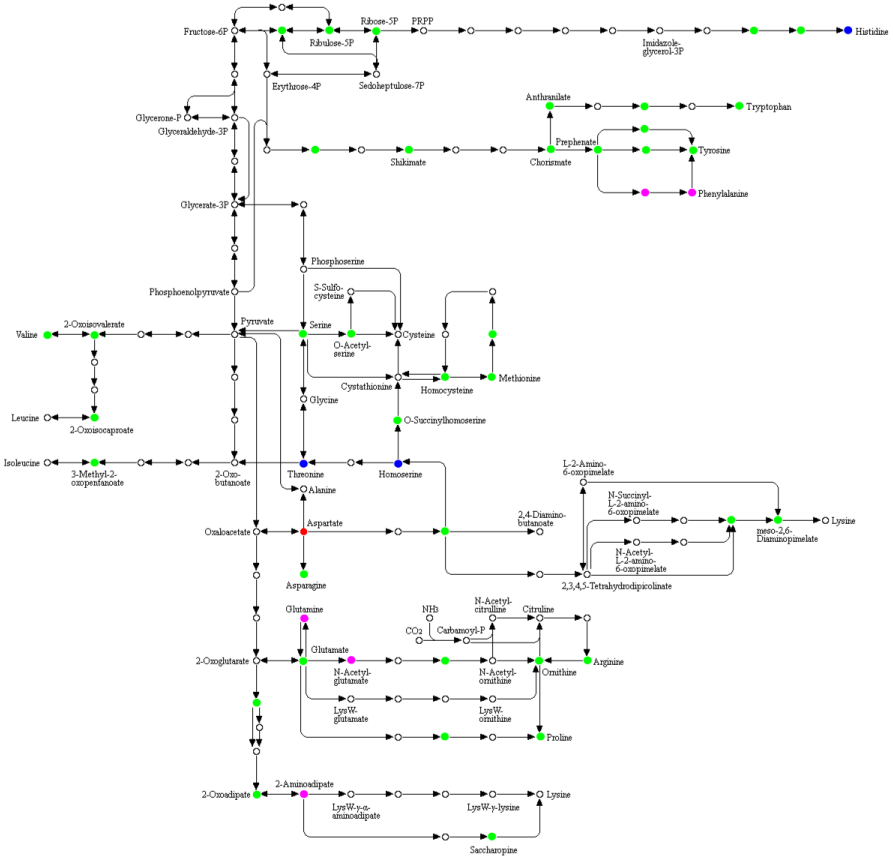


Figure 3.9. (Continued) KEGG metabolic pathway of amino acid biosynthesis, overlaid with ranked metabolic hits for samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method. C) Chloroform/methanol extraction and D) Pure Methanol extraction. Green circles indicate metabolites hit by all pathways, Fuchsia circles indicate metabolites hit by three methods, Blue circles indicate metabolites hit by two methods and Red circles indicate metabolites detected by only one method. A larger version of this Figure is available in Appendix F.

3.3. Discussion

The results of this study demonstrate that the choice of culturing technique as well as the metabolite extraction methodology can influence the quantity of the data observed within a metabolomics study and that this can be reflected in the quality of subsequent pathway analysis. Encouragingly, it is found that all prevalent methodologies do reflect true biological conditions. All methods tested were able to detect the most perturbed pathways, with variance existing only in the completeness or number of metabolite hits with which pathway intermediates were detected. However, as shown in Table 3.3., culturing *S. cerevisiae* in liquid broths rather than on solid media and by extracting metabolites using a 50:50 acetonitrile/water procedure returned the most predicted pathway hits, solidifying confidence in this methodology's ability to implicate the biologically relevant metabolic pathways tested in this study (Figure 3.8.).

3.3.1. Growth condition two: anaerobic respiration

Growth condition two failed to separate from our control leading us to concede that our method of establishing this growth condition was not efficiently maintained. However, the reflection of this in the statistical analysis of the data validates the robustness of the statistical analysis method employed and thus allows for greater confidence in the findings of subsequent studies. As previously mentioned, it could not be ruled out that the biochemical differences between the two conditions may not be sufficient to elicit a measurable response amongst the 'noise' of an untargeted metabolomic experiment. This would be possible to test using more advanced laboratory techniques for maintaining anaerobic growth conditions, such as CO_2 incubation, however such facilities were not available when this experiment was conducted.

3.3.2. Broth and filter cell culture

Results from Table 3.3. report that broth cell culture outperformed filter cell culture, with broth cell culture returning approximately 30% more relevant hits than filter growth. This highlights the importance of considering both cell culture and extraction methodology when undertaking metabolomics experimentation, highlighting the capability of cell culture to influence subsequent metabolic findings. As cells from each

culture format were collected at a time of active growth (OD_{600} = 0.6), had these cells been unable to synthesize the compounds identified on the minimal media provided, they would have failed to survive. These expected compounds, therefore, must have been present and yet remained undetected within filter cell culture.

This suggests that the growth of cells using a filter cell culture method acted as a limiting factor here, with respect to the metabolites detected. Although it may be that cellular trafficking behaviour have influenced results for filter culture. As despite much being known about the uptake of amino acids on alternative media sources, much less is known about cellular responses of amino acid excretion and cellular trafficking on alternative media types (Velasco *et al.* 2004). This highlights a limitation of the study, given that only a single predictable growth condition was achieved, conclusions formed on this basis are indicative as opposed to truly conclusive.

3.3.3. Establishing a methodology for use in *S. cerevisiae*

The premise of using predictable metabolic changes to assess the capability of any given methodology has sound scientific reasoning, although practically these known changes, irrespective of findings within the metabolomics field, are difficult to establish. Growth condition three (ammonium as a sole nitrogen source) allowed for valid assessment in this study. However, it is possible that the method favoured here is one that favours those metabolites involved in amino acid biosynthesis and their intermediates, as opposed to being a general method that represents global metabolome-wide perturbations. Given our criticism of method comparison papers to reach widescale conclusions off of the back of less stringent gradation and assessment of their results, it seems only appropriate to suggest that this, broth cell culture and 50:50 acetonitrile/water method, be adopted when conducting targeted metabolomic experiments aimed at detecting changes within the biosynthetic pathways of the amino acids.

Given the drastically different biochemical conditions present within cellular architecture, multiple methods may be advisable dependent upon which metabolic area is of interest. When examining results here it is evident that maximisation of pathway hits could be achieved by using not one but two extraction methods, as some

researchers currently argue (Riekeberg and Powers, 2017; Canelas *et al.* 2009; Oldiges and Takors 2005; Goodacre *et al.* 2004; Mashego *et al.* 2003). By closely examining Figure 3.9. and focussing on red and blue highlighted metabolites in particular, 50:50 acetonitrile/water extraction method (A) reveals 6 unique (to that method) red hits and 5 shared (by two of the methods) blue hits. The next most successful grouping is that of the chloroform/methanol extraction (C), with 4 unique red hits and 2 shared (by two of the methods) blue hits. The other two extraction methods (B and D) each have 1 unique hit only and 3 shared (by two) between them. This strongly implicates the use of both 50:50 acetonitrile/water and chloroform/methanol extraction for maximum coverage within amino acid biosynthesis.

3.3.4. Limitations and future research

The results within this chapter, whilst encouraging regarding the premise used throughout, will have been restricted by the LC-MS conditions and data analysis. Results were only acquired in positive ionisation mode, using reversed phase chromatography. It would have been preferable to acquire data in both positive and negative ionisation mode, using both reversed phase and HILIC chromatography. However, experiments of this magnitude were not possible for this study. Whilst this will have had an impact on the number and nature of metabolites detected, this effect should have been a constant and present among all samples. Within the context of amino acid detection, reversed phase LC-MS is known to have a high selectivity and detection of amino acids and their intermediates and able to resolve most amino acids (Dahl-Lassen *et al.* 2018; Aviram *et al.* 2016; Jander *et al.* 2004) There can be unwanted signal suppression when using reversed phase due to buffers being added to the eluent, which the use of HILIC can aid in avoiding, allowing for enough sensitivity to quantify metabolites reliably (Kambhampati *et al.* 2019). However, quantification was never an aim of these experiments.

The use of a single data analysis program and the impact that this has on pathway analysis will be explored in chapter four. However, Li *et al.* (2019) and Myers *et al.* (2017) results indicate that the effect of this may be considerable and that using

multiple programmes may be preferable. This is not however currently common practise within the field.

Together, these factors may well have impacted the findings of this chapter. Metabolites that were undetected, would have remained undetected in all samples. Adopting all the suggestions above may indicate another sample methodology as being the most effective at reflecting biological change; however, this will only be the case if the LC-MS conditions and data analysis employed are the same. If research is to be conducted using the same LC-MS conditions and data analysis employed here, then culturing *S. cerevisiae* in liquid broths and extracting metabolites using a 50:50 acetonitrile/water is the most effective methodology.

The most valuable contribution of these experiments is testing for biological interpretation permitting the ranking of the effectiveness of a metabolomics methodology. Future research elaborating on this experimental design and aimed at detecting predictable changes within multiple areas of the cell, would produce clear metabolic protocols for use in targeted metabolomics experimentation. Although this would only be relevant for the organism in question, it would be possible to conduct similar studies in a variety of organisms. In addition, by investigating particular cellular networks present within all cells, a universality that is not currently available may be achieved. Combining and comparing these ‘metabolic area’ results would then provide guidance for those conducting untargeted metabolomics, as well as giving further insight into the ‘blind spots’ of such experimentation.

3.3.5. Data analysis issues and the method use in subsequent chapters

The analysis as presented here was previously not in as presented here. Upon analysing this data and reaching, what appeared to be a satisfactory solution, experimental work begun for the subsequent chapters within this thesis. However, upon collection and comparison of the data for subsequent chapters, it became evident that during the original data acquisition stage of our original metabolomics experiments, the data files had been corrupted in a way that was unclear without comparison. This made all subsequent analysis and hence subsequence conclusions about preferred methodology, incorrect (defined as original chapter three analysis in

chapter two). The data presented within this chapter is due to subsequent re-analysis of the original data files and is, to the best of our knowledge, a correct interpretation of the data. The decision was made to keep this chapter in its current position within the thesis so as to prevent unnecessary confusion. The methods used in subsequent chapters are regrettable, although it is important to note that all of the comparisons (outlined in objective three of this chapter) showed significant differences and identified the most perturbed pathways to be within amino acid biosynthesis. Whilst the experiments in explained in subsequent chapters may have limited the number of hits per pathway, there is no evidence to suggest that the pathways identified are incorrect and, as explained within the introduction to this chapter, all of these methods are routinely used within the field.

Chapter Four - The native role of Rnq1p

This chapter attempts to elucidate the native role of Rnq1p in and provide details of the 'loss of function' effects on metabolism caused by Rnq1p in its prion form. Identification of metabolites perturbed solely by the loss of function of Rnq1p can then be subtracted from subsequent metabolomic studies identifying prion-specific perturbations. These experiments will be achieved by determining the metabolic differences between cells that no longer possess a working copy of the *RNQ1* gene (Δ*rnq1*) cells and cells that possess the *RNQ1* gene but the resulting protein is present in its non-prion form [*rnq*$^{-}$]. In addition, a more thorough evaluation of the data analysis methods employed was investigated to ensure robustness of the results obtained. This involved the use of two separate data analysis programmes, MZmine2 and XCMS, and a comparative analysis using overlapping features of the two programmes. Experiments were conducted as detailed in chapter two.

4.1. Introduction

Rnq1p in its misfolded prion form acts as a facilitator to prion formation. The most well described of these so-called heterologous prion interactions is in its role as a cross seeding mechanism. This is the process by which any one of several novel [*RNQ*$^{+}$] variants allow for *de novo* [*PSI*$^{+}$] or other prion formation. Determining the strength of the [*RNQ*$^{+}$] variant in use is laborious experimentally, although Bardill and True (2010) have offered an alternative to the typical SDD-AGE / Western Blot procedure of Kryndushkin *et al.* (2003), but this is not without issue (discussed in Section 1.1.11). [*RNQ*$^{+}$] variants form *in vivo* at a spontaneous rate approximately five times more frequently than the estimated formation rate of [*PSI*$^{+}$] and differ greatly in their [*PSI*$^{+}$] induction abilities (Stein and True, 2011; Bradley *et al.* 2002; Derkatch *et al.* 2001). Huang *et al.* (2013) suggest that this demonstrates the extensive variability of the structures that Rnq1p can assume and strongly indicates, by lack of correlation between [*RNQ*$^{+}$] number and the role of variants, that Rnq1p has a functional role within the cell independent of cross seeding. Despite investigation in this area, very little is known about the native role of Rnq1p (Liebman and Chernoff, 2012).

RNQ1 deletant strains (Δ*rnq1*, strains that do not contain the entire open reading frame of the *RNQ*1 gene) are viable and have been investigated to determine the loss of function phenotype to little avail. Sondhiemer and Lindquist (2000) show *Δrnq1* strains to have comparable growth rates to wild-type strains on a variety of carbon and nitrogen source medias, with similar sensitivity to a number of metal assays (cadmium, cobalt, copper). They argue that the absence of a clear phenotype associated with a loss of function of Rnq1p is not as surprising as it would first appear. Although the cellular role of the native conformations of many of the other well-described prions are known, this is largely due to the significant amount of experimentation conducted on the role of the native protein itself. Whilst it is true to say that the particular genetic backgrounds that support the expected phenotypic changes of any of the known prions are now common place in many laboratories, it should not be underestimated that these are phenotypic assays that only function under an unusual set of circumstances, often relying on very specific genetic mutations (Liebman and Chernoff, 2012; Sondhiemer and Lindquist, 2000).

For example, the interaction between the regulator of nitrogen catabolism Ure2p and the transcriptional activators Gln3p and Gat1p within cells containing the prion form of the Ure2 protein ([*URE3*]) leads to these transcriptional activators being free to engage in their transcriptional programme, thus allowing poor nitrogen sources to be used by prion carrying cells. A simple growth assay can therefore be used to determine if the [*URE3*] prion is present in cells by their ability to grow on media containing ureidosuccinate (USA) (Cunningham *et al.* 2000; Courchesne and Magasnik, 1988). This method only works in Ure2 mutant cells that cannot ordinarily synthesize USA. Similarly, the atypical colour assay used to determine the presence of the [*PSI*$^{+}$] prion, is only functional in cells that possess particular mutations within the adenine biosynthetic pathway as discussed in section 1.1.10. (Hong *et al.* 2011; Cunningham *et al.* 2000; Courchesne and Magasnik, 1988). The native conformations of the most well-described yeast prions often play key cellular roles in DNA binding, RNA binding and cell signalling functions (Chernova *et al.* 2014; Liebman and Chernoff, 2012; Tuite and Serio, 2010), suggesting an important role for these native proteins within the information flow and architecture of the cell.

A metabolomics analysis could provide a fresh insight into the cellular function of Rnq1p. Representing the ultimate functional read out of the state of a cell and providing a direct link between the mechanics of the cellular architecture/biochemistry and the phenotype of cells, metabolomics has the potential to act as a fundamental tool in uncovering the unknown function of genes and proteins as well as discovering new ones (Patti *et al.* 2012; Prosser *et al.* 2014). When applied to studying proteins this has largely involved enzyme annotation, using isotopic labelling to determine molecule fate and allowing for the assignment of a mechanistic class and function to enzymes which had yet to be defined due to lack of sequence homology with class members (Prosser *et al.* 2014; Borodovsky *et al.* 2002).

Enzymes act as good examples of the investigative power of metabolomics. They are often poor candidates for *in vitro* studies due their inability to function in isolation, whereas *in vivo* metabolomics studies allow for observations of the complex and dynamic interactions key to their functions within the cell (Zhang *et al.* 2011). Aiming to report on the relative metabolite concentrations within a cell at any given time, the sensitive and powerful technique of metabolomics enables the mapping of metabolites within a cell and analysis into any metabolic pathways that have been significantly perturbed by this relative change. This relative or comparable nature of metabolomics alongside the use of delicate analytical machinery and strong statistical modelling software enables both small and large changes to be monitored simultaneously (Misra, 2018; Paglia *et al.* 2018; Patti et al, 2012). Therefore, due to the inherent sensitivity and yet global reporting on subtle cellular changes that metabolomics offers, the aim of these studies was to compare the metabolic profiles of *Δrnq1* and [*rnq*$^{-}$] cells, given that the only difference between these two cell types is the presence or absence of Rnq1p. Key metabolic upregulation and downregulation events that present between the two groups would allow for the identification of significantly altered metabolic pathways and reveal the functional consequences of loss of Rnq1p.

When considering this experimental strategy, the delicate nature of our query was best suited to LC-MS, known to be a cornerstone of metabolomics by virtue of its chromatographic resolution, high sensitivity, unparalleled throughput and good metabolite coverage (Misra, 2018; Myers *et al.* 2017; Li *et al.* 2013). However, given that our LC-MS data was to be acquired without the spiking of expensive internal

standards and the high false discovery rate (FDR) reported for LC-MS studies (Jeong et al. 2012), it seemed reasonable to consider alternate ways with which to streamline our data analysis.

Recent literature within the field suggests that stringent consideration of the data analysis programs used are capable of controlling for these undesirable variables (Li *et al.* 2019; Myers *et al.* 2017). The increasing size and complexity of metabolomics (especially untargeted metabolomics) experimental data sets have created a need for faster, more accurate and reliable computational pre-processing. This has led to the availability of a wide variety of programmes used to construct extracted ion chromatograms (EIC), detect and annotate EIC peaks, align samples, identify and relatively quantify analytes (Myers *et al.* 2017). Many of these bioinformatic processing tools are open-source software such as XCMS, MZmine2, MS-Dial including several python and R packages. In addition, there are also commercial software available such as MarkerLynx, Progenesis etc. With a variety of algorithm processes now underpinning data analysis via any of these methods, concerns have arisen about the capabilities and performance of these programmes relative to each other (Misra, 2018; Li *et al.* 2019; Myers *et al.* 2017).

Myers *et al.* (2017) demonstrate that the variance in the results between one programme to the another can be large. Having identified this challenge, Li *et al.* (2019) in direct response to the criticisms outlined by Myers *et al.* (2017) suggest the use of a multiprogram comparative analysis, with their study indicating that this method removes a large proportion of false positives from any given data set. This would allow for the mining of true features and increasing confidence in biomarker investigation. The most effective method of comparative analysis tested was achieved by cross correlating detected m/z values of the two most widely used programmes, XCMS and MZmine2. To evidence these findings, here both XCMS and MZmine2 independent analysis and an XCMS/MZmine2 comparative analysis will be performed. Combined with our current biological understanding of Rnq1p this approach will address our FDR and LC-MS concerns and will provide key information about the role of Rnq1p in its native state, as well as key information pertinent to our understanding of prion formation and interaction.

4.2. XCMS results

4.2.1. Feature detection and normalisation

Data mining, following the parameters outlined in section 2.8.4., revealed 4454 aligned features in positive ionisation mode (PIM) with negligible retention time deviation (Figure 4.1. (A)). Total Ion Chromatograms (TIC) revealed very little variance between samples, and automated normalization of the data was conducted (using XCMS) to ensure feature alignment and allow comparison of the two sample groups to take place (Figure.4.1. (C + E)).

Similar treatment of the negative ionisation mode (NIM) data, following the parameters outlined in section 2.8.4., revealed 3228 aligned features with very little retention time deviation with the exception of two of the samples very early on in the run (Figure 4.1. (B)). Fortunately, despite this early variation in retention time, automated normalisation of the TICs, visually assessed, suggest that these features had little bearing on either the original or corrected TIC (Figure 4.1. (D + F)).

4.2.2. Multivariate modelling of XCMS metabolomic data differentiates *Δrnq1* and [*rnq*$^-$] samples

To assess if significant differences could be seen between the two groups, *Δrnq1* and [*rnq*$^-$], PCA modelling in SIMCA was conducted using all 12 samples. This revealed visible separation between the two groups, both for positive and negative ionisation mode with no visible outliers. Upon closer examination, however, the score plots respective R^2 and Q^2 values were less than ideal, but are considered biologically relevant, with values $R^2 = 0.560$ $Q^2 = 0.308$ and $R^2 = 0.729$, $Q^2 = 0.149$ (Figure 4.2. (A + B)). R^2 values were considered to be acceptable but the Q^2 values were the more disconcerting, suggesting that the ability of the model to predict new data may be hampered and that the data set itself may contain too much noise. Loadings were assessed but with no outliers, little could be done to improve these scores.

To address this OPLS-DAs were built, again including all 12 samples, and focused on comparing *Δrnq1* to [*rnq*$^-$]. PIM data produced models showing strong significant separation of *Δrnq1* and [*rnq*$^-$] sample groupings with excellent R^2 and Q^2 values, $R^2(X) = 0.55$, $R^2(Y) = 0.992$, $Q^2 = 0.953$ (Figure 4.2. C). In addition, CV-ANOVA gave a *p*-

value = 9.41 x 10^{-5} giving confidence that within this data existed metabolic differences pertinent to the two groups and thus the function of Rnq1p. NIM data was found to support this, itself too having excellent R^2 and Q^2, $R^2(X) = 0.662$, $R^2(Y) = 0.996$, $Q^2 = 0.951$ and a CV-ANOVA *p-value* = 7.35 x 10^{-4}, indicating excellent model predictability (Figure 4.2. D).

4.2.3. Statistical analysis of XCMS metabolomic data in Metaboanalyst finds significantly altered features

Group XCMS data from the 12 samples were then subjected to pairwise comparison within Metaboanalyst, using the settings described in section 2.9. Resultant T-tests with the adjusted *p-value* of =>0.01 found 2427 positive and 1439 negative features that had been significantly altered (*p-value*=>0.01) when comparing *Δrnq1* samples to [*rnq*$^-$] samples.

4.2.4. Tentative feature ID and pathway analysis of XCMS metabolomic data via Metaboanalyst

To establish the identity of the increased and decreased features present within the respective 2427 and 1439 features and to identify this within particular metabolic pathways, Metaboanalyst's Mummichog *Peaks to Pathways* function was used, as indicated by Figure 4.1. A *p-value* of $1.0\ x\ 10^{-4}$ allowed for the acquisition of 589 PIM results and 458 NIM results (closest possible to the required between 300~700 for algorithm functioning). BioCyc's Fungi pathway library for *Saccharomyces cerevisiae* (yeast) was selected. This provided details of the most significantly altered metabolites, the directionality of their perturbation (whether they were up- or down-regulated) and the most significantly disrupted pathways, including the total number of hits within stated pathways (Tables 4.1., 4.2., 4.3. and 4.4.). It would have been possible given these results to move into more in depth pathway analysis, however given that one of the purposes of this chapter, discussed in the introduction, was to perform comparative analysis between the two programmes and so findings past this point ceased to be accurate information available and hence were excluded.

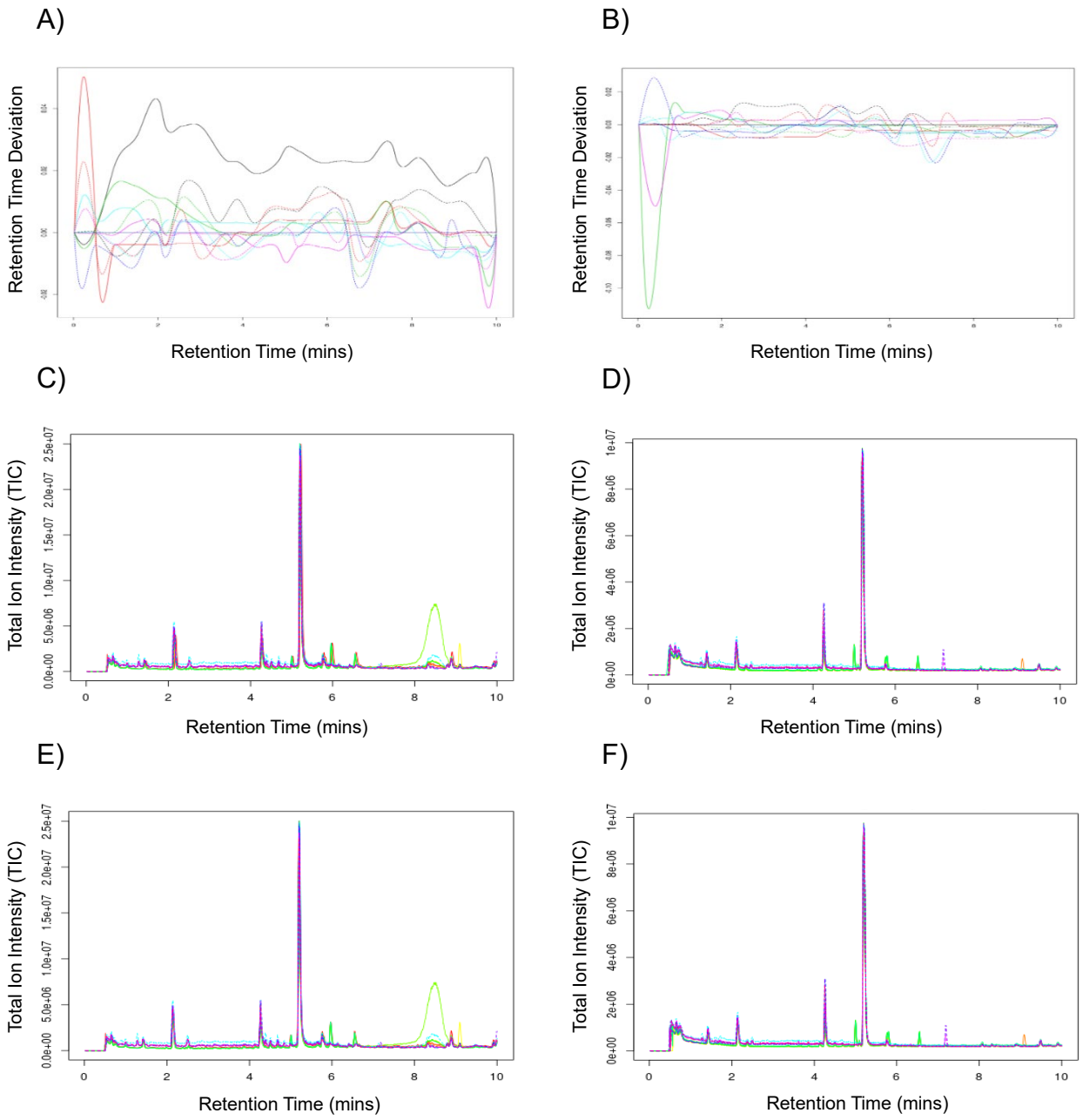


Figure 4.1. Retention time deviation, uncorrected and corrected total ion chromatograms from XCMS of *Δrnq1* and [*rnq*$^-$] samples. XCMS online generated processing information regarding all of the detected LC-MS peaks observed within all 6 *Δrnq1* (samples 03, 16, 21, 23, 27 and 29) and all 6 [*rnq*$^-$] (samples 11, 14, 18, 19, 22 and 28) samples uploaded for analysis and peak detection. *Δrnq1* are shown as solid lines and [*rnq*$^-$] are shown as dashed lines. X axis in all graphs is retention time in minutes A) PIM observed retention time deviation B) NIM observed retention time deviation C) PIM original total ion chromatogram D) NIM original total ion chromatogram E) PIM retention time corrected total ion chromatogram F) NIM retention time corrected total ion chromatogram.

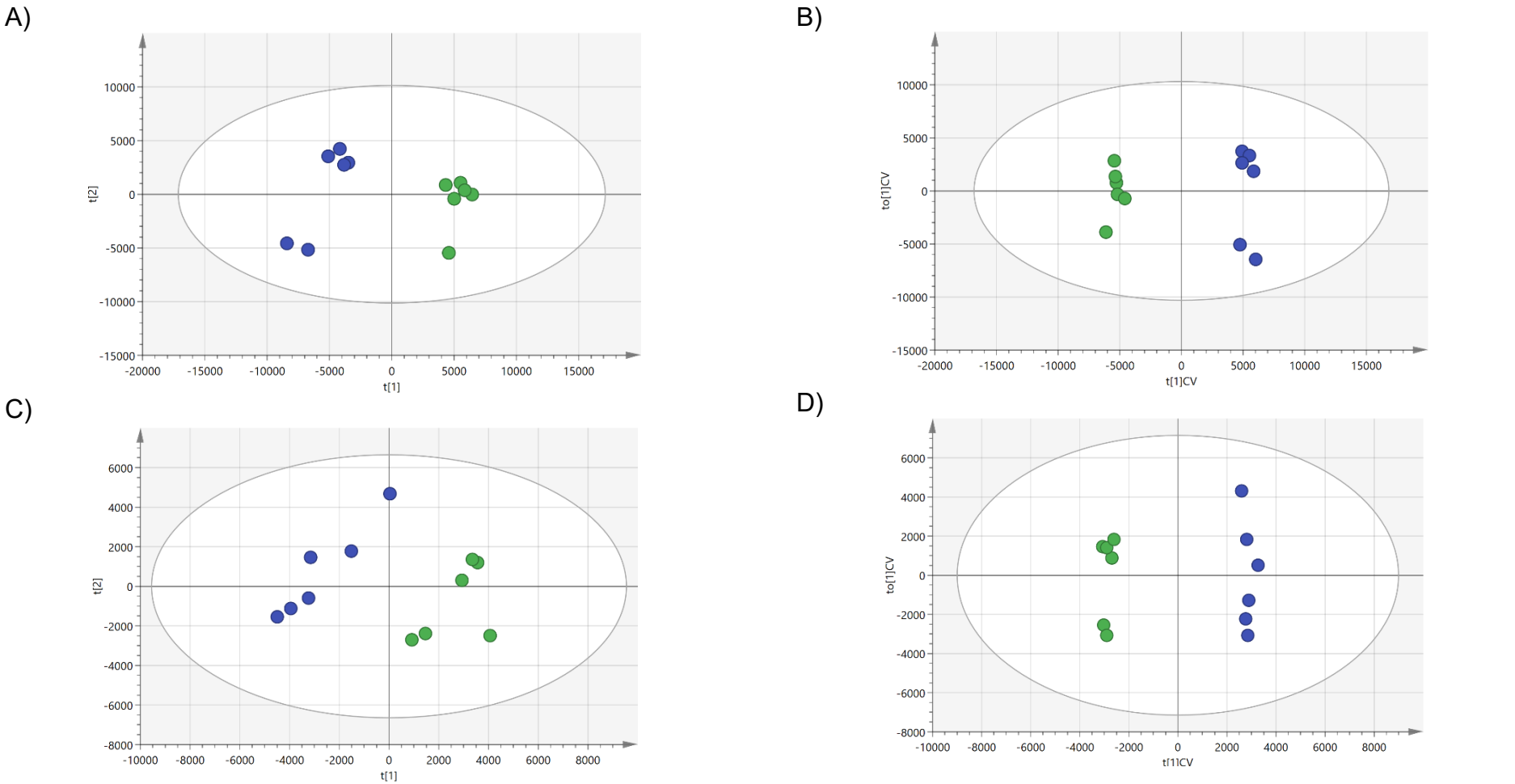


Figure 4.2. PCA (A and C) and OPLS-DA (B and D) models of XCMS comparison of *Δrnq1* and [*rnq*$^-$] samples. Green circles represent *Δrnq1* samples and blue circles represent [*rnq*$^-$] samples. Overview of the data shows no outlying samples within the 95% confidence range within each model. (A) XCMS analysed PIM data PCA scores with $R^2 = 0.560$ $Q^2 = 0.308$ values. (B) XCMS analysed PIM data OPLS-DA scores with $R^2(X) = 0.55$, $R^2(Y) = 0.992$, $Q^2 = 0.953$ and CV-ANOVA *p-value* = 9.41×10^{-5}. (C) XCMS analysed NIM data PCA scores with $R^2 = 0.729$, $Q^2 = 0.149$ values. (D) XCMS analysed NIM data OPLS-DA scores with $R^2(X) = 0.662$, $R^2(Y) = 0.996$, $Q^2 = 0.951$ and a CV-ANOVA *p-value* = 7.35×10^{-4}.

Table 4.1. The top twenty unique and available proposed identities of the most significant PIM XCMS features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Fold Change	p-value	FDR	Mass difference
355.2250	5.00	*all-trans-retinol*	-0.98	2.86E-15	4.25E-12	0.0006
369.2370	5.01	*(9S,10S)-10-hydroxy-9-(phosphooxy) octadecanoate*	-0.98	4.38E-14	1.87E-11	0.0027
234.0780	1.90	*N2-succinyl-L-glutamate 5-semialdehyde*	96.42	1.93E-13	4.79E-11	0.0007
292.1010	1.28	*canavaninosuccinate*	98.87	7.57E-13	1.05E-10	0.0006
213.1030	5.00	*geraniol*	-0.89	8.25E-13	1.11E-10	0.0010
270.1170	1.28	*S-(hydroxymethyl)glutathione*	559.76	1.31E-12	1.54E-10	0.0023
111.0470	2.42	*O-succinyl-L-homoserine*	5.44	2.08E-12	2.11E-10	0.0005
399.1210	2.74	*S-adenosyl-4-methylthio-2-oxobutanoate*	188.65	3.07E-12	2.58E-10	0.0001
251.1040	1.29	*5'-deoxyadenosine*	166.71	3.38E-12	2.76E-10	0.0018
183.0810	6.16	*coniferyl alcohol*	20.49	4.31E-12	3.11E-10	0.0012
289.1400	0.89	*hypoglycin B*	107.75	5.24E-12	3.43E-10	0.0006
237.0860	0.90	*pyridoxamine*	147.56	6.58E-12	4.10E-10	0.0012
253.0920	2.49	*2'-deoxyinosine*	16.8	6.80E-12	4.10E-10	0.0008
252.0870	2.50	*S-acetyldihydrolipoamide*	45.6	7.69E-12	4.51E-10	0.0022
149.0110	1.50	*imidazole acetol-phosphate*	12.52	8.22E-12	4.69E-10	0.0000
267.0980	0.67	*adenosine*	109.98	8.80E-12	4.87E-10	0.0012
256.0590	2.36	*N-acetyl-α-D-glucosamine 1-phosphate*	25.1	1.11E-11	5.79E-10	0.0007
273.0870	3.22	*L-tryptophan*	3.91	1.31E-11	6.42E-10	0.0025
137.0710	0.90	*L-canaline*	154.2	1.80E-11	8.25E-10	0.0013
440.2270	5.96	*sphinganine 1-phosphate*	-0.86	2.10E-11	9.26E-10	0.0033

m/z values, retention times, fold change ((B-A)/A), log2 fold change, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table 4.2. PIM pathway hits via Metaboanalyst's Mummichog using XCMS data with three or more significant hits.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	6	5	0.09457	0.016791	0.001701
tryptophan degradation I (via anthranilate)	9	5	4	0.20286	0.043635	0.002779
trans, trans-farnesyl diphosphate biosynthesis	5	4	3	0.40064	0.109	0.007017
biotin biosynthesis from 7-keto-8-aminopelargonate	15	7	4	0.42055	0.16963	0.007724
isoleucine biosynthesis	15	5	3	0.5325	0.20705	0.013436
hexaprenyl diphosphate biosynthesis	6	5	3	0.5325	0.20705	0.013436
salvage pathways of pyrimidine deoxyribonucleotides	17	5	3	0.5325	0.20705	0.013436

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst. All top ten hits are available in appendix A.

Table 4.3. The top twenty unique and available proposed identities of the most significant NIM XCMS features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Fold Change	p-value	FDR	Mass difference
181.0700	7.02	*coniferyl alcohol*	-0.9	4.07E-14	2.63E-11	0.0018
611.3840	5.80	*3-methoxy-4-hydroxy-5-all-trans-hexaprenylbenzoate*	-0.98	5.20E-14	2.80E-11	0.0024
515.1770	5.00	*5,10-methenyltetrahydrofolate mono-L-glutamate*	-0.96	1.81E-13	4.51E-11	0.0002
267.0880	1.28	*coniferyl acetate*	476.22	8.44E-13	1.23E-10	0.0006
172.0350	2.50	*3-dehydroshikimate*	10.76	1.26E-12	1.63E-10	0.0012
233.0540	4.26	*geraniol*	51.85	3.32E-12	2.82E-10	0.0000
218.0410	2.49	*L-tyrosine*	18.89	1.00E-11	6.23E-10	0.0007
231.0710	1.84	*N2-succinyl-L-glutamate 5-semialdehyde*	429.97	1.39E-11	7.86E-10	0.0008
190.0420	2.49	*3-dehydroquinate*	12.13	1.60E-11	8.76E-10	0.0015
268.0950	1.28	*2'-deoxyuridine*	361.2	1.70E-11	9.01E-10	0.0011
234.0350	2.49	*monodehydroascorbate radical*	12.18	1.82E-11	9.19E-10	0.0022
158.0540	4.55	*2-isopropylmaleate*	57.2	1.98E-11	9.69E-10	0.0003
205.0630	2.90	*3-hydroxy-L-kynurenine*	7.57	2.02E-11	9.74E-10	0.0007
202.0460	2.50	*L-phenylalanine*	9.11	4.02E-11	1.58E-09	0.0015
214.0410	3.02	*O-ureido-L-homoserine*	6.74	4.32E-11	1.68E-09	0.0001
234.0420	4.07	*kynurenate*	58.83	6.05E-11	2.10E-09	0.0018
220.0510	2.50	*L-histidinol phosphate*	16.36	8.40E-11	2.77E-09	0.0017
838.5450	6.27	*trehalose-cis-methoxy-mono-mycolate*	-0.94	1.22E-10	3.79E-09	0.0024
327.9830	3.38	*N2-succinylglutamate*	-0.78	1.32E-10	4.01E-09	0.0024
303.9800	2.50	*4-amino-4-deoxychorismate*	10.78	1.48E-10	4.33E-09	0.0022

m/z values, retention times, fold change ((B-A)/A), log2 fold change, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table 4.4. NIM pathway hits via Metaboanalyst's Mummichog using XCMS data with three or more significant hits.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
phosphopantothenate biosynthesis I	16	5	4	0.52968	0.21092	0.010671
leucine biosynthesis	16	4	3	0.71247	0.34181	0.022197
pyridoxal 5'-phosphate salvage pathway	13	4	3	0.71247	0.34181	0.022197
ubiquinol-6 biosynthesis (eukaryotic)	21	5	3	0.83433	0.53814	0.039989

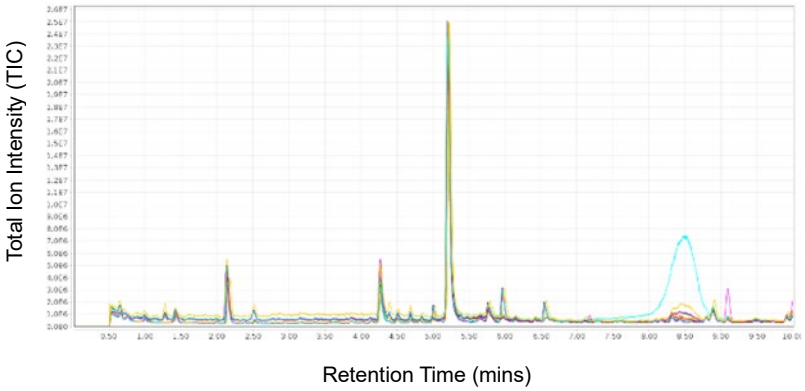
The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst. All top ten hits are available in appendix A.

4.3. MZMine2 results

4.3.1. Feature detection and normalisation

MZmine2 data analysis of the same 12 samples previously analysed in XCMS was conducted following the parameters outlined in section 2.8. This revealed 8295 aligned features in PIM with very little retention time deviation. The TICs produced in MZmine2 showed very little variance between samples and bear a striking resemblance to those produced in XCMS (Figures 4.3. A). In contrast to XCMS more features were found in NIM data, revealing 12672 aligned features with very little retention time deviation. The TICs produced in MZmine2 showed very little variance between samples and align with those produced in XCMS well (Figures 4.3. B).

A)



B)

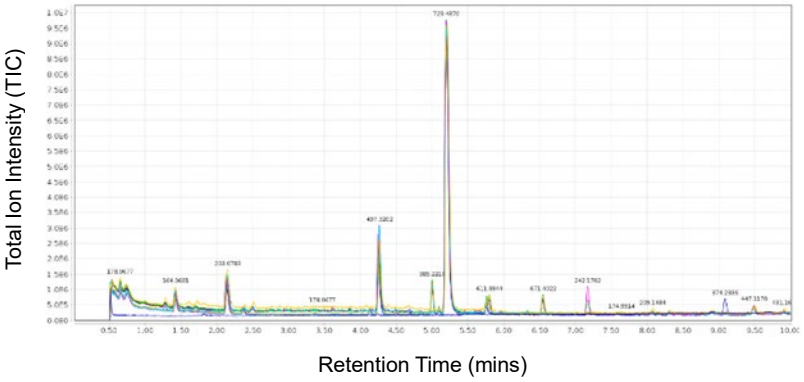


Figure 4.3. Total ion chromatograms from MZmine2 of *Δrnq1* and [*rnq*$^{-}$] samples. MZmine2 processing information regarding all of the detected LC-MS peaks observed within all 6 *Δrnq1* (samples 03, 16, 21, 23, 27 and 29) and all 6 [*rnq*$^{-}$] (samples 11, 14, 18, 19, 22 and 28) samples imported for analysis and peak detection. A) PIM original total ion chromatogram B) NIM original total ion chromatogram

4.3.2. Multivariate modelling of MZmine2 metabolomic data reveals some differences between *Δrnq1* and [*rnq*$^-$] samples

To assess if significant differences could be seen between the two groups, *Δrnq1* and [*rnq*$^-$], PCA modelling in SIMCA was conducted, this time using all 12 samples of MZmine2 pre-processed data. This showed some tentative, observable separation between the two groups both for PIM and NIM. However, the spatial separation of the NIM model was arguably questionable, existing as a single large cluster as opposed to two individual clusters. Model scores were poor, giving low levels of confidence with their respective R^2 and Q^2 values less than acceptable for biological relevance, $R^2 = 0.323$ $Q^2 = 0.156$ and $R^2 = 0.512$, $Q^2 = 0.067$ (Figure 4.4. A + C). When examining such variability between XCMS and MZmine2 findings, it was is evident that the number of detected features was larger within MZmine2 than XCMS and may introduce additional levels of noise to the MZmine2 models hindering model quality and predictability within SIMCA.

OPLS-DAs were built in an effort to improve on these poor models including all 12 samples and focusing on comparing *Δrnq1* to [*rnq*$^-$]. With PIM these showed strong significant separation of *Δrnq1* and [*rnq*$^-$] samples with $R^2(X) = 0.708$, $R^2(Y) = 1$, $Q^2 = 0.933$ and a CV-ANOVA *p-value* = 3.23×10^{-2} (Figure 4.4. B). This demonstrates that metabolic changes between the two groups were significantly different from each other. However, in NIM although the $R^2(X) = 0.624$, $R^2(Y) = 0.999$, $Q^2 = 0.847$ were acceptable (evidence that the model itself was of good quality), the CV-ANOVA *p-value* showed no significant difference at 0.0576 despite the appearance of the model to show clear separation (Figure 4.4. D). Although relatively weak PCA models were produced here, some degree of separation is seen, and so OPLS-DA models were appropriate. It was recognised, that these models have been built using weaker sources of variation than their PCA counterparts (Worley and Powers, 2016). Visible separation in OPLS-DA is expected and so CV-ANOVA scores will better represent the significance of these separations. Despite what was found in XCMS data and working off MZmine2 NIM data alone, the clear conclusion is that there is no significant difference seen among the two groups. However, it is known that OPLS-DA statistics are rapidly decreased with added noise within a data set, with this having been a concern during PCA analysis it is possible that this could be a limiting factor here. Due to the presence of a known “noise factor”, the decision to continue into Metaboanalyst with this data set, despite these ‘insignificant’ findings was taken.

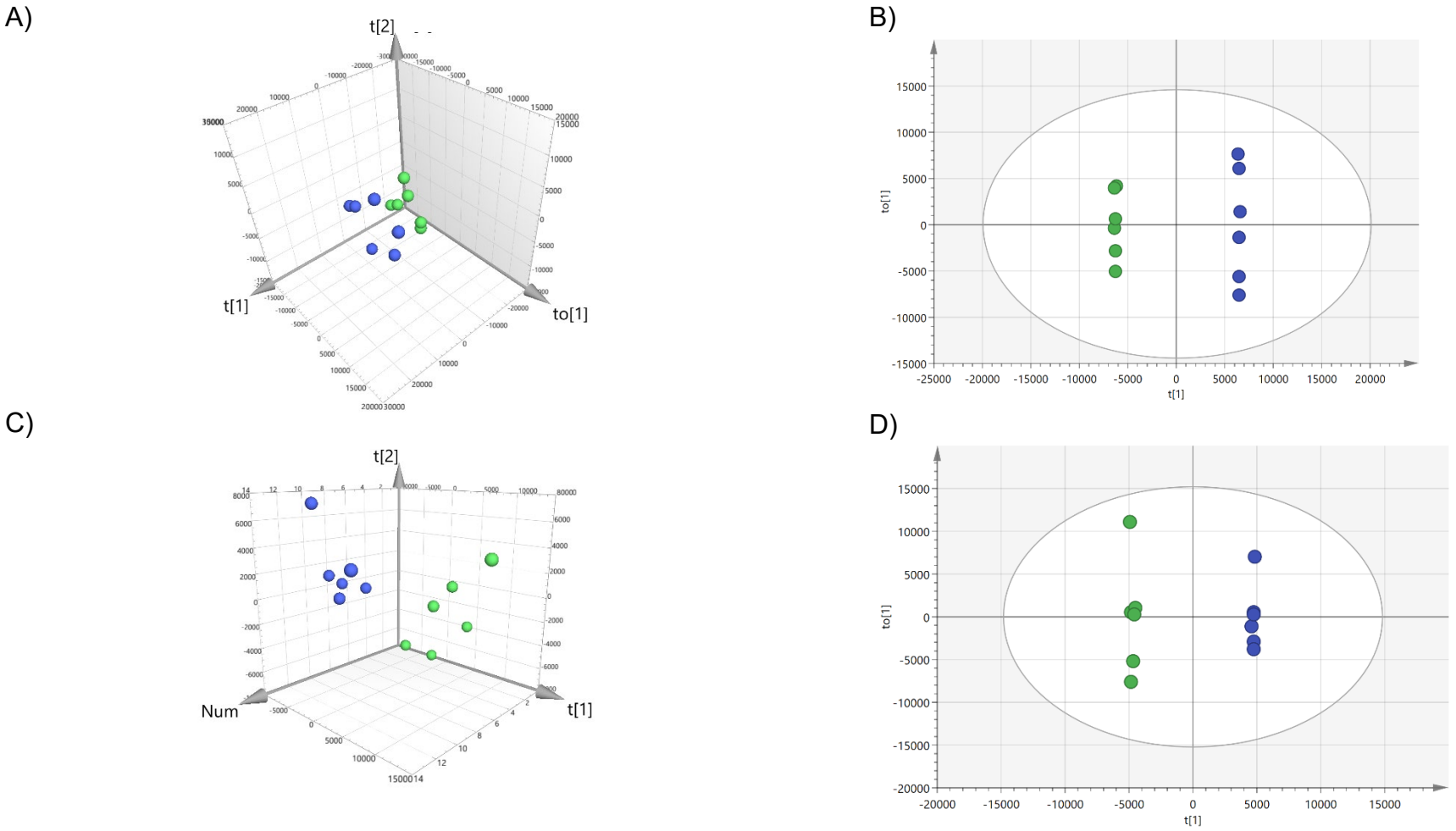


Figure 4.4. PCA (A and C) and OPLS-DA (B and D) models of MZmine2 comparison of *Δrnq1* and [*rnq*$^-$] samples. Overview of the data shows no outlying samples within the 95% confidence range within each model. Green circles represent *Δrnq1* samples and blue circles represent [*rnq*$^-$] samples (A) MZmine2 analysed PIM data 3D PCA scores with R^2 = 0.323 Q^2 = 0.156 values. (B) MZmine2 analysed PIM data OPLS-DA scores with $R^2(X)$ = 0.708, $R^2(Y)$ = 1, Q^2 = 0.933 and a CV-ANOVA *p-value* = 3.23 x 10^{-2}. (C) MZmine2 analysed NIM data 3D PCA scores with R^2 = 0.512, Q^2 = 0.067 values. (D) MZmine2 analysed NIM data OPLS-DA scores with $R^2(X)$ = 0.624, $R^2(Y)$ = 0.999, Q^2 = 0.847 and a CV-ANOVA *p-value* = 0.0576.

4.3.3. Statistical analysis of MZmine2 metabolomic data in Metaboanalyst finds significantly altered features

Much like XCMS data, group MZmine2 data of all 12 samples was subjected to pairwise comparison within Metaboanalyst, using the settings described in section 2.9. Resultant T-tests with the adjusted *p-value* of =>0.01 found 1807 positive and 3676 negative features that had been significantly altered (*p-value*=>0.01) when comparing *Δrnq1* samples to [*rnq*$^{-}$] samples.

4.3.4. Tentative feature ID and pathway analysis of MZmine2 metabolomic data via Metaboanalyst

In order to establish the identity of these increased and decreased features detected within the respective 1807 and 3676 features and run pathway analysis, Metaboanalyst's, Mummichog, *Peaks to Pathways* function was used, as indicated by Figure 4.1. A *p-value* of 1.0 x 10^{-5} allowed for the acquisition of 293 PIM results and 916 NIM results (closest possible to the required between 300~700 for algorithm functioning). BioCyc's Fungi pathway library for *S. cerevisiae* (yeast) was selected. This provided details of the most significantly different metabolites, the direction of the disturbance (whether they were up or down regulated), and the most significantly disrupted pathways including the total number of hits within specified pathways (Tables 4.5., 4.6., 4.7. and 4.8.). For reasons outlined previously in section 4.2.4., the decision was taken to progress onto comparative analysis.

Table 4.5. The top twenty unique and available proposed identities of the most significant PIM MZmine2 features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Fold Change	p-value	FDR	Mass difference
258.0760	2.80	*5'-chloro-5'-deoxyadenosine*	2.23	6.44E-11	1.34E-07	0.0006
252.0880	1.80	*S-acetyldihydrolipoamide*	6.45	1.39E-10	1.68E-07	0.0004
151.0880	1.70	*pyridoxamine*	10.28	1.41E-09	5.33E-07	0.0010
235.1090	1.70	*hypoglycin B*	10.66	1.59E-09	5.73E-07	0.0016
293.0660	1.93	*5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide*	2.64	2.73E-09	8.40E-07	0.0013
369.2390	5.08	*(9S,10S)-10-hydroxy-9-(phosphooxy) octadecanoate*	0.12	4.96E-09	1.25E-06	0.0013
199.0580	3.04	*(2R,3S)-3-isopropylmalate*	0.82	6.29E-09	1.49E-06	0.0001
161.0630	5.57	*2-isopropylmaleate*	12.7	1.18E-08	2.33E-06	0.0004
213.0760	1.98	*L-arginine*	0.26	1.58E-08	2.91E-06	0.0012
369.2410	5.06	*stearate*	0.98	1.66E-08	3.00E-06	0.0000
355.2250	4.14	*all-trans-retinol*	0.98	1.96E-08	3.46E-06	0.0004
241.0610	1.98	*dUMP*	7.78	2.07E-08	3.54E-06	0.0001
179.0820	1.71	*3-hydroxy-L-kynurenine*	9.94	2.09E-08	3.54E-06	0.0008
193.1000	2.00	*geraniol*	0.73	2.32E-08	3.83E-06	0.0018
175.0850	4.59	*(indol-3-yl)acetamide*	0.95	2.36E-08	3.83E-06	0.0013
253.1200	1.64	*hypoglycin B*	3.34	2.99E-08	4.43E-06	0.0016
125.0710	1.38	*histidinol*	13.88	3.29E-08	4.70E-06	0.0003
251.1010	1.75	*5'-deoxyadenosine*	14.32	3.36E-08	4.72E-06	0.0011
179.0820	1.71	*3-hydroxy-L-kynurenine*	9.94	3.42E-08	4.72E-06	0.0002
248.1100	5.05	*thiamine*	1.23	3.61E-08	4.91E-06	0.0012

m/z values, retention times, fold change ((B-A)/A), log2 fold change, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table 4.6. PIM pathway hits via Metaboanalyst's Mummichog using MZmine2 data with three or more significant hits.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
salvage pathways of pyrimidine deoxyribonucleotides	17	6	4	0.26337	0.075142	0.001591
tryptophan biosynthesis	17	9	5	0.27948	0.10584	0.00173
4-amino-2-methyl-5-diphosphomethylpyrimidine biosynthesis	11	4	3	0.35673	0.088631	0.002596
thiamin diphosphate biosynthesis IV (eukaryotes)	11	4	3	0.35673	0.088631	0.002596
pyridoxal 5'-phosphate salvage pathway	13	4	3	0.35673	0.088631	0.002596
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	4	3	0.35673	0.088631	0.002596
tryptophan degradation I (via anthranilate)	9	4	3	0.35673	0.088631	0.002596
methionine biosynthesis	18	8	4	0.46127	0.2066	0.004561

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst. All top ten hits are available in appendix A.

Table 4.7. The top twenty unique and available proposed identities of the most significant NIM MZmine2 features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Fold Change	p-value	FDR	Mass difference
268.0940	1.31	*2'-deoxyuridine*	0.09	2.33E-14	2.88E-11	0.0003
231.0700	1.32	*N2-succinyl-L-glutamate 5-semialdehyde*	-0.16	6.74E-14	4.74E-11	0.0001
218.0400	2.49	*L-tyrosine*	16.98	1.33E-13	7.66E-11	0.0000
611.3840	5.80	*3-methoxy-4-hydroxy-5-all-trans-hexaprenylbenzoate*	-0.98	2.03E-13	1.03E-10	0.0024
267.0870	1.38	*coniferyl acetate*	-0.01	4.57E-13	1.81E-10	0.0003
515.1750	5.01	*5,10-methenyltetrahydrofolate mono-L-glutamate*	-0.35	4.93E-13	1.86E-10	0.0013
165.0600	0.75	*2-deoxy-D-glucose 6-phosphate*	0.45	1.39E-12	3.92E-10	0.0012
195.0730	1.14	*6-(hydroxymethyl)-7,8-dihydropterin*	-0.06	1.90E-12	4.92E-10	0.0015
221.9790	4.24	*(1E)-4-oxobut-1-ene-1,2,4-tricarboxylate*	-0.57	3.52E-12	7.97E-10	0.0003
202.0430	2.49	*L-phenylalanine*	11.96	5.08E-12	1.04E-09	0.0016
335.0720	1.38	*S-formylglutathione*	-0.02	1.01E-11	1.91E-09	0.0026
214.0420	3.04	*O-ureido-L-homoserine*	0.05	1.20E-11	2.18E-09	0.0007
176.0270	3.19	*L-ascorbate*	0.08	1.33E-11	2.26E-09	0.0008
219.0420	2.49	*phosphocholine*	-0.16	1.35E-11	2.26E-09	0.0005
268.0730	2.73	*1,2-dibutyrin*	-0.59	1.61E-11	2.58E-09	0.0008
359.0860	3.92	*4'-phosphopantetheine*	-0.06	2.06E-11	3.15E-09	0.0009
193.0710	2.95	*(R)-2,3-dihydroxy-3-methylpentanoate*	1.86	2.06E-11	3.15E-09	0.0001
237.0310	2.58	*4-hydroxy-2-nonenal*	40.08	2.11E-11	3.15E-09	0.0000
334.0140	2.41	*7,8-dihydroneopterin*	-0.54	3.37E-11	4.55E-09	0.0010
493.1100	4.21	*5,10-methylenetetrahydropteroyl mono-L-glutamate*	-0.9	3.75E-11	5.00E-09	0.0023

m/z values, retention times, fold change ((B-A)/A), log2 fold change, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table 4.8. NIM pathway hits via Metaboanalyst's Mummichog using MZmine2 data with three or more significant hits.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
6-hydroxymethyl-dihydropterin diphosphate biosynthesis I	14	4	3	0.14625	0.019586	0.000531
folate transformations	26	7	3	0.36067	0.11365	0.001703
phosphopantothenate biosynthesis I	16	7	3	0.36067	0.11365	0.001703
folate interconversions	22	8	3	0.4297	0.15899	0.002512

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst. All top ten hits are available in appendix A.

4.4. Comparative data analysis results

The top twenty proposed variable features identified by T-tests and the subsequent pathway analyses reveal that much discrepancy exists between the two analysis programmes. As discussed earlier (Section 4.1.), it has been suggested by Li *et al.* (2019) that cross comparison of the overlapping m/z values from a variety of programmes, specifically XCMS and MZmine2, provides a level of confidence in analysis that cannot be achieved via a single analysis method. This approach has been taken here correlating both our positive and negative ionisation mode data sets in an attempt to provide a level of confidence in our findings that could not be achieved via single analysis programme.

4.4.1. Feature detection and normalisation

As instructed via Myers *et al.* (2017), cross correlation of m/z values from XCMS and MZmine2 was achieved via the methods outlined in section 2.8.4. Within PIM data 2748 m/z values were found to have been reported by both programmes and within NIM data 2769 overlapping features were found (Figure 4.5.(A+B)).

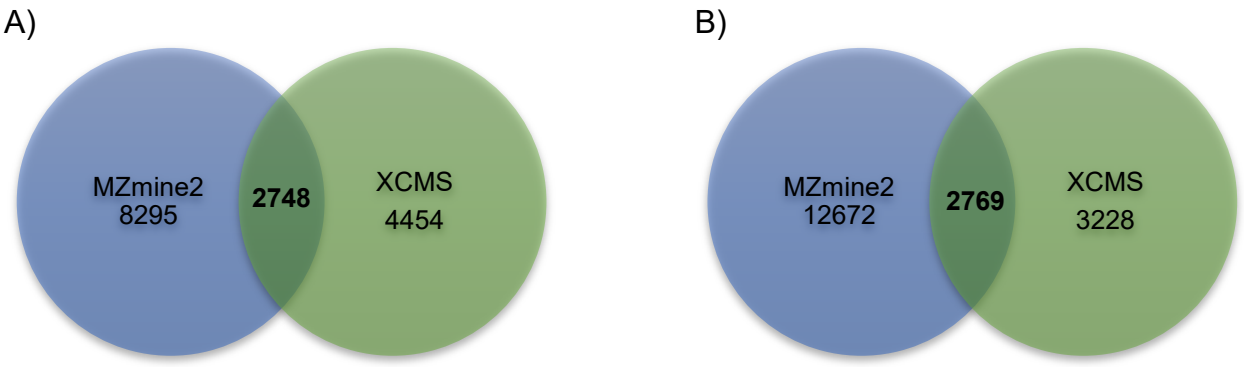


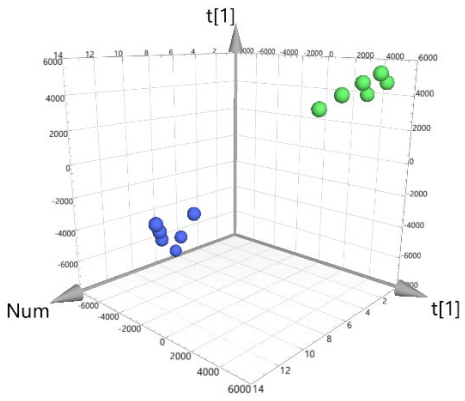
Figure 4.5. Venn diagrams showing the total number of overlapping features reported by MZmine2 and the total number of features reported by XCMS when comparing *Δrnq1* and [*rnq*$^-$] data. Mzmine2 features are shown in blue and XCMS features are shown in green, the overlapping section in the centre of the two depicts the total number of m/z values reported by both analysis programmes in bold A) Venn diagram of PIM m/z values of MZmine2 and XCMS B) Venn diagram of NIM m/z values of MZmine2 and XCMS

4.4.2. Multivariate modelling of XCMS/MZmine2 comparative metabolomic data in SIMCA differentiates *Δrnq1* and [*rnq*$^-$] samples

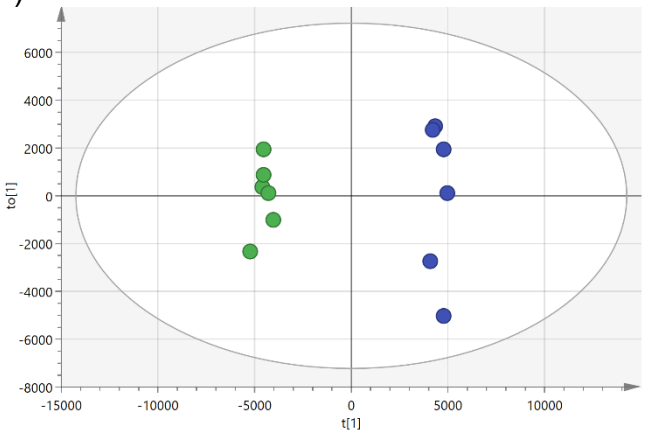
Statistical models were built in SIMCA to see if clustering and separation between the two groups, *Δrnq1* and [*rnq*$^-$], could be seen between the newly established 2748 PIM m/z features and the 2769 NIM m/z features. Visible group separation and group clustering was observed, although model confidence was only moderate due to the respective PIM and NIM, R^2 = 0.451 Q^2 = 0.312 and R^2 = 0.622, Q^2 = 0.316 (Figure 4.6. (A + C)). R^2 values however were considered acceptable but Q^2 values were slightly less than ideal, suggesting that the model's capability to predict traits would have been unreliable. When examining Q^2 values, it was considered that this may have been as a result of a limited *n* number. More repeats are always preferable but given the hypothesis generating nature of the research, it was decided to restrict the *n* number of all sample groups to six. This decision was based on personal communication suggesting that an *n* of five was considered valid for this type of experimentation, the prevalence of an *n* of five in the literature and the financial implications of increasing the number of samples, with respect to machine time (Personal communication with Dr Jake Bundy, 2016).

Encouragingly, despite some trepidation after evaluating our PCAs, OPLS-DAs gave great scores, with PIM showing $R^2(X)$ = 0.544, $R^2(Y)$ = 0.994, Q^2 = 0.959 and NIM showing $R^2(X)$ = 0.685, $R^2(Y)$ = 0.998, Q^2 = 0.953. Giving great confidence in the models and their respective excellent CV-ANOVA *p-value*s of 5.85 x 10^{-5} and 6.61 x 10^{-4} respectively (Figure 4.6. (B+D)). It seemed as though comparative analysis or 'data trimming' had removed a large amount of the noise from the system, hence strengthening our OPLS-DA models, allowing confidence in the data set and subsequent models, and therefore allowing for the conclusions that significant metabolic differences were present between *Δrnq1* and [*rnq*$^-$] samples.

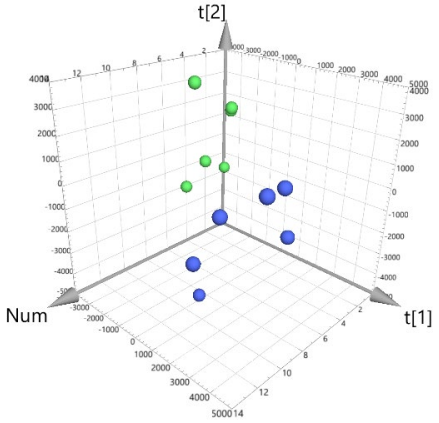
A)



B)



C)



D)

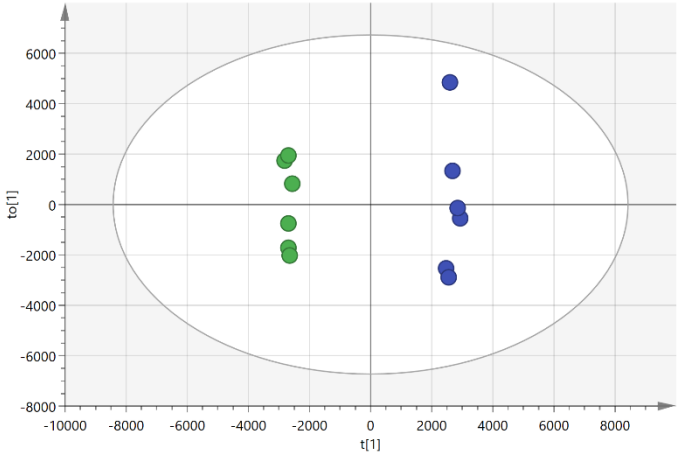


Figure 4.6. PCA (A and C) and OPLS-DA (B and D) models of XCMS/MZmine2 comparative data of *Δrnq1* and [*rnq*$^-$] samples. Overview of the data shows no outlying samples within the 95% confidence range within each model. Green circles represent *Δrnq1* samples and blue circles represent [*rnq*$^-$] samples (A) Comparative analysis PIM data 3D PCA scores with $R^2 = 0.451$ $Q^2 = 0.312$ values. (B) Comparative analysis PIM data OPLS-DA scores with $R^2(X) = 0.544$, $R^2(Y) = 0.994$, $Q^2 = 0.959$ and a CV-ANOVA *p-value* = 5.85×10^{-5}. (C) Comparative analysis NIM data 3D PCA scores with $R^2 = 0.622$, $Q^2 = 0.316$ values. (D) Comparative analysis NIM data OPLS-DA scores with $R^2(X) = 0.685$, $R^2(Y) = 0.998$, $Q^2 = 0.953$, and a CV-ANOVA *p-value* = 6.61×10^{-4}.

4.4.3. Statistical analysis of XCMS/MZmine2 comparative metabolomic data in Metaboanalyst finds significant differences between *Δrnq1* and [*rnq*$^-$] samples

Comparative data were treated in an identical manner to single data analysis, with all 12 samples being subjected to pairwise comparison within Metaboanalyst, using the settings described in section 2.9. The resultant T-tests, with the adjusted *p-value* of =>0.01, found 1620 positive and 1349 negative features that had been significantly altered (*p-value*=>0.01) when comparing *Δrnq1* samples to [*rnq*$^-$] samples. This was encouraging given the reduction in overall number of features caused by m/z value comparison, suggesting that this 'data trimming' exercise had little effect on the significantly altered metabolites.

4.4.4. Tentative feature ID and pathway analysis of XCMS/MZmine2 comparative metabolomic data via Metaboanalyst

To establish the identity and run pathway analyses on the increased and decreased features detected by both XCMS and MZmine2, the 1620 PIM features and 1349 NIM features where subject to Metaboanalyst's, Mummichog, *Peaks to Pathways* function, as indicated by Figure 4.1. A *p-value* of $1.0\,x\,10^{-7}$ allowed for the acquisition of 368 PIM results and 456 NIM results (closest possible to the required between 300~700 for algorithm functioning). BioCyc's Fungi pathway library for *Saccharomyces cerevisiae* (yeast) was selected, providing details of the most significantly different metabolites, the direction of the disturbance (whether they were up or down regulated), and the most significantly disrupted pathways including the total number of hits within specified pathways (Tables 4.9., 4.10., 4.11. and 4.12.).

From examining Tables 4.9. and 4.11. it is evident that approximately 80% of the changes observed were due to individual metabolites being upregulated in strains with the native Rnq1p conformation with only 7 PIM and 6 NIM down regulation events. In isolation, these individual results are only indicative, and conclusions based on single changes unreliable. However, taken within the context of pathway analysis, if one metabolite is known to change considerably more or in a more accessible or measurable way, this may reveal areas of interest regarding the function of Rnq1p. When considered as part of a pathway analysis, the only significantly disturbed pathway is that of tryptophan degradation. It was considered that the limitations of Metaboanalyst with regard to the individual treatment of both PIM and NIM results may be providing the weak EASE results seen within pathway analysis and so the decision was made to continue analysis within BioCyc to enable PIM and NIM results to be combined and subsequent pathways further examined.

Table 4.9. The top twenty-five unique and available proposed identities of the most significant PIM comparative features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Fold Change	p-value	FDR	Mass difference
355.2250	5.00	*all-trans-retinol*	-0.98	2.38E-15	4.91E-12	0.0006
369.2373	5.01	*(9S,10S)-10-hydroxy-9-(phosphooxy) octadecanoate*	-0.98	9.14E-14	1.93E-11	0.0027
234.0780	1.90	*N2-succinyl-L-glutamate 5-semialdehyde*	96.42	2.54E-13	3.70E-11	0.0007
270.1165	1.28	*S-(hydroxymethyl)glutathione*	559.76	6.87E-13	7.02E-11	0.0023
336.2296	5.81	*leukotriene B4*	-0.91	6.89E-13	7.02E-11	0.0005
111.0466	2.42	*O-succinyl-L-homoserine*	5.44	8.60E-13	8.15E-11	0.0005
213.1026	5.00	*geraniol*	-0.89	1.16E-12	9.94E-11	0.0010
399.1206	2.74	*S-adenosyl-4-methylthio-2-oxobutanoate*	188.65	2.95E-12	2.06E-10	0.0001
251.1036	1.29	*5'-deoxyadenosine*	166.71	3.41E-12	2.06E-10	0.0018
253.0924	2.49	*2'-deoxyinosine*	16.8	3.53E-12	2.06E-10	0.0008
237.0858	0.90	*pyridoxamine*	147.56	4.61E-12	2.48E-10	0.0012
252.0866	2.50	*S-acetyldihydrolipoamide*	45.6	4.75E-12	2.51E-10	0.0022
81.0334	1.50	*3-methyl-2-oxobutanoate*	22.06	1.29E-11	5.28E-10	0.0001
273.0871	3.22	*L-tryptophan*	3.91	1.49E-11	5.79E-10	0.0025
149.0111	1.50	*imidazole acetol-phosphate*	12.52	1.83E-11	6.63E-10	0.0000
311.0775	1.91	*5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide*	341.54	2.33E-11	7.54E-10	0.0025
167.0825	0.66	*1-octanal*	41.17	2.52E-11	7.86E-10	0.0004
440.2270	5.96	*sphinganine 1-phosphate*	-0.86	3.15E-11	9.42E-10	0.0033
258.0741	2.90	*5'-chloro-5'-deoxyadenosine*	15.72	3.40E-11	9.83E-10	0.0010
210.9926	1.50	*dimethylallyl diphosphate*	18.81	6.41E-11	1.63E-09	0.0007
197.0923	0.90	*3-hydroxy-L-kynurenine*	201.1	8.39E-11	2.01E-09	0.0003
108.0456	1.84	*D-serine*	22.83	8.52E-11	2.02E-09	0.0001
613.4265	5.81	*ubiquinone-6*	-0.97	9.86E-11	2.28E-09	0.0037
166.0180	5.01	*dUMP*	-0.96	1.93E-10	3.96E-09	0.0007
287.1250	0.68	*7,8-dihydropteroate*	78.04	2.00E-10	4.01E-09	0.0000

m/z values, retention times, fold change ((B-A)/A), log2 fold change, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table 4.10. PIM pathway hits via Metaboanalyst's Mummichog using comparative data with three or more significant hits.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	6	6	0.035235	0.0034343	0.0006108
UDP-N-acetylglucosamine biosynthesis	12	5	5	0.078028	0.0089517	0.0007525
UDP-N-acetyl-D-glucosamine biosynthesis II	12	5	5	0.078028	0.0089517	0.0007525
tryptophan degradation I (via anthranilate)	9	5	5	0.078028	0.0089517	0.0007525
phenylalanine biosynthesis	11	6	5	0.16361	0.036539	0.0011471
tyrosine biosynthesis	11	6	5	0.16361	0.036539	0.0011471
hexaprenyl diphosphate biosynthesis	6	4	4	0.16654	0.023216	0.0011639
trans, trans-farnesyl diphosphate biosynthesis	5	4	4	0.16654	0.023216	0.0011639
leucine biosynthesis	16	9	6	0.24341	0.088884	0.0017097
biotin biosynthesis from 7-keto-8-aminopelargonate	15	7	5	0.26891	0.087449	0.0019449

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst.

Table 4.11. The top twenty-five unique and available proposed identities of the most significant NIM comparative features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Fold Change	p-value	FDR	Mass difference
181.0704	7.02	*coniferyl alcohol*	-0.9	2.11E-14	1.17E-11	0.0018
611.3844	5.80	*3-methoxy-4-hydroxy-5-all-trans-hexaprenylbenzoate*	-0.98	6.51E-14	2.89E-11	0.0024
515.1766	5.00	*5,10-methenyltetrahydrofolate mono-L-glutamate*	-0.96	4.83E-13	9.55E-11	0.0002
267.0875	1.28	*coniferyl acetate*	476.22	1.03E-12	1.51E-10	0.0006
172.0345	2.50	*3-dehydroshikimate*	10.76	1.75E-12	2.02E-10	0.0012
613.3840	5.80	*3-methoxy-4-hydroxy-5-all-trans-hexaprenylbenzoate*	-0.97	2.62E-12	2.51E-10	0.0003
233.0541	4.26	*geraniol*	51.85	5.64E-12	4.23E-10	0.0000
218.0405	2.49	*L-tyrosine*	18.89	9.93E-12	6.79E-10	0.0007
190.0423	2.49	*3-dehydroquinate*	12.13	1.32E-11	7.92E-10	0.0015
231.0712	1.84	*N2-succinyl-L-glutamate 5-semialdehyde*	429.97	2.45E-11	1.18E-09	0.0008
158.0538	4.55	*2-isopropylmaleate*	57.2	2.46E-11	1.18E-09	0.0003
205.0626	2.90	*3-hydroxy-L-kynurenine*	7.57	2.79E-11	1.25E-09	0.0007
234.0353	2.49	*monodehydroascorbate radical*	12.18	2.94E-11	1.29E-09	0.0022
268.0950	1.28	*2'-deoxyuridine*	361.2	3.43E-11	1.46E-09	0.0011
214.0408	3.02	*O-ureido-L-homoserine*	6.74	4.18E-11	1.73E-09	0.0001
202.0464	2.50	*L-phenylalanine*	9.11	5.59E-11	2.15E-09	0.0015
838.5454	6.27	*α, α'-trehalose 6-α-mycolate*	-0.94	6.74E-11	2.46E-09	0.0024
220.0510	2.50	*L-histidinol phosphate*	16.36	9.07E-11	3.04E-09	0.0017
234.0420	4.07	*kynurenate*	58.83	9.12E-11	3.04E-09	0.0018
327.9831	3.38	*N2-succinylglutamate*	-0.78	9.41E-11	3.07E-09	0.0024
303.9799	2.50	*4-amino-4-deoxychorismate*	10.78	9.57E-11	3.07E-09	0.0022
174.0513	2.90	*(S)-ureidoglycolate*	1.98	1.34E-10	4.08E-09	0.0007
165.0598	1.17	*2-deoxy-D-glucose 6-phosphate*	16.1	3.05E-10	7.98E-09	0.0014
237.0318	2.58	*4-hydroxy-2-nonenal*	183.18	4.25E-10	1.03E-08	0.0005
219.0414	2.49	*phosphocholine*	18.63	1.96E-10	5.54E-09	0.0014

m/z values, retention times, fold change ((B-A)/A), log2 fold change, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table 4.12. NIM pathway hits via Metaboanalyst's Mummichog using comparative data with three or more significant hits.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
ubiquinol-6 biosynthesis (eukaryotic)	21	6	4	0.51914	0.22761	0.0068221
leucine biosynthesis	16	4	3	0.57758	0.21942	0.0088931
pyridoxal 5'-phosphate salvage pathway	13	4	3	0.57758	0.21942	0.0088931
phosphopantothenate biosynthesis I	16	5	3	0.71593	0.37775	0.017462
salvage pathways of pyrimidine deoxyribonucleotides	17	7	3	0.88128	0.65561	0.046963

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst. All top ten hits are available in appendix A.

