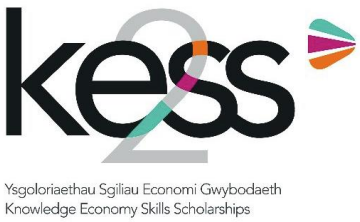
PRIFYSGOL ABERYSTWYTH UNIVERSITY

Investigating Metabolic Changes Associated with Human Oncology

Rachel Stafford BSc (Hons)

Submitted in full candidature for the degree of Masters of Philosophy, Institute of Biological, Environmental and Rural Sciences, Aberystwyth University.

September 2020



ABSTRACT

Cancer is a major global health problem which, although it arises through genetic changes, has consequences which are reflected in cellular metabolism. Metabolomics is a method through which thousands of metabolites may be assessed in a high-through-put mode. Metabolomics was employed to seek out potential biomarkers for diagnosing and staging lung and gastric cancer malignancies. For lung cancer, saliva was selected as the biofluid for investigation as it is non-invasive and few salivary biomarkers have been defined Investigations for potential biomarkers for gastric tumours focused on gastric fluid which has not been explored as a biofluid in metabolomic analyses. In terms of novel biomarkers for the early diagnosis of lung cancer, 43 statistically significant ($P<0.05$) metabolites were identified, of which 40 met the Area Under the Curve (AUC) threshold for sensitivity and specificity (≥ 0.7) to make them of clinical interest. For staging of lung cancers 16 metabolites were found show statistically significant ($P<0.05$, ANOVA) difference between the stages. The gastric fluids could be separated into clusters using multivariate statistics but no biomarkers were found which were both statistically (using ANOVA) and clinically significant (as indicated by AUC). However, in terms of gaining insight into the pathways and mechanisms associated with malignancy we have found metabolites representing significant changes in methylation, inflammation, redox change and amino acid processing. These experiments were all carried out using comparatively small data sets so the studies would need to be investigated using larger cohorts from multiple hospitals and blinded. Then the potential biomarkers found in this study could be tested for their robustness and repeatability, as well as their sensitivity and specificity.

ACKNOWLEDGEMENTS

I wish to thank my supervisor Professor Luis Mur for his expert advice and guidance throughout my project. I also wish to express my heartfelt gratitude to Tenovus for allowing me this opportunity and being overwhelmingly encouraging and supportive throughout this project. I also wish to express my gratitude to KESS, without whom I would not have been able to do this project and lastly to my family and friends for their continuing support and motivation.

KESS Funding Statement

Knowledge Economy Skills Scholarships (KESS 2) is a pan-Wales higher level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for West Wales and the Valleys.

Contents

LIST OF FIGURES

CHAPTER 3 | Results

Saliva for Lung Cancer Diagnosis

Saliva for Lung Cancer Staging

Gastric Fluid for Gastric Cancer Diagnosis

LIST OF TABLES

CHAPTER 1 | Introduction

CHAPTER 3 | Results

CHAPTER 1 | Introduction

1.1 | Metabolomics in Clinical Research

There have been many significant technological advances in the biological sciences over the past few decades bringing us into a new era of research, including the emerging field of systems biology (Dettmer, *et al.*, 2007). Kirschner (2005) suggests that systems biology offers the opportunity to study how the phenotype is generated from the genotype and a small glimpse of how evolution crafted the phenotype. There are many types of "Omics" technologies, such as metabolomics, genomics and proteomics. These methods allow for many different types of biological samples to be analysed, such as sputum, blood, saliva, serum and urine (Wheelock, et al., 2013).

Metabolomics is described as the comprehensive assessment of endogenous metabolites in attempts to identify and quantify the metabolites in a biological sample (Zang, *et al.*, 2013). This offers a unique top-down approach for studying complex biological systems (Kosmides, *et al.*, 2014). A metabolite itself is defined as a product or intermediate of cellular metabolism (Teunissen, *et al.*, 2018). The concepts of metabolomics have been documented in scientific literature since the 1970s, with the word itself being a hybridisation of the words 'metabolites' and 'genomics' reflecting the shift into the new era of biological research (Nikolau, 2007). Metabolites are the end products of regulatory cellular processes, and their levels co-respond to the biological systems response to genetic or environmental changes (Fiehn, 2002). Metabolites have been described as "broadcasting signals from the genetic architecture and the environment" and to give out a "functional readout of the physiological state" of an organism (Gieger, et al., 2008).

Metabolomics has four critical steps, preparation, extraction, detection and quantification (Falco & Lanzotti, 2018). Sample preparation relies on quick sampling and metabolism quenching to rapidly interrupt the metabolism during the sampling process. Quenching aims to stop the metabolism by inhibiting the enzymes in the sample, to prevent changes in the metabolomic profile during the sampling process. Metabolite extraction then aims to efficiently release the enzymes from the sample, remove interferents, such as salts or proteins, and concentrate the trace metabolites ready for analysis (Alvarez-Sanchez, *et al.*,

2010). There are several analytical approaches which may then be used to detect the metabolites, chromatographic methods; such as liquid chromatography (LC), high-performance liquid chromatography (HPLC) and gas chromatography (GC), mass spectroscopy methods (MS) or spectroscopy methods; such as nuclear magnetic resonance (NMR), Fourier transform infra-red (FT-IR) and ultraviolet (UV) (Falco & Lanzotti, 2018). Often analytic techniques are linked in so call hyphenated approaches. Thus, for example, with GC-MS, there is a prior separation with GC before entry into the MS. This can extend the number of metabolites that can be detected in a given sample. The quantification stage of metabolomics is critical for understanding biological processes. There are two different strategies for quantification, absolute and relative. Relative quantification is when the metabolite signal is normalised to that of a relative metabolite or an internal standard, typically used in large-scale, non-targeted metabolomics, whereas absolute quantification utilises external standards or internal isotopically labelled standards and is mostly used in targeted metabolomics (Lei, *et al.*, 2011). A major obstacle in the quantification of unknown metabolites is that the metabolites peak intensity not only reflects its concentration but also its chemical structure (Chalcraft, *et al*, 2009 : Lei, *et al.*, 2011).

As well as different analytical techniques, there are also four conceptual techniques commonly used in metabolomics, metabolic fingerprinting, target analysis and metabolite profiling (Fiehn, 2002). Traditionally metabolomics has used complementary analytical methods, such as GC-MS, LC-MS and/or NMR, to determine and quantify as many either identified or unknown metabolites as possible. Metabolic fingerprinting detects is the metabolites in a sample but does not idenitfy them. Thus, the outputs provides a signature of the concentrations of and array of different metabolites present at any given time can be thought of as a metabolic “fingerprint” (Trygg, *et al.*, 2006). Fingerprinting is a non-invasive, rapid way to analyse samples provide a rapid means of characterising phenotypes and making distinctions in metabolic states due to, for example, altering environments (Liang, *et al.*, 2015). Target analysis is the concept of determining and quantifying a small set of known metabolites, known as targets, using the one analytical technique with the best performance for the known compounds. Whereas, metabolite profiling aims to analyse much larger sets of compounds and attempts to identify them (Roessner & Bowne, 2018).

The advancements in metabolomics over the past decade, has led to an ever growing number of biomedical applications, in particular identifying biomarkers for disease diagnosis, understanding disease mechanisms, suggesting novel drug targets, and monitoring treatment outcomes (Wishart, 2016).

One of the many ways metabolomics is being used in human health and disease is in the idea of a new-born screening. Impaired activity of enzymes, cofactors and transporters may lead to the accumulation of metabolites that demonstrate a serious health condition for a neonate. Inherited metabolic disorders are relatively rare, but there are over 500 of them. Early detection of potentially lethal toxic metabolites or lack of an important end product can provide better outcomes and save lives (Ficicioglu, 2017). Metabolite measurement to detect phenylalanine and its metabolites leading to early diagnosis of phenylketonuria was established 5 decades ago. Then ten years ago the screening was extended beyond bacterial inhibition assays, radioimmunoassays and enzyme -immunoassays with the use of GC-MS (Ismail, *et al.*, 2019). One of the biggest problems highlighted in this study has been over diagnosis and the inability to distinguish between those with these harmful conditions and those without, leading to parental anxiety and potential overtreatment. Genetic screening is used as a secondary screening method; however, it can cause unnecessary delays and has cost implications. The accuracy and robustness of findings from targeted and untargeted metabolomics therefore requires further study to fulfil its clinical potential (Ismail, *et al.*, 2019).

Interest in the use of 'omics' technologies for biomarker discovery has been increasing, but there are currently only 21 FDA (US Food and Drug Administration) qualified biomarkers for diagnosis, prognosis or use in monitoring (U.S. Food and Drug Administration, 2020). Metabolomics is utilised to analyse exactly what takes place within an organism exposed to genetic changes, pathophysiological stimuli, and environmental stress (Noreldeen, *et al.*, 2019). A biomarker is a measured characteristic that indicates a normal or abnormal biological process or the effect of an alteration to the process e.g. drug administration (National Institute of Health, 2001). The National Cancer Institute (2018), further defines a biomarker as a "biological molecule" discovered within "blood or bodily fluid" which indicates a condition, disease or process (National Cancer Institute, 2020). There is an ever growing

demand for the identification of more novel biomarkers with the hope of developing more personalised, stratified therapies (Lodi, et al., 2013). An example of a biomarker is troponin which is used to diagnose acute coronary syndrome. Troponin meets the criteria for a good biomarker of enabling biomonitoring during therapy to indicate improvement, showing abnormal physiology, prognostic, used to guide therapy (Wettersen & Maisel, 2015) and is both specific, sensitive and reproducible (Holland, 2016).

Metabolomics is also used in clinical research for identifying drug targets. For decades drug discovery methods have comprised of whole genome sequencing, cloning of identified genes, purification of target proteins and then high throughput screening to identify potential drug targets, which can then be optimised and investigated in animal models with the hope of reaching human trials (Wishart, 2016). These current programmes are lengthy and costly (Hughes, *et al.*, 2011), with only a small percentage making it to phase 1 clinical trials (Bains, 2004) and an even smaller number making it past the stage 3 trials (Mullard, 2014). One of the issues associated with the current method of discovery is that relatively few diseases have a strong genetic basis (Cuatrecasas, 2006), instead they are often the result of exposures (Rappaport, *et al.,* 2014). Metabolomics may offer the solution to some of these problems, offering a far more cost-effective and efficient pathway to drug discovery. As most diseases are the consequence of genetic and environmental interactions focusing only on the genetic aspects is not enough. Metabolomics gives a unique insight into the genetic and environmental contributions to disease development, making of great potential in the identification of novel drug targets (Ma, *et al.*, 2018). Secondarily, patient biomarkers can be used to group disease into subtypes to allow more focused approaches to trial enrolment procedure. This is likely to improve response rates to treatments as more appropriate disease subgroups are likely to be better affected by candidate drugs (Beyoglu & Idle, 2013).

Another problem associated with drug discovery and development is drug toxicity. Drug toxicity is a costly and very challenging problem, primarily caused by the variation is patients' response to the drug (Powers, 2009). One reason for the late detection of severe adverse side effects is that there are very limited injury biomarkers taken from the preclinical trials to the clinical trials. If new and specific translational biomarkers for kidney and liver injury caused by the drug toxicity then it could significantly reduce the time and cost associated with drug

discovery and approval (Beger, *et al.*, 2010). The kidney is one of the most susceptible organs to toxicants as they interfere with its ability to excrete body waste and maintain body fluid levels. Metabolomics has since been applied to nephrotoxicity to attempt to identify novel specific biomarkers for biological changes associated with nephrotoxicity (Zhao & Lin, 2014). Toxicometabolomics is a powerful tool both clinically and in industry (Garcia-Sevillano, *et al.*, 2015), which looks to determine pathways of toxicity as well as toxicity signatures, with the aim of making it possible to predict the toxicity response *in vitro* and *in vivo* of a specific molecular target or drug formulation (Bueno, 2015).

Metabolomics also has the ability to help categorise drugs into classes based on mechanism of action, as shown in the study by Verhoeckx, *et al.,* (2004) where the changes in metabolites caused by anti-inflammatory compounds were analysed. The results showed that different kinds of anti-inflammatory drugs resulted in distinct and characteristic mRNA, protein, and lipid patterns, which can be used to classify both current and future molecules by function. Metabolomics can also be used to analyse pathogen metabolomes looking for biomarkers specific to pathogens with a certain trait, for example antibiotic resistance. Niemiec, *et al.,* (2012) used metabolite profiling of *Streptococcus aureus* to identify molecules that would function as non-competitive inhibitors of pyruvate catabolism, which could be used in the future to combat MRSA.

Metabolomics is also promising for treatment monitoring. Cancer, for example, has specific cellular metabolism changes that seem to drive the disease, the monitoring of these changes allows for potential to track progression of the disease and effectiveness of treatments on it, along with helping predict the prognosis for a patient. Patient samples that are analysed can take several forms, direct biopsies from tumours can be very effective but are invasive and probably not suitable for routine monitoring. Thus, less invasive patient samples are becoming more heavily examined for biomarkers that can test for disease along with staging the progression of disease. These less invasive samples are easier to obtain and can be easier to ethically justify talking from control groups meaning they are easier to study. However as they are extracellular, the metabolic profile is dependent on cell uptake and excretion from all bodily processes, not just the tumour cells and these need to be considered when analysing samples. The type of sample taken depends on the kind of disease, for example saliva could be used to assess lung cancer but urine for bladder or kidney cancers (Armitage & Southam,

2016). Same patient resampling, through the use of longitudinal metabolomics it's possible to monitor progression, relapse, and remission of diseases like cancer. For example TCA cycle intermediates and RNA degradation products decrease in serum when colorectal cancer is remising or at a stable level (Zhu, et al., 2015). An issue with progression monitoring is differentiating between metabolomic changes linked to the disease from the treatment as drugs and therapies to treat a disease will have metabolic impacts (Lodi, et al., 2013). Phenylketonuria requires lifelong dietary management to prevent neurocognitive impairment in those affected, metabolomics offers the ability to monitor the diseases progression and urine sample metabolomics offers a less invasive method than using blood specimens, but a more accurate method than self-reported dietary records, as urinary biomarkers could identify poor diet adherence via elevated levels of catabolites showing phenylalanine intoxication (Wild, *et al.,* 2019).

1.2 | Lung Cancer

Lung cancer is the leading cause of cancer-related death worldwide, with 2.09 million cases worldwide (World Health Organsiation, 2018). The 5-year survival rate for lung cancer is 13% (Chen, et al., 2016), however when the cancer is detected early before it has had a chance to grow and metastasise (stage 1 and some stage 2) this rate increases to 57% (Siegel, *et al.*, 2020). The high mortality rate of this disease reflects the fact that it is often misdiagnosed or detected only in its advanced stages (Song, et al., 2015). Three quarters of lung cancer cases could be prevented if patients did not smoke. Smoking cannabis and passive smoking are also risk factors (NHS, 2020). There is 84% mortality, with low- and middle-income countries having the highest death rates (Adjei, 2019). Between ten and twenty five percent of cases are in never smokers, more often seen in Asians and women, and often arising at a younger age. A large proportion of these cases are never explained by genetic or environmental risk factors (Llandi, *et al.*, 2019). Akthar *et al.,* (2017) implicates environmental exposure, mutations, air pollution and single nucleotide polymorphism as non-smoking causes of lung cancer (Akhtar, *et al.*, 2017).

Early diagnosis and detection will dramatically improve a patient's long-term prognosis and recovery relies heavily on detecting the primary malignancy before the cancer spreads

(Hofman, 2017). There are already some screening programmes in place for specific types of malignancies, such as tumour markers and cervical smear tests, but many screening systems currently in place carry potential harms as well as benefits. In 2012, the potential risks were published as part of a systematic review and found that the main concern was the number of false positives (Bach, *et al.*, 2012). When a patient receives a positive result, it can be devastating for not just the patient but their whole family, therefore it is imperative to minimise the number of false positives wherever possible. There is also the matter of cost, the cost of equipment, and clinicians' time. For lung cancer screening is only cost-effective when criteria are put in place to assess someone's eligibility for being screened (Haaf, et al., 2017). Criteria such as, age, smoking status and family history are used to determine a patient's suitability for the chest x-ray. The gold standard of screening states that for any screening programme, potential benefits must outweigh any risks associated with the screening process itself. In terms of lung cancer screening this means chest x-rays are only suitable for "at risk" groups (American Cancer Society, 2015).

Histologically lung cancer is subtyped into non-small cell (NSCLC) and small cell lung cancer (SCLC). NSCLC is far more prevalent and accounts for 85% of cases, whereas SCLC is far less common accounting for just 15%. NSCLC includes adenocarcinomas, squamous cell carcinomas, and large cell carcinomas (Oser, *et al.*, 2015). The sub-types are characterised by their location of origin. Adenocarcinomas develop in the peripheral bronchi and make up 40% of all lung cancer cases. Squamous cell carcinomas are found in the bronchial tubes arising from early mutations in the squamous cells of the airway and account for approximately 25-30% of all cases. The origin of large cell carcinomas is still not fully known they make up around 10% of all cases (Kenfield, *et al.*, 2008). SCLC originate in the Kulchitsky cells found in the lung and are known to rapidly invade the submucosal lymphatic vessels and into the lymph nodes (Lemjabbar-Alaoui, *et al.*, 2015). Our knowledge of the differing cellular metabolisms in these different LC forms is limited (Lim, et al., 2018).

Staging lung cancer currently involves evaluation of a tumour size, the level of penetration into surrounding tissues and the presence or absence of any metastases either in the lymph nodes or into other organs (Table 1) (Detterbeck, *et al.*, 2017). Staging is hugely important at it is directly used in determining the most suitable treatment plan for a patient, since most lung-cancer therapies are geared towards the specific stages (Gajdhane & Deshpande, 2014).

An example of this is when deciding whether to put a patient forward for surgery, in the early stages (stages 1 and 2) surgery is the main curative treatment, whereas locally advanced stages are rarely cured with local treatment alone, therefore one important advantage of accurate staging is to select whether a patient should be a candidate for primary surgery or for multidisciplinary action (Stamatis, 2015).

Primary Tumour (T)			
T0	No primary tumour		
T1	Tumour size ≤ 3cm		
T1a	Tumour size ≤ 1cm		
T1b	Tumour size > 1cm but ≤ 2cm		
T1c	Tumour size > 2cm but ≤ 3cm		
T2	Tumour size > 3cm but ≤ 7cm	OR	a tumour which: invades the visceral pleura, involves the main bronchus (≥2cm distally to the carina) or is characterised by atelectasis or obstructive pneumonia which extends to the hilum but does not involve the entire lung
T2a	Tumour size > 3cm but ≤ 5cm		
T2b	Tumour size > 5cm but ≤ 7cm		
T3	Tumour size > 7cm	OR	Directly invading the chest wall, diaphragm, phrenic nerve, mediastinal pleura or parietal pericardium
		OR	Is in the main bronchus <2cm distal to the carina
		OR	atelectasis or obstructive pneumonia of the entire lung
		OR	Separate tumour nodules in the same lobe
T4	Tumour of any size with invasion spreading to the heart, great vessels, trachea, recurrent laryngeal nerve, oesophagus, vertebral body or carina	OR	Separate tumour nodules in a different ipsilateral lobe
Regional Lymph Nodes (N)			
N0	No metastasis to regional nodes		
N1	Metastasis to ipsilateral peribronchial and/or perihilar lymph nodes and intrapulmonary nodes, including involvement by direct extension		
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph nodes		
N3	Metastasis in contralateral mediastinal lymph nodes	OR	Contralateral hilar lymph nodes
		OR	Ipsilateral or contralateral scalene lymph nodes
		OR	Supraclavicular lymph nodes
Distant Metastasis (M)			
M0	No distant metastasis		
M1a	Separate tumour nodes in a contralateral lobe	OR	A tumour with pleural nodules
		OR	Malignant pleural dissemination
M1b	Distant metastasis		

Table 1: The current staging criteria for lung cancer (Detterbeck, *et al.*, 2017).

For staging chest x-ray is the most common initial tool, but this is not suitable to stage early non-small cell lung cancer. Pre-operative assessment of non-small cell lung cancer can involve ultrasound, Computerised Tomography (CT), Magnetic Resonance Imagery (MRI) and 18-Fluorodeoxyglucose positron emission tomography to assess pleural invasion. Ultrasound is

cheap, non-invasive, and widely available and means the mass can be observed whilst the patient is breathing and loss of movement against the pleura suggest chest wall involvement. MRI is used similarly, but with improved accuracy. PET is used to assess nodes and metastasis. CT can assess the depth of any chest wall invasion. Invasion of the parietal pleura is considered shallow and deep invasion involves soft tissue and ribs (Jhih-Hao, *et al.,*2019). CT can further assess by opacity with solid component thought to indicate an invasive lesion. Further evaluations can add information such as the Tumour Disappearance Ratio, assessing tumour borders (Hsu, et al., 2017), collating size/contact with pleura/angle of contact and pleural thickening (Quint, *et al.,* 1995). Chest x-ray and CT remain the most widely used imaging modalities. After CT the NICE guidance (2019) advises when PET-CT should be used and when endobrachial ultrasound should be employed in the United Kingdom (NICE, 2019). A number of methods are used to obtain a tissue diagnosis dependent on the size of tumour, site of tumour, patient's comorbidities and overall health status (Latimer & Mott, 2015). It remains the case that a rapid cost-effective means of lung cancer staging could have significant benefits in terms of diagnosis and patient care.

1.3 | The Use of Saliva as a Biofluid

There are many different biofluids which can be used for metabolomics, all of them have advantages and disadvantages and some are better than others in specific circumstances so it is important to consider which biofluid will provide the most accurate, reliable and importantly repeatable results. The most common biofluids used for metabolomic analyses are blood, both serum and plasma, and urine. The collection of these biofluids is far less invasive than traditional methods such as tissue biopsy or cerebrospinal fluid collection so are suitable for use in screening and patient monitoring, which is an important advantage. Others like sputum, specifically induced sputum, have unique advantages specifically when investigating certain conditions, such as chronic lung diseases. One of these advantages is the amount of information already known about its characteristics in both normal participants and patients with disease, especially related to inflammatory content and bacterial load (Nicholas & Djukanovic, 2009). However, one of the disadvantages of using these fluids is that their metabolite content is influenced by many factors such as diet or activity level of the participants or recent physical stimuli, the most well-known being that prostate massage can

lead to increased release of prostate-specific antigen into the urine (Giskeodegard, *et al.*, 2019).

Saliva is a clear, slightly acidic, complex biofluid composed of secretions from the major salivary glands. Human salivary glands combine water, salts, salivary proteins and molecules which arise from the blood to produce 1-1.5 L of serous and mucinous saliva daily (Humphrey & Williamson, 2001). These molecules such as hormones, enzymes and antibodies enter the saliva from the blood by either transcellular (passive intracellular diffusion and active transport) or paracellular routes (extracellular ultrafiltration) (Jusko & Milsap, 1993). This means that most compounds found in serum or plasma will also be present in the saliva and so the saliva should also be a good reflection of the body's physiological state (Lee & Wong, 2009). There was however concern that although the saliva contained these molecules as found in blood that they would be found at significantly smaller concentrations and therefore may be much more difficult to detect (Miller S. , 1994) though due to recent developments in specifically the sensitivity or equipment the low concentrations of molecules in saliva is no longer a limitation (Lee & Wong, 2009).

The non-invasive nature of sampling saliva is one of its major advantages, along with its dramatic changes in concentrations of expressed metabolites during disease. However, one major disadvantage is that there are various factors can influence saliva constituents, for example age, gender, general health, smoking status, general oral hygiene, and nutrition, and these have to be taken into account when conducting metabolomics on these samples. On top of this there are several different ways to collect saliva, unstimulated, where a patient passively drools into a graduated tube. Where larger volumes are required stimulation can be applied via masticatory stimulation (Mikkonen, et al., 2015) but 44 of 137 metabolites in stimulated saliva were found in higher concentrations than in unstimulated saliva (Okuma, et al., 2017). Based on groups of cancer patients who fasted for 1.5 hours, 3 hours and 12 hours before saliva collection, where use of oral hygiene products like toothpaste were forbidden for 1 hour before sample collection, it was found that 51 metabolites differed significantly, it was found that 12 hours is the optimal amount of time to fast prior to collecting saliva for metabolite analysis (Ishikawa, et al., 2017). In recent years, the use of saliva as a biofluid has been suggested for the detection biomarkers associated with lung cancer (LI, et al., 2012).

1.4 |Gastric Cancer

Gastric cancer is often diagnosed too late for curative treatment and represents a leading cause of cancer related death worldwide (Van Cutsem, *et al.*, 2016) accounting for 783,000 deaths worlwide in 2018 (Rawla & Barouk, 2018) and 4,333 deaths between 2016-2018 in the United Kingdom (Cancer Research UK, 2021). The main causes can be divided into infection/damage, hereditary and environmental factors. *Helicobacter pylori* infection is the primary cause of sporadic distal gastric cancer. The infection causes inflammation after which carcinogenesis, bacterial, host, environmental and other factors facilitate damage repair. The change in cell production, programmed cell death and inherited genetic modifications may then play a part in the evolution of dysplasia or adenocarcinoma. Epstein-Barr virus is also implicated as a cause of gastric cancer and this pathogen is found in tumour cells, but not normal epithelial cells. Environmental factors such as smoking, low dietary fibre intake, high salt diet, high pickled food intake (Shivappa, *et al.*, 2016) and obesity (Lauby-Scretan, et al., 2016) are associated with high risk of gastric cancer.

Gastric cancers can be classified by anatomical position, histologically and further by genomics. 95% of stomach cancers are adenocarcinomas. Rotondo *et al.* (2019) describe the origins of these tumours from glandular epithelium. The remainder are soft tissue sarcomas, gastrointestinal tumours, mucosal associated stromal tumours, lymphomas, and carcinoid tumours. These are from multiple morphologies, molecular backgrounds and histogenesis (Rontondo, et al., 2019). The cancer genome atlas divides gastric cancers into four groups according to the genomics of the tumour chromosomally unstable tumours, genomically stable tumours, Epstein-Barr virus infected tumours and microsatellite instability tumours (Sohn, et al., 2017).

Patients with gastric malignancy are likely to present with weight loss, nausea, vomiting, early feeling of fullness, indigestion, and epigastric pain. In the later stages palpable nodes and masses may be present such as abdominal masses, Virchow's node (left supraclavicular) or Sister Mary and Joseph's node (umbilical). Diagnosis is achieved by gastroscopy and biopsy. Endoscopic ultrasound is employed with an 85% specificity in identifying T1/2 tumours or T3/4 tumours. CT/PET scans are used to evaluate for local spread, nodes, and metastasis. Gastric tumours are staged according to the tumour, nodes, and metastasis system (TNM).

Laparoscopic assessment of peritoneal progression and further biopsies are invaluable (Johnston & Beckman, 2019).

Most gastric cancers are not diagnosed in the early stages as patients are commonly asymptomatic until the tumours are advanced. The specificity and sensitivity of serum biomarkers remains poor; CEA and CA-19 are the most used tumour markers in gastric cancer diagnosis and can be useful for prognostic estimation but do not display enough sensitivity or specificity for screening or early diagnosis, whereas CA72-4 has been shown to have better sensitivity and accuracy but there is little evidence for its relevance in screening or early diagnostics (Necula, et al., 2019). Photofluorography and other screening methods are said to be unpractical in the western world due to the relatively low incidence making them cost-prohibitive, but they are used effectively in Asian countries (Johnston & Beckman, 2019). In terms of previous metabolomic investigation into gastric cancers in a study of malignant vs normal gastric tissue, it was found that 18 metabolites were detected differently, and there were 5 differences between invasive and non-invasive tumours, these metabolites allowed construction of a diagnostic model that had an AUC of 0.9629 when differentiating normal tissue from tumours (Wu, et al., 2010). Gastric cancer metabolites were found to have 77 differing metabolites from controls, allowing creation of a discriminatory biomarker panel of 3 molecules with AUC of 0.95 (Chan, et al., 2016).

In terms of using gastric fluid, a review of studies into oesophago-gastric cancers and the effectiveness of metabolomic analysis found a variety of biofluids have been used to varying levels of success. 11 studies used tumour vs normal tissue, 8 used serum, 1 used urine and 1 used gastric contents (Abbassi-Ghadi, *et al.*, 2013). It is thought that gastric content is an important factor in the development of gastric diseases including cancer, a ROC analysis found a panel of 4 metabolites (acetaldehyde, formaldehyde, hydrogen sulphide and methyl phenol) able to differentiate cancer from healthy controls with an AUC of 0.9 (Kumar, et al., 2012). Gastric juices were capable of differentiating between chronic superficial gastritis, and Intestinal metaplasia and gastric cancer, and also monitoring its progress accurately based on metabolomic analysis studying the conversion of histidine to histamine, as well as five bile acids, suggesting progression of gastric cancers is linked to gut metabolic processes (Lee, et al., 2020).

1.5 | Aims and Objectives

In the past cancer diagnoses have been based on the clinical presentation of the disease, otherwise known as the disease phenotype, then procedures such as biopsies or computerized tomography (CT) and positron emission tomography (PET) scans before a definitive diagnosis can be given. This delay in diagnosis attributes to a more advanced stage at diagnosis, which is strongly related to the cancer mortality rate (Rutherford, et al., 2013). To improve survival there is a focus on early presentation, reducing the delay, to improve the stage at diagnosis (Neal, 2009). In recent years, and with the increased knowledge from significant scientific developments, most notably the Human Genome Project (Collins, *et al.*, 2003), there have been significant advances in the use of functional genomics to detect novel biomarkers (Das, *et al.*, 2017), which may be used for early detection of cancers in the future.

