Investigating the Role of Cigarette Smoke and Alcohol in Oesophageal Squamous Cell Carcinoma

Raabie'ah Gamieldien

Supervisor: Prof. Denver Hendricks
Co- Supervisor: Assoc. Prof. M. Iqbal Parker

SUBMITTED in fulfilment of the requirements for the degree

MSc (Med) in Medical Biochemistry
Department of Integrative Biomedical Sciences
Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

February 2022

Declaration

I hereby declare that all the work presented in this thesis is my own and that it has not been previously submitted to UCT or any other university for any degree.
I understand the meaning of plagiarism and declare that all the work in this dissertation, except for what is properly acknowledged, is my own.

I allow the university to reproduce, for the purpose of research, either parts or the whole of this thesis in any manner.

Raabie’ah Gamieldien

February 2022

Acknowledgements

I would like to thank my supervisor, Denver, thank you for being truly kind and for your guidance. I would also like to thank my co-supervisor Prof. Parker and his lab members, Victoria, Hendrina and Didi, all of you have helped contribute to my learning and development.

I’d also like to thank the bioinformatician, Dr. Peter at the CPGR and Dr. Wei for going above and beyond helping me with the bioinformatics analysis.

Special mention to all the post-grad students on the sixth floor of Falmouth building, Caryn, Nonku, Gaby Arad and especially Bianca, Humaira and Shabs, I cannot express the gratitude I have for your support, both scientifically and emotionally. Science is a team sport and you have helped me more than you realise.

I’d also like to thank my family, to my little sister Ath-raa who knows nothing and doesn’t want to know anything about science, thank you for listening to my rants.

Abbreviations

ADH	Alcohol Dehydrogenase
ASR	Age-standardised Incidence Rate
AHR	Aryl Hydrocarbon Receptor
ALDH	Aldehyde Dehydrogenase
ATCC	American Type Culture Collection
bp	base pairs
CPGR	Centre for Proteomic and Genomic Research
CSC	Cigarette smoke condensate
cDNA	Complementary deoxyribonucleic acid
Ct	Cycle threshold
CYP P450	Cytochrome P450 genes
DEGs	Differentially Expressed Genes
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxy-Nucleotide Tri-Phosphate
EtOH	Ethanol
FACS	Fluorescence-activated cell sorting
GO	Gene Ontologyy
GSTs	Glutathione S-transferases
HBTE	Human Bronchial/Trachea Epithelial cells
HIFBS	Heat Inactivated Fetal Bovine Serum
MMPs	Matrix Metalloproteinases
MMR	Mis-Match Repair enzymes
MTHFR	methylenetetrahydrofolate reductase
MTT	3-(4,5-dimethylthiazol-2yl-)-2,5-diphenyl tetrazolium bromide
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NNN	N′-nitrosonornicotine
NTC	No template control
OAC	Oesophageal Adenocarcinoma
OC	Oesophageal Cancer
OSCC	Oesophageal Squamous Cell Carcinoma
PAHs	Polycyclic Aromatic Hydrocarbons
PCR	Polymerase chain reaction
PI	Propidium Iodide
P/S	Penicillin /Streptomycin
PBS	Phosphate Buffered Saline
QC	Quality Control
ROS	Reactive Oxygen Species
RT-qPCR	Quantitative Reverse Transcriptase-Polymerase Chain Reaction
RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing
SNPs	Single Nucleotide Polymorphisms
TA	Annealing temperature
Trypsin-EDTA	0.05% trypsin-10 mM EDTA
UGTs	Uridine diphosphate glucuronosyltransferases
UV	Ultraviolet
WHO	World Health Organisation

Figures and Tables

Figures

Tables

Abstract

The World Health Organisation (WHO) has estimated cancer to be one of the leading causes of death to people under the age of 70. Every year there are about 18.1 million new cancer cases worldwide; oesophageal cancer (OC) ranks seventh out of all incidence cases and sixth in mortality. OC has a poor 5-year overall survival rate of about a 15% (Arnal et al., 2015). This occurs as OC is largely asymptomatic and patients often seek medical assistance at a late stage of their cancer. This late diagnosis and a lack of efficient treatment has rendered OC a serious world health problem. Oesophageal Squamous Cell Carcinoma (OSCC) is the more common subtype of and accounts for about 90% of incident oesophageal cancers every year (Abnet et al., 2018) with the majority of OSCC cases occur in developing countries. The objective of this study was to investigate cigarette smoking and alcohol consumption as they have widely been reported as risk factors for OSCC. The project explored the impact of cigarette smoke condensate (CSC) treatment and EtOH on the expression of genes in cultured oesophageal cancer cells. Prior to treating the cells for expression analysis, cytotoxicity experiments were conducted to determine treatment conditions of CSC and EtOH which were sub-lethal, for the cell types and time-frames investigated. It was found that concentrations of 40 µg/ml CSC and 50 mM EtOH did not cause cell death for the time period of three days. Furthermore, we also showed that cell viability was maintained up to 10 days of treatment. RNA-sequencing then revealed a wide variety of genes that were differentially expressed in the OSCC cells treated with these selected concentrations of CSC and EtOH. One gene found to be differentially expressed in two RNA-Seq analyses was confirmed to be up-regulated by RT-qPCR. Seven genes, AHRR, ALDH3A1, CYP1A1, CYP1B1, GSTM3, GSTM4 and UGT1 A6, involved in xenobiotic metabolism and a number of other metabolic pathways were also altered in response to CSC and EtOH treatment. However, these genes' involvement require further confirmation by RT-qPCR. This result of this study confirms that we have designed a reliable experimental system to investigate the role of EtOH and CSC in the development of OSCC. These results gave us a deeper insight into the genes and pathways affected by CSC and EtOH which may contribute to OSCC. Additionally, the cytotoxicity data can be use for future experimental work and the RNA-seq data can be used for further investigation into the development of OSCC.

Table of Contents

Chapter 1: Literature Review

1.1 Introduction

Every year there are about 18.1 million new cancer cases worldwide (Bray et al., 2018). Incident cases increased by 28% in a ten year period between 2006 to 2016 with the least developed countries seeing the highest increase in cases (Fitzmaurice et al., 2018). The World Health Organisation (WHO) has estimated cancer is the first or second leading cause of death of people under the age of 70 in 112 out of 183 countries (Sung et al., 2021). According to the Global Cancer Statistics 2020, oesophageal cancer (OC) ranks seventh out of all incidence cases and sixth in mortality overall (Sung et al., 2021). In regions of sub-saharan Africa it ranks as the third most common cancer (Liu et al., 2017). In South Africa it has been estimated that there are about 40 000 deaths from cancer yearly (58 000 cases) and one of the leading causes of deaths in males is oesophageal cancer (11%) (Mayosi et al., 2009).

There are two subtypes of oesophageal cancer, Oesophageal Squamous Cell Carcinoma (OSCC) and Oesophageal Adenocarcinoma (OAC)- of the two distinct histologic subtypes, OSCC is the more common (Arnold et al., 2015). OSCC accounts for about 90% of incident oesophageal cancers every year (Abnet et al., 2018). Between the two subtypes there is quite a substantial geographic variation as well as different aetiologies (Sung et al., 2021) with the majority of OSCC cases occurring in developing countries with 79% of OSCC cases occurring in South and East-Asia. Conversely, OAC is more common in developed regions such as North America, Europe and Oceania. This is depicted in figure 1.1 below showing the Age-standardised Incidence Rate (ASR) per 100 000 of (A) oesophageal adenocarcinoma (OAC) and (B) oesophageal squamous cell carcinoma (OSCC) in men (Arnold et al., 2015).

The high mortality rate indicates the poor prognosis for patients diagnosed with oesophageal cancer. Of the two subtypes, both have a poor 5-year overall survival rate of about a 15% (Arnal et al., 2015). Patients tend to have poor prognosis due to the late stage diagnosis of their cancer (Abnet et al., 2018) and this occurs as OC is largely asymptomatic and patients tend to seek healthcare at a late stage of their cancer (Hendricks & Parker, 2002). Late diagnosis and a lack of efficient treatment has rendered OC a serious world health problem.

Some studies hypothesize that the large geographical variation in the incidence of OC may be due to differences in environmental factors (Sung *et al.*, 2021; Talukdar *et al.*, 2018). Interestingly, the two subtypes of OC have been reported to have different risk factors, for example case studies have shown that smoking and consuming alcohol are associated with an increased risk of OSCC but are not associated with oesophageal adenocarcinoma (Lagergren et al., 2000). Instead, OAC is associated with chronic gastric reflux and Barrett's Oesophagus (Arnal, Arenas and Arbeloa, 2015; Wheeler and Reed, 2012). Obesity and smoking have also been shown to play a role in OAC (Wheeler & Reed, 2012). In developed countries OAC cases are rising, possibly due to increased excess body weight and increasing gastroesophageal reflux disease (Sung *et al.*, 2021). In contrast there has been an observed decline in cases of OSCC in Asia, possibly resulting from economic gain and dietary improvement, whereas the decline of OSCC incidence in high-income areas in the west is thought to be due to declines in smoking (Sung *et al.*, 2021).

OSCC displays a large geographic distribution variation in the incidence of the disease. For both men and women the highest incidence rates are seen in Eastern Asia followed by Southern Africa, Eastern Africa, Northern Europe, and South Central Asia (Sung *et al.*, 2021). In the rural Eastern Cape area of South Africa, oesophageal cancer was the most commonly diagnosed cancer in males from 1998 to 2012 (Somdyala et al., 2015). The highest incidence rates occur in Eastern Asia, where in Mongolia and China rates are in the top 5 in the world. There are also high rates in various African countries, for example: It is the leading cause of cancer mortality in Kenyan men and some of the highest incidence rates worldwide are seen for both men and women in Malawi (Bray et al., 2018). In Eastern Africa male incidence rates rank third by region (Bray et al., 2018). Due to the high incidence of the disease in Southern Africa, the intention of this study is to better understand OSCC. These variations depend on many factors, which include the level of socio-economic development and associated social and life style factors (Bray et al., 2018). The following sections will review some of the risk factors associated with OSCC.

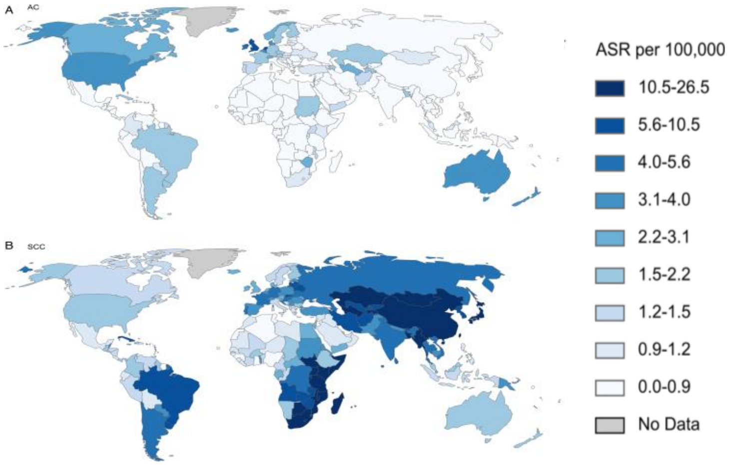

Figure 1.1: Map of the world showing distribution *Age-standardised incidence rate (ASR) per 100 000 of (A) oesophageal adenocarcinoma (OAC) and (B) oesophageal squamous cell carcinoma (OSCC) in men* (Arnold et al., 2015).

1.1 OSCC Risk Factors

Many factors have been identified as potential risk factors for OSCC. The main factors explored in this study are cigarette smoking and alcohol which have, along with their synergistic effects, been shown to be key risk factors for OSCC in western countries (Sung *et al.*, 2021). As previously mentioned, there is striking geographic variation in the prevalence of OSCC suggesting environmental factors may play a key role in the development of OSCC (Matejcic & Parker, 2015) and that the risk factors of OSCC may be different across geographic locations. In Western countries it is proposed that the major risk factors of oesophageal squamous cell carcinoma is heavy drinking, smoking and their cumulative effects (Bray et al., 2018; Lagergren et al., 2000; Wheeler & Reed, 2012; Zambon et al., 2000). In lower income countries such as regions of sub-Saharan Africa and Asia other risk factors which have been investigated are dietary factors such as nutritional deficiencies, pickled vegetables and very hot food and beverages (Sung *et al.*, 2021). Specifically in Africa, risk factors that have been identified to be associated with the development of OSCC are smoking, alcohol consumption, diets poor in fresh fruit and vegetables, consumption of foods contaminated with *Fusarium verticillioides* (Hendricks & Parker, 2002), HPV infection (Matsha et al., 2002), low socioeconomic status, genetic factors and HIV infection (Asombang et al., 2019).

A striking characteristic of OC aetiology is the gender disparity, about 70% of all patients are men (Middleton et al., 2019). Furthermore, an epidemiological study done in South Africa found that when looking at risk factors, males displayed the risk factors commonly associated with OSCC, such as cigarette smoking and alcohol consumption. However women with OSCC did not display the same trends with smoking and alcohol consumption, suggesting men and women may have different risk factors (Loots et al., 2017). These results were reflected in another study in African populations, a meta-study which found alcohol to be risk factor and went on to further suggest that men are more at risk of OSCC because it was found in these countries men tend to consume alcohol more than their female counterparts (Asombang et al., 2019).

Differences in risk factors among countries may be due to lifestyle factors and particulars of those specific environments. However, because environmental factors like smoking and alcohol use are not only present in regions with a high incidence of oesophageal cancer, the disease aetiology likely reflects a complex interplay between a number of environmental and genetic factors (Hendricks & Parker, 2002). Environmental lifestyle factors shown to be implicated in OSCC in Malawi include firewood cooking, use of white maize flour and cigarette smoking (Mlombe et al., 2015). Other external factors considered have been diet, nutrients and genetics (Arnal et al., 2015) and drinking high-temperature beverages is also associated with increased risk of OC, possibly due to thermal injury to the oesophageal mucosa (Islami et al., 2009). An OSCC risk factor that was of high interest in the South African context is folate deficiency (Van Helden et al., 1987). It was found that populations with high incidence of OSCC correlated with low serum levels of vitamin A, E and folate. Folate metabolism will also be further discussed with its impact in cancer development with regards to alcohol further on in this review. A meta study found that multiple studies have reported tobacco smoking, tobacco chewing and alcohol consumption to be a risk factor for OSCC, across multiple African countries, including South Africa (Asombang et al., 2019).

Genetic factors may increase susceptibility to cancer because of the heterogeneity arising from polymorphisms within populations, this can determine whether individuals are at a greater or lesser risk of cancer. In one study it was found that certain single nucleotide polymorphisms (SNPs) in DNA mis-match repair enzymes (MMR) genes were associated with an increased risk of OSCC in mixed ancestry population but not in African populations (Vogelsang et al., 2012). A specific SNP in the CHEK2 gene involved in cell cycle regulation repair showed an increased risk of OSCC in African populations, however other variants found in Asian and European populations were not found to have significant associations with OSCC (Chen et al., 2019). This may indicate that genetic differences in populations play an important role as different populations have different gene variant risk factors associated with susceptibility to OSCC. Similar results were observed with a TRAK2 gene variant found to be associated with OSCC in mixed ancestry populations of South Africa, but not with black populations, providing further evidence that the role of genetic susceptibility to OSCC is population specific (Matejcic et al., 2019). In South Africa, genetic variants found to be associated with OSCC in the mixed ancestry population, included ADH1B (rs1229984), COX-2 (rs689466), CASP8 (rs1045485) and MGMT (rs12917), as well as ALDH2 (rs886205). ADH1B and ALDH2 is of particular interest as they are involved in the metabolism of alcohol (Bye et al., 2011). These genetic variants were also not associated with the Black population, indicating the difference in aetiology between populations, may be due to ancestry, differences in environmental exposures or a combination of these (Bye et al., 2011).

Gene-environment interactions may play a key role in susceptibility to OSCC as polymorphisms of xenobiotic enzymes can affect the efficacy of those genes and influence the individual’s susceptibility to disease (Matejcic & Parker, 2015). These enzymes involved in the detoxification of environmental and dietary carcinogens which have been found to be associated with OSCC in South African populations are the glutathione S-transferases (GSTs)- GSTP1 341T variant (Li et al., 2010), N-acetyltransferases (NATs)- NAT1 and NAT2 (Matejcic et al., 2015), cytochrome P450s (CYPs)- CYP3A5 (Dandara et al., 2005) and CYP2E1*6 (Li et al., 2005), sulfotransferases (SULTs)- SULT1A1*2/*2 and CYP3A5 (Dandara et al., 2006), UDP-glucuronosyltransferases (UGTs), and epoxide hydrolases (EHs), (Matejcic & Parker, 2015).

Studies in South African Black and Mixed Ancestry populations found environmental factors such as smoking and alcohol consumption associated with increased risk of OSCC and they found that different gene variant may increase susceptibility to OC (Dandara et al., 2006). It has been found that gene variants SULT1A1*2/*2 and CYP3A5 are associated with OSCC and conferred higher risk with smokers compared to non-smokers (Dandara et al., 2006). The study linking GSTP1 341T to OSCC also found that this variant increased risk for OSCC in tobacco smokers, alcohol consumers and those using wood or charcoal for cooking and heating, further evidence linking population genetics to environmental risk factors (Li et al., 2010). NAT1 and NAT2 are involved in the metabolism of compounds found in cigarette smoke and red meat, and variants of these genes have been found to be associated with an increased or decreased risk of OSCC in Black and Mixed ancestry populations that are smokers or frequent alcohol consumers and frequent red meat consumers (Matejcic et al., 2015).

In addition to environmental factors, population genetics are also likely to play a role. In a study conducted by Chang-claude *et al* (1997) in China, familial aggregation was studied to determine if the high incidence of OC was due to environmental factors or due to a hereditary component (Chang-claude et al., 1997). This study found an increased rate of OC among relatives, proving familial aggregation. However, a stronger association was found between parent-parent pairwise comparisons compared to parent-offspring and between siblings, suggesting that the environmental factors have a stronger effect on OSCC development compared to genetics.

In the United States, it was found that smoking was a key impactor on the OSCC burden in this population (Engel et al., 2003). This study also reported that increasing levels of alcohol consumption has an increased association with disease burden. Although low dietary fruit and vegetable intake is a risk factor, the contribution to risk is much lower than both smoking and alcohol consumption. These results confirmed other case-studies which showed that OSCC is associated with alcohol and smoking and the combination of these factors revealed an even greater risk (Asombang et al., 2019; Lagergren et al., 2000; Zambon et al., 2000).

Although some studies seem to indicate that the field is unclear on whether smoking and alcohol is directly involved (Arnal et al., 2015); there seems to be substantial evidence implicating smoking and alcohol in the development of OSCC. In some developing regions, including Southern Africa, these factors have also been shown to be implicated in OSCC incidence (Sewram et al., 2016). Given this evidence, the following sections will review the involvement of cigarette smoke and alcohol in OSCC in further detail.

1.2 Risk factor: Cigarette Smoke

Carcinogens found within tobacco products are the main cause for the many cancers associated with tobacco use (Hecht, 2003). Cigarette smoke contains around 4000 compounds, of which 200 are poisonous and 60 are carcinogenic (T. Wang *et al.*, 2017). Scientists have found that multiple fractions of cigarette condensate have the ability to cause frameshift mutations (Kier et al., 1974). Smoking has been widely studied in relation to cancer and the genotoxic potential of cigarettes has been investigated extensively. Smoke exposure can result in the formation of DNA adducts, point mutations and the induction of double-stranded DNA breaks (DeMarini, 2004; Hecht, 2003; Lee & Pausova, 2013; Sobus & Warren, 2014; T. Wang et al., 2017). Moreover, tobacco smoke has been associated with epigenetic changes as it has been found to induce changes in the activity of chromatin-modifying enzymes (T. Wang et al., 2017). These cigarette-induced genotoxic effects have been found in various tissue types including oesophagus, oral nasal region, pharynx/larynx area, lung, pancreas, myeoloid organs, bladder/ureter and uterine cervix (DeMarini, 2004). In this section we will go through the various effects of cigarette smoke on cells.

Multiple epidemiological studies have shown that cigarette smoking is a risk factor of OSCC. One study carried out in the Eastern Cape of South Africa, found that participants, regardless of gender, who smoke more than 14g of tobacco have a 4-times greater chance of developing OC (Sewram et al., 2016). Other population based studies have found that for various gastro-intestinal cancers, especially OSCC, there is poor survival associated with both former and current smokers (Mcmenamin et al., 2017). Additionally, it was found that heavy smoking was associated with shorter survival, compared to light or moderate smokers. A study conducted in Italy found that the number of daily cigarettes and long-duration smoking was associated with higher risks of OSCC and that people who start smoking below the age of 17 years was associated with a moderately greater risk compared with later ages among current smokers (Zambon et al., 2000). Their data also displayed a clear decrease of oesophageal cancer risk followed smoking cessation.

Despite epidemiological studies reporting cigarette smoking as a risk factor for OSCC, studies determining mutational signatures involved in OSCC have been less clear. Mutational signatures are a particular pattern of gene mutations linked to a mutagenic process. These mutational signatures can be used as a biomarker for risk factors of cancer. Aging and apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) activity are just two examples of common mutational signatures found in many different cancers, including OSCC. However, the smoking mutational signature has very rarely (if ever) been associated with OSCC. The smoking tobacco use mutational signature has not been found to be associated with OSCC in Chinese populations (Zhang et al., 2015) or the sub-Saharan population of Malawi (Liu et al., 2017). This study (Liu et al., 2017) carried out in an OSCC cohort in a Malawian population directly contradicts an epidemiological study which found OSCC to be five-times higher in people with a smoking history compared to non-smokers in a Malawian population (Mlombe et al., 2015). To investigate this discrepancy, it would be essential to investigate mutational signatures in cell culture.

1.2.1 Cigarette Smoke: Effects on Cells

Consistent with the epidemiological studies, *in vitro* tissue culture has shown treating cultured cells with cigarette smoke causes an increase in cancer-like properties. *In vitro* studies have shown that treating normal breast epithelial cells with cigarette smoke can transform them such that the emerging cell type has cancerous properties (Mohapatra et al., 2014). Cancer-like properties mentioned in this study include a greater capacity for anchorage-independent growth, higher colony forming ability and when placed in mouse models the transformed cells were shown to form solid tumours. Cigarette smoke condensate (CSC), which is smoke that has been extracted from cigarettes and solubilised in a solvent such as Dimethylsulfoxide (DMSO), has also been shown to affect the proliferation and doubling time of mouse embryonic stem cells (Assadollahi et al., 2019). Similar results have also been seen in OSCC cells where high concentrations of CSC increased cell proliferation (Li et al., 2009).

1.2.2 Cigarette Smoke: Effect of Carcinogens

Carcinogens found in cigarettes are characterised as either strong carcinogens or weak carcinogens and make up about 1-3mg per cigarette (Hecht, 2003). Strong carcinogens are found in smaller proportions in cigarettes but of those the most typically present are polycyclic aromatic hydrocarbons (PAHs) and nitrosamines, particularly the tobacco-specific nitrosamines 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and *N*′-nitrosonornicotine (NNN). NNN has been shown to cause oesophageal tumour growth in rats and is the most common nitrosamine found in cigarette smoke. However, it is noted that no single constituent of cigarette smoke causes cancer, rather the combination of carcinogens (Hecht, 2003). In addition to PAHs and NNNs, other carcinogens found in cigarette smoke are arsenic, chromium, formaldehyde and they are all known to cause DNA damage by causing double-stranded breaks (Lee & Pausova, 2013).

Commonly studied carcinogens found in cigarette smoke are polycyclic aromatic hydrocarbons (PAHs), such as benzo[a]pyrene (BaP) (Chen et al., 2019). A mediator of the effects of PAHs are the aryl hydrocarbon receptor (AHR). The AHR is ligand-activated to mediate cellular responses triggered by compounds in the environment and diet. The AHR is a transcription factor which binds ligands such as

PAHs and is responsible for many homeostatic functions such as cell proliferation, inflammation and cell motility and migration (Sobus & Warren, 2014). One of the important results of engaging the AHR system is the activation of the xenobiotic metabolic system which includes the cytochrome P450 genes. These CYP genes encode enzymes which are involved in the metabolism of endogenous and exogenous compounds, including drugs and pollutants (Ingelman-sundberg, 2002) and play an important role in converting PAHs into toxic compounds.

CYP1A1 is a cytochrome P450 gene known to be involved in bioactivating certain AHR ligands, such as benzo[a]pyrene (Chen et al., 2019). A study has been shown that when normal human bronchial/trachea epithelial (HBTE) cells are treated with CSC but the AHR is supressed there is decreased levels of CYP1A1 gene expression and decreased levels of CYP1A1 enzyme activity (Chen et al., 2019). Carcinogens such as PAHs and NNNs enter the body via the cigarette and most need to be metabolically activated. This paper found that BaP, one of the key pro-carcinogenic AHR ligands found in CSC, leads to carcinogenesis after being metabolised to become the carcinogenic compound (Chen et al., 2019). Most of the compounds which undergo these reactions are de-toxified and excreted, however, some are electrophilic and react with DNA to form the DNA adducts (Hecht, 2003).

The main mechanism through which carcinogens cause cancer are through the formation of DNA adducts (Hecht, 2003). DNA adducts are the products formed when cancer-inducing chemicals binds to segments of DNA. Usually, the DNA adducts are removed by the cell's DNA repair mechanisms and returned to its previous state but if the DNA adducts are missed mutations occur (Hecht, 2003). If these mutations arise in oncogenes or tumour-suppressor genes, this could result in altered key regulatory processes in the cell (like cell proliferation and apoptosis) eventually resulting in tumour formation. Further studies have shown links between CSC and the cytochrome P450 genes. In human oral cells it has been shown that CSC increases the expression of cytochrome P450 genes: CYP1A1, CYP1B1; and aldo-keto reductase genes: AKR1C1, AKR1C3, AKR1B10 in oral cells (Nagaraj et al., 2006). Carcinogenesis via this method is thought to occur via the formation of DNA adducts and production of reactive oxygen species (ROS).

The importance of ROS and inflammation has been observed in other studies as well on human monocytic cells (Rao *et al.*, 2016) as well as alveolar epithelial cells (Yadav *et al.*, 2016). These studies have found that CSC treatment and PAHs increases ROS production and this change is associated with an increase in CYP1A1 gene expression. Oxidative stress (ROS) and inflammation has been known to be important in tumour progression. It has been suggested cigarette smoke can cause oxidative damage because it contains free radicals such as nitric oxide and various levels of hydroquinones, semiquinones and quinones which can induce redox cycling (Hecht, 2003). Inflammation is thought to be involved in the early stages of neoplastic tissue development, where it encourages cell proliferation, survival and migration as well as encouraging invasion and metastasis (Coussens and Werb, 2002). Studies have shown that CSC increases the expression of various cytokines that promote inflammation (Parsanejad *et al.*, 2008).