4.4.5. BioCyc 'omics dashboard and cellular overview enables mapping of most significant metabolic changes indicating key metabolic pathways of interest

To prepare for pathway analysis within BioCyc, all significant punitively identified metabolites (as established by Metaboanalyst's Mummichog), their relative abundances and average relative abundances (within groupings) were collated (this included all PIM and NIM omics data). Relative abundances of the most significantly altered metabolites (established via Metaboanalyst's Mummichogg) from the six *Δrnq1* samples and the six [*rnq*$^{-}$] samples were then uploaded to BioCyc's omics dashboard. This allowed for an overall assessment of the key metabolic regions influenced by this change (Figure 4.7) and visualisation of the scale of specific metabolic regions altered between groups, shown in Figure 4.8.

As shown by Figure 4.7. the most significantly perturbed metabolic pathways identified were involved in biosynthesis, with cofactor and amino acid biosynthesis being the largest effected groups. It was noted that these results correlated well with previous Mummichog observations. Given these changes in biosynthesis, it was a surprise to see how little the metabolic pathways involved in energy metabolism had been perturbed, although changes in this metabolic area may result in a knockout mutant such as the *Δrnq1* strain being inviable. The second largest perturbation was shown to be within pathways and of compounds (Cpd) that are yeast specific. Relating this result to the lack of an identifiable homolog of Rnq1p other organisms suggests the role of this protein is likely to be yeast specific and changes in the presence or absence of Rnq1p are likely to influence yeast specific mechanisms. The identification of perturbations in both degradation pathways and in amino acid synthesis pathways on loss of Rnq1p suggests that either the use of amino acids or amino acid preference within the cells may have changed or that the communication mechanisms within the cell that regulate amino acid production or degradation has been altered.

Upon interpretation of Figure 4.7 it was considered difficult to form solid conclusions as the data provides a far more global overview of cellular metabolic perturbations than that required to understand the implications of the presence or absence of Rnq1p. Given the complexity of cellular metabolism, a slow, step by step analysis would allow for better interpretation and understanding of the results.

Figure 4.7. gives an overview of metabolic regions that had been perturbed but lacks information about the direction of these changes. Results were then visualised within their

samples groups, *Δrnq1* and [*rnq*$^-$], in Figure 4.8 and are shown by metabolic area. By way of comparison, both average relative abundance and logarithmic scales have been used to visualise the relationship of these changes. Generally, it appears that [*rnq*$^-$] samples have higher levels of activity in most metabolic regions (in all but cofactor biosynthesis) with activation/inactivation and interaction giving higher readings in this sample group. Again, the lack of fine detail regarding these changes and the large error bars presented on these graphs suggests that deeper mining of the data is needed to unpick the metabolic regions influenced.

This same data was subsequently uploaded to BioCyc's specific *S. cerevisiae* cellular overview and pathway hits can be seen in Table 4.13. Considerable overlap is seen between the pathways observed and those indicated via Mummichogg are shown in Table 4.10. and 4.12. This consolidation of the PIM and NIM data strengthened the argument for the pathways indicated, allowing for greater confidence in these findings. Pathways are ranked by DPPS score (a measure of the pathway perturbation within multiple data sets), and all of the pathways listed gave scores over 100,000 strongly indicating their perturbation.

Table 4.13. Top Ten most perturbed pathways between *Δrnq1* and [*rnq*$^-$].

Pathway Name
ubiquinol-6 biosynthesis from 4-hydroxybenzoate (eukaryotic)
ubiquinol-6 bypass biosynthesis
superpathway of chorismate metabolism
superpathway of ubiquinol-6 biosynthesis
inosine-5'-phosphate biosynthesis II
tryptophan biosynthesis
formaldehyde oxidation II (glutathione-dependent)
phospholipids degradation
urea degradation
histidine biosynthesis

When comparing the most significant *Δrnq1* relative abundances to the most significant [*rnq*$^-$] relative abundance, via BioCyc's specific *S. cerevisiae* cellular overview.

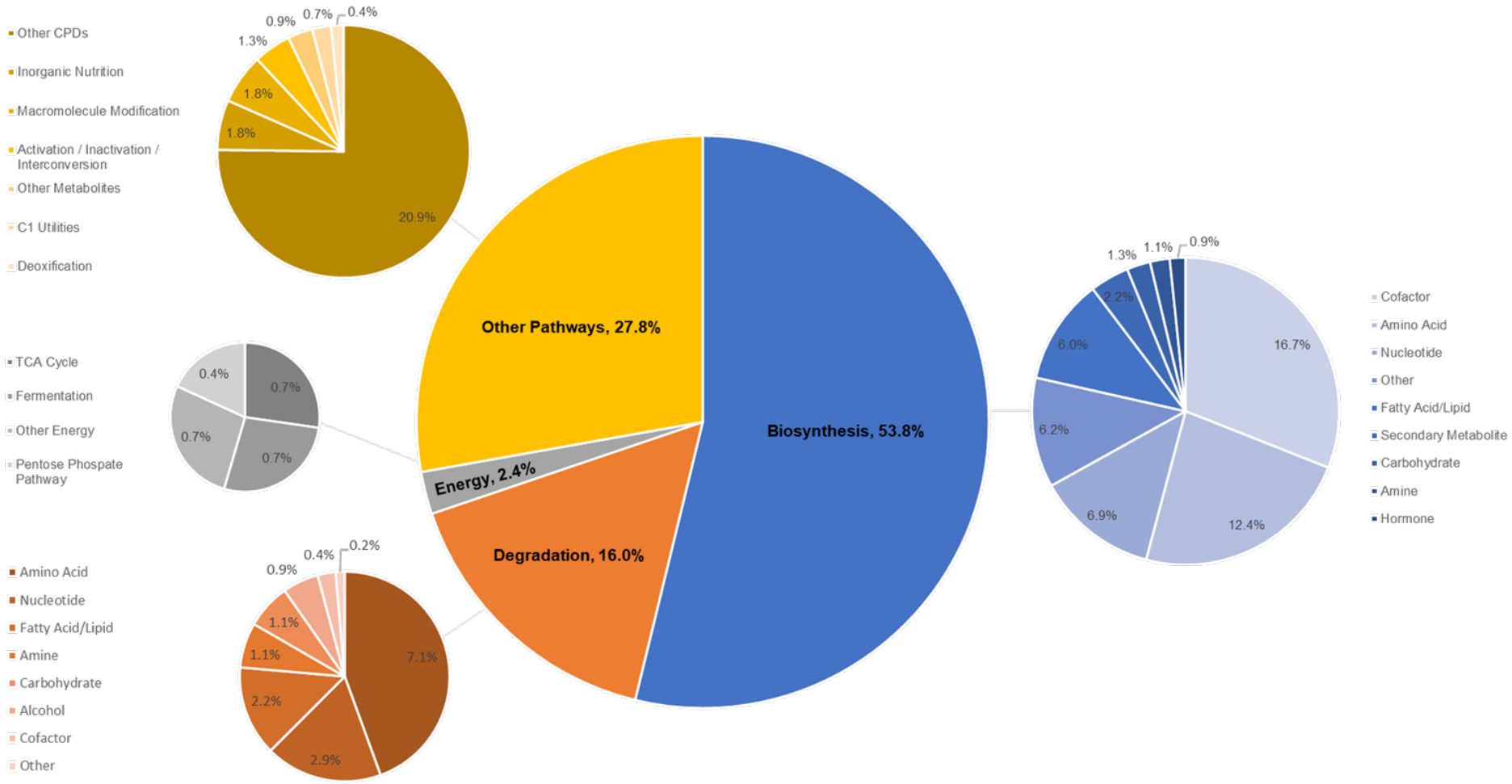


Figure 4.7. Graphical representation of the metabolic perturbations detailed by region when comparing *Δrnq1* and [*rnq*⁻] samples. The largest pie chart shows a general overview of the key areas detected as changed between groups and the percentage of the change assigned to those areas, as determined via BioCyc's omics dashboard. The smaller pie charts are colour coded with the small blue pie chart showing the detailed metabolic regions (by percentage) that are contributing to the overall biosynthesis percentage seen in the largest pie chart. The orange pie chart provides these details for the degradation percentage, the grey pie chart provides these details for the energy percentage and the yellow pie chart provides these details for the other pathway's percentage.

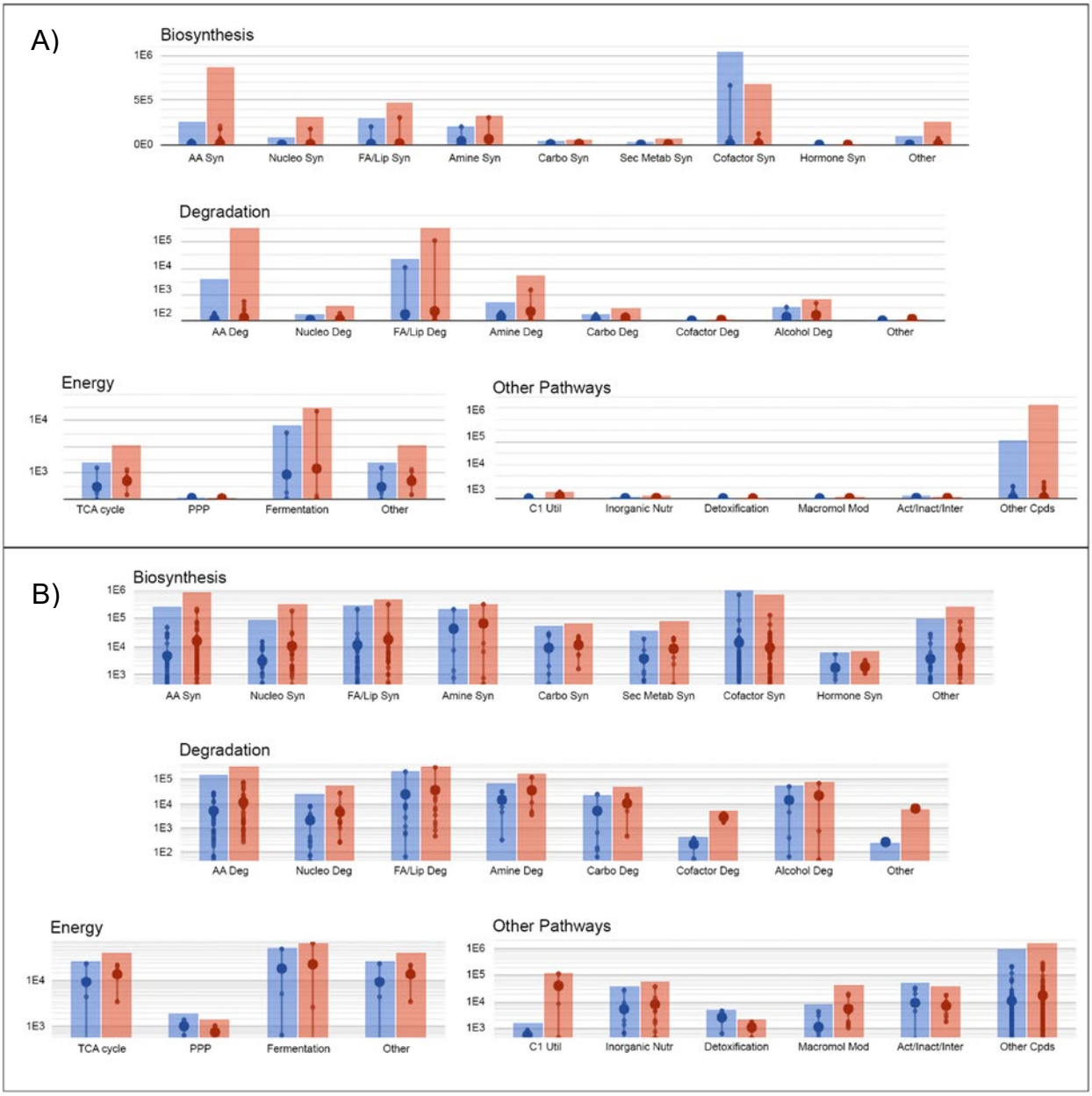


Figure 4.8. Graphical comparisons of the average relative abundances between *Δrnq1* and [*rnq*$^-$] samples within key metabolic regions. A) Charts average comparisons of the relative abundances between *Δrnq1* samples (shown in blue) and [*rnq*$^-$] (shown in red), data is first separated by key metabolic region (shown as a graph title) and then by specific metabolic area (shown on the x axis). The Y axis is set to a linear scale, appropriate to the relative abundances. Individual data points are shown as small circles within the larger summated bars. B) Shows the same information with a logarithmic scale on the Y axis.

4.4.6. Data overlay onto implicated pathways, via BioCycs pathway collage, reveals the direction of the up and down regulation between groups

The information provided by Table 4.13. gave specific areas of inquiry to investigate and the decision was made to overlay the raw data from each of the samples onto the specific pathways, to further visualize and interpret findings. To overlay LC-MS data onto highlighted pathways within BioCyc, relative abundances were subject to standardization in order to adjust the scale of the changes relative to the metabolite in question. This was achieved via Microsoft Excels STANDARDIZE function, using the raw relative abundance of the sample in question, the average mean of all data, and the standard deviation of all data. This provided a normalized value known as a z-score (or standard score) which represents the number of standard deviations a given data point is from the mean. Positive z-scores indicate a value greater than the mean and negative scores indicate a value less than the mean. This allowed for a visually accessible directionality that would have otherwise been unobtainable, given the large range of values that were present among the relative abundances. It is imperative when engaging in this type of pathway analysis, that there is a continuous relation of our findings to existing biological knowledge and the conditions tested and that caution is implemented given the opportunity for misinterpretation at this level of cellular analysis.

All of the ubiquinol pathways shown on Table 4.13 were grouped together with standardized data overlaid onto them (shown in Figure 4.9.). Here a consistent upregulation of these pathways in *Δrnq1* samples, and/or a downregulation in [*rnq*$^{-}$] samples is seen. This pattern can be seen throughout the entirety of the super-pathway of ubiquinol-6 biosynthesis, the ubiquinol-6 biosynthesis from 4-hydroxybenzoate (eukaryotic) and the ubiquinol-6 bypass biosynthesis.

The overlay of the other pathways listed in Table 4.13 do not make for such simple analysis. The chorismate biosynthesis pathway feeds directly into the tryptophan pathway and so finding both identified is encouraging. However, the linking metabolites occurring at the end and start of these pathways respectively appear to give variable results (shown in Figure 4.10.). Initially the chorismate pathway begins upregulated (in the absence of Rnq1p) but this quickly shifts to a relative downregulation. As stated, the pattern is somewhat lost during the connecting nodes of these pathways but appears to be steadily downregulated (in the absence of Rnq1p) throughout the remainder of the tryptophan pathway. It is possible that the overall measure of chorismate within the cell (the measure used here) is highly variable,

given that chorismate is involved in a range of other metabolic pathways and may obscure the true value for this single metabolite in this particular pathway. It is also possible that any one m/z value may be misappropriated and so assessing the trends present across pathways and among their connections should be considered a more reliable source of information when addressing our aims than individual metabolite identities alone. Focusing on a single metabolite change, is ill advised, as the likelihood of a single mislabelled metabolite is high, whereas the likelihood of the unbiased labelling of all of the metabolites within a single pathway all showing the same directionality is low.

The urea degradation and formaldehyde oxidation II pathways (shown in Figure 4.10.) both show clear indications of down regulation (in the absence of Rnq1p). Much like the inosine pathway (shown in Figure 4.10.) it must be acknowledged that these pathways are small and so without further targeted experiments aimed at these metabolites' conclusions drawn on this basis alone would be prone to error. Likewise, the same is true of the phospholipid degradation pathway (shown in Figure 4.11.) where clear evidence for the downregulation of choline and phosphocholine is seen. Although inherently linked to this pathway, there is not a metabolite within it and so it is entirely possible that this pathway has been considered significant due to its relative size and that the influence of the choline and phosphocholine changes have been misappropriated. Early in the histidine pathway (shown in Figure 4.10.) there is a clear trend in downregulation in the absence of Rnq1p. This trend becomes more convoluted at the branches of this pathway and may be due to further mitigating connections that dilute or counteract the overall trend observed. Further targeted investigation is needed to elucidate on this finding.

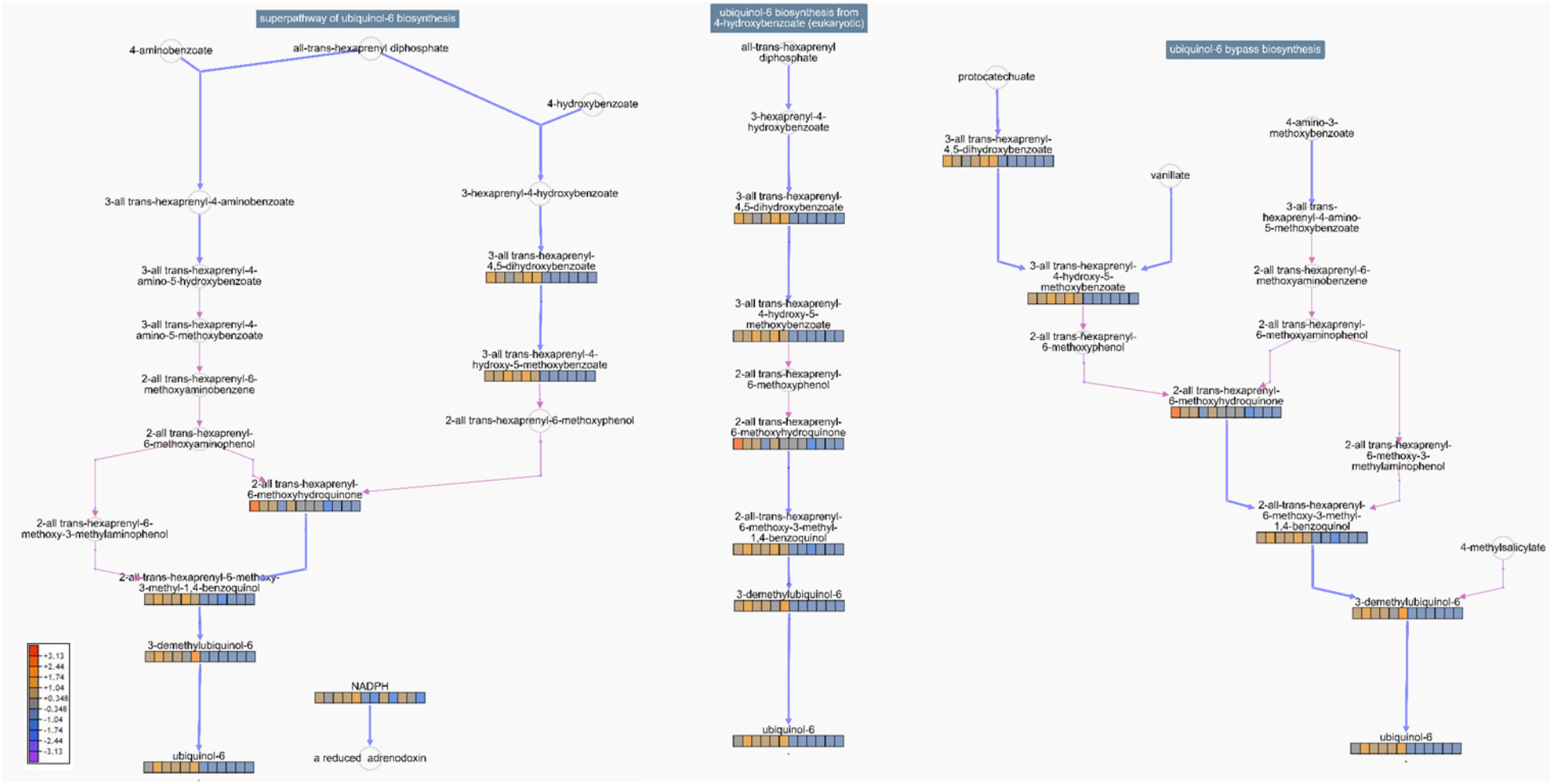


Figure 4.9. Pathway collage of three of the top 10 implicated pathways. (via BioCycs cellular overview) with standardised omics *Δrnq1* and [*rnq*$^-$] data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^-$] samples. A larger version of this Figure is available in Appendix G.

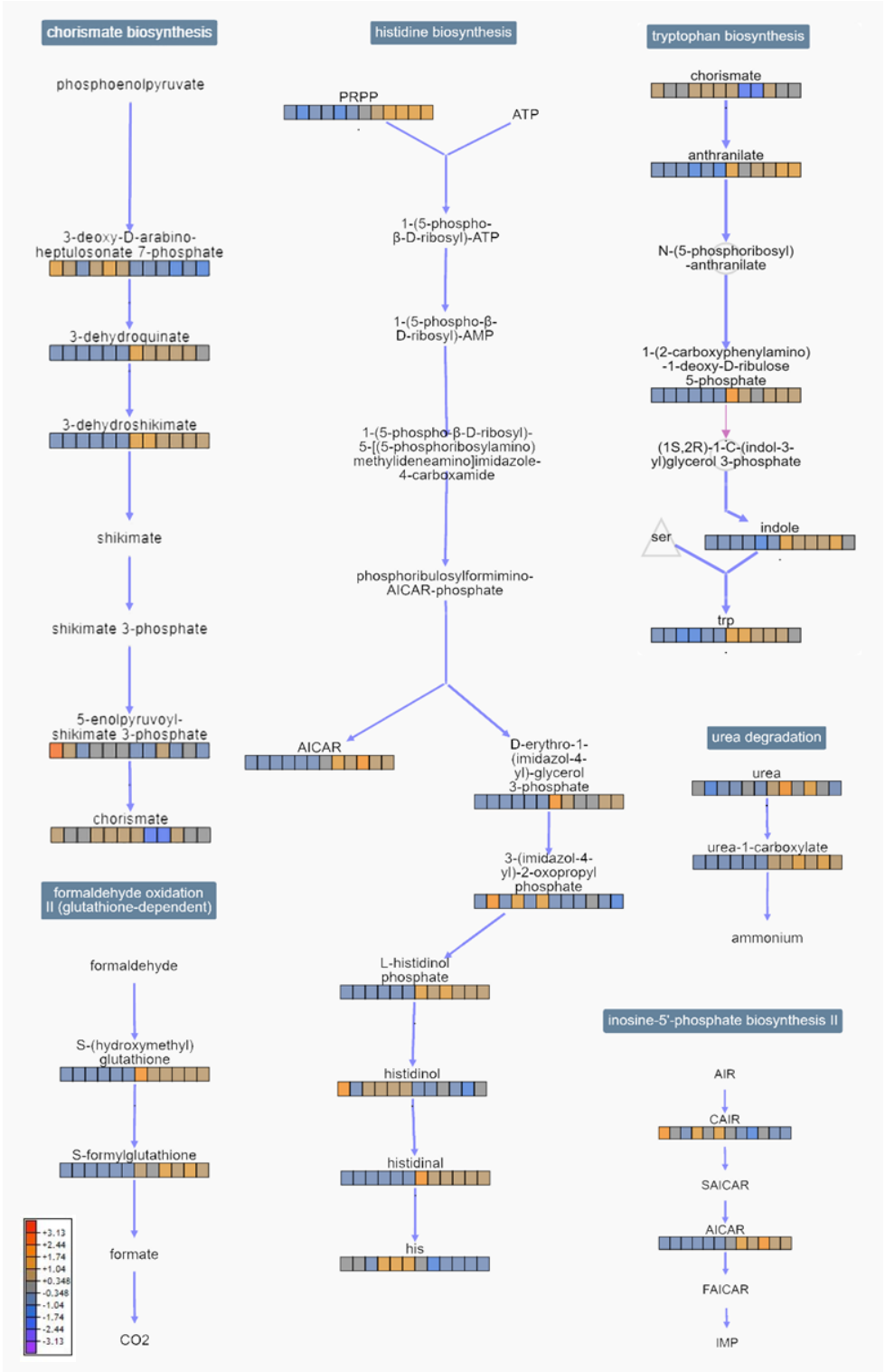


Figure 4.10. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics *Δrnq1* and [*rnq*$^{-}$] data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples. A larger version of this Figure is available in Appendix G.

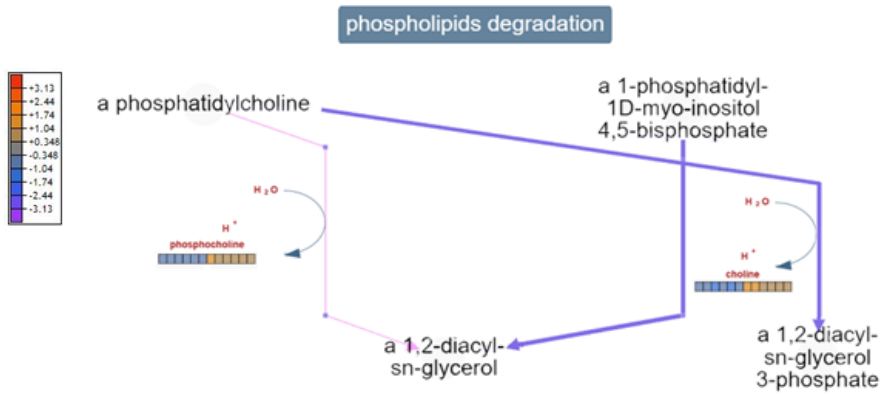


Figure 4.11. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics *Δrnq1* and [*rnq*$^-$] data of individual metabolites overlaid. A colour legend is shown in the middle left to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^-$] samples A larger version of this Figure is available in Appendix G.

4.4.7. Examination of the upstream connections of the ubiquinol pathways, via BioCycs pathway collage, reveals other potential interactions of Rnq1p

The identification of the ubiquinol pathway as being perturbed (shown in Figure 4.9.) is interesting but presents issues. Rnq1p, is known to reside within the cytoplasm (Huh *et al.* 2003) and confirmatory bioinformatic tests using PSORT (https://wolfpsort.hgc.jp/) have shown that the likelihood of this Rnq1p to be localised to the mitochondria as very low. Given that the ubiquinol pathways largely operate within the mitochondria this leads to the conclusion that observations are the downstream effect of an earlier change. Prior to investigating these pathways upstream, there was a desire to rule out the possibility of alternate influencers. The only known influencer other than those pathways inherently linked to ubiquinol pathways is glucose depression. Szkopinska (2000) states that glucose depression in yeast increases the level of ubiquinone-6 synthesised. To ensure that this was not in operation here, glycolysis, glycogen biosynthesis, glycogen degradation and glucosamine biosynthesis pathways were overlaid with the same standardised data (shown in Figure 4.12.). Given that this data set only contained the most significant findings, only one metabolite hit was shown. This provided substantial evidence to rule out glucose depression as being a mitigating issue. In addition, given the presence of pyruvate, (a metabolite that is often falsely identified as a positive hit due to its repetition in many

metabolic pathways), a further analysis was performed under the same conditions, but encouragingly no significant hits were found.

Ubiquinone pathways are known to be upregulated upon various types of oxidative stress and are downregulated with age. This is intriguing given the known relationship between Rnq1p and oxidative stress (Chiti and Dobson, 2006; Szkopinska, 2000). Ubiquinone pathways were indicated as most perturbed (Table 4.13) and considering the cellular location of Rnq1p, the decision was taken to trace the source of these metabolic pathways. Both the superpathway of ubiquinol-6 biosynthesis and the ubiquinol-6 biosynthesis from 4-hydroxybenzoate (eukaryotic) feed into the hexaprenyl diphosphate biosynthesis pathway. This in turn, links into the superpathway of geranylgeranyldiphosphate biosynthesis I (via mevalonate) which initiates at the pyruvate dehydrogenase complex. The standardised data was overlaid onto all the pathways listed (shown in Figure 4.13.) and a clear downregulation trend (in the absence of Rnq1p) is seen in the hexaprenyl diphosphate biosynthesis pathway. Within the geranylgeranyldiphosphate biosynthesis I (via mevalonate) the picture is a little more complex. The pathway terminates with a clear downregulation trend (in the absence of Rnq1p) but between DMAPP and (R)-mevalonate diphosphate, this appears to switch. Unfortunately, due to either lack of measurement/ID or the lack of significant hit data for this metabolite, information for IPP was not available.

Above this point in the pathway, the downregulation trend switches direction, although limited information is available for early stages of this pathway. It was considered that the influence seen within the ubiquinone pathways may well be an area particularly sensitive to change. However, following the downstream effect of the changes observed in the geranylgeranyldiphosphate biosynthesis I (via mevalonate) pathway may well point to other key areas of perturbation. Figure 4.14. reveals the connection between this pathway and the superpathway of ergosterol biosynthesis. Overlaying our data onto this pathway shows interesting findings, but we must note that our experimental methodology (untargeted and metabolite focused in nature) does not have the specificity to have reliably extracted and measured the type of relatively large lipid molecules that are found towards the latter end of this pathway.

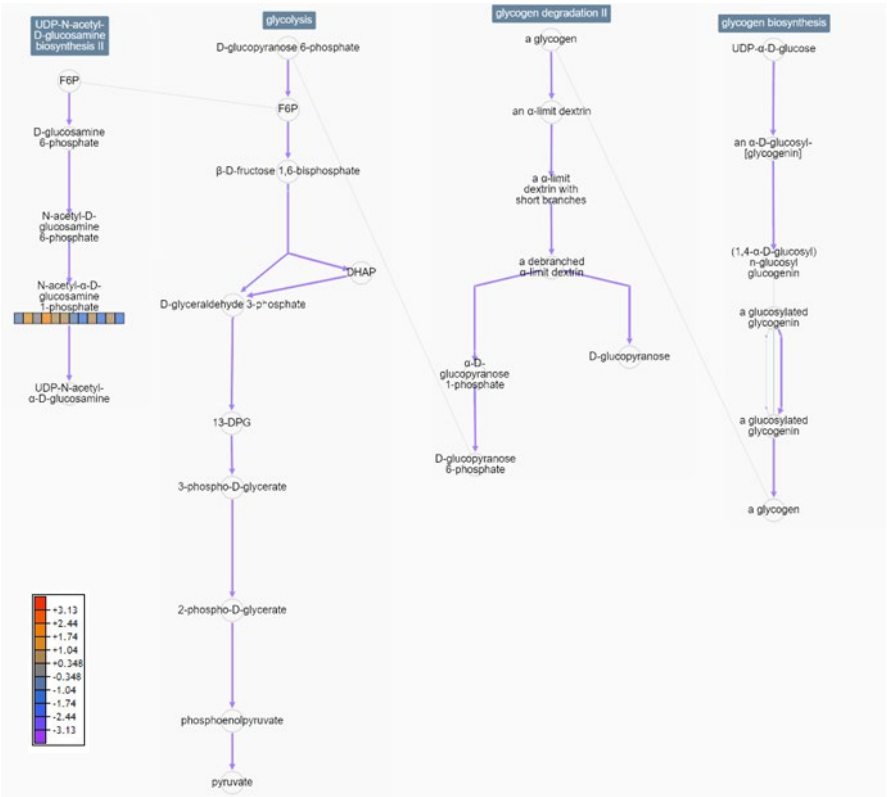


Figure 4.12. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics *Δrnq1* and [*rnq*$^-$] data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^-$] samples. A larger version of this Figure is available in Appendix G.

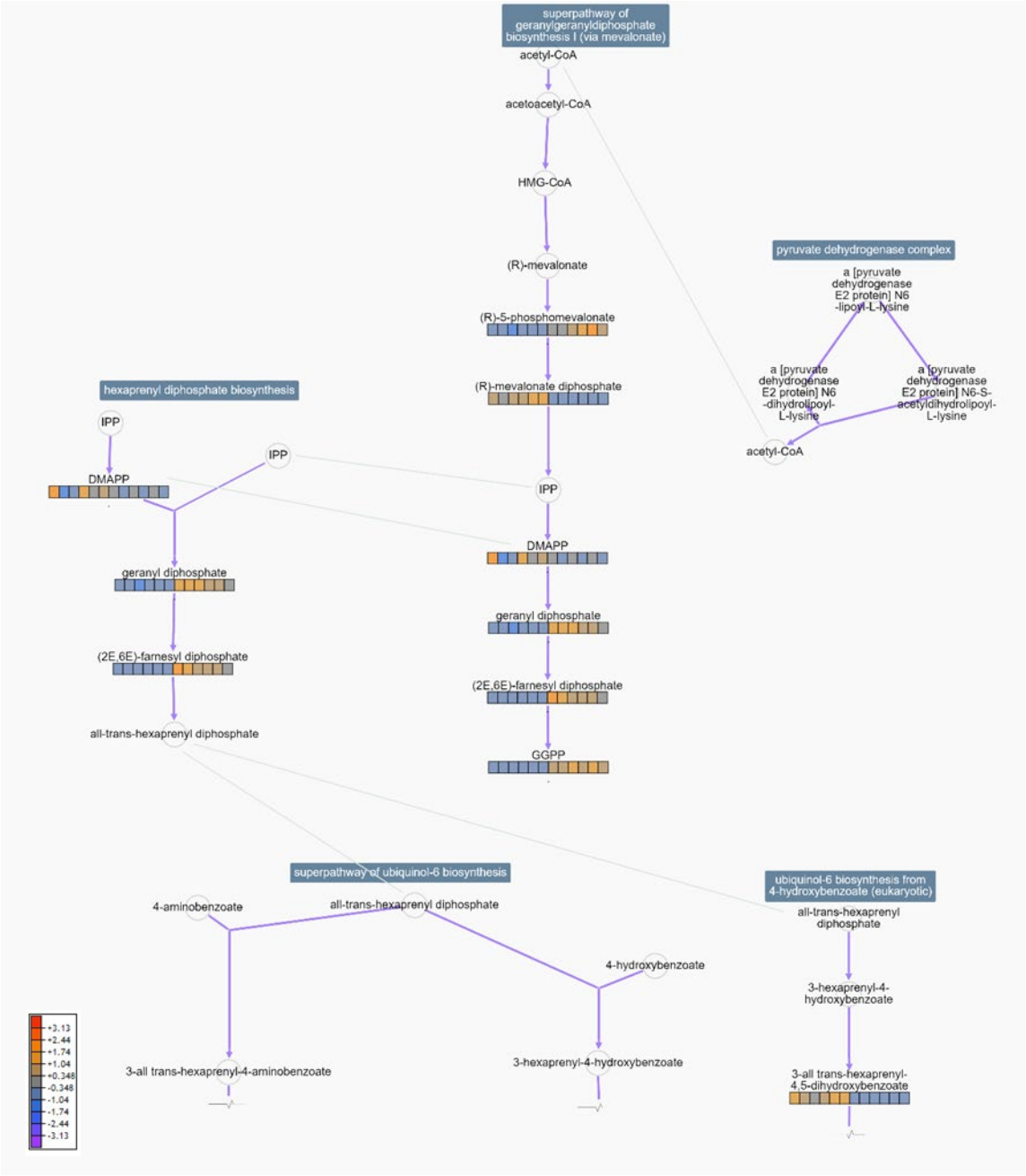


Figure 4.13. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics *Δrnq1* and [*rnq*$^{-}$] data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples. A larger version of this Figure is available in Appendix G.

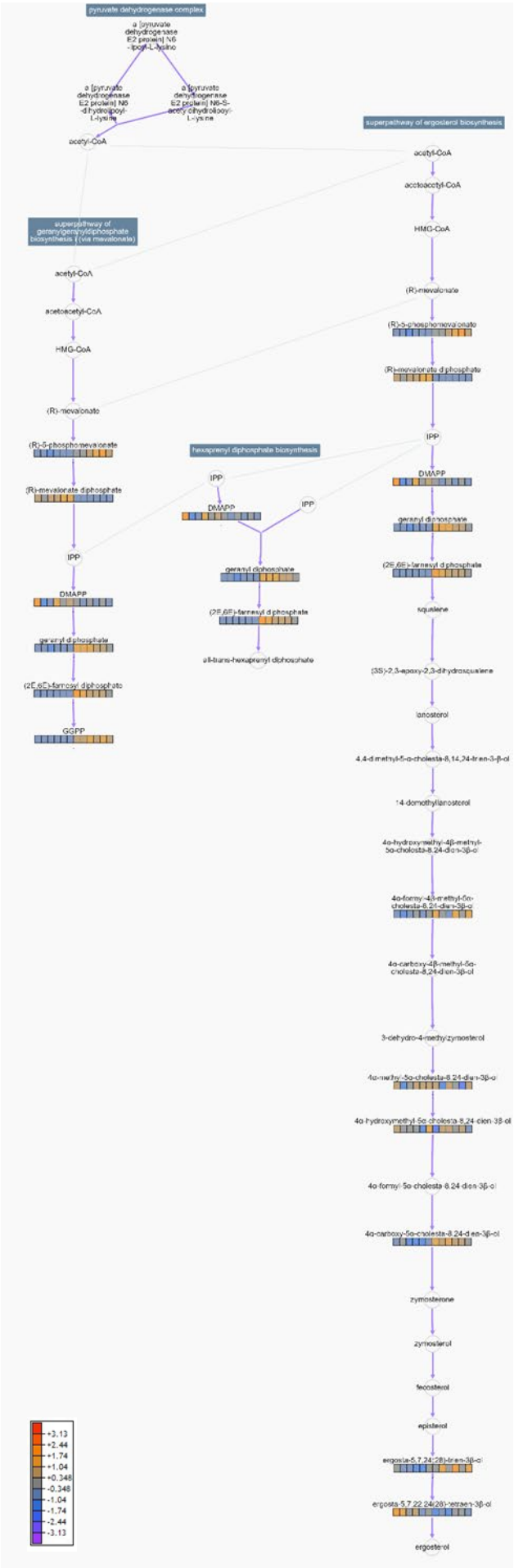


Figure 4.14. Pathway collage of metabolic implications upstream from three of the top 10 implicated pathways with standardised omics *Δrnq1* and [*rnq*$^-$] data of individual metabolites overlaid. A colour legend is shown to the centre left of the Figure to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^-$] samples A larger version of this Figure is available in Appendix G.

4.5. Discussion

The results of the analysis of the data as documented in section 4.2. and 4.3. confirms the assertions of Myers *et al.* (2017) that the variance between metabolomics analysis programs can be considerable. The use of cross comparison of these two programs, as outlined by Li *et al.* (2019), has proved to be a valuable addition to our bioinformatic methodology and one which will be continued for subsequent analyses in order to improve confidence in our data and robustness to our conclusions. Encouragingly, Simca analysis found significant differences between the *Δrnq1* samples and [*rnq*$^-$] samples, enabling analysis in Metaboanalyst and BioCyc. Observations on this scale can at first be difficult to interpret due the complexity and extent of the information available. However, by combining PIM and NIM results within BioCyc has allowed for strong and reliable DPPS scores that have added an extra dimension of stringency and validating our findings.

4.5.1. Exploring the Ubiquinol pathways, searching for Rnq1p influences

Our analyses strongly indicate that perturbations within the ubiquinol 6 biosynthetic pathways and their connecting pathways maybe key to elucidating Rnq1p's cellular role. Ubiquinol-6 pathways show a relative down regulation when Rnq1p is present. However, as this is a mitochondrial based pathway and Rnq1p is cytoplasmic localised, further investigation into this pathway and its interactions was sought.

Ubiquinol-6 is a mitochondrial pathway whose enzymatic components consist of the Coenzyme Q group of enzymes. Indeed, the pathways listed in Figure 4.10 involve the function of COQ2, COQ3, COQ5, COQ6 and COQ7 (BIOCYC database, 2019). A search of the literature regarding this pathway and enzymes suggest that the quinonoid nucleus of the COQ's is derived from the shikimate pathway via chorismate within bacteria and via tyrosine with the higher eukaryotes (Meganathan, 2001; Olsen and Rudney, 1983). For information regarding COQ nomenclature please see Stefely and Pagliarini (2018) comprehensive paper. In *S. cerevisiae* if the shikimate pathway is present (i.e. not a knockout strain) cells can derive tyrosine from the chorismate pathway *in vivo* and studies have shown that wild-type mutants favoured method of deriving Q is via the chorismate intermediate (Meganathan, 2001; Olsen and Rudney, 1983).

Given that in our analyses the chorismate pathway has also been identified as significantly perturbed, this area was investigated further. These investigations did not provide definitive

evidence of up or downregulation within the chorismate pathway with no data available concerning shikimate metabolites. The DPPS score of 287000 given by Biocyc of the super pathway of chorismate is indicative of the entire pathway's perturbation and clearly indicates an upregulation within this pathway in the absence of Rnq1p.

The upregulation in ubiquinol-6 pathways in the absence of Rnq1p may be due to an increased availability in the component parts for the Coenzyme Q enzymes. However, the rate limiting steps and the set of reactions for the biosynthesis and increased availability of the component metabolites (those which are incorporated within the enzymes) are not currently known (Awad *et al.* 2018; Stefely and Pagliarini, 2018; Gonzalez-Mariscal *et al.* 2014; Meganathan, 2001). Further research using UHPLC-MS to quantify the relative availability of the COQs and their redox state within *Δrnq1* and [*rnq*$^{-}$] cells is therefore advised.

Ubiquinone's (coenzyme Q) primary role is as a component of the mitochondrial respiratory chain but in its reduced form it also acts as an antioxidant, protecting membrane phospholipids and lipoproteins from lipid peroxidation. It also has a role in protecting other cell components from free radical-induced oxidative damage (Awad *et al.* 2018; Szkopinska, 2000; Genova *et al.* 1999; Mikosovska *et al.* 1999). These associated pathways are known to be upregulated upon various types of oxidative stress and downregulated with age (Szkopinska, 2000). This regulation is actioned via the mevalonate pathway, where there is a delicate interplay between the major biosynthetic products of mevalonate metabolism. The precise details as to how this interplay is controlled and monitored (and hence how the subsequent up and down regulation of these metabolites is achieved) is, as yet, unknown (Mohamed *et al.* 2015; Szkopinska, 2000; Szkopinska *et al.* 2000; Grabowska *et al.* 1998).

What is known is that the biosynthesis of these Coenzyme Q enzymes and the activity of these pathways are regulated to match these two key cellular processes: bioenergetic metabolism and antioxidant defence (González-Mariscal *et al.* 2014). Assuming that no bioenergetic change has occurred (Figure 4.13.), we can conclude that antioxidant defence is key to the role of Rnq1p: when Rnq1p is present, ubiquinone production is low and when Rnq1p is not present, ubiquinone production is high. This suggests that Rnq1p is a negative controller of ubiquinone production.

4.5.2. The biological function of Rnq1p as a regulator of ubiquinone production and suggestions for further research to confirm this role

How Rnq1p acts upon these pathways is still unclear due to its known cellular location. Given that ubiquinone production is regulated via the mevalonate pathways and so too are dolichol and sterol, the action of Rnq1p on these pathways may occur further upstream of the ubiquinol pathways (Mohamed et al. 2015; Szkopinska, 2000). Within the hexaprenyl pathway a relative upregulation in the presence of Rnq1p is observed. This indicates some disconnect between the hexaprenyl pathway, and both, the superpathway of geranylgeranyldiphosphate biosynthesis I (via mevalonate) and the super pathway of ergosterol biosynthesis, that appears to only be occurring in the presence of Rnq1p. The Rnq1p is cytoplasmic, and curiously this connection between the ubiquinol and hexaprenyl pathways is a connection from the cytoplasm into the mitochondria. The observed restriction of the downstream flow of all-trans-hexaprenyl diphosphate into the ubiquinol pathways in the presence of Rnq1p may provide evidence of a cellular role for Rnq1p in limiting the trafficking of these metabolites to the mitochondria. The observed upregulation of the hexaprenyl and connected pathways may be a consequence of this halt in the downstream connection, allowing these metabolites to build up within cells. However, overlay of metabolic data shown in Figures 4.14 and 4.15 demonstrate that the observed upregulation may have diverted these resources to the ergosterol superpathway, although evidence for this is limiting due the nature of experimentation.

The yeast homolog of the human lipid cholesterol, ergosterol has long been considered to play a role in the regulation of membrane fluidity and structure. The most abundant sterol in fungal cell membranes, ergosterol is thought to participate in cytoprotective roles, interkingdom interactions and immune response (Rodrigues, 2018; Koselny *et al.* 2018). Despite this already large list of the beneficial roles, new biologically relevant roles for this lipid are still being discovered (Rodrigues, 2018; Koselny *et al.* 2018). Of relevance to this study is the use of this sterol as a target for many of the antifungal treatments (Rodrigues, 2018) and a suggestion for future research could be to compare the effectiveness of antifungals such as the azoles on the strains used within this study. A prediction would be that strains containing native, soluble Rnq1p ([*rnq*$^{-}$]) would exhibit a higher azole sensitivity due to their apparent higher concentration of ergosterols. Conversely strains where Rnq1p is absent (Δ*rnq1* and potentially [*RNQ*$^{+}$]) may exhibit a lower azole sensitivity.

The sterol composition of membranes is known to have a strong correlation with the tolerance of organisms to a variety of stressors (Bhattacharya et al. 2018; Lui et al. 2017; Mohamed et al. 2015). However, there is a systematic lack in understanding among the relationship between cellular resistance and sterol composition (Bhattacharya *et al.* 2018; Lui *et al.* 2017). Various modifications and recombinantly engineered enzymes have been produced within these pathways in an attempt to enhance stress resistance, but with variable results (Bhattacharya *et al.* 2018; Lui *et al.* 2017). Interestingly, both Bhattacharya *et al.* (2018) and Lui *et al.* (2017) note that duration of lag growth phase can be significantly altered due to changes in sterol composition. During the course of this research it has routinely been found that starter cultures of [*rnq*$^-$] have a considerably longer lag growth phase than the other strains tested. This observation has not been previously reported in the literature but may be indicative of the role of Rnq1p. Additional studies of the lag phases of Rnq1p strains may provide evidence for a link to sterol metabolism supported by targeted lipidomic studies focusing on sterol composition and quantification.

4.5.3. Limitations of this study and suggestions for further research

Initially there was some concern about the impact that this variability in lag phase would have on the growth rate and cell number of our two biological classes, and subsequently the metabolic perturbations observed. To ensure parity between different strains at the time of sampling, the total protein content was to be determined as an approximation of metabolite concentration (Tredwell *et al.* 2011) and cell counting using a haemocytometer were considered. However, due to the time sensitivity of quenching the samples, the time-intensive nature of cell counting and the lack of availability of instrumentation for rapid protein quantification, this was unrealistic. In addition, considering that experimental evidence for a general normalisation of metabolite concentration to cell number or protein concentration appeared to be lacking (Muschet *et al.* 2016), neither techniques were added to the experiments conducted.

Instead, all cells were harvested at the same OD (0.6 OD_{600}), indicative of exponential growth and cell number in yeast. Despite our early observations, no notable growth effects were reported for the strains used and indeed, we did not see any growth effects outside of the lag phase observation reported for [*rnq*$^-$]. These observations support the assumption that growth rate and cell number was consistent between the two classes. For future experimentation, it may be advisable to monitor the growth rate of the strains independently alongside metabolomic experimentation and so ensure growth rate is not a variable. An

alternative solution would be the inclusion of a fluorescence-based DNA quantification of metabolite harvested cells. Muschet *et al.* (2016) suggest this method as fast, sensitive, and robust and demonstrate that DNA quantification, unlike cell or protein quantification, is indicative of metabolite concentration (for 82-97% of metabolites).

The most significant and uncontrollable factor within this research is the bottleneck between features detected and the features identified (Monge *et al.* 2019). Although regrettable, the impact of this is unknown. More feature identity may have led to other undiscovered pathway perturbations being revealed. Alternatively, it may have added valuable information about uncharacterised metabolites within the pathways highlighted. Given that this is a factor in any metabolomic research, metabolite libraries are constantly evolving and being added too (Monge *et al.* 2019). However, for this work to continue, demand from researchers must exist. Therefore, conducting metabolomic experiments, even with the knowledge that some information will be missed, is preferable as new perturbation information will be gained and the demand and work on improvements will continue. Specifically, within this research, greater feature ID would have been possible using Progenesis, or any other platform that would have granted access to the higher-energy collision data. Such data, containing valuable fragmentation information, was acquired and stored but was not accessible throughout this research.

4.5.4. Findings and implications

This work has revealed that Rnq1p may play a lipid/mevalonate based cytoprotective role as a regulator of ubiquinone production. Present in wild populations (Halfmann *et al.* 2012; Kelly *et al.* 2012a; Nakayashiki *et al.* 2005), the native conformation of tRnq1p in times of limited stress interrupts/interferes with the regulation of ubiquinol 6, having a relative downregulation effect. This down regulation effect on ubiquinol-6 masks the cells capacity to produce high levels of ubiquinone. It is suggested that these cellular resources are diverted to the production of sterols, upregulating the production of these more long-term membrane bound cytoprotective lipids. Rnq1p however is intrinsically disordered and so when acted upon by a stress condition, will readily misfold into its prion form, behaving much like a switch. Conversion to the prion form downregulated the production of membrane bound cytoprotective lipids and allows for a fast and effective upregulation of the ubiquinol pathways which needed for intercellular antioxidant defence. Mechanistically how this is achieved is yet unclear, although further targeted metabolic experimentation aimed at these

pathways may reveal details. In addition, experimentation that investigated these roles biochemically would add evidence to these theories and direct future studies.

If confirmed, the implications of these findings may well be far reaching. As of yet, no DNA sequence homolog has currently been identified for Rnq1p but the potential for another intrinsically disordered protein to play a similar role within mammalian cells is likely. Even without the identification of such a protein, the switch from lipid to ubiquinol biosynthesis occurring in the presence of misfolding proteins as part of a cellular response against stress, would provide an opportunity with which to track and monitor the progress of a related disease using specific biomarkers.

Chapter Five - The metabolic perturbations associated with the presence of the [*RNQ*$^{+}$] prion

This chapter attempts to identify the metabolic pathways that are perturbed in the presence of the [*RNQ*$^{+}$] prion. This is achieved by comparing the metabolites profiles from [*rnq*$^{-}$] and [*RNQ*$^{+}$] cells together with those identified from previous studies involving *Δrnq1* (as detailed in the previous chapter). Comparison with ∆*rnq1* data allows for the removal of 'loss of function' effects observed for Rnq1p. In addition to these biological classes, [*rnq*$^{-}$] cells and [*RNQ*$^{+}$] cells were also subjected to a mild oxidative stress with the metabolite profiles obtained acting as controls as a general stress response within a prion-containing and a prion-free background. These studies will enable the subtraction of a general stress response factor from differences seen between [*rnq*$^{-}$] and [*RNQ*$^{+}$] cells and so identify changes only due to [*RNQ*$^{+}$] prion presence. Experiments were conducted as detailed in chapter two.

5.1. Introduction

The self-propagating and transmissible proteinaceous agents, known as prions, are responsible for causing a host of devastating and fatal degenerative diseases in mammals (Liebman and Chernoff, 2012; Tuite and Serio, 2010). Despite the mechanistic similarity between the prions of both kingdoms, yeast prions are largely benign and so are not associated with a disease phenotype. As a result, yeast prions provide a unique vantage point from which to view the underlying cellular chemistry of prion presence and subsequent amyloid formation and their influence on cellular functioning (Liebman and Chernoff, 2012; Tuite and Serio, 2010).

In these experiments, metabolic comparisons of prion free cells ([*rnq*$^{-}$]) with prion containing cells ([*RNQ*$^{+}$]) were conducted. It was hypothesised that inference towards a pseudo 'non-disease' vs 'disease' comparison may reveal information pertinent to the understanding of amyloid formation and relevant interactions with normal cellular biochemistry, without the experimental restrictions imposed by mammalian cell death. Whilst at first this may seem contradictory, due to the lack of a disease phenotype *per se*, yeast models have long been used to study protein mis-folding and the underlying cellular pathomechanisms thereof (Gregorio and Duennwald, 2018; Wickner *et al.* 2015).

The role of the [*RNQ*$^{+}$] prion as a facilitator to prion formation places this prion in the initial stages of what is a complex and downstream process of protein misfolding. Although experimentation on the much more readily studied [*PSI*$^{+}$] prion would have been possible, it would have been difficult to rule out the implications of the presence of [*RNQ*$^{+}$] prions, even in strains which no longer contained this particular prion (given that [*RNQ*$^{+}$] is required for formation not maintenance of [*PSI*$^{+}$]) (Bradley et al. 2002).

The structure and function of the protein domains of Rnq1p are also an area of interest. Rnq1p-PFD, found within the C-terminal region of the protein, is atypically Q/N rich; however, some discrepancies exist as to the relationship between prion status and other protein domains present within the protein (Stein and True, 2011; Vitrenko *et al.* 2007). The N-terminal domain has been implicated in both positive and negative regulation of [*RNQ*$^{+}$] prion maintenance (Bardill and True, 2010). What is clear is that the Rnq1p-PFD is more complex than other prion counterparts, with loosely defined oligopeptide repeats over four Q/N rich regions. Hence, it is posited that any work which aims to define or elucidate the role of these regions may too be crucial in our understanding of amyloid formation (Stein and True, 2011; Vitrenko *et al.* 2007).

Irrespective of a mechanistic interest in amyloidosis, the [*RNQ*$^{+}$] prion has many other intriguing qualities. Linked to its role in heterologous appearance and the presence of multiple variant conformations in its prion form, it has been argued that [*RNQ*$^{+}$] prions demonstrate a Lamarckian inheritance pattern via prion state switching, which appears to allow access to complex traits in a single generation (Stein and True 2011; Halfmann *et al.* 2010). This switching is capable of inferring both negative and positive survivability outcomes for the yeast cells depending on the stress condition. When considered alongside the discovery of [*RNQ*$^{+}$] prions in wild populations, it is now thought that this function plays a pivotal role in the adaptive abilities of yeast and the evolution of new traits (Halfmann *et al.* 2010). Certainly, this is not without precedent; many other organisms contain amyloids which are known to play a variety of positive biological roles, including protection from stress, silk production, multicellular growth and community development (Upadhyay and Mishra, 2018; Dragos *et al.* 2017; Greenwald and Riek, 2010).

Apart from the protists, prions or amyloids are present throughout all of the kingdoms of life (Upadhyay and Mishra, 2018), with evidence for the presence of amyloid like proteins in over 30 species (Upadhyay and Mishra, 2018). This has increased our understanding of the

prions and amyloids, encouraging further debate about the function of prions within the fungal kingdom. Halfmann *et al.* (2010) argues that prions provide a specific mechanism with which cells can induce phenotypic variability within fluctuating environments. This has been likened to that of an evolutionary capacitor, capable of hiding the effects of genetic polymorphisms. This would allow for the storage or silence of many benign differences between cells, which when acted upon by changes in the environment would then allow for a sudden exposure of these differences. Typically, this presentation of diverse ecological niches leads to natural selection, driving the stabilization of one or more of these variations (Nelson and Masel, 2018; Halfmann *et al.* 2010). Arguably the extent to which prions or amyloids are subject to evolution or have a role to play in natural selection (even in yeast alone) is debatable. It has been demonstrated that these natural phenomena present mechanistic similarities, despite inferring a range of both beneficial and deadly effects and so present a conflicting and intriguing quality as inducers of phenotypic plasticity (Allwein et al. 2019; Nelson and Masel, 2018; Upadhyay and Mishra, 2018).

Although it is possible to report on the distinct strain of [*RNQ*$^{+}$] present (using the system discussed in section 1.1.11), is not ideal, as the level of genetic manipulation to induce this assay may well inhibit or mask the effect of [*RNQ*$^{+}$] (Stein and True, 2011). Hence an understanding of the role of the metabolic perturbations as a result of the presence of the [*RNQ*$^{+}$] prion as carried out in this study may point to a biochemical test that may be indicative of the strain strength. Such information could potentially inform a valuable new biomarker or biochemical assay, or even a phenotypic assay, via genetic manipulation when armed with this information.

The use of metabolomic studies as a tool to investigate amyloidosis and prion formation is not without precedent. Many human blood and plasma studies have focused on biomarker recognition, largely for Alzheimer's, as well as Parkinson's and other well-known neurodegenerative disorders (Jiang *et al.* 2019; Pena-Bautista *et al.* 2019; Varma *et al.* 2018; Phelan *et al.* 2017; Trushina and Mielke, 2014). The rigour of the studies is improving all the time aided by the ever-increasing availability of identifiable metabolites within the Human Metabolome Database. Despite this potential, metabolomics studies remain largely hindered by the time scales required to conduct such experimentation and a general inconsistency between studies. In a recent review of pooled data, Jiang *et al.* (2019) found that there were no metabolites that were consistently identified in all studies, but they did note that lipids and amino acids clearly played a pivotal role in the progression of dementia,

calling for a standardisation within experimentation and the routine use of larger more diverse participant groups to address this variation. It is noteworthy that the acknowledgment of models outside of the relatively small group of mammals that are typically used in this type of experimentation is lacking, even among those who appear to actively encourage multidisciplinary approaches (Brandner and Jaunmuktane, 2017). Narayan *et al.* (2014) argue that the shortfalls of prion or amyloid research that focuses on the higher organisms, results from a general lack of understanding regarding the underlying disease biology, highlighting the pivotal role that relatively biochemically simple model organisms, such as yeast, can have in unravelling such complexities.

It was proposed that metabolic comparisons of [*rnq*$^{-}$] and [*RNQ*$^{+}$], may identify biomarkers and/or metabolic pathways that were indicative of prion presence (detailed in Table 5.1.). In an effort to ensure that the observed metabolic changes were specific for prion presence rather than simply being indicative of a general cellular stress response, the experimentation included the induction of an atypical stress response in *S. cerevisiae* as a control. Previous stress response studies in *S. cerevisiae* suggest that a near universal metabolic stress response is seen via heat shock, oxidative stress and high pH (Kang *et al.* 2012). Given that the other stress responses tested appeared to give a variety of results and the known biochemical associations between heat/oxidative stress and amyloidosis the decision to use a mild oxidative stress in our control [*rnq*$^{-}$] was taken (detailed in Table 5.1.). This detracted from atypical stress response influencers, ensuring that our results reflected the pseudo-disease state in question, without stressing cells enough to induce protein misfolding.

As a matter of curiosity, a mild-oxidative stress response was also induced in [*RNQ*$^{+}$] cells to investigate if or how, metabolically, the presence of the prion effected the cells ability to cope with this stress (detailed in Table 5.1.). Consideration was also be given to the loss of function of Rnq1p so as not to confuse the effect of the prion itself with this loss of function. Although these factors are inherently linked, extrapolation of the metabolic effects of loss of protein to our understanding of amyloidosis as a whole may be hindered without ample consideration of this factor.

Table 5.1. Experimental sample groups used in chapter five.

Sample Group	Prion Status	Stress	Function	Purpose
[*rnq*$^-$]	-	n/a	Negative Control	Establish metabolite baseline
[*rnq*$^-$] with mild oxidative stress	-	H_2O_2 (0.2mM)	Positive Control	Detect metabolite changes due to stress Establish baseline (when comparing to [*RNQ*$^+$] stress)
[*RNQ*$^+$]	+	n/a	Test Group	Detect metabolite changes due to prion presence
[*RNQ*$^+$] with mild oxidative stress	+	H_2O_2 (0.2mM)	Additional Test Group	Detect metabolite changes due to prion presence

Their name, prion status, function, and details of their intended experimental role of those conditions.

5.2. Results – All samples

5.2.1. Feature detection and normalisation

As instructed from the literature and described in chapter 4, cross correlation of m/z values from XCMS and MZmine2 was achieved via the methods outlined in section 2.8.4. PIM data reported 4898 m/z values and NIM data reported 3196 by both programmes (Figure 5.1.(A+B)). Comparative analysis was conducted as detailed in section 4.5. and overlapping features were used to build statistical models via SIMCA and perform pathway analysis via Metaboanalyst and BioCyc.

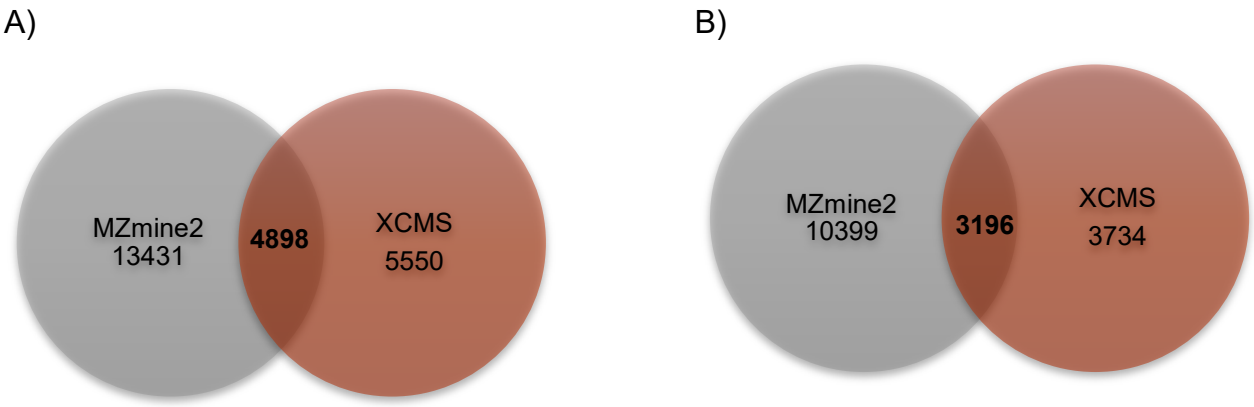


Figure 5.1. Venn diagrams showing the total number of features reported by MZmine2 and the total number of features reported by XCMS when considering [*rnq*$^-$], [*rnq*$^-$] with a mild oxidative stress, [*RNQ*$^+$] and [*RNQ*$^+$] with a mild oxidative stress data. Mzmine2 features are shown in grey and XCMS features are shown in red, the overlapping section in the centre of the two depicts the total number of m/z values reported by both analysis programmes in bold A) Venn diagram of PIM m/z values of MZmine2 and XCMS B) Venn diagram of NIM m/z values of MZmine2 and XCMS

5.2.2. PCA and PLSDA modelling of comparative metabolomic data in SIMCA differentiates and shows relationships between groups

All 24 samples were used to build initial PCA and PLS-DA models, consisting of 6 samples from each of the sample groups: [*rnq*$^-$], [*rnq*$^-$] with mild oxidative stress, [*RNQ*$^+$] and [*RNQ*$^+$] with mild oxidative stress. This would allow us to interrogate the connectivity and provide an insight into the relationship between our 'non-disease' ([*rnq*$^-$]) and 'disease' phenotype ([*RNQ*$^+$]) as well as the stresses tested.

PIM data gave a PCA model with good confidence levels, providing acceptable R^2 and Q^2 values of 0.718 and 0.509 respectively. The NIM data demonstrated less confidence with the PCA model reporting an acceptable R^2 value of 0.667 but a lower Q^2 value of 0.378. This indicated that the predictivity of the model may not be reliable but was deemed to provide fair confidence levels overall given the number of variables present within the data.

Visual assessment of these models (Figure 5.2.) revealed that the strongest separation components of the model, t[1] vs t[2], for both PIM and NIM data showed distinct clustering of all sample's groups. This was slightly more evident in PIM data (Figure 5.2. A and B). Very strong separation was observed between prion-containing and prion-free samples, however one of the [*rnq*$^-$] samples sat outside of the 95% confidence region. Interestingly, both of the stress sample groups appeared to cluster together indicating a closer relationship to each other than to either of the prion 'status' sample groups, that both appeared somewhere in the middle of this variation.

Observations between the model components, t[1] vs t[3] and t[1] vs t[4], reinforced the t[1] separation of the [*RNQ*$^+$] samples from the rest of the groups with varying levels of separation seen between these other groupings (Figure 5.2. C-F). This was evidence that the strongest source of variation appeared to be dependent on [*RNQ*$^+$] status. Although differences between the other sample's groups were observable, they were not as compelling. A discrepancy appears here in that one of the stress conditions was also prion containing but suggests that metabolic differences are potentially put aside in order to cope with more prevailing stresses. Alternatively, given the findings of chapter four, a switch in ubiquinone/sterol production due to oxidative stress has brought metabolite levels within these cells closer to each other. This action on [*rnq*$^-$] with oxidative stress samples may not be sufficient enough to cause protein misfolding to occur in cells, given the difference between these and stressed [*RNQ*$^+$] samples, the stress however is deliberately mild, and so potentially is on the brink of initiating this process. Switching view to consider the other most important variables, t[2] vs t[3], give new perspective to these relationships now revealing distinctions between the stressed and non-stressed conditions, with very little heed to the prion status other than some moderate clustering (Figure 5.2. G and H).

To further elucidate on the variation seen here PLSDA models were built using both PIM and NIM data. The models gave excellent scores indicative of strong reliable models: $R^2(X)$ =

0.659, $R^2(Y) = 0.939$, $Q^2 = 0.866$ and $R^2(X) = 0.667$, $R^2(Y) = 0.987$, $Q^2 = 0.923$ respectively (Figure 5.3.). CV-ANOVAs also suggest statistically significant differences had been observed between groupings, PIM data giving a *p-value* = 4.23×10^{-8} and NIM data a *p-value*s of *p-value* = 1.59×10^{-11}. Clustering within models appeared more defined between groups, although arguably the models themselves gave no further insight into the differences between groups, visually supporting the PCA observations of Figure 5.3.

Group modelling and observation confirmed that each of the sample groups were distinct from each other, providing clear evidence that each of our conditions had been maintained successfully. These established that the differences between our negative control, [*rnq*$^-$], and our positive control, [*rnq*$^-$] with mild oxidative stress, had in fact produced a distinct stress response from the stress response observed as a result of prion presence (Figure 5.2.and Figure 5.3.). The response to mild oxidative stress varied, but was not dissimilar, between prion-containing and prion cells (Figure 5.2.and Figure 5.3.). An unexpected result, given that the separation of the samples, especially within NIM data, is that [*RNQ*$^+$] with a mild oxidative stress is behaving much more like the prion-free samples than the prion-containing samples.

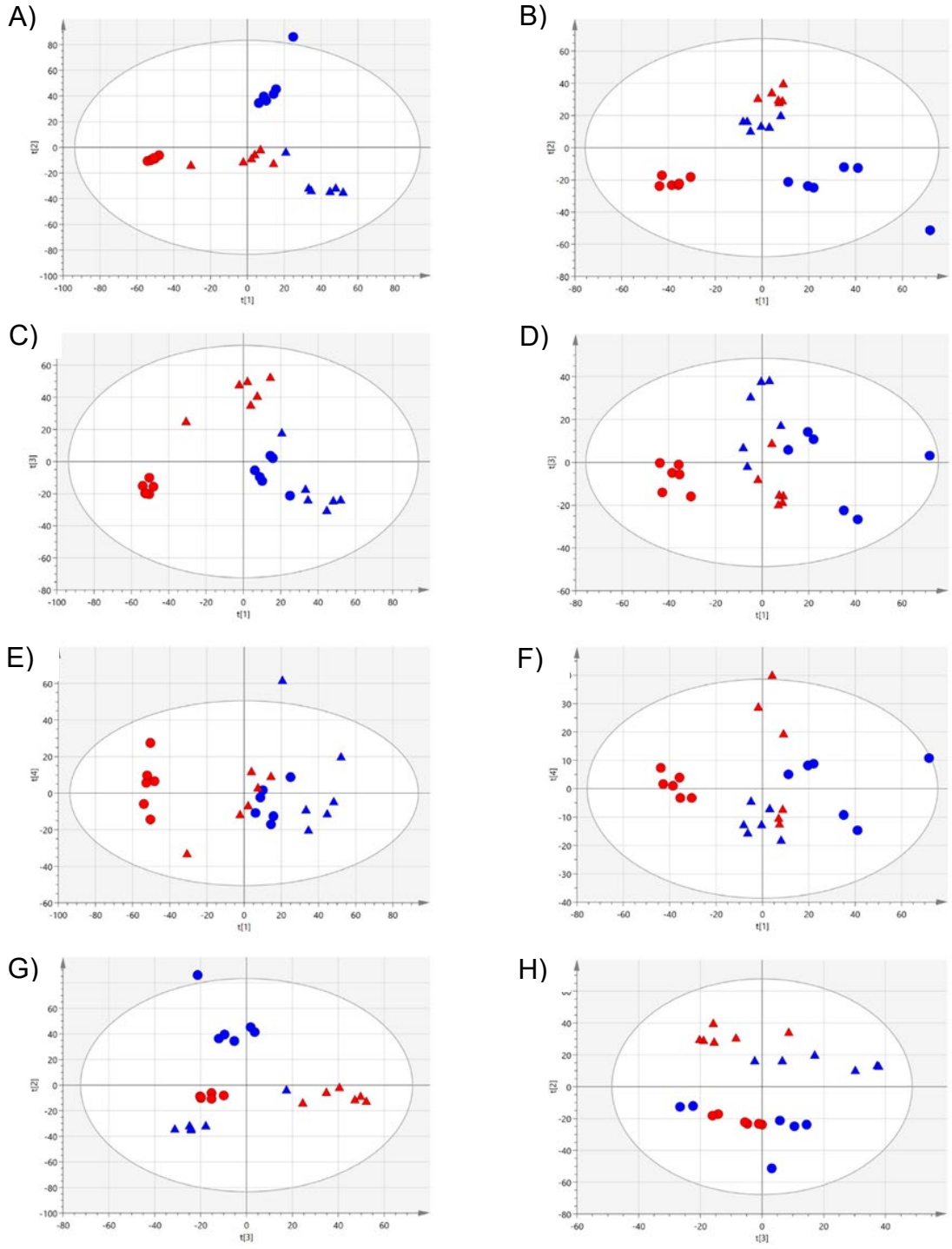


Figure 5.2. PCA models of XCMS/MZmine2 comparative data plotting [*rnq*$^-$], [*rnq*$^-$] with a mild oxidative stress, [*RNQ*$^+$] and [*RNQ*$^+$] with a mild oxidative stress (multiple components shown). Overview of the data shows no more than one outlying sample within the 95% confidence range within each model. The PIM PCA model, with R^2 = 0.451 Q^2 = 0.312 values, is shown on the left (A, C, E and G) and the NIM PCA model, with R^2 = 0.451 Q^2 = 0.312 values, on the right (B, D, F and H). Shown here are a variety of the contributary components from each of the PIM and NIM PCA models. Blue circles represent [*rnq*$^-$], blue triangles show [*rnq*$^-$] with a mild H_2O_2 stress, Red circles represent [*RNQ*$^+$] and red triangles show [*RNQ*$^+$] with a mild H_2O_2 stress. A and B) Show component t[1] on the X axis and component t[2] on the Y axis, C and D) Show component t[1] on the X axis and component t[3] on the Y axis, E and F) Show component t[1] on the X axis and component t[4] on the Y axis, G and H) Show component t[2] on the X axis and component t[3] on the Y axis.

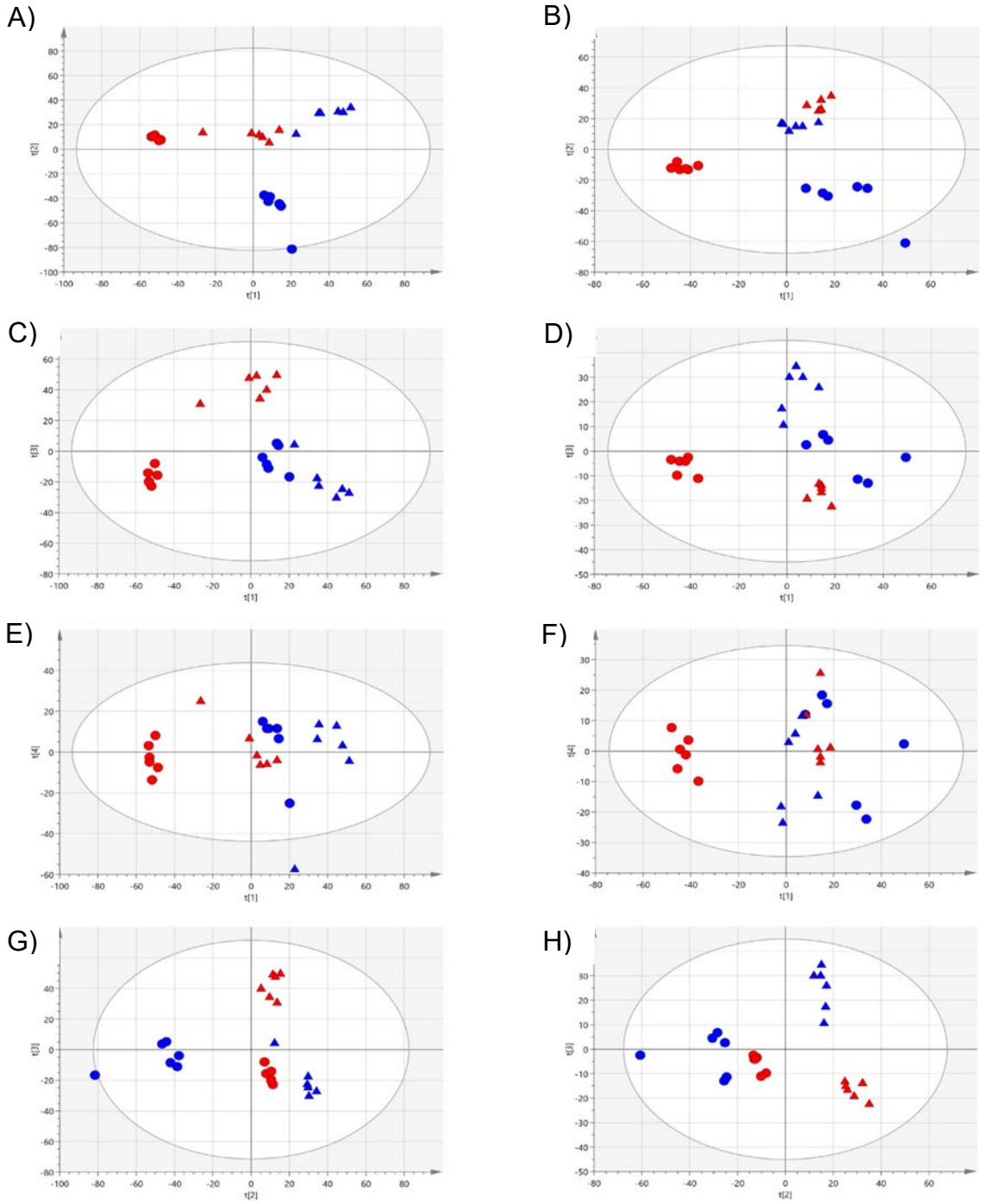


Figure 5.3. PLS-DA models of XCMS/MZmine2 comparative data plotting [*rnq*$^-$], [*rnq*$^-$] with a mild oxidative stress, [*RNQ*$^+$] and [*RNQ*$^+$] with a mild oxidative stress (multiple components shown). Overview of the data shows no more than one outlying sample within the 95% confidence range within each model. The PIM PLS-DA scores plot a model with $R^2(X) = 0.659$, $R^2(Y) = 0.939$, $Q^2 = 0.866$ and a CV-ANOVA *p-value* = 1.60×10^{-11}, shown on the left (A, C, E and G) and the NIM PLS-DA scores plot a model with $R^2(X) = 0.687$, $R^2(Y) = 0.987$, $Q^2 = 0.923$ and a CV-ANOVA *p-value* = 4.23×10^{-8}, shown on the right (B, D, F and H). Shown here are a variety of the contributary components from each of the PIM and NIM PCA models. Blue circles represent [*rnq*$^-$], blue triangles show [*rnq*$^-$] with a mild H_2O_2 stress, Red circles represent [*RNQ*$^+$] and red triangles show [*RNQ*$^+$] with a mild H_2O_2 stress. A and B) Show component t[1] on the X axis and component t[2] on the Y axis, C and D) Show component t[1] on the X axis and component t[3] on the Y axis, E and F) Show component t[1] on the X axis and component t[4] on the Y axis, G and H) Show component t[3] on the X axis and component t[2] on the Y axis.

