Institute of Systems, Molecular and Integrative Biology

Department of Molecular and Clinical Cancer Medicine

Plasma Extracellular Vesicle MicroRNA Biomarkers for Lung Cancer Diagnosis

Thesis submitted in accordance with the requirements of the

University of Liverpool for the degree of Doctor in Philosophy by

Xiaolei Yang

June 2022

Supervisors: Dr. Triantafillos Liloglou, Prof. John K Field and Dr. Michael Davies

Declaration

I, Xiaolei Yang, confirm that the research included within this thesis is my own work, or that where it has been carried out in collaboration, or supported by others, this is duly acknowledged. I declare that I have exercised reasonable care to ensure that the work is original and does not, to the best of my knowledge, break any UK law, infringe any third party copyright or other intellectual property rights. I accept that the University of Liverpool has the right to use plagiarism detection software to check the electronic version of this thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this, or any other university.

Name: Xiaolei Yang

Signature: Xiaolei Yang

Date: 23/06/2022

Scientific Acknowledgements

From the bottom of my heart, I am grateful to all of those who have helped me throughout my PhD programme.

First and foremost, I would like to express my utmost gratitude and respect to my supervisory team, Dr. Triantafillos Liloglou, Prof. John K Field and Dr. Michael Davies, for allowing me this opportunity and being fantastic role models. I feel so lucky and thankful that I have had the privilege of working alongside such supportive, insightful, and dedicated supervisors. They continuously inspired me through their immense scholarly knowledge and skills, and their invaluable experience, innovative ideas, and modest attitude never cease to amaze me even to this day.

I wish to express my deepest gratitude to Dr. Triantafillos Liloglou for his unstinting support throughout my doctoral program, especially for his guidance and mentorship in all aspects of my research and the writing of this thesis. He has always been modest, approachable and willing to listen and help, and I am truly grateful for his understanding, encouragement, and care, which helped me get through some tough times during my PhD.

I am also profoundly grateful to Dr. Michael Davies for his continuous guidance and support of my PhD study. Dr. Michael Davies has helped me enormously in the writing of this thesis with his vast knowledge, specialist expertise, especially in the field of statistics and his unfailing ability to explain clearly, as well as his meticulous guidance and tireless patience. Through his supervision, two principles were imparted upon me, which were that if a thing was worth doing, it is worth doing well and to think logically.

My sincere thanks also go to Prof. John K Field, whose optimism, positivity, and conscientious attitude I especially admire. My PhD project was conducted based on the outcomes of a pilot study of the Liverpool Lung Project (LLP), which is led by Prof. John K Field. The implementation of this pioneering translational research would not have been possible without his support and guidance. I would like to express my gratitude to the wider LLP team for their tremendous help and support, as well as access to the biorepository. Special thanks also to Mr Nial Hodge for his support in LLP data collection and reduction and Dr. Michael Marcus for his great support in the original statistical analysis of the HTG EV miRNA data.

My supervisory team has taken care of my wellbeing, especially during the pandemic. Aside from my research, we discussed a variety of issues and topics in life from which I learnt the art of looking at things as an “optimist”. Despite not everything being easy, my supervisors

never gave up on me and always maintained their confidence in my ability to complete the PhD. Every time I felt like things were getting on top of me, they taught me that every day is a new beginning. I believe another good quality I learned from my supervisors, which will benefit me for a lifetime, is how to “keep calm and carry on” at work and in day-to-day life.

My heartfelt gratitude goes to our lab technician, Stephanie Tate, for her unfailing support and help in my experiments (when one pair of hands was simply not enough) and other aspects of my research project, both technical and through her unstinting moral and personal support. Stephanie has helped me with EV isolation and RNA extraction procedures and qRT-PCR, as it was a great amount of repetitive work which benefitted from a team approach to ensure quality control. Stephanie has become a close friend; she has always been by my side, selflessly encouraging me and supporting me. I will always treasure the compassion and kindness she has shown me.

I am grateful to the Roy Castle Lung Cancer Foundation (RCLCF) for the grant funding to perform the discovery phase of the project and their long-term support for the LLP, without which samples and data would not have been available. I would also like to express my sincere appreciation to all the lung cancer patients and control subjects who very kindly consented to donate their blood samples to the LLP Biobank and made their clinical data available, making it possible to conduct pioneering translational research.

My sincere gratitude also goes to Dr. Ghaliah Obaid F Alnefaie, Dr. Amelia Acha-Sagredo, Dr. Caroline McCarthy, Dr. Yue Cheng, Dr. Alda Cunha, Dr. Yamini Krishna and Dr. Sophie Thornton for their encouragement, support and help during my PhD study and the great time we spent together, which I will forever treasure.

I am grateful to the Institute of Systems, Molecular and Integrative Biology (ISMIB), University of Liverpool, especially for the support from the PGR administration team. My sincere thanks also go to the International Advice and Guidance (IAG) team, the Wellbeing Advice & Guidance team and the Student Administration Centre for their professionalism and continued support during both my MRes and PhD studies at the University of Liverpool. Thanks are due to them all for giving me such a memorable student experience at the University of Liverpool.

Personal Acknowledgements

On a personal level, I would like to thank a few people who have been an enormous source of support and encouragement for me in everything I have attempted, and I am deeply blessed and grateful to have had them by my side.

Brian McMinn, who serves as the Senior Counsellor at the University of Liverpool, I am thankful for his inspirational wellbeing sessions, through which I gradually regained my confidence and started to understand and accept things from various perspectives. Dr James Reilly, a Clinical Psychologist at the Student Health Centre. Words of thanks are not enough to express my appreciation for all his support. He has provided me with regular sessions over the past few years despite the pandemic and has always been extremely understanding, considerate and patient. Thanks also to Dr Abbas, GP at Brownlow Health for whose support and kind words I am indebted.

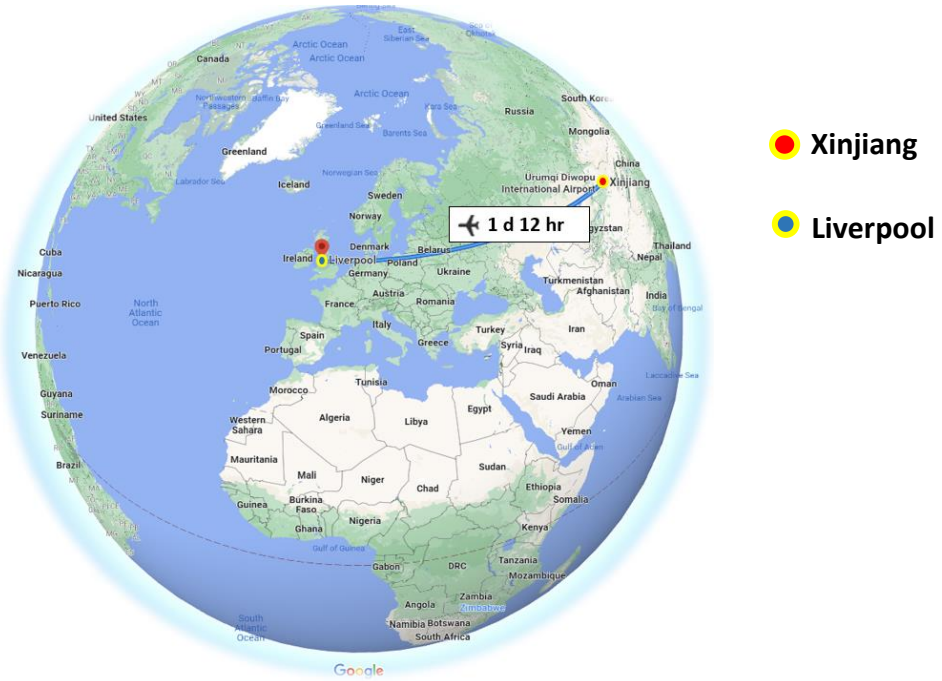
I thank my friends who are all so precious to me; Nana, we have been friends and confidantes for over 25 years. We are witnesses to our lives, the ups and downs and comings and goings. Even though you are 4,000 miles away, you have always been my lighthouse. You trust me more than I trust myself, even when I lose faith in myself. I can always cool down and bounce back after hearing you say, “You will get through this.” Hope to meet you again soon in our hometown and tell you I could not have come this far without your support in person with a big hug. Dr. Thompson Gana, Dr. Richard Malcolm Douglas Greensmith, and Dr. Abdelhafid Masoud, I am truly blessed to have such amazing friends like you, you have made my life in Liverpool a lot more enjoyable and fulfilled. Roland Liu, Johnny Jiang and James Wang, you have always believed in me and been there for me, I can't thank you enough for all that you have done for me.

Six years ago, I arrived in the UK after travelling for over 30 hours, it was my birthday, and I recall it rained heavily. I still remember it as if it were yesterday. For a number of reasons, I went back to China to visit my parents only once in these years, when my mum suffered a spine injury as a result of an accident. As an only child, being away from them for such a long time has not been easy for us, with it being much more difficult for them, especially given the family bonds of Chinese culture. I am deeply indebted to them for their unending support, whilst also having a sense of regret compounded by a shared family shyness that makes it harder for us to share our feelings. I once came across a piece of Chinese poetry that particularly struck a chord with me, "when children leave to pursue their dreams, parents

watch them leave with solicitousness and hope and pray they come home safe and sound."

Upon reading this poem, I was moved to tears.

Mum and Dad, I know it must be hard for any parent when their daughter flies the nest, I want you to know that, in my way, I love and appreciate you very much.



My physical and intellectual journey in pursuit of knowledge and self-discovery

To my family and friends, I dedicate this thesis in token of my love and gratitude.

谨以此博士论文献给我的家人和朋友，以表达对他们的爱和感激。

Abstract

Plasma Extracellular Vesicle MicroRNA Biomarkers for Lung Cancer Diagnosis

Xiaolei Yang

Background Globally, lung cancer represents the most common cause of cancer-related death. The high mortality rate of lung cancer is attributed mainly to the late stage of the disease at diagnosis. Blood-based biomarkers provide a method for early detection, but any biomarkers need to be rigorously validated at a pre-clinical and clinical level. Extracellular vesicles (EVs) are nanosized functional vacuolar structures released by cells, which play a pivotal role in intercellular communication. They are released in plasma, where they can be measured in a minimally invasive manner. It is hypothesized that a panel of EV miRNAs may provide a sensitive and specific biomarker to detect lung cancer at an early stage.

Aims and objectives This thesis aims to identify panels of miRNAs in the human plasma EV fraction that may serve as biomarkers for early diagnosis of lung cancer. The specific objectives include: (a) qPCR validation of a panel of 18 pre-selected miRNAs, identified by logistic regression analysis; (b) further analysis of the EV fraction data by alternative statistical approaches; and (c) comparison of miRNA expression between whole plasma and the EV fraction.

Methods The HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA) was used to measure the expression of 2,083 human miRNA transcripts from (a) plasma EV fractions of 60 cases and 60 controls and (b) plasma samples of 26 cases and 24 controls. Differential expression was assessed using HTG EdgeSeq Reveal data analysis suite. For validation, the exoEasy Midi Kit (Qiagen) was used to isolate EVs from a 2ml plasma sample. In total, EVs were extracted from 188 cases and 187 age/sex matched control plasma samples. Direct-zol RNA Miniprep Kit (Zymo Research) was used for EV RNA isolation. The TaqMan Advanced miRNA cDNA Synthesis Kit (Applied Biosystems) was used to reverse transcribe miRNA. qPCR, with miRCURY LNA Probe PCR Kits (Qiagen) and TaqMan™ Advanced miRNA Assays (Applied Biosystems), was used to analyse miRNA expression.

Results Validation of the 18 miRNAs previously identified by qPCR demonstrated that only a limited number of miRNAs were differentially expressed when applying an alternative EV isolation method and measurement technology. Reanalysis of the HTG EV miRNA data using updated quality control and alternative statistical methods identified batch effects for EV isolation and revealed alternative differentially expressed miRNAs. Differential expression of miRNAs in whole plasma was more robust than in EVs, with higher fold-changes and greater statistical significance (despite a smaller sample size). Comparison to whole plasma miRNA profiles identified both enrichment and depletion of miRNAs in EVs.

Conclusions We have shown that some miRNAs, in both EV and whole plasma, are up- or down-regulated in lung cancer. Furthermore, we could identify miRNAs differentially expressed in relation to other characteristics, such as COPD, smoking status and sex. The results clearly demonstrate the impact of EV isolation techniques and statistical analysis methods in the selection and validation of miRNAs. Despite a clear biological relevance to EV miRNA expression and the demonstration of differential enrichment of miRNAs, there are clear technological challenges to identification and validation of EV miRNA expression patterns. Plasma and EV miRNAs show great promise as clinically useful biomarkers, but great care must be taken to overcome the practical challenges of assay reproducibility and utility.

Table of Contents

List of Figures

List of Tables

Abbreviations

3D	Three-dimensional
3D-CRT	Three-dimensional conformal radiation therapy
AAMR	Age-adjusted lung cancer mortality rate
ACS	American Cancer Society
ADC	Adenocarcinoma
AJCC	American Joint Committee on Cancer
ALK	Anaplastic lymphoma kinase
ASCO	American Society of Clinical Oncology
ASRs	Age-standardized rates
AUC	Area under the ROC Curve
AYAs	Adolescents and young adults
Bap	Benzo(a)pyrene
BEC	Bronchial epithelial cells
BMs	Brain metastases
CAT	Computerized axial tomography
CCO	Cancer Care Ontario
CEA	Carcinoembryonic antigen
CFR	Case fatality rate
Chemo	Chemotherapy
CHART	Continuous hyperfractionated accelerated radiotherapy
ChT	Chemotherapy
circRNAs	Circular RNAs
CIS	Carcinoma in situ
CMT	Combined-modality therapy
CNB	Core-needle biopsy
CNS	Central nervous system
CO	Carbon monoxide
COPD	Chronic obstructive pulmonary disease
CPGs	Clinical Practice Guidelines
CPM	Counts per Million
CRT	Chemoradiotherapy
CRUK	Cancer Research UK
CSF	Cerebrospinal fluid
CT	Computed tomography
Ct	Cycle threshold
CTCs	Circulating tumour cells
ctDNA	Circulating tumour DNA
CTECs	Circulating tumour vascular endothelial cells
CXR	Chest x-ray
DLCST	Danish lung cancer screening trial
EBC	Exhaled breath condensate
EBRT	External beam radiation therapy
EBUS	Endobronchial ultrasound
EBUS-TBNA	Endobronchial ultrasound-guided transbronchial needle aspiration

ECOG	Eastern Cooperative Oncology Group
ECS	Extracapsular spread
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
EPA	Environmental Protection Agency
ESCRT	Endosomal sorting complexes required for transport
ESMO	European Society for Medical Oncology
ES-SCLC	Extensive-stage SCLC
ETS	Environmental tobacco smoke
EU	European Union
EVs	Extracellular vesicles
FDA	Food and Drug Administration
FDR	False discovery rate
FGF2	Fibroblast growth factor 2
FNAB	Fine-needle aspiration biopsy
GAS5	Growth arrest specific transcript 5
GF	Growth fraction
GPCRs	G protein-coupled receptors
HKGs	Housekeeping genes
HN2	Nitrogen mustard
IARC	International Agency of Research on Cancer
IASLC	International association for the study of lung cancer
ICIs	Immune checkpoint inhibitors
ILVs	Intraluminal vesicles
IMRT	Intensity-modulated radiation therapy
IO	Immuno-oncology
IPNs	Indeterminate pulmonary nodules
IV	Intravenous
KRAS	Kirsten rat sarcoma virus
LA-NSCLC	Locally advanced NSCLC
LASSO	Least Absolute Shrinkage and Selection Operator
LCC	Large cell carcinoma
LDCT	Low-dose computed tomography
LEO	Low-Earth orbit
LLP	Liverpool Lung Project
LNA	Locked nucleic acid
lncRNAs	Long non-coding RNAs
LS-SCLC	Limited-stage SCLC
MCCs	Multidisciplinary Cancer Conferences
MDT	Multidisciplinary team
MET	Mesenchymal-epithelial transition
miRNAs	MicroRNAs
MPE	Malignant pleural effusion
MRI	Magnetic resonance imaging

mRNA	Messenger RNA
MVBs	Multivesicular bodies
MVs	Microvesicles
NCI	National Cancer Institute
ncRNA	Non-coding RNA
NCRS	National Cancer Registration Service
NLCA	National lung cancer audit
NLST	National Lung Screening Trial
NO_2	Nitrogen dioxide
NPPs	Nuclease protection probes
NPV	Negative predictive value
NSCLC	Non-small cell lung carcinoma
NTA	Nanoparticle tracking analysis
NTRK	Neurotrophic tyrosine receptor kinase
O_3	Ozone
OH	Ontario Health
OS	Overall survival
PAHs	Polycyclic aromatic hydrocarbons
Pb	Lead
PCA	Principal component analysis
PCI	Prophylactic cranial irradiation
PDT	Photodynamic therapy
PD-L1	Programmed death-ligand 1
PDT	Photodynamic therapy
PET	Positron emission tomography
PFTs	Pulmonary function tests
PM	Particulate matter
PORT	Postoperative radiation therapy
PPV	Positive predictive value
PS	Performance status
PSM	Positive surgical margins
QC	Quality control
QOL	Quality of life
RATS	Robotic-assisted thoracoscopic surgery
RCLCF	Roy Castle Lung Cancer Foundation
RCT	Randomised controlled trial
RET	Rearranged during transfection
RFA	Radiofrequency ablation
RNA-seq	RNA-sequencing
ROC	Receiver operating characteristic
ROS1	ROS proto-oncogene 1
RQ	Relative quantity
RSD	Relative standard deviation
RT	Radiotherapy
SABR	Stereotactic ablative body radiation
SBRT	Stereotactic body radiation therapy
SCLC	Small cell lung carcinoma
SD	Standard deviation
SEER	Surveillance, Epidemiology, and End Results
SHS	Second-hand smoke
SO_2	Sulfur dioxide

SPNs	Solitary pulmonary nodules
SqCCa	Squamous cell carcinoma
SRS	Stereotactic radiosurgery
TBNA	Transbronchial needle aspiration
TBs	Tumour boards
TEM	Transmission electron microscopy
TEPs	Tumour-educated platelets
TEVs	Tumour-derived extracellular vesicles
TKIs	Tyrosine kinase inhibitors
TME	Tumour microenvironment
TRT	Thoracic radiation therapy
TWR	Thoracoscopic wedge resection
UICC	Union for International Cancer Control
UK	United Kingdom
U.S.	United States
US CDC	United States Centres for Disease Control and Prevention
USPSTF	United States Preventive Services Task Force
US	United States
VATS	Video-assisted thoracoscopic surgery
VALSG	Veterans' Administration Lung Study Group
VEGFA	Endothelial growth factor
WBRT	Whole-brain RT
Wg	Wingless
WHO	World health organisation
WTA	Whole transcriptome assay

Chapter 1 General introduction

1.1 Cancer

According to the World Health Organization (WHO), cancer is the second leading cause of death worldwide, following cardiovascular disease (WHO 2018). Despite the fact that human beings have been plagued by cancer for centuries, it is often considered a modern disease for a number of reasons. Firstly, the prevalence of cancer has surged substantially in recent decades, which leads to a much more significant impact on modern society (Faguet 2015). In addition, it is widely known that cancer is much more common in older people. However, there is mounting evidence proving that cancer incidence is on the rise in young generations. A retrospective, serial cross-sectional, United States (U.S.) population-based study investigating the epidemiology of cancer in 497,452 adolescents and young adults (AYAs) aged 15 to 39 years (diagnosed between 1973 and 2015) indicated that the cancer rate in US AYAs increased by nearly 30% in just 42 years (Scott, Stoltzfus et al. 2020). As reported by Cancer Research UK (CRUK) (2015-2017), since the early 1990s, incidence rates for cancers in children and young people have escalated by 12% and 24%, respectively (www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers/incidence last accessed: 2022, www.cancerresearchuk.org/health-professional/cancer-statistics/young-people-cancers/incidence last accessed: 2022). The upward trend of cancer incidence in children has been observed in many studies conducted on childhood cancer in various countries worldwide in recent decades. Public awareness and concern have raised dramatically due to the growing burden of childhood cancer globally (Steliarova-Foucher, Colombet et al. 2017, Force, Abdollahpour et al. 2019). Furthermore, cancer is the leading cause of disease-related death in adolescents (CDC , Bleyer A 2006).

Finally, many studies on cancer history suggested different reasons for the rarity of cancers in past centuries compared with contemporary society. Nerlich et al. concluded that cancers were rare in previous centuries can be principally attributed to shorter life expectancy rather than the relative influence of primary environmental or genetic factors (Nerlich, Rohrbach et al. 2006). In contrast, a study by David and Zimmerman indicated that the rarity of malignancies in past centuries can be linked to the absence of negative factors found in modern lifestyles, such as exposure to carcinogens in tobacco smoke and industry pollution (David and Zimmerman 2010).

According to the GLOBOCAN-2020 report, the cancer burden worldwide is projected to be 28.4 million cases in 2040, a 47% escalation from 2020, in which there were 19.3 million new cancer cases (Sung, Ferlay et al. 2021). Correspondingly, unless the increasing number of cancers are treated and managed appropriately with adequate resources in healthcare, the rise in the incidence of these cancers will most likely take place in parallel with the escalation in mortality rates (Lortet-Tieulent, Georges et al. 2020). Given the significant and escalating impact of cancer on society, an increasing body of research has been undertaken to investigate the principles of cancer. In a seminal publication, Hanahan and Weinberg proposed that cells acquire a set of six cancer-associated biological capabilities in the multi-step process of carcinogenesis (Hanahan and Weinberg 2000). A decade later, in its sequel, the authors elucidated the involvement of the dynamic tumour microenvironment (TME) that emerges during the course of tumorigenesis, signalling interactions in the TME, dysregulated cellular energetics, immune evasion, genomic instability and mutability, and inflammation in cancer development and progression (Fig. 1.1) (Hanahan and Weinberg 2011). Based on research to date, future studies are likely to reveal and identify the crucial mediators of

internal communications in cancer. One aim will be to utilise them as potential biomarkers for cancer prevention, early detection, and cancer therapies in the future.

Figure 1.1. The hallmarks of cancer. In addition to the six acquired hallmark capabilities of cancer, Hanahan and Weinberg proposed four more novel cancer hallmarks which have extended the list to ten hallmarks of cancer [adapted from (Hanahan and Weinberg 2000, Hanahan and Weinberg 2011)].

1.2 Lung cancer

1.2.1 Lung cancer: epidemiology

Globally, lung cancer represents the most common cancer in terms of combined incidence and mortality. The WHO revealed the top 10 causes of death worldwide in 2019, which accounted for 55% of the 55.4 million global deaths. The number of deaths from trachea, bronchus and lung cancers increased by approximately 0.6 million in 2000 to nearly 1.8 million

in 2019 and now ranks 6th among the leading causes of death globally (www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death last accessed: 2022).

The GLOBOCAN 2018 database shows that lung cancer is responsible for 18.4% of the total cancer deaths worldwide, despite accounting for 11.6% of all new cancer cases (Bray, Ferlay et al. 2018). Furthermore, lung cancer appears to be the most commonly diagnosed type of cancer and the leading cause of cancer deaths in males. On the other hand, lung cancer ranks 3rd and 2nd for incidence and mortality among women, respectively (Bray, Ferlay et al. 2018).

Based on recently published data by GLOBOCAN 2018, lung cancer incidence rates are variable among geographical regions, with the highest incidence rates observed in Micronesia/Polynesia, followed by Northern America, Eastern Asia, Western Europe, and Eastern Europe. The lowest incidence rates were observed in Western Africa, closely followed by Eastern and middle Africa. It is also worth mentioning that incidence rates of lung cancer not only vary among geographical regions but also within the same area. For instance, in Eastern Europe, rates as high as 77.4/100,000 males and above 40/100,000 females have been recorded in Hungary based on age-standardised rates (ASRs) compared to the entire Eastern European region that was estimated at 49.3/100,000 males and 11.9/100,000 females, respectively (Bray, Ferlay et al. 2018).

Lung cancer incidence is also variable between sexes, with women tending to have lower incidence rates generally compared to men (Bray, Ferlay et al. 2018). Among men, the highest incidence rates were seen in Micronesia/ Polynesia (except for Hungary, ASR, 77.4/100,000 males). The incidence rates among men were generally low in Africa with the lowest incidence rates observed in Western Africa. With the exception of Hungary (ASR, 41.4/100,000 females), Northern America has the highest incidence rates among women (ASR, above

30.7/100,000 females), where Western Africa has the lowest incidence rates (ASR, 1.2/100,000 females) (Fig. 1.2) (Bray, Ferlay et al. 2018).

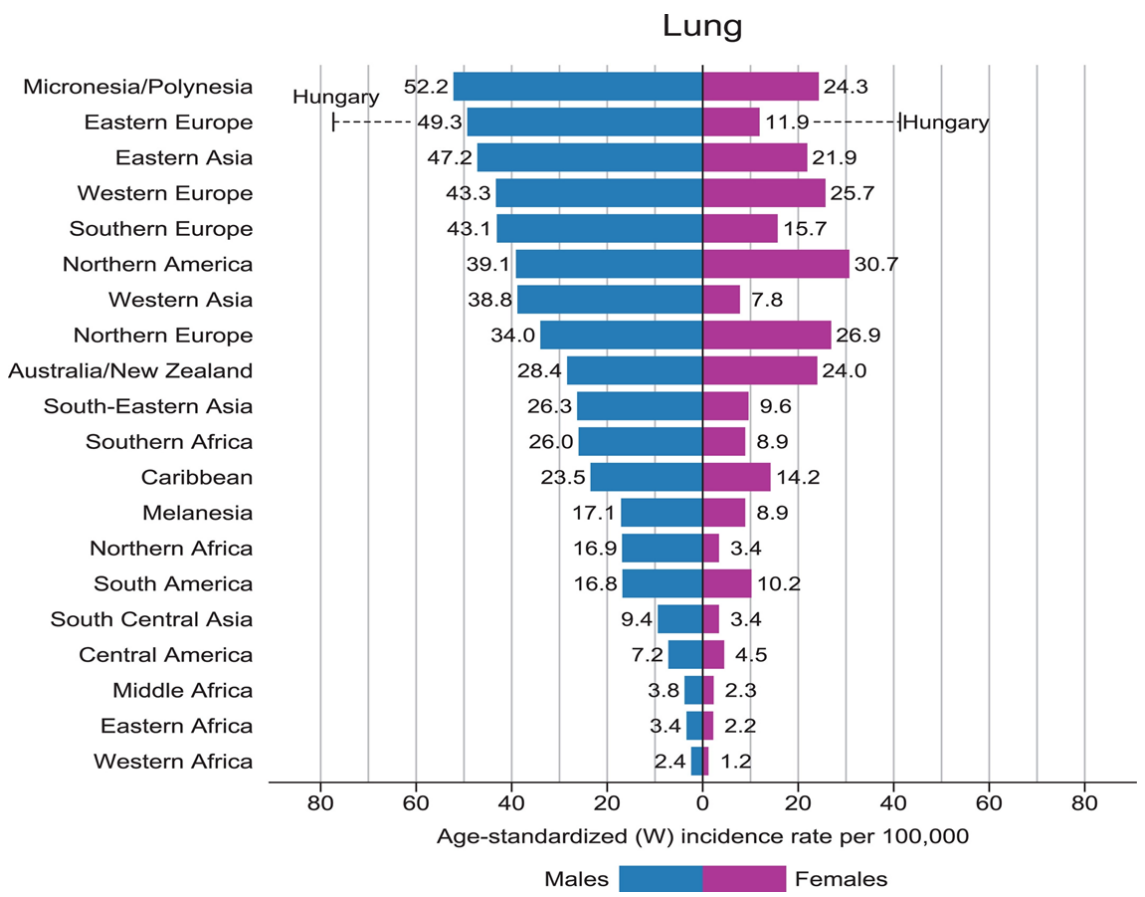


Figure 1.2. Bar graph of region-specific incidence by sex for lung cancer in 2018 (ASRs per 100,000 person-years). The incidence rates are displayed in descending order of the world (W) ASRs among men (Source: GLOBOCAN 2018) (Bray, Ferlay et al. 2018).

The cancer statistics from CRUK reveal that around 48,000 new cases of lung cancer were recorded in the United Kingdom (UK) in 2017, accounting for 13% of all new cases of cancer. Lung cancer ranks as the 2nd most common cancer in both males and females in the UK (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero last accessed: 2022). Based on 2016-2018 data from CRUK, around 35,100 lung cancer deaths were registered in the UK every year. Among both males

and females in the UK, lung cancer was the most common cause of cancer death in 2018, and it is responsible for 21% of all cancer deaths (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-One last accessed: 2022).

1.2.2 Lung cancer: aetiology

Tobacco smoking

According to Cancer Research UK and WHO, in 2018, lung cancer was the number one cause of cancer-related death in the UK, in Europe and worldwide (gco.iarc.fr/today/data/factsheets/populations/908-europe-fact-sheets.pdf last accessed: 2022, www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-One last accessed: 2022). There has been a considerable amount of research on the risk factors for lung cancer. Tobacco smoking is considered to be the leading cause of lung cancer and is closely linked to lung cancer mortality.

Cancer Research UK reported that over 70% of lung cancer cases in the UK are caused by smoking (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Three last accessed: 2022). It is also estimated that around 84% and 81% of lung cancer deaths in men and women in the US are caused by tobacco smoking (Health 2014). It is noteworthy that these mortality rates are associated with active smoking but not passive (involuntary) smoking [means inhaling second-hand smoke (SHS), or environmental tobacco smoke (ETS)], suggesting that the mortality of smoking-related lung cancer could be much higher when smoking as a whole including active and passive smoking is taken into account. A study predicting the ongoing effect of tobacco controls in the US (Jeon, Holford et al. 2018) uses accumulated data on smoking collected from 1964-2015 and

lung cancer mortality from 1969-2010 for adults aged 30-84. The study predicts a 79% reduction in the smoking-related age-adjusted lung cancer mortality rate (AAMR) from 2015-2065, and thus further highlights the impact of smoking on trends in lung cancer mortality.

It is widely known that exposure to polycyclic aromatic hydrocarbons (PAHs) from various sources is associated with a risk of cancer (Boffetta, Jourenkova et al. 1997, Chen and Liao 2006). Tobacco smoking is one of the most important sources of human exposure to PAHs (Goldman, Enewold et al. 2001). The carcinogenic PAHs formed during cigarette smoking are neutral, nonpolar and lipophilic compounds that are able to penetrate the pulmonary cell membrane through passive diffusion and undergo several metabolic reactions within the cells resulting in generating harmful reactive metabolites with the ability to covalently bind to DNA molecules and form DNA adducts that are capable of altering DNA replication and function (Yang, Ma et al. 2012). Among the smoking-related susceptibility genes that are subject to the effect of smoking-related metabolites is the tumour suppressor p53, where mutations in the p53 gene have been found to be more prevalent in lung cancers from smokers compared to non-smokers (Hainaut and Pfeifer 2001, Gibbons, Byers et al. 2014). The function of the tumour suppressor p53 is known to be vital in cell cycle regulation and apoptosis (Williams and Schumacher 2016, Kastenhuber and Lowe 2017). A research conducted by Greenblatt et al. revealed that for lung cancers from smokers, G→T transversions are the predominant type of p53 mutation (Greenblatt, Bennett et al. 1994). Moreover, several of the most frequently mutated codons of p53 are in correspondence with the sites of DNA adduct formation *in vitro* by PAHs metabolites. For instance, codons 157, 248, and 273 are among the most common mutational hot spots of the p53 gene in lung cancer, as well as the specific sites for DNA adduct formation by benzo(a)pyrene (BaP) metabolites in cultured cells (Ronai, Gradia et al.

1993, Denissenko, Pao et al. 1996). It is worth mentioning that these codons are not frequently mutated in non-smokers with lung cancer (Hernandez-Boussard and Hainaut 1998). Additionally, based on the International Agency of Research on Cancer (IARC) p53 mutation database, it was observed that the majority of the base changes at common hot spot codons 248, 249, and 273 in lung cancer diverge from those frequently seen at these codons in other cancer types (Hernandez-Boussard and Hainaut 1998). Hence, a unique p53 mutation spectrum in lung cancer from smokers is distinct from lung cancer in non-smokers, which also differs from all other cancers. These findings also indicated the strong association between active tobacco smoking, exposure to PAHs, and the development of lung cancer (Hernandez-Boussard and Hainaut 1998).

Given the critical role of smoking in lung cancer development, among men and women, the rates of incidence and mortality tend to be higher in men compared to women (Bray, Ferlay et al. 2018). This appears to be closely related to cigarette smoking patterns. Internationally, smoking rates peaked initially in men before women, mirroring smoking patterns some 20 years earlier (Proctor 2001). Lung cancer incidence has been declining since the 1990s (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero last accessed: 2022) following a steady decrease in smoking rates from the 1970s. Smoking rates have declined further, following the establishment of comprehensive tobacco control programs in developed countries, including the US and the UK. In contrast to developed countries, smoking rates in developing countries, including the BRICS countries (Brazil, Russia, India, China, and South Africa), increased more recently and remain high, with lung cancer incidence peaking more recently or continuing to rise (Bray, Ferlay et al. 2018).

Ambient (outdoor) air pollution

Outdoor air pollution has been classified by IARC as carcinogenic to humans (WHO 2013). There is considerable evidence that ambient air pollution exposure contributes to lung cancer (Cohen and Pope 1995, Nafstad, Haheim et al. 2003, Raaschou-Nielsen, Andersen et al. 2013). A study undertaken by Boffetta revealed that approximately 11% of lung cancers in Europe is attributable to urban air pollution (Boffetta 2006). As reported by Cancer Research UK, exposure to ambient air pollution causes nearly 1 in 10 cases of lung cancer (10%) in the UK (www.cancerresearchuk.org/about-cancer/lung-cancer/risks-causes#:~:text=We%20know%20that%20air%20pollution last accessed: 2022). It is believed that outdoor air pollution is responsible for approximately 5% of lung cancers in men and 3% in women in the US (WHO 2013).

Many studies have indicated an escalating risk of lung cancer ascribed to outdoor air pollution. The risk depends on the type of air pollutant, degree, and time of exposure, etc. (Vallero 2014). The US EPA (Environmental Protection Agency) has identified six common air pollutants (also called "criteria air pollutants"), which are carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO_2), ground-level ozone (O_3), particle pollution [often referred to as particulate matter (PM), such as PM_{10}, $PM_{2.5}$, PM_1 and $PM_{0.1}$ (graded by size)] (Brzezina, Kobolova et al. 2020), and sulphur dioxide (SO_2). A meta-analysis by Chen et al. suggests that long-term exposure to NO_2, SO_2 and $PM_{2.5}$ from exhaust emissions (primary sources of outdoor air pollution globally) considerably increases lung cancer risk (Chen, Wan et al. 2015). A large prospective study using data from 17 European cohort studies that contained 312,944 cohort members shows that long-term exposure to PM air pollution is associated with lung cancer (particularly adenocarcinoma) incidence in Europe (Raaschou-Nielsen, Andersen et al.

2013). Furthermore, some carcinogens are often found in air pollution, for instance, benzene, formaldehyde and coal ash (WHO 2013).

Occupational risks

As with silica dust and asbestos fibre, exposure to carcinogenic substances can occur among a variety of occupations, which may cause or contribute to the development of lung cancer. Apart from asbestos and silica, the IARC has identified a number of occupational exposures associated with lung cancer, such as arsenic, cadmium, chromium compounds, coal gasification, soot, coke oven fumes, nickel refining, foundry substances, diesel exhaust, tars, oils, radon, beryllium (or beryllium compounds), chloromethyl ethers (CME) and vinyl chloride (ATSDR 2009). A study by Doll and Peto indicated that 15% and 5% of lung cancers in American men and women respectively could be attributed to occupational exposures (Doll and Peto 1981). Besides, exposure to arsenic may lead to around 5,297 cases of lung cancer in the US each year (Putila and Guo 2011). In addition, it has been well documented that exposure to radon gas and its decay products is associated with the development of lung cancer in cohort studies of uranium miners (Roscoe, Steenland et al. 1989).

Previous lung diseases

Chronic obstructive pulmonary disease (COPD) describes a collection of progressive chronic lung illnesses that includes emphysema and chronic bronchitis. It is known that COPD is closely linked to lung cancer, explicitly squamous cell lung carcinoma (Papi, Casoni et al. 2004). It has also been reported that COPD is a significant risk factor for the development of lung cancer and predates lung cancer in as many as 70–80% of cases among smokers (Young and Hopkins 2011). Furthermore, a large South Korea cohort study of 338,548 subjects aged 40 to 84 suggests that COPD is a potent independent risk factor for lung cancer among never

smokers (Park, Kang et al. 2020). Additionally, lung cancer incidence in never smokers with COPD was over 2.6 times higher than never smokers without COPD (Park, Kang et al. 2020). Another important implication of this study is that regardless of smoking status, the COPD population is at high risk of lung cancer (Park, Kang et al. 2020).

Genetic predisposition to lung cancer

The genetic basis for lung cancer susceptibility has been extensively investigated over past decades. Numerous studies show that susceptibility to the effects of similar levels and types of exposure to all sorts of lung carcinogens vary among individuals in the development of lung cancer (Li and Hemminki 2004, Hung, McKay et al. 2008). Moreover, the findings of several important familial aggregation studies provide compelling evidence supporting genetic susceptibility to lung cancer. A previous study by Tokuhata and Lilienfeld revealed that among smokers, the lung cancer mortality in relatives of lung cancer patients was 2 to 2.5 times greater than in relatives of control subjects, which was not accounted for by age, gender, race, generation, residence, and cigarette smoking factors (Tokuhata and Lilienfeld 1963). Cassidy and colleagues (2008) observed a significantly elevated risk for those with a first-degree early-onset (i.e., under 60 years of age at diagnosis) family history of lung cancer (Cassidy, Myles et al. 2006). It is possible that host genetic factors at least partially determine susceptibility to lung cancer. Thus, the population with genetic susceptibility to lung cancer may be more susceptible if they are exposed to lung carcinogens.

1.2.3 Lung cancer: histological classification, staging and grading

1.2.3.1 Lung cancer classification

The majority of lung cancers are thought to originate from genetic alterations at the molecular level in the epithelial cells lining the airway (Wistuba and Gazdar 2006, Herbst, Heymach et al. 2008, Travis, Brambilla et al. 2011). Lung cancer is histologically categorised into two main types: small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC), which account for 15% and 85% of all lung cancers, respectively (Paez, Janne et al. 2004, Sher, Dy et al. 2008, George, Lim et al. 2015). There are three major histologic types of NSCLC: adenocarcinoma (ADC), squamous cell carcinoma (SqCCa, also known as epidermoid carcinoma), and large cell carcinoma (LCC) (Fig. 1.3, Fig. 1.4) (Gridelli, Rossi et al. 2015).

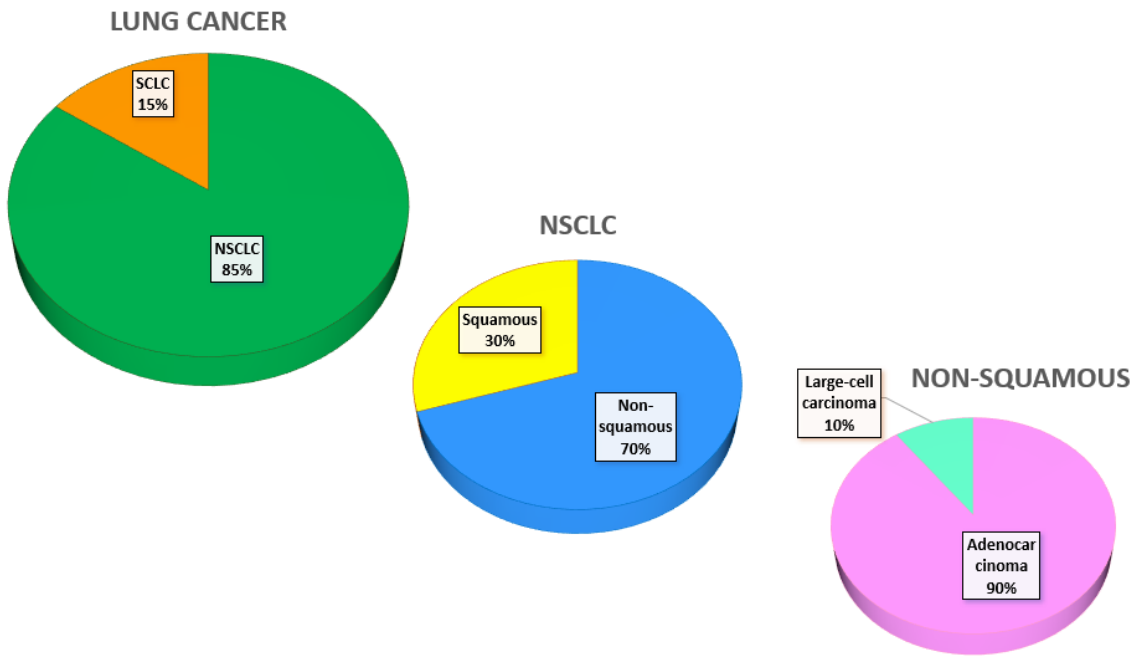


Figure 1.3. Lung cancer classification. [adapted from (Travis, Brambilla et al. 2013)]

These subtypes differ in cellular origin, morphology or staining pattern, location within the lung, and growth pattern, they are grouped together because of their similarities in biological

behaviour, therapeutic approaches, response to treatment, prognosis, etc. (Travis, Brambilla et al. 2015). There are also some less common histologies of lung cancer such as pleomorphic, carcinoid tumour, salivary gland carcinoma, adenosquamous carcinoma, sarcomatoid carcinoma and unclassified carcinoma (www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq last accessed: 2022).

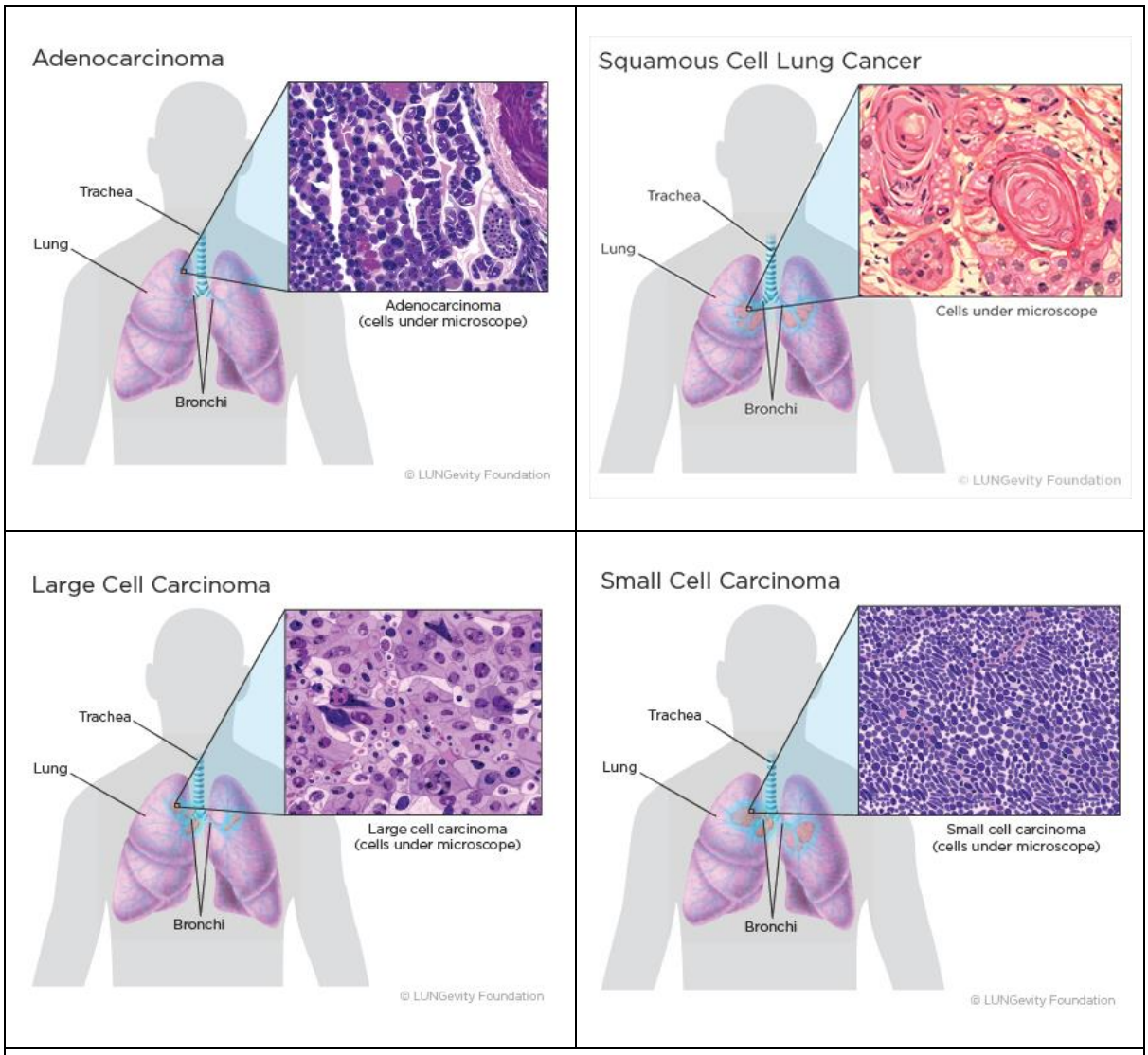


Figure 1.4. Histological classification scheme for lung cancers presented in microscopic images. (Source: LUNGevity Foundation) (www.lungevity.org/for-patients-caregivers/lung-cancer-101/types-of-lung-cancer/small-cell-lung-cancer-sclc last accessed: 2022).

Adenocarcinoma arises from glandular cells, which secrete substances such as mucus. Adenocarcinoma of the lung is the most prevalent histologic type of lung cancer, accounting for around 40% of all lung cancers (Travis, Brambilla et al. 2011, Zappa and Mousa 2016). It is highly heterogeneous at the histological level, with five histologic patterns: lepidic, acinar, papillary, micropapillary, and solid (Fig. 1.4) (Solis, Behrens et al. 2012). The majority of lung cancers found in never smokers are adenocarcinomas (www.cancer.org/cancer/lung-cancer/about/what-is.html 2022). According to the United States Centres for Disease Control and Prevention (US CDC), approximately 50%-60% of lung cancers in people who have never smoked are adenocarcinomas (www.cdc.gov/cancer/lung/nonsmokers/index.htm last accessed: 2022). Adenocarcinoma occurs more often in women than men, and it is also the most common type of lung cancer seen in young people compared to other lung cancer types (www.cancer.org/cancer/lung-cancer/about/what-is.html 2022).

Squamous cell lung cancer originates from bronchial epithelial cells (BEC) through multiple preneoplastic stages, such as basal cell hyperplasia, squamous metaplasia and dysplasia (Perez-Moreno, Brambilla et al. 2012). It is recognised by its most characteristic differentiation features: keratinisation, intercellular bridges, and keratin pearl formation (Fig. 1.4) (Perez-Moreno, Brambilla et al. 2012, Zamay, Zamay et al. 2017). It tends to occur in the central part of the lung or near/in the main airways, such as the bronchus. Squamous cell carcinoma is often linked to smoking, with over 90% of SqCCa reported to be attributable to tobacco smoking (IARC 2004) (Ettinger, Akerley et al. 2013).

Large cell carcinoma arises from lung epithelial cells and represents a minority of NSCLC that is devoid of the cytological and architectural characteristics of SCLC and histological differentiation of ADC and SqCCa (atlasgeneticsoncology.org/solid-tumor/5141/lung-non-

small-cell-carcinoma last accessed: 2022). LCC is often found peripherally in the lungs and appears necrotic. It is characterised by its large, undifferentiated (or poorly differentiated), polymorphic, polygonal cells with vesicular nuclei (Fig. 1.4) (Muller 1984).

Small cell lung carcinoma is also known as oat cell cancer since the tumour cells look small (smaller than normal, healthy cells and cells of other types of NSCLC), round or oval, and resemble oats under the microscope (Fig. 1.4). SCLC is thought to arise from neuroendocrine cells in the lung and often found in the bronchi near the central part of the lung. Furthermore, it is a highly malignant and aggressive disease that commonly develops and spreads more rapidly and widely than NSCLCs (www.nccn.org/ last accessed: 2022). Smoking is a significant risk factor for SCLC, and it is almost found exclusively in smokers (Fig. 1.5) (Muscat and Wynder 1995, Ettinger and Aisner 2006).

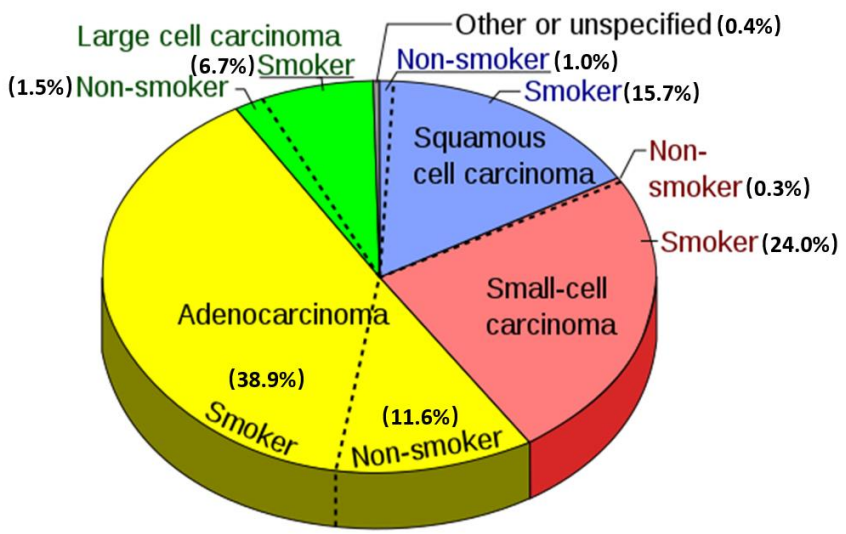


Figure 1.5. Pie chart of the incidence of lung cancer histological types and smoking. The incidence of SCLC is highlighted in red, and it is clearly shown that tobacco smoking is the predominant cause of small cell lung cancer [adapted from (Kenfield, Wei et al. 2008, commons.wikimedia.org/wiki/File:Pie_chart_of_lung_cancers.svg last accessed: 2022)].

1.2.3.2 Lung cancer staging

The staging of lung cancer is an integral part of deciding the treatment options and the prognosis for individual patients. The Union for International Cancer Control (UICC) and the American Joint Committee on Cancer (AJCC) TNM stage classification is the global standard to assess and classify the size and extent of the primary tumour (T stage, T1-T4), lymph node involvement (N stage, N0-N3), and metastasis spread (M stage, M0 or M1). Combining the specific T, N, and M indicates the stage of the disease for the patient. The TNM staging system described above is the system used for NSCLC (Fig. 1.6), it comprises five stages (stage 0, I, II, III, and IV) with multiple subgroups of NSCLC. As a rule, the lower the stage, the less advanced and metastatic the lung cancer is (Fig. 1.7, Fig. 1.8) (Detterbeck, Boffa et al. 2017).

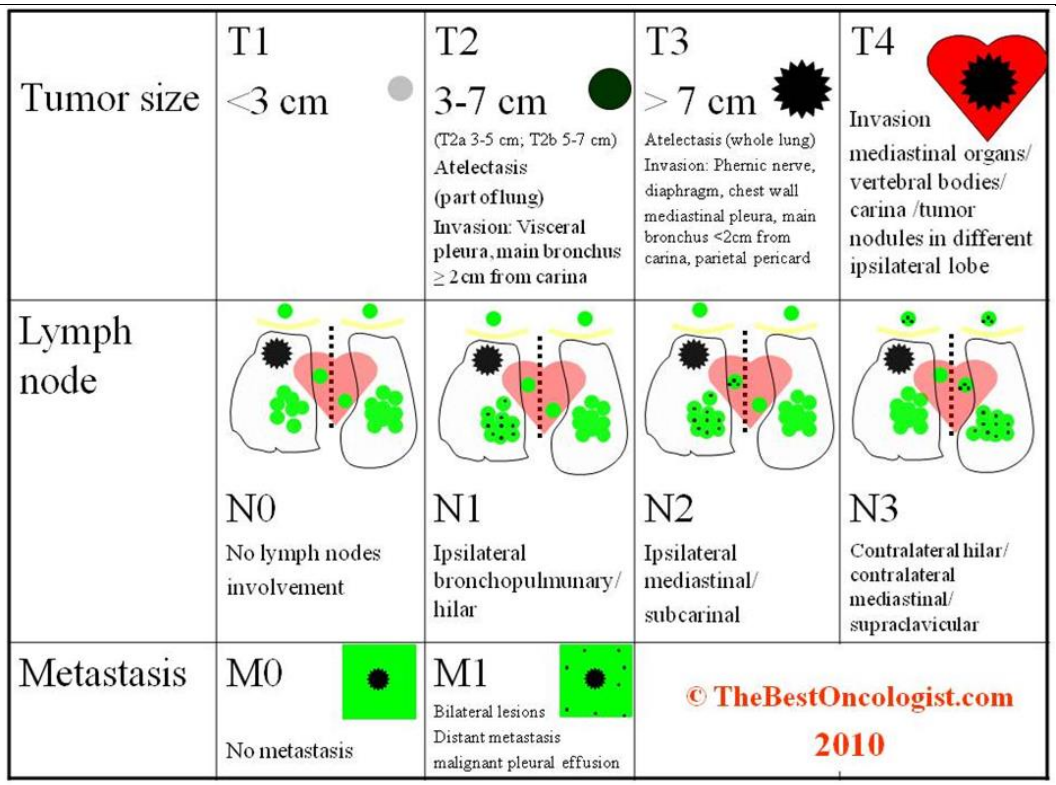


Figure 1.6. TNM stages of NSCLC. [adapted from (www.thebestoncologist.com/Cancer_Diseases/Lung_Cancer/Staging_of_Lung_Cancer.html last accessed: 2022)].

A 2-stage system is mostly adopted to classify SCLC. It is defined as limited-stage (limited to one hemithorax) and extensive-stage (cancer has spread outside the lung where it began) (Fig. 1.7) (www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/staging-sclc.html last accessed: 2022). However, the TNM staging system is now widely used for SCLC as well.

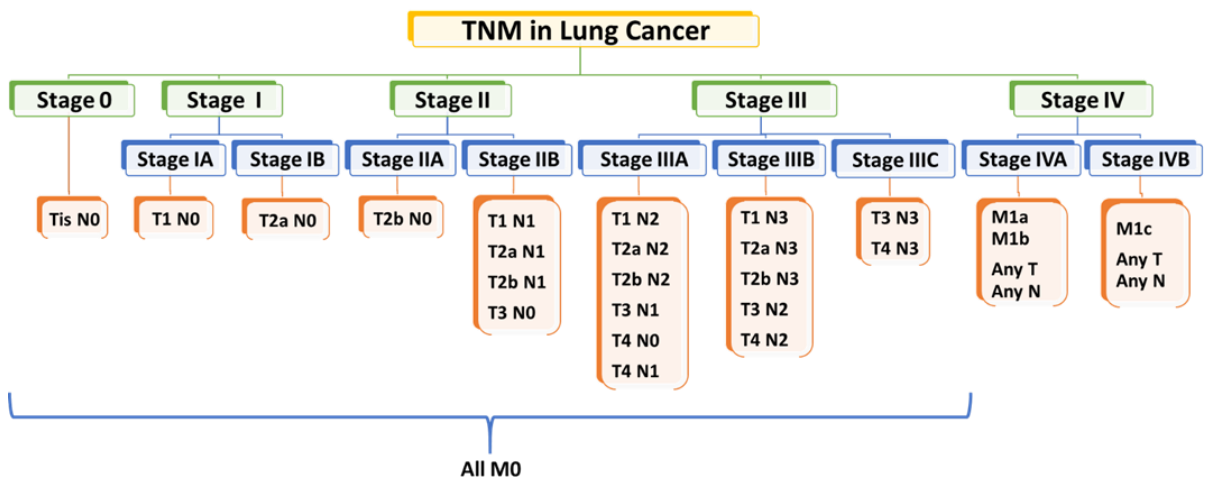
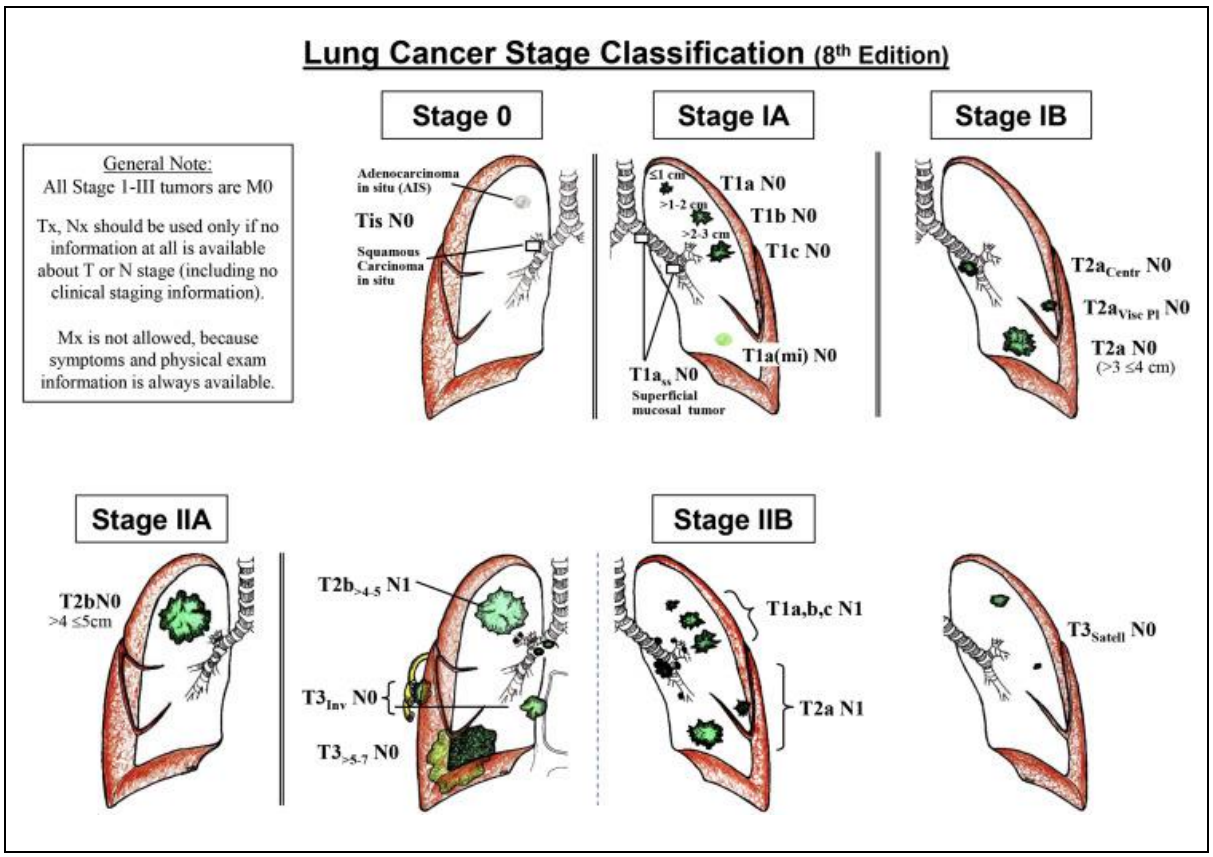


Figure 1.7. Staging of lung cancer. [adapted from (Detterbeck, Boffa et al. 2017)]

Stage 0 NSCLC is also known as carcinoma *in situ* (CIS); it has not spread or invaded into surrounding lung tissue, but remains within the alveoli or bronchiole where it originally developed. Not strictly a cancer, this is the earliest recognised stage in the development of lung carcinoma. The next stage in development is for the cancer to grow and invade surrounding tissue; this is reflected in the T stage. Initially these are T1 cancers, but larger cancers are classified as T2, T3 or T4 (although this classification also depends on location and other lung symptoms, such as atelectasis/pneumonitis). Another aspect of cancer development is its ability to spread to other tissues (metastasise). Often the first site of metastasis is to the local (regional) lymph nodes; this is reflected in nodal status (N stage), with N1 denoting nodes closer to the primary site, while N2 or N3 denoting distal nodes.