In addition to a directly causing DNA mutation as previously discussed, it has been shown smoking is involved in cancer progression via epigenetic mechanisms (Lee & Pausova, 2013). Epigenetics is a gene expression regulatory mechanism which occurs via the chemical modification of DNA strands or histones. In non-small cell lung carcinoma, methylation of the promoter region of the tumour suppressor gene TSLC1/IGSF4 was observed to a greater extent in smokers and this hypermethylation was associated with years-smoked and packs per day (Kikuchi et al., 2005). In this study greater methylation was observed in males compared to females, which may be of interest as OSCC is more common in males compared to females (Kikuchi et al., 2005). Mechanisms of epigenetics include DNA methylation, histone modification, miRNA silencing (T. Wang et al., 2017).

CSC has been shown to affect all these mechanisms. DNA methylation is an essential type of epigenetic regulation, it is a highly active form of gene regulation during embryonic development and thereafter remains relatively constant. Environmental factors can affect methylation patterns and cigarette smoke is believed to be one of the strongest environmental modifiers of DNA methylation (Lee & Pausova, 2013). Studies on adenosquamous lung cancer (Vaz et al., 2017) and ESCC cells (Huang et al., 2011) have shown CSC causes stem cell like cancer properties via affecting methylation patters. Additionally, it has been shown that CSC may alter gene expression through modulating histone acetylation activity. In lung alveolar epithelial cells it has been shown that increasing concentrations of CSC increases histone acetyl transferase (HAT) activity (Moodie et al., 2005). CSC has also been shown affect epigenetic regulatory mechanisms via microRNAs (miRNAs). miRNAs are non-coding RNAs which interact with their complementary mRNAs to degrade transcripts or inhibit translation and thus regulate signalling network. In lung cancers it has been shown miRNAs are abnormally expressed, largely due to cigarette smoking (Xi et al., 2013). This study has shown that CSC exposure causes the downregulation of miR-487b, resulting in the upregulation of various signalling pathways and thereby causing increased proliferation, invasion and metastatic potential of lung cancer cells (Xi et al., 2013).

Knowing the pathways directly affected by CSC allows us better understanding of the mechanisms occurring and opens pathways for studying diagnostic biomarkers or possible therapeutic treatments.

1.3 Risk Factor: Alcohol

After tobacco smoking, alcohol is one of the most significant known causes of cancer (Boffetta & Hashibe, 2006). It has been reported that alcohol consumption is responsible for 3.6% of all cancers worldwide- 1.7% in women and 5.2% in men (Cao & Giovannucci, 2016). The association between alcohol and cancer has been found in many cancers including oral, pharynx, larynx, oesophagus, breast, colorectum, liver, pancreas and colorectal cancer (T. Wang et al., 2017).

1.3.1 Alcohol Consumption: Epidemiology

The literature reports that alcohol has been shown to be associated with OSCC and not with oesophageal adenocarcinoma (Boffetta & Hashibe, 2006; Lagergren et al., 2000; Wheeler & Reed, 2012). Alcohol as a risk factor for OSCC has also been shown in South African populations where a study reported that subjects who drink more than 53 g of ethanol a day have a 5-times greater chance of OC development than non-drinkers and an 8.5 times greater chance than people who smoke more than 14 g of tobacco (Sewram et al., 2016). In addition to being identified as a risk factor for developing OSCC, studies on gastro-intestinal (GI) cancers have found that there was decreased survival for patients who had ever drank alcohol, and this was observed especially in OSCC (Mcmenamin et al., 2017). These studies also reported that there is better prognosis for patients who previously consumed alcohol, compared to current drinkers (Mcmenamin et al., 2017).

In contrast, a study conducted in Malawi concluded that alcohol consumption was not found to be associated with OSCC (Mlombe et al., 2015). This was observed after adjusting for age, gender, socioeconomic status, cooking methods, and cigarette smoking. They suggested alcohol consumption may be found related to OSCC due to confounding factors (Mlombe et al., 2015). However, the paper conducted by Sewram *et al* (2016) which showed alcohol to be a risk factor for OSCC did account for confounding factors such as age and gender. This may also highlight the issue epidemiological studies face when determining alcohol consumption by self-reporting, in an environment where there may be concerns of stigmatization.

Case control studies have also found an inconsistent association between lung cancer and alcohol consumption. Multiple studies found there is a j-shaped association between alcohol consumption and lung cancer; but a positive association between certain alcoholic drinks and lung cancer, for example, evidence was found for an association between beer drinking and squamous cell carcinoma (Brenner et al., 2019). The authors of this paper concluded that the inconclusive results are due to possible confounding factors, inadequate statistical power caused by problems in case control studies heavily relying on questionnaires that had large variation in the way they determined alcohol intake among subjects. Although epidemiology shows a strong case can be made for the role of alcohol consumption in cancer, they cannot be used to assign cause and effect relationships (Zakhari, 2015). This is especially pertinent in alcohol- intake epidemiological studies as the data relies on self-reporting of alcohol consumption and issues related to combining studies started decades ago to current or recent studies on alcohol consumption (Zakhari, 2015). Thus, there is certainly scope to determine the biological effects of alcohol experimentally.

1.3.2 Alcohol: Impact on Cells

The issues regarding data collection for alcohol consumption studies increases the importance of understanding the biological underpinnings of how alcohol may cause carcinogenesis. The mechanisms of the carcinogenic effects of alcoholic drinks are not completely understood and it has been suggested they are different across different target organs (Boffetta & Hashibe, 2006). The impact of alcohol on carcinogenesis may in part be due to genetic factors and it is suspected that alcohol may assert the observed associations by impacting ethanol metabolism enzymes; folate metabolism; increased oestrogen concentration, relating to breast cancer; acting as a solvent for tobacco carcinogens and by the production of reactive oxygen species and nitrogen species (Boffetta & Hashibe, 2006).

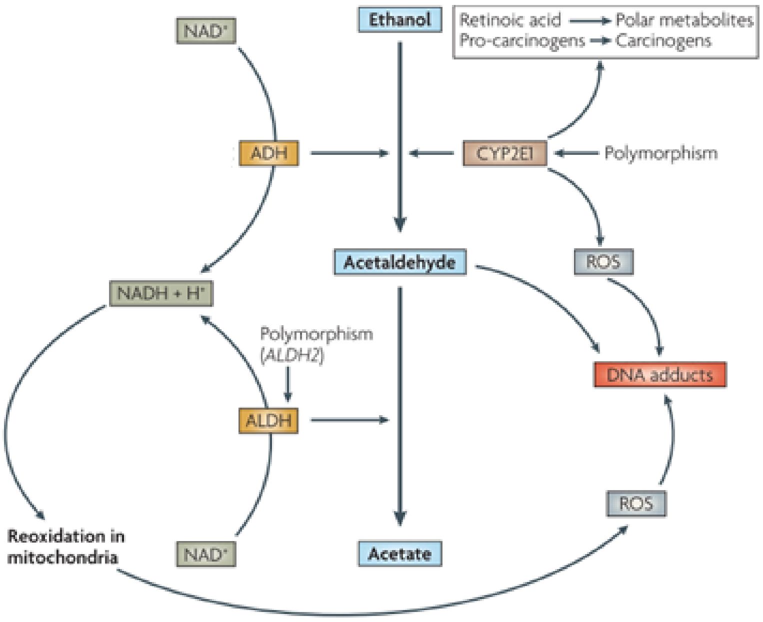

Figure 1.2: Chemical reaction displaying the conversion of ethanol into acetaldehyde and subsequently, acetaldehyde into a non-carcinogenic acetate. These reactions primarily occur via alcohol dehydrogenase (ADH) and acetaldehyde dehydrogenase (ALDH). A highly ative ALDH enzyme can lead to an accumulation of Acetaldehyde which results in multiple tumorigenic effects, notably the formation of DNA adducts. Ethanol metabolism may also occur via CYP2E1 which result in the formation of ROS which can cause DNA adducts to form. Schematic taken from (Seitz & Stickel, 2007).

The severity of the effects of alcohol is genetic, likely as a result of different variations of the genes which metabolise alcohol. The alcohol dehydrogenase (ADH) enzymes and aldehyde dehydrogenase (ALDH) enzymes, as seen in figure 1.2 above, are key enzymes- their functions being to oxidize ethanol into acetaldehyde and acetaldehyde to the non-toxic acetate, respectively. It is suggested people who carry ADH alleles with high enzymatic activity or ALDH alleles with low activity are at increased risk of alcohol-related cancers (Cao & Giovannucci, 2016). These common functional genetic variants are existing differences between individuals that would result in different acetaldehyde exposure (Boffetta & Hashibe, 2006). Thus, the same alcohol intake between individuals would result in differences in potential tumour formation. Importantly it has been reported that 40-50% of Asians have the ALDH2 variant with low activity, designated ALDH*2, and these people are at a higher risk of developing OSCC (Seitz & Stickel, 2007). People homozygous for this allele are unable to oxidize acetaldehyde and after ingesting even small amounts of alcohol symptoms of nausea, vomiting and facial flushing are induced, termed flushing syndrome. Individuals heterozygous for the allele have a marked decrease in ALDH activity, can metabolise small amounts of alcohol and can thus tolerate the effects of alcohol despite the flush reaction. These individuals are also at a high risk of developing OSCC (Seitz & Stickel, 2007). Interestingly cytochrome P-450 2E1 (CYP2E1), another enzyme induced by ethanol which oxidises ethanol into acetaldehyde and activates tobacco procarcinogens, also has polymorphisms but these have not been found to be associated with cancer (Boffetta & Hashibe, 2006; Seitz & Stickel, 2007).

1.3.3 Alcohol: Role of Acetaldehyde

Acetaldehyde is the primary endogenous ethanol metabolite and it is understood to be a cancer-inducing compound (Cao & Giovannucci, 2016; Seitz & Stickel, 2007). Acetaldehyde induces cancer development through multiple mechanisms, such as the formation of DNA adducts, induction of DNA damage and DNA replication interference (Cao & Giovannucci, 2016; Lopes et al., 2012). Acetaldehyde is also capable of inducing deletions, chromosomal aberrations, and generating protein adducts (Scoccianti et al., 2014). Acetaldehyde has also been shown to damage and thus increase the proliferation of hepatocytes (Boffetta & Hashibe, 2006). Mechanisms involving acetaldehyde metabolism has involved the production of inflammatory cytokines, immunologic and metabolic pathways derangements and signal transduction changes (Zakhari, 2015). Animal studies have shown that acetaldehyde inhalation causes tumours of the respiratory tract, particularly in the nasal mucosa of rats and laryngeal carcinomas in hamsters (Boffetta & Hashibe, 2006). Further studies have shown evidence for the formation of DNA adducts by acetaldehyde in human cells *in vitro* and in rats who had lifetime exposure to ethanol. *In vivo*, a study has shown alcohol users had 7-times higher the mean number of acetaldehyde–DNA adducts in lymphocytes, compared to controls (Boffetta & Hashibe, 2006).

Acetaldehyde is thought to play a large role in alcohol induced liver cancer (Zakhari, 2015) and breast cancer, where an association has been found with alcohol consumption in a linear dose–response manner (Scoccianti et al., 2014). *In vitro* cell line studies suggest alcohol is involved in breast carcinogenesis through an association with increased hormonal receptor levels, increased cell proliferation, a direct stimulatory effect, DNA adduct formation, increase cyclic adenosine monophosphate, change in potassium channels and modulation of gene expression (Cao & Giovannucci, 2016). Damage caused by acetaldehyde is proposed to underlie the association between alcohol and breast cancer are (Y. Wang *et al.*, 2017). It has been suggested that breast cells have a limited ability to detoxify acetaldehyde and therefore it will persist in these tissues and have a greater carcinogenic effect (Y. Wang *et al.*, 2017). These effects of acetaldehyde are likely to play a role in OSCC as a study on oesophageal cancer patient biopsies had shown increased ADH IV activity and normal aldehyde dehydrogenase activity, which could lead to an increased accumulation of acetaldehyde and increase carcinogenesis (Jelski et al., 2009).

1.3.4 Alcohol: Role of Reactive Oxygen Species

Acetaldehyde may also play a role in liver cancer by inducing ROS formation (Zakhari, 2015) . At high alcohol concentration levels, alcohol is oxidised by CYP2E1 and acetaldehyde is produced. Thereafter it is metabolised mainly by mitochondrial aldehyde dehydrogenase (ALDH2) which leads to increased formation of highly reactive oxygen species (ROS) such as hydrogen peroxide, hydroxyethyl, peroxynitrite and hydroxyl radicals (Zakhari, 2015). It has also been shown in breast cancer that alcohol can be metabolised by cytochrome P4502E1 (CYP2E1), resulting in the oxidation of NADPH and formation of ROS (Y. Wang et al., 2017).

Another mechanism alcohol may contribute to carcinogenesis with ROS production is via epigenetic regulation (Y. Wang et al., 2017). Similar to cigarette smoke exposure, alcohol has also been shown to induce carcinogenesis through epigenetic changes such as DNA methylation, histone acetylation, and histone methylation (Cao & Giovannucci, 2016; Y. Wang et al., 2017). Moreover, ROS has been found to stimulate hypermethylation of the promoter region of the tumour suppressor, E-cadherin, a regulator of the epithelial-to-mesenchymal transition (Zakhari, 2015).

ROS is a large contributor to carcinogenesis, mainly through promoting DNA damage and altering cellular signalling pathways. Additionally, oxidative stress contributes to metabolic dysregulation by interfering with repair mechanisms. Oxidative stress caused by ROS and nitrogen species can also cause DNA damage and react with lipids to produce compounds that react with DNA to form exocyclic DNA adducts and reactive aldehydes; a mechanism of interest in liver carcinogenesis (Boffetta & Hashibe, 2006).

1.3.5 Alcohol: Role of Folate Metabolism

Another mechanism whereby alcohol may promote cancer is via an interaction between alcohol and folate metabolism. Folate, otherwise known as vitamin B, is involved in cell growth, DNA synthesis, repair and methylation (Boffetta & Hashibe, 2006). The link between alcohol consumption and folate metabolism is a consequence of alcohol being a folate antagonist which affects folate bioavailability (Álvarez-Avellón et al., 2017). Folate deficiency may contribute to cancer development as it can lead to misincorporation of uracil, chromosomal instability, and DNA hypomethylation, this can result in altered gene expression and thus cancer development (Scoccianti et al., 2014). Additionally, a function of methylenetetrahydrofolate reductase (MTHFR) enzyme involved in folate metabolism is to synthesize the major carbon donor in the remethylation of homocysteine to methionine, and this provides a link between epigenetic deregulation and the initiation and progression of breast cancer among drinkers (Scoccianti et al., 2014).

In support of this, numerous cohort studies have observed that high blood folate levels may play a protective role on breast cancer risk related to alcohol consumption (Scoccianti et al., 2014). Various studies have also reported that dietary folate may curb breast cancer risk by affecting DNA methylation during the early stages of mammary gland development (Scoccianti et al., 2014). In addition to breast cancer, an associated risk has been found in colorectal cancer (Boffetta & Hashibe, 2006) and the link between alcohol, MTHFR and folate metabolism has been observed in lung cancer (Álvarez-Avellón et al., 2017). A case-control study has observed that consumption of 0.1–9.9 g/day of alcohol combined with a high consumption of vegetables and fruits decreases lung cancer risk, while consumption of alcohol at 10–29.9 g/day increases lung cancer risk (Álvarez-Avellón et al., 2017). In OSCC, a study has revealed an inverse association between dietary folate intake and the risk of death after esophagectomy and thus concluded that folate can have beneficial effects on prognosis but this may also suggest alcohol can have detrimental effects via its interaction with folate metabolism (Lu et al., 2011).

Literature reports a wide range of effects of alcohol on cellular functioning and all these effects may lead to carcinogenesis. Out of all the potential mechanisms discussed here, the direct carcinogenic effect of acetaldehyde seems to be the most plausible reason for the observed associated between alcohol consumption and cancer.

1.4 Combination of Alcohol and Cigarette Smoke Condensate

Alcohol consumption and tobacco smoking when combined, work synergistically to contribute to carcinogenesis- this interaction has become a model of how two environmental factors may contribute to carcinogenesis (Boffetta & Hashibe, 2006). Additionally, there is evidence for synergistic interactions with alcohol and other chemical compounds, such as 2-Amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP), a compound found in cooked red meat, which results in increased carcinogenic effects (Malik et al., 2018). Various epidemiological papers have reported that the combination of cigarette smoking and alcohol consumption increase risk for OSCC, however the underlying biological mechanisms need to be understood. These mechanisms that will be discussed are pathways which both alcohol and smoking affects such as cytochrome activity (Cao & Giovannucci, 2016) or mechanisms where one factor may enhance the carcinogenicity of the other factor.

Various epidemiological papers have reported an increased risk of OSCC with the combined consumption of tobacco and alcohol. OSCC has been shown to be strongly associated with tobacco use and only moderately with alcohol, while the combined use of tobacco and alcohol resulted in a stronger increased risk for OSCC (Lagergren et al., 2000). Another case-controlled study on OSCC patients in Italy also looked at smoking and alcohol consumption, individually and in combination, and found long-term smoking was associated with a greater risk of developing OSCC and alcohol intake increased oesophageal cancer risk drastically with increasing amounts of alcohol consumption (Zambon et al., 2000). Unlike with smoking there was no trend for duration of alcohol drinking or age at start (Zambon et al., 2000). In combination, the risk of oesophageal cancer increased with the increase of alcohol consumption in each category of smoking habit, including never-smokers and ex-smokers. Overall, these authors found that the links between OSCC and alcohol consumption are stronger compared with smoking based on exposure intensity but is not impacted by time variables the way smoking is. The biological mechanisms underlying the trends seen in case-control studies need to be investigated to determine if we can mitigate their effects. It is possible the increased risk of cancer associated with a combination of smoking and alcohol consumption reflects a biological interaction between the metabolites in cigarette smoke and alcohol (Brenner et al., 2019).

The synergistic effects of alcohol and smoking are observed in cancers of the upper aerodigestive tract, areas which come in contact with alcohol and cigarette smoke (Cao & Giovannucci, 2016). It is possible that alcohol and smoking can have different carcinogenic mechanisms, but they may interact to increase risk. For example, alcohol may increase the permeability of tissues thereby increasing their susceptibility to tobacco-specific carcinogens (Cao & Giovannucci, 2016). Combined ethanol and cigarette smoking may be particularly harmful as ethanol is a solvent which can render cell membranes more permeable to carcinogens (Lopes et al., 2012) and allow for these carcinogens to enter more easily through the mucosal lining of aerodigestive track (Boffetta & Hashibe, 2006). Similarly, it has also been postulated that the combination of cigarette smoking and alcohol work together to induce colon cancer because the ethanol solubilises liposoluble cigarette smoke components which can be carcinogenic (Sobus & Warren, 2014).

Alternatively, alcohol and cigarette smoking, when used in combination may increase the risk of cancer because the tissues are exposed to an increased load of a common carcinogen, such as acetaldehyde. It has been found that smokers have twice as much acetaldehyde in their saliva as non-smokers, even if they consume the same amounts of alcohol. Cigarette smoke has considerable amounts of acetaldehyde that dissolves into saliva during smoking and it has been proposed that smoking may cause an increase of acetaldehyde produced from ethanol by oral yeasts and bacteria (Cao & Giovannucci, 2016). Thus, the combination of smoking with alcohol consumption may increase the detrimental effects of either factor due to an increase in acetaldehyde.

Another carcinogenic mechanism affected by both alcohol and smoking is via the cytochrome p450 pathway. The induction of cytochrome P450-2E1 (CYP2E1) in the mucosa of the upper aerodigestive tract may increase the activation of procarcinogens in tobacco, such as PAHs and various nitrosamines (Cao & Giovannucci, 2016). As discussed previously, both alcohol and smoking affect CYP2E1 enzyme activity suggesting combined consumption would result in an increased cumulative effect. Studies have observed a positive interaction between high pack-years of smoking and elevated levels of drinking , where there has been an increased risk of lung cancer with people in the highest alcohol and tobacco consumption groups (Brenner et al., 2019). These authors concluded that the interaction between alcohol and tobacco smoke may increase carcinogenicity by inducing the activity of cytochrome P-450 enzymes (Brenner et al., 2019).

The increase in ROS production that is associated with alcohol consumption may increase the risk of tumour formation when used in combination with smoking, as has been observed for other chemical combinations (Malik et al., 2018). In these studies, the interaction of ethanol and PhIP, a compound found in cooked red meat, lead to a greater increase in CYP2E1 expression compared to the factors alone and resulted in ROS generation (Malik et al., 2018). These results provided evidence for a mechanism of how combined consumption of ethanol and PhIP can trigger genetic damage, which suggests a possible mechanism which may underpin how the combined synergistic effects of cigarette smoke and alcohol consumption can trigger genetic damage and maintain the tumour microenvironment.

In addition, both alcohol and smoking have been shown to affect several matrix metalloproteinases (MMPs). Thus, the combination of alcohol and smoking may have a cumulative effect that results in an increased risk for the development of tumours via the MMP enzymes. Studies have found that when exposed to CSC malignant oral cells have increased production and enzymatic activity of several MMPs, which they conclude is a manner in which CSC increases the invasiveness of metastasizing oral squamous cell carcinoma cell lines. (Allam et al., 2011). This is possible as MMPs play a role in degrading the extracellular matrix of cells allowing them to dissociate and invade other tissues (Y. Wang *et al.*, 2017). Similar effects in breast cancer cells have been observed where alcohol increases the production of MMPs (Y. Wang *et al.*, 2017). Future work still needs to be done to determine if the combined treatment of CSC and alcohol may affect the enzymatic activity of MMPs.

Another mechanism where the combination of drinking alcohol and smoking may cause increased carcinogenesis is via the cytotoxicity of ethanol. It has been shown ethanol increases cytotoxicity and may increase the amount of cell divisions in stem cells underlying at risk tissues (López-lázaro, 2016). These authors postulate that dividing cells are at an increased risk of mutation because as the cell divides, the DNA unwinds for DNA replication and the nuclear envelope disappears during mitosis. This may then facilitate the interaction of DNA-damaging agents, such as carcinogens present in CSC, with the DNA of the cell. This mechanism would provide a link between the local cytotoxic activity of ethanol and the synergistic effect of alcohol and tobacco consumption pertaining to carcinogenesis (López-lázaro, 2016).

1.5 Conclusion

After reviewing the literature there is evidence of a role for cigarette smoke and alcohol in cancer initiation and development. Epidemiology has shown evidence for cancer association with both cigarette smoke (Sewram et al., 2016) and alcohol (Boffetta & Hashibe, 2006) and when the factors are combined there seems to be synergistic effects which increase the risk of OSCC (Zambon et al., 2000). Cigarette smoke and alcohol have also been shown to have multiple effects on cellular pathways that may induce tumour formation- the formation of DNA adducts and ROS resulting from smoking (Hecht, 2003) and alcohol use (Boffetta & Hashibe, 2006; Cao & Giovannucci, 2016) are important mechanisms. However, there are questions surrounding the role of cigarette smoking as the cigarette smoking mutational signature has not been identified in populations with OSCC (Zhang et al., 2015). Given the information presented above we can identify gaps in our understanding regarding the role of alcohol and smoking in the development of OSCC, specifically whether smoking exposure results in mutations typically associated with smoking, and the involvement of various pathways inside the cell which may mediate the tumorigenesis associated with smoking and alcohol in oesophageal cancer. In this project we will only consider the latter question.

1.6 Aims and Objectives of the Study

The objective of this study was to investigate cigarette smoking and alcohol as risk factors in causing OSCC. This study investigated the molecular effects of these factors in cultured cells. Cultured OSCC cells were exposed to smoking via cigarette smoke condensate and alcohol, individually and in combination. The aims of the study were to:

1. Determine the conditions for treating oesophageal cancer cells with ethanol and cigarette smoke.
2. Determine the impact of ethanol and cigarette smoke on the expression pattern of genes in oesophageal cancer cells.

Chapter 2: Results

2.1 Optimisation of CSC and EtOH Treatment

2.1.1 Introduction

To investigate the role of CSC and alcohol in the development of OSCC in cell culture, cells were treated with sub-lethal concentrations of CSC and EtOH. To determine the concentrations of CSC and EtOH that cause cell death in OSCC cells, we measured cell proliferation as an indicator of the cell status. Cell proliferation was measured using the MTT [3-(4,5-dimethylthiazol-2yl-)-2,5-diphenyl tetrazolium bromide] cell proliferation assay. In addition, cell viability was measured separately using a trypan blue assay, measuring cell viability by counting live vs dead cells, over a longer time period.

2.1.2 MTT Assay: Optimisation of CSC and EtOH Treatment Condition

The first objective on this study was to determine the effects of a range of CSC and EtOH concentrations on cell proliferation as a proxy for cell number. This was accomplished by using a MTT [3-(4,5-dimethylthiazol-2yl-)-2,5-diphenyl tetrazolium bromide] cell proliferation assay. A variety of OSCC cells were treated with a range of CSC (0-80 µg/ml) and EtOH (0-50 mM). The CSC concentrations were selected based on literature reports which identified physiologically relevant concentrations (Allam et al., 2011; Assadollahi et al., 2019; Rao et al., 2016). The range of EtOH was selected based on the literature which reported 10-100 mmol/l as a physiologically relevant alcohol concentration range and 25 mmol/l EtOH to be close to a 0.08% blood alcohol level (BAL) (Dolganiuc & Szado, 2009). Cells were allowed to incubate for one day after treatment and after observing that this did not cause cell death, incubation was increased to three days. This experiment was carried out in four OSCC cell lines: WHCO1, WHCO6, KYSE150 and KYSE30's. To determine cell proliferation, the absorbance of MTT conversion to formazan crystals was measured in treated and non-treated cells. The average absorbance values of the treated cells were then normalised against the non-treated cell lines to determine the impact on cell proliferation.