The overarching aim of this study is to use metabolomic approaches to define different types of malignancy, using novel but easily accessible biofluids, with the hope of identifying potential biomarkers for diagnosis and staging, which are both statistically, and clinically significant in terms of assessing their sensitivity and specificity (Holland, 2016). Firstly, to analyse saliva samples from patients with lung cancer and from a healthy control group to see if they can be differentiated metabolically. If this is the case, then metabolites found to be at statistically significant levels ($P<0.05$) in lung cancer group can then be assessed for their value as potential clinical biomarkers. Then, to test saliva samples from patients with different stages of lung cancer for differences in the metabolome, with the aim of identifying metabolites able to discriminate between the stages that may be potential biomarkers which either individually or as a panel could provide very specific and sensitive detection. The objectives were to prepare the samples for mass spectrometry, run the mass spectrometry and then to analyse and interpret the results using statistical multi-variate analysis with respect to the patients' diagnosis.

Further to investigating lung cancer using saliva samples, to also assess the use of gastric fluid as a biofluid for metabolomic research, using samples from patients with gastric cancer and from a healthy control group looking for differences in their metabolomes, with the hope of identifying metabolites found at significantly different levels in the malignancy group which may be suitable for use as biomarkers. The objectives of the whole study are to prepare the

samples of varying types for mass spectrometry, run the mass spectrometry and then to analyse and interpret the results using statistical multi-variate analysis with respect to the patients' diagnosis.

CHAPTER 2 | Materials and Methods

2.1 | Sample Collection and Ethics

For the study, ethical approval for the project was sought from Aberystwyth University and the project was approved on 05/03/2019 (Reference: 12062). All samples were anonymised and given a unique code which corresponded to the participant information being held, to ensure patient confidentiality was maintained. The samples were also taken by trained medical professionals, to ensure patient safety and comfort during the sample collection.

2.1.1 | Saliva Samples of Lung Cancer and Control Samples

The samples used for assessing lung cancer samples against control samples were taken as part of an ongoing overarching project between our research group and a local health board. This project has received local and regional ethical approval from the Hywel Dda University Health Board (16/WA/0036, Novel technologies for diagnosing and monitoring pulmonary diseases). 30 samples were obtained, 27 in the lung cancer group and 3 in the control group, along with clinical data from each participant (Appendix 1) containing their age, sex, BMI, smoking history, medical history, family history and tumour stage (where applicable). The control participants were roughly matched in terms of BMI and age and none of them have had any form of malignancy in the past.

Patients were recruited from respiratory clinics or respiratory wards across three hospital sites around Wales, with the help of nurses and consultants. The patients were given information about the project before giving their valid and informed consent for the samples to be taken and for the storage and use of the information they provided on the clinical information sheet. The saliva samples were uninduced saliva samples, as soon as the samples were collected they were immediately put on ice and moved to the -80˚C freezer, the time taken to reach the freezer could vary by a few minutes depending on the site in which they were taken but not any longer than that. The samples were then transported from the hospital sites to the Institute of Biological, Environmental and Rural Sciences at Aberystwyth University by an NHS approved courier on dry ice.

2.1.2 | Saliva Samples of Different Lung Cancer Stages

For this study an application was filed to the Welsh Cancer Bank (WCB) to issue these samples, after assessment by the WCB board the project was approved for the samples to be released (Reference: 18/017). The Wales Cancer Bank is licensed by the Human Tissue Authority (licence 12107) to store human tissue for research and has ethics approval from Wales Research Ethics Committee 3 as a research tissue biobank to collect and issue biomaterials for cancer related research. Having received these samples from the WCB, approval from National Research Ethics Service for the use of these samples was not required, and the samples were provided anonymised with only minimal associated data. 54 samples were collected along with their clinical data (Appendix 2) including the sub-type of lung cancer, their sex, smoking status and staging information.

The saliva samples were collected in respiratory clinics at various hospital sites around Wales, by doctors and research nurses using the following protocol:

Saliva Collection

1. Remove the lid and pass the saliva collection 15ml falcon tube to the patient
2. Ask the patient to spit saliva from the mouth into a 15ml Falcon collection tube.
3. Place the lid onto the collection tube and attached the unique barcoded label to the sample tube
4. Record the time and date the sample was taken from the patient on the saliva collection form
5. Place the sample in -80^{o}C freezer as soon as possible recording the time and date

The time between collection and freezing would vary slightly between each hospital site depending on the location of the clinic in relation to the laboratory containing the -80˚C freezer. The samples were then couriered to the Welsh Cancer Bank on dry ice and stored there in -80˚C freezers before being issued and couriered to Aberystwyth University on dry ice and transferred into the -80˚C freezers where they were stored until ready to be processed.

2.1.3 | Gastric Fluid Samples

These samples were sent over from Brazil for analysis, having been collected as part of a wider study by the Medical Genomics Laboratory at the AC Camargo Cancer Centre, Brazil. The study was approved by the Institutional Review Board of the A.C Camargo Cancer Centre (Reference: 2134/15). The gastric fluid samples were obtained after the participants had undergone an 8-10 hour fast. 50 gastric fluid samples were obtained, 20 with gastric cancer and 10 in the control group, along with minimal clinical data (Appendix 3).

2.2 | Technical Information

2.2.1 | Saliva Samples

2.2.1.1 | Processing of Samples

For each sample, the saliva supernatant was thawed at 4°C, and 200µl then transferred into a sterile 2ml microcentrifuge tube. To this microcentrifuge tube 30 mg of ≤ 106 µM glass beads washed in acetone (Sigma-Aldrich, Dorset, UK) were added. To this 1520µl of the solvent mix of HPLC grade methanol and chloroform, in a ratio of 4:1, was added. Following this the sample was vortexed for 5s and sonicated for 30s at 30Hz to homogenise the mixture. It was then shaken for 20 minutes at 4°C and stored at -20°C for 20 minutes to allow time for protein precipitation. The sample was then centrifuged for 6 minutes at 11 000 x g with the temperature set to 4°C and the resulting supernatant was then transferred into a new sterile 2ml microcentrifuge tube.

2.2.1.2 | LQT-MS

From this, 70 µL was transferred into an LTQ-MS vial and sealed, then stored at -20 °C until run. The samples were then run in a randomised order using an autosampler with the tray kept at a constant 15°C. From each vial, 20 µL was then injected into a flow volume of 60µL/minute water-methanol (ratio of 7:3) using a Surveyor liquid chromatography system (Thermo Scientific, MA, USA). The data obtained from each individual sample was conducting, in alternating positive and negative ionisation mode, over four scan ranges (15-110 *m/z*, 100-

220 *m/z,* 210-510 *m/z,* 500-1200 *m/z*) on an LTQ linear ion trap (Thermo Electron Corporation, CA, USA), with an acquisition time of five minutes. Individual metabolite *m/z* value was normalised as a percentage of the total ion count for each sample.

2.2.2 | Gastric Fluid Samples

2.2.2.1 | Processing of Samples

For each sample, the gastric fluid was placed on ice for two hours to ensure the sample had completely thawed, before being spun at 13,000rpm for 10 minutes in the centrifuge. 100µl of the gastric fluid was diluted with 900µl of methanol and 100µl of ultrapure water and mixed with a vortex for 5 seconds. For the gastric samples it was decided to use just methanol and water, rather than the Bligh and Dyer protocol used for the saliva samples as it favours polar molecules. The samples were then left on ice for 10 minutes to allow precipitation of the proteins before being centrifuged once again for 10 minutes at 13,000rpm and the supernatant decanted into new microcentrifuge tubes. 50µl of the supernatant was then transferred into a glass vial with a clean glass insert and 250µl of 70% aqueous methanol (made of HPLC grade methanol and ultrapure water) with a temperature of 4°C added before vortexing for 5 seconds to ensure the solutions was thoroughly mixed. Each vial was then stored at -20°C until ready to be transferred into the mass spectrometers.

2.2.2.2 | Mass Spectrometry

The samples were then run in a randomised order using an autosampler with the tray kept at a constant 15°C. From each vial, 20 µL was then injected into a flow volume of 60µL/minute water-methanol (ratio of 7:3) using a Surveyor liquid chromatography system (Thermo Scientific, MA, USA). This generated metabolite fingerprints in both, positive and negative ionisation mode, in a single run. Samples (20 µL volume) were injected into a flow of 100 µl.min-1methanol: water (70:30, v/v). Ion intensities were acquired between m/z50 and 1000 for 3.5 min at a resolution setting of 100,000 (at m/z200) resulting in 3 (± 1) ppm mass accuracy. Mass spectra around the apex of the infusion peak were combined in a single intensity matrix (runs x m/z) for each ion mode.

2.3 | Data Analysis and Statistics

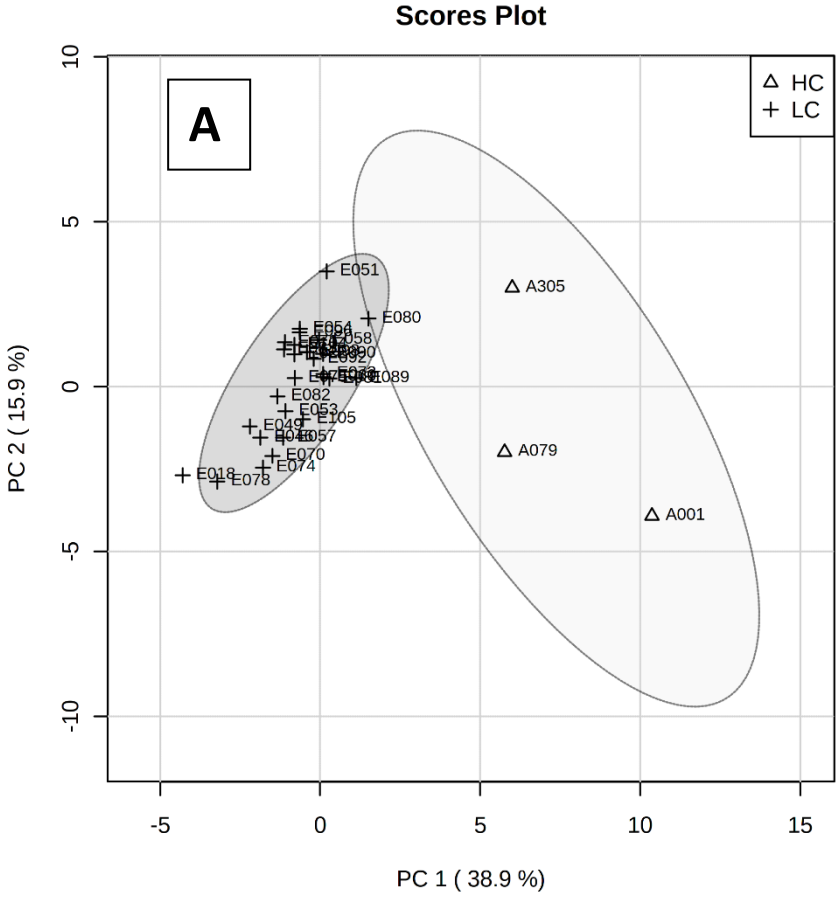
The metabolite data was then analysed using MetaboAnalyst 4.0 (available at: www.metaboanalyst.ca/MetaboAnalyst/home.xhtml), a web-based analysis programme for omics data. Initially the 'statistical analysis' module was used, which implemented a data integrity check, normalisation, and overview. The data was then normalised using the auto-scaling method of transformation, by mean-centring and division by the standard variation of each variable. The distribution was then checked to ensure that a normal distribution was followed.

MetaboAnalyst was then used to carry out multivariate approach statistical analysis. Principle component analysis and partial least squares discriminant analysis (PLSDA) was used to determine which samples showed similarities and differences in metabolite abundance patterns, both within and between groups. Further statistical analysis was then carried out using a t-test or one-way ANOVA and creating volcano plots to assess the differential concentration patterns. The biomarker analysis module was then used to run ROC (Receiving operating capacity) curve analysis looking at the specificity and sensitivity of the metabolites which appeared to show significant differences between groups. The area under the curve (AUC) was calculated and heat maps used to visualise the data. The value used as the accepted AUC threshold for clinical significance for this project was 0.7, as a biomarker with an AUC of 0.5-0.7 is considered a fail or weak biomarker, whereas 0.7-0.8, 0.8-0.9 and 0.9-1.0 are considered fair, good and excellent respectively (Xia, *et al.*, 2013).

3.1 | Saliva for Lung Cancer Diagnosis

3.1.1 | Cancer vs. Non-Cancer

The data derived from the LC-MS was uploaded to MetaboAnalyst 4.0 and the statistical analysis module was used to produce a principle component analysis (PCA) biplot and a partial least squares discriminant analysis (PLS-DA), which showed similarities and differences in metabolite expression between the groups (Figure. 1).



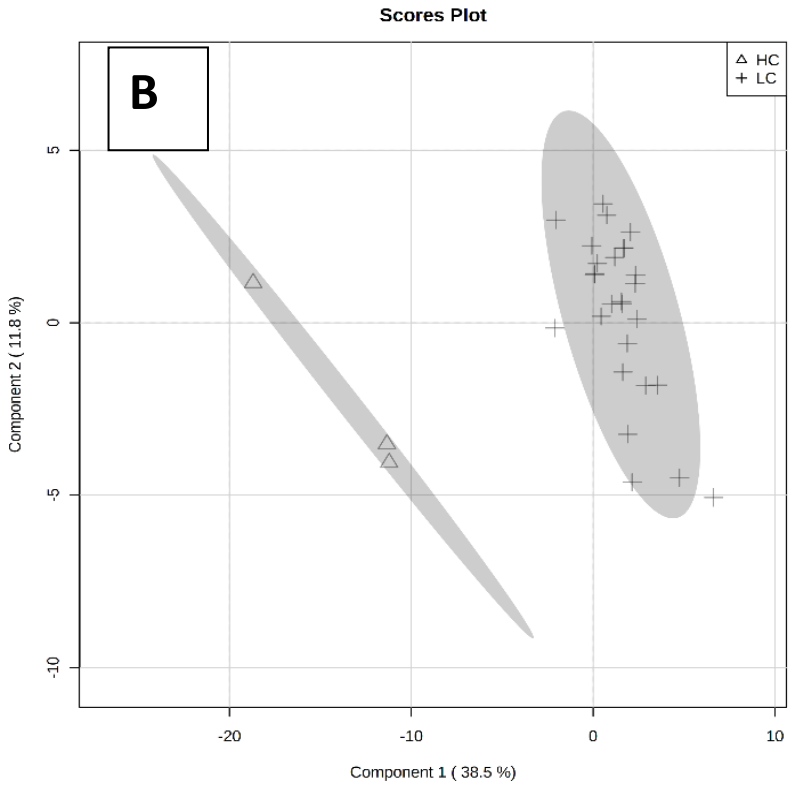


Figure 1: (A) Principal component analysis (PCA) and (B) partial least squares discriminant analysis (PLS-DA) biplots showing the degree of separation both between the individual samples and between the lung cancer (LC) and healthy control (HC) groups. Each plot represents a sample, the triangle plots are associated with the healthy control group and the crosses are associated with the lung cancer group and the ellipses indicate 95% confidence.

The PCA biplot (Figure 1a) showed that variation appeared to be linked to the experimental groups with little overlap in the 95 % confidence ellipses. Some of the samples in the lung cancer group form a particularly tight cluster, whereas some are more spread out. The healthy control group consisted of a small number of samples and is far less clustered, showing higher variety in the metabolites expressed. PLS-DA (Figure 1b) is a supervised approach where the net effect is seeking variables which discriminate between the categories assigned to the analysis by the investigator. Using this analysis, the groups did not overlap, indicating that they have discrete metabolite expression patterns. However there were also two samples in the lung cancer group which fell outside of the imposed grouping, implying that they had higher levels of variation in the metabolite expression compared to the rest of the group.

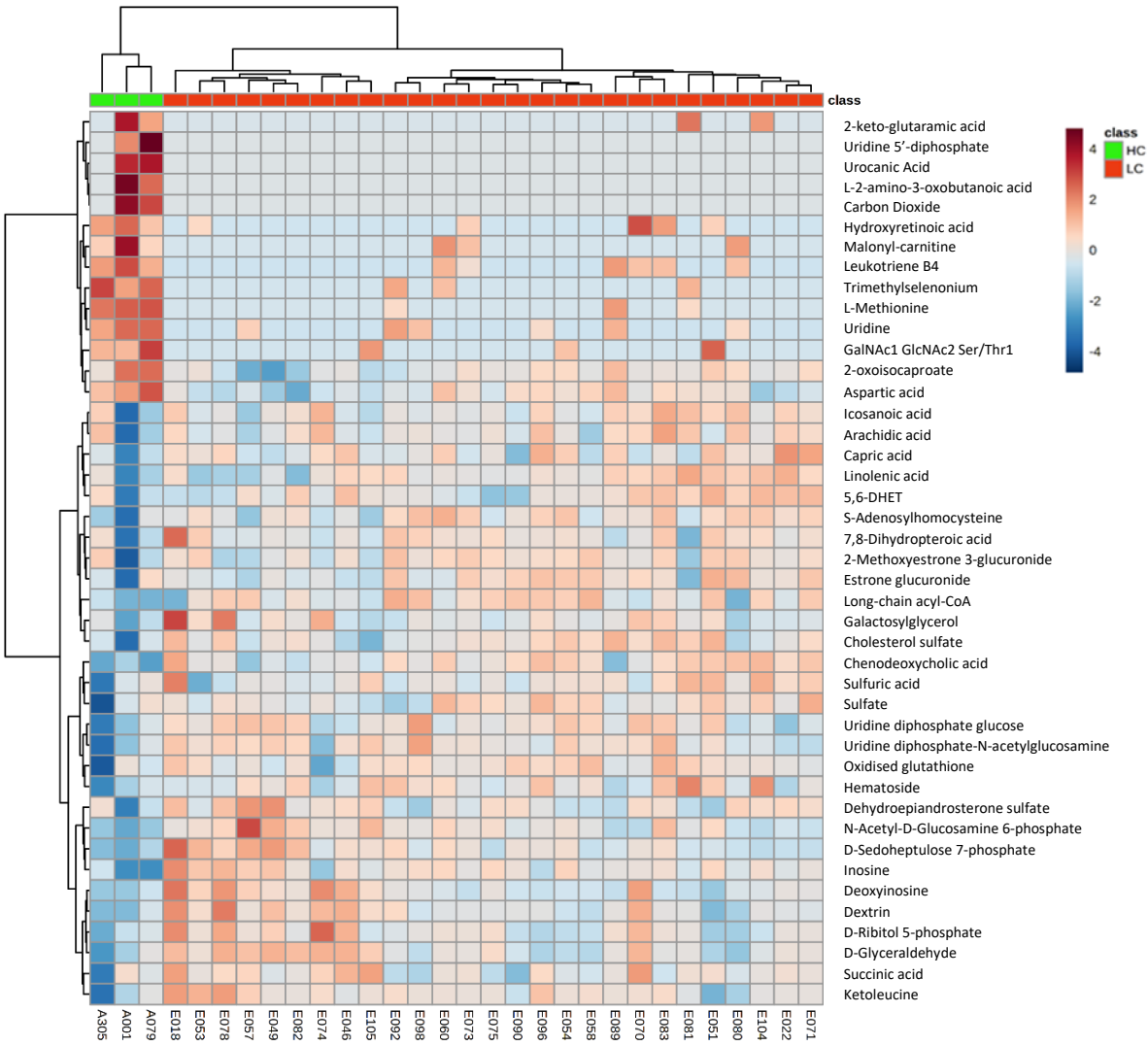


Figure 2: Heatmap comparing metabolite accumulation patterns of the 30 patient samples from both the lung cancer and healthy control groups. The sample codes are along the x-axis and metabolites are on the y-axis. Cells which are dark red have relatively lower content , whilst cells which are dark blue relatively lower content. The dendrograms represent the relationship between the metabolites and the samples. The key (top right) shows the diagnosis groups of healthy controls (HC) and lung cancer (LC).

The major sources of variation between the groups were identified by t-test and displayed using a heat map (Figure 2). The length of the dendrogram branches shows the degree of difference between the clusters, the longer the length, the greater the degree of difference. It was noted that the HC group form a distinct cluster from the LC group further indicating that this formed a distinctive patient grouping, but that the LC group then branches off into two distinct branches on the x-axis, showing variation within the group. The y-axis dendrogram was used to compare the metabolites, the dendrogram shows two distinct

groups, the top 14 metabolites form one of these branches and from the heatmap itself it is clear that they had far higher levels of accumulation in the healthy control group compared to the lung cancer.

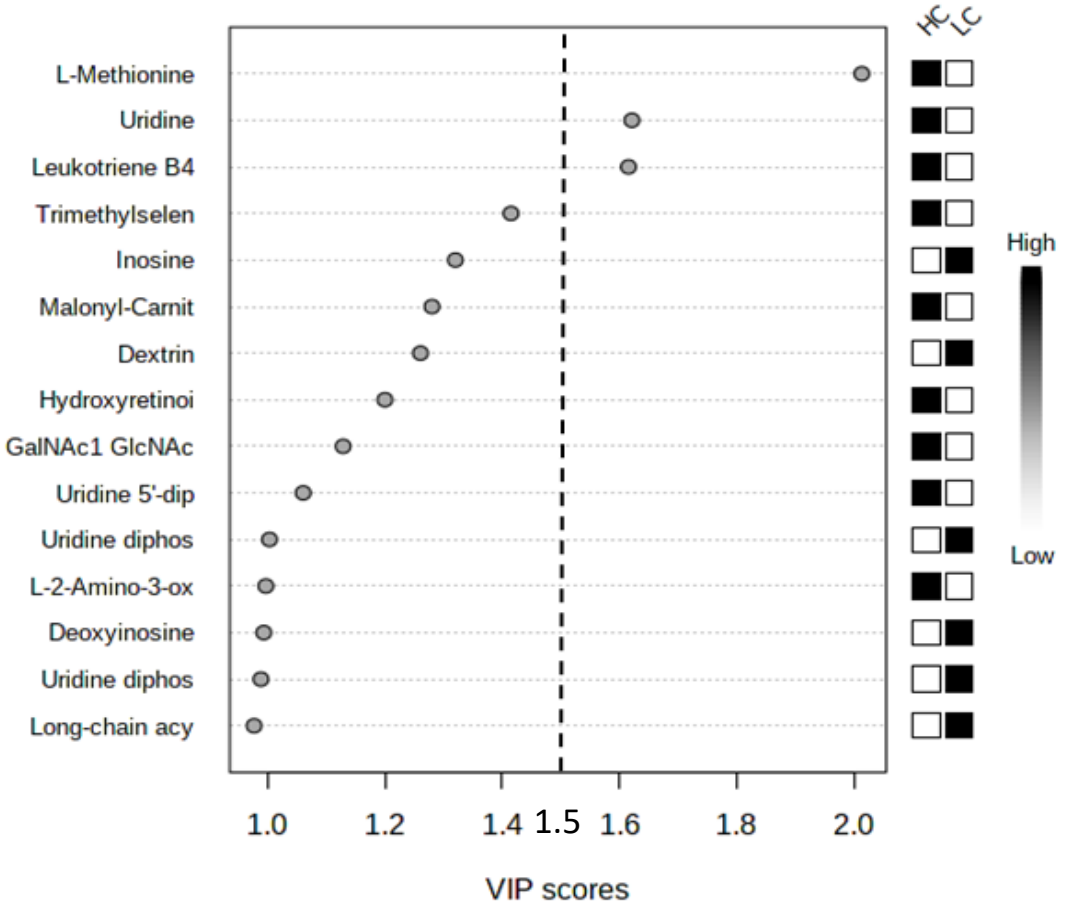


Figure 3: Variable importance in projection (VIP) scores for metabolites in the PLS-DA showing in Fig. 1b. VIPs are a weighted sum of squares that considers the amount of explained Y-variance of each component. The metabolites are listed from highest to lowest VIP score. The black squares represent high levels of the metabolite and white representing low levels of the metabolite.

The PLS-DA biplot showing complete separation for the HC and LC groups (Fig. 1b) was interrogated to define the major sources of variation as indicated by variable importance in projection (VIP) scores. A value of 1.5 is usually the accepted VIP for significance (Akarachantachote, Chadcham, & Saithanu, 2014) and there are 3 metabolites (L-methionine, Uridine and Leukotriene B4) which have VIP scores above this threshold. All were found at high levels in the healthy control group and far lower levels in the lung cancer group.


L-Methionine
L-2-Amino-3-oxobutanoic
Trimethylselenonium
Aspartic acid
2-Oxoisocaproate
Sulfuric acid
Succinic acid
2-Keto-glutaramic acid
Malonyl-Carnitin
Ketoleucine
Uridine diphosphate-N-acetylglucosamine
Oxidized glutathione
GalNAc1 GlcNAc2 Ser/Thr1
Urocanic acid
Long-chain acyl-CoA
Carbon dioxide
D-Glyceraldehyde
Sulfate
Linolenic acid
Hydroxyretinoic acid
Inosine
7,8-Dihydropteroic acid
Leukotriene B4
Dehydroepiandrosterone sulphate
S-Adenosylhomocysteine
5,6-DHET
Galactosylglycerol
Arachidic acid
Icosanoic acid
Hematoside
HC
LC

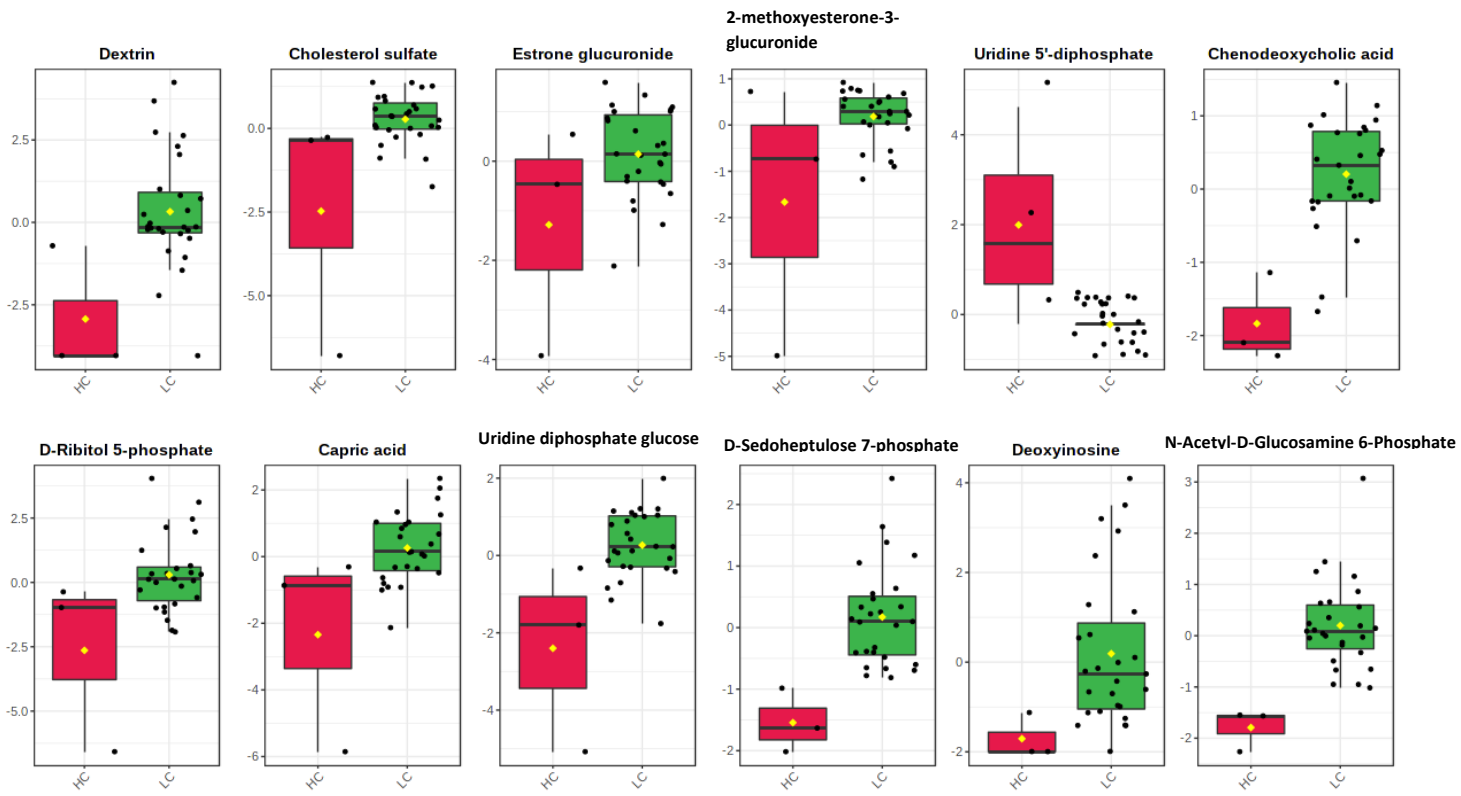


Figure 4: Box plots for the metabolites where significantly different levels of the metabolite in the saliva of the healthy controls and lung cancer groups. The red plots representing the healthy control group and the green plot representing the lung cancer group ($p < 0.05$). The groups are shown on the x-axis and the expression levels are shown on the y-axis.

All of the metabolites which were found to significantly differ between HC and LC groups or to have high VIP scores were also presented using boxplots (Figure 4). The box plots (Figure 4) of 10 metabolites (L-methionine, Trimethylselenonium, malonyl-Carnitin, GalNAo1 GloNAo2 Ser/Thr1, Hydroxyretinoic acid, leukotriene B4, uridine 5'-biphosphate, D-Sedoheptulose 7-phosphate, acetyl-D-glucosamine 6-phosphate) showed no overlap between the control group and the lung cancer group. These could therefore be clinically significant for differential diagnosis of malignancy. The other 14 metabolite boxplots either showed overlap between the expression levels or had overlapping error bars, so their expression levels were not as different between the two groups.