Ultimately, cancers can also spread to organs in more distant sites, e.g. brain, bones, liver; this is reflected in the M stage (with subcategories based on location and extent of metastatic spread).

The overall stage (Fig. 1.8) is a combination or size/location (T), nodal status (N) and distant metastasis (M). For example, a tumour of stage IIB might be a larger (T3) cancer with no metastatic spread (N0M0) or a smaller tumour (T1 or T2) with N1 metastasis. A tumour with any distant metastasis (M1) is always stage IV, the highest/latest stage (Detterbeck, Boffa et al. 2017).



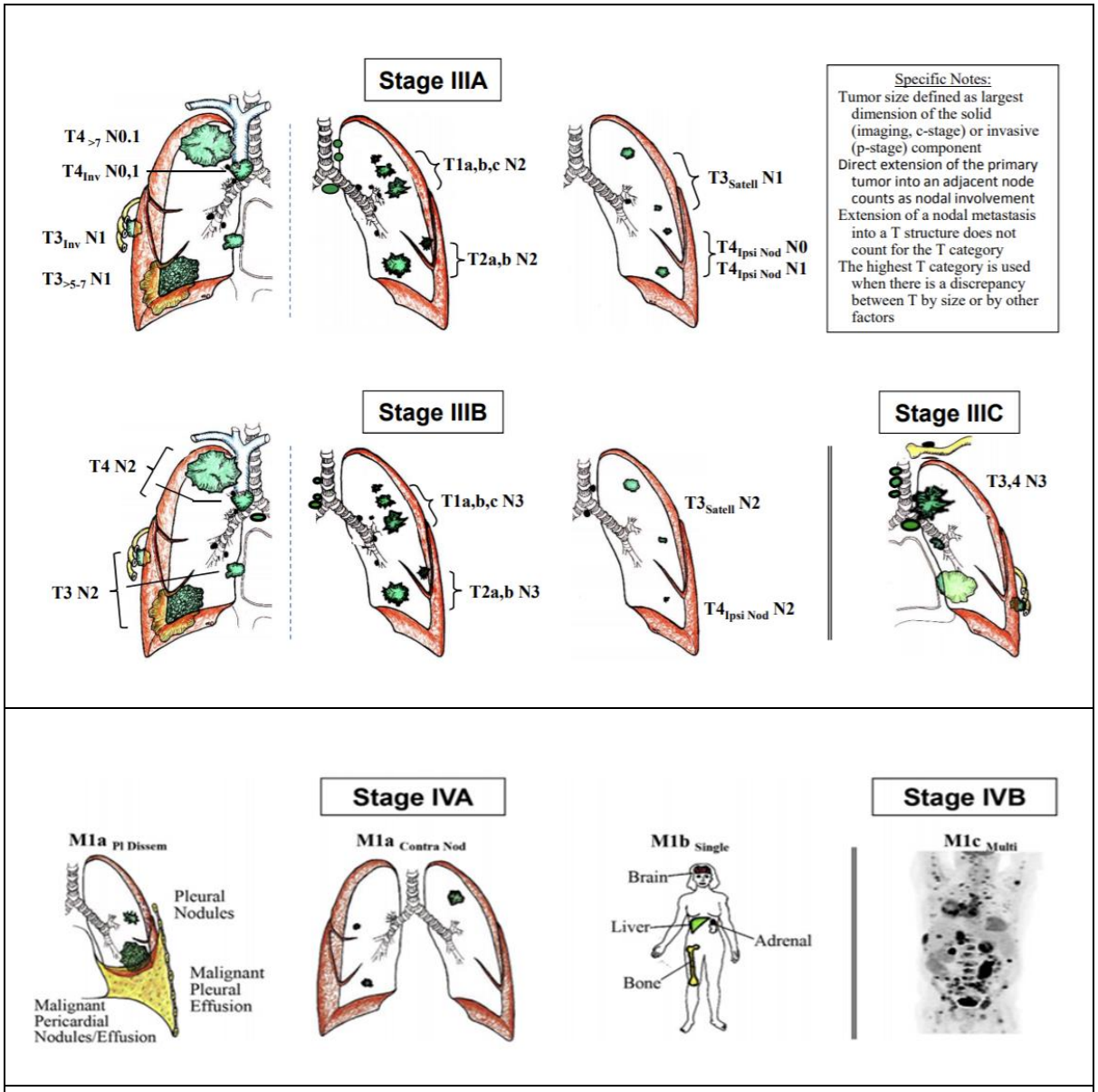


Figure 1.8. The 8th edition AJCC/UICC TNM stage classification for lung cancer. It is shown the graphic illustration of stage 0, I, II, III, and IV. Each T (tumour), N (node), and M (metastasis) component is divided into several categories (e.g., T1, T2) with increasing severity. Various characteristics, known as descriptors, define what is included within a T, N, or M category, e.g. size and location. Specific combinations of T, N, and M categories are grouped together into stage groups, which closely reflect outcomes. (Detterbeck, Boffa et al. 2017).

1.2.3.3 Grades of lung cancer

Conventional histological grading is based on the degree of tumour cell differentiation. In other words, the resemblance between the microscopic appearance of the tumour cells and the cells of origin, and how normal the tumour tissue structure appears to be. The grading of lung cancer (Travis, Brambilla et al. 2016, www.cancer.gov/about-cancer/diagnosis-staging/prognosis/tumor-grade-fact-sheet#r1 last accessed: 2022) is broken down into a spectrum from low grade (G1, well differentiated) to G2 (moderately differentiated), G3 (poorly differentiated) and high grade (G4, undifferentiated) (Fig. 1.9). Where the grade is not determined it is recorded as GX.

Based on morphological features, the grading of lung cancer is used to predict prognosis and develop individual treatment plans. In general, a lower grade indicates a greater degree of differentiation and more favourable behaviour of a tumour (i.e., better outcomes are related to the tumour cells more closely resembling normal tissue). High-grade tumours tend to be more aggressive and spread more rapidly than those with lower grades (as less differentiated cells are more plastic, better able to adapt and divide) (Travis, Brambilla et al. 2016).

Grading of Malignant Neoplasms		
Grade	**Definition**	**Microscopic appearance**
I	Well differentiated	
II	Moderately differentiated	
III	Poorly differentiated	
IV	Nearly anaplastic	

Figure 1.9. An adenocarcinoma is utilised as an example to illustrate the principles of grading. [Source (webpath.med.utah.edu/NEOHTML/NEOPL069.html last accessed: 2022)]

1.2.4 Lung cancer: diagnosis

Some lung cancers are found by screening, and some can be incidental findings. A study by Quadrelli et al. indicated that incidental findings of NSCLC occurred more frequently in smokers and in those with a previous history of malignancy. Moreover, lung cancer as an incidental medical finding is not uncommon even in non-smokers (Quadrelli, Lyons et al. 2015). However, most lung cancers are found because of the symptoms, which do not necessarily have to be specific signs and symptoms of lung cancer as the cancer cells might have already spread beyond their site of origin.

As stated by the United States Preventive Services Task Force (USPSTF) and CDC, annual lung cancer screening with low-dose computed tomography (LDCT, low-dose CT, also called low-dose CT scan) is recommended for older adults (aged 50-80 years) who have a 20 pack-year or more smoking history and who currently smoke or have quit in the past 15 years (Force, Krist et al. 2021, www.cdc.gov/cancer/lung/basic_info/screening.htm last accessed: 2022).

For patients with symptoms that signal lung cancer, the diagnostic evaluation usually includes medical history, physical exam, imaging tests, biopsy, sputum cytology, etc. Furthermore, blood tests and pulmonary function tests (PFTs) are frequently performed to evaluate overall health and how well the lungs are working, respectively.

More recently, biomarker testing has become increasingly important in both the diagnosis and management of lung cancer. Diagnostic biomarkers include those that help determine the histological classification of lung cancer (e.g. adenocarcinoma or squamous cell carcinoma) or the origin of the disease (primary lung cancer as opposed to metastatic spread to the lungs).

Molecular tests for gene changes [e.g. KRAS, epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), ROS1, RET, BRAF, MET, etc.] and tests for certain proteins on tumour cells [e.g. programmed death-ligand 1 (PD-L1)] are not yet commonly used in diagnosis, however, are important in helping guide treatment (www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html last accessed: 2022). Biomarker testing will be discussed in detail in subsequent sections 1.2.7.

Imaging tests

The chest radiograph [also known as the chest x-ray (CXR)] has typically been the first test used for detecting any lung abnormalities and is an essential first-line imaging exam in patients suspected of having lung cancer (Rajpurkar, Irvin et al. 2018, Bradley, Hatton et al. 2021).

A computed tomography (CT) scan [also known as computerised axial tomography (CAT) scan] of the chest is often used as an initial imaging tool when a lung mass (or nodule) or something else suspicious is seen on a chest x-ray. More recently, CT scans have been used as the first imaging modality at rapid access chest clinics in the UK. A CT scan uses x-rays to take pictures from multiple different angles to create detailed cross-sectional images of the lungs and the inside of the chest. Unlike an X-ray, a CT scan can deliver a three-dimensional (3D) view of the chest and hence better detect both acute and chronic changes in the airways and lung parenchyma (Johny A. Verschakelen 2018). CT scan is the most frequently used imaging modality for clinical T staging of lung cancer (Purandare and Rangarajan 2015). It is more likely to detect small lesions in the lungs than routine chest x-rays. It should also be able to show the size, shape, and location of the lesion, which may help understand the nature of its formation (www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-

diagnosed.html last accessed: 2022). The advantages of CT imaging over X-ray have also been explored for early detection of lung cancer (low-dose CT screening, section 1.2.6.2). Additionally, magnetic resonance imaging (MRI) scan, positron emission tomography (PET) scan, PET/CT scan, and nuclear medicine scans such as bone scan and thyroid scan are supplementary imaging modalities that are often utilised to improve the accuracy of staging based on clinical data, CXR or CT findings, histopathology types of lung cancer, etc.

Tests needed to provide a definitive diagnosis of lung cancer

The gold standard for conclusive lung cancer diagnosis is considered histological examination (Mukhopadhyay 2012, Mehic, Duranovic Rayan et al. 2016). The cells can come from various sources, for instance, lung secretions, pleural effusions, and small samples from a suspect area of the airways or the lung parenchyma. The choice of which test to be used depends on the situation of the individual patient.

Sputum cytology

Sputum cytology is a non-invasive and cost-effective cytological examination. It may provide a means to reveal the presence of cancerous cells in a sample of thick mucus or phlegm. It is more likely to assist in finding central malignancy, for instance, squamous cell lung cancers (Oswald, Hinson et al. 1971, www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html last accessed: 2022).

Thoracentesis with pleural fluid cytology

Thoracentesis is frequently performed as a diagnostic or therapeutic procedure when the cause of pleural effusions remains unclear or excess pleural fluid removal is required. Thoracentesis is commonly recommended in patients with suspected lung cancer who have

accessible pleural effusions (Committee 1985). It is a minimally invasive and high-yield procedure to distinguish features of pleural effusions. A study by Porcel et al. shows that 37% of malignant pleural effusion (MPE) cases are secondary to lung cancer, followed by breast cancer (16%) (Porcel, Esquerda et al. 2014). Moreover, in general, the presence of MPE connotes a poor overall prognosis for lung cancer patients. The international association for the study of lung cancer (IASLC) staging indicated that the NSCLC patients with pleural dissemination had significantly lower median survival and 5-year survival rate (8 months, 2%) in comparison to the comparator clinical T4 group (13 months, 14%) attributed to other T4 factors rather than MPE (Rami-Porta, Ball et al. 2007).

Needle biopsy

In the diagnosis of lung cancer, cytological or histological specimens are often obtained from lung nodules or pleural membrane by image-guided needle biopsy (for instance, using fluoroscopy, CT, or ultrasound). Fine-needle aspiration biopsy (FNAB) and core-needle biopsy (CNB) are the two minimally invasive transthoracic biopsy techniques commonly used to diagnose lung cancer (www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html last accessed: 2022). Based on the findings of many diagnostic accuracy studies, the overall sensitivity and specificity of FNAB and CNB for lung cancer are relatively high (Milman 1995, Wallace, Krishnamurthy et al. 2002). The greater diagnostic yield of CNB is mainly linked to the size of the sample. Unlike FNAB, which provides samples with limited tissue architecture, CNB is capable of obtaining larger intact tissue fragments with preserved architecture to be sufficient for the classification (i.e. be more likely to discriminate *in situ versus* invasive cancer compared to FNAB) and molecular analysis of lung cancer (Yao, Gomes et al. 2012, Ocak, Duplaquet et al. 2016). The most commonly reported

complications of FNAB and CNB for lung cancer are pneumothorax and pulmonary haemorrhage. An initial meta-analysis by Yao et al. showed no significant differences in complication rates with either FNAB or CNB for lung cancer (Yao, Gomes et al. 2012).

Bronchoscopy

Bronchoscopy is considered one of the most essential procedures for diagnosing lung cancer and obtaining anatomical information of the airways. Endobronchial ultrasound (EBUS) has revolutionised bronchoscopy as it enhances the diagnosis and staging of lung cancer by enabling real-time ultrasonic guidance in minimally invasive sampling mediastinal/hilar lymph nodules and paratracheal and peribronchial lung masses by transbronchial needle aspiration (TBNA) (Du Rand, Barber et al. 2011). Endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) was initially reported by Hurter and Hanrath in 1992, and has been extensively used in clinical practice over the last decade (Hurter and Hanrath 1992). More recently, Navani et al. (2015) carried out an open-label, randomised controlled trial (RCT) involving 133 patients from six UK centres with suspected stage I to IIIA lung cancer based on CT scans. They examined lung cancer diagnosis and staging with EBUS-TBNA compared with conventional approaches such as bronchoscopy, CT guided biopsy, conventional TBNA, mediastinoscopy and PET-CT scan. Their findings suggested that EBUS-TBNA could be considered as an initial investigation technique for patients with suspected lung cancer as it is an accurate diagnostic modality for lung cancer diagnosis and nodal staging, and it facilitates faster treatment decision-making in comparison to conventional diagnosis and staging techniques (Navani, Nankivell et al. 2015).

Video-assisted thoracoscopic surgery (VATS)

Jacobaeus initially described thoracoscopy in 1910 (Jacobaeus 1910). Relying on the rapid development of micro cameras and endoscopic instruments, VATS has undergone a stepwise evolution and became a widely used minimally invasive surgical procedure for diagnosing and treating lung cancer. VATS is utilised in various aspects of lung cancer diagnosis, such as evaluating indeterminate pulmonary nodules and pleural effusions, staging of mediastinal lymph nodes, assessing pleural cavity with subsequent biopsy of visually abnormal areas, and thoracoscopic wedge resection (TWR) for diagnosing indeterminate solitary pulmonary nodules (SPNs) (Celik, Halezeroglu et al. 1998, Stoica and Walker 2000).

1.2.5 Lung cancer: treatment

Globally, lung cancer has the highest mortality rate of all cancers. According to the WHO, lung cancer is the leading cause of cancer death (1.80 million deaths) in 2020, followed by colon and rectum Cancer (916,000 deaths) and liver cancer (830,000 deaths) (www.who.int/news-room/fact-sheets/detail/cancer last accessed: 2022). Providing appropriate treatment at the earliest possible stage for lung cancer patients on the basis of accurate staging and evidence-based protocols may improve outcomes.

Multidisciplinary team (MDT) meetings [also known as multidisciplinary case conferences (MCCs) or tumour boards (TBs)] in thoracic oncology have been introduced and endorsed in the UK over the past two decades with the aim of optimising treatment strategy, providing collective evidence-based recommendations on the treatment of lung cancer patients, shortening the time between clinical presentation, diagnosis and treatment initiation, prolonging survival and improving quality of life (QOL) for patients, etc. (Wales 1995, Whitehouse 1995, Patkar, Acosta et al. 2011, Powell and Baldwin 2014). Findings in the National Lung Cancer Audit (NLCA) Report 2012 reveal that nearly 96% of patients with

primary lung cancer were discussed at MDT meetings in the UK in 2011 (HQIP 2011). Moreover, multiple studies have explored the effects of MDT meetings on clinical practice for lung cancer. A study conducted in Australia reporting on the survival of patients with inoperable NSCLC showed a significant improvement in survival of those whose cases had been discussed at MDT meetings (Bydder, Nowak et al. 2009). In addition, the post hoc analysis of prospectively collected data obtained from 1,197 lung cancer cases diagnosed (with tissue confirmation) at a single institution in Australia between 2006 and 2012 suggested that MDT presentation is associated with prolonged adjusted survival for lung cancer at 1, 2 and 5 years after diagnosis (Stone, Rankin et al. 2018).

Several fundamental factors that are considered vital for lung cancer treatment planning, including types of tumour cells, tumour grade, the location, size and spread of cancer, diagnosis and clinical staging, with or without specific genetic/genomic/epigenetic alterations, as well as overall health, performance status (PS) and lung function of individual patients. Moreover, suppose lung cancer progresses while on treatment or relapses after treatment. In that case, apart from the factors mentioned above, further treatment will be based mainly on prior therapies, desire to receive more treatments, understanding of the goals, risks, and benefits of any further treatment, etc. (www.cancer.gov/about-cancer/diagnosis-staging/staging last accessed: 2022).

Surgery

The main treatment options for lung cancer include surgery, chemotherapy (ChT, also called chemo), radiation therapy (RT, also called radiotherapy), immunotherapy, and targeted therapy (www.cancer.gov/about-cancer/diagnosis-staging/staging last accessed: 2022). Several surgical approaches can be used to treat lung cancer, including wedge resection,

segmentectomy, sleeve resection, lobectomy, bilobectomy, and pneumonectomy (Fig. 1.10) (www.cancer.org/cancer/lung-cancer/treating-non-small-cell/surgery.html last accessed: 2022). However, surgery is not commonly used as a primary treatment for SCLC, as cancer has already metastasized at the time of diagnosis in more than 60% of patients with SCLC (Carter, Glisson et al. 2014, Tammela and Sage 2020). Sleeve resection (also called sleeve lobectomy) is a relatively complex, technically demanding procedure. It is performed to remove the tumours involving one lobe and the main airways to the lung. The cancerous lobe and the affected part of the airway are removed, and the ends of the bronchus are rejoined. Therefore, the remaining lung lobes can be reattached to the residual bronchus (www.who.int/news-room/fact-sheets/detail/cancer last accessed: 2022). If technically feasible (adequate tumour-free margin), sleeve resection should always be considered over pneumonectomy to maximize preservation of pulmonary function (Fig. 1.10) (www.who.int/news-room/fact-sheets/detail/cancer last accessed: 2022). Additionally, some valid minimally invasive alternatives to open surgery have been increasingly applied to treat NSCLC with the advantage of less pain, blood loss, discomforts, complications, shorter postoperative stay, and improved QOL and long-term functional results. For instance, VATS and robotic-assisted thoracoscopic surgery (RATS) (Postmus, Kerr et al. 2017, Ma, Li et al. 2021).

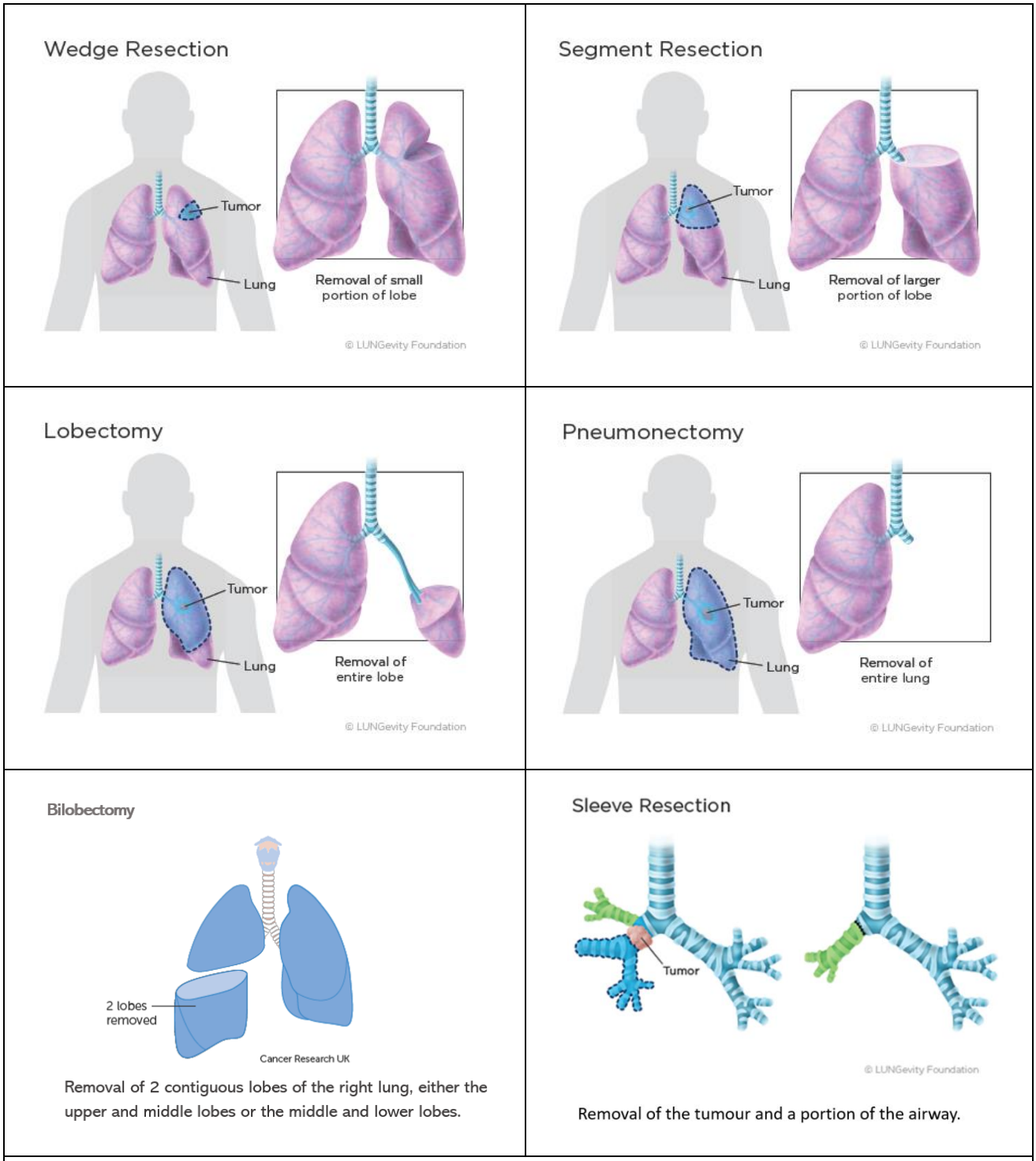


Figure 1.10. Types of lung surgery. (Detterbeck, Boffa et al. 2017) (www.cancerresearchuk.org/about-cancer/lung-cancer/treatment/surgery/types last accessed: 2022).

Chemotherapy

Chemotherapy is a predominantly oral/intravenous (IV) drug therapy that uses cytotoxic drugs to kill rapidly growing and multiplying cells. The history of cancer chemotherapy can be traced back to the early 20th century (DeVita and Chu 2008). Moreover, some of the chemotherapy pioneers, David Karnofsky and colleagues, published their research in the journal Cancer in 1948 investigating the effects of nitrogen mustard (HN2) on bronchogenic carcinoma. This marked the start of a period during which chemotherapy has been used to treat lung cancer (Craver 1948).

Depending on the stage of cancer and a variety of other factors, chemotherapy can be used as a type of systemic treatment for lung cancer in several ways with different treatment goals. Chemotherapy may be given on its own or along with other treatments to treat lung cancer. Neoadjuvant chemotherapy is a type of cancer treatment that may be administrated prior to the primary treatment to decrease or debulk cancerous tumours. Numerous studies have shown the value of neoadjuvant chemotherapy. Surgically, it allows otherwise impossible operations to be performed and facilitates less extensive procedures. In addition, research confirms early eradication of micrometastases and improved survival outcomes for patients with potentially resectable NSCLC (Berghmans, Paesmans et al. 2005, De Marinis, Gebbia et al. 2005). Adjuvant chemotherapy, on the other hand, is delivered after the primary therapy to lower the risk of recurrence and prolong overall survival. The findings of two meta-analyses of individual patient data carried out by NSCLC Meta-analysis Collaborative Group indicate that adjuvant chemotherapy after operation [whether with or without postoperative radiation therapy (PORT)] improves survival for patients with operable NSCLC (Group, Arriagada et al. 2010). Combined-modality therapy (CMT) for cancer refers to the sequential

or concurrent use of two or more treatment options. It is gradually becoming a widely used approach to the treatment of lung cancer. A number of studies confirmed the superior survival of chemoradiotherapy (CRT) over radiotherapy alone in locally advanced NSCLC (LA-NSCLC) (Zatloukal, Petruzelka et al. 2004, Dawe, Christiansen et al. 2016). The most common side effects of chemo for lung cancer include hair loss, nausea and vomiting, constipation or diarrhoea, and fatigue. Most of the adverse effects caused by chemotherapy are temporary and will gradually improve or disappear after the treatment finishes or can be managed symptomatically.

Radiation therapy

Radiotherapy plays a particularly prominent role in both curative and palliative treatments for lung cancer. Radiotherapy uses ionizing radiation (high-energy), including x-rays, gamma rays, electron beams, proton beams, neutron beams, carbon ion, alpha particles, and beta particles, to kill malignant cells and shrink tumours by damaging DNA in actively dividing cells (www.cancer.org/content/dam/CRC/PDF/Public/6151.00.pdf last accessed: 2022). There are two main types of radiotherapy, external beam radiation therapy (EBRT) and internal radiation therapy. EBRT is most often used to treat lung cancer and its spread to other parts of the body. For instance, stereotactic body radiation therapy [SBRT, also known as stereotactic ablative body radiation (SABR), or stereotactic ablative radiotherapy], three-dimensional conformal radiation therapy (3D-CRT) and intensity-modulated radiation therapy (IMRT). Internal radiation therapy is a type of cancer treatment in which the source of radiation is placed inside the body, in or close to the tumour. In lung cancer patients, it can be administered if the tumour is blocking or partly blocking the airway to relieve symptoms and to reduce the risk of shortness of breath, a collapsed lung, and other breathing problems

(www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html last accessed: 2022).

Depending on the stage of cancer and other factors, radiotherapy might be given to treat lung cancer in different ways for various purposes. For example, patients with early stage, medically inoperable NSCLC who are poor candidates for surgery or unwilling to undergo surgical procedures might be recommended radiotherapy with curative intent (radical radiotherapy or curative radiotherapy) (Price 2003). In addition, many studies confirm the superiority of continuous hyperfractionated accelerated radiotherapy (CHART) regimen over conventional radiotherapy for patients with LA-NSCLC in achieving local tumour control and survival with no significant evidence of a difference in long-term toxicity (Mauguen, Le Pechoux et al. 2012, Sanganalmath, Lester et al. 2018). Neoadjuvant radiotherapy (usually combined with chemotherapy) aims to improve resectability, downstage nodal disease, and to reduce the rate of local failures and distant metastasis (Sonett, Suntharalingam et al. 2004). For patients with potentially operable lung cancer, many studies have shown that neoadjuvant CRT can effectively facilitate complete resection and ensure safety, enhance pathologic complete response, and improve mediastinal clearance (Sonett, Suntharalingam et al. 2004, Pless, Stupp et al. 2015). Besides that, PORT (adjuvant) might be administered (alone or along with chemotherapy) to NSCLC patients with the presence of either close or microscopically positive resection margins, extracapsular spread (ECS), or involvement of multiple nodal stations to reduce the risk of local relapse and improve survival (Lally, Zelterman et al. 2006, Rasing, Peters et al. 2022). Radiotherapy may also be used to treat lung cancer that has spread to other areas with the aim of achieving pain relief and local tumour control (www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html

last accessed: 2022). For instance, whole-brain radiotherapy (WBRT) and stereotactic radiosurgery (SRS) are the two most commonly used treatments for brain metastases (BMs) (Mulvenna, Nankivell et al. 2016, Gondi, Meyer et al. 2021). Furthermore, palliative radiotherapy can help relieve a variety of tumour-related symptoms of advanced lung cancer and problems caused by cancer that has metastasized to other organs, such as bone or brain. Additionally, the side effects of RT for lung cancer depend on the treatment site and radiation dose. Some common side effects include tiredness, weakness, and skin changes in the treated area, such as redness, blistering, and peeling. Medicines might be administered to prevent or treat the side effects of radiotherapy (www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html last accessed: 2022).

Targeted therapy and immunotherapy

Other than chemotherapy, immunotherapy and targeted therapy are the two main forms of systemic therapy for lung cancer, which often present only mild and manageable adverse effects. These two emerging therapies have had an increasing role in the management of patients with advanced NSCLC.

As a cornerstone of precision medicine, targeted therapy uses drugs designed to target and interfere with specific molecules (known as molecular targets, such as genes and proteins) that are involved in the growth, progression, spread and survival of the cancer cells to stop the cancer from growing and spreading (www.cancer.gov/about-cancer/treatment/types/targeted-therapies last accessed: 2022). Unlike chemotherapies, targeted therapies are predominantly cytostatic rather than cytotoxic. In addition, targeted therapies are deliberately chosen or designed to act on molecular targets that are associated with cancer, whereas chemotherapies are identified due to being toxic to rapidly growing and

dividing cells, which means damage to normal cells (for example, the cells lining the lumen of the digestive tract) is sometimes inevitable (www.cancer.gov/about-cancer/treatment/types/targeted-therapies last accessed: 2022).

Over the past two decades, significant progress has been made in the discovery, identification, and characterization of oncogenic driver mutations in lung cancer that has led to the expansion of targeted therapeutic options for NSCLC, and many more potential targeted therapies are under investigation for both NSCLC and SCLC. Current evidence suggests that around 60% of lung ADC and approximately 50-80% of SCC harbour a known oncogenic driver mutation (Fig. 1.11) (Chan and Hughes 2015). To date, there have been around 20 FDA-approved targeted drugs for the treatment of NSCLC since the approval of gefitinib (Iressa, AstraZeneca Pharmaceuticals) in 2003 (for advanced NSCLC patients after failure of both platinum-based doublet chemotherapy and docetaxel chemotherapy), which was voluntarily withdrawn attributed to an inability to verify clinical benefit in the subsequent confirmatory trials; the current approval was granted in July 2015 for a different patient population (advanced NSCLC, previously untreated, EGFR mutation-positive, Table 1.1) (www.asco.org/research-guidelines/cancer-progress-timeline/lung-cancer last accessed: 2022). Examples of FDA-approved drugs include epidermal growth factor receptor (EGFR) inhibitors, anaplastic lymphoma kinase (ALK) inhibitors, ROS proto-oncogene 1 (ROS1) inhibitors, BRAF inhibitors, mesenchymal-epithelial transition (MET) inhibitors, neurotrophic tyrosine receptor kinase (NTRK) inhibitors, kirsten rat sarcoma virus (KRAS) inhibitors and rearranged during transfection (RET) inhibitors (www.lungcancerresearchfoundation.org/research/why-research/treatment-advances/ last accessed: 2022). Among these, the majority of the targeted drugs for lung cancer are used to

treat locally advanced or metastatic NSCLC (Table 1.1). The FDA and the EMA have also approved osimertinib (TAGRISSO, AstraZeneca Pharmaceuticals LP) as an adjuvant treatment for resected IB-IIIA NSCLC with EGFR exon 19 deletions or exon 21 L858R substitution mutations (Table 1.1) (www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-osimertinib-adjuvant-therapy-non-small-cell-lung-cancer-egfr-mutations last accessed: 2022). Furthermore, a meta-analysis by Chen et al. confirmed the definitive effect of targeted therapy on stage IIIA EGFR mutated NSCLC and superiority over chemo in respect of toxicity and response rate in the neoadjuvant setting (Chen, Jin et al. 2021). In addition to the preceding, at least two of these are indicated to treat solid tumours (including NSCLC) with specific biomarkers. For instance, larotrectinib was approved to treat solid tumours that harbour an NTRK gene fusion (Table 1.1) (www.lungcancerresearchfoundation.org/research/why-research/treatment-advances/ last accessed: 2022).

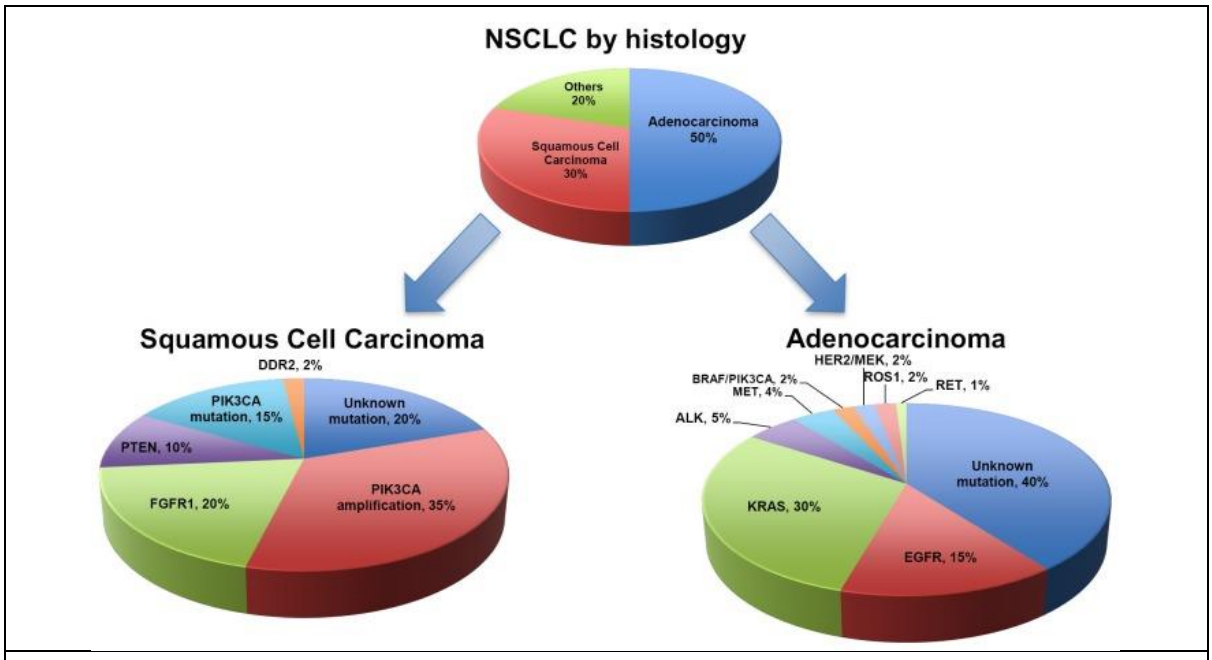


Figure 1.11. Proportion of driver mutations in NSCLC by histology (Chan and Hughes 2015).

Table 1.1. FDA-approved targeted therapies for NSCLC

Targeted therapy	FDA-approval date	Approved for	Biomarker
erlotinib(Tarceva, OSI Pharmaceuticals, Inc.)	18/11/2004	metastatic NSCLC	EGFR
gefitinib (Iressa, AstraZeneca)	13/07/2015	metastatic NSCLC (first-line treatment)	EGFR exon 19 deletions or exon 21 (L858R) mutations
osimertinib (Tagrisso, AstraZeneca Pharmaceuticals LP)	18/04/2018	metastatic NSCLC (first-line treatment)	EGFR exon 19 deletions or exon 21 (L858R) mutations
dacomitinib (VIZIMPRO, Pfizer Pharmaceutical Company)	27/09/2018	metastatic NSCLC (first-line treatment)	EGFR exon 19 deletion or exon 21 (L858R) mutations
erlotinib(Tarceva, OSI Pharmaceuticals, Inc.)	29/05/2020	metastatic NSCLC [in combination with ramucirumab (CYRAMZA, Eli Lilly and Company) for first-line treatment]	EGFR exon 19 deletions or exon 21 (L858R) mutations
osimertinib (Tagrisso, AstraZeneca Pharmaceuticals LP)	18/12/2020	NSCLC (adjuvant therapy after tumour resection)	EGFR exon 19 deletions or exon 21 (L858R) mutations
rybrevant (amivantamab-vmjw)	21/05/2021	NSCLC	EGFR exon 20 insertion mutations
mobocertinib (Exkivity, Takeda Pharmaceuticals, Inc.)	15/09/2021	metastatic NSCLC	EGFR exon 20 insertion mutations
dabrafenib and trametinib (TAFINLAR and MEKINIST, Novartis Pharmaceuticals Inc.)	22/06/2017	metastatic NSCLC	BRAF V600E mutation
lorlatinib (LORBRENA, Pfizer, Inc.)	02/11/2018	metastatic NSCLC (second- or third-line treatment)	ALK
brigatinib (ALUNBRIG, ARIAD Pharmaceuticals Inc.)	22/05/2020	metastatic NSCLC	ALK
lorlatinib (LORBRENA, Pfizer, Inc.)	03/03/2021	metastatic NSCLC (first-line treatment)	ALK
larotrectinib (VITRAKVI, Loxo Oncology Inc. and Bayer)	26/11/2018	NSCLC	NTRK gene fusion
entrectinib (ROZLYTREK, Genentech Inc.)	15/08/2019	NSCLC (NTRK+); metastatic NSCLC (ROS1+)	NTRK gene fusion or ROS1
capmatinib (TABRECTA, Novartis)	06/05/2020	metastatic NSCLC	a mutation that leads to MET exon 14 skipping
tepotinib (Tepmetko, EMD Serono Inc.)	03/02/2021	metastatic NSCLC	MET exon 14 skipping alterations
selpercatinib (RETEVMO, Eli Lilly and Company)	08/05/2020	metastatic NSCLC	RET fusion
pralsetinib (GAVRETO, Blueprint Medicines Corporation)	04/09/2020	metastatic NSCLC	RET fusion
sotorasib (Lumakras, Amgen)	28/05/2021	advanced NSCLC (at least one prior systemic therapy)	KRAS G12C

Source: (www.fda.gov/ last accessed: 2022, www.lungcancerresearchfoundation.org/about/ last accessed: 2022)

Immunotherapy is a category of cancer therapies that stimulates or boosts the body's immune system to recognise, attack and eliminate specific cancer cells. As a relatively new approach to cancer treatment, immuno-oncology (IO) is becoming an important field. Tremendous advances in immunotherapy for lung cancer have been made in immune checkpoint inhibitors (ICIs), and additional approaches are under investigation in clinical trials (such as cancer vaccines and adoptive T-cell therapy) (www.cancer.org/cancer/lung-cancer/treating-non-small-cell/immunotherapy.html last accessed: 2022). So far, four ICIs have been approved for advanced NSCLC in the UK, including two PD-1 inhibitors (nivolumab and pembrolizumab) and two PD-L1 inhibitors (atezolizumab, durvalumab) (www.cancerresearchuk.org/about-cancer/lung-cancer/treatment/immunotherapy-targeted last accessed: 2022). In addition to the preceding, the U.S. Food and Drug Administration (FDA) has approved cemiplimab (PD-1 inhibitor) and nivolumab/ipilimumab (CTLA-4 inhibitor) to treat patients diagnosed with advanced NSCLC (www.fda.gov/ last accessed: 2022). Most recently, nivolumab (Opdivo, Bristol-Myers Squibb Company) in combination with platinum-doublet chemo has been granted FDA approval as neoadjuvant treatment for adults with early stage NSCLC based on the results of the phase III CheckMate-816 trial (NCT02998528). This marks the first-ever FDA approval of neoadjuvant therapy for resectable NSCLC (www.fda.gov/ last accessed: 2022). Moreover, the FDA has approved atezolizumab (Tecentriq, Genentech, Inc.) as an adjuvant treatment (following surgery and platinum-based chemo) for stage II to IIIA NSCLC expressing PD-L1 ≥ 1% based on the findings of the phase III IMpower010 trial (NCT02486718), which makes atezolizumab the first and only adjuvant immunotherapy available for patients with NSCLC (www.fda.gov/ last accessed: 2022). Both the FDA and the European Medicines Agency (EMA) have approved the use of atezolizumab or durvalumab in combination with platinum-based chemo in extensive-stage

SCLC (ES-SCLC) (www.fda.gov/ last accessed: 2022). Nevertheless, the indications for nivolumab (Opdivo, Bristol-Myers Squibb Company) and pembrolizumab (Keytruda, Merk & Co.) have been withdrawn from the U.S. market for some patients with metastatic SCLC as the confirmatory studies failed to meet their primary endpoints of overall survival (OS) (www.fda.gov/drugs/resources-information-approved-drugs/withdrawn-cancer-accelerated-approvals last accessed: 2022).

Along with the development of technology and cancer research, more innovative approaches become complementary and alternative therapies for lung cancer treatment aiming to benefit poor candidates for the main treatments, prevent or slow progression and recurrence, relieve symptoms and side effects of cancer treatment, improve quality of life for cancer patients, etc. Examples of alternative therapies for lung cancer include photodynamic therapy (PDT), laser therapy, radiofrequency ablation (RFA), cryosurgery and electrocautery (www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html last accessed: 2022, www.cancer.org/cancer/lung-cancer/treating-small-cell/by-stage.html last accessed: 2022).

Among various factors discussed above that affect treatment decision-making, the type and the extent (stage) of the cancer are the essential elements required to guide and determine treatment strategy (www.cancer.gov/about-cancer/treatment/types last accessed: 2022). Lung cancer is a remarkably heterogeneous disease, and it is broadly broken down into two main histological groups: NSCLC and SCLC. Their therapeutic strategies differ substantially because of their marked differences in the aggressiveness, the doubling time, the growth fraction (GF), the propensity for early and widespread metastasis, the case fatality rate (CFR), etc. (Carter, Glisson et al. 2014). Furthermore, although the TNM staging system is now used

for both NSCLC and SCLC, for treatment purposes and practical reasons the modified Veterans' Administration Lung Study Group (VALSG) two-stage classification scheme [limited-stage SCLC (LS-SCLC) and extensive-stage SCLC (ES-SCLC)] once widely used, is still in use in some hospitals for the clinical staging of SCLC (Jett, Schild et al. 2013, Carter, Glisson et al. 2014). For the reasons given above, treatment options for different stages of NSCLC and SCLC will be demonstrated in the following sections.

Treatment choices for NSCLC

Early and locally advanced NSCLC

Stage 0 (Tis, N0, M0) is the earliest stage of NSCLC that can be detected. For stage 0 NSCLC, surgical resection is the treatment of choice if the patient is a good candidate for surgery. Moreover, minimally invasive thoracic surgery (for example, VATS and RATS) has emerged as the preferred surgical approach for early stage NSCLC (Postmus, Kerr et al. 2017). Alternative treatments for poor surgical candidates include laser therapy, PDT, SBRT, and brachytherapy (Postmus, Kerr et al. 2017).

Radical surgery remains the mainstay of treatment for patients with early stage NSCLC. Notably, complete resection is recommended for patients with multifocal lung cancer, and it should be discussed at MDT for all potential candidates (Postmus, Kerr et al. 2017). In addition, depending on the location, size and other factors, adjuvant ChT after surgical resection might be needed to lower recurrence risk and improve overall survival (www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html last accessed: 2022). In the presence of positive surgical margins (PSM), additional surgeries (which might be followed by ChT) or PORT may be required to ensure the remaining cancer cells are removed (Crino, Weder et al. 2010, Postmus, Kerr et al. 2017). Besides that, curative RT could

be an alternative treatment for potentially resectable lung cancer, such as SABR and hypofractionated high-dose RT (Postmus, Kerr et al. 2017). Moreover, SABR is a preferred treatment for patients with inoperability or peripherally located stage I NSCLC, or those who prefer non-surgical therapy (Lindberg, Nyman et al. 2015, Verstegen, Lagerwaard et al. 2015, Postmus, Kerr et al. 2017). Multiple studies investigating the role of SABR in early stage NSCLC have demonstrated that local control rates at 5 years post-SABR were approximately 90% based on imaging follow-up data (Lindberg, Nyman et al. 2015, Verstegen, Lagerwaard et al. 2015, Postmus, Kerr et al. 2017). Additionally, RFA may be a valid alternative for stage I NSCLC patients with contraindications for surgery or SABR (Ambrogi, Fanucchi et al. 2015, Postmus, Kerr et al. 2017). Furthermore, emerging findings from multiple ongoing clinical trials suggest that (neo)adjuvant immunotherapy might play a promising role in multimodality therapy for early stage NSCLC (Jia, Xu et al. 2020, Felip, Altorki et al. 2021).

LA-NSCLC represents a highly heterogeneous subtype of lung cancer in terms of tumour characteristics, the spectrum of disease distribution, patient population, treatment options, prognosis, etc. (Daly, Singh et al. 2022). Therefore, an experienced and skilled MDT could be of paramount importance in therapeutic decision-making. The European Society for Medical Oncology (ESMO) Clinical Practice Guidelines (CPGs) indicate that platinum-based ChT plays an essential role in the treatment of LA-NSCLC regardless of resectability (Postmus, Kerr et al. 2017). Furthermore, platinum-based ChT (preferably cisplatin) combined with RT (either concurrent CRT or sequential ChT followed by definitive RT) might be a valid option for the treatment of unresectable LA-NSCLC in the absence of contraindications (Postmus, Kerr et al. 2017, www.bccancer.bc.ca/books/lung/management/non-small-cell-lung-cancer-nsclc/combined-modality-therapy-for-unresectable-stage-iii 2022). Three meta-analyses

involving over 50 trials revealed that the combination of platinum-based ChT and RT demonstrates superior survival benefit over RT alone in unresectable LA-NSCLC (Marino, Preatoni et al. 1995, Pritchard and Anthony 1996). In addition, a meta-analysis reviewing 6 clinical trials confirms both the 3-year and 5-year survival benefit (primarily attributed to decreased locoregional failures) of concomitant CRT over sequential CRT (induction ChT followed by RT) in LA-NSCLC (Auperin, Le Pechoux et al. 2010, Postmus, Kerr et al. 2017, www.bccancer.bc.ca/books/lung/management/non-small-cell-lung-cancer-nsclc/combined-modality-therapy-for-unresectable-stage-iii 2022). Moreover, surgery might play a role in the multimodal management of LA-NSCLC. However, medical opinion remains divided (especially regarding stage IIIA-N2 NSCLC) due in large part to the lack of definitive randomised evidence proving the superior survival of surgical resection following induction treatment over radical radiotherapy and adequate innovative radiotherapy techniques (Thomas, Rube et al. 2008, van Meerbeeck 2008, Postmus, Kerr et al. 2017, Evison and AstraZeneca 2020, www.bccancer.bc.ca/books/lung/management/non-small-cell-lung-cancer-nsclc/combined-modality-therapy-for-unresectable-stage-iii 2022). Several local and international guidelines have acknowledged a few possible treatment regimens for potentially resectable stage III NSCLC (Postmus, Kerr et al. 2017, Evison and AstraZeneca 2020). For instance, induction concurrent CRT followed by surgical resection and pre-operative ChT followed by surgery (Postmus, Kerr et al. 2017, Evison and AstraZeneca 2020). According to a UK study of 6,276 patients diagnosed with stage III NSCLC (3,827 stage IIIA and 2,449 stage IIIB) in 2016 (using the NLCA data), less than 20% of patients with stage III NSCLC underwent curative intent multimodality treatment (either surgery or RT combined with ChT) (Adizie, Khakwani et al. 2019). Stage III NSCLC has a 1-year survival rate of 32.9%, and the highest survival rates were seen among those who received multimodality therapy with ChT and surgery (Adizie,

Khakwani et al. 2019). The findings of this study indicate that LA-NSCLC remains a challenging disease primarily because of the lack of timely access to accurate staging and integrated multidisciplinary lung cancer care (Adizie, Khakwani et al. 2019, Daly, Singh et al. 2022). It is also important to note, that along with the continued expansion of drugs approved by the FDA for immunotherapy and targeted therapy in LA-NSCLC, that IO and targeted therapy are predicted to further revolutionise the therapy landscape for stage III NSCLC (see 1.2.5 Lung cancer: treatment/Immunotherapy and targeted therapy for detailed information).

Advanced/metastatic NSCLC

Stage IV is the most advanced stage of NSCLC, commonly with a complex and arduous treatment process. According to the Surveillance, Epidemiology, and End Results (SEER)-18 data, the incidence of stage IV NSCLC declined from 21.7 to 19.6/100,000 between 2010 and 2017 in the U.S. (ascopost.com/news/november-2021/nsclc-in-the-united-states-update-on-incidence-prevalence-and-survival/ last accessed: 2022, seer.cancer.gov/data/ last accessed: 2022). Nevertheless, improving survival rates in metastatic NSCLC is still complicated and requires multifactorial interventions. A recent study conducted in Canada which included 24,729 patients with NSCLC diagnosed between April 2010 and March 2015 showed that nearly half of the patients (49.2%) were diagnosed with stage IV disease, the survival of whom remained poor, particularly amongst those who received no active treatment [at a median of 2.3 and 2.2 months for patients with squamous (1,704 patients) and nonsquamous (8,608 patients), respectively)] (Seung, Hurry et al. 2020).

Multiple factors should be taken into consideration in determining treatment strategies for stage IV NSCLC, such as the spread and location of the cancer, molecular pathology, histologic type, PS, age, and comorbidities (Planchard, Popat et al. 2018) (www.cancer.org/cancer/lung-

cancer/treating-non-small-cell/by-stage.html last accessed: 2022). It is usually necessary for patients with stage IV disease to shift their treatment objectives from curative to palliative intent for various reasons. Furthermore, the ESMO CPGs indicate that systemic therapy should be recommended for all stage IV patients with an Eastern Cooperative Oncology Group (ECOG) PS of 0 to 2 (Masters, Temin et al. 2015, Planchard, Popat et al. 2018). Historically, PS, toxicity profiles and comorbidities are primarily considered in chemotherapy choice (Socinski, Evans et al. 2013). Based on sufficient evidence, the current FDA-approved indications for both pemetrexed and bevacizumab have been updated, which reflects the significance of accurate characterization of histology in systemic therapies (Socinski, Evans et al. 2013). For instance, FDA approved carboplatin and paclitaxel ChT along with antiangiogenesis agent bevacizumab which should be used only in patients with advanced nonsquamous NSCLC, due to a higher incidence of major hemoptysis in patients with SqCCa (Johnson, Fehrenbacher et al. 2004, Sandler, Gray et al. 2006, Socinski, Evans et al. 2013). Additionally, pemetrexed can be used (either on its own or in combination) as an initial treatment, a second-line treatment, and a maintenance therapy for patients with advanced NSCLC, and its use should be limited to nonsquamous histology (Socinski, Evans et al. 2013). Moreover, the American Society of Clinical Oncology (ASCO) and the Ontario Health [OH, or Cancer Care Ontario (CCO)] NSCLC Expert Panel also updated the recommendations for stage IV NSCLC treatment (Hanna, Robinson et al. 2021). It is suggested that all patients with nonsquamous NSCLC should test for their potentially targetable alterations before treatment initiation for advanced NSCLC (Masters, Temin et al. 2015, Hanna, Robinson et al. 2021). Patients with molecular driver alterations, such as EGFR, ALK, ROS-1, BRAF V600e, RET, MET, and NTRK, should be offered the corresponding targeted therapy as initial or second-line therapy if not administered as first-line treatment (Masters, Temin et al. 2015, Hanna, Robinson et al. 2021). In addition to

the above-mentioned therapies, immunotherapy plays an increasingly important role in the treatment of advanced NSCLC. Pembrolizumab and atezolizumab have both been granted FDA approval as first-line treatments for advanced NSCLC (EGFR- and ALK-negative) with high PD-L1 expression (≥ 50%) (Planchard, Popat et al. 2018). Additionally, EBRT is the most common form of radiotherapy used for symptom control of metastases in stage IV NSCLC. For instance, relief of pain and symptomatic airway obstruction (Planchard, Popat et al. 2018). WBRT or SRS might be used for brain metastases with MRI follow-up scans (Planchard, Popat et al. 2018). Concerning asymptomatic brain metastases, next-generation TKIs might be recommended for patients with driver mutations to restore control of brain disease and reduce the risk of new central nervous system (CNS) metastases (Planchard, Popat et al. 2018). Furthermore, early palliative care intervention in the cancer treatment for patients with stage IV NSCLC is strongly recommended to improve QOL and prolong survival (Socinski, Evans et al. 2013, Masters, Temin et al. 2015, Planchard, Popat et al. 2018).

Treatment choices for SCLC (two-stage system)

SCLC accounts for around 13% of all lung cancer cases, and fewer than one-third of affected patients are diagnosed with LS-SCLC (Murray and Turrisi 2006, Byers and Rudin 2015, Senan 2021) (Jemal, Murray et al. 2005). SCLC represents the most aggressive and deadly subtype of lung cancer, and the prognosis remains dismal with a 5-year survival rate of approximately 7% (Semenova, Nagel et al. 2015). In addition, SCLC has been estimated to be responsible for nearly 4% of cancer death (Jemal, Murray et al. 2005, Murray and Turrisi 2006).

Unlike NSCLC, there has been a lack of substantial progress in the treatment of SCLC despite multiple ongoing investigations, and the standard treatments have remained unchanged for decades (Semenova, Nagel et al. 2015, Senan 2021). Chemotherapy remains the cornerstone

in the treatment of SCLC and the standard treatment in both first- and second-line settings. Etoposide-platinum based chemotherapy can be used for both stages (Osterlind 2001, Yang, Zhang et al. 2019). Furthermore, lurbinectedin (an inhibitor of RNA polymerase II), as a novel cytotoxic chemotherapy agent, has been granted FDA approval for the treatment of metastatic SCLC based on the results of the phase II multicentre multicohort study (NCT02454972) (Yang, Zhang et al. 2019, www.fda.gov/drugs/drug-approvals-and-databases/fda-grants-accelerated-approval-lurbinectedin-metastatic-small-cell-lung-cancer last accessed: 2022). Besides that, in spite of high response rates to initial chemotherapy, nearly 80% of LS-SCLC patients develop relapse or progression of disease (Kalemkerian, Akerley et al. 2013). Topotecan (an inhibitor of topoisomerase I), another cytotoxic drug, is the only available FDA-approved agent for the treatment of relapsed SCLC (Carter, Glisson et al. 2014). More positively, progress has been made in immunotherapy for SCLC. For instance, atezolizumab or durvalumab, either of which combined with chemo has led to significantly improved survival outcomes in ES-SCLC clinical trials and have been granted FDA approval (www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-atezolizumab-extensive-stage-small-cell-lung-cancer last accessed: 2022, www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-durvalumab-extensive-stage-small-cell-lung-cancer last accessed: 2022). Additionally, many studies have demonstrated the superior survival benefit of thoracic radiation therapy (TRT) in combination with etoposide-cisplatin chemo in patients with LS-SCLC (Perry, Eaton et al. 1987, Pignon, Arriagada et al. 1992). In addition, prophylactic cranial irradiation (PCI) might be used in either LS-SCLC or ES-SCLC to reduce the risk of brain metastases as approximately 50–60% develop brain metastases during the course of the disease (Auperin, Arriagada et al. 1999, Carter, Glisson et al. 2014, Edelman 2020). Lastly, early integrated palliative care for patients with SCLC is highly recommended.

1.2.6 The rationale for developing minimally invasive biomarkers for lung cancer early detection

Despite significant advances in diagnostics and therapy, survival or prognosis of lung cancer remains low compared to other cancer types. According to UK Cancer Statistics and Data, in 2017, there were 35,148 deaths from lung cancer in the UK, representing 21% of all cancer deaths (male deaths: 18,810, 21% of all male cancer deaths; female deaths: 16,338, 21% of all female cancer deaths) (www.wcrf-uk.org/preventing-cancer/uk-cancer-statistics/ last accessed: 2022). Within the European Union (EU), lung cancer represents 20.5% of all cancer deaths and 5.3 % of all deaths in 2016 (ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics_-_specific_cancers last accessed: 2022).

Given the high levels of lung cancer morbidity and mortality, the substantial socioeconomic and human burden of lung cancer and globally, the relatively high prevalence of smoking with a marked rise observed in the youth of many countries (smoking has been established as the single biggest risk factor for lung cancer as discussed in detail in Section 1.2.2)(Reitsma, Flor et al. 2021, tobaccoatlas.org/ last accessed: 2022), lung cancer remains a major health challenge. It is therefore evident that a more coordinated and comprehensive approach to tackling lung cancer with primary prevention and screening as priorities is most urgently required.

1.2.6.1 Early detection is key to reducing lung cancer mortality

The high mortality rate of lung cancer is known to be attributed mainly to the advanced stage of the disease at diagnosis as the 5-year survival for lung cancer drops dramatically from a stage I diagnosis (more than 55%) to a stage IV diagnosis (less than 5%) (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-

type/lung-cancer/survival%22 last accessed: 2022). Therefore, early detection of lung cancer is pivotal to increasing survival rates.