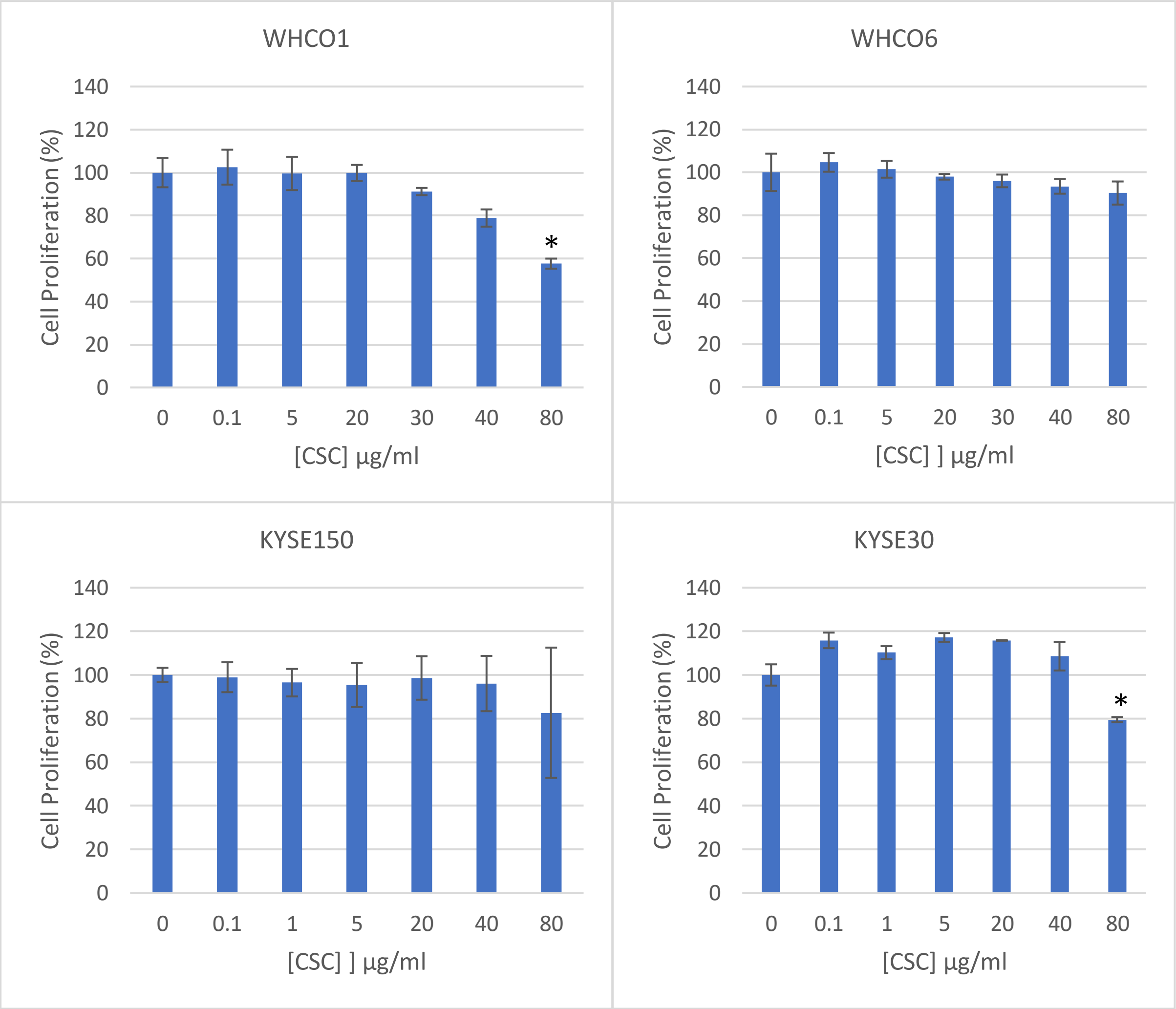

Figure 2.3: Graphs showing the effect of CSC on OSSC cells (WHCO1, WHCO6, KYSE150 and KYSE30) individually. An MTT assay was carried out to determine cell proliferation. The cells were treated with a range of CSC (0-80μg/ml). Cells were allowed to incubate for 3 days, and cell proliferation was measured as an average of three technical replicates. The error bars represent the standard deviation and statistical significance is indicated by an asterix (*) where * $p<0.05$. These results were confirmed in separate biological triplicates.

As seen in figure 2.1 above, CSC at 80 μg/ml caused a significant decrease in WHCO1 and KYSE30 cell lines ($P < 0.05$). In the KYSE150 cell line there was no significant decrease in cell growth when the cells are treated with 80 μg/ml CSC, although this may be due to the larger error bars observed in the sample. In WHCO6, a slight decrease in cell growth was also observed at 80 μg/ml CSC, however it was not significant, probably due to the slightly larger error bars apparent in the untreated sample. Therefore, from this data it can be concluded that treatment with CSC at 40 μg/ml is non-lethal after three days, whereas 80 μg/ml CSC causes a decrease in cell proliferation. A consistent, small increase in cell proliferation can be seen in the KYSE30 cell lines with CSC treatment, indicating CSC may cause cell growth.

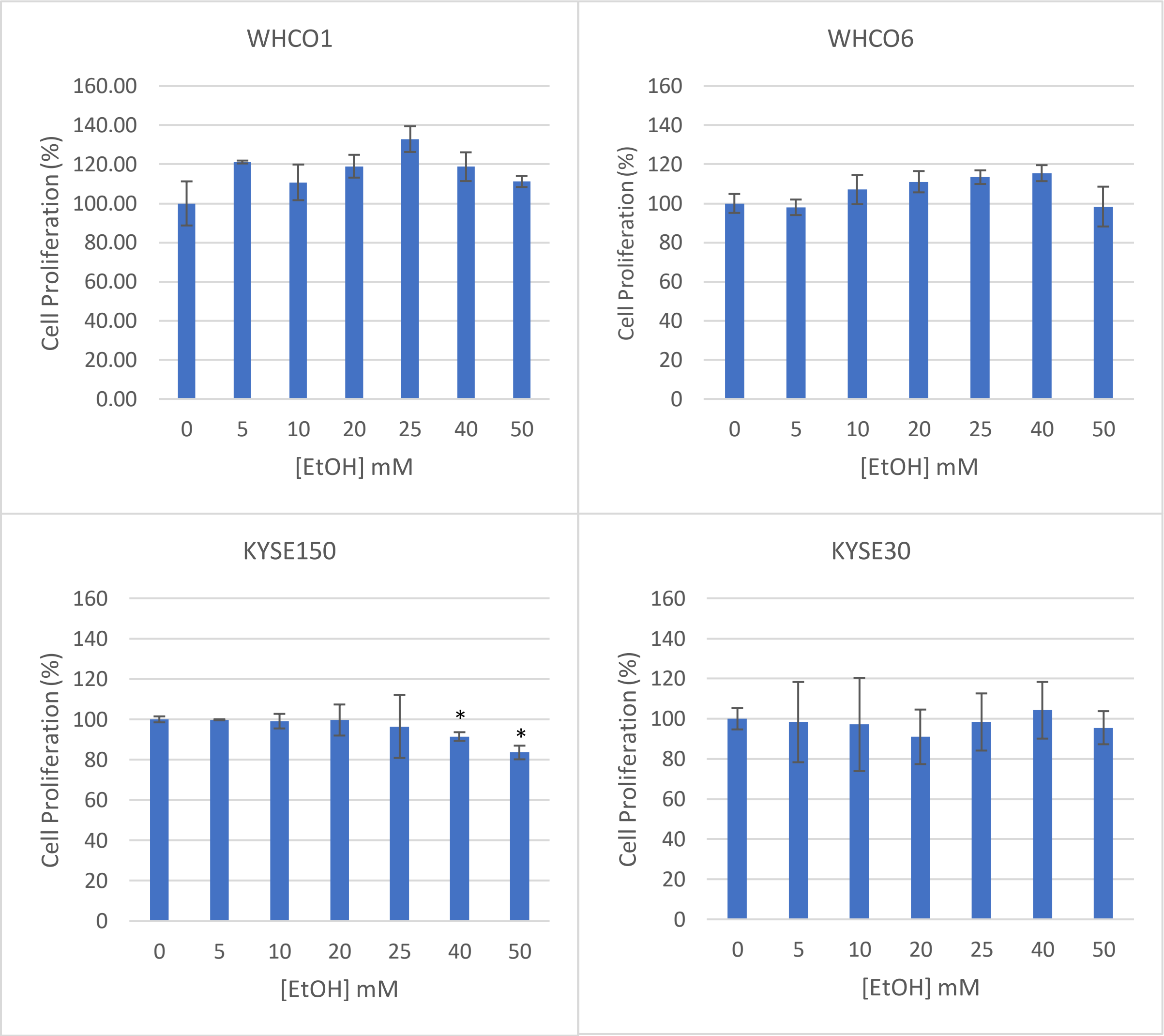

Figure 2.2: Graphs showing the effect of EtOH on OSSC cells (WHCO1, WHCO6, KYSE150 and KYSE30) individually. An MTT assay was carried out to determine cell proliferation. The cells were treated with a range of EtOH (0-50mM). Cells were allowed to incubate for 3 days, and cell proliferation was measured as an average of three technical replicates. The error bars represent the standard deviation and statistical significance is indicated by an asterix (*) where * $p<0.05$. These results were confirmed in separate biological triplicates.

As seen in figure 2.2, there was no decrease in cell proliferation observed in response to EtOH in cell lines WHCO1, WHCO6 and KYSE30. However, there was a significant ($p<0.05$) decrease in cell proliferation observed in KYSE150 cells in response to 40 mM and 50 mM EtOH, showing a 8.5% and 16.4% reduction in cell proliferation, respectively.

Combination treatments were investigated to determine if combining alcohol and cigarette smoke may have altered the cytotoxic profile. Cells were treated with a range of CSC (0-80 µg/ml) in the presence or absence of 25 mM and 50 mM EtOH. Cells were then treated with a range of EtOH (0-50 mM EtOH) in the presence or absence of 40 µg/ml or 80 µg/ml CSC.

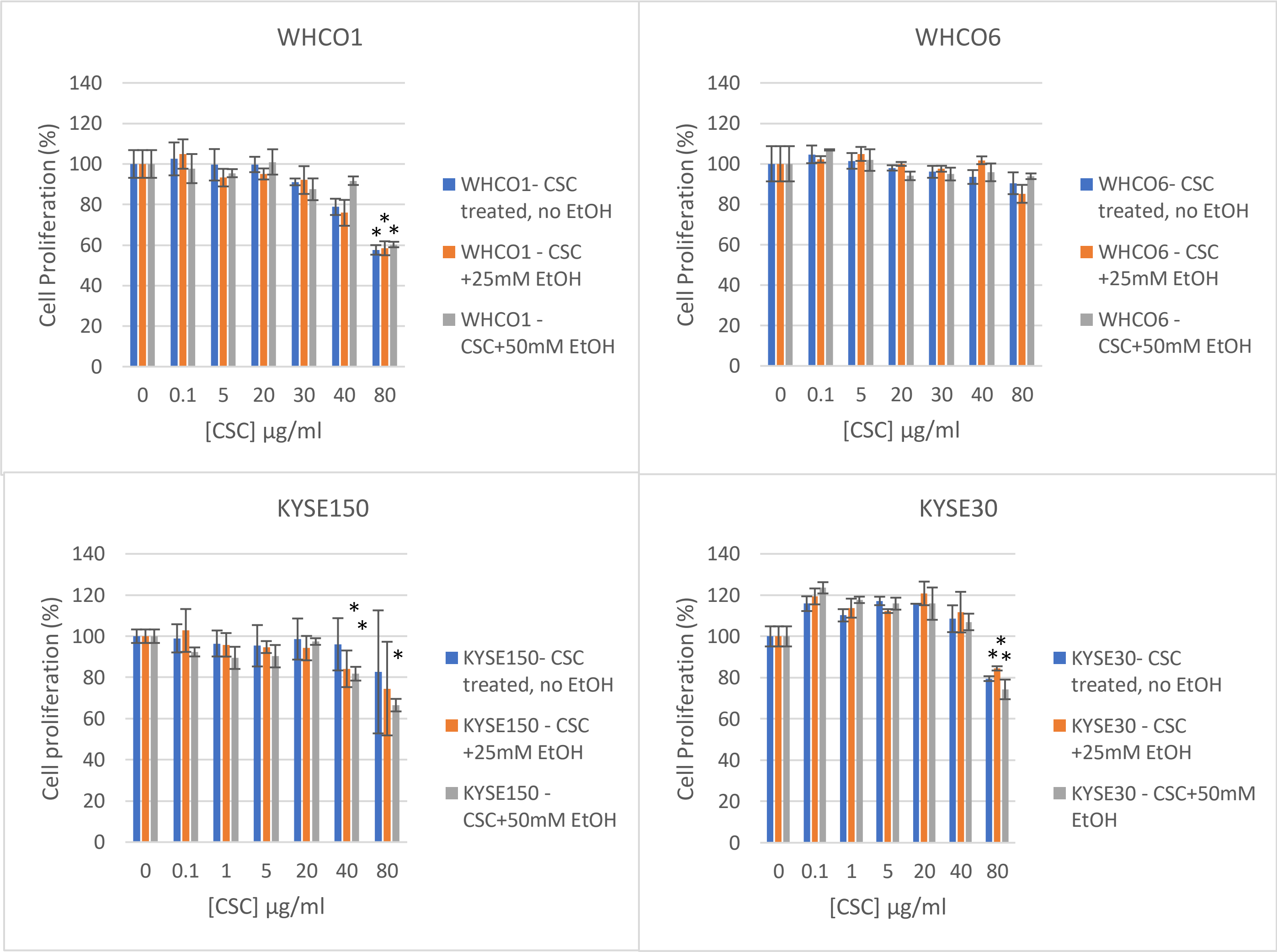

Figure 4.3: Graphs showing the effect of CSC on OSSC cells (WHCO1, WHCO6, KYSE150 and KYSE30) proliferation in the presence and absence of EtOH. An MTT assay was carried out to determine cell number. The cells were treated with a range of CSC concentrations (0-80 µg/ml) in the absence of EtOH and in presence of EtOH (25 mM or 50 mM). Cells were allowed to incubate for 3 days, and absorbance was measured at OD 595nm. The average of three technical replicates was used to determine relative cell proliferation (%). The error bars represent the standard deviation and statistical significance is indicated by an asterix (*) where * $p<0.05$. These results were confirmed in separate biological triplicates.

In WHCO1 and KYSE30 cells, a decrease in cell number was observed with combined treatments of CSC (80 µg/ml) and 25 mM and 50 mM EtOH, as seen in figure 2.3. In WHCO6 cell line there was no significant decrease in cell number observed in any of the treatment combinations. This is likely due to the large standard deviations seen in the untreated samples, despite this, a slight trend of decreasing cell number can be seen in the graph with increasing concentrations of CSC. In KYSE150, there was an observed significant decrease in cell number ($P < 0.05$) when cells were treated with 80 µg/ml CSC and 50 mM EtOH as well as 40 µg/ml CSC with 25 mM EtOH and 50 mM EtOH. Thus 40 µg/ml CSC was selected as a sub-lethal concentration of CSC to be validated by the trypan blue assay and in further downstream experiments. The decreased proliferation observed in WHCO1, KYSE150 and KYSE30 at 80 ug CSC µg /ml probably reflects the effect of CSC only since a similar result was observed with CSC treatment in the absence of EtOH.

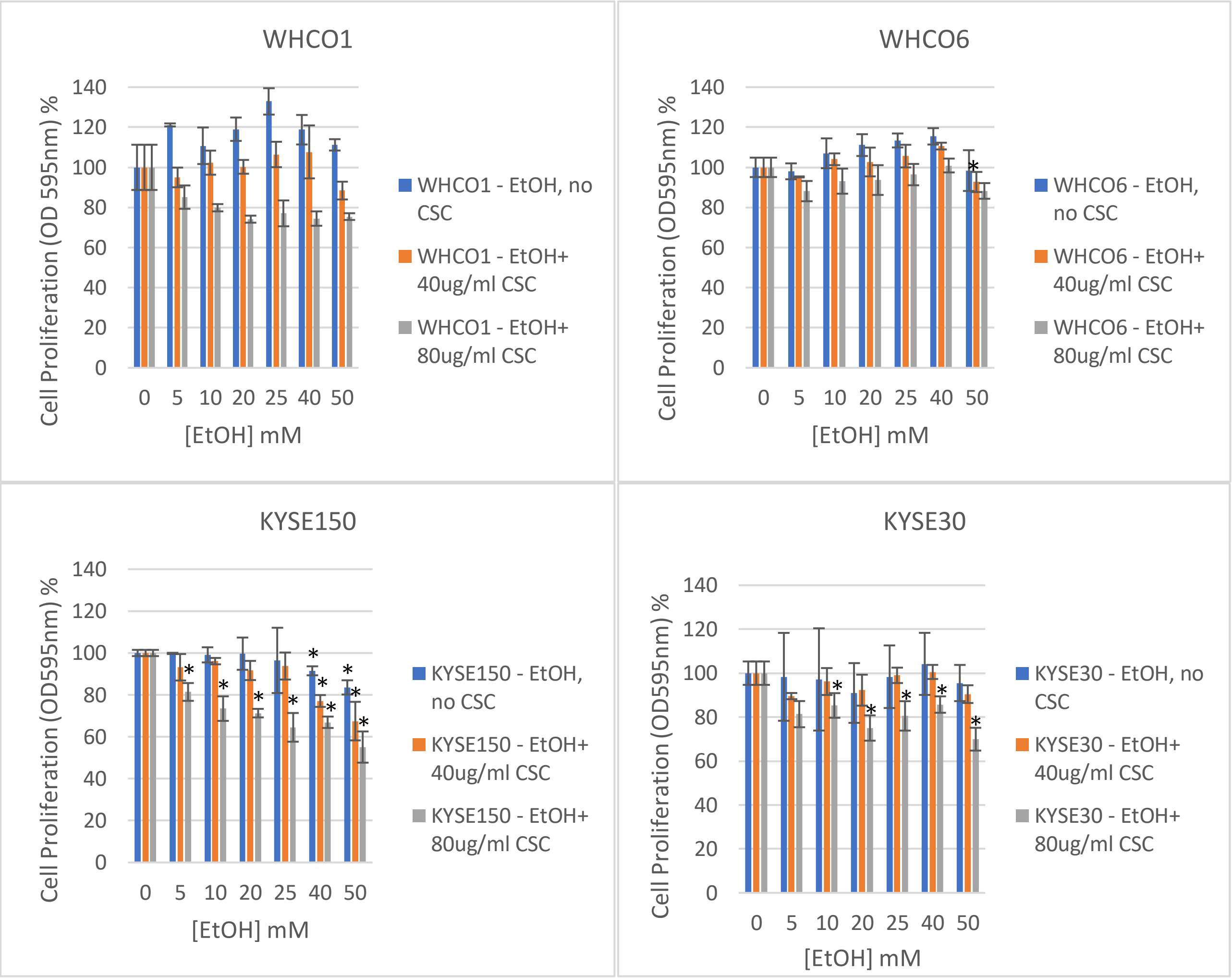

Figure 2.4: Graphs showing the effect of EtOH on OSSC cells (WHCO1, WHCO6, KYSE150 and KYSE30) proliferation in the presence and absence of CSC. An MTT was carried out to determine cell proliferation. The cells were treated with a range of EtOH (0-50 mM) in the absence of CSC and in presence of CSC (40 µg/ml or 80 µg/ml). Cells were allowed to incubate for 3 days, and the absorbance was measured at OD595 nm. The average of three technical replicates was used to determine relative cell proliferation (%). The error bars represent the standard deviation and statistical significance is indicated by an asterix (*) where * p<0.05. These results were confirmed in separate biological triplicates.

As seen in figure 2.4, WHCO1 cells treated with a range of EtOH, a decrease in cell number was observed in the presence of 80 µg/ml CSC, across all concentrations of EtOH (5-50 mM). However, this was not significant ($p > 0.05$), possible due to the slightly larger error bars seen in the untreated sample. In the WHCO6 cell line, a significant decrease in cell number ($p < 0.05$) was observed in the cells treated with 50 mM EtOH and 80 µg/ml CSC. In KYSE150 cells, a significant decrease in cell number was observed in the presence of 80 µg/ml CSC and EtOH (10-50 mM). Decreased cell proliferation was also observed in the cells treated with 40 mM and 50 mM EtOH in the absence of EtOH and in the presence of 40 µg/ml CSC. Thus, it can be seen that 50 mM EtOH does seem to cause a small decrease in proliferation in KYSE150 cells, and this effect is greater when treated together with CSC at 80 ug/ml. However, these results were not observed in the other cell lines. In KYSE30, the cells treated with 10 mM-50 mM EtOH and 80 µg/ml CSC showed significantly reduced cell proliferation ($p < 0.05$). Confirming the results observed in figure 2.3- 80 µg/ml CSC and EtOH cause decreased cell proliferation. Hence, due to these results and the information presented by Dolganiuc & Szado (2009), concentrations of 50 mM EtOH was used in the trypan blue assay and downstream experiments.

2.1.3 Trypan Blue Assay: Effect of CSC and EtOH Treatment on Cell Viability

The concentrations of CSC and EtOH found to be sub-lethal in OSCC cells using the MTT assay were further investigated using the trypan blue (TB) reagent to ensure they did not cause a decrease in cell viability. In a TB assay, cell viability is determined by differentiating between living cells and dead cells in culture. This is done by mixing an aliquot of cells with trypan blue dye and visualising the cells under a microscope. Living cells will be seen as clear, while nonviable cells will have a blue cytoplasm. This occurs as nonviable cells have compromised plasma membranes which causes cell uptake of the dye, whereas viable cells exclude the dye. Cell viability is then measured as a percentage of average living cells as a proportion of average total cells (living cells and dead cells) using a haemocytometer. This assay was carried out in three cell lines: KYSE150, WHCO1 and WHCO6. KYSE30 was excluded from this analysis because the KYSE30 cell line is more dependent on cell density for healthy growth and thus, the low cell plating number required for measuring viability until day 10 caused the cell line to die before being able to carry out the experiment.

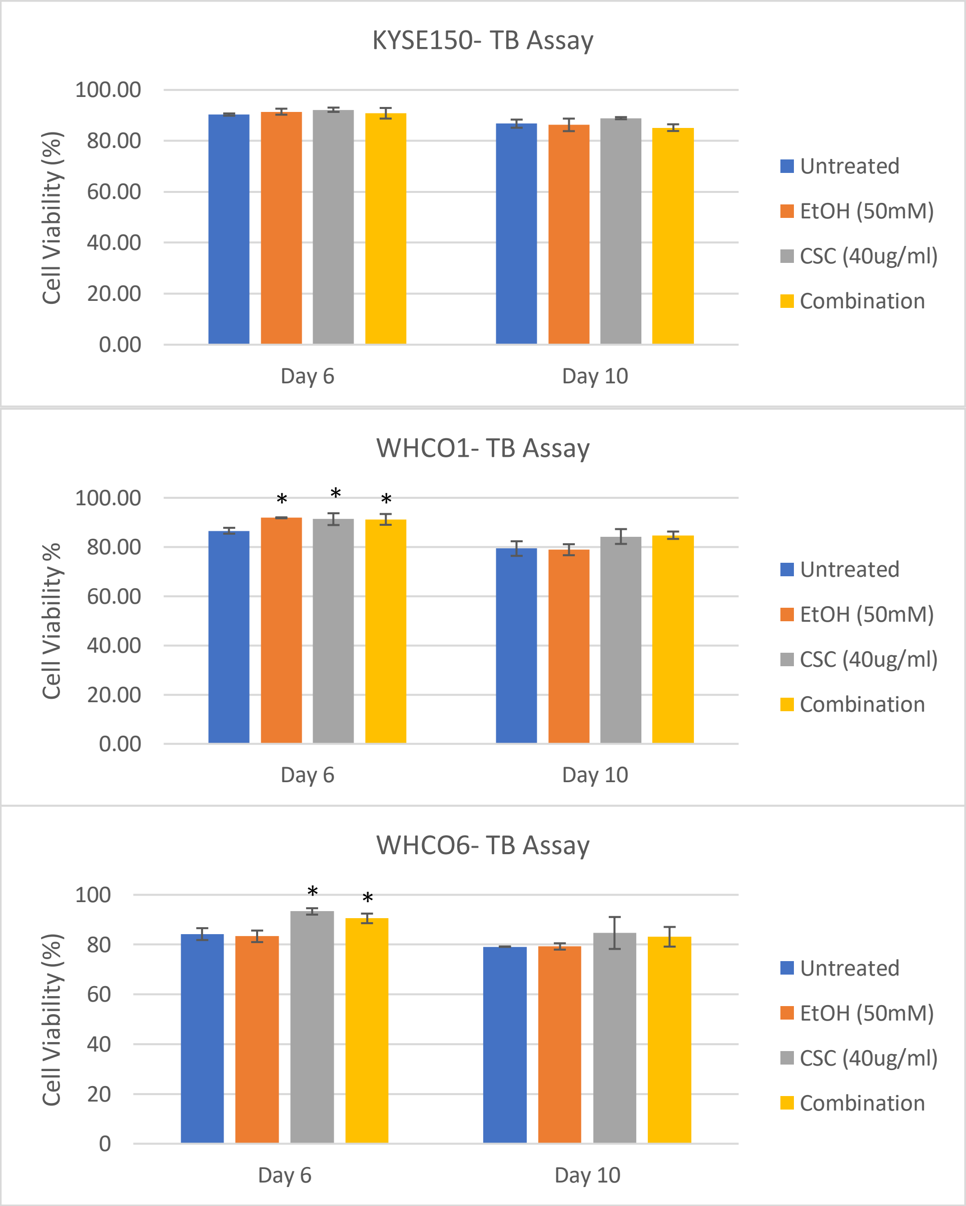

Figure 2.5: Graph showing cell viability in response to CSC, EtOH and the combination of CSC(40µg/ml) and EtOH(50mM). Living cells are plotted as a percentage of total cells in response to treatment. Cells were allowed to incubate for 6 and 10 days and living and dead cells were counted using trypan blue reagent and a haemocytometer. This was done in duplicate in three cell lines (KYSE150, WHCO6, WHCO1). The error bars represent the standard deviation between three technical replicates. Significant differences were determined by comparing untreated to treated cells. * $p<0.05$.

The trypan blue assay confirmed that in three cell lines, KYSE150, WHCO1 and WHCO6, there was no significant reduction in cell viability for longer CSC and EtOH treatment times. For the trypan blue assay, cells were treated with 40 µg/ml CSC, 50 mM EtOH and a combination of these concentrations. This was done in triplicate. Untreated cells were incubated with 0.2% DMSO and allowed to incubate for 6 and 10 days after treatment. Media was changed every three days with treated media. In the KYSE150 cell line the cell viability stayed consistent for all treatment conditions for 6 days and 10 days. In the WHCO1 cell line, the cell viability remained constant for the treatments after 10 day incubation, however, there was a small significant increase (p<0.05) in cell viability for the treated cells compared to the untreated cells of about 10% after 6 days incubation. In the WHCO6 cell line, cell viability remained constant for all treatments except cells treated with 40 µg/ml CSC and a combination of 40 µg/ml CSC and 50mM EtOH for 6 days, which showed a small significant (p<0.05) increase of around 5% in cell viability. These results confirmed that there was no resultant cell death after treatment, as observed in the MTT assay results.

Thus, it was concluded that these concentrations of 40µg/ml CSC and 50mM EtOH individually and in combination are sub-lethal dosages of CSC and EtOH in OSCC cells. They were selected for further downstream experiments investigating differential gene expression.

2.2 Differential Gene Expression in Response to CSC and EtOH

2.2.1 Introduction

Cancer develops as a result of an accumulation of gene mutations and epigenetic changes which alter cellular function leading to tumour formation. Considering that CSC and EtOH are mutagenic, it is possible that these compounds may cause mutations in oncogenes and tumour suppressor genes which lead to OSCC development (Qin et al., 2016). In addition, studying differential gene expression in response to treatment with CSC and alcohol (EtOH) may allow us to understand the biological underpinnings of how those factors are involved in tumorigenesis. It was hypothesized that xenobiotic enzymes, in addition to other metabolic signalling pathways may be differentially expressed in response to treatment with CSC and EtOH *in vitro*. It is very likely that these altered metabolic and signaling pathways in response to EtOH and CSC treatment also play an important role in the development of OSCC, in addition to the mutational events triggered by these mutagenic compounds.