5.3. Results – Comparison of [*rnq*$^-$] cells in the presence or absence of mild oxidative stress

5.3.1. PCA and OPLSDA modelling of comparative metabolomic data in SIMCA differentiates between [*rnq*$^-$] and [*rnq*$^-$] with mild oxidative stress

In order to account for the metabolic effects of a generalised stress condition, the negative and positive control groups, [*rnq*$^-$] and [*rnq*$^-$] with oxidative stress, were modelled together. This is important as the identification of the metabolic changes that occur due to a stress response would inform our subsequent interpretations of the metabolic responses caused solely to prion presence and allow us to distinguish between general stress responses and those unique to prion presence.

PCA scores plot reasonable models for both PIM and NIM data, with $R^2 = 0.570$, $Q^2 = 0.206$, $R^2 = 0.843$, $Q^2 = 0.160$ values respectively (shown in Figure 5.4.). Acceptable R^2 values are given but Q^2 values indicate poor predictability within the data set. This may be due to the large number of variables. OPLS-DAs plot models with good levels of confidence, with the PIM model giving an $R^2(X) = 0.649$, $R^2(Y) = 0.990$, $Q^2 = 0.882$ and CV-ANOVA *p-value* = 9.45×10^{-3}, and the NIM model reporting an $R^2(X) = 0.564$, $R^2(Y) = 0.977$, $Q^2 = 0.743$ and a CV-ANOVA *p-value* = 8.13×10^{-2} (shown in Figure 5.4.). All models, except for the NIM PCA, show distinct clustering and separation as a result of sample group. NIM PCA shows significantly more overlap than ideal but given the strong separation within the PIM PCA and the significant CV-ANOVA values, this is acceptable. These results provide confidence that [*rnq*$^-$] samples are metabolically distinct from [*rnq*$^-$] samples with oxidative stress.

5.3.2. Statistical analysis and tentative feature ID of comparative metabolomic data via Metaboanalyst

The methodology and details that relate to the subsequent Mummichog analysis (including statistical analysis, tentative feature ID and pathway analysis) can be found within appendix B.

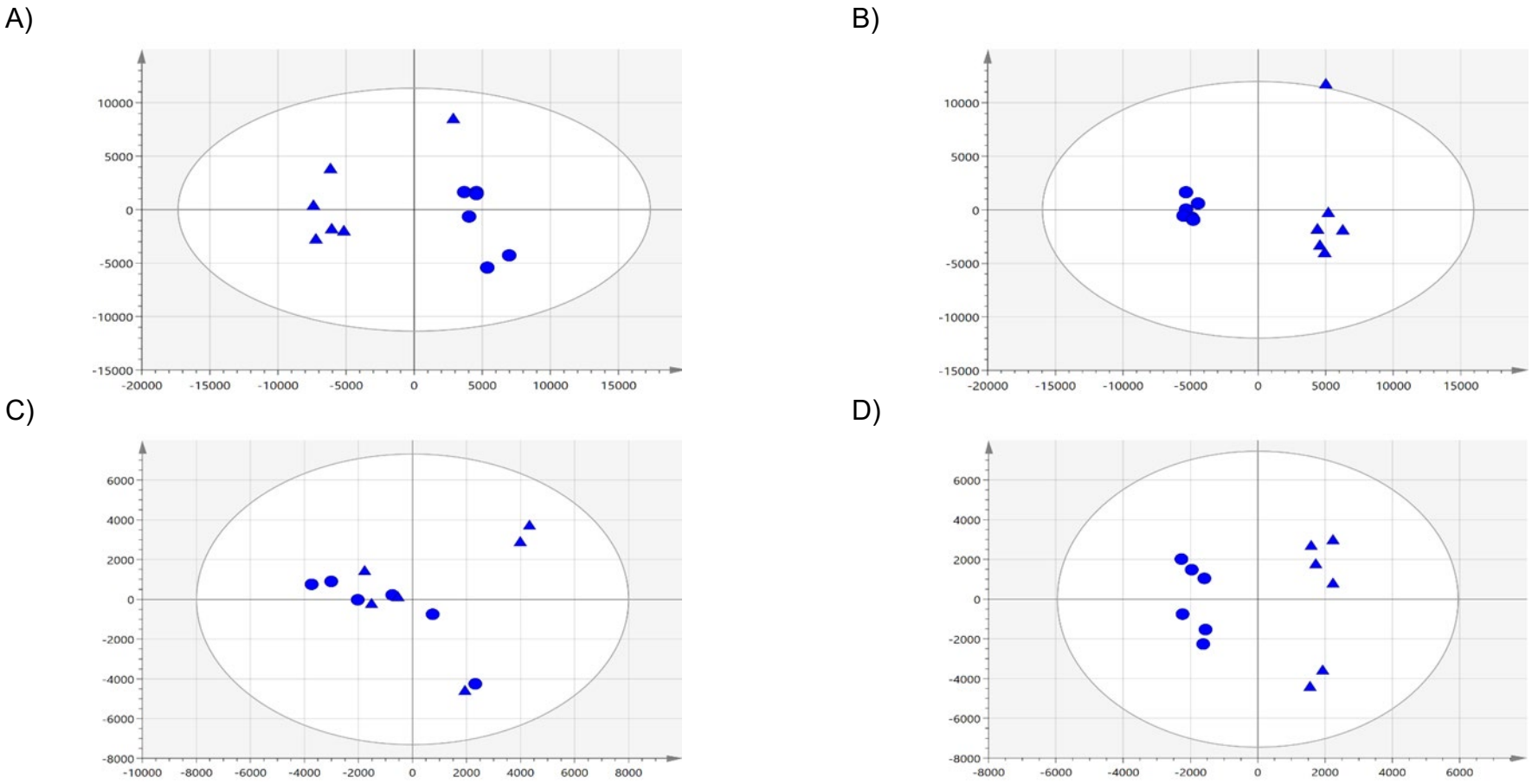


Figure 5.4. PCA (A and C) and OPLS-DA (B and D) models of XCMS/MZmine2 comparative data of [*rnq*$^-$] and [*rnq*$^-$] with a mild oxidative stress. Overview of the data shows one outlying samples within the 95% confidence range within each model. Blue circles represent [*rnq*$^-$] and blue triangles show [*rnq*$^-$] with a mild H_2O_2 stress (A) Comparative PIM data PCA scores with $R^2 = 0.570$ $Q^2 = 0.206$ values. (B) Comparative PIM data OPLS-DA scores with $R^2(X) = 0.649$, $R^2(Y) = 0.990$, $Q^2 = 0.882$ and CV-ANOVA *p-value* = 9.45×10^{-3}. (C) Comparative NIM data PCA scores with $R^2 = 0.843$, $Q^2 = 0.160$ values. (D) Comparative NIM data OPLS-DA scores with $R^2(X) = 0.564$, $R^2(Y) = 0.977$, $Q^2 = 0.743$ and a CV-ANOVA *p-value* = 8.13×10^{-2}..

5.3.3. BioCyc omics dashboard and cellular overview enables mapping of most significant metabolic changes indicating key metabolic pathways of interest

To prepare for pathways analysis within BioCyc, all significant punitively identified metabolites (as established by Metaboanalyst's Mummichog), their relative abundances and average relative abundances (within groupings) were collated. This included all PIM and NIM 'omics data. Relative abundances of the most significantly altered metabolites (established via Metaboanalyst's Mummichogg) from the six [*rnq*$^{-}$] samples and the six [*rnq*$^{-}$] with mild oxidative stress samples were uploaded to BioCyc's Omics Dashboard. This permitted an overall, non-directional, assessment of the key metabolic regions influenced by the stress condition (Figure 5.5.) and a directional visualisation of the scale of specific metabolic regions altered between groups, shown in Figure 5.6.

Figure 5.5 reveals that pathways involving biosynthesis are the most widely affected within cellular metabolism, particularly cofactor and amino acid biosynthesis pathways. This was found to correlate with the information from Metaboanalyst (shown in appendix B). Kang *et al.* (2012) study was used to design the mild-stress condition used in this experiment, they did not perform a subsequent pathway analysis but identified amino acid as the most significantly altered metabolites. However, Kang *et al.* (2012) do state that the confines of their experiment and the metabolic burden of the stresses used made identifying key perturbed metabolites implausible.

Another key area is energy metabolism, which shows significant change in several of the key energy production cycles and pathways. It appears that Glycolysis and ATP production have remained relatively unaltered, but the TCA cycle, other energy and fermentation cycles have changed, collectively contributing over half of the changes seen within energy metabolism. The pathways involved in degradation shown in Figure 5.6. is also considerably larger than previously observed in chapter 4. Current thinking suggests that under stress conditions *S. cerevisiae* and other organisms' ability to control waste is hampered, leading to an effect known as stress induced recycling (Parrou *et al.* 1997). This may explain why approximately a 25% increase within this area is seen, without a significant change within biosynthesis. Yeast specific/other pathways make a significant contribution to the overall spread of the changes seen Figure 5.5., representing a 30 % decrease in activity when compared to chapter four.

Figure 5.6 divides the data into [*rnq*$^-$] and [*rnq*$^-$] with mild oxidative stress samples respectively to give a greater insight into the direction of these changes. Biosynthesis pathways and degradation pathways are found to be downregulated in both [*rnq*$^-$] samples and [*rnq*$^-$] with mild oxidative stress, but this downregulation is not equal. Biosynthesis appears affected more than degradation and this lack of correlation between the two processes may be intrinsically linked to stress induced recycling. It may also be linked to the increase in reactive oxygen species (ROS) that is known to occur under oxidative stress. ROS although common by-products of normal cellular metabolism, can serve as signalling molecules and mediate cellular damage (Ali *et al.* 2010; D'Autreaux and Toledano, 2007). Their tendency to donate oxygen to other substances due to their unstable and highly reactive properties makes them capable of attacking important macromolecules leading to homeostatic disruption (Lobo *et al.* 2010). Given the common compartmentation of cellular degradation, it stands to reason that biosynthesis pathways may be more susceptible to the cellular presence of ROS and then the degradation pathways, hence creating this disconnect in downregulation. Essentially biosynthesis pathways may simply be more susceptible to the metabolic burden presented by stress conditions due to their free availability within cells.

Within the pathways related to energy metabolism, upregulation of the PPP and Glycolysis pathways can be seen within [*rnq*$^-$] with mild oxidative stress samples, an effect that has been noted by other researchers as a response to oxidative stress within *S. cerevisiae.* This suggests that our mild oxidative stress conditions were maintained throughout the experiment (Gonzalez-Siso *et al.* 2009; Magherini *et al.* 2009). The most significant change seen within the remaining pathways is within the C1 utilities pathway, where a considerable downregulation is observed in the presence of oxidative stress. As a common source of biomarkers for the occurrence of oxidative stress and associated diseases, this finding is expected and offers further support that our mild oxidative conditions were maintained (Vona *et al.* 2019). Despite providing a general overview our analyses (Figure 5.5. and 5.6.) did not reveal which specific pathways were most perturbed as a result of the mild oxidative stress condition. As the purpose of this experimental stage was to allow us to separate the pathways implicated as playing a key role in prion presence from those involved in a general stress response, combined PIM and NIM data was uploaded to BioCyc's specific *S. cerevisiae* cellular overview. This enabled the identification of pathways perturbed by oxidate stress at a pathway levels and are listed in Table 5.2.

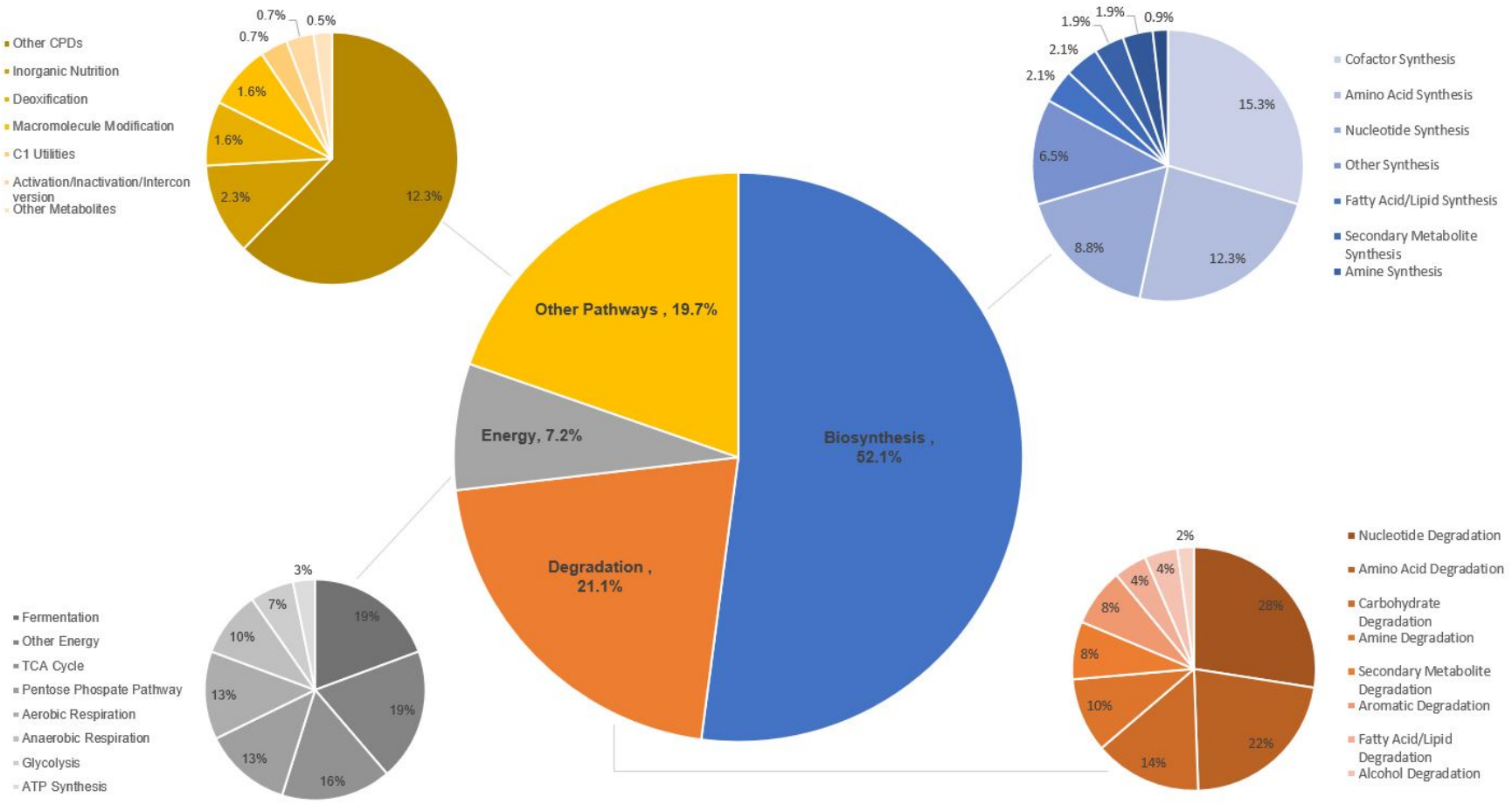
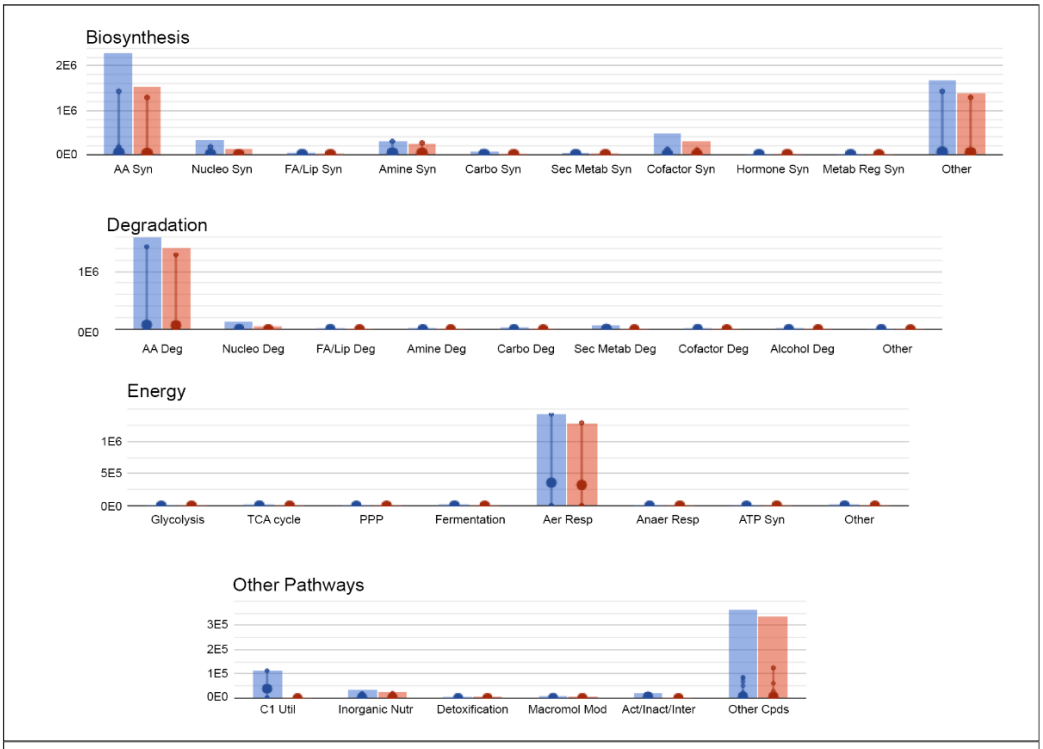


Figure 5.5. Graphical representation of the metabolic perturbations detailed by region when comparing [*rnq*$^+$] and [*rnq*$^-$] with a mild oxidative stress. The largest pie chart shows a general overview of the key areas detected as changed between [*rnq*$^-$] and [*rnq*$^-$] with mild oxidative stress, and the percentage of the change assigned to those areas, as determined via BioCyc's omics dashboard. The smaller pie charts are colour coded with the small blue pie chart showing (by percentage) the metabolic region contributing to the overall biosynthesis percentage seen in the largest pie chart. The orange pie chart provides these details for the degradation percentage, the grey pie chart provides these details for the energy percentage and the yellow pie chart provides these details for the other pathway's percentage.

A)



B)

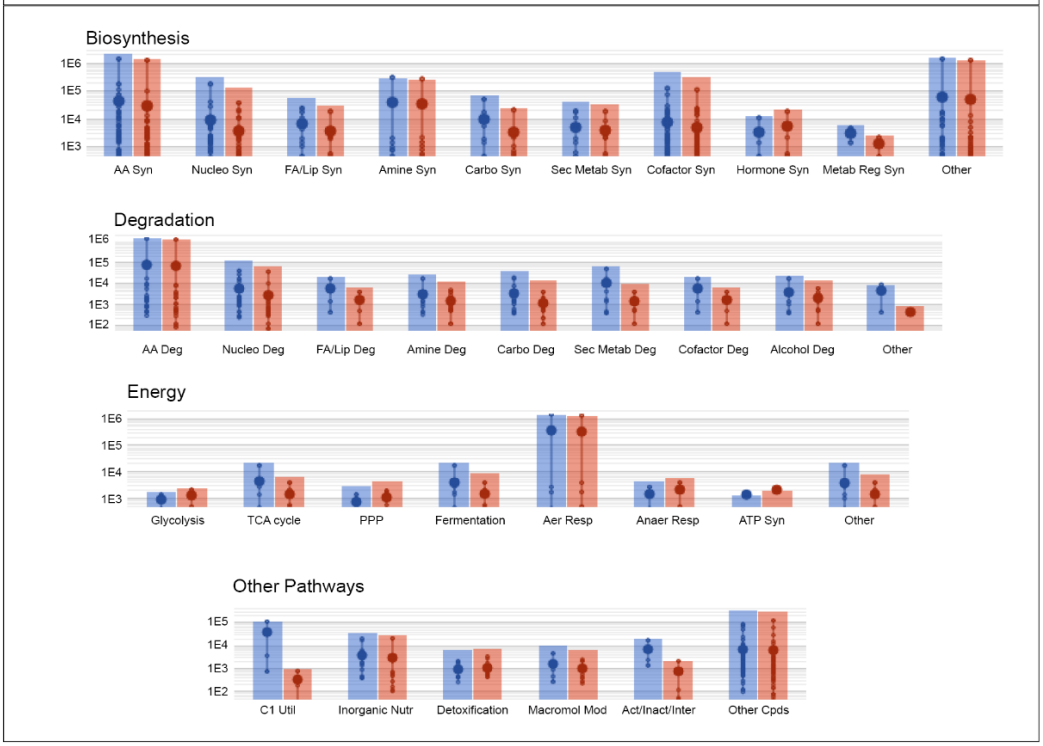


Figure 5.6. Graphical comparisons of the average relative abundances between [*rnq*$^-$] and [*rnq*$^-$] with a mild oxidative stress samples within key metabolic regions. A) Charts average comparisons of the relative abundances between [*rnq*$^-$] samples (shown in blue) and [*rnq*$^-$] with mild oxidative stress (shown in red), data is first separated by key metabolic region (shown as a graph title) and then by specific metabolic area (shown on the x axis). The Y axis is set to a linear scale, appropriate to the relative abundances. Individual data points are shown as small circles within the larger summated bars. B) Shows the same information with a logarithmic scale on the Y axis.

Table 5.2. Top Ten most perturbed pathways between [*rnq*$^-$] and [*rnq*$^-$] with a mild oxidative stress.

Pathway Name
proline degradation
arginine degradation VI (arginase 2 pathway)
proline biosynthesis
inosine-5'-phosphate biosynthesis II
tryptophan biosynthesis
citrulline biosynthesis
formaldehyde oxidation II (glutathione-dependent)
histidine biosynthesis
superpathway of phenylalanine, tyrosine and tryptophan biosynthesis
aerobic respiration

When comparing the most significant [*rnq*$^-$] relative abundances to the most significant [*rnq*$^-$] with mild oxidative stress relative abundances, via BioCyc's specific *S. cerevisiae* cellular overview, as ranked by DPPS.

5.3.4. Data overlay onto implicated pathways, via BioCycs pathway collage, reveals the direction of the up and down regulation between groups

Table 5.2 lists the key perturbed pathways identified using standardized data (as described in section 4.5.6). Using BioCyc pathway collage, this data was overlaid onto pathways of interest to further visualize and interpret these results. The pathways listed in Table 5.2. are presented within Figures 5.7 and 5.8. The data shown is optimized to permit efficient presentation as opposed to presenting them by DPPS rank (as shown in Table 5.6.).

The data within Figures 5.7 and 5.8 indicates that in the presence of mild-oxidative stress there is a downregulation in pathways that involve proline, arginine, and the aromatic amino acids. These residues have been identified as the preferred targets for ROS attack (Stadtman and Levine (2003) and that within proline and arginine metabolism, ROS will result in the conversion of these amino acids to the oxidation-species glutamic semialdehyde. This metabolite is an intermediate within all of the pathways shown in Figure 5.8 but was not identified in this study. In addition, no amino acids containing thiol groups were identified which may have been expected as these are commonly associated with ROS. Methionine and cysteine are thought to be particularly susceptible to oxidation under H_2O_2 stress resulting in disruption to intracellular signaling (Vona *et al.* 2019). These effects may be more subtle and not included in the ten most perturbed pathways shown here. Indeed, from data shown of the PPP pathway seen in Figure 5.6 this would suggest so. Work in this area has demonstrated that the mechanism of upregulation within PPP is intrinsically linked with alterations to methionine concentrations (Campbell *et al.* 2016) highlighting the value of an overall general comparison prior to individual pathway analysis.

A further explanation as to why these pathways do not appear within the ten most perturbed pathways could be explained by the cyclic, reversible oxidation of thiol groups which may be overshadowed by unidirectional effects (Stadtman and Levine, 2003). Further examination of this data will be conducted with respect to the aims of this chapter, within the following sections.

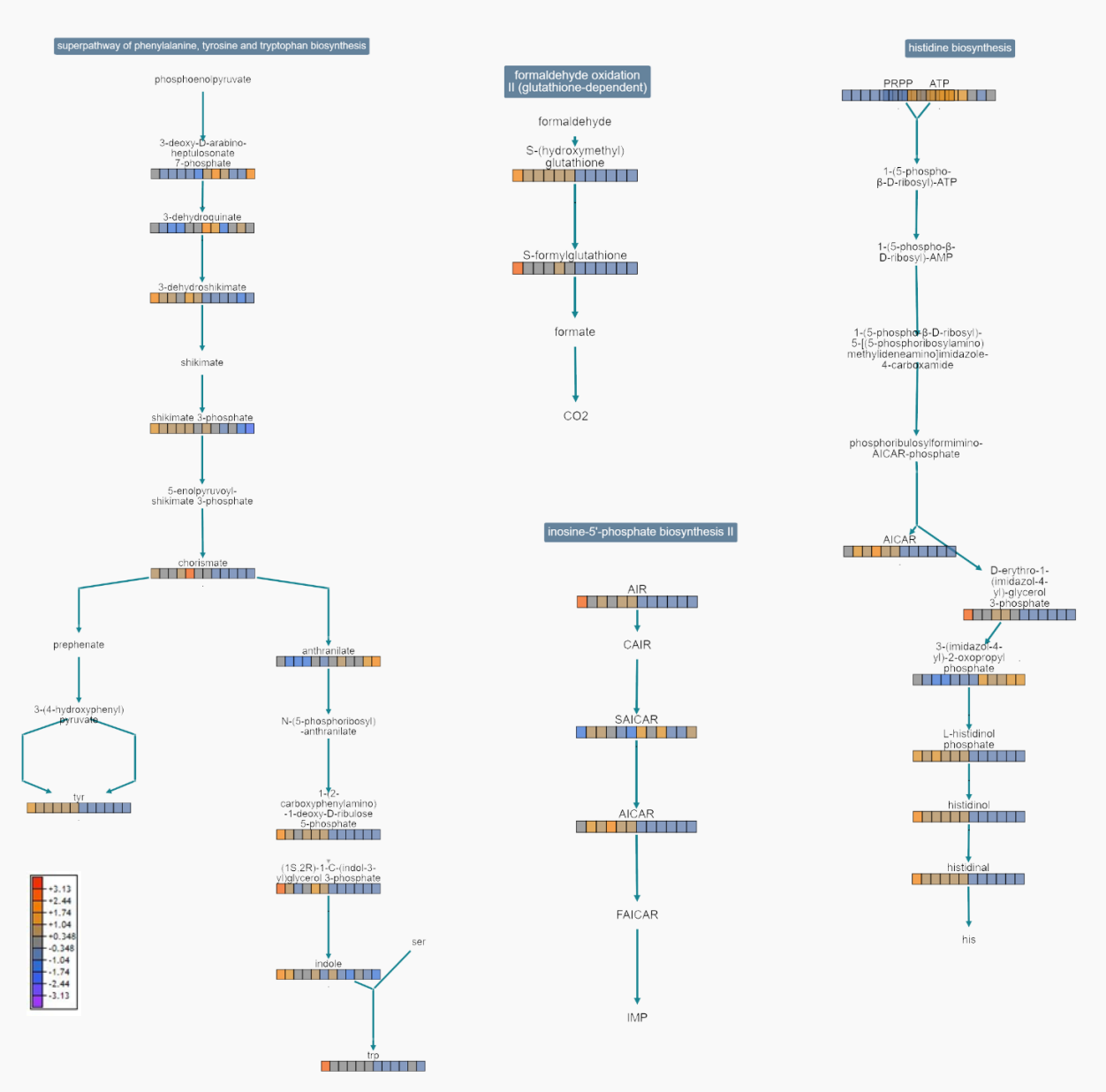


Figure 5.7. Pathway collage of 5 of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics [*rnq*$^+$] and [*rnq*$^-$] with a mild oxidative stress sample data of individual metabolites overlaid. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data where available. Reading from left to right, the first six boxes in any 'heat block' belong to [*rnq*$^+$] samples and the last six boxes belong to [*rnq*$^-$] with mild oxidative stress samples. A larger version of this Figure is available in Appendix H.

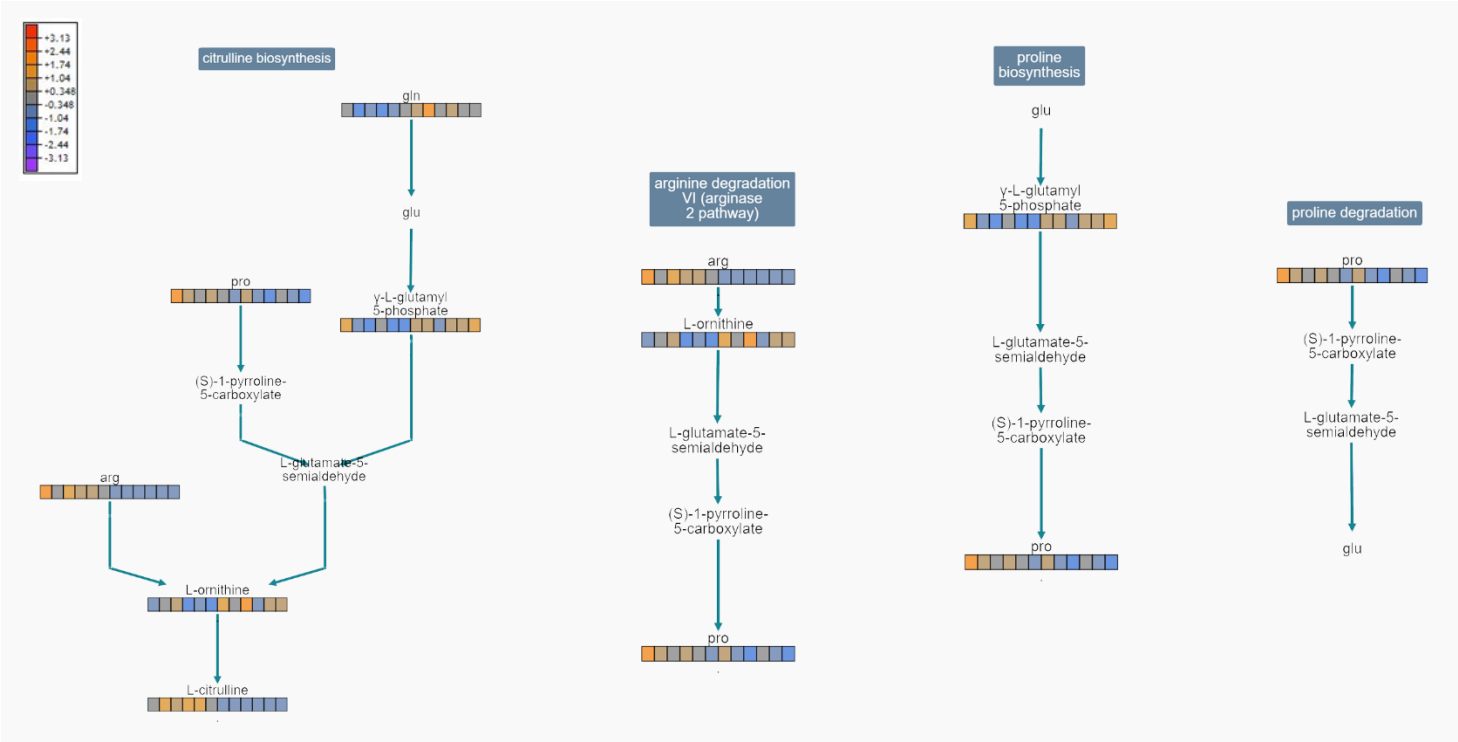


Figure 5.8. Pathway collage of 4 of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics [*rnq*$^-$] and [*rnq*$^-$] with a mild oxidative stress sample data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data where available. Reading from left to right, the first six boxes in any 'heat block' belong to [*rnq*$^-$] samples and the last six boxes belong to [*rnq*$^-$] with mild oxidative stress samples. A larger version of this Figure is available in Appendix H.

5.4. Results – Comparison of the Metabolomics Perturbations Observed between [*rnq*$^-$] and [*RNQ*$^+$]

5.4.1. PCA and OPLSDA modelling reveals significant difference between prion free [*rnq*$^-$] and prion containing cells [*RNQ*$^+$]

To establish if a metabolic difference had been observed between prion free, [*rnq*$^-$] cell samples and prion containing, [*RNQ*$^+$] cell samples, PCA plots for both PIM and NIM data were built as shown in Figure 5.9. The PCAs showed no outlying samples and gave desirable scores for both PIM and NIM, with R^2 = 0.712 Q^2 = 0.523 and R^2 = 0.807, Q^2 = 0.556 values, respectively. Clustering of the sample types was clearly seen, although [*RNQ*$^+$] samples appeared to cluster much more closely than [*rnq*$^-$] samples indicating that the [*RNQ*$^+$] samples present less variability among the group than the [*rnq*$^-$] samples, nevertheless clear metabolic separation was observed.

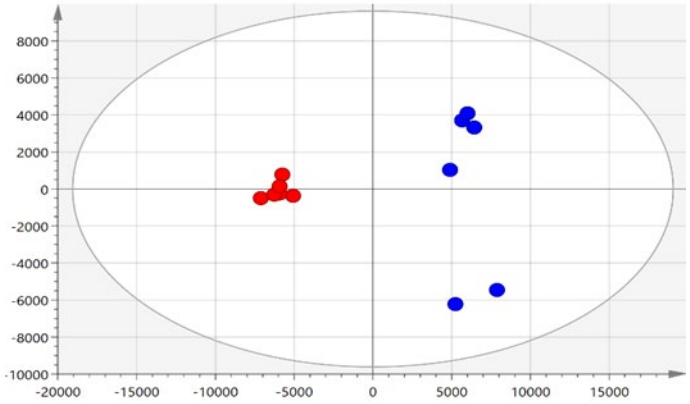
Subsequent OPLS-DAs plot models with good levels of confidence with PIM data scoring R^2(X) = 0.748, R^2(Y) = 0.999, Q^2 = 0.976 and NIM data scoring R^2(X) = 0.776, R^2(Y) = 0.997, Q^2 = 0.936. Although the same level of spread was seen in the [*rnq*$^-$] samples, CV-ANOVAs revealed significant differences between the sample groups for both PIM and NIM data sets, CV-ANOVA *p-value* = 9.26 x 10^{-5} and CV-ANOVA *p-value* = 1.57 x 10^{-3} (Figure 5.9.).

These analyses provide confidence in the data and confirm the generally commonly considered notion that [*rnq*$^-$] samples are metabolically distinct from [*RNQ*$^+$] samples.

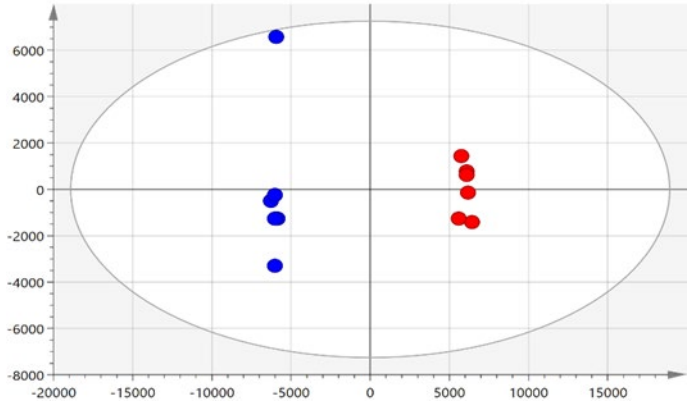
5.4.2. Statistical analysis and tentative feature ID of comparative metabolomic data via Metaboanalyst

The methodology and details that relate to the subsequent Mummichog analysis (including statistical analysis, tentative feature ID and pathway analysis) can be found within appendix C.

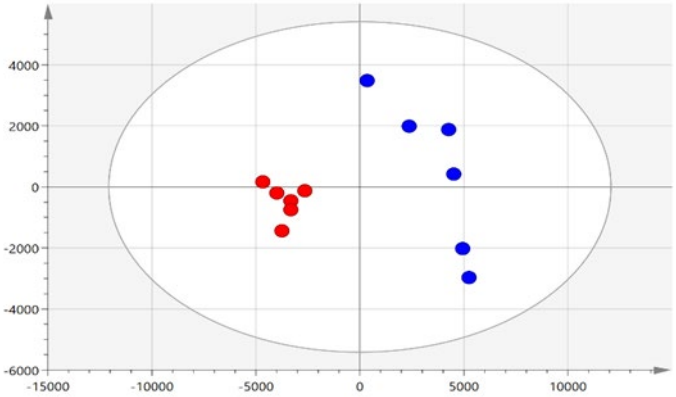
A)



B)



C)



D)

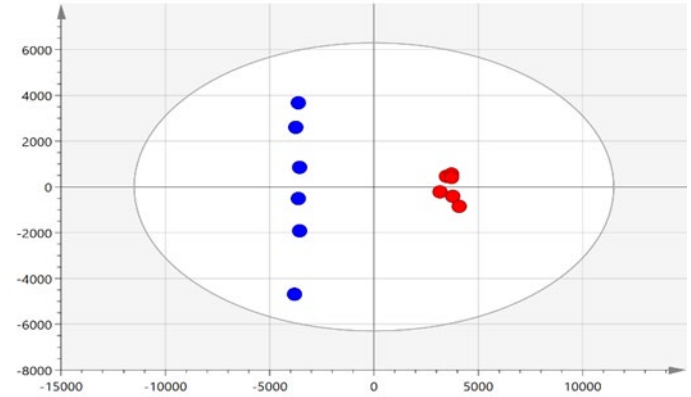


Figure 5.9. PCA (A and C) and OPLS-DA (B and D) models of XCMS/MZmine2 comparative data of [*rnq*$^-$] and [*RNQ*$^+$] samples. Overview of the data shows no outlying samples within the 95% confidence range within each model. Blue circles represent [*rnq*$^-$] and Red circles represent [*RNQ*$^+$]. (A) Comparative PIM data PCA scores with $R^2 = 0.712$ $Q^2 = 0.523$ values. (B) Comparative PIM data OPLS-DA scores with $R^2(X) = 0.748$, $R^2(Y) = 0.999$, $Q^2 = 0.976$ and CV-ANOVA *p-value* = 9.26×10^{-5}. (C) Comparative NIM data PCA scores with $R^2 = 0.807$, $Q^2 = 0.556$ values. (D) Comparative NIM data OPLS-DA scores with $R^2(X) = 0.776$, $R^2(Y) = 0.997$, $Q^2 = 0.936$ and a CV-ANOVA *p-value* = 1.57×10^{-3}.

5.4.3. BioCyc omics dashboard and cellular overview enables mapping of most significant metabolic changes indicating key metabolic pathways of interest

Data throughout this section has been processed as described in section 5.3. Figure 5.10. which shows metabolic perturbation by region when comparing [*rnq*$^{-}$] and [*RNQ*$^{+}$] samples, when compared with Figure 4.7. reveals similar trends in the contributary percentage of biosynthesis, degradation, and other pathways. Indeed, even the detailed metabolic regions under influence are largely the same. This was not found for energy metabolism, with Figure 5.10. energy metabolism percentage influence at 4.4%, somewhere between the findings of Figure 4.7. (2.4%) and those in Figure 5.5. (7.2%), with the new areas of energy metabolism seen in Figure 5.5. making a reappearance.

The same general impression is provided by Figure 5.11 with a striking similarity to the regulation events seen in Figure 4.8. relative to the data's relationship to [*rnq*$^{-}$]. By this, it is meant that within Figure 4.8. [*rnq*$^{-}$] is seen on the left due to comparison with *Δrnq1*, however within Figure 5.11. [*rnq*$^{-}$] appears on the right due to comparison with [*RNQ*$^{+}$]. Once again, the only significantly different observation is within energy metabolism with fermentation and glycolysis both showing upregulation in the presence of the [*RNQ*$^{+}$] prion.

Whilst this number of shared features may appear surprising, this was entirely anticipated as the loss of function effects shown previously (chapter four) should also be observed in the [*RNQ*$^{+}$] samples. This is because a metabolic effect of prion presence is a loss of function due the now non-functional misfolded Rnq1 protein. Crossover between the stress effects was also expected, via a general cellular stress response.

For the expectations stated above, the decision was made to include the top 20 most significantly perturbed pathways as identified in BioCyc. This would permit overlap with our previous findings as well as allow for the identification of new ones. The results of this search (Table 5.3) found that most if not all of the pathways perturbed appeared to be linked to either loss of function or a generalized stress response. The ten most significantly perturbed had DPPS scores above 100000 and all pathways listed gave DPPS scores greater than 29,000.

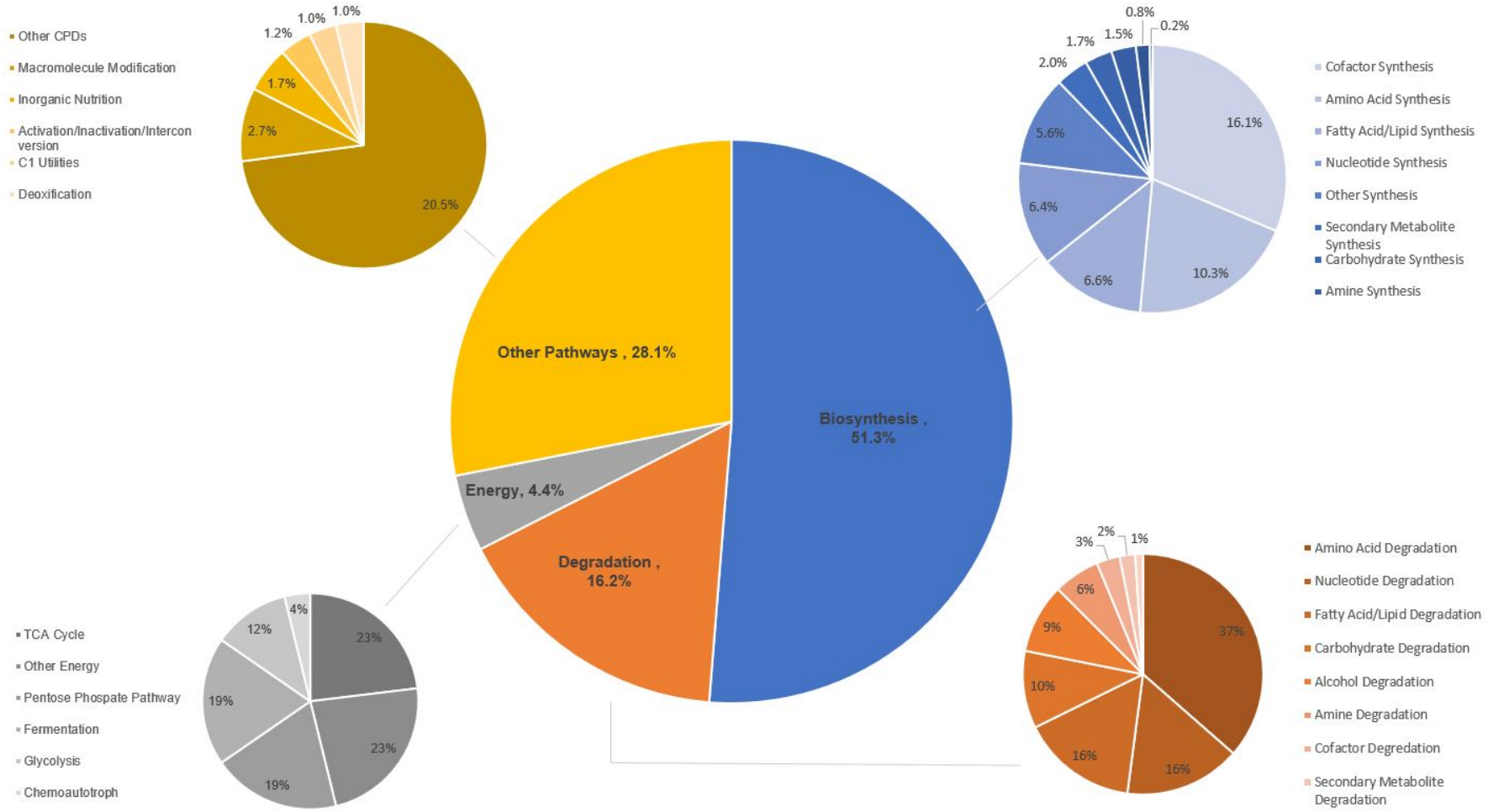
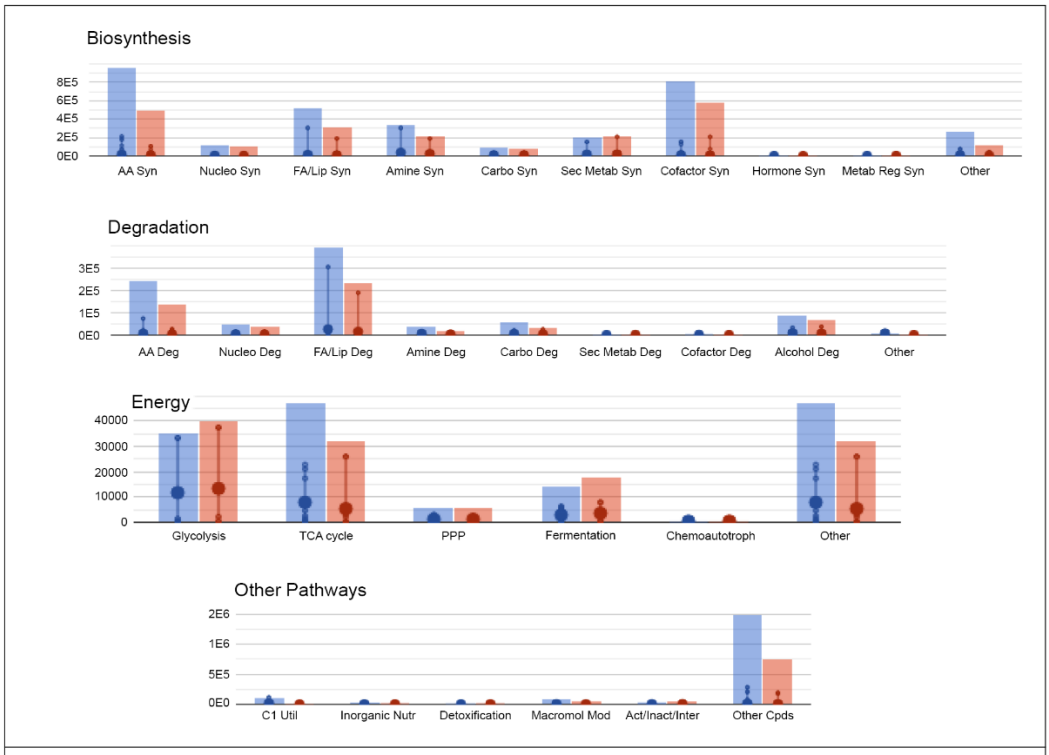


Figure 5.10. Graphical representation of the metabolic perturbations detailed by region when comparing [*rnq*$^-$] and [*RNQ*$^+$].The largest pie chart shows a general overview of the key areas detected as changed between groups and the percentage of the change assigned to those areas, as determined via BioCyc's omics dashboard. The smaller pie charts are colour coded with the small blue pie chart showing the detailed metabolic regions (by percentage) that are contributing to the overall biosynthesis percentage seen in the largest pie chart. The orange pie chart provides these details for the degradation percentage, the grey pie chart provides these details for the energy percentage and the yellow pie chart provides these details for the other pathway's percentage.

A)



B)

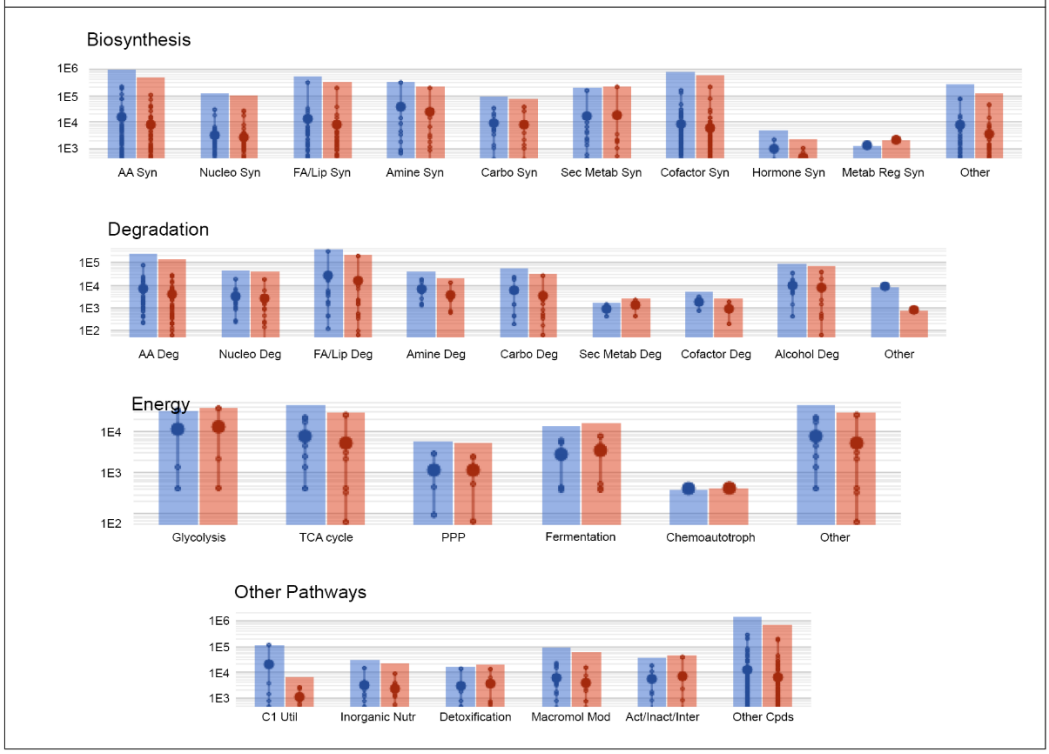


Figure 5.11. Graphical comparisons of the average relative abundances between [*rnq*$^-$] and [*RNQ*$^+$] samples within key metabolic regions. A) Charts average comparisons of the relative abundances between *Δrnq1* samples (shown in blue) and [*rnq*$^-$] (shown in red), data is first separated by key metabolic region (shown as a graph title) and then by specific metabolic area (shown on the x axis). The Y axis is set to a linear scale, appropriate to the relative abundances. Individual data points are shown as small circles within the larger summated bars. B) Shows the same information with a logarithmic scale on the Y axis.

Table 5.3. Top Ten most perturbed pathways between [*rnq*$^-$] and [*RNQ*$^+$].

Pathway Name
tryptophan biosynthesis
geranylgeranyldiphosphate biosynthesis
formaldehyde oxidation II (glutathione-dependent)
phospholipids degradation
histidine biosynthesis
superpathway of phenylalanine, tyrosine and tryptophan biosynthesis
phospholipid biosynthesis (Kennedy pathway)
choline biosynthesis
phosphatidylcholine resynthesis via glycerophosphocholine
4-hydroxyphenylpyruvate biosynthesis
tyrosine biosynthesis
tyrosine degradation
siroheme biosynthesis
superpathway of geranylgeranyldiphosphate biosynthesis I (via mevalonate)
ubiquinol-6 biosynthesis from 4-hydroxybenzoate (eukaryotic)
homocysteine and cysteine interconversion
L-lysine biosynthesis IV
superpathway of chorismate metabolism
ubiquinol-6 bypass biosynthesis
leucine biosynthesis

When comparing the most significant [*rnq*$^-$] relative abundances to the most significant [*RNQ*$^+$] relative abundance, via BioCyc's specific *S. cerevisiae* cellular overview.

5.4.4. Data combination and overlay onto implicated pathways, via BioCycs pathway collage, reveals loss of function effects

The most significant perturb pathways provided by Table 5.3. was cross correlated with the information from chapter 4, to enable the 'removal' of those pathways linked to loss of function. Data from chapter 4 was combined with [*RNQ*$^+$] data and any pathways implicated, or those pathways linked to pathways which had been implicated were overlaid with this information. Examination of Figure 5.13. and 5.14 shows that a consistent and similar response can be seen in *Δrnq1* and [*RNQ*$^+$] samples, validating the initial interpretation of these pathway hits as being as a result of the loss of function of the Rnq1 protein. However, the response seen within the ubiquinol pathways (shown in Figure 5.13) is not as strongly defined as implied within the discussion in chapter 4 indicating that there are other factors involved that are not yet known.

5.4.5. Data combination and overlay onto implicated pathways, via BioCycs pathway collage, reveals generalised stress effects

Cross correlation of the [*rnq*$^-$] relative abundances to the most significant [*RNQ*$^+$] relative abundances from section 5.3 confirmed that a generalized stress response can be seen in both [*rnq*$^-$] with mild-oxidative stress samples and [*RNQ*$^+$] samples. Both conditions were found to give comparable data with the unaffected [*rnq*$^-$] samples. The perturbed metabolites identified as a result of a generalized stress response include the amino acid tryptophan and many of the intermediates within its biosynthetic pathway (Figure 5.14.). Closer inspection of Figure 5.14 reveals that tyrosine biosynthesis and tyrosine degradation show common features in [*rnq*$^-$] and [*rnq*$^-$] with mild oxidative stress and only appear to be (generally) downregulated in our [*RNQ*$^+$] samples. However, given the relationship of these pathways to pathways known to be implicated by stress, this result would require further validation.

5.4.6. Data combination and overlay onto implicated pathways, via BioCycs pathway collage, reveals loss of function effects

Three of the pathways, namely homocysteine and cysteine interconversion, L-lysine biosynthesis and leucine biosynthesis, shown on Table 5.3 were not present as hits or strongly related to hits within those pathways that had previously been implicated by either loss of Rnq1 function or stress response. These pathways were subsequently overlaid with data from all groups to ensure all potential hits were included. Figure 5.15 reveals perturbations within the leucine biosynthesis pathway and that fluctuations within this pathway could be attributed to both loss of Rnq1 function and stress response. A similar situation is also apparent within the L-lysine biosynthesis pathway. How these perturbations have manifest is not clear, although a possible explanation could be that there is further induction of additional pathways under stress conditions as a result of a knockout strain like *Δrnq1*.

The perturbations seen within the homocysteine and cysteine interconversion pathways are mixed, showing up and down regulation in a disordered pattern and irrespective of sample condition, making it difficult to come to any definitive conclusion.

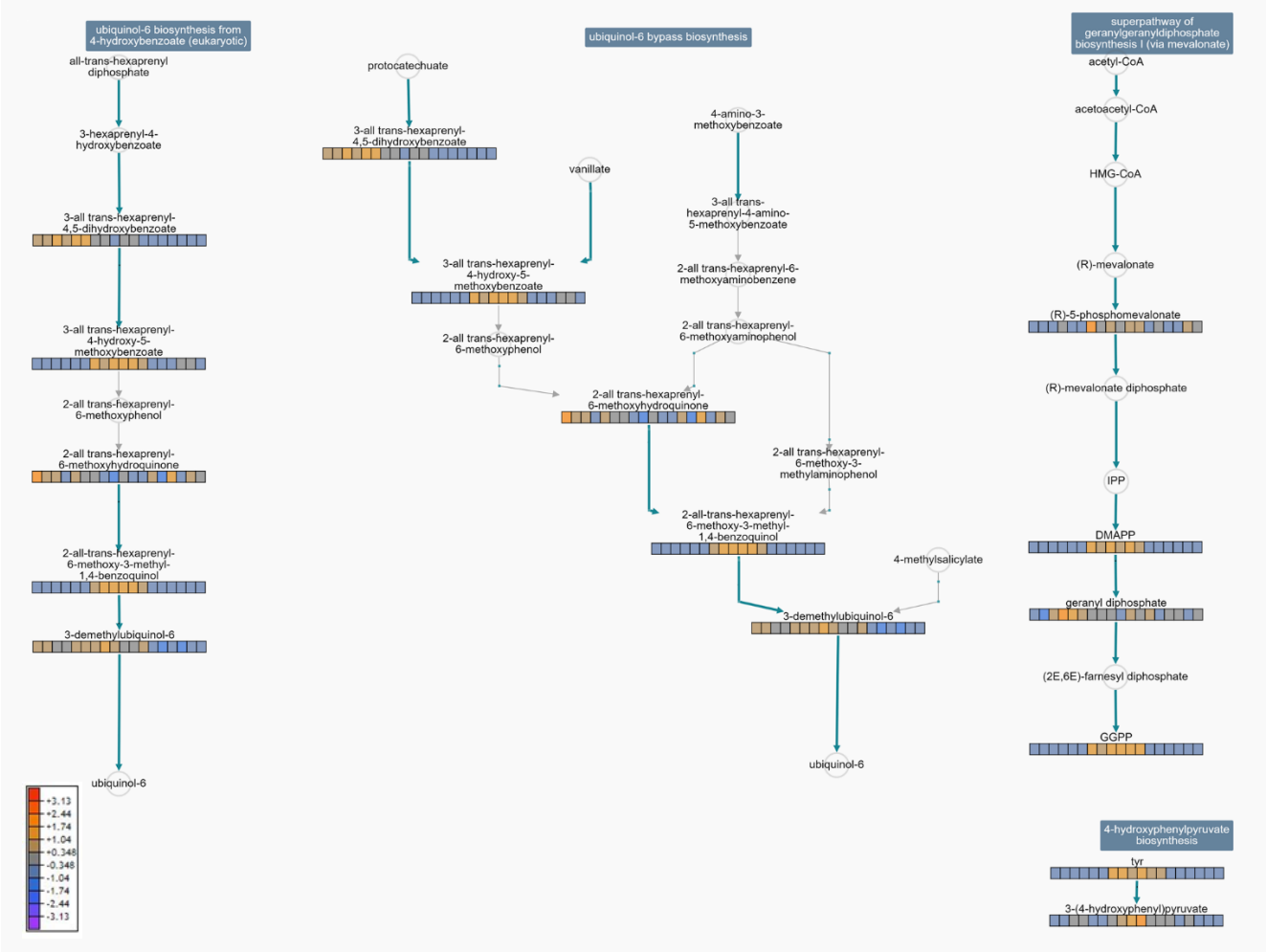


Figure 5.12. Pathway collage of three of the top 10 implicated pathways in the loss of function effect (via BioCycs cellular overview) overlaid with *Δrnq1*, [*rnq*$^{-}$] and [*RNQ*$^{+}$] data. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the middle six boxes belong to [*rnq*$^{-}$] samples and the last six boxes to [*RNQ*$^{+}$] samples. A larger version of this Figure is available in Appendix H.

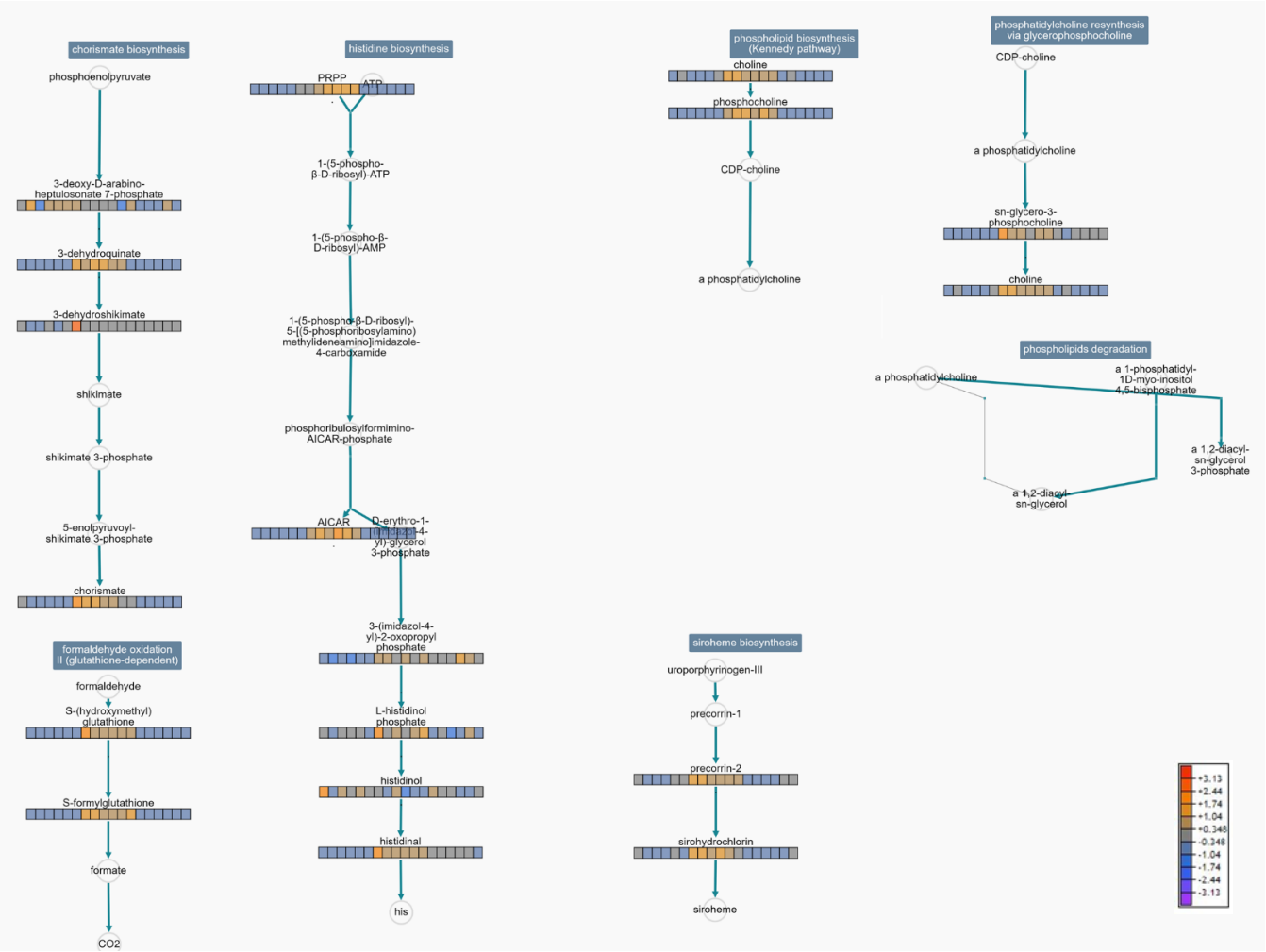


Figure 5.13. Pathway collage of seven of the top 10 implicated pathways in the loss of function effect (via BioCycs cellular overview) overlaid with *Δrnq1*, [*rnq*$^-$] and [*RNQ*$^+$] data. A colour legend is shown in the bottom right-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples, the middle six boxes belong to [*rnq*$^-$] samples and the last six boxes to [*RNQ*$^+$] samples. A larger version of this Figure is available in Appendix H.

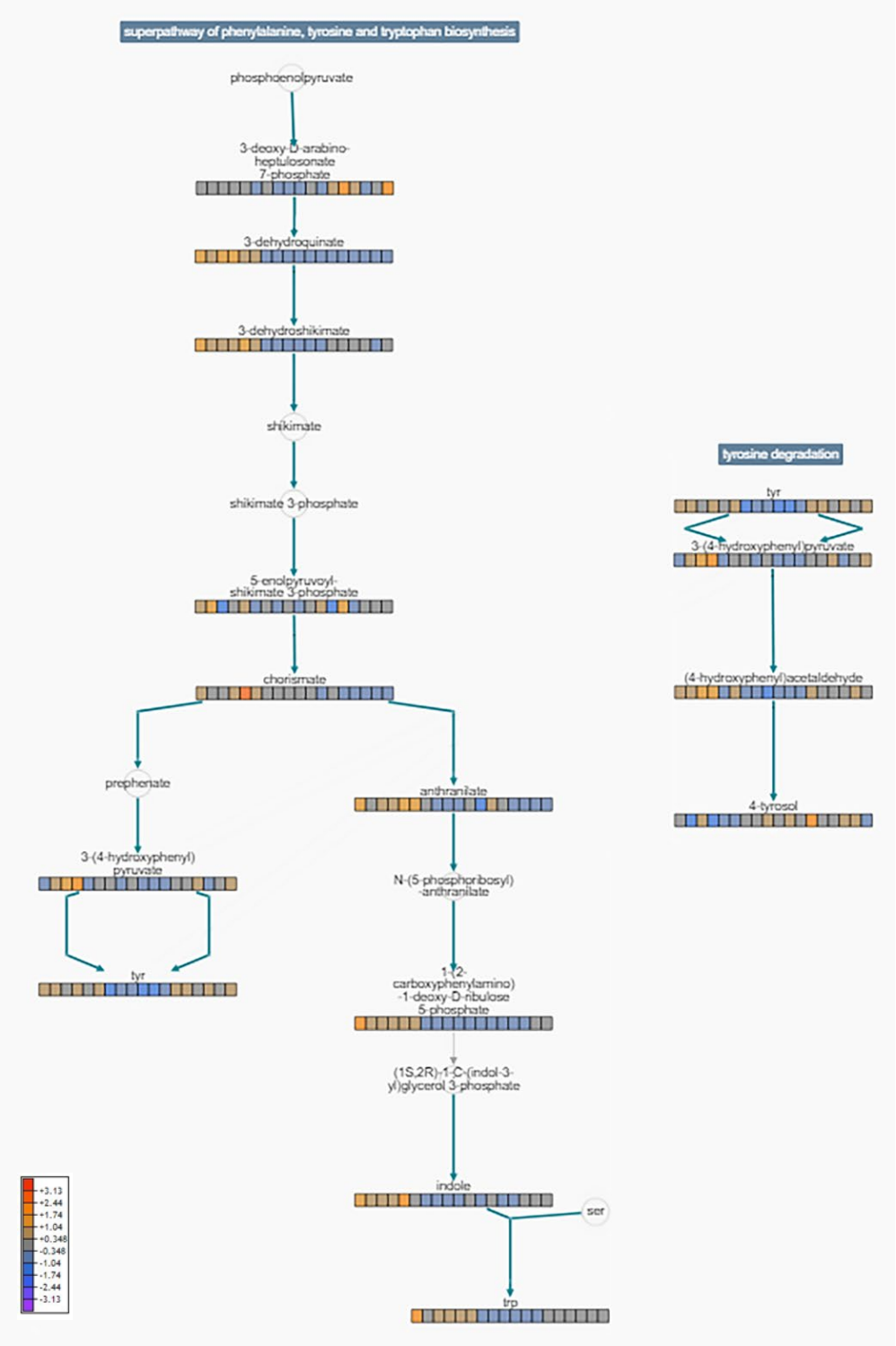


Figure 5.14. Pathway collage of four of the top 10 implicated pathways in general stress response (via BioCycs cellular overview) overlaid with [*rnq*$^-$], [*RNQ*$^+$], [*rnq*$^-$] with mild oxidative stress data. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to [*rnq*$^-$] samples, the middle six boxes belong to [*RNQ*$^+$] samples and the last six boxes to [*rnq*$^-$] with mild oxidative stress samples. A larger version of this Figure is available in Appendix H.

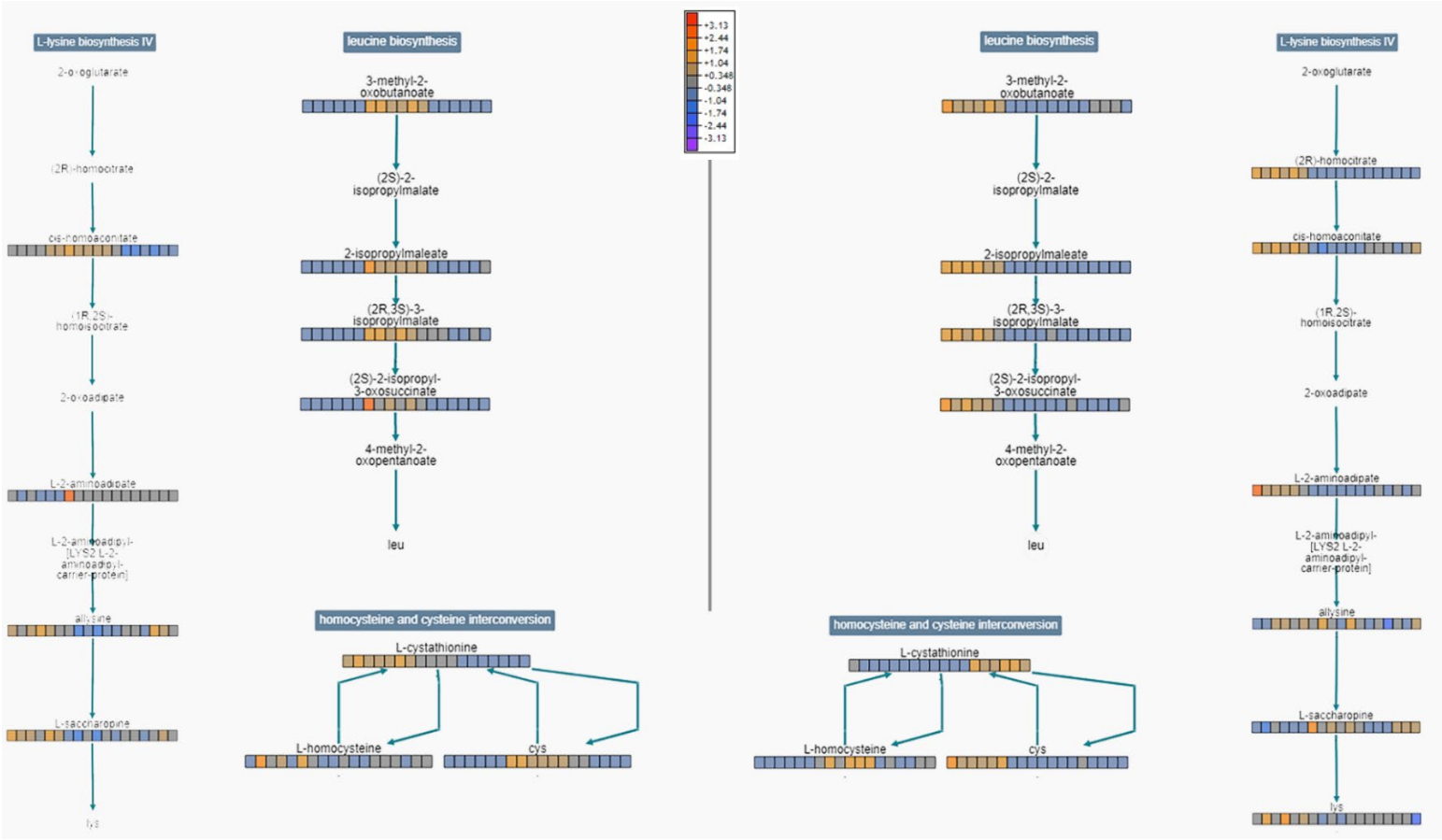


Figure 5.15. Pathway collage of three of the top 20 implicated pathways in metabolic comparisons of [*rnq*$^-$] and [*RNQ*$^+$] (via BioCycs cellular overview) overlaid on the left with *Δrnq1*, [*rnq*$^-$] and [*RNQ*$^+$] and on the right with [*rnq*$^-$], [*rnq*$^-$] with mild oxidative stress and [*RNQ*$^+$] data. A colour legend is shown in the top middle to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' left of the page belong to *Δrnq1* samples and the middle six boxes belong to [*rnq*$^-$] samples and the last six boxes to [*RNQ*$^+$] samples. The first six boxes in any 'heat block' right of the page belong to [*rnq*$^-$] samples and the middle six boxes belong to [*rnq*$^-$] with mild oxidative stress samples and the last six boxes to [*RNQ*$^+$] samples. A larger version of this Figure is available in Appendix H.

5.4.7. Further pathway and metabolite investigations reveal another potential prion specific perturbation

The most perturbed pathways identified by BioCyc were hits to sphingolipid metabolism (Table 5.3). However, although the DPPS scores for these hits were high, one of the metabolites, sphinganine, dominated this result. It was suspected that sphinganine may be falsely identified and that this single metabolite should be removed from the data set owing to the large effect it appeared to be having on pathway hits. Once sphinganine was removed, analysis was repeated. With the exception of the first two hits, the pathways most perturbed in this new analysis remained the same. After attributing pathway hits to either loss of Rnq1 function or generalized stress response, it became clear that specific metabolites and pathway perturbations linked to prion formation or influence could not be clearly identified. As a consequence, these removed sphingolipid metabolism hits were investigated further, requiring further experimentation for validation.