Reports of poor prognosis of lung cancer have been attributed to various challenging factors, which are highly relevant to the late diagnosis of lung cancer and its impact on the resulting severity of the disease and limited treatment options. Recently published statistics investigated the incidence, treatment, and survival of lung cancer patients based on the SEER database, accounting for 26% of the US population, which includes registries collecting data on 1,148,341 lung cancer patients (646,662 males and 501,679 females) in the United States from 1973 to 2015 (Lu, Yang et al. 2019). The findings of the study revealed that the survival rate of the lung cancer patients diagnosed at a localised stage have increased significantly from 1988 to 2010 (45.3% to 58.2%). In contrast, the average 5-year relative survival rate of those diagnosed at a distant stage remained poor (2.9%) over the same period, albeit with a slight improvement (Lu, Yang et al. 2019). Additionally, a UK study of 152,821 newly diagnosed cancer patients in 2012, analysing data retrieved from the registration data set of the National Cancer Registration Service (NCRS), indicated that of 34,997 lung cancer patients, 49% were diagnosed at stage IV, 20% at stage III, 7.5% at stage II, and only 13.2% were diagnosed at stage I (Fig. 1.12) (McPhail, Johnson et al. 2015). It should be noted that the percentage of 1-year survival rate declined along with the advancement of the lung cancer disease at the time of diagnosis, where those diagnosed at stage I have a higher 1-year relative survival (84%) compared to those diagnosed at stage II (69.5%), stage III (44.6%) and stage IV (17.1%) (Fig. 1.12) (McPhail, Johnson et al. 2015).

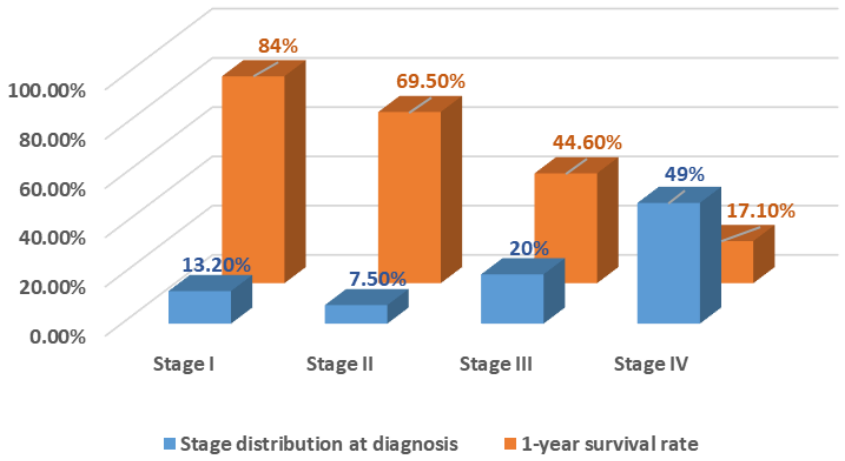
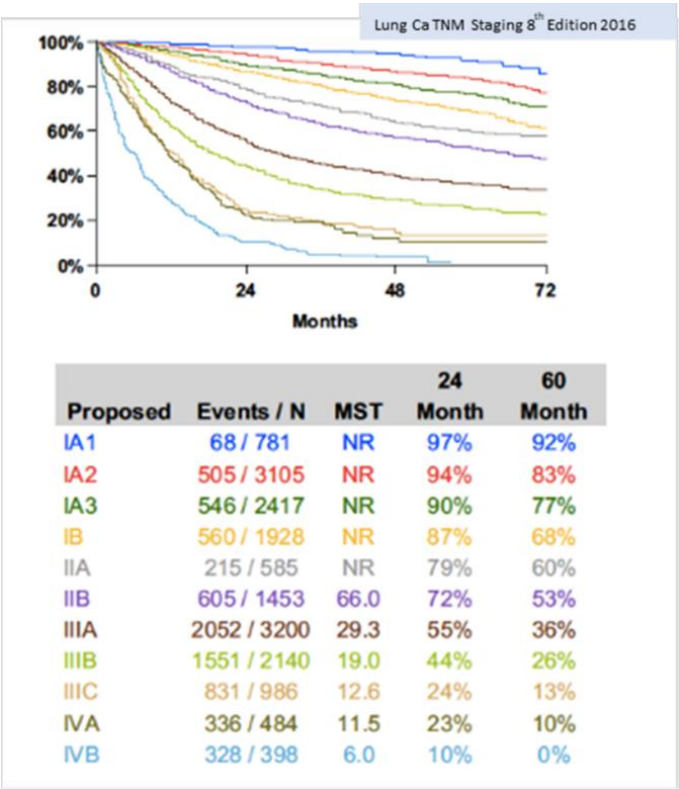


Figure 1.12. Column chart showing stage distribution at diagnosis and 1-year survival rate by stage [adapted from (McPhail, Johnson et al. 2015)].



Proposed	Events / N	MST	24 Month	60 Month
IA1	68 / 781	NR	97%	92%
IA2	505 / 3105	NR	94%	83%
IA3	546 / 2417	NR	90%	77%
IB	560 / 1928	NR	87%	68%
IIA	215 / 585	NR	79%	60%
IIB	605 / 1453	66.0	72%	53%
IIIA	2052 / 3200	29.3	55%	36%
IIIB	1551 / 2140	19.0	44%	26%
IIIC	831 / 986	12.6	24%	13%
IVA	336 / 484	11.5	23%	10%
IVB	328 / 398	6.0	10%	0%

Figure 1.13. Overall survival by clinical stage according to the 8th edition of the TNM classification for lung Cancer [adapted from (Nicholson, Chansky et al. 2016)].

Note: MST = median survival time.

Further data illustrating that survival is inversely proportional to the stage at diagnosis has been evidenced using the latest TNM staging system for lung cancer (AJCC 8th edition). This shows that the 5-year overall survival rate of the patients diagnosed with stage IA1 lung cancer (92%) is significantly higher compared to those with stage IVB disease (0%) (Fig. 1.13) (Nicholson, Chansky et al. 2016). Thus, detecting lung cancer at an early stage is the paramount factor affecting the overall survival rate in lung cancer patients. Improving early detection methods would positively impact long-term survival, increase range of treatment options and improve quality of life.

Furthermore, US-based collaborative research between the American Cancer Society (ACS) and the National Cancer Institute (NCI) reveals that only 21% of lung cancer cases are diagnosed at an early stage (stage I), and 61% of cases are not detected until a late stage (stage III or IV) (Fig. 1.14) (Miller, Nogueira et al. 2019). In 2019, based on the SEER database, the investigation reported that in 2015, the percentage of lung cancer patients in the US diagnosed at an early, localised stage increased to 23.6% compared to those diagnosed in 1988 (16.6%); a relatively small increase over nearly three decades (Lu, Yang et al. 2019). These findings highlight the ongoing challenges in early diagnosis of lung cancer, which are mainly due to the lack of noticeable symptoms in the early stages of lung cancer as well as poor diagnostic tools for early detection.

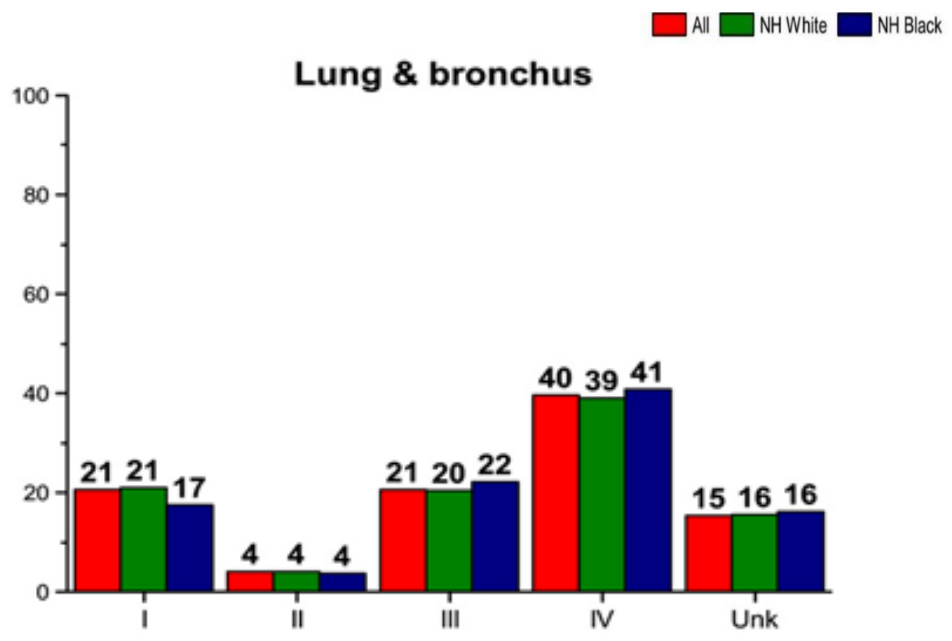


Figure 1.14. Lung cancer stage distribution at diagnosis (%) by race [adapted from (Miller, Nogueira et al. 2019)].

Note: NH = non-Hispanic; Unk = unknown stage

Over the past decade, intensive research has been carried out to understand the mechanisms behind carcinogenesis and tumour progression leading to the identification of a number of dysregulated transcripts in the human malignancy (Vogelstein, Papadopoulos et al. 2013). Along with the continuous evolution of cancer biology research and the emergence of new technologies, extensive studies have been undertaken. These studies have considered the development of novel diagnostic strategies and the discovery and validation of potential diagnostic biomarkers for lung cancer early detection, as mortality remains high and challenging despite the current advances in lung cancer management.

1.2.6.2 Lung cancer screening with low-dose computed tomography (LDCT)

The advantages of CT imaging over X-ray for routine diagnosis of lung cancer have been discussed previously (Section 1.2.4). However, these are pertinent only for those patients

with a high likelihood of a lung cancer diagnosis. The high doses of radiation and the use of contrast agents associated with the diagnostic CT technique are not suitable for screening.

In comparison with traditional chest X-Rays, CT scans, without a doubt, provide more comprehensive images. However, with more than 100 times higher radiation dose of a chest CT compared to that of a routine chest radiograph, using CT as an early screening tool in lung cancer diagnosis remains hazardous (Mettler, Huda et al. 2008, Linet, Kim et al. 2009). To overcome the limitations of CT scans, the usage of LDCT appears to be advantageous, notably that an LDCT is reported to utilise an average of 78% less radiation dose compared to that of a standard chest CT scan (Larke, Kruger et al. 2011). Additionally, LDCT is a non-invasive and painless (does not require administration of radiocontrast agents) and an extremely rapid (the actual scan takes less than 1 minute) approach to lung cancer screening (www.cedars-sinai.org/programs/imaging-center/exams/ct-scans/lung.html last accessed: 2022). The safety of LDCT as a screening tool in lung cancer early diagnosis has been investigated in several clinical trials exploring the effect of LDCT screening on lung cancer mortality rates in asymptomatic, high-risk persons.

A large NCI-sponsored national lung screening trial (NLST) enrolled 53,454 individuals aged 55 to 74 years at the time of randomization considered to be at high risk for lung cancer in the U.S. between 2002 and 2004, where 26,722 and 26,732 participants were screened annually over three years (including three specific screenings; T0, T1 and T2) by LDCT and single-view posteroanterior chest radiography, respectively (National Lung Screening Trial Research, Aberle et al. 2011). The findings of the study published in 2011 demonstrate a considerably higher rate of positive screening tests in the LDCT group compared to the radiography group, not only in the outcome of overall three rounds (LDCT, 24.2% *versus* chest radiography, 6.9%)

but also of every round of the screening (LDCT, T0: 27.3%, T1: 27.9% and T2: 16.8% *versus* chest radiography, T0: 9.2%, T1: 6.2% and T2: 5.0%) (National Lung Screening Trial Research, Aberle et al. 2011). In addition, 7.5% of all screening tests in the LDCT group identified an abnormality with clinical significance other than an abnormality indicative of lung cancer, which was three times higher than that of the radiography group (2.1%) (National Lung Screening Trial Research, Aberle et al. 2011). Furthermore, a significant relative reduction in mortality (20%) from lung cancer was observed in the LDCT group compared to the radiography group in a large population of current and former heavy smokers (≥ 30 pack-years, < 15 years quit time) (National Lung Screening Trial Research, Aberle et al. 2011).

Although the NLST positive outcomes are higher in terms of the rate of positive screening tests and substantially reduced mortality rates in the LDCT group compared to the radiography group, a higher percentage of false positive results were evident among those flagged positive by LDCT screening (96.4%) compared to those identified by chest radiography (94.5%) across all rounds (National Lung Screening Trial Research, Aberle et al. 2011). In the NLST, any non-calcified pulmonary nodule measuring 4mm or greater in diameter was classified as positive by LDCT. Nevertheless, by increasing the nodule volume to > 500 mm^3 (diameter ~ 9.8 mm) in another sufficiently powered trial, the Dutch-Belgian lung cancer screening trial (NELSON), explored the outcomes of LDCT screening (n = 7, 915) over three years *versus* no screening (n = 7,909) in lung cancer high-risk participants (aged 50 ~ 75 years, former smoker, ≥ 18.75 pack-years; current smoker, ≥ 15 pack-years or < 10 years quit time).

This trial undertook three annual rounds of screening and the results suggest a significantly lower percentage of false positive screening results (59.4%) compared to that of NLST (96.4%) and another trial which also used volumetry-based screening strategy [Danish lung cancer

screening trial (DLCST), 65.2%] (Horeweg, van der Aalst et al. 2013). Additionally, an increasing number of studies investigated the impact of false positive results in lung cancer screening on quality of life, health care resource utilisation and cost-effectiveness of LDCT screening, highlighting that a high rate of false-positive results might lead to substantially increased health care costs, potential risks of radiation exposure in diagnostic follow-ups, such as repeated LDCT scans and PET-CT scans, and other unnecessary invasive diagnostic procedures, for instance, lung biopsies (Goulart and Ramsey 2013, Sozzi and Boeri 2014). It further emphasises the importance of integrating other screening techniques with LDCT to identify and manage instances of false positive results that are clearly associated with LDCT.

Utilising potential biomarkers for preliminary screening to identify possible individuals at high risk for lung cancer before LDCT screening would be highly likely to improve early detection and management of lung cancer, whilst additional biomarkers to address false positive imaging results would also be highly beneficial.

1.2.7 Biomarkers in lung cancer early detection

Although there have been revolutionary new advances in the oncological management of lung cancer over recent years, especially for NSCLC, with substantial progress in screening, diagnosis and therapy, it remains true that most lung cancer patients are diagnosed at a later stage and often have a poor prognosis. Lung cancer screening with LDCT appears to benefit patients. However, the high false positive rate associated with LDCT screening and the subsequent issues, as previously discussed, underline the rationale for and the importance of integrating potential biomarkers to augment LDCT screening to remedy its limitations. Given that a high degree of inter-patient molecular heterogeneity has long been recognised in cancer (Vogelstein, Papadopoulos et al. 2013), multiple predictive biomarkers could play a

key role in identifying asymptomatic individuals at high risk of developing lung cancer as an initial test to help reveal essential information regarding the presence, phenotype and aggressiveness of tumours. Furthermore, such approaches may result in identifying tumours that could respond to, or patients likely to benefit from individualised, targeted therapies, possibly leading to a paradigm shift in lung cancer diagnosis (Kerr, Bubendorf et al. 2014). Additionally, biospecimens for biomarker detection in early stage lung cancer are diverse and most of them are easily accessible (Fig. 1.15).

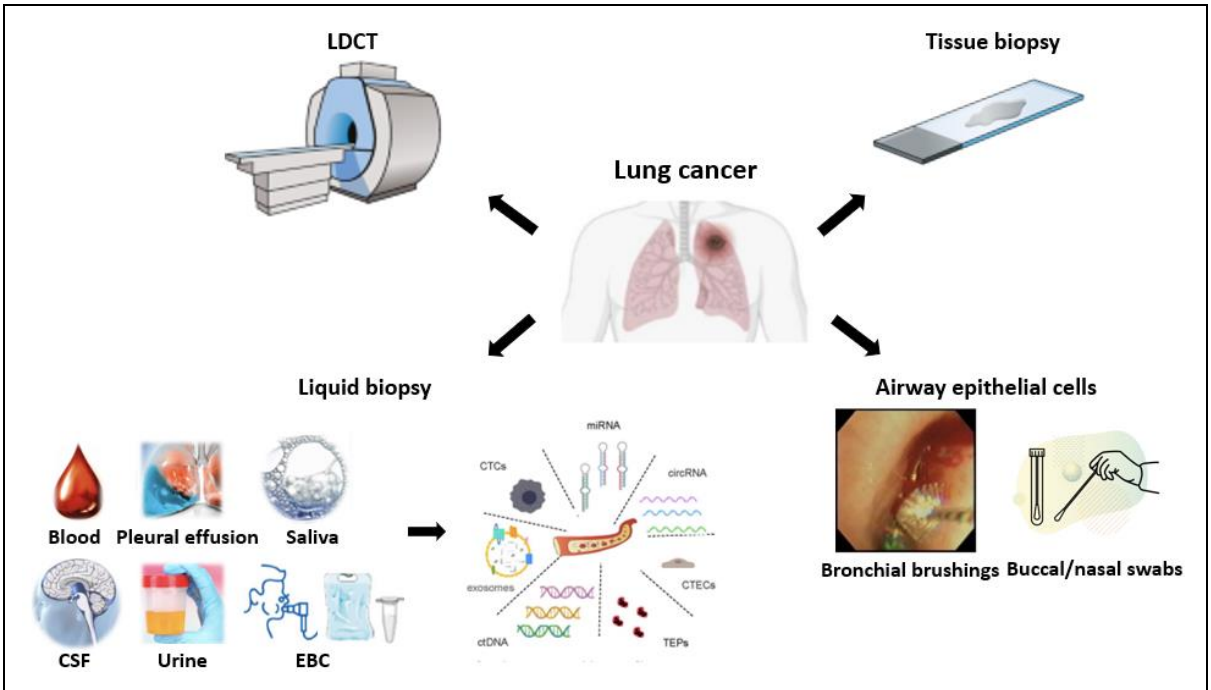


Figure 1.15. Early detection of lung cancer with alternative approaches and specimens.

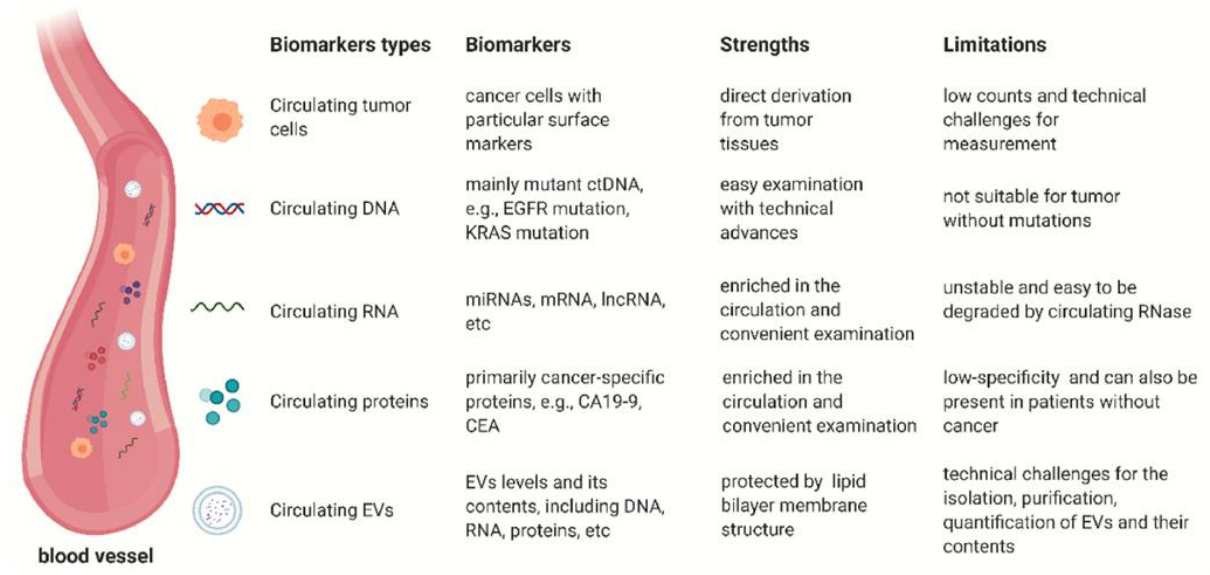
Note: CSF = cerebrospinal fluid, CTCs = circulating tumour cells, ctDNA = circulating tumour DNA, TEPs = tumour-educated platelets, CTECs = circulating tumour vascular endothelial cells, circRNAs = circular RNAs, miRNA = microRNAs [adapted from (Rolfo and Russo 2020, Perez-Sanchez, Barbarroja et al. 2021, Li, Liu et al. 2022, encyclopedia.pub/entry/683 last accessed: 2022)].

Biospecimens can be taken from lung or bronchial tissue, a variety of biofluids [such as peripheral blood and its components, sputum and exhaled breath condensate (EBC)], and also

from airway epithelial cells through bronchial brushings and buccal and nasal swabs, etc. (Davis, Montpetit et al. 2012, Hassanein, Callison et al. 2012, Brothers, Hijazi et al. 2013).

Identification of biomarkers for the early detection of lung cancer has been explored in a wide variety of approaches (Hassanein, Callison et al. 2012). For instance, detection of circulating molecules [e.g. ctDNA, extracellular vesicle (EV)-associated miRNAs, CTCs, proteins and autoantibodies] in blood or blood components (Patz, Campa et al. 2007, Bianchi, Nicassio et al. 2011, Yu, Chen et al. 2013), gene expression biomarker measurements in airway epithelium (Blomquist, Crawford et al. 2009), analysis of DNA methylation signatures in ctDNA (Ponomaryova, Rykova et al. 2013) and splice variant analysis (Higgins, Roper et al. 2012). For example, a study by Higgins et al. showed that an alternative splice variant of the nuclear matrix protein Ciz1 has been widely detected in malignant but not adjacent lung tissues (Higgins, Roper et al. 2012). Nevertheless, significant challenges remain to be overcome before these biomarkers can be regarded as clinically useful diagnostic biomarkers for lung cancer (Zhou, Li et al. 2021). Furthermore, the usage of blood-borne and sputum biomarkers and such have gained much attention attributed to their minimal invasiveness, safety and cost-effectiveness. A further benefit is the possibility to repeat sampling and thereby monitor disease progression and therapeutic efficacy. It remains true, however, that the candidate sputum and blood-borne biomarkers have not yet been clinically validated in lung cancer. In fact, none of the candidate molecular biomarkers for early detection of lung cancer have been validated clinically nor integrated into routine clinical practice despite the enormous efforts devoted to the identification of molecular biomarkers in various biospecimens and the availability of a broad range of these markers in the laboratory (Rodriguez, Ajona et al. 2021). Currently, a blood-based autoantibody biomarker EarlyCDT-Lung test together with two

miRNA signature assays in serum and plasma respectively are in advanced validation stages (Sozzi and Boeri 2014).



Biomarkers types	Biomarkers	Strengths	Limitations
Circulating tumor cells	cancer cells with particular surface markers	direct derivation from tumor tissues	low counts and technical challenges for measurement
Circulating DNA	mainly mutant ctDNA, e.g., EGFR mutation, KRAS mutation	easy examination with technical advances	not suitable for tumor without mutations
Circulating RNA	miRNAs, mRNA, lncRNA, etc	enriched in the circulation and convenient examination	unstable and easy to be degraded by circulating RNase
Circulating proteins	primarily cancer-specific proteins, e.g., CA19-9, CEA	enriched in the circulation and convenient examination	low-specificity and can also be present in patients without cancer
Circulating EVs	EVs levels and its contents, including DNA, RNA, proteins, etc	protected by lipid bilayer membrane structure	technical challenges for the isolation, purification, quantification of EVs and their contents

Figure 1.16. Illustrating strengths and limitations of circulating biomarkers in cancer (Zhou, Li et al. 2021).

Non-invasive or minimally invasive biomarkers based on miRNAs or long non-coding RNAs (lncRNAs) have been of interest due to their roles in several human malignancies (Bhan, Soleimani et al. 2017). Circulating miRNAs as potential biomarkers for lung cancer early detection has attracted much attention, particularly as miRNAs are frequently dysregulated in cancers and play a role in lung cancer growth, invasion, metastasis and recurrence (Lin, Yu et al. 2010, Wang, Ling et al. 2011, Chen, Xu et al. 2012). miRNAs are small (approximately 22 nucleotides in length), single-stranded, evolutionarily conserved, endogenous, non-coding RNA (ncRNA) molecules (Garzon, Calin et al. 2009, Lin, Yu et al. 2010, Wang, Ling et al. 2011). It has long been known that miRNAs are capable of influencing the stability and translation of their target messenger RNAs (mRNAs) through complementary base-pairing despite the fact that their precise mode of action remains the subject of debate (Garzon, Calin et al. 2009,

Wilczynska and Bushell 2015). Multiple studies have shown that dysregulation of miRNA expression plays an essential role in the pathogenesis of most cancers (Di Leva, Garofalo et al. 2014, Hayes, Peruzzi et al. 2014). These small molecules are present and remarkably stable in various body fluids, such as plasma, serum and urine, are therefore good candidates for biomarkers (Mitchell, Parkin et al. 2008, Di Leva, Garofalo et al. 2014, Hayes, Peruzzi et al. 2014).

Lately, a large amount of research has revealed that lncRNAs play a key role in the development and progression of human cancers and some lncRNAs have been proven to be capable of affecting the hallmarks of lung cancer (Bhan, Soleimani et al. 2017, Chen, Zitello et al. 2021). For instance, lncRNA growth arrest-specific transcript 5 (GAS5) is implicated in proliferation and apoptosis in NSCLC. A study by Shi et al. revealed that GAS5 is down-regulated in NSCLC tissues compared to the adjacent noncancerous tissues (Shi, Sun et al. 2015, Rahmani, Mojarrad et al. 2020). Moreover, some lncRNAs are detectable, relatively stable and vastly expressed in a variety of body fluids within CTCs or EVs, highlighting the importance of lncRNAs as potential biomarkers in early stage lung cancer detection (Arita, Ichikawa et al. 2013, Shi, Gao et al. 2016, Naderi-Meshkin, Lai et al. 2019, Rahmani, Mojarrad et al. 2020).

Another important class of biomarkers associated with lung cancer are lung cancer-related antigens, such as carcinoembryonic antigen (CEA). Serum levels of CEA are informative in the prognosis of recurrence and death in NSCLC patients (Grunnet and Sorensen 2012). However, the usage of CEA as a prognostic marker in lung cancer remains in doubt (Grunnet and Sorensen 2012).

1.3 Extracellular Vesicle-associated miRNAs as diagnostic biomarkers in lung cancer early detection

1.3.1 Extracellular Vesicles

Extracellular vesicles (EVs) are lipid bilayer-delimited submicron-sized vesicles that are naturally secreted into the extracellular space by virtually all cell types. EVs have been found in diverse body fluids, including plasma, CSF, bile, lymph, amniotic fluid (Caby, Lankar et al. 2005, Akers, Ramakrishnan et al. 2013, Milasan, Tessandier et al. 2016, Yoon and Chang 2017, Dixon, Sheller-Miller et al. 2018). EVs are categorised into three main subtypes: exosomes, microvesicles (MVs) and apoptotic bodies, based primarily on their size, contents, route of formation, function, release pathways, mode of biogenesis, etc. (Fig. 1.17) (Crescitelli, Lasser et al. 2013, Yanez-Mo, Siljander et al. 2015, Zaborowski, Balaj et al. 2015, Doyle and Wang 2019).

<u>Exosomes</u>

Exosomes are 30-150 nm diameter membrane-encapsulated vesicles of endocytic origin that arise as intraluminal vesicles (ILVs) within the endosomal network (Fig. 1.17) and exhibit characteristic cup-shaped morphology under electron microscopy (Yanez-Mo, Siljander et al. 2015, Zaborowski, Balaj et al. 2015). Fusion of multivesicular bodies (MVBs) with the plasma membrane subsequently releases ILVs into the extracellular microenvironment, which, once extracellular, are termed ‘exosomes’ (Yanez-Mo, Siljander et al. 2015, Zaborowski, Balaj et al. 2015). Exosomes contain numerous bioactive molecules, for instance, nucleic acids (such as DNA, mRNA, miRNA and lncRNAs), various proteins (e.g. receptors and enzyme), lipids and metabolites (Valadi, Ekstrom et al. 2007, Guescini, Genedani et al. 2010, Mashouri, Yousefi et al. 2019). Initially, exosomes were considered to be a mechanism whereby cells eliminate

unneeded materials and waste products (Yanez-Mo, Siljander et al. 2015). However, it has since been revealed in multiple studies that exosomes play a critical role in transporting abundant quantities of a variety of components during exosome-mediated cell-cell communication to transfer information to recipient cells. Additionally, the endosomal sorting complexes required for the transport (ESCRT) pathway is considered one of the most important pathways in both formation and release of exosomes (Doyle and Wang 2019).

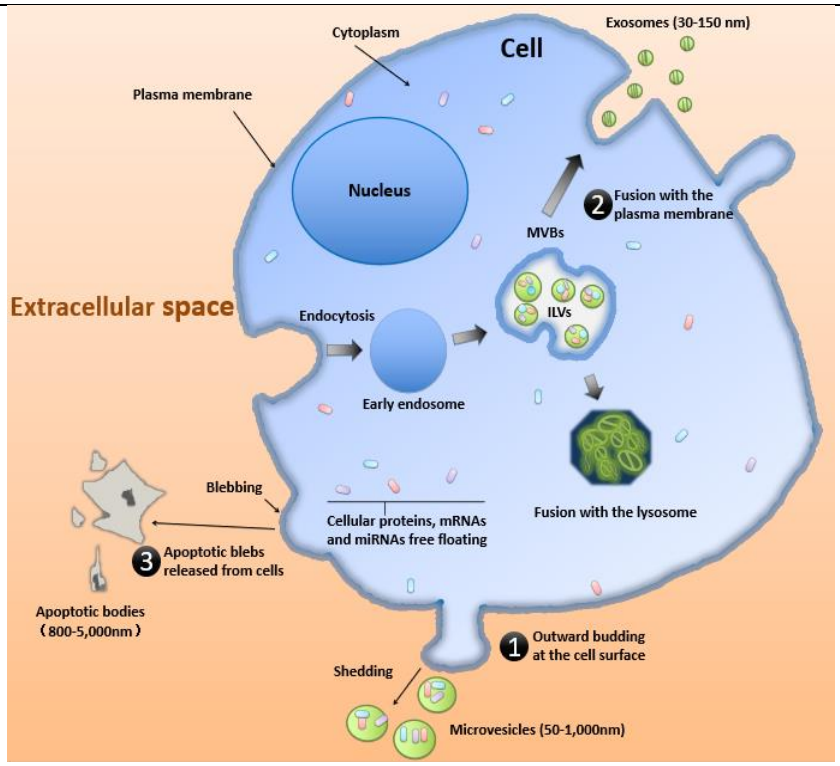


Figure 1.17. Secretion of extracellular vesicles. Based on the generation process, EVs can be classified as three main groups: (1) microvesicles released by outward budding and shedding from plasma membrane; (2) within multivesicular bodies (MVBs), exosomes form as intraluminal vesicles (ILVs) which are released upon MVBs fusion with plasma membrane; (3) the release of apoptotic bodies during apoptosis (Akers, Ramakrishnan et al. 2013, Crescitelli, Lasser et al. 2013, Yanez-Mo, Siljander et al. 2015, Zaborowski, Balaj et al. 2015, Doyle and Wang 2019).

Microvesicles

MVs, also referred to as ectosomes or shedding macrovesicles, are 50-1,000 nm diameter membranous vesicles that are shed from cells and formed by outward blebbing of the plasma membrane (Fig. 1.17) (Crescitelli, Lasser et al. 2013, Yanez-Mo, Siljander et al. 2015, Zaborowski, Balaj et al. 2015, Doyle and Wang 2019). The contents and biological purpose of MVs are similar to those of exosomes. Like exosomes, MVs can also be released by most, if not all, living cells (Crescitelli, Lasser et al. 2013, Yanez-Mo, Siljander et al. 2015, Zaborowski, Balaj et al. 2015, Doyle and Wang 2019). Furthermore, multiple studies have indicated that the number of MVs produced and consumed depends on the physiological state and microenvironment of the donor cell and the recipient cells, respectively (Zaborowski, Balaj et al. 2015). Moreover, there is some evidence that the uptake of MVs (and exosomes) is temperature dependent (suppressed at lower temperatures) and Ca2+ dependent. These findings showed that the uptake of MVs (and exosomes) is mediated by endocytosis in an energy dependent process (Morelli, Larregina et al. 2004, Escrevente, Keller et al. 2011, Christianson, Svensson et al. 2013).

Apoptotic bodies

Apoptotic bodies (800-5,000 nm diameter) are released by fragmented apoptotic cells into the extracellular environment. These vesicular apoptotic bodies are formed from the plasma membrane and cellular fragments in the process of programmed cell death, representing therefore, the remnants of dying cells (Fig. 1.17) (Stahl, Johansson et al. 2019, Battistelli and Falcieri 2020). Under specific conditions, apoptotic bodies can be more abundant than MVs or exosomes and their appearance can vary in terms of size, composition and structure (Battistelli and Falcieri 2020). Moreover, they may contain a broad range of cellular

components, including intact or degraded proteins and/or organelles, DNA fragments, RNA, cytosol portions, chromatin and lipids (Doyle and Wang 2019, Stahl, Johansson et al. 2019, Battistelli and Falcieri 2020).

Functions of EVs in normal physiology

It is widely accepted that EVs deliver encapsulated biomolecules between cells as a mechanism of intercellular communication (Stahl, Johansson et al. 2019). Therefore, the functions of EVs vary depending on the cargos that have been transferred from the parent cells to the recipient cells (Stahl, Johansson et al. 2019). It is reported that many EV protein and lipid cargos are implicated in cell signalling by interacting with receptors on target cells (Greco, Hannus et al. 2001). For instance, a study by Greco et al. indicated that morphogen protein wingless (Wg) exposed by microvesicles partakes in the formation of a morphogen gradient, which is critical for tissue patterning, by binding to G protein-coupled receptors (GPCRs) (Greco, Hannus et al. 2001). Apart from that, some proteins and lipids transported by EVs are associated with the modulation of target cells (Stahl, Johansson et al. 2019). It has also been established that EVs can facilitate cell signalling in the original absence of the necessary receptors or boost the number of receptors by transporting functional receptors to target cells (Baj-Krzyworzeka, Majka et al. 2002). It is reported that the adhesion capability, proliferation and survival of hematopoietic cells can be modulated by delivering adhesion molecules and receptors from platelets to hematopoietic cells (Baj-Krzyworzeka, Majka et al. 2002). In addition, through the transportation of mRNA and miRNA cargos, EVs are capable of exerting a substantial influence on the phenotype of the recipient cells (Stahl, Johansson et al. 2019). There is evidence that EVs can transfer functional miRNAs to the recipient cells to inhibiting the roles of mRNA, therefore EVs may implicated in the regulation of mRNA levels

in recipient cells (Pegtel, Cosmopoulos et al. 2010, Mittelbrunn, Gutierrez-Vazquez et al. 2011, Montecalvo, Larregina et al. 2012). Moreover, some studies show EVs may contribute to homeostasis restoration by stimulating compensatory proliferation and repair in neighbouring recipient cells when cell injury occurs (Bussolati and Camussi 2017, Gupta, Goldufsky et al. 2017).

1.3.2 EVs in cancer

EVs have been reported to play a role in a broad variety of diseases, such as neurodegenerative diseases (Perez-Gonzalez, Gauthier et al. 2012), cardiovascular diseases (Su, Li et al. 2020), lung diseases (Yin, Shelke et al. 2020), diabetes (Wu, Noren Hooten et al. 2020), kidney diseases (Gildea, Seaton et al. 2014) and cancers (Hasan, Sohal et al. 2022).

The biological functions of EVs in cancer have captured the interest of many scientists in recent years and emerging evidence suggests that EVs are involved in the multistep process of cancer development, including proliferation, angiogenesis, invasion, progression, pre-metastatic niche generation, migration and immune escape of cancer cells (Peinado, Aleckovic et al. 2012, Lopatina, Gai et al. 2016).

In many cancer types, EVs are found to facilitate cell proliferation, invasion and drug resistance in an autocrine manner through activating signalling pathways, for instance, the phosphatidylinositol-3-kinase (PI3K)/Akt and the mammalian target of rapamycin (mTOR) signalling pathways, which was reported in NSCLC research (Choi, You et al. 2014). Nucleophosmin (NPM), a ubiquitously expressed nucleolar phosphoprotein, which is also a remarkably enriched oncoprotein detected in EVs in a number of cancers, has been reported to involved in pathways associated with growth suppression and proliferation (Carvalho,

Baeta et al. 2020). Studies also revealed that EVs isolated from tumours were found to condition the tumour microenvironment (Conde-Vancells, Rodriguez-Suarez et al. 2008), indicating possible distant signalling for metastatic spread (Salido-Guadarrama, Romero-Cordoba et al. 2014) and involvement in cancer growth and development ascribed to the composition of the tumour microenvironment, including fibroblasts, endothelial and immune cells.

In many in vivo and in vitro studies EVs have been reported to play an essential role in cell migration in a number of cancers, including melanoma (Hao, Ye et al. 2006), nasopharyngeal carcinoma (Aga, Bentz et al. 2014) and breast cancer (McCready, Sims et al. 2010). Tumour-derived EVs may prime distant organs through the formation of pre-metastatic niche and colonisation to facilitate metastasis by modulating the tumour microenvironment (Hoshino, Costa-Silva et al. 2015). EVs can also modulate cell invasion and metastasis by transferring molecules that enhance migration. For example, tumour-derived EVs were shown to contribute to metastasis by transferring miRNAs (e.g. miR-145 and miR-542-3P) that can affect cell–cell adhesion molecules and angiogenesis-promoting proteins (Aleckovic and Kang 2015, Ma, Wu et al. 2020).

EVs also participate in tumour development through promoting acquisition of the other "hallmarks of cancer" (Hanahan and Weinberg 2000, Hanahan and Weinberg 2011, Hanahan 2022). Tumour-derived EVs play an important role in tumour progression by partaking in the inhibition of immune surveillance and the evasion of immune destruction. For example, it is reported that tumour-derived EVs are associated with the evasion of the immune destruction of tumour cells through suppressing T cell immune responses, promoting regulatory T cells (Treg) expansion and inducing T cell apoptosis (Wieckowski, Visus et al. 2009, Ma, Vayalil et

al. 2021). Apart from this, tumour-derived EVs may also influence tumour angiogenesis, which is key to cancer progression. It is found that tumour cells tend to release more EVs in a hypoxic tumour microenvironment than cells in a normoxic environment highly likely associated with uncontrollable tumour proliferation (Kuriyama, Yoshioka et al. 2020). In addition, tumour-derived EVs contain pro-angiogenic mediators, such as vascular endothelial growth factor (VEGFA) and fibroblast growth factor 2 (FGF2) (Burgos-Ravanal, Campos et al. 2021).

1.3.3 EV miRNAs in lung cancer

Much focus has been given to circulating miRNAs as biomarkers in lung cancer diagnosis (Bianchi, Nicassio et al. 2011, Boeri, Verri et al. 2011, Hennessey, Sanford et al. 2012, Hannafon and Ding 2013). Several circulating miRNAs appeared to be differentially expressed in lung cancer patients compared to control subjects (Wozniak, Scelo et al. 2015), highlighting the potential of EV miRNAs as biomarkers in lung cancer diagnosis. For instance, miRNAs including miRNA-21, -126, -210, and 486-5p were reported to be differentially expressed in both plasma and tissue samples of NSCLC patients compared to healthy individuals (plasma only)(Shen, Todd et al. 2011). Moreover, the panel of the four miRNAs has the potential to identify NSCLC patients with high specificity (96.55%) and sensitivity (73.33%) in the identification of NSCLC patients at early stage (stage I) (Shen, Todd et al. 2011). Other circulating miRNAs in plasma, including miR-944 and miR-3662, were also reported to be up-regulated in NSCLC patients compared to healthy individuals. It is worth mentioning that the expression levels of these two miRNAs were not only up-regulated in NSCLC but also correlated with the tumour stages (Powrozek, Krawczyk et al. 2015). Besides, miRNA expression analysis to improve lung cancer diagnosis has been explored in LDCT screening studies, where two panels of miRNAs were identified for prior to lung cancer detection by CT

scan (15 miRNAs) and at the time of lung cancer detection by CT scan (13 miRNAs) (Bianchi, Nicassio et al. 2011, Boeri, Verri et al. 2011). The sensitivity and the specificity of the identified miRNAs were found to be significant with 80% sensitivity and 90% specificity for the 'plasma samples collected prior to lung cancer detection by CT scan' group and 75% and 100% respectively for the 'plasma samples collected at surgery or at the time of lung cancer detection by CT scan' group (Boeri, Verri et al. 2011). The lower sensitivity observed for the "at the time of detection" group was assumed to be associated with the presence of more small, indolent, and/or early stage nodules in this analysis (Boeri, Verri et al. 2011). So far, 109 circulating miRNAs were reported to be statistically significant for lung cancer diagnosis (Powrozek, Krawczyk et al. 2015). Among those 109 circulating miRNAs, 9 were not part of any panels, the median sensitivity and specificity of the multiple miRNAs-based panels appeared higher compared to those of individual miRNAs (Yu, Guan et al. 2018). However, it remains unclear if circulating miRNAs are protein bound (stabilised) nucleic acids or protected within exosomes, as until recently, most extraction techniques used did not discriminate between different forms.

In the study carried out by Cazzoli et al. (Cazzoli, Buttitta et al. 2013, Zhang, Qin et al. 2019) investigating EV-derived miRNAs from 30 plasma samples obtained from patients with lung adenocarcinomas (10pt), lung granulomas (10pt), and ten healthy smokers as negative controls. miRNAs analysis resulted in the identification of a wide-range of miRNAs, among those confirmed miRNAs, miR-378a, miR-379, miR-139-5p, and miR-200b-5p were found to be useful in the identification of nodule (lung adenocarcinomas + carcinomas) and non-nodule (healthy former smokers). Furthermore, miR-151a-5p, miR-30a-3p, miR-200b-5p, miR-629,

miR-100, and miR-154-3p were beneficial in discriminating those of lung adenocarcinoma and granuloma (Cazzoli, Buttitta et al. 2013).

Differential expression patterns of EVs-derived miRNAs have been reported among lung cancer patients with lung abnormalities detected by CT scan, including patients whose CT scans revealed solid or ground-glass nodules, with the latter differentiated based on their structure as pure or mixed (Zhang, Qin et al. 2019). Interestingly, patients with pure ground-glass nodules appeared to be distinct from the other two groups. This discovery is believed to benefit the diagnosis of small pulmonary nodules on CT with ground-glass opacity (GGO). It is well known that the determination of this type of nodules remains quite a challenge due to their potential malignancy and heterogenous characteristics (Zhang, Qin et al. 2019). A recently published report (Song, Wang et al. 2019) from meta-analysis investigating the diagnostic accuracy of liquid exosomes for lung cancer in 13 eligible published articles comprising 1338 lung cancer patients and 1075 paired controls suggests that exosomes have potential as a novel marker in lung cancer diagnosis. Additionally, a higher level of circulating miRNAs are reported in lung cancer patients compared to healthy individuals (Rabinowits, Gercel-Taylor et al. 2009).

Taken together, there is significant evidence that the usage of EV miRNAs and other circulating miRNAs as minimally invasive tools to be integrated with LDCT screening in early detection of lung cancer appears to be promising. However, more large-scale clinical trials are required to validate and optimise the candidate miRNA biomarkers for lung cancer diagnosis. The most up-to-date potential biomarkers reported in lung cancer are summarised in table 1.2.

Table 1.2. Potential biomarkers reported in lung cancer

Study	Type of Lung Cancer	Types of biomarkers DNA/RNA/protein	Sample	Method	Sensitivity	Specificity	Potential utility
(Powrozek, Krawczyk et al. 2015)	LC	miR-944, -3662	Plasma	qRT-PCR	82	92	Histologic type-specific early lung cancer biomarkers
	NSCLC				92	86	
(Powrozek, Krawczyk et al. 2016)	LC	miR-48,-4478	Plasma	qRT-PCR	89	79	Non-invasive LC biomarkers
	NSCLC				90	76	
(Zaporozhchenko, Morozkin et al. 2016)	LC	miR-19b, -183	Plasma	qRT-PCR	94.70%	95.20%	Discriminate histological subtypes of LC and with higher reliably
(Thai, Statt et al. 2013)	LC	LncRNA SCAL1	Whole blood	qRT-PCR	NA	NA	Diagnostic LC biomarker
(Weber, Johnen et al. 2013)	NSCLC	LncRNA MALT1	Whole blood	qRT-PCR	56%	96%	Diagnostic LC biomarker
(Guo, Yu et al. 2015)	LC			qRT-PCR	NA	NA	Diagnostic NSCLC biomarker
(Deng, Feng et al. 2017)	Any LC	LncRNA loc146880 and lc3b	Serum	qRT-PCR	NA	NA	Diagnostic NSCLC biomarker
(Tantai, Hu et al. 2015)	NSCLC	LncRNA XIST	Serum	qRT-PCR	NA	NA	Predictive biomarker for NSCLC screening
(Xie, Zhang et al. 2018)	NSCLC	LncRNA SOX2OT and ANRIL	Serum	qRT-PCR	77.10%	79.20%	Diagnostic and prognostic biomarkers in NSCLC patients
(Zhu, Zhang et al. 2017)	LC	LncRNA 16	Plasma	qRT-PCR	NA	NA	Biomarker in early diagnosis of lung cancer
(Liang, Lv et al. 2016)	NSCLC	LncRNA GAS5	Plasma	qRT-PCR	82.20%	72.70%	Biomarker for the diagnosis of NSCLC
(Wang, Lu et al. 2015)	NSCLC	LncRNA-UCA1	Plasma	qRT-PCR	NA	NA	Clinical diagnosis as a predictive biomarker in NSCLC patients
(Hu, Bao et al. 2016)	NSCLC	LncRNA SPRY4-IT1	Plasma	qRT-PCR	NA	NA	Diagnostic and prognostic biomarker in NSCLC Patients
(Hu, Bao et al. 2016)	NSCLC	LncRNA ANRIL	Plasma	qRT-PCR	82.80%	92.30%	Diagnostic and prognostic biomarker in NSCLC Patients
(Tang, Ni et al. 2015)	NSCLC	LnRNA RP11-397D12.4	Plasma	qRT-PCR	NA	NA	Non-invasive screening biomarker for NSCLC
(Tang, Ni et al. 2015)	NSCLC	LnRNA-AC007403.1	Plasma	qRT-PCR	NA	NA	Non-invasive screening biomarker for NSCLC
(Tang, Ni et al. 2015)	NSCLC	LnRNA-ERICH1-AS1	Plasma	qRT-PCR	NA	NA	Non-invasive screening biomarker for NSCLC

1.4 Project background

The Liverpool Lung Project (LLP), a case/control and population cohort (Field, Smith et al. 2005, Cassidy, Myles et al. 2008), was used previously to select a panel of 18 EV miRNAs on the basis of their expression in lung cancer cases compared to controls. This was achieved by differential expression analysis using the HTG EdgeSeq platform [as a service by HTG Molecular Diagnostics, Inc. (AZ, USA)] as part of a Roy Castle Lung Cancer Foundation (RCLCF)-funded pilot study undertaken by Dr. Lakis Liloglou and Dr. Michael Marcus in 2016.

EVs were isolated from 2ml plasma samples of 60 lung cancer patients and 60 age/sex matched control subjects recruited in the LLP (Field, Smith et al. 2005, Cassidy, Myles et al. 2008) using a standard differential ultracentrifugation protocol, half the EV fraction was used for HTG EdgeSeq analysis and half for quality control. EV quality was assessed by nanoparticle tracking analysis (NTA), transmission electron microscopy (TEM) and western blot analysis for known EV and exosome proteins (Flot-2, CD63, Tsg101, EGFR, SORT1, Alix) and negative controls (Cyc C1 and Calnexin) (Fig. 1.18). Following HTG EdgeSeq quantitation of 2,083 miRNA, plus control probes, 23 samples were discounted based on ANT QC (HTG 2019), leaving 49 cases and 48 controls.

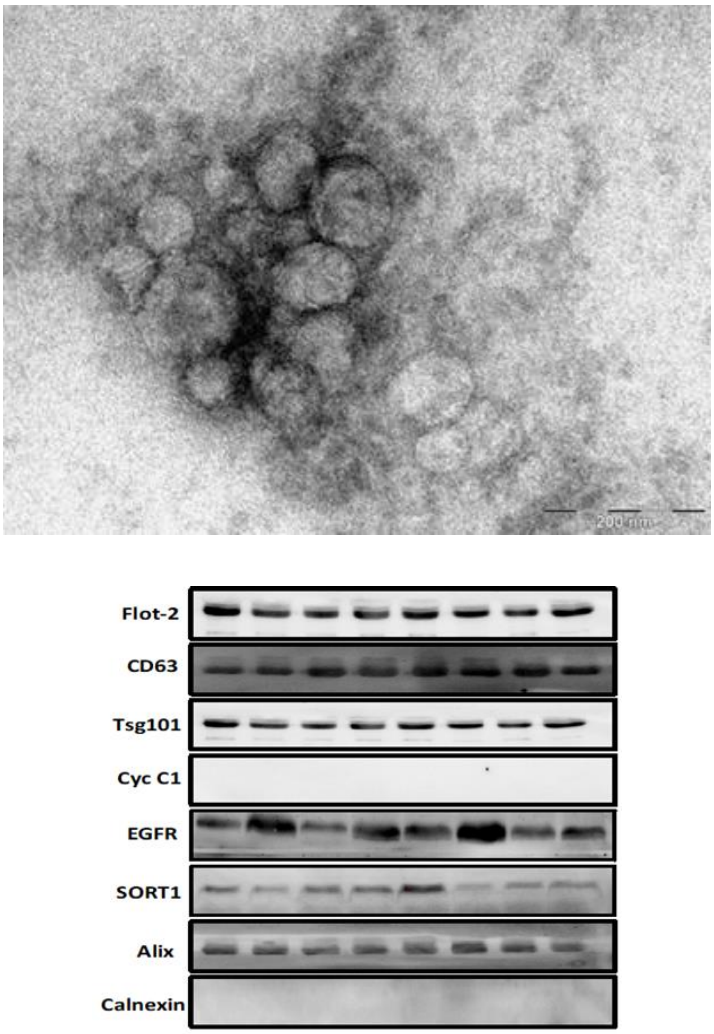


Figure 1.18. EV validation by electron microscopy and western blot. Electron micrograph of EVs isolated from plasma specimens of lung patients by differential ultracentrifugation, as used for HTG EdgeSeq analysis. Western blot analysis of EV fractions isolated by differential ultracentrifugation. The data was generated by Dr. Amelia Acha-Sagredo in the pilot study undertaken by Dr. Lakis Liloglou in 2016.

The epidemiology, clinical data and follow-up data were acquired from the LLP. Aiming to select the panel of miRNAs that best predict lung cancer, the Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression was implemented to provide model fitting and model building for generalized linear models. Based on the bioinformatic analysis using Mann-Whitney U test, the best predicting model of lung cancer was composed of four

miRNAs (miR-1277-5p, miR-509-3-5p, miR-514b-5p, miR-582-5p) with an AUC of 0.84 and a bias corrected AUC of 0.82 on ROC analysis and internal validation by ROC analysis with leave-one-out cross-validation. It should be noted that there are further steps required to translate the promising outcomes of this pilot study into a clinical tool. For instance, it is well established in biomarker studies that corroborating the results of the discovery phase in a larger independent cohort of patients is essential for validation, ideally by a different methodological approach. This is especially true when the initial approaches taken (both the differential ultracentrifugation and the HTG EdgeSeq assay) are not suitable for routine clinical use. It was therefore proposed to employ column-based EV miRNA isolation and qRT-PCR as a methodology for miRNA detection in the validation phase. In addition, it was decided that the panel to be tested should be expanded beyond the four miRNAs identified in the best predicting model, to allow for substitution if any miRNA biomarkers dropout or to provide a larger panel, in case that proves more robust with the alternative quantitation technology.

Subsequently, 18 miRNA targets were selected to be assayed for their diagnostic accuracy, including the miRNAs that contributed to the best predicting model of lung cancer obtained from the pilot study and the next best predictive miRNAs, whilst minimising the redundancy by excluding any miRNAs that are closely correlated with those selected already. We also exclude those miRNAs that are strongly associated with sex, age, smoking status, and COPD, which might lead to spurious associations, as these are risk factors for lung cancer.

Apart from the validation by qPCR, we have also re-analysed the original HTG EV miRNA data, as the analysis software was updated by HTG during my PhD, providing improved quality control checks and differential expression methods (integrated into an online tool: REVEAL software: www.htgmolecular.com/reveal). In addition, this new software was used to check

whether batch effects and statistical analysis methods have an impact on the selection and validation of miRNA panels for lung cancer diagnosis. We performed a comparison between the three EV miRNA datasets (HTG EV miRNA original analysis and re-analysis, qPCR validation results of EV miRNA) to address technical issues. Furthermore, we obtained an HTG dataset using whole plasma and performed some comparisons to the EV miRNA data, providing biological insight, as well as investigating the relative enrichment and depletion of miRNA between whole plasma and its EV fraction.

1.5 Thesis hypotheses

It is hypothesized that miRNA and other types of RNA cargo of cancer-EVs may provide a sensitive and specific biomarker panel that can be used to detect lung cancer at an early stage and/or assist in distinguishing cancerous from non-cancerous nodules detected by LDCT.

1.6 Aims and objectives

This thesis aims to identify panels of miRNAs in human plasma EV fractions which might be used as minimally invasive biomarkers for screening and early detection of lung cancer.

The specific objectives include:

a. qPCR validation of a panel of 18 preselected miRNAs in the plasma EV fraction, identified by logistic regression analysis.
b. Further analysis of the EV fraction data by alternative statistical approaches.
c. Comparison of miRNA expression between whole plasma and EV fractions.

Chapter 2 Materials and Methods

2.1 Validation of preselected miRNAs in plasma EV as diagnostic biomarkers for lung cancer by qRT-PCR

2.1.1 Optimisation and validation of assay conditions

It is essential to optimise and validate the assay conditions before testing the targeted miRNAs using our matched samples. In the initial stage of the workflow, a selection of assay conditions was assessed, optimised and validated.

Reagents

TaqMan™ Advanced miRNA Assays and the TaqMan® Advanced miRNA cDNA Synthesis Kit were purchased from Applied Biosystems. The miRCURY LNA Probe PCR Kit and QuantiNova Multiplex RT-PCR Kit were purchased from Qiagen. The exoEasy Midi Kit was purchased from Qiagen.

Instruments

GeneAmp PCR System 9700 (Applied Biosystems) was used for RT reaction. 7500 Real-Time PCR System and 7500 Fast Real-Time PCR System (Applied Biosystems) were used for performing qPCR reactions. NanoDrop (Thermo Scientific) was used to measure the concentration of the samples and the calibrators. miVac DNA concentrator (Genevac) was used to dry down RNA samples to be used for qRT-PCR.

2.1.1.1 Testing QuantiNova PCR reaction mixes and pre-amplification

For the qPCR assay to be used for the validation cohort, we tested the primers and probes from the TaqMan™ Advanced miRNA Assays with two types of reaction mix: 2x QuantiNova

Probe PCR Master Mix (miRCURY LNA Probe PCR Kit) and 4x QuantiNova Multiplex RT-PCR Master Mix (QuantiNova Multiplex RT-PCR Kit).

Total EV RNA extracted from cell lines using the exoEasy Midi Kit (provided by Dr Liloglou) was used in 4-fold serial dilutions for RT reactions and subsequent miR-Amp reactions. The concentration of the extracted RNA was 40 ng/µl, therefore, the concentrations of the dilutions were: 10 ng/µl; 2.5 ng/µl; 0.625 ng/µl and 0.15625 ng/µl, respectively. The TaqMan® Advanced miRNA cDNA Synthesis Kit was used to prepare the cDNA template. The products of RT reaction (cDNA) and the subsequent miR-Amp reaction (miR-Amp reaction product) were utilised to test miRCURY LNA RT Kit and QuantiNoMultiplex RT-PCR Kit, with the TaqMan™ Advanced miRNA Assay probes and primers for miR-99a-5p.

<u>The poly[A] tailing reaction</u>

1 Thawed samples and cDNA synthesis reagents on ice, gently vortexed, then centrifuged briefly.

2 In a 1.5ml microcentrifuge tube, prepared sufficient Poly(A) Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
10X Poly(A) Buffer	0.5 µL	2.2 µL	5.5 µL
ATP	0.5 µL	2.2 µL	5.5 µL
Poly(A) Enzyme	0.3 µL	1.3 µL	3.3 µL
RNase-free water	1.7 µL	7.5 µL	18.7 µL
Total Poly(A) Reaction Mix volume	**3.0 µL**	**13.2 µL**	**33 µL**

[1] Note: Volumes include 10% overage.

3 Vortexed the Poly(A) Reaction Mix, then centrifuged briefly.

4 Added 2µl of RNA sample to RNase-free, low-bind reaction tubes, then transferred 3µl of Poly(A) Reaction Mix to each tube. The total volume was 5µl per reaction tube.

5 Sealed the reaction tubes, then vortexed briefly to thoroughly mix the contents.

6 Centrifuged the reaction tubes briefly to spin down the contents and eliminate air bubbles.

7 Placed the reaction tubes into a thermal cycler, then incubate using the following settings and standard cycling:

Step	Temperature	Time
Polyadenylation	37°C	45 minutes
Stop reaction	65°C	10 minutes
Hold	4°C	Hold

<u>The adaptor ligation reaction</u>

1 In a 1.5ml microcentrifuge tube, prepared sufficient Ligation Reaction Mix for the required number of reactions according to the following table, 50% PEG 8000 was used at room temperature (15-25°C) and aspirated and dispensed slowly to ensure accurate pipetting.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
5X DNA Ligase Buffer	3 µL	13.2 µL	33 µL
50% PEG 8000[2]	4.5 µL	19.8 µL	49.5 µL
25X Ligation Adaptor	0.6 µL	2.6 µL	6.6 µL
RNA Ligase	1.5 µL	6.6 µL	16.5 µL
RNase-free water	0.4 µL	1.8 µL	4.4 µL
Total Ligation Reaction Mix volume	**10 µL**	**44 µL**	**110 µL**

[1] Note: Volumes include 10% overage.