Previous studies have used RNA sequencing techniques to study molecular mechanisms of OSCC pathogenesis with the intent to gain knowledge on potential early biomarkers for diagnosis (Zeng et al., 2017). To further investigate the role of cigarette smoking and alcohol in the development of OSCC, RNA-seq was used to determine differential gene expression in response to treatment with sub-lethal concentrations of CSC and EtOH. These genes were then further analysed to investigate which pathways were being affected by CSC and alcohol treatment. To confirm these results a select few genes were chosen from the resulting list for further analysis using Quantitative Real-Time PCR (RT-qPCR).

2.2.2 RNA Extraction

RNA analysed for gene expression analysis was extracted from three OSCC cells lines (KYSE150, WHCO6, KYSE180), one breast cancer cell line (MDA-MB-213) and one cervical cancer line (SiHa). All cells were treated in triplicate with 40 µg/ml CSC, 50mM EtOH or the combination of CSC and EtOH combination.

The untreated sample was treated with 0.2% DMSO only. The RNA extracted from the KYSE150 cells were used for RNA-sequencing (RNA-seq) and RT-qPCR, whereas the RNA from the other cells was only used for RT-qPCR analysis. KYSE150 was chosen for the RNA-Seq analysis as these cells were obtained from ATCC and so is therefore a recognised OC cell line.

Good quality RNA is required for both these assays. RNA quality was assessed on the nanodrop while RNA integrity was checked by formaldehyde MOPS gel electrophoresis. A260/280 and A260/230 ratios can be seen in table 2.1. The 260/280 ratios and 260/230 ratios were between 1.6-2.09 indicating good RNA purity. The RNA yields for all samples were above 1000 ng/ul RNA, more than what was required for RNA-Seq and RT-qPCR. Minimal degradation was seen in the RNA extracted from KYSE150, KYSE180 and WHCO6 in the formaldehyde MOPS gel (figure 2.6-2.8). As seen in the gel images below, clear 28S and 18S RNA bands were visualised indicating good quality RNA. The RNA isolated from cell lines WHCO6, KYSE180, MDA-MB-213, SiHa were not used for RNA-seq, only RT-qPCR, but were in similar ranges for concentration and absorbance ratios, A260/280 and A260/230.

Twelve RNA samples isolated from KYSE150 OSCC cells treated as previously described were submitted for RNA-seq at the Centre for Proteomic and Genomic research (CPGR). At the CPGR, samples were reanalysed for quality control; this included using the NanoDrop ND 8000 (Thermo Fisher Scientific) to assess the purity and quality of the samples, Qubit® 4.0 Fluorometer for absolute quantification and the Agilent TapeStation RNA ScreenTape (Agilent Technologies) to assess RNA integrity using the RIN values. The purity ratios of the RNA samples fell between 2.04 – 1.98 for the A260/280 ratio, and 1.98 – 1.43 for the A260/230 ratio. For quantification the concentrations from their nanodrop readings to be slightly lower than the concentrations as seen in table 2.1, however they were found to be similar to their Qubit readings and a sufficient yield to carry on with the mRNA isolation and library prep steps of the RNA-seq. For RNA integrity, RIN numbers were between 8.3 and 10.0, indicating good quality RNA.

Table 2.1: Table displaying nanodrop results for RNA extracted for RNA-seq and RT-qPCR. A260/280 and A260/230 absorbance ratios, concentration of RNA (ng/ul) and total RNA in 30 ul DEPC-H_20 (elution solution).

RNA	**Sample**	**Nucleic Acid Conc. (ng/ul)**	**260/280**	**260/230**	**Total (ug)**
1	**KYSE150 CSC 1**	1355,1	2,04	2,09	40,653
2	**KYSE150 CSC 2**	1839,7	2,01	1,81	55,191
3	**KYSE150 CSC 3**	3385	2,02	2,12	101,55
4	**KYSE150 EtOH 1**	1525,6	2,03	2,03	45,768
5	**KYSE150 EtOH 2**	1674,2	2,04	1,96	50,226
6	**KYSE150 EtOH 3**	2076,1	1,98	1,91	62,283
7	**KYSE150 Combo 1**	2164,2	2	2,08	64,926
8	**KYSE150 Combo 2**	1547,6	2,04	2,03	46,428
9	**KYSE150 Combo 3**	1506,7	2,03	2,02	45,201

10	**KYSE150 Untreated 1**	1835	2.01	1.95	55.05
11	**KYSE150 Untreated 2**	1656.1	1.99	1.6	49,68
12	**KYSE150 Untreated 3**	3118.4	2.02	1.66	93.55

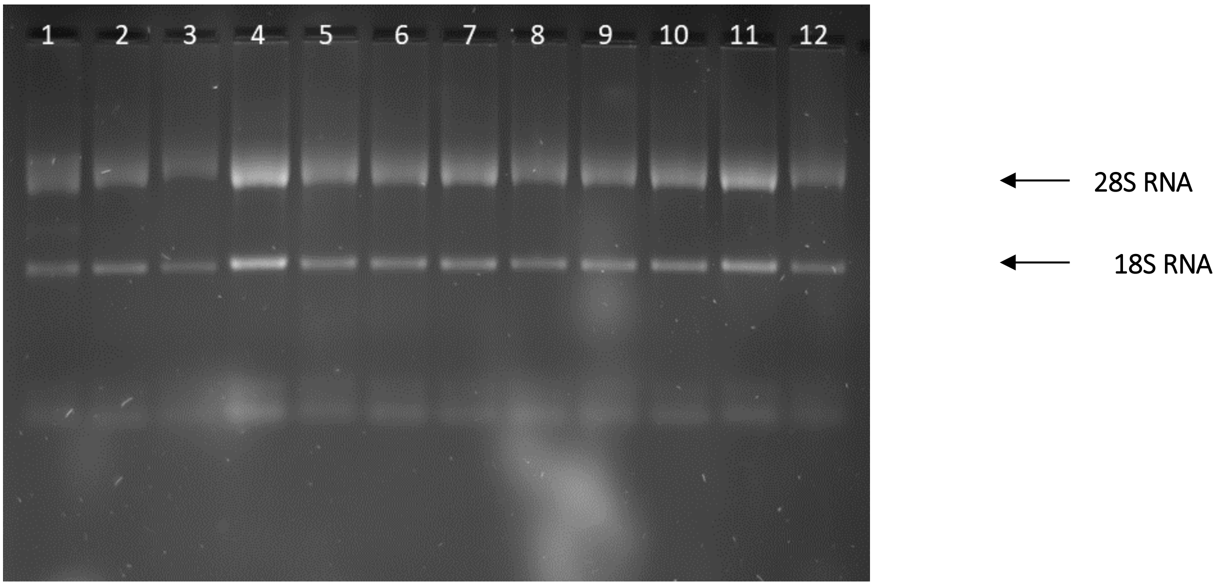

Figure 2.6: Extracted RNA from treated and control cultured KYSE150 OSCC cells, run on a MOPS gel at 80V for 45 minutes. Lane 1-3 RNA from cells treated with 40 ug/ml CSC. Lane 4-6 RNA from cells treated with 50mM EtOH. Lane 7-9: RNA from cells treated in combination with 40 ug/ml CSC and 50mM. Lane 10-12: RNA from untreated cells (0.2% DMSO).

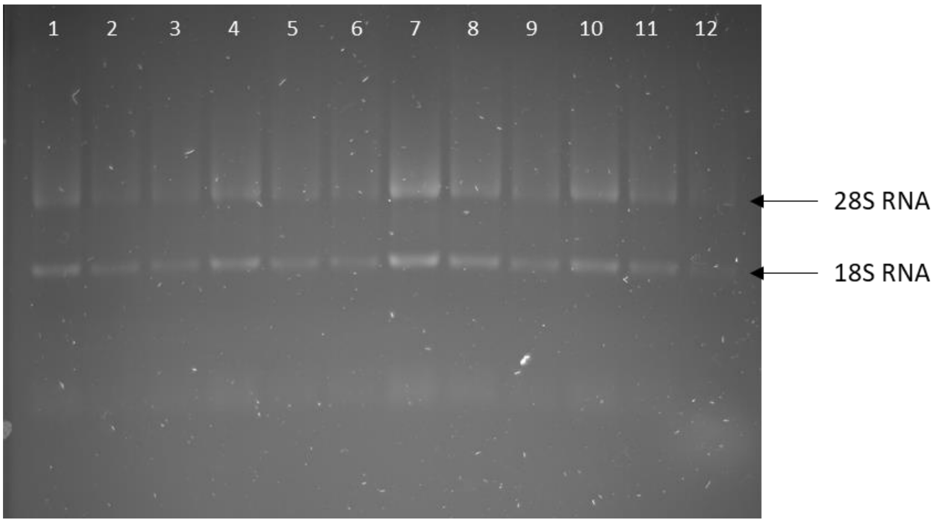

Figure 2.7: Extracted RNA from treated and control cultured KYSE180 OSCC cells, run on a MOPS gel at 80V for 45 minutes. Lane 1-3 RNA from cells treated with 40 ug/ml CSC. Lane 4-6 RNA from cells treated with 50 mM EtOH. Lane 7-9: RNA from cells treated in combination with 40 ug/ml CSC and 50mM. Lane 10-12: RNA from untreated cells (0.2% DMSO).

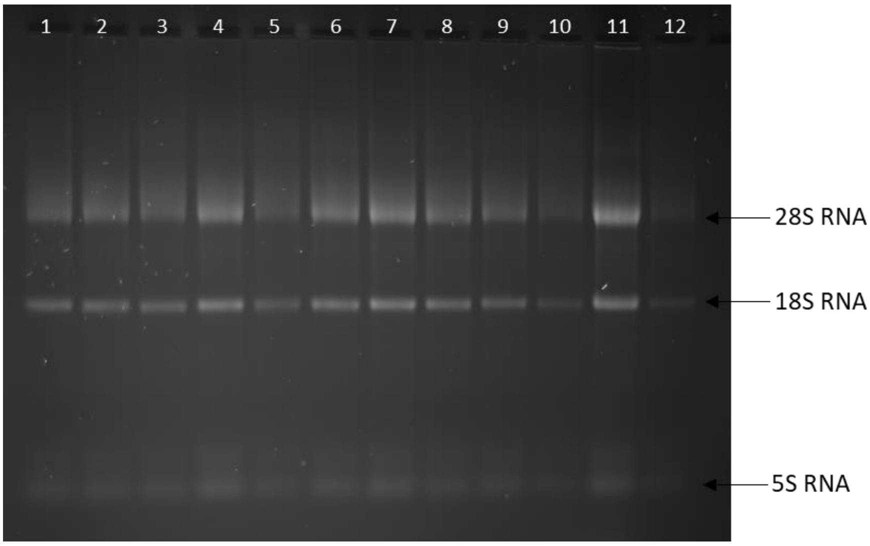

Figure 2.8: Extracted RNA from treated and control cultured WHCO6 OSCC cells, run on a MOPS gel at 80V for 45 minutes. Lane 1-3 RNA from cells treated with 40 ug/ml CSC. Lane 4-6 RNA from cells treated with 50 mM EtOH. Lane 7-9: RNA from cells treated in combination with 40 ug/ml CSC and 50mM. Lane 10-12: RNA from untreated cells (0.2% DMSO).

2.2.3 RNA-Seq Gene Expression Analysis

RNA-Seq was used to identify genes differentially expressed in OSCC cells treated with CSC and EtOH, individually and in combination, compared to untreated cells. The CPGR carried out the sequencing and conducted the bioinformatic analysis. The output FastQ files were subject to quality control and the reads were subsequently aligned to the refence genome using HiSAT2. The reads were then quantified using HTSeq and differential expression analysis was performed using DESeq2. This data was analysed to determine the impact of the treatment conditions on gene expression.

A list of genes differentially expressed can be seen in table 2.2-2.6. A total of 22 genes were found to be up-regulated; 16 genes were up-regulated in response to the combination of CSC and EtOH, 10 genes were up-regulated in response to CSC and 4 genes were up-regulated in response to EtOH. A total of 17 genes were found to be down-regulated; 7 genes were down-regulated in response to the combination of CSC and EtOH, 10 genes were down-regulated in response to CSC and 7 genes were down-regulated in response to EtOH. The number of genes which were differentially expressed was lower that what was anticipated. When investigating the differentially expressed genes (DEGs), genes which were differentially expressed with a p-value<0.1 were also considered for further analysis, this was because the number of DEGs were so unexpectedly low. There was considerable overlap between genes that were up and downregulated in response to the combination treatment and the CSC treatment, as seen in figure 2.9. In total, 27 were protein coding genes and 12 were characterised as non-coding RNAs.

Table 2.2: Significantly differentially expressed (DE) genes as detected using RNA-Seq in KYSE150 cells treated with CSC and the combination of CSC and EtOH. This table shows a list of genes as an output of the assay, the NCBI official full name and/or description (if available), log2Fold change values for the treatments.

Differentially Expressed Genes in Combo and CSC		
UpRegulated	**NCBI Gene Name/Description**	**Log2FoldChange**
C10orf55	Chromosome 10 Putative Open Reading Frame 55, lncRNA	1.43- Combo 1.27- CSC ; p<0.05
CAMK4	Calcium/Calmodulin Dependent Protein Kinase IV	1.12- Combo 1.37- CSC ; p<0.05
FCMR	Fc fragment of IgM receptor (antibody fragment)	7.82- Combo 8.23- CSC ; p<0.05
MALSU1	Mitochondrial Assembly Of Ribosomal Large Subunit 1	0.73- Combo 0.88- CSC ; p<0.05
RMDN2	Regulator of microtubule dynamics 2	1.74- Combo 1.95- CSC ; p<0.05
SERPINB2	Serpin Peptidase Inhibitor B2	4.48- Combo 4.73- CSC ; p<0.1
LOC105374836	ncRNA	0.81- Combo; p<0.05 0.65-CSC; p<0.1
CLVS1	Clavesin 1	0.8- Combo; p<0.1 0.96- CSC; p<0.05
DownRegulated		
CHRM4	Cholinergic receptor muscarinic 4 (G protein-coupled receptor)	-1.94- Combo -2.07- CSC ; p<0.05
GPR39	Ghrelin receptor family	-1.62- Combo -1.81- CSC ; p<0.05
LINC01764	Long Intergenic Non-Protein Coding RNA 1764	-0.84- Combo -0.76- CSC ; p<0.05
LOC105373623	ncRNA	-2.07- Combo -1.96- CSC ; p<0.05
LOC112268170	ncRNA	-0.57- Combo -0.67- CSC ; p<0.05
SPATA9	Spermatogenesis-Associated Protein 9	-0.84- Combo; p<0.05 -0.94- CSC p<0.1

Table 2.3: Significantly differentially expressed genes as detected using RNA-Seq in KYSE150 cells treated with EtOH and the combination of CSC and EtOH. This table shows a list of genes as an output of the assay, the NCBI official full name and/or description (if available), log2Fold change values for the treatments.
* WTAPP1 was upregulated in response to combination treatment and downregulated in response to EtOH.

Differentially Expressed Genes in Combo and EtOH		
Upregulated	**NCBI Gene Name/Description**	**Log2FoldChange**
RUSC1-AS1	ncRNA	0.66- Combo 0.83- EtOH ; p<0.05
*WTAPP1	Wilms tumor 1 associated protein pseudogene 1	1.43- Combo -1.41- EtOH; p<0.05

Table 2.4: Significantly differentially expressed genes as detected using RNA-Seq in KYSE150 cells treated with the combination of CSC and EtOH. This table shows a list of genes as an output of the assay, the NCBI official full name and/or description (if available), log2Fold change values for the treatments.

Differentially Expressed Genes in Combo		
Upregulated	**NCBI Gene Name/Description**	**Log2FoldChange**
LINC01527	Long Intergenic Non-Protein Coding RNA 1527	6.87 ; p<0.05
TCP1	Chaperonin containing TCP1 subunit (molecular chaperone, Subunit of CCT8)	0.74 ; p<0.05
TRNS1	Mitochondrially encoded tRNA serine 1	5.98 ; p<0.05
ARSG	sulfatase enzyme family	1.19 ; p<0.1
CBR4	Carbonyl Reductase 4	0.88 ; p<0.1
MTERF2	Mitochondrial Transcription Termination Factor 2	0.69; p<0.1
SH3PXD2A	SH3 and PX domains 2A	1.89 ; p<0.1
Downregulated		
RNR2	mtRNA- mitochondrially encoded 16S RNA (small subunit of ribonucleotide)	-1.02 ; p<0.05

Table 2.5: Significantly differentially expressed genes as detected using RNA-Seq in KYSE150 cells treated with CSC. This table shows a list of genes as an output of the assay, the NCBI official full name and/or description (if available), log2Fold change values for the treatments.

Differentially Expressed Genes in CSC		
Upregulated	**NCBI Gene Name/Description**	**Log2FoldChange**
RPE65	Retinoid isomerohydrolase	2.21; p<0.05
USP47	Ubiquitin-specific Peptidase 47	1.31; p<0.1
Downregulated		
KRT7-AS	KRT7 Antisense RNA 1- ncRNA	-0.84; p<0.05
LOC107985034	ncRNA	-0.66; p<0.05
DNASE1	Deoxyribonuclease 1	-0.56; p<0.1
THRIL	TNF And HNRNPL Related Immunoregulatory lncRNA	-0.63; p<0.1

Table 2.6: Significantly differentially expressed genes as detected using RNA-Seq in KYSE150 cells treated with EtOH. This table shows a list of genes as an output of the assay, the NCBI official full name and/or description (if available), log2Fold change values for the treatments.

Differentially Expressed Genes in EtOH		
Upregulated	**NCBI Gene Name/Description**	**Log2FoldChange**
SHISA8	Family of single-transmembrane proteins	0.76; p<0.05
WDR38	WD Repeat-Containing Protein 38	0.93; p<0.05
LOC100130283	ncRNA	0.78; p<0.1

Downregulated		
CEBPZOS	CEBPZ Opposite Strand	-0.96; p<0.05
GLIPR1	Glioma Pathogenesis-Related Protein 1	-0.94; p<0.05
IQCK	IQ motif containing K	-1.43; p<0.05
KRR1	KRR1 small subunit processome component homolog	-1.38; p<0.05
LOC107985304	ncRNA	-0.7; p<0.05
SLC66A3	Solute Carrier Family 66 Member 3 (Transmembrane protein)	-0.74; p<0.1

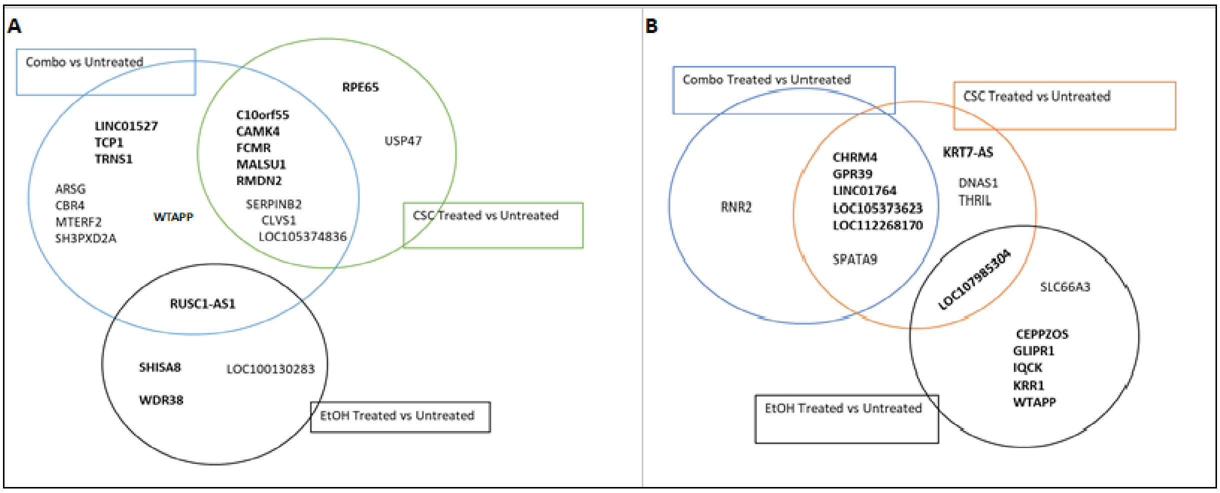

Figure 2.9: (A) Venn diagram displaying the RNA sequences upregulated in response to CSC and EtOH treatment compared to untreated cells. (B) Genes downregulated in response to treatment with CSC and EtOH compared to untreated cells. Genes in bold were significantly differentially expressed p<0.05, genes not in bold were differentially expressed with a p-value <0.1.

The graph, seen in figure 2.10, shows the relative fold change of genes that were differentially expressed in response to CSC and EtOH relative to the untreated sample. More genes were affected by CSC than EtOH. Most genes had a smaller log2fold change difference of approximately between 1 and 2; and some genes had a larger fold change, between 4.7 and 8.2. Six genes (and their log2FoldChange values) were selected for further investigation using RT-qPCR: SERPINB2, 4.48 (Combination) and 4.73 (CSC); GPR39, -1.62 (Combination) and -1.81 (CSC); SH3PXD2A, 1.89 (Combination); GLIPR1, -0.94 (EtOH); USP47, 1.31 (CSC) and FCMR, 7.82 (Combination) and 8.23 (CSC). These are highlighted with red ovals in figure 2.10.

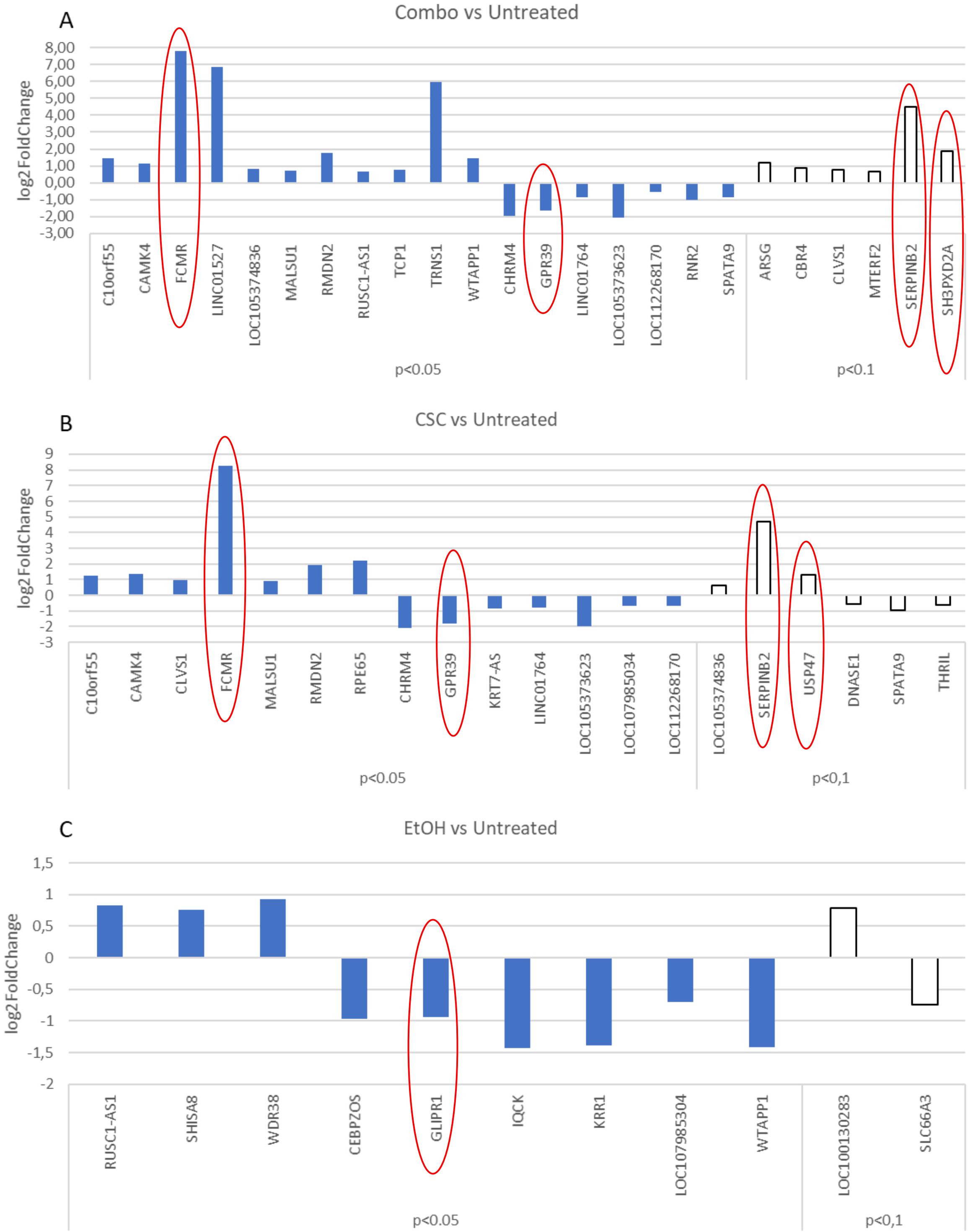

Figure 2.10: RNA-Seq data- Graphs displaying the relative log2Fold change of genes differentially expressed in KYSE150 OSCC cells, where the genes were up or downregulated with a p<0.05 or p<0.1. (A) Graph showing genes up or downregulated in response to the combination of CSC and EtOH compared to untreated cells (0.2% DMSO). (B) Graph showing genes up or downregulated in response to CSC treatment compared to untreated cells (0.2% DMSO). (C) Graph showing genes up or downregulated in response to EtOH treatment compared to untreated cells (0.2% DMSO). Genes circled were selected for further analysis using RT-qPCR.

2.2.4 RT-qPCR Gene Expression Analysis

RT-qPCR was used to confirm the results of the RNA-Seq to determine the effects of CSC and EtOH treatment on cultured OSCC cells. Six genes were selected from the RNA-seq data of genes found to be differentially expressed in response to the treatments: SERPINB2, GPR39, SH3PXD2A, GLIP1R, FCMR and USP47. They were chosen based on their potential involvement in cancer development, as reported by previous literature. Apart from the other genes, FCMR was specifically chosen as it was highly upregulated in the RNA-Seq results, as seen in figure 2.10, with a log2 fold change of 7.8 and 8.3 in combination and CSC treated cells, respectively.

cDNA was prepared from the same sample of RNA which was submitted for RNA-Seq analysis, as described earlier, in addition to RNA from four other cell lines (WHCO6, KYSE180, MDA-MB-213 and SiHa). Two levels of RT-qPCR analysis were performed to confirm the differential expression observed in the RNA-Seq analysis. (1) The exact RNA samples (KYSE150) which were submitted for RNA-Seq analysis were subjected to RT-qPCR. (2) In addition , the OC cell lines: WHCO6, KYSE180, and breast cancer cell line, MDA-MB-213 and cervical cancer cell line, SiHa were treated with CSC and EtOH separately and in combination. This was done to determine the reproducibility of the observed changes in other oesophageal cancer cell lines and other cancer types.