3.1.2 | Identification of Clinically Relevant Biomarkers

Metabolite Name	Discriminant	AUC
L-Methionine	HC-LC	1
Trimethylselenonium	HC-LC	1
D-Sedoheptulose 7-phosphate	HC-LC	1
N-Acetyl-D-Glucosamine 6-Phosphate	HC-LC	1
Uridine	HC-LC	0.987654321
Aspartic acid	HC-LC	0.975308642
Inosine	HC-LC	0.975308642
Leukotriene B4	HC-LC	0.975308642
Chenodeoxycholic acid	HC-LC	0.975308642
GalNAc1 GlcNAc2 Ser/Thr1	HC-LC	0.950617284
Hydroxyretinoic acid	HC-LC	0.938271605
Long-chain acyl-CoA	HC-LC	0.938271605
Malonyl-Carnitin	HC-LC	0.925925926
Uridine diphosphate glucose	HC-LC	0.925925926
Dextrin	HC-LC	0.925925926
Deoxyinosine	HC-LC	0.919753086
Uridine diphosphate-N-acetylglucosamine	HC-LC	0.919753086
Hematoside	HC-LC	0.919753086
Ketoleucine	HC-LC	0.913580247
2-Oxoisocaproate	HC-LC	0.901234568
Cholesterol sulfate	HC-LC	0.901234568
S-Adenosylhomocysteine	HC-LC	0.888888889
D-Glyceraldehyde	HC-LC	0.864197531
Capric acid	HC-LC	0.839506173
D-Ribitol 5-phosphate	HC-LC	0.839506173
L-2-Amino-3-oxobutanoic acid	HC-LC	0.833333333
Uridine 5'-diphosphate	HC-LC	0.833333333
Urocanic acid	HC-LC	0.833333333
Carbon dioxide	HC-LC	0.833333333
Galactosylglycerol	HC-LC	0.827160494
Oxidized glutathione	HC-LC	0.814814815
2-Keto-glutaramic acid	HC-LC	0.796296296
Linolenic acid	HC-LC	0.796296296
7,8-Dihydropteroic acid	HC-LC	0.790123457
Dehydroepiandrosterone sulfate	HC-LC	0.783950617
Sulfate	HC-LC	0.75308642
Sulfuric acid	HC-LC	0.740740741
5,6-DHET	HC-LC	0.740740741
Icosanoic acid	HC-LC	0.728395062
Estrone glucuronide	HC-LC	0.728395062
Arachidic acid	HC-LC	0.691358025
2-Methoxyestrone 3-glucuronide	HC-LC	0.691358025
Succinic acid	HC-LC	0.654320988

Table 2: ROC (Receiving Operating Capacity) curve analysis showing the AUC (Area Under the Curve) value for each of the 43 significant metabolites found in saliva and discriminating between healthy control and lung cancer patients. The green cells show the metabolites with AUC values which exceed the accepted 0.7 threshold. HC and LC represent the healthy control and lung cancer groups respectively.

All of the metabolites discriminating between HC and LC (Figure 4) were assessed using AUC-ROC as potential biomarkers. There were 40 metabolites which had an AUC value over the 0.7 limit, so may be useful as potential clinical biomarkers. Therefore, the other 3 metabolites did not reach the threshold for sensitivity and specificity to be good biomarkers. This number of significant biomarkers showed the extent to which the malignancy affected the biological make up of not just the localised tissues and fluids but also as far as the saliva.

3.1.3 | Pathway Analysis

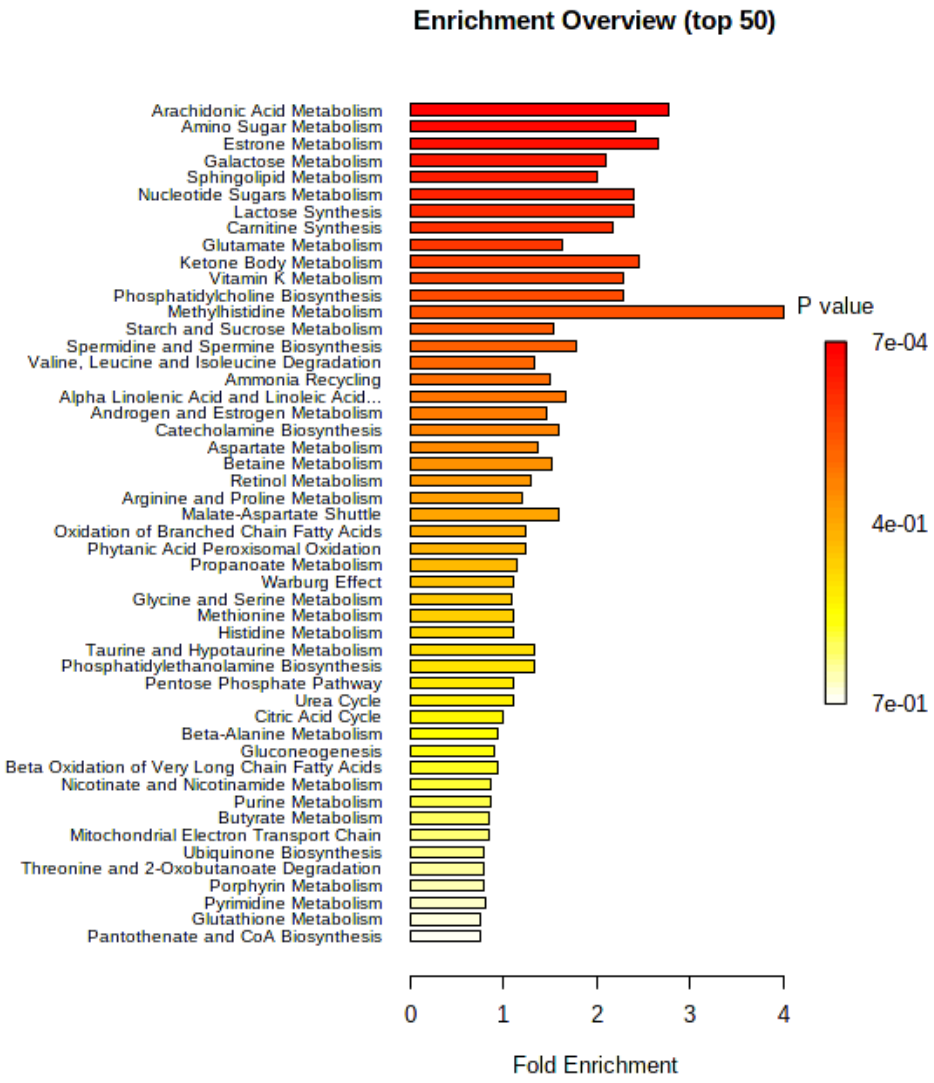


Figure 5: Summary plot for Over Representation Analysis (ORA) of the metabolites found at significantly different levels in the lung cancer group compared to the healthy control group. ORA is used to evaluate whether a particular metabolite set is represented more than would be expected by chance within the given compound list.

3.2 | Saliva for Lung Cancer Staging

3.2.1 | Combined LC forms: Stages 1-4

A comparison of all LC stages was undertaken in an attempt to find some discriminatory metabolites which were significantly different between the groups.

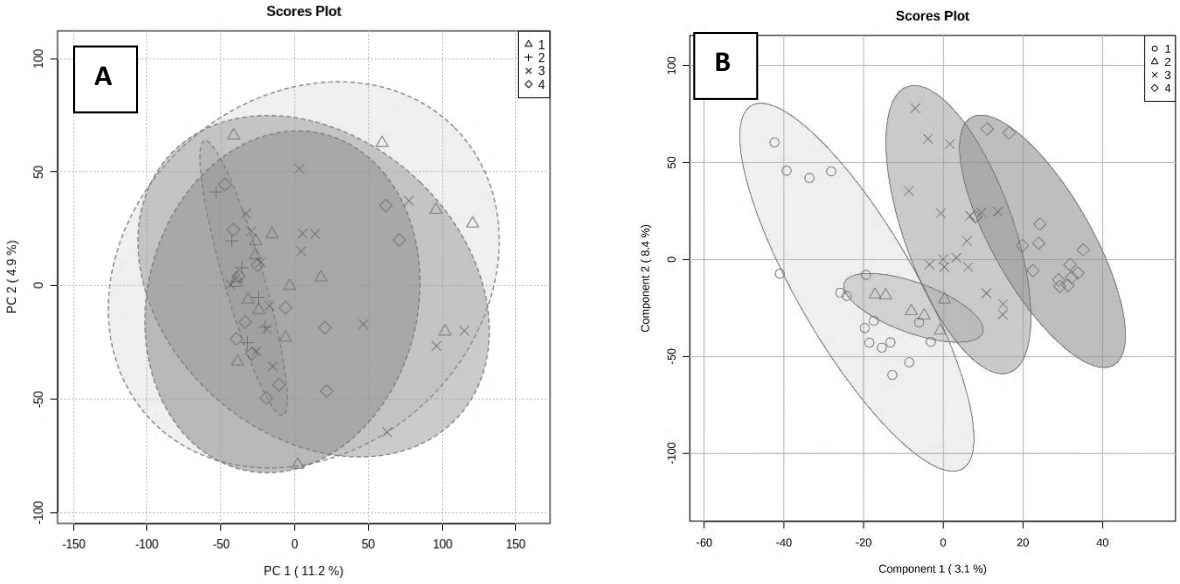


Figure 6: (a) Principal component analysis (PCA) biplot and (b) Partial least squares discriminant analysis (PLS-DA) of saliva metabolomes of patients at different stage of lung cancer (both adenocarcinoma and squamous cell carcinoma). Each plot represents a sample, the triangle plots are associated with the stage 1 samples, the cross plots represent stage 2 samples, x plots with the stage 3 samples and diamond plots with the group 4 samples.

The results of the PCA biplot (Figure 6a) showed that the metabolomes show some evidence of clustering, with the stage 2 group having made a particularly tight cluster however it was not possible to distinguish this from the other groups. The four groups overlap almost entirely suggesting there are strong similarities in the metabolomes of all stages. Whereas the PLS-DA plot in Figure 6b showed only a small amount of overlap between groups 1 and 3, stage 2 overlapped with both group 1 and 3. There was also overlap between stages 3 and 4. Taken together, this indicates that there may be differences in the metabolomes between some of

the stage groups it may not be possible to differentiate between all four of the different stages.

ANOVA was used to define the major sources of variation (Figure 7). There were 16 metabolites found to have significantly different expression levels between the 4 groups, these were shown to be statistically significant but for them to be useful clinical biomarkers they must also reach the threshold for being both sensitive and specific.

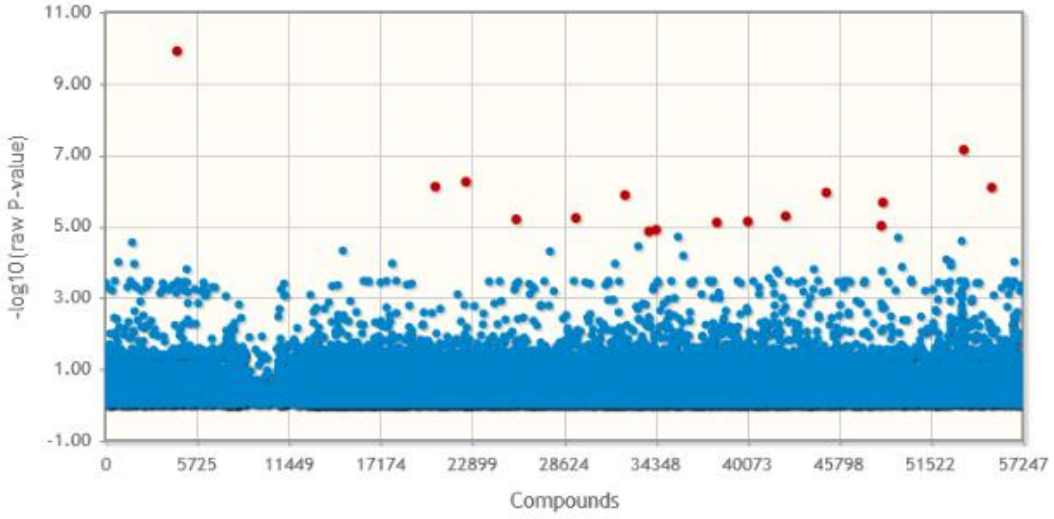


Figure 7: ANOVA off saliva metabolomes of patients at different stage of lung cancer with a P-value cut-off of 0.05, blue plots represent plots which are not significant, and the red plots represent metabolites which show significantly different expression levels between the 4 groups.

Using the data produced by the ANOVA (P values and t scores), enrichment analysis was carried out using the MS Peaks to Pathways module of MetaboAnalyst 4.0 (https://www.metaboanalyst.ca/MetaboAnalyst/Secure/mummichog/LibraryView.xhtml) using a molecular weight tolerance of 10ppm, the negative analytical mode and ranking by P value. The enrichment analysis selected was Mummichog with a P value cutoff of 0.05 and the pathway library used was *Homo sapiens* (human) [BioCyc]. The compound hits were then downloaded and annotated using a set of default ionisation product mass calculation rules (Draper, et al., 2009). The three identified metabolites and the unidentified metabolites with their concentrations from the original matrix were tested using the Biomarker Analysis module and for each discriminant produced the results below (Table 3).

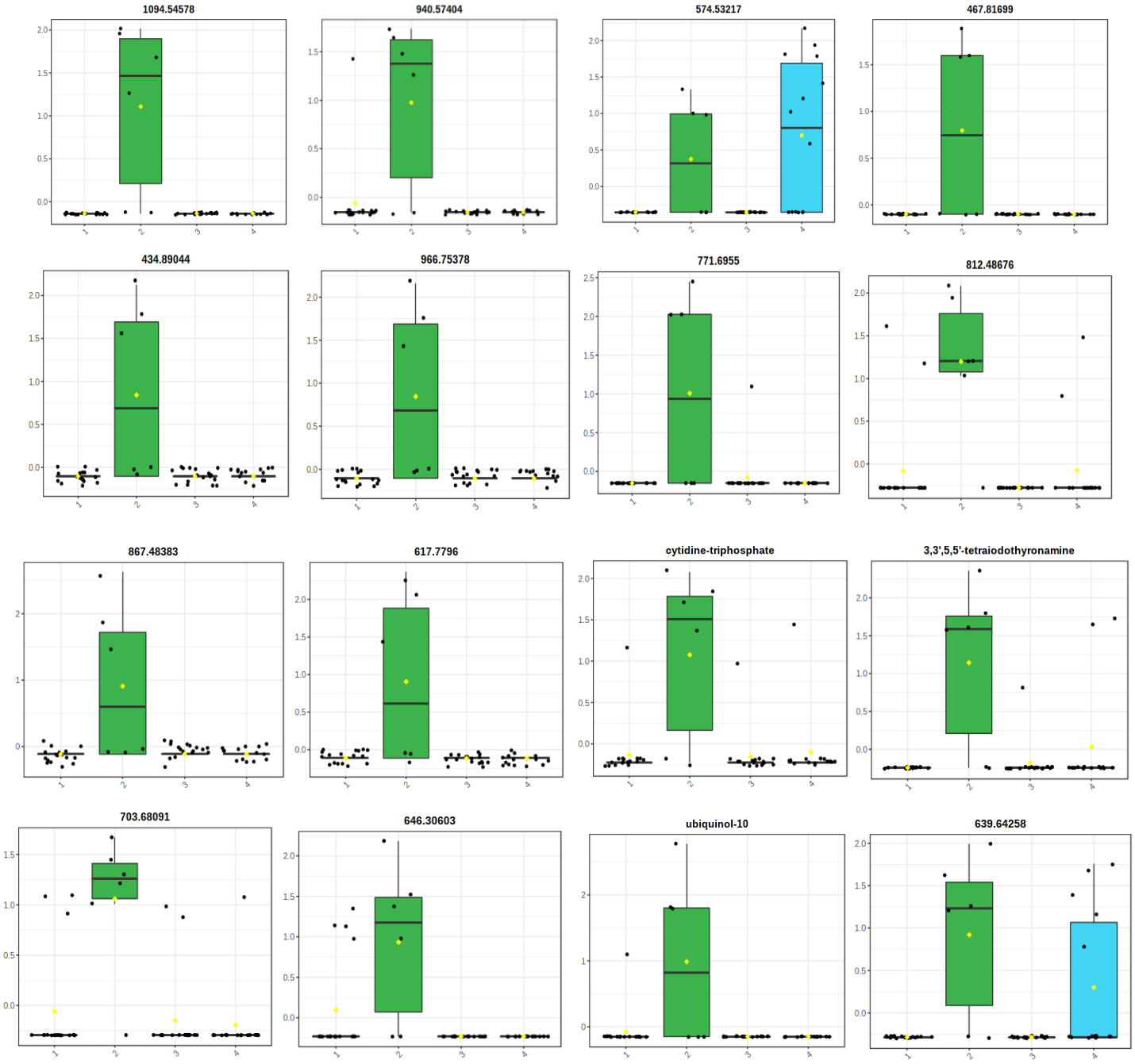


Figure 8: Box plots for the metabolites where significantly different ($p < 0.05$) levels of the metabolite were found between the groups, the red plots representing the stage 1 group, green plots representing the stage 2 group, dark blue representing the stage 3 group and light blue representing the stage 4 group. The groups are shown on the x-axis and the expression levels are shown on the y-axis.

	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC
3,3',5,5'-tetraiodothyronamine	Group 1 - 2	0.83333	Group 1 - 3	0.52941	Group 1 - 4	0.57143	Group 2 - 3	0.82353	Group 2 - 4	0.7619	Group 3 - 4	0.54622
cytidine-triphosphate	Group 1 - 2	0.82353	Group 1 - 3	0.50173	Group 1 - 4	0.5084	Group 2 - 3	0.82353	Group 2 - 4	0.80952	Group 3 - 4	0.5084
ubiquinol-10	Group 1 - 2	0.73529	Group 1 - 3	0.52941	Group 1 - 4	0.52941	Group 2 - 3	0.75	Group 2 - 4	0.75	Group 3 - 4	
1094.54578	Group 1 - 2	0.83333	Group 1 - 3		Group 1 - 4		Group 2 - 3	0.83333	Group 2 - 4	0.83333	Group 3 - 4	
940.57404	Group 1 - 2	0.81373	Group 1 - 3	0.52941	Group 1 - 4	0.52941	Group 2 - 3	0.83333	Group 2 - 4	0.83333	Group 3 - 4	
467.81699	Group 1 - 2	0.75	Group 1 - 3		Group 1 - 4		Group 2 - 3	0.75	Group 2 - 4	0.75	Group 3 - 4	
434.89044	Group 1 - 2	0.75	Group 1 - 3		Group 1 - 4		Group 2 - 3	0.75	Group 2 - 4	0.75	Group 3 - 4	
966.75378	Group 1 - 2	0.75	Group 1 - 3		Group 1 - 4		Group 2 - 3	0.75	Group 2 - 4	0.75	Group 3 - 4	
812.48676	Group 1 - 2	0.86765	Group 1 - 3	0.55882	Group 1 - 4	0.5084	Group 2 - 3	0.91667	Group 2 - 4	0.86905	Group 3 - 4	0.57143
617.7796	Group 1 - 2	0.75	Group 1 - 3		Group 1 - 4		Group 2 - 3	0.75	Group 2 - 4	0.75	Group 3 - 4	
867.48383	Group 1 - 2	0.75	Group 1 - 3		Group 1 - 4		Group 2 - 3	0.75	Group 2 - 4	0.75	Group 3 - 4	
771.6955	Group 1 - 2	0.75	Group 1 - 3	0.52941	Group 1 - 4		Group 2 - 3	0.73529	Group 2 - 4	0.75	Group 3 - 4	0.52941
574.53217	Group 1 - 2	0.75	Group 1 - 3		Group 1 - 4	0.78571	Group 2 - 3	0.75	Group 2 - 4	0.61905	Group 3 - 4	0.78571
703.68091	Group 1 - 2	0.88235	Group 1 - 3	0.53633	Group 1 - 4	0.55462	Group 2 - 3	0.90686	Group 2 - 4	0.89881	Group 3 - 4	0.5084
646.30603	Group 1 - 2	0.76471	Group 1 - 3	0.61765	Group 1 - 4	0.61765	Group 2 - 3	0.83333	Group 2 - 4	0.83333	Group 3 - 4	
639.64258	Group 1 - 2	0.83333	Group 1 - 3		Group 1 - 4	0.67857	Group 2 - 3	0.83333	Group 2 - 4	0.67857	Group 3 - 4	0.67857

Table 3: ROC (receiving operating capacity) curve analysis showing the AUC (area under the curve) value for each of the discriminants for saliva metabolomes of patients at different stage of lung cancer. The green cells show the discriminant and corresponding AUC values which exceed the accepted 0.7 threshold.

Out of the 16 significant metabolites all of them had an AUC value over the 0.7 limit for sensitivity and specificity to be a good biomarker for at least two of the discriminant groups. There were 14 metabolites which had AUC values above the threshold for at least 3 discriminants, and one metabolite (*m/z* 574.53217) which had an AUC value above the threshold for 4 of the 6 discriminants.

The discriminatory metabolites were also plotted using box and whisker plots (Figure 8). This indicated that the Stage 2 metabolomes were particularly distinctive with only two metabolites showing a difference in stage 4 samples. Stages 1 and 3 did not display on change in the levels of discriminatory metabolites.

3.2.2 | Combined LC forms: Early vs. Late Stages

Having noted that stages 2 and 4 appeared to be the most different (Figure 8), the saliva samples were redesignated as either early (stage 1 and 2) and late (stages 3 and 4) to assess if this approach could provide further discriminatory metabolites.

The PCA plot (Figure 9a) showed little evidence of any discrimination between the groups. However, the supervised PLS-DA plot (Figure 9b) showed no overlap between the groups, indicating that they have discrete metabolites which can discriminate between the groups.

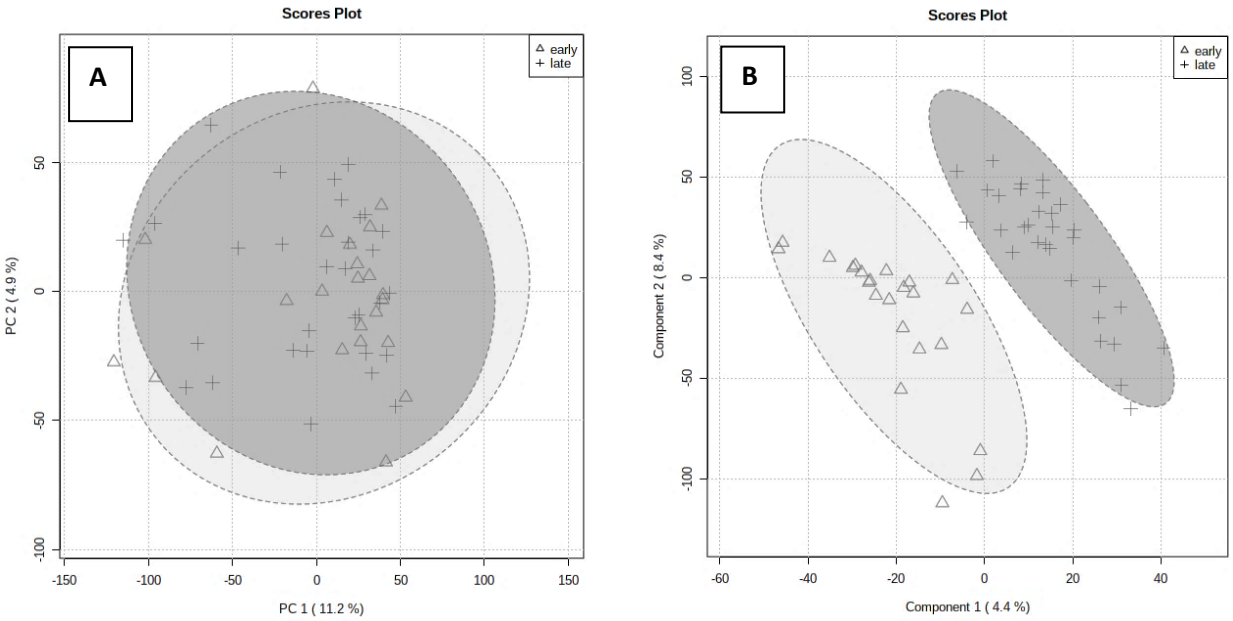


Figure 9: (a) Principal component analysis (PCA) biplot and (b) Partial least squares discriminant analysis (PLS-DA) of the salivary metabolomes of LC patients at early (stage 1 + 2) vs. late (stage 3 + 4) stages. Each plot represents a sample, the triangle plots are associated with the early stage samples and the cross plots represent the late stage samples (stages 3 and 4)

3.2.3 | LC subtype: adenocarcinoma stages 1-4

Further assessments focused on attempting to find metabolites which could be linked to the stages of different subtypes of NSCLC.

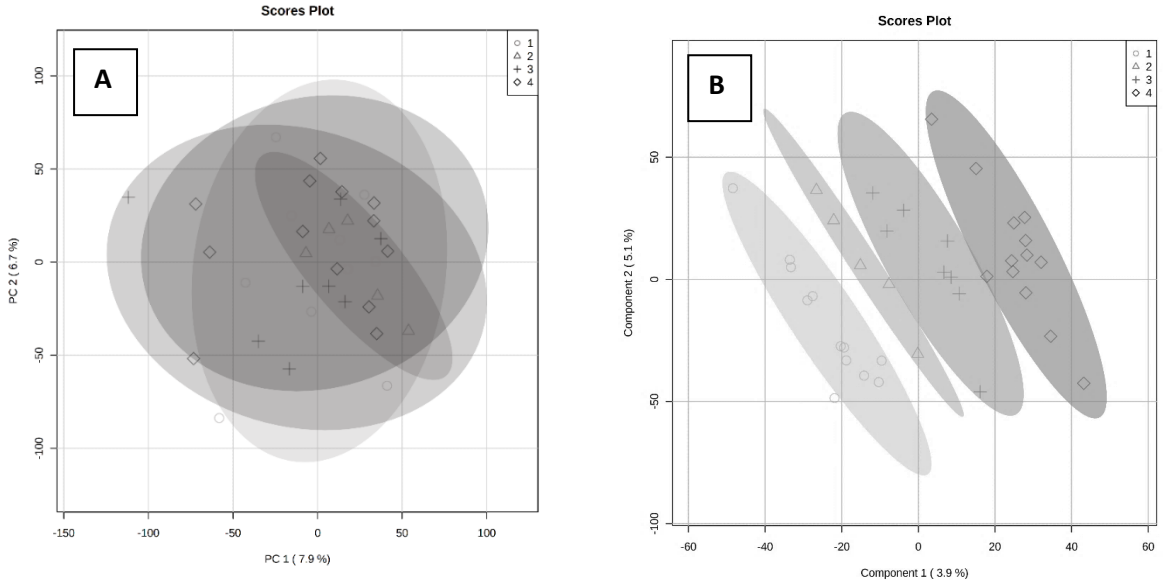


Figure 10: (a) Principal component analysis (PCA) biplot and (b) partial least squares discriminant analysis (PLS-DA) of saliva metabolomes from patients with different stages of lung adenocarcinomas. Each plot represents a sample, the circle plots are associated with the stage 1 samples, the triangle plots represent stage 2 samples, cross plots with the stage 3 samples and diamond plots with the group 4 samples.

The results of the PCA biplot (Figure 10a) showed that the saliva metabolomes of patients with different stages of LC could not be distinguished. PLS-DA plot (Figure 10b) presents the data differently using partial least squares regression. This is a supervised approach where variables which discriminate between the assigned groups are sought out. This analysis showed some overlap in the groups between stages 2 from stage 3, and between stages 3 and 4 however the groups showed some separation into distinct stages. This indicated that discrete metabolites could discriminate between the groups.

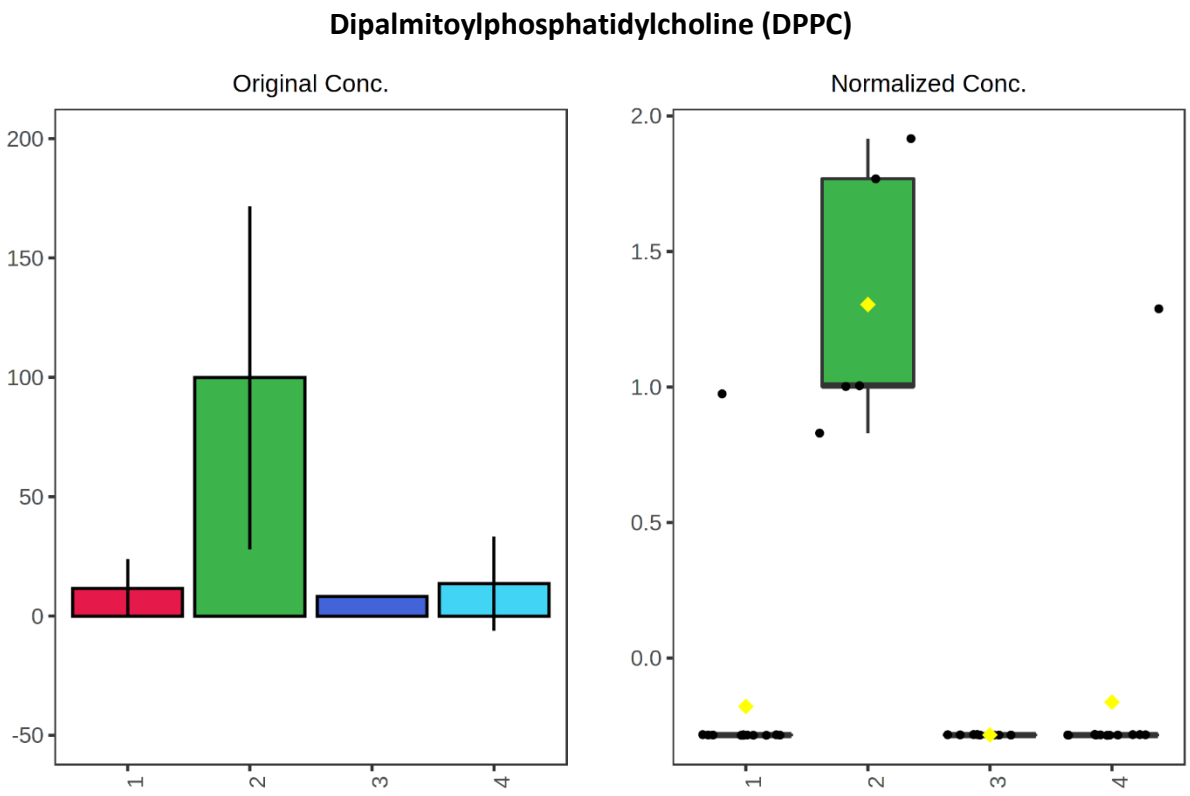


Figure 11: Box plots for the original concentrations and the concentration levels after normalisation for the one metabolite in the saliva metabolome whose content significantly differed. ($p < 0.05$). between patients with different stages of lung cancer. The red, green, dark blue and light blue bars represent the stage 1, 2, 3 and 2 samples, respectively. The groups are shown on the x-axis and the expression levels are shown on the y-axis.