2 Vortexed the Ligation Reaction Mix, then centrifuged briefly.

3 Transferred 10µl of the Ligation Reaction Mix to each reaction tube containing the poly(A) tailing reaction product. The total volume was 15µl per reaction tube.

4 Sealed the reaction tubes, then vortexed briefly to thoroughly mix the contents.

5 Centrifuged the reaction tubes briefly to spin down the contents.

6 Placed the reaction tubes into a thermal cycler, then incubated using the following settings and standard cycling:

Step	Temperature	Time
Ligation	16°C	60 minutes
Hold	4°C	Hold

<u>The reverse transcription (RT) reaction</u>

1 In a 1.5ml microcentrifuge tube, prepared sufficient RT Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
5X RT Buffer	6 µL	26.4 µL	66 µL
dNTP Mix (25 mM each)	1.2 µL	5.3 µL	13.2 µL
20X Universal RT Primer	1.5 µL	6.6 µL	16.5 µL
10X RT Enzyme Mix	3 µL	13.2 µL	33 µL
RNase-free water	3.3 µL	14.5 µL	36.3 µL
Total RT Reaction Mix volume	**15 µL**	**66 µL**	**165 µL**

[1] Note: Volumes include 10% overage.

2 Vortexed the RT Reaction Mix, then centrifuged briefly.

3 Transferred 15µl of the RT Reaction Mix to each reaction tube containing the adaptor ligation reaction product. The total volume was 30µl per reaction tube.

4 Sealed the reaction tubes, then vortexed briefly to thoroughly mix the contents.

5 Centrifuged the reaction tubes briefly to spin down the contents.

6 Placed the reaction tubes into a thermal cycler, then incubated using the following settings and standard cycling:

Step	Temperature	Time
Reverse transcription	42°C	15 minutes
Stop reaction	85°C	5 minutes
Hold	4°C	Hold

<u>The miR-Amp reaction</u>

1 In a 1.5ml microcentrifuge tube, prepared sufficient miR-Amp Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
2X miR-Amp Master Mix	25 µL	110 µL	275 µL
20X miR-Amp Primer Mix	2.5 µL	11 µL	27.5 µL
RNase-free water	17.5 µL	77 µL	192.5 µL
Total miR-Amp Reaction Mix volume	**45 µL**	**198 µL**	**495 µL**

[1] Note: Volumes include 10% overage.

2 Vortexed the miR-Amp Reaction Mix, then centrifuged briefly.

3 Transferred 45µl of the miR-Amp Reaction Mix to a new reaction tube.

4 Added 5µl of the RT reaction product to each reaction tube. The total volume was 50µl per reaction tube.

5 Sealed the reaction tubes, then vortexed briefly to thoroughly mix the contents.

6 Centrifuged the reaction tubes briefly to spin down the contents.

7 Placed the reaction tubes into a thermal cycler, then incubated using the following settings MAX ramp speed, and standard cycling:

Step	Temperature	Time	Cycles
Enzyme activation	95°C	5 minutes	1
Denature	95°C	3 seconds	14
Anneal/Extend	60°C	30 seconds	
Stop reaction	99°C	10 minutes	1
Hold	4°C	Hold	1

Real-time PCR

1 Thawed the assays on ice, gently vortexed and then centrifuged briefly.

2 Prepared 1:5 dilution of RT reaction product and 1:10 dilution of miR-Amp reaction product.

3 In a 1.5ml microcentrifuge tube, prepared sufficient PCR Reaction Mix for the required number of reactions according to the following table.

Using 2x QuantiNova Probe PCR Master Mix

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
2 × QuantiNova Probe PCR Master Mix	10 µL	44 µL	110 µL
miRNA assay miR-99a-5p	1 µL	4.4 µL	11 µL
Rox Reference Dye	0.1 µL	0.44 µL	1.1 µL
molecular biology grade water	5.9 µL	25.96 µL	64.9 µL
Total PCR Reaction Mix volume	**17 µL**	**74.8 µL**	**187 µL**

[1] Note: Volumes include 10% overage.

Using 4x QuantiNova Multiplex RT-PCR Master Mix

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
QuantiNova Multiplex RT-PCR Master Mix	5 µL	22 µL	55 µL
miRNA assay miR-99a-5p	1 µL	4.4 µL	11 µL
Rox Reference Dye	0.1 µL	0.44 µL	1.1 µL
molecular biology grade water	10.9 µL	47.96 µL	119.9 µL
Total PCR Reaction Mix volume	**17 µL**	**74.8 µL**	**187 µL**

[1] Note: Volumes include 10% overage.

4 Vortexed the PCR Reaction Mix, then centrifuged briefly.

5 Transferred 17µl of the PCR Reaction Mix to each well of a PCR reaction plate.

6 Added 3 µl of the diluted RT reaction product/miR-Amp product to each reaction well of the plate. The total volume was 20µl per reaction well. All reactions were run in duplicate; no DNA controls were also included.

7 Thoroughly mixed the contents by pipetting up and down a few times, then sealed the reaction plate.

8 Centrifuged the reaction plate briefly to spin down the contents.

9 Set up and ran the real-time PCR instruments (7500 Fast systems).

(1) Loaded the reaction plate in the real-time PCR instrument.

(2) Set the appropriate experiment settings and PCR thermal cycling conditions.

Step	Temperature	Time	Cycles
Enzyme activation	95°C	2 mins	1
Denature	95°C	5 secs	45
Anneal / Extend	60°C	30 secs	

(3) Set the reaction volume appropriate for the reaction plate.

(4) Started the run.

(5) Applied Biosystems 7500 Software v2.3 (Life Technologies) was used for data collection, with the appropriate filter setting for the FAM TaqMan probes used and the Rox reference dye.

(6) The same software was used for data analysis, with automated baseline subtraction and thresholds. Occasionally baseline start and stop positions and thresholds were adjusted to avoid artefacts such as unexplained spikes or baselines drift.

2.1.1.2 Testing efficiency, sensitivity, and linearity of TaqMan qRT-PCR

The same RT reaction products prepared in the experiment described in Section 2.1.1.1 (RNA extracted from cell line and its 4-fold serial dilutions were used for RT reaction) were used to test efficiency, sensitivity, and linearity of TaqMan™ Advanced miRNA Assays for real-time PCR with miR-99a-5p. The whole process of preparing cDNA template was the same as Section 2.1.1.1, except that no miR-Amp step was included.

Real-time PCR

1 Thawed the assays on ice, gently vortexed and then centrifuged briefly.

2 In a 1.5ml microcentrifuge tube, prepared sufficient PCR Reaction Mix for the required number of reactions according to the following table.

Using TaqMan™ Advanced miRNA Assays

Component	1 Rxn	4 Rxns[1]
TaqMan® Fast Advanced Master Mix (2X)	10 µL	44.0 µL
TaqMan® Advanced miRNA Assay (20X)	1 µL	4.4 µL
RNase-free water	4 µL	17.6 µL
Total PCR Reaction Mix volume	**15 µL**	**66 µL**

[1] Note: Volumes include 10% overage.

3 Vortexed the PCR Reaction Mix, then centrifuged briefly.

4 Transferred 15µl of the PCR Reaction Mix to each well of a PCR reaction plate.

5 Added 5µl of the diluted (1:5) cDNA template to each reaction well of the plate. The total volume was 20µl per each reaction well. All reactions were run in duplicate; no DNA controls were also included.

6 Thoroughly mixed the contents by pipetting up and down a few times, then sealed the reaction plate.

7 Centrifuged the reaction plate briefly to spin down the contents.

8 Set up and ran the real-time PCR instruments (7500 Fast systems).

(1) Loaded the reaction plate in the real-time PCR instrument.

(2) Set the appropriate experiment settings and PCR thermal cycling conditions.

Step	Temperature	Time	Cycles
Enzyme activation	95°C	2 mins	1
Denature	95°C	3 secs	45
Anneal / Extend	60°C	30 secs	

(3) Set the reaction volume appropriate for the reaction plate.

(4) Started the run.

(5) Analysed as above.

2.1.1.3 Testing efficiency, sensitivity, linearity, and reproducibility of QuantiNova qRT-PCR

The 2x QuantiNova Probe PCR Master Mix was retested (as described in Section 2.1.1.1, but only RT reaction products were used, therefore the only difference compared to Section 2.1.1.1 was no miR-Amp reaction included) to investigate whether the RT products were contaminated and to investigate reproducibility. 5 miRNAs (miR-342-5p, miR-26a-1-3p, miR-1185-5p, miR-146b-5p and miR-451a) were tested with the 2x QuantiNova Probe PCR Master Mix.

2.1.1.4 Assessment of miRNA levels in plasma and normal lung tissue using two PCR methods

Three miRNAs (miR-342-5p, miR-1185-5p, and miR-451a) were tested with 2x QuantiNova Probe PCR Master Mix and the other three (miR-146b-5p, miR-99a-5p, and miR-26a-1-3p) with 4x QuantiNova Multiplex RT-PCR Master Mix using RNA extracted from 6 plasma samples (P1-P6) and 4 normal lung tissue samples (N1-N4), using the exoEasy Midi Kit (provided by Dr Liloglou). Additionally, a 2-Step qPCR method (initial denaturation at 95°C for 2 minutes followed by 45 cycles of 95°C for 5 seconds and 60°C for 30 seconds) and a 3-Step qPCR method (initial denaturation at 95°C for 2 minutes followed by 45 cycles of 95 °C for 5 seconds, 59 °C for 15 seconds, and 60 °C for 40 seconds.) were tested.

The concentration of the RNA samples is shown below and 10µl of samples P1-P6, 2µl of samples N1-N3 and 3µl of N4 were dried down using the miVac DNA concentrator for 1 hour.

The whole process of preparing cDNA template and PCR reaction was the same as Section 2.1.1.1 apart from the poly[A] tailing reaction (as the samples were dried down, 2µl of extra

molecular biology grade water was added to the reaction mix per sample and samples carefully resuspended) and no miR-Amp step was included.

Sample No.	Sample type	Concentration by NanoDrop (ng/µl)	Dilution	Concentration after dilution (ng/µl)	Volume in qRT-PCR (µl)
P1	Isolated RNA from plasma	10.3	NA	10.3	10
P2	Isolated RNA from plasma	10.5	NA	10.5	10
P3	Isolated RNA from plasma	11.9	NA	11.9	10
P4	Isolated RNA from plasma	14.6	NA	14.6	10
P5	Isolated RNA from plasma	15.8	NA	15.8	10
P6	Isolated RNA from plasma	17.5	NA	17.5	10
N1	Isolated RNA from lung tissue	53.8	1 in 6	8.97	2
N2	Isolated RNA from lung tissue	98.1	1 in 10	9.81	2
N3	Isolated RNA from lung tissue	69.9	1 in 7	19.97	2
N4	Isolated RNA from lung tissue	7.6	NA	7.6	3

The poly[A] tailing reaction

1 Thawed the isolated EV RNA samples and cDNA synthesis reagents on ice, gently vortexed, then centrifuged briefly.

2 In a 1.5ml microcentrifuge tube, prepared sufficient Poly(A) Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
10 × Poly[A] Buffer	0.5 µL	2.2 µL	5.5 µL
ATP	0.5 µL	2.2 µL	5.5 µL
Poly[A] Enzyme	0.3 µL	1.32 µL	3.3 µL
molecular biology grade water	3.7 µL	16.28 µL	40.7 µL
Total PCR Reaction Mix volume	**5 µL**	**22 µL**	**55 µL**

[1] Note: Volumes include 10% overage.

3 Vortexed the Poly(A) reaction mix, then centrifuged briefly.

4 Transferred 5µl of the Poly(A) reaction mix to the tubes containing dried-down RNA samples.

5 Sealed the reaction tubes, then vortexed briefly to thoroughly mix the contents.

6 Centrifuged the reaction tubes briefly to spin down the contents and eliminate air bubbles.

7 Placed the reaction tubes into a thermal cycler, then incubate using the following settings and standard cycling:

Step	Temperature	Time
Polyadenylation	37°C	45 minutes
Stop reaction	65°C	10 minutes
Hold	4°C	Hold

2.1.1.5 Assessment of miR-Amp amplification of miRNAs

Three miRNAs (miR-99a-5p, miR-26a-1-3p, and miR-1185-5p) were tested with 2x QuantiNova Probe PCR Master Mix using RNA extracted from 6 plasma samples and 4 normal lung tissue samples (same samples described in Section 2.1.1.4). Pre-prepared 1:5 dilution of RT reaction product and 1:10 dilution of miR-Amp reaction product (Section 2.1.1.4) and 2-step qPCR method were used for the qPCR reaction. The process of this experiment was similar to Section 2.1.1.4 apart from the miR-Amp reaction was performed and the miR-Amp reaction product was used for the subsequent PCR reaction (same as Section 2.1.1.1).

2.1.1.6 Validation of calibrator samples to be used for normalization

<u>Preparing plasma calibrator</u>

The Plasma calibrator EV RNA was prepared using 20 pooled plasma samples from an anonymised blood bank (collected as part of LLP), using the exoEasy Midi Kit. The frozen plasma samples were thawed at room temperature, then spun at 14,000 x g for 10 minutes. The supernatants of all samples were pooled, then 10ml of the sample was transferred into each of 4 falcon tubes (50ml). 30ml of TRIzol was added into each falcon tube, vortexed for 5

minutes. 5.5ml Chloroform was then added, vortexed for 5 minutes, then spun at 4,000 x g for 5 minutes at room temperature. 12.5ml of upper aqueous phase was transferred to a new falcon tube, 2 volumes of 100% ethanol (25ml) was added, then mixed well by vortexing in fridge (4°C) for 15 minutes. Up to 19ml of sample was transferred into a column (exoEasy Midi Spin Column) which was placed on the vacuum manifold using a sterile Vac Connector. The tap was switched on to allow liquid to flow through, then switched off when the column was empty. 700µl buffer RWT was added to the column under vacuum, then 500µl buffer RPE was used to wash the column twice under vacuum; all flow-through was discarded. The column was then placed in a collection tube, spun at 14,000 x g for 4 minutes at room temperature to remove any remaining wash buffer. The column was placed in a new collection tube, 20µl molecular biology grade water was added to the centre of the spin column membrane. The column was left incubating at room temperature for 2 minutes, then spun at 12,000 x g for 1 minute at room temperature to elute the RNA. All RNA extraction products were pooled and the concentration was measured by NanoDrop. The extracted plasma RNA sample (measured as 553 ng/µl) was diluted at 1:100 to be used for the EV miRNA validation cohort as plasma calibrator (5.5 ng/µl). The cell line calibrator was RNA extracted from cell line SK-LU-1, which was prepared by our group previously using exoEasy Midi Kit. The extracted cell line RNA sample (20 ng/µl) was diluted at 1:4 to be used for the EV miRNA validation cohort as cell line calibrator (5 ng/µl).

To investigate the amount of the calibrators suitable for our EV miRNA qPCR validation cohort, the serial dilutions (the stock, 1:2, 1:4, 1:8, 1:16) of the miR-Amp reaction products of plasma calibrator and cell-line calibrator for miR-92a-3p was tested in duplicate. The whole process

of preparing cDNA template and PCR reaction was the same as Section 2.1.1.1 and 2µl of each sample was used as the starting material.

2.1.1.7 Potential artefacts in qPCR following RT and miR-Amp reactions

To investigate whether the signals of no RNA controls are because of miR-Amp reactions or the RT, the RT and miR-Amp reaction products of 2 tissue RNA samples, cell line calibrator (with the same starting amount) and 3 no RNA controls were tested for miR-1228-3p by qPCR. The tissue samples (concentration measured by NanoDrop: 135.4 ng/µl and 99.5 ng/µl) were diluted at 1:27 (5 ng/µl) and 1:20 (5 ng/µl) respectively. 2µl of the tissue samples, cell line calibrator (5 ng/µl) and molecular biology grade water (for the 3 no RNA controls) were used as the starting material for the RT reactions. The whole process of preparing cDNA template and PCR reaction was the same as Section 2.1.1.1. The RT reaction products and miR-Amp reaction products were diluted at 1:5 and 1:10 respectively and tested for miR-1228-3p by qPCR in duplicate.

2.1.2 Measurement of EV miRNAs in clinical plasma samples

<u>Experiment flow</u>

This is a retrospective case-control study. All blood specimens had already been collected, plasma isolated by centrifugation, aliquoted and stored at -80°C (Fig. 2.1).

The workflow process of my experiment included three main steps (Fig. 2.1):

1. EV isolation and RNA extraction from the plasma samples.
2. TaqMan probe-based qRT-PCR for miRNA level measurement.
3. Data analysis.

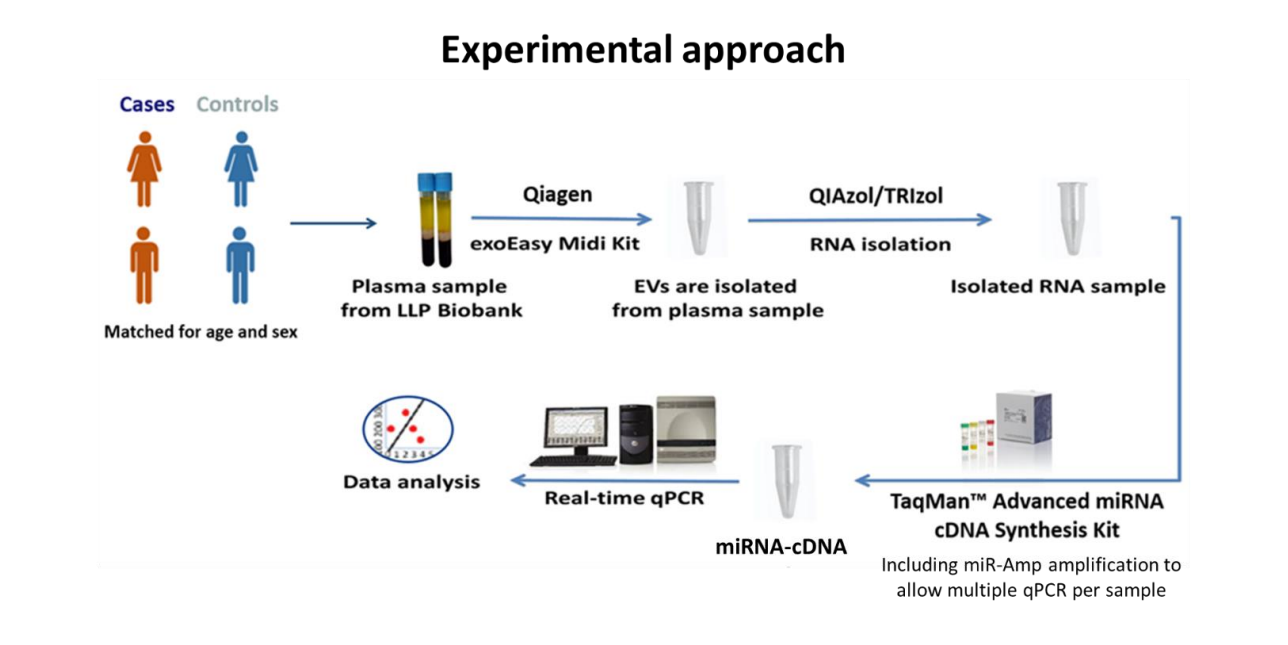


Figure 2.1. Sample process flow diagram.

<u>Collection, processing and storage of plasma samples and clinical data collection</u>

The plasma samples used for this validation cohort were collected from lung cancer patients and control subjects recruited in the Liverpool Lung Project (LLP). The lung cancer cases were mainly from Liverpool Heart and Chest Hospital and Aintree University Hospital, and the control subjects were mainly from the population cohort clinics (n=136), with some from hospital clinics (n= 51) with a negative lung cancer diagnosis and no lung cancer on follow-up. Samples were spun on-site (hospital) in EDTA-gel tubes within an hour of collection, minimising plasma contamination from blood cell products, aliquots of each sample were stored in our -80°C laboratory freezer. All subjects included in this study provided informed voluntary consent for the use of samples and collection of subsequent lung cancer-related clinical data. The LLP biobank operation and the research conduction were under the LLP ethical approval (Liverpool Research Ethics Committee REC 97/141).

Patient cohort

The samples used for the validation cohort case/control study (Section 3.3.1) are independent from the cohort used for the initial HTG identification (Section 4.1). Cases included only NSCLC (n = 60 for discovery; n = 188 for validation) and controls (n= 60 for discovery; n = 187 for validation) who did not have the cancer we matched for age and sex, and additionally matched on smoking duration for the discovery cohort.

Reagents

The exoEasy Midi Kits were purchased from Qiagen. The Direct-zol™ RNA MiniPrep Kits were purchased from Zymo Research. TRIzol™ Reagent was purchased from Invitrogen.

All 18 TaqMan® Advanced miRNA Assays (Table 2.1) tested in the plasma EV miRNA validation cohort and TaqMan™ Advanced miRNA cDNA Synthesis Kit were purchased from Applied Biosystems. The miRCURY LNA Probe PCR Kits were purchased from Qiagen.

Table 2.1. TaqMan™ advanced miRNA assays detected in the plasma EV validation cohort

Assay name	Mature miRNA Sequence
hsa-miR-3149	UUUGUAUGGAUAUGUGUGUGUAU
hsa-miR-1277-5p	AAAUAUAUAUAUAUAUGUACGUAU
hsa-miR-514b-5p	UUCUCAAGAGGGAGGCAAUCAU
hsa-miR-5093	AGGAAAUGAGGCUGGCUAGGAGC
hsa-let-7a-3p	CUAUACAAUCUACUGUCUUUC
hsa-miR-103a-3p	AGCAGCAUUGUACAGGGCUAUGA
hsa-miR-107	AGCAGCAUUGUACAGGGCUAUCA
hsa-miR-1178-5p	CAGGGUCAGCUGAGCAUG
hsa-miR-1205	UCUGCAGGGUUUGCUUUGAG
hsa-miR-1228-3p	UCACACCUGCCUCGCCCCCC
hsa-miR-1247-3p	CCCCGGGAACGUCGAGACUGGAGC
hsa-miR-6872-5p	UCUCGCAUCAGGAGGCAAGG
hsa-miR-301b-3p	CAGUGCAAUGAUAUUGUCAAAGC
hsa-miR-1185-5p	AGAGGAUACCCUUUGUAUGUU
hsa-let-7c-3p	CUGUACAACCUUCUAGCUUUCC
hsa-miR-26a-1-3p	CCUAUUCUUGGUUACUUGCACG

hsa-miR-146b-5p	UGAGAACUGAAUUCCAUAGGCUG
hsa-miR-342-5p	AGGGGUGCUAUCUGUGAUUGA

Instruments

Three types of centrifuges were used in the experimental process: microcentrifuge (Labnet International), Prism™ R refrigerated microcentrifuge (Labnet International) and Sigma 4K15 laboratory centrifuge (Sigma). miVac DNA concentrator (Genevac) was used to dry down the EV RNA samples isolated from the plasma samples. Vortex mixer (Starlab) and a rotator mixer (Starlab) were used in the experimental process. GeneAmp PCR System 9700 (Applied Biosystems) was used for RT reaction. 7500 Real-Time PCR System and 7500 Fast Real-Time PCR System (Applied Biosystems) were used in parallel for performing qPCR reactions in the duplicate plates. A class 1 fume hood (Labcaire) was used in the process of EV RNA isolation. Multichannel pipette (Starlab) was used to transfer the required volume of the reaction mix from the reservoir to each well of the 96-well plate.

Software used for data analysis

IBM SPSS Statistics 27 was used for analysing qRT-PCR results of plasma EV miRNA validation cohort. Applied Biosystems 7500 software (v2.3) was used for collection and analysis of qPCR data.

2.1.2.1 EV isolation and RNA extraction

Isolation of EV fraction from plasma

Lysed EVs were isolated from plasma using exoEasy Midi Kits. 2ml of frozen plasma sample was thawed at room temperature, the sample was then spun at 14,000 x g for 10 minutes and

filtered through a 0.45µm syringe filter (Starlab). To collect the residue of the plasma in the filter, 150µl PBS was added and pushed through the filter to combine with the previous filtrate. The following process was then used with the exoEasy Midi kit: added 1 volume buffer XBP to 1 volume of filtered sample, mixed well by gently inverting the tube for 5 minutes at room temperature on the rotating mixer, then spun the sample briefly. Placed the exoEasy column on the vacuum manifold using a sterile Vac Connector. Added the sample/XBP mix onto the exoEasy column and switched tap on to allow liquid to flow through, switched tap off when the column was empty. Added 3.5ml buffer XWP, switched tap on. Switched tap off when the column was empty. Transferred the column to the original collection tube and spun at 4,000 x g for 3 minutes. Transferred the column to a fresh collection tube, added 700µl QIAzol/TRIzol to the membrane (from this step, the work was done in the fume hood), incubated for 2 minutes at room temperature, then spun at 4,000 x g for 5 minutes to collect the lysate. Transferred the lysate completely to a 2ml tube, incubated at room temperature for 5 minutes, then spun briefly. Added 700µl of absolute ethanol to the tube, vortexed for 15 seconds, then incubated on ice for 5 minutes. At this stage, samples can be stored at -20°C indefinitely (Fig. 2.2).

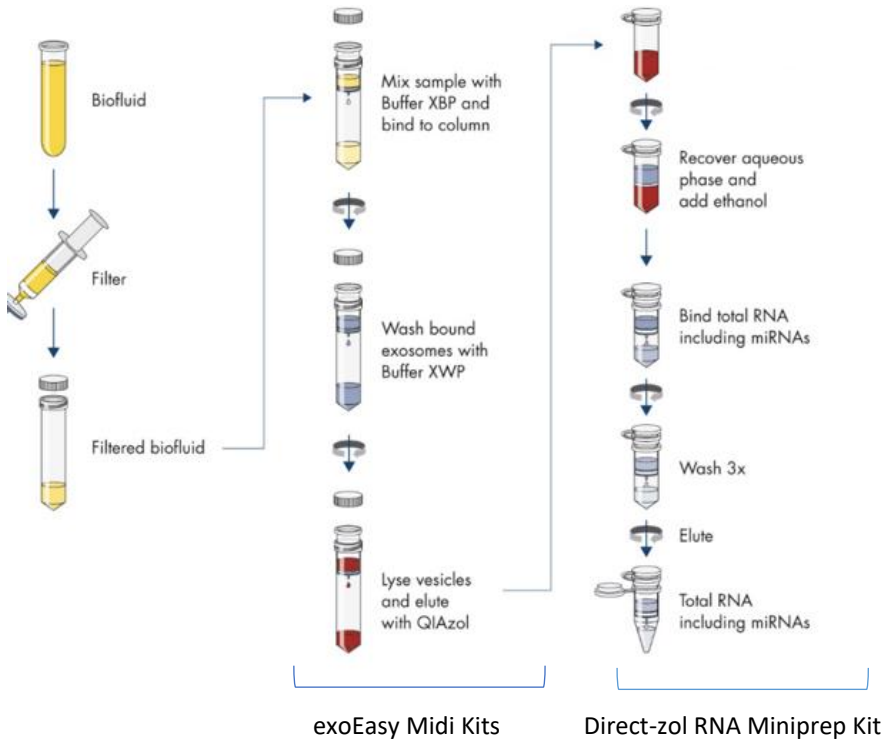


Figure 2.2. EV purification and total RNA isolation flow diagram (Qiagen 2019).

Total RNA isolation from plasma EV fraction

RNA was extracted from the lysed EVs using the Direct-zol™ RNA Miniprep Kit. In a fume hood, pipetted up to 700µl sample into a Zymo-Spin IIC™ Column (Direct-zol™ RNA Miniprep Kit) which was placed in a 2ml collection tube. Centrifuged at 6,000 x g for 1 minute at room temperature, then discarded the flow-through in phenol waste. Pipetted the second 700µl sample batch into the column, spun and discarded as above. Added 400µl RNA Pre-wash buffer into the column, spun at 6,000 x g for 1 minute at room temperature. Discarded the flow-through in buffer waste and tapped the tube upside down to remove any residue on blue roll, then repeated this step (from this step, the work was done outside the fume hood). Added 700µl RNA wash buffer, spun at 6,000 x g for 1 minute at room temperature. Discarded the flow-through in buffer waste and tapped the tube upside down to remove any residue on

blue roll. Performed dry spin at 14,000 x g for 3 minutes at room temperature. Placed the spin column in a new 1.5ml low retention tube. Added 25µl molecular biology grade water directly to the centre of the column membrane, incubated for 2 minutes at room temperature, then spun at 12,000 x g for 1 minute to elute the RNA. Stored the RNA sample immediately on ice. The RNA samples can be stored at -20°C for short periods or -80°C for longer periods.

Given the large sample size (380 samples were included, 1 sample failed at extraction step due to a clogged column), multiple 96-well plates were used. 10µl of the EV RNA sample was transferred to each 96-well plate. Consequently, plates 2-6 contained 92, 92, 92, 89 and 14 samples, respectively. Apart from the samples, 4 wells were loaded with cell line calibrator (3µl of 5 ng/µl), plasma calibrator (3µl of 5.5 ng/µl) and 2 no RNA controls. According to the protocol, for blood, serum, or plasma samples, we should use 2µl of sample eluent (from the sample isolation procedure) per reaction. Given the limited absolute quantity of the EV miRNA isolated from blood, we decided to dry down the samples, so that 10µl rather than 2µl of the EV RNA sample could be used as the starting material. Samples and the calibrators were dried down using the miVac DNA concentrator for 1hour and 10 minutes, and then stored at -80°C.

2.1.2.2 Measuring EV miRNA expression using qRT-PCR in isolated EV fractions from human plasma samples

Applied Biosystems TaqMan® Advanced miRNA Assays together with the Applied Biosystems™ TaqMan Advanced miRNA cDNA Synthesis Kit enable highly sensitive and specific quantification of mature miRNAs using qPCR.

Due to the large sample size and multiple-step process, we transferred the reaction mix to a disposable reagent reservoir (Starlab) and used multichannel pipette to transfer the required

volume of the reaction mix from the reservoir to each well, so that the process could be more efficient as well as to reduce operator errors and process variability to some extent.

The poly[A] tailing reaction

1 Thawed the dried, isolated EV RNA samples and cDNA synthesis reagents on ice, gently vortexed, then centrifuged briefly.

2 In a 1.5ml microcentrifuge tube, prepared sufficient Poly(A) Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
10 × Poly[A] Buffer	0.5 µL	2.2 µL	5.5 µL
ATP	0.5 µL	2.2 µL	5.5 µL
Poly[A] Enzyme	0.3 µL	1.3 µL	3.3 µL
molecular biology grade water	3.7 µL	16.28 µL	40.7 µL
Total PCR Reaction Mix volume	**5 µL**	**22 µL**	**55 µL**

[1] Note: Volumes include 10% overage.

3 Vortexed the Poly(A) reaction mix, then centrifuged briefly.

4 Transferred 5µl of the Poly(A) reaction mix to each well of the reaction plate.

5 Thoroughly mixed the contents by pipetting up and down a few times, then sealed the reaction plate.

6 Centrifuged the reaction plate briefly to spin down the contents and eliminate air bubbles.

7 Placed the reaction plate into a thermal cycler, then incubated using the following settings and standard cycling:

Step	Temperature	Time
Polyadenylation	37°C	45 minutes
Stop reaction	65°C	10 minutes
Hold	4°C	Hold

The adaptor ligation reaction

1 In a 1.5ml microcentrifuge tube, prepared sufficient Ligation Reaction Mix for the required number of reactions according to the following table, 50% PEG 8000 was used at room temperature and aspirated and dispensed slowly to ensure accurate pipetting.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
5X DNA Ligase Buffer	3 µL	13.2 µL	33 µL
50% PEG 8000[2]	4.5 µL	19.8 µL	49.5 µL
25X Ligation Adaptor	0.6 µL	2.6 µL	6.6 µL
RNA Ligase	1.5 µL	6.6 µL	16.5 µL
RNase-free water	0.4 µL	1.8 µL	4.4 µL
Total Ligation Reaction Mix volume	**10 µL**	**44 µL**	**110 µL**

[1] Note: Volumes include 10% overage.

2 Vortexed the Ligation Reaction Mix, then centrifuged briefly.

3 Transferred 10µl of the Ligation Reaction Mix to each well of the reaction plate containing the poly(A) tailing reaction product, the total volume was 15µl per well.

4 Thoroughly mixed the contents by pipetting up and down a few times, then sealed the reaction plate.

5 Centrifuged the reaction plate briefly to spin down the contents.

6 Placed the reaction plate into a thermal cycler, then incubated using the following settings and standard cycling:

Step	Temperature	Time
Ligation	16°C	60 minutes
Hold	4°C	Hold

The reverse transcription (RT) reaction

1 In a 2ml microcentrifuge tube, prepared sufficient RT Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
5X RT Buffer	6 µL	26.4 µL	66 µL
dNTP Mix (25 mM each)	1.2 µL	5.3 µL	13.2 µL
20X Universal RT Primer	1.5 µL	6.6 µL	16.5 µL
10X RT Enzyme Mix	3 µL	13.2 µL	33 µL
RNase-free water	3.3 µL	14.5 µL	36.3 µL
Total RT Reaction Mix volume	**15 µL**	**66 µL**	**165 µL**

[1] Note: Volumes include 10% overage.

2 Vortexed the RT Reaction Mix, then centrifuged briefly.

3 Transferred 15µl of the RT Reaction Mix to each well of the reaction plate containing the adaptor ligation reaction product, the total volume was 30µl per well.

4 Thoroughly mixed the contents by pipetting up and down a few times, then sealed the reaction plate.

5 Centrifuged the reaction plate briefly to spin down the contents.

6 Placed the reaction plate into a thermal cycler, then incubated using the following settings and standard cycling:

Step	Temperature	Time
Reverse transcription	42°C	15 minutes
Stop reaction	85°C	5 minutes
Hold	4°C	Hold

<u>The miR-Amp reaction</u>

1 In a 5ml microcentrifuge tube, prepared sufficient miR-Amp Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
2X miR-Amp Master Mix	25 µL	110 µL	275 µL
20X miR-Amp Primer Mix	2.5 µL	11 µL	27.5 µL
RNase-free water	17.5 µL	77 µL	192.5 µL
Total miR-Amp Reaction Mix volume	**45 µL**	**198 µL**	**495 µL**

[1] Note: Volumes include 10% overage.

2 Vortexed the miR-Amp Reaction Mix, then centrifuged briefly.

3 Transferred 45µl of the miR-Amp Reaction Mix to each well of a new reaction plate.

4 Added 5µl of the RT reaction product to each reaction well, the total volume was 50µl per well.

5 Thoroughly mixed the contents by pipetting up and down a few times, then sealed the reaction plate.

6 Centrifuged the reaction plate briefly to spin down the contents.

7 Placed the reaction plate into a thermal cycler, then incubated using the following settings MAX ramp speed, and standard cycling:

Step	Temperature	Time	Cycles
Enzyme activation	95°C	5 minutes	1
Denature	95°C	3 seconds	14
Anneal/Extend	60°C	30 seconds	
Stop reaction	99°C	10 minutes	1
Hold	4°C	Hold	1

8 Added 20µl of the miR-Amp reaction product and 180µl of molecular biology grade water to a new plate to dilute the miR-Amp reaction product 1 in 10. Thoroughly mixed the contents by pipetting up and down a few times, then sealed the plate, spun at 6,000 x g for 1 minute.

9 For each targeted miRNA, we needed to perform duplicate PCR experiments; therefore, 10µl of diluted miR-Amp reaction product of each sample was transferred to a new PCR plate. The diluted samples were stored at -20°C.

Real-time PCR

1 Thawed the previously prepared plate containing 10µl of the diluted miR-Amp reaction product and PCR assays on ice, gently vortexed and then centrifuged briefly.

2 In a 5ml microcentrifuge tube, prepared sufficient PCR Reaction Mix for the required number of reactions according to the following table.

Component	1 Rxn	4 Rxns[1]	10 Rxns[1]
2 × QuantiNova Probe PCR Master Mix	10 µL	44 µL	110 µL
miRNA assay	1 µL	4.4 µL	11 µL
Rox Reference Dye	0.1 µL	0.44 µL	1.1 µL

Diluted miR-Amp product	5 µL	22 µL	55 µL
molecular biology grade water	3.9 µL	17.16 µL	42.9 µL
Total PCR Reaction Mix volume	**20 µL**	**88 µL**	**220 µL**

[1] Note: Volumes include 10% overage.

3 Vortexed the PCR Reaction Mix, then centrifuged briefly.

4 Transferred 30µl of the PCR Reaction Mix to each well of the previously prepared PCR reaction plate containing 10µl of diluted miR-Amp reaction product. Thoroughly mixed the contents by pipetting up and down a few times, sealed the plate, then spun briefly.

5 Unsealed the PCR plate, then aliquoted 20µl mix of the PCR reaction into a second PCR plate as the duplicate plate.

6 Sealed the reaction plate with a StarSeal Advanced Polyolefin Film (Starlab), centrifuged the reaction plate briefly to spin down the contents.

7 Set up and ran the real-time PCR instruments (7500 and 7500 Fast systems) in parallel.

(1) Loaded the reaction plate in the real-time PCR instrument.

(2) Set the appropriate experiment settings and PCR thermal cycling conditions.

Step	Temperature	Time	Cycles
Enzyme activation	95°C	2 mins	1
Denature	95°C	5 secs	45
Anneal / Extend	60°C	30 secs	

(3) Set the reaction volume appropriate for the reaction plate.

(4) Started the run.

(5) Applied Biosystems 7500 Software v2.3 (Life Technologies) was used for data collection, with the appropriate filter setting for the TaqMan probes used and the Rox internal reference.

(6) The same software was used for data analysis, with automated baseline subtraction and thresholds. Occasionally baseline start and stop positions and thresholds were adjusted to avoid artefacts such as unexplained spikes or baselines drift.

2.1.2.3 Statistical analysis of the plasma EV miRNA qRT-PCR data

The qPCR data from samples on plates 2-6 was exported to excel spreadsheets respectively, each spreadsheet contains the raw data of the samples from the duplicate plates tested for a single targeted miRNA. Median Ct value was used for normalisation (ΔCt = Ct value of individual sample – median Ct value) of the samples on the duplicate plates. $2^{-\Delta\Delta Ct}$ method (assuming a uniform PCR amplification efficiency of 100% across all samples) was used to calculate the relative quantification (RQ). The average of the Ct values for duplicate RT-PCRs was used to calculate the RQ value for each sample. RQ values were used for the statistical analysis of the qPCR data from all samples. A master spreadsheet was used to combine the ratios of all tested samples for the targeted miRNAs and the clinical data. Mann-Whitney U test analysis was used to compare differential expression of miRNAs (or other continuous variables, such as LLPv2 risk score) between lung cancer cases and control subjects, or other categorical factors (e.g. COPD, smoking status, histology, sex). Receiver Operating Characteristic (ROC) curve analysis was performed to assess predictive accuracy (sensitivity and specificity). Spearman's correlations were performed to examine the correlations between miRNAs and between miRNAs and other continuous variables, such as smoking duration.

2.2 EV miRNA HTG EdgeSeq data analysis

HTG EdgeSeq Chemistry Overview

HTG EdgeSeq assays employ nuclease protection to measure RNA. As shown in Fig. 2.3B (www.htgmolecular.com/systems/workflow last accessed: 2022), target-specific protection probes with universal priming sites (wings) hybridise to target transcripts and wingman (Step 1). S1 nuclease is added to digest non-hybridised probes and RNA, resulting in 1-to-1 ratios of probes and target RNAs (Step 2). After S1 nuclease inactivation (Step 3), the remaining probes are amplified with primers carrying sequencing adaptors and molecular barcodes, used to identify which sample the tagged probes come from after pooling for sequencing (Step 4). The resulting PCR products are purified, quantitated, and combined to make a sequencing ready library (Step 5). HTG EdgeSeq assays do not require processing steps such as reverse transcription, adenylation, or adaptor ligation (Fig. 2.3B). It is noteworthy that steps 1 - 3 are fully automated on the HTG EdgeSeq system.

HTG quality control (QC) metrics

The original HTG QC metric use five internal negative controls (ANTs), not found in human miRNA, but spiked into the hybridisation mix at very low level; when miRNA quantity or quality is low, these generate a higher cpm. Under normal circumstances, the vast majority of the probes bind to human miRNA and the proportion of signal (cpm) for ANT controls is very low.

New post-sequencing quality control (QC) metrics for the HTG EdgeSeq miRNA WTA were established by HTG Molecular to identify samples that should be removed from subsequent analysis due to one or a combination of factors; these were implemented as part of HTG

Reveal software. Three QC metrics that address specific post-sequencing failure modes were established (Table 2.2).

QC0; detects poor quality RNA within the sample (degraded or low concentration). Samples with poor quality RNA are expected to have higher relative read depth for the positive (POS) control probes. The POS read depth is inversely proportional to the quality/quantity of miRNA contained within a sample; i.e., samples with high sample quality result in low read depth for POS and vice versa.

QC1; detects samples with insufficient read depth, given that sample total counts vary according to the sample size in each plate. Samples that failed to meet the established read depth requirement are excluded from further analysis.

QC2; uses relative standard deviation (RSD) of log2-transformed counts of all probes of the HTG EdgeSeq miRNA WTA Panel, minus the control probes to detect samples with low variation of counts across probes not reflective of biological expression variability.

Table 2.2. Summarising the three post-processing QC metrics for the HTG EdgeSeq miRNA WTA

HTG EdgeSeq miRNA Whole Transcriptome Assay QC metrics		
Metric	Corresponding Failure Mode	Final Cut-off Value
QC0	Degraded RNA or poor quality of sample	POS% ≥ 14% as failure
QC1	Insufficient read depth	Read depth ≤ 500K as failure
QC2	Minimal expression variability	RSD ≤ 0.08 as failure
Source: A Research Use Only White Paper for HTG EdgeSeq miRNA Whole Transcriptome Assay QC Metrics (HTG 2019)		

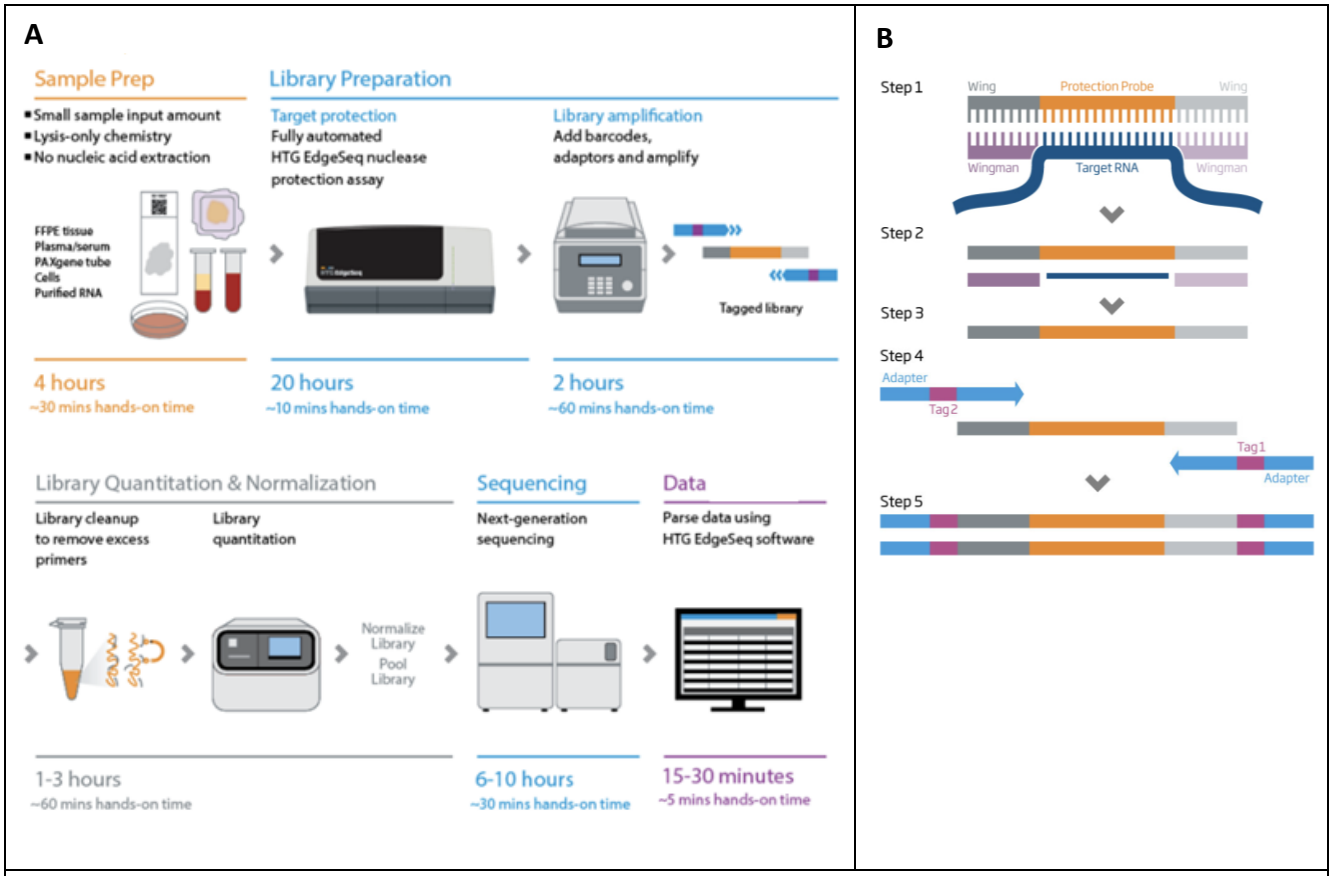


Figure 2.3. HTG EdgeSeq system workflow (A) and HTG EdgeSeq chemistry overview (B) (www.htgmolecular.com/systems/workflow last accessed: 2022)

<u>EV miRNA HTG EdgeSeq data analysis</u>

The HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA) was used to measure the expression of 2,083 human miRNA transcripts [plus 13 housekeeping (HK) genes] in the isolated EV fractions from plasma samples (plasma EV miRNA HTG EdgeSeq data).

In the original analysis of the EV miRNA HTG EdgeSeq data, only one type of QC was applied, which was based on ANT controls. The target miRNAs for qPCR validation were selected based on bioinformatic analysis using Mann-Whitney U test. The best predicting model of lung cancer was established based on model building for generalised linear models using Lasso

penalized logistic regression and internal validation by ROC analysis with leave-one-out cross-validation.

HTG EdgeSeq Reveal statistical analysis package, including improved quality control and integration of differential expression analysis (based on edgeR), was used to reanalyse the EV miRNA HTG data. A combination of ANT-QC, QC0, QC1 and QC2 was applied in the reanalysis. PCA and volcano plots were used to analyse the distribution of the samples and FDR adjusted p values were used to identify differentially expressed miRNAs. Further reanalysis of HTG data was performed in batches 1 and 2 separately, and compared to each other or to the combined analysis, to reach a consensus on differentially expressed miRNAs.

HTG EdgeSeq differential expression analysis was performed in HTG Reveal software using the edgeR package (version 3.16.5) available from Bioconductor. The edgeR package provides methods for estimating and testing differential expression using negative-binomial generalized linear models. Empirical Bayes methods are used to estimate dispersion and log2 (fold change) with data-driven prior distributions. See edgeR: differential analysis of sequence read count data User's Guide (bioconductor.org) for more information. No pre-filtering is applied to the data prior to analysis. The edgeR model corrects for differences in library size using the Trimmed-mean of M-values (TMM) method from Robinson and Oshlack (Robinson and Oshlack 2010). Dispersions are estimated with the Cox Reid-adjusted profile likelihood method developed by McCarthy et al. (McCarthy, Chen et al. 2012). Log2 fold change is estimated via Tikhonov/ridge regularization with a zero-centred normal prior distribution with variance calculated using the observed distribution of maximum likelihood coefficients (see edgeR documentation for details).

Chapter 3 Validation of plasma extracellular vesicle (EV) miRNAs as diagnostic biomarkers for lung cancer

In this chapter, we describe experiments to validate the 18 miRNAs preselected as potential diagnostic lung cancer biomarkers, in human plasma EV fractions of an independent set of samples, using alternative isolation and quantitation methods (which are more easily applied in a clinical setting). This includes optimisation and validation of the qRT-PCR, description of validation cohort, measurement of the miRNAs in clinical samples and assessment of differential expression, between cases and controls and with other characteristics (e.g. smoking, COPD).

3.1 Study design

This part of the project aimed to validate a panel of miRNAs in the plasma extracellular vesicle fraction as a diagnostic molecular signature of lung cancer patients. This panel of 18 miRNA markers had been selected from a previous analysis of HTG EdgeSeq data (Section 1.4) derived from a pilot study funded by the Roy Castle Lung Cancer Foundation (RCLCF) in 2016. The samples and associated clinical information for the cases and controls was sourced from the Liverpool Lung Project (LLP).

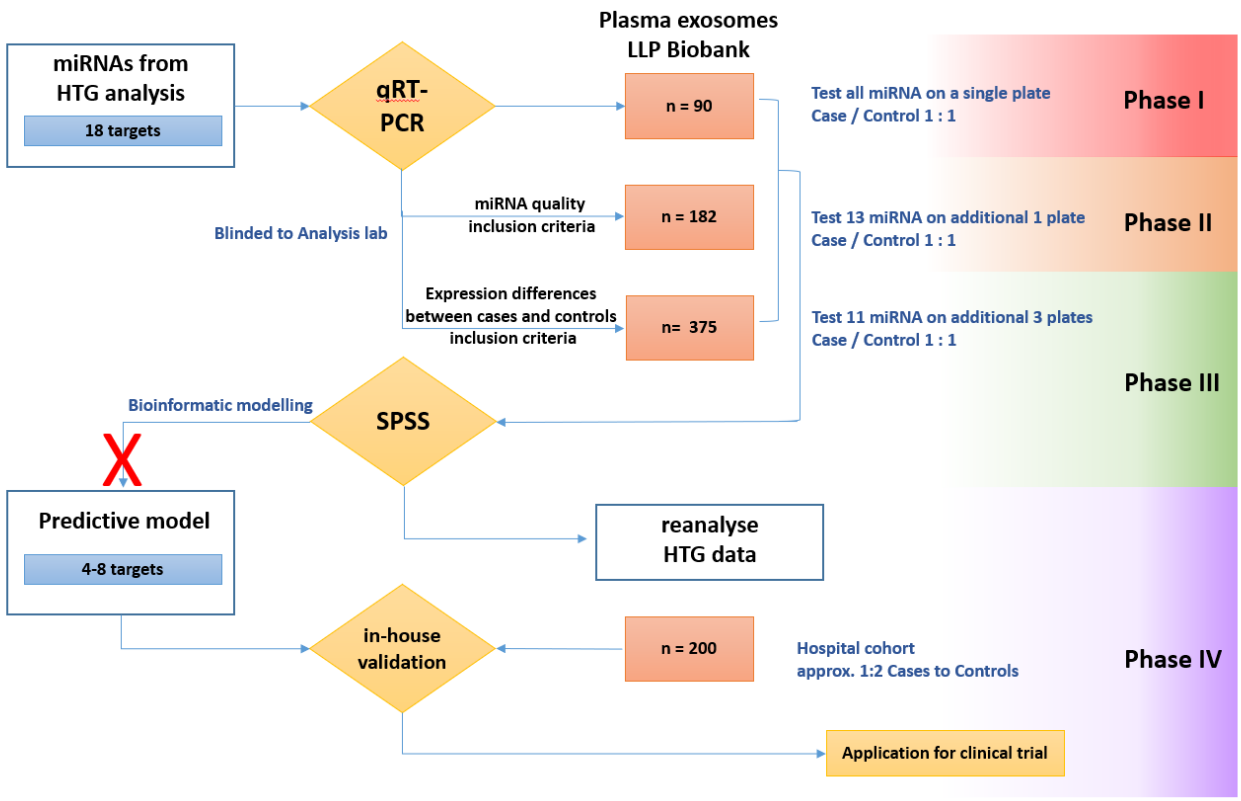


Figure 3.1. Workflow diagram of the project. The project initially aimed to validate the 18 microRNAs from the original analysis of HTG EdgeSeq data in the discovery phase, using qRT-PCR. The plan, following validation of the individual microRNAs (in Phases I, II & III, with increasing numbers of samples), was to generate the best predictive panel and test this signature in an independent set of samples (Phase IV). As the validation process resulted in few EV microRNAs with high enough diagnostic potential, a reanalysis of the HTG data was performed using Reveal bioinformatics software.

3.2 Optimisation and validation of assay conditions

3.2.1 Testing QuantiNova PCR reaction mixes and pre-amplification

Before starting the validation with our matched samples, we needed to decide the qPCR assay platform; the options at the time were TaqMan™ Advanced miRNA Assays (Applied Biosystems) and miRCURY LNA Probe PCR Kits (Qiagen); the relevant assays and experimental methods are listed in Materials and Methods, Section 2.2. We used total RNA extracted from cell lines in 4-fold serial dilutions for RT reaction and the subsequent miR-Amp reaction. The products of the two above reactions (cDNA and miR-Amp reaction product) were utilised to test miRCURY LNA Probe PCR Kits and QuantiNova Multiplex PCR Kits. The results showed that cDNA-multiplex reaction conditions were slightly better than cDNA-Probe reaction conditions: the R^2 was closer to 1, meaning better reproducibility and the cycle threshold (Ct) values were slightly lower meaning better sensitivity. However, both provided good linear responses with 2 Ct value difference between 4-fold dilutions (Fig. 3.2). This indicates that the RT reaction was equally efficient across all RNA concentrations and the qPCR was highly efficient.

Using the miR-Amp reaction products, the probe reaction conditions were better than the multiplex conditions with both better reproducibility and a closer relationship between amount and Ct values (steeper gradient) (Fig. 3.2). However, for both reaction conditions, the Ct change with 4-fold dilutions was significantly less than expected 2 Cts (Fig. 3.2). This indicates that the amplification reaction is not quantitative.

Despite the amplification products demonstrating a lack of efficiency of the amplification (a flatter gradient), there clearly was some amplification as the Ct values were lower, for example, using the probe reaction mix: the highest RNA concentration gave Ct of 27 without

amplification and a Ct of 23 with amplification (a difference of 4 Cts); the lowest RNA concentration gave Ct of 34 without amplification and a Ct of 26 with amplification (a difference of 8 Cts, Fig. 3.2). This indicates that amplification was more efficient at lower RNA starting concentrations, but was still less efficient than expected given 14 cycles of application: approx. 57% efficient at lowest starting amount and approx. 29% efficient at highest starting amount.

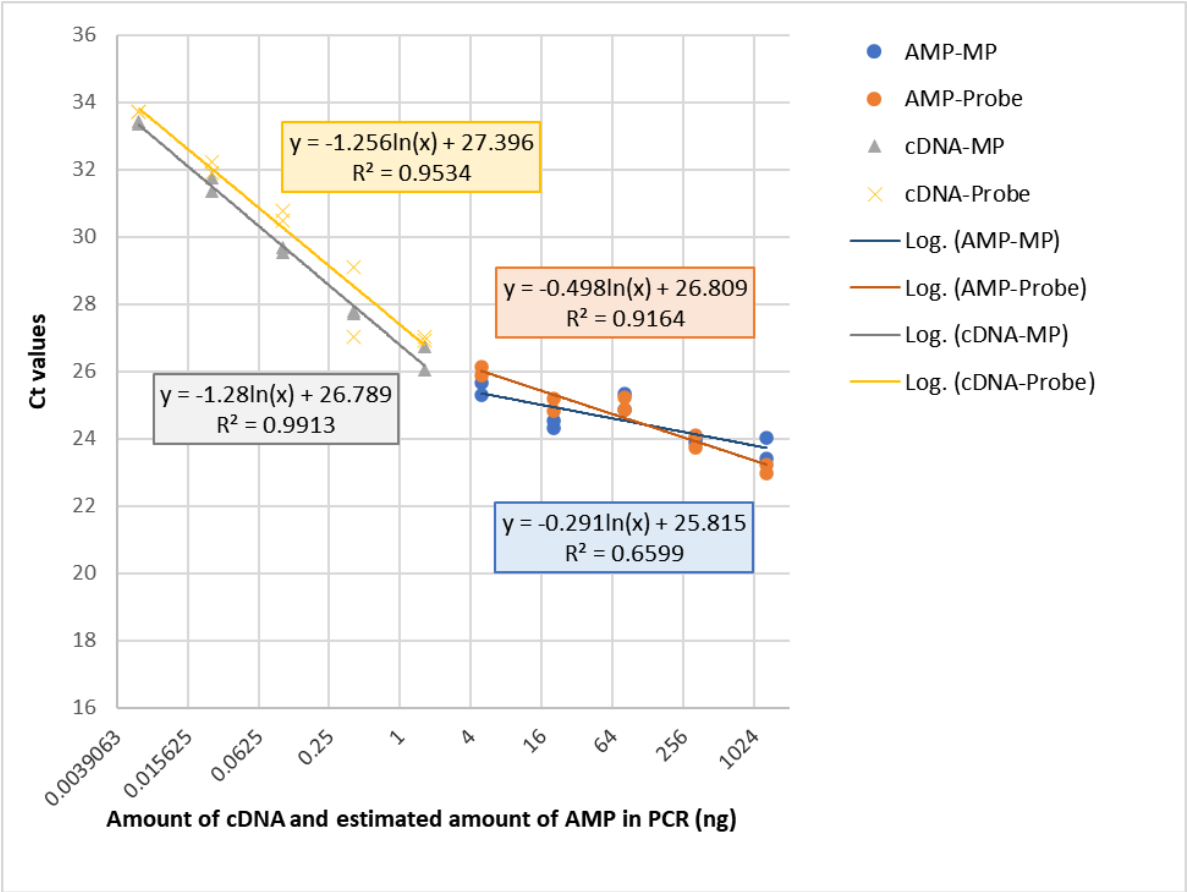


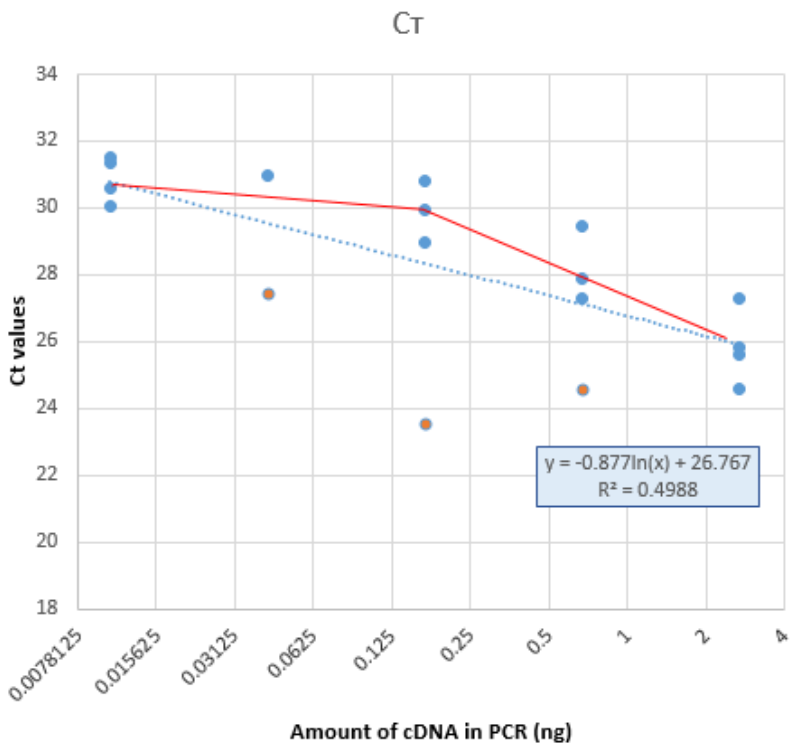
Figure 3.2. miRCURY LNA Probe PCR Kits and QuantiNova Multiplex PCR Kits for miR-99a-5p on serial dilutions of cell line cDNA and miR-Amp reaction product. miRNA extracted from cell line and its 4-fold serial dilutions were used in RT reactions and subsequent qPCR without amplification (cDNA) or with amplification by 14 rounds of PCR with TaqMan™ Advanced miRNA Assays kit miR-Amp reaction (AMP). For the AMP samples, the cDNA amount was estimated based on 100% efficiency.