RT-qPCR was performed as described using the qPCR primers shown in table 5.3, chapter 5. The fold change is the change in RNA levels of the particular gene, representing changes in gene expression. This was calculated from the average Ct values of three technical replicates, using the $2^{-\Delta\Delta CT}$ method for determining relative quantification (Livak & Schmittgen, 2001). GAPDH was used as an internal housekeeping gene for normalisation and statistical significance was determined by student t-test, comparing treated cells to untreated cells.

As seen in figure 2.11, SERPINB2 was upregulated in cells treated with CSC and the combination of CSC and EtOH in both KYSE150 and WHCO6 cell lines, confirming the results of the RNA-Seq assay. This was observed in all three biological replicates. In the KYSE180 cell line, SERPINB2 was upregulated in cells treated with CSC, this was confirmed in two biological replicates, however no significant change in gene expression was observed in the third replicate. In MDA-MB-213 breast cancer cells no differential expression was observed. In SiHa, cervical cancer cells, SERPINB2 was upregulated in cells treated with CSC and EtOH. This was confirmed in two biological replicates.

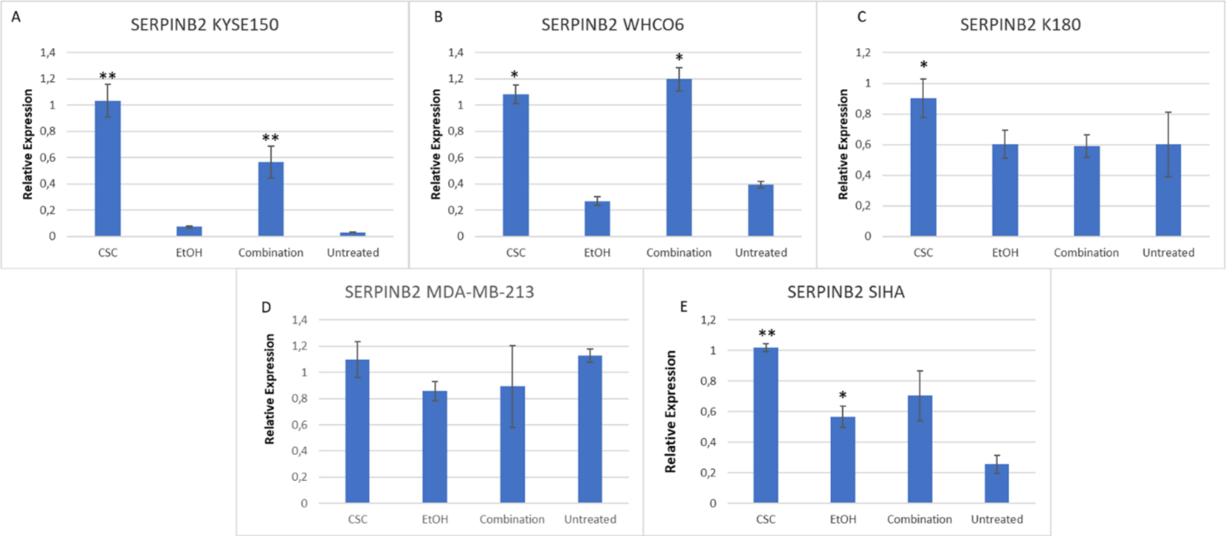

Figure 2.11.: Quantitative real time PCR of SERPINB2 in (A) KYSE150, (B) WHCO6, (C) KYSE180, (D) MDA-MB-213 and (E) SiHa cells. Expression was normalised to GAPDH reference gene and shown as relative values. Statistical analysis was determined using the t-test, comparing treated to untreated samples (** p<0.005; * p<0.05). Error bars indicate standard deviation.

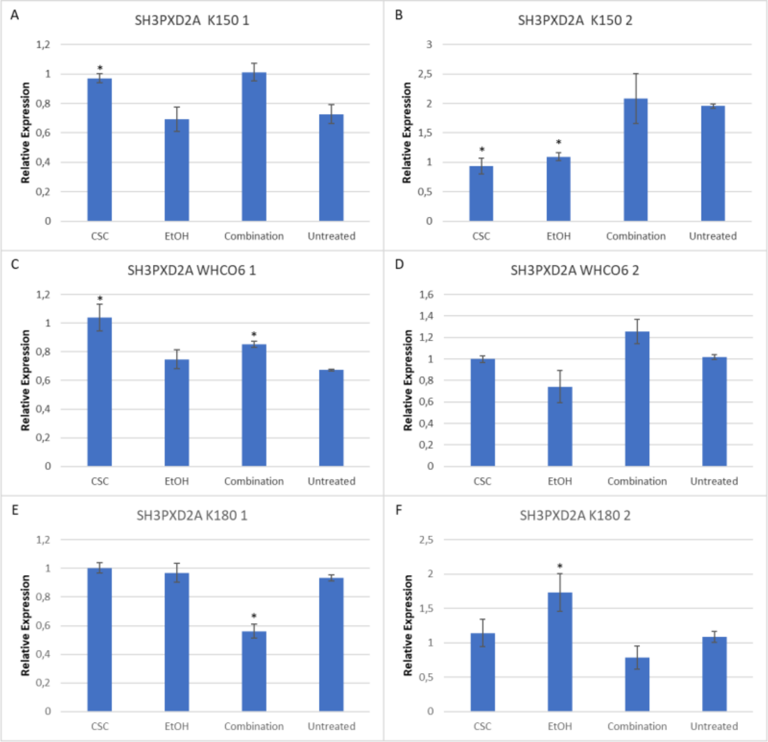

Figure 2.12: Quantitative real time PCR of SH2PXD2A in OSCC cells (A) KYSE150 1, (B) KYSE150 2 (C) WHCO6 1, (D) WHCO6 2, (E) KYSE180 1, (F) KYSE180 2. Expression was normalised to GAPDH reference gene and shown as relative values. Statistical analysis was determined using the t-test, comparing treated to untreated samples (* p<0.05). Error bars indicate standard deviation.

SH3PXD2A which was found to be overexpressed in the RNA-Seq analysis in CSC treated cells, with a p<0.1, however the Rt-qPCR results on the relative expression of SH3PXD2A was highly variable. As seen in figure 2.12 above, SH3PXD2A was found to be upregulated CSC treated cells, this was confirmed by a secondary biological replicate. However, in the third biological replicate (data not shown) SH3PXD2A was found to be significantly downregulated in both EtOH and CSC treated KYSE150 cells. In the WHCO6 cell line, SH3PXD2A was found to be significantly upregulated in cells treated with CSC and combination treatment in one of the biological replicates, whereas no significant difference was observed in the other two biological replicates. In the KYSE180 cell line SH3PXD2A was found to be significantly downregulated in cells treated with the combination of CSC and EtOH in two biological replicates and significantly upregulated in one biological replicate treated when treated with EtOH.

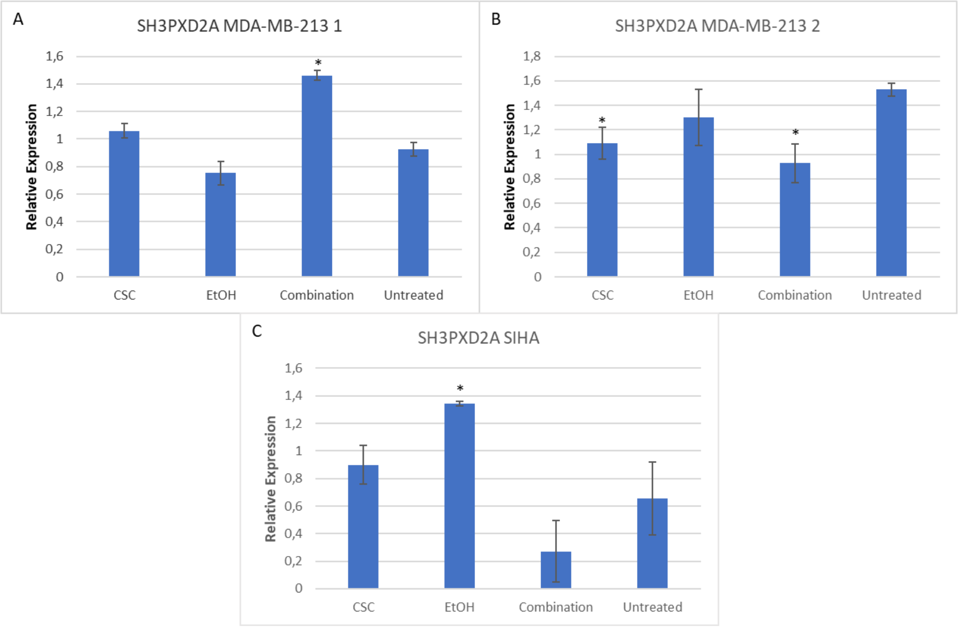

Figure 2.13: Quantitative real time PCR of SH2PXD2A in (A) MDA-MB-213 1, (B) MDA-MB-213 2 and (C) SiHa cells. Expression was normalised to GAPDH reference gene and shown as relative values. Statistical analysis was determined using the t-test, comparing treated to untreated samples (* p<0.05). Error bars indicate standard deviation.

In the non-OSCC cell lines, the variation among the biological replicates was continued to be observed, as seen in figure 2.13 above. In the MDA-MB-213 breast cancer cells, one biological replicate showed a significant upregulation in the cells treated with combination of CSC and EtOH, the second biological showed a slight but significant downregulation in the combination and CSC treated cells; while the third biological replicate showed no change in expression in treated cells compared to untreated cells. In the SiHa cervical cells, one cell line showed a significant upregulation in the cells treated with EtOH, the second biological replicate showed no differential gene expression. The third biological replicate had Ct values below the no template control, indicating that SH3PXD2A was not present in this cell. The RT-qPCR assay on the third sample was repeated twice for confirmation. Given the variability in the results observed, the results appear erratic, and not consistent with the RNA Seq data, thus implying that it cannot be concluded that there is a change in gene expression.

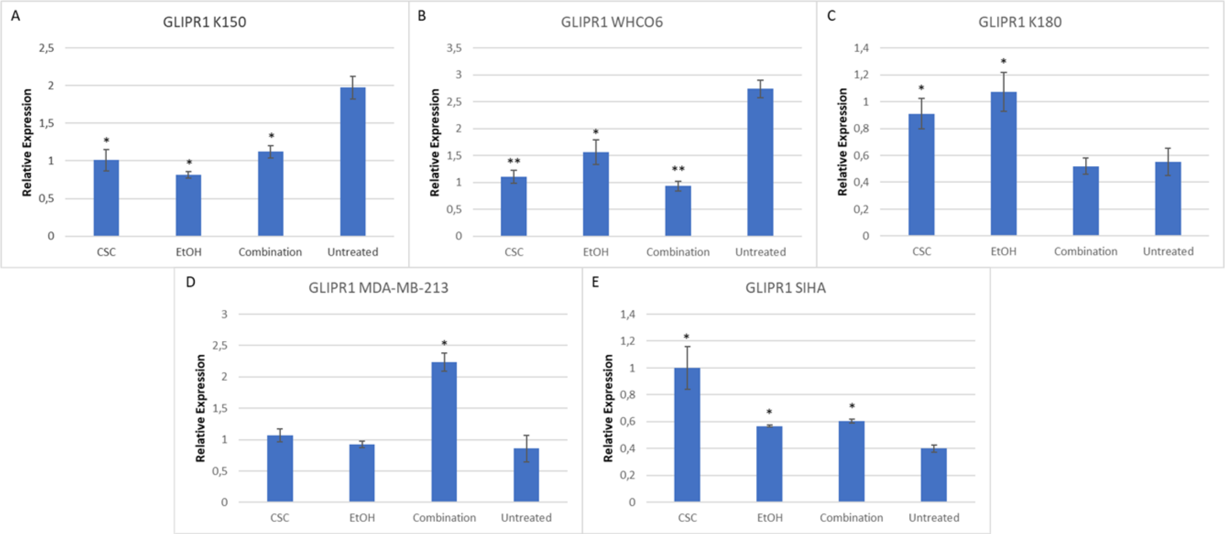

Figure 2.14: Quantitative real time PCR of GLIPR1 in (A) KYSE150, (B) WHCO6, (C) KYSE180, (D) MDA-MB-213 and (E) SiHa cells. Expression was normalised to GAPDH reference gene and shown as relative values. Statistical analysis was determined using the t-test, comparing treated to untreated samples (** $p<0.005$; * $p<0.05$). Error bars indicate standard deviation.

The RNA-seq data showed GLIPR1 was down-regulated in response to EtOH treatment, and this was confirmed by most RT-qPCR assays we performed. In the KYSE150 cell, line GLIPR1 was shown to be down regulated with all treatments as seen in figure 2.14. This was confirmed by second biological replicate which showed expression was downregulated when treated with CSC and EtOH individually, however in the last replicate no change in expression was observed. In the WHCO6 cell line, GLIPR1 was significantly downregulated in response to CSC, EtOH and the combination treatments, and this observed in all biological replicates. In KYSE180 cells, GLIPR1 was shown to be upregulated with CSC and EtOH treatment in two biological replicates. However, in the third biological replicate (not shown), GLIPR1 was upregulated in CSC only. In the MDA-MB-231 cells, GLIPR1 was shown to be significantly upregulated in the combination treated cells, as seen in figure 2.14(D). In the second biological replicate, expression was upregulated in EtOH treated cells, however in the third biological replicate no change in expression was observed. In the SiHa cell line, GLIPR1 was shown to be upregulated in all treatments for one assay (figure 2.14 (E)), downregulated in CSC treated cells for one biological replicate and showed no differential gene expression in the other biological replicate.

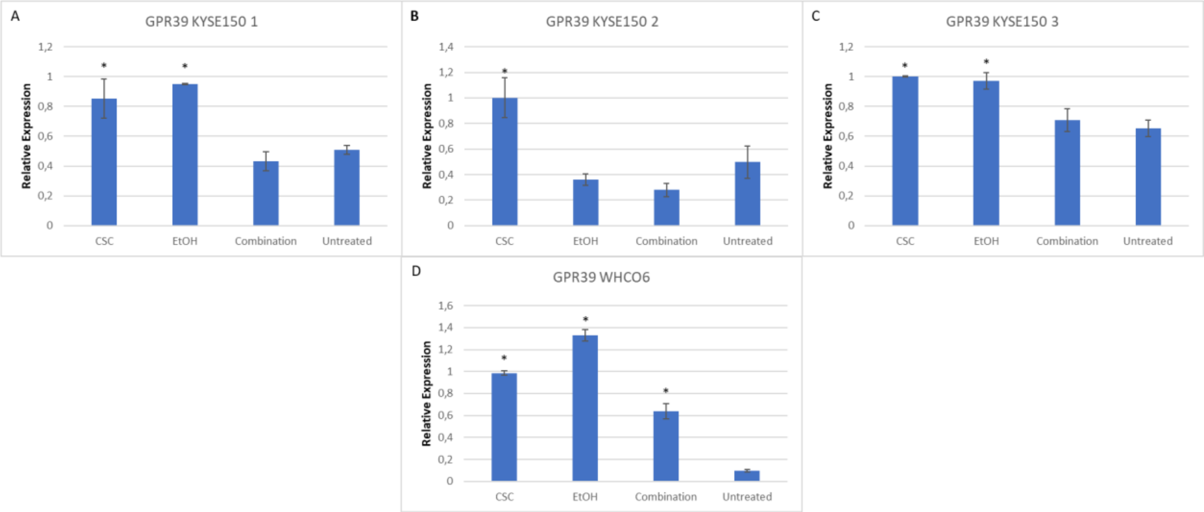

Figure 2.15: Quantitative real time PCR of GPR39 in (A) KYSE150 1, (B) KYSE150 2, (C) KYSE150 3 biological replicates and in (D) WHCO6 cells. Expression was normalised to GAPDH reference gene and shown as relative values. Statistical analysis was determined using the t-test, comparing treated to untreated samples (* p<0.05). Error bars indicate standard deviation.

GPR39 was downregulated in the RNA-seq data and this was not confirmed by the RT-qPCR assay. As seen in figure 2.15, GPR39 was upregulated in response to CSC and EtOH treatment in two biological replicates. The other biological replicate showed an upregulation in CSC treatment only. In WHCO6 cells, GPR39 was upregulated in response to all treatments in one assay, in the other two biological replicates no differential expression was observed. As the data was not reflecting the results observed in the RNA-Seq, this gene was not further investigated in the other cell lines.

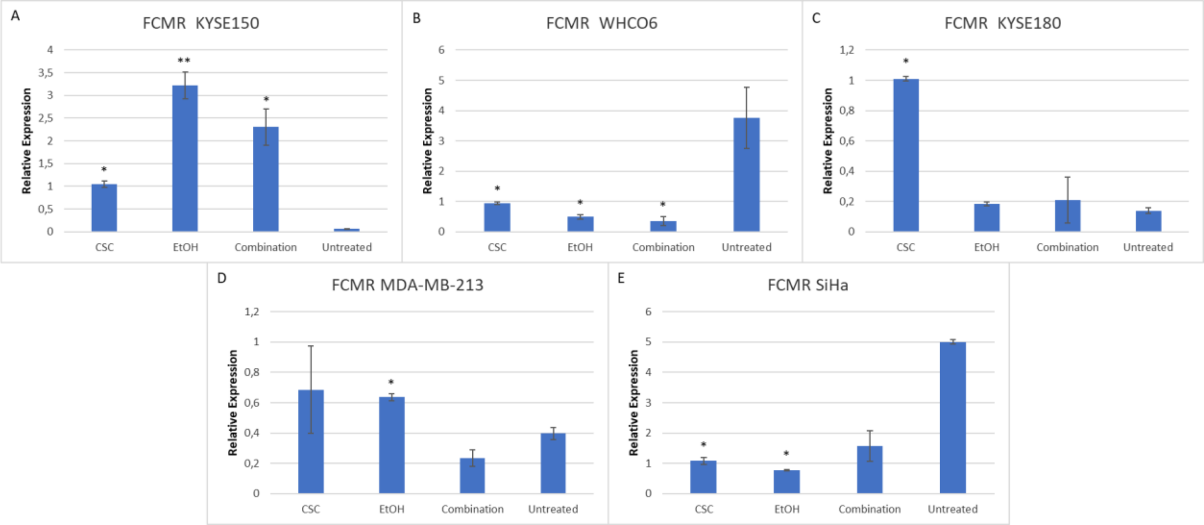

Figure 2.16: Quantitative real time PCR of FCMR in (A) KYSE150, (B) WHCO6, (C) KYSE180, (D) MDA-MB-213 and (E) SiHa cells. Expression was normalised to GAPDH reference gene and shown as relative values. Statistical analysis was determined using the t-test, comparing treated to untreated samples (* p<0.05). Error bars indicate standard deviation.

FCMR was highly upregulated in the RNA-seq data and thus was expected to be upregulated in the RT-qPCR data. In the KYSE150 cells FCMR was upregulated among all treatment groups, and this was reflected in two biological replicates, confirming the RNA-Seq data. The third biological replicate did not show any change in gene expression. In WHCO6 cells, FCMR was downregulated in two biological replicates. In KYSE180 cells, FCMR was upregulated in CSC treated samples of one biological replicate (*fig. 2.16(C)*), whereas the other two biological replicates showed no change in gene expression. In the MDA-MB-213 cells, FCMR was significantly upregulated in EtOH and can be seen to be upregulated in CSC treated cells, however this was not statistically significant likely due to the large standard deviation observed. Biological replicate two showed downregulation in combination, however this was not statistically significant and biological replicate three showed no change in gene expression. In SiHa cells, FCMR showed significantly decreased expression in CSC and EtOH treated cells, but not the combined treatment, (although the trend showed a decrease). Biological replicate two and biological three showed no statistically significant change in gene expression.

The last gene chosen for further investigation, USP47, could not be successfully amplified using RT-qPCR. As seen in figure 2.17*,* the melt curve had multiple peaks for this gene of interest, indicating there was non-specific amplification of the cDNA. GAPDH melt curve had one peak for all treatments at the same temperature, as expected, indicating one amplicon was produced. In further optimisation experiments, amplification temperatures were altered with no improvement in the specificity of the PCR reaction.

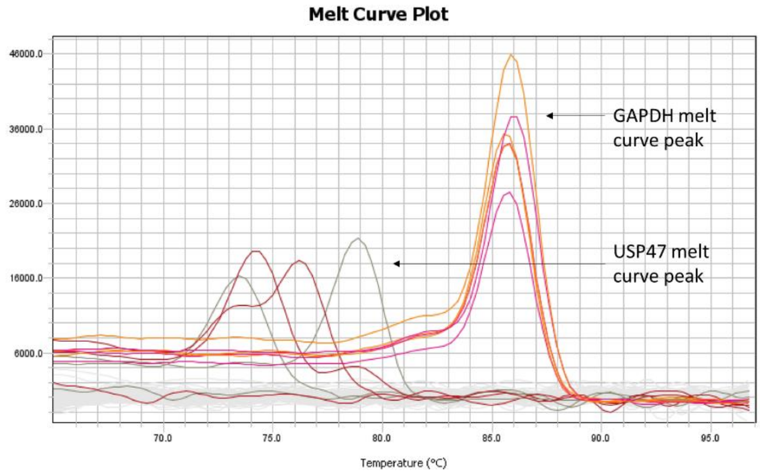

Figure 2.17: Melt curve plot of USP47 optimisation RT-qPCR. Multiple peaks indicate non-specific amplification. Melt curve plots show data collected during the melt curve stage, where the peaks indicate the melting temperature (T_m) of the amplified product and can identify non-specific PCR amplification.

2.2.5 Re-analysis of RNA-Seq Data

After assessing the data, it was decided the data required re-analysis using a different approach. This was done because the analysis of the RNA-Seq data using the HiSAT2/HTSeq approach at CPGR only yielded a low number of differentially expressed genes. Furthermore, it was anticipated xenobiotic and alcohol metabolising genes would be upregulated in response to CSC and alcohol, none of which were present in the original dataset.

The second bioinformatic analysis approached was conducted with assistance of a previous lab member (Wei Wei). As previously mentioned, the CPGR originally employed a strategy using HiSAT2 and HTSeq for alignment and quantification of the obtained sequence reads. Unfortunately, when re-analysing the data the internal QC suggested approximately 80% of the reads were not assigned to any transcriptomic annotation. For the re-analysis the CPGR provided the transcript quantification through an alignment-free algorithm implemented by Salmon. Re-analysis with the new Salmon algorithm gave QC results which suggested 84%±1% reads were assigned to the transcriptome, addressing the concern and giving greater confidence in the second set of results.

After the analysis of the RNA-seq data and sequence annonation, DE analysis was performed. Only those genes which were significantly altered in expression, both over- and under-expressed, were plotted into the heatmap illustrated in figure 2.18. The heatmap clearly displays consistency of results between biological replicates for the different treatment conditions, confirming a high level of reproducibility with the experiments. Another observation drawn from the heat map, is similar to the first analysis, genes over or underrepresented are grouped together in the CSC and combination treated samples, whereas genes over or underrepresented in response to EtOH and untreated samples are grouped together. These clusters of genes grouping together indicates that the dominant factor altering gene expression is CSC treatment.

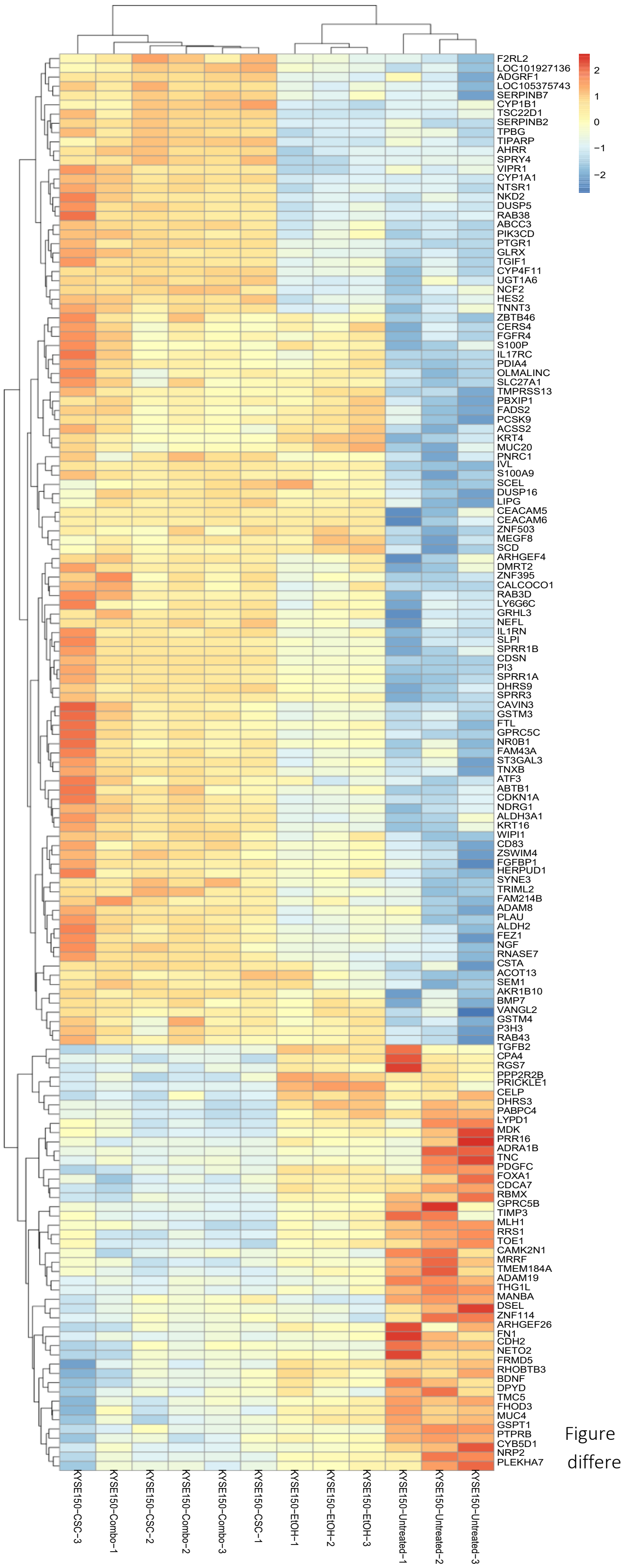

Figure 2.18: Heat-map showing genes differentially expressed after re-analysis.

As seen in figure 2.19*,* a total of 153 genes were over or underrepresented with statistical significance. Seventy-five genes were reported as being overrepresented in the CSC and combination samples, while being underrepresented in the untreated samples. 46 genes were underrepresented in CSC and combination treated samples, while being over-represented in the untreated samples. 32 genes can be seen to be overrepresented in all three treatment groups, while being underrepresented in the untreated sample. The results of the second analysis found a much larger set of affected genes compared to the 39 genes found to be differentially expressed in the first analysis, shown in *section 2.2.3.* Only one gene was common to both datasets, SERPINB2, which was initially shown to be upregulated in response to CSC and combination as seen in *table 2.2*. In the re-analysis, SERPINB2 was shown to be overrepresented in both the CSC and combination treatment groups and all three biological replicates showed this same result.