Metabolite	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC	Discriminant	AUC
Dipalmitoylphosphatidylcholine	Group 1 - 2	0.98333	Group 1 - 3	0.54167	Group 1 - 4	0.5	Group 2 - 3	1	Group 2 - 4	0.95385	Group 3 - 4	0.53846

Table 4: ROC (receiving operating capacity) curve analysis showing the AUC (area under the curve) value for each of the discriminants, for the one significant metabolite found. The green cells show the discriminant and AUC values which exceed the accepted 0.7 threshold.

Assessment of the sources of variation in the saliva metabolomes using ANOVA identified only a single metabolite dipalmitoylphosphatidylcholine (DPPC) that was significantly different (Figure 11). Levels appeared to be particularly elevated in the stage 2 salivary metabolomes.

DPPC was then assessed using ROC-AUC and had an AUC of over 0.7 for three of the stage comparisons (staged 1 vs 2; 2 vs 3, 2 vs 4) so it may be useful as potential biomarker for distinguishing between these groups. However, it did not reach the threshold for sensitivity and specificity when comparing stage 1 vs 4.

3.2.4 | Adenocarcinoma LC: Early vs. Late stages

The salivary metabolome of patients with the adenocarcinoma form of LC at different stages were assessed. PCA shows little evidence of any clustering and both groups are almost entirely overlapped, showing strong similarities in the metabolomes (Figure 12a). However, the supervised PLS-DA plot (Figure 12b) could separate between the groups which indicate some metabolites could discriminate between the two groups. A t-test was used to identify these variables, but none were shown to be significantly different using this form of assessment.

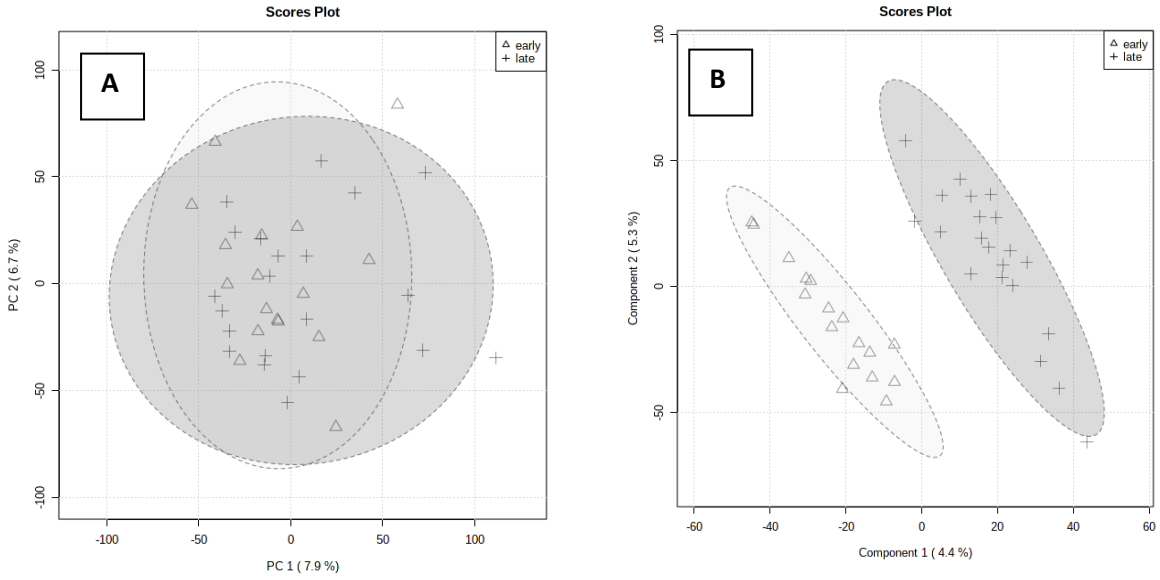


Figure 12: (a) Principal component analysis (PCA) biplot and (b) partial least squares discriminant analysis (PLS-DA) plot showing of saliva metabolomes from adenocarcinoma lung cancer patients at different stages. Each plot represents a sample, the triangle plots are associated with the early stage samples (stages 1 and 2), the cross plots represent the late stage samples (stages 3 and 4).

3.2.5 | Squamous LC: Early vs. Late Stage

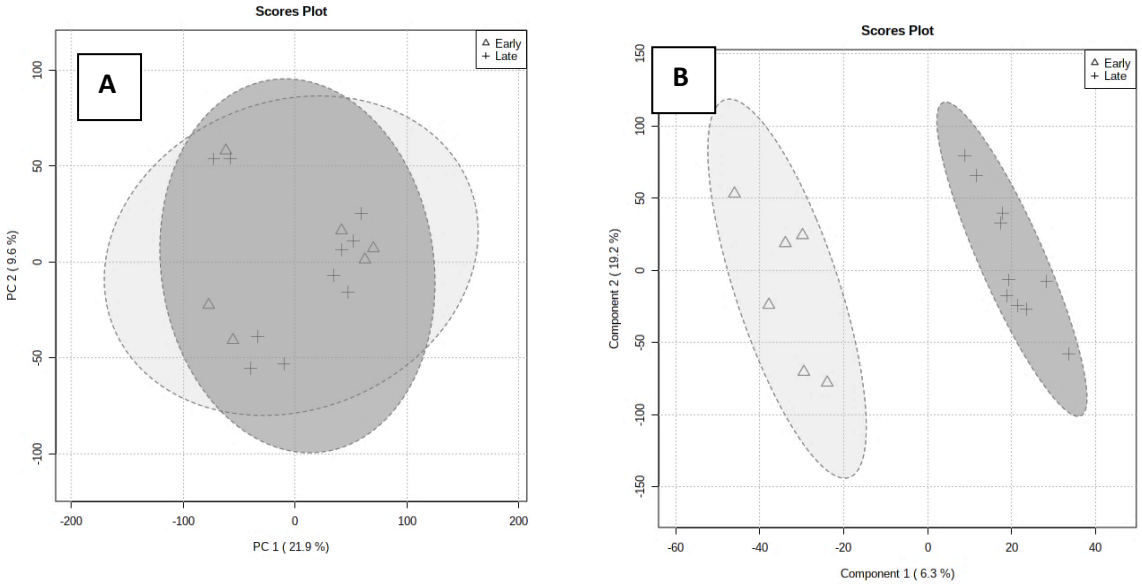


Figure 13: (a) Principal component analysis (PCA) biplot and (b) partial least squares discriminant analysis (PLS-DA) plot of saliva metabolomes from patients with the squamous cell carcinoma form of LC. Each plot represents a sample, the triangle plots are associated with the early stage samples (stages 1 and 2), the cross plots represent the late stage samples (stages 3 and 4).

The salivary metabolomes of patients with the squamous cell carcinoma form of LC at different stages were initially assessed using PCA. Figure 13a shows little evidence of any separate clustering of the early (stages 1+2) and the late (stage 3+4) stages of LC indicating a close similarity in the metabolomes. Whereas, the supervised PLS-DA plot showed no overlap between the groups (Figure 13b). This indicated that they have are some metabolites that could be discriminate between the discrete metabolite expression patterns. A t-test was used to attempt to identify these variables, but none were shown to be significantly different using this form of assessment.

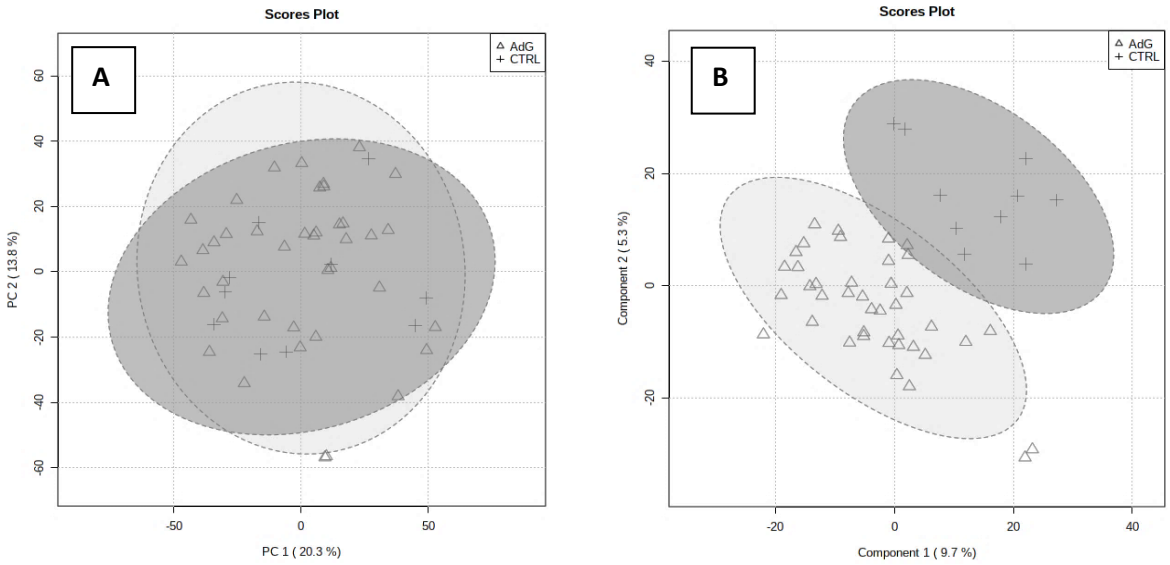


Figure 14 (a) *Principal component analysis (PCA) biplot and (b) Partial least squares discriminant analysis (PLS-DA) of gastric juice metabolomes. Each plot represents a sample, the triangle plots are associated with the gastric cancer group and the cross plots are associated with the healthy control group.*

The results of the PCA biplot (Figure 14a) showed little evidence of separate clustering and there was significant overlap between the two groups. The supervised PLS-DA plot (Figure 14b) showed that the groups could be separated into different clusters, with only minor overlap in the 95% confidence ellipses, suggesting that they show significant features.

Metabolites which significantly differed between the groups were identified using a t-test. These were identified using the Mummichog programme with a P value cut-off of 0.05 and the pathway library used was *Homo sapiens* (human) [BioCyc]. These metabolites are displayed using a heatmap and dendrogram (Figure 15). The dendrograms showed the relationship between the expression patterns of each sample (x- axis) and the concentration levels of the metabolites (y- axis). A longer branch represents a greater degree of difference between the clusters. The x-axis dendrogram showed that samples in the gastric cancer group mostly formed a distinct cluster, which suggested that within the group samples share similar expression patterns. However, there were samples from the other group mixed in among those in the cancer group suggesting that as shown in the PLS-DA plot there is overlap in metabolite concentrations between the groups.

A more focused heatmap showed the relationship between only the top five metabolites and the average concentrations within the two groups (Figure 16). 5-formyl-5,6,7,8-tetrahydromethanopterin, angiotensin I, urocanic acid, prostaglandin, 12-OPDA were the most important metabolites identified, with 5-formyl-5,6,7,8-tetrahydromethanopterin and Angiotensin I being relatively increased in the gastric cancer group and reduced in the control group, and the converse relationship seen with urocanic acid, Prostaglandin and 12-OPDA.

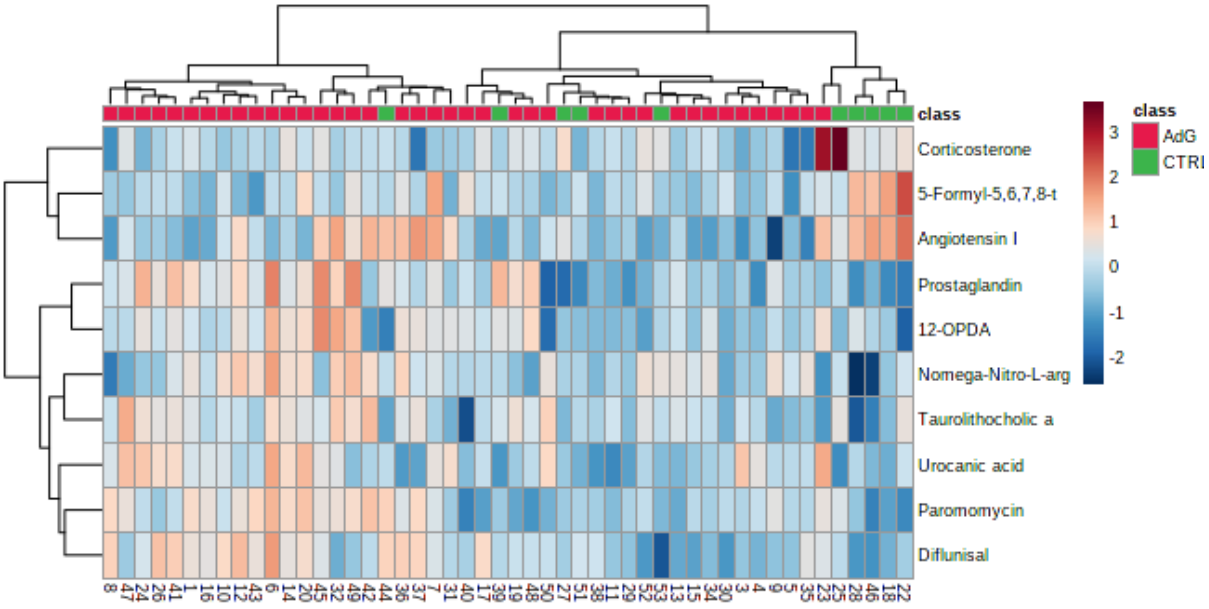


Figure 15: Heatmap comparing metabolite expression across the 50 patient samples of the top 10 metabolites. The sample codes are along the x-axis and metabolites are on the y-axis. Cells which are dark red are highly upregulated, whilst cells which are dark blue are highly downregulated. The key (top right) shows the diagnosis groups of AdG and CTRL, representing the gastric cancer group and the control group respectively.

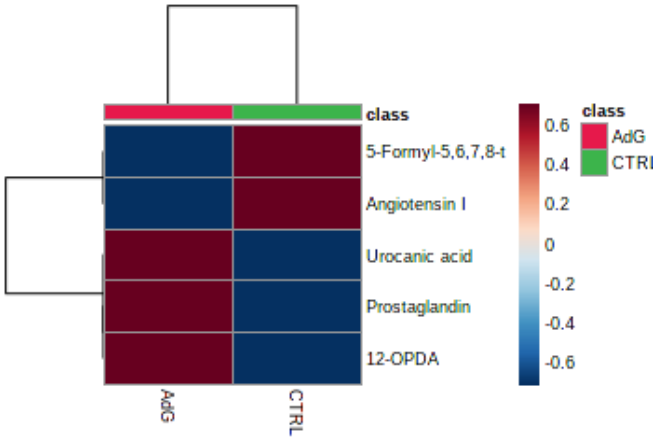


Figure 16: Heatmap comparing metabolite expression as an average of the two groups for the top 5 metabolites. The group names are along the x-axis and metabolites are on the y-axis. Cells which are dark red are highly upregulated, whilst cells which are dark blue are highly downregulated.0The key (top right) shows the diagnosis groups of AdG and CTRL, representing the gastric cancer group and the healthy control group, respectively.

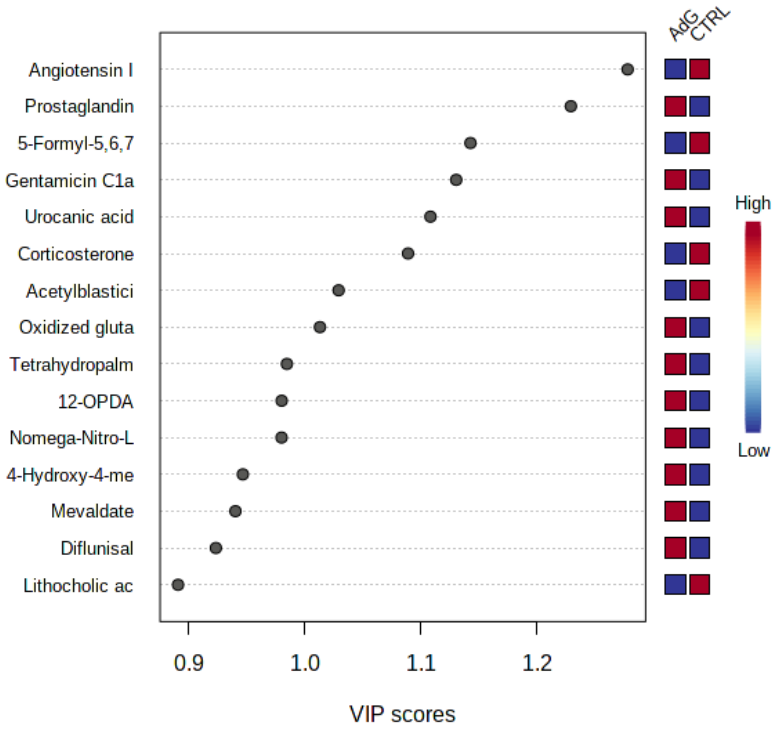


Figure 17: Variable importance in projection (VIP) scores are a weighted sum of squares that considers the amount of explained Y-variance of each component. The metabolites are listed from highest to lowest VIP score. The metabolite expression levels are shown on the right with dark red squares representing high levels of the metabolite and dark blue squares representing low levels of the metabolite.

Focusing on PLS-DA, the key metabolites were assessed based on VIP scores. None of the metabolites met the VIP score value of 1.5 which is the value usually accepted VIP for significance (Akarachantachote, Chadcham, & Saithanu, 2014), however, angiotensin I, prostaglandin, 5-formyl-5,6,7,8-tetrahydromethanopterin, urocanic acid and 12-OPDA also all feature in the top metabolites for their VIP scores as well as being shown to be important from the heatmap. VIP scores also highlighted Gentamicin C1a in the top five most important metabolites based on their VIP score.

A volcano plot was used to identify any important features with a fold-change threshold of 2 and t-tests threshold of 0.05. The data produced from the volcano plot (fold change and t scores) was then used to run enrichment analysis, carried out using the MS Peaks to Pathways module of MetaboAnalyst 4.0 (https://www.metaboanalyst.ca/MetaboAnalyst/Secure/mummichog/LibraryView.xhtml) using a molecular weight tolerance of 5ppm and the negative analytical mode. The enrichment analysis selected was Mummichog with a P value cut-off of 0.05 and the pathway library used was Homo sapiens (human) [BioCyc]. The compound hits were then downloaded and annotated using a set of default ionisation product mass calculation rules (Draper, et al., 2009). The enrichment analysis was used to produce a metabolite set enrichment overview.

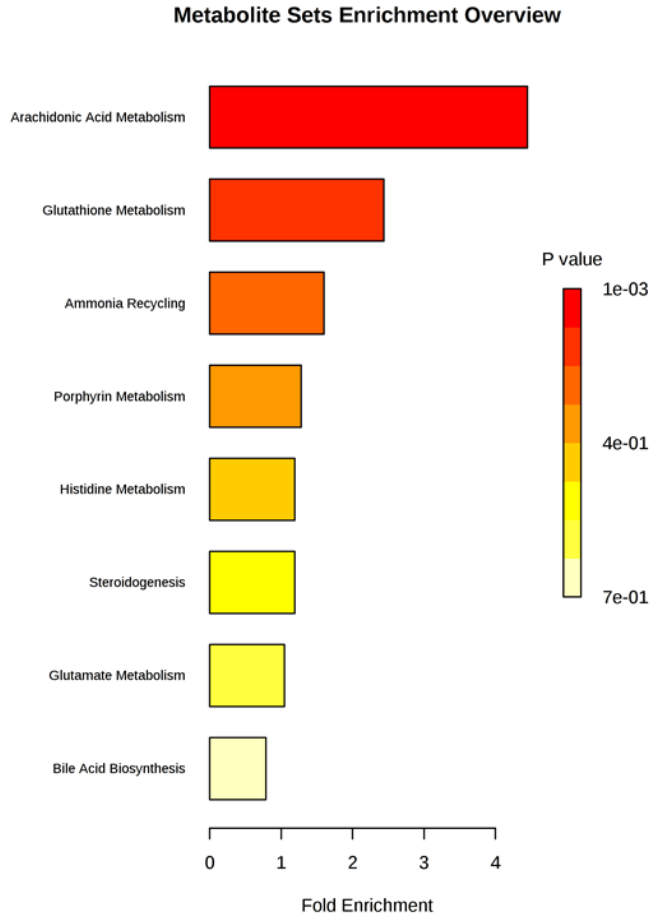


Figure 18: Summary plot for Over Representation Analysis (ORA) using metabolites found at significantly different levels in gastric fluid from patients with gastric cancer than without. ORA is used to evaluate whether a particular metabolite set is represented more than would be expected by chance within the given compound list.

	total	expected	hits
Arachidonic Acid Metabolism	69	1.55	6
Glutathione Metabolism	21	0.47	1
Ammonia Recycling	32	0.72	1
Porphyrin Metabolism	40	0.90	1
Histidine Metabolism	43	0.97	1
Steroidogenesis	43	0.97	1
Glutamate Metabolism	49	1.10	1
Bile Acid Biosynthesis	65	1.46	1

Table 5: Results from Over Representation Analysis.

3.4 | Important Metabolites (Entire Study)

LC vs HC Clinically Significant Biomarkers		LC Staging Clinically Significant Biomarkers				GC vs HC Heatmap	
Higher levels in Lung Cancer	**Higher levels in Control Samples**	**Upregulated in Stage 1**	**Upregulated in Stage 2**	**Upregulated in Stage 3**	**Upregulated in Stage 4**	**Upregulated in Gastric Cancer**	**Upregulated in the Control Group**
D-Sedoheptulose 7-phosphate	L-Methionine		3,3',5,5'-tetraiodothyronamine		574.53217	Urocanic acid	5-Formyl-5,6,7,8-tetrahydromethanopterin
N-Acetyl-D-Glucosamine 6-Phosphate	Trimethylselenonium		cystidine-triphosphate		639.64258	Prostaglandin	Angiotensin I
Inosine	Uridine		ubiquinol-10			12-OPDA	
Chenodeoxycholic acid	Aspartic acid		1094.54578				
Long-chain acyl-CoA	Leukotriene B4		940.57404				
Uridine diphosphate glucose	GalNAc2 GlcNAc2 Ser/Thr1		467.81699				
Dextrin	Hydroxyretinoic acid		43489044				
Deoxyinosine	Malonyl-Carnitin		966.75378				
Uridine diphsophate-N-acetylglucoseamine	2-Oxoisocaproate		812.48676				
Hematoside	L-2-Amino-3-oxobutanoic acid		617.7796				
Ketoleucine	Uridine 5'-diphosphate		867.48383				
Cholesterol sulfate	Urocanic acid		771.6955				
S-Adenosylhomocysteine	Carbon Dioxide		574.53217				
D-Glyceraldehyde	2-keto-glutaramic acid		703.68091				
Capric acid			646.30603				
D-Ribitol 5-phosphate			639.64258				

Galactosylglycerol							
Oxidized glutathione							
Linolenic acid							
7,8-Dihydropteroic acid							
Dehydroepiandrosterone sulphate							
Sulfate							
Sulfuric acid							
5,6-DHET							
Icosanoic acid							
Estrone glucuronide							

Table 6: Biomarkers identified as significant in each of the parts of this study: Clinically Significant Biomarkers in distinguishing lung cancer saliva samples from healthy controls, Clinically Significant Biomarkers in distinguishing different stages of lung cancer and Metabolites which appeared significant by observation of a heatmap produced to investigate gastric cancer against control samples

CHAPTER 4 | Discussion

4.1 | Addressing the hypothesis

This study used metabolomics to find statistically and clinically relevant biomarkers for both Lung cancer diagnosis and staging, as well as exploring the most significant pathways associated with the lung cancer and gastric cancer metabolome. Using metabolomics to not only analyse large data sets quickly and accurately but also to represent not only the genome but also the input from the environment (Wishart, 2016). This study has explored the similarities and differences in metabolite concentration patterns both within and between groups, using multivariate approaches, to produce heatmaps, biplots and boxplots, before identifying and calculating ROC curve AUCs to isolate potentially significant biomarkers.

Firstly, for the use of saliva in identifying lung cancer samples against non-malignant samples 43 metabolites were identified as statistically significant ($P<0.05$). Then to test the sensitivity and specificity of the metabolites found, ROC curves were plotted to find the AUC value for each of the metabolites. There were 40 metabolites which were both statistically significant ($P<0.05$) and with an AUC value over 0.7 to be accepted as potential biomarkers (Table 2). Over representation analysis was then run on the statistically significant metabolites identifying potentially important pathways associated with lung cancer, such as arachidonic acid metabolism, amino sugar metabolism and estrone metabolism (Figure 5).

For investigating the use of saliva in staging lung cancer the analysis was separated into staging adenocarcinoma samples, squamous cell carcinoma samples and combining the two subtypes. For the combined analysis 16 metabolites were identified as statistically significant but none of the metabolites had AUC values high enough when distinguishing between all 4 stages to make them clinically significant in terms of sensitivity and specificity (Table 3). For the adenocarcinoma group DPPC was identified as statistically significant but also failed to reach the required AUC across all 4 stages (Table 4). For both the squamous cell carcinoma group and the adenocarcinoma group when exploring the ability to distinguish early against late stage samples, no metabolites were identified as statistically significant although PLS-DA plots showed good levels of separation (Figure 13).

For the gastric fluid samples, while unable to identify any metabolites which appear to be statistically significant for diagnosis of gastric adenocarcinomas, has explored the metabolites which may be important in differentiating between malignant and non-malignant gastric fluid samples and looked at some of the pathways implicated in gastric cancers based on the metabolites found in the samples, such as arachidonic acid metabolism, glutathione metabolism and ammonia recycling (Figure 18).

Across the study Urocanic acid was found to be a clinically significant metabolite in diagnosis of lung cancer as well as an important feature in the heatmap produced when investigating diagnosis of gastric cancer (Table 6). The pathways shown as potentially important in gastric cancer (Figure 18) also feature in the pathways associated with lung cancer (Figure 5), with 7 out of 8 of the gastric cancer pathways being shared with lung cancer.

4.2| Statistical Analysis

4.2.1 | Lung Cancer vs. Control

The PCA biplot of the saliva metabolomes (Figure 1a) showed two distinct groups. This reflected the many mechanisms specifically associated with LC malignancy (Perlikos, *et al.*, 2013) which appeared to be reflected in saliva. The LC group was seen to have formed a tight cluster, whereas the healthy control group was far more spread out illustrating higher variety in the metabolite expression between the samples, this was likely due to the associated increase in variance having such a limited sample size (n=3). The groups did have a small amount of overlap, although only one sample seemed to have shared characteristics (E080), this could be that the sample was an early stage sample so have expression patterns which are more closely related to the healthy control group due to it being a small, localised tumour (Knight, et al., 2017) looking at the patient data associated with the sample, it is a localised tumour but is in fact a stage 3 lung cancer (appendix 1). With PLS-DA (Figure 1b) the groups did not overlap, indicating that they had discrete metabolite expression patterns, however there were also two samples in the lung cancer group which fell outside of the imposed

grouping, implying that they had higher levels of variation in the metabolite expression compared to the rest of the group.

The significant sources of variation were then identified by t-test and displayed using a heat map (Figure 2). It showed that some of the metabolites were clearly upregulated in the HC group but some only in two of the three control samples. Carbon dioxide, L-2-amino-3-oxobutanoic acid, urocanic acid and uridine-5-diphosphate appeared highly upregulated in two of the HC samples and did not seem to occur in any of the other samples. The following nine metabolites seemed to show more promise as although they were slightly upregulated in a few of the LC samples they were found at much higher concentrations in all samples in the control group. Dendrograms on the x-axis showed that the HC group form a distinct cluster from the LC group further indicating that there is distinct differences between the patient grouping, whereas the LC group then branches off into two distinct branches on the x-axis, with one branch further branching creating three distinct clusters (left, middle and right), showing variation within the group, this may have been due to differing mechanisms depending on the stage of the individual lung cancer samples (Perlikos, *et al.*, 2013). In relation to the patient data the mean BMI values of the participants in the 3 clusters is almost equal though the left hand cluster is slightly higher at 26.43 (left), 24.11 (middle) and 23.59 (right), the mean age of the groups is also very similar at 72.7 (left), 71.1 (middle) and 67.5 (right) as are their mean pack years at 46.1, 40.5 and 42.38 (using numbers from appendix 1). However of the 9 participants in the left hand cluster 8 of them have previously been treated either surgically, pharmacologically or with radiotherapy, compared to 5 out of 9 having been treated in the middle cluster and only 5 out of 9 having been treated in the right hand cluster. This higher proportion of treatment may be what is leading to the separation on the x-axis as the mechanisms of the malignancy have been disrupted.