3.2.2 Testing efficiency, sensitivity, and linearity of TaqMan qRT-PCR

We used the same RT products (miRNA extracted from cell line and its 4-fold serial dilutions) to test TaqMan™ Advanced miRNA Assays for real-time PCR. The results showed that the PCR is not quantitative when Ct is above 30 and the amount of cDNA in PCR is less than 0.1667 ng (Fig. 3.3A). This non-linearity of the TaqMan qPCR reaction contrasts with previous QuantiNova results, which were linear at all concentrations tested (and up to a Ct of 34).

Notably the Ct values for replicates were more variable in this TaqMan experiment than in the previous QuantiNova experiment. When all datapoints are included, the R^2 was less than 0.5, meaning poor reproducibility and a poor fit to the trendline (Fig. 3.3A). R^2 improves when the three outliers are removed and the lowest 2 dilutions excluded, with the resultant observed Ct change with 4-fold dilutions closer to the expected 2 Cts (Fig. 3.3B). This indicates that although the PCR is not quantitative if the RNA starting concentrations were too low, it was reasonably quantitative at higher starting cDNA amounts. That the Ct values for the lowest 2 cDNA amounts have similar Ct values to the next highest amount could be due to contamination or a PCR artefact.

A



B

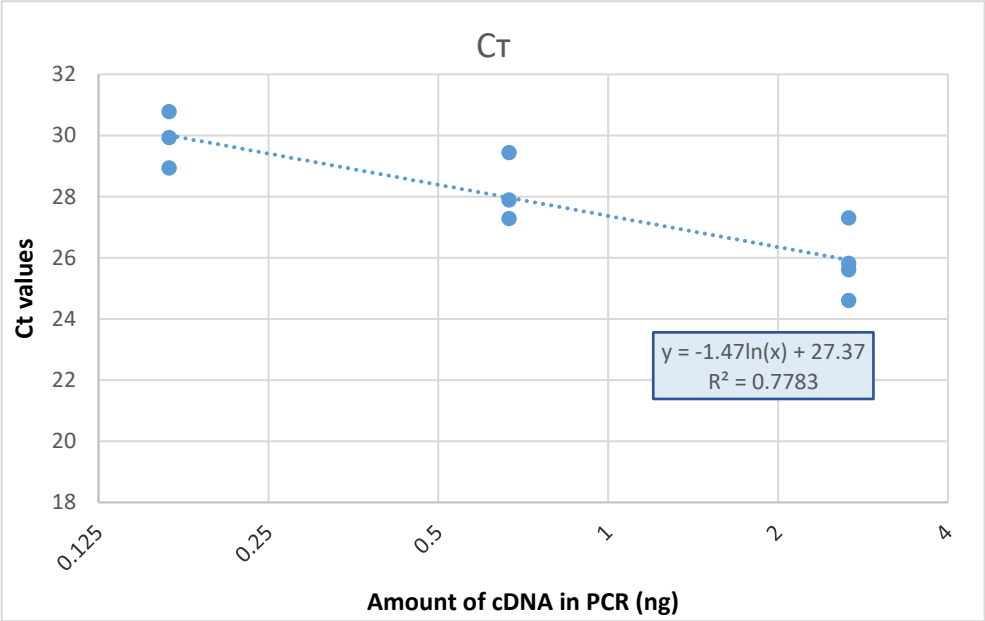


Figure 3.3. TaqMan™ Advanced miRNA Assays for miR-99a-5p on serial dilutions of cell line cDNA. miRNA extracted from cell line and its 4-fold serial dilutions were used in RT reactions and subsequent qPCR. All qPCR results are shown in A, notably Ct values for the lowest 3 dilutions are not significantly different; the lowest 2 dilutions were removed in B, along with 2 outliers.

3.2.3 Testing efficiency, sensitivity, linearity, and reproducibility of QuantiNova qRT-PCR

Having seen a lack of linearity for TaqMan PCR, the QuantiNova probe reaction mix was retested to investigate if the RT products were contaminated and to investigate reproducibility (by comparison to the previous QuantiNova probe reaction mix results). The results show that between the two runs with QuantiNova probe reaction mix, run 2 has slightly better within-run reproducibility (between replicates in the same experiment). Run 1 has steeper gradient, meaning a more efficient quantitation and a closer relationship between the amount and Ct values (Fig. 3.4).

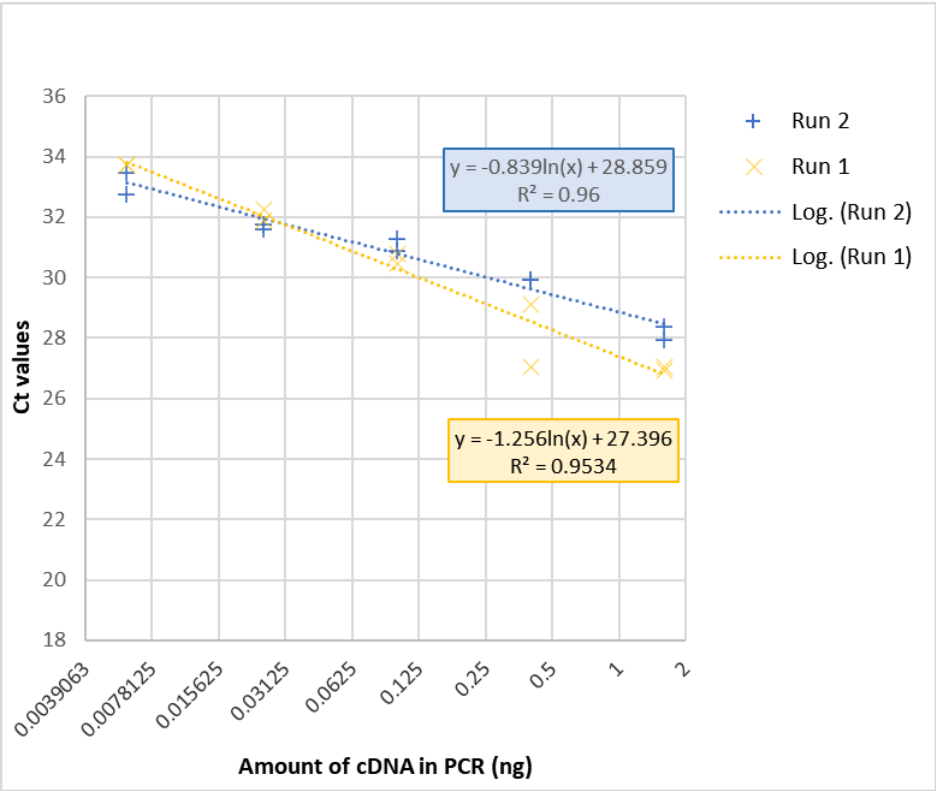


Figure 3.4. miRCURY LNA Probe PCR Kits for miR-99a-5p on serial dilutions of cell line cDNA. miRNA extracted from cell line and its 4-fold serial dilutions were used in RT reactions and subsequent qPCR. Run 1 as shown in Fig. 3.2, Run 2 repeated PCR for the same RT reaction products four days later.

Subsequently, run 1 has better sensitivity than run 2 when Ct is below 32 and the amount of cDNA in PCR is more than 0.025 ng (Fig. 3.4). Furthermore, QuantiNova probe PCR is more reproducible and with linearity over a wider dynamic range compared to TaqMan PCR, as the same RT products were used for all PCRs (Fig. 3.3, 3.4).

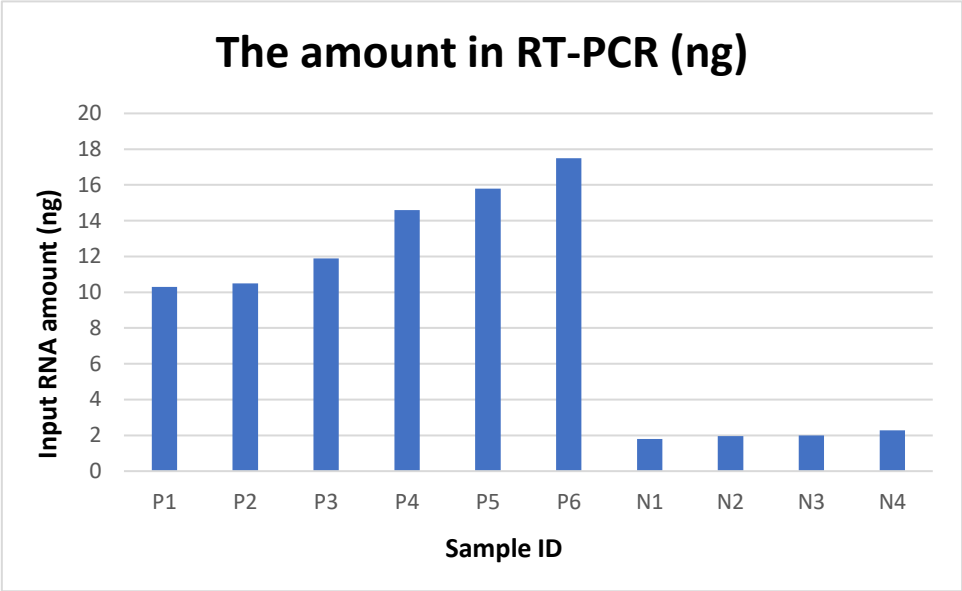
An additional 5 miRNAs (miR-342-5p, miR-26a-1-3p, miR-1185-5p, miR-146b-5p and miR-451a) were similarly tested with the QuantiNova PCR mix, but none were detectable above background readings, mostly giving Ct values in the region of 34-40. This is most probably because these miRNAs were not expressed in the cell line used; therefore, we used a range of normal lung tissues and plasma samples for further testing.

3.2.4 Assessment of miRNA levels in plasma and normal lung tissue using two PCR methods

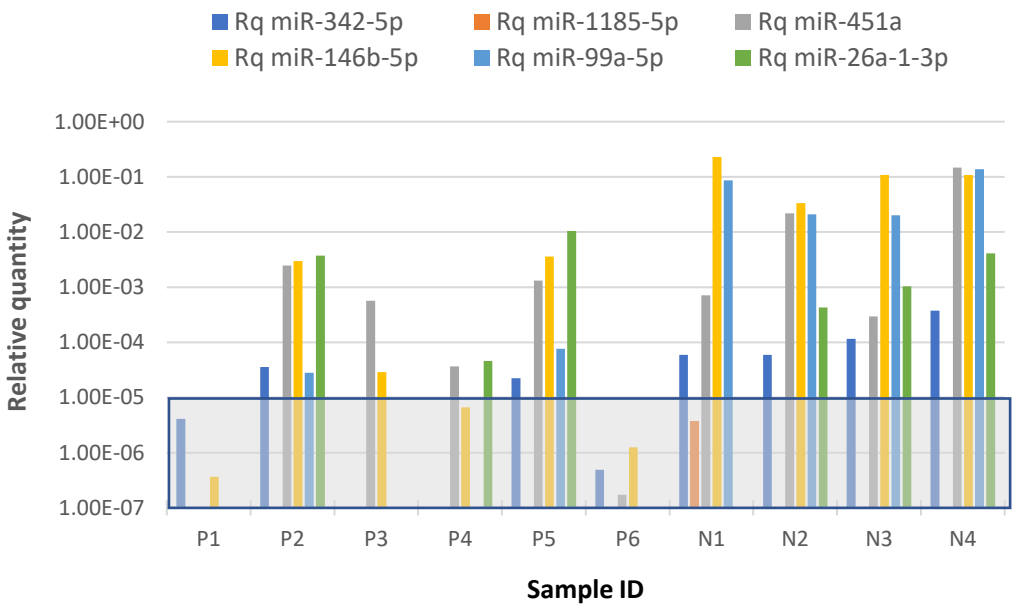
Three miRNAs (miR-342-5p, miR-1185-5p, and miR-451a) were tested with QuantiNova probe reaction mix and the other three (miR-146b-5p, miR-99a-5p, and miR-26a-1-3p) with QuantiNova multiplex reaction mix using RNA extracted from 6 plasma samples and 4 normal lung tissue samples (Fig. 3.5). Despite the low amount of starting RNA for normal tissue samples N1 to N4 (corresponding approximately to 2ng per qPCR reaction, Fig. 3.5A), the relative quantity (RQ) of all tested miRNAs was generally higher than for plasma (Fig. 3.5B & C) for which there was between 5- and 9-fold higher amounts of RNA added (approx. 10-18ng per qPCR).

miR-1185-5p was undetectable for all tested samples except N1, most likely indicating it is not expressed in the plasma and tissue used (Fig. 3.5B). Similar sporadic positive results with Ct values > 40 were found when we changed the run method from a 2-step to a 3-step protocol (Fig. 3.5C).

A


The amount in RT-PCR (ng)
Input RNA amount (ng)
20
18
16
14
12
10
8
6
4
2
0
P1
P2
P3
P4
P5
P6
N1
N2
N3
N4
Sample ID

B


Rq miR-342-5p
Rq miR-1185-5p
Rq miR-451a
Rq miR-146b-5p
Rq miR-99a-5p
Rq miR-26a-1-3p
Relative quantity
1.00E+00
1.00E-01
1.00E-02
1.00E-03
1.00E-04
1.00E-05
1.00E-06
1.00E-07
P1
P2
P3
P4
P5
P6
N1
N2
N3
N4
Sample ID

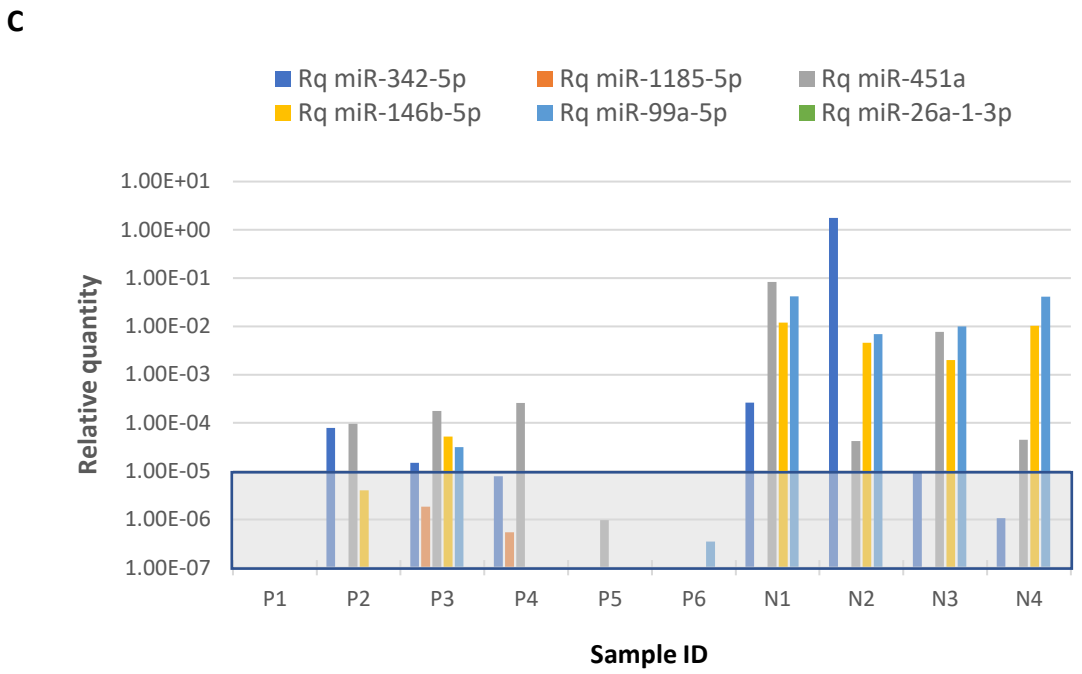


Figure 3.5. miRCURY LNA Probe PCR Kits for miR-342-5p, miR-1185-5p, and miR-451a and QuantiNova Multiplex PCR Kits for miR-146b-5p, miR-99a-5p, and miR-26a-1-3p on plasma and normal lung tissue cDNA. RNA extracted from 6 plasma samples (P1-P6) and 4 normal lung tissue samples (N1-N4) were used in RT reactions and subsequent qPCR. The amount of 10 cDNA samples for qPCR as shown in A, the relative quantity from 2-step qPCR is shown in B and 3-step qPCR in C. 2- Step qPCR method: initial denaturation at 95°C for 2 minutes followed by 45 cycles of 95 °C for 5 s and 60 °C for 30 s. 3- Step qPCR method: initial denaturation at 95°C for 2 minutes followed by 45 cycles of 95 °C for 5 s, 59 °C for 15 s, and 60 °C for 40 s. The grey box indicates the relative quantities based on Ct values approximately above 35 (which are considered no longer be fully quantitative).

When the PCR method is changed from a 2-step to a 3-step protocol, it significantly affects the quantitation of certain miRNAs (Fig. 3.5C compared to Fig. 3.5B, Fig. 3.6). The 3-step PCR method did not work for miR-26a-1-3p, whilst there was little correlation in relative quantities for miR-342-5p or miR-451a. Relative quantities of miR-99a-5p and miR-146b-5p appear to correlate better, although there is relatively little data to base this on (Fig. 3.6). Overall, the 2-step method seems to be more sensitive.

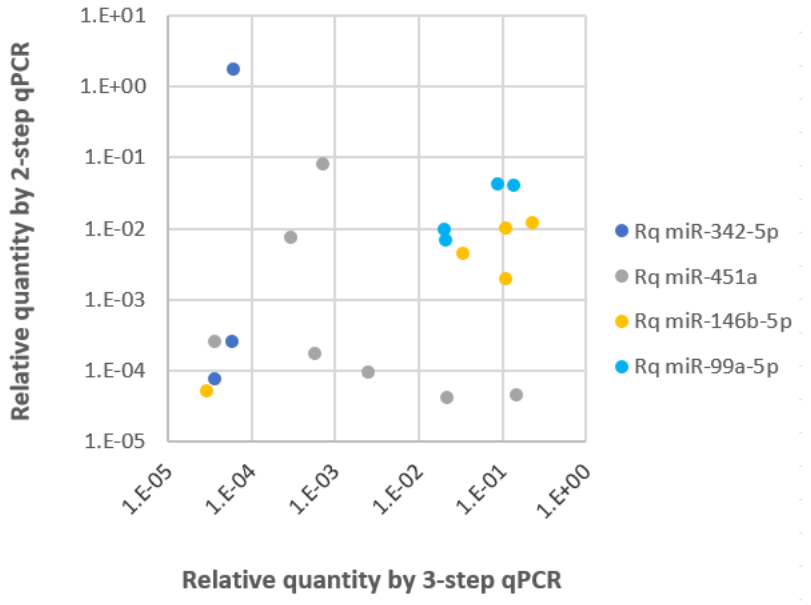
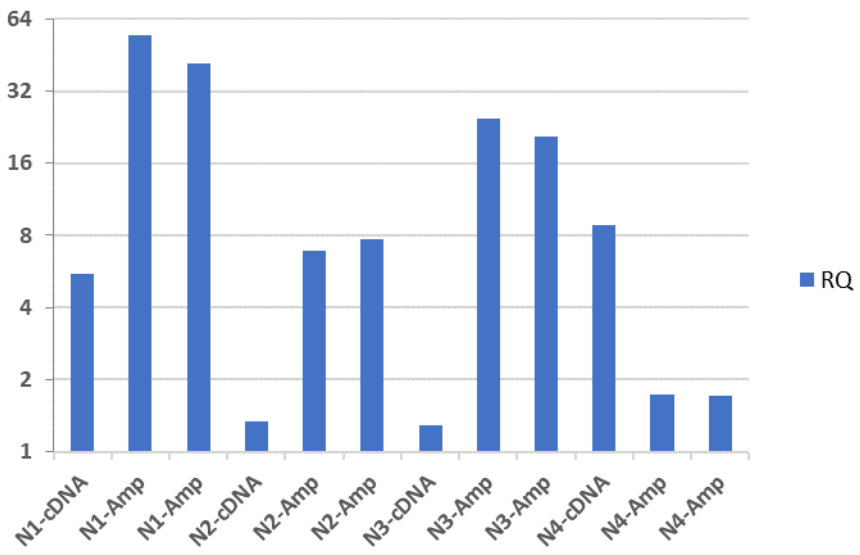


Figure 3.6. Comparison of relative quantities by 2-step qPCR and 3-step qPCR. The relative quantities less than 10^{-6} were removed (as they were based on Ct values greater than 35).

3.2.5 Assessment of miR-Amp amplification of miRNAs

Three miRNAs (miR-99a-5p, miR-26a-1-3p, and miR-1185-5p) were tested with QuantiNova probe reaction mix using RNA extracted from 6 plasma samples and 4 normal lung tissue samples (Fig. 3.7). miR-Amp reaction products and 2-step qPCR method were used for the qPCR reaction. Compared to QuantiNova Probe PCR for 6 miRNAs using normal lung tissue cDNA, there is clearly some enrichment by miR-Amp reaction, which allows multiple miRNAs to be measured from limited starting material (as expected from EV purification). However, for the miRNA extracted from normal tissue used here, there was some variation in the enrichment achieved.



	N1-Amp	N2-Amp	N3-Amp	N4-Amp
Relative SQ	0.01	0.01	0.01	0.01
Relative measurement	8.7	5.4	17.6	0.2
Enrichment by Amp	870	540	1760	20

SQ = starting quantity

Figure 3.7. Comparison of enrichment of miR-99a-5p by miR-Amp reaction (Amp) and non-amplified RT products (cDNA) for 4 normal lung tissue samples (N1-N4).

3.2.6 Validation of calibrator samples to be used for normalization

As we are testing multiple samples across numerous RT-PCR plates, we planned to include a calibrator sample, which could be used for relative quantitation normalisation. A calibrator was produced by mixing 24 plasma samples from control subjects. We tested the plasma calibrator alongside RNA extracted from cell line SK-LU-1 (cell line calibrator) for miR-99a-5p and miR-92a-3p with miR-Amp amplification and QuantiNova probe PCR reaction mix. miR-99a-5p was undetectable for the plasma calibrator but was positive for cell line (Ct average = 26.0). miR-92a-3p was detectable for both the plasma calibrator (Ct average = 30.1) and cell line (Ct average = 21.4).

We tested serial dilutions (the stock, 1:2, 1:4, 1:8, 1:16) of miR-Amp reaction products of plasma calibrator and cell line calibrator for miR-92a-3p (Fig. 3.8). Notably, Ct values for the serial dilutions of plasma calibrator (Fig. 3.8A) and cell line calibrator (Fig. 3.8B) decrease in a linear relationship with dilution, only at dilutions above 1 in 4 (0.25 to 0.0625). Lower dilutions (undiluted and 2-fold) correspond to Ct values similar or greater than for the 1 in 4 (i.e., 31 for plasma calibrator and 19.8 for cell line calibrator, Fig. 3.8A&B).

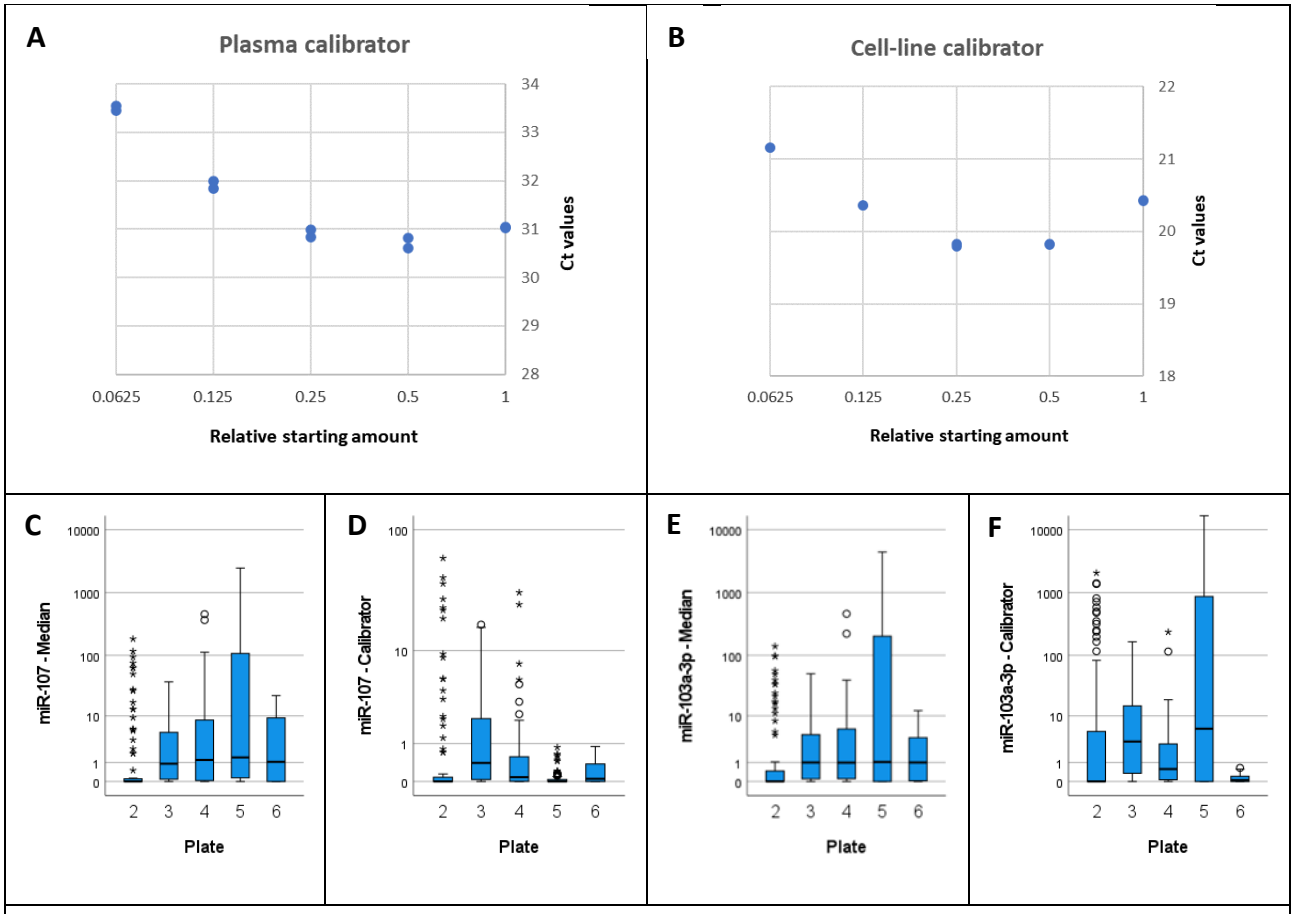


Figure 3.8. Assessment of the relative quantitation of miR-Amp reaction products of plasma calibrator (A) and cell line calibrator (B), using serial dilutions for miR-92a-3p qPCR reaction and use of the calibrator as a normalisation between plates (D & F), comparted to use of the median (C & E).

This indicates that the qPCR reaction for both plasma calibrator and cell line calibrator are quantitative only if the miR-Amp reaction product is diluted appropriately. We therefore

chose a 10-fold dilution for subsequent experiments for measuring miRNA in clinical samples. However, while analysing the qPCR data for the targeted miRNAs, it was noticed that the Ct values of the calibrators vary substantially between the plates for different miRNAs (Fig. 3.8D&F), resulting in more variation between plates than if using no calibrator. We therefore divided by the median expression for each miRNA to produce a normalisation between the plates (Fig. 3.8C&E), these were more similar between plates.

Having analysed the Ct values for calibrator samples in the qPCR analysis of cases and controls (see section 3.3.2), it was determined that they were not suitable for normalisation between plates, as use of the single-sample calibrator values introduced too much variability. Therefore, normalisation was performed using median Ct value per PCR plate, based on up to 92 samples (with an even number of cases and controls).

3.2.7 Potential artefacts in qPCR following RT and miR-Amp reactions

Occasionally we noted signals in qPCR at later Ct values (35 - 45), even in no RNA controls. Presumably these are due to artefacts created within either the RT reaction, or the miR-Amp reaction, as we did not see them in our no cDNA controls (i.e., qPCR with water instead of RT-miR-Amp products). We investigated whether the signals of no RNA controls are because of miR-Amp reactions or the RT. We tested RT and miR-Amp reaction products of 2 tissue RNA samples, cell line calibrator (with the same starting amount) and 3 no RNA controls for miR-1228-3p by qPCR (Fig. 3.9). All of the no RNA controls resulted in undetermined signal, for either RT or miR-Amp reaction products, although signals above threshold appear in the region of Ct of 39-41 (Fig. 3.9). These signals occurred whether or not the miR-Amp reaction was performed, so therefore are likely to be due to RT artefacts. It should be noted that, unlike the tissue and cell line samples, the signals for the no RNA controls are not shifted

significantly to the left (lower Ct) by the miR-Amp reaction, indicating that this does not amplify the artefact to the same extent as true miRNA products (Fig. 3.9).

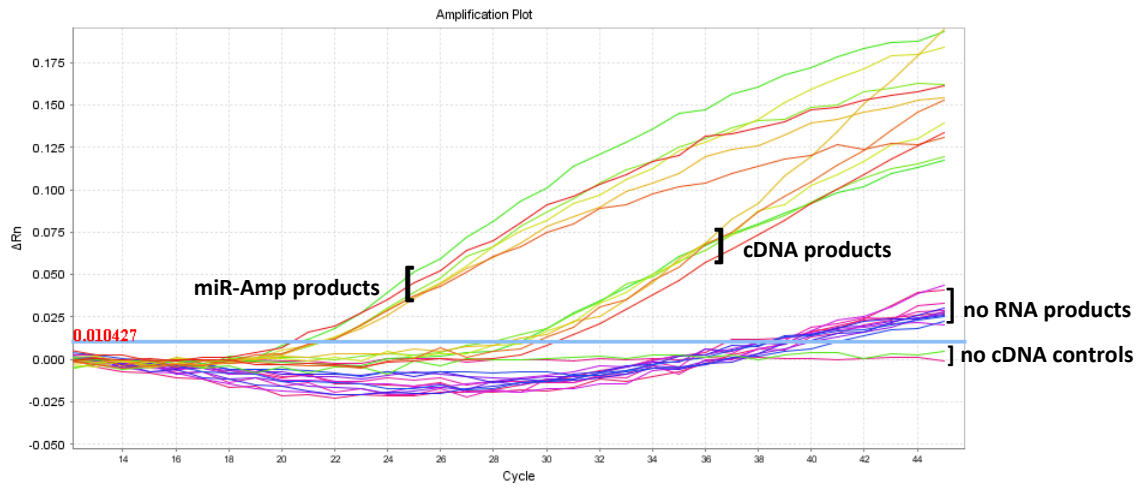


Figure 3.9. Exploration of potential artefacts in qPCR following RT and miR-Amp reactions by testing RT and miR-Amp reaction products of 2 tissue RNA samples, cell line calibrator (with the same starting amount) and 3 no RNA controls for miR-1228-3p by qPCR. No RNA products include both no RNA in the RT reaction and miR-Amp reaction.

To further explore the most possible reason for artefacts in our previous experiments, we tested miR-146b-5p using a positive control cell line calibrator, a yeast tRNA sample and a no RNA control. Based on the fact that yeast tRNA gives a strong product in the RT (without miR-Amp) it is clear that some random priming can take place in the RT reaction that is subsequently detected as miRNA in the qPCR, despite this being a TaqMan-based assay with the additional specificity provided by the probe.

We further investigated the nature of the artefacts by performing a QIAquick column-based PCR purification after amplification and before qPCR. The results showed that the PCR quantitation using undiluted, purified amplification reactions was approximately the same as

for 1/10 diluted unpurified RT-amplification products, indicating significant losses on purification, making this impractical for inclusion in the quantitation of miRNAs. No RNA controls that gave a signal without purification (in the region of Ct 35-40), still gave a signal after purification, indicating that any artefact due to RT or amplification was not easily purified away from true RT product.

3.3 Measurement of EV miRNAs in clinical plasma samples

In the original HTG analysis (preceding my PhD study), 18 miRNAs were identified as differentially expressed between lung cancer cases and controls in the plasma EV fractions (ultracentrifugation protocol). In addition, a 4-miRNA panel was identified by lasso regression; however, one of the miRNAs (miR-582-5p) had no suitable qRT-PCR assay available, the other 3 miRNAs were included.

Here we seek to validate these miRNAs as potential lung cancer biomarkers, using a combination of column-based EV isolation and quantitative RT-PCR, on an independent cohort of cases and controls, matched for age, sex and smoking status.

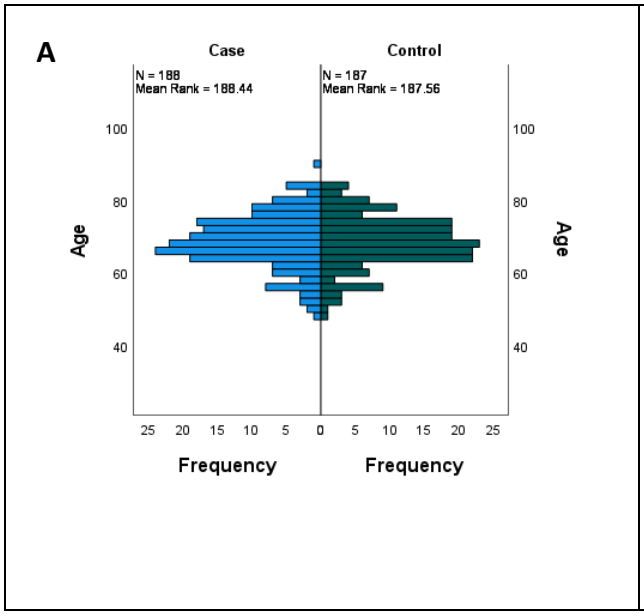
3.3.1 Clinical cohort

The Liverpool Lung Project has collected plasma samples from lung cancer cases and from controls subjects (healthy individuals or those who received a negative diagnosis for lung cancer in local hospitals). Informed consent for use of samples and collection of subsequent lung cancer diagnosis and outcome data was provided; research was conducted under the LLP ethical approval (Liverpool Research Ethics Committee). Plasma samples taken within 3 months of a lung cancer diagnosis, but before any treatment, were identified and histological type, stage, age at diagnosis, sex and other clinical and epidemiological information collated

(e.g. smoking, COPD). The same data (with the exception of cancer diagnosis) was collated for controls and a single control was matched to each case, having the same sex, age. Where possible, smoking status was the same and case and control had approximately the same duration of smoking (although these were not as closely matched as in the discovery HTG cohort). This helps to control for confounding factors when performing analysis of e.g. case *versus* control. Ages ranged from 48 to 89 (mean = 68) and 60.3% were male; Ages, storage time and smoking duration are given in Fig. 3.10. Age (Mann-Whitney P = 0.94) and storage time (Mann-Whitney P = 0.31) were not significantly different between cases and controls, and neither was sex (chi squared P = 0.95). However, despite similar profiles (Fig. 3.10E), smoking duration was significantly higher (Mann-Whitney P < 0.001) in those with lung cancer (median = 43, interquartile range = 19) than controls (median = 35, interquartile range = 37). This is related to the greater proportion of never smokers in controls and fewer current smokers (Table 3.1). Notably, as might be expected there was a trend for more cases of COPD in those with lung cancer (chi squared P = 0.06). Further details of the cohort are given in Fig. 3.10, 3.11, 3.12 and Table 3.1, 3.2.

Table 3.1. Clinical characteristics by case/control status

Characteristics	Case (n, %)	Control (n, %)	Total (n, %)	Chi Sq. Test
COPD	53 (33.1%)	33 (23.4%)	86 (28.6%)	3.5
No COPD	107 (66.9%)	108 (76.6%)	215 (71.4%)	P = 0.06
Current	70 (40.2%)	47 (25.5%)	117 (32.7%)	
Quit_under3y_Duration_over30y	23 (13.2%)	12 (6.5%)	35 (9.8%)	
Quit_over3y_Duration_over30y	47 (27.0%)	50 (27.2%)	97 (27.1%)	25.2
Quit_over3y_Duration_under30y	23 (13.2%)	39 (21.2%)	62 (17.3%)	$P = 5 \times 10^{-5}$
Never	11 (6.3%)	36 (19.6%)	47 (13.1%)	
Male	113 (60.1%)	113 (60.4%)	226 (60.3%)	0.004
Female	75 (39.9%)	74 (39.6%)	149 (39.7%)	P = 0.95
Total	188	187	375	



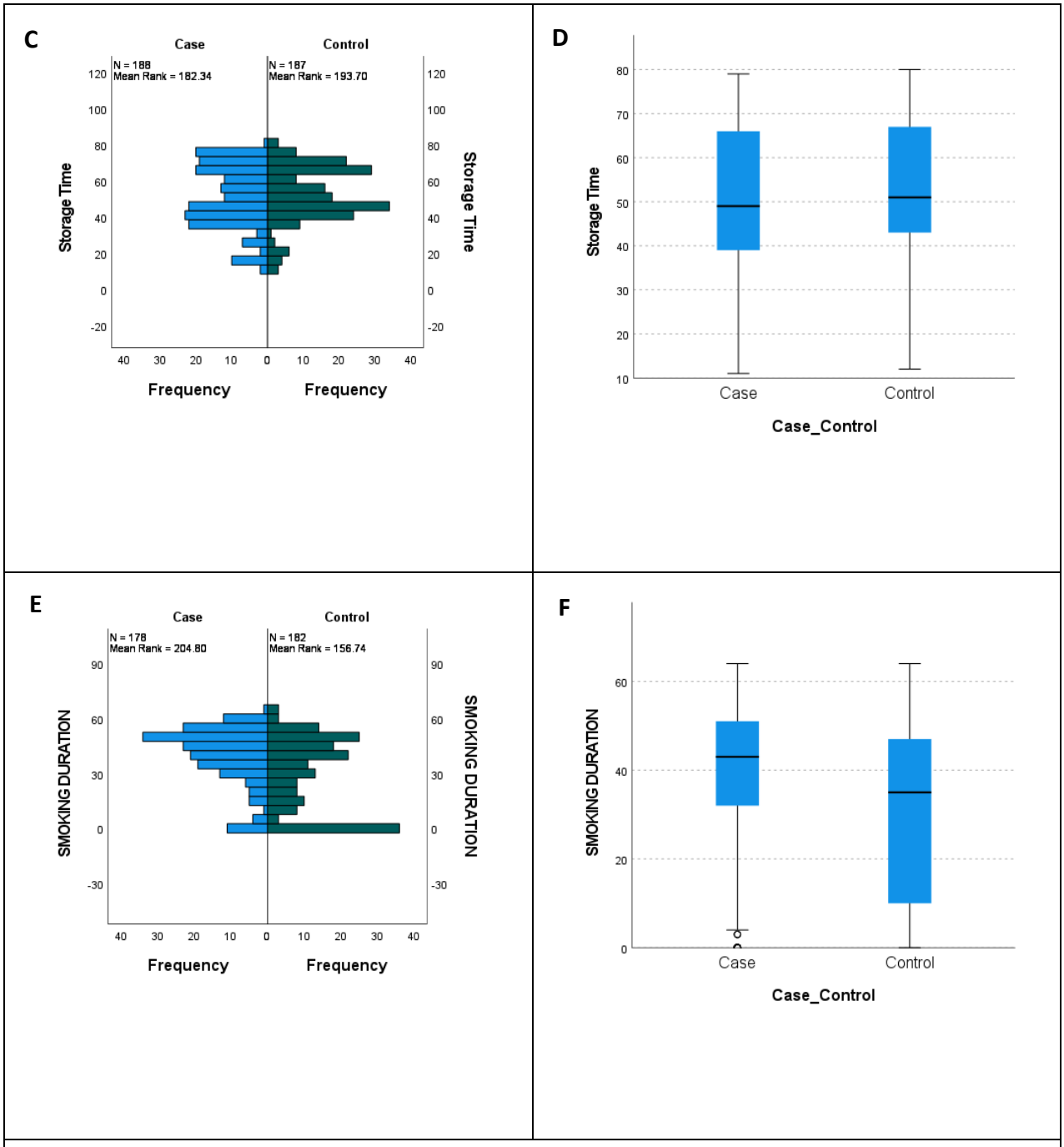


Figure 3.10. Distribution of age (A, B), storage time (C, D) and smoking duration (E, F) of the validation cohort; histograms (A, C, E) and boxplots (B, D, F).

The 188 NSCLC cases consisted of predominantly early stage ADC and SqCCa (Fig. 3.11, Table 3.2).

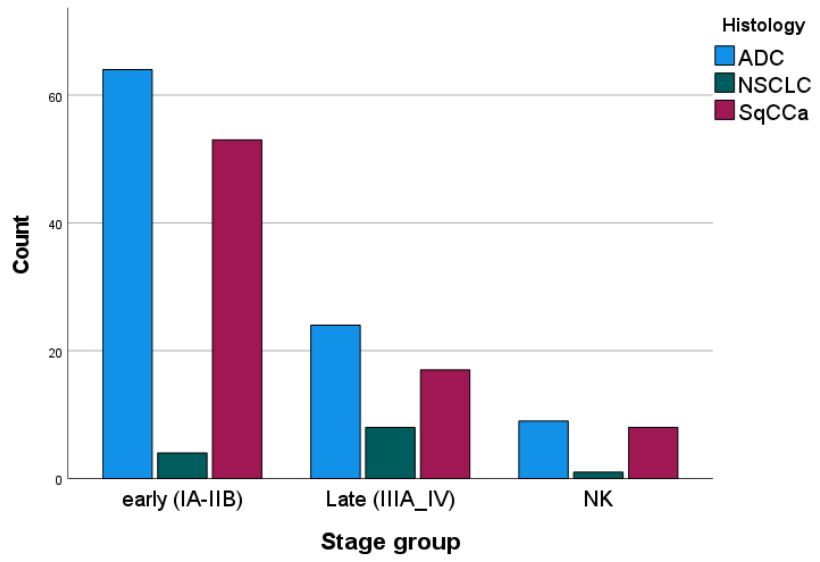


Figure 3.11. Bar chart of histology and stage group for cases used in validation.

Table 3.2. The histology of early stage group (IA-IIB), late stage group (IIIA_IV) and not known (NK)

Characteristics			**Histology**			**Total**
			ADC	**NSCLC**	**SqCCa**	
Stage group	*Early (IA-IIB)*	*Count*	64	4	53	121
		% within Histology	66.0%	30.8%	67.9%	64.4%
	Late (IIIA_IV)	*Count*	24	8	17	49
		% within Histology	24.7%	61.5%	21.8%	26.1%
	NK	*Count*	9	1	8	18
		% within Histology	9.3%	7.7%	10.3%	9.6%
Total		*Count*	97	13	78	188

Note: There are more cases without specific histology (NSCLC) in the late-stage group, which can be attributed to the limited clinical sampling for advanced lung cancer (e.g. EBUS cytology samples rather than tumour resection).

In terms of all-cause mortality (Fig. 3.12), this was significantly worse for cases than controls and poor outcome was associated with later stages (Kaplan-Meier P = 10^{-29}).

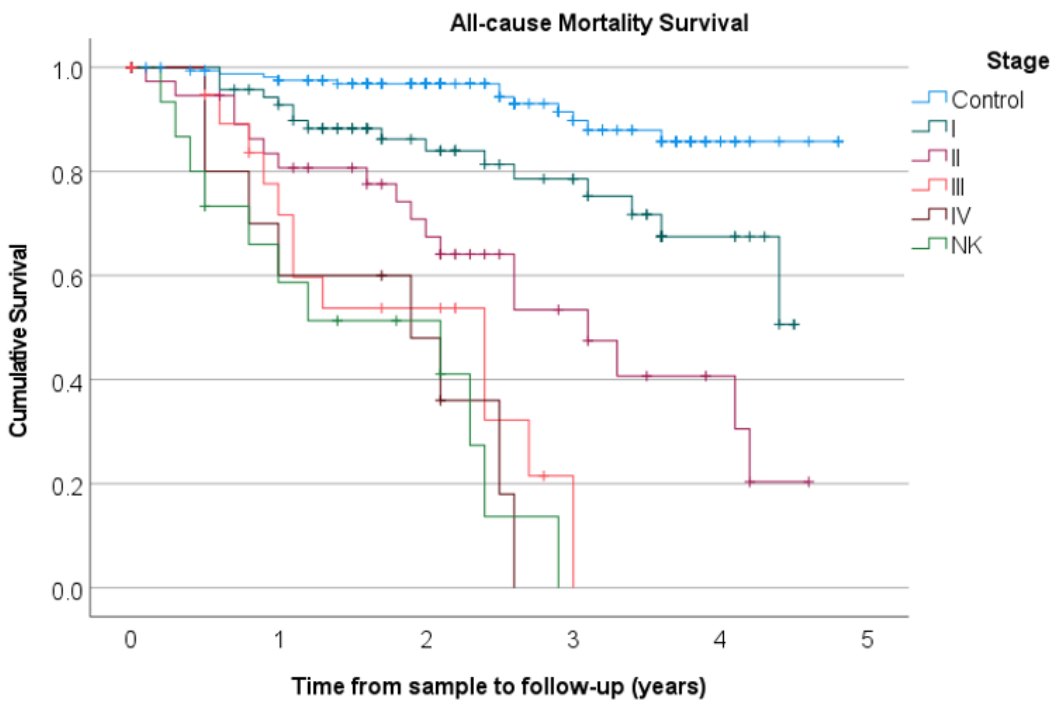


Figure 3.12. Survival for all-cause mortality from time of sample, including both cases (by stage, NK = not known) and controls.

3.3.2 Analysis of miRNA expression by qRT-PCR across 375 samples

In total, 190 pairs of blood samples from lung cancer patients and control subjects were included in the qRT-PCR validation, EV miRNAs were extracted from all but 1 case sample (after filtration through a 0.45 µm syringe filter, there was not enough material for the next step due to a clogged column). Three controls and one case were excluded from the qPCR data analysis as during follow up; the case was found to be a potential metastasis rather than primary lung cancer, and the controls had some evidence of a subsequent lung cancer or other cancer diagnosis. Therefore, eventually, the qPCR data of 375 samples (188 cases and 187 controls) were included in the statistical analysis.

The matched cases and controls were assigned to adjacent wells on the RT plates, with 46 pairs analysed per RT-PCR plate; one miRNA was analysed in each single reaction. Also,

included on the RT plate was the cell line calibrator, the plasma calibrator and the no RNA control. Duplicate qPCR plates were run in parallel using the Applied Biosystems 7500 and 7500 Fast instruments, with the same plate layout as the RT plate, but including a no DNA control. Median Ct values were used to normalise each plate for each miRNA.

miR-103a-3p was measured for 188 cases and 187 controls (plates 2-6) and was significantly higher in cases (Mann-Whitney P = 0.001, Fig. 3.13A, Table 3.3). The other 3 most significant up-regulated miRNAs are let-7c-3p (Mann-Whitney P = 0.007), miR-107 (Mann-Whitney P = 0.008) and miR-3149 (Mann-Whitney P = 0.010) (Fig. 3.13B, C & D, Table 3.3). Two miRNAs showed some trend towards significance (miR-1228-3p, Mann-Whitney P = 0.057; miR-301b, Mann-Whitney P = 0.086, Fig. 3.13E & F, Table 3.3).

Table 3.3. Comparative qPCR measurements (case *vs* control) for 11 miRNAs measured across plates 2-6

miRNA	Control (n)	Control median (IQ range)	Case (n)	Case median (IQ range)	FC (median)	Significance* (P)
miR-103a-3p	187	0.2 (0 - 3.57)	188	1.02 (0.03 - 9.13)	5.1	0.001
let-7c-3p	187	1.03 (0.18 - 3.18)	188	1.92 (0.4 - 5.16)	1.86	0.007
miR-107	187	0.21 (0 - 5.22)	188	1.23 (0.02 - 16.24)	5.86	0.008
miR-3149	187	1.02 (0.27 - 3.16)	188	2.07 (0.44 - 5)	2.03	0.010
miR-1228-3p	187	0.97 (0.35 - 2.43)	188	1.25 (0.44 - 5.02)	1.29	0.057
miR-301b	187	0.93 (0.15 - 4.33)	188	1.41 (0.36 - 5.64)	1.52	0.086
miR-146b-5p	187	1.27 (0.17 - 8.7)	188	1.03 (0.22 - 5.14)	0.81	0.487
miR-514b-5p	187	1.18 (0.36 - 2.97)	188	1.17 (0.4 - 3.3)	0.99	0.488
miR-342-5p	187	0.96 (0.19 - 3.17)	188	1.02 (0.22 - 3.94)	1.06	0.585
miR-1178-5p	187	0.89 (0.15 - 2.88)	188	0.86 (0.2 - 3.06)	0.97	0.818
miR-26a-1-3p	187	1.2 (0.62 - 64.44)	188	1.2 (0.69 - 53.37)	1	0.981

* Kruskal-Wallis test
FC = fold change.

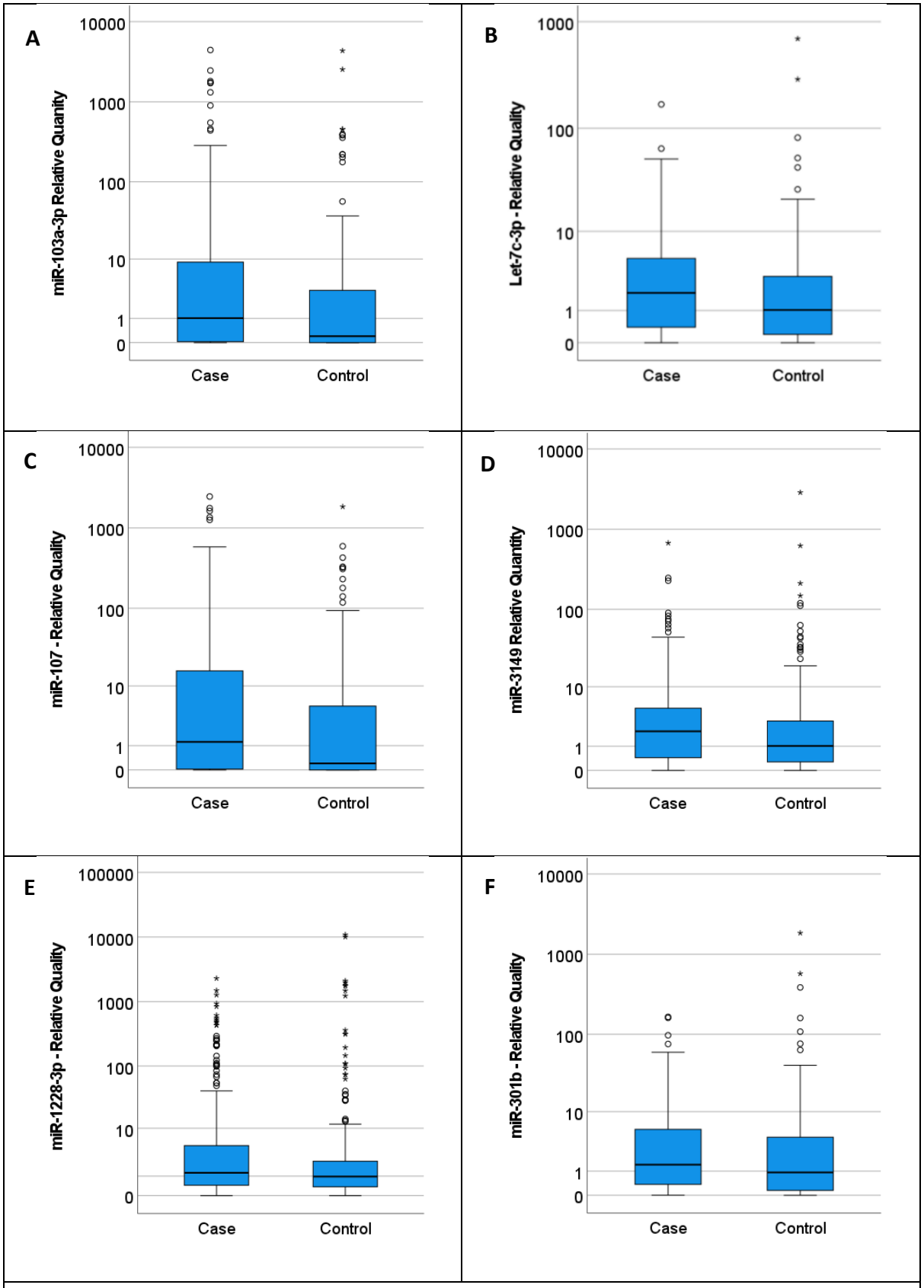


Figure 3.13. Boxplots of the relative qualities of the top 6 most significant miRNAs in cases and controls.

3.3.3 Analysis of miRNA expression by qRT-PCR across 182 or 90 samples

Seven miRNAs were measured for 182 or 90 samples and showed no significant difference in miRNA expression between cases and controls (Table 3.4). As the P values were so high, these miRNAs were not analysed on further RT-PCR plates.

Table 3.4. Comparative qPCR measurements (case *vs* control) for 7 miRNAs measured across plates 2-3 or plate 2 only

miRNA	Control (n)	Control median (IQ range)	Case (n)	Case median (IQ range)	FC (median)	Significance* (P)
miR-1277-5p	45	0 (0 - 0.87)	45	0 (0 - 1.81)	NA	0.409
miR-1205	91	0.83 (0.23 - 4.38)	91	1.06 (0.24 - 5.2)	1.28	0.436
let-7a-3p	45	0.83 (0.4 - 4.33)	45	0.91 (0.4 - 1.59)	1.1	0.539
miR-1247-3p	45	0 (0 - 1.01)	45	0 (0 - 1.3)	NA	0.687
miR-1185-5p	45	0.97 (0.48 - 1.96)	45	1.13 (0.51 - 1.54)	1.16	0.701
miR-5093	45	0.32 (0 - 1.73)	45	0 (0 - 1.06)	NA	0.753
miR-6872-5p	91	1.01 (0.36 - 2.88)	91	1.01 (0.35 - 3.62)	1	0.967

* Kruskal-Wallis test
FC = fold change, NA = not available.

Given that the selected EV miRNAs did not validate by qRT-PCR, one of the quality control measures taken was to check the assay IDs of the reagents used against the original statistical analysis output. At this stage, it was realised that rather than using assays for miR-509-3-5p, we had used an assay for miR-5093. Hence not only are we unable to validate miR-509-3-5p, but we are unable to recapitulate the 4-miRNA panel identified originally. Notably only one of the miRNAs in this panel was validated by qRT-PCR, so it is unlikely that the 4-miRNA panel would have been validated anyway.

3.3.4 miRNAs for prediction of case status

Receiver Operating Characteristic (ROC) curve analysis for differentially expressed miRNAs (higher in cases, miR-103a-3p, let-7c-3p, miR-107, miR-3149, miR-1228-3p, miR-301b) indicates that some miRNAs (miR-103a-3p, let-7c-3p, miR-107, miR-3149) are significant

predictors of case status (Fig. 3.14 and Table 3.5). However, Cox Regression analysis indicated that none of the miRNAs were significantly associated with case status.

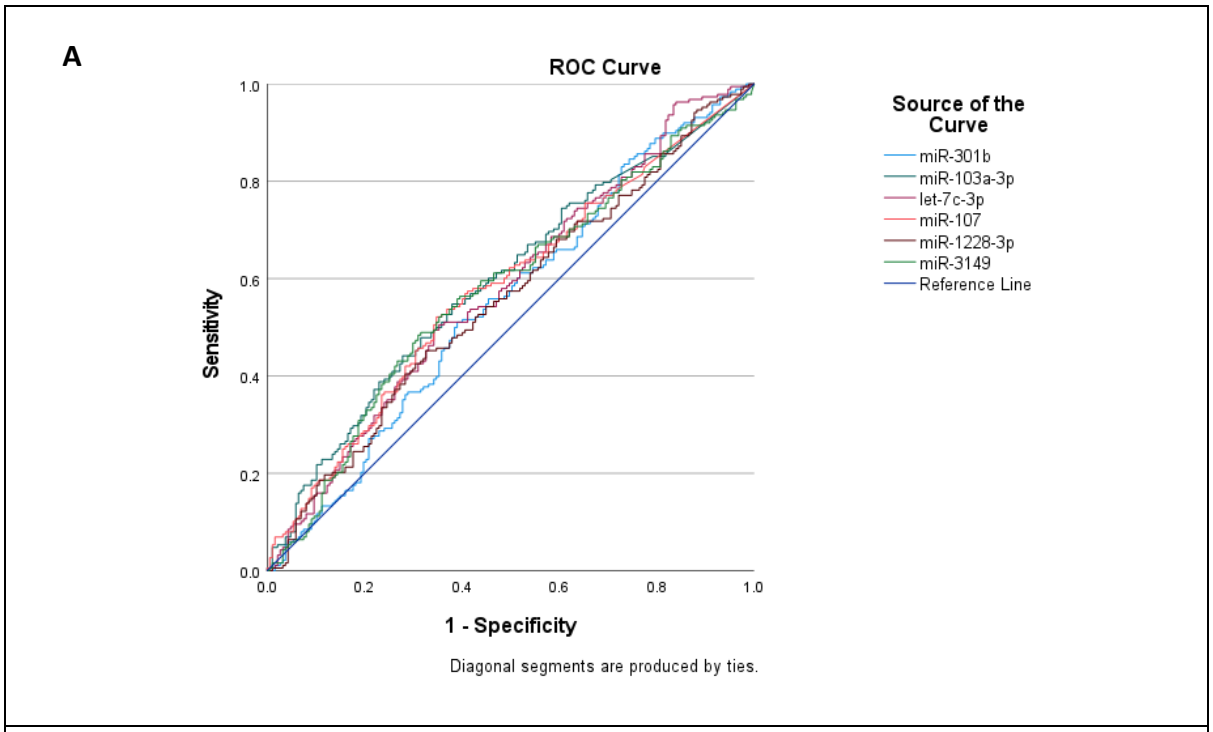


Figure 3.14. ROC curve analysis for those miRNAs expressed at higher levels in cases than controls (A) and the most significant EV miRNA miR-103a-3p (B).