Attempts to combine all of the sample data, similar to that shown in Figure 5.15 were made; however, *Δrnq1* and [*rnq*$^{-}$] with mild-oxidative stress samples lacked data, with only one metabolite hit within the sphingolipid pathways. Figure 5.17 shows the pathway overlay with data from [*rnq*$^{-}$] and [*RNQ*$^{+}$] samples, showing a clear downregulation in the [*RNQ*$^{+}$] samples. Interestingly the relationship between the sphingolipid pathways and protein misfolding is well documented (Varma *et al.* 2018; Wilkins and Trushina, 2018; Laurens *et al.* 2015; Han *et al.* 2011; Oresic *et al.* 2011). These results shown here demonstrate the predicted directionality due to this metabolic perturbation. This type of clear molecular mechanistic similarity between yeast prions and mammalian prions/amyloids has not been documented, despite the common inference that they are indeed present. Data for only one metabolite within these pathways was present in all biological class and comparisons of this in Figure 5.18 show that only [*RNQ*$^{+}$] samples appear to have significant and change in sphinganine reading.

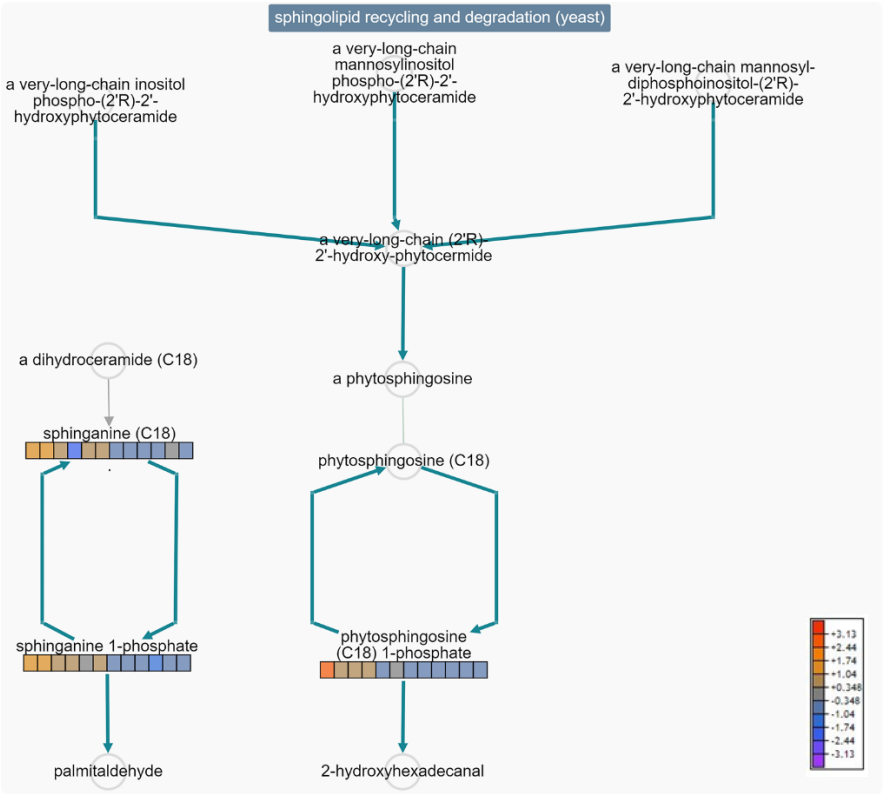


Figure 5.16. Pathway collage of sphingolipid recycling and degradation in yeast (via BioCycs cellular overview) with standardised omics [*rnq*$^-$] and [*RNQ*$^+$] data of individual metabolites overlaid. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to [*rnq*$^-$] samples and the last six boxes belong to [*RNQ*$^+$] samples. A larger version of this Figure is available in Appendix H.

Figure 5.17. Box and whisker plot of sphinganine relative abundances with [*rnq*$^-$], [*RNQ*$^+$], [*rnq*$^-$] with mild oxidative stress and *Δrnq1* data. Darker lines in the body of the box depict the mean line and error bars show quartile ranges from the mean. Legend shown below details the sample group reading represented by each colour.

5.5. Results – Comparison of the metabolomics perturbations observed between [*rnq*$^-$] with mild-oxidative stress and [*RNQ*$^+$] with mild-oxidative stress

5.5.1. PCA and OPLS-DA modelling reveals no significant difference between [*rnq*$^-$] with mild oxidative stress *vs* [*RNQ*$^+$] with mild oxidative stress

Rather unexpectedly, [*RNQ*$^+$] with a mild-oxidative stress appears much more closely 'related' to [*rnq*$^-$] oxidative stress than to [*RNQ*$^+$] samples. Modelling these sample groups would permit the identification of any significant differences in metabolic response due to the presence of absence of the [*RNQ*$^+$] prion under mild-oxidative stress.

Figure 5.18. (A) shows that PIM PCA modelling positioned just one of the [*rnq*$^-$] mild oxidative stress samples alongside the [*RNQ*$^+$] mild oxidative stress samples rather than grouped with the other [*rnq*$^-$] oxidative stress samples. Given the propensity for [*rnq*$^-$] cells under stress to randomly acquire [*RNQ*$^+$] prion, it was considered that this sample had acquired [*RNQ*$^+$] status, however this was not seen in Figure 5.18. (C) NIM PCA models. This sample was considered acceptable and the relative clustering of all other samples into groups was considered encouraging. However, the PIM PCA model plotted reasonable scores with R^2 = 0.691, Q^2 = 0.339, whereas the NIM PCA model was relatively acceptable with R^2 = 0.494, but poor Q^2 = -0.145, indicating that this model was not predictive.

Assessment of CV-ANOVA scores were used to evaluate the significance of these differences given the poor PCA models. Strong OPLS-DAs models were achieved, with the PIM model giving an $R^2(X)$ = 0.624, $R^2(Y)$ = 0.979, Q^2 = 0.695 and CV-ANOVA *p-value* = 0.126, and the NIM model reporting an $R^2(X)$ = 0.496, $R^2(Y)$ = 0.970, Q^2 = 0.665 and a CV-ANOVA *p-value* = 0.160 (shown in Figure 5.18. B and D). CV-ANOVA values reveal that although some clustering was seen in the PIM PCA model, no significant difference in metabolic profile exists between the stressed groups irrespective of prion status. As a result of this, no further pathway analysis was conducted on this data set.

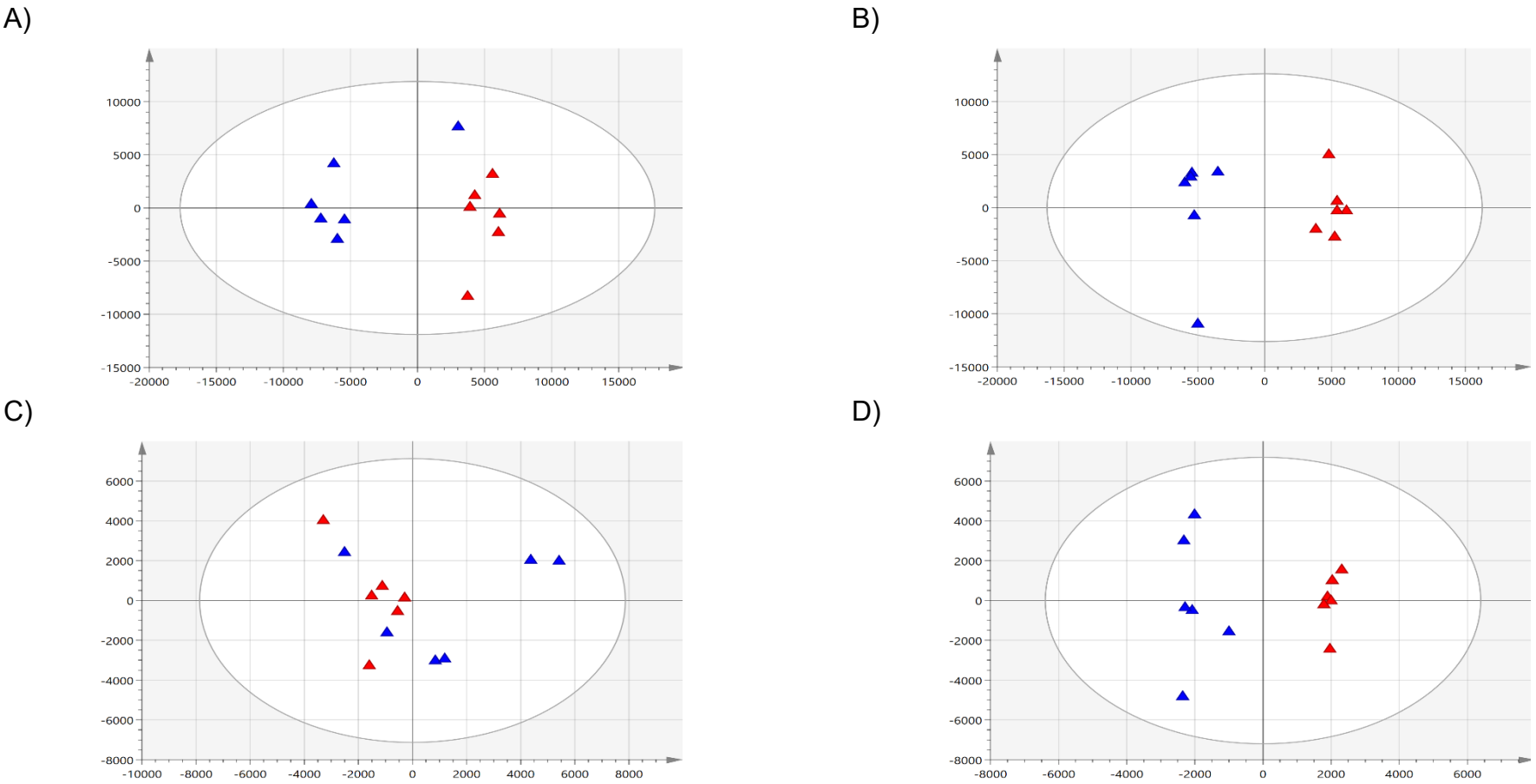


Figure 5.18. PCA (A and C) and OPLS-DA (B and D) models of XCMS/MZmine2 comparative data of [*rnq*$^-$] with a mild oxidative stress and [*rnq*$^-$] with a mild oxidative stress. Overview of the data shows no outlying samples within the 95% confidence range within each model. Blue triangles show [*rnq*$^-$] with a mild H_2O_2 stress and red triangles show [*RNQ*$^+$] with a mild H_2O_2 stress. (A) Comparative PIM data PCA scores with $R^2 = 0.691$ $Q^2 = 0.339$ values. (B) Comparative PIM data OPLS-DA scores with $R^2(X) = 0.624$, $R^2(Y) = 0.979$, $Q^2 = 0.695$ and CV-ANOVA *p-value* = 0.126. (C) Comparative NIM data PCA scores with $R^2 = 0.494$, $Q^2 = -0.145$ values. (D) Comparative NIM data OPLS-DA scores with $R^2(X) = 0.496$, $R^2(Y) = 0.970$, $Q^2 = 0.665$ and a CV-ANOVA *p-value* = 0.160.

5.6. Discussion

The experiments described in this chapter attempt to further elucidate the effect that presence of the [*RNQ*$^{+}$] prion had on the underlying cellular functioning of yeast cells. However, it was clear that certain factors had the potential to confuse results, namely loss of function metabolic effects and generalised stress response. Therefore, in order to allow for discrimination between these results, an appropriate stress control was conducted (section 5.3.) and data from chapter four was utilised where appropriate. Loss of Rnq1 function and a generalised stress response were found to have had a significant impact on the pathways perturbed between [*rnq*$^{-}$] cells and [*RNQ*$^{+}$] cells. Whereas pathways involved in sphingolipid metabolism (specifically a downregulation in sphingolipid recycling and degradation) appear to be perturbed solely due to prion presence.

5.6.1. Overlapping effects of loss of function and generalised stress response, their implications and potential for future research

Cross comparison with data from chapter four revealed that approximately 70% of the pathways perturbed when comparing [*rnq*$^{-}$] cells to [*RNQ*$^{+}$] cells are as a direct result of the loss of function of the Rnq1 protein. This may be the case for any of the yeast prions and these results substantiate the findings of chapter four. Knowledge of loss of function of effects within the prion field have been used to develop easy to use assays capable of reporting the strength of the strain and the roles of distinct sections of the protein. Such knowledge may be extremely valuable in the development of an assay or knockout strain for Rnq1 protein similar to those which exist for [*PSI*$^{+}$] and [*URE3*].

Examining the metabolic differences between [*rnq*$^{-}$] cells to [*RNQ*$^{+}$] cells, such an assay may be possible via a number of different experimental approaches. Firstly, in section 4.5. the possibility of using azole compounds to distinguish between [*rnq*$^{-}$] cells and *Δrnq1* based on their changeable sterol composition due to upstream changes with the ubiquinol pathways was discussed. As these same changes have been observed in [*RNQ*$^{+}$] cells (whether due to loss of function or not) these sterol changes and hence variable azole sensitivity could be used to select for [*RNQ*$^{+}$] variants. With [*rnq*$^{-}$] cells having a higher concentration of ergosterols, the target for the azoles (Rodrigues, 2018), their tolerance of azoles should be low. Whereas in [*RNQ*$^{+}$] cells, having a lower concentration of the ergosterol target, their tolerance for the azoles should be much higher. This potential variability in azole sensitivity could result in ‘doses’ being established, which would enable the separation of prion and

non-prion strains. Secondly, the changes in tryptophan and tyrosine biosynthesis may be of interest, as both were downregulated in [*RNQ*$^{+}$] cells. Colorimetric and fluorometric assays are available for both amino acids (Sigma) and establishing the content of either or both of these amino acids may be a reliable way to distinguish between [*RNQ*$^{+}$] cells and [*rnq*$^{-}$] cells. However, this would rely on sufficient concentrations of these amino acids for the assays to be effective.

Amino acid restriction in *S. cerevisiae* is beneficial to the overall health of cells, promoting longevity (Hu *et al.* 2014; Wu *et al.* 2013; Eisenberg *et al.* 2012). The results of this work show a downregulation in histidine, leucine, and choline in [*RNQ*$^{+}$] cells, possibly instigating the types of dietary restrictions known to have positive effects on longevity. Wang *et al.* (2017) found that the presence of the [*PSI*$^{+}$] prion prolonged the chronological lifespan of *S. cerevisiae*. The downregulation of amino acid pathways shown here in [*RNQ*$^{+}$] cells appears to correlate well with these findings. Hence further investigation into the comparative chronological lifespan of [*rnq*$^{-}$] and [*RNQ*$^{+}$] cells is recommended.

The shared pathways between the mild-oxidative stressed condition and [*RNQ*$^{+}$] samples, suggests that the role prions play within yeast cells is not as benign as previously thought. Perhaps evidence confirming this observation can be found within the pathways that are not implicated within [*RNQ*$^{+}$] samples. Comparisons of Table 5.2. (stress condition pathways) and Table 5.3. (prion presence pathways) reveal that proline and arginine degradation are unique to the stress condition. This may point to potential biomarkers for distinguishing between damage done as a result of oxidative stress and damage as a result of misfolded protein presence, although further research is needed to validate this.

5.6.2. [*RNQ*$^{+}$] specific perturbations, limitations, and further research

The main perturbation that occurred due to the [*RNQ*$^{+}$] prion appeared to be the down regulation of sphingolipid recycling and degradation.

The sphingolipids, alongside the phospholipids and the sterols, are synthesised mainly within the endoplasmic reticulum (ER) (Hebert and Molinari, 2007) an organelle that plays a key role in the promotion of the correct assembly and folding of newly synthesised proteins. Due to this essential role, monitoring of the ER is crucial to cellular success and when subject to stresses (such as the accumulation of misfolded proteins and perturbations in lipid

synthesis) the ER is acted heavily upon by the unfolded protein response or UPR, which aims to assist protein folding, remove misfolded proteins, and promote lipid synthesis (Travers *et al.* 2000). ER stresses are known to initiate a “diffusion barrier response”, which prevents damaged proteins and metabolites from entering daughter cells during budding (Clay *et al.* 2014). Experiments on Hsp70 revealed that this barrier prevents misfolded proteins from entering daughter cells and confirm that the sphingolipids play a key and vital role in this compartmentalization and barrier formation (Clay *et al.* 2014). Posited to form the bottom part of a fatty molecules layer, the sphingolipids, are thought to act as a skeleton to allow this diffusion barrier to form (Clay *et al.* 2014). Given that the mechanism which prions use to proliferate from mother to daughter cells is now understood to involve ‘an unknown’ mechanism of partitioning (Ness *et al.* 2017). It seems reasonable, based on the observations of changes in sphingolipid degradation, to suggest that these diffusion barrier forming, sphingolipids may have a critical role within this partitioning between mother to daughter cells, although further investigation is needed.

Our results show that when prions are present, a downregulation event occurs within the degradation and recycling of the sphingolipids, not a downregulation of sphingolipid production itself. This could be due to the identification of common metabolites within the degradation pathways or that metabolites within the synthesis pathways were not measured and identified, either due to either their lipid nature or chance. The result being that the highlighted pathway for the known or identified metabolites would always revert to the smallest pathway known to contain them, hence the degradation pathways. Alternatively, it is possible that degradation of the sphingolipids is lower in [*RNQ*$^{+}$] samples due to the cells simply using all the sphingolipid resources available to them to deal with the number of misfolded proteins. Indeed, within amyloid biology it is well documented that microdomains are formed by the sphingolipids and cholesterols which act as sites for the binding and oligomerisation of amyloidogenic proteins (Fantini and Yahi, 2010).

Within cells carrying the mammalian prion (PrP^{SC}), sphingolipid rafts present in cell plasma membranes, are known to decrease in concentration (Naslavsky *et al.* 1999). It is proposed that sphingolipid rafts are used by cells to aid in the formation of the PrP^{SC} from its normal conformer PrP^{C} (Naslavsky *et al.* 1999). This observation by Naslavsky *et al.* (1999) helps to support the observations seen in this study, where sphingolipid recycling and degradation are downregulated in the presence of [*RNQ*$^{+}$], it is reasoned that this downregulation may be as a result of continued use or containment of sphingolipid rafts within the process of

misfolding and subsequent aggregation of [*RNQ*$^{+}$]. Despite the common inference that connects the mechanistic actions of mammalian prions/amyloids with the yeast prions, this has not previously been observed in yeast and provides new evidence of homology between yeast prions and mammalian prions/amyloids. The results from this research may further encourage the use of this easy to use and readily available model organism within a wider range of neurodegeneration studies, hence rapidly increasing our fundamental understanding of these mechanisms.

This research has established that metabolic differences exist between [*rnq*$^{-}$] and [*RNQ*$^{+}$] cells indicating that prion presence is not without metabolic consequence. Similar experiments using the 74D-strain variants of [*RNQ*$^{+}$] (which include low, medium, high, and very high strength [*RNQ*$^{+}$]) may aid in the discovery of biomarkers for the quantification of the strength of prion variants. As there are many other known yeast prions and several model organisms commonly in use for amyloid studies, such as *Caenorhabditis elegans*, *Drosophila melanogaster* and *Danio rerio* (Carlo, 2012), this work could be extended to examine if homologous metabolic perturbations are seen for other prion proteins, potentially identifying a specific set of metabolic pathways unique for prion presence. As perturbations in phospholipids and sterols pathways have been identified here the use of revised methodology such as GC-MS for sterol capture and lipidomics is recommended to obtain better quality information.

In addition to the prion-specific downregulation observed within sphingolipid degradation and recycling, we also observed a prion-specific down regulation in tyrosine metabolism. Hilaly *et al.* (2016) in their work on PD (Parkinson's Disease) found *in vitro* that tyrosine residues in patients with PD are sequestered into covalently crosslinked homodimers that appear to play a key role in the formation of seed and oligomeric species. Without confirmation of the presence of such a dimer in yeast, it is possible to only speculate that such an event may be occurring within yeast prions too and so enabling the formation of the oligomeric stages that precede amyloid formation. Further research to investigate if the downregulation of tyrosine in [*RNQ*$^{+}$] cells is due to similar crosslinked homodimer being formed using tyrosine residues would be advisable. This would provide further evidence of the mechanistic links of oligomeric formation between yeast prions and amyloid formation.

5.6.3. Stress induction in [*rnq*$^-$] and [*RNQ*$^+$] cells requires further research and highlights the need for prion assays

The lack of an observable difference between stressed conditions regardless of their prion status, is a puzzling and unanticipated result, one which may be inaccurate. This contradicts the literature regarding the effect of prion formation on cells and their capacity under stress. Where have the loss of function effects of the prion gone? Does Rnq1 not lose its function entirely? A plausible explanation for such an observation is that the increase in oxidative stress to cells although mild had a direct influence on the number of Hsp and UPR within the cell. These response mechanisms may have been able to cope with the misfolded forms of proteins at a rate sufficient to reduce the templating action to daughter cells and essentially act as a curing agent. If this is the case though, it begs the question is mild oxidative stress a curing agent? This seems incredibly unlikely, due to the extensive experimentation done by Doronina *et al.* (2015) which repeatedly evidenced an increase in the devo formation of [*PSI*$^+$] due to oxidative stress, however perhaps the influence of this condition varies depending on the prion studied (Grant, 2015).

Upon reassessment of the PCAs in Figure 5.18. the PIM PCA does show significant levels of separation, although the NIM one does not. The decision to not analyse this data any further based on the *p-values* of OPLSDAs CV ANOVAS may have been an inaccurate one given the PCA separation and the contradictory findings here. It would therefore be preferable to repeat this analysis, especially considering a new feature of Metaboanalyst which allows for the combination of PIM and NIM data.

The decision not to assay for prion formation during this experiment was due to confidence in the prion status of the cell lines used with their status confirmed of within other work alongside or prior to this research (data not shown). However, it would have been advisable to assay for prion formation in all the biological classes used within this experiment, given the reactive nature of prion formation with stress. Common assays which had been conducted on these strains in other experimentation include western blotting and the addition of GFP protein tags with subsequent visualisation of foci.

Kryndushkin *et al.* (2003) procedure for the western blotting of prions requires the growth of cells in broth until an OD_{600} of 0.6 is reached. These cells are then lysed using lysis buffer, glass beads and centrifugation. Protein concentration of lysates using a BSE Bradford assay is established to normalise cell concentrations. Lysates are then run on agarose gels, SDD-

AGE, as opposed to the common polyacrylamide SDS-PAGE. Proteins from these agarose gels, which can separate large aggregates from smaller ones, are then transferred to PDVF or nitrocellulose membrane using an electric current. The membrane now containing the proteins, is first probed with a weak milk solution to 'fill the gaps' and then between washes a primary antibody specific to the Rnq1 protein and a secondary antibody (bound to horse radish peroxidase (HRP)) capable of binding to the primary antibody are added. The membrane is then imaged via enhanced chemiluminescence (ECL) of HRP, thus allowing for a semi quantitative detection of the Rnq1 protein and its aggregates. The size of the bands on the image are used to establish the presence of aggregates and hence the prion status.

Alternatively, cells may be transformed with a plasmid containing *RNQ1-GFP* insert, using the transformation protocol described in section 2.4.3. Fluorescence microscopy of individual transformed cells can be imaged for the presence of diffuse GFP, as would be expected of [*prion*$^{-}$] variants, and distinct foci, as would be expected of [*prion*$^{+}$] variants (Vitrenko *et al.* 2007). Another possible option would be to use the common amyloid diagnostic dye thioflavin-T, which binds to amyloid aggregates and is detectable via fluorescence (excitation 450 nm, emission 485 nm) (Douglas *et al.* 2008).

Overall, our results supported by the current literature indicate that [*RNQ*$^{+}$] prions have an effect and/or are interacting with the sphingolipid pathways. Confirmation of the findings of our loss of function experimentation in chapter four have been acquired and new metabolic information regarding both general stress response and specific stress responses of oxidative stress on cells both with and without prions has been obtained.

Chapter Six - The metabolic perturbations associated with toxicity on overexpression of Rnq1 protein

The experiments within this chapter attempt to identify the metabolic pathways that are perturbed by induced toxicity caused by the overexpression of the *RNQ1* gene in a [*RNQ*$^{+}$] background. This was achieved by comparing four biological classes: [*rnq*$^{-}$], [*RNQ*$^{+}$], [*rnq*$^{-}$] with overexpression of Rnq1 and [*RNQ*$^{+}$] with overexpression of Rnq1. The pYES2 plasmid containing the *RNQ1* gene under the control of the strong, inducible GAL promoter (Table 2.6) was transformed via methods detailed in section 2.4. into [*rnq*$^{-}$] and [*RNQ*$^{+}$] prior to experimentation and maintained using synthetic complete -ura media (see methods section 2.2.1.). Galactose induction of the *RNQ1* gene was initiated by a switch to synthetic complete - ura media containing galactose as a carbon source when cultures had reached an OD_{600} of 0.5. All biological classes were then allowed to grow on galactose media for up to four hours, with a sampling of cells being removed at one-hour intervals. Metabolite extraction, UHPLC-MS analysis, and data analysis was conducted as detailed in chapter two.

6.1. Introduction

A long-standing question within the amyloid field has been the role that amyloids play within disease pathology. It has been well established that amyloid formation can result in degradation of affected tissues which quickly led to the conclusion that the assembly of amyloid fibres is a pathological process (Jackson and Hewitt, 2017; Chiti and Dobson, 2006). However, throughout the preceding decade of study, many functional amyloids have been discovered that play host to a large range of physiological process (Jackson and Hewitt, 2017; Audas *et al.* 2016). As such, these discoveries have cast doubt over conclusions that had previously been made regarding the toxicity of amyloid formation.

The formation of amyloid fibrils is as a result of a cascading event via multiple oligomeric species (detailed in section 1.1.5.) and hence identifying the culprits of toxicity and elucidating their actions has remained a challenging priority (Sipe *et al.* 2016). These oligomers are often transient in nature, with multiple different forms present at any one time and despite containing the same peptide coding, these proteins are capable of producing both toxic and non-toxic effects (Campioni *et al.* 2010). What is clear about the mechanisms

of toxicity is that (the permeable) lipid membranes appear to be a major target for these prefibrillar oligomers (Bucciantini *et al.* 2004).

Given the relationships with the major lipid metabolites (see section 4.5. and 5.6.), it was possible that one may be able to further explore these toxic effects within yeast via the [*RNQ*$^{+}$] prion. The major problem being that, as explained in section 5.1., the yeast prions are not thought to present a disease phenotype *per se*, and so toxic, cell-death inducing effects were not available to monitor. However, Douglas *et al.* (2008) found that despite being considered generally benign, the moderate (approximately 5-10 fold) overexpression of the Rnq1 protein in cells which carried the [*RNQ*$^{+}$] prion was moderately toxic to cells.

Alongside this Douglas *et al.* (2008) found that the same effects could not be seen in deletant strains or in cells with the non-prion conformation of Rnq1, [*rnq*$^{-}$]. These findings were replicated in a number of different strains that consistently exhibited a number of extreme growth defects, with overexpression in a [*RNQ*$^{+}$] background resulting in approximately 25% cell death within four hours. Most intriguingly, Treusch and Lindquist's (2012) experimentation detailed how inherent toxicity of this overexpression is not as a result of generalised proteomic stress but is instead part of a highly specific and orchestrated mitotic arrest.

Further experimentation involving the molecular chaperones known to be involved with the conversion of soluble Rnq1 into amyloid fibrils, confirmed that upregulation of these chaperones eliminated the observed toxicity in [*RNQ*$^{+}$] cells with Rnq1 protein overexpression (Sondheimer *et al.* 2001). This provided evidence that insufficient chaperone conversion to amyloid allows for the formation of other toxic Rnq1 conformers. Such toxic Rnq1 conformers do not form in [*RNQ*$^{+}$] cells as molecular chaperones sequester the misfolded protein into amyloid aggregates; however, when overexpressed, the molecular chaperone system becomes overwhelmed allowing for the formation of these toxic conformers (Douglas *et al.* 2009; Douglas *et al.* 2008).

Additionally, a similar effect was observed using polyQ-containing proteins, highlighting the role of the [*RNQ*$^{+}$] prion as a mediator to prion formation (Meriin *et al.* 2002). However, the lack of a mammalian homolog has meant that the details of this relationship have remained elusive. It seemed reasonable to ask if the key metabolic disruptions that arise as a result of the [*RNQ*$^{+}$] prion are conducive to polyQ aggregation and toxicity. Furthermore, are there

disruptions to mammalian counterparts that may be induced by an entirely different protein, albeit one acting on the same metabolic networks.

Given the complexity surrounding the manifestation of toxicity, it is hoped that by inducing this toxic effect of overexpression the detection of key specific metabolic perturbations that are intrinsically linked to cell death may be possible. Such information could provide information regarding the specific metabolic process that are being disrupted by these oligomeric protein species and may identify regions of the metabolome that could act as potential targets to eliminate these effects.

In order to conduct these experiments, the Rnq1 protein was overexpressed in [*RNQ*$^{+}$] cells. Using a kind gift from the Kent Fungal Group, the pYES2 plasmid with GAL promoter, URA3 selective marker and *RNQ1* insert. The GAL promoter is induced by the presence of galactose in the media. These growth conditions are significantly different from previous experimentation and so metabolic perturbation data obtained previously could not be used for comparative purposes. Thus, to ensure parity, all test strains and experiments were grown under identical conditions to control for metabolic changes caused by using galactose as a carbon source. The need for such controls within metabolic experimentation is stressed via the recent discovery by Alam *et al.* (2016) whose study found that the detectable metabolic disruption attributed to a variety of the commonly used gene deletions in yeast was large and widespread across many areas of the metabolome.

6.2. Results

6.2.1. Feature detection and normalisation

Data was placed into the four-time groupings, shown in Table 6.1., as the complexity and size of the data files did not allow for an entire single group analysis to be conducted. Cross correlation of the m/z values from each time-groupings XCMS and MZmine2 analysis was achieved via the methods outlined in section 2.8.4. Reported values from each program individually and comparatively are shown in Figure 6.1. (A-D). Overlapping comparative features were used to build statistical models via SIMCA and perform pathway analysis via Metaboanalyst and BioCyc.

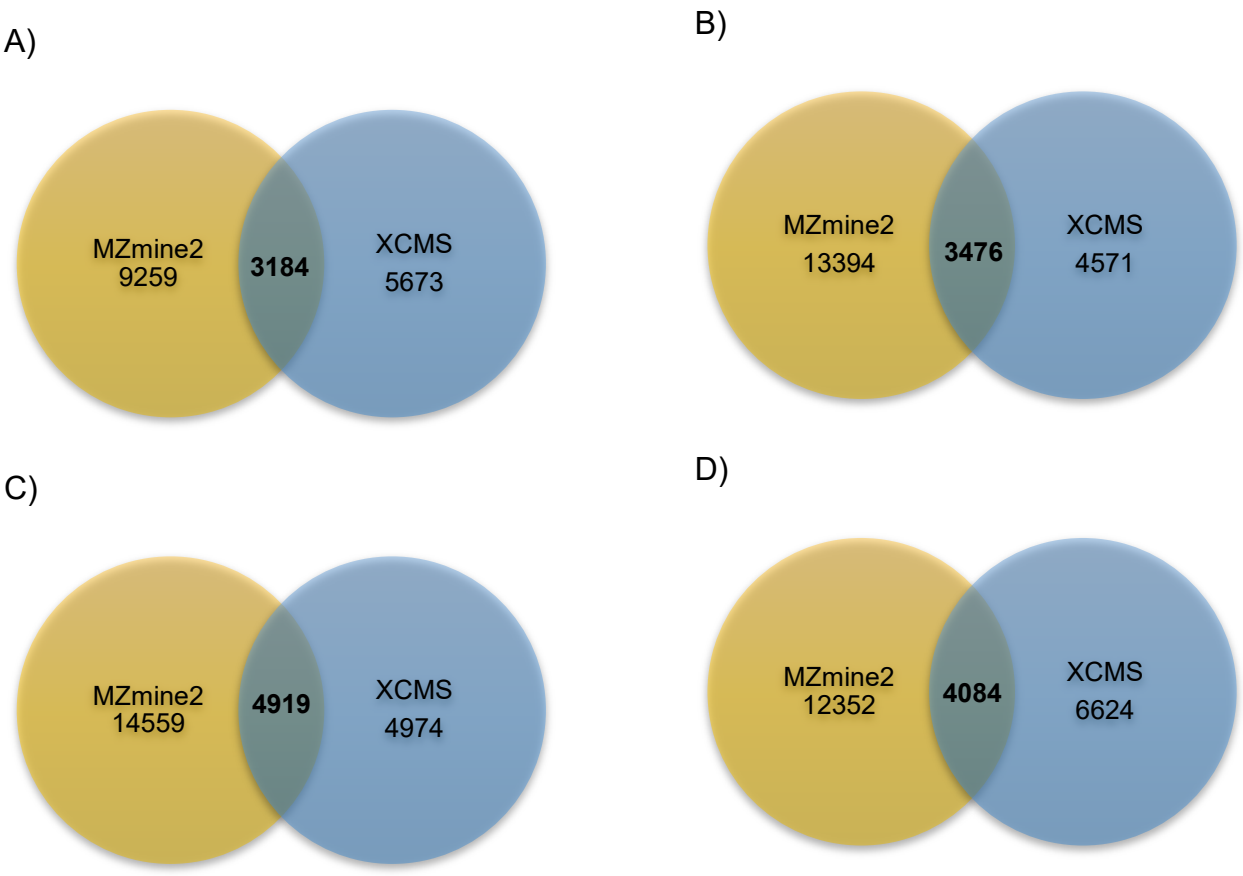


Figure 6.1. Venn diagrams showing the total number of features reported by MZmine2 and the total number of features reported by XCMS when considering [*rnq*$^-$], [*rnq*$^-$] with overexpression of Rnq1, [*RNQ*$^+$] and [*RNQ*$^+$] with overexpression of Rnq1 data from four separate time points. The overlapping section in the centre of the two depicts the total number of m/z values reported by both analysis programmes in bold A) Venn diagram of T1 m/z values of MZmine2 and XCMS A) Venn diagram of T2 m/z values of MZmine2 and XCMS A) Venn diagram of T3 m/z values of MZmine2 and XCMS A) Venn diagram of T4 m/z values of MZmine2 and XCMS

6.2.2. PCA and PLSDA group modelling of XCMS/MZmine2 comparative metabolomic data in SIMCA highlights time as strongest metabolic divider

All 96 samples were used to build initial PCA and PLS-DA models, consisting of 6 samples from each of the sample groups: [*rnq*$^-$], [*rnq*$^-$] with overexpression of Rnq1, [*RNQ*$^+$] and [*RNQ*$^+$] with overexpression of Rnq1. This would establish the important factors and relationships present within the dataset as a whole.

These initial models were strong, with good confidence levels. The PCA reported R^2 and Q^2 values of 0.718 and 0.509 respectively, and the PLS-DA reported $R^2(X) = 0.659$, $R^2(Y) = 0.939$, $Q^2 = 0.866$ and CV-ANOVA of 0. Upon examination of the data, it was apparent that the strongest metabolic differences were founded on a similarity in time point, with T1 samples clustering together irrespective of the group the sample belonged to. The same was observed for data from samples for time points T2, T3 and T4 (Figure 6.2). Some differentiation was seen between samples within their time group clusters but this was only visible in PLS-DAs and not sufficiently distinct to establish the relationship between the sample groups with respect to their prion or overexpression status. There appeared to be a relationship between samples from the same collection time, and to investigate this further collection times were modelled independently. To simplify this analysis, symbols were assigned to each sample group as shown in Table 6.2.

Table 6.1. Symbol assignment for sample groups within chapter six.

Sample group	Assigned symbol
[*rnq*$^-$]	W
[*rnq*$^-$] with overexpression of *Rnq*1	X
[*RNQ*$^+$]	Y
[*RNQ*$^+$] with overexpression of *Rnq*1	Z

Describes experimental group conditions and assigns symbols to those groups for future reference.

6.2.3. PCA and PLSDA time separated group modelling of XCMS/MZmine2 comparative metabolomic data in SIMCA differentiates between groups

Visual assessment of the models (Figure 6.3.), confirm the prediction that as time progresses so do the number of metabolic differences between groups, accentuating their group separation. T1's PCA gives a good $R^2 = 0.670$ value but the $Q^2 = 0.386$ is sub-optimal for biological interpretation and shows some moderate group separation. The forced

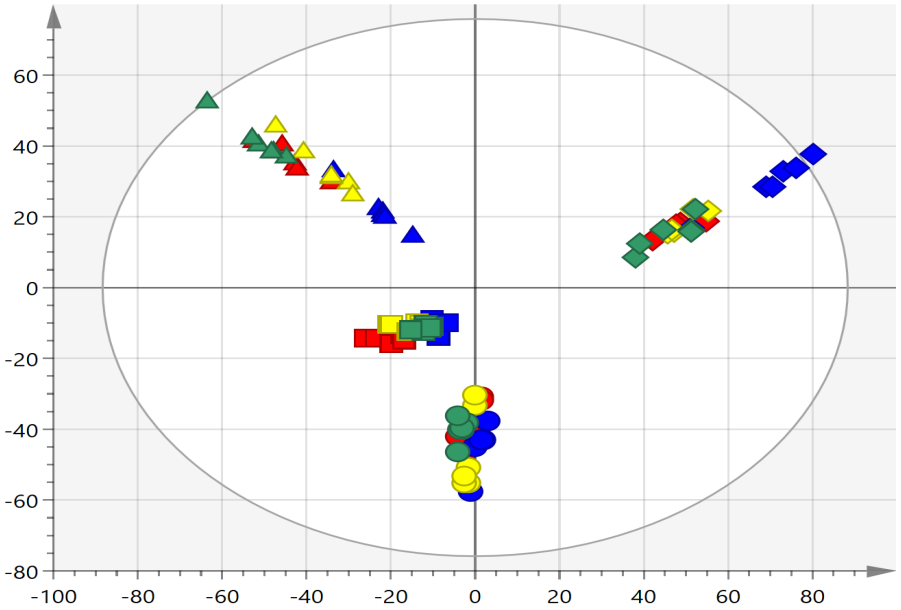
separation of the PLSDA gives a strong model (R2(X) = 0.407, R2(Y) = 0.807, Q2 = 0.508) and a CV-ANOVA *p-value* = 2.71 x 10-2 indicating that even at this early timepoint, significant differences are seen between groups. Explanations for these differences can be attributed to these groupings being different; some containing plasmids, some containing prions.

Continuing to T2, the PCA some moderate, albeit slightly more convincing (than T1) clustering occurs, with group X and Z plasmid-containing samples appearing closely linked and W and Y groups appearing within their own clusters. The model itself suffers similar issues as T1 with respect to its scores, $R^2 = 0.668$, $Q^2 = 0.338$. T2's PLSDA with its $R^2(X) = 0.636$, $R^2(Y) = 0.964$, $Q^2 = 0.863$ and a CV-ANOVA *p-value* = 3.23×10^{-6} showing that significant differences between groups are present at only two hours after plasmid induction. These may suggest differences between [*rnq*$^-$] and [*RNQ*$^+$], but the increase in the clustering and significance suggests that other factors may be present.

T2 appeared to possess the largest pooling of samples due to the presence of the plasmid. At T3, the presence of the prion appears to have become dominant with regard to the characteristic causing clustering, with [*rnq*$^-$] and [*RNQ*$^+$] appearing closely grouped to each other. The T3 PCA score is stronger but the predictability of the model is still less than desired with $R^2 = 0.832$, $Q^2 = 0.440$. T3 PLSDA gave much stronger score plots with $R^2(X) = 0.691$, $R^2(Y) = 0.979$, $Q^2 = 0.830$ and a CV-ANOVA *p-value* = 5.64×10^{-5}. By T4, PCA score plots indicate biological relevance with $R^2 = 0.790$, $Q^2 = 0.528$. Interestingly, this model has group W appearing alone with a closer clustering between the remaining groups than had previously been observed. The T4 PLSDA values indicate a strong significant difference between groups, with clustering of individual groups now clearly observable, $R^2(Y) = 0.967$, $Q^2 = 0.907$ and a CV-ANOVA *p-value* = 1.77×10^{-9}.

Group modelling and observation confirmed that each of the sample groups were distinct from each other; evidence that each of the conditions had been maintained successfully. Clustering behaviour based on time intervals has started to reveal the cellular conditions that most influence the metabolome; however, in order to investigate this further, detailed two-group comparisons will be needed to extract the significant metabolic changes between sample groups.

A)



B)

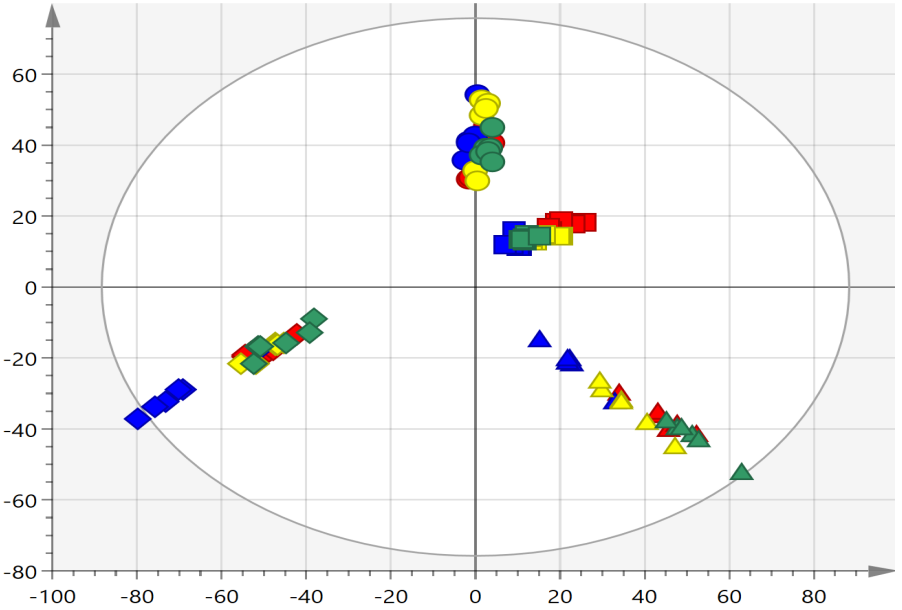


Figure 6.2. PCA (A) and PLS-DA (B) models of XCMS/MZmine2 comparative data of [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X), [*RNQ*$^+$] (Y) and [*RNQ*$^+$] with overexpression of Rnq1 (Z) data from four separate time points. Overview of the data shows no more than two outlying sample within the 95% confidence range within each model. The PCA model, with $R^2 = 0.820$ $Q^2 = 0.707$ values, is shown on the left and the PLS DA model, with $R^2(X) = 0.757$, $R^2(Y) = 0.715$, $Q^2 = 0.540$ and a CV-ANOVA *p-value* = 0×10^{-7}, shown on the right. Circles represent T1, squares T2, triangles T3 and diamonds T4, green shows W, yellow shows X, red show Y and blue shows Z.

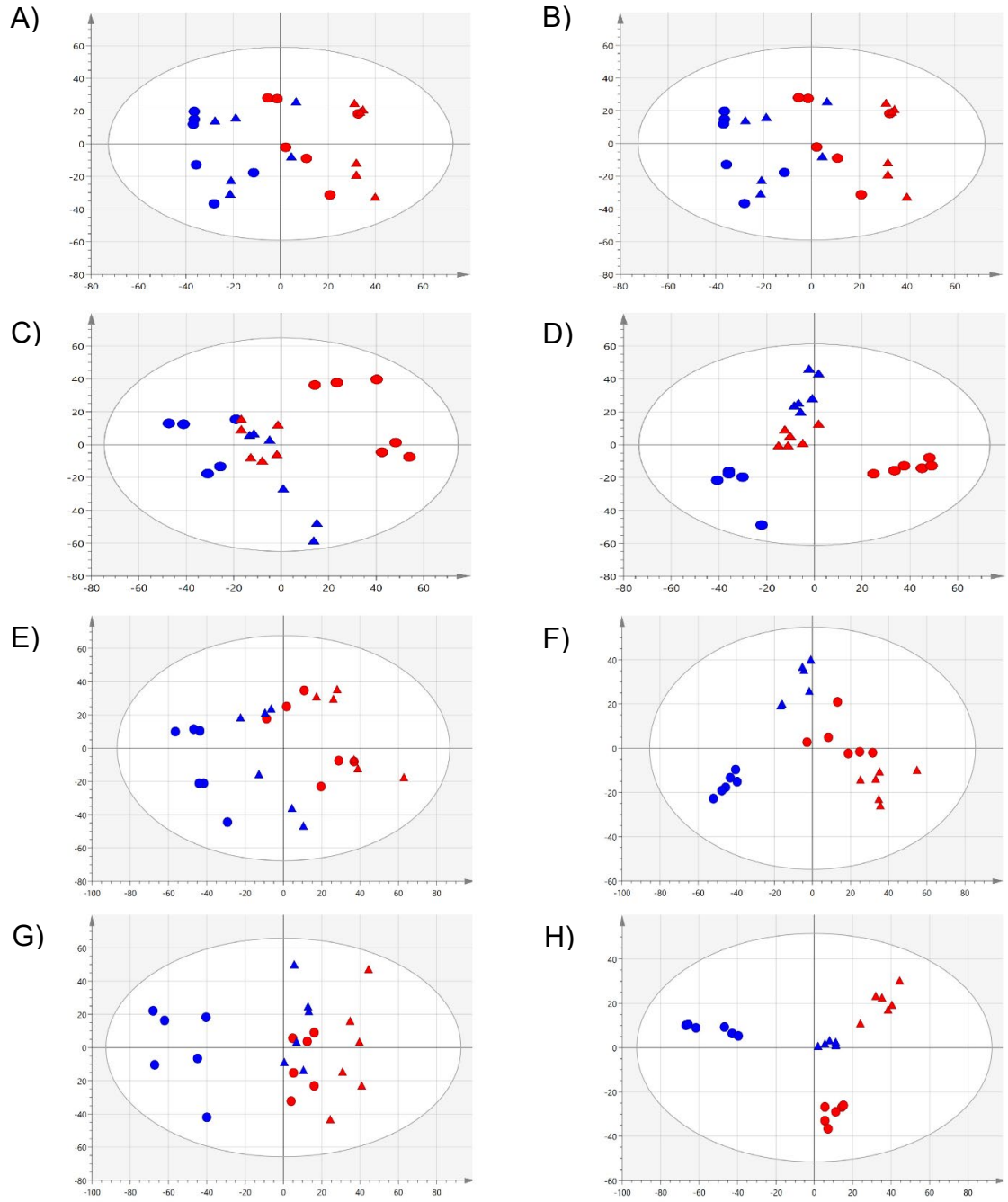


Figure 6.3. PCA (A,C,E and G) and PLS-DA (B,D,F and H) models of XCMS/MZmine2 comparative data of [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X), [*RNQ*$^+$] (Y) and [*RNQ*$^+$] with overexpression of Rnq1 (Z) data at individual time points. Overview of the data shows no more than one outlying sample within the 95% confidence range within each model. On the left are PCA model in order of time, with T1 appearing at the top all the way to T4. PCA scores for models A, C, E and G were $R^2 = 0.670$, $Q^2 = 0.386$, $R^2 = 0.668$, $Q^2 = 0.338$, $R^2 = 0.832$, $Q^2 = 0.440$ and $R^2 = 0.790$, $Q^2 = 0.528$ respectively. On the right are PLS DA models in order of time, with T1 appearing at the top (B) all the way to down T4 (H). PLS DA scores for models B, D, F and H were $R^2(X) = 0.407$, $R^2(Y) = 0.807$, $Q^2 = 0.508$ and a CV-ANOVA *p-value* = 2.71×10^{-2}, $R^2(X) = 0.636$, $R^2(Y) = 0.964$, $Q^2 = 0.863$ and a CV-ANOVA *p-value* = 3.23×10^{-6}, $R^2(X) = 0.691$, $R^2(Y) = 0.979$, $Q^2 = 0.830$ and a CV-ANOVA *p-value* = 5.64×10^{-5} and $R^2(X) = 0.598$, $R^2(Y) = 0.967$, $Q^2 = 0.907$ and a CV-ANOVA *p-value* = 1.77×10^{-9} respectively. Blue circles show W, blue triangles show X, red circles show Y and red triangles show Z.

6.3. Results - [*rnq*$^-$] (W) and [*rnq*$^-$] with overexpression of Rnq1 (X) comparison

6.3.1. PCA and OPLSDA modelling of comparative metabolomic data in SIMCA differentiates between sample group W and sample group X

To rule out the metabolic effects presented as a result of plasmid presence, sample groups W and X were subject to PCA and OPLSDA comparison. This detailed examination of plasmid influence would inform our interpretation of the metabolic influencer of toxicity without the side-effects of plasmid presence, allowing the identification of any metabolic variation caused as a result of induced toxicity.

Initially groups W and X do not separate well in the T1 PCA, the model itself is gave a good $R^2 = 0.664$ value but poor $Q^2 = 0.288$ (Figure 6.4). The forced separation of the T1 OPLSDA suggests that there is a significance to this divide as determined by statistically significant CV-ANOVA *p-value* = 2.79×10^{-1} with good model scores of $R^2(X) = 0.455$, $R^2(Y) = 0.883$, $Q^2 = 0.475$. T2 to T4 PCAs show an increasing amount of significance in group separation and clustering. The T2 PCA model is comparable to the T1, with $R^2 = 0.549$, $Q^2 = 0.193$. However, the strength of the models for T3 and T4 are much improved, with their $R^2 = 0.714$, $Q^2 = 0.544$ and $R^2 = 0.756$, $Q^2 = 0.521$ respectively. OPLSDAs for T2, T3 and T4 reaffirm these findings with their respective model scores of $R^2(X) = 0.621$, $R^2(Y) = 0.993$, $Q^2 = 0.863$ and a CV-ANOVA *p-value* = 3.21×10^{-2}, $R^2(X) = 0.776$, $R^2(Y) = 0.984$, $Q^2 = 0.901$ and a CV-ANOVA *p-value* = 5.68×10^{-3} and $R^2(X) = 0.617$, $R^2(Y) = 0.990$, $Q^2 = 0.969$ and a CV-ANOVA *p-value* = 2.36×10^{-5}.

From the similarity in the model scores and subsequent observations to the group models, it was decided to remove T1 and T2 from any further analysis owing to the lack of strength in PCA model scores and the lack of significant CV-ANOVA *p-values.* The changes caused by carbon-source change as well as the overexpression of Rnq1 from the GAL1 promoter from time point 0 take would time to influence the metabolome and this is apparent in the data (Hovland *et al.* 1989).

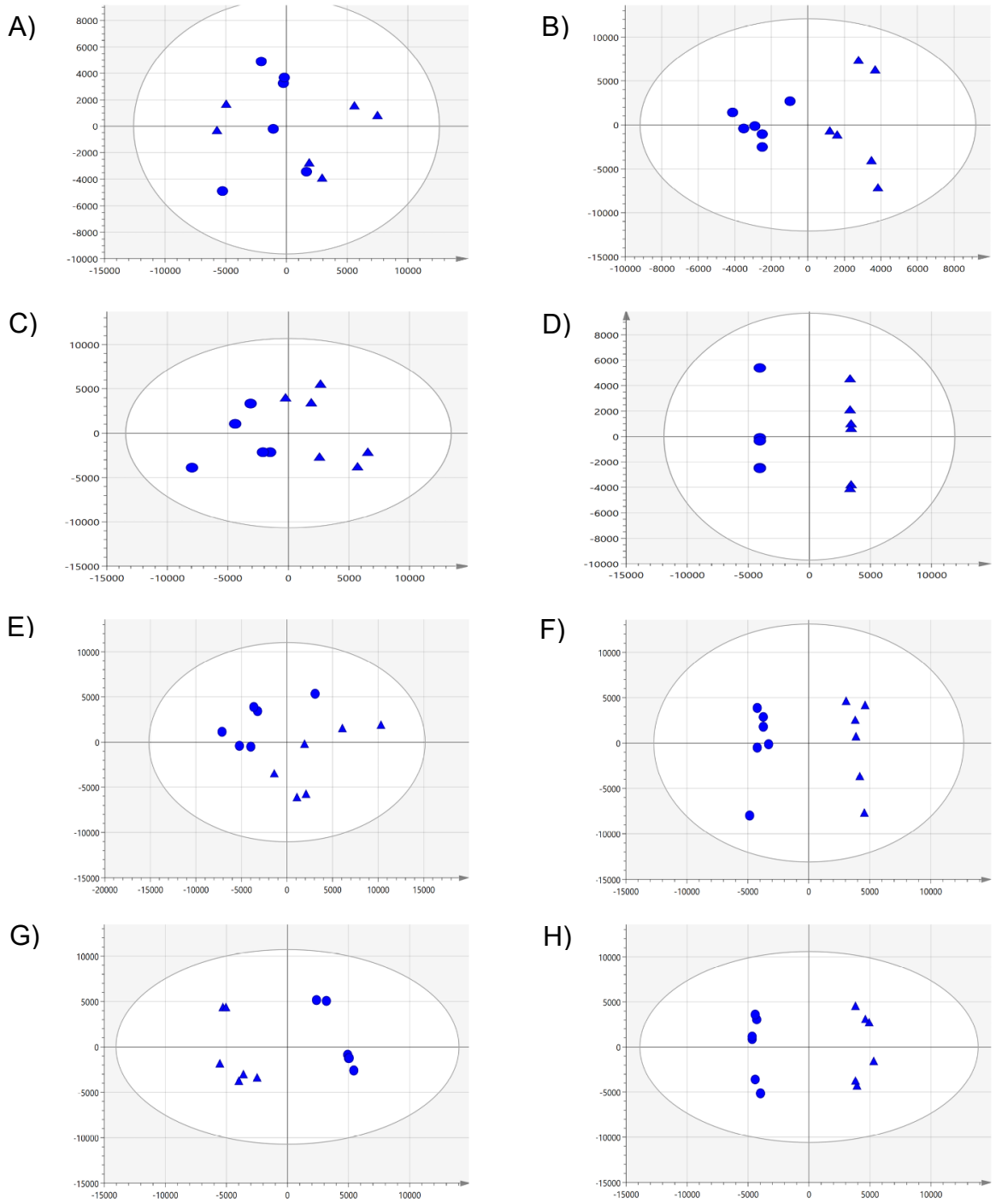


Figure 6.4. PCA (A,C,E and G) and OPLS-DA (B,D,F and H) models of XCMS/MZmine2 comparative data of [*rnq*$^-$] (W) and [*rnq*$^-$] with overexpression of Rnq1 (X) data at individual time points. Overview of the data shows no more than one outlying sample within the 95% confidence range within each model. On the left are PCA models in order of time, with T1 appearing at the top all the way to T4. PCA scores for models A, C, E and G were $R^2 = 0.664$, $Q^2 = 0.288$, $R^2 = 0.549$, $Q^2 = 0.193$, $R^2 = 0.714$, $Q^2 = 0.544$ and $R^2 = 0.756$, $Q^2 = 0.521$ respectively. On the right are OPLS DA models in order of time, with T1 appearing at the top (B) all the way to T4 (H). OPLS DA scores for models B, D, F and H were $R^2(X) = 0.455$, $R^2(Y) = 0.883$, $Q^2 = 0.475$ and a CV-ANOVA *p-value* = 2.79×10^{-1}, $R^2(X) = 0.621$, $R^2(Y) = 0.993$, $Q^2 = 0.863$ and a CV-ANOVA *p-value* = 3.21×10^{-2}, $R^2(X) = 0.776$, $R^2(Y) = 0.984$, $Q^2 = 0.901$ and a CV-ANOVA *p-value* = 5.68×10^{-3} and $R^2(X) = 0.617$, $R^2(Y) = 0.990$, $Q^2 = 0.969$ and a CV-ANOVA *p-value* = 2.36×10^{-5} respectively. Blue circles depict sample group W and blue triangles depict sample group X.

6.3.2. Further PCA and OPLSDA modelling of comparative metabolomic data in SIMCA addressing the prion formation issue

Given the nature of prion formation, the presence of overexpression of Rnq1 protein may well have induced prion formation. This presents a challenge to control for; overcome by using multiple comparisons. Comparing group W with group Y provides lists of pathways effected over time by prion presence. Whilst in previous chapters the analyses have commonly relied on preceding data to inform on such issues, this is not possible due to the change of media that all of the cells have been subjected to. This means that the predictability of the metabolic changes based on our previous findings, whilst not entirely inaccurate, cannot serve as a guide.

Samples X and Y will also be compared to provide some perspective as to the prion status of our group X samples and allow for plasmid effects to be distinguished. If samples from group X have not taken on a prion conformation, they will reveal the effects of the plasmid via comparison with group W and the effects of the prion and the plasmid when compared to group Y. If Rnq1 protein has been sequestered into a prion form, then when compared to group Y, the data will only reveal the effects of the plasmid. When the data is compared to group W, it will reveal the effects of the prion. It is planned that this cross comparison of multiple groups will ascertain the distinct metabolic influencers of the plasmid presence to remove these influencers from our toxicity study.

Group W vs Group Y comparisons for T3 and T4 gave strong PCA models, giving $R^2 = 0.952$, $Q^2 = 0.567$, $R^2 = 0.835$, $Q^2 = 0.593$ respectively. PCAs showed defined clustering, increasing in strength with time, a similar trend is observed in our OPLSDA models with good scores and significant CV-ANOVA *p-value*s, $R^2(X) = 0.821$, $R^2(Y) = 0.999$, $Q^2 = 0.918$ and a CV-ANOVA *p-value* = 4.60×10^{-2}, $R^2(X) = 0.635$, $R^2(Y) = 0.991$, $Q^2 = 0.966$ and a CV-ANOVA *p-value* = 3.12×10^{-5}. This establishes metabolic differences that are separating the two groups W and Y from each other, giving confidence that these groups are distinct from each other. This confirms the findings of chapter 5, namely that a change in prion status alone has a measurable bearing on the metabolic fingerprint of cells.

Groups X and Y comparisons for T3 and T4 gave similar results, although the predictability of the T3 model was less than desirable the other model scores, observations of this model show moderate clustering but one of each sample group that deviated. This was thought to have caused the lack of predictability or low Q^2 value within this model, $R^2 = 0.666$, $Q^2 =$

0.281 and R^2 = 0.809, Q^2 = 0.445. OPLSDAs gave high scores and showed similar clustering as before but on a much larger scale, demonstrating that the differences between the groups where governed by stronger metabolic characteristic changes than other models, , $R^2(X)$ = 0.679, $R^2(Y)$ = 0.997, Q^2 = 0.856 and a CV-ANOVA *p-value* = 4.96×10^{-2} and $R^2(X)$ = 0.620, $R^2(Y)$ = 0.998, Q^2 = 0.962 and a CV-ANOVA *p-value* = 3.56×10^{-4}. These values demonstrate with some confidence that these two groups are metabolically distinct from each other.

This distinction alone is expected, as the two groups contain within them samples which are distinct from each other with regard to both prion status and plasmid addition. Therefore, in order to ascertain the metabolic features that are responsible for these changes' further analysis via Metaboanalyst and Mummichogg as previously conducted was needed. However, given the number of two-way comparisons now needed to prepare the data for BioCyc pathway analysis and to simplify the analysis within this chapter, Metaboanalyst statistical analysis, tentative feature ID and pathway details / results will be provided in Appendix D.

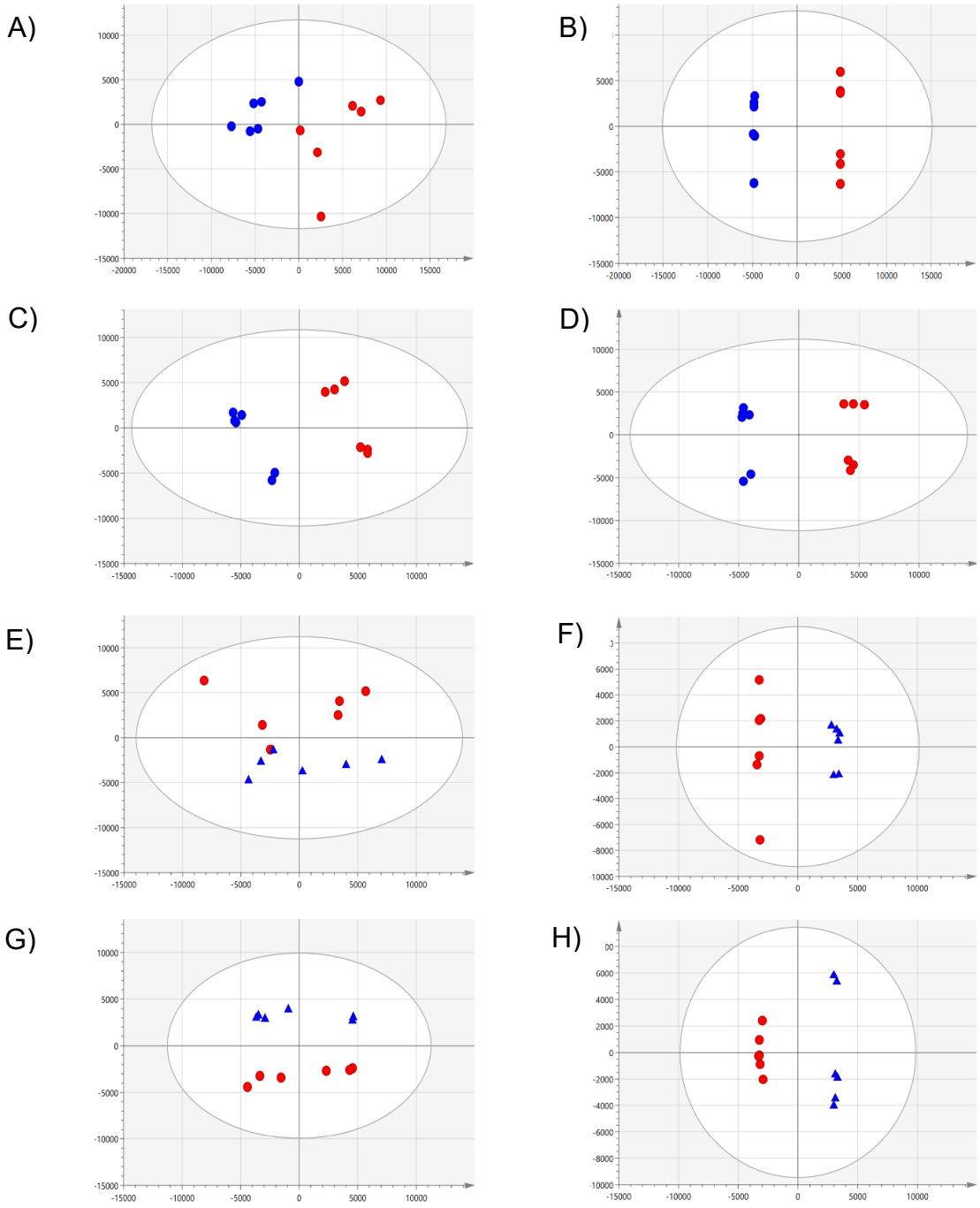


Figure 6.5. PCA (A,C,E and G) and OPLS-DA (B,D,F and H) models of XCMS/MZmine2 comparative plotting of [*rnq*$^-$] (W) vs [*RNQ*$^+$] (Y) (A-D) and [*rnq*$^-$] with overexpression of Rnq1 (X) vs [*RNQ*$^+$] (Y) (E-H) at T3 and T4. Overview of the data shows no more than one outlying sample within the 95% confidence range within each model. On the left are PCA models T3 and T4 for W vs Y (A and C) and T3 and T4 for X vs Y (E and G). PCA scores for models A, C, E and G were $R^2 = 0.952$, $Q^2 = 0.567$, $R^2 = 0.835$, $Q^2 = 0.593$, $R^2 = 0.666$, $Q^2 = 0.281$ and $R^2 = 0.809$, $Q^2 = 0.445$ respectively. OPLS-DA scores for models T3 and T4 for W vs Y (B and D) and T3 and T4 for X vs Y (F and H). OPLS-DA scores for models B, D, F and H were $R^2(X) = 0.821$, $R^2(Y) = 0.999$, $Q^2 = 0.918$ and a CV-ANOVA *p-value* = 4.60×10^{-2}, $R^2(X) = 0.635$, $R^2(Y) = 0.991$, $Q^2 = 0.966$ and a CV-ANOVA *p-value* = 3.12×10^{-5}, $R^2(X) = 0.679$, $R^2(Y) = 0.997$, $Q^2 = 0.856$ and a CV-ANOVA *p-value* = 4.96×10^{-2} and $R^2(X) = 0.620$, $R^2(Y) = 0.998$, $Q^2 = 0.962$ and a CV-ANOVA *p-value* = 3.56×10^{-4} respectively. Blue circles depict sample group W, blue triangles depict sample group X and red circles depict sample group Y.

6.3.3. BioCyc omics dashboard and cellular overview enables mapping of most significant metabolic changes indicating key metabolic pathways of interest

To prepare for pathways analysis within BioCyc, those metabolites considered significant (as established by Metaboanalyst's Mummichog) together with their relative abundances and average relative abundances (within groupings) were collated. The relative abundances of the most significantly altered metabolites (established via Metaboanalyst's Mummichogg) from each of the six W, X and Y samples were then uploaded to BioCyc's Omics Dashboard. Each time stamp, T3 and T4, were treated individually to allow for assessment over time. This three-way comparison made the value of the non-directional assessment (used in chapter 4 and 5) null and so the analysis centred on the directional visualisation of specific metabolic regions that altered between groups (Figure 6.6 and 6.7).

The linear version of the graphs (Figures 6.6 and 6.7.) show differences between the two-time stamps. A significant downregulation in amino acid synthesis was observed occurring in sample groups X and Y between T3 and T4, although sample group W's amino acid biosynthesis appears unchanged. Likewise, a significant downregulation in cofactor biosynthesis was observed for all groups but was more prominent for group W than other groups. In contrast, both the biosynthesis and degradation of fatty acid and lipids appears to be significantly upregulated from T3 to T4. There appears more similarity between groups X and Y than between group W and X within biosynthesis and degradation. However, when considering pathways involved in energy metabolism, there is a similarity between groups W and Y with group X appearing to be substantially different. This could be as a direct result of plasmid presence. It is known that controlling the flux in the galactose utilization pathways can be difficult and this may explain the changes seen in energy metabolism between T3 and T4 for group X (Ostergaard *et al.* 2000).

The logarithmic graphs within Figures 6.6. and 6.7., appear to show a far greater similarity between group X and group Y, with group W consistently appearing as different. This is interesting due to the concern that overexpression of the Rnq1 protein may lead to [*RNQ*$^{+}$] formation. This data leads to the conclusion that the overexpression of Rnq1 protein has led to the formation of prions within the formally prion-free cells and that any metabolic differences left between these two groups are likely a result of the plasmid. However, further pathway analysis will be needed to confirm these findings.

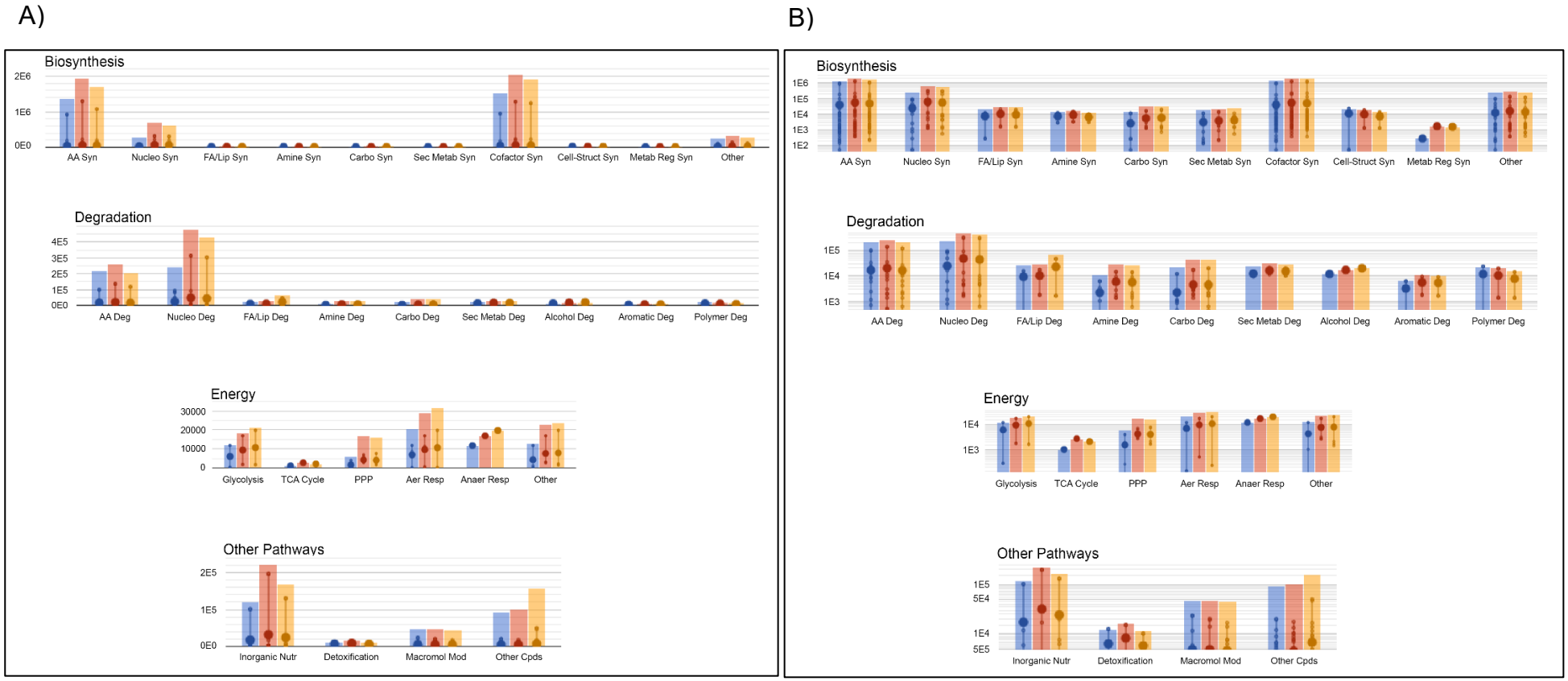


Figure 6.6. Graphical comparisons of the average relative abundances between [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) samples within key metabolic regions at T3. A) Charts average comparisons of the T3 relative abundances between group W samples (shown in blue), group X (shown in red), and group Y (shown in yellow) data is first separated by key metabolic region (shown as a graph title) and then by specific metabolic area (shown on the x axis). The Y axis is set to a linear scale, appropriate to the relative abundances. Individual data points are shown as small circles within the larger summated bars. B) Shows the same information with a logarithmic scale on the Y axis.

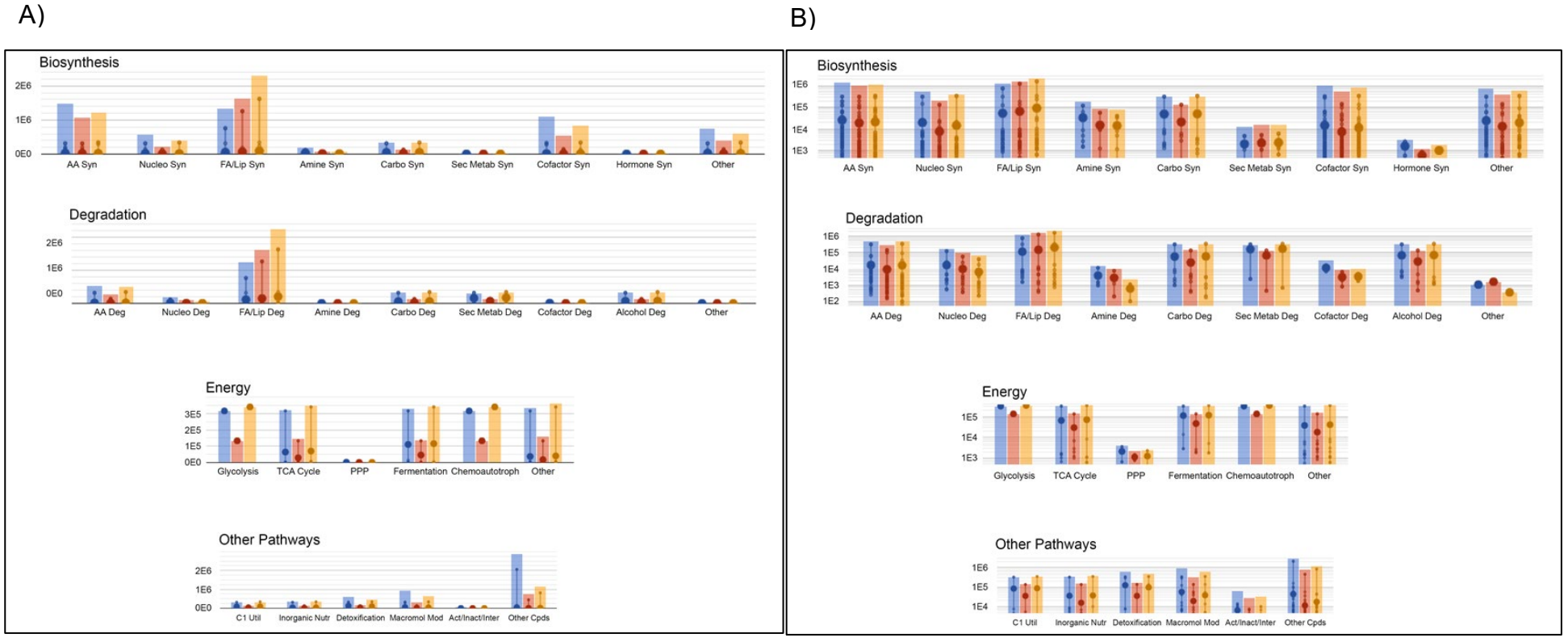


Figure 6.7. Graphical comparisons of the average relative abundances between [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) samples within key metabolic regions at T4. A) Charts average comparisons of the T4 relative abundances between group W samples (shown in blue), group X (shown in red), and group Y (shown in yellow) data is first separated by key metabolic region (shown as a graph title) and then by specific metabolic area (shown on the x axis). The Y axis is set to a linear scale, appropriate to the relative abundances. Individual data points are shown as small circles within the larger summated bars. B) Shows the same information with a logarithmic scale on the Y axis.