4.2.2 | Staging of all lung cancer saliva samples

The full data set was then analysed containing samples from both patients with adenocarcinomas and patients with squamous cell carcinomas. The ANOVA (Figure 7), produced using multivariant analysis of the 4 stages of both adenocarcinoma and squamous cell carcinoma, shows that 16 of the metabolites are statistically significant ($P<0.05$) across

the groups. The PCA (Figure 6a) shows the four groups almost entirely overlapping suggesting there are strong similarities of the metabolite expression between the groups, this is expected as genetic and epigenetic events such as multiple genes and enzymes associated with DNA and histone modifications being altering during both cancer initiation and progression (Chakravarthi, *et al.* 2016) . The PLS-DA plot presents the data differently using partial least squares regression, a supervised approach where variables which discriminate between the assigned groups are sought out (Gromski, et al., 2015). The advantages of PLS-DA include its ability to handle noisy data, which is very common in outputs from metabolomics experiments (Want & Masson, 2011) as well as producing a scores plot to illustrate the separation between different groups which is easily interpretable (Worley, *et al.*, 2013). The PLS-DA plot produced here for the four stages (Figure 6b) shows while that there may be differences in the expression of the metabolites between some of the groups, there are also similarities so while there may be significant differences in the concentrations of some of the metabolites between the groups they may not be able to differentiate between all four of the different stages.

4.2.3 | Adenocarcinoma Staging

The PCA biplot (Figure 6a) showed significant overlap between all four groups showing similarities in the metabolite concentration patterns between the groups, this could be explained by the fact that although the tumours are at different stages they are all adenocarcinomas, this is supported by the fact Hori *et al.* (2011) produced PLS-DA plots which exhibited discrimination among the groups of each histological type and similarities within the groups (Hori, et al., 2011). However, there have been several metabolites found in serum which were found to be differentially expressed in early stage compared to advanced stage lung cancers indicating alteration in metabolite expression during disease metastasis and increased inflammation (Hori, et al., 2011) therefore it would be expected for there to at least be discrimination between the early stages compared to the advanced stages. The PLS-DA plot (Figure 6b) which uses the groupings given by the investigator to seek out components which do separate out the groups looks very different, showing mostly tight clustering between all four groups, with the stage 1 groups showing no overlap with the others and the stage 2 group showing only a small amount of overlap with the stage 3 group, which appears

to be driven mostly by one sample which can be seen bordering the two groups, this may be due to the fact that while lung cancers are grouped into discrete stages, depending on how long the tumour has been at that stage and the rate at which the tumour is growing, the tumour in that sample may be preparing to metastasise and therefore share some metabolite expression with the samples in the stage 3 group. The discrete metabolite concentrations shown in the group 1 compared to the group 2 are potentially linked to the increased levels of proliferation as the tumour grows in size associated with the stage 2 samples compared to the stage 1 samples where it would be expected to find more metabolites linked to initial malignancy pathways.

The metabolite concentrations in the stage 2 group (Figure 6a and 5b) were the most clustered and therefore had least variation in the metabolite concentrations within the group, this could be due to a number of factors such as the samples being taken which the progression of the stage 2 tumours was most similar, the patients from which the samples were taken may have more similar habits, there are many things which can affect the metabolite concentration in the saliva such as BMI, biological sex, common eating habits and even the time of day when the samples are taken (Mikkonen, et al., 2015) and while in this study we have tried to account for as many of these variants as possible, having the samples collected at the same time of day, asking the participants not to eat the day the samples were collected, with human sampling it is not possible to negate them all.

The PCA biplot for the adenocarcinoma group split into early against late stage (Figure 12a) shows considerable overlap of the two groups, much the same as the between the four groups within the adenocarcinoma group (Figure 6a), which again may be due to the overarching similarities from them all being adenocarcinomas. The PLS-DA plot (Figure 12b) analysing the labelled data shows no overlap between the early and late stages, which suggests that they do exhibit differences in metabolite patterns between the two groups.

4.2.4 | Squamous cell carcinomas

In comparison the squamous cell carcinoma samples were only able to be separated into early (stages 1 and 2) and late (stages 3 and 4) due to the numbers of patients who presented with squamous carcinomas in stage 1 and 4 and agreed to take part in the study were too low to carry out the statistical analysis. The PCA biplot (Figure 13a) for the early and late stage

squamous cell carcinoma groups shows similarities with that of the PCA for the adenocarcinoma samples (Figure 12a), with both groups having significant overlap representing similarities in the metabolite expression found in the two groups. This may also be attributed to the fact that although the tumours are at different stages they are all the same type, in this case squamous cell carcinomas, which is again supported the fact Hori *et al.* (2011) produced PLS-DA plots which exhibited discrimination among the groups and clustering within the histological groups which shows similarities within the groups which appears to be independent of the stages of the malignancies (Hori, et al., 2011). The PLS-DA biplot however does once again show that when taking into account the groupings assigned to the samples, in this case early and late, to seek out components which discriminate between the groups, the groups show no overlap at all (Figure 13b).

4.2.5 | Diagnosis of Gastric Cancer

The PCA biplot (Figure 14a) showed many similarities in concentration patterns between the groups. This may be due to a significant part of the gastric fluid's biochemistry being associated with its function in the body and there being a large variety of compounds found in gastric juice, such as, peptides, nucleic acids and mucin (Freeman & Kim, 1978) in addition to salivary components arising from swallowing, inflammatory mediators or blood from damaged gastric walls (Hsu, et al., 2007). The PLS-DA plot (Figure 14b) allowed some discrimination between the groups and the key metabolites were identified. On the heatmap (Figure 15) the x-axis dendrogram shows samples in the gastric cancer group mostly form a distinct cluster meaning that the groups have unique patterns and although the PLS-DA plot does appear to show overlap, when removing the ellipses the groups are distinctive.

4.3| Biomarker Analysis

4.3.1 | Lung Cancer vs. Control

Three metabolites with VIP scores above the 1.5 threshold were all found at higher concentrations in the healthy control group than in the lung cancer group. L-methionine, uridine and leukotriene B4 may therefore be potential biomarkers for distinguishing between

lung cancer and non-lung cancer samples but this cannot be confirmed from the VIP scores alone. Methionine could be an important biomarker as it has established roles in cancer formation via a range of biochemical activities including influencing methylation and nucleotide metabolism (Sanderson, *et al.*, 2019). LC cells are associated with abnormal DNA methylation (Greenberg, et al., 2007). It is thought that the abnormal DNA methylation reaction could be associated with decreased level of methionine in lung cancer (Ni, *et al.,* 2019), which would explain the results found in this study. An impact on nucleotide metabolism could be reflected in the nucleotide base uridine being also targeted in our analyses.

From the multivariate statistical analysis carried out on the data, 43 metabolites were found to be significant ($p<0.05$). Of those 43 metabolites, 3 of them were found to be unsuitable for use as biomarkers due to their specificity and sensitivity being too low, this was found by running ROC (receiving operating capacity) curve analysis and finding the AUC values for each metabolite (Table 2). The remaining metabolites which were both statistically significant ($P<0.05$) and had potential for use as biomarkers were used not only for investigation of potential biomarkers but also for furthering our knowledge of the pathways and mechanisms to which these metabolites belong and their place in lung cancer.

These included methionine with a 100 % accuracy in distinguishing between HC and LC samples. As well as L-methionine, S-Adenosylhomocysteine (SAH) was also both statistically ($P<0.05$) and clinically significant as a potential biomarker due to its AUC value of 0.889 (Table 2) and it was found in higher abundance in the lung cancer group (Figure 4). It is found in the methionine cycle, where methionine undergoes demethylation to form homocysteine and produces Adenosylmethionine for methylation reactions (Barroso, *et al.*, 2017). S-adenosylhomocysteine is one of the two intermediate compounds and is formed from the reaction in which the methyl group of adenosylmethionine (SAM) is donated to a methyl acceptor, it is then hydrolysed into adenosine and L-homocysteine by adenosylhomocysteine hydrolase (Chiang, et al., 1996). S-Adenosylhomocysteine has also already been found as a potential biomarker, both in oral cancer (Ishikawa, et al., 2016) and in vascular disease (Castro, et al., 2003). The L-homocysteine produced in this cycle is then either remethylated back to methionine or condensed with serine by cystathionine ß-

synthase to form cystathionine which when further metabolised in the associated transsulfuration pathway gives rise to glutathione (Chiang, et al., 1996). In our study we have found oxidized glutathione to be both statistically ($P<0.05$) and clinically (AUC = 0.815) (Table 2) significant as a potential biomarker, and that it occurs in higher concentrations in the lung cancer group compared to the control group (Figure 4); this corroborates findings published by Tsao, *et al.* in 2007, where they also found it to be significantly increased in patients with non-small cell lung cancer (Tsao, *et al.*, 2007). As well as some of the L-homocysteine giving rise to the glutathione, the adenosine released from the cycle is then either phosphorylated to become AMP by adenosine kinase or deaminated to form inosine by adenosine deaminase (Chiang, et al., 1996).

In this study inosine was also found to discriminate between the LC and HC groups, with an accuracy of 97.5% (Table 2), further supporting the link between NSCLC development and alterations in DNA methylation. Inosine has immunomodulatory and neuroprotective effects. It assists in mast cell degranulation to release antimicrobial peptides, slows the production of pro-inflammatory mediators by macrophages, lymphocytes and neutrophils. It has shown to be protective in sepsis, ischemia-reperfusion, and autoimmunity in animal models. Inosine also helps preserve glial and neuronal cells during hypoxia and encourage axonal regrowth post injury. Its action is thought to work via adenosine receptors. Inosine is thought to be usable to treat tissue damage resulting from ischemia or inflammation (Hasko, *et al.*, 2004). Inosine was one of 13 influential (p-value <0.05) biomarkers from serum samples for lung cancer. 6 biomarkers (hypoxanthine, inosine, L-tryptophan, indoleacrylic acid, acyl-carnitine C10:1, and lysoPC (18:2)) selected together from serum samples were found to have a combined AUC of 0.99, sensitivity of 0.98 and specificity of 0.99 (Ruiying, et al., 2020). The major LC-subtype non-small cell lung cancer (NSCLC) group had <10 % the serum inosine of the control group which corroborates other research about lung cancer where it has also been shown that Inosine is down regulated in lung cancer patients (Kumar, *et al.*, 2017). However, this study, also investigated NSCLCs and it was found that amongst our samples there was a higher concentration of inosine in the LC group compared to the control group (Figure 4). This may have been due to the high variance arising from such a small number of healthy control samples or the fact that the concentrations of inosine associated with lung cancer may differ from serum to saliva. As well as inosine, deoxyinosine, another product of the conversion of

adenosine and deoxyadenosine, catalysed by T-lymphocyte enzyme adenosine deaminase (ADA) (Tang, et al., 2019) was also both statistically ($P<0.05$) and clinically (AUC = 0.920) significant in this lung cancer study (Table 2) and was also found at higher concentrations in the lung cancer group compared to the control (Figure 4).

Carbon dioxide (CO_2) was also found to be both statistically significant and have an AUC of 0.833 (Table 2). Patients' with lung cancer often present with dyspnoea, which is known to be associated with increased serum levels of CO2 (Hayen, *et al.*, 2015), this is likely to be because their lungs are no longer able to effectively excrete CO2 in the breath, therefore, CO2 would be found at lower levels in the saliva of patients with lung cancer compared to healthy controls.

Other metabolites are considered below.

Butyric acid is the functional parent of L-2-amino-3-oxobutanoic acid (ChEBI, 2020). In terms of diagnostic properties the use of a propionic acid/butyric acid ratio in stool samples showed a significant difference between subjects with and without irritable bowel syndrome (IBS) (Farup, *et al.*, 2016). Within a cancer malignancy context,-amino-n-butyric acid was one of twenty-five amino acids found to occur at significantly lower levels in prostate cancer serum and urine samples compared to controls with an AUC value of 0.932 (Derezinski, *et al.*, 2017). In this current study L-2-amino-3-oxobutanoic acid was found to have an AUC slightly lower at 0.833 (Table 2) but still well above the 0.7 threshold. However, it was found to be at higher concentrations the saliva of the lung cancer group (Figure 4) as compared to the lower concentrations in the prostate cancer biofluids (Derezinski, *et al.*, 2017).

Uridine 5'-diphosphate (UDP) was found to be reduced at a significantly slower rate in healthy liver than in the Morris 7793 hepatoma (Blocker & Roth, 1977). In this current work, uridine 5'-diphosphate was found in higher concentrations in the saliva in the HC group compared to the LC (Figure 4). The differences between the concentrations was statistically significant (0.05) and had an AUC of 0.833 for specificity and sensitivity (Table 2). Uridine diphosphate glucose and uridine diphosphate-N-acetylglucosamine were also statistically ($P<0.05$) and clinically significant metabolites in our data. Uridine diphosphate glucose had an AUC of 0.926 and uridine diphosphate-N-acetylglucosamine an AUC of 0.920 (Table 2), making both of

them biomarker candidates for sensitivity and specificity but these would need to be validated in a much larger study (Holland, 2016). The concentrations of uridine diphosphate-N-acetylglucosamine have been found to be higher in cells treated by cancer-drug cisplatin, indicating it as a possible biomarker for monitoring treatment response in lung cells (Duarte, et al., 2013). The boxplots showed that both uridine diphosphate-N-acetylglucosamine and uridine diphosphate glucose are both found in higher concentrations in the lung cancer group, whereas uridine 5'-diphosphate was found at higher concentrations in the healthy control group (Figure 4). Uridine itself was also found to be statistically ($P<0.05$) and clinically significant (AUC = 0.988) (Table 2) and was found to be at higher concentrations in the healthy control group compared to the lung cancer group.

Selenium is known to be a growth regulatory, too high a concentration of selenium and cell growth is inhibited but too little and the cell is unable to survive (Selenius, *et al.,* 2010). It is thought that selenium may possess cancer preventative affects by inducing apoptosis in initiated cells (Combs & Gray, 1998). Hydrogen selenide is a precursor for the synthesis of selenocysteine and can be methylated to trimethylselenonium cation (CH3)3Se+. Trimethylselenonium cation is believed to be a volatile metabolite which when the capacity for it is exceeded is exhaled from the lungs (Mudipalli & Zelikoff, 2017). In this study trimethylselenonium concentrations were found to be higher in the healthy control group than in the lung cancer group (Figure 4), which may mean the healthy control group were exceeding the Trimethylselenonium limit and so more was being exhaled and dissolved into the saliva. It was found to be a statistically and clinically significant biomarker with an AUC of 1.0 (Table 2).

Carnitine is involved in transporting fatty acids across the mitochondrial membrane. It is linked to the fatty acid acetyl groups and transported by carnitine palmitoyltransferase I and carnitine palmitoyltransferase II (Flanagan, *et al.*, 2010). This transport is linked to b-oxidation and the formation of acetyl-CoA. Malonyl carnitine was also found to be both statistically ($P<0.05$) significant and have an AUC of 0.926, well above the 0.7 cut-off and was found at a higher concentration in the healthy control group compared to the lung cancer group. Previously implicated in malignancy malonyl carnitine is one of several metabolites' indicative of apoptotic processes in cell cultures and it is these specific metabolic features of cancer cells

make good targets for drug therapies with metabolites promoting apoptosis in tumours considered one of the strongest tools for cancer therapeutics. (Halama, *et al.*, 2011). Long-chain acyl-CoA had an AUC of 0.938 (Table 2) and was also found at higher concentrations in the lung cancer group compared to the control group (Figure 4). This suggested an increase in lipid metabolism in malignancy as both carnitine and acyl-coA are constituents of the process, with acyl-CoA being converted to acylcarnitine with the addition of carnitine by the carnitine palmitoyltransferase I enzyme. The acyl-carnitine is then translocated into the mitochondria where the acyl-carnitine is converted back to acyl-CoA which then undergoes ß-oxidation to form Acetyl-CoA ready to enter the tricarboxylic acid (TCA) cycle which is essential for adenosine 5-triphosphate (ATP) production as a source of energy for the cells (Nakagawa, et al., 2018). This also tied in with the second theory of cancer, which suggests that cancer cells survive through symbiosis, that one cancer cell produces lactate with ATP production by consuming glucose (Warburg effect) so that their neighbouring cells can take in the lactate to produce its own ATP through the TCA cycle and oxidative phosphorylation (Kim, 2018).

(GalNAc)1 (GlcNAc)2 (Ser/Thr)1 is a glycan associated with mucin type O-glycan biosynthesis (KEGG, 2020). Glycoproteins with o-linked carbohydrate chains are often found in secretions on the surface of cancer cells, with the unusual structures documented as contributing to the biology of the cancer cell. Some of the mechanisms in the O-glycan biosynthesis pathways have been determined in cancer models but it is a complex process which is not well understood. This is important as therapeutic options have been produced to target cells via their glycans and cancer diagnosis has been based on appearance of certain epitopes (Brockhausen, 1999). In this current study, it was found at significantly ($P<0.05$) lower concentrations (Figure 4) in the lung cancer group and had an AUC for sensitivity and specificity of 0.951 (Table 2).

Hematoside, a ganglioside, NeuAc(α2-3)Gal(β1-4)Glc(β1-1)Cer, that is found in red blood cells (Cammack, et al., 2008) also showed itself to be significant. It was in higher concentration in the lung cancer group (Figure 4) and was both statistically significant ($P<0.05$) and had an AUC of 0.920 for sensitivity and specificity (Table 2).

Aspartic acid was another potential biomarker for the diagnosis of lung cancer, at higher concentrations in the control group (Figure 4), it was statistically significant ($P<0.05$) and had an AUC of 0.975 (Table 2). This was not unexpected as it had previously been found as a potential biomarker for lung cancer, both as one of 13 influential (p-value <0.05) biomarkers from serum and as one of 27 influential biomarkers (p-value <0.05) in plasma (Kumar, *et al.,* 2017), however, it was exciting to also find it in saliva as further evidence that saliva is a good biofluid for biomarker discovery metabolomics.

Chenodeoxycholic acid has been found to promote carcinogenesis and could encourage invasion of colon cancer cells as exposure to chenodeoxycholic acid led to increased PGE2 synthesis which has been linked to tumour invasion (Wu, *et al.*, 2018). This fits well with the results in Figure 4, where chenodeoxycholic acid was found at higher levels in the lung cancer group. Lui, *et al.,* (2019) also found chenodeoxycholic acid in higher concentrations in non-small cell lung cancer than the control and found it to be linked with tumour cell migration (Lui, *et al.*, 2019). However, in this study, we have also now shown that as well as being implicated in malignancies chenodeoxycholic acid may be a significant novel biomarker for lung cancer ($P<0.05$) with an AUC of 0.975 (Table 2).

Capric acid was also found in higher concentrations in the lung cancer group (Figure 4) and is also both statistically ($P<0.05$) and clinically (AUC = 0.840) (Table 2) significant in this investigation. Capric acid has previously been implicated in malignancy, with concentrations of capric acid found at significantly lower levels in polycystic ovary syndrome and endometrial cancer compared to controls (Shafiee, et al., 2020). As well as basal cell carcinoma patients having lower levels of Capric acid in their red blood cells than the controls (Rahrovani, et al., 2018).

Ketoleucine has been previously found as a saliva-based biomarker. In oral squamous cell carcinoma patients ketoleucine was found to be up regulated and useful as a biomarker (Song, et al., 2020). In our study, ketoleucine was found in higher concentrations in the LC group (Figure 4) and was also both statistically ($P<0.05$) and clinically (AUC = 0.914) (Table 2) significant suggesting it will also be useful as a potential biomarker for LC from saliva. Ketoleucine is already a serum-based biomarker for LC, where Ketoleucine was found to be at lower abundance in the cancer group (Ros-Mazurczyk, et al., 2017).

D-Glyceraldehyde is one of the glycolytic metabolites elevated by sphingosine kinase 1 (SK1) treatment, which suggests that it accumulates due to inhibition of the glycolytic pathway. This was found in the first study which reported SK1 involvement in regulating the Warburg effect in cancer cells (Watson, et al., 2013). The data in this section showed a higher concentration of D-glyceraldehyde in the lung cancer group (Figure 4). It also showed it to be both statistically ($P<0.05$) and clinically (AUC = 0.864) (Table 2) significant.

N-acetyl-D-glucosamine 6-phosphate is an amino sugar (Inokuma, *et al.*, 2018). The results of this work showed that N-acetyl-D-glucosamine 6-phosphate was at higher concentrations in the lung cancer group (Figure 4) and that it was also both statistically ($P<0.05$) and clinically (AUC = 1.0) (Table 2) significant.

Cholesterol sulfate was also found at higher concentrations in the lung cancer group (Figure 4) and it was also both statistically ($P<0.05$) and clinically (AUC = 0.901) (Table 2) significant. Cholesterol sulfate has previously been shown to be increased in metastases, which indicates cholesterol sulfate is involved with propagating metastasis (Johnson, et al., 2017).

Linolenic acid was found to be both statistically ($P<0.05$) and clinically (AUC = 0.796) (Table 2) significant and was at higher concentrations in the lung cancer group (Figure 4). However, linolenic acid has previously been found to inhibit HIF1α so lead to suppression of hypoxia-induced proliferation and invasion in non-small cell lung cancers (Wang, *et al.*, 2020).

7,8-dihydropteroic acid was also found at higher concentrations in the lung cancer group (Figure 4) and it was also both statistically ($P<0.05$) and clinically (AUC = 0.790) (Table 2) significant as a potential biomarker. 7,8-Dihydropteroic acid was previously used as part of a metabolite panel to discriminate between bladder cancer and renal cell carcinoma in plasma samples (Liu, et al., 2020).

Sulphate was also found at higher concentrations in the lung cancer group (Figure 4) and it was also both statistically ($P<0.05$) and clinically (AUC = 0.753) (Table 2) significant as a potential biomarker. However as the AUC value was slightly lower it may not be as useful in practise than some of the other metabolites with higher AUC values, as it would give more false results. This stated., it can be made more sensitive and specific when used in

corroboration with other metabolite biomarkers. Sulfuric acid, like sulfate, was found at higher concentrations in the lung cancer group (Figure 4) and it was also both statistically ($P<0.05$) and clinically (AUC = 0.741) (Table 2) significant but with an even lower AUC value, which again while above the threshold for significant biomarkers may limit is uses in practise (Askenazi, et al., 2012). Dehydroepiandrosterone sulfate, like sulfate and sulfuric acid, was also found in higher concentrations in the lung cancer group (Figure 4) and is both a significantly ($P<0.05$) and potentially clinically useful biomarker as it had an AUC value of 0.784 (Table 2).

Arachidic acid has been implicated with the respiratory system, Seo, *et al.,* (2018) showed that levels of stearic acid and arachidic acid were significantly lower in serum of mice with induced asthma (Seo, et al., 2018). This study showed that arachidic acid is found at higher levels in the lung cancer group than the control group (Figure 4) but does not have good potential for use as a biomarker as it's AUC value fell below the accepted 0.7 cut-off (Table 2).

2-oxoisocaproate however has an AUC of 0.901 (Table 2) so is both statistically significant ($P<0.05$) and potentially a useful clinical biomarker as its AUC value is over the threshold. This metabolite is found in lower concentrations in the lung cancer group compared to the healthy control group. Hao, *et al.,* (2016) found that metabolites which included as 2-opxoisocaproate were relatively higher in patients with adenocarcinomas, whereas lactate was comparatively lower. Comparing adenocarcinomas with squamous cell carcinomas suggested that there is a relationship between 2-oxoisocaproate and malignancy (Hao, et al., 2016).

Hydroxyretinoic Acid was also found at lower concentrations in the lung cancer group compared to the healthy control group (Figure 4) but is still statistically significant ($P<0.05$) and it has an AUC value of 0.938 (Table 2).

D-sedoheptulose 7-phosphate was found at higher concentrations in the lung cancer group (Figure 4) and it is also both statistically ($P<0.05$) and clinically (AUC = 1.0) (Table 2) significant as a robust biomarker with a very high AUC value. D-sedoheptulose 7-phosphate is a key intermediate molecule of primary glucose metabolism. It is an important intermediate of the non-oxidative pentose phosphate pathway (PPP) and its bioavailability contributes to cellular

carbon flux (Nagy & Haschemi, 2013). Malignant cells are known for altering their metabolisms in order to support their rapid growth and spread across the body. In particular a switch to aerobic glycolysis is often observed, the increased amount of glucose then used in the glycolytic pathway generates larger numbers of metabolic intermediates, one of these intermediates is glucose 6-phosphate which is then utilised in the PPP to allow rapid nucleic acid synthesis and therefore increases the speed of DNA replication (Annibaldi & Widmann, 2010). This explains the increased levels of D-sedoheptulose 7-phosphate in the lung cancer group (Figure 4).

D-Ribitol 5-phosphate was also found at higher concentrations in the lung cancer group (Figure 4) and it is also both statistically ($P<0.05$) and clinically (AUC = 0.840) (Table 2) significant.

Galactosylglycerol was also found at higher concentrations in the lung cancer group (Figure 4) and is also both statistically ($P<0.05$) and clinically (AUC = 0.827) (Table 2) significant. Estrone glucuronide was also found at higher concentrations in the lung cancer group (Figure 4) and is also both statistically ($P<0.05$) and clinically (AUC = 0.728) (Table 2) significant, but it has the lowest AUC value of the metabolites which made it over the threshold.

2-keto-glutaramic acid is found at lower concentrations in the lung cancer group compared to the healthy control group (Figure 4) but is still statistically significant ($P<0.05$) and it has an AUC value of 0.796 (Table 2). 2-keto-glutaramic acid is found in the Alanine, aspartate and glutamate metabolism, part of amino acid metabolism, as an intermediary between the oxoglutaric acid from the TCA cycle and L-Glutamine (Li, et al., 2017). Amino acids are not only direct substrates for protein synthesis, essential for cancers to sustain their high proliferation rate, but also have roles in energy generation and maintenance of cellular redox homoeostasis so it is important for cancers to have an abundant supply (Vettore, *et al.*, 2020).

Leukotriene B4 was also found at lower concentrations in the lung cancer group compared to the healthy control group (Figure 4) but is still statistically significant ($P<0.05$) and it has an AUC value of 0.975 (Table 2). Leukotriene B4 is already associated with severe inflammation and is has been found previously as a potential biomarker for community acquired pneumonia in children found in their exhaled air (Carraro, et al., 2008).

5,6-Dihydroxy-8,11,14-icosatrienoic acid (5,6-DHET) was also found at higher concentrations in the lung cancer group (Figure 4) and is also both statistically ($P<0.05$) and clinically (AUC = 0.741) (Table 2) significant, however as the AUC value is slightly lower it may not be as useful in practise than some of the other metabolites with higher AUC values. 5,6-DHET is one of the dihydroxyeicosatrienoic acids (DHETs) formed from the arachidonic acid (AA) cascade, there are 3 pathways through AA metabolism, the cyclooxygenase (COX) pathway, the lipoxygenase (LOX) pathway and the cytochrome P450 (CYP) pathway. The DHETs are formed through the CYP pathway when epoxyeicosatrienoic acids (EETs) are further metabolised (Tacconelli & Patrignani, 2014). The arachidonic acid pathway itself plays a key role in carcinogenesis (Yarla, et al., 2016) so it follows that there is increased AA metabolism in malignancy and that is why we have found higher levels of 5,6-DHET in the lung cancer group.

4.3.2 | Staging for all lung cancer subtypes

Of the 16 statistically significant ($P<0.05$) metabolites determined by the ANOVA (Figure 7), table 3 shows the AUC values for each of the six discriminants. Out of the 16, all of them have an AUC value over the 0.7 limit for sensitivity and specificity to be robust biomarkers for at least two of the discriminant groups. There are 14 metabolites which have AUC values above the threshold for at least 3 discriminants, and one metabolite (*m/z* 574.53217) which has AUC values above the threshold for 4 of the 6 discriminants. Of these 14 metabolites only three of them were found to be identifiable in the *Homo sapiens* metabolite database (BioCyc), these were 3,3',5,5'-tetraiodothyronamine, cytidine-triphosphate and ubiquinol-10.

3,3',5,5'-tetraiodothyronamine, is an iodoaromatic compound which occurs in thyronamine and iodothyronamine metabolism (HumanCyc, 2020). Thyronamines (TAM) are still a fairly new group to have been identified and are described as a novel class of iodothyronine-like endogenous signalling compounds (Scanlan, et al., 2004) but the pathways of biosynthesis to produce thyronamine remain unknown (Hoefig, et al., 2012). They do however show promising therapeutic potential, representing the only endogenous compounds to elicit responses such as inducing hypothermia as a prophylactic. They may be used as an acute treatment for stroke, the fact they are endogenous is important in that they may then cause fewer side effects compared to the traditional synthetic compounds (Piehl, *et al.*, 2011). While

cancers metastasising to the thyroid are rare and when it happens the primary tends to be in the breast or kidneys, there have been a few very rare cases where a primary lung malignancy has metastasised to the thyroid gland (Albany, *et al.*, 2011). As well as malignancy, thyroid function and in turn dysfunction should be explored, while screening for novel markers of thyroid function is lacking in humans, the marked effect of thyroid dysfunction on many metabolomes and proteomes have been investigated using rodent models (Pietzner, et al., 2017). The fact that in this study the stage 2 lung cancer group has appeared so markedly distinctive compared to the other groups is unexpected, and it may be that with such a small sample size there are one or two patients in the stage 2 group who have multiple-comorbidities. Given this marked change in the concentration of 3,3',5,5'-tetraiodothyronamine (T4AM) among the group there is a possibility that as well as the primary lung tumour some of the patients also have either thyroid malignancy or more likely a form of thyroid dysfunction, this is due to the fact that it is produced in one of the alternative pathways for thyroid hormone degradation rather than deiodination (McDougall, 2013).