Table 3.5. Receiver Operating Characteristic (ROC) curve analysis for differentially expressed miRNAs

Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
miR-103a-3p	0.595	0.029	0.001	0.538	0.652
let-7c-3p	0.580	0.029	0.007	0.522	0.637
miR-107	0.579	0.029	0.008	0.521	0.637
miR-3149	0.577	0.03	0.010	0.519	0.635
miR-1228-3p	0.557	0.03	0.057	0.499	0.615
miR-301b	0.551	0.03	0.086	0.493	0.609

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

3.3.5 Summary of case *vs* control miRNA validation

18 miRNAs initially identified as differentially expressed on HTG analysis were tested using alternative EV isolation and miRNA quantitation methodology; 11 out of 18 miRNAs were measured for samples on plates 2-6 (188 cases, 187 controls), 2 miRNAs were measured for samples on plate 2 and 3 (91 cases, 91 controls), 5 miRNAs were measured for samples on plate 2 only (45 cases, 45 controls). Only 4 were found to be differentially expressed, being higher in cases than controls, but 2 further miRNAs had a trend for overexpression (Table 3.3).

3.3.6 Correlation between the tested miRNAs and correlation with smoking duration

Most of the correlations between the tested 18 miRNAs are significant at either the P = 0.01 level (2-tailed) or the P = 0.05 level (2-tailed) (Table 3.6). However, some correlations, e.g. miR-107 vs miR-103a-3p (Spearman R 0.91, R^2 = 0.832, Fig. 3.15A) were stronger than others, e.g. miR-107 *vs* miR-3149 (Spearman R 0.17, R^2 = 0.019, Fig. 3.15B).

We have also performed the correlation between the tested miRNAs and smoking duration. Although smoking duration is slightly higher in cases, none of the miRNAs that were associated with case status correlated (all $P > 0.134$) with smoking duration. The miRNA most closely correlated with smoking duration was miR-6872-5p, however, the correlation was poor and not statistically significant (correlation coefficient = -0.099, $P = 0.19$, Fig. 3.16).

Table 3.6. Spearman's correlations between the tested miRNAs

	miR-301b	miR-103a-3p	let-7c-3p	miR-514b-5p	miR-26a-1-3p	miR-107	miR-146b-5p	miR-1178-5p	miR-1228-3p	miR-342-5p	miR-3149	miR-1205	miR-6872-5p	let-7a-3p	miR-1185-5p	miR-1247-3p	miR-5093
miR-103a-3p	0.36																
let-7c-3p	0.59	0.51															
miR-514b-5p	0.67	0.44	0.65														
miR-26a-1-3p	0.23	0.26	0.33	0.32													
miR-107	0.36	0.91	0.47	0.41	0.19												
miR-146b-5p	0.34	0.31	0.35	0.42	0.20	0.35											
miR-1178-5p	0.59	0.54	0.62	0.63	0.32	0.49	0.29										
miR-1228-3p	0.61	0.39	0.60	0.67	0.29	0.39	0.29	0.54									
miR-342-5p	0.69	0.49	0.66	0.73	0.29	0.45	0.42	0.69	0.60								
miR-3149	0.47	0.18	0.45	0.46	0.21	0.17	0.23	0.35	0.44	0.41							
miR-1205	0.60	0.43	0.59	0.60	0.36	0.43	0.40	0.57	0.61	0.64	0.35						
miR-6872-5p	0.41	0.32	0.39	0.55	0.26	0.32	0.34	0.48	0.59	0.42	0.30	0.49					
let-7a-3p	0.27	0.23	0.20	0.40	0.16*	0.22	0.17*	0.17*	0.24	0.21	0.25	0.22	0.19*				
miR-1185-5p	0.53	0.43	0.59	0.57	0.41	0.54	0.53	0.62	0.50	0.66	0.42	0.54	0.33	0.19*			
miR-1247-3p	0.52	0.45	0.50	0.59	0.45	0.55	0.40	0.59	0.33	0.55	0.46	0.50	0.26	0.18	0.59		
miR-5093	0.57	0.50	0.41	0.57	0.33	0.53	0.54	0.57	0.38	0.67	0.36	0.48	0.20*	0.16*	0.57	0.61	
miR-1277-5p	0.42	0.53	0.32	0.54	0.22	0.58	0.69	0.37	0.31	0.47	0.18*	0.43	0.36	0.19*	0.41	0.43	0.52

*. Correlation is NOT significant at the 0.05 level (2-tailed).

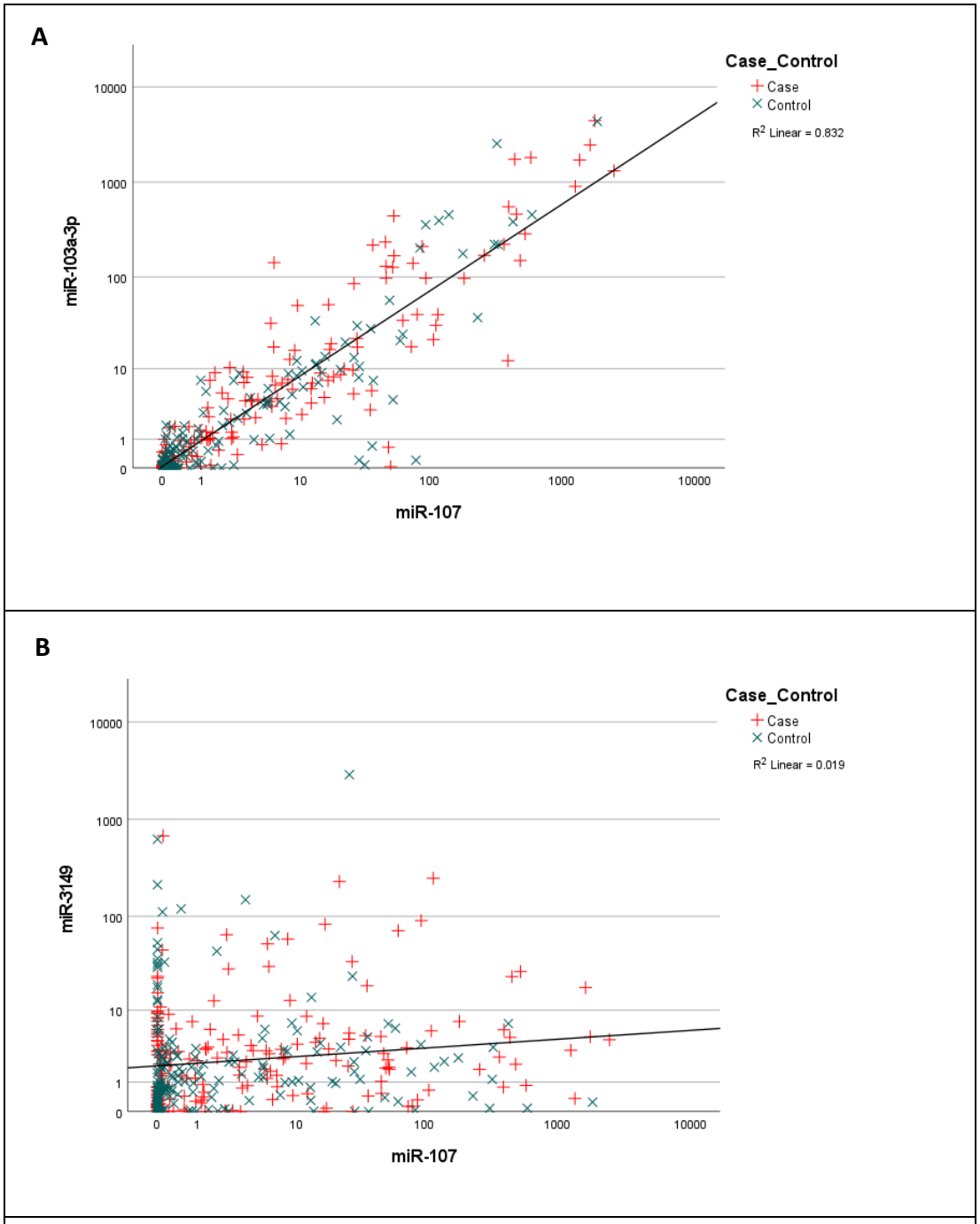


Figure 3.15. Correlations between miRNAs: miR-107 *vs* miR-103a-3p (A) and miR-107 *vs* miR-3149 (B).

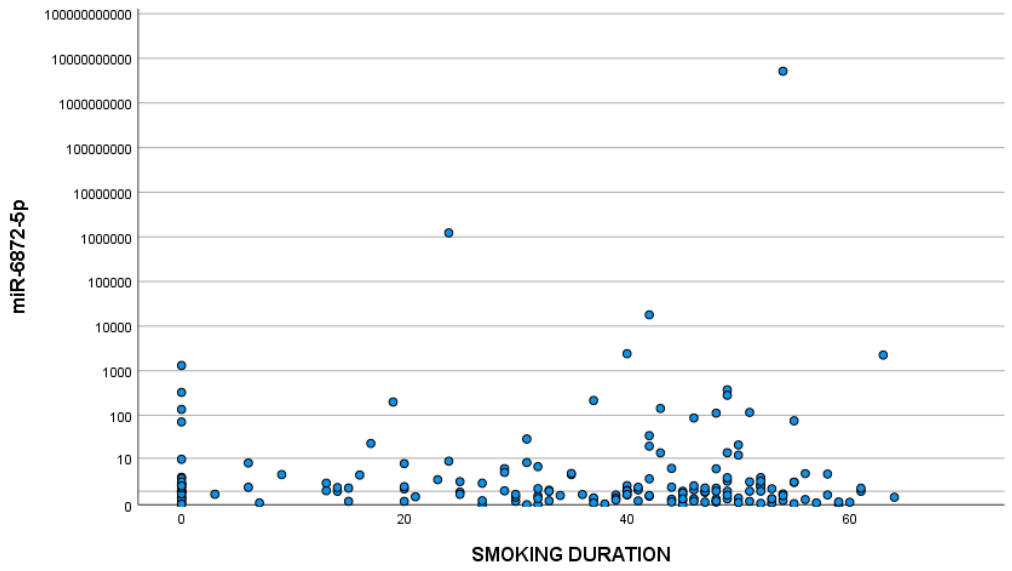


Figure 3.16. Correlations between smoking duration and miR-6872-5p.

3.3.7 Associations between miRNAs and COPD

Some miRNAs were shown to be associated with COPD status, either higher in those with COPD (miR-1277-5p, Mann Whitney P = 0.041, Fig. 3.17A) or lower (miR-1228-3p, Mann Whitney P = 0.055, Fig. 3.17B).

Furthermore, the levels of these miRNAs could be used to predict COPD status, see Area under Curve (AUC) analysis (Fig. 3.17C): miR-1277-5p, AUC = 0.638 (95% CI 0.492 – 0.785), P = 0.070; miR-1228-3p, AUC = 0.347 (95% CI 0.198 – 0.496), P = 0.046. As these miRNA associations were in the opposite direction, a ratio miR-1277-5p/miR-1228-3p gave a better prediction, however, it still suggested a poor discrimination [AUC = 0.667 (95% CI 0.518 – 0.816), P = 0.029].

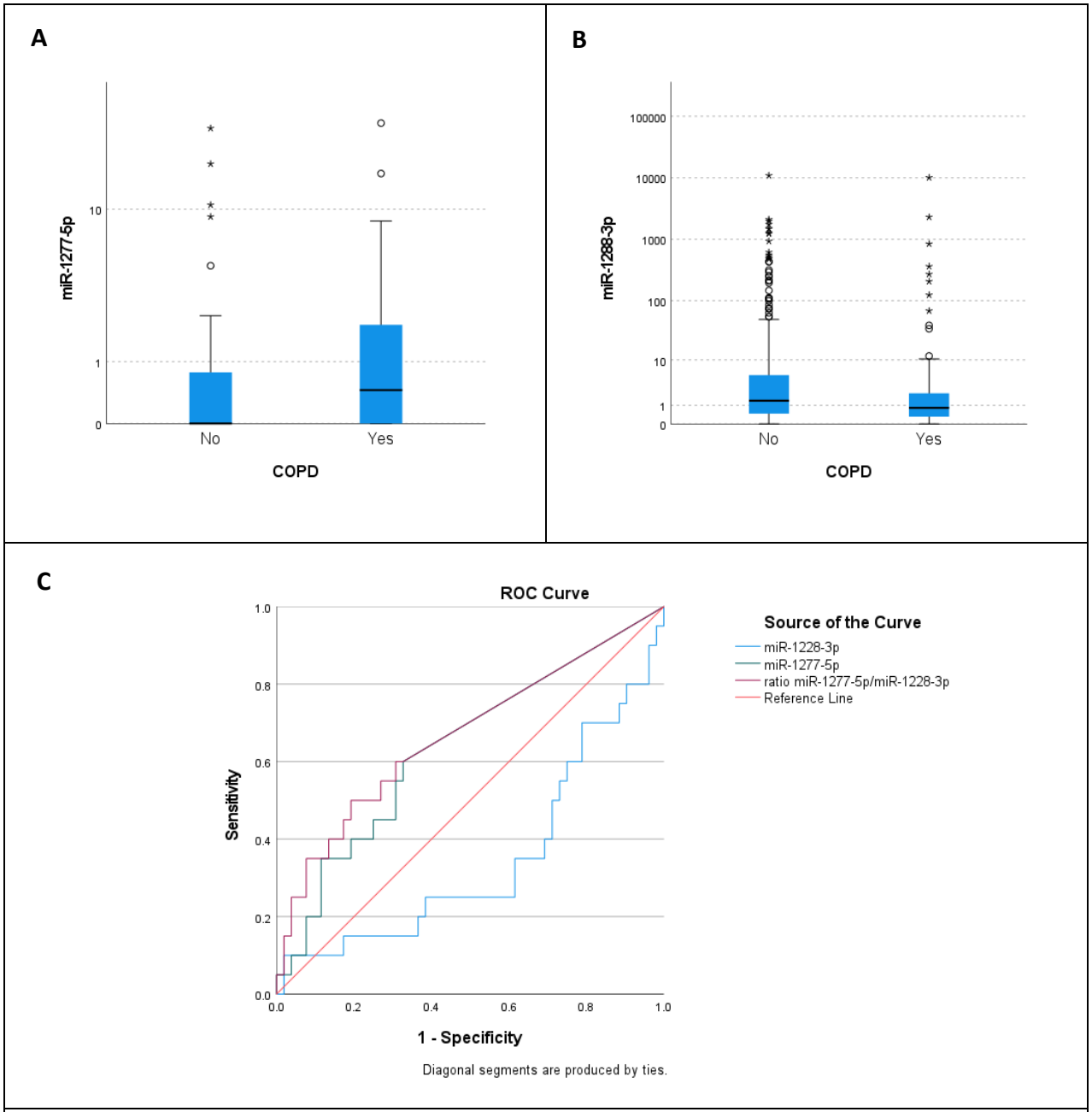


Figure 3.17. Boxplots of miR-1228-3p (A) and miR-1277-5p (B) levels in those with and without COPD. AUC of miR-1228-3p, miR-1277-5p and the ratio of miR-1277-5p to miR-1228-3p (C).

3.3.8 LLPv2 lung cancer risk model

The Liverpool Lung Project (LLP) lung cancer risk model (version 2, LLPv2) provided the percentage risk of lung cancer over the next 5 years for individuals, based on a series of questions relating to known risk factors (age, sex, smoking duration, prior respiratory disease, asbestos exposure, family history of lung cancer and personal history of cancer) (Cassidy, Myles et al. 2008, Field, Vulkan et al. 2021). It was used in the UK Lung Cancer Screening (UKLS) trial to select individuals for low-dose CT screening (Field, Duffy et al. 2016, Field, Duffy et al. 2016); it is also being used in the NHS Targeted Lung Health Check programme (with a threshold of an LLPv2 predicted 2.5% risk of lung cancer within the next 5 years (NHS 2019). This threshold was selected to stratify the screened population, to limit unnecessary imaging, whilst providing good sensitivity for lung cancer detection and a "number needed to screen" that was considered cost-effective.

Within the validation cohort, a high risk score was associated with case status, as expected, (Mann Whitney P = 0.0005, Fig 3.18B). However, some controls also had a high risk score (>2.5%, Fig 3.18A).

Only let-7a-3p was associated with LLPv2 risk score, being lower in high-risk subjects (Mann Whitney P = 0.019, Fig. 3.18C). However, there was not a significant correlation to LLPv2 value.

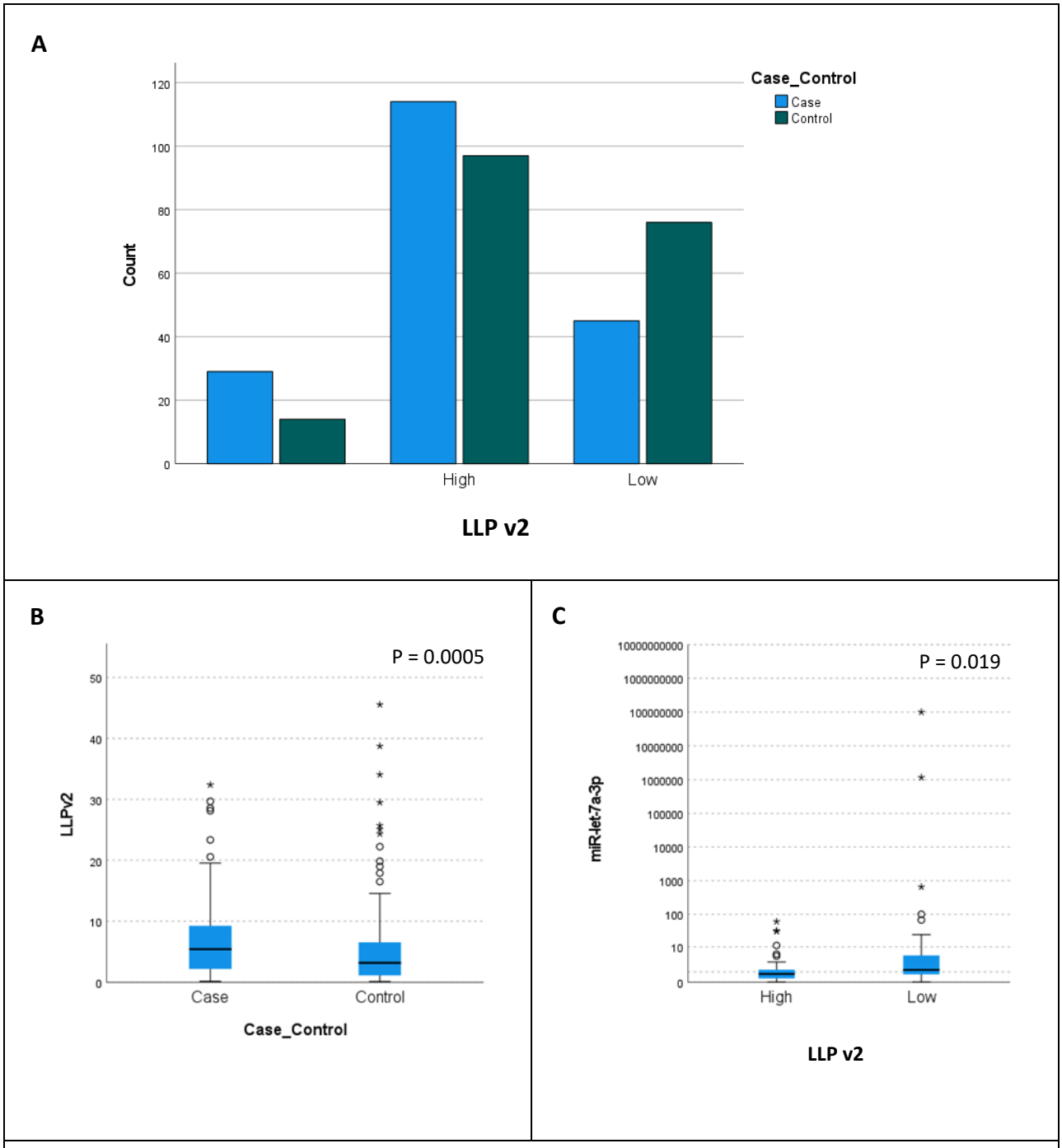


Figure 3.18. The association of LLP risk score with case status (A, B) and let-7a-3p (C). LLP risk score (LLPv2, % risk of lung cancer within 5 years) was classified as high (≥2.5%) or low (A, C) or treated as a continuous variable (B).

Chapter 4 HTG EV miRNA data reanalysis

Since the HTG-based differential expression of the selected miRNAs, as a diagnostic molecular signature of lung cancer patients, was not consistently validated by the qRT-PCR of the plasma EVs validation cohort, we considered a reanalysis of the original HTG data on EV miRNA, in order to understand the possible reasons for this. Furthermore, a new software tool HTG REVEAL analysis became available, including improved quality control and integration of differential expression analysis (based on EdgeR).

Performing this reanalysis of EV miRNA HTG data aims to:

(1) Address technical challenges by using alternative statistical approaches and a new software tool with improved quality control and comprehensive integration of differential expression analysis; (2) To explore the possible reasons for not being able to validate the preselected miRNAs fully.

4.1 Study design

The HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA) was used to measure the expression of 2,083 human miRNA transcripts [plus 13 housekeeping (HK) genes] from EV fractions isolated from plasma samples of 60 cases and 60 controls by differential centrifugation (also called differential velocity centrifugation) in two batches (Table 4.1, Fig. 4.1).

Cases and controls were selected from the Liverpool Lung Project and matched on age and sex. Ages ranged from 45.9 to 88.6 (mean = 67.5) and 65% were male, all were smokers (mean smoking duration was 42.9 years for cases and 38.8 years for controls); further details for

those that passed quality control on reanalysis are provided below (Section 4.3). Additional clinical data became available during the project; following review, it was determined that 6 of the controls had potential lung cancer diagnosis (either confirmed or suspected). These were subsequently removed from the analysis as "questionable controls".

Table 4.1. Clinical characteristics of the original selected sample set of 60 matched pairs by case/control status and HTG analysis batch

		Status		**Batch**		
Group		Case	Control	1	2	Total
Total	***n***	60	60	40	80	120
Sex	*Female*	21	21	18	24	42
	Male	39	39	22	56	78
Age	*Mean (SD)*	67.5 (8.6)	67.5 (8.7)	66.1 (10.7)	68.2 (7.3)	
COPD_YN	*COPD*	30	30	15	45	60
	No COPD	28	30	24	34	58
	Unknown	2	0	1	1	2
Histology		60	0	20	40	60
	ADC	30	0	10	20	30
	SqCCa	30	0	10	20	30
Smoking Duration	*Mean (SD)*	42.9 (13.9)	38.8 (12.9)	36.1 (12.7)	43.2 (13.3)	
Time since quit	*Mean (SD)*	7.7 (12.2)	11.7 (13.2)	11.8 (13.8)	8.7 (12.3)	
Smoking status detailed	*1_Current*	25	19	12	32	44
	2_Quit<3y Duration >30y	12	6	7	11	18
	3_Quit >3y Duration >30y	12	22	10	24	34
	4_Quit >3y Duration <30y	9	13	11	11	22
	Unknown	2	0	0	2	2
Smoking status	*Current*	25	19	12	32	44
	Former	34	41	28	47	75
	Unknown	1	0	0	1	1

SD = Standard deviation

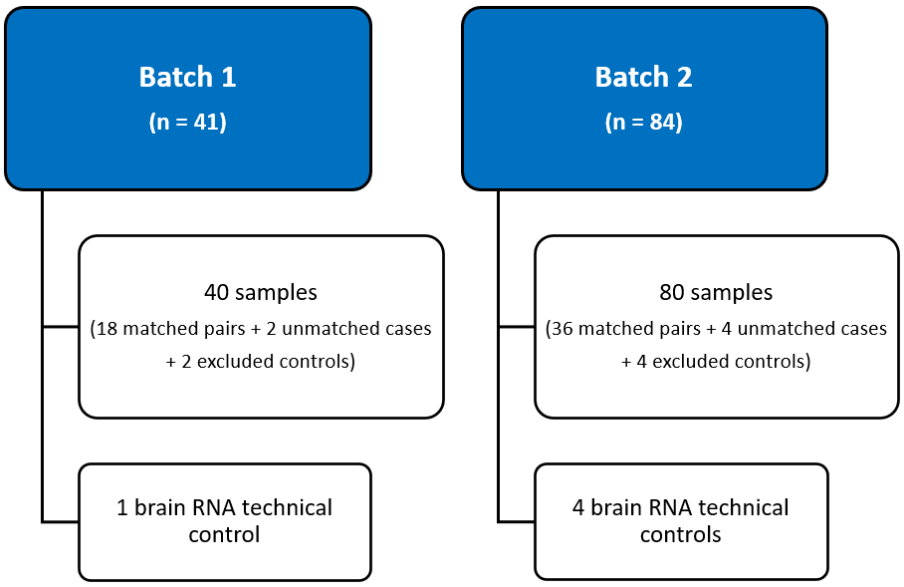


Figure 4.1. Both sets of HTG data on EV miRNA were imported. Data acquisition was performed in two batches. Batch 1 testing was conducted in June 2016, while batch 2 was in October 2016. There were 120 samples in total and 5 brain RNA technical controls. Matched pairs consisted of plasma samples from individuals of matched age and sex, with a similar smoking history. Following the clinical data update, 6 controls were excluded.

4.2 Assessing the reproducibility of the assay

In order to test how robust the HTG EdgeSeq miRNA WTA is in terms of reproducibility, we examined the brain RNA technical controls, which represent identical, high-quality samples that were run in both batches (Fig. 4.2). Brain RNA is supplied by HTG and used as a technical control, to validate assay reproducibility within and between runs, as it has a known quality and optimised quantity. This shows clearly that the assay reproducibility was very good, including between the batch 1 run (X25ng.Brain_1) and the four on the batch 2 run (Pearson correlation coefficient = 0.96 - 0.98).

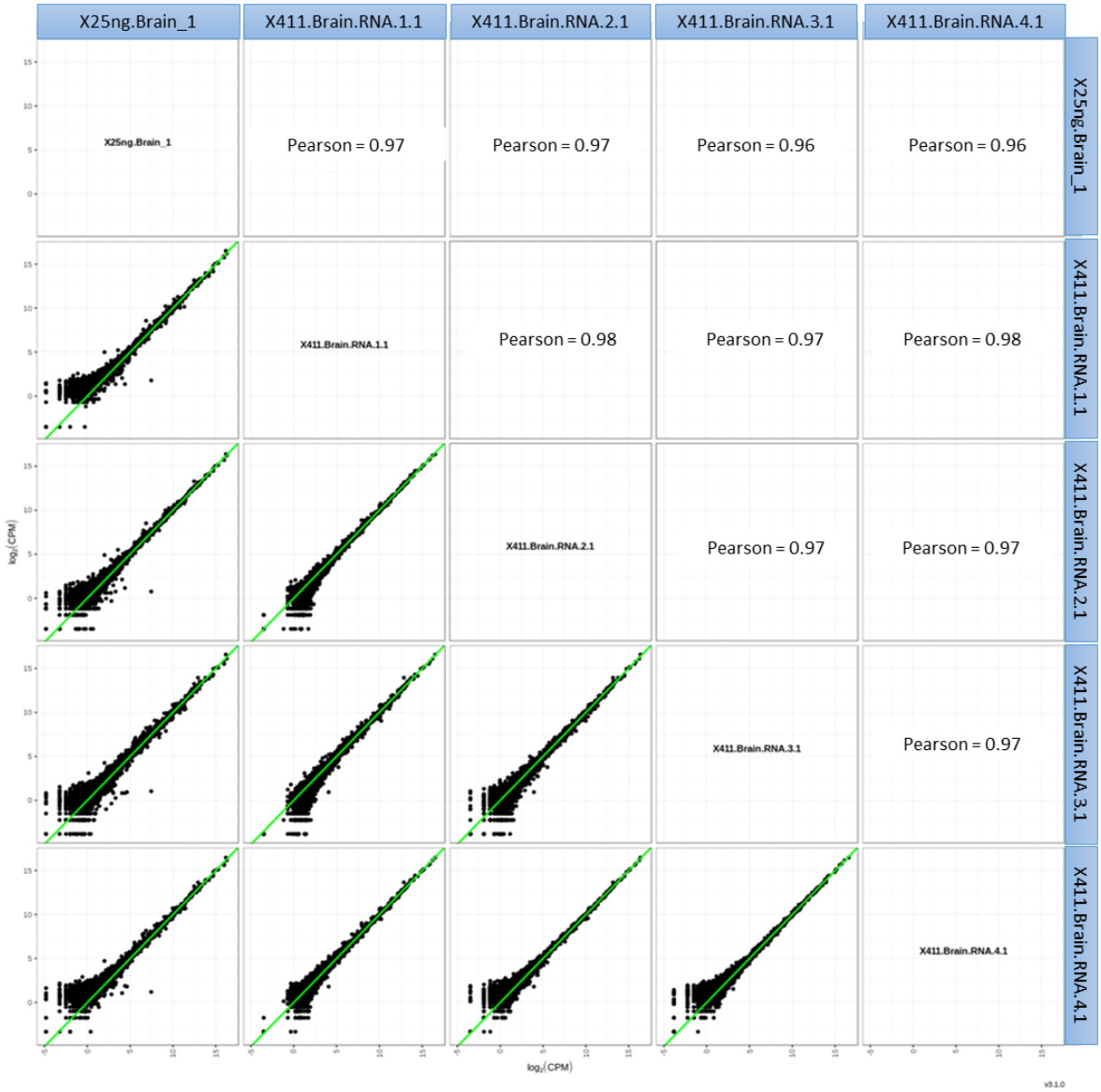


Figure 4.2. The brain RNA controls demonstrate the assay reproducibility.

4.3 Quality control (QC) metrics

The HTG EdgeSeq miRNA Whole Transcriptome Assay contains 2,102 nuclease protection probes (NPPs), including 13 housekeeping genes, one internal positive control (POS), and five internal negative controls (ANTs) (Source: A Research Use Only White Paper for HTG EdgeSeq miRNA Whole Transcriptome Assay QC Metrics) (HTG 2019).

In the original analysis, only one type of QC was applied, which was based on ANT controls; samples were considered of poor quality if the read counts for the ANT controls were high (greater than mean+2SD of 96 plasma control samples, as calculated by HTG Molecular Diagnostics during process validation). This resulted in 23 QC failures (Table 4.2), leaving 49 cases and 48 controls (Fig. 4.3).

In the reanalysis, three types of QC [QC0 = based on the percentage of the total counts that are due to the positive control; QC1 = the total count number; QC2 = the relative standard deviation (RSD) of counts per miRNA; methods section 2.2, Table. 2.2] were applied, resulting in 27 QC0 and 5 QC2 failures (Table 4.2), leaving 49 cases and 39 controls (Fig. 4.3). Hereinafter, "new QC" will be used to imply the combination of QC0, QC1 and QC2 to differentiate from the original ANT QC. Those 49 cases and 39 controls that passed the new QC are not the same as those that passed the original QC. For example, of the 88 samples that passed the new QC, 13 failed the original QC; and of the 97 samples that passed the original QC, 22 failed the new QC (Fig. 4.3).

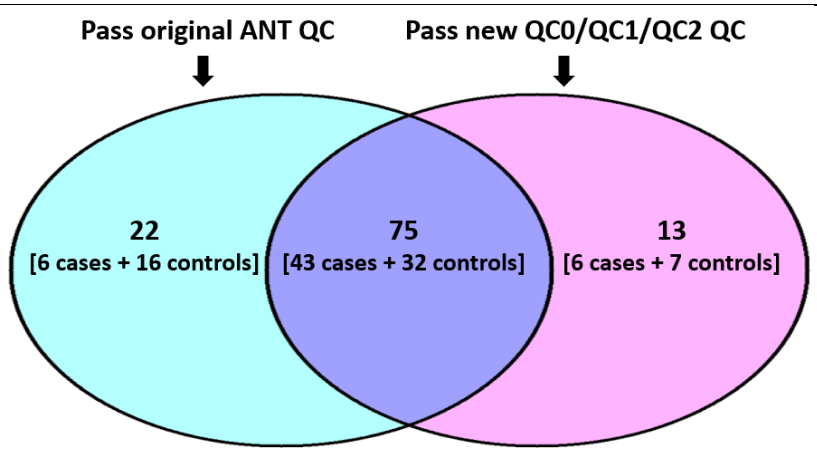


Figure 4.3. Venn diagram of quality control of HTG EV miRNA data. Samples that passed the original ANT QC, the new QC0/QC1/QC2 QC, or both.

Table 4.2. Different types of QC failures by batch

		QC Failures			
		ANT QC	QC0	QC1	QC2
Batch 1	cases	1	3	0	0
	controls	2	10	0	0
	total	3	13	0	0
Batch 2	cases	12	5	0	3
	controls	8	9	0	2
	total	20	14	0	5
Combined QC failures		23	27	0	5

In terms of quality, there were more failures in batch 2 (25% batch 2 *versus* 7.5% batch 1) using original ANT QC, but less (23.75% batch 2 *versus* 32.5% batch 1) using new QC (Table 4.2). Whilst the proportion of failures is similar using both QC measures in batch 2, it is quite different for batch 1, where the more stringent new QC identifies more issues with the data.

The principal component analysis (PCA) plot for the samples that passed the new QC shows that the samples that failed the original ANT-based QC are all clustered to the right (Fig. 4.4A). Additionally, the small group to the top left are the brain RNA controls. Note that these good quality samples cluster as far as possible from the ANT QC failures. Of the 23 original ANT QC failures, all 3 in batch 1 passed the new QC along with 10 of batch 2 (hence the 13 blue dots on the PCA plot). Furthermore, the volcano plot of samples that passed the new QC shows differential expression between the samples that passed both QCs and those that failed the original ANT QC (Fig. 4.4B).

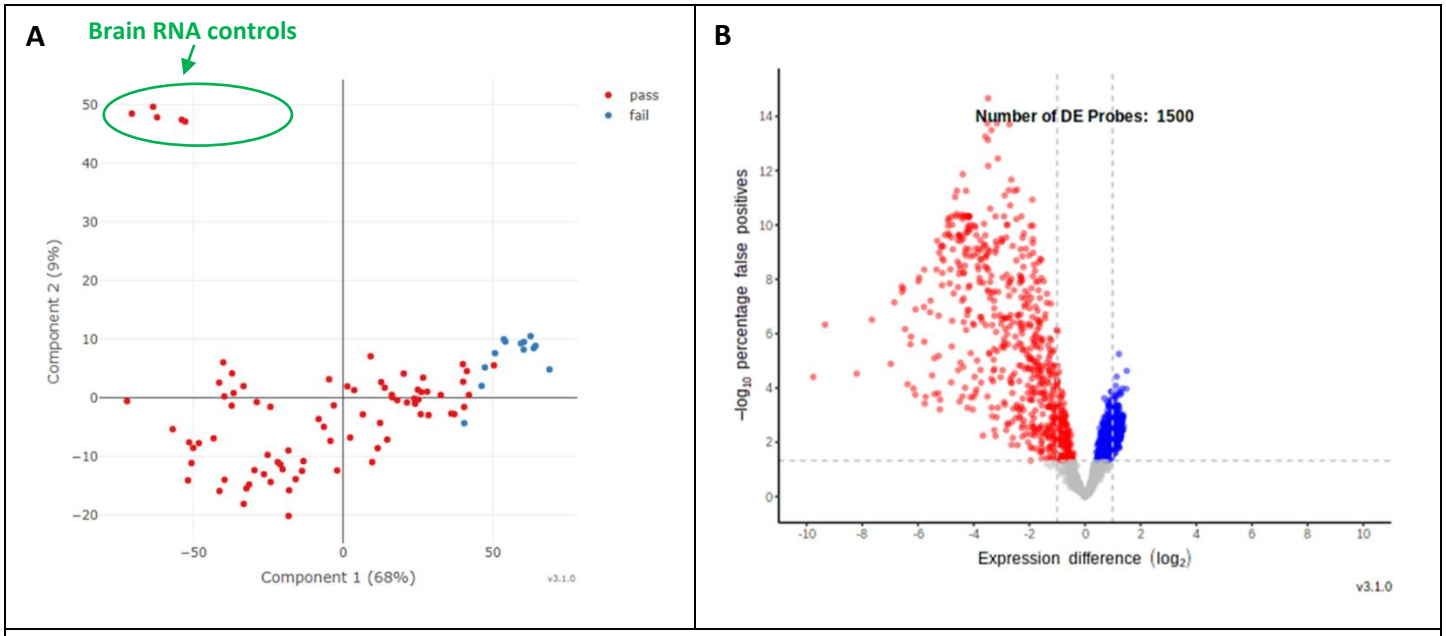


Figure 4.4. The PCA plot (A) and the volcano plot (B) for the samples that passed the new QC. This indicates differential expression between those that passed the ANT QC (pass, red in PCA plot) and those that failed the ANT QC (fail, blue in PCA plot) is a major component of PC1. There are more miRNAs with high counts when the sample pass ANT QC (red in the volcano plot) than when they fail (blue in the volcano plot).

When QC failures are included in the PCA plot, they appeared to cluster (Fig. 4.5A). ANT failures cluster to the right (positive for the first principal component, PC1), with those that also fail QC2 to the extreme right (all QC2 failures are also ANT failures). QC0 failures tend to cluster to the top left; if samples also fail for ANT (yellow) then they cluster either with QC0 (purple) or with ANT QC failures (green).

To reproduce the original EV miRNA HTG data analysis and investigate the effect of different QC metrics, we produced a PCA plot for disease status (case *vs* control) on the REVEAL analysis platform, based on the 4-miRNA panel, using only ANT-QC to exclude samples of low quality (Fig. 4.5B), as was the original analysis. Note that the separation of cases and controls is better on the left hand side and worse on the right hand side. It is interesting to note that principal component 1 seems to be related in part to sample quality, with poorer ANT-QC scores to the

right (Fig. 4.4A). Sample quality may therefore be impacting the performance of the 4-marker panel.

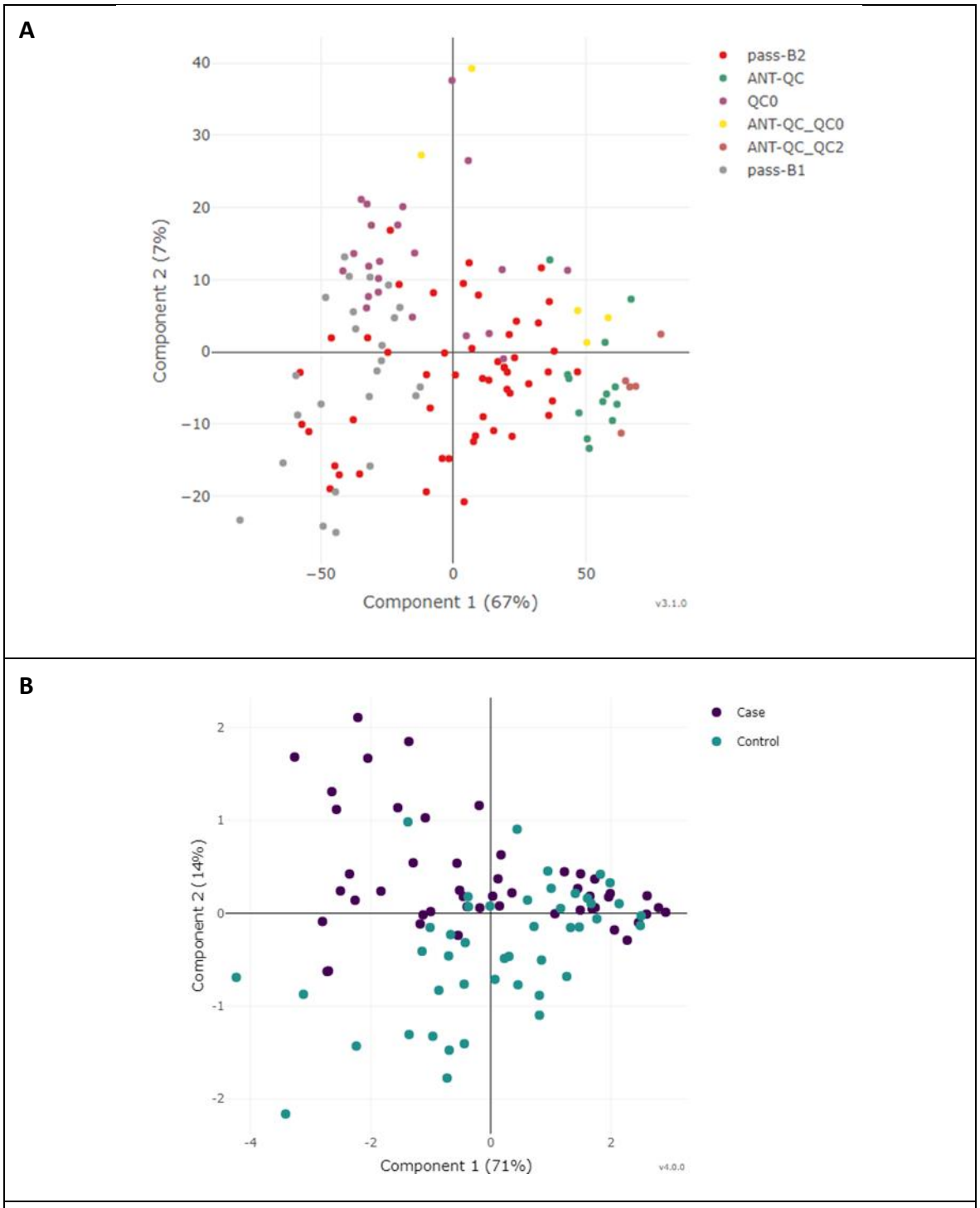


Figure 4.5. (A) PCA plot of all samples coloured by QC pass or failure. Key: pass B1 = batch 1 samples passing both QCs (n = 24); pass B2 = batch 2 samples passing both QCs (n = 51); ANT-QC = samples failing the original ANT QC (n = 13); QC0 = samples failing

QC0 (n = 22); ANT-QC_QC0 = samples failing both ANT-QC and QC0 (n = 5); ANT-QC_QC2 = samples failing both ANT-QC and QC2 (n = 5).

(B) PCA plot of EV miRNA HTG data excluding ANT-QC failures and questionable controls only and clustering based on the 4-miRNA panel.

4.4 Analysis of the batch effects

We explored the batch effects by checking the PCA plot for the data from the combined batch 1 and batch 2 that passed the new QC (without excluding the ANT QC failures, Fig. 4.6). There is clearly a batch effect, with blue batch 1 clustering differently to most of the red batch 2. This is partly related to those samples that failed the original ANT QC (clustering to the right in Fig. 4.4A) – note specifically the three batch 1 samples that cluster to the right in Fig. 4.6 are all original ANT QC failures.

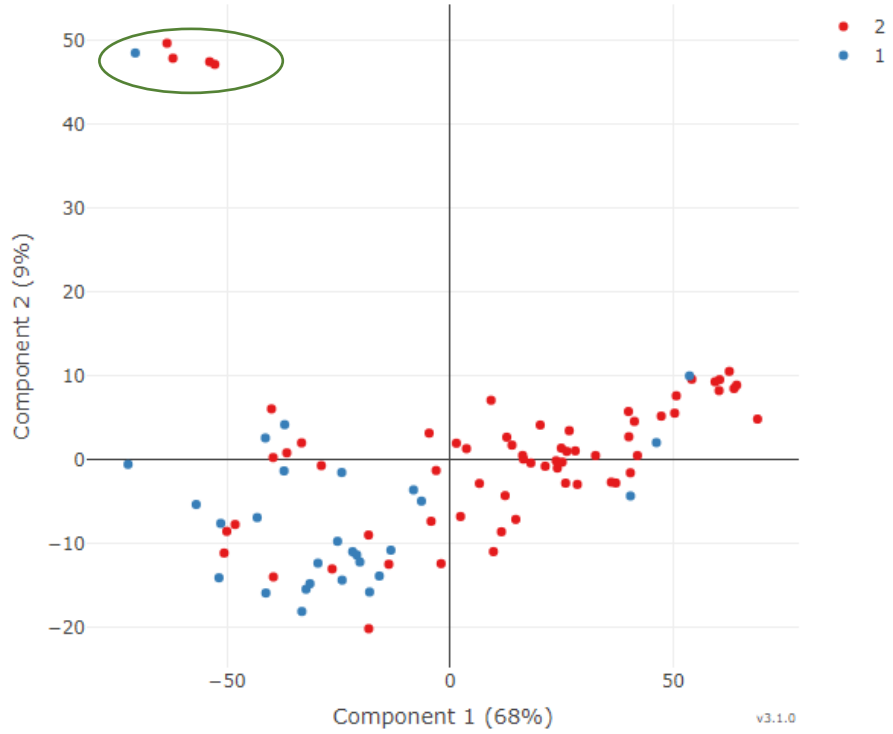


Figure 4.6. PCA plot demonstrating batch effects between batch 1 and batch 2.

We then examined the data from the combined batch 1 and batch 2 after removing the original ANT QC failures, new QC failures, brain RNA controls and additionally three control samples that were found on case note review to be potential cases during follow up (questionable controls) (Fig. 4.7). These samples are those that will be used in subsequent analysis, e.g. cases *versus* controls: 7 controls *versus* 16 cases in batch 1 and 22 controls *versus* 27 cases in batch 2.

The PCA plot (Fig. 4.7A) clearly demonstrates clustering in principle component 1 based on the batch, indicating significant variation in miRNA measurement potentially related to when or how the samples were processed. The volcano plot (Fig. 4.7B) demonstrates that the majority of the miRNAs are differentially expressed between batches 1 and 2.

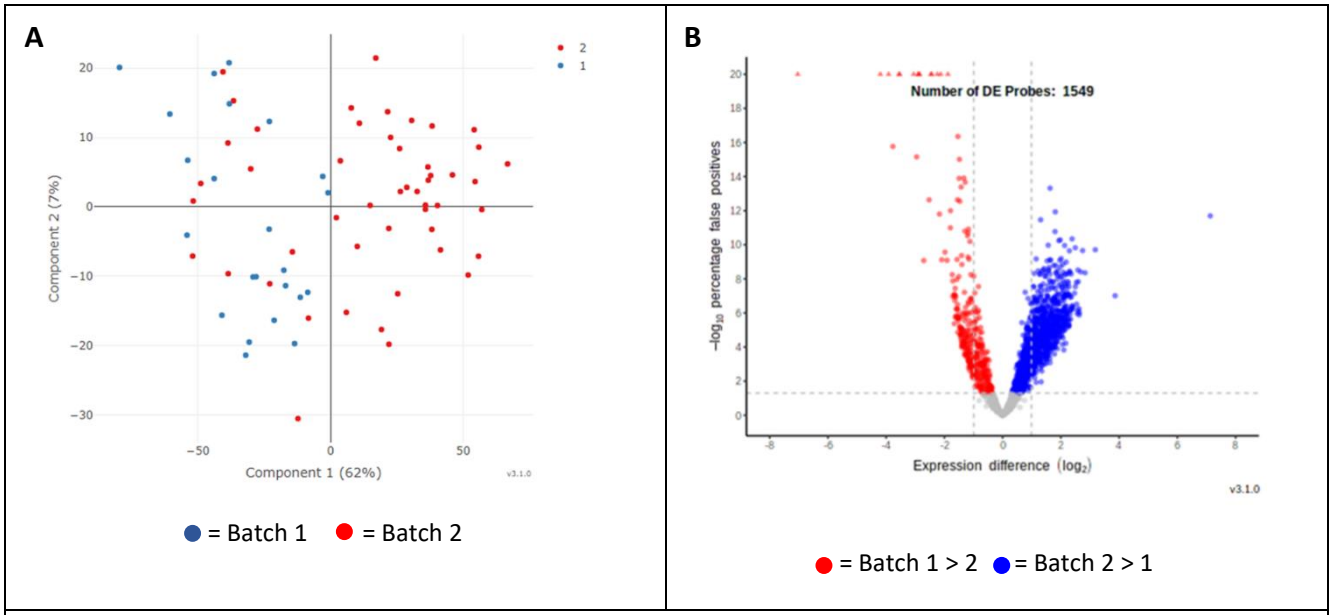


Figure 4.7. PCA plot (A) and volcano plot (B) for the data to be used in subsequent analysis, demonstrating batch effects and differential expression between the batches. Dots in Figure A are the samples, coloured by batch 1 (blue) and batch 2 (red). Dots in Figure B represent individual miRNAs, with the differentially expressed miRNAs coloured red (higher in Batch 1) or blue (higher in Batch 2).

The differences between batch 1 and batch 2 are also seen when we look at the PCA plot for miRNA expression in controls only (Fig. 4.8). This indicates that the batch effects (Fig. 4.7A) are not necessarily related to the balance of cases and controls in the two batches, or the nature of the cases in each batch.

In the original analysis of the combined data, there were differences in expression identified between cases and controls, including those miRNAs that we attempted to validate. However, it is now clear that combining the batches for analysis may not be appropriate given the batch effects. Alternatively, we could analyse the batches separately to look at EV miRNA differences between cases and controls or other parameters (e.g. smoking, COPD). An initial PCA plot of the case/control status for the combined analysis (Fig. 4.9) indicates no evidence of clustering on status, compared to that seen for batch effects (Fig. 4.7).

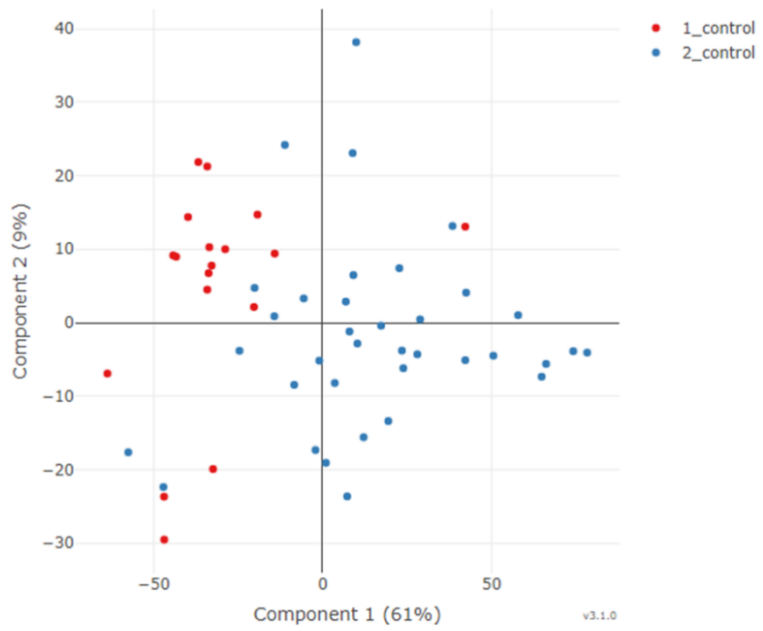


Figure 4.8. PCA plot for the data that comes from controls for batch 1 and batch 2.

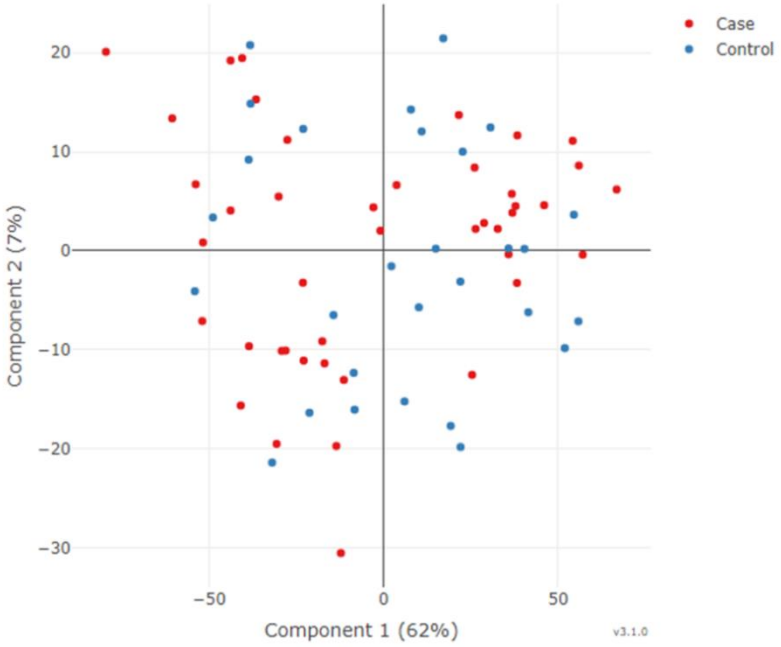


Figure 4.9. Lung cancer status analysis of the data that only passes the new QC, limiting to samples to be used for subsequent analysis.

4.5 Reanalysis of HTG data for differential expression

As significant differences between batch 1 and batch 2 were seen, we are now taking each batch separately for the analysis.

4.5.1 Differential expression of miRNAs in relation to clinical characteristics

Further reanalysis of HTG data was performed in batches 1 and 2 separately, and compared to each other or to the combined analysis, to reach a consensus on differentially expressed miRNAs. This analysis was performed only on samples that passed both QC analyses and excluded brain RNA controls and six questionable controls (i.e., selected as controls but with some subsequent evidence of cancer).

Table 4.3. Clinical characteristics of samples which passed all quality control, by HTG analysis batch

Characteristics	Batch 1 (n, %)	Batch 2 (n, %)	Total (n, %)	Chi Sq. Test
Case	16 (69.6%)	27 (55.1%)	43 (59.7%)	1.36 P = 0.24
Control	7 (30.4%)	22 (44.9%)	29 (40.3%)	
Adenocarcinoma	7 (43.8%)	14 (51.9%)	21 (48.8%)	0.26 P = 0.61
Squamous cell Carcinoma	9 (56.3%)	13 (48.1%)	22 (51.2%)	
COPD	7 (30.4%)	28 (57.1%)	35 (48.6%)	4.51 P = 0.10
No COPD	15 (65.2%)	20 (40.8%)	35 (48.6%)	
Unknown	1 (4.3%)	1 (2.0%)	2 (2.8%)	
Current Smoker	7 (30.4%)	24 (49.0%)	31 (43.1%)	2.91 P = 0.23
Former Smoker	16 (69.6%)	24 (49.0%)	40 (55.6%)	
Unknown	0	1 (2.0%)	1 (1.4%)	
Current smoker	7 (30.4%)	24 (49.0%)	31 (43.1%)	8.16 P = 0.086
Quit_under3y_Duration_over30y	6 (26.1%)	4 (8.2%)	10 (13.9%)	
Quit_over3y_Duration_over30y	4 (17.4%)	13 (26.5%)	17 (23.6%)	
Quit_over3y_Duration_under30y	6 (26.1%)	6 (12.2%)	12 (16.7%)	
Unknown	0	2 (4.1%)	2 (2.8%)	
Male	15 (65.2%)	33 (67.3%)	48 (66.7%)	0.032 P = 0.86
Female	8 (34.8%)	16 (32.7%)	24 (33.3%)	
Total	23	49	72	

For the samples that passed all quality control, batch 2 (n = 49) had a larger sample size than batch 1 (n = 23). The percentages of male and female were nearly equal in batch 1 and batch 2. The proportions of cases and SqCCa were higher in batch 1, as was the proportion of former smokers. Batch 2 had a larger proportion of the subjects with COPD. However, none of the differences in relative proportions were significant (Table 4.3).

Table 4.4. Clinical characteristics of samples which passed all quality control, by case or control status

Characteristics	**Case (n, %)**	**Control (n, %)**	**Total (n, %)**	**Chi Sq. Test**
COPD	20 (46.5%)	15 (51.7%)	35 (48.6%)	1.4 P = 0.49
No COPD	21(48.8%)	14 (48.3%)	35 (48.6%)	
Unknown	2 (4.7%)	0	2 (2.8%)	
Current Smoker	18 (41.9%)	13 (44.8%)	31 (43.1%)	0.71 P = 0.70
Former Smoker	24 (55.8%)	16 (55.2%)	40 (55.6%)	
Unknown	1 (2.3%)	0	1 (1.4%)	
Current smoker	18 (41.9%)	13 (44.8%)	31 (43.1%)	6.8 P = 0.15
Quit_under3y_Duration_over30y	9 (20.9%)	1 (3.4%)	10 (13.9%)	
Quit_over3y_Duration_over30y	8 (18.6%)	9 (31.0%)	17 (23.6%)	
Quit_over3y_Duration_under30y	6 (14.0%)	6 (20.7%)	12 (16.7%)	
Unknown	2 (4.7%)	0	2 (2.8%)	
Male	26 (60.5%)	22 (75.9%)	48 (66.7%)	1.8 P = 0.17
Female	17 (39.5%)	7 (24.1%)	24 (33.3%)	
Total	43	29	72	

For the cases and controls from batch 1 and batch 2 that passed all quality control, the ages of the subjects ranged from 45.9 to 88.6 (mean = 66.9). There was a slightly higher proportion of COPD and current smokers in controls, but not significantly so. In addition, the percentage of males was higher in controls, again not significantly (Table 4.4).

4.5.2 Case or control status

Differential expression of miRNAs between plasma exosome fractions isolated from lung cancer cases and matched controls was analysed in order to identify which exosome miRNAs were detected at significantly higher or lower levels in lung cancer patients. Analysis was performed in two batches (relating to when exosomes were isolated and HTG analysis performed). Results in each batch were compared to identify miRNAs where expression patterns were consistent across batches (using each batch as a validation for results from the other); a combined analysis (both batch 1 and 2) was also performed.

For batch 1 (Fig. 4.10), despite no clear pattern of clustering on the PCA plot (Fig. 4.10A), 17 miRNAs were significantly differentially expressed at a false discovery rate (FDR) adjusted $P < 0.05$ (Fig. 4.10B). There were 39 miRNAs where Case > Control at a raw $P < 0.01$, of which 11 had an FDR adjusted $P < 0.05$ (Table 4.5); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.10C. There were 23 miRNAs where Case < Control at a raw $P < 0.01$, of which 6 had an FDR adjusted $P < 0.05$ (Table 4.5); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.10D.