Further comparison of the two datasets shows that although the first gene expression analysis revealed no xenobiotic genes to be differentially expressed, the re-analysis found 7 genes that were significantly overrepresented in the CSC and combination treated samples, as seen in figure 2.20. Indicating that CSC exposure did result in an increase in the gene expression of xenobiotic metabolising enzymes.

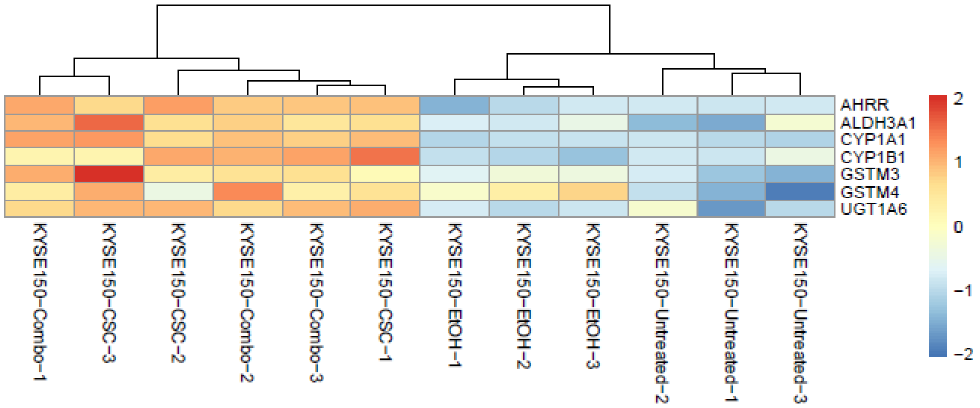

Figure 2.20: Heatmap showing xenobiotic genes overrepresented or underrepresented in response to CSC, EtOH, combination treatments and untreated samples.

Chapter 3: Discussion

3.1 Optimisation of CSC and EtOH Treatment

CSC and ethanol have been shown to cause cell death *in vitro* and this has implications when doing studies in cell culture investigating their effects as it has the potential to introduce bias on any observable results. Previous researchers have shown that treatment of cultured cells with CSC can cause cell death at different concentrations depending on the cell type being tested. In U937 monocytic cells after treatment with CSC for 24-hours, there was a significant loss of about 15% cell viability ($p \leq 0.05$) when treated with 50 µg/ml CSC (Rao et al., 2016). This was confirmed for the same concentration of CSC after a 48-hour incubation which resulted in a significant increase of 10 to 30% apoptosis in monocytic cells, measured using an Annexin V staining assay. In normal breast epithelial and breast cancer cells, cell viability decreased by 75% at doses of 100 µg/mL after 48-hours (Mohapatra et al., 2014). Oral squamous cell carcinoma cells treated with concentrations of 0 to 100 µg/ml CSC for time periods of up to 24-hours was examined by MTT assay to determine cell viability. Decreasing cell viability was observed with increasing concentration of CSC, and with increasing exposure time at each dose (Nagaraj et al., 2006). In mouse embryonic stem cells (mESCs) 48-hour treatments of 0.4, 4, and 40 µg/mL CSC resulted in a significant reduction in cell viability of 20-40% at all concentrations in the 2-C57 ESC cells ($P < 0.0001$), in a non-dose dependent manner. The 6-C57 ESC cells had significant increase in cell viability ($P < 0.05$) at 0.4 and 4 µg/mL CSC and significantly decreased viability of 40% at the 40 µg/mL concentration ($P < 0.0001$). A study on NMRI ESC cells showed slightly decreased viability at the 40 µg/mL concentration CSC, however, these changes were not significant (Assadollahi et al., 2019). In alveolar epithelial cells, A549 cell type, it was found that cell survival decreased with increasing concentration of CSC after 24-hour incubations, with significant reduction in cell viability of 15-25% from 100 µg/ml to 200 µg/ml of CSC. Whereas in WI26 cell type, no significant change was observed in survival rate up to 100µg/ml concentration of CSC. However, a reduced cell viability was observed at 150 µg/ml and 200 µg/ml CSC to around 95% and 62%, respectively (Yadav et al., 2016).

In addition to being a sub-lethal dose of CSC and EtOH in culture, the treatment concentrations also had to be in the range of what is considered physiologically relevant to what people might typically consume or present in the blood of subjects. Problems arising from trying to place a value on this is that it is dependent on factors such as how much people consume and could also be dependent on the brand of cigarette being smoked since all cigarettes are different. From the literature it has been suggested that physiologically relevant concentrations of CSC is around 50 µg/mL CSC although some studies have gone as high as using 200 µg/mL CSC in *in vitro* experiments (Rao et al., 2016). Another study used concentrations of CSC ranging from 0 to 400 mg/ml, stating that this was based on what is relevant to human exposure according to previous reports (Allam et al., 2011). Another study used 0.4, 4, and 40 µg/mL concentrations of CSC based on previous literature which showed concentrations higher than 40 µg/mL was the most optimal dose and higher concentrations led to extensive cell death and caused problems with gene expression assays (Assadollahi et al., 2019). These authors also state that according to the literature concentrations of 0.4, 4 and 40 µg/mL CSC overlap with plasma levels of nicotine found in smokers.

With regards to the relevant dosage of alcohol an article has been written as a guideline on how to investigate acute alcohol exposure in an *in vitro* model (Dolganiuc & Szado, 2009). In previous studies using a range of alcohol concentrations, from 1 - to 500 mmol/L, concentrations of above 100 mmol/L alcohol caused a direct cytotoxic effect on cells. Concentrations below 100 mmol/L has been shown to affect the redox status of cells, change intercellular junctions, increase membrane fluidity and affect the lipid rafts of cells which all affect cell permeability and thus cell function (Dolganiuc & Szado, 2009). In the US the blood alcohol level (BAL) of 0.08% is equivalent to four or five drinks which is equivalent to around 70 g of alcohol as 1 drink is defined as 14 g of alcohol (Dolganiuc & Szado, 2009). This BAL of above or equal to 0.08% has been established as a meaningful level of alcohol exposure both *in vitro* and *in vivo*. The effects of alcohol can be observed on physiology and/or behaviour, however with variations in individuals depending on age, gender, metabolic particularities, genetic background and species. For acute alcohol abuse the physiological range of ethanol is considered to be 10-100 mmol/L ethanol, with 25 mmol/L ethanol being close to 0.08% BAL achieved *in vivo* after 4-5 drink equivalents (Dolganiuc & Szado, 2009).

In our study we found a consistent pattern of decreased cell viability at 80 µg/ml CSC in various cell lines of OSCC cell types after a 72-hour incubation. Although, previous literature had found decreased cell viability at concentrations as low as 40 µg/mL CSC in ESCs (Assadollahi et al., 2019) and at higher concentrations of 100 µg/mL after only 24 hours, as seen in normal breast epithelial cells (Mohapatra et al., 2014). These differences may be due to the cell lines investigated, it is possible stem cells or normal cells might be more sensitive to the effects of CSC. Regarding the EtOH concentration to be investigated, 50 mM was as expected well within the range the literature reported to not be a lethal (Dolganiuc & Szado, 2009). Since 10-100 mmol/l ethanol was reported to being considered a physiological range of acute alcohol abuse an experimental concentration of 50 mM EtOH was fitting. Hence, 40 µg/mL CSC and 50 mM EtOH was chosen for further experiments.

For further investigation, to confirm that the chosen concentrations of CSC and EtOH did not cause cell death, a trypan blue assay which measures cell death as a fraction of total cell number was carried out. As seen in figure 2.4, no increased cell death was observed in the treated cells compared to the non-treated cells. There was a slight increase in cell number which might suggest the treatments increase cell growth. Based on these results and physiological concentrations reported by the literature, 40 µg/mL CSC and 50 mM EtOH concentrations were chosen to investigate how these risk factors affect gene expression in OSCC cells.

3.2 Gene Expression in response to CSC and EtOH

3.2.1 RNA-Seq Analysis

To investigate the role CSC and alcohol may play in the development of OSCC, gene expression in response to treatments were investigated using RNA-seq, this was subsequently confirmed using RT-qPCR. The inital RNA-Seq assay analysis found a total of 22 genes upregulated and 17 genes downregulated across treatment groups. This was much less genes than what was anticipated to be differentially expressed, the RNA-Seq data was then re-analysed used a different method. The new analysis resulted in a much larger set of genes shown to be differentially expressed compared to the first dataset. A total of 153 genes were found to be over or underrepresented with statistical significance. Seventy-five genes were reported as being overrepresented in the CSC and combination samples, 46 genes were underrepresented in CSC and combination treated samples and 32 genes were overrepresented in all three treatment groups. The new methodology was able to yield approximately 85% annotation of the sequences, compared to the first analysis which only resulted in 10% annotation of the sequences and consistency was seen among the biological triplicates, giving confidence in the reproducibility of the results.

The second bioinformatic analysis approached was conducted with the assistance of a previous lab member (Wei Wei). For each sample, RNA-seq generated approximately 25 million reads, these short sequencing reads were assigned to their original transcripts to infer gene expression. As previously mentioned, the CPGR originally employed a strategy by first aligning the sequencing reads to human reference genome using HiSAT2, following by using HTSeq to obtain relative expression levels for each gene. Unfortunately, the internal QC suggested approximately 80% of the reads were not assigned to any transcriptomic annotation. For the re-analysis the CPGR provided the transcript quantification through an alignment-free algorithm implemented by Salmon. The Salmon algorithm avoids base-to-base alignment of the reads and provides a faster and more accurate quantification directly at the transcript level (Patro et al., 2017). The QC suggested 84%±1% reads were assigned to the transcriptome, addressing the concern and giving greater confidence that the second differential expression analysis would be more accurate than the first. The large discrepancy in annotated reads between the two analyses were due to the method in which the two programs map reads. HTSeq discards both ambiguous and multi-mapping reads whereas in Salmon, both ambiguous and multimapping reads are counted. Thereafter, the gene quantification results were subjected to differential expression analysis using DESeq2 to determine the impact of the treatment conditions on gene expression.

One similarity observed between the two datasets was that there was a substantial overlap between the genes differentially expressed in the CSC treated samples and the combination treated samples. For dataset one this can be seen in figure 2.9 and for dataset 2, this can be seen in figure 2.17. This indicates that the effect on gene expression is primarily caused by the CSC treatment. There was however a subset of 32 genes that were overrepresented in response to CSC, combination and EtOH treatment alone, while being underrepresented in the untreated samples. This indicates that the concentration of EtOH used in the experiments was sufficient enough to cause an observable change. Further investigation is required to determine why these genes were altered by all three treatments.

When comparing our DEG dataset to genes most commonly mutated in an OSCC database (https://portal.gdc.cancer.gov/), one gene was found in the top 5 mutated genes. It was reported that the most mutated gene in OSCC is TP53. TP53 is a transcription factor known to be dysregulated in OSCC that plays a central role in genome integrity and cellular growth signals (Hendricks & Parker, 2002). MUC4 was the third most commonly mutated gene, found in approximately 30% of cases in this database and was found to overrepresented in the CSC and combination treated cells in re-analysis of the RNA-Seq data.

To determine functional pathways affected by the treatments, the data was investigated using Gene Set Enrichment Analysis- following the algorithm developed by MIT (Subramanian et al., 2005). When looking at functional pathways enriched in this dataset, several genes (n=26) related to reactive oxygen species metabolic process (GO:0072593) were activated upon treatment, 8 out of these 26 genes also mediate reactive nitrogen species metabolic process (GO:2001057).

In addition to resulting in only a small list of DEGs, the initial RNA-seq analysis did not find any xenobiotic metabolising genes to be differentially expressed. In the first analysis, to investigate whether xenobiotic genes were altered, genes containing the prefix: "CYP","SULT"," NAT","GST","UGT","ADH" and "ALDH" were searched for. They were chosen as previous literature have found South African populations with variants of these genes: GSTs, NATs, CYPs, SULTs, UGTs and EHs, were at higher risk of OSCC (Matejcic

& Parker, 2015). However, none of them were found to be significantly differentially expressed. In the second analysis, seven xenobiotic genes were overrepresented (figure 2.19), including: AHRR, ALDH3A1, CYP1A1. CYP1B1, GSTM3, GSTM4 and UGT1A6. All of these genes were overrepresented in CSC and combination treated samples and one gene, GSTM4, was overrepresented in all three treatment samples, including the EtOH treated samples.

One of the xenobiotic metabolising enzymes shown to be upregulated was the AHRR gene, which encodes for the the arylhydrocarbon receptor repressor, its expression is ubiquitous among tissues and can be altered by various AhR ligands (Haarmann-Stemmann & Abel, 2006), such as BaP, found in cigarette smoke (Chen et al., 2019). As previously discussed in chapter 1, the AHR is involved in cell proliferation, inflammation and cell motility and migration (Sobus & Warren, 2014) and activates the xenobiotic metabolic system, including the activation of cytochrome P450 genes.

The CYP P450 genes encode enzymes involved in the metabolism of endogenous and exogenous compounds (Ingelman-sundberg, 2002). CYP1A1 and CYP1B1 are both CYP P450 genes involved in xenobiotic metabolism and were found to be overrepresented in both CSC and combination treated samples. CYP1A1 is involved in the metabolism of PAHs leading to the formation of carcinogenic intermediates. The function of CYP1B1 is similar to CYP1A1 as it functions to activate several environmental mutagens (such as PAHs), but it also catalyses reactions involving estrogens, which have an important physiological role, but may also be involved in hormonally driven carcinogenesis (Piotrowska et al., 2013). These genes are also involved in the formation of ROS, which has been shown to be enriched pathway in the RNA-Seq data analysis, as previously mentioned.

Other xenobiotic genes shown to be differentially expressed were glutathione S-transferases (GSTs) and uridine diphosphate glucuronosyltransferases (UGTs). GSTM3 and GSTM4 encode for GSTs, which are detoxification enzymes that catalyse the conjugation of glutathione to a variety of compounds, including products of oxidative stress and environmental carcinogens. This reaction renders the metabolites less reactive by neutralizing their electrophilic sites so they can be excreted from the body (Matejcic & Parker, 2015). UGTs are enzymes involved in the detoxification of various environmental carcinogens, including PAHs (Matejcic & Parker, 2015). UGT1A6 was found to be overrepresented in response to CSC and combination treatments in this study, but previously, it has been shown to be downregulated in oesophageal tumour tissues compared to healthy oesophageal tissue (Matejcic & Parker, 2015).

ALDH3A1 encodes for aldehyde dehydrogenase, an enzyme involved in alcohol metabolism, specifically the metabolization of acetaldehyde. It has previously been found to be involved in lung adenocarcinoma (Fan et al., 2021) and OSCC (Y. Y. Jiang et al., 2020), where it has been shown to be upregulated and promote tumour development. In OSCC, the findings was suggested as a possible link between alcohol consumption as a risk factor for the development of this disease (Y. Y. Jiang et al., 2020).

3.2.1 RT-qPCR Discussion

Five genes were selected from the initial list of differentially expressed genes resulting from the first RNA-Seq analysis, to be further investigated using RT-qPCR: SERPINB2, SH3PDX2A, GLIPR1, GPR39, FCMR and USP47. These genes were selected based on previous research which reported they may play an important role in the development of OSCC or other cancers.

SERPINB2 is a protein coding gene which codes for “Serpin Peptidase Inhibitor B2” protein. SERPINB2 was chosen as it has been shown to be indirectly regulated by the AhR, which has been shown to play a role in tumour formation and progression (Brauze et al., 2019). As previously discussed in chapter 1, the AhR is an important receptor involved in the link between tumour formation and xenobiotic metabolism. SERPINB2 was found to be upregulated in both CSC and combination treatments and had one of the largest fold changes out of the differentially regulated genes with a log2FoldChange of 4.73 in response to CSC and 4.48 in response to combination treatment. This was then confirmed by RT-qPCR to be upregulated in KYSE150, WHCO6, KYSE180 and SIHA cells. It was also the only gene in the first list of DEGs to be shown to have altered gene expression in re-analysis of the RNA-Seq data. In the second analysis, SERPINB2 was shown to be overexpressed in CSC and combination samples and this was reflected in all three biological replicates.

SH3PDX2A is an adaptor protein, the SH3 and PX domains are important domains in protein binding of the tks protein (previously known as fish protein) (Gianni et al., 2011). It has been shown to localise in podosomes- implicated in translocation, growth factor processing, cell adhesion, motility and invasiveness (Abram et al., 2003) and play a role in ROS generation (Gianni et al., 2011). SH3PDX2A was shown to be upregulated in cells treated with the combination of CSC and EtOH with a significant log2 Fold Change of 1.89. The upregulation of SH3PDX2A was then shown to upregulated in KYSE150 cells, WHCO6 cells, KYSE180, MDA-MB-213 and SIHA cells. However, with the exception of two biological replicates with the KYSE150 cells, this result was not reproduced in the biological replicates.

GLIPR1 is a gene which encodes for Glioma Pathogenesis-Related Protein 1. It has been shown to be of significance in cancer as it is a p53 target gene and a cancer drug (SB225002) target (De Vasconcellos et al., 2015). In osteosarcoma, GLIPR1 overexpression has been found to inhibit the proliferation, migration and invasion of cancer cells and induced the differentiation of cancer-initiating cells (Dong et al., 2016) and in prostate and bladder cancer cells GLIPR1 silencing has shown decrease ROS formation (De Vasconcellos et al., 2015). GLIPR1 has been found to be downregulated in human cancers (Dong et al., 2016) and in this study it was found to be downregulated in the RNA-Seq assay, with a log2FoldChange of -0.94 in cells treated with EtOH. This RNA-seq result was then observed by RT-qPCR, where GLIPR1 was downregulated in two OSCC cells, KYSE180 and WHCO6 and in one biological replicate of the cervical cancer cell line, SIHA.

GPR39 codes for a Ghrelin receptor family protein, it has strong tumorigenic ability and has been found to be overexpressed in ESCC (Xie et al., 2011) and oral squamous cell carcinoma (Y. Jiang et al., 2020). GPR39 has been shown to promote cell proliferation and enhance cell motility and invasiveness in ESCC cell lines , increase foci formation and colony formation in soft agar, and promote tumour formation in nude mice (Xie et al., 2011). In oral squamous cell carcinoma high GPR39 expression was associated with malignant progression and poor survival in patients, suggesting GPR39 as a therapeutic target and biomarker (Y. Jiang et al., 2020). These studies found higher expression of GPR39 to be associated with cancer, however the RNA-Seq study found GPR39 to be downregulated in cells treated with CSC and the combination of CSC and EtOH ($P<0.05$). Furthermore, when we tried to confirm the RNA-Seq result using RT-qPCR, the results displayed an up-regulation in response to CSC or no change in gene expression. These results suggests that GPR39 may be up-regulated in cancer and the RNA-seq analysis reported an inaccurate result.

USP47 encodes for a ubiquitin-specific peptidase (USP). USPs have been widely studied in relation to disease and are thought to be important in cancer progression (Young et al., 2019). It has been reported USP47 promotes colorectal cancer epithelial mesenchymal transition (EMT) and malignancy and has been shown to play a role in promoting gastric cancer growth (Young et al., 2019). USP47 was found to be differentially expressed in cells treated with CSC, it had a positive log2FoldChnage of 1.31. Unfortunately, this was unable to be confirmed by RT-qPCR as non-specific amplification was observed. However, the RNA-seq assay suggests this pathway may be affected by cigarette smoking and lead to an increase in tumorigenesis.

FCMR encodes for the Fc fragment of the IgM antibody receptor. As opposed to the other genes which were selected for further investigation because of their role in cancer development as reported by the literature, FCMR was chosen as it was shown to be upregulated by a large margin. There has, however, been a couple of studies to show FCMR involvement in tumour growth (Kubli et al., 2019), promoting cervical cancer (J. Jiang et al., 2020) and one study also showed the promoter region of FCMR had recurrent mutations in OSCC (Chang et al., 2017). In the initial RNA-Seq analysis FCMR was one of the most affected genes with a log2 fold change greater than 9 in both the CSC treated sample and combined CSC and EtOH treated samples, hence, it was expected to be highly up-regulated in the RT-qPCR assays as well. In the KYSE150 samples, it was confirmed to be up-regulated in all three treatments, and this was reflected in two biological replicates. An up-regulation of FCMR was also observed in the KYSE180 CSC treated cells, however this could not be confirmed in any biological replicates. In the other cell lines, more variation was observed.

Therefore, after compiling all the RT-qPCR data, it can be observed that although the technical replicates appear to be consistent the RT-qPCR results are highly variable among the biological replicates. In some cases, the genes appear to have been differentially expressed in the wrong direction and in most cases there is no statistically significant change in gene expression. This can be said for all genes investigated, except for SERPINB2, which was the only gene shown to be altered in the second RNA-Seq analysis. This provided further evidence that the analysis of the RNA-Seq data needed to be re-done and the second results are more accurate.

Chapter 4: Conclusions and Future Work

Oesophageal squamous cell carcinoma is an important disease of interest in South Africa as it has a high incidence and a relatively low survival rate. To investigate risk factors of OSCC to mitigate their effects, two risk factors shown to be involved in South African aetiology, cigarette smoking and alcohol consumption were investigated in this study. After finding and confirming concentrations of CSC and EtOH that do not cause cell death in treated cells compared to untreated cells, the effects of CSC and EtOH on gene expression were investigated using RNA-Seq and RT-qPCR.

The concentrations found to be sub-lethal dosages of CSC and EtOH were 40 µg/mL CSC and 50 mM EtOH. These concentrations showed a consistent pattern of not causing a decrease in cell proliferation across four oesophageal cell lines: WHCO1, WHCO6, KYSE150 and KYSE30. This was initially shown via an MTT assay, measuring cell proliferation as a proxy for cell death, after a 72-hour treatment incubation period. The results were then confirmed using a trypan blue assay, which measured living cell count as a percentage proportion of total living cells. This assay confirmed 40 µg/mL CSC and 50 mM EtOH does not cause cell death, for an increased period of six and ten days. These concentrations of CSC (40 µg/mL) and EtOH (50 mM) were then used in downstream gene expressions analysis. Further work researching the role of CSC and EtOH in cell death can investigate whether these factors cause apoptosis, necrosis or whether the factors inhibit apoptosis, and allow the cells to persist in conditions of increased exposure to potentially damaging compounds.

RNA-Seq analysis was then conducted on samples treated with CSC and EtOH, individually and in combination for three days, alongside an untreated sample that was supplemented with 0.2% DMSO, to determine genes differentially expressed in response to the respective treatments. The initial RNA-seq analysis found a total of 39 genes differentially expressed, 22 upregulated genes and 17 downregulated genes across treatment groups, while a secondary re-analysis of the RNA-Seq data found a total of 153 genes were found to be over or underrepresented with statistical significance, 107 genes were reported as being overrepresented and 46 genes were found to be underrepresented. Future work on these genes is needed to determine why they have been affected by the treatments and what pathways they individually affect. A “Gene Set Enrichment Analysis” found reactive oxygen species metabolic process and reactive nitrogen species metabolic process to be enriched in this dataset. Future work needs to be done to confirm that CSC and/or EtOH treatment causes an increase in ROS or nitrogen species in OSCC cells.

As seen in the results of both sets of RNA-Seq analysis there was substantial overlap between DEGs in CSC and combination treated cells and more genes were affected by cells exposed to CSC compared to

EtOH alone, indicating that the effect of alcohol was smaller compared to CSC. A subset of genes was also found to be altered across all treatments. It would be interesting to find out why some genes cluster together in some treatments and whether these genes affect the same pathway. Interestingly, it was not observed that more genes were affected by the combination treatments compared to CSC and EtOH alone, indicating the synergistic effects of the two factors did not have an increased effect in this study. Future work needs to be done to determine if changing treatment conditions, such as increased EtOH concentrations or longer incubation times, may increase the effects of combined CSC and EtOH treatment.

A caveat of the RNA expression analysis was that OSCC cancer lines were used for the experiments and not oesophageal normal cells. This is a drawback as cancer cell lines are already altered, therefore as we are comparing treated cancer cells to untreated cancer cells, there may be genes that can be affected by CSC and EtOH that we may have missed. It would be very useful to perform these experiments using an immortalised oesophageal squamous cell line. Such an immortalised cell line may allow the identification of genes which may be altered by CSC and EtOH which may be missed in a cancer cell line. However, immortalised cells may have significant differences from normal squamous cells. Another caveat of working with oesophageal cultured cancer cells is that there are differences to how CSC may naturally come into contact with oesophageal tissue. This lab model does not incorporate mucosal levels or differences that may occur as a result of being exposed to mainstream smoke or environmental smoke.

Genes from the first RNA-Seq data analysis were then further investigated using RT-qPCR, where a select number of genes which previous literature had reported to have a role in cancer were chosen for investigation. The genes selected were SERPINB2, GPR39, SH3PXD2A, GLIP1R, FCMR and USP47. Besides SERPINB2, none of these genes were found to have altered expression patters in the second RNA-Seq analysis and the RT-qPCR assays conducted could not find a consistent altered pattern of expression in three biological replicates. Further work on genes identified in the second RNA-Seq analysis needs to be done to confirm their altered expression in response to the treatments. It would also be interesting to include functional studies on these new genes, such as knockdown experiments to investigate their cellular functions and role in cancer development. As the second RNA-Seq analysis did identify xenobiotic genes differentially expressed in response to the treatments, it would be interesting to further investigate these specific genes and their role in OSCC. Additionally, it would be interesting to investigate if these genes are differentially expressed in patient tumours.

Future work investigating cigarette smoking and alcohol in OSCC can be to investigate the mutational signatures identified in response to CSC and EtOH treatment. Mutational signatures are a pattern of gene mutations associated with a mutagenic process and can be used to increase our understanding of the pathology of different cancers and can be used as potential biomarkers to identify early mutagenesis (Boot et al., 2018). For example, mutational signature 1 is associated with age of cancer diagnosis of the subjects and is found in all cancer types and most cancer tissues. As previously discussed there are epidemiological studies on the association between cigarette smoking and alcohol and an increased risk of developing OSCC (Wheeler & Reed, 2012) and multiple studies reporting the carcinogenic effects of smoking and alcohol on cells which contributes to tumour formation, however, the mutational signatures for smoke use has not been identified in populations at high risk for OSCC (Liu et al., 2017; Zhang et al., 2015). To further investigate smoke and alcohol as risk factors for OSCC, it may be useful to know that these factors cause their respective mutational signatures in oesophageal cells in response to treatment in a cell culture system.

In conclusion the aim of this study was to investigate the role of CSC and EtOH in OSCC; this was accomplished by identifying genes that have been affected by treatment with these risk factors. The list of genes found as a result of the RNA-Seq assay provides data that can be used as a base for future work investigating the role of CSC and EtOH in OSCC, relating to cellular function as well as xenobiotic metabolism.