Ubiquinol-10 and ubiquinone-10 make up the coenzyme Q10 which plays a key role as an enzyme in ATP synthesis and the electron transport chain (Overvad, *et al.*, 1999). Another role of CoQ10 is its involvement with membrane stability, control of cell growth and apoptosis (Ruiz-Jimenez, *et al.,* 2007). In this section Ubiquinol-10 has been found in significantly higher concentrations in the stage 2 lung cancer group (Figure 8), this is unexpected as Cobanoglu (2011) found significantly lower coQ10 levels when investigating the erythrocyte levels in patients with lung cancer (Cobanoglu, et al., 2011), however they did not investigate how the levels change depending on the stage of the tumour. If our result could be replicated in a much larger study then from the preliminary results we have Ubiquinol-10 appears to be both a statistically significant (Figure 7) and clinically significant (Table 3) biomarker for distinguishing stage 2 lung cancers. We have shown it to be specific and sensitive but for a biomarker to have good clinical relevance, it needs to be both specific and sensitive, and repeatable (Holland, 2016).

Cytidine-triphosphate is important in phospholipid biosynthesis and the biosynthesis of nucleic acids (Molina-Romero, et al., 2017). Upregulation of cytidine-triphosphate (CTP) is described as indispensable for tumour proliferation (Meng, et al., 2012). The enzyme CTP synthase catalyses the conversion of uridine triphosphate (UTP) to CTP and this pathway is

implemented in the synthesis of CTP in proliferating and malignant tissues. CTP synthase, and therefore CTP, have previously been found at elevated levels associated with numerous malignancies such as acute lymphocytic leukaemia (ALL), acute myeloid leukaemia (AML), hepatoma, colon cancer and renal carcinoma, as well as non-Hodgkin's lymphoma (NHL) (Verschuur, 2007). This was then taken further for use as a drug target and gemcitabine was found to directly inhibit cytidine-triphosphate synthesis (Galmarini, *et al.*, 2003). Gemcitabine (Gemzar®) is a chemotherapy agent employed to treat breast cancer, bladder cancer, pancreatic cancer and ovarian cancer, as well as small cell lung cancer (Macmillan, 2020). The idea that in this study cytidine-triphosphate was found in such high concentrations in the stage 2 lung cancer group (Figure 16) then comes as little surprise, especially when it has previously been shown. Warth *et al.* (2014) investigated the relationship between proliferative activity of non-small cell lung cancers with their respective stage and nodal status. They found that in adenocarcinomas the proliferative activity was significantly associated with tumour size, with the largest difference in proliferation activity found between the stage 1 and stage 2 tumours with more than 10% higher proliferation activity being found in the stage 2 group (Warth, et al., 2014). However, in terms of the aims of this study, while it is both an important metabolite for our understanding of stage 2 lung cancer and both a statistically significant ($P<0.05$) and clinically robust metabolite (Table 3) it is also able to, on its own, discriminate between stage 2 lung cancer and the other stages, therefore, in association with other metabolite biomarkers be a clinically useful biomarker for lung cancer staging in the future.

The remaining 13 metabolites shown to be both statistically ($P<0.05$) (Figure 7) and clinically significant (Table 3) remain unidentified. Eleven of these thirteen unknown metabolites are able to discriminate between the stages in the way 3,3',5,5'-tetraiodothyronamine, cytidine-triphosphate and ubiquinol-10 do, shown to be upregulated in the stage 2 groups only (Figure 8) and to have AUC values for the stage 2 discriminants above the 0.7 cut-off for clinical significance (Table 3). However, Unknown Metabolite 1 (UM1) (*m/z* = 574.53217) appears differently as it is shown to be upregulated in both the stage 2 and the stage 4 groups (Figure 8) and while unable to discriminate between stages 2-4 and stages 1-3 appears to be a robust biomarker for the other four discriminants (Table 3). Whereas Unknown Metabolite 2 (UM2) (*m/z* = 639.64258) has also been shown to be upregulated in both the stage 2 and stage 4

groups (Figure 8) but is instead a less useful biomarker as it is only able to discriminate between the group 2 discriminants like the identified metabolites but is also unable to discriminate between the stage 2-4 group (Table 3). Unknown metabolites have been impacting biomedical research since metabolomics started with a considerable number of high-ranking hits found among studies representing unidentified metabolites (Krumsiek, et al., 2012). This has been reported across many areas within metabolomics including studies into heart failure (Steffens, *et al.*, 2010), type 2 diabetes (Fiehn, *et al.*, 2010) and insulin resistance (Gall, et al., 2010), and for as long as the metabolites' chemical identities remain unknown these metabolites use as functional biomarkers remains limited (Krumsiek, et al., 2012). There are also benefits on untargeted metabolomics, the theory being that the analysis is unbiased and analyses a large population of the metabolome but limitations such as unidentified metabolites cannot be overlooked. Some unnamed metabolites can be found in databases of previously recorded masses but these metabolites have often not been validated which poses a further problem, but may be overcome in the future with more research into ancillary metabolites and further validating (Wood, 2013).

4.3.3 | Adenocarcinomas

In the adenocarcinoma group the one metabolite which was found to be significantly different ($P<0.05$) between the four stages was Dipalmitoyl-phosphatidylcholine (DPPC) which was significantly upregulated in the stage 2 group compared to the other three groups. It was also found that it had AUC values over the accepted 0.7 threshold for sensitivity and specificity for three of the six discriminants (Table 4) suggesting that it may be a robust biomarker for the staging of Stage 2 adenocarcinomas. DPPC is a phospholipid and is the main constituent of pulmonary surfactants, produced by most air-breathing animals they work to reduce the surface tension of the water during gas exchange, this low surface tension minimizes the magnitude of negative pressure in the airway liquid layer possibly reducing the tendency for airway wall collapse (Hohlfeld, *et al.*, 1997). The association of pulmonary surfactants with early stage lung cancer has been documented before, in 1988 Mitutani, *et al.* immunohistochemically studied eighty-nine primary lung carcinomas and twenty-three metastatic lung tumours for the expression of pulmonary surfactant apoproteins. Surfactant apoproteins were found in only the primary lung adenocarcinomas, none were found in any

of the other histological types of primary lung carcinoma or in any of the metastatic lung tumours, so it was then concluded that pulmonary surfactant apoproteins could be a useful immunohistochemical marker for differential diagnosis of whether lung adenocarcinoma are primary or metastatic (Mitutani, et al., 1988).

Based on Mitutani's findings in 1988 it would then be expected that if the adenocarcinoma group were then to be split into early (stage 1 and 2) and advanced stages (stage 3 and 4) that there would be differences in the metabolite concentrations including dipalmitoyl-phosphatidylcholine (DPPC), which although not found to be significant in the early against late stage, was found to be significant when investigating adenocarcinomas stages 1-4 (Figure 11). Unfortunately, unlike the ANOVA results for the 4-group analysis, the t-test for the early and late stages shows no significant metabolites ($P<0.05$). It has previously been documented that serum samples from early-stage patients can be categorised by an increase in lactate levels and all ketone bodies which are both commonly associated with cancer metabolism. This increase in lactate with a decrease in glucose concentration has been shown to be more prominent in advanced-stage patients (Puchades-Carrasco, et al., 2016) so it was expected that this study would also find metabolites which would discriminate between the early and late stages. The lack of significant metabolites may be due to the low numbers of samples in the study or the fact that they occur in the saliva at much lower levels than in the serum or that saliva samples are far more variable.

4.3.4 | Squamous cell carcinomas

Although the PLS-DA (Figure 13b) showed fairly tight clusters and no overlap between the groups the t-test showed that there were no significant ($P<0.05$) metabolites between the early and late stage groups. This was also found to be the case when attempting to use salivary metabolomics for oral cancer screening. Ishikawa, *et al.,* found that salivary metabolite markers did not show stage-specific difference, however, did report stage-independent elevation of certain metabolite biomarkers in the saliva, indicating that for oral cancers while salivary metabolite biomarkers may be clinically useful for screening and diagnosis they may not be for disease staging (Ishikawa, et al., 2016). Our results indicate that this is not the case in lung cancer as previously in this study it has been shown that there is one significant ($P<0.05$) metabolomic biomarker found between the adenocarcinoma groups, this may

indicate that there are more noticeable metabolomic changes associated with adenocarcinomas, compared to squamous cell carcinomas. There are many differences between adenocarcinomas and squamous cell carcinomas so differences in metabolite expression is to be expected, patients with squamous cell carcinomas are known to have poorer prognoses when compared to those with adenocarcinomas following surgical resection (Asamura, et al., 2008). One aspect of this may be that squamous cell carcinomas mostly develop in smokers who often either are suffering or go on to suffer from life-threatening co-morbidities, but, the exact differences in biological aggressiveness of the two histological types is still not well understood (Kawase, et al., 2012).

4.3.5 | Gastric Fluid for Diagnosis of Gastric Cancer

The second heatmap (Figure 16) illustrates the similarities and differences in metabolite concentrations using the averages for the two groups against the top 5 metabolites for importance. Two of the metabolites (5-Formyl-5,6,7,8-tetrahydromethanopterin and Angiotensin I) are shown to be higher in the gastric cancer group and lower in the control group. Angiotensin I has been implicated in malignancy and specifically gastric tumours for several years, ACE inhibitors inhibit the ACE-mediated conversion of angiotensin II from angiotensin I and studies have shown that ACE inhibitors have inhibitory effects on both tumour progression, and metastasis (Sugimoto, et al., 2006), and that stimulation of angiotensin II type 2 receptors inhibits the development of cancer (Lever, *et al.*, 1998). 5,6,7,8-tetrahydromethanopterin is a coenzyme of the methanogenic archaea (Scheucher, et al., 1992).

The other three metabolites (urocanic acid, prostaglandin and 12-OPDA) were upregulated in the control group and downregulated in the gastric cancer group. *Cis*-urocanic acid has been shown to be immunosuppressive with anti-inflammatory properties (Konkol, et al., 2016) and in terms of biomarker use, urocanic acid has been shown to be a robust biomarker for detecting uterine cervix cancer using urine samples and returned an AUC of 0.997 (Chen, et al., 2013).

Prostaglandin E_2 (PGE_2) is the most abundant prostaglandin in the body and is particularly interesting as it is involved with all the processes which give rise to the redness, swelling and pain classically associated with inflammation. The redness and swelling are caused by increased blood flow to the affected tissue caused by PGE_2-mediated increase of arterial dilation and increased microvascular permeability, and the pain comes from the effect of PGE_2 on peripheral sensory neurons in the nervous system (Ricciotti & FitzGerald, 2011). PGE_2 is found in gastric tissue, is involved with numerous functions of the gastric mucosal cells (Uefuji, *et al.*, 2000) and has been implicated in carcinogenesis, using mechanisms of immunosuppression, inhibiting apoptosis, increasing metastatic potential of epithelial cells, and promoting angiogenesis (Jang, 2004 : Jones, *et al.,* 1999).

None of the metabolites have VIP scores above 1.5 which is the value usually accepted VIP for significance (Akarachantachote, *et al.*, 2014), however, Angiotensin I, prostaglandin, 5-Formyl-5,6,7,8-tetrahydromethanopterin, Urocanic acid and 12-OPDA also all feature in the top metabolites for their VIP scores as well as being shown to be important from the heatmap. As well as these, the VIP scores also highlight Gentamicin C1a in the top five most important metabolites based on their VIP score. Gentamicin C1a is found at higher concentrations in the gastric cancer group and lower concentrations in the control group. Gentamicin is an aminoglycoside broad spectrum antibiotic (British National Formulary, 2020). It is possible that subjects had been given antibiotics for reasons other than their cancer for example for a urine infection or if their breathlessness had been attributed to a respiratory infection.

4.3.6 | Biomarkers in common

Along with the metabolites which were found to be significant in the individual sections of this study, Table 6 shows the clinically relevant biomarkers found in the lung cancer diagnosis and lung cancer staging sections and the metabolites from the gastric cancer section which were highlighted by the heatmap as having some significance although not statistically. Urocanic Acid was the one metabolite found to be common to both the lung cancer diagnosis and gastric cancer diagnosis analysis.

Urocanic acid results from the deamination of histidine by histidase, this forms trans-urocanic acid and then when exposed to UV radiation it is converted to *cis*-urocanic acid (Uusi-Oukari,

et al., 2000). Huan, *et al.,* (2018) assessed the use of salivary metabolomics in differentiating cognitively normal vs mild cognitive impairment vs Alzheimer's disease. This metabolomic analyses could distinguish between the 3 states with an AUC of above 0.77, and in the case of Alzheimer's vs cognitively normal, an AUC of 1 (Huan, et al., 2018). In terms of biomarker use, urocanic acid has been used to detect uterine and cervical cancer via urine sample with an AUC of 0.997 (Chen, et al., 2013). In lung cancer urocanic acid is also present in two biomarker panels in sweat that were able to differentiate between smokers and those with lung cancer with a 95% specificity and 95% sensitivity (Delgardo-Povedano, *et al.,* 2016). In this study, urocanic acid was found to be statistically significant ($P<0.05$) and have an AUC of 0.833 (Table 2), well above the accepted 0.7 threshold, this time in the saliva of LC patients.

4.4 | Pathway Analysis of Lung Cancer and Gastric Cancer

Following investigation of potentially significant biomarkers enrichment analysis was then used to produce metabolite set over representation analysis (ORA) summary plots for the Lung Cancer diagnosis group (Figure 5) and the Gastric Cancer diagnosis group (Figure 18). There is a huge degree of similarity between the two sets of pathways, with all but one of the gastric cancer pathways also featuring in the lung cancer group. The pathways highlighted for both Lung and Gastric Cancer in Figures 5 and 18 suggest proinflammatory involvement given the clear involvement of arachidonic acid metabolism, as well as clear evidence of redox involvement given that glutathione metabolism appears prevalent as well as the associated glutamate metabolism and ammonia recycling.

Arachidonic acid metabolism features as most overrepresented in both groups showing clear association with malignancy propagation. Arachidonic acid (AA) is sourced in the body either through direct consumption of foods with high levels of AA like whole eggs and lean meat and tuna or through its parent molecule linoleic acid (LA), abundant in vegetable oils like sunflower oil. AA is normally found incorporated in phospholipids in the cell membrane, particularly high levels are present in the skeletal muscle, brain, liver, spleen, and retina phospholipids (Hanna & Hafez, 2018). The enzyme cyclooxygenase (COX) is known for its ability to metabolise arachidonic acid into metabolites such as prostaglandins and thromboxane A_2 (TXA_2), known collectively as prostanoids (Hoxha, 2018). COX exists in two isoforms, COX-1 and COX-2. COX-1 is known to be expressed in most cells and facilitates the

production of the prostanoids needed for homeostasis, when there is then an inflammatory stimuli COX-2 is induced and is the more important source of prostanoids in inflammation and proliferative diseases, such as cancer (Ricciotti & FitzGerald, 2011). The arachidonic acid pathway itself plays a key role in carcinogenesis (Yarla, et al., 2016). At a cellular level phospholipases act on phospholipids to release the esterified AA. Once liberated the free functional AA may enter other cells, be reincorporated into phospholipids, or be metabolised. The free AA also has an important role in apoptosis, as its accumulation causes arachidonoyl CoA transferase inhibition which activate enzymes within the cell and trigger degradation of sphingolipids (Hanna & Hafez, 2018). AA metabolism is made up of complex, partially interconnected pathways and a number of the constituents in the AA cascade have not yet been sufficiently explored (Hyde & Missailidis, 2009). Also potentially related, bile acid biosynthesis, was found to be of importance (Figure 18) is the main pathway for excreting excess cholesterol in humans, however as many bile acid metabolites are cytotoxic the genes which encode the enzymes required for bile acid synthesis are under strict regulatory control (Pandak, et al., 2001). Bile acids have been found to be associated with the development of cancers in the digestive tract, specifically relating to growth and transformation signalling, which is believed to be related to the induced COX-2 expression and increase in PGE_2 production (Wu, *et al.*, 2018). While bile acid biosynthesis was only found associated with gastric cancer, this is also important as this pathway analysis not only gives rise to pathways potentially involved with all types of malignancy but also some unique to a certain type of cancer.

The first two stages in the glutathione cycle are biosynthesis steps of glutathione by two sequential enzymes, the glutathione undergoes two ATP-dependent stages, the glutathione formed can then be degraded to yield 5-oxoproline and cysteinyl glycine, the 5-oxoproline formed is cleaved by a 5-oxoprolinase to yield glutamate, while the cysteinyl glycine is cleaved to yield cysteine and glycine, if necessary the glutamate, cysteine and glycine can be put back in to the synthesis of glutathione, thus completing the cycle (Bachhawat & Yadav, 2018). Glutathione metabolism is the pathway with the second highest number of constituent hits found in the lung cancer analysis (Figure 18) but also featured in the gastric cancer analysis although not as highly. As the most abundant natural antioxidant found in humans, glutathione (GSH) has multiple roles. Most of these are related to redox homeostasis. GSH

maintains levels of cysteine and xenobiotics are detoxified by GSH whilst conferring resistance to cancer cells. Some malignancies see a beneficial role for GSH, while others see GSH in a pathogenic role. Targeting GSH may make tumour cells more amenable to radiotherapy and chemotherapy (Bansal & Simon, 2018). This also links to the third pathway identified by the enrichment analysis, ammonia recycling, as ammonia is considered a by-product of the conversion of glutamine to glutamate, as well as the glutamate metabolism also identified in Figure 18. Glutamate, an abundant amino acid, has been researched as an important nutrient in cancer. It has the ability to donate its nitrogen and carbon to many pathways promoting growth. Cancer cells often display addictions to glutamine in laboratory conditions (Hensley, *et al.*, 2013). Whereas, the consequences of ammonia production, in cancer, are not fully comprehended but Spinelli *et al.,* discovered that ammonia is not merely as waste product but is used in central amino acid metabolism to enhance nitrogen use. The data demonstrated that ammonia is not merely a waste product, but a fundamental nitrogen source that is able to support tumour biomass (Spinelli, *et al.,* 2017). The importance of considering not just the tumour itself but its interactions with the rest of the host body and between any other tumours has also been highlighted (Gouirand, *et al.*, 2018).

Other pathways to be considered are steroidogenesis, porphyrin metabolism and histidine metabolism, also found in both the lung cancer (Figure 5) and gastric cancer analysis (Figure 18). Histidine is an amino acid, which humans are obligated to obtain from their diet as the body is unable to synthesise it, decreasing the gene that encodes forminidoyltransferase cyclodeaminase (FTDC) was found to reduce the sensitivity of cancer cells to methotrexate (Kanarek, *et al.*, 2018). The authors also suggested that methotrexate is more effective against leukaemia in patients who had a higher level of histidine ammonia lysis enzyme. They proposed that this was due to high levels of HAL boosting the degradation of histidine (Frezza, 2018).

Another pathway which we have highlighted is porphyrin metabolism (Figure 18). Naturally occurring porphyrins have been described in increased amounts in the blood of patients with various malignant tumours. The authors concluded that tumour cells amass a level of endogenous porphyrin sufficient for diagnosis (Lualdi & Cavalli, 2018). Porphyrin metabolism was investigated in a study exploring the effects on porphyrins of Hodgkin's lymphomas and

acute lymphoblastic leukaemia against healthy controls and found that some dramatic changes were found in porphyrin metabolism in adult patients suffering from acute lymphoblastic leukaemia (El-Sharabasy, *et al.*, 1992).

Steroidogenesis is the processes by which cholesterol is converted to steroid hormones by the transport of enzymes, proteins and redox partners and cofactors (Miller & Auchus, 2011). CYP450 and hydroxysteroid dehydrogenases are the main two steroidogenic enzymes. CYP450 are divided into type 1 mitochondrial and type 2 found in endoplasmic reticulum. Hydroxysteroid dehydrogenase is part of the aldo-ketoreductase or short-chain reductase/dehydrogenase families. The enzymes are controlled by post translational modifications and cofactors particularly the redox partners, which donate electrons. The study of steroidogenesis is fundamental to understanding sexual differentiation disorders, fertility, reproduction, obesity, hypertension and homeostasis in physiological terms (Miller & Auchus, 2011). Adenocarcinomas in the gastrointestinal tract predominantly occur in males (Chang, et al., 2017), which has led to people hypothesising about the role of sex hormones in gastric cancer development, although the literature surrounding this idea is very split (Korenaga, *et al.,* 1998 : Cronin-Fenton, *et al.,* 2010), however earlier this year a paper was published which validated that in regards to tumour-associated neutrophils (TANs), part of the tumour immune microenvironment (TIME), there are differences between the sexes in gastric cancers (Clausen, *et al.*, 2020).

As well as bile acid biosynthesis only being found in the gastric cancer analysis (Figure 18) there were many more pathways only highlighted by the lung cancer study (Figure 5). The most significant of these being the arachidonic acid metabolism, amino sugar metabolism and estrone metabolism. Arachidonic acid metabolism as previously mentioned shows the significance of the proinflammatory response in the propagation of lung cancer, while amino sugar metabolism is likely linked to the increased levels of glucose required by the malignant cells. All malignancies are known for causing cachexia as tumours remotely cause metabolic disturbances, potentially this isn't as important in gastric cancers as the tumours location enables easier attainment of energy from its surroundings rather than having to draw from the systemic system. Heber *et al.* (1985) found that alanine was being released at an enhanced rate and being utilised faster in lung cancer patients compared to controls for the

production of glucose (Heber, *et al.*, 1985). Estrone metabolism was also highlighted in the LC analysis uniquely, lung cancer is known for being very complex with many factors contributing to the development of lung tumours. Although some of these connections such as the link between lung cancer and cigarette smoke is very well documented there is also evidence that the presence of lung cancer can be affected by gender-specific factors such as estrogens with females being at higher risk (Slowikowski, *et al.*, 2016). In 2012, Drzewiecka and Jagodzinski studied the conversion of estrone to 17-beta-estradiol and found that NSCLC cells were able to independently produce 17-beta-estradiol from esterone, and so its link to estrone metabolism (Drzewiecka & Jagodzinski, 2012).

4.5 | Conclusions and Future Work

Overall, many robust potential biomarkers have been identified, however further study is required with a much larger sample cohort and further work to explore the individual biomarkers identified and their potential uses. This study has explored the pathways and mechanisms found to be related to the metabolites which were found at significantly different levels and their potential to expand the knowledge base regarding lung tumours and gastric tumours, including metabolites involved in methylation, fatty acid transport and glucose metabolism. There is clear proinflammatory and redox involvement specifically related to arachidonic acid metabolism, and the results also appear to support the ongoing hypotheses that gastric and lung cancer may be sexually dimorphic, with gastric cancer associated with sex hormones produced in steroidogenesis, as well as lung cancer being linked to estrone metabolism. It has also highlighted the need for further identification and investigation into unidentified metabolites.

In the future these results should be analysed further paying close attention to the clinical data associated with the samples. Further analysis could be carried out into the metabolic differences between participants of different sexes or on those taking different pharmacological agents, the similarities and differences of metabolite expression between different sub-types of lung cancer and for the gastric carcinoma samples the metabolic profiles of the group who did respond to treatment against the group which did not. For the data on staging lung cancer group 2 seems to show unprecedented results and could be

skewing the analysis, the data should be reanalysed with the stage 2 group removed to see how this affects the results.

This study shows preliminary pathway analysis but the enrichment overview was performed on all of the statistically significant metabolites, not grouped into which ones were upregulated or downregulated in the lung cancer group compared to the controls, in the future these should be separated and undergo further statistical analysis, perhaps presented as box and whisker plots to show the relevance of each pathway to the groups. The analysis carried out could be looked at in much more detail, exploring all of the pathways identified in the lung cancer samples, if they are pathways which have been associated with general malignancy or more specifically lung cancer in the past. The considerable amount of overlap in the pathways identified in the gastric cancer study and the lung cancer study also gives rise to future questions around running pathway analysis like this on other types of malignancy such as ovarian cancer, colorectal carcinomas etc. to get more detail into which pathways are shared between types of cancers and which are unique depending on their origin and location. The ORA could also be carried out on the metabolites identified in lung cancer staging to see if there are common pathways unique to say initiation or progression of the cancer. This will not only enhance our understanding of cancer formation, growth and metastases but also potentially identify treatment targets in the future.

Abbassi-Ghadi, N., Huang, K., Takats, G., & Hanna, G. (2013). Metabolomic profiling of oesophago-gastric cancer: A systematic review . *European Journal of Cancer*, Volume 49, Issue 17, Pages 3625-3637.

Adjei, A. (2019). Lung Cancer Worldwide. *Journal of Thoracic Oncology*, vol 14, issue 6, page 956.

Akarachantachote, N., Chadcham, S., & Saithanu, K. (2014). Cutoff threshold of variable importance. *International Journal of Pure and Applied Mathematics*. Volume 94, Issue 3, Pages 307-322.

Akhtar, N., Gupta, J., & Bansal, G. (2017). Risk factors onlung cancer in non smoker. *Current problems in cancer*, Volume 41, Issue 5, Pages 328-339.

Albany, C., Jain, A., Ulbright, T., & Einhorn, L. (2011). Lung cancer, thyroid cancer or both: An unusual case presentation. *Journal of Thoracic Disease*, Volume 3, Issue 4, Pages 271-273.

Alvarez-Sanchez, B., Priego-Capote, F., & Luque de Castro, M. (2010). Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC Trends in Analytical Chemistry*, Volume 29, Issue 2, Pages 120-127.

American Cancer Society. (2015). *American Cancer Society Guidelines for the Early Detection of Cancer: Lung Cancer.* American Cancer Society.

Annibaldi, A., & Widmann, C. (2010). Glucose metabolism in cancer cells. *Current Opinion in Clinical Nutrition and metabolic Care*, 13.4: 466-470.

Armitage, E., & Southam, A. (2016). Monitoring cancer prognosis, diagnosis and treatment efficiacy using metabolomics and lipidomics. *Metabolomics*, doi: 10.1007/s11306-016-1093-7.

Askenazi, D., Koralkar, R., Hundley, H., Montesanti, A., Parwar, P., Sonjara, S., & Ambalavanan, N. (2012). Urine Biomarkers Predict Acute Kidney Injury in Newborns. *The Journal of Pediatrics*, Volume 161, Issue 2, Pages 270-275.e1

Bach, P., Mirkin, J., & Oliver, T. (2012). Benifits and harms of CT screening for lung cancer: a systematic review. *JAMA*, 307 (22), 2418-2429.

Bachhawat, A., & Yadav, S. (2018). The Glutathione Cycle: Glutathione MetabolismBeyond thec-Glutamyl Cycle. *International Union of Biochemistry and Molecular Biology*, *70*(7), 585-592.

Bains, W. (2004). Failure rates in drug discovery and development: will we ever get better? *Drug Discovery*, 9-18.

Bansal, A., & Simon, C. (2018). Glutathione metabolism in cancer progression and treatment resistance. *Journal of Cell Biology*, *217*(7), 291-2298.

Barroso, M., Handy, D., & Castro, R. (2017). TheLink Between Hyperhomocysteinemia and Hypomethylation: Implications for Cardiovascular Disease. *Journal of Inborn Errors of Metabolism and Screening*, *5*, 2326409817698994.

Baylin, S., & Chen, W. (2005). Aberrant gene silencing in tumor progression: Implications for control of cancer. *Cold Spring Harbor Symposia on Quantitative Biology*, Volume 70, Pages 427-433.

Beger, R., Sun, J., & Schnackenberg, L. (2010). Metabolomics approaches for discovering biomarkers of drug-induced hepatotoxicity and nephrotoxicity. *Toxicity and Applied Pharmacology*, *243*(2), 154-166.

Beyoglu, D., & Idle, J. (2013). Metabolomics and its potential in drug devolopment. *Biochemical Pharmacology*, *85*(1), 12-20.

Blocker, R., & Roth, J. (1977). The reduction of uridine 5'-diphosphate and uridine 5'-triphosphate in some transplantable rat hepatomas . *Cancer Research*, *37*(6), 1923-1928.

British National Formulary. (2020, March 3). *Gentamicin*. Retrieved from BNF: https://bnf.nice.org.uk/drug/gentamicin.html

Brockhausen, I. (1999). Pathways of O-glycan biosynthesis in cancer cells. *Biochimica et Biophysica Acta (BBA) - , 1473*(1), 67-95.

Bueno, J. (2015). Metabolomics in Antimicrobial Drug Discovery: The Success of the Chemical Diversity. *Microbial and Biochemical Technology*, 7, 380-383.

Cammack, R., Atwood, T., Campbell, P., Parish, H., Smith, A., Vella, F., & Stirling, J. (2008). *Oxford Dictionary of Biochemistry and Molecular Biology (2 ed.) .* Oxford University Press.

Cancer Research UK. (2021, January 13). *Stomach cancer statistics*. Retrieved from Cancer Research UK: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/stomach-cancer

Carraro, S., Andreola, B., Alinovi, R., Corradi, M., Freo, L., Dalt, L., & Baraldi, E. (2008). Exhaled leukotriene B4 in children with community acquired pneumonia. *Pediatric pulmonology*, *43*(10), 982-986.

Casero, R., & Marton, L. J. (2007). Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. *Nature Reviews Drug Discovery*, *6*(5), 373-390.

Castro, R., Rivera, I., Struvs, E., Jansen, E., Ravasco, P., Camilo, M., . . . Almeida, I. (2003). Increased Homocysteine and S-Adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. *Clinical Chemistry*, *49*(8), 1292-1296.