For batch 2 (Fig. 4.11), no clear pattern of clustering was seen on the PCA plot (Fig. 4.11A), 1 miRNA was significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.11B). There were 6 miRNAs where Case > Control at a raw $P < 0.01$, none of which had an FDR adjusted $P < 0.05$ (Table 4.5); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.11C. There were 33 miRNAs where Case < Control at a raw $P < 0.01$, of which 1 had an FDR adjusted $P < 0.05$ (Table 4.5); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.11D.

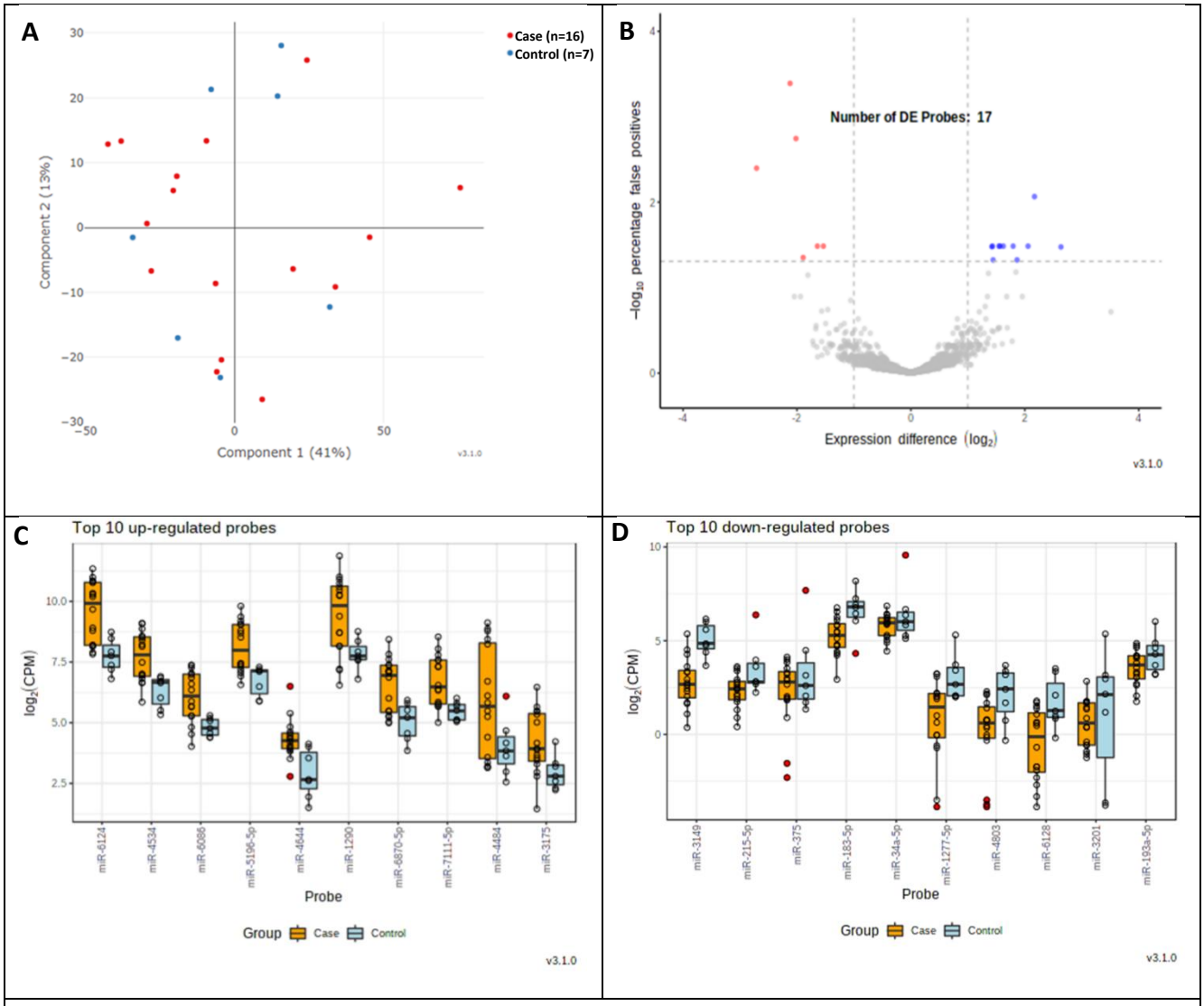


Figure 4.10. Case or control status - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for batch 1. Excluding the QC failures, questionable controls and brain RNA controls.

Table 4.5. miRNAs differentially expressed by lung cancer status (miR-3149 and miR-1277-5p are included in the validation by qRT-PCR)

	Batch 1					Batch 2					Both					
miRNA ID	**Mean normalized Case**	**Mean normalized Control**	**Fold Change Control *vs* Case**	**Raw P Control *vs* Case**	**adjP Control *vs* Case**	**Mean normalized Case**	**Mean normalized Control**	**Fold Change Control *vs* Case**	**Raw P Control *vs* Case**	**adjP Control *vs* Case**	**Mean normalized Case**	**Mean normalized Control**	**Fold Change Control *vs* Case**	**Raw P Control *vs* Case**	**adjP Control *vs* Case**	**Comparison Significant P values** Case > Control = Blue; Control > Case = Red; significant at an adjP < 0.05 in bold
miR-3149	**75**	**296**	**-4.35**	**1.94E-07**	**0.0004**	308	838	-3.47	0.0001	0.061	**194**	**639**	**-3.75**	**9.12E-08**	**0.0002**	**Control > Case B1 B2 Both**
miR-6124	**8574**	**1723**	**4.50**	**1.64E-05**	**0.0086**	11843	13190	-1.42	0.218	0.999	10205	9596	-1.07	0.769	0.951	**Case > Control B1**
miR-1290	**8351**	**1817**	**4.16**	**0.0002**	**0.0327**	10383	11299	-1.39	0.304	0.999	9295	8222	-1.01	0.983	0.997	**Case > Control B1**
miR-4534	**2103**	**619**	**3.08**	**0.0001**	**0.0327**	3290	3385	-1.31	0.283	0.999	2697	2529	-1.07	0.745	0.943	**Case > Control B1**
miR-4644	**202**	**63**	**2.92**	**0.0002**	**0.0327**	275	333	-1.54	0.047	0.829	245	244	-1.13	0.488	0.893	**Case > Control B1**
miR-5196-5p	**2735**	**836**	**2.96**	**0.0001**	**0.0327**	2074	2081	-1.28	0.292	0.999	2432	1694	1.26	0.215	0.828	**Case > Control B1**
miR-6086	**650**	**217**	**2.70**	**0.0001**	**0.0327**	793	820	-1.32	0.144	0.999	734	636	1.02	0.921	0.988	**Case > Control B1**
miR-6870-5p	**1030**	**269**	**3.47**	**0.0002**	**0.0327**	1173	1006	-1.09	0.732	0.999	1087	767	1.25	0.277	0.85	**Case > Control B1**
miR-7111-5p	**1017**	**343**	**2.68**	**0.0002**	**0.0332**	1104	1128	-1.30	0.226	0.999	1084	880	1.08	0.639	0.933	**Case > Control B1**
miR-4484	**1032**	**150**	**6.21**	**0.0002**	**0.0334**	1081	1097	-1.29	0.317	0.999	1123	778	1.27	0.319	0.873	**Case > Control B1**
miR-3175	**184**	**61**	**2.72**	**0.0004**	**0.0474**	183	148	-1.03	0.863	0.999	194	122	1.39	0.026	0.492	**Case > Control B1**
miR-4449	**2118**	**526**	**3.64**	**0.0004**	**0.0474**	1216	1597	-1.67	0.046	0.829	1650	1216	1.19	0.415	0.878	**Case > Control B1**
miR-215-5p	**44**	**160**	**-4.05**	**1.70E-06**	**0.0018**	189	161	-1.08	0.704	0.999	123	165	-1.53	0.026	0.492	**Control > Case B1**
miR-183-5p	**375**	**985**	**-2.90**	**0.0002**	**0.0327**	536	313	1.34	0.163	0.999	473	549	-1.32	0.158	0.826	**Control > Case B1**
miR-34a-5p	**493**	**1395**	**-3.12**	**0.0002**	**0.0327**	1067	655	1.28	0.336	0.999	788	882	-1.27	0.244	0.84	**Control > Case B1**
miR-375	**52**	**308**	**-6.54**	**5.74E-06**	**0.004**	179	145	-1.03	0.895	0.999	122	195	-1.82	0.0059	0.2904	**Control > Case B1 Both**
miR-1277-5p	**24**	**80**	**-3.71**	**0.0003**	**0.0445**	112	157	-1.79	0.033	0.763	70	132	-2.13	0.0013	0.1242	**Control > Case B1 Both**
miR-671-5p	409	234	1.58	0.0853	0.6439	**421**	**1078**	**-3.27**	**1.59E-05**	**0.034**	434	732	-1.92	0.0014	0.1242	**Control > Case B2 Both**
miR-6877-5p	1230	347	3.21	0.0015	0.1273	542	370	1.15	0.375	0.999	**939**	**375**	**2.20**	**6.03E-05**	**0.025**	**Case > Control B1 Both**
miR-4513	1733	137	11.41	0.0028	0.1921	309	329	-1.36	0.114	0.999	**1124**	**266**	**3.72**	**8.35E-05**	**0.027**	**Case > Control B1 Both**
miR-6750-5p	198	242	-1.35	0.2358	0.7962	395	791	-2.55	0.001	0.134	**307**	**565**	**-2.09**	**0.0002**	**0.045**	**Control > Case B2 Both**
miR-6080	36	43	-1.30	0.4286	0.8819	244	611	-3.18	0.0001	0.061	**146**	**385**	**-3.00**	**1.84E-05**	**0.018**	**Control > Case B2 Both**
miR-4519	215	180	1.08	0.6456	0.9289	454	821	-2.31	0.0001	0.061	**348**	**575**	**-1.88**	**5.59E-05**	**0.025**	**Control > Case B2 Both**
miR-4522	120	106	1.03	0.911	0.9848	285	686	-3.07	6.78E-05	0.061	**212**	**450**	**-2.42**	**2.57E-05**	**0.018**	**Control > Case B2 Both**
miR-3607-5p	99	104	-1.17	0.447	0.8862	419	837	-2.54	0.001	0.176	**268**	**578**	**-2.45**	**9.12E-05**	**0.027**	**Control > Case B2 Both**

adjP = FDR adjusted P, B1 = batch 1, B2 = batch 2

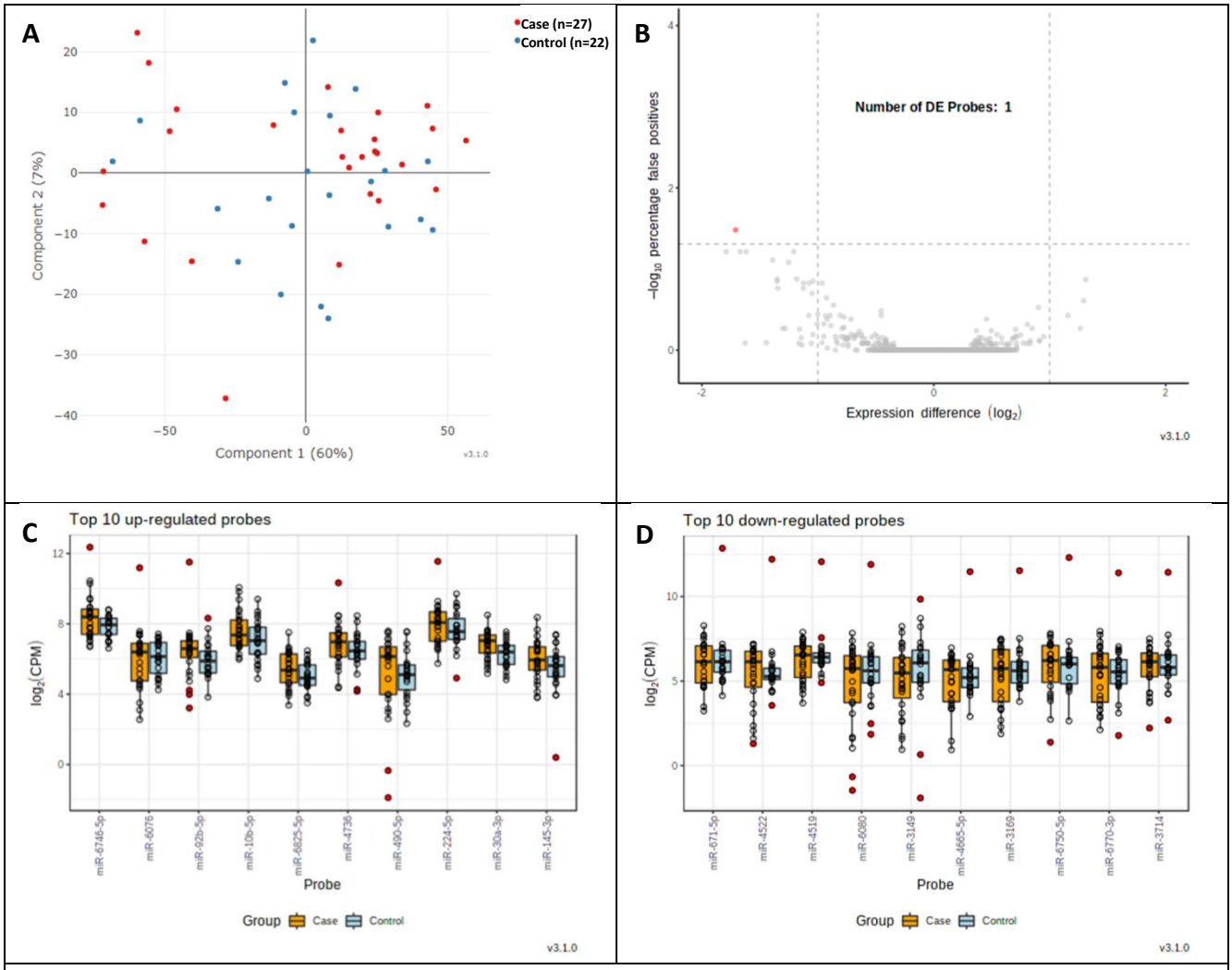


Figure 4.11. Case or control status - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for batch 2. Excluding the QC failures, questionable controls and brain RNA controls.

For combined batch 1 and batch 2 (Fig. 4.12), despite no clear pattern of clustering on the PCA plot (Fig. 4.12A), 8 miRNAs were significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.12B). There were 21 miRNAs where Case > Control at a raw $P < 0.01$, of which 2 had an FDR adjusted $P < 0.05$ (Table 4.5); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.12C. There were 37 miRNAs where Case < Control at a raw $P < 0.01$, of which 6 had an FDR adjusted $P < 0.05$ (Table 4.5); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.12D.

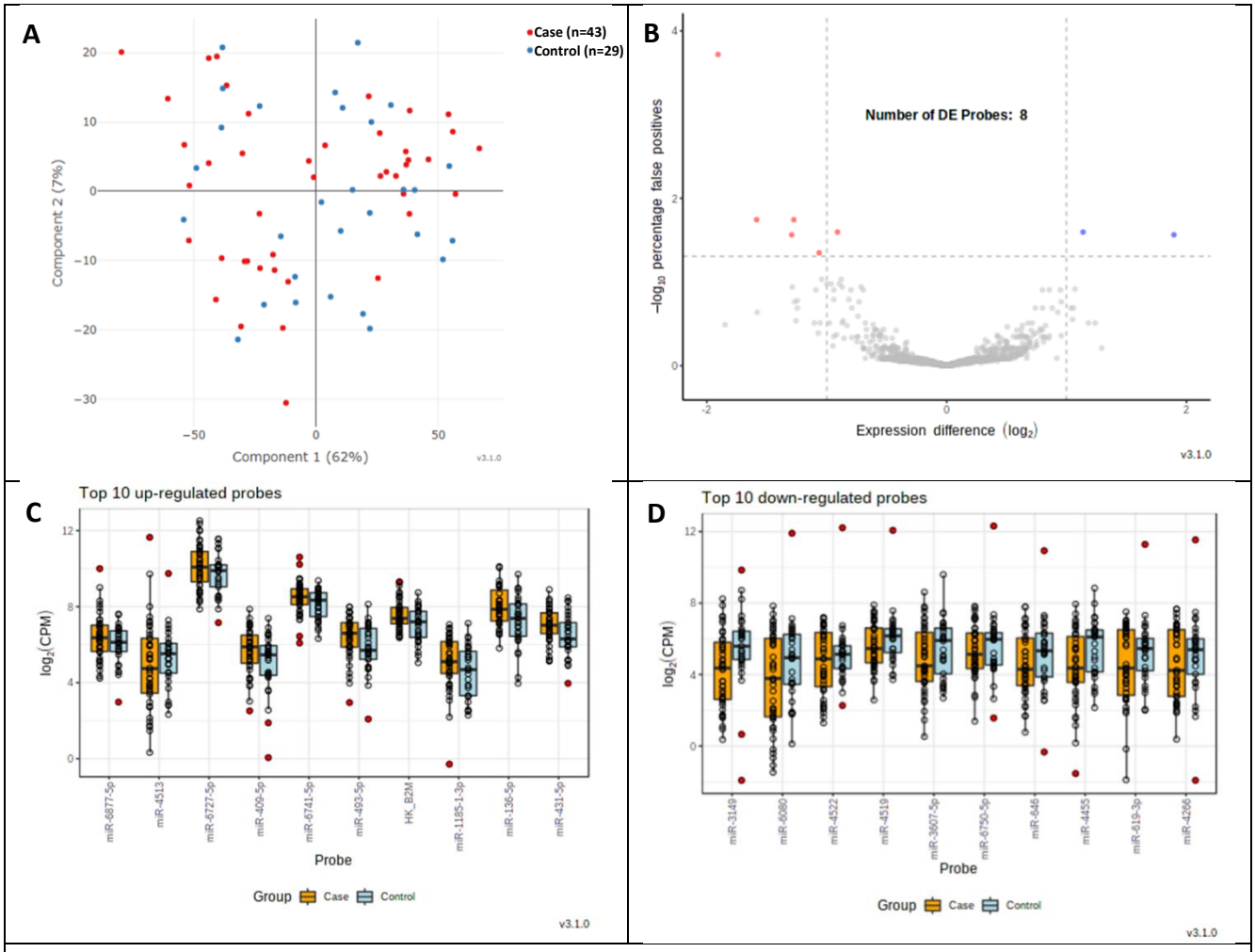


Figure 4.12. Case or control status - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for the combination of batch 1 and batch 2. Excluding the QC failures, questionable controls and brain RNA controls.

Comparison between differentially expressed miRNAs analysed by batch or combined are summarised in Table 4.5. No miRNA was consistently up-regulated in cases across all analyses at an FDR adjusted $P < 0.05$. However, 1 miRNA (miR-3149) was down-regulated (Case < Control) in batches 1, 2 and the combined analysis: raw $P = 1.9 \times 10^{-7}$ for batch 1, $P = 0.00015$ for batch 2, $P = 9.1 \times 10^{-8}$ for combined; FDR adjusted $P = 0.00041$ for batch 1, $P = 0.061$ for batch 2, $P = 0.00019$ in combined. These results indicate that there is only a limited overlap between batch 1 and batch 2. Interestingly, there were more miRNAs significantly differentially expressed in batch 1 than batch 2, despite batch 1 being smaller. Of the 17

miRNAs significant at an FDR adjusted P < 0.05 in batch 1 (16 cases *vs* 7 controls, Table 4.2), only 1 was significant when batch 2 was included in a combined analysis (43 cases *vs* 29 controls, Table 4.2), even though the number of cases and controls was higher. For batch 2, a single miRNA was significant at an FDR adjusted P < 0.05, but it did not maintain this significance in the combined analysis.

Two of the miRNAs (miR-3149 and miR-1277-5p) that were found to be differentially expressed between cases and controls in this reanalysis of the HTG data (with at least 1 FDR adjusted P < 0.05, Table 4.5) were amongst those chosen for validation by qRT-PCR in the original HTG data analysis. Whilst miR-3149 was significant in multiple analyses (see above), miR-1277-5p was only significant in batch 1 (FDR adjusted P = 0.045). Further comparison of the 18 miRNA validation set with original HTG analysis and HTG reanalysis is provided in Chapter 5.

Some miRNAs were not significantly differentially expressed in either batch, but when batches were combined for analysis, 2 miRNAs were significantly up-regulated (miR-6877-5p, FDR adjusted P = 0.025; miR-4513, FDR adjusted P = 0.027) and 5 down-regulated (miR-6750-5p, FDR adjusted P = 0.045; miR-6080, FDR adjusted P = 0.018; miR-4519, FDR adjusted P = 0.025; miR-4522, FDR adjusted P = 0.018; miR-3607-5p, FDR adjusted P = 0.027). Both the up-regulated miRNAs were significant at a raw P < 0.01 in batch 1 only and all the down-regulated miRNAs were significant at a raw P < 0.01 in batch 2 only. This indicates that significant differences in combined data, as used in previous analysis, are sometimes driven by samples only in one batch.

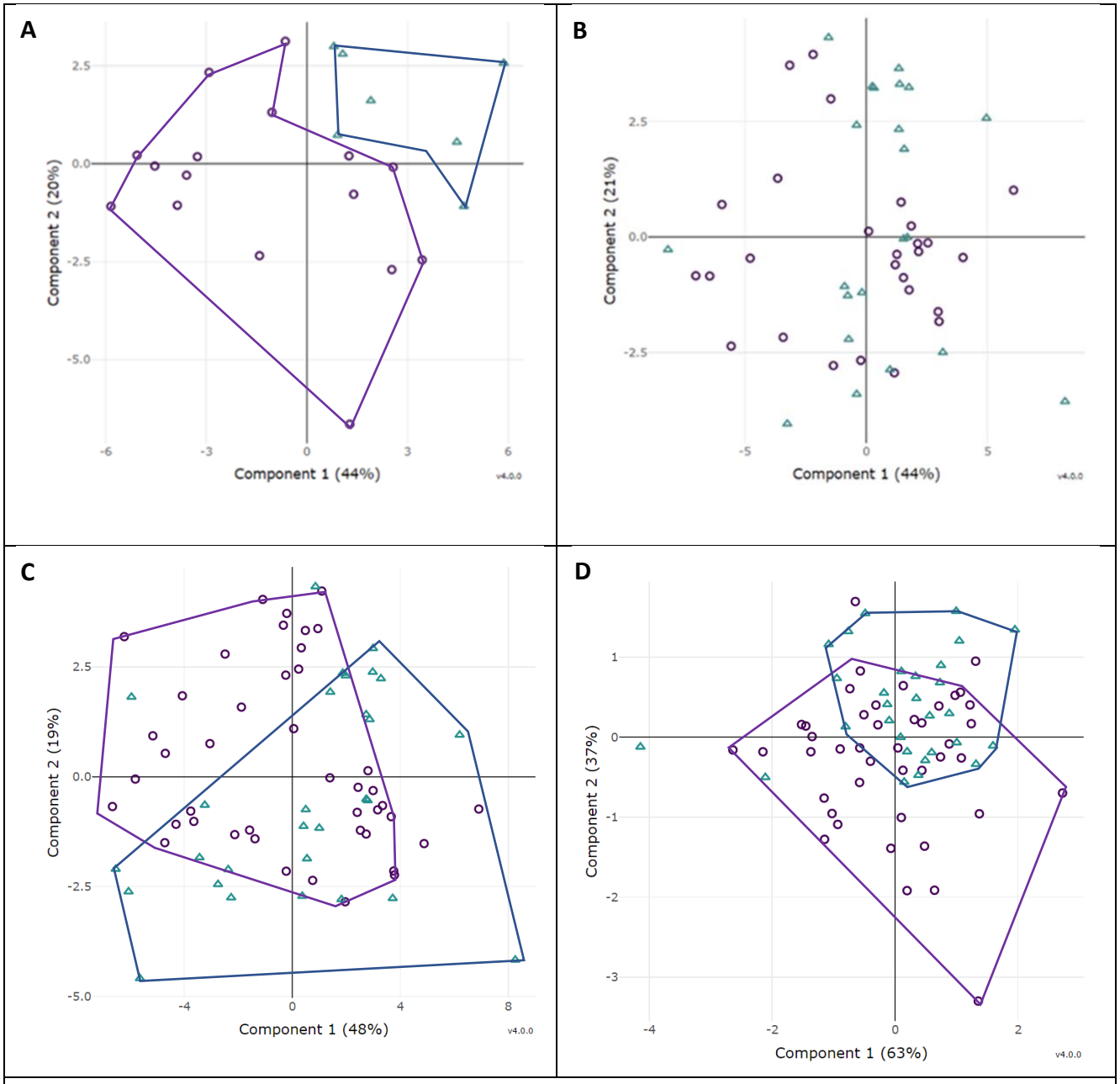


Figure 4.13. PCA plots for case or control status, based on 25 miRNAs with at least one FDR adjusted P < 0.05 (A, B, C) or the 2 most significant miRNAs in opposite directions [(miR-6877-5p (up-regulated) and miR-3149 (down-regulated)] (D), for batch 1 (A), batch 2 (B) or combined (C, D). Case = purple circle, control = teal triangle.

Note: outlines drawn round all samples for batch 1 or 2, but allowing 2 outliers for combined batches.

When we limit the miRNAs to those 25 with at least one FDR adjusted P < 0.05 (as per Table 4.5), we can see improved separation of cases and controls in batch 1 (Fig. 4.13A) and some clustering in batch 2 (Fig. 4.13B) and combined batches (Fig. 4.13C). However, we do not need

to include all 25 miRNAs to achieve some clustering in the PCA plot. If we take just the most significant up-regulated miRNA in the combined analysis (miR-6877-5p) and the most significant down-regulated (miR-3149), we get a similar pattern of clustering (Fig. 4.13D).

4.5.3 COPD

Chronic Obstructive Pulmonary Disease (COPD) is a common lung comorbidity in lung cancer patients. COPD is associated with many lung cancer risk factors, such as previous infectious disease and smoking history. It is therefore a potential confounding factor when trying to identify lung cancer biomarkers. It is also a major cause of morbidity and mortality, so it would be useful to identify new biomarkers specific for COPD. We therefore analysed the miRNA data in the same way (in 2 batches and combined) in order to identify any miRNA associated with COPD. These may be useful markers for COPD, including to help differentiate between lung cancer and COPD. The proportion of individuals in each batch with or without COPD are shown in Table 4.3. Overall, approximately half of the subjects had COPD; the proportion was higher for batch 2 (57%) than batch 1 (30%), but not significantly so ($P = 0.10$).

For batch 1 (Fig. 4.14), some association between PC2 and COPD status was seen in the PCA plot (Fig. 4.14A), and 50 miRNAs were significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.14B). There were 140 miRNAs where COPD > No COPD at a raw $P < 0.01$, of which 49 had an FDR adjusted $P < 0.05$ (Table 4.6); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.14C. There were 36 miRNAs where COPD < No COPD at a raw $P < 0.01$, of which 1 had an FDR adjusted $P < 0.05$ (Table 4.6); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.14D.

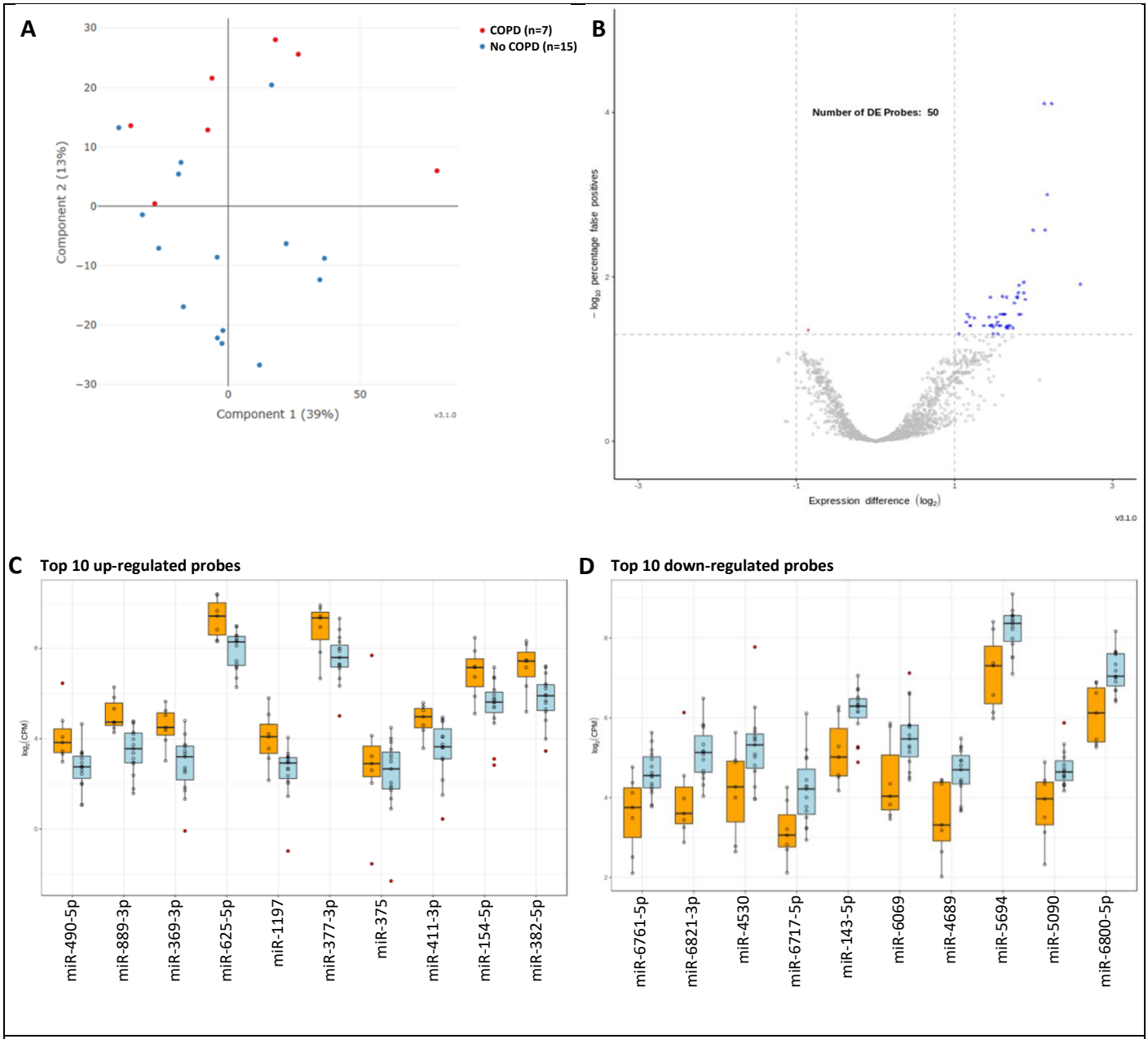


Figure 4.14. COPD - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for batch 1. Excluding the QC failures, questionable controls and brain RNA controls. Boxplots: orange = COPD, blue = no COPD.

Table 4.6. miRNA differentially expressed by COPD

	Batch 1					Batch 2					Both					
miRNA ID	Mean normalized noCOPD	Mean normalized COPD	Fold Change COPD *vs* noCOPD	Raw P COPD *vs* noCOPD	adjP COPD *vs* noCOPD	Mean normalized noCOPD	Mean normalized COPD	Fold Change COPD *vs* noCOPD	Raw P COPD *vs* noCOPD	adjP COPD *vs* noCOPD	Mean normalized noCOPD	Mean normalized COPD	Fold Change COPD *vs* noCOPD	Raw P COPD *vs* noCOPD	adjP COPD *vs* noCOPD	**Comparison Significant P values** COPD > noCOPD = Blue; noCOPD > COPD = Red; significant at an adjusted P < 0.05 in bold
let-7f-2-3p	**44**	**110**	**2.30**	**0.0007**	**0.0392**	166	185	1.10	0.6356	0.9985	104	167	1.37	0.0644	0.5111	**COPD > noCOPD B1**
miR-1185-5p	**86**	**254**	**2.71**	**0.0004**	**0.0308**	255	219	-1.18	0.4278	0.9985	166	224	1.15	0.408	0.7363	**COPD > noCOPD B1**
miR-1197	**53**	**231**	**3.99**	**6.44E-06**	**0.0027**	203	239	1.17	0.4674	0.9985	125	234	1.60	0.0105	0.2318	**COPD > noCOPD B1**
miR-134-5p	**341**	**1038**	**2.81**	**0.0008**	**0.0406**	615	537	-1.16	0.4832	0.9985	477	638	1.14	0.4552	0.7417	**COPD > noCOPD B1**
miR-136-3p	**217**	**703**	**3.00**	**0.0003**	**0.0285**	484	406	-1.20	0.3944	0.9985	344	467	1.16	0.4164	0.7401	**COPD > noCOPD B1**
miR-154-5p	**434**	**1728**	**3.67**	**6.94E-05**	**0.0156**	790	695	-1.15	0.5734	0.9985	606	923	1.30	0.2019	0.6508	**COPD > noCOPD B1**
miR-200b-3p	**140**	**317**	**2.08**	**0.0012**	**0.0496**	300	257	-1.18	0.4017	0.9985	220	274	1.06	0.6862	0.8663	**COPD > noCOPD B1**
miR-21-5p	**135533**	**413357**	**2.81**	**0.0012**	**0.0495**	139479	103528	-1.36	0.3109	0.9985	143508	177623	1.06	0.828	0.9283	**COPD > noCOPD B1**
miR-224-5p	**1169**	**4527**	**3.57**	**0.0003**	**0.0285**	2374	3734	1.56	0.2061	0.9985	1807	3527	1.67	0.0475	0.4653	**COPD > noCOPD B1**
miR-296-5p	**158**	**468**	**2.74**	**0.0001**	**0.0177**	203	208	1.01	0.9425	0.9985	185	268	1.24	0.2092	0.657	**COPD > noCOPD B1**
miR-299-3p	**131**	**316**	**2.22**	**0.0005**	**0.0355**	304	282	-1.09	0.5242	0.9985	216	288	1.14	0.2585	0.6884	**COPD > noCOPD B1**
miR-323a-3p	**186**	**639**	**3.16**	**0.0001**	**0.0177**	652	438	-1.50	0.0704	0.9549	409	473	-1.01	0.9524	0.9838	**COPD > noCOPD B1**
miR-323b-3p	**320**	**1013**	**2.92**	**0.0005**	**0.0355**	637	526	-1.22	0.3054	0.9985	475	643	1.16	0.4002	0.7313	**COPD > noCOPD B1**
miR-329-3p	**417**	**1469**	**3.25**	**0.0007**	**0.0392**	1001	715	-1.41	0.1614	0.9985	696	886	1.09	0.6910	0.8668	**COPD > noCOPD B1**
miR-335-5p	**7002**	**21893**	**2.89**	**0.0006**	**0.0385**	7990	7138	-1.13	0.6798	0.9985	7771	10653	1.17	0.516	0.7732	**COPD > noCOPD B1**
miR-342-5p	**156**	**386**	**2.28**	**0.0004**	**0.0308**	248	194	-1.29	0.1522	0.9985	206	245	1.02	0.9129	0.9623	**COPD > noCOPD B1**
miR-369-3p	**68**	**333**	**4.51**	**1.48E-06**	**0.001**	238	235	-1.02	0.9014	0.9985	149	259	1.49	0.0263	0.3693	**COPD > noCOPD B1**
miR-369-5p	**181**	**609**	**3.09**	**0.0003**	**0.0285**	391	312	-1.27	0.2495	0.9985	282	377	1.14	0.4577	0.7417	**COPD > noCOPD B1**
miR-375	**53**	**346**	**6.02**	**4.12E-05**	**0.0123**	157	167	1.05	0.8243	0.9985	105	214	1.73	0.0141	0.259	**COPD > noCOPD B1**
miR-376a-3p	**2755**	**9392**	**3.15**	**0.0008**	**0.0406**	4467	3346	-1.35	0.3147	0.9985	3557	4728	1.14	0.5991	0.8081	**COPD > noCOPD B1**
miR-376a-5p	**130**	**490**	**3.48**	**0.0001**	**0.0177**	365	249	-1.48	0.0673	0.9549	237	305	1.10	0.6139	0.8192	**COPD > noCOPD B1**
miR-376c-3p	**5695**	**19753**	**3.20**	**0.0009**	**0.0418**	10567	7930	-1.35	0.3426	0.9985	7956	10509	1.13	0.6279	0.8286	**COPD > noCOPD B1**
miR-377-3p	**1697**	**6757**	**3.67**	**3.32E-05**	**0.0116**	3134	2472	-1.28	0.4007	0.9985	2386	3432	1.23	0.3807	0.7227	**COPD > noCOPD B1**
miR-379-3p	**77**	**216**	**2.59**	**0.0006**	**0.0391**	234	220	-1.07	0.7483	0.9985	152	218	1.22	0.2553	0.6856	**COPD > noCOPD B1**
miR-379-5p	**325**	**1095**	**3.11**	**0.0003**	**0.0285**	881	703	-1.27	0.3076	0.9985	592	777	1.12	0.5481	0.7868	**COPD > noCOPD B1**
miR-382-3p	**197**	**584**	**2.73**	**0.0007**	**0.0392**	450	370	-1.23	0.2872	0.9985	316	410	1.11	0.5235	0.7772	**COPD > noCOPD B1**
miR-382-5p	**517**	**1960**	**3.50**	**7.45E-05**	**0.0156**	1102	956	-1.16	0.5357	0.9985	798	1180	1.26	0.2463	0.6812	**COPD > noCOPD B1**
miR-409-3p	**888**	**3043**	**3.16**	**0.0009**	**0.0418**	1854	1525	-1.23	0.4471	0.9985	1358	1830	1.15	0.5222	0.7768	**COPD > noCOPD B1**
miR-411-3p	**106**	**406**	**3.52**	**4.83E-05**	**0.0127**	351	301	-1.18	0.4025	0.9985	221	320	1.24	0.2184	0.6667	**COPD > noCOPD B1**
miR-4306	**6544**	**19692**	**2.78**	**0.0006**	**0.0385**	7250	4097	-1.79	0.0331	0.77	7311	8016	-1.07	0.7877	0.9186	**COPD > noCOPD B1**
miR-431-5p	**640**	**2033**	**2.93**	**0.0011**	**0.0495**	1224	963	-1.28	0.3403	0.9985	909	1181	1.11	0.624	0.8247	**COPD > noCOPD B1**
miR-432-5p	**462**	**1763**	**3.52**	**0.0003**	**0.0286**	990	713	-1.40	0.1795	0.9985	721	947	1.12	0.5956	0.808	**COPD > noCOPD B1**
miR-433-3p	**184**	**547**	**2.75**	**0.0008**	**0.0392**	438	331	-1.34	0.1477	0.9985	308	378	1.05	0.7865	0.9186	**COPD > noCOPD B1**

miR-4449	1001	3209	2.96	0.0007	0.0392	1699	1193	-1.44	0.1684	0.9985	1334	1657	1.06	0.7885	0.9186	COPD > noCOPD B1
miR-4800-3p	1054	2631	2.30	0.0008	0.0392	1255	1192	-1.06	0.7022	0.9985	1196	1571	1.12	0.4521	0.7417	COPD > noCOPD B1
miR-490-5p	52	248	4.40	4.37E-08	0.0001	303	273	-1.12	0.6267	0.9985	178	267	1.28	0.2242	0.6667	COPD > noCOPD B1
miR-493-5p	401	1389	3.20	0.0007	0.0392	756	617	-1.24	0.3807	0.9985	572	783	1.17	0.4448	0.7417	COPD > noCOPD B1
miR-494-3p	983	3199	3.00	0.0008	0.0392	1724	1482	-1.18	0.567	0.9985	1336	1848	1.18	0.4563	0.7417	COPD > noCOPD B1
miR-496	152	568	3.45	0.0001	0.0177	350	315	-1.12	0.5626	0.9985	248	369	1.27	0.1739	0.6273	COPD > noCOPD B1
miR-539-5p	160	534	3.07	0.0003	0.0285	399	327	-1.23	0.3005	0.9985	278	372	1.14	0.4506	0.7417	COPD > noCOPD B1
miR-548e-3p	67	173	2.38	0.0004	0.0316	200	211	1.04	0.7942	0.9985	134	203	1.29	0.0749	0.5337	COPD > noCOPD B1
miR-589-5p	154	375	2.24	0.0003	0.0285	317	243	-1.32	0.0307	0.7402	240	276	-1.02	0.8852	0.9513	COPD > noCOPD B1
miR-625-5p	2195	10526	4.42	5.58E-06	0.0027	2701	2685	-1.02	0.9522	0.9985	2540	4580	1.54	0.071	0.5223	COPD > noCOPD B1
miR-758-3p	175	581	3.05	9.14E-05	0.0174	384	346	-1.12	0.5138	0.9985	276	395	1.22	0.1921	0.6438	COPD > noCOPD B1
miR-215-5p	47	172	3.38	0.0002	0.0208	135	211	1.55	0.0476	0.8745	91	204	1.91	0.001	0.063	COPD > noCOPD B1 Both
miR-889-3p	96	488	4.69	7.45E-08	0.0001	248	283	1.13	0.5626	0.9985	167	331	1.69	0.003	0.114	COPD > noCOPD B1 Both
miR-6076	164	132	-1.34	0.2325	0.6046	339	1067	3.12	0.0003	0.1874	257	832	2.76	1.86E-05	0.0089	COPD > noCOPD B2 Both
miR-6761-5p	189	113	-1.80	0.001	0.0444	265	295	1.10	0.5698	0.9985	236	250	-1.10	0.4477	0.7414	noCOPD > COPD B1
miR-1247-3p	3394	2046	-1.80	0.009	0.1138	284	256	-1.12	0.3751	0.9985	2092	744	-3.29	2.79E-05	0.0089	noCOPD >COPD B1 Both
miR-150-5p	28790	13375	-2.33	0.0051	0.099	29687	15512	-1.93	0.0147	0.619	30974	15109	-2.40	3.33E-05	0.0089	noCOPD >COPD B1 Both
miR-2116-5p	739	438	-1.83	0.0056	0.1012	440	418	-1.06	0.6074	0.9985	636	430	-1.73	2.48E-05	0.0089	noCOPD >COPD B1 Both
miR-5587-3p	1223	713	-1.86	0.0051	0.099	203	266	1.29	0.0257	0.7279	802	391	-2.40	3.82E-05	0.0089	noCOPD >COPD B1 Both
miR-5694	2347	1351	-1.88	0.0038	0.091	436	454	1.03	0.7921	0.9985	1563	704	-2.60	6.02E-06	0.0089	noCOPD >COPD B1 Both
miR-3197	22572	91116	3.72	0.0001	0.0188	63819	30206	-2.13	0.0053	0.4715	42656	45985	-1.09	0.7263	0.891	COPD > noCOPD B1
miR-6798-5p	470	1700	3.34	0.0009	0.0418	708	392	-1.82	0.0034	0.402	578	717	1.06	0.7849	0.9176	COPD > noCOPD B1
miR-9-3p	78	243	2.88	0.0004	0.0308	1401	502	-2.82	0.008	0.481	717	445	-1.89	0.0506	0.4747	COPD > noCOPD B1
miR-1233-3p	1288	888	-1.57	0.0399	0.2295	506	525	1.03	0.8405	0.9985	984	626	-1.84	0.0002	0.0267	noCOPD >COPD Both
miR-144-3p	69794	40142	-1.88	0.0369	0.224	20308	14219	-1.44	0.2046	0.9985	50742	20854	-2.85	8.86E-05	0.0155	noCOPD >COPD Both
miR-144-5p	20798	12614	-1.79	0.0408	0.2318	7438	5994	-1.25	0.3851	0.9985	15686	7620	-2.41	0.0002	0.0267	noCOPD >COPD Both
miR-149-3p	50050	27634	-1.96	0.0381	0.225	25529	16064	-1.61	0.044	0.8392	40460	19512	-2.43	3.38E-05	0.0089	noCOPD >COPD Both
miR-3912-5p	1316	828	-1.72	0.0151	0.145	756	775	1.02	0.9061	0.9985	1119	800	-1.64	0.0003	0.0351	noCOPD >COPD Both
miR-541-3p	969	638	-1.65	0.0242	0.1778	182	203	1.10	0.4945	0.9985	644	324	-2.33	0.0001	0.0173	noCOPD >COPD Both
miR-6819-5p	1575	1031	-1.66	0.0174	0.1503	423	423	-1.01	0.928	0.9985	1112	598	-2.18	5.17E-05	0.0108	noCOPD >COPD Both
miR-6825-3p	3389	2284	-1.61	0.1044	0.4104	367	486	1.31	0.1131	0.9985	2138	953	-2.63	0.0003	0.0351	noCOPD >COPD Both
miR-6852-3p	822	523	-1.70	0.054	0.2735	226	218	-1.05	0.7368	0.9985	583	298	-2.29	3.13E-05	0.0089	noCOPD >COPD Both
miR-764	2075	1257	-1.79	0.0117	0.133	268	242	-1.12	0.4171	0.9985	1326	520	-2.99	1.72E-05	0.0089	noCOPD >COPD Both
miR-937-3p	875	564	-1.68	0.0162	0.1472	287	295	1.02	0.8799	0.9985	641	373	-2.01	6.18E-05	0.0118	noCOPD >COPD Both

adjP = FDR adjusted P, B1 = batch 1, B2 = batch 2

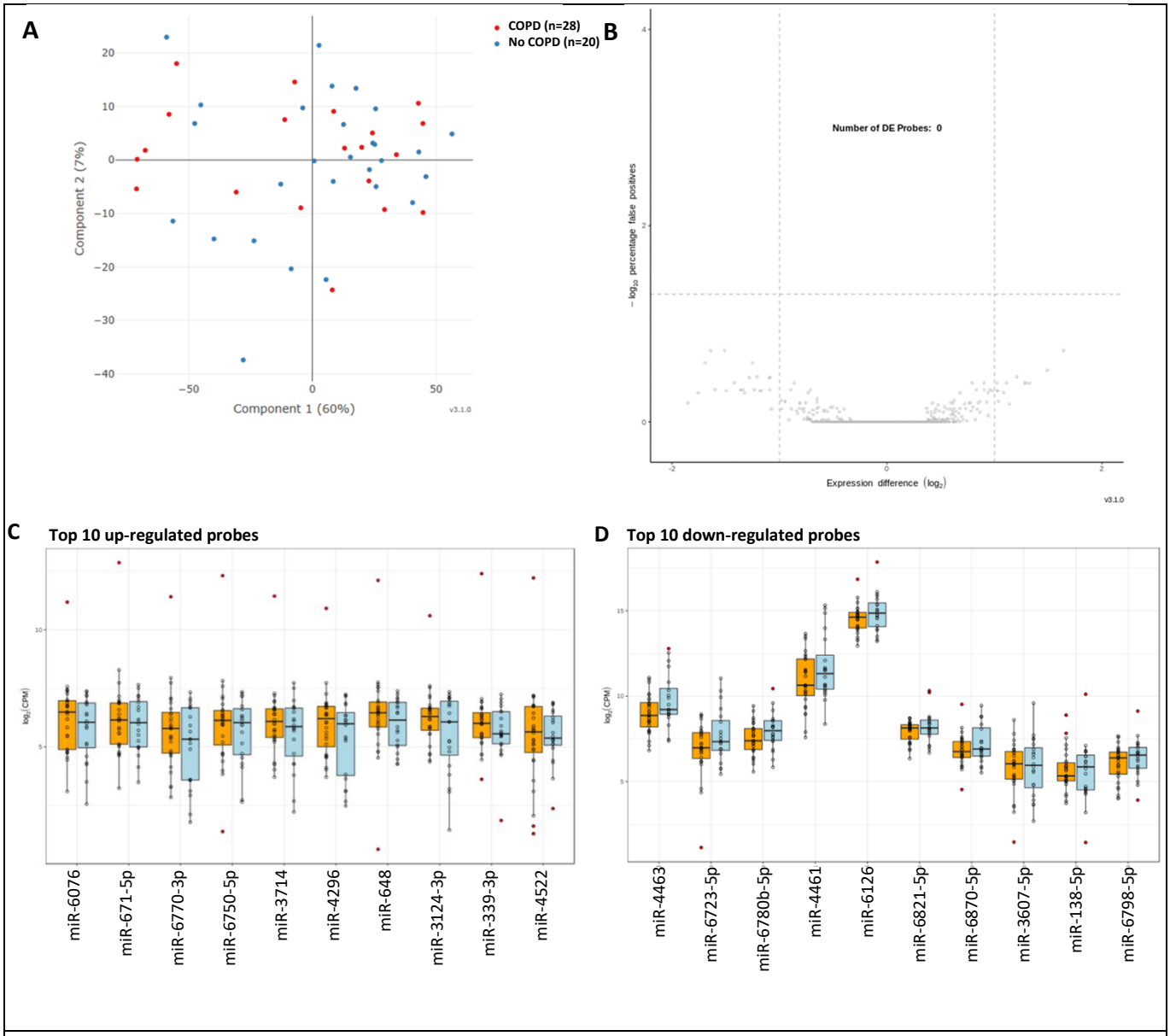


Figure 4.15. COPD - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for batch 2. Excluding the QC failures, questionable controls and brain RNA controls.

Boxplots: orange = COPD, blue = no COPD.

For batch 2 (Fig. 4.15), no clear pattern of clustering was seen on the PCA plot (Fig. 4.15A), no miRNA was significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.15B). There were 17 miRNAs where COPD > No COPD at a raw $P < 0.01$ (Table 4.6); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.15C. There were 24 miRNAs where COPD < No

COPD at a raw P < 0.01 (Table 4.6); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.15D.

For combined batch 1 and batch 2 (Fig. 4.16), despite no clear pattern of clustering on the PCA plot (Fig. 4.16A), 17 miRNAs were significantly differentially expressed at an FDR adjusted P < 0.05 (Fig. 4.16B). There were 24 miRNAs where COPD > No COPD at a raw P < 0.01, of which 1 had an FDR adjusted P < 0.05 (Table 4.6); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.16C. There were 67 miRNAs where COPD < No COPD at a raw P < 0.01, of which 16 had an FDR adjusted P < 0.05 (Table 4.6); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.16D.

Comparison between differentially expressed miRNAs analysed by batch or combined are summarised in Table 4.6. No miRNA was consistently up-regulated in COPD across all analyses at an FDR adjusted P < 0.05 and no miRNA down-regulated, or in the same direction for both batches 1 and 2. Additionally, no miRNA was consistently differentially expressed by COPD status across all analyses at a raw P < 0.01.

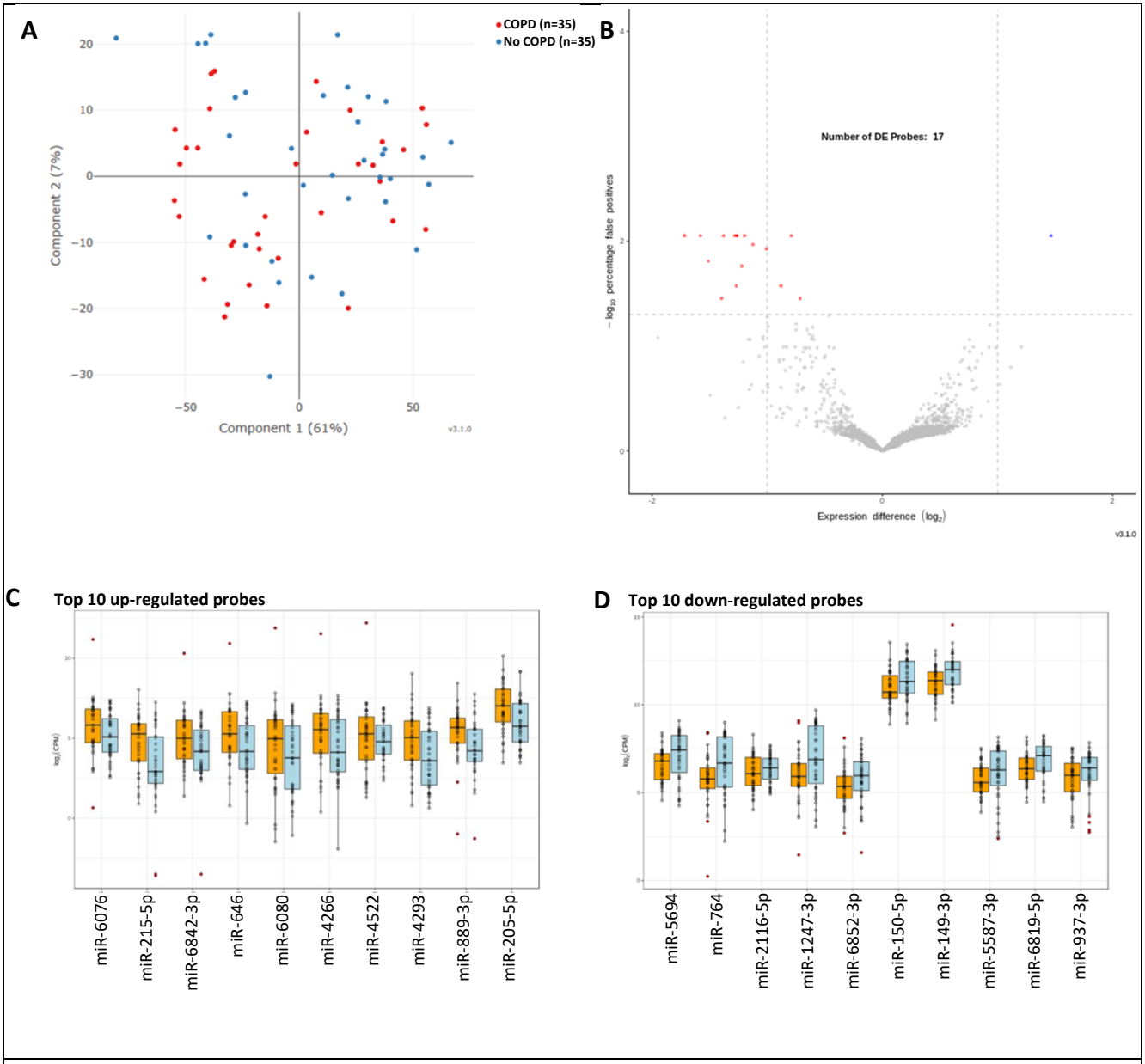


Figure 4.16. COPD - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for the combination of batch 1 and batch 2. Excluding the QC failures, questionable controls and brain RNA controls.

Boxplots: orange = COPD, blue = no COPD.

However, 4 miRNAs were up-regulated (COPD > No COPD) in both batch 1 and the combined at a raw P < 0.01; 2 of which were significantly up-regulated at an FDR adjusted P < 0.05 in batch 1. Additionally, 8 different miRNAs were up-regulated (COPD > No COPD) in both batch 2 and the combined at a raw P < 0.01, 1 of which was significantly up-regulated at an FDR

adjusted P < 0.05 in the combined. Besides that, 12 miRNAs were down-regulated (COPD < No COPD) in both batch 1 and the combined at a raw P < 0.01, 5 of which were significantly down-regulated at an FDR adjusted P < 0.05 in the combined. Similarly, 6 different miRNAs were down-regulated (COPD < No COPD) in both batch 2 and the combined at a raw P < 0.01.

Some miRNAs were not significantly differentially expressed in either batch at an FDR adjusted P < 0.05, but when batches were combined for analysis, 1 miRNA was significantly up-regulated (miR-6076, FDR adjusted P = 0.0089), this was significant with raw P = 0.00027 in batch 2, but not significant in batch 1. Also, 11 down-regulated miRNAs were significant at an FDR adjusted P < 0.05 for combined analysis only; none of these were significant in either batch, even at a raw P < 0.01. This indicates that significant differences in combined data, as used in previous analysis, are sometimes only seen when all data is combined.

When we limit the miRNAs to those with at least one FDR adjusted P < 0.05 (as per Table 4.6), we can see improved separation of COPD and No COPD in batch 1 (Fig. 4.17A), but no clear clustering in batch 2 (Fig. 4.17B). This is in keeping with the miRNAs being selected primarily due to differential expression in batch 1 or combined (Table 4.6).

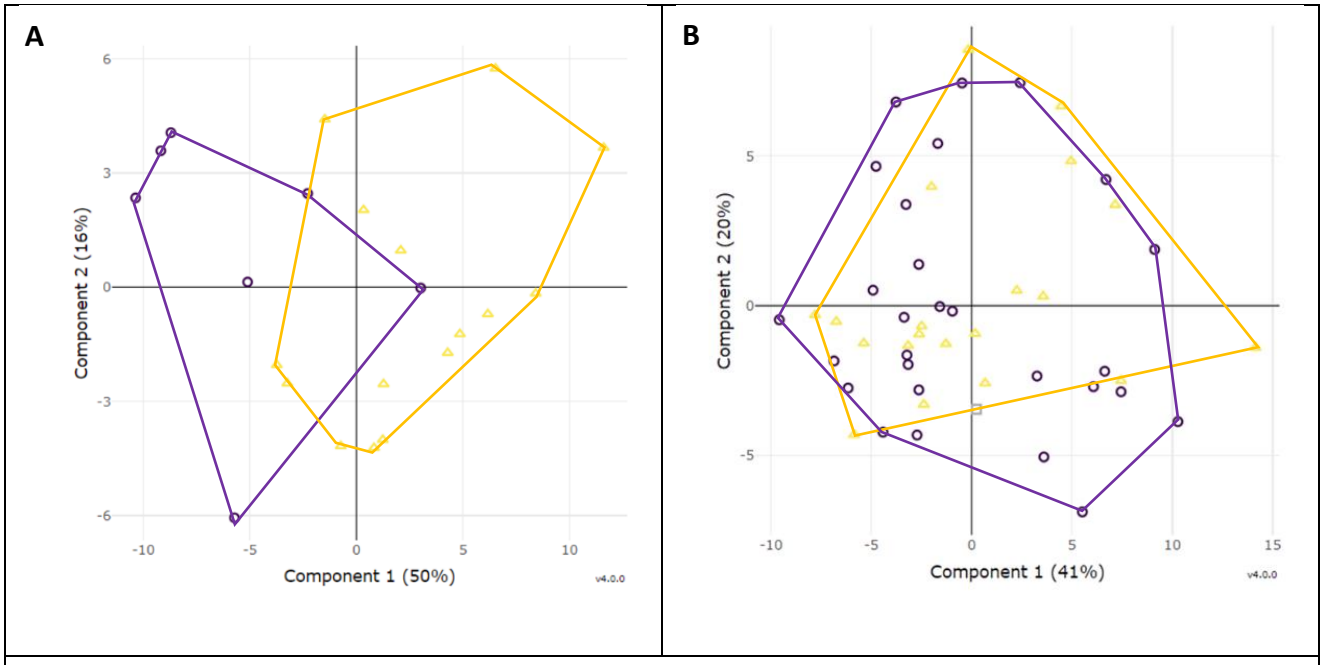


Figure 4.17. PCA plot based on 67 miRNAs with at least one FDR adjusted P < 0.05 when reanalysing HTG data for differential expression associated with COPD, batch 1 (A) and batch 2 (B). COPD = purple circle, no COPD = yellow triangle, NA = grey square.