Chapter 5: Materials and Methods

5.1 Cell Culture and Treatment

5.1.1 Cell lines

Five oesophageal cancer cell lines were used in this study: WHCO1, WHCO6, KYSE180 and KYSE150 and KYSE30; breast cancer cell line, MDA-MB-231 and cervical cell line, SiHa. The WHCO1 and WHCO6 cell lines were established from South African patients and were a gift from (Veale, 1984). The KYSE-150 (ACC 375), KYSE-180 (ACC379) and KYSE-30 (ACC 351) cells were obtained from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures. MDA-MB-231, (ATCC® HTB-26™) and SiHa cells (ATCC® HTB-35™)) came from the the American Type Culture Collection (ATCC). All cell culture work was conducted in a biosafety cabinet, that had been under ultraviolet (UV) light for 15-30minutes and subsequently sprayed with 70% ethanol.

5.1.2 Cell culture and maintenance

Cell lines were grown in 10cm plates in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat inactivated fetal bovine serum (HIFBS) and 1 U/ml penicillin and 1 µg/ml streptomycin (complete medium). Cells were maintained in a humidified environment of 5% CO_2 at 37˚C in a water-jacket incubator. The medium was replaced every 3 to 4 days, by aspirating the old medium and pipetting 10ml of fresh complete medium into the plate.

Cells were sub-cultured every 3 to 4 days, when the cells reached approximately 90% confluency. The media was aspirated, the plates briefly rinsed with 5ml of 0.05% trypsin-10 mM EDTA (trypsin-EDTA) in Phosphate Buffered Saline (PBS) and then fresh 5ml of trypsin-EDTA was added to the plated and allowed to incubate at 37˚C for 1-5 minutes. Cells were visualised under a microscope to ensure they had detached. Once the cells detached, the trypsin- was inactivated by adding 5 ml of complete medium to the plate. Cells were then pipetted into a 15 ml tube and pelleted by centrifugation at 4000 rpm for 1 minute, the cell pellet was then re-suspended in 5 ml of complete medium. Aliquots were then re-plated in new plates with 10 ml of fresh complete medium.

For the preparation of frozen stocks, the cells were aliquoted and re-suspended in 1 ml of complete medium. The cell suspension was then put into cryogenic tubes, 100 µl of DMSO was added and mixed by gently pipetting. The cell stocks were then stored in liquid nitrogen. This process was carried out on ice to keep the cells intact.

5.1.3 Cell Treatment

For all assays the cell lines were treated with a range of concentrations of cigarette smoke condensate (CSC) and ethanol (EtOH), individually and in combination. The CSC was purchased from Murty

Pharmaceuticals Inc, (Lexington, KY), produced from a research cigarette and solubilized in dimethylsulfoxide (DMSO) at a concentration of 40 mg/ml. The ethanol was obtained from Sigma-Aldrich.

Initially, for the MTT assay, CSC treatment was administered at a concentration range of 0-80 µg/ml, including concentrations of 0, 0.1, 5, 20, 30, 40 and 80 µg/ml CSC, as the literature reported 0-40 µg/ml to be physiological relevant (Assadollahi et al., 2019). For ethanol, a range of 10-100 mmol/l EtOH has been reported to be physiologically relevant and an acceptable range when studying acute alcohol exposure *in vitro* (Dolganiuc & Szado, 2009). Other studies have reported a range of 5-25 mM EtOH to be physiologically relevant and which yield carcinogenic effects for longer treatment times (Barnes et al., 2000). For this study a range of 0-50 mM EtOH was used to treat cells, including concentrations of: 0, 5, 10, 20, 25, 40 and 50 mM EtOH. To ensure all samples had a consistent level of DMSO exposure , non-treated cells were exposed to 0.2% DMSO when ran alongside it's treated counterparts. All CSC, EtOH and untreated preparations were diluted with DMEM.

5.2 Cytotoxicity

5.2.1 MTT Assay

To determine cytotoxicity an MTT cell proliferation assay was carried out on KYSE30, KYSE150, WHCO6 and WHCO1 cell lines. The cells were trypsinized as described in section 5.1.2, and the cells were counted using the automated cell counter, Countess II (Thermofischer Scientific). KYSE30 and KYSE150 cells were seeded at a concentration of 5000 cells per well and WHCO1 and WHCO6 cells were seeded at a concentration of 3000 cells per well in a 96-well tissue culture plate, in triplicate. Cells were incubated overnight in an incubator at 37°C and 5% CO_2. Cells were then treated with varying concentrations of cigarette smoke condensate (CSC) and ethanol, individually and in combination, as described in *section 5.1.3*, for 3 days, at 37°C and 5% CO_2. After the incubation period, 10 µl of the MTT labelling reagent (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) was added to each well and the plates were then placed back in the incubator for four hours. Thereafter 100 µl of the solubilization solution was added into each well and the plates were allowed to incubate overnight in the same conditions (37°C, 5% CO_2). The absorbance was then measured at 595 nm. Cell proliferation is then calculated as an average of the blank adjusted percentage of treated absorbance values relative to the untreated samples. Statistical significance was determined on Microsoft office, Excel using a student t-test.

5.2.2 Trypan Blue Assay

To confirm the results of the MTT assay a trypan blue assay was performed on KYSE150, WHCO6 and WHCO1 cells. Cell viability was determined as the percentage of living cells as a proportion of total cells. Living cells and dead cells can be differentiated as the dead cells take up the trypan blue dye and appear as blue under a microscope, while the living cells with intact cell membranes appear clear.

The cells were plated at a concentration of 20 000 cells per well in a 6-well plate in triplicate. Cells were then treated with 40 µg/ml CSC, 50 mM EtOH, individually and in combination, alongside untreated samples that were supplemented with 0.2% DMSO. Cells were allowed to incubate at 37°C and 5% CO_2 for 6 and 10 days. To determine cell death, the cells were trypsinized, pelleted by centrifugation and then resuspended in complete media. An aliquot of 50µl of cells was then mixed with equal volumes of Trypan Blue exclusion 0.4%, 0.85% NaCl (LONZA). 10ul of the cell suspension-trypan blue dye solution was placed on a Neubauer Improved Brightline haemocytometer (Superior) to determine the percentage of living cells as a fraction of total cells, i.e the cell viability (%). Four blocks of the haemocytometer was counted under a microscope (ZEISS Primovert) at 100X magnification and averaged. The number of living cells and total cells (living cells + dead cells) were then determined using the following equation:

Cell count = [(# of cells x 2)/0.0001] x dilution factor

Cell viability was then calculated as a percentage of average living cells as a fraction of average total cell count. Statistical significance was determined using a student t-test on Microsoft office, Excel.

5.3 Gene Expression Analysis

5.3.1 RNA Extraction

Following a 48-hour incubation RNA was extracted from cells which were treated in 10 cm plates with 40 µg/ml CSC, 50 mM EtOH, combined 40 µg/ml CSC and 50 mM EtOH and untreated cells supplemented with media containing 0.2% DMSO. The treated media was then aspirated out of the plate and 5 ml of PBS was added briefly, to rinse the plate. 1 ml of QIAzol Lysis Reagent (QIAGEN, Hilden, Germany) was added to each plate. A cell scraper was then used to lift and lyse the cells, and the cell-QIAzol solution was transferred to a 1.5 ml Eppendorf tube. 0.2 ml chloroform per 1 ml of QIAzol was added to the samples and they were inverted to mix and incubated for 10 minutes on ice. The samples were then separated into 3 phases by centrifuging at 8000rpm for 15 minutes at 4°C. The RNA present in the upper aqueous phase was then transferred to a new tube. Next, 0.5 ml isopropanol/ 1 ml QIAzol was added to the sample and mixed by inversion. The samples were then allowed to incubate overnight at -20°C to precipitate the RNA. After the overnight incubation, samples were centrifuged at 8000 rpm for 20 minutes at 4°C. The supernatant was removed and the RNA pellets were washed using 0.5 ml 75

% ethanol. After the ethanol was added, the pellet was centrifuged again at 8000 rpm for 10 min at 4°C. The ethanol was then aspirated and the RNA pellet was air-dried for 5 min on ice. The RNA pellet was then resuspended in 30 µl 0.01 % DEPC-containing dH2O. The RNA concentration and quality was then determined by measuring the concentration and 260/280 absorbance ratios with a Nanodrop® ND-1000 UV-Vis Spectrophotometer (Nanodrop technologies). RNA integrity was determined by visualizing 1.5 µg of RNA on a formaldehyde/agarose gel and running the gel with 1X MOPS buffer at 60 volts.

RNA was isolated from three biological replicates in five cell lines; three OSCC cell lines: KYSE150, WHCO6, KYSE180; a breast cancer cell line, MDA-MB-231 and a cervical cell line, SiHa.

5.3.2 Gene Expression: RNA Sequencing

A total of 12 RNA samples treated and extracted as previously described in section 5.3.1, from KYSE150 cells, were submitted to the Centre for Proteomic and Genomic Research (CPGR, Cape Town, South Africa) for RNA sequencing and bioinformatic analysis. At the CPGR the samples were re-analysed for quality control using the NanoDrop 8000 Spectrophotometer (ThermoFisher Scientific) to assess the purity and quality of the samples, Qubit® 4.0 Fluorometer for quantification and the Agilent TapeStation RNA ScreenTape (Agilent Technologies) to attain RIN values to assess RNA integrity.

Sequencing was done using the NEXTFLEX Rapid Directional RNA-Seq Kit 2.0 on the NextSeq 550 using a NextSeq 500/550 High Output Kit v2.5 (75 Cycles). The CPGR then carried out the bioinformatic analysis including quality control (QC), read alignment and expression analysis

5.3.2.1 Library Preparation pooling and QC

Poly(A) mRNA was purified from 5 µg total RNA for each sample using the NEXTFLEX Poly(A) Beads 2.0 (Perkin Elmer) according to the manufacturer’s instructions. The Poly(A) RNA was used for RNA-seq library preparation using the NEXTFLEX Rapid Directional RNA-Seq Kit 2.0 (Perkin Elmer) according to the manufacturer’s instructions. The Poly(A) RNA was fragmented and converted to cDNA. The cDNA was purified using NEXTFLEX Cleanup Beads XP, followed by adenylation and adapter ligation. The adapter ligated cDNA was purified by using NEXTFLEX Cleanup Beads XP and then amplified by PCR for 7 cycles. The libraries were purified using NEXTFLEX Cleanup Beads XP. The library size distribution for each library was determined using the Bioanalyzer High Sensitivity DNA Kit (Agilent). The purified RNA-seq libraries were quantified using the Qubit 1X dsDNA HS Assay Kit (ThermoFischer Scientific).

5.3.2.2 RNA-Seq library pooling and QC

The RNA-seq libraries were pooled in equimolar amounts for sequencing. The average fragment size for the pooled library was determined using the Bioanalyzer High Sensitivity DNA Kit. The concentration of adaptor-ligated DNA molecules was confirmed using the KAPA Library Quantification Kit for Illumina Platforms (Roche) on the QuantStudio 12K Flex System. The size adjusted library concentration was calculated using the KAPA Library Quantification Kit (Illumina Platforms) Data Analysis Template (v4.14).

5.3.2.3 Sequencing

The 4 nM pooled sequencing library pool was denatured using 0.2 N NaOH, diluted to 1.8 pM, and combined with the denatured PhiX positive control at a spike-in concentration of 1% v/v. The denatured library pool was loaded on the Illumina NextSeq 550 instrument and sequenced using one NextSeq 500/550 High Output kit v2.5 (75 cycles) (Illumina).

5.3.2.4 Bioinformatics

The Bioinformatics conducted by the CPGR included QC and sequence analysis. The output FastQ files were subject to QC using MultiQC and TrimGalore. HiSAT2 was then subsequently used to align the reads to the refence genome. The reads were then quantified using HTSeq and differential expression analysis was accomplished using DESeq2. The output was given as genes differentially expressed in treatment groups compared to the untreated sample. For the secondary re-analysis, the Salmon algorithm was used to align and quantify the sequencing reads used to determine the differential gene expression analysis. The raw reads were then imported without the need for the alignment step. Salmon then quantifies the transcript and the reads were then summarized to the gene level via the tximport package in R.

5.3.3 Gene Expression: RT-qPCR

5.3.3.1 cDNA Synthesis

RNA was converted to cDNA using the ImProm-II Reverse Transcription System (Promega). A total volume of 8 ul of RNA (2 µg) was added to 1 µl of random hexamer primers (Promega, USA). The reaction was then incubated at 70˚C for 10 min. A master mix containing the cDNA synthesis master mix reagents (Promega, USA) as seen in table 5.1 was added to the tubes to a final volume of 20 µl. The tubes were then placed in the PCR machine (GeneAmp PCR system 9700, Applied Biosystems) under the following cycling conditions: 42˚C for 2 hours and 70˚C for 10 minutes. To determine successful synthesis of cDNA, samples were run on a 1% (w/v) agarose gel at 80V for 40 min and visualised under UV. The cDNA samples were then stored at -20 ˚C. and used for gene expression analysis.

Table 5.1: cDNA master mix used to convert RNA to cDNA to be used in RT-qPCR gene expression analysis.

cDNA Master Mix	
5X Buffer	4ul
$MgCl_2$	1.6ul
dNTPs mix	2ul
RNAsin (RNase inhibitor)	0.5ul
Impromp II Reverse Transcriptase	1ul
DEPC-dH_2O	1.9ul

5.3.3.2 Primer Design

Primers used for RT-qPCR can be seen in table 5.2. The primers acquired from the literature were blasted against the NCBI database (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) to ensure they amplified the target gene of interest. The primers used to investigate the FCMR gene were designed using the primer-BLAST function on NCBI.

Table 5.2: Primers used for RT-qPCR gene expression analysis

Gene Primer	Primer Sequence	Amplicon length (bp)	Annealing T_A (°C)	Reference
SERPINB2_forward	5′-GAATGCTGTCTACTTCAA-3′	147	55°C	Brauze et al., 2017
SERPINB2_reverse	5′-TCTTCTATGTATCCAATGTT-3′			
GPR39_forward	5′-GCCACCGGGGTCTCACTTGC-3′	371	60°C	Xie et al., 2011
GPR39_reverse	5′-GGCCGCAGCCATGATCCTCC-3′			
SH3PXD2A_forward	5′-GGACCCCAAGCAAAGGATCAT-3′	115	55°C	Xiang et al. 2013
SH3PXD2A_reverse	5′-TGCCCGGCAGTATTCATCG-3′			
USP47_forward	5′-GGCAGGACGCTCATTAGGT-3′	113	60°C	de Vasconcellos et al., 2015
USP47_reverse	5′-GCACAACATGATTCCAAGTCAA-3′			
GLIPR1_forward	5′-AGCTGCACCCAAACTTCACT-3′	173	56°C	Yuan et al., 2021
GLIPR1_reverse	5′-ATCTGCCCAAACAACCTGAG-3′			
FCMR_forward	5′-CCACGGTACCACATGTTCCA-3′	161	55°C	Designed on NCBI
FCMR_reverse	5′-TCCTCTCCCTGCAGTATCGG-3′			
Control House-Keeping Gene				
GAPDH_forward	5′-GGCTCTCCAGAACATCATCC-3′	192	55°C - 60°C	
GAPDH_reverse	5′ -GCCTGCTTCACCACCTTC- 3′			

5.3.3.3.RT-qPCR

Quantitative Real-Time PCR was used to validate the expression of genes upregulated or down-regulated as found by the RNA-Seq. Genes were chosen based on previous literature investigating their role in tumorigenesis in OSCC or other cancers. These genes chosen were SERPINB2, GPR39, SH3PXD2A, USP47, GLIPR1 and FCMR.

RT-qPCR was performed on 12 RNA samples per cell line. CSC, EtOH and Combination treated samples and an untreated sample, as previously described. Three technical replicates of each treatment were run, in addition to the no template control (NTC). All samples were run alongside a house-keeping gene, GAPDH. The RT-qPCR was then repeated in two or three biological replicates of five cell lines; three OSCC cell lines: KYSE150, WHCO6, KYSE180; one breast cancer cell line, MDA-MB-231 and one cervical cell line, SiHa.

The RT-qPCR reaction was carried out using the KAPA SYBR fast master mix (KAPA Biosystems, South Africa). A master mix was made as per the manufactures instructions as seen in table 5.3, and 2 µl cDNA was added to the experimental reaction. The amplification was performed using the QuantStudio 3 (Applied Biosystems) with the following cycling parameters, based on the manufacturer's instruction: 95˚C for 3 min; 40 cycles of 95˚C for 10 sec, T_A (see table 5.3) for 20 sec, and 72˚C for 5 sec. A melt curve was then acquired using the parameters 95˚C for 5 seconds; 65˚C for 1 min, 97˚C for 7 min and finally a cooling step was included: 40˚C for 10 sec.

Table 5.3: Universal RT-qPCR master mix.

Component	Volume	Final Concentration
KAPA SYBR FAST qPCR Master Mix (2X) Universal	10 µL	1X
10 µM forward primer	0.4 ul	200 nM
10 µM reverse primer	0.4 µL	200 nM
PCR-grade water	Up to 20 µL	-

The RT-qPCR data was analysed using QuantStudio™ Design & Analysis Software (Applied Biosystems). Melt curves (seen in appendix) were used to ensure the primers were amplifying one target region. The relative fold gene expression levels was quantified using the $2^{-\Delta\Delta Cq}$ method. The fold change ratios are the averaged relative Ct values normalized to the internal reference gene, GAPDH. Statistical significance was determined using a student t-test on Microsoft office, Excel.

References

Abnet, C. C., Arnold, M., & Wei, W. (2018). *Epidemiology of Esophageal Squamous Cell Carcinoma*. 360–373. https://doi.org/10.1053/j.gastro.2017.08.023

Abram, C. L., Seals, D. F., Pass, I., Salinsky, D., Maurer, L., Roth, T. M., & Courtneidge, S. A. (2003). The adaptor protein fish associates with members of the ADAMs family and localizes to podosomes of Src-transformed cells. *Journal of Biological Chemistry*, *278*(19), 16844–16851. https://doi.org/10.1074/jbc.M300267200

Allam, E., Zhang, W., Al-shibani, N., Sun, J., Labban, N., Song, F., & Windsor, L. J. (2011). Effects of cigarette smoke condensate on oral squamous cell carcinoma cells. *Archives of Oral Biology*, *56*(10), 1154–1161. https://doi.org/10.1016/j.archoralbio.2011.03.008

Álvarez-Avellón, S. M., Fernández-Somoano, A., Navarrete-Munoz, E. M., Vioque, J., & Tardón, A. (2017). Effect of alcohol and its metabolites in lung cancer : CAPUA study. *Medicina Clinica*, *148*(12), 531–538. https://doi.org/10.1016/j.medcle.2016.12.072

Arnal, M. J. D., Arenas, Á. F., & Arbeloa, Á. L. (2015). *Esophageal cancer : Risk factors , screening and endoscopic treatment in Western and Eastern countries*. *21*(26), 7933–7943. https://doi.org/10.3748/wjg.v21.i26.7933

Arnold, M., Soerjomataram, I., Ferlay, J., & Forman, D. (2015). *Global incidence of oesophageal cancer by histological subtype in 2012*. 381–387. https://doi.org/10.1136/gutjnl-2014-308124

Asombang, A. W., Chishinga, N., Nkhoma, A., Chipaila, J., Nsokolo, B., Manda-Mapalo, M., Montiero, J. F. G., Banda, L., & Dua, K. S. (2019). Systematic review and meta-analysis of esophageal cancer in Africa: Epidemiology, risk factors, management and outcomes. *World Journal of Gastroenterology*, *25*(31), 4512–4533. https://doi.org/10.3748/wjg.v25.i31.4512

Assadollahi, V., Mohammadi, E., Fathi, F., Hassanzadeh, K., Bager Khadem Erfan, M., Soleimani, F., Banafshi, O., Yousefi, F., & Allahvaisi, O. (2019). Effects of cigarette smoke condensate on proliferation and pluripotency gene expression in mouse embryonic stem cells. *Journal of Cellular Biochemistry*, *120*, 4071–4080. https://doi.org/10.1002/jcb.27692

Barnes, S. L., Singletary, K. W., & Frey, R. (2000). Ethanol and acetaldehyde enhance benzo[a]pyrene-DNA adduct formation in human mammary epithelial cells. *Carcinogenesis*, *21*(11), 2123–2128. https://doi.org/10.1093/carcin/21.11.2123

Boffetta, P., & Hashibe, M. (2006). Alcohol and cancer. *Lancet Oncology*, *7*, 149–156.

Boot, A., Huang, M. N., Ng, A. W. T., Ho, S. C., Lim, J. Q., Kawakami, Y., Chayama, K., Teh, B. T., Nakagawa, H., & Rozen, S. G. (2018). In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Research*, *28*(5), 654–665. https://doi.org/10.1101/gr.230219.117

Brauze, D., Kiwerska, K., Bednarek, K., Grenman, R., Janiszewska, J., Giefing, M., & Jarmuz-Szymczak, M. (2019). Expression of Serpin Peptidase Inhibitor B2 (SERPINB2) is regulated by Aryl hydrocarbon receptor (AhR). *Chemico-Biological Interactions*, *309*(April). https://doi.org/10.1016/j.cbi.2019.06.013

Bray, F., Ferlay, J., & Soerjomataram, I. (2018). Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Brenner, D. R., Fehringer, G., Zhang, Z., Lee, Y. A., Meyers, T., Matsuo, K., Ito, H., Vineis, P., Stucker, I., Boffetta, P., Brennan, P., Christiani, D. C., Diao, N., Hong, Y., Landi, M. T., Morgenstern, H., Schwartz, A. G., Rennert, G., Saliba, W., … Straif, K. (2019). Alcohol consumption and lung cancer risk : A pooled analysis from the International Lung Cancer Consortium and the SYNERGY study. *Cancer Epidemiology*, *58*, 25–32. https://doi.org/10.1016/j.canep.2018.10.006

Bye, H., Prescott, N. J., Matejcic, M., Rose, E., Lewis, C. M., Parker, M. I., & Mathew, C. G. (2011). Population-specific genetic associations with oesophageal squamous cell carcinoma in South Africa. *Carcinogenesis*, *32*(12), 1855–1861. https://doi.org/10.1093/carcin/bgr211

Cao, Y., & Giovannucci, E. L. (2016). ALCOHOL AS A RISK FACTOR FOR CANCER. *Seminars in Oncology Nursing*, *32*(3), 325–331. https://doi.org/10.1016/j.soncn.2016.05.012

Chang-claude, J., Becher, H., Blettner, M., Songliang, Q., Guanrei, Y., & Wahrendorf, J. (1997). *Familial Aggregation of Oesophageal Cancer in a High Incidence Area in China*. *26*(6), 1159–1165.

Chang, J., Tan, W., Ling, Z., Xi, R., Shao, M., Chen, M., Luo, Y., Zhao, Y., Liu, Y., Huang, X., Xia, Y., Hu, J., Parker, J. S., Marron, D., Cui, Q., Peng, L., Chu, J., Li, H., Du, Z., … Lin, D. (2017). Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nature Communications*, *8*(17), 1–11. https://doi.org/10.1038/ncomms15290

Chen, J., Yakkundi, P., & Chan, W. K. (2019). Down-Regulation of p23 in Normal Lung Epithelial Cells Reduces Toxicities From Exposure to Benzo [a] pyrene and Cigarette Smoke Condensate via an Aryl Hydrocarbon Receptor-Dependent Mechanism. *Toxicological Sciences*, *167*(1), 239–248. https://doi.org/10.1093/toxsci/kfy234

Dandara, C., Ballo, R., & Parker, M. I. (2005). CYP3A5 genotypes and risk of oesophageal cancer in two South African populations. *Cancer Letters*, *225*(2), 275–282. https://doi.org/10.1016/j.canlet.2004.11.004

Dandara, C., Li, D. P., Walther, G., & Iqbal Parker, M. (2006). Gene-environment interaction: The role of SULT1A1 and CYP3A5 polymorphisms as risk modifiers for squamous cell carcinoma of the oesophagus. *Carcinogenesis*, *27*(4), 791–797. https://doi.org/10.1093/carcin/bgi257

De Vasconcellos, J. F., Laranjeira, A. B. A., Leal, P. C., Bhasin, M. K., Zenatti, P. P., Nunes, R. J., Yunes, R. A., Nowill, A. E., Libermann, T. A., Zerbini, L. F., & Yunes, J. A. (2015). SB225002 induces cell death and cell cycle arrest in acute lymphoblastic leukemia cells through the activation of GLIPR1. *PLoS ONE*, *10*(8), 1–19. https://doi.org/10.1371/journal.pone.0134783

DeMarini, D. M. (2004). Genotoxicity of tobacco smoke and tobacco smoke condensate : a review. *Mutation Research*, *567*, 447–474. https://doi.org/10.1016/j.mrrev.2004.02.001

Dolganiuc, A., & Szado, G. (2009). In vitro and in vivo models of acute alcohol exposure. *World Journal of Gastroenterology*, *15*(10), 1168–1177. https://doi.org/10.3748/wjg.15.1168

Dong, J., Bi, B., Zhang, L., & Gao, K. (2016). GLIPR1 inhibits the proliferation and induces the differentiation of cancer-initiating cells by regulating miR-16 in osteosarcoma. *Oncology Reports*, *36*(3), 1585–1591. https://doi.org/10.3892/or.2016.4949

Engel, L. S., Chow, W., Vaughan, T. L., Gammon, M. D., Risch, H. A., Stanford, J. L., Schoenberg, J. B., Mayne, S. T., Dubrow, R., Rotterdam, H., West, A. B., Blaser, M., Blot, W. J., Gail, M. H., & Fraumeni, J. F. (2003). Population Attributable Risks of Esophageal and Gastric Cancers AND. *Journal of the National Cancer Institute*, *95*(18), 1404–1413. https://doi.org/10.1093/jnci/djg047

Fan, F., Yin, R., Wang, L., Zhao, S., Lv, D., Yang, K., Geng, S., Yang, N., Zhang, X., & Wang, H. (2021).

ALDH3A1 driving tumor metastasis is mediated by p53/BAG1 in lung adenocarcinoma. *Journal of Cancer*, *12*(16), 4780–4790. https://doi.org/10.7150/JCA.58250

Fitzmaurice, C., Akinyemiju, T. F., Al Lami, F. H., Alam, T., Alizadeh-Navaei, R., Allen, C., Alsharif, U., Alvis-Guzman, N., Amini, E., Anderson, B. O., & Aremu, O. (2018). Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncology*, *4*(11), 1553–1568. https://doi.org/10.1001/jamaoncol.2018.2706

Gianni, D., DerMardirossian, C., & Bokoch, G. M. (2011). Direct interaction between Tks proteins and the N-terminal proline-rich region (PRR) of NoxA1 mediates Nox1-dependent ROS generation. *European Journal of Cell Biology*, *90*(2–3), 164–171. https://doi.org/10.1016/j.ejcb.2010.05.007

Haarmann-Stemmann, T., & Abel, J. (2006). The arylhydrocarbon receptor repressor (AhRR): structure, expression, and function. *Biological Chemistry*, *387*(9), 1195–1199.