Table 6.2. Top Twenty most perturbed pathways during T3 when comparing [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) samples.

Pathway
L-leucine biosynthesis
siroheme biosynthesis
superpathway of branched chain amino acid biosynthesis
purine ribonucleosides degradation
guanine and guanosine salvage
xanthine and xanthosine salvage
lipoate biosynthesis and incorporation (pyruvate dehydrogenase and oxoglutarate dehydrogenase)
lipoate biosynthesis and incorporation I
lipoate biosynthesis and incorporation (glycine cleavage system)
superpathway of lipoate biosynthesis and incorporation (PDH, KGDH, GCV)
biotin biosynthesis from 8-amino-7-oxononanoate I
β-alanine biosynthesis
inosine-5'-phosphate biosynthesis II
salvage pathways of adenine, hypoxanthine and their nucleosides
salvage pathways of pyrimidine ribonucleotides
adenine and adenosine salvage IV
superpathway of purine nucleosides salvage
L-histidine biosynthesis
7-(3-amino-3-carboxypropyl)-wyosine biosynthesis
superpathway of purine nucleotides de novo biosynthesis I

The most significant relative abundances of samples groups W, X and Y, via BioCyc's specific *S. cerevisiae* cellular overview.

Table 6.3. Top Twenty most perturbed pathways during T4 when comparing [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) samples.

Pathway
sphingolipid biosynthesis (yeast)
sphingolipid recycling and degradation (yeast)
dehydro-D-arabinono-1,4-lactone biosynthesis
formate oxidation to CO2
pyruvate decarboxylation to acetyl CoA
L-malate degradation II
(R,R)-butanediol degradation
ethylene glycol degradation
glycine biosynthesis II
leucine degradation III
leucine degradation
mannitol degradation II
glycine cleavage
sorbitol degradation
10-trans-heptadecenoyl-CoA degradation (reductase-dependent)
xylitol degradation
L-glutamate biosynthesis IV
10-cis-heptadecenoyl-CoA degradation
oleate β-oxidation (isomerase-dependent)
glycolysis

The most significant relative abundances of samples groups W, X and Y, via BioCyc's specific *S. cerevisiae* cellular overview.

Data was loaded onto to BioCyc's specific *S. cerevisiae* cellular overview and pathway hits can be seen in Table 6.2. and 6.3. The most perturbed pathways appear to change according to time point. Pathways are ranked by DPPS score which is a measure pathway perturbation within multiple data sets. All the pathways listed gave scores over 100,000 for Table 6.2. and over 400,000 for Table 6.3. indicating perturbation. Due to the very high DPPS score for Table 6.3. and the variance between the two the lower ranking perturbations of Table 6.3. were subsequently investigated. The perturbations seen within Table 6.2. are still present in Table 6.3. with approximately the same DPPS scores. However, within 6.3. these perturbations appear relatively insignificant when compared to the DPPS scores of the pathways shown in Table 6.3., knocking the pathways shown in Table 6.2. below the top 20 ranking. Clearly significant changes to the pathways present in Table 6.3. must have occurred in the hour between T3 cell harvest and T4 cell harvest.

6.3.4. Data overlay onto implicated pathways, via BioCycs Pathway collage, reveals the direction of the up and down regulation between sample groups

The information provided by Table 6.2. and 6.3. highlighted areas for further investigation. Standardised LC-MS data (as described in section 4.5.6.) was overlaid onto these specific pathways to further visualise and interpret the findings, and so separate the plasmid-only and prion-only metabolic indicators. T3 pathway overlays (Figure 6.8. and 6.9.) indicated that metabolic characteristics are shared between group X and Y, defining group W as our outlying sample group. Interestingly this same pattern was observed in the logarithmic version of the graphs (Figure 6.6. and 6.7.). This indicates that overexpression of the Rnq1 protein within sample group X, may have induced prion formation as previously suggested. It is regrettable that western blots were not conducted to elucidate this, as discussed in section 5.6.3. However, owing to the presence of distinct variants of prions; it may be possible that the slight metabolic differences between the two groups may have presented themselves regardless, owing to a different strength in the prion variant formed as a result of overexpression (Bardill and True, 2010).

The data within Figure 6.9 show a reliable and notable upregulation of the pathways listed, yet within Figure 6.8, the overlay is not so robust. The presence of a gradient within the pyrimidine salvage pathway between samples W, X and Y (from upregulation in sample group W to downregulation in sample group Y) is perplexing. Also, there are clear metabolic differences apparent within sample group X. These differences may be attributed to the presence of the plasmid, as similar inconsistencies within salvage pathways are also seen

within Figure 6.8., appearing to show sample group X as an outlier group. This suggests that the presence of the plasmid is influencing nucleotide and amino acid salvage within cells.

Previous studies have shown that adenine and nucleotide metabolism can be altered due to change in carbon and nitrogen source within bacterial strains, and that within yeast strains the salvage of nucleotides aids in their survival during nutrient starvation and oxidative stress (Xu *et al.* 2013; Nygaard *et al.* 1996). Interestingly, within yeast, this dysregulation of nucleotide pathways when at high levels can lead to cytotoxicity (Kowalski *et al.* 2008). These observations would help support a plasmid-caused effect here if a clear upregulation of these pathways was seen in sample group X. However, upregulation is not consistent enough to draw firm conclusions. Further research is required, potentially using another plasmid which does not overexpress Rnq1 alongside this plasmid, both with GAL promoters, to aid in picking apart these differences.

Figure 6.10 shows the pathways indicated within Table 6.3. as a result of T4 and displays a sub selection of the pathways listed. The sub selection is since many of the pathways listed on Table 6.3. are influenced or modified by small add on pathways that are not easily presentable within such a pathway diagram. The decision has been made to focus on those pathways that have been directly influenced by changes in the metabolome between the groups. Observations of these findings identify similarity between sample groups X and Y, leaving sample group W as an outlier. Encouragingly, despite the change in carbon source, similar metabolic influencers as seen previously (section 5.4.) are identified, with upregulation within the sphingolipid biosynthesis and degradation pathways alongside the presence of the prion (as is the case for group Y) and induced prion (as is reportedly the case for group X). Although, the same pathways have been implicated, the directionality of this perturbation has altered, with [*RNQ*$^{+}$] previously observed (section 5.4.) as causing downregulation in sphingolipid metabolism, and now [*RNQ*$^{+}$] showing upregulation of all sphingolipid pathways (Figure 6.10.). This difference may have been caused by media change and/or due to a variance in ‘age’ between the cells tested within these experiments and those tested previously (chapter five). However, this switch was not observed in the T3 pathway. It is possible that at T4, the cells are changing their growth phase from logarithmic to stationary phase. Degradation pathways are highlighted only within sample group X giving further credence to the plasmid influence on degradation pathways, especially those readily influenced by carbon source change.

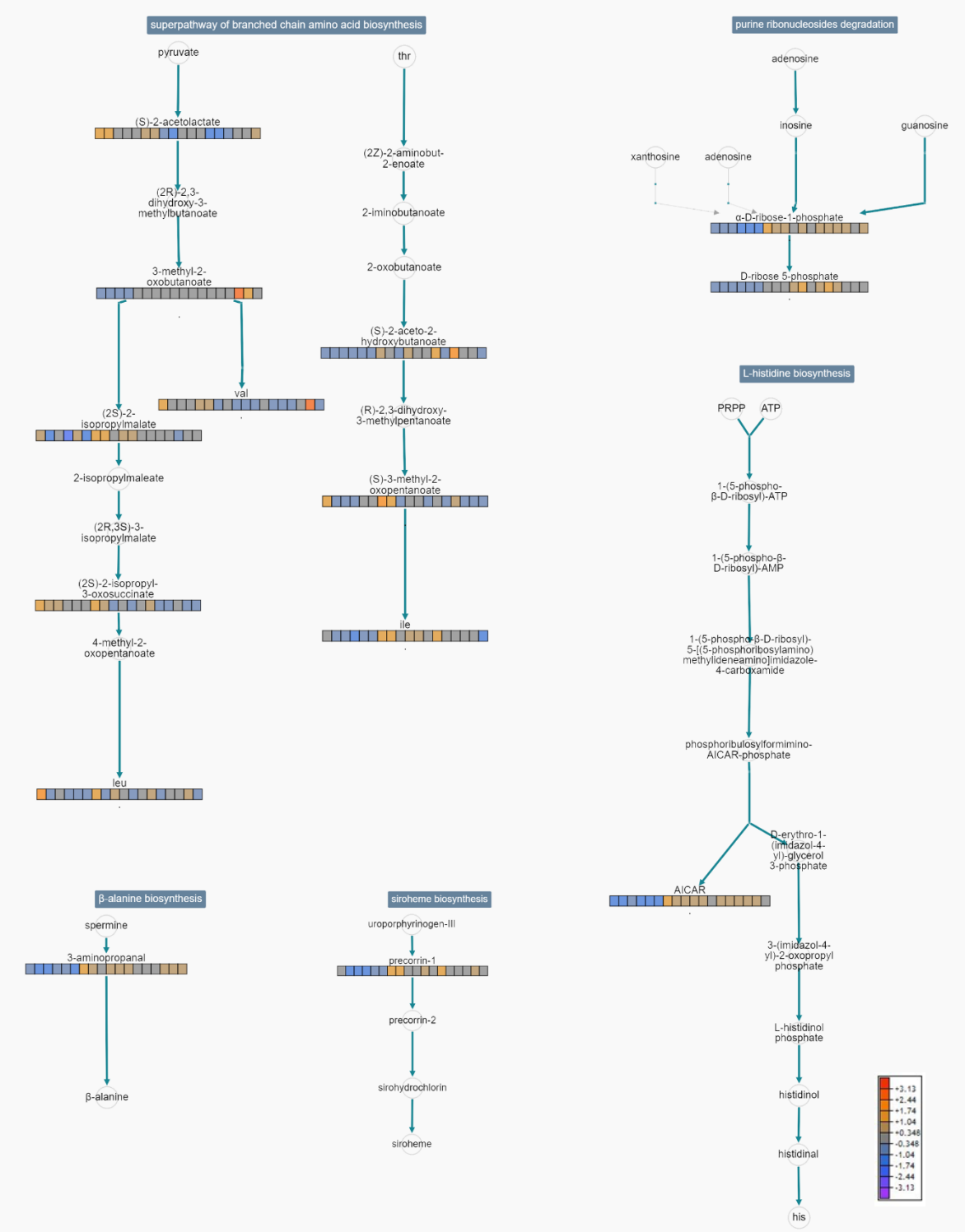


Figure 6.8. Pathway collage of four of the top 20 implicated pathways at T3 (via BioCycs cellular overview) with standardised omics [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) data of individual metabolites overlaid. A colour legend is shown in the bottom righthand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y. A larger version of this Figure is available in Appendix I.

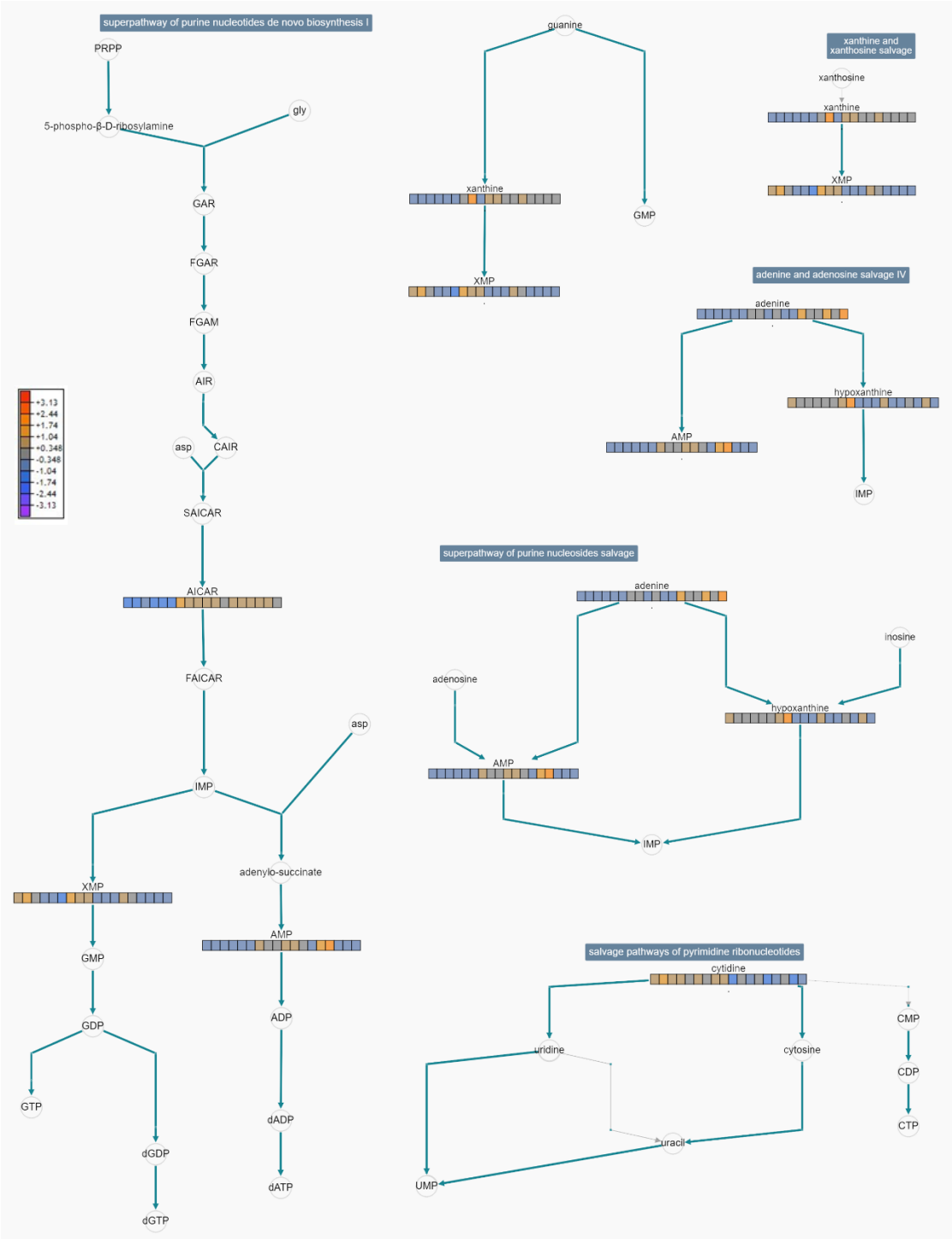


Figure 6.9. Pathway collage of five of the top 20 implicated pathways at T3 (via BioCycs cellular overview) with standardised omics [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) data of individual metabolites overlaid. A colour legend is shown on the middle left to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y. A larger version of this Figure is available in Appendix I.

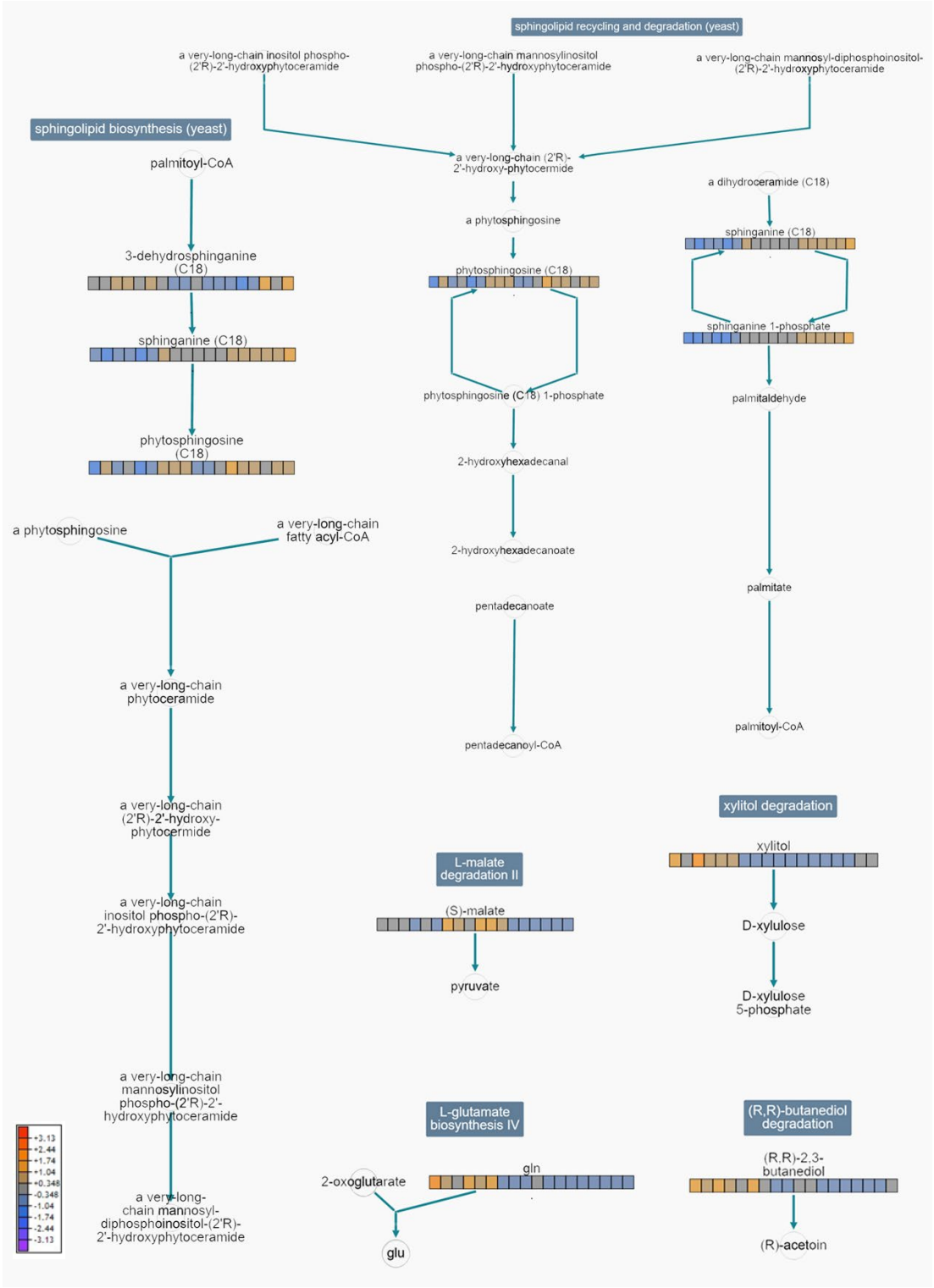


Figure 6.10. Pathway collage of six of the top 20 implicated pathways at T4 (via BioCycs cellular overview) with standardised omics [*rnq*$^-$] (W), [*rnq*$^-$] with overexpression of Rnq1 (X) and [*RNQ*$^+$] (Y) data of individual metabolites overlaid. A colour legend is shown in the bottom left hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y. A larger version of this Figure is available in Appendix I.

6.4. Results – Comparison of the Metabolomic Perturbations of [*RNQ*$^{+}$] and [*RNQ*$^{+}$] with overexpression of Rnq1

6.4.1. PCA and OPLSDA modelling of comparative metabolomic data in SIMCA differentiates between sample group Y and sample group Z at T4

Initial scrutiny of the PCAs reveal samples appear moderately clustered at T4. PCA scores for T3 gave good R^2 values at 0.654, however the Q^2 values of 0.279 are not indicative of a model with good predictability. Clustering within Figure 6.11. (A) is limited, with considerable overlap between samples. By T4 a visible divide is apparent and model scores improve considerably: $R^2 = 0.813$, $Q^2 = 0.478$.

Given the lack of separation in Figure 6.11. (A) for T3, the resultant OPLSDA, shows reasonable model scores but no significant difference within the CV-ANOVA, $R^2(X) = 0.690$, $R^2(Y) = 0.998$, $Q^2 = 0.808$ and a CV-ANOVA *p-value* = 9.50×10^{-2}. This is not apparent at T4 (D) where a significant metabolic difference between the two groups can be measured, with the model giving scores indicative of its reliability, $R^2(X) = 0.673$, $R^2(Y) = 0.987$, $Q^2 = 0.907$ and a CV-ANOVA *p-value* = 4.81×10^{-3}.

From this assessment, it is possible to conclude that sample groups appear to be metabolically similar until T4, and further analysis of this timepoint was conducted.

To determine the metabolic features responsible for these changes, further analysis via Metaboanalyst and Mummichogg was required. However, in order to prepare the data for BioCyc pathway analysis and to simplify the analysis within this chapter, Metaboanalyst statistical analysis, tentative feature ID and pathway details and results will be provided in appendix C, to better focus on the consequences of induced toxicity.

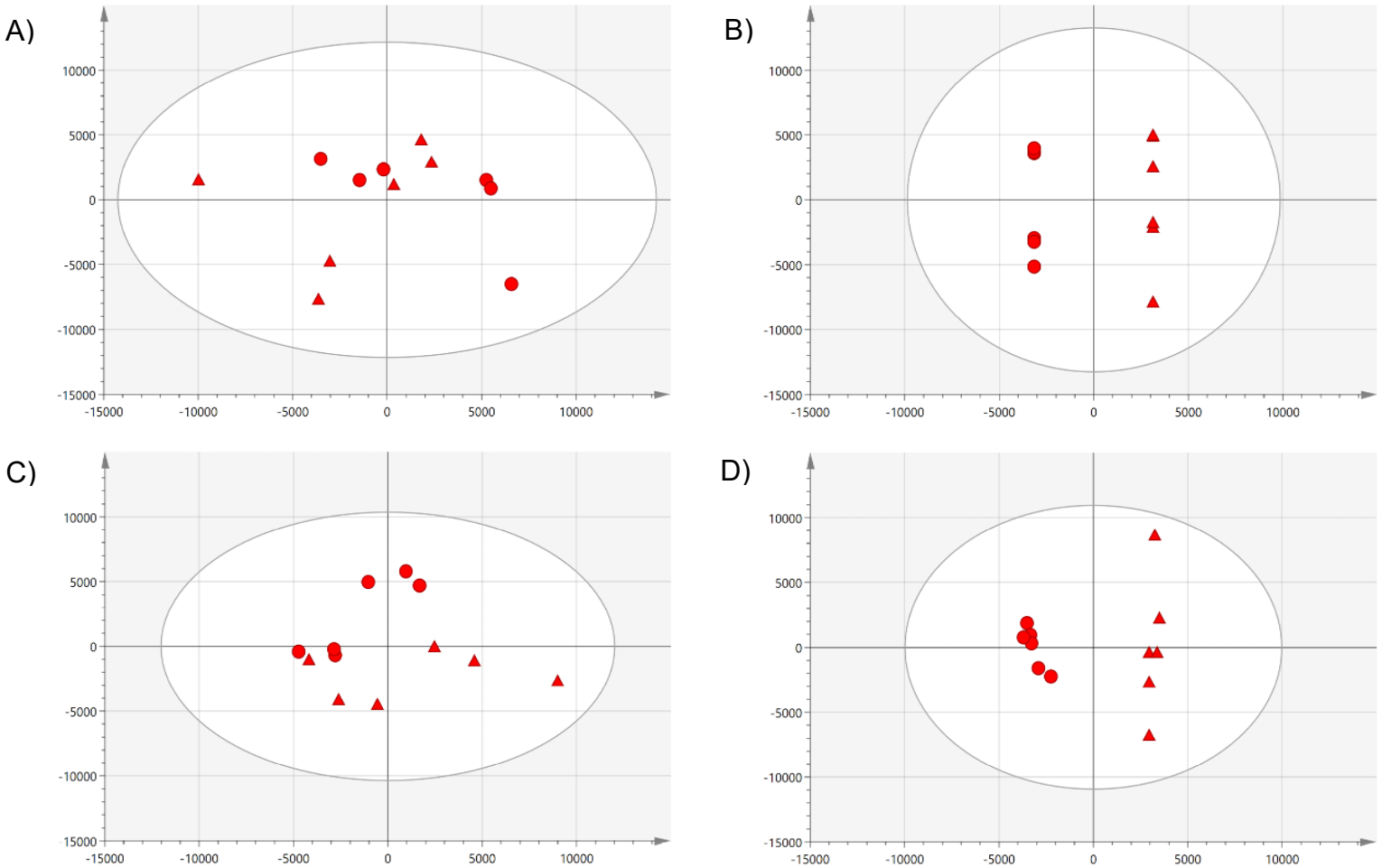


Figure 6.11. PCA (A and C) and OPLS-DA (B and D) models of XCMS/MZmine2 comparative data [*RNQ*$^{+}$] (Y) and [*RNQ*$^{+}$] with overexpression of Rnq1 (Z) data from T3 (A and B) and T4 (C and D). Overview of the data shows no more than one outlying sample within the 95% confidence range within each model. On the left are PCA model in order of time, with T3 appearing at the top (A) and T4 at the bottom (B). PCA scores for models A and C, were $R^2 = 0.654$, $Q^2 = 0.279$ and $R^2 = 0.813$, $Q^2 = 0.478$. On the right are OPLSDA models in order of time, with T3 appearing at the top (B) and T4 appearing at the bottom (D). OPLSDA scores for models B and D were $R^2(X) = 0.690$, $R^2(Y) = 0.998$, $Q^2 = 0.808$ and a CV-ANOVA *p-value* = 9.50×10^{-2} and $R^2(X) = 0.673$, $R^2(Y) = 0.987$, $Q^2 = 0.907$ and a CV-ANOVA *p-value* = 4.81×10^{-3}. Red circles depict sample group Y and red triangles depict sample group Z.

6.4.2. BioCyc omics dashboard and cellular overview enables mapping of most significant metabolic changes indicating key metabolic pathways of interest

Data was treated identically to that of section 6.4. and direct visualization of the specific metabolic regions altered between groups was made possible by way of Figure 6.12.

Considerable downregulation is seen in group Z when compared to group Y, with a clear trend observed in both linear and logarithmic graphs. This downregulation appears to be affecting all areas of the metabolism, including biosynthesis, degradation, energy, and other pathways. This is likely the effect of toxicity induction causing cellular damage. The only notable exceptions from this trend in group Z were secondary metabolite synthesis, amine degradation and “other degradation”, in which there is a perceived upregulation. Many of the secondary metabolites act to reduce the effect of cytotoxicity and thereby protect the cells (Roze *et al.* 2012). The increase in degradation may (as previously discussed in section 6.3.) be as a result of the action of the plasmid: a side effect of the change in carbon source and the leaky GAL promoter.

To enable deeper pathway analysis, this data was loaded onto to BioCyc’s specific *S. cerevisiae* cellular overview and pathway hits can be seen in Table 6.4. Pathways are ranked by DPPS score (that is a measure of the pathway perturbation within multiple data sets), all the pathways listed gave scores over 400,000 strongly indicating their perturbation.

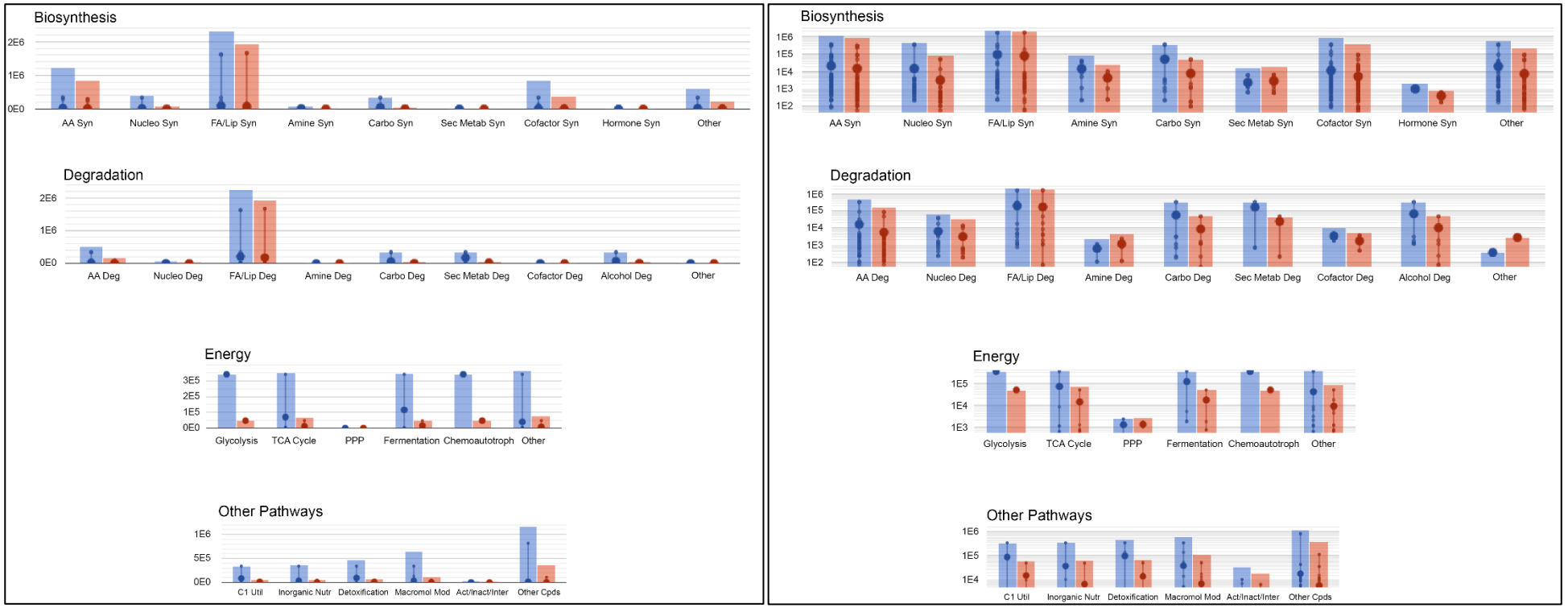


Figure 6.12. Graphical comparisons of the average relative abundances between [*RNQ*$^{+}$] (Y) and [*RNQ*$^{+}$] with overexpression of Rnq1 (Z) samples within key metabolic regions at T3 (A) and T4 (B). A) Charts average comparisons of the T3 relative abundances between group Y samples (shown in blue) and group Z (shown in red), data is first separated by key metabolic region (shown as a graph title) and then by specific metabolic area (shown on the x axis). The Y axis is set to a linear scale, appropriate to the relative abundances. Individual data points are shown as small circles within the larger summated bars. B) Shows the same information with a logarithmic scale on the Y axis.

Table 6.4. Top Twenty most perturbed pathways during T4 when comparing [*RNQ*$^{+}$] (Y) and [*RNQ*$^{+}$] with overexpression of Rnq1 (Z) samples.

Pathway
sphingolipid biosynthesis (yeast)
sphingolipid recycling and degradation (yeast)
dehydro-D-arabinono-1,4-lactone biosynthesis
formate oxidation to CO2
pyruvate decarboxylation to acetyl CoA
L-malate degradation II
(R,R)-butanediol degradation
ethylene glycol degradation
glycine biosynthesis II
leucine degradation III
leucine degradation
mannitol degradation II
glycine cleavage
sorbitol degradation
10-trans-heptadecenoyl-CoA degradation (reductase-dependent)
10-trans-heptadecenoyl-CoA degradation (MFE-dependent)
xylitol degradation
9-cis, 11-trans-octadecadienoyl-CoA degradation (isomerase-dependent)
L-glutamate biosynthesis IV
10-cis-heptadecenoyl-CoA degradation

The most significant relative abundances of samples groups Y and Z, via BioCyc's specific *S. cerevisiae* cellular overview.

6.4.3. Data overlay onto implicated pathways, via BioCycs Pathway collage, reveals the direction of the up and down regulation between sample groups

The information provided by Table 6.4. indicated areas of enquiry. Standardised LC-MS data (as described in section 4.5.6.) was overlaid onto these specific pathways to further visualise and interpret the findings in an attempt to isolate the toxic effects of Rnq1 overexpression without the influence of the effects of the prion or the plasmid. However, the pathways shown in Table 6.4. overlap considerably with those found in Table 6.3. Whilst these effects could be largely prion mitigated, the changes within each group were further examined with the only two more Coenzyme A degradation pathways being identified.

Figure 6.13. details the most prevalent of these changes, a figure largely based on the pathways available for this type of visualisation within BioCyc and the ranking of the DPPS score. Pathways shown here do not represent whether up or downregulation is strongly linked to either group. Given that this is not a process whereby overexpression routinely

causes cell death, one would expect that there must be some yet undiscovered factor that compels prion formation down a cytotoxic route.

The unpredictability of trend within the sphingolipid pathways between sample groups Y and Z (Figure 6.13.) implies that these pathways may be especially sensitive to perturbations of this kind. Further research into these pathways using targeted lipidomics may be able to provide some clarity with regards to the up and downregulation seen here within biological classes. Between sample groups Y and Z, a relative downregulation (Y>Z) is seen in the Xylitol degradation, (R-R) butanediol degradation and L-glutamate biosynthesis pathways (Figure 6.13.). It is reasoned that this too may be due to the presence of the plasmid and the known effects on energy and secondary metabolism.

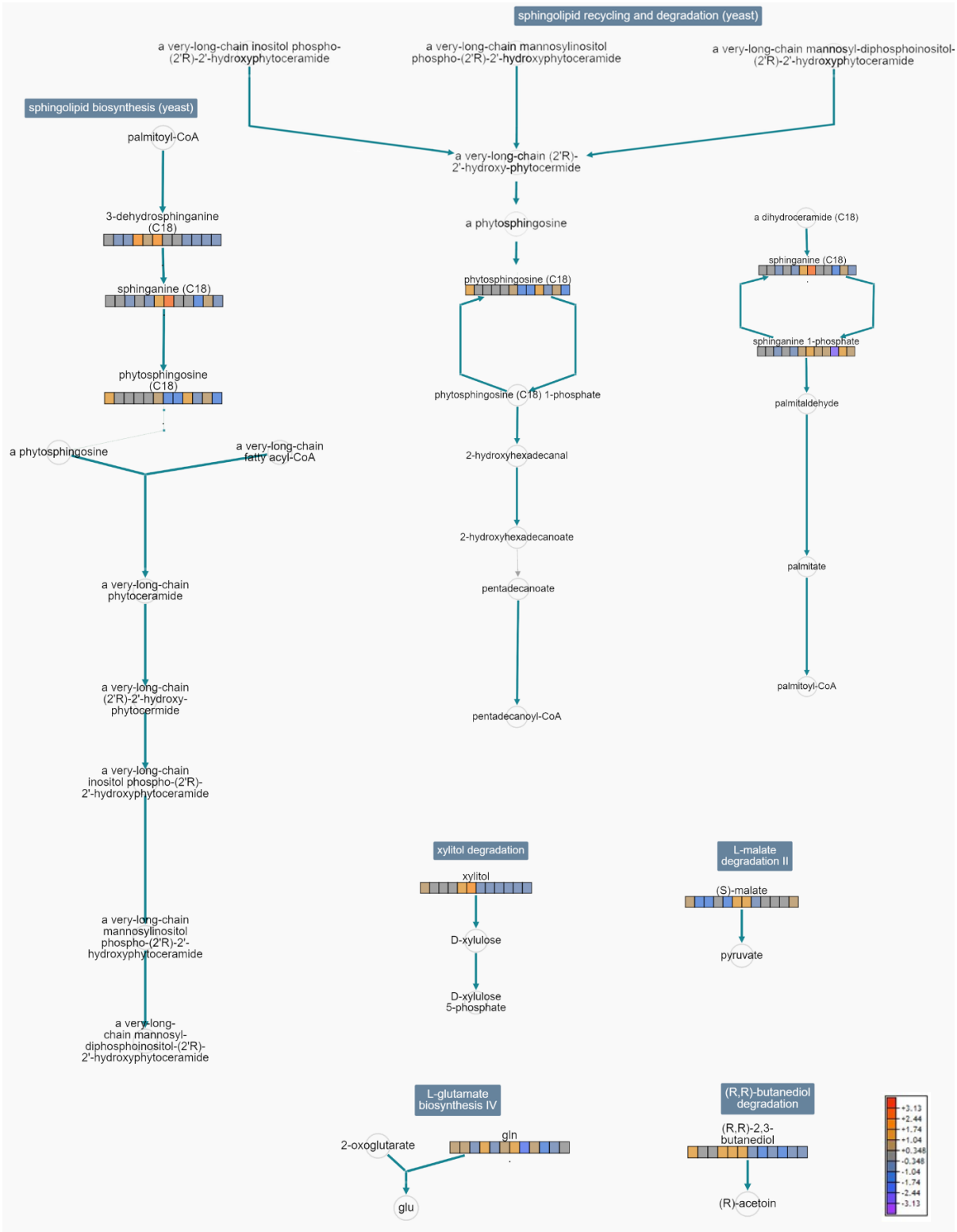


Figure 6.13. Pathway collage of four of the top 20 implicated pathways at T4 (via BioCycs cellular overview) with standardised omics [*RNQ*$^+$] (Y) and [*RNQ*$^+$] with overexpression of Rnq1 (Z) data of individual metabolites overlaid. A colour legend is shown in the bottom righthand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y. A larger version of this Figure is available in Appendix I.

6.5. Discussion

The experiments performed with this study have not been able to robustly identify the metabolic disruptions that arise as a result of induced toxicity. Despite efforts to control for the variables of prion and plasmid mitigated effects, clear evidence about the individual influence of these factors has been difficult to acquire. These experiments have revealed further specifics metabolomic effects regarding plasmid influence on nucleotide and secondary metabolite pathways as well as perturbations within the sphingolipid pathways, although the directionality of these results remains unclear.

6.5.1. The problem determining the influence of plasmid presence and possible solutions

Many of the pathways analysed appear to indicate that nucleotide, secondary metabolite, and amino acid degradation may have been affected by the presence of the plasmid. However, the overlap between these pathways and those affected by the presence of the prion itself is regrettable and makes for difficult analysis. To successfully pick apart the many factors at play here, it would be preferable to change the GAL promoter used in this study for another inducible promoter which does not have the same issues with 'leakage'. Metabolic engineering studies have shown that promoter selection can be key in the design of experimentation and that without tightly controlled promoter strength and application, toxicity within the growth phase can occur (Blazek and Alper, 2012). This questions whether the perceived toxicity of overexpression observed by Douglas *et al.* (2008) was in fact as a result of intricate relationships between the number of prion proteins or was indeed as the result of unpredictable plasmid effects. However, if this had been the case, the cell death effect would be seen in all cells as opposed to only 25% of them. We repeated the experiments of Douglas *et al.* (2008) to confirm these cell losses and indeed our numbers support this (data available in appendix E).

The GAL promoter is widely employed as an inducible promoter in pathway engineering applications, although perhaps this is a unique case where the promoter is not suitable for cellular conditions (Da Silva and Srikrishnan, 2012; Blazek and Alper, 2012). Potentially if this is the case then the use of a more tightly regulated inducible promoter, such as *CUP1*, would be advisable. *CUP1* is easily inducible via addition of copper sulphate to the media with expression levels related to the concentration of copper sulphate added. This provides the potential to control a range of protein expression during experimentation (Da Silva and

Srikrishnan, 2012; Mascorro-Gallardo *et al.* 1996). However, the use of this promoter might cause issues of its own as copper is known to induce prion formation in PrP monomers. There are also links between copper exposure and PrP neurotoxicity, but no evidence that this is relevant with regards to yeast prions (Yen *et al.* 2016).

6.5.2. Loss of function does not present as a perturbation in this experimentation

The loss of Rnq1 function perturbations (which is described in chapters four and five to have a considerable effect on the most perturbed pathways) were not identified in the pathways highlighted here. This could be due to either the native role of Rnq1 changing when on alternative carbon sources or that the influence or role of Rnq1 changes over time, given that the cells tested were considerably more aged than those in previous experiments. Without the inclusion of data from a *Δrnq1* strain in these experiments, it is difficult to ascertain the answer to this question. Such a control was not considered given the focus on toxicity and that the loss of function would be observed in both [*RNQ*$^{+}$] and [*RNQ*$^{+}$] with overexpression.

6.5.3. The toxicity of overexpression of the Rnq1 protein, findings, limitations, and further research

The results shown here provide additional evidence of a connection between prion presence and a perturbation of sphingolipid metabolism, as described in chapter five. However, Figure 6.13. shows considerable variation between samples within biological classes (Y and Z) and so drawing distinctions between the two groups is difficult. This may be that the perturbations that occur as a result of the prion (described in chapter five) are simply exaggerated in cells with overexpression. However, if this were the only factor here then a clear upregulation of these pathways would be seen from Y to Z, instead of the sporadic pattern observed. An explanation of this may be that data here (appendix E) and Douglas *et al.* (2008) demonstrate that many of the cells in the Z group are dead. As the metabolic profile of live cells is likely to be very different from dead cells, what we may observe in group Z is a portrayal of this imbalance. However, this should have been approximately represented by all group Z samples given that the extractions are made from a large 2mL sample of cells not individual cells. However, there is considerable variance between the observed percentage of cell death between data here (appendix E) and Douglas *et al.* (2008). With data here revealing a 76% reduction in live colony number between samples Y and Z (respectively) whereas Douglas *et al.* (2008), much larger experiment, found approximately

a 25% reduction. In addition, we observe a difference between the live colony count of all [*rnq*$^{-}$] samples and [*RNQ*$^{+}$], an unanticipated result. It is possible that this difference may have been due to the (undetermined) strength of the [*RNQ*$^{+}$] strain used here and within Douglas *et al.* (2008) study.

It is possible that another factor may have limited the results seen here, in chapters four and five, due to the small number of samples both PIM and NIM MS data were collected. However, for the experiments here, due to a much larger number of samples only PIM MS data was collected. Regrettably, it is possible that this uncollected NIM data may have provided a clarity to pathway analysis here that has not been obtainable. For now, the only conclusions that can be drawn from this data is that perturbations to sphingolipid metabolism are present in both in [*RNQ*$^{+}$] and [*RNQ*$^{+}$] with overexpression of the Rnq1 protein, and that these perturbations appear to be distinct from each other, at least some of the time.

Potentially the use of NMR as a non-destructive analytical method may have been a better methodology to employ for these experiments. As NMR is capable of monitoring the same samples over time, this may have been able to overcome the variance and overlap seen between the Y and Z samples. As this is a non-destructive method, having live samples at the end of the experiment would permit viable colony counts to derive an estimation of cell death within the culture. This would facilitate appropriate corrections to the subsequent relative abundances of metabolites or permit the ruling out that some Z cell samples suffered large losses and others did not. An ideal methodology would be the rapidly advancing metabolic experiments which aim to monitor single cell metabolic profiles and so be able to measure difference between individual cells within this single sample grouping and determine if toxicity had therefore been induced (Duncan *et al.* 2019; Rubakhin *et al.* 2013). Even when limited to the MS technologies used in this experimentation, targeted lipidomics using standards to enable the quantification of the metabolites within the pathways (Figure 6.13.) may bring more clarity to the findings here and aid in understanding the influence of the toxicity associated with overexpression.

The extensive lag time to acquire definite separation between groups as seen in all PCAs, suggests that induction of *RNQ1* via switching the media to galactose at an earlier point may have been preferable. This timing was selected based on the previously established four-hour time scale at which toxicity is reportedly found (Douglas *et al.* 2008). With hindsight, extending induction beyond this time point and into the increasingly cytotoxic cells timeline

would have been useful. (section 5.6.3.), The use of SDDAGE and western blotting to test for prion formation and thereby determine prion strength would also have confirmed the prion status of cells, removing any doubts regarding this status. This was considered at the time and indeed lysates from each of the cell lines were acquired with a view to test these after the large metabolic experiments had been conducted. Sadly, a large power cut led to a mass loss of frozen samples, the lysates among them. Further studies about the rate of cell death with overexpression of Rnq1 in cell lines that are known to already have issues with their lipid membranes, could provide information about the target and effect of toxicity.

These experiments have determined that all biological classes tested here are metabolically distinct and act as “proof of concept”. Future metabolic studies, albeit perhaps using NMR or one of the other analytical methods discussed in this chapter, would reveal more details about the metabolic effects of Rnq1 induced toxicity. In addition, it is suggested that other yeast models that have been developed for the study of cellular toxicity be included. In particular, polyglutamine (polyQ) expansions of the huntingtin (Htt) protein (the presence of which results in Huntington’s disease in humans) (Walker, 2007). Fragments of Huntington exon I with varying length polyQ expansions are available for use in *S. cerevisiae* (Meriin *et al*. 2002; Krobitsch and Lindquist, 2000). Such fragments recapitulate aggregation in S. cerevisiae models, with Meriin et al. (2002) polyQ model capable of inducing cellular toxicity. Meriin et al. (2002) found that the polyQ polypeptide 103Q only presented toxicity in the [*RNQ*$^{+}$] cells, and in the absence of this prion both aggregation and toxicity were dramatically suppressed; suggesting that it would be a good candidate for further study here. Other yeast models such as α-Synuclein and PrP are available and may also provide valuable information about aggregation and toxicity.

All biological classes within this chapter presented distinct metabolic profiles, with perturbations to the sphingolipid pathways appearing between groups Y and Z. However, details of this perturbation have been difficult to extract from the data and further research is needed to provide a better understanding of the implications of toxicity on cellular metabolism.

Chapter Seven - Discussion

7.1. Overview of the project

A common question encountered throughout this research from members of the metabolomics community has been “why use metabolomics?” It was understood that this was not asked to be deliberately obtuse, but to encourage us to consider if the use of this complex technique for this area of enquiry was justified. Given that it would be possible to ‘do’ metabolomics on a plethora of subjects, the question sought more to address why metabolomics was suitable for investigating the metabolic perturbations of protein misfolding in *S. cerevisiae*?

As discussed in the introductory chapter, despite great improvements to our knowledge within the field of amyloid biology over the last 50 years, there remains a lack of fundamental understanding regarding the mechanistic actions of protein misfolding and the subsequent disease states they cause (Knowles *et al.* 2014). This, combined with the recent knowledge which presents a lack of distinction between the amyloid diseases and prion diseases, suggests that prions may be causing the same cellular disruptions as their amyloid counterparts (Jaunmuktane *et al.* 2015). However, despite great efforts to elucidate the cellular disruptions that take place due to prion presence, little is known or understood about these events and quite how these lead to cell death (Valastyan and Lindquist, 2014; Dobson, 2003). The prions of yeast present an opportunity to study this intricate mechanism without the complexities of mammalian brain chemistry (Liebman and Chernoff, 2012). It was therefore considered that investigating the metabolic disruptions of this process was both a novel and worthwhile endeavour and one which carried with it the possibility of increasing our fundamental understanding about the cellular changes that occur in response to prion formation and the metabolic pathways acted on by the presence of such promiscuous proteins.

7.2. Developing a reliable metabolomic methodology in *S. cerevisiae*

In order to achieve our aim, it was understood that a reliable, accessible (to us with our limited laboratory resources) and well evidenced technique for use with our model organism would be needed. Given the popularity of *S. cerevisiae* and the reliability with which excellent experimental protocols are usually available, surprisingly no single metabolomics protocol was found; instead a large number of varied protocols were available with the methods of

rating protocols and establishing a consensus within the field appearing vague (Canelas *et al.* 2009; Villas-Bôas *et al.* 2005; Gonzalez *et al.* 1999; de Koning and van Dam,1992). Therefore, experiments to determine the most effective protocol for use within untargeted metabolomics for the model organism, *S. cerevisiae*, were investigated. It was considered important when conducting these experiments that due attention was appropriated to all stages of sample preparation and that a ranking system that prioritised the biological relevance of findings was established. This included testing alternative methods of cell growth and extraction. However, due to the large number of samples needed to test these factors, issues such as quenching and the preparation of samples for LCMS, such as freeze drying, were not considered. Instead current literature, and the disputes within, were duly acknowledged and used as a guide for quenching and LCMS preparation.

With regards to the biological relevance and rankings of findings, the ability to accurately report on pre-established metabolic changes was thought to be an appropriate measure of the suitability of a methodology. The cellular conditions of aerobic respiration, anaerobic respiration and ammonium as a sole nitrogen source were decided upon as our input metabolic conditions for detection. Alongside two different growth conditions, the commonly used broth growth and the alternative filter growth method, and four different extraction techniques, BE, CM, PM and ACN (Kim *et al.* 2013; Tredwell *et al.* 2011; Boer *et al.* 2010; Canelas *et al.* 2009; Villas-Bôas *et al.* 2005; Prasad Maharjan and Ferenci 2003; Gonzalez *et al.* 1999; de Koning and van Dam 1992).

Multivariate analysis revealed that the anaerobic conditions had not been maintained within our experiment, revealing significant overlap with aerobic conditions when PCA models were produced. The ammonium as a sole nitrogen source, did however, show considerable differences when compared to the aerobic or control condition. The results of this study demonstrate that the choice of culturing technique as well as the metabolite extraction methodology can influence the quantity of the data observed within a metabolomics study, and that this can be reflected in the quality of subsequent pathway analysis. LCMS analysis and the subsequent ranking of methodologies revealed that using broth growth and ACN extraction method were the most effective at reporting the predicted changes in amino acid biosynthetic pathways, given the conditions used.

Given the drastically different biochemical conditions present within cells and their organelles, multiple methods of extraction may be advisable dependent upon which

metabolic area is of interest. Many researchers argue that maximisation of pathway hits could be achieved by using not one but two extraction methods (Riekeberg and Powers, 2017; Canelas *et al.* 2009; Oldiges and Takors 2005; Goodacre *et al.* 2004; Mashego *et al.* 2003). Close examination of the results from chapter three reveals that the use of both ACN and CM extraction would allow for maximum coverage of metabolite extraction from within amino acid biosynthesis.

This experiment introduced the workflow of metabolomics and enabled familiarisation with MZmine2, Simca and Metaboanalyst (Chong *et al.* 2019a; Chong *et al.* 2019b; Pluskal *et al.* 2010; Ni *et al.* 2016). For some time, the issue of converting the RAW data files into a format usable by MZmine2, caused issue; however, this was overcome using Proteowizard. Originally identification via MZmine2 using the YMDB was attempted, however the comparatively small number of IDs available within this library created a considerable bottleneck for our analysis. The use of Chemspider and Kegg libraries were then experimented with, however the lack of organism-specific selection created a considerable number of false IDs and so this method of identification was abandoned, with preference given to the organism-specific IDs that are possible via Metaboanalyst. First attempts at Simca analysis revealed the plethora of statistical analyses that are possible via Simca. S-plots were used to examine the most contributary factors of the score plots, however lack of IDs at this stage made this information difficult to interpret and so once Metaboanalyst's compound and pathways analysis had been conducted, the choice was made not to continue with additional Simca analysis. Within the pathway analysis of chapter three, KEGG pathway was used to overlay information. Although effective, an alternative was sought owing to the detail of the pathway models available and the technical difficulties in altering and colour coding models. This required a considerable amount of user time to produce the figures used throughout this chapter (Kanehisa and Goto, 2000).

As apparent from the discussion in chapter three, the use of broth growth with ACN is not the prevailing method throughout this thesis. Despite our experimental efforts, issues surrounding the peak selection and capture mode of our original data created a situation that, where analysis was performed unknowingly on a corrupt data set. The information provided throughout chapter three is as a result of having returned to the original data and corrected for this issue with automatic peak detection for the correct mode settings and recalculation of all of the subsequent analyses, hence we are confident about the conclusions drawn on this data. It was unfortunate that this issue was not apparent until the

analysis had been conducted and that the subsequent experiments for chapters four, five and six was scheduled prior to this realisation. Encouragingly, our findings indicate that all methods accurately reported on the predicted changes. If our flawed approach had been detrimental to subsequent experiments, this most likely would have resulted in the under reporting of pathway hits rather than change the overall findings.

The most valuable part of chapter three is the method by which the protocols were assessed, namely the reporting of pre-established inputs. It could be criticised however given that in reality our results may simply reflect the best method for the measurement of amino acid biosynthesis. It would be preferable to take heed of a variety of biosynthesis, degradation, energy and other pathway metabolism and determine predictable metabolic determinants of these pathways for use in a similar application. Such experiments although very large, have the potential not only to point to one or two excellent methods for use in untargeted metabolomics within yeast, but to provide specific and detailed guidance with which to conduct targeted metabolomics. This would allow researchers to achieve the best metabolite recovery possible for their intended metabolites or pathways of investigation. Such a gold standard protocol for any model organisms would be of high value to the metabolomics community and contribute to the extended study of the field.

The use of untargeted metabolomics as a generator for hypotheses which explain the cause of variation between groups, was used in the subsequent chapters to investigate the Rnq1 protein and its prion. These finding were combined with our current biological knowledge of [*RNQ*$^{+}$] behaviour to develop a better understanding of the mechanism and cellular consequences of the Rnq1 protein's presence, prion formation and the induction of cellular toxicity.

7.3. The native role of Rnq1 protein

Chapter four sought to investigate the native role of the Rnq1 protein and the effect of its functional loss on cells. As explained in detail within this chapter, the native role of the Rnq1 protein has long remained unknown - despite detailed descriptions of the native role of other prion proteins and the often-important roles these proteins are known to play cellularly (Chernova *et al.* 2014; Liebman and Chernoff, 2012).

Additionally, this study aimed to improve data analysis via recommendations within the literature to combine the results from two of the most popular freely available data analysis programmes, MZmine2 and XCMS, addressing concerns regarding FDR and variance in algorithmic processes underpinning data processing (Misra, 2018; Li *et al.* 2017; Myers, 2017). Similarly to that of Lui *et al.* (2017), we observed that large variances were present between the top IDs and pathways hits for each individual analysis program and so efforts were made (based on Lui *et al.* (2017) recommendation) to combine the results from both programs and use the overlapping information to extend the analysis. This proved a valuable addition to the study in removing a lot of the noise from our data set and only allowing analysis to take place on the remaining overlapping features. From our observations, we would strongly favour the use of XCMS over MZmine2, due to the often large number of peaks identified within MZmine2 and the highly editable user controls of MZmine2, all of which need to be preprogramed prior to data analysis.

Using this comparative data analysis method revealed several perturbed pathways linked to the ubiquinol pathways. Initially this created some confusion owing to the cytoplasmic nature of the Rnq1 protein and the mitochondrial location of these pathways. It was therefore considered that these may be the downstream effects of an earlier upstream influencer and so the upstream pathways of ubiquinol were investigated. This revealed a potential role for the Rnq1 protein as a negative controller for ubiquinone production. Known to be upregulated with oxidative stress and downregulated with age, these effects link well with the known influences of prion formation and oxidative stress. The secondary role of ubiquinone's COQ as an antioxidant, protecting the membrane phospholipids and lipoproteins from lipid peroxidation was of particular interest given that the details of this delicate interplay are not currently understood (Awad *et al.* 2018; Mohamed *et al.* 2015; Szkopinska, 2000; Genova *et al.* 1999; Mikosovska *et al.* 1999). Regrettably, hits were not available for many of the lipids further upstream of these changes, largely owing to the metabolite focused range of our m/z capture and our metabolite streamlined preparation protocols. Targeted metabolomic and lipidomic studies would be beneficial to confirm results found here.

Essentially chapter four reveals that the Rnq1 protein may play a cytoprotective role with respect to the lipid/mevalonate pathways. Whilst in its native conformation in times of limited stress, Rnq1 causes a relative downregulation of ubiquinone, diverting resources to the longer term cytoprotective membrane bound lipids. However as Rnq1 is readily able to

misfold when acted upon by stress, Rnq1 loses this proposed function and thus the suppression of ubiquinone is reversed, leading to high concentrations of ubiquinone (a much needed shorter-term protective agent) to engage in antioxidant defence whilst simultaneously downregulating the production of longer term cytoprotective sterols.

Further experimentation involving the sterol composition of *Δrnq1* cells, [*RNQ*$^{+}$] and [*rnq*$^{-}$] cells, including the addition of rnq1 mutants (either carrying known sterol mutations or pre-screening for their influence on sterol composition) may aid in confirming these findings. Comparing the effectiveness of the ergosterol targeting azoles (Rodrigues, 2018), common antifungal treatments, may show key differences in response between rnq1 mutants, *Δrnq1*, [*RNQ*$^{+}$] and [*rnq*$^{-}$] cells. In addition, detailed studies of the lag phase of the biological classes used within the experiments would be advised as changes in sterol composition are known to alter the lag phase of organisms considerably (Bhattacharya *et al.* 2018; Lui *et al.* 2017). To support this, targeted lipidomic studies focusing on sterol composition and quantification would also be desirable.

As briefly discussed in section 4.6. if confirmed, the potential for the presence of a homologous (by behaviour) protein in mammalian cells here is highly likely. The discovery of such a protein would have considerable potential to provide biomarkers to track and monitor disease progress of many of the amyloid diseases in a way that is not currently possible. Such a protein may also aid in understanding the often-mitochondrial nature of the disruptions within diseases such as Alzheimer's (Paglia *et al.* 2012).

7.4. The metabolic perturbations associated with the presence of the [*RNQ*$^{+}$] prion

The experiments described in chapter five were to investigate the metabolic perturbations of the presence of the [*RNQ*$^{+}$] prion. The results of this work were compared to those found previously (detailed in Chapter four) as the expected loss of function effects caused by [*RNQ*$^{+}$] state will be present in both knockout and [*RNQ*$^{+}$] cells. These needed to be determined so that these loss of function effects could be accounted for when comparing [*RNQ*$^{+}$] and [*rnq*$^{-}$] data to reveal the metabolic effects of prion presence alone. In addition, it was proposed that cells containing prions represent a pseudo 'disease' state that cells may respond to with a specific stress response. To distinguish between general stress responses

and those caused by the presence of the [*RNQ*$^{+}$] prion, a further analysis of cells under mild-oxidative stress were included in the study.

Stress mitigated pathway responses correlate well with the findings of Kang et al. (2012), providing pertinent information for our subsequent prion-free and prion-containing analyses. Subsequent comparisons between [*rnq*$^{-}$] and [*RNQ*$^{+}$] metabolic profiles confirmed many of the loss of function pathways hits from chapter four, permitting the removal of 70% of the pathways implicated by these experiments as loss of Rnq1 function. Considerable overlap was also seen between our mild-oxidative stress control and our prion samples, suggesting that the [*RNQ*$^{+}$] prion is having a deleterious effect on cells. It would be interesting to see if the same conditions that [*RNQ*$^{+}$] cells respond positively to also occurred when cells are under oxidative stress. This would suggest that an increase in the generalised stress response and upregulation of many of the cytoprotective elements is responsible for these observations, as opposed to any direct influence on the part of the prion. However, without such experimentation, the reasons for the shared relationship between prion stress and oxidative stress cannot not be explained. Comparisons between stress-condition pathways and prion-presence pathways reveal that the proline and arginine degradation pathways are unique to the stress-condition. This may be useful as a potential biomarker for distinguishing between damage caused as a result of oxidative stress and damage caused as a result of misfolded protein, although further research is needed to validate this.

Comparison of the metabolites from [*rnq*$^{-}$] and [*RNQ*$^{+}$] cells show downregulation within the degradation and recycling of the sphingolipids in [*RNQ*$^{+}$] cells. Although it is posited that this may be due to a lack of hits and that the shortest of these pathways (the degradation pathways) are more prominent compared to the larger synthesis pathways. Alternatively, the degradation of the sphingolipids may be reduced in [*RNQ*$^{+}$] samples due to the cells simply using all the sphingolipid resources available to them to deal with the number of misfolded proteins. Indeed, within amyloid biology, it is well documented that microdomains are formed by the sphingolipids and cholesterols which act as sites for the binding and oligomerisation of amyloidogenic proteins (Fantini and Yahi, 2010). Within cells carrying the mammalian prion, scrapie, sphingolipid rafts are known to decrease in concentration. It is posited that they are used by cells to aid in the formation of the PrP^{SC} from its normal conformer PrP^{C} (Naslavsky *et al.* 1999). This downregulation event in sphingolipid pathways, having not been previously observed in yeast prions, provides new and further evidence of homology between yeast prions and mammalian prions/amyloids.

The pathways highlighted with this chapter are supported by the results of chapter four, with consistencies in the up- and downregulations that occur within the sphingolipid, phospholipid, and sterol pathways. Known to maintain balance in the membrane, these pathways are linked via shared metabolites that are commonly used for cellular signalling (Pina *et al.* 2018; Han *et al.* 2010). It is understood that these pathways play a key role within protein quality control and membrane homeostasis, however understanding of the mechanistic nature of this maintenance is in its infancy. Han *et al.* (2010) found that the Orm proteins in yeast work to control membrane biogenesis coordinating and regulating sphingolipid synthesis. Testing multiple stress conditions on mutant phenotypes for one or more of the Orm proteins, revealed the nature of this mechanism. Further research investigating how mutant phenotypes of Orm proteins respond to prion presence could support the links to these pathways.

Given the number of distinct variants present within the prions, it would be interesting within this experimentation to have included a variety of variant strengths from weak to very strong (Stein and True, 2011; Kalastavadi and True, 2010; Bardill and True, 2010), to inform on the question whether the strength of a prion strain has a different metabolic fingerprint. This information would aid the yeast community with knowledge surrounding the presence of distinct strains and their cellular actions, as well as uncovering biomarkers that may be present within other amyloid species carrying with them the potential to confer a more exact prognosis for patients.

Further research is required to investigate if homologous perturbations are present among the different yeast prions and if a universality of metabolomic perturbations can be observed across the eukaryotic domain within other model organisms and within cell culture. Given the focus of our findings on the phospholipids and sterols throughout experimentation, it is recommended that these experiments use GC-MS for sterol capture and lipidomic. Within this preliminary work, although some changes have been observed, these has been limited due to the nature of experimentation.

Curiously, observations of downregulation in some of the amino acid pathways were observed in [*RNQ*$^{+}$] cells. These seem to correlate well with amino acid restriction studies, which demonstrate extended longevity under these conditions (Hu et al. 2014; Wu et al. 2013; Eisenberg et al. 2012). Links between these findings and the recent finding that the presence of [*PSI*$^{+}$] prolonged chronological lifespan of *S. cerevisiae*, bring about questions

regarding the comparative chronological lifespan of [*rnq*$^{-}$] and [*RNQ*$^{+}$] cells and if this amino acid restriction seen in [*RNQ*$^{+}$] cells may explain this observation in [*PSI*$^{+}$] and [*RNQ*$^{+}$].

In addition, the results of this work indicate that prion presence causes downregulations within the tyrosine pathways and similar findings have been observed by researchers working on other misfolding protein diseases such as PD (Hilaly *et al.* 2016). This suggests that there is a commonality here regarding metabolic disruptions, with yeast demonstrating the same metabolic responses to the presence of prions that have observed in mammalian cells. Indeed, the connections between the role of Rnq1 and the influence of the prions to lipid membranes and sphingolipid rafts share some commonality to the well-known surface glycoprotein PrP, which is localised in the membrane lipid rafts, interacting with a variety of proteins and regulating synapse transmission (Bieberich, 2018; Glich *et al.* 2006; Collinge, 2005; Unterberger *et al.* 2005). There is a growing body of evidence suggesting the importance between misfolded proteins, lipid membranes and the interactions between the two and yeast and their prions would provide a convenient and replicable model system in which to undertake these investigations.

7.5. The metabolic perturbations associated with toxicity of overexpression of Rnq1 protein

Chapter six aimed to investigate the induction of toxicity to cells caused by Rnq1 overexpression but also reaffirmed many of the previous findings from chapter five, strengthening our argument for the universality of prion mitigated metabolic effects. However, the results within this chapter did not reliably identify the metabolic disruptions that arise as a result of induced Rnq1 toxicity. Considerable overlap between the pathways effected by [*RNQ*$^{+}$] and the pathways affected by the presence of the overexpression plasmid were observed. The merits of changing the GAL promoter on the plasmid used for the overexpression of Rnq1 were discussed (Section 6.6.) to aid in mitigating the effects of the plasmid from those cause by Rnq1-induced toxicity (Da Silva and Srikrishnan, 2012; Blazek and Alper, 2012).

Analyses performed on samples of [*RNQ*$^{+}$] and [*RNQ*$^{+}$] with over expression highlighted the sphingolipid pathways as being the most perturbed. However, upon closer examination of this data, a sporadic pattern of up and downregulation is seen across both sample groups, making it difficult to draw conclusions from the data. It was considered that this may be due

to the lack of NIM data in this study when compared to the previous experiments. Potentially this NIM data may have provided a clarity to the pathways overlays that was not possible using only PIM data. Furthermore, the change of media from glucose to galactose in this experiment prevented comparisons with previous studies and highlighted the requirement for the inclusion of a Δrnq strain within this experiment. The loss of Rnq1 function effects may have changed on alternative media and were not seen in the top pathway hits as might have been expected, although other explanations could have been the different growth stage that these cells were sampled at.

These experiments have provided valuable information about how to investigate toxicity in cells using a metabolomics approach and many of these are discussed in Section 6.5.3. PCAs and OPLS-DA provide ample evidence that all biological classes are metabolically distinct, providing a proof of concept. It would be advised that further research used NMR instead of, or alongside LCMS, given its non-destructive nature and that other yeast models for toxicity including Meriin *et al.* (2002) polyQ polypeptide 103Q be used.

7.6. Conclusion

This research has revealed that yeast prions share many of the metabolic perturbations associated with prion and amyloid disease in mammals (Bourgognon *et al.* 2018; Varma et al. 2018; Wilkins and Trushina, 2018; Laurens et al. 2015; Han et al. 2011; Oresic et al. 2011; Pushie *et al.* 2011; Allison *et al.* 2008; Allison *et al.* 2007). These include observations of significant perturbations within lipid metabolism (the sphingolipid pathways, geranylgeranyldiphosphate biosynthesis I (via mevalonate) and ergosterol), amino acid metabolism and some disruption to energy metabolism. Having confirmed these similarities, questions arise regarding the impact of this knowledge. Given that most cells are known to be subject to considerable harm from misfolded proteins, how yeast circumnavigate this harm, if indeed they are harmful at all and if not why not, may provide valuable information about the pathology of these diseases and their influence on metabolism.

The native role of the Rnq1 protein downregulates the ubiquinone biosynthesis pathways within cells, suggesting that Rnq1 protein may play a lipid/mevalonate-based cytoprotective role as a regulator of ubiquinone production. The intrinsic links between the native role and realities of prion formation are not surprising, given the Lamarckian or epigenetic nature of the [*RNQ*$^{+}$] prion and its beneficial effects in some environmental conditions (Chakravarty *et*

al. 2020; Itakura *et al.* 2020; Jarosz et al. 2014a; Halfmann *et al.* 2012). Are these beneficial effects of prion formation present in some of the mammalian prions; simply masked in later life by the devastating effects of their presence and potentially providing a similar plasticity to changing environmental conditions? Or are yeast cells simply able to circumnavigate the toxic effects of prion formation due to their continual transmission of those prions to daughter cells and their relatively short life spans? A common technique within wine making may be able to help with this inquiry, Potassium sorbate added to a growing yeast culture has the ability not to kill the yeast therein, but to effectively make all the yeast present sterile, impairing their ability to reproduce. Could such an addition to cells containing [*RNQ*$^{+}$] prions, or any other yeast prion for that matter, cause the induction of the cytotoxic effects so commonly seen in mammalian cells? With regard to prion proliferation with yeast cells, could the observations of perturbations within the sphingolipid pathways here indicative of barrier diffusion (Clay *et al.* 2014) or unequal (mother to daughter cell) partitioning mechanism (Cox and Tuite, 2019; Ness *et al.* 2017).

Paglia *et al*. (2018) argue that the use of metabolomics within the amyloid field has not been given the attention it deserves, while Narayan *et al*. (2014) state that the use of appropriate models to increase fundamental understanding of amyloid disease should be overlooked at our peril. Hence still in its infancy, metabolomics experimentation that does exist within the field, much like this present study, tend to propose more questions than they answer, further strengthening the argument for more research (Shao and Le, 2019; Paglia *et al.* 2018; Varma et al. 2018). With an aging global population and amyloid disease set to be the ‘next great plague’, all opportunities to increase our understanding of these diseases must be explored (Olshansky *et al.* 2005; World Alzheimer report, 2016; World Alzheimer report, 2016). Even if effective treatments were found, it is unlikely that they will simply be made available to everyone. Efficient and effective early diagnosis will still be a requirement of these treatments, if they are to aid in the prevention of these devastating diseases to as many people as possible, and so it seems there is an important place for metabolomics firmly rooted in this field (Jiang *et al*. 2019; Paglia *et al.* 2018).

Chapter Eight - References

Aguilar, M.I. (2004) 'Reversed-phase high-performance liquid chromatography', In: Aguilar MI. (eds) HPLC of Peptides and Proteins. Methods in Molecular Biology, vol 251. Springer, Totowa, NJ

Alam, M.T., Zelezniak, A., Mulleder, M., Shiliaha, P., Schwarz, R., Capuano, F., Vowinckel, J., Radmanesfahar, E., Kruger, A., Calvani, E., Michel, S., Borno, S., Christen, S., Patil, K.R., Timmermann, B., Lilley, K.S. and Ralser, M. (2016) 'The metabolic background is a global player in *Saccharomyces* gene expression epistasis', Nat Microbiol., 1(3), 15030, doi: 10.1038/nmicrobiol.2015.30

Ali, S.S., Marcondes, M.C.G., Bajova, H., Dugan, L.L. and Conti, B. (2010) 'Metabolic depression and increased reactive oxygen species production by isolated mitochondria at moderately lower temperatures', J Biol Chem., 285 (42), pp. 32522-32528, doi: 10.1074/jbc.M110.155432

Allen, D.R. and McWhinney, B.C. (2019) 'Quadrupole time-of-flight mass spectrometry: a paradigm shift in toxicology screening applications', Clin Biochem Rev., 40(3), pp. 135–146, doi: 10.33176/AACB-19-00023

Allison, G. G., Horton, R. A., Rees Stevens, P., Jackman, R., and Moorby, J. M. (2007) 'Changes in plasma metabolites and muscle glycogen are correlated to bovine spongiform encephalopathy in infected dairy cattle', Res. Vet. Sci., 83, pp. 40–46, doi: 10.1016/j.rvsc.2006.11.008

Allison, G. G., Rees Stevens, P., Heasman, L., Davies, A., Jackman, R., and Moorby, J. M. (2008) 'Effect of scrapie incubation on the concentrations of plasma amino acids and L-lactate in infected lambs', Vet. Res. Commun., 32, pp. 591–97, doi: 10.1007/s11259-008-9067-5

Allwein, B., Kelly, C., Kammoonah, S., Mayor, T. and Cameron, D.M. (2019) 'Prion-dependent proteome remodeling in response to environmental stress is modulated by prion variant and genetic background', Prion, 13 (1), pp. 53-64, doi: 10.1080/19336896.2019.1583041

Alper, T. (1993) 'The scrapie enigma: insights from radiation experiments', Radiat Re., 135, pp. 283-292, doi: 10.2307/3578866

Alper, T., Cramp, W.A., Haig, D.A. and Clarke, M.C. (1967) 'Does the agent of scrapie replicate without nucleic acid?', Nature, 214, pp. 764-766, doi: 10.1038/214764a0

Alper, T., Haig, D.A. and Clarke, M.C. (1966) 'The exceptionally small size of the scrapie agent', Biochem Biophy Res Comm., 22, pp. 278-284, doi: 10.1016/0006-291x(66)90478-5

Anfinsen, C.B., Haber, E., Sela, M., White, F.H. (1961) 'The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain', Proc Natl Acad Sc. USA, 47, pp. 1309-1314, doi: 10.1073/pnas.47.9.1309

Antoniewicz, M.R. (2015) 'Methods and advances in metabolic flux analysis: a mini-review', J Ind Microbiol Biotechnol., 42 (3), pp. 317-325, doi: 10.1007/s10295-015-1585-x

Audas, T.E., Audas, D.E., Jacob, M.D., Ho, D.J.J., Khacho, M., Wang, M., Perera, J.K., Gardiner, C., Bennett, C.A., Head, T., Kryvenko, O.N., Jorda, M., Daunert, S., Malhorta, A., Trinkle-Mulcahy, L., Gonzalgo, M.L. and Lee, S. (2016) 'Adaptation to Stressors by Systemic Protein Amyloidogenesis', Dev Cell., 39 (2), pp. 155-168, doi: 10.1016/j.devcel.2016.09.002

Auer, S., Meersman, F., Dobson, C. M. and Vendruscolo, M. (2008) 'A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates', PLoS Comp. Biol., 4, e1000222, doi: 10.1371/journal.pcbi.1000222

Aviram, L.Y., McCooeye, M. and Mester, Z. (2016) 'Determination of underivatized amino acids in microsamples of a yeast nutritional supplement by LC-MS following microwave assisted acid hydrolysis', Anal. Methods, 8, pp. 4497-4503, doi: 10.1039/C6AY00407E

Awad, A.M., Bradley, M.C., Fernandez-del-Rio, L., Nag, A., Tsui, H.S. and Clarke, C.F. (2018) 'Coenzyme Q10 deficiencies: pathways in yeast and humans', Essays Biochem, 62 (3), pp. 361-376, doi: 10.1042/EBC20170106

Babu, M.M., Kriwacki, R.W. and Pappu, R.V. (2012) 'Structural biology. Versatility from protein disorder', Science, 337 (6101), pp. 1460-1461, doi: 10.1126/science.1228775

Baidoo, E.E.K., Benke, P.I. and Keasling, J.D. (2012) 'Mass spectrometry-based microbial metabolomics', In: Navid A. (eds) Microbial Systems Biology. Methods in Molecular Biology (Methods and Protocols), Humana Press, Totowa, NJMicrobial Systems Bio, 881, pp. 215-278, doi: 10.1007/978-1-61779-827-6_9

Baldwin, A.J., Knowles, T.P.J., Tartaglia, G.G., Fitzpatrick, A.W., Devlin, G.L., Shammas, S.L., Waudby, C.A., Mossuto, M.F., Meehan, S., Gras, S.L., Christodoulou, J., Anthony-Baldwin, R.L. (2011) 'Metastability of native

proteins and the phenomenon of amyloid formation', J. Am. Chem. Soc., 133 (36), pp. 14160-14163, doi: 10.1021/ja2017703

Balser, K., Oesch, B., Scott, M., Westaway, D., Walchli, M., Growth, D.F., McKinley, M.P., Pruisner, S.P. and Weissmann, C. (1986) 'Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene', Cell, 46 (3), pp. 417-428, doi: 10.1016/0092-8674(86)90662-8

Bamji-Stocke, S., van Berkel, V., Miller, D.M. and Frieboes, H.B. (2018) 'A review of metabolism-associated biomarkers in lung cancer diagnosis and treatment', Metabolomics, 14, pp. 81, doi: 10.1007/s11306-018-1376-2

Banerjee, S. and Mazumdar, S. (2012) Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte, International Journal of Analytical Chemistry, 2012, pp. 1-40, doi.org/10.1155/2012/282574

Bardill, J.P. and True, H.L. (2010) 'Heterologous prion interactions are altered by mutations in the prion protein Rnq1 protein', J Mol Biol., 388 (3), pp. 583-596, doi: 10.1016/j.jmb.2009.03.036

Bennett, J.H., Rhodes, F.A. and Robson, H.N. (1959) 'A possible genetic basis for Kuru', Aust. An. Med., 11, pp. 169, PMC1931994

Benton, H.P., Ivanisevic, J., Mahieu, N., Kurczy, M., Johnson, C., Franco, L., Rinehart, D., Valentine, E., Gowda, H., Ubhi, K., Tautenhahn, R., Gieschen, A., Fields, M., Patti, G.J. and Siuzdak, G. (2014) 'Autonomous metabolomics for rapid metabolite identification in global profiling', Anal. Chem., 87, pp. 884-891, doi: 10.1021/ac5025649

Bhattacharya, S., Esquivel, B. D. and White, T.C. (2018) 'Overexpression or deletion of ergosterol biosynthesis genes alters doubling time, response to stress agents, and drug susceptibility in *Saccharomyces cerevisiae*', mBio, 9:4, e01291-18, doi: 10.1128/mBio.01291-18

Bieberich, E. (2018) 'Sphingolipids and lipid rafts: Novel concepts and methods of analysis', Chem. Phys. Lipids, 216, pp. 114-131, doi: 10.1016/j.chemphyslip.2018.08.003.