Chakravarthi, B., Nepal, S., & Varambally, S. (2016). Genomic and Epigenomic Alterations in Cancer. *The American Journal of Pathology*, *186*(7), 1724-1735.

Chalcraft, K., Lee, R., Mills, C., & Britz-McKibbin, P. (2009). Virtual Quantification of Metabolites by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry: Predicting Ionization Efficiency Without Chemical Standard. *Analytical chemistry*, *81*(7), 2506-2515.

Chan, A., Mercier, P., Schiller, D., Bailey, R., Robbins, S., Eurich, D., . . . Broadhurst, D. (2016). 1H-NMR urinary metabolomic profiling for diagnosis of gastric cancer. *British Journal of Cancer*, *114*(1), 59-62.

Chang, W., Huang, S., Lee, Y., Lai, H., Cheng, B., Cheng, W., . . . Ma, W. (2017). Cholesterol import and steroidogenesis are biosignatures for gastric cancer patient survival. *Oncotarget*, *8*(1), 692-704.

ChEBI. (2020, 08 25). *CHEBI:17844 - 2-amino-3-oxobutanoic acid*. Retrieved from ChEBI: https://www.ebi.ac.uk/chebi/searchId.do;jsessionid=2E01D4A5998748E2B0B495EF5214EBE4?chebiId=CHEBI%3A17844

Chen, D., Cao, G., & Chen, H. (2019). Identification of serum metabolites associating with chronic kidney disease progression and anti-fibrotic effect of 5-methoxytryptophan. *Nature Communications*, *10*(1), 1-15.

Chen, W., Zheng, R., Baade, P., Zhang, S., Zeng, H., Bray, F., . . . He, J. (2016). Cancer statistics in China. *CA-A CANCER JOURNAL FOR CLINICIANS*, *66*(2), 115-132.

Chen, Y., XU, J., Zhang, R., Shen, G., Gong, Y., Sun, J., . . . Abliz, Z. (2013). Assessment of data pre-processing methodsfor LC-MS/MS-based metabolomics of uterine cervix cancer. *Analyst*, *138*(9), 2669-2677.

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G., Doctor, B. P., Pardhasaradhi, K., & McCann, P. (1996). S-Adenosylmethionine and methylation. *Federation of American Societies for Experimental Biology journal*, *10*(4), 471-480.

Clausen, F., Behrens, H., Kruger, S., & Rocken, C. (2020). Sexual dimorphism in gastric cancer: tumor-associated neutrophils predict patient outcome only for women. *Journal of Cancer Research and Clinical Oncology*, *146*(1), 53-66.

Cobanoglu, U., Demir, H., Cebi, A., Sayir, F., Alp, H., Akan, Z., . . . Bakan, E. (2011). Lipid Peroxidation, DNA Damage and CoenzymeQ10 in Lung Cancer Patients - Markers for Risk Assessment? *Asian Pacific Journal of Cancer Prevention*, *12*(6), 1399-1403

Collins, F., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from Large-Scale Biology. *Science*, *300*(5617), 286-290.

Combs, G., & Gray, W. (1998). Chemopreventive agents: selenium. *Pharmacology & Therapeutics*, *79*(3), 179-192.

Cronin-Fenton, D., Murray, L., Whiteman, D., Cardwell, C., Webb, P., Jordan, S., . . . Lagergren, J. (2010). Reproductive and sex hormonal factors and oesophageal and gastric junction adenocarcinoma: a pooled analysis . *European Journal of Cancer*, *46*(11), 2067-2076.

Cuatrecasas, P. (2006). Drug discovery in jeopardy. *Journal of clinical investigation*, *116*(11), 2837-2842.

Das, V., Kalita, J., & Pal, M. (2017). Predictive and prognostic biomarkers in colorectal cancer: A systematic review of recent advances and challenges. *Biomedicine and Pharmacotherapy*, *87,* 8-19.

Delgardo-Povedano, M., Calderon-Santiago, M., Prieggo-Capote, F., Jurado-Gamez, B., & Lungue de Castro, M. (2016). Recent advances in human sweat metabolomics for lung cancer screening. *Metabolomics*, *12*(11), 1-12.

Denkert, C., Budczies, J., Kind, T., Weichert, W., Tablack, P., & Sehouli, J. (2006). Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Research* , *66*(22), 10795-10804.

Derezinski, P., Klipczynska, A., Sawicki, W., Palka, J., & Kokot, Z. (2017). Amino Acid Profiles of Serum and Urine in Search for Prostate Cancer Biomarkers: a Pilot Study. *International Journal of Medical Sciences*, *14*(1), 1-12.

Detterbeck, F., Boffa, D., Kim, A., & Tanoue, L. (2017). The Eighth Edition Lung Cancer Stage. *Chest*, *151*(1), 193-203.

Dettmer, K., Aronov, P., & Hammock, B. (2007). Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, *26*(1), 51-78.

Draper, J., Enot, D., Parker, D., Beckmann, M., Snowdon, S., Lin, W., & Zubair, H. (2009). Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'. *BMC Bioinformatics.*, *10*(1), 1-16

Drzewiecka, H., & Jagodzinski, P. (2012). Conversion of estrone to 17-beta-estradiol in human non-small-cell lung cancer cells in vitro. *Biomedicine & Pharmacotherapy*, *66*(7), 530-534.

Duarte, I., Duarte, A., Ladeirinha, A., Lamego, I., Gil, A., Carvalho, L., . . . Melo, J. (2013). Potential Markers of Cisplatin Treatment Response Unveiled by NMR Metabolomics of Human Lung Cells. *Molecular Pharmaceutics*, *10*(11), 4242-4251.

El-Sharabasy, M., El-Waseef, A., Hafez, M., & Salim, S. (1992). Porphyrin metabolism in some malignant diseases. *British Journal of Cancer - Nature*, 65, 409-412.

Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene*, *21*(35), 5427-5440.

Falco, B., & Lanzotti, V. (2018). NMR spectroscopy and mass spectrometry in metabolomics of saliva . *Phytochemistry Reviews* , *17*(5), 951-972.

Fan, Y., Li, Y., & Chen, Y. (2016). Comprehensive Metabolomic Characterization of Coronary Artery Diseases. *JACC: Journal of the American College of Cardiology*, *68*(12), 1281-1293.

Farup, P., Rudi, K., & Hestad, K. (2016). Faecal short-chain fatty acids - a diagnostic biomarker for irritable bowel syndrome? *BMC Gastroenterology*, *16*(1), 1-7.

Ficicioglu, C. (2017). New tools and approaches to newborn screening: ready to open Pandora's box? *Cold Spring Habor molecular case studies* , *3*(3), a001842.

Fiehn, O. (2002). Metabolomics - the link between genotypes and phenotypes . *plant molecular biology*, 155-171.

Fiehn, O., Garvey, W., Newman, J., Lok, K., & Hoppel, C. (2010). Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS One*, *5*(12), e15234

Flanagan, J., Simmons, P., Vehige, J., Wilcoz, M., & Garret, Q. (2010). Role of Carnitine in Disease. *Nutrition and Metabolism*, *7*(1), 1-14.

Freeman, H., & Kim, Y. (1978). Digestion and absorption of protein. *Annual Review of Medicine* , 99-116.

Frezza, C. (2018). Histidine metabolism boosts cancer therapy. *Nature*, 559, 484-485.

Gajdhane, V., & Deshpande, M. (2014). Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques . *IOSR Journal of Computer Engineerin*, *16*(5), 28-35.

Gall, W., Beebe, K., Lawton, K., Adam, K., Mitchell, M., Nakhle, P., . . . Ferrannini, E. (2010). alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One*, *5*(5), e10883.

Galmarini, C., Jordheim, L., & Dumontet, C. (2003). Pyrimidine nucleoside analogs in cancer treatment. *Expert Review of Anticancer Therapy*, *3*(5), 717-728.

Gandy, J. (2019). *Manual of Dietetic Practice, 6th Edition.* Wiley-Blackwell, 311.

Garcia-Sevillano, M., Garcia-Barrera, T., Navarro, F., Abril, N., & Pueyo, C. (2015). Combination of direct infusion mass spectrometry and gas chromatography mass spectrometry for toxicometabolomic study of red blood cells and serum of mice Mus musculus after mercury exposure. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *985*, 75-84.

Gevi, F., Belardo, A., & Zolla, L. (2020). A metabolomics approach to investigate urine levels of neurotransmitters and related metabolites in autistic children. *Biochimica et Biophysica Acta (BBA)- Molecular Basis of Disease*, *1866*(10), 165859

Gieger, C., Geistlinger, E., Altmaier, M., Hrabé de Angelis, F., Kronenberg, T., Meitinger, H., & Mewes, H. (2008). Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLOS Genetics*, 4, no. 11 (2008): e1000282

Giskeodegard, G., Andreassen, T., Bertilsson, H., Tessem, M., & Bathen, T. (2019). The effect of sampling procedures and day-to-day variations in metabolomics studies of biofluids. *Analytica Chimica Acta*, *1081*, 93-102.

Gouirand, V., Guillaumond, F., & Vasseur, S. (2018). Influence of the Tumor Microenvironment on Cancer Cells Metabolic Reprogramming. *Frontiers in Oncology*, *8*, 117.

Greenberg, A., Rimal, B., Felner, K., Zafar, S., Hung, S., Eylers, E., . . . Crawford, B. (2007). S-adenosylmethionine as a biomarker for the early detection of lung cancer. *Chest*, *132*(4), 1247-1252.

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis–a marriage of convenience or a shotgun wedding. *Analytica chimica acta*, *879*, 10-23.

Haaf, K., Tammemagi, M., Bondy, S., McGregor, E., Nicholas, G., & Paszat, L. (2017). Performance and cost effectiveness of computed tomography lung cancer screening. *PLoS Medicine*, *14*(2), e1002225.

Halama, A., Moller, G., & Adamski, J. (2011). Metabolic Signatures in Apoptotic Human Cancer Cell Lines. *OMICS: A Journal of Integrative Biology*, Vol 15, Issue 5, 325-335.

Hanna, V., & Hafez, E. (2018). Synopsis of arachidonic acid metabolism: A review. *Journal of Advanced Research*, *11*, 23-32.

Hao, D., Sarfaraz, O., Farshidfar, F., Bebb, D., Lee, C., Card, C., . . . Weljie, A. (2016). Temporal characterization of serum metabolite signatures in lungcancer patients undergoing treatment. *Metabolomics*, *12*(3), 58.

Hasko, G., Sitkovsky, M., & Szabo, C. (2004). Immunomoulatory and neuroprotective effects of inosine. . *Trends in Pharmacological Sciences*, *25*(3), 152-157.

Hayen, A., Herigstad, M., Weich, K., & Pattinson, K. (2015). Subjective evaluation of experimental dyspnoea – Effects of isocapnia and repeated exposure. *Respiratory Physiology & Neurobiology*, *208*, 21-28.

Heber, D., Byerly, L., & Chlebowski, R. (1985). Metabolic Abnormalities in the Cancer Patient. *Cancer*, *55*(S1), 225-229.

Hensley, C., Wasti, A., & De Barardinis, R. (2013). Glutamine and cancer:cell biology, physiology, and clinical opportunities. . *Journal of clinical investigations*, *123*(9), 3678-3684.

Hoefig, C., Renko, K., Piehl, S., Scanlan, T., Bertoldi, M., & Opladen, T. (2012). Does the aromatic L-amino acid decarboxylase contribute to thyronamine biosynthesis? . *Molecular and Cellular Endocrinology*, *349*(2), 195-201.

Hofman, P. (2017). Liquid biopsy for early detection of lung cancer . *Current opinion in oncology*, *29*(1), 73-78.

Hohlfeld, J., Fabel, H., & Hamm, H. (1997). The role of pulmonary surfactant in obstructive airways disease. *European Respiratory Journal*, *10*(2), 482-491.

Holland, R. (2016). What makes a good biomarker? *Advances in Precision Medicine*, *1*(1), 66-77.

Hori, S., Nishiumi, S., Kobayahi, K., Sinohara, M., Hatakeyamay, Y., Kotani, Y., . . . Bamba, T. (2011). A metabolic approach to lung cancer . *Lung cancer*, *74*(2), 284-292.

Hoxha, M. (2018). A systematic review on the role of eicosanoid pathways in rheumatoid arthritis. *Advances in Medical Sciences*, *63*(1), 22-29.

Hsu, J., Jaw, T., Yang, C., Lin, S., Shih, M., Chou, S., . . . Chiang, I. (2017). Convex border of peripheral non-small cell lung cancer on CT images as a potential indicator of pleural invasion. *Medicine* , *96*(42).

Hsu, P., Chen, C., Hsieh, C., Chang, W., Lai, K., G., L., . . . Lu, P. (2007). A1-Antitrypsin Precursor in Gastric Juice Is a Novel Biomarkerfor Gastric Cancer and Ulcer. *Clinical Cancer Research*, *13*(3), 876-883.

Huan, T., Tran, T., Zheng, J., Straddha, S., MacDonald, S., Camicioli, R., . . . Li, L. (2018). Metabolomics Analyses of Saliva Detect Novel Biomarkers of Alzheimer's Disease. *Journal of Alzheimers disease*, *65*(4), 1401-1416.

Hughes, J., Rees, S., Kalindjian, S., & Philpott, K. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239-1249.

HumanCyc. (2020, August 5). *3,3',5,5'-tetraiodothyronamine*. Retrieved from HumanCyc: https://biocyc.org/compound?orgid=HUMAN&id=CPD-13023#tab=showAll

Humphrey, S., & Williamson, R. (2001). A review of saliva: normal composition, flow, and function. *Journal of Prosthetic Dentistry*, *85*(2), 162-169.

Hyde, C., & Missailidis, S. (2009). Inhibition of arachidonic acid metabolism and its implication on cell proliferation and tumour-angiogenesis. *International Immunopharmacology*, *9*(6), 701-715.

Inokuma, K., Matsuda, M., & Sasaki, D. (2018). Widespread effect of N-acetyl-d-glucosamine assimilation on the metabolisms of amino acids, purines, and pyrimidines in Scheffersomyces stipitis. *Microbial cell factories*, *17*(1), 1-11.

Ishikawa, S., Sugimoto, M., Kitabatake, K., Sugano, A., Nakamura, M., Kaneko, M., . . . Lino, M. (2016). Identification of salivary metabolomic biomarkers for oral cancer screening. *Scientific reports*, *6*(1), 1-7.

Ishikawa, S., Sugimoto, M., Kitabatake, K., Tu, M., Sugano, A., Yamamori, I., . . . Lino, M. (2017). Effect of timing of collection of salivary metabolomic biomarkers on oral cancer detection. *Amino Acids*, *49*(4), 761-770.

Ismail, I., Showalter, M., & Fiehn, O. (2019). Inborn Errors of Metabolism in the Era of Untargeted Metabolomics and Lipidomics. *Metabolites*, *9*(10), 242.

Jang, T. (2004). Expression of proteins related to prostaglandin E2 biosynthesis is increased in human gastric cancer and during gastric carcinogenesis. *Virchows Archiv: European Journal of Pathology*, *445*(6), 564-571.

Jhih-Hao, B., Hsieh, M., Liao, H., Lin, M., & Chen, J. (2019). Prediction of pleural invasion using different imaging tools in non-small cell lung cancer. *Annals of Translational Medicine*, *7*(2).

Johnson, C., Santidrian, A., LeBoeuf, S., Kurczy, M., Rattray, N., Pattray, Z., . . . Siuzdak, G. (2017). Metabolomics guided pathway analysis reveals link between cancer metastasis, cholesterol sulfate, and phospholipids. *Cancer and Metabolism* , *5*(1), 1-9.

Johnston, F., & Beckman, M. (2019). Updates on Management of Gastric Cancer. *Current Oncology Reports*, *21*(8), 1-9.

Jones, M., Wang, H., Peskar, B., Levin, E., Itani, R., Sarfeh, I., & Tarnawski, A. (1999). Inhibition of angiogenesis by nonsteroidal anti-inflammatory drugs: insight into mechanisms and implications for cancer growth and ulcer healing. *Nature Medicine*, *5*(12), 1418-1423.

Jusko, W., & Milsap, R. (1993). Pharmacokenetic principles of drug distribution in saliva . *Annals of the New York Academy of Sciences*, *694*(1), 36-47.

Kamil, Z., & Rashid, K. (2019). Reliability of salivary alpha amylase as a biomarker for obesity. *Indian Journal of Public Health Research & Development*, *10*(10), 3013-3017.

Kanarek, N., Keys, H., & Jason, R. (2018). Histine catabolism is a major determinant of methotrexate sensitivity . *Nature*, *559*(7715), 632-636.

Kasirga, E., Coker, I., & Aydogdu, S. (1999). Increased gastric juice leukotriene B4, C4, and E4 concentrations in children with Helicobacter pylori colonization. *The Turkish Journal of Pediatrics*, *41*(3), 335-339.

Kawase, A., Yoshida, J., Ishii, G., Nakao, M., Aokage, K., Hishida, T., . . . Nagai, K. (2012). Differences Between Squamous Cell Carcinoma and Adenocarcinoma of the Lung: Are Adenocarcinoma and Squamous Cell Carcinoma Prognostically Equal? *Japanese Journal of clinical oncology*, *42*(3), 189-195.

KEGG. (2020, July 15). *Glycan: G00029*. Retrieved from KEGG: https://www.kegg.jp/dbget-bin/www_bget?G00029

Kenfield, S., Wei, E., Stampfer, M., & Rosner, B. (2008). Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco Control*, *17*(3), 198-204.

Kim, S. (2018). Cancer Energy Metabolism: Shutting Power off Cancer Factory. *Biomolecules and Therapeutics* , *26*(1), 39-44.

kirschner, M. (2005). The meaning of systems biology. *Cell*, *121*(4), 503-504.

Knight, S., Crosbie, P., Balata, H., Chudziak, J., Hussell, T., & Diva, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, *7*(9), 170070.

Konkol, Y., Bernouli, J., Streng, T., Jaaskelainen, K., Laihia, J., & Leino, L. (2016). Intravesical treatment with cis-urocanic acid improves bladder function in rat model of acute bladder inflammation. *Neurourology and urodynamics* , *35*(7), 786-791.

Korenaga, D., Orita, H., Okuyama, T., Kinoshita, J., Maekawa, S., Ikeda, T., & Sugimachi, K. (1998). Sex hormone-receptor-negative tumors have a higher proliferative activity than sex hormone-receptor-positive tumors in human adenocarcinomas of the gastrointestinal tract. *Surgery Today*, *28*(10), 1007-1014.

Kosmides, A., Kamisoglu, K., Calvano, S., Corbett, S., & Androulakis, I. (2013). Metabolomic Fingerprinting: Challenges and Opportunities. *Critical Reviews in Biomedical Engineering*, 2 *41*(3), 205-221.

Krumsiek, J., Suhre, K., Evans, A., Mitchell, M., Mohney, R., Milburn, M., . . . Kastenmuller, G. (2012). Mining the Unknown: A Systems Approach to Metabolite Identification Combining Genetic and Metabolic Information. *PLoS Genetics*, *8*(10), e1003005.

Kumar, N., Shahjaman, M., Mollah, M., Islam, S., & Hoque, M. (2017). Metabolomic Biomarkers for Lung Cancer. *Bioinformation*, *13*(6), 202-208.

Kumar, S., Huang, J., Cushnir, J., Spanel, P., Smith, D., & Hanna, G. (2012). Selected Ion Flow Tube-MS Analysis of Headspace Vapor from Gastric Content for the Diagnosis of Gastro-Esophageal Cancer. *Analytical Chemistry*, *84*(21), 9550-9557.

Latimer, K., & Mott, T. (2015). Lung Cancer: Diagnosis, Treatment Principles, and Screening. *American Family Physician*, *91*(4), 250-256.

Lauby-Scretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., & Straif, K. (2016). Body fatness and cancer - viewpoint of the IARC working group. *New England Journal of Medicine* , *375*(8), 794-798.

Lee, W., Um, J., Hwang, B., Lee, Y., Chung, B., & Hong, J. (2020). Assessing the progression of gastric cancer via profiling of histamine, histidine, and bile acids in gastric juice using LC-MS. *The Journal of Steroid Biochemistry amd Molecular Biology*, *197*, 105539.

Lee, Y., & Wong, D. (2009). Saliva: An emerging biofluid for early detection of diseases. *American Journal of Dentistry*, *22*(4), 241.

Lei, Z., Huhman, D., & Sumner, L. (2011). MassSpectrometryStrategiesin Metabolomics. *The journal of biological chemistry*, *286*(29), 25435-25442.

Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y. W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options . *Biochimica et Biophysica Acta (BBA)*, *1856*(2), 189-210.

Lever, A., Hole, D., & Gillis, C. R. (1998). Do inhibitors of angiotensin-I-convertingenzyme protect against risk of cancer? . *Lancet*, *352*(9123), 179-184.

Li, C., Song, H., Wang, X., Ding, X., Liu, S., Dai, G., . . . Ju, W. (2017). Urinary metabolomics reveals the therapeutic effect of HuangQi Injections in cisplatin-induced nephrotoxic rats. *Scientific reports*, *7*(1), 1-12.

LI, X., Asmitananda, T., Gao, L., Gai, D., Song, Z., Zhang, Y., . . . Chen, M. (2012). iomarkers in the Lung Cancer Diagnosis: A Clinical Perspective. *Neoplasma*, *59*(5), 500-507.

Li, X., Yang, S., & Qiu, Y. (2010). Urinary Metabolomics as a Potentially Novel Diagnostic and Stratification Tool for Knee Osteoarthritis. *Metabolomics*, *6*(1), 109-118.

Liang, Q., Wang, C., Li, B., & Zhang, A. (2015). Metabolic fingerprinting to understand therapeutic effects and mechanisms of silybin on acute liver damage in rat. *Pharmacognosy magazine* , *11*(43), 586-593.

Lim, S., Jia, Z., Zhang, H., Teng, C., Bay, B., Shen, H., & Ong, C. (2018). Metabolic signatures of four major histological types of lung cancer cells. *Metabolomics*, *14*(9), 1-13.

Liu, X., Zhang, M., Cheng, X., Liu, X., Sun, H., Guo, Z., . . . Ji, Z. (2020). LC-MS-Based Plasma Metabolomics and Lipidomics Analyses for Differential Diagnosis of Bladder Cancer and Renal Cell Carcinoma. *Frontiers in Oncology*, *10*, 717.

Llandi, M., Zhang, T., Garcia-Closas, M., & Bosse, Y. (2019). Abstract SY26-02: Sherlock-Lung : Tracing lung cancer mutational processes in never smokers. *American Association for Cancer Research Annual Meeting.* Atlanta: Cancer Research.

Lodi, A., Tiziani, S., Khanim, F., Gunther, U., Viant, M., Morgan, G., . . . Drayson, M. (2013). Proton NMR-Based Metabolite Analyses of Archived Serial Paired Serum and Urine Samples from Myeloma Patients at Different Stages of Disease Activity Identifies Acetylcarnitine as a Novel Marker of Active Disease. *PLoS One*, *8*(2), e56422.

Lualdi, M., & Cavalli, A. (2018). Early detection of colorectal adnecarcinoma: a clinical decision support tool based on plasma porphyrin accumulation and risk factors. *BMC Cancer* , *18*(1), 1-9.

Lui, X., Xue, S., & Jiang, H. (2019). CDCA promotes non-small-cell lung cancer (NSCLC) migration by regulating Akt/Erk1/2 signaling pathways. *European Respiratory Journal*.

Ma, J., Zheng, N., Xie, G., & Li, H. (2018). Emerging Applications of Metabolomics in Drug Development. *Journal of Metabolomics and Systems Biology*, *1*(3), 1-3.

Macmillan. (2020, August 25). *Gemcitabine (Gemzar®)*. Retrieved from Macmillan Cancer Support: https://www.macmillan.org.uk/cancer-information-and-support/treatments-and-drugs/gemcitabine

Maynard, D., Groger, H., Dierks, T., & Dietz, K. (2018). The function of the oxylipin 12-oxophytodienoic acid in cell signaling, stress acclimation, and development. *Journal of Experimental Botany*, *69*(22), 5341-5354.

McDougall, R. (2013). Metabolism of thyroid hormones. In R. McDougall, *Thyroid Disease in Clinical Practice* (pp. 23-24). Springer.

Meng, L., Zhongmin, P., Quingwei, L., Sun, J., Yao, S., & Liu, Q. (2012). Value of 11C-choline PET/CT for lung cancer diagnosis and the relation between choline metabolism and proliferation of cancer cells. *Oncology Report*, *29*(1), 205-211.

Mikkonen, J., Singh, S., Herrala, M., Lappalainen, R., Myllymaa, S., & Kullaa, A. (2015). Salivary metabolomics in the diagnosis of oral cancer and peridontal diseases. *Journal of peridontal research*, *51*(4), 431-437.

Milburn, M., & Lawton, K. (2013). Application of Metabolomics to Disgnosis of Insulin Resistance . *Annual Review of Medicin*, *64*, 291-305.

Miller, S. (1994). Saliva testing - a nontraditional diagnostic tool. *Clinical laboratory science: journal of the American Society for Medical Technology*, *7*(1), 39-44.

Miller, W., & Auchus, R. (2011). The Molecular Biology, Biochemistry, and Physiology of human Steroidogenesis and Its Disorders. *Endocrine Reviews*, *32*(1), 81-151.

Mitutani, Y., Nakajima, T., Morinaga, S., Gotoh, M., Shimosato, Y., Akino, T., & Suzaki, A. (1988). Immunohistochemical Localization of Pulmonary Surfactant Apoproteins in Various Lung Tumors . *Cancer*, *61*(3), 532-537.

Molina-Romero, C., Rangel-Escareno, C., Ortega-Gomez, A., Alanis-Funes, G., Aviles-Salas, A., Avila-Moreno, F., . . . Arrieta, O. (2017). Differential gene expression profiles according to the Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society histopathological classification in lung adenocarcinoma subtypes. *Human Pathology*, *66*, 188-199.

Motta, J., Steffen, R., Lobato, C., Mendonca, V., & Silva, J. (2020). Endobronchial ultrasound-guided transbronchial needle aspiration verses mediastinoscopy for mediastinal staging of lung cancer: a systematic review of economic evaluation studies. *PLoS One*, *15*(6), e0235479.

Mudipalli, A., & Zelikoff, J. (2017). *Essential and Non-essential Metals.* Springer International Publishing.

Mullard, A. (2014). A new drug costs US $2.6 billion to develop. *Nature Reviews Drug Discovery*, *13*(12), 877.

Nagy, C., & Haschemi, A. (2013). Sedoheptulose kinase regulates cellular carbohydrate metabolism by sedoheptulose 7-phosphate supply. *Biochemical Society Transactions*, *41*(2), 674-680.

Nakagawa, H., Hayata, Y., Kawamura, S., Yamada, T., Fujiwara, N., & Koike, K. (2018). Lipid Metabolic Reprogramming in Hepatocellular Carcinoma. *Cancers*, *10*(11), 447.

National Cancer Institute. (2020, February 18). *NCI Dictionary of Cancer Terms*. Retrieved from National: https://www.cancer.gov/publications/dictionaries/cancerterms/def/biomarker

National Institute of Health. (2001). Biomarkers and surrogate endpoints: preferred definitions and. *Clincal Pharmacolog*, *69*(3), 89-95.

Neal, R. (2009). Do diagnostic delays matter? *British Journal of Cancer*, *101*(2), S9-S12.

Necula, L., Matej, L., Neagu, A., Mambet, C., Nedeianu, S., Bleotu, C., . . . Chiyu-Economescu, M. (2019). Recent advances in gastric cancer early diagnosis. *Worl Journal of Gastroenterology*, *25*(17), 2029-2044.

NHS. (2020, February 13). *Lung Cancer*. Retrieved from NHS Inform: https://www.nhsinform.scot/illnesses-and-conditions/cancer/cancer-types-in-adults/lung-cancer#causes-of-lung-cancer

Ni, J., Xu, L., Li, W., Zheng, C., & Wu, L. (2019). Targeted metabolomics for serum amino acids and acylcarnitines in patients with lung cancer. *Experimental and Therapeutic Medicine* , *18*(1), 188-198.

NICE. (2019, March 28). Lung cancer: diagnosis and management. *NICE guideline [NG122]*. National Institute for Health and Care Excellence .

Nicholas, B., & Djukanovic, R. (2009). Induced sputum: a window to lung pathology . *Biochemical Society Transactions*, *37*(4), 868-872.

Nikolau, B. (2007). *Concepts in Plant Metabolomics.* The Netherlands: Springer.

Nishiumi, S., Shinohara, M., Ikeda, A., Yoshie, T., Hatano, N., & Kakuyama, S. (2010). Serum metabolomics as a novel diagnostic approach for pancreatic cancer. *Metabolomics*, *6*(4), 518-528.

Noreldeen, H., Liu, X., & Xu, G. (2019). Metabolomics of Lung cancer: analytical platforms and their applications. *Journal of separation science*, *43*(1), 120-133.

Okuma, N., Saita, M., Hoshi, N., Soga, T., Tomita, M., Sugimoto, M., & Kimoto, K. (2017). Effect of masticatory stimulation on the quantity and quality of saliva and the salivary metabolomic profile. *PLoS One*, *12*(8), e0183109.

Olsson, M., Gustadsson, O., Skogastierna, C., & Tolf, A. (2017). Regulation and expression of human CYP7B1 in prostate: Overexpression of CYP7B1 during progression of adenocarcimoma2017. *The Prostate*, *67*(13), 1439-1446.

Oser, M. G., Niederst, M. J., & Sequist, L. (2015). Transformation from non-smallcell lung cancer to small-cell lung cancer: molecular drivers and cells of origin. *The Lancet Oncology*, *16*(4), e165-e172.