Hecht, S. S. (2003). TOBACCO CARCINOGENS , THEIR BIOMARKERS AND TOBACCO-INDUCED CANCER. *Nature*, *3*(October), 733–744. https://doi.org/10.1038/nrc1190

Hendricks, D., & Parker, M. I. (2002). Oesophageal cancer in Africa. *IUBMB Life*, *53*(4–5), 263–268. https://doi.org/10.1080/15216540212643

Huang, Y., Chang, X., Lee, J., Cho, Y. G., Zhong, X., Park, I., Liu, J., Califano, J. A., Ratovitski, E. A., Sidransky, D., & Kim, M. S. (2011). Cigarette smoke induces promoter methylation of single-stranded DNA-binding protein 2 in human esophageal squamous cell carcinoma. *International Journal of Cancer*, *128*, 2261–2273. https://doi.org/10.1002/ijc.25569

Ingelman-sundberg, M. (2002). Polymorphism of cytochrome P450 and xenobiotic toxicity. *Toxicology*, *182*, 447–452.

Islami, F., Boffetta, P., Ren, J., Pedoeim, L., Khatib, D., & Kamangar, F. (2009). *High-temperature beverages and foods and esophageal cancer risk — A systematic review*. *524*, 491–524. https://doi.org/10.1002/ijc.24445

Jelski, W., Kozlowski, M., Laudanski, J., Niklinski, J., & Szmitkowski, M. (2009). The activity of class I, II, III, and IV Alcohol Dehydrogenase (ADH) Isoenzymes and Aldehyde Dehydrogenase (ALDH) in esophageal cancer. *Digestive Diseases and Sciences*, *54*(4), 725–730. https://doi.org/10.1007/s10620-008-0422-8

Jiang, J., Wu, R. H., Zhou, H. L., Li, Z. M., Kou, D., Deng, Z., Dong, M., & Chen, L. H. (2020). TGIF2 promotes cervical cancer metastasis by negatively regulating FCMR. *European Review for Medical and Pharmacological Sciences*, *24*(11), 5953–5962. https://doi.org/10.26355/eurrev_202006_21488

Jiang, Y., Li, T., Wu, Y., Xu, H., Xie, C., Dong, Y., Zhong, L., Wang, Z., Zhao, H., Zhou, Y., Li, J., Ji, N., Zeng, X., Feng, X., & Chen, Q. (2020). GPR39 Overexpression in OSCC Promotes YAP-Sustained Malignant Progression. *Journal of Dental Research*, *99*(8), 949–958. https://doi.org/10.1177/0022034520915877

Jiang, Y. Y., Jiang, Y., Li, C. Q., Zhang, Y., Dakle, P., Kaur, H., Deng, J. W., Lin, R. Y. T., Han, L., Xie, J. J., Yan, Y., Doan, N., Zheng, Y., Mayakonda, A., Hazawa, M., Xu, L., Li, Y. Y., Aswad, L., Jeitany, M., … Koeffler, H. P. (2020). TP63, SOX2, and KLF5 Establish a Core Regulatory Circuitry That Controls Epigenetic and Transcription Patterns in Esophageal Squamous Cell Carcinoma Cell Lines. *Gastroenterology*, *159*(4), 1311–1327. https://doi.org/10.1053/j.gastro.2020.06.050

Kier, L. D., Yamasaki, E., & Ames, B. N. (1974). Detection of Mutagenic Activity in Cigarette Smoke

Condensates. *Proceedings of the National Academy of Sciences*, *71*(10), 4159–4163.

Kikuchi, S., Yamada, D., Fukami, T., Maruyama, T., Ito, A., Asamura, H., Matsuno, Y., Onizuka, M., & Murakami, Y. (2005). Hypermethylation of the TSLC1 / IGSF4 Promoter Is Associated with Tobacco Smoking and a Poor Prognosis in Primary Nonsmall Cell Lung Carcinoma. *Cancer*, *106*(8), 1751–1758. https://doi.org/10.1002/cncr.21800

Koh, G., Zou, X., & Nik-Zainal, S. (2020). Mutational signatures: Experimental design and analytical framework. *Genome Biology*, *21*(1), 1–13. https://doi.org/10.1186/s13059-020-1951-5

Kubli, S. P., Vornholz, L., Duncan, G., Zhou, W., Ramachandran, P., Fortin, J., Cox, M., Han, S. J., Nechanitzky, R., Nechanitzky, D., Snow, B. E., Jones, L., Li, W. Y., Haight, J., Wakeham, A., Bray, M. R., & Mak, T. W. (2019). Fcmr regulates mononuclear phagocyte control of anti-tumor immunity. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10619-w

Lagergren, J., Bergstrom, R., Lindgren, A., & Nyren, O. (2000). The Role of Tobacco, Snuff and Alcohol Use in the Aetiology of Cancer of the Oesophagus and Gastric Cardia. *International Immunology*, *346*, 340–346.

Lee, K. W. K., & Pausova, Z. (2013). Cigarette smoking and DNA methylation. *Frontiers in Genetics*, *4*(132). https://doi.org/10.3389/fgene.2013.00132

Li, D., Dandara, C., & Parker, M. I. (2005). Association of cytochrome P450 2E1 genetic polymorphisms with squamous cell carcinoma of the oesophagus. *Clinical Chemistry and Laboratory Medicine*, *43*(4), 370–375. https://doi.org/10.1515/CCLM.2005.067

Li, D., Dandara, C., & Parker, M. I. (2010). The 341C/T polymorphism in the GSTP1 gene is associated with increased risk of oesophageal cancer. *BMC Genetics*, *11*(47). https://doi.org/10.1186/1471-2156-11-47

Li, P., Wu, W. K. K., Wong, H. P. S., Zhang, S. T., Yu, L., & Cho, C. H. (2009). Chloroform extract of cigarette smoke induces proliferation of human esophageal squamous-cell carcinoma cells : Modulation by ⍰ -adrenoceptors. *Drug and Chemical Toxicology*, *32*(3), 175–181. https://doi.org/10.1080/01480540902875253

Liu, W., Snell, J. M., Jeck, W. R., Hoadley, K. A., Wilkerson, M. D., Parker, J. S., Patel, N., Mlombe, Y. B., Mulima, G., Liomba, N. G., Wolf, L. L., Shores, C. G., Gopal, S., & Sharpless, N. E. (2017). Subtyping sub-Saharan esophageal squamous cell carcinoma by comprehensive molecular analysis. *JCI Insight*, *2*(22), 1–11. https://doi.org/10.1172/jci.insight.98457

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Loots, E., Sartorius, B., Madiba, T. E., Mulder, C. J. J., & Clarke, D. L. (2017). Oesophageal squamous cell cancer in a South African tertiary hospital: A risk factor and presentation analysis. *South African Journal of Surgery*, *55*(3), 42–46.

Lopes, C. F. B., de Angelis, B. B., Prudente, H. M., de Souza, B. V. G., Cardoso, S. V., Ribeiro, R. I. M. de A., & Azambuja, M. De. (2012). Concomitant consumption of marijuana , alcohol and tobacco in oral squamous cell carcinoma development and progression : Recent advances and challenges. *Archives of Oral Biology*, *57*, 1026–1033.

López-lázaro, M. (2016). A local mechanism by which alcohol consumption causes cancer. *Oral Oncology*, *62*, 149–152. https://doi.org/10.1016/j.oraloncology.2016.10.001

Lu, C., Xie, H., Wang, F., Shen, H., & Wang, J. (2011). Diet folate, DNA methylation and genetic polymorphisms of MTHFR C677T in association with the prognosis of esophageal squamous cell carcinoma. *BMC Cancer*, *11*(91). https://doi.org/10.1186/1471-2407-11-91

Malik, D. e. shahwa., David, R. M., & Gooderham, N. J. (2018). Ethanol potentiates the genotoxicity of the food-derived mammary carcinogen PhIP in human estrogen receptor-positive mammary cells: mechanistic support for lifestyle factors (cooked red meat and ethanol) associated with mammary cancer. *Archives of Toxicology*, *92*(4), 1639–1655. https://doi.org/10.1007/s00204-018-2160-9

Matejcic, M., Mathew, C. G., & Parker, M. I. (2019). The relationship between environmental exposure and genetic architecture of the 2q33 locus with esophageal cancer in South Africa. *Frontiers in Genetics*, *10*(MAY), 1–9. https://doi.org/10.3389/fgene.2019.00406

Matejcic, M., & Parker, M. I. (2015). Gene – environment interactions in esophageal cancer. *Critical Reviews in Clinical Laboratory Sciences*, *52*(5), 211–231. https://doi.org/10.3109/10408363.2015.1020358

Matejcic, M., Vogelsang, M., Wang, Y., & Parker, I. M. (2015). NAT1 and NAT2 genetic polymorphisms and environmental exposure as risk factors for oesophageal squamous cell carcinoma: A case-control study. *BMC Cancer*, *15*(150). https://doi.org/10.1186/s12885-015-1105-4

Matsha, T., Erasmus, R., Kafuko, A. B., Mugwanya, D., Stepien, A., & Parker, M. I. (2002). Human papillomavirus associated with oesophageal cancer. *Journal of Clinical Pathology*, *55*(8), 587–590. https://doi.org/10.1136/jcp.55.8.587

Mayosi, B. M., Flisher, A. J., Lalloo, U. G., Sitas, F., Tollman, S. M., & Bradshaw, D. (2009). The burden of non-communicable diseases in South Africa. *The Lancet*, *374*(9693), 934–947. https://doi.org/10.1016/S0140-6736(09)61087-4

Mcmenamin, Ú. C., Mccain, S., & Kunzmann, A. T. (2017). Do smoking and alcohol behaviours influence GI cancer survival? *Best Practice & Research Clinical Gastroenterology*, *31*(5), 569–577. https://doi.org/10.1016/j.bpg.2017.09.015

Middleton, D. R. S., Bouaoun, L., Hanisch, R., Bray, F., Dzamalala, C., Chasimpha, S., Menya, D., Gombé, C., Da, G. N., Woldegeorgis, M. A., Njie, R., Koulibaly, M., Buziba, N., Ferro, J., Nouhou, H., Ogunbiyi, F., Wabinga, H. R., Chokunonga, E., Borok, M. Z., … Mccormack, V. A. (2019). Esophageal cancer male to female incidence ratios in Africa : A systematic review and meta-analysis of geographic , time and age trends. *Cancer Epidemiology*, *53*(November 2017), 119–128. https://doi.org/10.1016/j.canep.2018.01.020

Mlombe, Y. B., Rosenberg, N. E., Wolf, L. L., Shaheen, N. J., Hosseinipour, M. C., & Shores, C. G. (2015). *Environmental risk factors for oesophageal cancer in Malawi : A case-control study*. *27*(September), 88–92.

Mohapatra, P., Preet, R., Das, D., Satapathy, S. R., Siddharth, S., Choudhuri, T., Wyatt, M. D., & Kundu, C. N. (2014). The contribution of heavy metals in cigarette smoke condensate to malignant transformation of breast epithelial cells and in vivo initiation of neoplasia through induction of a PI3K – AKT – NF κ B cascade. *Toxicology and Applied Pharmacology*, *274*(1), 168–179. https://doi.org/10.1016/j.taap.2013.09.028

Moodie, F. M., Matwick, J. A., Anderson, C. S., Patryk, S., Biswas, S. K., Bauter, M. R., Kilty, I., & Rahman, I. (2005). Oxidative stress and cigarette smoke alter chromatin remodeling but differentially regulate NF-kappaB activation and proinflammatory cytokine release in alveolar epithelial cells. *The FASEB Journal*, *18*(15), 1897–1899. https://doi.org/10.1096/fj.04-1506fje

Nagaraj, N. S., Beckers, S., Mensah, J. K., Waigel, S., Vigneswaran, N., & Zacharias, W. (2006). Cigarette smoke condensate induces cytochromes P450 and aldo-keto reductases in oral cancer cells. *Toxicology Letters*, *165*, 182–194. https://doi.org/10.1016/j.toxlet.2006.03.008

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature Methods*, *14*(4). https://doi.org/10.5040/9781635577068-1566

Piotrowska, H., Kucinska, M., & Murias, M. (2013). Expression of CYP1A1, CYP1B1 and MnSOD in a panel of human cancer cell lines. *Molecular and Cellular Biochemistry*, *383*(1–2), 95–102. https://doi.org/10.1007/s11010-013-1758-8

Qin, H., Liao, X., Chen, Y., Huang, S., Xue, W., Li, F., Ge, X., Liu, D., Cai, Q., Long, J., Li, X., Hu, Y., Zhang, S., Zhang, L., Lehrman, B., Scott, A. F., Lin, D., Zeng, Y., & Shugart, Y. Y. (2016). *Genomic Characterization of Esophageal Squamous Cell Carcinoma Reveals Critical Genes Underlying Tumorigenesis and Poor Prognosis*. 709–727. https://doi.org/10.1016/j.ajhg.2016.02.021

Rao, P. S. S., Ande, A., Sinha, N., Kumar, A., & Kumar, S. (2016). Effects of Cigarette Smoke Condensate on Oxidative Stress , Apoptotic Cell Death , and HIV Replication in Human Monocytic Cells. *PloS One*, *11*(5). https://doi.org/10.1371/journal.pone.0155791

Scoccianti, C., Lauby-secretan, B., Bello, P., Chajes, V., & Romieu, I. (2014). Female Breast Cancer and Alcohol Consumption. *American Journal of Preventive Medicine*, *46*(3), S16–S25. https://doi.org/10.1016/j.amepre.2013.10.031

Seitz, H. K., & Stickel, F. (2007). Molecular mechanisms of alcohol-mediated carcinogenesis. *Nature Reviews Cancer*, *7*(8), 599–612. https://doi.org/10.1038/nrc2191

Sewram, V., Sitas, F., Connell, D. O., & Myers, J. (2016). Tobacco and alcohol as risk factors for oesophageal cancer in a high incidence area in South Africa. *Cancer Epidemiology*, *41*, 113–121. https://doi.org/10.1016/j.canep.2016.02.001

Sobus, S. L., & Warren, G. W. (2014). The Biologic Effects of Cigarette Smoke on Cancer Cells. *Cancer*, *120*, 3617–3626. https://doi.org/10.1002/cncr.28904

Somdyala, N. I. M., Parkin, D. M., Sithole, N., & Bradshaw, D. (2015). Trends in cancer incidence in rural Eastern Cape Province; South Africa, 1998-2012. *International Journal of Cancer*, *136*(5), E470–E474. https://doi.org/10.1002/ijc.29224

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Talukdar, F. R., Di Pietro, M., Secrier, M., Moehler, M., Goepfert, K., Lima, S. S. C., Pinto, L. F. R., Hendricks, D., Parker, M. I., & Herceg, Z. (2018). Molecular landscape of esophageal cancer: Implications for early detection and personalized therapy. *Annals of the New York Academy of Sciences*, *1434*(1), 342–359. https://doi.org/10.1111/nyas.13876

Van Helden, P. D., Beyers, A. D., Bester, A. J., & Jaskiewicz, K. (1987). Esophageal Cancer: Vitamin and Lipotrope Deficiencies in an At-Risk South African Population. *Nutrition and Cancer*, *10*(4), 247–

255. https://doi.org/10.1080/01635588709513962

Vaz, M., Hwang, S. Y., Kagiampakis, I., Phallen, J., Patil, A., O'Hagan, H. M., Murphy, L., Zahnow, C. A., Gabrielson, E., Velculescu, V. E., Easwaran, H. P., & Baylin, S. B. (2017). Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations. *Cancer Cell*, *32*(3), 360-376.e6. https://doi.org/10.1016/j.ccell.2017.08.006

Veale, A. (1984). Atypical cytokeratins synthesized by human oesophageal carcinoma cells in culture. *South African Journal of Science*, *80*(6).

Wang, T., Hsia, S., Shih, Y., & Shieh, T. (2017). Association of Smoking , Alcohol Use , and Betel Quid Chewing with Epigenetic Aberrations in Cancers. *International Journal of Molecular Sciences*, *18*(6). https://doi.org/10.3390/ijms18061210

Wang, Y., Xu, M., Ke, Z., & Luo, J. (2017). Cellular and molecular mechanisms underlying alcohol-induced aggressiveness of breast cancer. *Pharmacological Research*, *115*, 299–308. https://doi.org/10.1016/j.phrs.2016.12.005

Wheeler, J. B., & Reed, C. E. (2012). Epidemiology of Esophageal Cancer. *Surgical Clinics of North America*, *92*(5), 1077–1087. https://doi.org/10.1016/j.suc.2012.07.008

Xi, S., Xu, H., Shan, J., Tao, Y., Hong, J. A., Inchauste, S., Zhang, M., Kunst, T. F., Mercedes, L., & Schrump, D. S. (2013). Cigarette smoke mediates epigenetic repression of miR-487b during pulmonary carcinogenesis. *The Journal of Clinical Investigation*, *123*(3), 1241–1261. https://doi.org/10.1172/JCI61271.Thejournal

Xie, F., Liu, H., Zhu, Y.-H., Qin, Y.-R., Dai, Y., Zeng, T., Chen, L., Nie, C., Tang, H., Li, Y., Fu, L., & Guan, X.-Y. (2011). Overexpression of GPR39 contributes to malignant development of human esophageal squamous cell carcinoma. *BMC Cancer*, *11*(86), 1471–2407. https://doi.org/10.1186/1471-2407-11-86

Yadav, S. K., Sharma, S. K., Farooque, A., Kaushik, G., Kaur, B., Pathak, C. M., Dwarakanath, B. S., & Khanduja, K. L. (2016). Cytosolic phospholipase A 2 (cPLA 2) IVA as a potential signature molecule in cigarette smoke condensate induced pathologies in alveolar epithelial lineages. *Lipids in Health and Disease*, *15*(129), 1–9. https://doi.org/10.1186/s12944-016-0300-x

Young, M. J., Hsu, K. C., Lin, T. E., Chang, W. C., & Hung, J. J. (2019). The role of ubiquitin-specific peptidases in cancer progression. *Journal of Biomedical Science*, *26*(42). https://doi.org/10.1186/s12929-019-0522-0

Zakhari, S. (2015). Chronic alcohol drinking : Liver and pancreatic cancer ? *Clinics Research in Hepatology and Gastroenterology*, *39*, 86–91.

Zambon, P., Talamini, R., La Vecchia, C., Dal Maso, L., Negri, E., Tognazzo, S., Simonato, L., & Franceschi, S. (2000). Smoking , Type of Alcoholic Beverage and Squamous-Cell Oesophageal Cancer in Northern Italy. *International Immunology*, *149*, 144–149.

Zeng, J.-H., Xiong, D.-D., Pang, Y.-Y., Zhang, Y., Tang, R.-X., Luo, D.-Z., & Chen, G. (2017). Identification of molecular targets for esophageal carcinoma diagnosis using miRNA-seq and RNA-seq data from The Cancer Genome Atlas: a study of 187 cases. *Oncotarget*, *8*(22), 35681–35699. https://doi.org/10.18632/oncotarget.16051

Zhang, L., Zhou, Y., Cheng, C., Cui, H., Cheng, L., Kong, P., Wang, J., Li, Y., Chen, W., Song, B., Wang, F., Jia, Z., Li, L., Li, Y., Yang, B., Liu, J., Shi, R., Bi, Y., Zhang, Y., … Zhang, X. (2015). Genomic Analyses Reveal Mutational Signatures and Frequently Altered Genes in Esophageal Squamous Cell Carcinoma. *The American Journal of Human Genetics*, *96*(4), 597–611.

https://doi.org/10.1016/j.ajhg.2015.02.017

Appendices

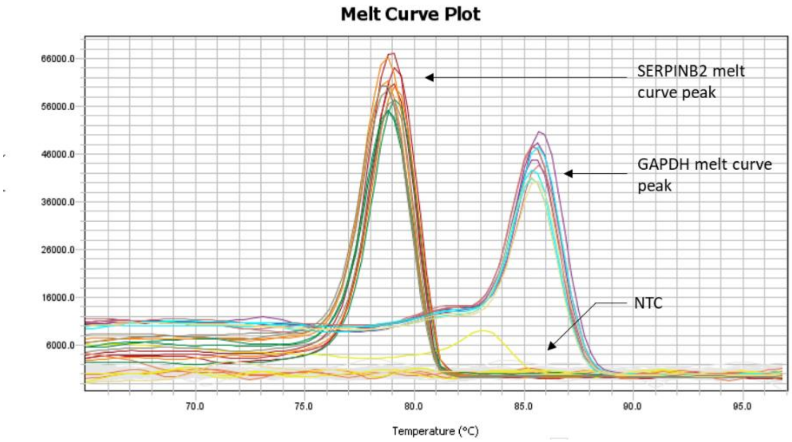

Figure A5: Melt curve analysis of SERPINB2 gene, alongside GAPDH reference gene.

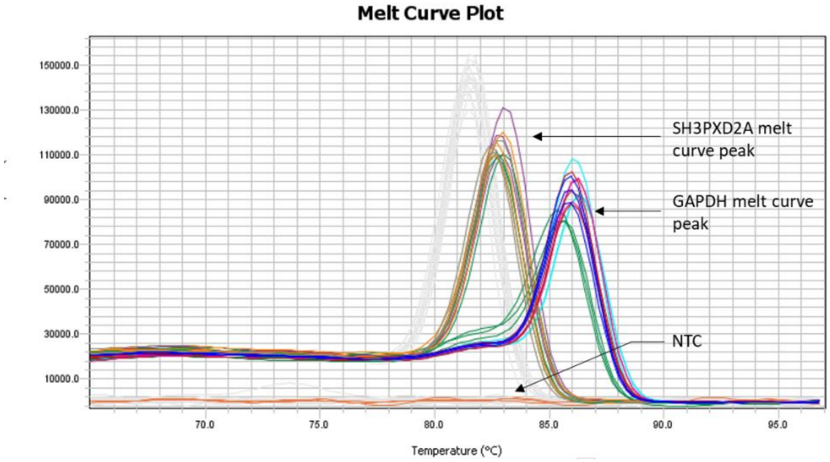

Figure A2: Melt curve analysis of SH3PXD2A gene, alongside GAPDH reference gene.

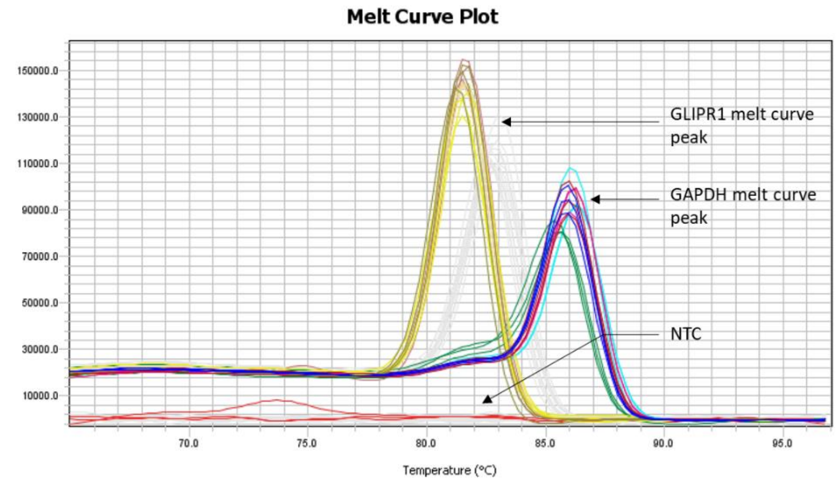

Figure A3: Melt curve analysis of GLIPR1 gene, alongside GAPDH reference gene.

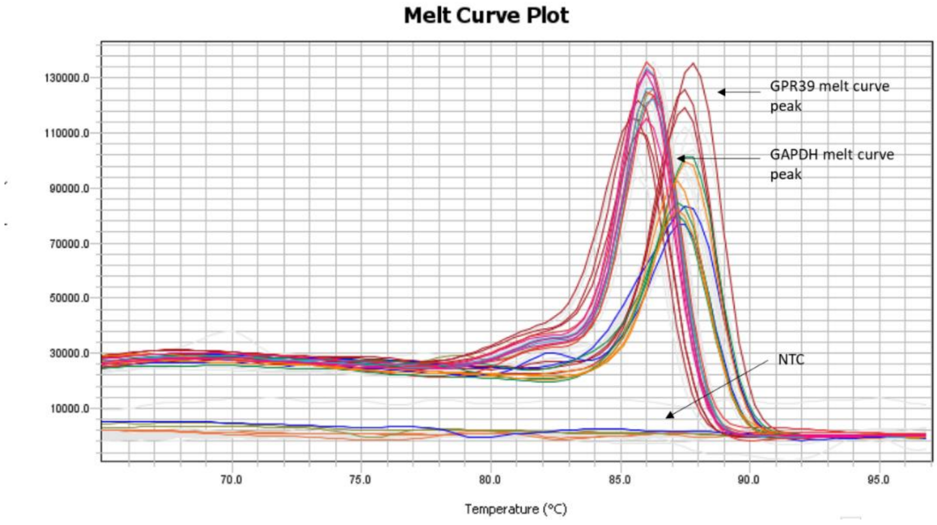

Figure A4: Melt curve analysis of GPR39 gene, alongside GAPDH reference gene.

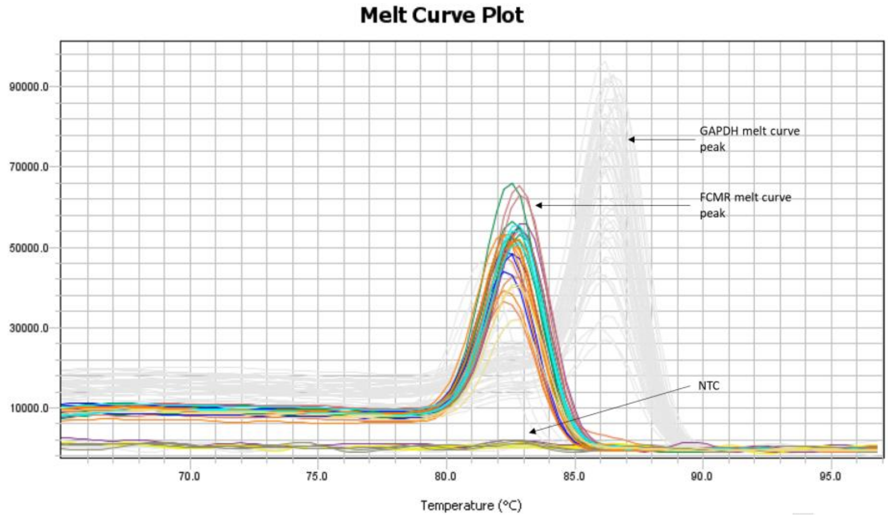

Figure A5:Melt curve analysis of FCMR gene, alongside GAPDH reference gene.