Bingol, K. (2018) 'Recent advances in targeted and untargeted metabolomics by nmr and ms/nmr methods', High Throughput, 7 (2), doi: 10.3390/ht7020009

Blazeck, J. and Alper, H.S. (2012) 'Promoter engineering: Recent advances in controlling transcription at the most fundamental level', Biotechnology Journal, 8 (1), pp. 46-58, doi: 10.1002/biot.201200120

Boer, V.M., Crutchfield, C.A., Bradley, P.H., Botstein, D. and Rabinowitz, J.D. (2010) 'Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations', Mol. Biol. Cell., 21 (1), pp. 198-211, doi: 10.1091/mbc.E09-07-0597

Bolton, D.C., McKinley, M.P. and Prusiner, S.B. (1982) 'Identification of a protein that purifies with the scrapie prion', Science, 218 (4579), pp. 1309-1311, doi: 10.1126/science.6815801

Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K.D., Ploegh, H.L. and Kessler, B.M. (2002) 'Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family', Chem Bio., 9 (10), pp. 1149-1159 , doi: 10.1016/s1074-5521(02)00248-x.

Bosley, K.S., Botchan, M., Brednoord, A.L., Carroll, D., Charo, R.A., Charpentier, E., Cohen, R., Corn, J., Doudna, J., Feng, G., Greely, H.T., Isasi, R., Ji, W., Kim, J., Knoopers, B., Lanphier, E., Li, J., Lovell-Badge, R., Martin, G.S., Moreno, J., Naldini, L., Pera, M., Perry, A.C.F., Venter, J.C., Zhang, F. and Zhou, Q. (2015) 'CRISPR germline engineering—the community speaks', Nat. Biotech., 33 (5), pp. 478-486, doi: 10.1038/nbt.3227

Botstein, D. and Fink, G.R. (2011) 'Yeast: an experimental organism for 21st Century biology', Genetics, 189 (3), pp. 695-704, doi: 10.1534/genetics.111.130765

Botstein, D., Chervitz, S.A. and Cherry, J.M. (1997) 'Yeast as a model organism', Science, 277 (5330), pp. 1259-60, doi: 10.1126/science.277.5330.1259

Bourgognon, J.M., Spiers, J.G., Scheiblich, H., Antonov, A., Bradley, S.J., Tobin, A.B. and Steinert, J.R. (2018) 'Alterations in neuronal metabolism contribute to the pathogenesis of prion disease', Cell Death Differ., 25 (8), pp. 1408-1425, doi: 10.1038/s41418-018-0148-x

Bowen, B.P. and Northern, T.R. (2010) 'Dealing with the unknown: metabolomics and metabolite atlases', J. Amer. Soc. Mas. Spec., 21 (9), pp. 1471-1476, doi: 10.1016/j.jasms.2010.04.003

Bradley, M.E., Edskes, H.K., Hong, J.Y., Wickner, R.B. and Liebman, S.W. (2002) 'Interactions among prions and prion "strains" in yeast', Proc Natl Acad Sc. USA, 99 (4), pp. 16392-16399, doi: 10.1073/pnas.152330699

Brandner, S. and Jaunmuktane, Z. (2017) ‘Prion disease: experimental models and reality’, Acta Neuropathol, 133 (2), pp. 197-222, doi: 10.1007/s00401-017-1670-5

Brauer, M.J., Yuan, J., Bennett, B.D., Lu, W., Kimball, E., Botstein, D. and Rabinowitz, J.D. (2006) ‘Conservation of the metabolomic response to starvation across two divergent microbes’, Proc Natl Acad Sc. USA, 103 (51), pp. 19302-19307, doi: 10.1073/pnas.0609508103

Broadhurst, D., Goodacre, R., Reinke, S.N., Kuligowski, J., Wilson, I.D., Lewis, M.R. and Dunn, W.B. (2018) ‘Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies’, Metabolomics, 14, doi: 10.1007/s11306-018-1367-3

Bucciantini, M., Calloni, G., Chiti, F., Formigli, L., Nosi, D., Dobson, C.M. and Stefani, M. (2004) ‘Prefibrillar amyloid protein aggregates share common features of cytotoxicity’, J Biol Chem, 279 (30), pp. 31374-31382, doi: 10.1074/jbc.M400348200

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M. and Stefani, M. (2002) ‘Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases’, Nature, 416, pp. 507-517, doi: 10.1038/416507a.

Bueler, H., Aguzzi, A., Sailer, A., Greiner, R.A., Autenried, P. Aguet, M. and Weissman, C. (1993) ‘Mice devoid of PrP are resistant to scrapie’, Cell, 73 (7), pp. 1339-1347, doi: 10.1016/0092-8674(93)90360-3

Byrne, L.J., Cole, D.J., Cox, B.S., Ridout, M.S., Morgan, B.J.T. and Tuite, M.F. (2009) ‘The number and transmission of [*PSI*$^{+}$] prion seeds (propagons) in the yeast *Saccharomyces cerevisiae*’, PLoS ONE, 4 (3), e4670, doi: 10.1371/journal.pone.0004670

Byrne, L.J., Cox, B,S, Cole, D.J., Ridout, M.S., Morgan, B.J.T. and Tuite, M.F. (2007) ‘Cell division is essential for elimination of the yeast [PSI+] prion by guanidine hydrochloride’, Proc Natl Acad Sc. USA, 104 (28), pp. 11688-11693, doi: 10.1073/pnas.0701392104

Cai, X., Chen, J., Xu, H., Lui, S., Jiang, Q., Halfmann, R. and Chen, Z.J. (2014) ‘Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation’, Cell, 156 (6), pp. 1207-1222, doi: 10.1016/j.cell.2014.01.063

Campbell, K., Vowinckel, J., Keller, M.A. and Ralser, M. (2016) ‘Methionine metabolism alters oxidative stress resistance via the pentose phosphate pathway’, Antioxid Redox Signal, 24 (10), pp. 543-547, doi: 10.1089/ars.2015.6516

Campioni, S., Mannini, B., Zampagni, M, Pensalfi, A., Parrini, C., Evangelisti, E., Relini, A., Stefani, M., Dobson, C.M., Cecchi, C. and Chiti, F. (2010) ‘A causative link between the structure of aberrant protein oligomers and their toxicity’, Nature Chem. Biol., 6, pp. 140–147, doi: 10.1038/nchembio.283.

Canelas, A.B., Harrision, N. and Nielsen, J. (2010) ‘Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains’, Nature comms., 1(9), 145, doi: 10.1038/ncomms1150

Canelas, A.B., Pierick, A.T., Ras, C., Seifar, R.M., van Dam, J.C., van Gulik, W.M. and Heijnen, J.J. (2009) ‘Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics’, Anal Chem., 81 (17), pp. 7379-7389, doi: 10.1021/ac900999t

Canelas, A.B., Ras, C., Pierick, A.T., van Dam, J.C., Heijen, J.J. and van Gulik, W.M. (2008) ‘Leakage-free rapid quenching technique for yeast metabolomics’, Metabolomics, 4 (3), pp. 226-239, doi: 10.1007/s11306-008-0116-4

Carlo, M.D. (2012) ‘Simple model systems: a challenge for Alzheimer's disease’, Immun Ageing, 9 (1), pp. 3, doi: 10.1186/1742-4933-9-3

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., Tisser, C., Walk, T.C., Zhang, P. and Karp, P.D. (2007) ‘The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases’, Nuc Acids Res, 36 (1), pp. D623-D631, doi: 10.1093/nar/gkm900

Castilla, J., Saa, P., Hetz, C. and Soto, C. (2005) ‘*In vitro* generation of infectious scrapie prions’, Cell, 121, pp. 195-206, doi: 10.1016/j.cell.2005.02.011

Castrillo, J.I., Hayes, A., Mohammed, S., Gaskell, S.J. and Oliver, S.G. (2003) ‘An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry’, Phytochemistry, 62 (6), pp. 929-937, doi: 10.1016/S0031-9422(02)00713-6

Caudy, A.A., Mulleder, M. and Ralsar, M. (2017) 'Metabolomics in Yeast', Cold Spring Harb. Protoc., 2017(9), pdb.top083576, doi: 10.1101/pdb.top083576

Caughey, B. and Lansbury, P. T. (2003) 'Protofibrils, pores, fibrils and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders', Annu. Rev. Neurosci., 26, pp. 267–298, doi: 10.1146/annurev.neuro.26.010302.081142

Chae, L., Kim, T., Nilo-Poyanco, R. and Rhee, S.Y. (2014) 'Genomic signatures of specialized metabolism in plants', Science, 344 (6183), pp. 510-513, doi: 10.1126/science.1252076

Chakrabortee, S., Kayatekin, C., Newby, G.A., Mendillo, M.L., Lancaster, A. and Lindquist, S. (2016) 'Luminidependens (LD) is an *Arabidopsis* protein with prion behavior', Proc Natl Acad Sc. USA, 113 (21), pp. 6065-6070, doi: 10.1073/pnas.1604478113

Chakravarty, A.K., Smejkal, T., Itakura, A.K., Garcia, D.M. and Jarosz, D.F. (2020) 'A non-amyloid prion particle that activates a heritable gene expression program', Mol. Cell., 77 (2), pp. 251-265, doi: 10.1016/j.molcel.2019.10.028

Chambers, M.C., MacLean, B., Burke, R., Amode, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L. and Mallick, P. (2012) 'A cross-platform toolkit for mass spectrometry and proteomics', Nature Biotechnology, 30, pp. 918-920, doi: 10.1038/nbt.2377

Chandler, R.L. (1961) 'Encephalopathy in Mice produced by inoculation with Scrapie brain material', Prelim. Comm., 277 (7191), pp. 1378-1379, doi: 10.1016/s0140-6736(61)92008-6

Cheon, M., Chang, I., Mohanty, S., Luheshi, L.M., Dobson, C.M., Vendruscolo, M. and Favrin, G. (2007) 'Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils', PLoS Comp. Biol., 3, pp. 1727–1738, doi: 10.1371/journal.pcbi.0030173

Chernoff, Y.O, Lindquist, S.L., Ono, B., Inge-Vechtomov, S.G. and Liebman, S.W. (1995) 'Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [*PSI*$^{+}$]', Science, 268 (5212), pp. 880-884, doi: 10.1126/science.7754373

Chernoff, Y.O. (2007) 'Stress and prions: Lessons from the yeast model', FEBS letters, 581 (19), pp. 3695-3701, doi: 10.1016/j.febslet.2007.04.075

Chernova, T.A., Wilkinson, K.D. and Chernoff, Y.O. (2014) 'Physiological and environmental control of yeast prions', FEMS Micro Biol Rev., 38 (2), pp. 326-344, doi: 10.1111/1574-6976.12053.

Chesebro, B., Race, R., Wehrly, K., Nishio, J., Bloom, M., Lechner, D., Bergstrom, S., Robbins, K., Mayer, L., Keith, J.M., Garon, C. and Haase, A. (1985) 'Identification of scrapie prion protein-specific mRNA in scrapie-infected and uninfected brain', Nature, 315, pp. 331-333, doi: 10.1038/315331a0

Chien, P. and Weissman, J.S. (2001) 'Conformational diversity in a yeast prion dictates its seeding specificity', Nature, 410, pp. 223-227, doi: 10.1038/35065632;

Chiti, F. and Dobson, C.M. (2006) 'Protein misfolding, functional amyloid, and human disease', Annu. Rev. Biochem., 75, pp. 333-366, doi: 10.1146/annurev.biochem.75.101304.123901

Chong, J. and Xia, J. (2018) 'MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data', Bioinformatics, 27, pp. 4313–4314, doi: 10.1093/bioinformatics/bty528

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S. and Xia, J. (2018) 'MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis', Nucl. Acids Res, 46, pp. W486-494, doi: 10.1093/nar/gky310

Chong, J., Wishart, D.S. and Xia, J. (2019a) 'Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis', Current Protocols in Bioinformatics 68, e86, doi: 10.1002/cpbi.86

Chong, J., Yamamoto M. and Xia, J. (2019b) 'MetaboAnalystR 2.0: From raw spectra to biological insights', Metabolites 2019, 9 (3), pp. 57, doi: 10.3390/metabo9030057

Clay, L., Caudron, F., Denoth-Lippuner, A., Boettcher, B., Frei, S.B., Snapp, E.L. and Barral, Y. (2014) 'A sphingolipid-dependent diffusion barrier confines ER stress to the yeast mother cell', eLife, e01883, doi: 10.7554/eLife.01883

Collinge, J. (2005) 'Molecular neurology of prion disease', J Neuro Neurosurg Psychiatry, 76 (7), pp. 906-919, doi: 10.1136/jnnp.2004.048660

Collinge, J. and Clarke, A.R. (2007) 'A general model of prion strains and their pathogenicity', Science, 318, pp. 930-936, doi: 10.1126/science.1138718

Comber T (1772) 'Real improvements in agriculture (on the principles of A Young Esq.)', In: Letters to Reade Peacock, Esq. and to Dr Hunter, Physician in York, concerning the rickets in sheep. Nicoll, London

Courchesne, W.E. and Magasanik, B. (1988) 'Regulation of nitrogen assimilation in *Saccharomyces cerevisiae*: roles of the URE2 and GLN3 genes', J Bacteriol., 170 (2), pp. 708-713, doi: 10.1128/jb.170.2.708-713.1988

Cox, B. and Tuite, M. (2018) 'The life of [PSI]', Curr. Genet., 64 (1), pp. 1-8, doi: 10.1007/s00294-017-0714-7

Cox, B., Ness, F. and Tuite, M. (2003) 'Analysis of the generation and segregation of propagons: entities that propagate the [*PSI*$^{+}$] prion in yeast', Genetics, 165 (1), pp. 23-33, PMC1462756

Cox, B.S., Tuite, M.F. and McLaughlin, C.S. (1988) 'The psi factor of yeast: a problem in inheritance', Yeast, 4 (3), pp. 159-178, doi: 10.1002/yea.320040302

Creek, D.J., Vincent, I.M. and Barrett, M.P. (2013) 'Pharmacological metabolomics in trypanosomes', Trypanosomatid Diseases, pp. 37-56, doi: 10.1002/9783527670383.ch3

Cremades, N., Cohen, S.I.A., Deas, E., Abramov, A.Y., Chen, A.Y., Orte, A., Sandal, M., Clarke, R.W., Dunne, P., Aprile, F.A., Bertoncini, C.W., Wood, N.W., Knowles, T.P.J., Dobson, C.M. and Klenerman, D. (2012) 'Direct observation of the interconversion of normal and toxic forms of α-synuclein', Cell, 149, pp. 1048–1059, doi: 10.1016/j.cell.2012.03.037

Cullie, J. and Chelle, P.L. (1936) 'La maladie dite tremblante du mouton est-elle inculable?', Com Rend l'Acad Sci., 203, pp. 1552-1554

Cullie, J. and Chelle, P.L. (1939) 'Experimental transmission of trembling to the goat', Com Rend l'Acad Sci., 208, pp. 1058-1160

Cunningham, T.S., Andhare, R. and Cooper, T.G. (2000) 'Nitrogen catabolite repression of DAL80 expression depends on the relative levels of Gat1p and Ure2p production in *Saccharomyces cerevisiae*', J Biol Chem., 275 (19), pp. 14408-14414, doi: 10.1074/jbc.275.19.14408

D'Autreaux, B. and Toledano, M.B. (2007) 'ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis', Nat Rev Mol Cell Biol., 8 (10), pp. 813-824, doi: 10.1038/nrm2256

Da Silva, N.A. and Srikrishnan, S. (2012) 'Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*', FEMS Yeast Res., 12 (2), pp. 197-214, doi: 10.1111/j.1567-1364.2011.00769.x

Dahl-Lassen, R., Hecke, J., Jorgensen, H., Bukh, C., Anderson, B. and Schjoerring, J.K. (2018) 'High-throughput analysis of amino acids in plant materials by single quadrupole mass spectrometry', Plant Methods, 14, doi: 10.1186/s13007-018-0277-8

Daskalaki, A., Vasiliadoum I.A., Bellou, S., Tomaszewske-Hetman, L., Chatzikotoula, C., Kompoti, B., Papanikolaou, S., Vayenas, D., Pavlou, S. and Aggelis, G. (2018) 'Data on cellular lipids of *Yarrowia lipolytica* grown on fatty substrates', Data in Brief, 21, pp. 1037-1044, doi: 10.1016/j.dib.2018.10.116

Dasuri, K., Zhang, L., Keller, J.N. (2013) 'Oxidative stress, neurodegeneration, and the balance of protein degradation and protein synthesis', Free Radical Biology and Medicine, 62, pp. 170-185, doi: 10.1016/j.freeradbiomed.2012.09.016

De Hoffmann, E. and Stroobant, V. (2007) 'Mass Spectrometry, principles and applications', Chichester, UK, John Wiley and Sons

de Koning, W. and van Dam, K. (1992) 'A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH', 204 (1), pp. 118-123, doi: 10.1016/0003-2697(92)90149-2

Debets, A.J.M., Dalstra, H.J.P., Slakhorst, M., Koopmanschap, B., Hoeksta, R.F. and Saupe, S.J. (2012) 'High natural prevalence of a fungal prion', Proc Natl Acad Sc. USA, 109 (26), pp. 10432-10437, doi: 10.1073/pnas.1205333109

Deng, C., Zhang, J., Yu, X., Zhang, W., Zhang, Z. (2004) 'Determination of acetone in human breath by gas chromatography-mass spectrometry and solid-phase microextraction with on-fiber derivatization', 810 (2), pp. 269-275, doi: 10.1016/j.jchromb.2004.08.013.

DePace, A.H., Santoso, A., Hillner, P. and Weissman, J.S. (1998) 'A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion', Cell, 93 (7), pp. 1241-1252, doi: 10.1016/S0092-8674(00)81467-1

Derkatch , I.L., Uptain, S.M., Outeiro, T.F., Krishnan, R., Lindquist, S.L. and Liebman, S.W. (2004) 'Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the *de novo* formation of the [*PSI*$^{+}$] prion in yeast and aggregation of Sup35 *in vitro*', Proc Natl Acad Sc. USA, 101 (35), pp. 12934-1293, doi: 10.1073/pnas.0404968101

Derkatch, I.L., Bradley, M.E., Hong, J.Y. and Liebman, S.W. (2001) 'Prions affect the appearance of other prions: The story of [*PIN*$^{+}$]', Cell, 106 (2), pp. 171-182, doi: 10.1016/s0092-8674(01)00427-5

Derkatch, I.L., Bradley, M.E., Masse, S.V., Zadorsky, S.P., Polozkov, G.V., Inge-Vechtomov, S.G. and Liebman, S.W. (2000) 'Dependence and independence of [*PSI*$^{+}$] and [PIN^{+}]: a two-prion system in yeast?', EMBO J., 19, pp. 1942–1952, doi: 10.1093/emboj/19.9.1942

Derkatch, I.L., Bradley, M.E., Zhou, P., Chernoff, Y.O. and Liebman, S.W. (1997) 'Genetic and environmental factors affecting the *de novo* appearance of the [*PSI*$^{+}$] prion in *Saccharomyces cerevisiae*', Genetics, 147, pp. 507–519, PMC1208174

Dettmer, K., Aronov, P.A. and Hammock, B.D. (2007) 'Mass spectrometry-based metabolomics', Mass Spectrom Rev, 26 (1), pp. 51-78, doi: 10.1002/mas.20108

Dill, K.A. and Chan, H.S. (1997) 'From Levinthal to pathways to funnels', Nature Struct. Bio., 4, pp. 10-19, doi: 10.1038/nsb0197-10

Dill, K.A. and MacCallum, J.L. (2012) 'The protein-folding problem, 50 years on', Science, 338 (6110), pp. 1042-1046, doi: 10.1126/science.1219021

Dobson, C. M. (1999) 'Protein misfolding, evolution and disease', Trends Bioch. Sci., 24, pp. 329–332, doi: 10.1016/s0968-0004(99)01445-0

Dobson, C.M. (2003) 'Protein folding and misfolding', Nature, 426, pp. 884-890, doi: 10.1038/nature02261

Dong, M.X., Feng, X., Xu, X.M., Hu, L., Liu, Y., Jia, S.Y., Li, B., Chen, W. and Wei, Y.D. (2018) 'Integrated analysis reveals altered lipid and glucose metabolism and identifies NOTCH2 as a biomarker for Parkinson's disease related depression', Frontiers in Molecular Neuroscience, 11, doi: 10.3389/fnmol.2018.00257

Doronina, V.A., Staniforth, G.L., Speldewinde, S.H., Tuite, M.F. and Grant, C.M. (2015) 'Oxidative stress conditions increase the frequency of *de novo* formation of the yeast [PSI+] prion', Mol Microbiol., 96 (1), pp. 163-174, doi: 10.1111/mmi.12930

Douglas, P.M., Summers, D.W. and Cyr, D.M. (2009) 'Molecular chaperones antagonize proteotoxicity by differentially modulating protein aggregation pathways', Prion, 3 (2), pp. 51-58, doi: 10.4161/pri.3.2.8587

Douglas, P.M., Treusch, S., Ren, H., Halfmann, R., Duennwald, M.L., Linquist, S. and Cyr, D.M. (2008) 'Chaperone-dependent amyloid assembly protects cells from prion toxicity', Proc Natl Acad Sc. USA, 105 (20), pp. 7206-7211, doi: 10.1073/pnas.0802593105

Dragos, A., Kovacs, A.T. and Claeesen, D. (2017) 'The role of functional amyloids in multicellular growth and development of gram-positive bacteria', Biomolecules, 7 (4), pp. 60 doi: 10.3390/biom7030060

Dumas, M.E. and Davidovic, L. (2013) 'Metabolic phenotyping and systems biology approaches to understanding neurological disorders', F1000 Prime Rep., 5 (18), doi: 10.12703/P5-18

Duncan, K.D., Fyrestam, J. and Lanekoff, I. (2019) 'Advances in mass spectrometry based single-cell metabolomics', Analyst, 144, pp. 782-793, doi: 10.1039/c8an01581c

Dunn, W.B. and Ellis, D.I. (2005) 'Metabolomics: Current analytical platforms and methodologies', Tre. Anal. Chem., 24 (4), pp. 285-294, doi: 10.1016/j.trac.2004.11.021

Dunn, W.B. and Winder, C.L. (2011) 'Sample preparation related to the intracellular metabolome of yeast methods for quenching, extraction, and metabolite quantitation', Methods Enzymol., 500, pp. 277-297, doi: 10.1016/B978-0-12-385118-5.00015-3

Dunn, W.B., Bailey, N.J. and Johnson, H.E. (2005) 'Measuring the metabolome: current analytical technologies', Analyst, 130 (5), pp. 606-625, doi: 10.1039/b418288j

Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R. and Griffin, J.L. (2011) 'Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy', Chem. Soc. Rev., 40, pp. 387-426, doi: 10.1039/B906712B

Duportet, X., Aggio, R.B.M., Carneiro, S. and Villas-Boas, S.G. (2011) 'The biological interpretation of metabolomic data can be misled by the extraction method used', Metabolomics, 8 (3), pp. 410-421, doi: 10.1007/s11306-011-0324-1

Eckert, G.P., Hooff, G.P., Strandjord, D.M., Igbavboa, U., Volmer, D.A., Muller, D.A., Muller, W.E. and Wood, W.G. (2009) 'Regulation of the brain isoprenoids farnesyl- and geranylgeranylpyrophosphate is altered in male Alzheimer patients', Neurobiol Dis., 35 (2), pp. 251–257, doi: 10.1016/j.nbd.2009.05.005

Edge, T. (2003) Chapter 4 - Turbulent flow chromatography in bioanalysis, Handbook of Analytical Separations, Elsevier, pp. 91-128, doi.org/10.1016/S1567-7192(03)80005-0

Eisenberg, D. and Jucker, M. (2012) 'The amyloid state of proteins in human diseases', Cell, 148, pp. 1188–1203, doi: 10.1016/j.cell.2012.02.022

Ellis, R.J. and Minton, A.P. (2006) 'Protein aggregation in crowded environments', Biological Chemistry, 387 (5), doi: 10.1515/BC.2006.064

Englander, S.W. and Mayne, L. (2014) 'The nature of protein folding pathways', Proc Natl Acad Sc. USA, 111 (45), pp. 15873-15880, doi: 10.1073/pnas.1411798111

Faijes, M., Mars, A.E. and Smid, E.J. (2007) 'Comparison of quenching and extraction methodologies for metabolome analysis of *Lactobacillus plantarum*', Microbial Cell Fact., 27 (6), doi: :10.1186/1475-2859-6-27

Fantini J. and Yahi, N. (2010) 'Molecular insights into amyloid regulation by membrane cholesterol and sphingolipids: common mechanisms in neurodegenerative diseases', Expert Rev Mol Med., 12, doi: 10.1017/S1462399410001602

Feldmann, H. (2012) 'Yeast: Molecular and cell biology', 2nd Edition, New Jersey, Wiley-Blackwell 2nd Edition

Ferreira, P.C., Ness, F., Edwards, S.R., Cox, B.S., and Tuite, M.F. (2001) 'The elimination of the yeast [PSI+] prion by guanidine hydrochloride is the result of Hsp104 inactivation', Mol Microbiol, 40, pp. 1357–1369, doi: 10.1046/j.1365-2958.2001.02478.x

Ferscht, A.R. (2000) 'Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism', Proc Natl Acad Sc. USA, 97, pp. 1525-1529, doi: 10.1073/pnas.97.4.1525

Fiehn, O. (2002) 'Metabolomics - the link between genotypes and phenotypes', Plant Mol. Biol., 48, pp. 155-171, doi: 10.1023/A:1013713905833

Finckh, U., Muller-Thomson, T., Mann, U., Eggers, C., Marksteiner, J., Meins, W., Binetti, G., Alberici, A., Sonderegger, P., Hock, C., Nitsch, R.M. and Gal, A. (2006) 'High frequency of mutations in four different disease genes in early-onset dementia', Ann. New York Aca. Sci., 920 (1), doi: 10.1111/j.1749-6632.2000.tb06910.x

Fitzpatrick, A.W.P, Debelouchina, G.T., Bayro, M.J., Clare, D.K., Caporini, M.A., Bajaj, V.S., Jaroniec, C.P., Wang, L., Ladizhansky, V., Müller, S.A., MacPhee, C.E., Waudby, C.A., Mott, H.R., Simone, A., Knowles, T.P.J., Saibil, H.R., Vendruscolo, M., Orlova, E.V., Griffin, R.G. and Dobson, C.M. (2013) 'Atomic structure and hierarchical assembly of a cross-β amyloid fibril', Proc Natl Acad Sc. USA, 110, pp. 5468–5473, doi: 10.1073/pnas.1219476110

Gabizon, R., McKinley, M.P., Groth, D. and Pruisner, S.B. (1988) 'Immunoaffinity purification and neutralization of scrapie prion infectivity', Proc Natl Acad Sc. USA, 85, pp. 6617-6621, doi: 10.1073/pnas.85.18.6617

Gajdusek, D.C. (1977) 'Unconventional viruses and the origin and disappearance of kuru', Science, 197 (4307), pp. 943-960, doi: 10.1126/science.142303

Galasko, D. and Golde, T.E. (2013) 'Biomarkers for Alzheimer's disease in plasma, serum and blood – conceptual and practical problems', Alzheimers Res Ther, 5, pp. 10, doi: 10.1186/alzrt164

Genova, M.L., Bonacorsi, E., D'Aurelio, M., Formiggini, G., Narudo, B., Cuccomarino, S., Turi, P., Pich, M., Lenaz, G. and Bovina, C. (1999) 'Protective effect of exogenous coenzyme Q in rats subjected to partial hepatic ischemia and reperfusion', Biofactors, 9 (2-4), pp. 345-349, doi: 10.1002/biof.5520090234

Gerszten, R.E. and Wang, T.J. (2008) 'The search for new cardiovascular biomarkers', Nature, 451, pp. 949-952, doi: 10.1038/nature06802

Gibbs, C.J.Jr., Gajausek, D.C., Asher, D.M., Alpers, M.P., Beck, E., Daniel, P.M., Matthews, W.B. (1968) 'Creutzfeldt-Jakob disease (spongiform encephalopathy): Transmission to the chimpanzee', Science, 161, pp. 388-389, doi: 10.1126/science.161.3839.388

Gieger, C., Geistlinger, L., Altmaier, E., Hrabe de Angelis, M., Kronenberg, F., Meitinger, T., Mewes, H., Wichmann, H.E., Wienberger, K.M., Adamski, J., Illig, T. and Suhre, K. (2008) 'Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum', PLOS Gen., 4 (11), doi: 10.1371/journal.pgen.1000282

Glich, S., Kehler, C. and Schatzl, H. (2006) 'The prion protein requires cholesterol for cell surface localization', Mol. Cell. Neurosci., 31 (2), pp. 346-353, doi: 10.1016/j.mcn.2005.10.008

Godzien, J., Alonso-Herranz, V., Barbas, C. and Armitage, E.G. (2015) 'Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample', Metabolomics, 11 (3), pp. 518-528, doi: 10.1007/s11306-014-0712-4

Golde, T.E. and Petrucelli, L. (2009) 'What kills neurons in neurodegenerative diseases?', Mol. Neurodeger., 4, pp. 7, doi: 10.1186/1750-1326-4-7

Goldfarb, L.G., Petersen, R.B., Tabaton, M., Brown, P., LeBlanc, A.C., Montagna, P., Cortelli, P., Julien, J., Vital, C. and Pendelbury, W.W. (1992) 'Fatal familial insomnia and familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism', Science, 258 (5083), pp. 806-808, doi: 10.1126/science.1439789

Gonzalez, B., Francois, J. and Renaud, M. (1999) 'A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol', Yeast, 13 (14), doi: 10.1002/(SICI)1097-0061(199711)13:14<1347::AID-YEA176>3.0.CO;2-O

Gonzalez-Mariscal, I., Garcia-Teston, E., Padilla, S., Martin-Montalvo, A., Viciana, T.P., Vazquez-Fonseca, L., Dominguez, P.G. and Santos-Ocana, C. (2014) 'The regulation of coenzyme q biosynthesis in eukaryotic cells: all that yeast can tell us', Molecular Syndromology, 5 (3-4), pp. 107-118, doi: 10.1159/000362897

Gonzalez-Siso, M.I., Garcia-Leiro, A., Tarrio, N. and Cerdan, M.E. (2009) 'Sugar metabolism, redox balance and oxidative stress response in the respiratory yeast *Kluyveromyces lactis*', Microb Cell Fact., 46 (8), doi: 10.1186/1475-2859-8-46

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. and Kell, D.B. (2004) 'Metabolomics by numbers: acquiring and understanding global metabolite data', 22 (5), pp. 245-252, doi: 10.1016/j.tibtech.2004.03.007

Gordon, W.S., (1966) 'Variation in susceptibility of sheep to scrapie and genetic implications', Report of Scrapie Seminar, ARS 91-53. Washington, pp. 53-67

Gowda, H., Ivanisevic, J., Johnson, C.H., Kurczy, M.E., Benton, H.P., Rinehart, D., Nguyen, T., Ray, J., Kuehl, J., Arevolo, B., Westenskow, P.D., Wang, J., Arkin, A.P., Deutschbauer, A.M., Patti, G.J. and Siuzdak, G. (2014) 'Interactive XCMS online: Simplifying data processing and statistical analyses', 86 (14), pp. 6931-6939, doi: 10.1021/ac500734c

Grabowska, D., Karst, F. and Szkopinska, A. (1998) 'Effect of squalene synthase gene disruption on synthesis of polyprenols in *Saccharomyces cerevisiae*', FEBS Letters, 434, pp. 406-408, doi: 10.1016/s0014-5793(98)01019-9

Grant, C.M. (2015) 'Sup35 methionine oxidation is a trigger for *de novo* [PSI+] prion formation', Prion, 9 (4), pp. 257-265, doi: 10.1080/19336896.2015.1065372

Green, M.R. and Sambrook, J. (2012) 'Molecular cloning: A laboratory manual', Fourth Edition, Cold Spring Harbour, N.Y.: Cold Spring Harbour Laboratory Press

Greenwald, J. and Riek, R. (2010) 'Biology of amyloid: Structure, function, and regulation', Sructure, 18 (10), pp. 1244-1260, doi: 10.1016/j.str.2010.08.009

Gregorio, S.E.D. and Duennwald, M.L. (2018) 'Yeast as a model to study protein misfolding in aged cells', FEMS Yeast Res., 18 (6), doi: 10.1093/femsyr/foy054

Griffin, J.L., Nicholls, A.W., Keun, H.C., Mortishire-Smith, R.J., Nicholson, J.K. and Kuehn, T. (2002) 'Metabolic profiling of rodent biological fluids via 1 H NMR spectroscopy using a 1 mm microlitre probe', 127 (5), pp. 582-584, doi: 10.1039/b201207c

Griffith, J.S. (1967) 'Self-replication and scrapie', Nature, 215, pp. 1043-1044, doi: 10.1038/2151043a0

Grimminger, V., Richter, K., Imhof, A., Buchner, J. and Walter, S. (2004) 'The prion curing agent guanidinium chloride specifically inhibits ATP hydrolysis by Hsp104', J. Biol. Chem., 279 (9), pp. 7378-7383, doi: 10.1074/jbc.M312403200

Haass, C. and Selkoe, D. J. (2007) 'Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide', Nature Rev. Mol. Cell. Biol., 8, pp.101–112, doi: 10.1038/nrm2101

Halfmann, R., Alberti, S. and Lindquist, S. (2010) 'Prions, protein homeostasis, and phenotypic diversity', Trends in Cell Bio., 20 (3), pp. 125-133, doi: 10.1016/j.tcb.2009.12.003

Halfmann, R., Jarosz, D.F., Jones, S.K., Chang, A., Lancaster, A.K. and Lindquist, S. (2012) 'Prions are a common mechanism for phenotypic inheritance in wild yeasts', Nature, 482, pp. 363-367, doi: 10.1038/nature10875

Han, S., Lone, M.A., Schneiter, R. and Chang, A. (2010) 'Orm1 and Orm2 are conserved endoplasmic reticulum membrane proteins regulating lipid homeostasis and protein quality control', Proc Natl Acad Sc. USA, 107 (13), pp. 5851-5856, doi: 10.1073/pnas.0911617107

Han, X., Rozen, S., Boyle, S.H., Hellegers, C., Cheng, H., Burke, J.R., Welsh-Bohmer, K.A., Doriaswamy, P.M. and Kaddurah-Daouk, R. (2011) 'Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics', PLoS One, 6 (7), pp. e21643, doi: 10.1371/journal.pone.0021643

Hardy, J. and Selkoe, D. J. (2002) 'Medicine - the amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics', Science, 297, pp.353–356, doi: 10.1126/science.1072994

Hartl, F.U., Bracher, A. and Hayer-Hartl, M (2011) 'Molecular chaperones in protein folding and proteostasis', Nature, 475 (7356), pp. 324-332, doi: 10.1038/nature10317

He, X., Ji, G., Jai, W. and Li, H. (2015) 'Gut microbiota and nonalcoholic fatty liver disease: insights on mechanism and application of metabolomics', Int. J. Mol. Sci., 17 (300), doi: doi:10.3390/ijms17030300

Hebert, D.N. and Molinari, M. (2007) 'In and out of the ER: protein folding, quality control, degradation, and related human diseases', Physiol Rev., 87 (4), pp. 1377-1408, doi: 10.1152/physrev.00050.2006

Hendricks, M.M.W.B., van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefslott, H.C.J. and Smilde, A.K. (2011) 'Data-processing strategies for metabolomics studies', Trends in Analy. Chem., 30 (10), pp. 1685-1698, doi: 10.1016/j.trac.2011.04.019

Henry, R.A. (2014) Impact of Particle Size Distribution on HPLC Column Performance, LCGC Special Issues, 32 (4), pp. 12-19,

Hilaly, Y.K., Biasetti, L., Blakeman, B.J., Pollack, S.J., Zibaee, S., Abdul-Sada, A., Thorpe, J.R., Xue, W.F. and Serpell, L.C. (2016) 'The involvement of dityrosine crosslinking in α-synuclein assembly and deposition in Lewy Bodies in Parkinson's disease', Scientific Reports, 6, doi: 10.1038/srep39171

HMDB (2019) available at http://www.hmdb.ca/, visited on 1st of June 2019

Hong, J.Y., Mathur, V. and Liebman, S.W. (2011) 'A new color assay for [URE3] prion in a genetic background used to score for the [PSI+] prion', 28 (7), pp. 555-560, doi: 10.1002/yea.1857

Hovland, P., Flick, J., Johnston, M. and Sclafani, R.A. (1989) 'Galactose as a gratuitous inducer of GAL gene expression in yeasts growing on glucose', Gene, 83 (1), pp. 57-64, doi: 10.1016/0378-1119(89)90403-4

Hsiao, K.K., Baker, H.F., Crow, T.J. (1989) 'Linkage of a prion protein missense variant to Gerstmann–Straussler syndrome', Nature, 338, pp. 342-345, doi: 10.1038/338342a0

Hsiao, K.K., Scott, M., Foster, D., Growth, D.F., DeArmond, S.J. and Prusiner, S.B. (1990) 'Spontaneous neurodegeneration in transgenic mice with mutant prion protein', Science, 250 (4987), pp. 1587-1590, doi: 10.1126/science.1980379

Hu, J., Wei, M., Mirzaei, H., Madia, F., Amparo, M., Chagoury, S., Kennedy, B. and Longo, V.D. (2014) 'Tor-Sch9 deficiency activates catabolism of the ketone body-like acetic acid to promote trehalose accumulation and longevity', Aging Cell, 13 (3), pp. 457-467, doi: 10.1111/acel.12202

Huang, V.J., Stein, K.C. and True, H.L. (2013) 'Spontaneous variants of the [RNQ+] prion in yeast demonstrate the extensive conformational diversity possible with prion proteins', PLoS ONE, 8 (10), doi: 10.1371/journal.pone.0079582

Huang, Z., Shao, W., Gu, J., Hu, X., Shi, Y., Xu, W., Huang, C. and Lin, D. (2015) 'Effects of culture media on metabolic profiling of the human gastric cancer cell line SGC7901', Mol Biosyst., 11 (7), pp. 1832-1840, doi: 10.1039/c5mb00019j.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. and O'Shea, E.K. (2003) 'Global analysis of protein localization in budding yeast', Nature, 425, pp. 686-691, doi: 10.1038/nature02026

Itakura, A.K., Chakravarty, A.K., Jakobson, C.M. and Jarosz, D.F. (2019) 'Widespread prion-based control of growth and differentiation strategies in *Saccharomyces cerevisiae*', Mol. Cell., 77 (2), pp. 266-278, doi: 10.1016/j.molcel.2019.10.027

Jackson, M.P. and Hewitt, E.W. (2017) ‘Why are functional amyloids non-toxic in humans?’, Biomolecules, 7 (4), doi: 10.3390/biom7040071

Jackson, W.S., Borkowski, A.W., Faas, H., Steele, A.D., King, O.D., Watson, N., Jasanoss, A. and Lindquist, S. (2009) ‘Spontaneous generation of prion infectivity in fatal familial insomnia knockin mice’, Neuron, 63 (4), pp. 438-450, doi: 10.1016/j.neuron.2009.07.026

Jander, G., Norris, S.R., Joshi, V., Fraga, M., Rugg, A., Yu, S., Li, L. and Last, R.L. (2004) ‘Application of a high-throughput HPLC-MS/MS assay to *Arabidopsis* mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality’, The Plant Journal, 39 (3), doi: 10.1111/j.1365-313X.2004.02140.x

Jandera, P. (2005) ‘liquid chromatography - normal phase. encyclopedia of analytical science’, 142–152. doi: 10.1016/b0-12-369397-7/00324-1

Jarosz, D.F., Brown, J.C.S., Walker, G.A., Datta, M.S., Ung, W.L., Lancaster, A.K., Rotem, A., Chang, A., Newby, G.A., Weitz, D.A., Bisson, L.F. and Lindquist, S. (2014b) ‘Cross-kingdom chemical communication drives a heritable, mutually beneficial prion-based transformation of metabolism’, Cell, 158 (5), pp. 1083-1093, doi: 10.1016/j.cell.2014.07.025

Jarosz, D.F., Lancaster, A.K., Brown, J.C.S. and Lindquist, S. (2014a) ‘An evolutionarily conserved prion-like element converts wild fungi from metabolic specialists to generalists’, Cell, 158 (5), pp. 1072-1082, doi: 10.1016/j.cell.2014.07.024

Jaunmuktane, Z., Mead, S., Ellis, M., Wadsworth, J.D.F., Nicoll, A.J., Kenny, J., Launchbury, F., Linehan, J., Richard-Leondt, A., Walker, A.S., Rudge, P., Collinge, J. and Brandner, S. (2015) ‘Evidence for human transmission of amyloid-β pathology and cerebral amyloid angiopathy’, Nature, 525, pp. 247-250, doi: 10.1038/nature15369

Jenkins, G.M. (2003) ‘The emerging role for sphingolipids in the eukaryotic heat shock response’, Cell Mol Life Sci., 6 (4), pp. 701-710, doi: 10.1007/s00018-003-2239-0

Jeong, K., Kim, S. and Bandeira, N. (2012) ‘False discovery rates in spectral identification’, BMC Bioinformatics, 13, doi: 10.1186/1471-2105-13-S16-S2

Jiang, Y., Zhu, Z., Shi, J., An., Y., Zhang, K., Wang, Y., Li, S., Jin, L., Ye, W., Cui, M. and Chen, X. (2019) ‘Metabolomics in the development and progression of dementia: A systematic review’, Front. Neurosci., doi: 10.3389/fnins.2019.00343

Jung, G.M., Jones, G., and Masison, D.C. (2002) ‘Amino acid residue 184 of yeast Hsp104 chaperone is critical for prion-curing by guanidine, prion propagation, and thermotolerance’, Proc Natl Acad Sci USA, 99, pp. 9936–9941, doi: 10.1073/pnas.152333299.

Kadnar, M.L., Articov, G. and Derkatch, I.L. (2010) ‘Distinct type of transmission barrier revealed by study of multiple prion determinants of Rnq1’, PLoS Gen., 6 (1), doi: 10.1371/journal.pgen.1000824

Kalastavadi, T. and True, H.L. (2010) ‘Analysis of the [RNQ+] prion reveals stability of amyloid fibers as the key determinant of yeast prion variant propagation’, J. Biol. Chem., 285, pp. 20748-20755, doi: 10.1074/jbc.M110.115303

Kambhampati, S, Li, J., Evans, B.S. and Allen, D.K. (2019) ‘Accurate and efficient amino acid analysis for protein quantification using hydrophilic interaction chromatography coupled tandem mass spectrometry’, Plant Methods, 15, doi: 10.1186/s13007-019-0430-z

Kanehisa, M. and Goto, S. (2000) ‘KEGG: kyoto encyclopedia of genes and genomes’, Nucleic Acids Res, 28 (1), pp. 27-30, doi: 10.1093/nar/28.1.27

Kang, Y., Park, Y.S., Park, Y.C., Yoon, S.M., JongAhn, H., Kim, G. and Kwon, S.W. (2012) ‘UPLC/Q-TOF MS based metabolomics approach to post-mortem-interval discrimination: mass spectrometry based metabolomics approach’, J. Pharma Inv., 42 (1), pp. 41-46, doi: 10.1007/s40005-012-0006-7

Kapoore, R.V., Coyle, R., Staton, C.A., Brown, N.J. and Vaidyanathan, S. (2017) ‘Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-MB-231’, Analyst, 142, pp. 2038-2049, doi: 10.1039/C7AN00207F

Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S.M. and Subhraveti, P. (2017) ‘The BioCyc collection of microbial genomes and metabolic pathways’, Briefings in Bioinformatics, 20 (4), pp. 1085-1093, doi: 10.1093/bib/bbx085

Karran, E., Mercken, M. and Strooper, B. (2011) ‘The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics’, Nature Rev. Drug Disc., 10, pp. 698-712, doi: 10.1038/nrd3505

Kelly, A.C., Busby, B. and Wickner, R.B. (2012a) ‘Effect of domestication on the spread of the [PIN+] prion in *Saccharomyces cerevisiae*’, Genetics, 197 (3), pp. 1007-1024, doi: 10.1073/pnas.1213449109

Kelly, A.C., Shewmaker, F.P., Kryndushkin, D. and Wickner, R.B. (2012b) ‘Sex, prions, and plasmids in yeast’, Proc Natl Acad Sc. USA, 109 (40), pp. E2683-E2690, doi: 10.1073/pnas.1604478113

Keun, H.C., Beckonert, O., Griffin, J.L., Richter, C., Moskau, D., Lindon, J.C. and Nicholson, J.K. (2002) ‘Cryogenic probe 13C NMR spectroscopy of urine for metabonomic studies’, J. Anal. Chem., 74 (17), pp. 4588-4593, doi: 10.1021/ac025691r

Kim, Y.E., Hipp, M.S., Hayer-Hartl, A.B.M. and Hartl, F.U. (2013) ‘Molecular chaperone functions in protein folding and proteostasis’, Annual Rev. Biochem., 82, pp. 323-355, doi: 10.1146/annurev-biochem-060208-092442

Klatzo, I. (1959) ‘Pathology of Kuru’, Lab Inves., 8 (4), pp. 799-847

Knowles, T.P.J., Vendruscolo, M. and Dobson, C.M. (2014) ‘The amyloid state and its association with protein misfolding diseases’, Nature Rev. Mol. Cell Biol., 15, pp. 384-396, doi: 10.1038/nrm3810

Konermann, L., Ahadi, E., Rodriguez, A.D. and Vahadi, S. (2013) Unraveling the Mechanism of Electrospray Ionization, Anal. Chem., 85(1), pp. 2-9, doi.org/10.1021/ac302789c

Koselny, K., Mutlu, N., Minard, A.Y., Kumar, A., Krysan, D.J. and Wellington, M. (2018) ‘A genome-wide screen of deletion mutants in the filamentous *Saccharomyces cerevisiae* background identifies ergosterol as a direct trigger of macrophage pyroptosis’, American Society for Microbiology, doi: 10.1128/mBio.01204-18.

Kowalski, D., Pendyala, L., Daignan-Fornier, B., Howell, S.B. and Huang, R.Y. (2008) ‘Disregulation of purine nucleotide biosynthesis pathways modulates cisplatin cytotoxicity in *Saccharomyces cerevisiae*’, Mol Pharmacol., 74 (4), pp. 1092-1100, doi:10.1124/mol.108.048256

Krobitsch, S. and Lindquist, S. (2000) ‘Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins’, Proc Natl Acad Sc. USA, 97 (4), pp. 1589-1594, doi: 10.1073/pnas.97.4.1589

Kryndushkin, D.S., Alexandrov, I.M., Ter-Avanesyan, M.D. and Kushnirov, V.V. (2003) ‘Yeast [*PSI*$^{+}$] prion aggregates are formed by small sup35 polymers fragmented by Hsp104’, J. Biol. Chem., 278, pp. 49636-49643, doi: 10.1074/jbc.M307996200

Kujawska, M. and Jodynis-Leibert, J. (2018) ‘What is the evidence that parkinson's disease is a prion disorder, which originates in the gut?’, Int J Mol Sci, 19 (11), doi: 10.3390/ijms19113573

Laurens, B., Constantinescu, R., Freeman, R., Gerhard, A., Jellinger, K., Jeromin, A., Krismer, F., Mollenhauer, B., Schlossmacher, M.G., Shaw, L.M., Verbeek, M.M., Wenning, G.K., Winge, K., Zhang, J. and Meissner, W.G. (2015) ‘Fluid biomarkers in multiple system atrophy: A review of the MSA Biomarker Initiative’, Neurobiology of Disease, 80, pp. 29–41, doi: 10.1016/j.nbd.2015.05.004

Leopoldt, J.G. (1750) ‘Nützliche und auf die erfahrung gegründete einleitung zu der land-wirthschafft’, Volume 5, Johann Gottlieb Rothen, Sorau

Levinthal, C. (1968) ‘Are there pathways for protein folding?’, J. Chim. Phys., 65 (1), pp. 44

Lewis, G.D., Asnani, A. and Gerzten, R.E. (2008) ‘Application of metabolomics to cardiovascular biomarker and pathway discovery’, J. Amer. Col. Cardiol., 2, pp. 117-123, doi: 10.1016/j.jacc.2008.03.043

Lewitt, P.E., Li, J., Lu, M., Beach, T.G., Adler, C.H. and Guo, L. (2013) ‘3-hydroxykynurenine and other Parkinson's disease biomarkers discovered by metabolomic analysis’, Movement Disorders, 28 (12), pp. 1653–1660, doi: 10.1002/mds.25555

Li, S., Park, Y., Duraisingham, S., Strobel, F.H., Khan, N., Soltow, Q.A., Jones, D.P. and Pulendran, B. (2013) ‘Predicting network activity from high throughput metabolomics’, PLOS ONE, 9 (7), doi: 10.1371/journal.pcbi.1003123

Li, Z., Li, Y., Tang, Y.J. and Shui, W. (2019) ‘Exploiting high-resolution mass spectrometry for targeted metabolite quantification and 13c-labeling metabolism analysis’, Meth. Mol. Biol., doi: 10.1007/978-1-4939-8757-3_9

Liang, Q., Lui, H., Li, X. and Zhang, A. (2016) ‘High-throughput metabolomics analysis discovers salivary biomarkers for predicting mild cognitive impairment and Alzheimer's disease’, RSC Advances, 79 (6), pp. 75499-75504, doi: 10.1039/C6RA16802G

Liebman, S.W. and Chernoff, Y.O. (2012) ‘Prions in yeast’, Genetics, 191 (4), pp. 1041-1072, doi: 10.1534/genetics.111.137760

Liesenfeld, D.B., Habermann, N., Owen, R.W., Scalbert, A. and Ulrich, C.M. (2013) 'Review of mass spectrometry–based metabolomics in cancer research', Canc. Epid. Biom. Prev., 22 (12), doi: 10.1158/1055-9965.EPI-13-0584

Lima-Filho, R.A.S. and Oliveira, M.M. (2018) 'A role for cellular prion protein in late-onset alzheimer's disease: evidence from preclinical studies', J Neurosci., 38 (9), pp. 2146-2148, doi: 10.1523/JNEUROSCI.3307-17.2018

Linden, R. (2017) 'The biological function of the prion protein: a cell surface scaffold of signaling modules', Front. Mol. Neurosci., 10, doi: 10.3389/fnmol.2017.00077

Lindenbaum, S. (2008) 'Understanding kuru: the contribution of anthropology and medicine', Phil. Trans. R. Soc. B., 363, pp. 3715-3720, doi: 10.1098/rstb.2008.0072

Lindquist, S., Krobitsch, S., Liming, L. and Sondheimer, N. (2001) 'Investigating protein conformation-based inheritance and disease in yeast', Philos. Trans. R. Soc. Lond. B. Biol. Sci., 356 (1406), pp.169-176, doi: 10.1098/rstb.2000.0762

Lindquist, S.L. and Kelly, J.W. (2011) 'Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases–progress and prognosis', Cold Spring Harb. Perspect. Bio., doi: 10.1101/cshperspect.a004507

Linsley, G. (1951) 'Excess and restraint. Social control among a New Guinea mountain people', Patrol Report No 8 of 1950/51, Cited in Berndt RM, p. 279. Chicago, IL: University of Chicago Press.

Lobo, V., Patil, A., Phatak, A. and Chandra, N. (2010) 'Free radicals, antioxidants and functional foods: Impact on human health', Pharmacogn Rev., 4 (8), pp. 118-126, doi: 10.4103/0973-7847.70902

Lui, G., Chen, Y., Faergeman, N.J. and Nielsen, J. (2017) 'Elimination of the last reactions in ergosterol biosynthesis alters the resistance of *Saccharomyces cerevisiae* to multiple stresses', FEMS Yeast Research, 17 (6), doi: 10.1093/femsyr/fox063

Magherini, F., Carpentieri, A., Amoresano, A., Gamberi, T., De Filippo, C., Rizzetto, L., Biagini, M., Pucci, P. and Modesti, A. (2009) 'Different carbon sources affect lifespan and protein redox state during *Saccharomyces cerevisiae* chronological ageing', Cell Mol Life Sci, 66 (5), pp. 933-947, doi: 10.1007/s00018-009-8574-z

Makhatadze, G.I. (2017) 'Linking computation and experiments to study the role of charge–charge interactions in protein folding and stability', Physical Biology, 14 (1), doi:10.1088/1478-3975/14/1/013002

Mani, K., Cheng, F., Havsmark, B., Jonsson, M., Belting, M. and Frannson, L.A. (2003) 'Prion, amyloid beta-derived Cu(II) ions, or free Zn(II) ions support S-nitroso-dependent autocleavage of glypican-1 heparan sulfate', J Biol. Chem., 278 (40), pp. 38956-38965, doi: 10.1074/jbc.M300394200

Mascorro-Gallardo, J.O., Covarrubias, A.A. and Gaxiola, R. (1996) 'Construction of a CUP1 promoter-based vector to modulate gene expression in *Saccharomyces cerevisiae*', Gene, 172 (1), pp. 169-170, doi: 10.1016/0378-1119(96)00059-5

Mashego, M.R., van Gulik, W.M., Vinke, J.L. and Heijnen, J.J. (2003) 'Critical evaluation of sampling techniques for residual glucose determination in carbon-limited chemostat culture of *Saccharomyces cerevisiae*', Biotech and Bioeng., 83 (4), doi: 10.1002/bit.10683

Masters, C.L., Gajdusek, D.C. and Gibbs, C.J. (1981) 'Creutzfeldt-Jakob disease virus isolations from the Germstmann-Straussler syndrome with an analysis of the various forms of amyloid plaque deposition in the virus induced spongiform encephalopathies', Brain, 104, pp. 559-588, doi: 10.1093/brain/104.3.559

Mastrianni, J.A. (2003) 'Genetic prion diseases', Gen. Revs., University of Washington, Seattle

Mathew, A.K. and Padmanaban, V.C. (2013) 'Metabolomics: the apogee of the omics trilogy', Int. J Pharma. Sci., 5, pp. 45-48

Meganathan, R. (2001) 'Ubiquinone biosynthesis in microorganisms', FEMS Microbiology Letters, 203 (2), pp. 131-139, doi: 10.1111/j.1574-6968.2001.tb10831.x

Meriin, A.B., Zhang, X., He, X., Newnam, G.P., Chernoff, Y.O. and Sherman, M.Y. (2002) 'Huntington toxicity in yeast model depends on polyglutamine aggregation mediated by a prion-like protein Rnq1', J Cell Biol., 157 (6), pp. 997-1004, doi:

Miksovska, J., Schiffer, M., Hanson, D. and Sebban, K. (1999) 'Proton uptake by bacterial reaction centers: The protein complex responds in a similar manner to the reduction of either quinone acceptor', Proc Natl Acad Sci USA, 96, pp. 14348-14353, doi: 10.1073/pnas.96.25.14348

Milo, R., Jorgensen, P., Moran, U., Weber, G. and Springer, M. (2010) 'BioNumbers—the database of key numbers in molecular and cell biology', Nucleic Acids Res, 38, pp. D750-D753, doi: 10.1093/nar/gkp889

Misra, B.B. (2018) 'New tools and resources in metabolomics: 2016–2017', Electrophoresis, 39 (7), doi: 10.1002/elps.201700441

Mitchell, T.N. and Cortisella,B. (2004) 'NMR – from spectra to structures; an experimental approach', Springer-Verlag, Berlin, Heidelberg

Mohamed, A., Smith, K., and Posse de Chaves, E. (2015) 'The mevalonate pathway in alzheimer's disease — cholesterol and non-sterol isoprenoids', Alzheimer's disease - challenges for the future, doi: 10.5772/59904

Moldoveanu, S.C. and David, V. (2013) Chapter 5 - Retention Mechanisms in Different HPLC Types, Essentials in Modern HPLC Separations, Elsevier, pp. 145-190, doi.org/10.1016/B978-0-12-385013-3.00005-7

Monge, M.E., Dobbs, J.N., Baker, E.S., Edison, A.S. and Fernandez, F.M. (2019) 'Challenges in Identifying the Dark Molecules of Life', Ann. Rev. Analy. Chem., 12, pp. 177-199, doi: 10.1146/annurev-anchem-061318-114959

Muschet, C., Moller, G., Prehn, C., de Angelis, M.H., Adamski, J. and Tokarz, J. (2016) 'Removing the bottlenecks of cell culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell harvesting method', Metabolomics, 12 (10), pp. 151, doi: 10.1007/s11306-016-1104-8

Mushtaq, M.Y., Choi, Y.H., Verpoorte, R. and Wilson, E.G. (2014) 'Extraction for metabolomics: access to the metabolome', Phyto Analy., 25 (4), doi: 10.1002/pca.2505

Myers, O.D., Sumner, S.J., Li, S., Barnes, S. and Du, X. (2017) 'Detailed investigation and comparison of the xcms and mzmine 2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data', Anal. Chem., 89 (17), pp. 8689-8695, doi: 10.1021/acs.analchem.7b01069

Nakamura, T. and Lipton, S.A. (2017) ''SNO'-Storms compromise protein activity and mitochondrial metabolism in neurodegenerative disorders', Trends in Endocrinology & Metabolism, 28 (12), pp. 879-892, doi: 10.1016/j.tem.2017.10.004

Nakayashiki, T., Kurtzman, C.P., Edskes, H.K. and Wickner, R.B. (2005) 'Yeast prions [URE3] and [PSI+] are diseases', Proc Natl Acad Sc. USA, 102 (30), pp. 10575-10580, doi: 10.1073/pnas.0504882102

Narayan, P., Ehsani, S. and Lindquist, S. (2014) 'Combating neurodegenerative disease with chemical probes and model systems', Nat. Chem. Biol., 10, pp. 911-920, doi: 10.1038/nchembio.1663

Naslavsky, N., Shmeeda, H., Friedlander, G., Yanai, A., Futerman, A.H., Barenholz, Y. and Taraboulos, A. (1999) 'Sphingolipid depletion increases formation of the scrapie prion protein in neuroblastoma cells infected with prions', Journal of Biochem, 274 (30), pp. 20763-20771, doi: 10.1074/jbc.274.30.20763

Naz, S., Vallejo, M., Garcia, A. and Barbas, C. (2014) 'Method validation strategies involved in non-targeted metabolomics', J. Chromato. A, 1353, pp. 99-105, doi: 10.1016/j.chroma.2014.04.071

Nazor, K.E., Kuhn, F., Seward, T., Green, M., Zwald, D., Purro, M., Schmid, J., Biffiger, K., Power, A.M., Oesch, B., Raeber, A.J. and Telling, G.C. (2005) 'Immunodetection of disease-associated mutant PrP, which accelerates disease in GSS transgenic mice', EMBO J., 24, pp. 2472-2480, doi: 10.1038/sj.emboj.7600717

Nedelsky, N.B., Todd, P.K. and Taylor, J.P (2008) 'Autophagy and the ubiquitin-proteasome system: collaborators in neuroprotection', Biochim. Biophys. Acta., 1782 (12), pp. 691-699, doi: 10.1016/j.bbadis.2008.10.002

Nelson, P. and Masel, J. (2018) 'Evolutionary capacitance emerges spontaneously during adaptation to environmental changes', Cell Reports, 25 (1), pp. 249-258, doi: 10.1016/j.celrep.2018.09.008

Ness, F., Cox, B.S., Wongwigkarn, J., Naeimi, W.R. and Tuite, M.F. (2017) 'Over-expression of the molecular chaperone Hsp104 in *Saccharomyces cerevisiae* results in the malpartition of [PSI+] propagons', Mol. MicroBiol., 104 (1), pp. 125-143, doi: 10.1111/mmi.13617

Ni, Y., Su, M., Qui, Y., Jia, W., and Du, X. (2016) 'ADAP-GC 3.0: Improved peak detection and deconvolution of co-eluting metabolites from GC/TOF-MS Data for metabolomics studies', Anal. Chem., 88 (17), pp. 8802-8811, doi: 10.1021/acs.analchem.6b02222

Nielsen, J. and Oliver, S. (2005) 'The next wave in metabolome analysis', Trends Biotechnol., 23 (11), pp. 544-546, doi: 10.1016/j.tibtech.2005.08.005

Nikitas, P., Pappa-Louisi, A., and Agrafiotou, P. (2002). Effect of the organic modifier concentration on the retention in reversed-phase liquid chromatography. Journal of Chromatography A, 946(1-2), pp. 33–45. doi:10.1016/s0021-9673(01)01536-9

Noble Prize (2018) available at https://www.nobelprize.org/prizes/medicine/1997/prusiner/lecture/, visited on 1st September 2018

Nygaard, P., Duckert, P. and Saxild, H.H. (1996) 'Role of adenine deaminase in purine salvage and nitrogen metabolism and characterization of the ade gene in *Bacillus subtilis*', Journal of Bacteriology, 178 (3), pp. 846-853, doi: 10.1128/jb.178.3.846-853.1996

Oates, M.E., Romero, P., Ishisa, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztanyi, Z., Uversky, V.N., Obradovic, Z., Kurgan, L., Dunker, A.K. and Gough, J. (2013) 'D2P2: database of disordered protein predictions', Nucleic Acids Res., 41, pp. 508-516, doi: 10.1093/nar/gks1226

Oesch, B., Westaway, D., Walchli, M., McKinley, M.P., Kent, S.B.H., Aebersold, R., Barry, R.A., Tempst, P., Teplow, D.B., Hood, L.E., Prusiner, S.B. and Weissmann, C. (1985) 'A cellular gene encodes scrapie PrP 27-30 protein', Cell, 40 (4), pp. 735-746, doi: 10.1016/0092-8674(85)90333-2

Oldiges, M. and Takors, R. (2005) 'Applying metabolic profiling techniques for stimulus-response experiments: Chances and pitfalls', Tech. Trans. Biotech., 92, pp. 173-196, doi: 10.1007/b98913

Oleander, A., Hall, D.R., Bray, D.P. and Burman, J.P.J. (2019) 'Identification of female sex pheromone for monitoring the barred tooth striped moth, Trichopteryx polycommata, a priority conservation species', J Chem Ecol, 45 (8), pp. 649-656, doi: 10.1007/s10886-019-01093-1

Oliver, S.G., Winson, M.K., Kell, D.B. and Baganza, F. (1998) 'Systematic functional analysis of the yeast genome', Tre. Biotech., 16 (9), pp. 373-378, doi: 10.1016/s0167-7799(98)01214-1

Olshansky, S.J., Passaro, D.J., Hershow, R.C., Layden, J., Carnes, B.A., Brody, J., Hayflick, L., Butler, R.N., Allison, D.B. and Ludwig, D.S. (2005) 'A potential decline in life expectancy in the united states in the 21st century', N. Engl. J. Med., 352, pp. 1138-1145, doi: 10.1056/NEJMsr043743

Olson, R.E. and Rudney, H. (1983) 'Biosynthesis of ubiquinone', Vitamins and Hormones, 40, pp. 1–43, doi: 10.1016/S0083-6729(08)60431-8

Oresic, M., Hyotylainen, T., Herukka, S.K., Sysi-Aho, M., Mattila, I., Seppanan-Laakso, T., Julkunen, V., Gopalacharyulu, P.V., Hallikainen, M., Koikkalainen, J., Kivipelto, M., Helisalmi, S., Lotjonen, J. and Soininen, H. (2011) 'Metabolome in progression to Alzheimer's disease', Trans. Psych., 1, pp. e57, doi: 10.1038/tp.2011.55

Oscherovich, L.Z. and Weissman, J.S. (2001) 'Multiple Gln/Asn-Rich prion domains confer susceptibility to induction of the yeast [*PSI*$^{+}$] prion', Cell, 106 (2), pp. 183-194, doi: 10.1016/s0092-8674(01)00440-8

Ostergaard, S., Olsson, L., Johnston, M. and Nielsen, J. (2000) 'Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network', Nat Biotechnol, 18, pp. 1283-1286, doi: 10.1038/82400

Paglia, G., Fabiola Del Greco, M., Sigurdsson, B.B., Rainer, J., Volani, C., Hicks, A.A., Pramstaller, P.P. and Smarason, S.V. (2018) 'Influence of collection tubes during quantitative targeted metabolomics studies in human blood samples', Clinica Chimica Acta., 486, pp. 320-328, doi: 10.1016/j.cca.2018.08.014

Paglia, G., Magnúsdóttir, M., Thorlacius, S., Sigurjónsson, Ó.E., Guðmundsson, S., Palsson, B.Ø. and Thiele, I. (2012) 'Intracellular metabolite profiling of platelets: Evaluation of extraction processes and chromatographic strategies', Journal of Chromatography B, 898, pp. 111–120, doi: 10.1016/j.jchromb.2012.04.026

Paglia, G., Stocchero, M., Cacciatore, S., Lai, S., Angel, P., Alam, M.T., Keller, M., Ralser, M. and Astarita, G. (2016) 'Unbiased metabolomic investigation of Alzheimer's disease brain points to dysregulation of mitochondrial aspartate metabolism', J Proteome Res, 15 (2), pp. 608–618, doi: 10.1021/acs.jproteome.5b01020

Pan, X., Elliott, C.T., McGuinness, B., Passmore, P., Kehoe, P.G., Holscher, C., McClean P.L., Graham, S.F. and Green, B.D. (2017) 'Metabolomic profiling of bile acids in clinical and experimental samples of Alzheimer's disease', Metabolites, 7 (2), pp. 28–39, doi: 10.3390/metabo7020028

Park, Y.N., Morales, D., Rubinson, E.H., Masison, D., Eisenberg, E., and Greene, L.E. (2012) 'Differences in the curing of [PSI+] prion by various methods of Hsp104 inactivation', PLoS ONE, 7, pp. e37692, doi: 10.1371/journal.pone.0037692

Parrou, J.L., Teste, M.A. and Francois, J. (1997) 'Effects of various types of stress on the metabolism of reserve carbohydrates in *Saccharomyces cerevisiae*: genetic evidence for a stress-induced recycling of glycogen and trehalose', Microbiology, 143 (6), pp. 1891-1900, doi: 10.1099/00221287-143-6-1891

Patel, B.K. and Liebman, S.W. (2007) '"Prion-proof" for [PIN+]: Infection with *in vitro*-made amyloid aggregates of Rnq1 protein-(132–405) induces [*PIN*$^{+}$]', J. Mol. Biol., 365 (3), pp. 773-782, doi: 10.1016/j.jmb.2006.10.069

Patti, G.J., Yanes, O. and Siuzdak, G. (2012) 'Metabolomics: the apogee of the omics trilogy', Nat. Rev., 13, pp. 263-269, doi: 10.1038/nrm3314

Paushkin, S.V., Kushnirov, V.V., Smirnov, V.N. and Ter-Avanesyan, M.D. (1996) 'Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor', EMBO J, 15 (12), pp. 3127-3134, PMC450255

Pena-Bautista, C., Roca, M., Hervas, D., Cuevas, A., Lopez-Cuevas, R., Vento, M., Baquero, M., Garcia-Blanco, A. and Chafer-Pericas, C. (2019) 'Plasma metabolomics in early Alzheimer's disease patients diagnosed with amyloid biomarker', J. Proteomics, 200, pp. 144-152, doi: 10.1016/j.jprot.2019.04.008

Perera, R.M.M., Mariott, P.J. and Galbally, I.E. (2002) 'Headspace solid-phase microextraction - Comprehensive two-dimensional gas chromatography of wound induced plant volatile organic compound emissions', 127 (12), pp. 1601-1607, doi: 10.1039/b208577a

Perutz, M.F., Pope, B.J., Owen, D., Wanker, E.E. and Scherzinger, E. (2002) 'Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β-peptide of amyloid plaques', Proc Natl Acad Sc. USA, 99 (8), pp. 5596-5600, doi: 10.1073/pnas.042681599

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F. and Tycko, R. (2002) 'A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR', Proc Natl Acad Sc. USA, 99, pp. 16742–16747, doi: 10.1073/pnas.262663499

Phelan, M.M., Caamano-Gutierrez, E., Gant, M.S., Grosman, R.X. and Madine, J. (2017) 'Using an NMR metabolomics approach to investigate the pathogenicity of amyloid-beta and alpha-synuclein', Metabolomics, 13 (151), doi: 10.1007/s11306-017-1289-5

Pina, F., Yagisawa, F., Obara, K., Gregerson, J.D., Kihara, A. and Niwa, M. (2018) 'Sphingolipids activate the endoplasmic reticulum stress surveillance pathway', The Journal of Cell Biology, 217 (2), pp. 495–505, doi: 10.1083/jcb.201708068

Pinu, F.R. and Villas-Boas, S.G. (2017) 'Extracellular microbial metabolomics: The state of the art', Metabolites, 7 (3), doi: 10.3390/metabo7030043

Pinu, F.R., Beale, D.J., Paten, A.M., Kouremenos, K., Swarup, S., Schirra, H.J. and Wishart, D. (2019) 'Systems biology and multi-omics integration: viewpoints from the metabolomics research community', Metabolites, 9 (4), pp. 76, doi: 10.3390/metabo9040076

Pinu, F.R., Villas-Boas, S.G. and Aggio, R. (2017) 'Analysis of intracellular metabolites from microorganisms: quenching and extraction protocols', Metabolites, 7 (4), doi: 10.3390/metabo7040053

Plumb, R., Castro-Perez, J., Granger, J., Beattie, I., Joncour, K. and Wright, A. (2004) 'Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry', Rapid Commun Mass Spectrom, 18 (19), pp. 2331-2337, doi: 10.1002/rcm.1627

Pluskal, T., Castillo, S., Villar-Briones, A. and Oresic, M. (2010) 'Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data', BMC Bioinformatics, 11, doi:10.1186/1471-2105-11-395

Prasad Maharjan, R. and Ferenci, T. (2003) 'Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*', Anal Bioc, 313 (1), pp. 145-154, doi: 10.1016/s0003-2697(02)00536-5

Prosser, G.A., Larrouy-Maumus, G. and Carvelho, L.P. (2014) 'Metabolomic strategies for the identification of new enzyme functions and metabolic pathways', EMBO Rev., 15 (6), pp. 657-669, doi: 10.15252/embr.201338283

Pruisner, S.B. (1982) 'Novel proteinaceous infectious particles cause scrapie', Science, 216, pp. 136-144, doi: 10.1126/science.6801762

Pruisner, S.B. (1998) 'Prions', Proc Natl Acad Sc. USA, 95, pp. 13363-13383, doi: 10.1073/pnas.95.23.13363

Prusiner, S.B., Woermana, A.L., Mordes, D.A., Watts, J.C., Rampersaud, R., Berry, D.B., Patel, S., Oehler, A., Lowe, J.K., Kravitz, S.N., Geschwind, D.H., Glidden, D.V., Halliday, G.M., Middleton, L.T., Gentleman, S.M., Grinberg, L.T. and Giles, K. (2015) 'Evidence for α-synuclein prions causing multiple system atrophy in humans with parkinsonism', Proc Natl Acad Sc. USA, 112 (38), doi: 10.1073/pnas.1514475112

Pushie, M.J., Shaykhutdinov, R., Nazyrova, A., Graham, C. and Vogel, H.J. (2011) 'An NMR metabolomics study of elk inoculated with chronic wasting disease', Journal of Toxicology and Environmental Health Part A, 74 (22-24), pp. 1476–1492, doi: 10.1080/15287394.2011.618977

Rabinowitz, J.D. (2007) 'Cellular metabolomics of *Escherchia coli*', Ex. Rev. Prot., 4 (2), pp. 187-198, doi: 10.1586/14789450.4.2.187

Rabinowitz, J.D. and Kimball, E. (2007) 'Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*', Anal Chem., 79 (16), pp. 6167-6173, doi: 10.1021/ac070470c

Ramanan, V.J. and Saykin, A.J. (2013) 'Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders', Am. J. Neurodeger. Dis., 2 (3), pp. 145-147, PMC3783830

Ramierez-Gaona, M., Marcu, A., Pon, A., Guo, A.C., Sajed, T., Wishart, N.A., Karu, N., Djoumbou-Feunang, Y., Arndt, D. and Wishart, D.S. (2017) 'YMDB 2.0: a significantly expanded version of the yeast metabolome database', Nucleic Acids Res., 4 (45), pp. 440-445, PMC3783830

Raskatov, J.A. and Teplow, D.B. (2017) 'Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins', Sci. Rep., 7, doi: 10.1038/s41598-017-10525-5

Riekeberg, E. and Powers, R. (2017) 'New frontiers in metabolomics: from measurement to insight', F1000 Res., 6, doi: 10.12688/f1000research.11495.1

Roberts, L.D., Souza, A.L., Gerszten, R.E. and Clish, C.B. (2013) 'Targeted metabolomics', Curr Protoc Mol Biol, 30 (2), doi: 10.1002/0471142727.mb3002s98

Rodrigues, M.L. (2018) 'The multifunctional fungal ergosterol', American Society for Microbiology, doi: 10.1128/mBio.01755-18