Overvad, K., Diamant, B., & Holm, L. (1999). Coenzyme Q10 in health and disease. *European Journal of clinical nutrition*, *53*(10), 764-770.

Pandak, W., & Kakiyama, G. (2019). The acidic pathway of bile acid synthesis:Not just an alternative pathway. *US department of veteran affairs*, *3*(2), 88-98.

Pandak, W., Schwarz, C., Hylemon, P., Mallonee, D., Valerie, K., Heuman, D., . . . Vlahcevic, Z. (2001). Effects of CYP7A1 overexpression on cholesterol and bile acid homeostasis. *American Journal of Physiology Gastrointestinal and Liver Physiology*, *281*(4), G878-G889.

Perlikos, F., Harrington, K., & Syrigos, K. (2013). Key molecular mechanisms in lung cancer invasion and metastasis: A comprehensive review. *Critical Reviews in Oncology/Hematology*, *87*(1), 1-11.

Piehl, S., Hoefig, C., Scanlan, T., & Hohrle, J. (2011). Thyronamines—Past, Present, and Future. *Endocrine Reviews*, *32*(1), 64-80.

Pietzner, M., Engelmann, B., Kacprowski, T., Golchert, J., Dirk, A., Hammer, E., . . . Brabant, G. (2017). Plasma proteome and metabolome characterization of an experimental human thyrotoxicosis model. *BMC Medicine*, *15*(1), 1-18.

Powers, R. (2009). NMR metabolomics and drug discovery. *Magnetic Resonance in Chemistry* , *47*(S1), S2-S11.

Puchades-Carrasco, L., Jantus-lewintre, E., Perez-rambia, C., Garcia-Garcia, F., Lucas, R., Calabuig, S., . . . Pineda-Lucena, A. (2016). Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer. *Oncotarget*, *7*(11), 12904.

Quint, L., Francis, I., Wahl, R., Gross, B., & Glazer, G. (1995). Preoperative staging of non-small-cell carcinoma of the lung: imaging methods. *American Journal of Roentgenology*, *164*(6), 1349-1359.

Rahrovani, F., Javanbakht, M., Ehsani, A., Esrafili, A., Mohammadi, H., Ghaedi, E., . . . Djalali, M. (2018). Erythrocyte membrane saturated fatty acids profile in newly diagnosed Basal Cell Carcinoma patients. *Clinical Nutrition ESPEN*, *23*, 107-111.

Rappaport, S., Barupal, D., Wishart, D., Vineis, P., & Scalbert, A. (2014). The blood exposome and its role in discovering causes of disease. *Environmental Health Perspective*, *122*(8), 769-774.

Rawla, P., & Barouk, A. (2018). Epidemiology of gastric cancer: global trends, risk factors and prevention. *Przeglad Gastroenterologiczny*, *14*(1), 26.

Reynolds, M., Kastury, K., Groskopf, J., Schalken, J., & Rittenhouse, H. (2007). Molecular markers for prostate cancer. *Cancer Letters*, *249*(1), 5-13.

Ricciotti, E., & FitzGerald, G. (2011). Prostaglandins and Inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *31*(5), 986-1000.

Roessner, U., & Bowne, J. (2018). What is metabolomics all about? *Biotechniques*, *46*(5), 363-365.

Rontondo, R., Rizzolio, F., Perin, T., Berreta, M., Zanconati, F., Giordano, A., & Canzonieri, V. (2019). Pathological Diagnosis and Classification of Gastric Epithelial Tumours. *Gastric Cancer In The Precision Medicine Era*, 53-82.

Ros-Mazurczyk, M., Wojakowska, A., Marczak, L., Polanski, K., Pietrowska, M., Polanska, J., . . . Widlak, P. (2017). Panel of serum metabolites discriminates cancer patients and healthy participants of lung cancer screening – a pilot study. *The Journal of the Polish Biochemical Society and of the Polish Academy of Sciences*, *64*(3), 513-518.

Ruiying, C., Zeyun, L., Yonglang, Y., Zijia, Z., Ji, Z., & al., e. (2020). A comprehensive analysis of metabolomics and transcriptomics on small cell lung cancer. *PLoS One*, *15*(5), e0232272.

Ruiz-Jimenez, J., Priego-Capote, F., Mata-Granados, J., Quesada, J., & Castro, L. (2007). Determination of the ubiquinol-10 and ubiquinone-10 (coenzyme Q10) in human serum by liquid chromatography tandem mass spectrometry to evaluate the oxidative stress. *Journal of chromatography A*, *1175*(2), 242-248.

Rutherford, M., Hinchliffe, S., Abel, G., Lyratzopoulos, G., Lambert, P., & Greenberg, D. (2013). How much of the deprivation gap in cancer survival can be explained by variation in stage at

diagnosis: An example from breast cancer in the East of England . *International Journal of Cancer*, *133*(9), 2192-2200.

Sanderson, S., Gao, X., & Dai, Z. (2019). Methionine metabolism in health and cancer: a nexus of diet and precision medicine. *Nature Reviews Cancer*, *19*(11), 625-637.

Scanlan, T., Suchland, K., Hart, M., Chiellini, G., Huang, Y., & Kruzich, P. (2004). 3-Iodothyronamine is an endogenous and rapid-acting derivative of thyroid hormone. *10*(6), 638-642.

Scheucher, J., Schworer, B., Zirngibl, C., Koch, U., Weber, W., Egert, E., . . . Gresinger, C. (1992). Determination of the relative configuration of 5,6,7,8-tetrahydromethanopterin by two-dimensional NMR spectroscopy. *FEBS Letters*, *314*(3), 440-444.

Selenius, M., Rundlof, A., Olm, E., Fernandes, A., & Bjornstedt, M. (2010). Selenium and the selenoprotein thioredoxin reductase in the prevention, treatment and diagnostics of cancer. *Antioxidants & Redox Signaling*, *12*(7), 867-880.

Seo, S., Kim, E., Park, S., Byun, S., Lee, S., Bok, S., . . . Son, H. (2018). GC/MS-Based Metabolomics Reveals Biomarkers in Asthma Murine Model Modulated by Opuntia humifusa. *Evidence-Based Complementary and Alternative Medicine*, https://doi.org/10.1155/2018/1202860.

Seyfried, T., Ando, S., & Yu, R. (1978). Isolation and characterization of human liver hematoside. *Journal of Lipid Research*, *19*(5), 538-543.

Shafiee, M., Ortori, C., Barrett, D., Mongan, N., Abu, J., & Atiomo, W. (2020). Lipidomic Biomarkers in Polycystic Ovary Syndrome and Endometrial Cancer. *International Journal of Molecular Sciences*, *21*(13), 4753.

Shivappa, N., Herbert, J., Ferraroni, M., La Vecchia, C., & Rossi, M. (2016). Association between dietry inflammatory index and gastric cancer risk in an italian case-control study. *Nutrition and cancer* , *68*(8), 1262-1268.

Siegel, R., Miller, K., & Jemal, A. (2020). Cancer statistics 2020. *CA: A Cancer Journal for Clinicians* , *70*(1), 7-30.

Slowikowski, B., Lianeri, M., & Jagodzinski, P. (2016). Exploring estrogenic activity in lung cancer. *Molecular Biology Reports* , 35-50.

Sohn, B., Hwang, J., Jang, H., Lee, H., Oh, S., Shim, J., . . . al., e. (2017). Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project. *Clinical cancer research*, *23*(15), 4441-4449.

Song, T., Zhang, J., Li, B., Zhou, H., Jiang, P., Yang, J., . . . Guo, Q. (2015). Misdiagnosis of a Small Cell Lung Cancer Resulting From Inaccurate Pathology. *Annuals of Thoracic Surgery*, *99*(5), e125-e127.

Song, X., Yang, X., Narayanan, R., Shankar, V., Ethiral, S., Wang, X., . . . Zare, R. (2020). Oral squamous cell carcinoma diagnosed from saliva metabolic profiling. *PNAS*, *117*(28), 16167-16173.

Spinelli, J., Yoon, H., Ringel, A., Jeanfavre, S., & Clish, C. (2017). Metabolic recycling of ammonia via glutamate dehydrogenase supports breast cancer biomass. . *Science*, *358*(6365), 941-946.

Stamatis, G. (2015). Staging of lung cancer: the role of noninvasive, minimally invasive and invasive techniques. *European Respiratory Journal*, *46*(2), 521-531.

Steffens, D., Jiang, W., Karoly, E., & Mitchell, M. (2010). Metabolomic differences in heart failure patients with and without major depression. *Journal of Geriatric Psychiatry and Neurology*, *23*(2), 138-146.

Sugimoto, M., Furuta, T., Shirai, N., Ikuma, M., Sugimura, H., & Hishida, A. (2006). Influences of Chymase and Angiotensin I-Converting EnzymeGene Polymorphisms on Gastric Cancer Risks in Japan. *Cancer Epidemiology, Biomarkers & Prevention*, *15*(10), 1929-1934.

Tacconelli, S., & Patrignani, P. (2014). Inside epoxyeicosatrienoic acids and cardiovascular disease. *Frontiers in Pharmacology*, *5*, 239.

Tang, Y., Zhang, J., Huang, H., He, X., Zhang, J. O., Li, G., . . . Zhang, G. (2019). Pleural IFN-γ release assay combined with biomarkers distinguished effectively tuberculosis from malignant pleural effusion. *BMC Infectious Diseases*, *19*(1), 1-8.

Teunissen, C., Verheul, C., & Willemse, E. (2018). Cerebrospinal Fluid in Neurologic Disorders. In *Handbook of clinical neurology* (pp. Volume 146, Pages 3-20). Elsevier.

Thomas, T., & Thomas, T. (2003). Polyamine metabolism and cancer. *Journal of Cellular and Molecular Medicine*, *7*(2), 113-126.

Trygg, J., Holmes, E., & Lundstedt, T. (2006). Chemometrics in Metabonomics. *Journal of Proteome Research* , *6*(2), 469-479.

Tsao, S., Yin, M., & Liu, W. (2007). Oxidant Stress and B Vitamins Status in Patients With Non-Small Cell Lung Cancer. *Nutrition and Cancer*, *59*(1), 8-13.

U.S. Food and Drug Administration. (2020, 04 14). *FDA - U.S. Food and Drug Administration*. Retrieved from List of Qualified Biomarkers: https://www.fda.gov/drugs/biomarker-qualification-program/list-qualified-biomarkers

Uefuji, K., Ichikura, T., & Mochizuki, H. (2000). Cyclooxygenase-2 expression is related to prostaglandin biosynthesis and angiogenesis in human gastric cancer. *Clinical Cancer Research*, *6*(1), 135-138.

Uusi-Oukari, M., Soini, S., Heikkila, J., Koivisto, A., Neuvonen, K., Pasanen, P., . . . Korpi, E. (2000). Sterospecific modulation of GABAA receptor function by urocanic acid isomers. *European Journal of Pharmacology*, *400*(1), 11-17.

Van Cutsem, E., Sagaert, X., Topal, B., Haustermans, K., & Prenen, H. (2016). Gastric cancer . *The Lancet*, *388*(10060), 2654-2664.

Verschuur, A. (2007). Cytidine triphosphate synthetase (CTP synthetase) as a druggable target in cancer. *Drugs of the future*, *32*, 1071-1080.

Vettore, L., Westbrook, R., & Tennant, D. (2020). New aspects of amino acid metabolism in cancer. *British Journal of Cancer*, *122*(2), 150-156.

Vinayavekhin, N., Homan, E., & Saghatelian, A. (2010). Exploring disease through metabolomics. . *American Chemical Society*, *5*(1), 91-103.

Wallace, H., Fraser, A., & Hughes, A. (2003). A perspective of polyamine metabolism. *Biochemical Journal*, *376*(1), 1-14.

Wang, Y., Shi, J., & Liya, G. (2020). Gamma linolenic acid suppresses hypoxia-induced proliferation and invasion of non-small cell lung cancer cells by inhibition of HIF1α. *Genes and Genomics*, *42*(8), 927-935.

Want, E., & Masson, P. (2011). Processing and analysis of GC/LC–MS-based metabolomics data. *Methods in Molecular Biology*, 277-298.

Warth, A., Cortis, J., Soltermann, A., Meister, M., Budczies, J., Stenzinger, A., . . . Weichert, W. (2014). Tumour cell proliferation (Ki-67) in non-small cell lung cancer: a critical reappraisal of its prognostic role. *British Journal of Cancer*, *111*(6), 1222-1229.

Watson, D., Tonelli, F., Alossaimi, M., Williamson, L., Chan, E., Gorshkova, I., . . . Pyne, S. (2013). The roles of sphingosine kinases 1 and 2 in regulating the Warburg effect in prostate cancer cells. *Cellular Signalling*, *25*(4), 1011-1017.

Wettersen, N., & Maisel, A. (2015). Role of cardiac troponin levels in Acute Heart Failure. . *Cardiac Failure Review*, *1*(2), 102.

Wheelock, C., Goss, V., Balgoma, D., Nicholas, B., Skipp, P., & Rossios, C. (2013). Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *European Respiratory Journal*, *42*(3), 802-825.

Wild, J., Shanmuganathan, M., Hayashi, M., Potterb, M., & Britz-McKibbin, P. (2019). Metabolomics for improved treatment monitoring of phenylketonuria: urinary biomarkers for non-invasive assessment of dietary adherence and nutritional deficiencies. *Analyst*, *144*(22), 6595-6608.

Wishart, D. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Reviews Drug Discovery*, *15*(7), 473.

Wood, P. (2013). Mass Spectrometry Strategies for Clinical Metabolomics and Lipidomics in Psychiatry, Neurology, and Neuro-Oncology. *Neuropsychopharmacology Reviews*, *39*(1), 24-33.

World Health Organsiation. (2018, September 12). *Cancer*. Retrieved from World health organisation: https://www.who.int/news-room/fact-sheets/detail/cancer

Worley, B., Halouska, S., & Powers, R. (2013). Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical of Biochemistry*, *433*(2), 102-104.

Wu, H., Xue, R., Tang, Z., Deng, C., Lui, T., Zeng, H., . . . Shen, X. (2010). Metabolomic investigation of gastric cancer tissue using gas chromatography/mass spectrometry. *Analytical and Bioanalytical Chemistry*, *396*(4), 1385-1395.

Wu, Y., Chiu, C., Hsueh, C., & Hsueh, C. (2018). The role of bile acids in cellular invasiveness of gastric cancer. *Cancer Cell International*, *18*(1), 1-8.

Xia, J., Broadhurst, D., Wilson, M., & Wishart, D. (2013). Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics*, *9*(2), 280-299.

Yarla, N., Bishayee, A., Sethi, G., Reddanna, P., Kalle, A., Dhananjaya, B., . . . Duddukuri, G. (2016). Targeting arachidonic acid pathway by natural products for cancer prevention and therapy. *Seminars in Cancer Biology*, Vol. 40, 48-81.

Zang, A. H., Sun, H., Qiu, S., & Wang, X. (2013). Metabolomics in non-invasive breast cancer. *Clinica Chimica Acta* , *424*, 3-7.

Zhang, X., Li, Q., Xu, Z., & Dou, J. (2020). Mass spectrometry-based metabolomics in health and medical science: a systematic review. *RSC Advances*, *10*(6), 3092-3104.

Zhao, Y., & Lin, R. (2014). Metabolomics in Nephrotoxicity. *Advances in clinical chemistry*, *65*, 69-89.

Zhu, J., Djukovic, D., Deng, L., Gu, H., Himmati, F., Zaid, M., . . . Raftery, D. (2015). Targeted serum metabolite profiling and sequential metabolite ratio analysis for colorectal cancer progression monitoring. *Analytical and Bioanalytical Chemistry*, *407*(26), 7857-7863.

Appendices

Appendix 1 – Patient data associated with the Saliva Samples for Lung Cancer Diagnosis

Sample Code	Diagnosis	Age	Sex	BMI	Smoking History	Pack years	Medical History	Drug History	Clinical Diagnosis Lung Cancer	Past History of Cancer	Family History of Cancer	If yes to family history of cancer, specify	Final LC Diagnosis	T	N	M	Stage (I – IV)
A305	HC	47	m	23.2	current	53	obstructive apnoea	varenicline; calcipotril	no	no	yes	sister LC (maybe)	n/a	n/a	n/a	n/a	n/a
A001	HC	48	m	26.4	never	n/a	hypertension	valsartan	no	no	no	n/a	n/a	n/a	n/a	n/a	n/a
A079	HC	76	m	34.7	never	n/a	atrial fibrillation; lv systolic dysfunction; hypertension; previous CVA; hypertension; rem sleep disorder	carvedilol; clonazepam; warfarin; ramipril; simvastatin; furosemide; spironolactone	no	no	no	n/a	n/a	n/a	n/a	n/a	n/a
E018	LC	68	f	23.1	ex	50	rheumatoid arthritis, osteopenia, herpes simplex	methotrexate, celebrex, hydroxychloroquine, co-codamol, zimovave, strontium, folic acid, prednisolone	yes	-	no	n/a	NSCLC	3 or 4	0	0	-

E053	LC	77	m	21.7	ex	40	pleural effusion	statins?	yes - lobectomy	-	no	n/a	NSCLC	3	0	0	-
E078	LC	67	m	22.8	ex	14	inguinal hernia	salbutamol	yes - cisplatin/ etopside, radiotherapy	-	no	n/a	NSCLC	4	1	0	IIIA
E057	LC	81	f	19.1	current	60	supraglottic cancer, emphysema, COPD	chemo (2017),co-codamol, amitryptyline, aspirin, du...?, tiottropium, salbutamol	yes - treatment received	-	yes	supraglottic squamous carcinoma	NSCLC	-	-	-	-
E049	LC	72	m	30.6	ex	63	asbestos exposure	-	yes - cisplatin/ etoposide 26/9/2016, radiotherapy 2016	-	no	n/a	NSCLC	4	1	0	-
E082	LC	71	m	40	ex	100	COPD, chest wall hernia, type 2 diabetes	prednisolone, frusamide, tamsulosin, inhalers?, alondronate	yes - SABR radiotherapy	-	no	n/a	NSCLC	1b	0	0	1A
E074	LC	61	f	32.1	ex	23	penicillin allergy, osteoarthritis	amtriptyline, oramorph, etodolac	yes - pembrolymeb? dec 2018	-	no	n/a	NSCLC	4	3	1a	-

E046	LC	77	m	27.6	ex	55	ischemic heart disease, AF, disliademia, hypertension, COPD	rivaroxaban, peridopril, felodipine, bisoprolol, simvastatin, aspirin	yes - cisplatin/ etoposide 20/3/2018, radiotherapy	-	no	n/a	NSCLC	2	2	0	-
E105	LC	80	m	20.9	ex	10	COPD, bronchiectasis, benign prostatic hyper trophy, reflux, dyslipaenemia	tamsulosin, tolterodine, finasteride, omeprazole, trimethoprim, cephalexin, salbutamol	yes - lobectomy	-	yes	mother - breast, father - throat	NSCLC	2a	0	0	-
E092	LC	70	f	-	ex	30	CABG, lung ca, hysterectomy, cholerysterectomy, polymyalgia rheumatica, Ca oesophagus	-	yes	-	yes	-	NSCLC	2a	2	0	-
E098	LC	59	f	20.5	current	43	depression	citalopram	yes	-	yes	lung	NSCLC	4	2	1b	IV
E060	LC	71	-	21.3	ex	57	MI, coronary artery stent, ischemic heart disease	clopidogrel, bisoprolol, atorvastatin	yes	-	no	n/a	NSCLC	4	1	0	IIIA

E073	LC	79	m	-	ex	8	B3 thymona, gout, hypertension, osteoporosis, asbestos exposure, lung ca (2015)	digoxin, lansoprolol, lisinopril, tamsulosin, bisoprolol, solofonacin, colecalciferol	yes - radiotherapy 2015	-	no	n/a	NSCLC	1b	0	0	IA
E075	LC	78	m	23.1	ex	20	chrones, IDA (iron deficiency aenemia), myeloprofeature disorder	aspirin, ferrous fumerate, fru...?, domperidone, zaparin, spirolactone, omeprazole	yes - stereotactic radiotherapy	-	no	n/a	NSCLC	1a	0	0	I
E090	LC	68	f	22	current	66	IBS, arthritis, diverticulosis, depression	mebeverine, morphine, citalopram, ventolin, omeprazole	yes - carboplatin/peclitaxed?	-	yes	lung, colorectal	NSCLC	4	0	0	-
E096	LC	80	f	21	ex	60	lung nodule, alzhemers dis...?, DVT 2016	quinine sulphate, donepezil, bisoprolol, simvastatin, omeprazole	yes	-	no	n/a	NSCLC	3	2	0/1a	-
E054	LC	67	f	25.6	ex	40	COPD, due to have a nephrectomy, IBS, tonsilectomy, hysterectomy	ellipta inhaler	yes - lobectomy	-	yes	father - testicular and bladder	NSCLC	2	0	0	-
E058	LC	68	f	35.3	ex	-	carpel tunnel, nasal polyposis,	mebeverine, flixonase,	yes - right	-	yes	ovarian, breast,	NSCLC	2b	2	x	-

							IBS, post concussion syndrome, osteoarthritis, allergies?, varicose eczema	fexofenadine, naproxen, blink eye drops, glucosamine, evening primrose, oxicodone, paracetamol	upper lobectomy			cervical, facial					
E089	LC	69	m	25	current	28	type 2 diabetes, bowel cancer, asthma	salbutamol	yes - chemo, immunotherapy and radiotherapy	-	yes	cervical	NSCLC	3	0	0	-
E070	LC	64	f	30.1	never	n/a	benign goitre, TAH (total abdominal hysterectomy)	-	yes - erotinib?	-	no	n/a	NSCLC	2a	0	1b	-
E083	LC	57	f	20.8	ex	26	VATS lobectomy, multiple sclerosis, depression	clonazepam, interferon, baclofen	yes - lobectomy, chemotherapy carboplatin/vibrobrin?	-	-	n/a	NSCLC	2a	0	x	IB
E081	LC	76	m	16.7	current	40	COPD, hypertension, aortic aneurism, reflux	ramapril, omeprazole, zapanin, finasteride, frumiside, aspirin	yes - radiotherapy L3-5 vertebrae	-	no	n/a	NSCLC	2b	2	1b	-

E051	LC	-	f	23.2	ex	10	diverticulosis?, EBUS, hysterectomy	prednisolone	yes	-	no	n/a	NSCLC	3	3	1a	-
E080	LC	67	m	28.2	ex	40	BPH (benign prostatic hyperplasia), diabetes, hypertension	alfuzosin, afervastatin, fuesteside, ramapril, metformin, co-codamol	yes	-	no	n/a	NSCLC	3	0	0	IIB
E104	LC	69	f	22.2	ex	44	COPD, appendectomy, chronic kidney disease	atenolol, salbutamol, serverent, ramipril	yes - lobectomy	-	yes	aunt - stomach, lymph nodes	NSCLC	1c	0	0	I
E022	LC	65	f	20.3	ex	104	gout, carpel tunnel, depression	none	yes	-	yes	-	NSCLC	2b	0	0	-
E071	LC	73	f	25.8	ex	47	hysterectomy	-	yes	-	yes	mother - cervix	SCLC	3	3	0	-

Appendix 2 – Patient data associated with the Saliva samples for Lung cancer Staging

Sample code	Diagnosis	Sex	TnmType	TnmT	TnmN	TnmM	Stage	Smoking status
X0160	adeno	M	Clinical	T1a	N2	M0	1	Ex Smoker (22 years free)
X1337	adeno	F	Pathological	T2	N1	M1	4	Current smoker - 20-30/day
X1436	adeno	M	Clinical	T4	N0	M1b	4	Ex Smoker
X1470	adeno	M	Clinical	T2a	N0	M0	1	Ex Smoker
X1534	adeno	M	Clinical	T4	N2	M1a	4	Ex Smoker
X1563	adeno	M	Clinical	T4	N2	M0	3	Heavy Smoker
X1586	adeno	M	Pathological	T3	N1	MX	3	Non Smoker
X1587	adeno	M	Clinical	T4	N3	M1b	4	Ex Smoker (15 years free)
X1588	adeno	M	Clinical	TX	NX	M1b	4	-
X1593	adeno	M	Pathological	T1b	N1	M1b	4	Ex Smoker (25 years free)
X1616	adeno	F	Clinical	T4	N0	M0	3	Ex Smoker
X1633	adeno	F	Clinical	T2a	N2	M1a	4	Non Smoker
X1643	adeno	F	Pathological	T2a	N0	MX	1	-
X1645	adeno	M	Pathological	T2b	N2	MX	3	Smoker
X1733	adeno	M	Clinical	T1b	N2	M1b	4	Ex-smoker of 50 years
X1736	adeno	F	Pathological	T3	N1	M0	3	Current smoker
X1740	adeno	M	Pathological	T2b	N2	M0	3	Lifelong smoker
X1748	adeno	M	Pathological	T4	N2	M1a	4	Current smoker
X1769	adeno	F	Pathological	T2a	N0	M0	1	Ex-smoker - stopped at diagnosis
X1777	adeno	F	Pathological	T2b	N0	MX	1	Ex-smoker
X1782	adeno	F	Clinical	T3	N1	M1c	4	Ex-smoker - stopped at diagnosis
X2398	adeno	M	Clinical	T3	N2	M1c	4	Stopped smoking 20 years ago - smoked 20-40/day
X2701	adeno	M	Pathological	T3	N0	M0	2	Ex-smoker of 33 years
X2704	adeno	F	Pathological	T2b	N2	MX	1	Ex-smoker - stoppped at diagnosis - was 20/day since age 14
X2780	adeno	F	Pathological	T3	N0	M1b	4	-

X4303	adeno	F	Pathological	T2b	N0	M not applicable	2	Ex Smoker (30 years free)
X4304	adeno	M	Pathological	T1b	N0	MX	1	-
X4305	adeno	F	Pathological	T2a	N1	M not applicable	2	Ex Smoker (less than 10 years)
X4315	adeno	M	Pathological	T2a	N1	MX	2	-
X4333	adeno	F	Pathological	T1b	N0	M not applicable	1	-
X4334	adeno	F	Pathological	T1b	N0	MX	1	-
X4335	adeno	F	Pathological	T2a	N0	MX	2	-
X4859	adeno	F	Pathological	T2b	N1	M0	3	-
X5174	adeno	F	Clinical	T3	N2	M0	3	-
X5180	adeno	F	Clinical	T4	N3	M1c	4	Ex-smoker of 1 month
X5184	adeno	F	Pathological	T2b	N0	MX	1	Ex-smoker - stopped at diagnosis
X5185	adeno	F	Pathological	T1b	N0	M0	1	Ex-smoker
X5195	adeno	F	Pathological	T1b	N0	MX	1	-
X1581	squam	M	Clinical	T2b	N2	M0	3	Ex-smoker
X1590	squam	M	Clinical	T3	N1	M0	3	-
X1599	squam	F	Clinical	T4	N2	M0	3	Ex-smoker - stopped at diagnosis
X1649	squam	M	Clinical	T4	N2	M0	3	Lifelong smoker
X1658	squam	M	Pathological	T2a	N0	MX	1	-
X1766	squam	M	Clinical	T2b	N0	M1c	4	Ex-smoker of 30 years
X1770	squam	M	Pathological	T3	N0	MX	2	Ex-smoker of 3 years
X1780	squam	M	Pathological	T4	N1	M0	3	Ex-smoker
X2787	squam	M	Clinical	T4	N2	M0	3	Ex-smoker of 5 years
X4301	squam	F	Pathological	T2a	N0	MX	1	-
X4302	squam	F	Pathological	T1a	N1	M not applicable	3	-
X4307	squam	F	Pathological	T1b	N0	M not applicable	1	-
X4326	squam	F	Pathological	T2a	N0	MX	1	-
X5173	squam	F	Clinical	T3	N1	M0	3	Ex-smoker - stopped at diagnosis

X5179	squam	F	Pathological	T1c	N0	MX	1	-
X5199	squam	F	Pathological	T3	N2	MX	3	Current smoker - 10/day

Appendix 3 – Patient data Associated with the Gastric Fluid samples

Sample Code	Diagnosis group	Treatment Responder	Gender	Age At Diagnosis
1	AdG	NR	M	67
3	AdG	R	F	41
4	AdG	NR	M	55
5	AdG	R	M	66
6	AdG	R	M	64
7	AdG	NR	F	77
8	AdG	NR	M	47
9	AdG	R	M	37
10	AdG	NR	F	69
11	AdG	NR	F	69
12	AdG	R	F	45
13	AdG	R	M	63
14	AdG	NR	M	53
15	AdG	NR	F	63
16	AdG	NR	M	74
17	AdG	R	F	53
18	CTRL	N/A	F	NA
19	AdG	NR	M	60
20	AdG	NR	M	80
22	CTRL	N/A	M	NA
23	AdG	NR	M	74
24	AdG	R	M	63
25	CTRL	N/A	F	NA
26	AdG	R	M	37
27	CTRL	N/A	M	NA
28	CTRL	N/A	F	NA
29	AdG	R	M	63
30	AdG	NR	M	60
31	AdG	R	M	63
32	AdG	R	M	62
34	AdG	R	M	55
35	AdG	NR	M	63
36	AdG	NR	F	57
37	AdG	NR	F	57
38	AdG	R	M	70
39	CTRL	N/A	M	NA
40	AdG	NR	M	78
41	AdG	NR	M	53
42	AdG	NR	F	77
43	AdG	NR	M	68

44	CTRL	N/A	M	NA
45	AdG	NR	F	37
46	CTRL	N/A	M	NA
47	AdG	NR	F	50
48	AdG	R	F	41
49	AdG	NR	M	78
50	AdG	R	M	61
51	CTRL	N/A	M	NA
52	AdG	NR	M	86
53	CTRL	N/A	M	NA