Roessner, U. and Bowne, J. (2018) 'What is metabolomics all about?', Biotechniques, 46 (5), doi: 10.2144/000113133

Rohn, T.T., Kim, N., Isho, N.F. and Mack, J.M. (2018) 'The potential of CRISPR/Cas9 gene editing as a treatment strategy for Alzheimer's disease', J. Alz. Park., 8 (3), pp. 439, doi: 10.4172/2161-0460.1000439

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. and Dunker, A.K. (2001) 'Sequence complexity of disordered protein', Proteins, 42 (1), pp. 38-48, doi: 10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3

Roser, M. (2017) 'Life expectancy', accessed on 03/03/2017, available at http://ourworldindata.org/data/population-growth-vital-statistics/life-expectancy/

Ross, E.D., Minton, A. and Wickner, R.B. (2005) 'Prion domains: sequences, structures and interactions', Nature Cell Biol., 7, pp. 1039-1044, doi: 10.1038/ncb1105-1039

Roza, L.V., Chanda, A. and Linz, J.E. (2012) 'Compartmentalization and molecular traffic in secondary metabolism: a new understanding of established cellular processes', Fungal Genet. Biol., 48 (1), pp. 35-48, doi:10.1016/j.fgb.2010.05.006

Roze, L.V., Chanda, A. and Linz, J.E. (2012) 'Compartmentalization and molecular traffic in secondary metabolism: a new understanding of established cellular processes', Fungal Genet Biol, 48 (1), pp. 35–48, doi: 10.1016/j.fgb.2010.05.006

Rubakhin, S.S., Lanni, E.J. and Sweedler, J.V. (2013) 'Progress toward single cell metabolomics', Curr Opin Biotechnol, 24 (1), pp. 95–104, doi: 10.1016/j.copbio.2012.10.021

Saborio, G.P., Permanne, B. and Soto, C. (2001) 'Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding', Nature, 411, pp. 810-813, doi: 10.1038/35081095

Safar, J., Wille, H., Itri, V., Growth, D., Serban, H., Torchia, M., Cohen, F.E. and Prusiner, S.B. (1998) 'Eight prion strains have PrPSc molecules with different conformations', Nat. Med., 4, pp. 1157-1165, doi: 10.1038/2654

Salazar, S.V. and Strittmatter, S.M. (2017) 'Cellular prion protein as a receptor for amyloid- oligomers in Alzheimer's disease', Biochem Biophys Res Commun, 483, pp.1143–1147, doi: 10.1016/j.bbrc.2016.09.062

Sanchez, B.J. and Nielsen, J. (2015) 'Genome scale models of yeast: towards standardized evaluation and consistent omic integration', 7 (8), pp. 846-858, doi: 10.1039/c5ib00083a

Sasidharan, K., Amariei, C., Tomita, M. and Murray, D.B. (2012) 'Rapid DNA, RNA and protein extraction protocols optimized for slow continuously growing yeast cultures', Yeast, 29 (8), doi: 10.1002/yea.2911

Sauerschnig, C., Doppler, M., Bueschl, C. and Schuhmacher, R. (2018) 'Methanol generates numerous artifacts during sample extraction and storage of extracts in metabolomics research', Metabolites, 8 (1), doi: 10.3390/metabo8010001

Ser, Z., Lui, X., Tang, N.N. and Locasale, J.W. (2015) 'Extraction parameters for metabolomics from cell extracts', Anal Biochem, 475, pp. 22-28, doi: 10.1016/j.ab.2015.01.003

Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., Moslehi, J.J., Serpell, L., Arnsdorf, M.F. and Lindquist, S. (2000) 'Nucleated conformational conversion and the replication of conformational information by a prion determinant', Science, 289, pp. 1317-1321, doi: 10.1126/science.289.5483.1317

Shao, Y. and Le, W. (2019) 'Recent advances and perspectives of metabolomics-based investigations in Parkinson's disease', Mol. Neurodeg., 3, doi: 10.1186/s13024-018-0304-2

Sherman, F. (2002) 'Getting started with yeast', Meth. in Enzy., 350, pp. 3-41, doi: 10.1016/S0076-6879(02)50954-X

Si, K., Choi, Y.B., White-Grindley, E., Majumdar, A. and Kandel, E.R. (2010) 'Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation', Cell, 140, pp. 421-435, doi: 10.1016/j.cell.2010.01.008

Sigurdson, C.J., Nilsson, K.P.R., Hornemann, S., Heikenwalder, M., Manco, G., Schwarz, P., Ott, D., Rulicke, T., Liberski, P.P., Julius, C., Falsig, J., Stitz, L., Wuthrich, K. and Aguzzi, A. (2009) '*De novo* generation of a transmissible spongiform encephalopathy by mouse transgenesis', Proc Natl Acad Sc. USA, 106 (1), pp. 304-309, doi: 10.1073/pnas.0810680105

Sipe, J.D., Benson, M.D., Buxbaum, J.N., Ikeda, S.I., Merlini, G., Saravina, M.J. and Westermark, P. (2016) 'Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines', Amyloid, 23 (4), pp. 209-213, doi: 10.1080/13506129.2016.1257986

Smolinska, A., Blanchet, L., Buydens, L.M.C. and Wijmenga, S.S. (2012) 'NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review', Anal. Chim. Acta., 750, pp. 82-97, doi: 10.1016/j.aca.2012.05.049

Sondheimer, N. and Lindquist, S. (2000) '*Rnq*1: an epigenetic modifier of protein function in yeast', Mol Cell, 5 (1), pp. 163-72, doi: 10.1016/s1097-2765(00)80412-8

Sondheimer, N., Lopez, N., Craig, E.A. and Lindquist, S. (2001) 'The role of Sis1 in the maintenance of the [RNQ+] prion', EMBO J., 20 (10), pp. 2435-2442, doi: 10.1093/emboj/20.10.2435

Stadtmann, E.R. and Levine, R.L. (2003) 'Free radical-mediated oxidation of free amino acids and amino acid residues in proteins', Amino Acids, 25 (3-4), pp. 207-218, doi: 10.1007/s00726-003-0011-2

Stanbury, P. F., Whitaker, A., & Hall, S. J. (2017). The recovery and purification of fermentation products. Principles of Fermentation Technology, 619–686. doi:10.1016/b978-0-08-099953-1.00010-7

Steele, A.D., Emsley, J.G., Ozdinler, P.H., Linquist, S. and Macklis, J.D. (2006) 'Prion protein (PrPc) positively regulates neural precursor proliferation during developmental and adult mammalian neurogenesis', Proc Natl Acad Sc. USA, 103 (9), pp. 3416-3421, doi: 10.1073/pnas.0511290103

Stefely, J.A. and Pagliarini, D.J. (2018) 'Biochemistry of mitochondrial coenzyme q biosynthesis', Trends Biochem Sci., 42 (10), pp. 824-843, doi: 10.1016/j.tibs.2017.06.008

Stein, K.C. and True, H.L. (2011) 'The [RNQ$^+$] prion: A model of both functional and pathological amyloid', Prion, 5 (4), pp. 291-298, doi: 10.4161/pri.18213

Steinhart, J.R. (2015) 'Prion protein as a mediator of synaptic transmission', Commun. Integr. Biol., 8, e1063753, doi: 10.1080/19420889.2015.1063753

Stephan, J.S., Fioriti, L., Lamba, N., Colnaghi, L., Karl, K., Derkatch, I.L. and Kandel, E.R. (2015) 'The CPEB3 protein is a functional prion that interacts with the actin cytoskeleton', Cell Reports, 11 (1), pp. 1772-1785, doi: 10.1016/j.celrep.2015.04.060

Szkopinska, A. (2000) 'Ubiquinone. biosynthesis of quinone ring and its isoprenoid side chain. Intracellular localization', Acta Biochimica Polonica, 47 (2), pp. 469-480, doi: 10.18388/abp.2000_4027

Szkopinska, A., Swiezewska, E. and Karst, F. (2000) 'The regulation of activity of main mevalonic acid pathway enzymes: Farnesyl diphosphate synthase, 3-hydroxy-3-methylglutaryl-CoA reductase and squalene synthase in yeast *Saccharomyces cerevisiae*', Biochem. Biophys. Res. Commun., 267 (4), pp. 473-477, doi: 10.1006/bbrc.1999.1981

Tambellini, N.P., Zaremberg, V., Turner, R.J. and Weljie, A.M. (2013) 'Evaluation of extraction protocols for simultaneous polar and non-polar yeast metabolite analysis using multivariate projection methods', Metabolites, 3 (3), pp. 592-605, doi: 10.3390/metabo3030592

Tang, J. (2011) 'Microbial metabolomics', Curr. Genomics, 12 (6), pp. 391-403, doi: 10.2174/138920211797248619

Tautenhahn, R., Cho, K.Y., Uritboonthai, W., Zhu, Z., Patti, G.J. and Siuzdak, G. (2012a) 'An accelerated workflow for untargeted metabolomics using the METLIN database', Nat. Biotechnol., 30 (9), pp. 826-828, doi: 10.1038/nbt.2348

Tautenhahn, R., Patti, G.J., Rinehart, D. and Siuzdak, G. (2012b) 'XCMS online: a web-based platform to process untargeted metabolomic data', Anal. Chem., 84 (11), pp. 5035-5039, doi: 10.1021/ac300698c

Taylor, D.R., Whitehouse, I.J. and Hooper, N.M. (2009) 'Glypican-1 mediates both prion protein lipid raft association and disease isoform formation', PLoS Pathog., 5 (11), e1000666, doi: 10.1371/journal.ppat.1000666

Telling, G.C., Haga, T., Torchia, M., Tremblay, P., DeArmond, S.J. and Prusiner, S.B. (1996a) 'Interactions between wild-type and mutant prion proteins modulate neurodegeneration in transgenic mice', Genes Dev., 10. Pp. 1736-1750, doi: 10.1101/gad.10.14.1736

Telling, G.C., Parchi, P., DeArmond, S.J., Cortelli, P., Montagna, P., Gabizon, R., Mastrianni, J., Lugaresi, E., Gambetti, P. and Prusiner, S.B. (1996b) 'Evidence for the conformation of the pathologic isoform of the prion protein enciphering and propagating prion diversity', Science, 274 (5295), pp. 2079-2082, doi: 10.1126/science.274.5295.2079

The Van Deemter equation (2021) The Van Deemter Equation, Chemistry LibreTexts, Available at: https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Supplemental_Modules_(Analytical_Chemistry)/Chromedia/01Gas_Chromotography_(GC)/Gas_Chromotography%3A_Basic_Theory/13The_Van_Deemter_equation. Accessed 4th January 2021

Thermo Fisher Scientific (2014) HILIC Separations Technical Guide, Available at: https://assets.thermofisher.com/TFS-Assets/CMD/brochures/TG-21003-HILIC-Separations-TG21003-EN.pdf. Accessed 5th January 2021

Toledo, J.B., Arnold, M., Kastenmuller, G., Chang, R., Baillie, R.A., Han, X., Thambisetty, M., Tenenbaum, J.D., Suhre, K., Thompson, J.W., St John-Williams, L., MahmoudianDehkordi, S., Rotroff, D.M., Jack, J.R., Motsinger-Reif, A., Risacher, S.L., Blach, C., Lucas, J.E., Massaro, T., Louie, G., Zhu, H., Dallmann, G., Klavins, K., Koal, T., Kim, S., Nho, K., Shen, L., Casanova, R., Varma, S., Legido-Quigley. C., Moseley, M.A., Zhu, K., Henrion, M.Y.R., van der Lee, S.J., Harms, A.C., Demirkan, A., Hankemeier, T., van Duijn, C.M., Trojanowski, J.Q., Shaw, L.M., Saykin, A.J., Weiner, M.W., Doraiswamy, P.M. and Kaddurah-Daouk, P. (2017) 'Metabolic network failures in Alzheimer's disease: A biochemical road map', Alzheimers Dement., 13 (9), pp. 965-984, doi: 10.1016/j.jalz.2017.01.020.

Tompa, P. (2012) 'Intrinsically disorder proteins: a 10-year recap', Trend Biochem. Sci., 37 (12), pp. 509-519, doi: 10.1016/j.tibs.2012.08.004

Toombs, J.A., McCarty, B.R. and Ross, E.D. (2010) 'Compositional determinants of prion formation in yeast', Mol. Cell Biol., 30 (1), pp. 319-332, doi: 10.1128/MCB.01140-09

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S. and Walter, P. (2000) 'Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation', Cell, 101 (3), pp. 249-258, doi: 10.1016/s0092-8674(00)80835-1

Tredwell, G.D., Jones, B.E., Leak, D.J. and Bundy, J.G. (2011) 'The development of metabolomic sampling procedures for *Pichia Pastoris*, and baseline metabolome data', PLOS ONE, 6 (1), doi: 10.1371/journal.pone.0016286

Treusch, S. and Lindquist, S. (2012) 'An intrinsically disordered yeast prion arrests the cell cycle by sequestering a spindle pole body component', J Cell Biol., 197 (3), pp. 369-379, doi: 10.1083/jcb.201108146

Treusch, S., Cyr, D.M. and Lindquist, S. (2009) 'Amyloid deposits: Protection against toxic protein species?', 8 (11), pp. 1668-1674, doi: 10.4161/cc.8.11.8503

Trevino, S.J., Scholtz, J.M. and Pace, C.N. (2007) 'Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa', J. Mol. Biol., 366 (2), pp. 449-460, doi: 10.1016/j.jmb.2006.10.026

Trushina, E. and Mielke, M.M. (2014) 'Recent advances in the application of metabolomics to Alzheimer's disease', Biochim. Biophys. Acta., 1842 (8), pp. 1232-1239, doi: 10.1016/j.bbadis.2013.06.014

Tuite, M.F. (2015) 'Yeast prions: Paramutations at the protein level?', Semin. Cell Dev. Biol., 44, pp. 51-61, doi: 10.1016/j.semcdb.2015.08.016

Tuite, M.F. and Cox, B.S. (2003) 'Propagation of yeast prions', Nature Reviews Molecular Cell Biology, 4, pp. 878-890, doi: 10.1038/nrm1247

Tuite, M.F. and Serio, T.R. (2010) 'The prion hypothesis: from biological anomaly to basic regulatory mechanism', Nat Rev Mol Cell Biol., 11 (12), pp. 823-833, doi: 10.1038/nrm3007

Tuite, M.F., Mundy, C.R. and Cox, B.S. (1981) 'Agents that cause a high frequency of genetic change from [*PSI*$^{+}$] to [*PSI*$^{-}$] in *Saccharomyces cerevisiae*', Genetics, 98 (4), pp. 691-711, PMC1214469

Ugolini, S. and Bruschi, C. (1996) 'The red/white colony color assay in the yeast *Saccharomyces cerevisiae*: epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells', Curr. Genet., 30 (6), pp. 485-492, doi: 10.1007/s002940050160

Unterberger, U., Voigtlander, T. and Budka, H. (2005) 'Pathogenesis of prion diseases', Acta Neuropathol, 109 (1), pp. 32-48, doi: 10.1007/s00401-004-0953-9

Upadhyay, A. and Mishra, A. (2018) 'Amyloids of multiple species: are they helpful in survival?', BIol Rev, 93 (3), doi: 10.1111/brv.12399

Valastyan, J.S. and Lindquist, S. (2014) 'Mechanisms of protein-folding diseases at a glance', Diseases Models and Mechanisms, 7, pp. 9-14, doi: 10.1242/dmm.013474

Valkó, K., Snyder, L. R., and Glajch, J. L. (1993) Retention in reversed-phase liquid chromatography as a function of mobile-phase composition. Journal of Chromatography A, 656(1-2), pp. 501–520, doi:10.1016/0021-9673(93)80816-q

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, R.W., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V. and Tompa, P. (2014) 'Classification of intrinsically disordered regions and proteins', Chem. Rev., 114 (3), pp. 6589-6631, doi: 10.1021/cr400525m

Varma,V.R., Oommen, A.M., Varma, S., Casanova, R., An, Y., Andrews, R.M., O'Brien, R., Pletnikova, O., Troncoso, J.C., Toledo, J., Baillie, R., Arnold, M., Kastenmueller, G., Nho, K., Doraiswamy, P.M., Saykin, A.J., Kaddurah-Daouk, R., Legido-Quigley, C. and Thambisetty, M. (2018) 'Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study', PLoS ONE, doi: 10.1371/journal.pmed.1002482

Velasco, I., Tenreiro, S., Calderon, I.L. and Andre, B. (2004) '*Saccharomyces cerevisiae* Aqr1 is an internal-membrane transporter involved in excretion of amino acids', Eukaryot Cell., 3 (6), pp. 1492-1503, doi: 10.1128/EC.3.6.1492-1503.2004

Vendruscolo, M., Zurdo, J., MacPhee, C.E. and Dobson, C.M. (2003) 'Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems', Phil. Trans. R. Soc., 361, pp. 1205-1222, doi: 10.1098/rsta.2003.1194

Villas-Boas, S.G., Hojer-Pedersen, J., Akesson, M., Smedsgaard, J. and Neilsen, J. (2005) 'Global metabolite analysis of yeast: evaluation of sample preparation methods', Yeast, 22 (14), pp. 1155-1169, doi: 10.1002/yea.1308

Villas-Boas, S.G., Roessner, U., Hansen, M.A.E., Smedsgaard, J. and Nielsen, J. (2007) 'Metabolome analysis: An introduction', 1 ed. John Wiley and Sons Inc., Hoboken, New Jersey, USA.

Vitrenko, Y.A., Gracheva, E.O., Richmond, J.E. and Liebman, S.W. (2007) 'Visualization of aggregation of the Rnq1 prion domain and cross-seeding interactions with Sup35NM', J. Biol. Chem., 282, pp. 1779-1787, doi: 10.1074/jbc.M609269200

Vitrenko, Y.A., Pavon, M.E., Stone, S.I. and Liebman, S.W. (2007) 'Propagation of the [*PIN*$^{+}$] prion by fragments of Rnq1 fused to GFP', Curr. Gene., 51 (5), pp. 309-319, doi: 10.1007/s00294-007-0127-0

Vona, R., Gambardella, L., Cittadini, C., Straface, E. and Pietraforte, D. (2019) 'Biomarkers of oxidative stress in metabolic syndrome and associated diseases', Oxidative Med Cell Longe, 2019, doi: 10.1155/2019/8267234

Walker, F.O. (2007) ‘Huntington’s disease’, The Lancet, 369 (9557), pp. 218–228, doi: 10.1016/s0140-6736(07)60111-1

Wang, K., Melki, R., Kabani, M. and Vorbuerg, I.M. (2017) ‘A prolonged chronological lifespan is an unexpected benefit of the [PSI+] prion in yeast’, PLoS One, 12 (9), pp. 0184905, doi: 10.1371/journal.pone.0184905

Waters (2021a) HPLC Separation Modes – Polarity, Phases, & Chromatography in HPLC | Waters. [online] Available at: www.waters.com/waters/en_US/HPLC-SeparationModes/nav.htm?cid=10049076&locale=en_US#:~:text=In%20normal%2Dphase%20chromatography%2C%20the,phase%20and%20may%20not%20elute. Accessed 2nd January 2021

Waters (2021b) The Promise of Small Particles | Waters. [online] Available at: www.waters.com/waters/en_US/The-Promise-of-Small-Particles/nav.htm?cid=134804750&locale=en_US. Accessed 2nd January 2021

Weckwerth, W. (2010) ‘Metabolomics: an integral technique in systems biology’, Bioanalysis, 2 (4), doi: 10.4155/bio.09.192

Westergard, L., Christensen, H.M. and Harris, D.A. (2007) ‘The cellular prion protein (PrP(C)): its physiological function and role in disease’, Biochim Biophys Acta., 1772, pp. 629-644, doi: 10.1016/j.bbadis.2007.02.011

White, D.A., Buell, A.K., Knowles, T.P., Welland, M.E. and Dobson, C.M. (2010) ‘Protein aggregation in crowded environments’, J. Am. Chem. Soc., 132 (14), pp. 5170-5175, doi: 10.1021/ja909997e

Wickner, R.B. (1994) ‘[URE3] as an altered Ure2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*’, Science, 264 (5158), pp. 566-569, doi: 10.1126/science.7909170

Wickner, R.B., Shewmaker, F.P., Bateman, D.A., Edskes, H.K., Gorkovskiy, A., Dayani, Y. and Bezsonov, E.E. (2015) ‘Yeast prions: structure, biology, and prion-handling systems’, 79 (1), pp. 1-17, doi: 10.1128/MMBR.00041-14

Wickner. R.B., Edskes, H.K., Son, M., Wu, S. and Niznikiewicz, M. (2020) ‘How do Yeast cells contend with prions?’, Int. J. Mol. Sci., 21 (13), pp. 4742-4759, doi: 10.3390/ijms21134742

Wilkins, J.M. and Trushina, E. (2018) ‘Application of metabolomics in Alzheimer’s disease’, Front. Neurol, 8, pp. 710, doi: 10.3389/fneur.2017.00719

Winder, C.L., Dunn, W.B., Schuler, S., Broadhurst, D., Jarvis, R., Stephens, G.M. and Goodacre, R. (2008) ‘Global metabolic profiling of *Escherichia coli* cultures: an evaluation of methods for quenching and extraction of intracellular metabolites’, Anal. Chem., 80 (8), pp. 2939-2948, doi: 10.1021/ac7023409

World Alzheimer report (2010) ‘The global economic impact of dementia’, Alzheimer’s Disease International

World Alzheimer report (2016) ‘Improving healthcare for people living with dementia: coverage, quality and costs now and in the future’, Alzheimer’s Disease International, available at http://eprints.lse.ac.uk/67858/

Worley, B. and Powers, R. (2016) ‘PCA as a practical indicator of OPLS-DA model reliability’, Curr Metabolomics, 4 (2). Pp. 97-103, doi: 10.2174/2213235X04666160613122429

Wu, Z., Song, L., Liu, S.Q. and Huang, D. (2013) ‘Independent and additive effects of glutamic acid and methionine on yeast longevity’, PloS One, 8 (11), pp. e79319, doi: 10.1371/journal.pone.0079319

Xia, J. and Wishart, D.S. (2010a) ‘MetPA: a web-based metabolomics tool for pathway analysis and visualization’, Bioinformatics, 26, pp. 2342-2344, doi: 10.1093/bioinformatics/btq418

Xia, J. and Wishart, D.S. (2010b) ‘MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data’, Nucleic Acids Research, 38, pp. W71-77, doi: 10.1093/nar/gkq329

Xia, J. and Wishart, D.S. (2011a) ‘Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst’, Nature Protocols, 6 (6), pp. 743-760, doi: 10.1038/nprot.2011.319

Xia, J. and Wishart, D.S. (2011b) ‘Metabolomic data processing, analysis, and interpretation using MetaboAnalyst’, Current Protocols in Bioinformatics, 34 (1), doi: 10.1002/0471250953.bi1410s34

Xia, J. and Wishart, D.S. (2016) ‘Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis’, Current Protocols in Bioinformatics, 55 (1), doi: 10.1002/cpbi.11

Xia, J., Broadhurst, D., Wilson, M. and Wishart, D. (2013) ‘Translational biomarker discovery in clinical metabolomics: an introductory tutorial’, Metabolomics, 9, pp. 280-299, doi: 10.1007/s11306-012-0482-9

Xia, J., Mandal, R., Sinelnikov, I., Broadhurst, D., and Wishart, D.S. (2012) ‘MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis’, Nucl. Acids Res, 40, pp. W127-133, doi: 10.1093/nar/gks374

Xia, J., Psychogios, N., Young, N. and Wishart, D.S. (2009) 'MetaboAnalyst: a web server for metabolomic data analysis and interpretation', Nucl. Acids Res, 37, pp. W652-660, doi: 10.1093/nar/gkp356

Xia, J., Sinelnikov, I., and Wishart, D.S. (2011) 'MetATT - a web-based metabolomic tool for analyzing time-series and two-factor data sets', Bioinformatics, 27, pp. 2455-2456, doi: 10.1093/bioinformatics/btr392

Xia, J., Sinelnikov, I., Han, B., and Wishart, D.S. (2015) 'MetaboAnalyst 3.0 - making metabolomics more meaningful', Nucl. Acids Res, 43, pp. W251-257, doi: 10.1093/nar/gkv380

Xu, Y., Letisse, F., Absalan, F., Lu, W., Kuznetsova, E., Brown, G., Caudy, A.A., Yakunin, A.F., Broach, J.R. and Rabinowitz, J.D. (2013) 'Nucleotide degradation and ribose salvage in yeast', Mol Syst Biol, 9, doi:10.1038/msb.2013.21

Yen, C., Harischandra, D.S., Kanthasamy, A. and Sivasankar, S. (2016) 'Copper-induced structural conversion templates prion protein oligomerization and neurotoxicity', Sciences Advances, 2 (7), doi: 10.1126/sciadv.1600014

Yuan. J., Bennett, B. and Rabinowitz, J. (2008) 'Kinetic flux profiling for quantitation of cellular metabolic fluxes', Nature Protoc, 3, pp. 1328–1340, doi: 10.1038/nprot.2008.131

Zhang, A., Sun, H. and Wang, X. (2017) 'Emerging role and recent applications of metabolomics biomarkers in obesity disease research', RSC Adv., 25 (7), pp. 14966-14973, doi: 10.1039/C6RA28715H

Zhang, A., Sun, H., Xu, H., Qiu, S. and Wang, X. (2013) 'Cell metabolomics', Omics, 17 (10), pp. 495-501, doi: 10.1089/omi.2012.0090

Zhang, G., Sadhukhan, S., Tochtrop, G.P. and Bruenengraber, H. (2011) 'Metabolomics, pathway regulation, and pathway discovery', J. Biol. Chem., 286, pp. 23631-23635, doi: 10.1074/jbc.R110.171405

Zhu, Z.J., Schultz, A.W., Wang, J., Johnson, C.H., Yannone, S.M. Patti, G.J. and Siuzdak, G. (2013) 'Quadropole-time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database', Nat. Proto., 8, pp. 451-460, doi: 10.1038/nprot.2013.004

Appendixes

Appendix A

Details the top ten pathway hits for Chapter 4 analysis, specifically relating to Table 4.2., 4.4., 4.6, 4.8. and 4.12. respectively.

Table A.1. Top Ten PIM pathway hits via Metaboanalyst's Mummichog using XCMS data.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	6	5	0.09457	0.016791	0.001701
tryptophan degradation I (via anthranilate)	9	5	4	0.20286	0.043635	0.002779
trans, trans-farnesyl diphosphate biosynthesis	5	4	3	0.40064	0.109	0.007017
biotin biosynthesis from 7-keto-8-aminopelargonate	15	7	4	0.42055	0.16963	0.007724
isoleucine biosynthesis	15	5	3	0.5325	0.20705	0.013436
hexaprenyl diphosphate biosynthesis	6	5	3	0.5325	0.20705	0.013436
salvage pathways of pyrimidine deoxyribonucleotides	17	5	3	0.5325	0.20705	0.013436
homoserine biosynthesis	12	2	2	0.55112	0.10985	0.01477
nicotinate riboside salvage pathway I	5	2	2	0.55112	0.10985	0.01477
dolichol and dolichyl phosphate biosynthesis	13	2	2	0.55112	0.10985	0.01477

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst.

Table A.2. Top Ten NIM pathway hits via Metaboanalyst's Mummichog using XCMS data.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
phosphopantothenate biosynthesis I	16	5	4	0.52968	0.21092	0.010671
leucine biosynthesis	16	4	3	0.71247	0.34181	0.022197
pyridoxal 5'-phosphate salvage pathway	13	4	3	0.71247	0.34181	0.022197
nicotinamide riboside salvage pathway II	7	2	2	0.76657	0.26946	0.028388
thiazole biosynthesis III (eukaryotes)	11	2	2	0.76657	0.26946	0.028388
xylose metabolism	8	2	2	0.76657	0.26946	0.028388
sphingolipid recycling and degradation (yeast)	22	2	2	0.76657	0.26946	0.028388
sphingosine and sphingosine-1-phosphate metabolism	14	2	2	0.76657	0.26946	0.028388
xylitol degradation	8	2	2	0.76657	0.26946	0.028388
ubiquinol-6 biosynthesis (eukaryotic)	21	5	3	0.83433	0.53814	0.039989

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst.

Table A.3. Top Ten PIM pathway hits via Metaboanalyst's Mummichogg using MZmine2 data.

Pathway Name	*Pathway total*	*Hits total*	*Hits sig*	*EASE*	*FET*	*Gamma*
salvage pathways of pyrimidine deoxyribonucleotides	17	6	4	0.26337	0.075142	0.001591
tryptophan biosynthesis	17	9	5	0.27948	0.10584	0.00173
4-amino-2-methyl-5-diphosphomethylpyrimidine biosynthesis	11	4	3	0.35673	0.088631	0.002596
thiamin diphosphate biosynthesis IV (eukaryotes)	11	4	3	0.35673	0.088631	0.002596
pyridoxal 5'-phosphate salvage pathway	13	4	3	0.35673	0.088631	0.002596
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	4	3	0.35673	0.088631	0.002596
tryptophan degradation I (via anthranilate)	9	4	3	0.35673	0.088631	0.002596
methionine biosynthesis	18	8	4	0.46127	0.2066	0.004561
lipoate biosynthesis and incorporation I	9	2	2	0.51756	0.094284	0.006232
nicotinate riboside salvage pathway I	5	2	2	0.51756	0.094284	0.006232

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst.

Table A.4. Top Ten NIM pathway hits via Metaboanalyst's Mummichog using MZmine2 data.

Pathway Name	*Pathway total*	*Hits total*	*Hits sig*	*EASE*	*FET*	*Gamma*
6-hydroxymethyl-dihydropterin diphosphate biosynthesis I	14	4	3	0.14625	0.019586	0.000531
glycine biosynthesis II	8	2	2	0.32375	0.032041	0.001388
glycine biosynthesis from serine	5	2	2	0.32375	0.032041	0.001388
serine biosynthesis from glyoxylate	5	2	2	0.32375	0.032041	0.001388
nicotinamide riboside salvage pathway I	7	2	2	0.32375	0.032041	0.001388
glycine cleavage	11	2	2	0.32375	0.032041	0.001388
pyrimidine ribonucleosides degradation II	6	2	2	0.32375	0.032041	0.001388
folate transformations	26	7	3	0.36067	0.11365	0.001703
phosphopantothenate biosynthesis I	16	7	3	0.36067	0.11365	0.001703
folate interconversions	22	8	3	0.4297	0.15899	0.002512

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst.

Table A.5. Top Ten NIM pathway hits via Metaboanalyst's Mummichog using comparative data.

Pathway Name	*Pathway total*	*Hits total*	*Hits sig*	*EASE*	*FET*	*Gamma*
ubiquinol-6 biosynthesis (eukaryotic)	21	6	4	0.51914	0.22761	0.0068221
leucine biosynthesis	16	4	3	0.57758	0.21942	0.0088931
pyridoxal 5'-phosphate salvage pathway	13	4	3	0.57758	0.21942	0.0088931
sphingolipid recycling and degradation (yeast)	22	2	2	0.67657	0.18784	0.01429
xylitol degradation	8	2	2	0.67657	0.18784	0.01429
phosphopantothenate biosynthesis I	16	5	3	0.71593	0.37775	0.017462
mevalonate pathway	17	3	2	0.81695	0.40205	0.03071
xylose metabolism	8	3	2	0.81695	0.40205	0.03071
sphingosine and sphingosine-1-phosphate metabolism	14	3	2	0.81695	0.40205	0.03071
salvage pathways of pyrimidine deoxyribonucleotides	17	7	3	0.88128	0.65561	0.046963

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores EASE, FET and Gamma as calculated by Metaboanalyst.

Appendix B

B.1. Statistical analysis of comparative metabolomic data in Metaboanalyst finds significant differences between [rnq$^-$] and [rnq$^-$] with mild oxidative stress samples

Comparative data were treated in an identical manner to data in section 4.5., with all 12 samples being subjected to pairwise comparison within Metaboanalyst, using the settings described in section 2.9. This time however resultant T-tests, with the adjusted *p-value* of =>0.01, found 2189 positive and 1354 negative features that had been significantly altered (*p-value*=>0.01) when comparing [*rnq*$^-$] samples and [*rnq*$^-$] with mild oxidative stress samples. Notably more PIM features than NIM features, something which had not been previously observed using this programme.

B.2. Tentative feature ID and pathway analysis of comparative metabolomic data via Metaboanalyst

As in chapter 4, to establish the identity and run pathway analysis on the increased and decreased features detected by both XCMS and MZmine2, the 2189 PIM features and 1354 NIM features where subject to Metaboanalyst's, Mummichog, Peaks to Pathways function, as indicated by Figure 5.1. A new algorithm available within Metaboanalyst, known as GSEA, believed to have been adapted from the GSEA algorithm commonly used in genetic analysis, described by Metaboanalyst as 'a cutoff-free method using the overall rank based on t.score', was selected. This decision was made based on a personal dislike of the arbitrary cut off points that are required by the previous method, in requiring a limited number of significant hits for algorithm function. Once again, BioCyc's Fungi pathway library for *Saccharomyces cerevisiae* (yeast) was selected, providing details of the most significantly different metabolites, the direction of the disturbance (whether they were up or down regulated), and the most significantly disrupted pathways including the total number of hits within specified pathways (Tables A.1., A.2., A.3. and A.4.). Curiously within the PIM results show only downregulation for the top 25 hits, this is something that has not been previously observed. However, more correlation is observed between PIM and NIM results than ever before.

Table B.1. The top twenty-five unique and available proposed identities of the most significant PIM comparative features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Relative change [*rnq*$^-$] > [*rnq*$^-$] stress	p-value	FDR	Mass difference
204.0643	2.52	*indole pyruvate*	↓	1.11E-13	1.86E-10	0.0012
251.1036	1.3	*ch33ado*	↓	1.14E-13	1.86E-10	0.0018
234.078	1.91	*cpd-822*	↓	4.05E-13	3.06E-10	0.0007
121.0683	0.95	*homo-ser*	↓	4.37E-13	3.06E-10	0.0006
399.1206	2.76	*s-adenosyl-4-methylthio-2-oxobutanoate*	↓	1.12E-12	4.56E-10	0.0001
310.1084	1.17	*s-hydroxymethylglutathione*	↓	1.36E-12	5.07E-10	0.0018
137.0709	0.89	*c08270*	↓	1.45E-12	5.07E-10	0.0013
253.0924	2.52	*c05512*	↓	3.65E-12	9.40E-10	0.0008
311.0775	1.92	*c04677*	↓	4.40E-12	9.81E-10	0.0025
183.0805	6.17	*coniferyl-alcohol*	↓	4.76E-12	1.00E-09	0.0012
292.0789	2.83	*cpd-5923*	↓	4.90E-12	1.00E-09	0.0028
274.0688	2.51	*n-acetyl-d-glucosamine-1-p*	↓	5.72E-12	1.07E-09	0.0003
267.098	0.67	*c00212*	↓	5.92E-12	1.07E-09	0.0012
269.1132	1.29	*7-8-dihydropteroate*	↓	6.62E-12	1.16E-09	0.0014
149.0111	1.51	*imidazole-acetol-p*	↓	7.45E-12	1.20E-09	0.0000
167.0825	0.86	*cpd-371*	↓	1.34E-11	1.89E-09	0.0004
310.093	1.5	*cpd-10809*	↓	1.62E-11	2.07E-09	0.0018
237.0858	0.9	*c00534*	↓	1.62E-11	2.07E-09	0.0012
120.0476	2.52	*amino-oxobut*	↓	1.71E-11	2.07E-09	0.0006
218.0817	2.91	*cpd-12676*	↓	4.27E-11	4.19E-09	0.0010
179.0811	0.9	*3-hydroxy-l-kynurenine*	↓	4.82E-11	4.63E-09	0.0005
175.0623	2.81	*indole_acetate_auxin*	↓	1.28E-10	1.04E-08	0.0010
292.1013	1.29	*canavaninosuccinate*	↓	1.42E-10	1.11E-08	0.0006
185.0938	0.86	*c00526*	↓	1.58E-10	1.14E-08	0.0017
252.0866	2.52	*c01136*	↓	2.48E-10	1.60E-08	0.0022

Their m/z values, retention times, up or down regulation when comparing [*rnq*$^-$] to [*rnq*$^-$] with oxidative stress, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table B.2. Top Ten PIM pathway hits via Metaboanalyst's Mummichogg using comparative data.

Pathway Name	Pathway total	Hits total	P-value	NES
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	29	9	0.01124	1.736
UDP-N-acetylglucosamine biosynthesis	12	5	0.02353	1.577
UDP-N-acetyl-D-glucosamine biosynthesis II	12	5	0.02353	1.577
tryptophan degradation I (via anthranilate)	3	1	0.025	1.229
phenylalanine biosynthesis	16	8	0.03448	1.49
tyrosine biosynthesis	12	7	0.03571	1.487
hexaprenyl diphosphate biosynthesis	12	2	0.03704	1.341
trans, trans-farnesyl diphosphate biosynthesis	21	5	0.04706	1.515
leucine biosynthesis	9	1	0.04839	-1.334
biotin biosynthesis from 7-keto-8-aminopelargonate	9	1	0.04839	-1.334

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores *p-value* and NES as calculated by Metaboanalyst.

Table B.3. The top twenty-five unique and available proposed identities of the most significant PIM comparative features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Relative change [*rnq*$^-$] > [*rnq*$^-$] stress	p-value	FDR	Mass difference
218.0405	2.50	*tyr*	↓	1.07E-13	2.56E-10	0.0007
267.0875	1.29	*cpd-9875*	↓	1.60E-13	2.56E-10	0.0006
109.0152	2.50	*b-alanine*	↓	1.29E-12	6.72E-10	0.0001
181.0704	7.02	*coniferyl-alcohol*	↑	1.40E-12	6.72E-10	0.0018
190.0423	2.50	*dehydroquinate*	↓	1.47E-12	6.72E-10	0.0015
220.051	2.50	*l-histidinol-p*	↓	5.09E-12	1.48E-09	0.0017
268.0735	2.82	*cpd-13040*	↓	1.17E-11	2.87E-09	0.0015
174.0513	2.50	*cpd-1091*	↓	2.36E-11	4.61E-09	0.0007
172.0345	2.50	*3-dehydro-shikimate*	↓	2.39E-11	4.61E-09	0.0012
144.0409	2.50	*cpd-578*	↓	3.20E-11	5.44E-09	0.0005
269.047	7.02	*dihydrokaempferol-cmpd*	↑	4.78E-11	7.06E-09	0.0015
231.0712	1.84	*cpd-822*	↓	5.08E-11	7.06E-09	0.0008
233.0541	4.27	*c09871*	↓	1.63E-10	1.74E-08	0.0000
199.0217	4.28	*o-phospho-l-homoserine*	↓	2.27E-10	2.21E-08	0.0010
205.0563	1.38	*c00534*	↓	3.51E-10	2.98E-08	0.0005
219.0414	2.50	*phosphoryl-choline*	↓	3.68E-10	3.02E-08	0.0014
206.0395	2.50	*homo-cit*	↓	3.89E-10	3.11E-08	0.0007
335.0771	1.29	*cpd-548*	↓	4.42E-10	3.45E-08	0.0023
221.9796	4.26	*4-oxalomesaconate*	↑	5.17E-10	3.67E-08	0.0008
234.042	4.08	*kynurenate*	↓	8.63E-10	5.63E-08	0.0018
268.095	1.29	*c00526*	↓	1.00E-09	6.28E-08	0.0011
214.0408	3.02	*o-ureidohomoserine*	↓	1.18E-09	7.12E-08	0.0001
303.9799	2.50	*c11355*	↓	1.22E-09	7.21E-08	0.0022
270.0869	4.46	*dihydro-neo-pterin*	↓	1.44E-09	8.20E-08	0.0025
136.035	2.50	*c00262*	↓	1.51E-09	8.48E-08	0.0004

Their m/z values, retention times, up or down regulation when comparing [*rnq*$^-$] to [*rnq*$^-$] with oxidative stress, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table B.4. Top Ten NIM pathway hits via Metaboanalyst's Mummichog using comparative data.

Pathway Name	Pathway total	Hits total	P-value	NES
tryptophan degradation I (via anthranilate)	9	4	0.01493	1.604
alanine biosynthesis	4	3	0.01667	1.49
alanine degradation III	4	3	0.01667	1.49
beta-alanine biosynthesis	10	1	0.02174	1.361
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	7	0.02632	1.498
glycine biosynthesis from glyoxylate	4	2	0.0339	1.474
tyrosine biosynthesis	11	4	0.04478	1.445
tyrosine degradation	12	4	0.04478	1.445
thiazole biosynthesis III (eukaryotes)	11	2	0.05085	1.437
histidine biosynthesis	22	4	0.07463	1.401

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores *p-value* and NES as calculated by Metaboanalyst.

Appendix C

C.1. Statistical analysis of XCMS/MZmine2 comparative metabolomic data in Metaboanalyst finds significant differences between [rnq$^-$] and [RNQ$^+$] samples

Comparative data were treated in an identical manner to data in section 4.5., with all 12 samples being subjected to pairwise comparison within Metaboanalyst, using the settings described in section 2.9. This time however resultant T-tests, with the adjusted *p-value* of =>0.01, found 2880 positive and 1577 negative features that had been significantly altered (*p-value*=>0.01) when comparing [*rnq*$^-$] samples and [*RNQ*$^+$] samples.

C.2. Tentative feature ID and pathway analysis of XCMS/MZmine2 comparative metabolomic data via Metaboanalyst

As in chapter 4, to establish the identity and run pathway analysis on the increased and decreased features detected by both XCMS and MZmine2, the 2880 PIM features and 1577 NIM features where subject to Metaboanalyst's, Mummichog, Peaks to Pathways function, as indicated by Figure 5.1. The GSEA algorithm was selected alongside BioCyc's Fungi pathway library for *Saccharomyces cerevisiae* (yeast). This provided details of the most significantly different metabolites, the direction of the disturbance (whether they were up or down regulated), and the most significantly disrupted pathways including the total number of hits within specified pathways (Tables B.1., B.2., B.3. and B.4.). Curiously within the PIM results show only downregulation for the top 25 hits, this is something that has not been previously observed. However, more correlation is observed between PIM and results than ever before.

Table C.1. The top twenty-five unique and available proposed identities of the most significant PIM comparative features identified by Metaboanalyst's T-test.

m/z	RT (min)	Proposed Identity	Relative change [rnq-] > [RNQ+]	*p-value*	FDR	Mass difference
252.0866	2.5168	*c01136*	↓	2.90E-13	1.42E-09	0.0022
234.078	2.3707	*cpd-822*	↓	2.04E-12	3.38E-09	0.0007
270.1165	1.2872	*s-hydroxymethylglutathione*	↓	1.97E-11	1.10E-08	0.0023
183.0805	6.1744	*coniferyl-alcohol*	↓	2.95E-11	1.31E-08	0.0012
269.1132	1.2909	*7-8-dihydropteroate*	↓	9.88E-11	3.32E-08	0.0014
81.0334	1.509	*2-keto-isovalerate*	↓	1.06E-10	3.32E-08	0.0001
256.0588	3.7321	*n-acetyl-d-glucosamine-1-p*	↓	1.29E-10	3.32E-08	0.0007
399.1206	2.7587	*s-adenosyl-4-methylthio-2-oxobutanoate*	↓	1.87E-10	4.36E-08	0.0001
311.0775	1.9184	*c04677*	↓	2.28E-10	4.85E-08	0.0025
235.0807	2.5194	*ch33ado*	↓	2.85E-10	5.13E-08	0.0019
377.1456	2.3517	*c00255*	↓	2.87E-10	5.13E-08	0.0000
99.0458	5.0065	*amino-oh-hydroxymethyl-dihydropteridine*	↑	3.92E-10	6.40E-08	0.0010
292.0789	2.8304	*cpd-5923*	↓	4.59E-10	7.02E-08	0.0028
354.0989	3.7515	*cpd-548*	↓	7.28E-10	9.64E-08	0.0023
273.0871	1.8453	*trp*	↓	9.25E-10	1.13E-07	0.0025
108.0456	2.2337	*c00740*	↓	1.23E-09	1.26E-07	0.0001
204.0643	2.5193	*indole_pyruvate*	↓	1.36E-09	1.33E-07	0.0012
253.0924	2.5164	*c05512*	↓	1.67E-09	1.49E-07	0.0008
289.1401	0.8856	*cpd-9700*	↓	2.63E-09	2.13E-07	0.0006
149.0111	1.509	*imidazole-acetol-p*	↓	2.82E-09	2.13E-07	0.0000
441.1523	8.4308	*methylene-thf*	↑	3.67E-09	2.50E-07	0.0006
210.9926	1.5092	*cpd-4211*	↓	4.17E-09	2.71E-07	0.0007
292.1013	1.2898	*canavaninosuccinate*	↓	4.34E-09	2.76E-07	0.0006
102.0919	0.6209	*c01475*	↑	4.48E-09	2.78E-07	0.0005
287.0806	4.4028	*c01762*	↓	1.03E-08	5.37E-07	0.0019

Their m/z values, retention times, up or down regulation when comparing [*rnq*$^-$] to [*RNQ*$^+$], the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table C.2. Top Ten PIM pathway hits via Metaboanalyst's Mummichogg using comparative data.

Pathway Name	Pathway total	Hits	P_val	P_adj	NES
tetrapyrrole biosynthesis	11	1	0.04	0.8013	1.347
tetrapyrrole biosynthesis II	11	1	0.04	0.8013	1.347
fatty acid biosynthesis initiation	7	1	0.04	0.8013	1.347
fatty acid biosynthesis (concise)	10	1	0.04	0.8013	1.347
very long chain fatty acid biosynthesis	24	1	0.04	0.8013	1.347
palmitoleate biosynthesis	9	1	0.04	0.8013	1.347
ethanol degradation I	7	1	0.04	0.8013	1.347
pyruvate dehydrogenase complex	10	1	0.04	0.8013	1.347
carnitine shuttle	4	1	0.04	0.8013	1.347
UDP-N-acetylglucosamine biosynthesis	12	6	0.05769	0.8013	1.696

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores *p-value* and NES as calculated by Metaboanalyst.

Table C.3. The top twenty-five unique and available proposed identities of the most significant NIM comparative features identified by Metaboanalyst's T-test.

m/z	*RT (min)*	*Proposed Identity*	*Relative change [rnq⁻] > [RNQ⁺]*	*p-value*	*FDR*	*Mass difference*
144.0409	2.4967	cpd-578	↓	4.89E-14	5.83E-11	0.0005
218.0405	2.4968	tyr	↓	5.47E-14	5.83E-11	0.0007
267.0661	1.1675	c02291	↑	2.34E-11	1.08E-08	0.0010
267.0875	1.285	cpd-9875	↓	3.39E-11	1.08E-08	0.0006
231.0712	1.2858	cpd-822	↓	6.50E-11	1.60E-08	0.0008
303.9799	2.501	c11355	↓	1.34E-10	3.07E-08	0.0022
268.095	1.285	c00526	↓	1.45E-10	3.08E-08	0.0011
109.0152	2.4973	b-alanine	↓	2.03E-10	3.90E-08	0.0001
165.0598	1.1758	2-deoxy-d-glucose	↓	2.08E-10	3.90E-08	0.0014
220.051	2.4977	l-histidinol-p	↓	3.58E-10	6.03E-08	0.0017
172.0345	2.499	3-dehydro-shikimate	↓	4.44E-10	6.76E-08	0.0012
214.0408	3.0154	o-ureidohomoserine	↓	5.23E-10	7.60E-08	0.0001
369.0298	1.9388	pseudouridine-5-p	↑	7.57E-10	9.67E-08	0.0037
231.0712	1.8377	cpd-822	↓	1.41E-09	1.36E-07	0.0008
335.0771	1.2897	cpd-548	↓	1.43E-09	1.36E-07	0.0023
199.0217	4.278	o-phospho-l-homoserine	↓	1.79E-09	1.51E-07	0.0010
219.0414	2.4988	phosphoryl-choline	↓	2.38E-09	1.90E-07	0.0014
205.0563	1.3752	c00534	↓	2.76E-09	2.06E-07	0.0005
369.0298	1.7239	pseudouridine-5-p	↑	3.36E-09	2.34E-07	0.0037
268.0735	2.8159	cpd-13040	↓	7.72E-09	4.74E-07	0.0015
206.0395	2.5003	homo-cit	↓	7.94E-09	4.79E-07	0.0007
136.035	2.4951	c00262	↓	8.37E-09	4.95E-07	0.0004
269.0971	1.2908	cpd-13040	↓	8.74E-09	5.08E-07	0.0002
190.0423	2.4954	dehydroquinate	↓	9.48E-09	5.23E-07	0.0015
233.0541	4.2685	c09871	↓	9.49E-09	5.23E-07	0.0000

Their m/z values, retention times, up or down regulation when comparing [*rnq*⁻] to [*RNQ*⁺], the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table C.4. Top Ten NIM pathway hits via Metaboanalyst's Mummichog using comparative data.

Pathway Name	*Pathway total*	Hits	P_val	P_adj	NES
tyrosine biosynthesis	11	4	0.02	0.928	1.736
tyrosine degradation	12	4	0.02	0.928	1.736
urea degradation	9	1	0.02273	0.928	1.365
putrescine biosynthesis	4	1	0.03448	0.928	-1.311
beta-alanine biosynthesis	10	1	0.04545	0.928	1.346
formaldehyde oxidation II (glutathione-dependent)	12	1	0.04545	0.928	1.333
methionine degradation I (to homocysteine)	11	1	0.06897	0.928	-1.305
tryptophan biosynthesis	17	7	0.09302	0.928	-1.519
xylose metabolism	8	1	0.1136	0.928	1.288
xylitol degradation	8	1	0.1136	0.928	1.288

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the significance scores *p-value* and NES as calculated by Metaboanalyst.

Appendix D

D.1. Statistical analysis of comparative metabolomic data in Metaboanalyst finds significant differences between groups W, X, Y and Z

Analysis via Metaboanalyst was slightly different given that four groups were being compared, all 24 samples being subjected to group comparison within Metaboanalyst, using the settings described in section 2.9. This time however T-tests were not possible due the lack of pairwise comparison and so ANOVAs were used, with the adjusted *p-value* of =>0.01, found 1800 significant changes at T3 and 2361 significant changes at T4 (*p-value*=>0.01) when comparing samples in group W, X, Y and Z.

D.2. Tentative feature ID and pathway analysis of comparative metabolomic data via Metaboanalyst of groups W, X, Y and Z

As in chapter 4, to establish the identity and run pathway analysis on the increased and decreased features detected by both XCMS and MZmine2, the 1800 T3 features and 2361 T4 features where subject to Metaboanalyst's, Mummichog, Peaks to Pathways function, as indicated by Figure 5.1. The GSEA algorithm was not available for multiple comparisons and so the need to revert back to the algorithm used in chapter 4 was essential, A *p-value* of $1.0\ x\ 10^{-7}$ was selected for this algorithm functioning alongside BioCyc's Fungi pathway library for *Saccharomyces cerevisiae* (yeast). This provided details of the most significantly different metabolites, the direction of the disturbance (whether they were up or down regulated), and the most significantly disrupted pathways including the total number of hits within specified pathways (Tables C.1., C.2., C.3. and C.4.).

Table D.1. The top twenty-five unique and available proposed identities of the most significant T3 comparative features identified by Metaboanalyst's ANOVA.

m/z	*RT (min)*	*Proposed Identity*	*Relative change W > X*	*Relative change W > Y*	*Relative change X > Y*	*Relative change Y > Z*	*p-value*	*FDR*	*Mass difference*
369.2373	5.9355	*2-keto-3-methyl-valerate*	↑	↑	↓	↑	4.04E-07	4.05E-06	0.0002
74.0597	1.8879	*4-hydroxyphenyllactate*	↑	↓	↓	↓	1.37E-09	4.98E-08	0.0007
175.0854	0.9929	*allantoate*	↑	↑	↓	↓	9.72E-07	7.80E-06	0.0016
173.0757	2.3254	*amino-acetone*	↑	↑	↓	↑	8.68E-09	2.15E-07	0.0003
389.1824	3.0421	*anthranilate*	↑	↑	↑	↑	8.68E-07	7.22E-06	0.0001
110.06	4.2471	*c00193*	↑	↑	↑	↓	9.05E-07	7.45E-06	0.0007
102.0919	0.617	*c00250*	↑	↓	↓	↓	2.78E-09	8.30E-08	0.0005
231.2156	2.2425	*c00534*	↑	↓	↓	↓	2.89E-08	5.74E-07	0.0015
158.0629	0.995	*c00882*	↑	↑	↓	↓	8.38E-11	5.62E-09	0.0013
333.06	1.3453	*c02763*	↑	↑	↓	↑	4.01E-10	1.89E-08	0.0011
463.3517	5.0569	*c04525*	↓	↑	↑	↓	7.87E-11	5.41E-09	0.0016
123.0452	0.6415	*c04677*	↑	↑	↓	↑	4.79E-09	1.31E-07	0.0016
62.0604	0.6168	*c05382*	↑	↑	↓	↓	3.75E-10	1.79E-08	0.0006
216.0627	6.7033	*c09871*	↓	↑	↑	↑	4.13E-07	4.10E-06	0.0010
214.0721	7.7764	*c11355*	↓	↓	↑	↓	4.83E-07	4.61E-06	0.0004
336.2296	5.1849	*canavaninosuccinate*	↑	↑	↑	↑	2.06E-07	2.42E-06	0.0025
220.0325	0.6425	*cpd-10608*	↑	↑	↓	↑	5.63E-07	5.17E-06	0.0013
118.0666	0.6419	*cpd-11020*	↑	↑	↓	↑	1.17E-11	1.38E-09	0.0001
109.1006	3.8885	*cpd-195*	↓	↓	↑	↓	4.59E-12	7.20E-10	0.0008
130.0655	0.9942	*cpd-237*	↑	↑	↓	↓	2.06E-11	2.18E-09	0.0012
81.0334	1.4721	*cpd-255*	↑	↑	↓	↑	9.66E-07	7.80E-06	0.0001
183.0923	3.9537	*cpd-35*	↑	↑	↑	↑	8.47E-07	7.09E-06	0.0005
235.1343	5.9861	*cpd-375*	↑	↓	↑	↓	1.04E-08	2.45E-07	0.0008
91.0568	0.6418	*cpd-6082*	↑	↑	↓	↑	9.42E-11	6.05E-09	0.0003
398.3275	4.0253	*cpd-7682*	↑	↑	↑	↑	2.10E-10	1.13E-08	0.0009

Their m/z values, retention times, up or down regulation when comparing groups W, X, Y and Z, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table D.2. Top Ten T3 pathway hits when comparing groups W, X, Y and Z via Metaboanalyst's Mummichogg using comparative data.

Pathway Name	*Pathway total*	*Hits total*	*Hits sig*	*EASE*	*FET*	*Gamma*
leucine biosynthesis	16	8	7	0.20953	0.059754	0.032749
tryptophan degradation I (via anthranilate)	9	5	5	0.2476	0.048279	0.034922
methionine biosynthesis	18	7	6	0.30544	0.098258	0.038588
phenylalanine biosynthesis	11	4	4	0.38246	0.088888	0.044281
tetrahydrofolate biosynthesis	13	4	4	0.38246	0.088888	0.044281
phenylalanine degradation	12	4	4	0.38246	0.088888	0.044281
histidine biosynthesis	22	6	5	0.43119	0.15884	0.048468
isoleucine biosynthesis	15	6	5	0.43119	0.15884	0.048468
ubiquinol-6 biosynthesis (eukaryotic)	21	6	5	0.43119	0.15884	0.048468
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	6	5	0.43119	0.15884	0.048468

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the EASE, FET and Gamma scores as calculated by Metaboanalyst.

Table D.3. The top twenty-five unique and available proposed identities of the most significant T4 comparative features identified by Metaboanalyst's ANOVA.

m/z	RT (min)	Proposed Identity	Relative change W > X	Relative change W > Y	Relative change X > Y	Relative change Y > Z	p-value	FDR	Mass difference
148.075	8.6729	*1-keto-2-methylvalerate*	↓	↓	↓	↓	5.20E-12	9.15E-11	0.0014
195.0877	1.2127	*2-d-threo-hydroxy-3-carboxy-isocaproate*	↓	↓	↑	↓	1.34E-13	4.70E-12	0.0014
121.0491	2.9086	*2-oxobutanoate*	↓	↓	↑	↓	1.03E-12	2.39E-11	0.0005
442.1796	5.1176	*5-methyl-thf*	↓	↑	↑	↓	1.08E-13	4.03E-12	0.0037
243.0738	1.2764	*5-phospho-ribosyl-glycineamide*	↓	↓	↓	↓	1.80E-12	3.72E-11	0.0003
336.2296	6.5379	*6z8e10e14z-5s12r-512-dihydroxyi*	↑	↑	↑	↑	6.78E-14	2.89E-12	0.0005
269.1132	1.2735	*7-8-dihydropteroate*	↑	↑	↓	↑	5.74E-16	1.12E-13	0.0014
148.059	0.9987	*acetylserine*	↓	↓	↑	↓	7.22E-15	5.76E-13	0.0014
120.0452	3.8036	*amino-oxobut*	↓	↓	↑	↓	1.27E-13	4.51E-12	0.0004
92.0507	3.8032	*b-alanine*	↓	↓	↑	↓	2.23E-12	4.56E-11	0.0001
308.0904	0.5363	*c00051*	↓	↓	↑	↓	1.66E-12	3.46E-11	0.0007
107.0499	2.6535	*c00193*	↓	↓	↑	↓	3.72E-12	7.10E-11	0.0007
135.0296	0.8818	*c00242*	↓	↓	↑	↓	1.98E-13	6.27E-12	0.0006
177.0892	2.4958	*c00299*	↓	↓	↑	↓	1.84E-13	5.98E-12	0.0002
161.0952	0.5966	*c00526*	↓	↓	↑	↓	1.88E-14	1.18E-12	0.0007
141.101	1.5583	*c00534*	↓	↓	↑	↓	1.71E-13	5.70E-12	0.0012
74.0597	1.724	*c01475*	↓	↓	↑	↓	7.49E-13	1.83E-11	0.0003
198.1138	2.4247	*c01909*	↓	↓	↓	↓	2.28E-17	9.32E-15	0.0013
223.0793	0.6938	*c02052*	↓	↓	↑	↓	3.13E-13	8.94E-12	0.0020
223.0727	0.6043	*c02291*	↓	↓	↑	↓	2.92E-12	5.85E-11	0.0020
213.1026	3.5926	*c09871*	↓	↓	↑	↓	4.35E-12	7.99E-11	0.0010
252.1074	1.1062	*ch33ado*	↓	↓	↑	↓	5.20E-12	9.15E-11	0.0017
364.1559	2.074	*cpd-10279*	↓	↓	↓	↑	1.16E-13	4.29E-12	0.0024
328.1044	1.8014	*cpd-10809*	↓	↓	↑	↓	2.03E-14	1.25E-12	0.0028
291.0952	1.2759	*cpd-13040*	↑	↑	↓	↑	3.27E-13	9.22E-12	0.0018

Their m/z values, retention times, up or down regulation when comparing groups W, X, Y and Z, the *p-value*, false discovery rate and mass difference as calculated by Metaboanalyst.

Table D.4. Top Ten T4 pathway hits when comparing groups W, X, Y and Z via Metaboanalyst's Mummichogg using comparative data.

Pathway Name	Pathway total	Hits total	Hits sig	EASE	FET	Gamma
tryptophan biosynthesis	17	9	9	0.036781	0.18161	0.016493
histidine biosynthesis	22	12	11	0.077401	0.23357	0.018822
leucine biosynthesis	16	8	8	0.053308	0.23988	0.019129
lysine biosynthesis	22	8	8	0.053308	0.23988	0.019129
NAD biosynthesis from 2-amino-3-carboxymuconate semialdehyde	14	7	7	0.07718	0.31354	0.023169
tryptophan degradation VIII (to tryptophol)	12	7	7	0.07718	0.31354	0.023169
tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	7	7	0.07718	0.31354	0.023169
arginine biosynthesis	26	13	11	0.18911	0.39771	0.029041
NAD salvage pathway	17	6	6	0.11163	0.40464	0.029598
folate transformations	26	6	6	0.11163	0.40464	0.029598

The name of the pathway, the total number of metabolites known within the pathway, the total number of hits, the total number of significant hits, and the EASE, FET and Gamma scores as calculated by Metaboanalyst.

Appendix E

E.1. Colony counts from all sample groups in Chapter six experiments four hours after Galactose induction

Cells grown on filters on Galactose plates was rehydrated using 2ml of broth with vortexing and 100 ul was plated onto the appropriate solid media using glass beads. Solid media was incubated for four days at 30°C, at which point colonies were counted.

Table E.1. Number of colonies present from six individual filters for all groups.

[*rnq*$^-$] (W)	[*rnq*$^-$] with overexpression of *Rnq*1 (X)	[*RNQ*$^+$] (Y)	[*RNQ*$^+$] with overexpression of *Rnq*1 (Z)
Confluent	Confluent	227	36
Confluent	Confluent	202	45
Confluent	Confluent	212	55
Confluent	Confluent	196	60
Confluent	Confluent	222	51
Confluent	Confluent	231	42

The number of cells present in [*RNQ*$^+$] samples and the number of cells in [*RNQ*$^+$] with overexpression of *Rnq*1 samples reveals a 77.6% reduction in colonies as opposed to the 25% reported by Douglas et al. (2008). This may well have been due to the underlying and unknown strength of [*RNQ*$^+$] used here.

Appendix F

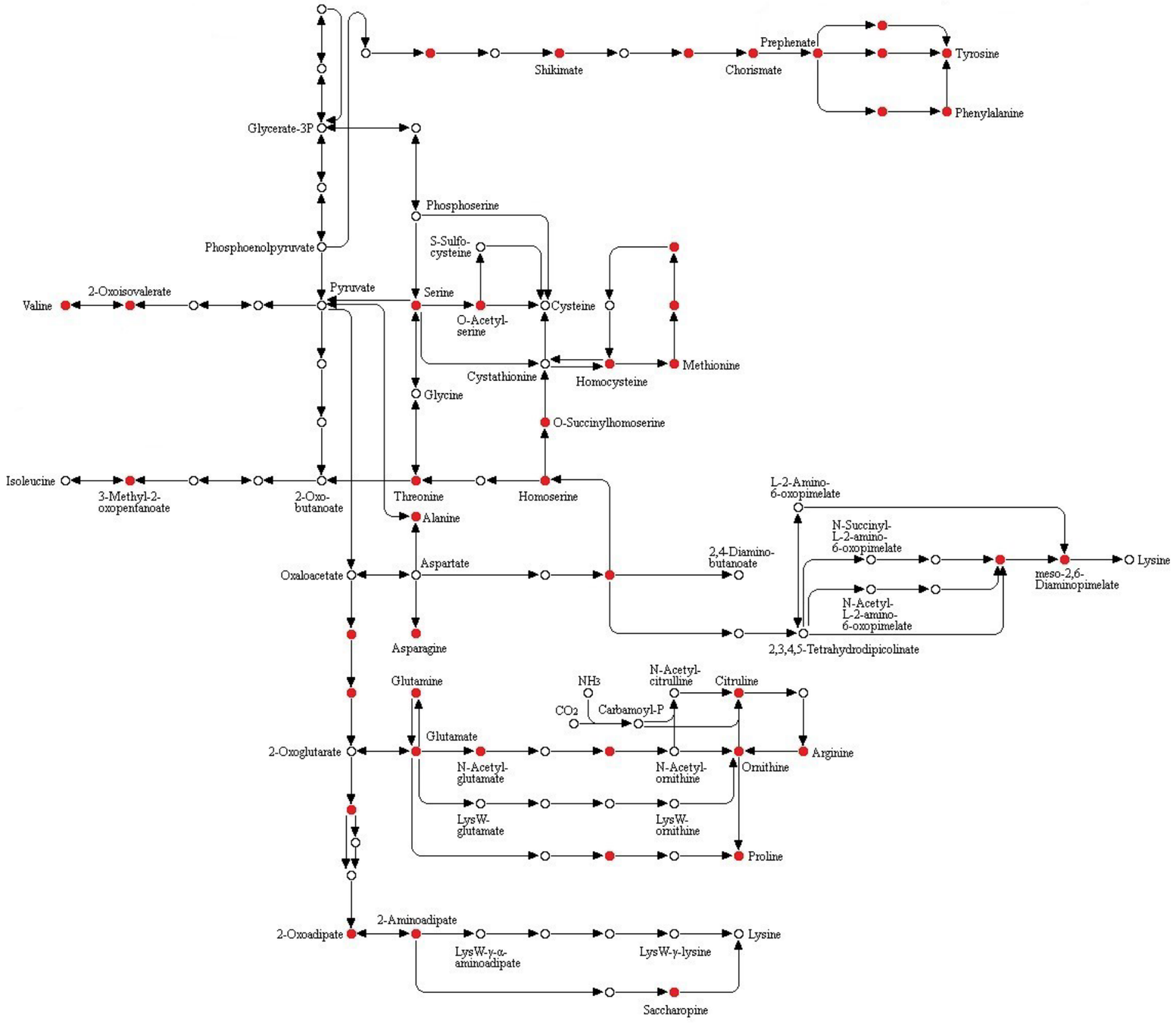


Figure F.1. KEGG Metabolic pathway of amino acid biosynthesis. Red indicates metabolites detected within samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method and acetonitrile extraction protocol.

A)

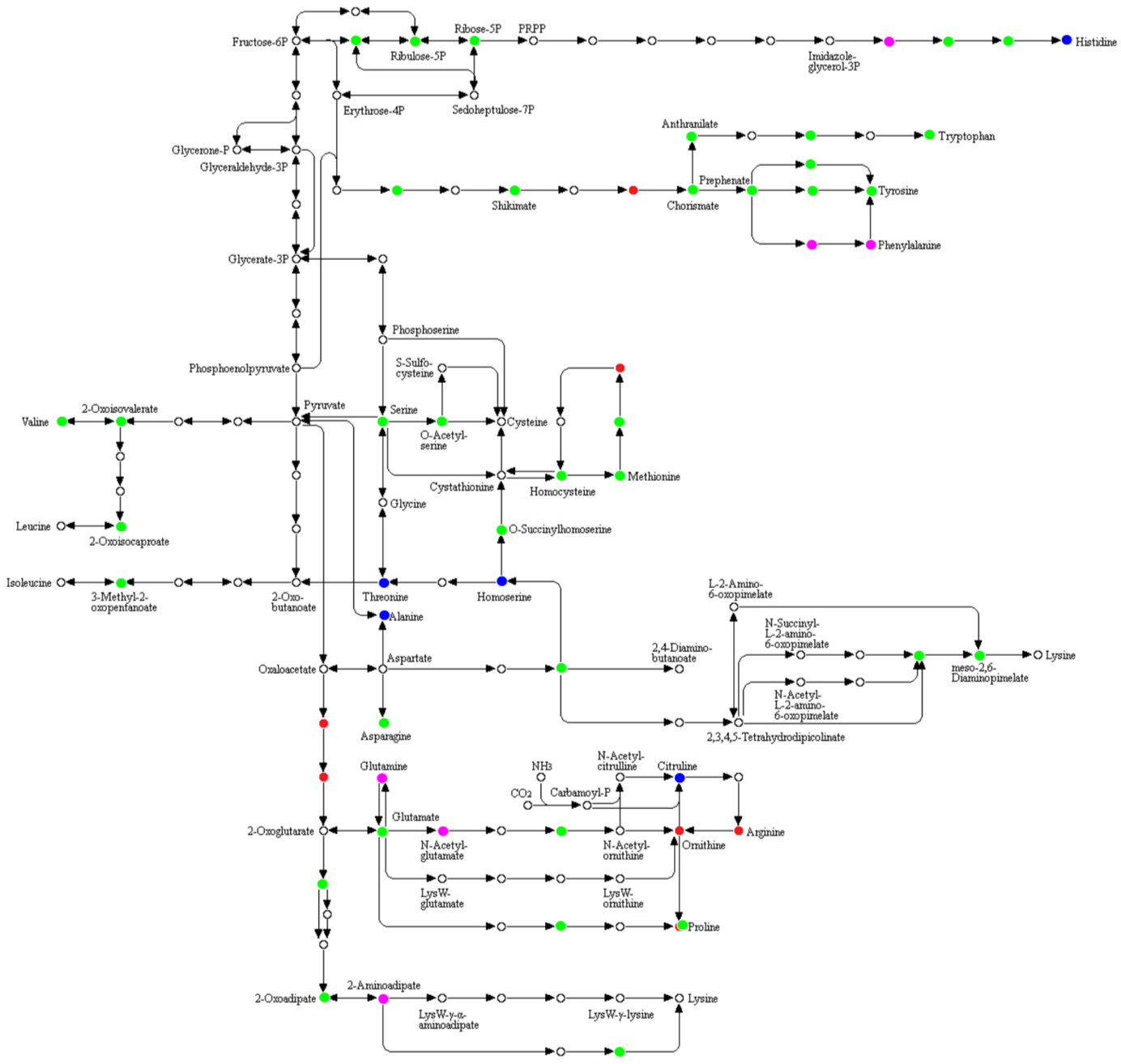


Figure F.2. Shows KEGG Metabolic pathway of amino acid biosynthesis, overlaid with the metabolic hits for samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method and A) 50:50 Acetonitrile/water extraction. Green circles indicate metabolites hit by all pathways, Fuchsia circles indicate metabolites hit by three methods, Blue circles indicate metabolites hit by two methods and Red circles indicate metabolites detected by only one method.

B)

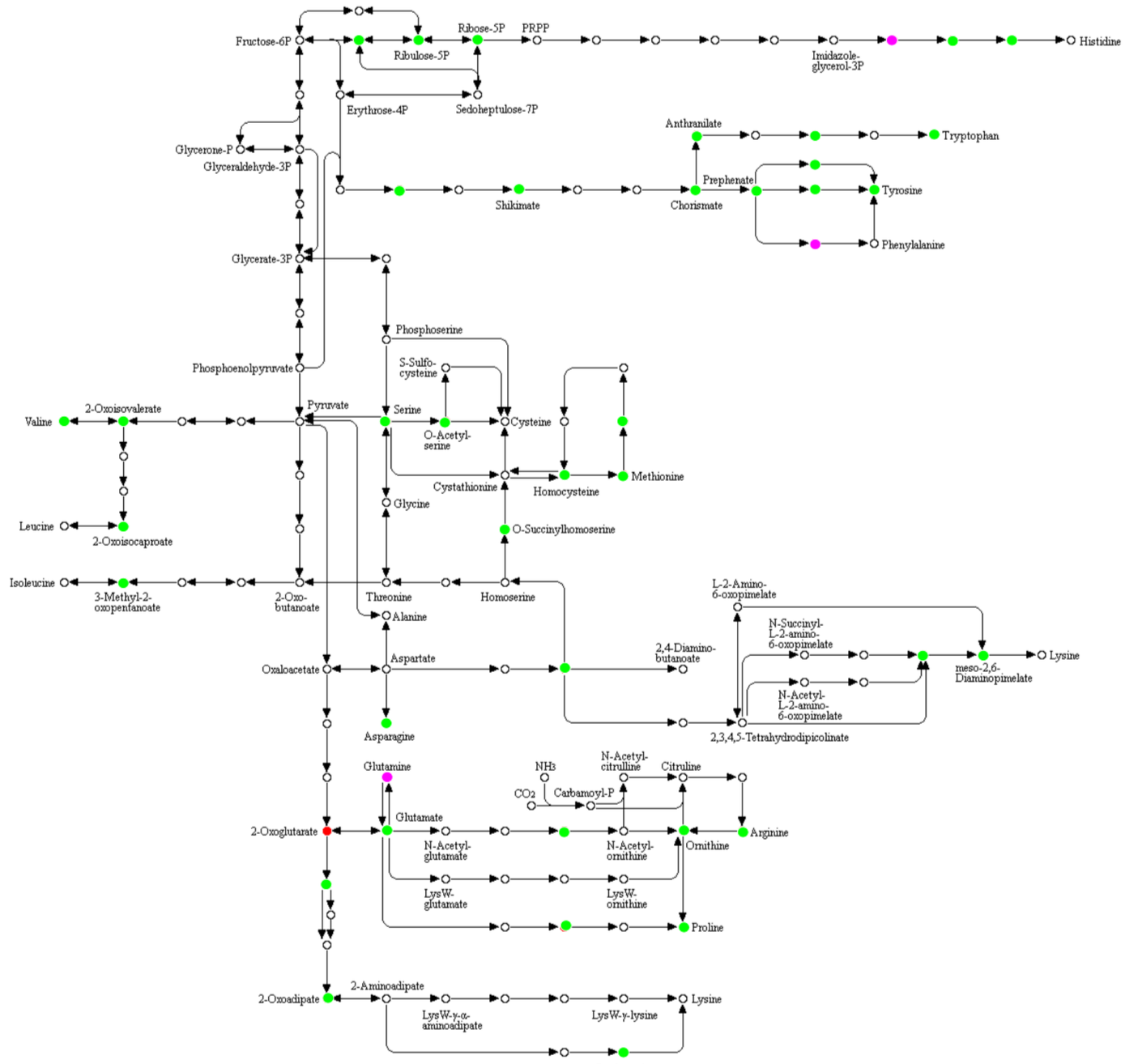


Figure F.2. (Continued) Shows KEGG Metabolic pathway of amino acid biosynthesis, overlaid with the metabolic hits for samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method and B) Boiling ethanol extraction. Green circles indicate metabolites hit by all pathways, Fuchsia circles indicate metabolites hit by three methods, Blue circles indicate metabolites hit by two methods and Red circles indicate metabolites detected by only one method.

C)

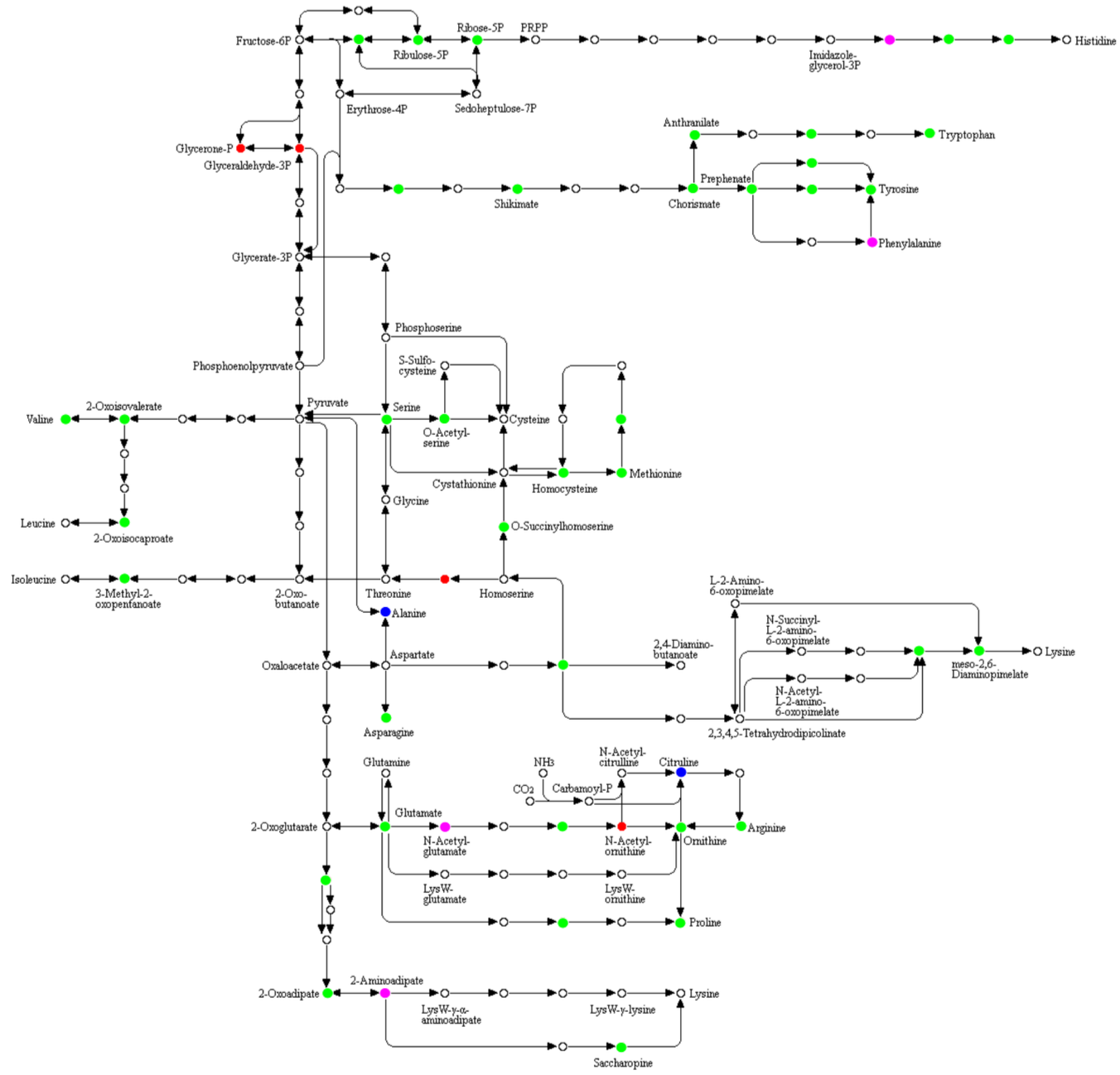


Figure F.2. (Continued) Shows KEGG Metabolic pathway of amino acid biosynthesis, overlaid with the metabolic hits for samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method and C) Chloroform/methanol extraction. Green circles indicate metabolites hit by all pathways, Fuchsia circles indicate metabolites hit by three methods, Blue circles indicate metabolites hit by two methods and Red circles indicate metabolites detected by only one method.

D)

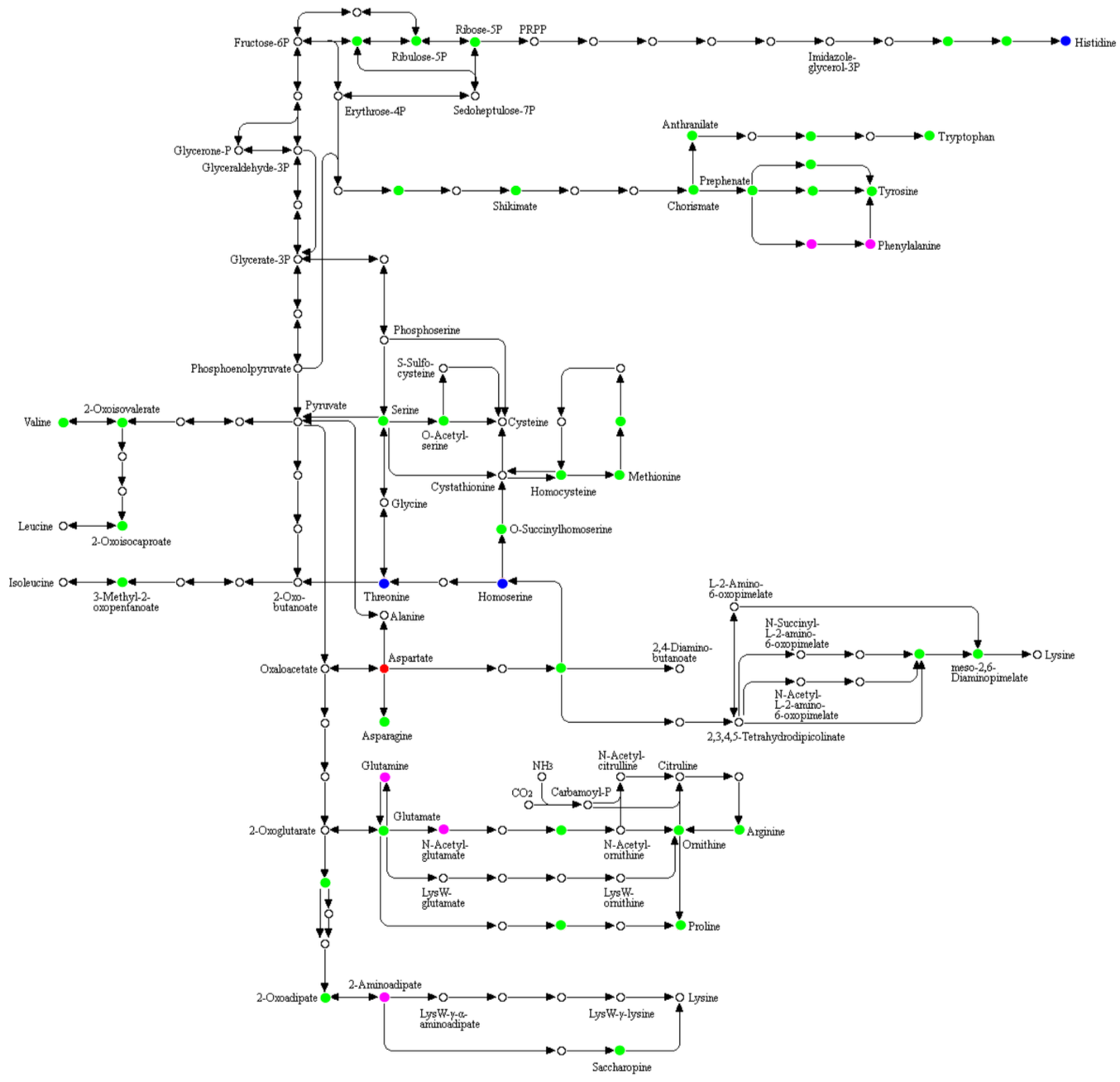


Figure F.2. (Continued) Shows KEGG Metabolic pathway of amino acid biosynthesis, overlaid with the metabolic hits for samples grown using growth condition three, ammonium as a sole nitrogen source, when applying a broth cell culture method and D) Pure Methanol extraction. Green circles indicate metabolites hit by all pathways, Fuchsia circles indicate metabolites hit by three methods, Blue circles indicate metabolites hit by two methods and Red circles indicate metabolites detected by only one method.

Appendix G

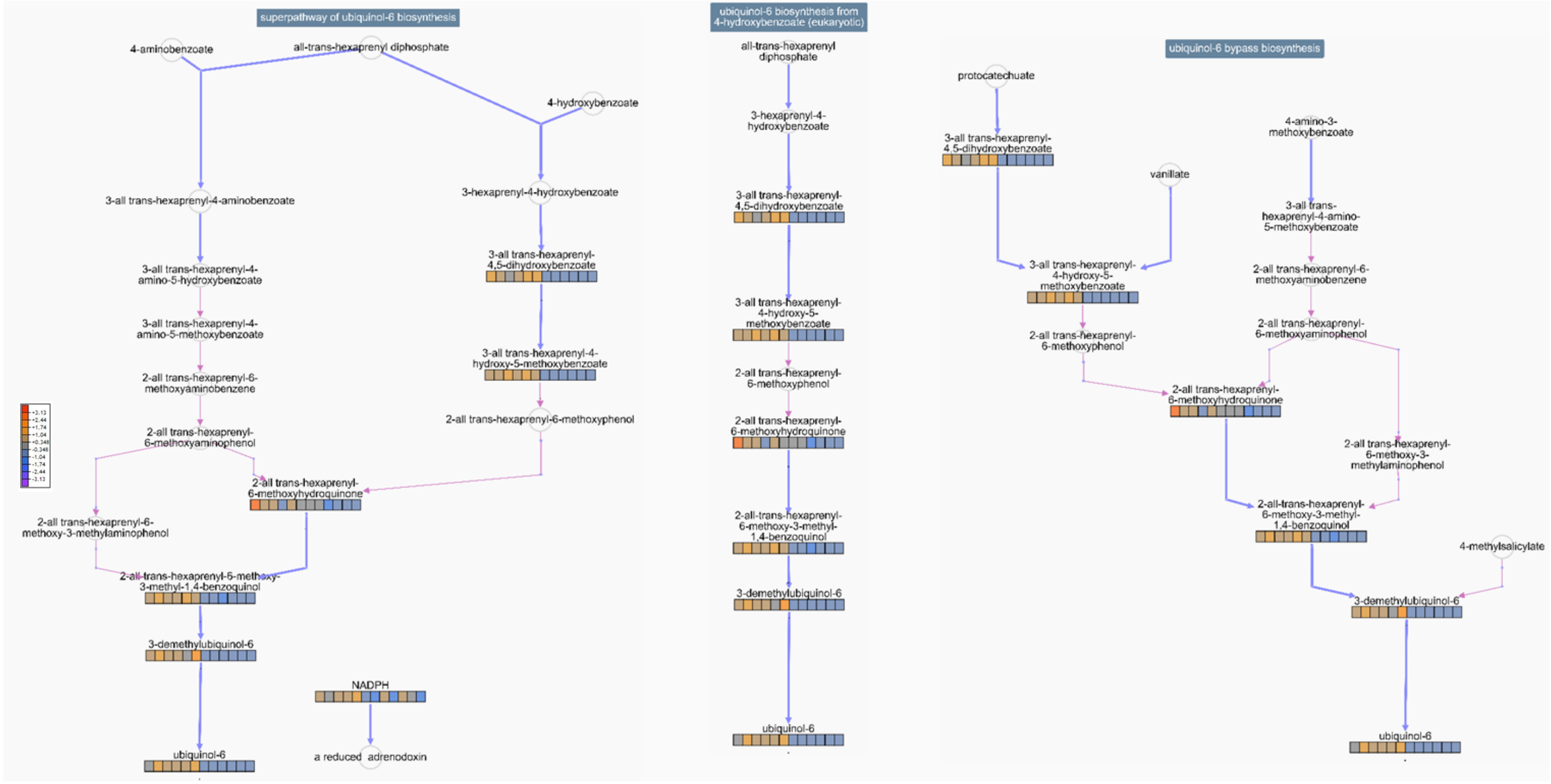


Figure G.1. Pathway collage of three of the top 10 implicated pathways. (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples

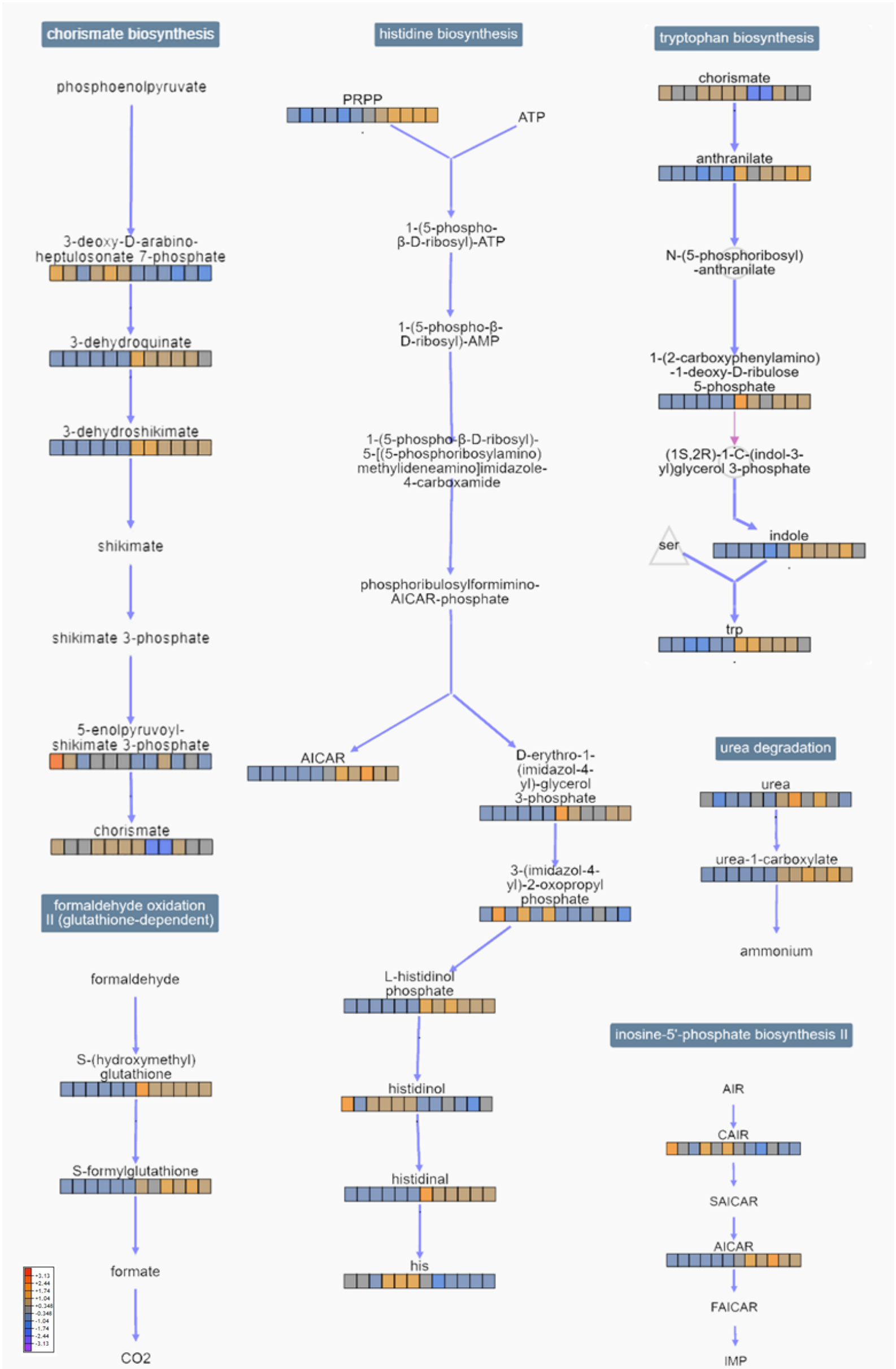


Figure G.2. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples.

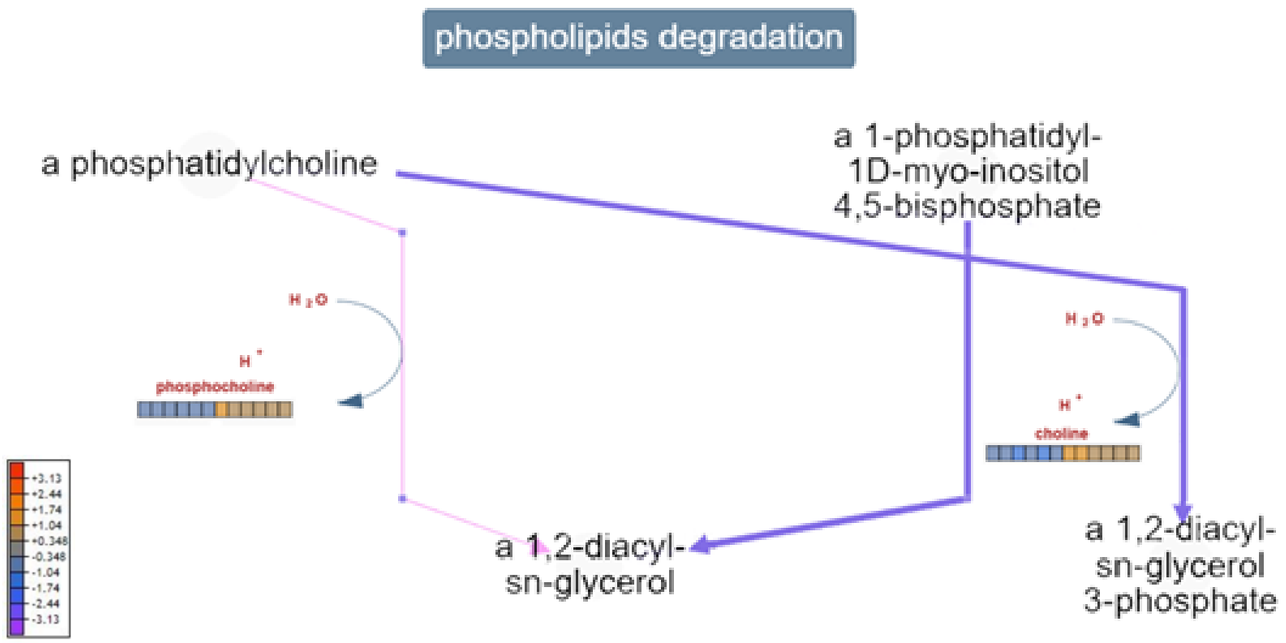


Figure G.3. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the middle left to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples.

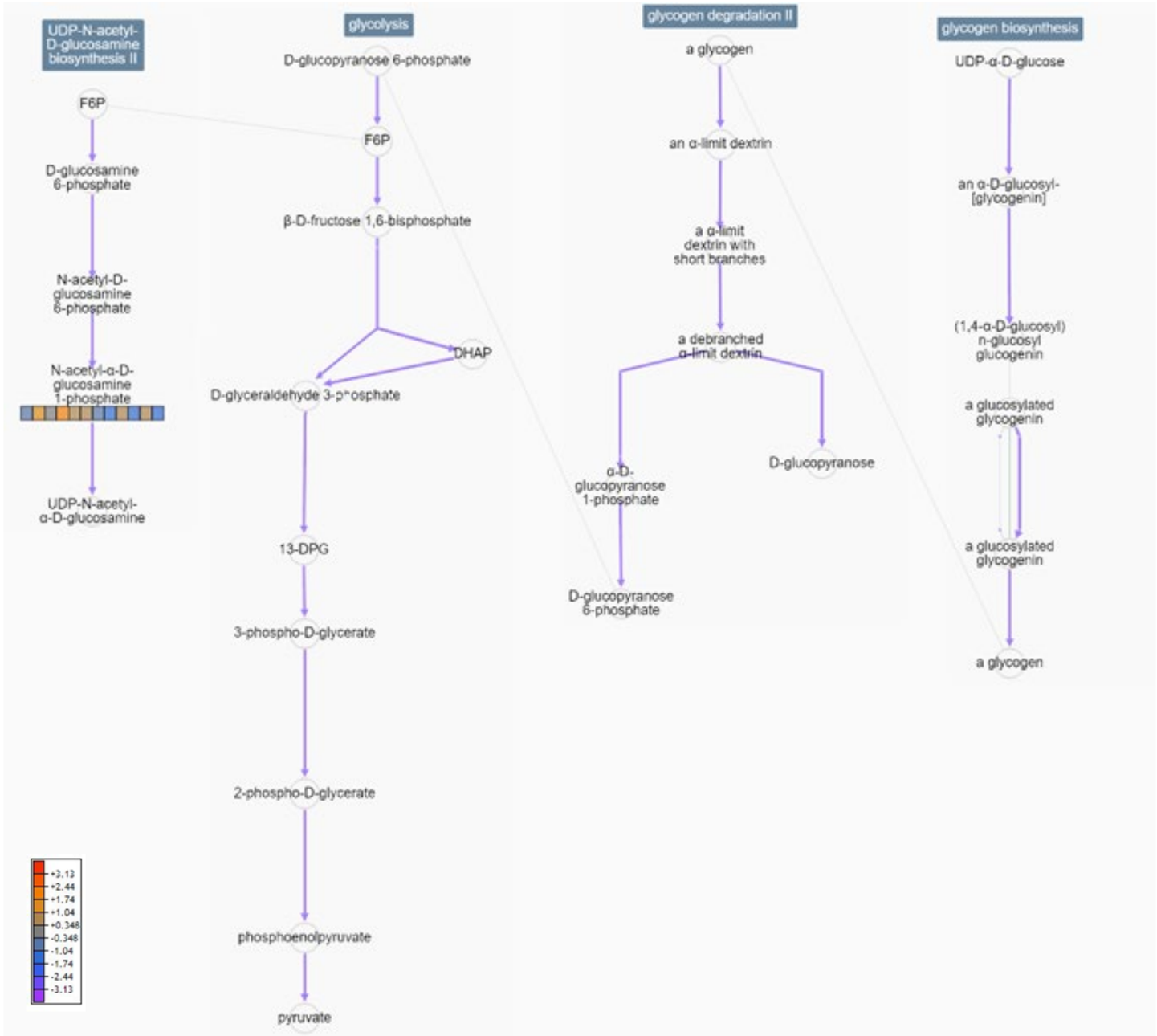


Figure G.4. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples.

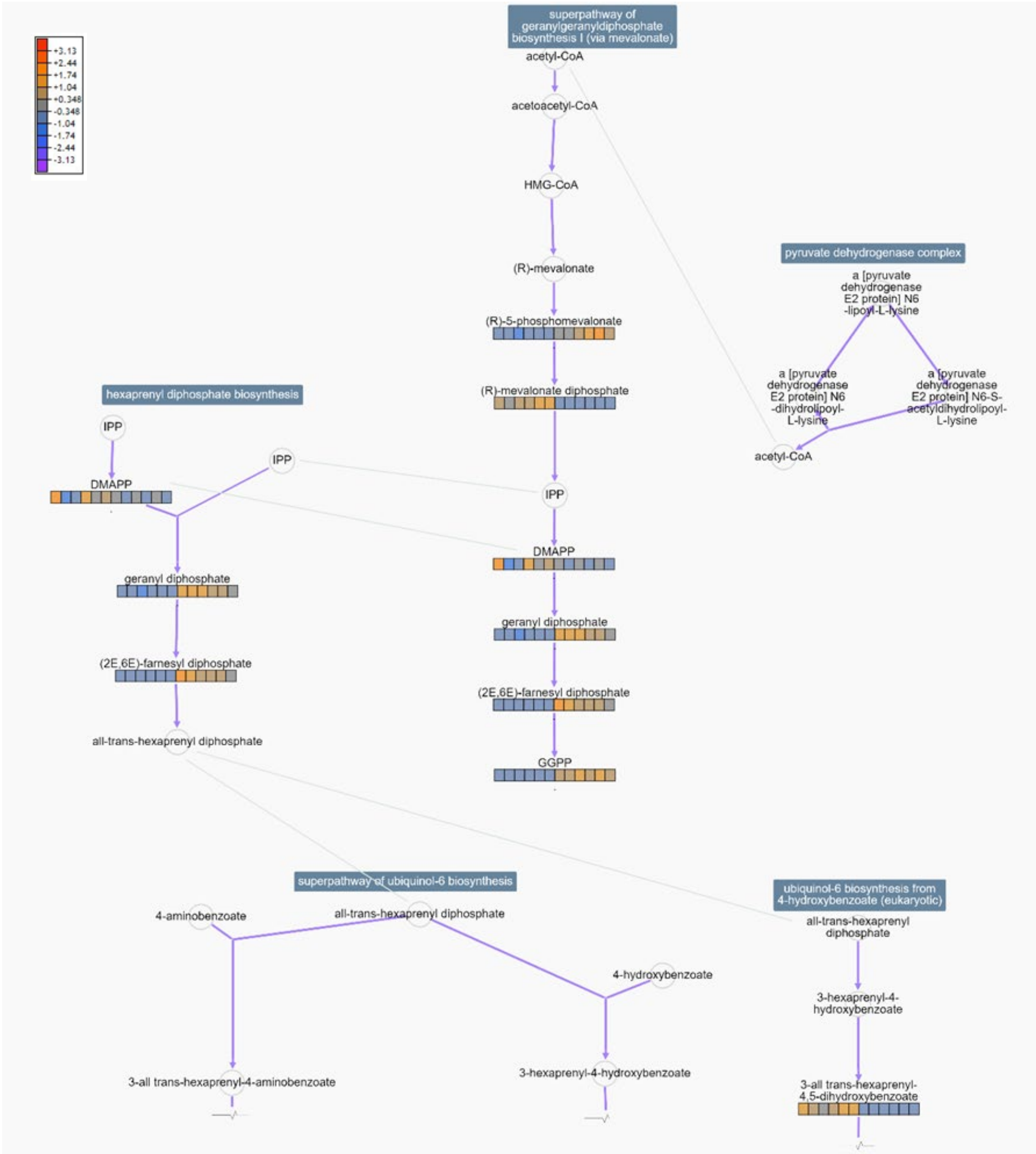


Figure G.5. Pathway collage of three of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^{-}$] samples.

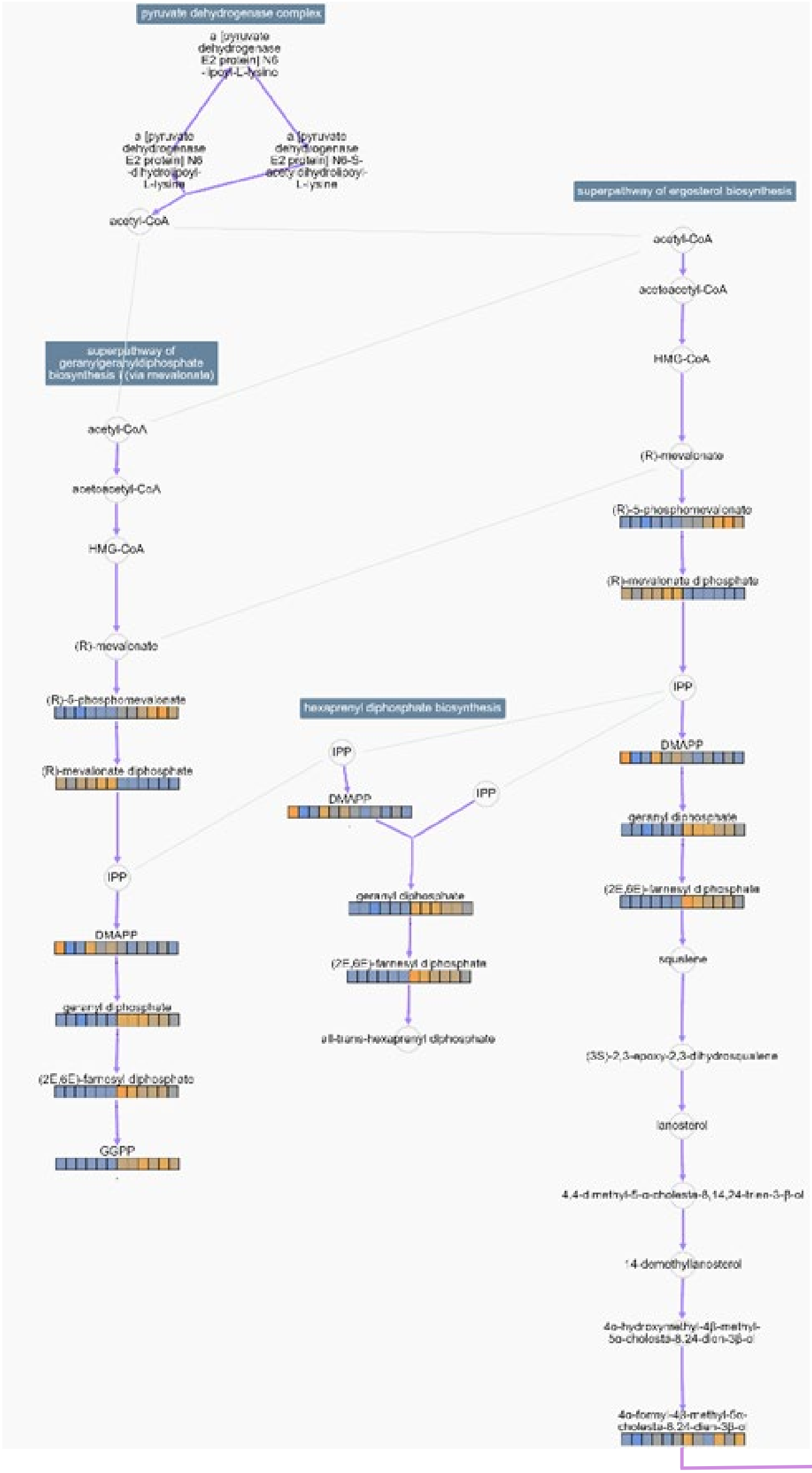


Figure G.6. Pathway collage of metabolic implications upstream from three of the top 10 implicated pathways with standardised omics data of individual metabolites overlaid. A colour legend is shown to the centre left of the figure to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, pink arrows depict spontaneous reactions, faded grey circles show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the last six boxes belong to [*rnq*$^-$] samples.

.

Appendix H

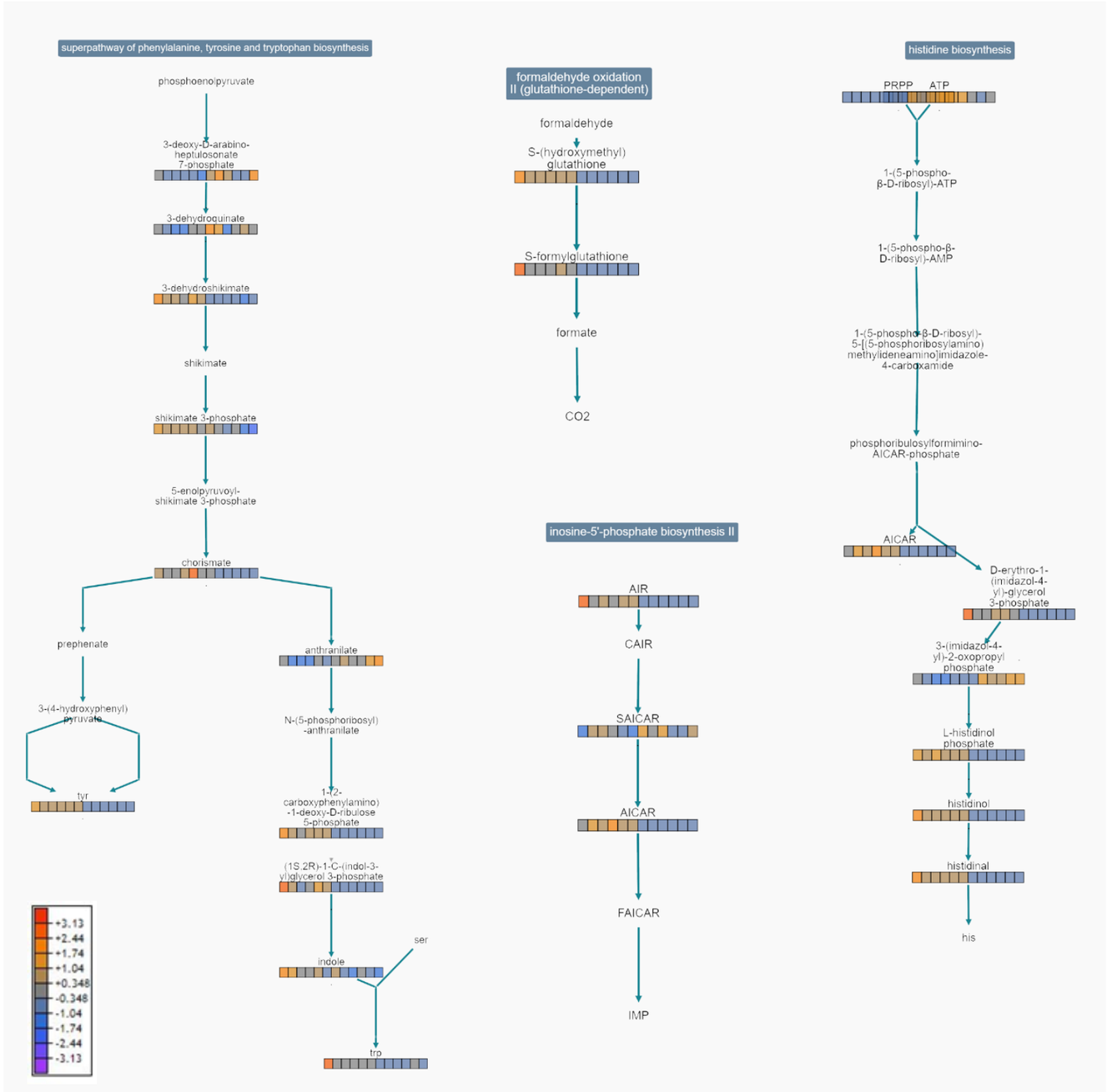


Figure H.1. Pathway collage of 5 of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data where available. Reading from left to right, the first six boxes in any 'heat block' belong to [*rnq*$^-$] samples and the last six boxes belong to [*rnq*$^-$] with mild oxidative stress samples.

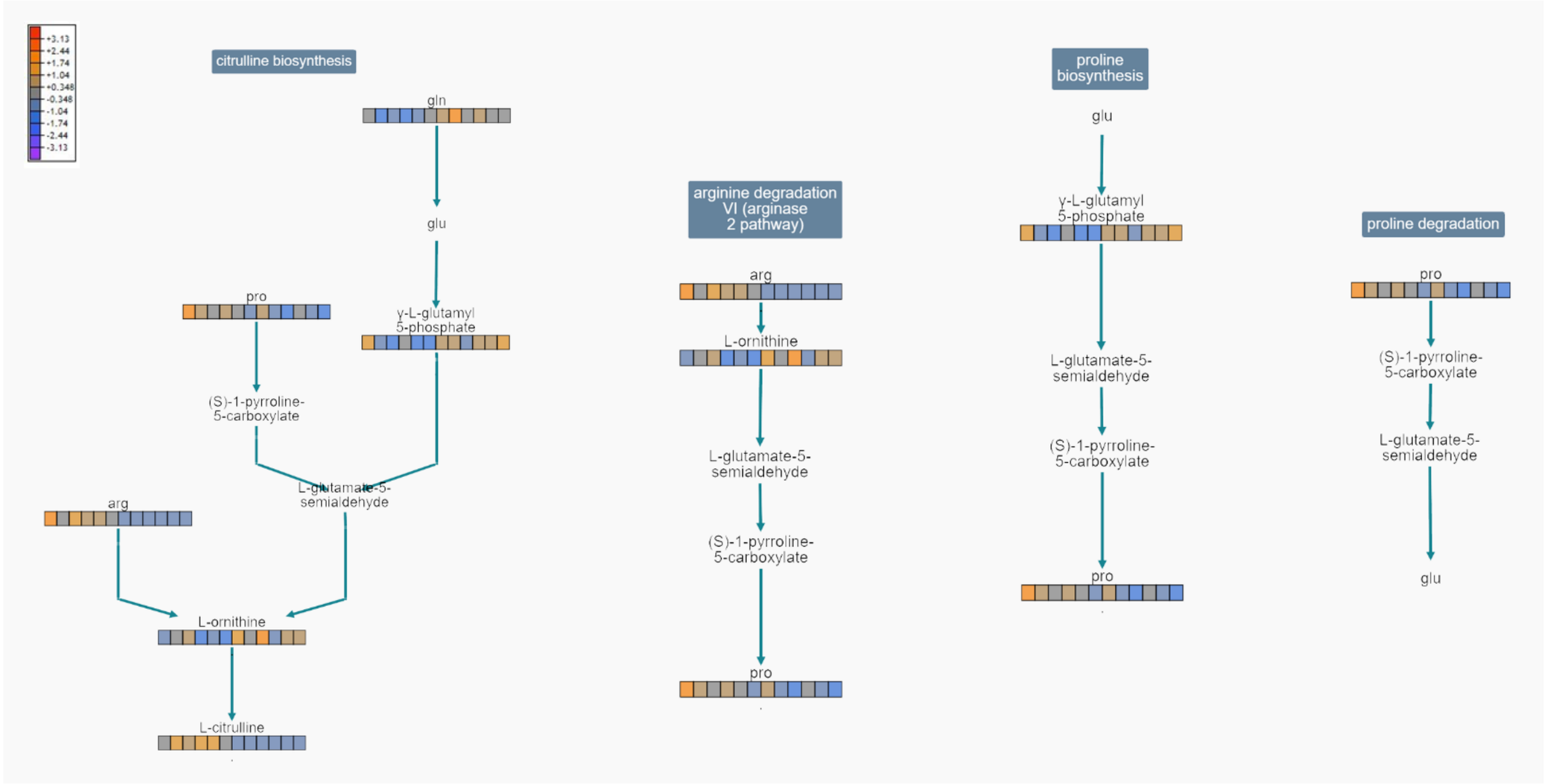


Figure H.2. Pathway collage of 4 of the top 10 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the top left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data where available. Reading from left to right, the first six boxes in any 'heat block' belong to [*rnq*$^-$] samples and the last six boxes belong to [*rnq*$^-$] with mild oxidative stress samples.

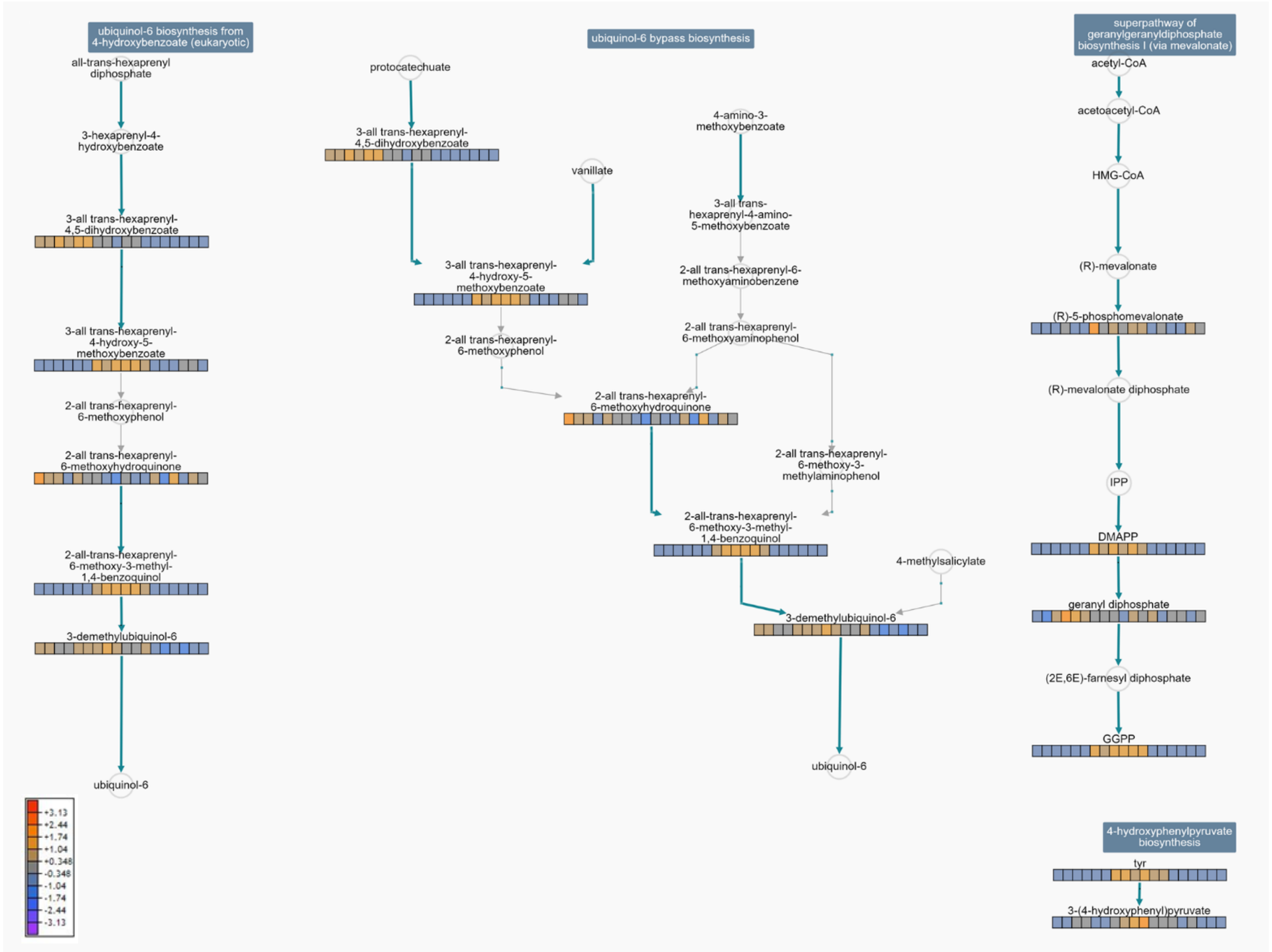


Figure H.3. Pathway collage of three of the top 10 implicated pathways in the loss of function effect (via BioCycs cellular overview) overlaid with *Δrnq1*, [*rnq*$^-$] and [*RNQ*$^+$] data. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples and the middle six boxes belong to [*rnq*$^-$] samples and the last six boxes to [*RNQ*$^+$] samples.

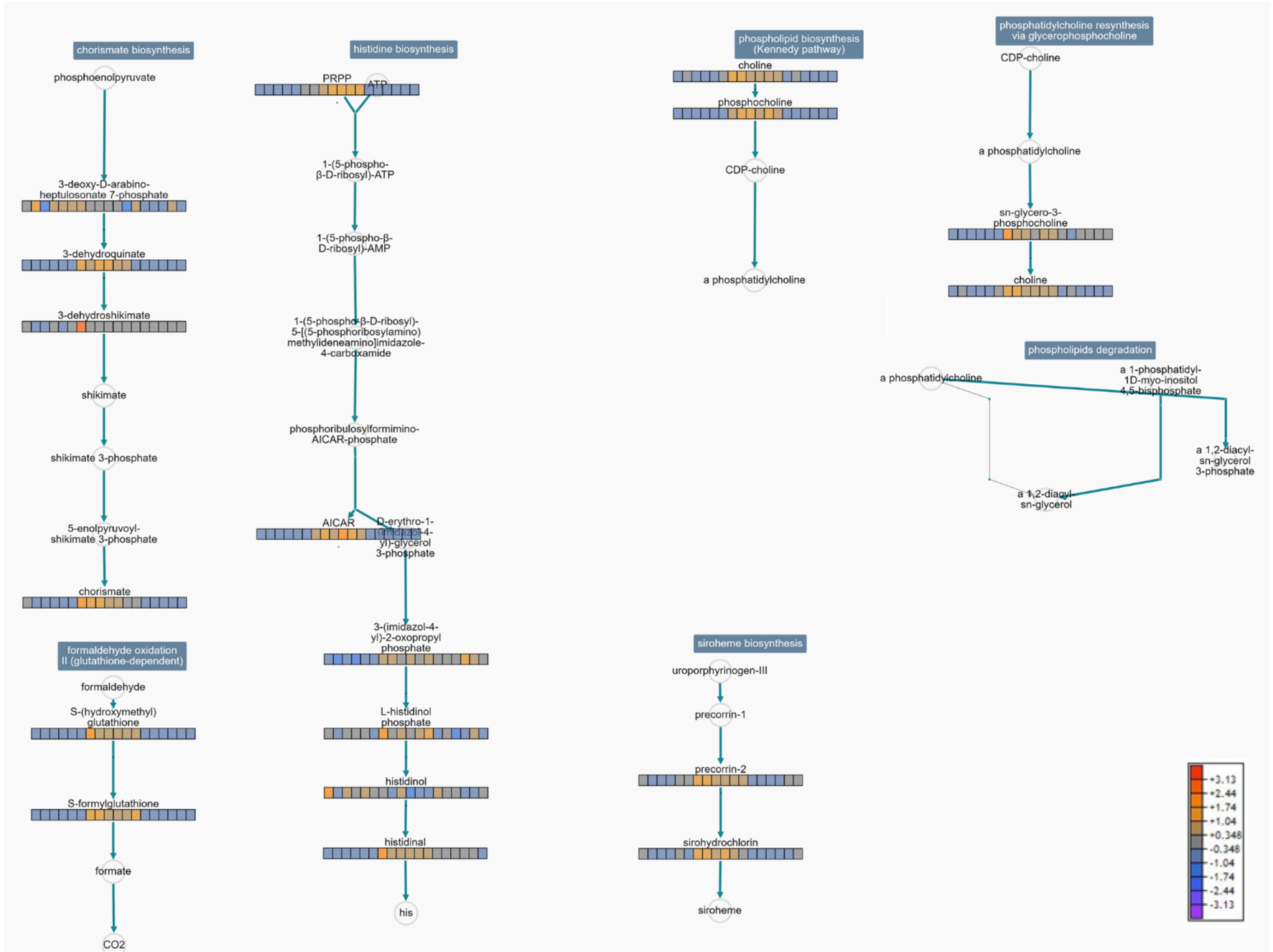


Figure H.4. Pathway collage of seven of the top 10 implicated pathways in the loss of function effect (via BioCycs cellular overview) overlaid with *Δrnq1*, [*rnq*$^{-}$] and [*RNQ*$^{+}$] data. A colour legend is shown in the bottom right-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to *Δrnq1* samples, the middle six boxes belong to [*rnq*$^{-}$] samples and the last six boxes to [*RNQ*$^{+}$] samples.

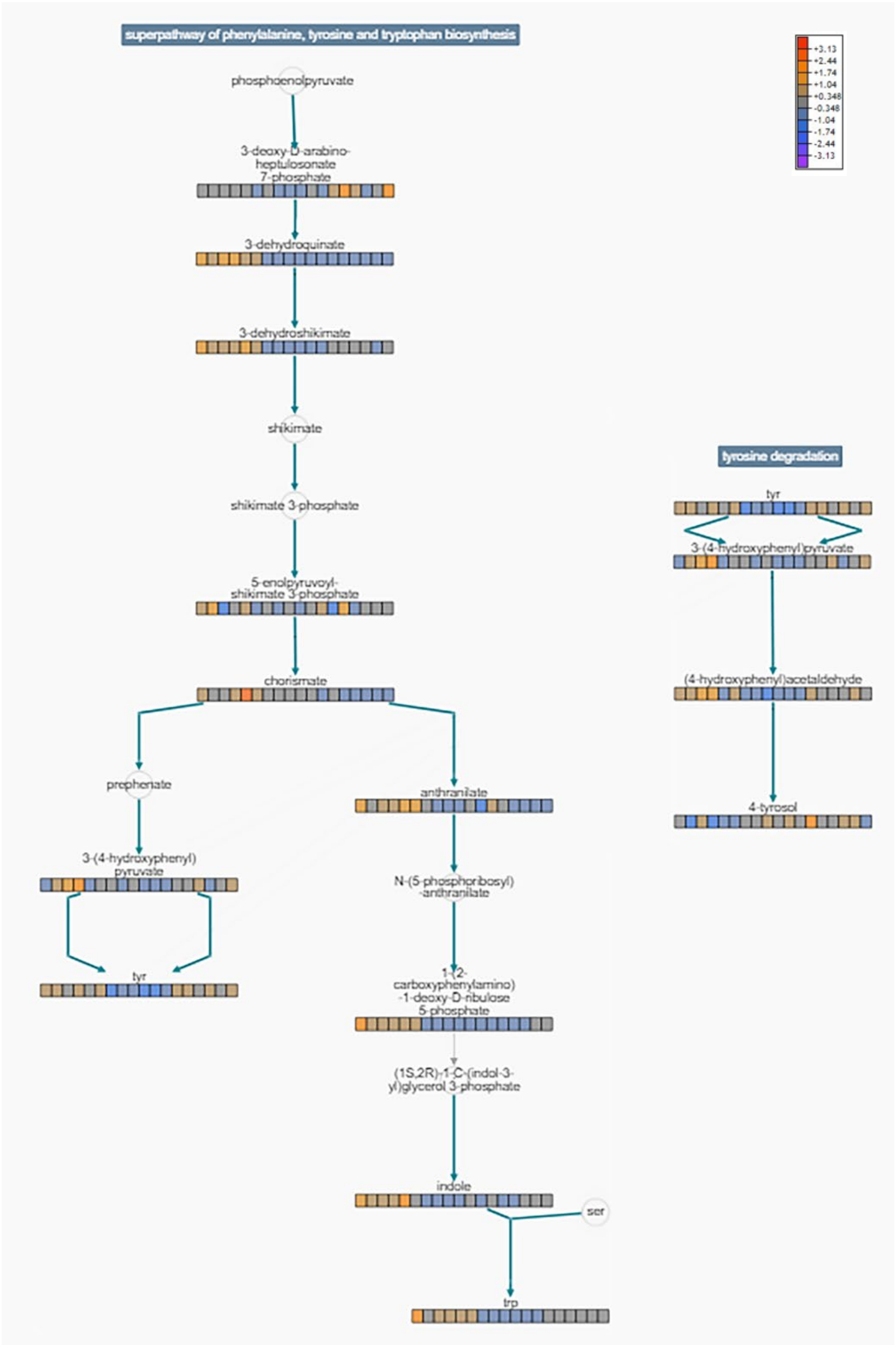


Figure H.5. Pathway collage of four of the top 10 implicated pathways in general stress response (via BioCycs cellular overview) overlaid with [*rnq*$^-$], [*RNQ*$^+$], [*rnq*$^-$] with mild oxidative stress data. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to [*rnq*$^-$] samples, the middle six boxes belong to [*RNQ*$^+$] samples and the last six boxes to [*rnq*$^-$] with a mild oxidative stress samples..

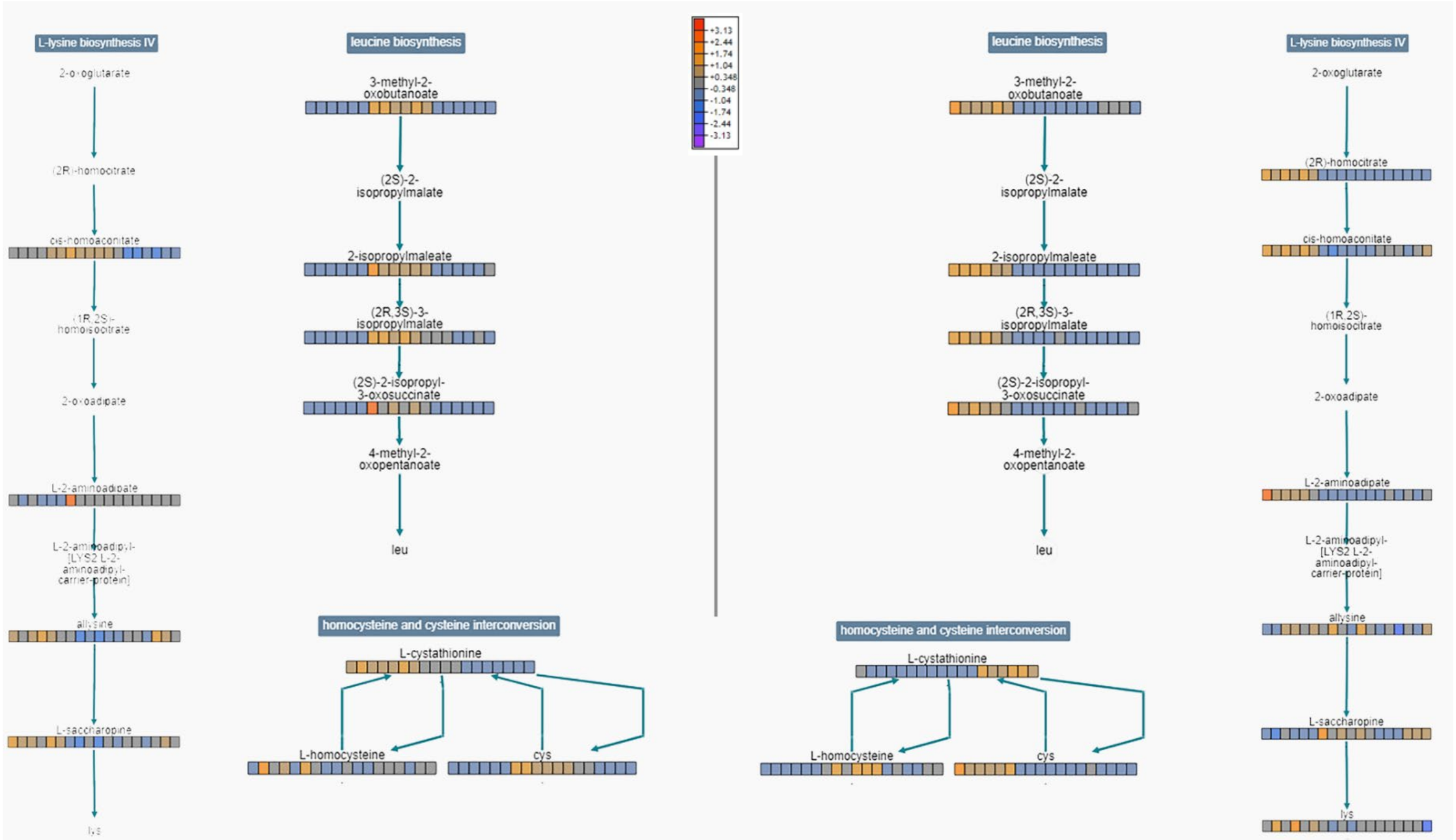


Figure H.6. Pathway collage of three of the top 20 implicated pathways in metabolic comparisons of [*rnq*$^-$] and [*RNQ*$^+$] (via BioCycs cellular overview) overlaid on the left with *Δrnq1*, [*rnq*$^-$] and [*RNQ*$^+$] and on the right with [*rnq*$^-$], [*rnq*$^-$] with mild oxidative stress and [*RNQ*$^+$] data. A colour legend is shown in the top middle to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' left of the page belong to *Δrnq1* samples and the middle six boxes belong to [*rnq*$^-$] samples and the last six boxes to [*RNQ*$^+$] samples. The first six boxes in any 'heat block' right of the page belong to [*rnq*$^-$] samples and the middle six boxes belong to [*rnq*$^-$] with a mild oxidative stress samples and the last six boxes to [*RNQ*$^+$] samples.

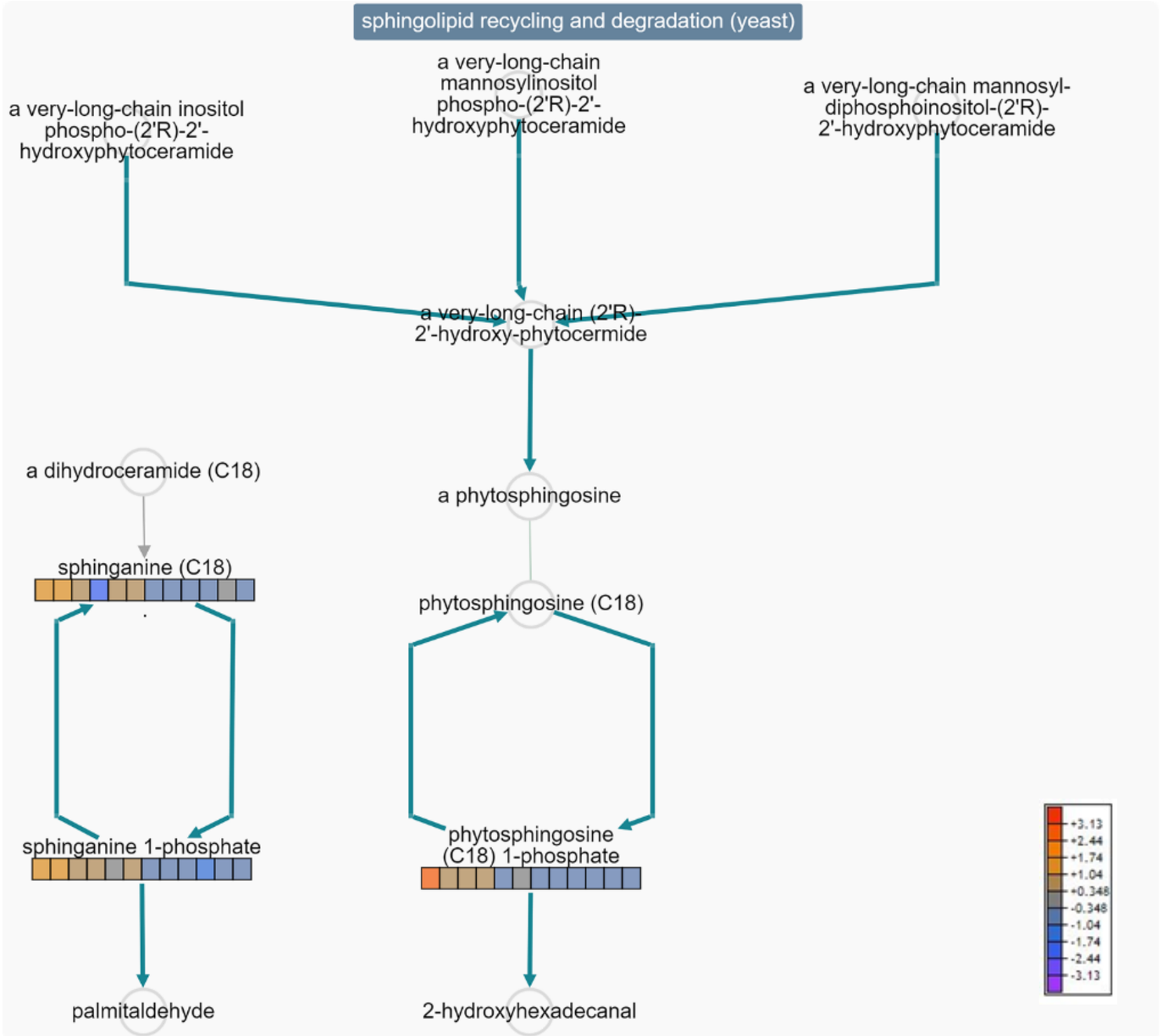


Figure H. 7. Pathway collage of sphingolipid recycling and degradation in yeast (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the bottom left-hand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to [*rnq*$^-$] samples and the last six boxes belong to [*RNQ*$^+$] samples.

Appendix I

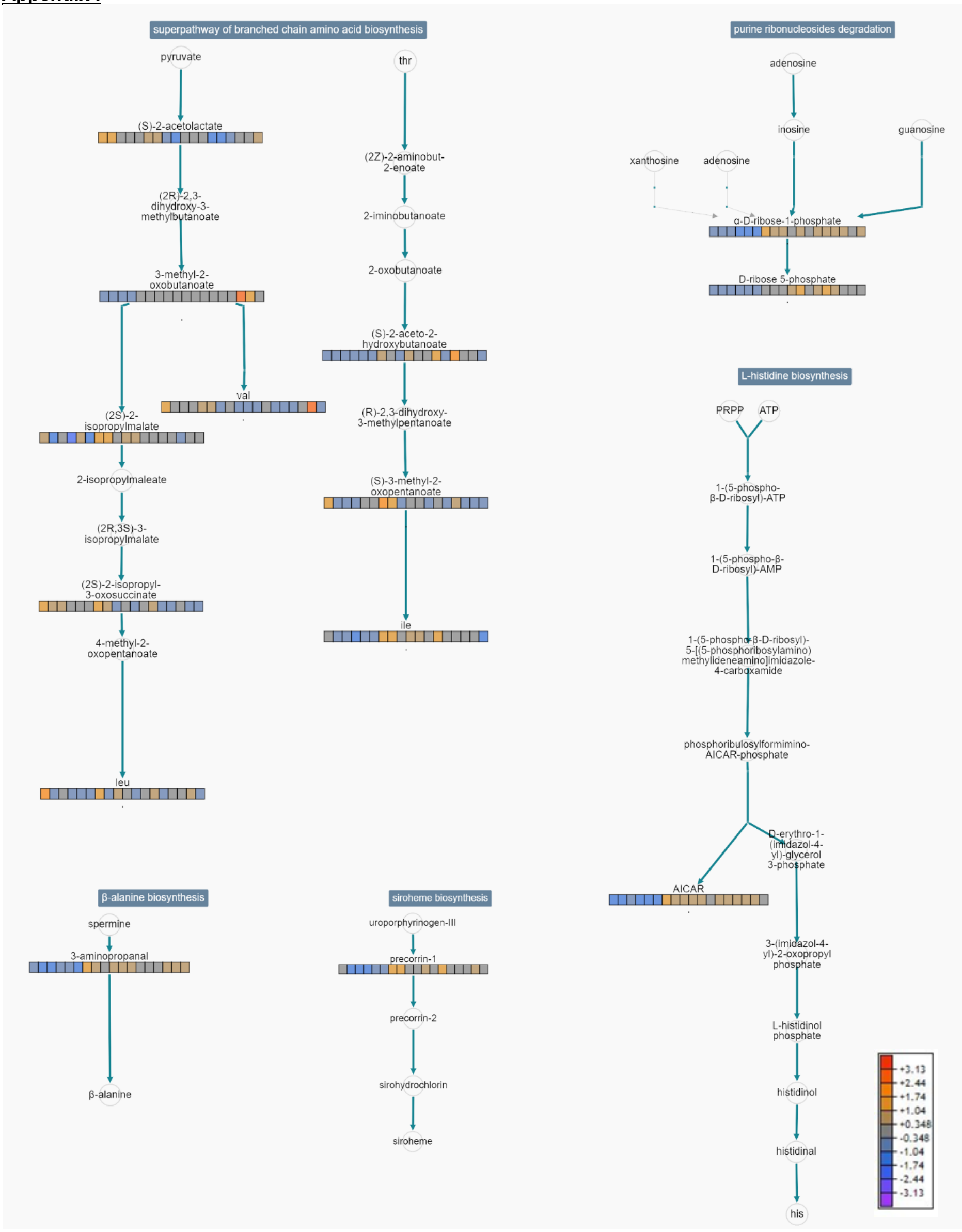


Figure I.1. Pathway collage of four of the top 20 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the bottom righthand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y.

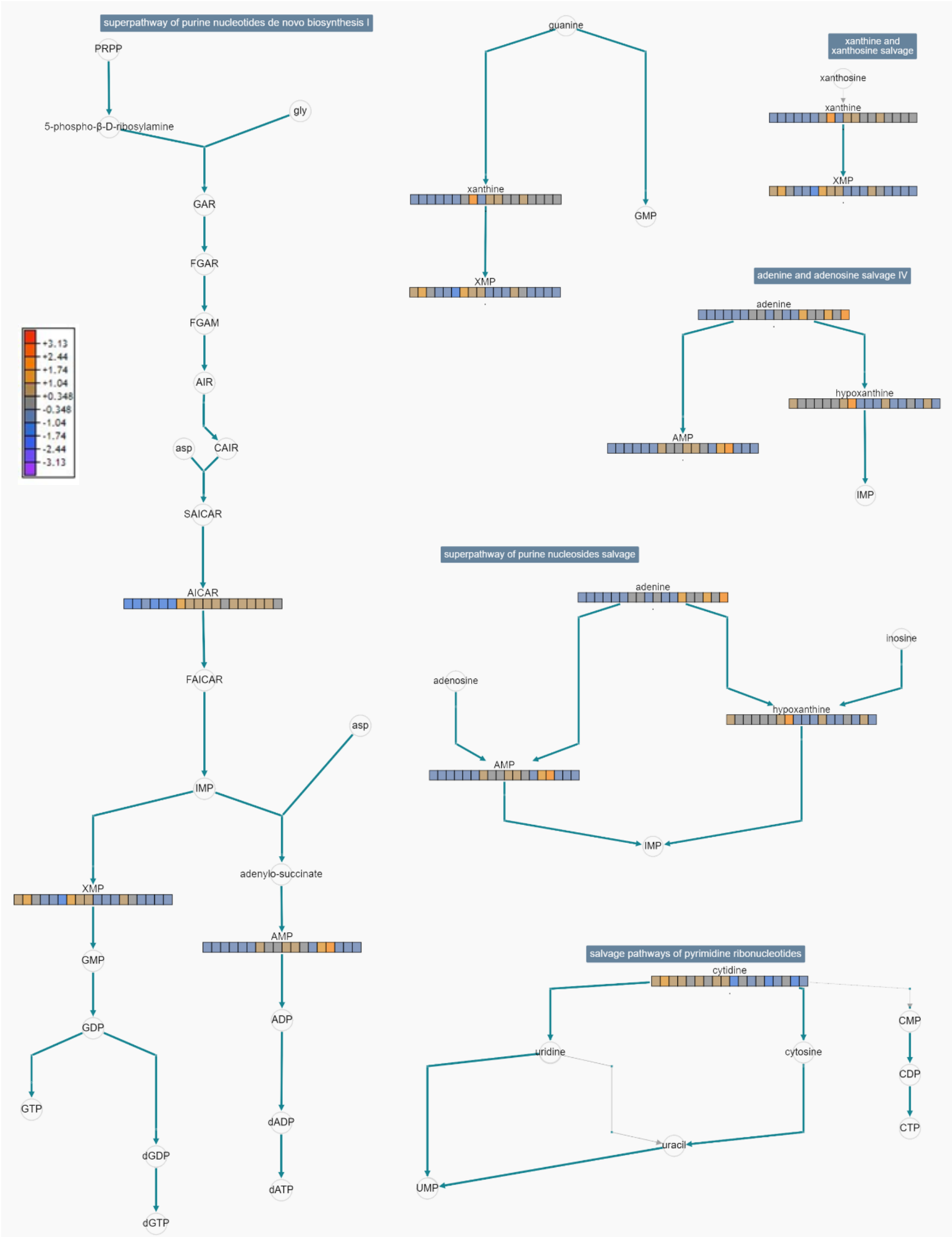


Figure I.2. Pathway collage of five of the top 20 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown on the middle left to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y.

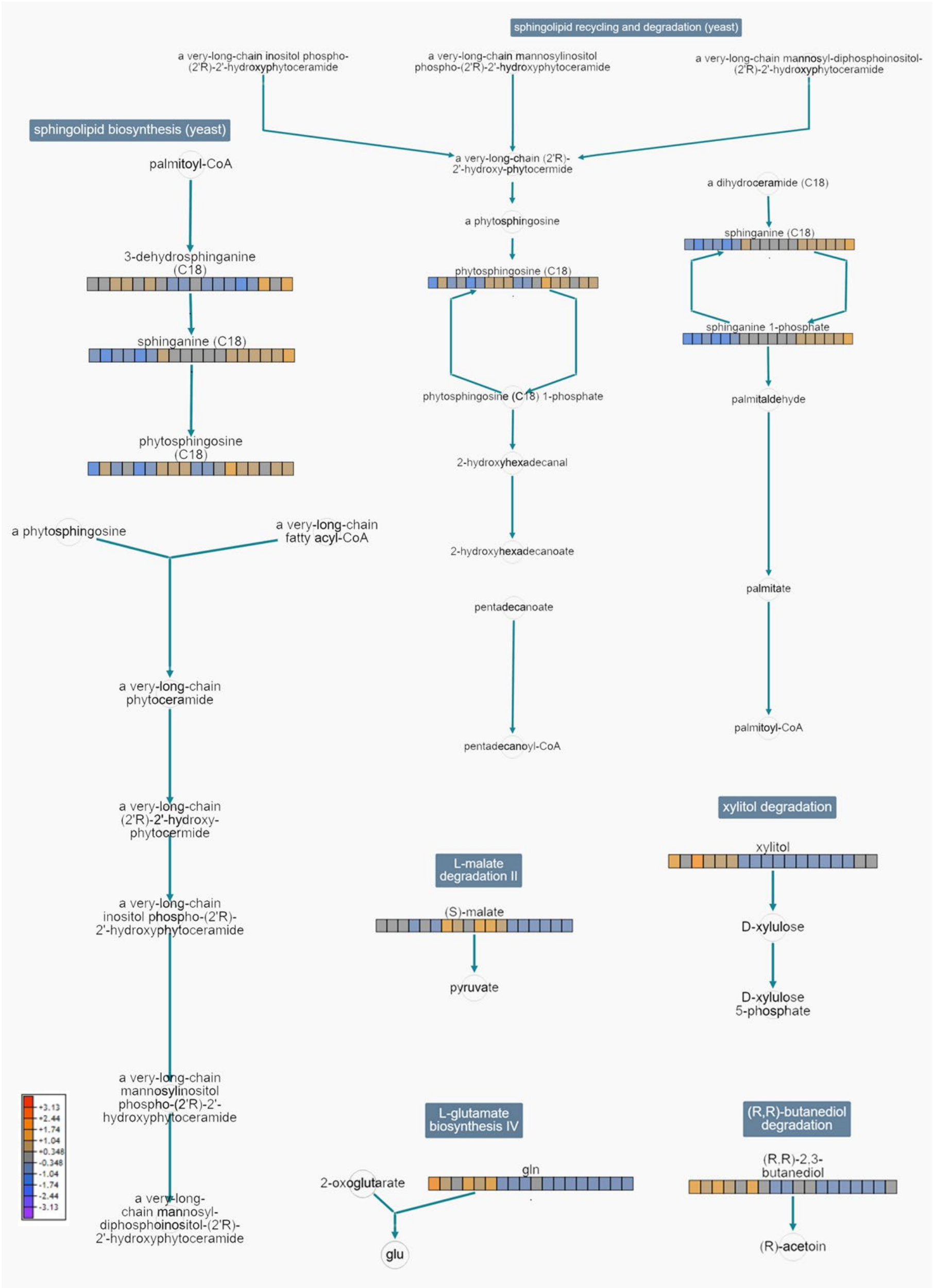


Figure I.3. Pathway collage of six of the top 20 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the bottom lefthand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured 'heat blocks' represent omics data. The first six boxes in any 'heat block' belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y.

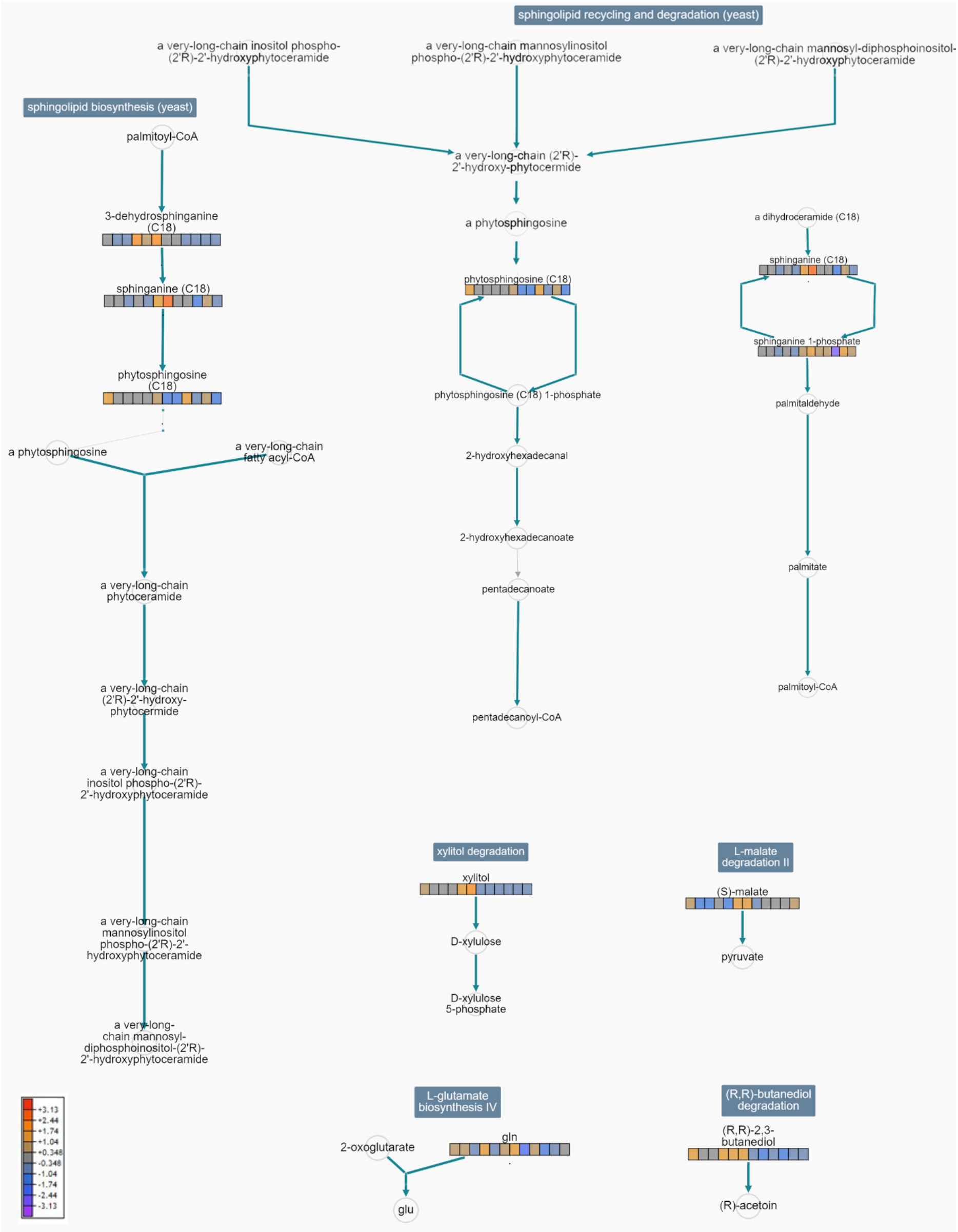


Figure I.4. Pathway collage of four of the top 20 implicated pathways (via BioCycs cellular overview) with standardised omics data of individual metabolites overlaid. A colour legend is shown in the bottom righthand corner to signify the direction of the change indicated by colour. Dark blue labels at the top of the pathway state the pathway name, black writing names individual metabolites, blue arrows depict reactions, grey text show metabolites for which no data is present, coloured ‘heat blocks’ represent omics data. The first six boxes in any ‘heat block’ belong to group W samples, the middle six belong to group X and the last six boxes belong to group Y.