Note: outlines drawn round all samples.

4.5.4 Sex

Lung cancer has been reported to be more common in males than females (although less so more recently due to changes in smoking habits), so might be a source of bias if cases and controls are not matched for sex. Whilst the initial samples were matched for sex, once we removed those that failed QC there was a slight imbalance (Table 4.4), with 60% male cases and 76% male controls, however this was not significant (Chi Square 1.8, P = 0.17).

For batch 1 (Fig. 4.18), despite no clear pattern of clustering on the PCA plot (Fig. 4.18A), 15 miRNAs were significantly differentially expressed at an FDR adjusted P < 0.05 (Fig. 4.18B). There were 67 miRNAs where Male > Female at a raw P < 0.01, of which 10 had an FDR adjusted P < 0.05; bar graphs for the top 10 miRNAs higher in males are shown in Fig. 4.18C.

There were 69 miRNAs where Male < Female at a raw P < 0.01, of which 5 had an FDR adjusted P < 0.05; bar graphs for the top 10 miRNAs higher in females are shown in Fig. 4.18D.

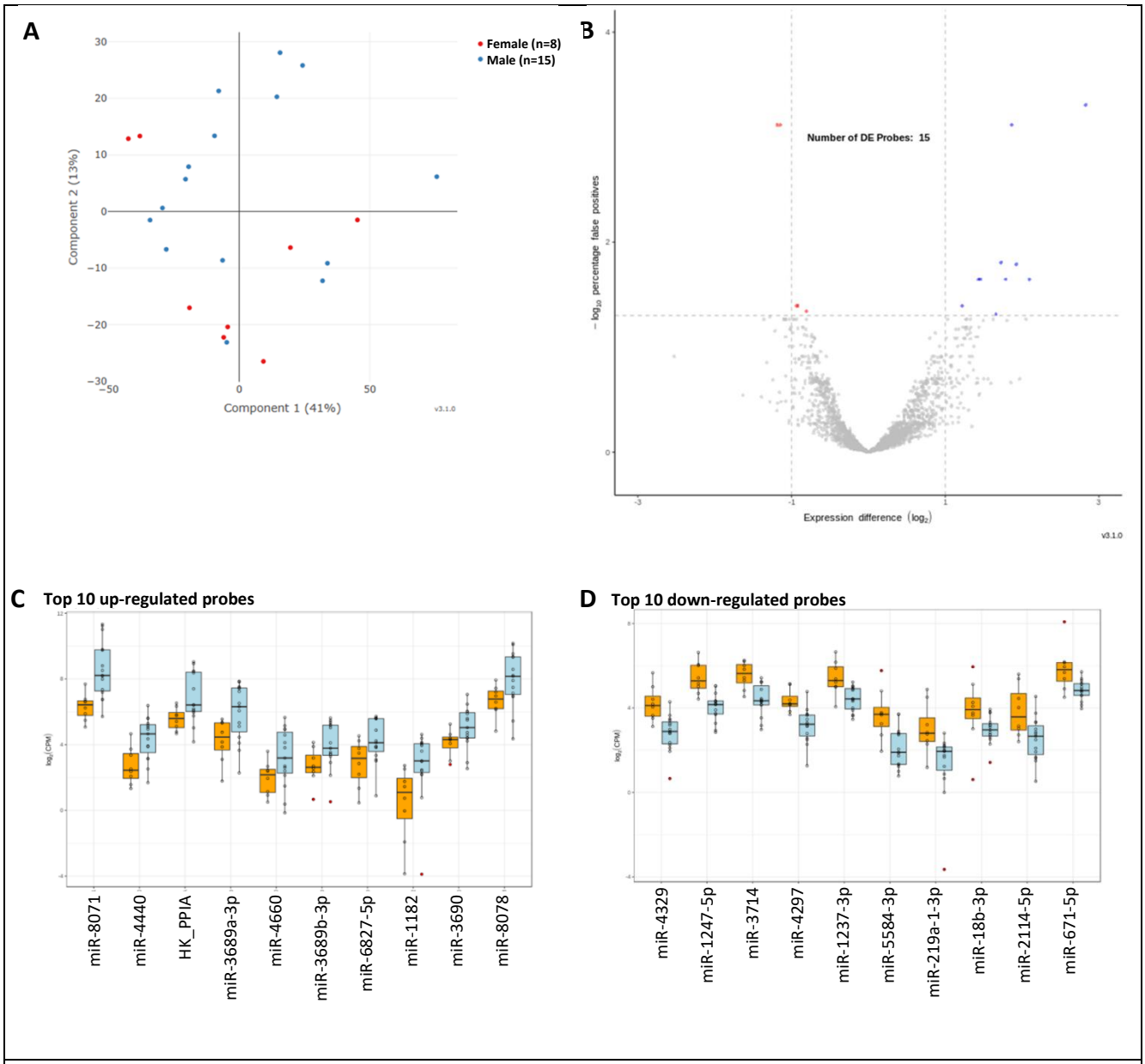


Figure 4.18. Sex - PCA plot (A), volcano plot (B), top 10 probes higher in males (C) and top 10 probes higher in females (D) for batch 1. Excluding the QC failures, questionable controls and brain RNA controls.

Boxplots: orange = female, blue = male.

For batch 2 (Fig. 4.19), despite no clear pattern of clustering was seen on the PCA plot (Fig. 4.19A), 28 miRNAs were significantly differentially expressed at an FDR adjusted P < 0.05 (Fig.

4.19B). There were 9 miRNAs where Male > Female at a raw P < 0.01, 2 of which had an FDR adjusted P < 0.05; bar graphs for the top 10 miRNAs higher in males are shown in Fig. 4.19C. There were 41 miRNAs where Male < Female at a raw P < 0.01, of which 26 had an FDR adjusted P < 0.05; bar graphs for the top 10 miRNAs higher in females are shown in Fig. 4.19D.

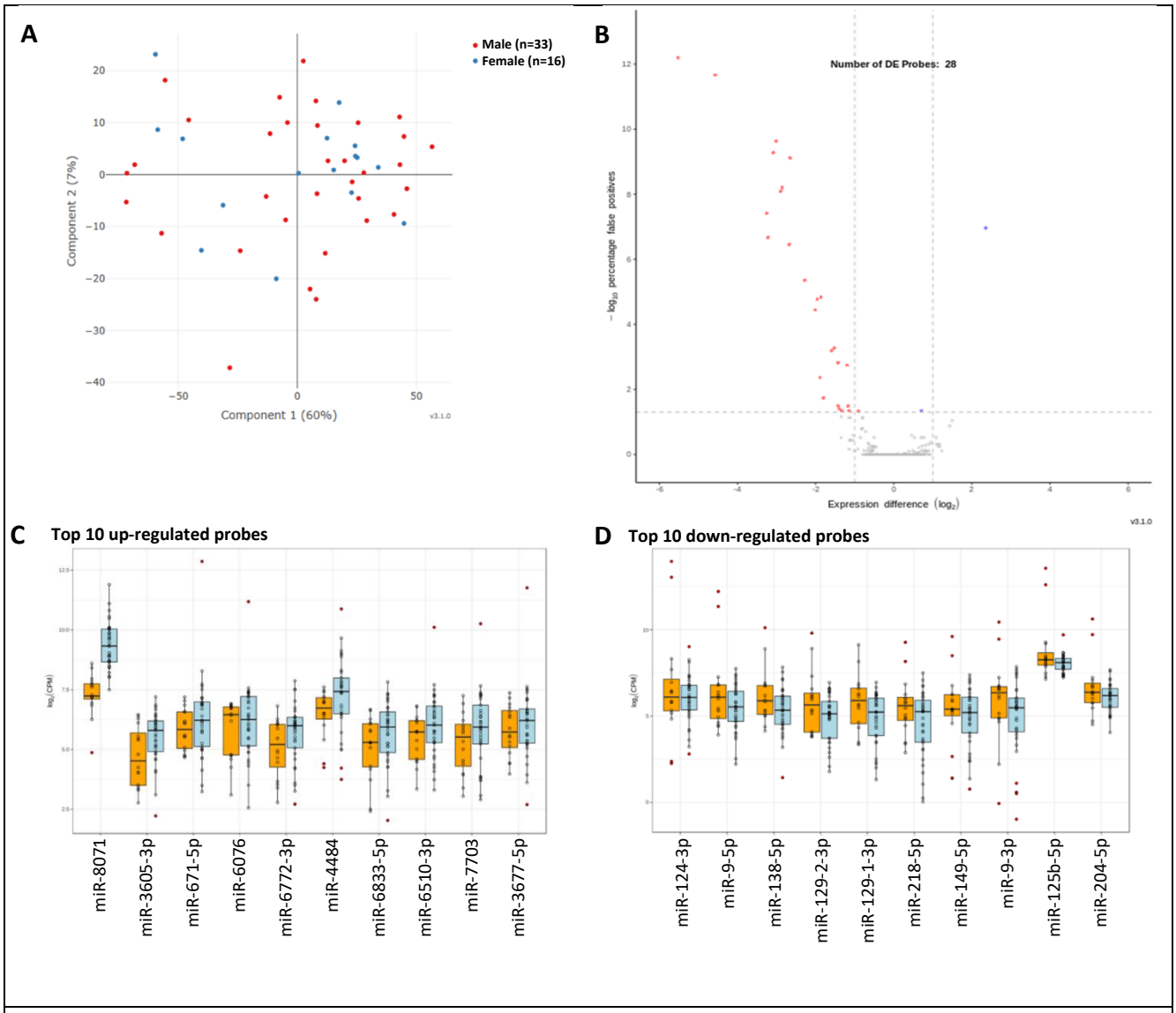


Figure 4.19. Sex - PCA plot (A), volcano plot (B), top 10 probes higher in males (C) and top 10 probes higher in females (D) for batch 2. Excluding the QC failures, questionable controls and brain RNA controls.

Boxplots: orange = female, blue = male.

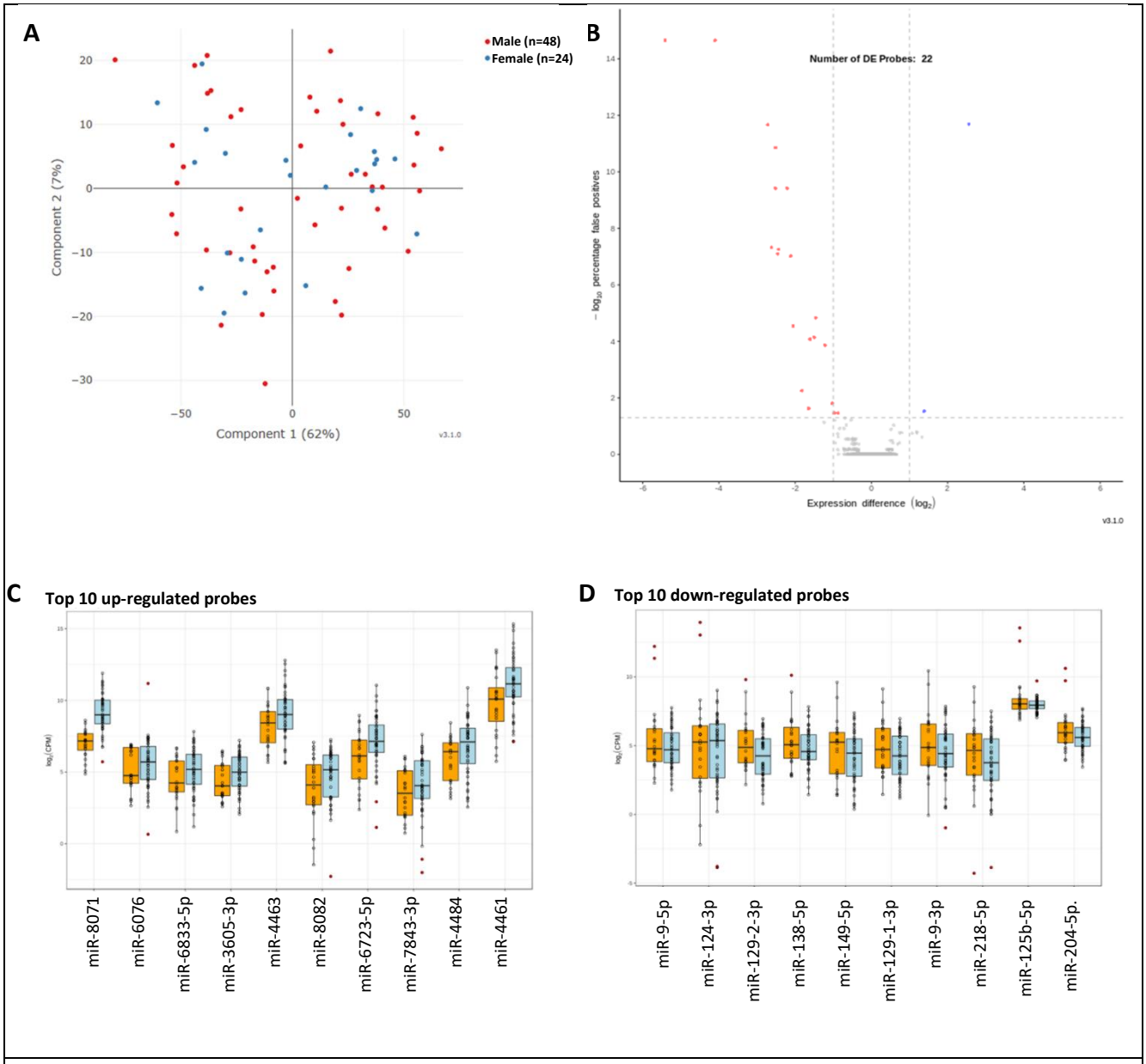


Figure 4.20. Sex - PCA plot (A), volcano plot (B), top 10 probes higher in males (C) and top 10 probes higher in females (D) for the combination of batch 1 and batch 2. Excluding the QC failures, questionable controls, and brain RNA controls.

Boxplots: orange = female, blue = male.

For combined batch 1 and batch 2 (Fig. 4.20), despite no clear pattern of clustering on the PCA plot (Fig. 4.20A), 22 miRNAs were significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.20B). There were 12 miRNAs where Male > Female at a raw $P < 0.01$, of which 2 had an FDR adjusted $P < 0.05$; bar graphs for the top 10 miRNAs higher in males are shown

in Fig. 4.20C. There were 39 miRNAs where Male < Female at a raw P < 0.01, of which 20 had an FDR adjusted P < 0.05; bar graphs for the top 10 miRNAs higher in females are shown in Fig. 4.20D.

Comparison between differentially expressed miRNAs analysed by batch or combined are summarised in Table 4.7, which shows only those that had an FDR adjusted P<0.05 for at least 2 comparisons. Only 1 miRNA (miR-8071) was consistently expressed at higher levels in males across all analyses at an FDR adjusted P < 0.05 (P = 0.00049 for batch 1, P = 1.1 $x10^{-7}$ for batch 2, P = 2.1 $x10^{-12}$ for the combined) and no miRNA consistently expressed at lower levels. Additionally, no miRNA other than miR-8071 was consistently differentially expressed by sex across all analyses at a raw P < 0.01.

Additionally, 4 other miRNAs were higher in males (Males > Females) in both batch 1 and the combined at a raw P < 0.01. Furthermore, 4 different miRNAs were higher in males (Males > Females) in both batch 2 and the combined at a raw P < 0.01, 1 of which was significantly higher in males at an FDR adjusted P < 0.05 in batch 2 (miR-3605-3p, p = 0.045) and another in the combined (miR-6076, p = 0.030), respectively. Besides that, 11 miRNAs were higher in females (Males < Females) in both batch 1 and the combined at a raw P < 0.01, 1 of which were significantly higher in females at an FDR adjusted P < 0.05 in batch 1 (miR-1247-5p, P = 0.00077); 2 of which were significantly higher in females at an FDR adjusted P < 0.05 in the combined (miR-4513, p = 0.0055; miR-6873-3p, p = 0.035). Similarly, 26 different miRNAs were higher in females (Males < Females) in both batch 2 and the combined at a raw P < 0.01, 22 of which were significantly higher in females at an FDR adjusted P < 0.05 in batch 2, and 18 of which were significantly higher in females at an FDR adjusted P < 0.05 in both batch 2 and the combined.

Table 4.7. miRNA differentially expressed by sex

	Batch 1					Batch 2					Both					
miRNA ID	**Mean normalized Female**	**Mean normalized Male**	**Fold Change Male *vs* Female**	**Raw P Male *vs* Female**	**adjP Male *vs* Female**	**Mean normalized Female**	**Mean normalized Male**	**Fold Change Male *vs* Female**	**Raw P Male *vs* Female**	**adjP Male *vs* Female**	**Mean normalized Female**	**Mean normalized Male**	**Fold Change Male *vs* Female**	**Raw P Male *vs* Female**	**adjP Male *vs* Female**	**Comparison Significant P values** Male > Female = Blue; Female > Male = Red; significant at an adjusted P < 0.05 in bold
miR-124-3p	49	28	-1.57	0.221	0.568	**23203**	**582**	**-46.09**	**3.04E-16**	**6.37E-13**	**12390**	**328**	**-42.58**	**2.12E-18**	**2.22E-15**	**Female > Male B2 Both**
miR-9-5p	99	114	1.27	0.36	0.685	**7101**	**345**	**-23.80**	**2.06E-15**	**2.16E-12**	**3834**	**251**	**-17.22**	**1.45E-18**	**2.22E-15**	**Female > Male B2 Both**
miR-8071	**715**	**4633**	**7.10**	**2.35E-07**	**0.0005**	**894**	**5291**	**5.12**	**4.68E-10**	**1.09E-07**	**812**	**5389**	**5.88**	**2.94E-15**	**2.06E-12**	**Male > Female B1 B2 Both**
miR-129-2-3p	106	58	-1.65	0.018	0.201	**1444**	**196**	**-8.51**	**1.01E-12**	**5.31E-10**	**821**	**141**	**-6.58**	**4.09E-15**	**2.14E-12**	**Female > Male B2 Both**
miR-138-5p	95	99	1.15	0.397	0.713	**1715**	**246**	**-8.07**	**3.35E-13**	**2.34E-10**	**963**	**189**	**-5.73**	**3.30E-14**	**1.39E-11**	**Female > Male B2 Both**
miR-129-1-3p	76	63	-1.10	0.662	0.863	**957**	**176**	**-6.28**	**1.82E-12**	**7.64E-10**	**551**	**134**	**-4.63**	**1.26E-12**	**3.76E-10**	**Female > Male B2 Both**
miR-149-5p	97	56	-1.58	0.079	0.368	**1218**	**189**	**-7.46**	**2.69E-11**	**8.06E-09**	**695**	**137**	**-5.73**	**1.20E-12**	**3.76E-10**	**Female > Male B2 Both**
miR-9-3p	95	149	1.71	0.121	0.452	**2183**	**265**	**-9.54**	**1.48E-10**	**3.87E-08**	**1211**	**221**	**-6.16**	**1.82E-10**	**4.76E-08**	**Female > Male B2 Both**
miR-218-5p	40	46	1.26	0.53	0.794	**996**	**159**	**-7.26**	**1.80E-11**	**6.27E-09**	**556**	**115**	**-5.42**	**2.37E-10**	**5.53E-08**	**Female > Male B2 Both**
miR-125b-5p	1882	1898	1.11	0.664	0.864	**18674**	**2316**	**-9.33**	**1.01E-09**	**2.11E-07**	**10781**	**2209**	**-5.50**	**3.81E-10**	**7.99E-08**	**Female > Male B2 Both**
miR-204-5p	312	263	-1.08	0.746	0.908	**2573**	**466**	**-6.38**	**1.83E-09**	**3.49E-07**	**1511**	**393**	**-4.33**	**5.06E-10**	**9.64E-08**	**Female > Male B2 Both**
miR-885-5p	303	198	-1.39	0.159	0.51	**1173**	**372**	**-3.64**	**9.01E-08**	**1.45E-05**	**771**	**316**	**-2.75**	**8.40E-08**	**1.47E-05**	**Female > Male B2 Both**
miR-219a-2-3p	35	24	-1.31	0.46	0.758	**844**	**201**	**-4.86**	**2.52E-08**	**4.39E-06**	**476**	**129**	**-4.15**	**1.78E-07**	**2.87E-05**	**Female > Male B2 Both**
miR-3609	87	112	1.42	0.075	0.356	**889**	**264**	**-3.89**	**1.12E-07**	**1.68E-05**	**521**	**207**	**-2.83**	**4.72E-07**	**7.07E-05**	**Female > Male B2 Both**
miR-3607-5p	92	105	1.26	0.235	0.582	**1173**	**337**	**-4.03**	**2.58E-07**	**3.60E-05**	**675**	**249**	**-3.05**	**6.01E-07**	**8.40E-05**	**Female > Male B2 Both**
miR-31-5p	221	118	-1.70	0.01	0.157	**686**	**295**	**-2.69**	**1.26E-05**	**0.0015**	**472**	**228**	**-2.33**	**1.04E-06**	**0.0001**	**Female > Male B2 Both**
miR-4513	2798	443	-5.76	0.006	0.123	259	357	1.19	0.4037	0.9999	**1334**	**424**	**-3.54**	**4.43E-05**	**0.0055**	**Female > Male B1 Both**
miR-1296-5p	183	178	1.07	0.829	0.938	**584**	**236**	**-2.87**	**4.05E-06**	**0.0005**	**392**	**217**	**-2.04**	**0.0001**	**0.0157**	**Female > Male B2 Both**
miR-137	20	19	1.06	0.908	0.964	**357**	**112**	**-3.70**	**4.07E-05**	**0.004**	**207**	**74**	**-3.13**	**0.0002**	**0.0236**	**Female > Male B2 Both**
miR-6076	114	168	1.62	0.031	0.236	302	947	2.70	0.0022	0.1309	**229**	**672**	**2.61**	**0.0003**	**0.0299**	**Male > Female B2 Both**
miR-132-3p	362	408	1.23	0.47	0.763	**1121**	**431**	**-3.01**	**5.13E-06**	**6.33E-04**	**752**	**430**	**-1.97**	**0.0004**	**0.0346**	**Female > Male B2 Both**
miR-6873-3p	4162	1638	-2.32	0.002	0.07	4215	3052	-1.60	0.0275	0.7705	**4228**	**2587**	**-1.84**	**0.0004**	**0.0346**	**Female > Male B1 Both**
miR-455-3p	136	91	-1.37	0.24	0.589	**645**	**338**	**-2.21**	**0.0006**	**0.0451**	419	243	-1.95	0.0007	0.0618	**Female > Male B2 Both**
miR-135b-5p	27	18	-1.36	0.417	0.723	**243**	**104**	**-2.69**	**0.0003**	**0.0323**	148	70	-2.37	0.0009	0.073	**Female > Male B2 Both**
miR-153-3p	255	293	1.26	0.419	0.725	**698**	**353**	**-2.29**	**1.64E-05**	**0.0018**	491	336	-1.65	0.0017	0.1203	**Female > Male B2 Both**
miR-3605-3p	111	106	1.05	0.777	0.923	**142**	**268**	**1.63**	**0.0006**	**0.0451**	129	208	1.43	0.0019	0.1308	**Male > Female B2 Both**
miR-1247-5p	**345**	**141**	**-2.22**	**1.46E-06**	**0.0008**	469	460	-1.18	0.2986	0.9999	407	332	-1.38	0.0079	0.3259	**Female > Male B1 Both**
miR-4329	**149**	**60**	**-2.28**	**7.79E-07**	**0.0008**	312	327	-1.10	0.4215	0.9999	245	220	-1.26	0.0863	0.9976	**Female > Male B1**
miR-4440	**59**	**196**	**3.64**	**1.17E-06**	**0.0008**	283	326	-1.00	0.982	0.9999	194	285	1.30	0.0835	0.9976	**Male > Female B1**

adjP = FDR adjusted P, B1 = batch 1, B2 = batch 2

Since there was some imbalance between the proportion of cases and controls that were male *vs* female (Table 4.4), we examined if those miRNAs that were differentially expressed between sexes for whether they were also expressed differentially between cases and controls. Of 29 miRNAs with an FDR adjusted P < 0.05 in any comparison, only 2 were significantly different between cases and controls (each only when looking at combined data, but not batch 1 or batch 2 separately). These miRNAs were both higher in females than males, but one was higher in cases, and one was lower. Therefore, in keeping with the Chi Squared test, we have found no evidence of a consistent sex-related bias that might confound our case *vs* control analysis.

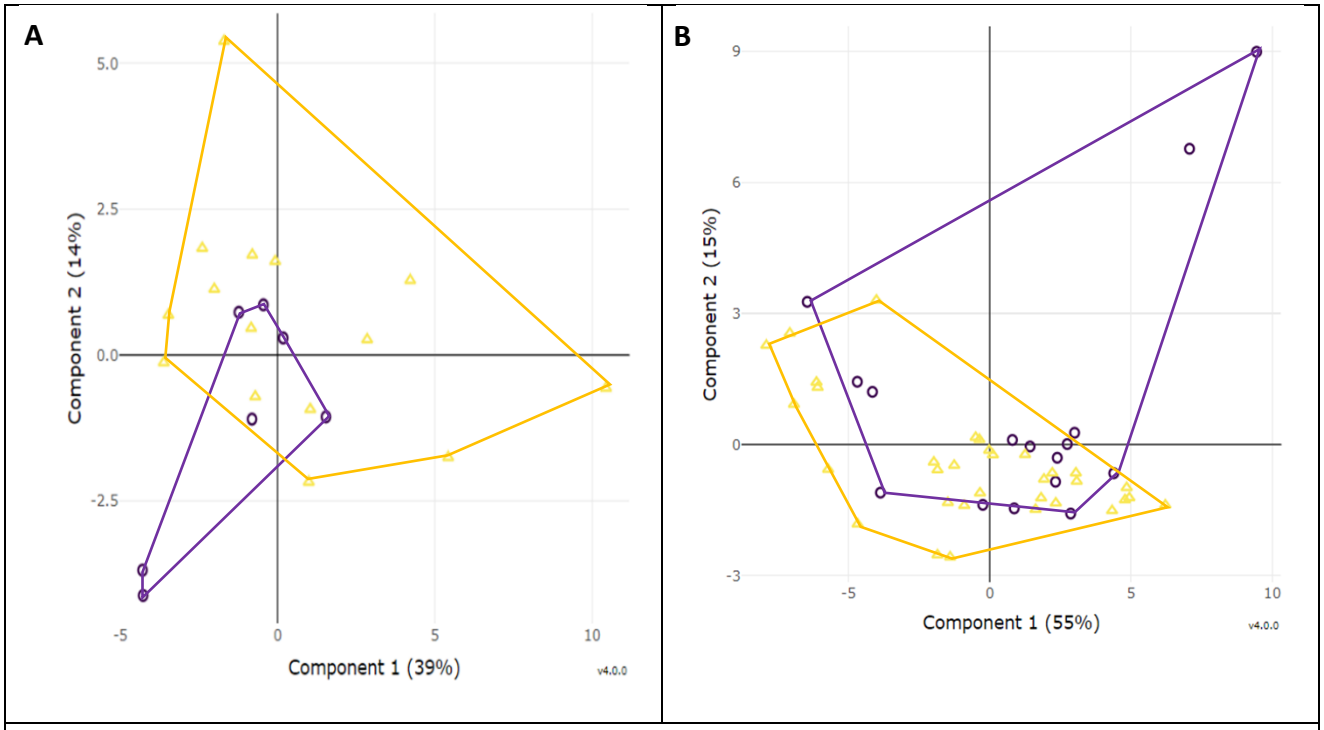


Figure 4.21. PCA plot based on 51 miRNAs with at least one FDR adjusted P < 0.05 when reanalysing HTG data for differential expression associated with sex, batch 1 (A) and batch 2 (B).

Female = purple circle, male = yellow triangle.

Note: outlines drawn round all samples.

When we limit the miRNAs to those with at least one FDR adjusted P < 0.05 (as per Table 4.7), there is no clear clustering in batch 1 (Fig. 4.21A), but we can see improved separation of male and female in batch 2 (Fig. 4.21B). This is in keeping with the miRNAs being selected primarily due to differential expression in batch 2 or combined (Table 4.7).

4.5.5 Smoking

Smoking causes significant stress to the lungs and might therefore lead to release of EVs containing miRNAs. There might therefore be differences in EV miRNA expression dependent on whether the subject was a current smoker. Other miRNAs might be differentially expressed due to long-term smoking damage; these might be associated with lung cancer risk, e.g. in those subjects that suffer most damage or are unable to repair that damage effectively. We can address these patterns by looking at smoking as current status, or a combination of duration of smoking, current status and years since quit smoking.

Current vs former smokers

For batch 1 (Fig. 4.22), no clear pattern of clustering was seen on the PCA plot (Fig. 4.22A), no miRNA was significantly differentially expressed at an FDR adjusted P < 0.05 (Fig. 4.22B). There were 3 miRNAs where Current > Former at a raw P < 0.01 (Table 4.8); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.22C. There were 2 different miRNAs where Current < Former at a raw P < 0.01 (Table 4.8); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.22D.

For batch 2 (Fig. 4.23), no clear pattern of clustering was seen on the PCA plot (Fig. 4.23A), 2 miRNAs were significantly differentially expressed at an FDR adjusted P < 0.05 (Fig. 4.23B).

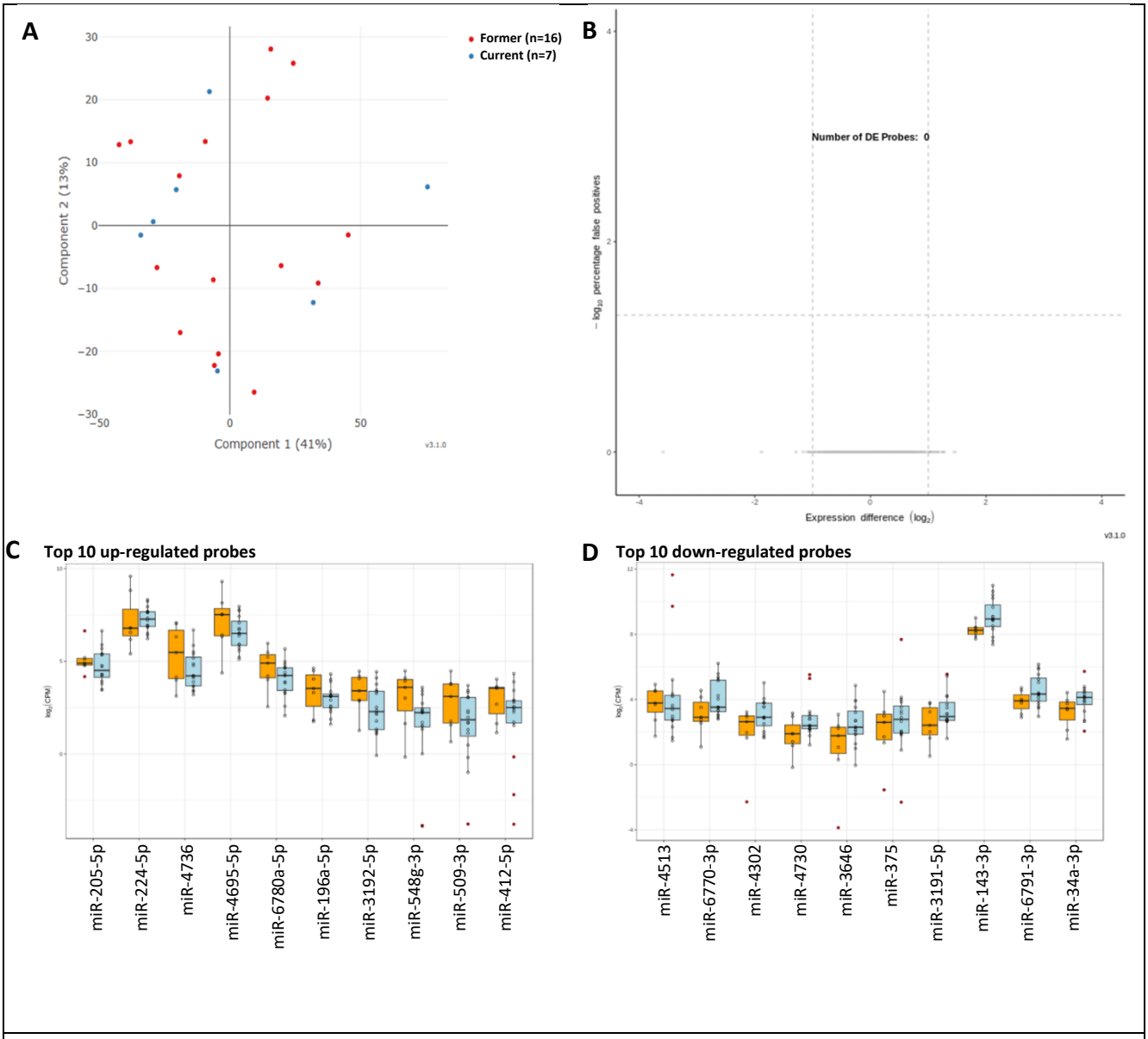


Figure 4.22. Current *vs* former smokers - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for batch 1. Excluding the QC failures, questionable controls and brain RNA controls.

Boxplots: orange = current, blue = former.

Table 4.8. miRNA differentially expressed by Current *vs* former smokers

	Batch 1					Batch 2					Both					
miRNA ID	**Mean normalized Former**	**Mean normalized Current**	**Fold Change Current *vs* Former**	**Raw P Current *vs* Former**	**adjP Current *vs* Former**	**Mean normalized Former**	**Mean normalized Current**	**Fold Change Current *vs* Former**	**Raw P Current *vs* Former**	**adjP Current *vs* Former**	**Mean normalized Former**	**Mean normalized Current**	**Fold Change Current *vs* Former**	**Raw P Current *vs* Former**	**adjP Current *vs* Former**	**Comparison Significant P values** Current > Former = Blue; Former > Current = Red; significant at an adjusted P < 0.05 in bold
miR-4513	1705	137	-12.02	0.0024	0.9992	376	272	-1.21	0.3495	0.9986	**1185**	**236**	**-4.79**	**1.53E-06**	**0.0016**	**Former > Current B1 Both**
miR-4695-5p	838	1540	1.91	0.0188	0.9992	1391	2415	1.99	0.0036	0.3583	**1139**	**2196**	**2.02**	**7.97E-05**	**0.0334**	**Current > Former B2 Both**
miR-4736	235	499	2.20	0.0126	0.9992	534	926	1.99	0.0011	0.2282	**397**	**812**	**2.15**	**1.54E-05**	**0.0107**	**Current > Former B2 Both**
miR-6076	150	153	1.06	0.8094	0.9992	**352**	**1114**	**3.63**	**9.19E-06**	**0.0096**	**260**	**830**	**3.35**	**6.84E-08**	**0.0001**	**Current > Former B2 Both**
miR-671-5p	391	267	-1.41	0.2072	0.9992	**1303**	**319**	**-3.57**	**9.04E-06**	**0.0096**	**772**	**308**	**-2.39**	**2.55E-05**	**0.0134**	**Former > Current B2 Both**

adjP = FDR adjusted P, B1 = batch 1, B2 = batch 2

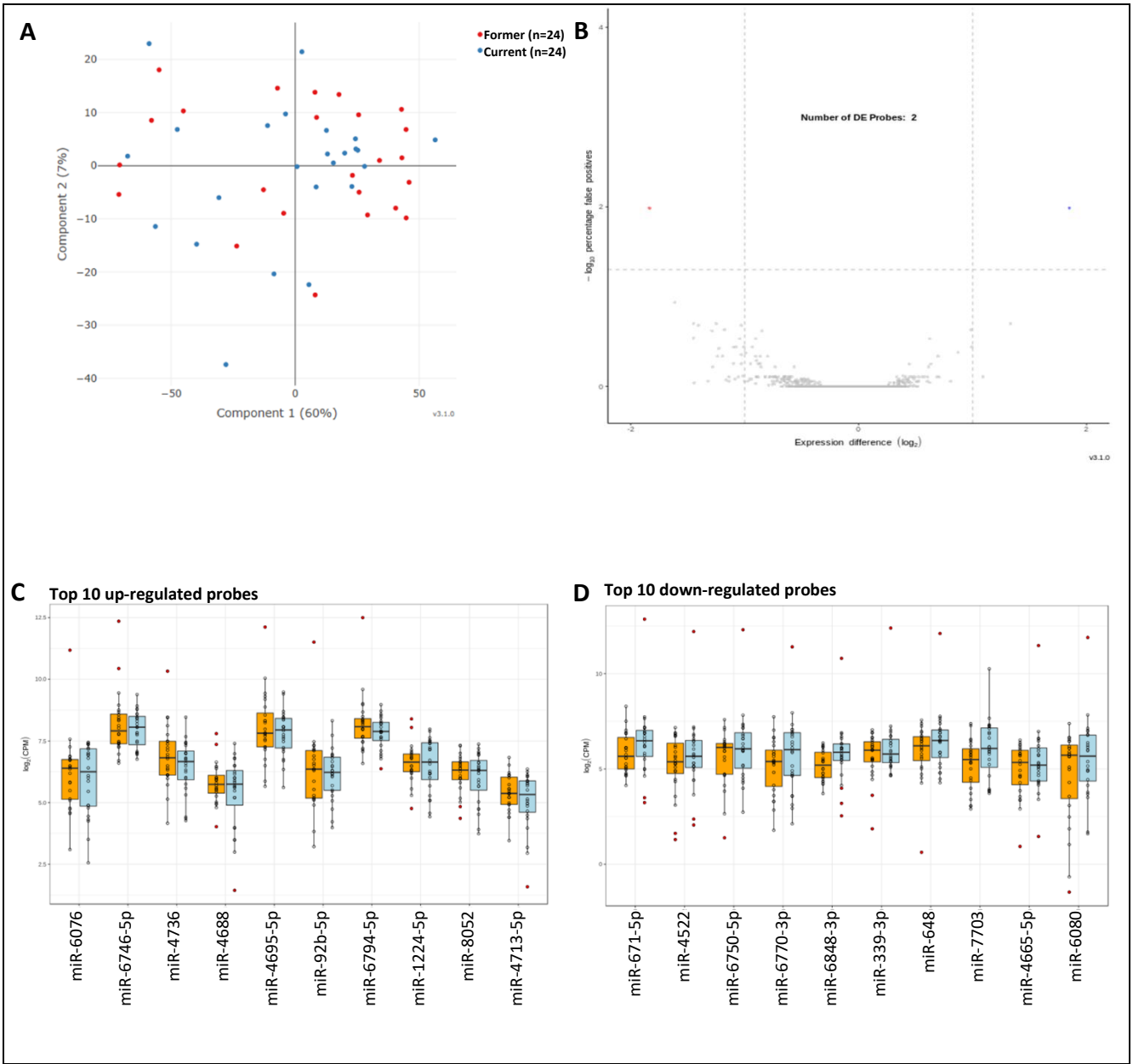


Figure 4.23. Current *vs* former smokers - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for batch 2. Excluding the QC failures, questionable controls, and brain RNA controls.

Boxplots: orange = current, blue = former.

There were 7 miRNAs where Current > Former at a raw P < 0.01, of which 1 had an FDR adjusted P < 0.05 (Table 4.8); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.23C. There were 29 miRNAs where Current < Former at a raw P < 0.01, of which 1 had an

FDR adjusted P < 0.05 (Table 4.8); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.23D.

For combined batch 1 and batch 2 (Fig. 4.24), despite no clear pattern of clustering on the PCA plot (Fig. 4.24A), 5 miRNAs were significantly differentially expressed at an FDR adjusted P < 0.05 (Fig. 4.24B). There were 12 miRNAs where Current > Former at a raw P < 0.01, of which 3 had an FDR adjusted P < 0.05 (Table 4.8); bar graphs for the top 10 up-regulated miRNAs are shown in Fig. 4.24C. There were 10 miRNAs where Current < Former at a raw P < 0.01, of which 2 had an FDR adjusted P < 0.05 (Table 4.8); bar graphs for the top 10 down-regulated miRNAs are shown in Fig. 4.24D.

Comparison between differentially expressed miRNAs analysed by batch or combined are summarised in Table 4.8. No miRNA was consistently up-regulated in current across all analyses at an FDR adjusted P < 0.05 and no miRNA down-regulated. However, 1 miRNA (miR-6770-3p) was consistently higher in former smokers across all analyses at a raw P < 0.01 (P = 0.0093 for batch 1, P = 0.0005 for batch 2, P = 0.001 for the combined) and no miRNA up-regulated.

Additionally, 1 miRNA was up-regulated (Current > Former) in both batch 1 and the combined at a raw P < 0.01. And 7 different miRNAs were up-regulated (Current > Former) in both batch 2 and the combined at a raw P < 0.01, 1 of which was significantly up-regulated at an FDR adjusted P < 0.05 in both batch 2 (miR-6076, p = 0.0096) and the combined (miR-6076, p = 0.00014).

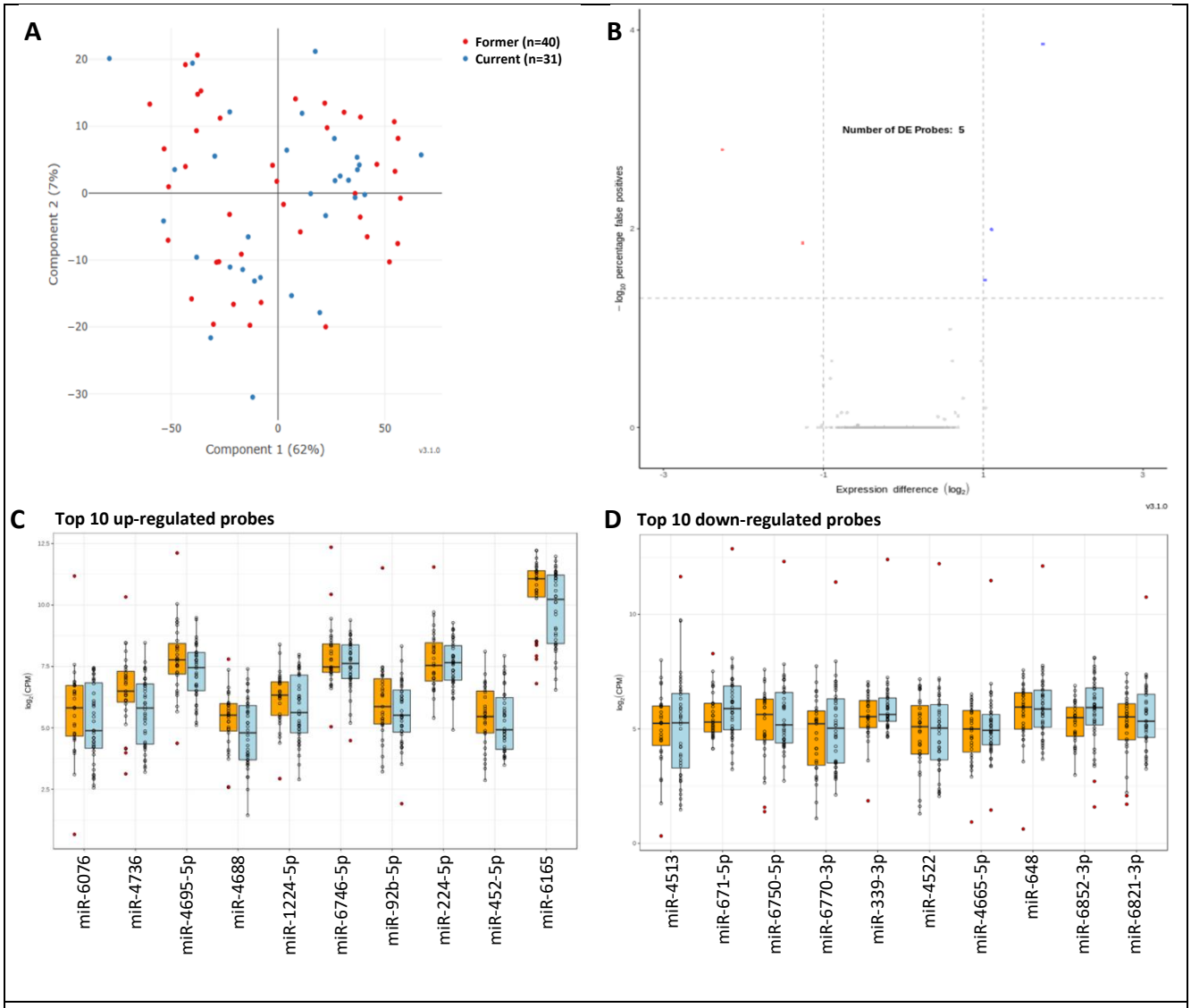


Figure 4.24. Current *vs* former smokers - PCA plot (A), volcano plot (B), top 10 up-regulated probes (C) and top 10 down-regulated probes (D) for the combination of batch 1 and batch 2. Excluding the QC failures, questionable controls and brain RNA controls.

Boxplots: orange = current, blue = former.

Besides that, 1 miRNA was down-regulated (Current < Former) in both batch 1 and the combined at a raw P < 0.01, and it was significantly down-regulated at an FDR adjusted P < 0.05 in the combined (miR-4513, P = 0.0016). Similarly, 6 different miRNAs were down-regulated (Current < Former) in both batch 2 and the combined at a raw P < 0.01, 1 of which

(miR-671-5p) was significantly down-regulated at an FDR adjusted $P < 0.05$ in both batch 2 ($p = 0.0096$) and the combined ($p = 0.0134$).

When we limit the miRNAs to those with at least one FDR adjusted $P < 0.05$ (as per Table 4.8), we can see improved separation of smoking-current and smoking-former in batch 1 (Fig. 4.25A) and some clustering in batch 2 (Fig. 4.25B). This is despite the miRNAs being selected primarily due to differential expression in batch 2 or combined (Table 4.8).

Detailed smoking history

As we classed smoking into 4 categories, we only looked at combined analysis (to avoid small group sizes). The categories were: (1) current smokers, all of whom had smoked over 30 years (n = 31, average smoking duration = 48 years; average pack years = 53 pack years); (2) former smokers who quit less than 3 years ago, i.e. recent quitters, all of whom had smoked over 30 years (n = 10, average smoking duration = 45; average pack years = 49 pack years); (3) former smokers who quit more than 3 years ago, i.e. long-term quitters, all of whom had smoked over 30 years (n = 17, average smoking duration = 43 years; average pack years = 45 pack years); (4) former smokers who quit more than 3 years ago, i.e. long-term quitters, but all of whom had smoked less than 30 years (n = 12, average smoking duration = 19; average pack years = 15 pack years).

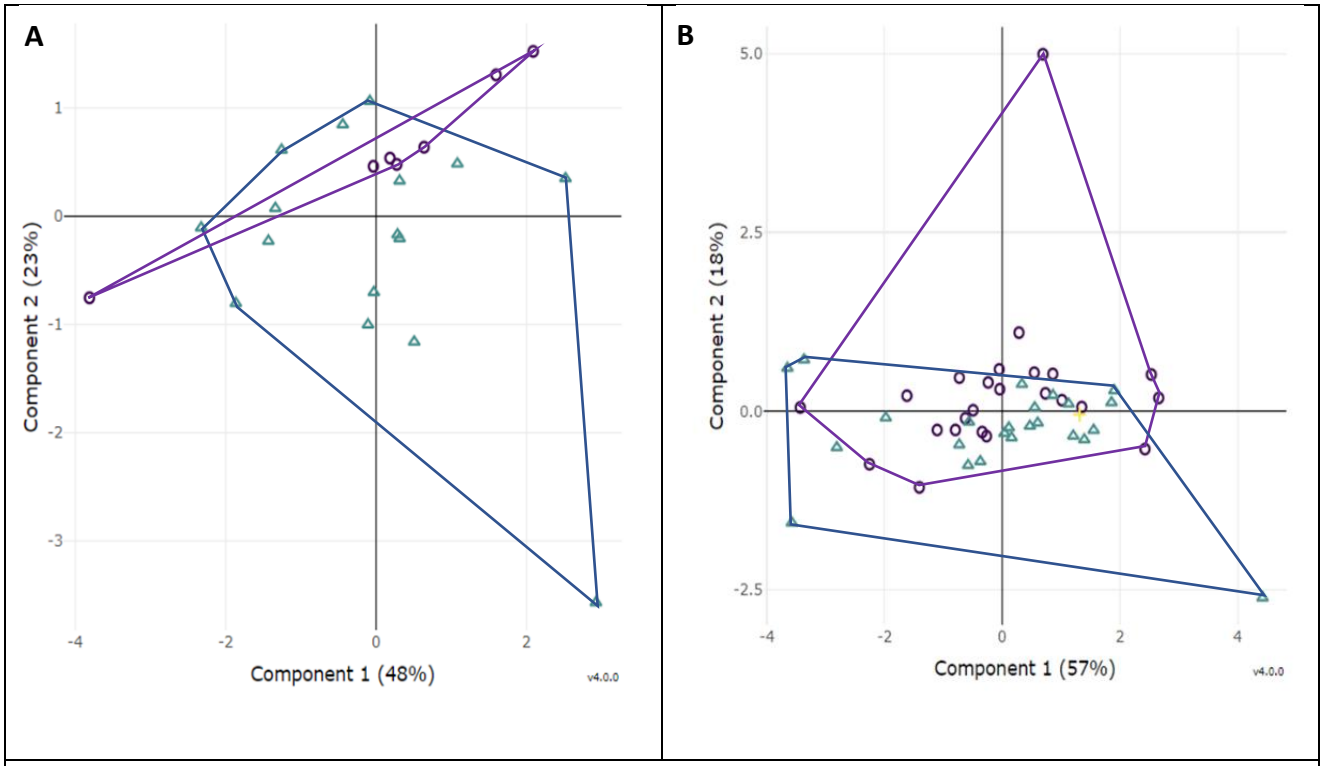


Figure 4.25. PCA plot based on 5 miRNAs with at least one FDR adjusted P < 0.05 when reanalysing HTG data for differential expression associated with smoking status, batch 1 (A) and batch 2 (B).

Current = purple circle, former = teal triangle, unknown = yellow cross.

Note: outlines drawn round all samples.

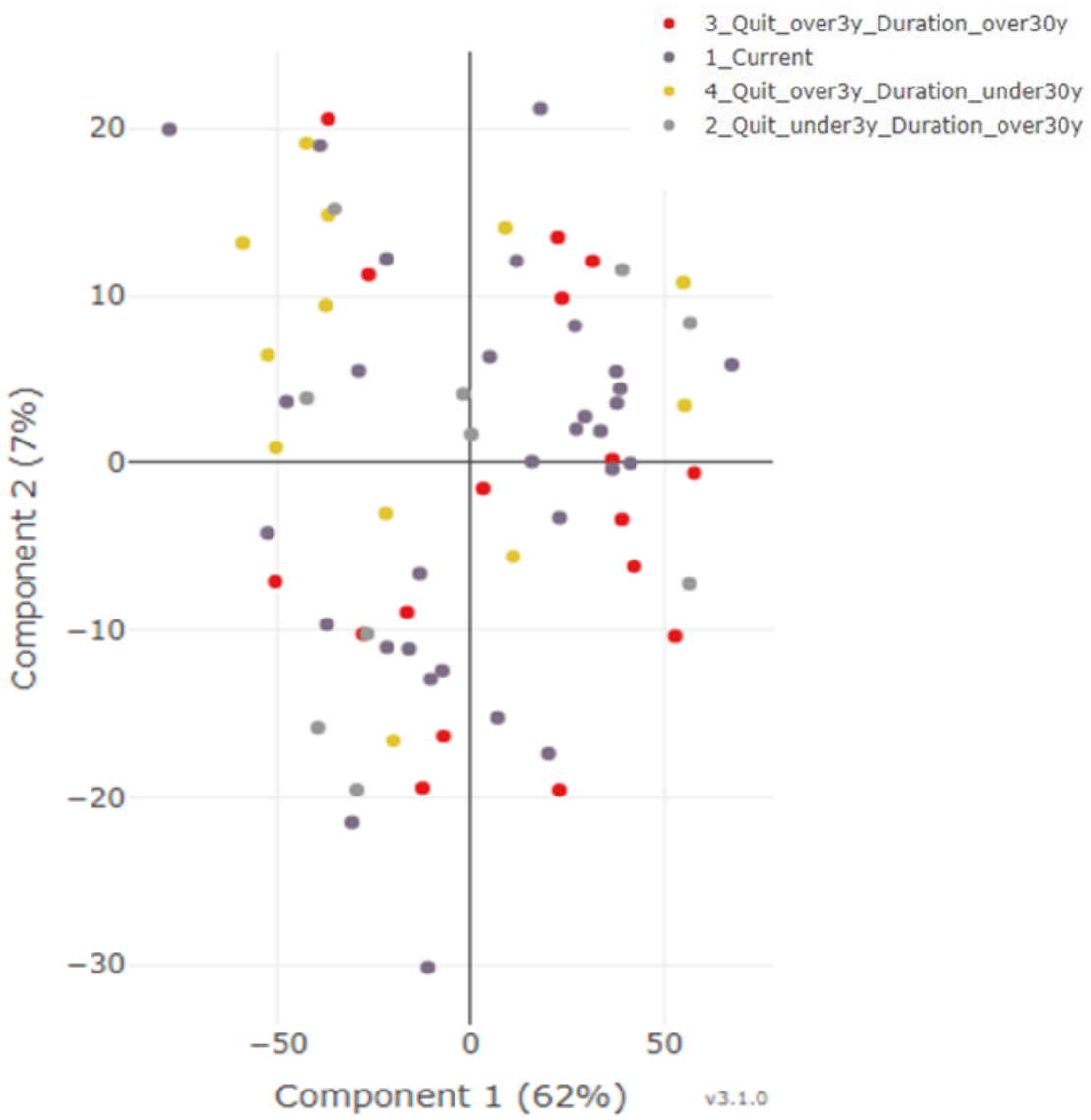


Figure 4.26. PCA plot of all samples coloured by the detailed smoking history category including years since quit smoking and duration of smoking.

Quit smoking less than 3 years ago compared to current smokers

We compared EV miRNA expression between current smokers and those who quit less than 3 years previously; we found no significant differences (Fig. 4.27).

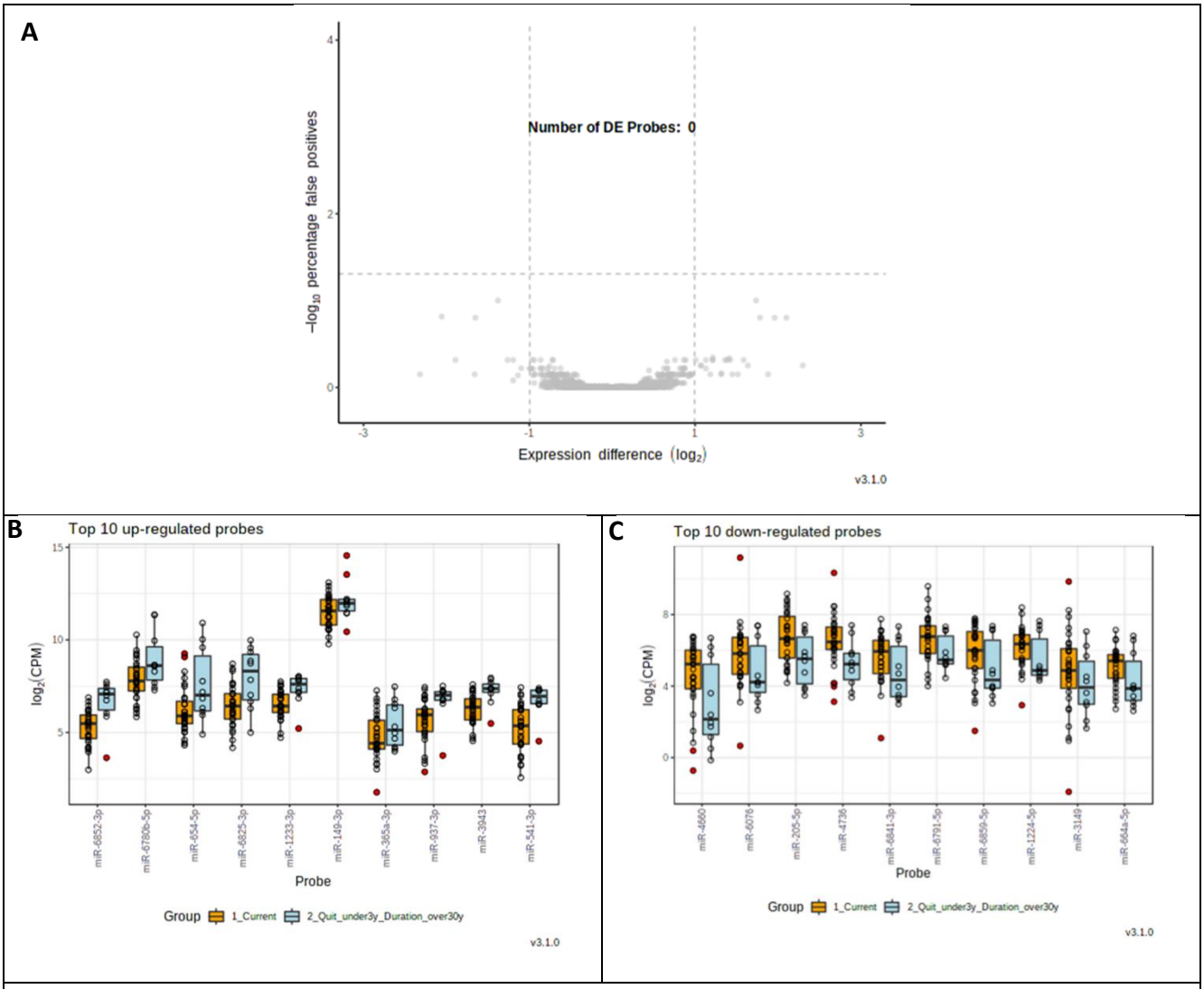


Figure 4.27. Volcano plot (A), top 10 up-regulated probes (B) and top 10 down-regulated probes (C) for those who quit smoking less than 3 years ago compared to current smokers (2_quit_under3y_Duration_over30y *versus* 1_Current).

Long-term smokers who quit smoking more than 3 years ago compared to current smokers

We compared EV miRNA expression between current smokers and those who quit more than 3 years previously and had smoked for over 30 years; we found 37 miRNAs significant differentially expressed at an FDR adjusted $P < 0.05$, 33 miRNAs of which were up-regulated in those who quit and 4 miRNAs of which were down-regulated (Fig. 4.28).

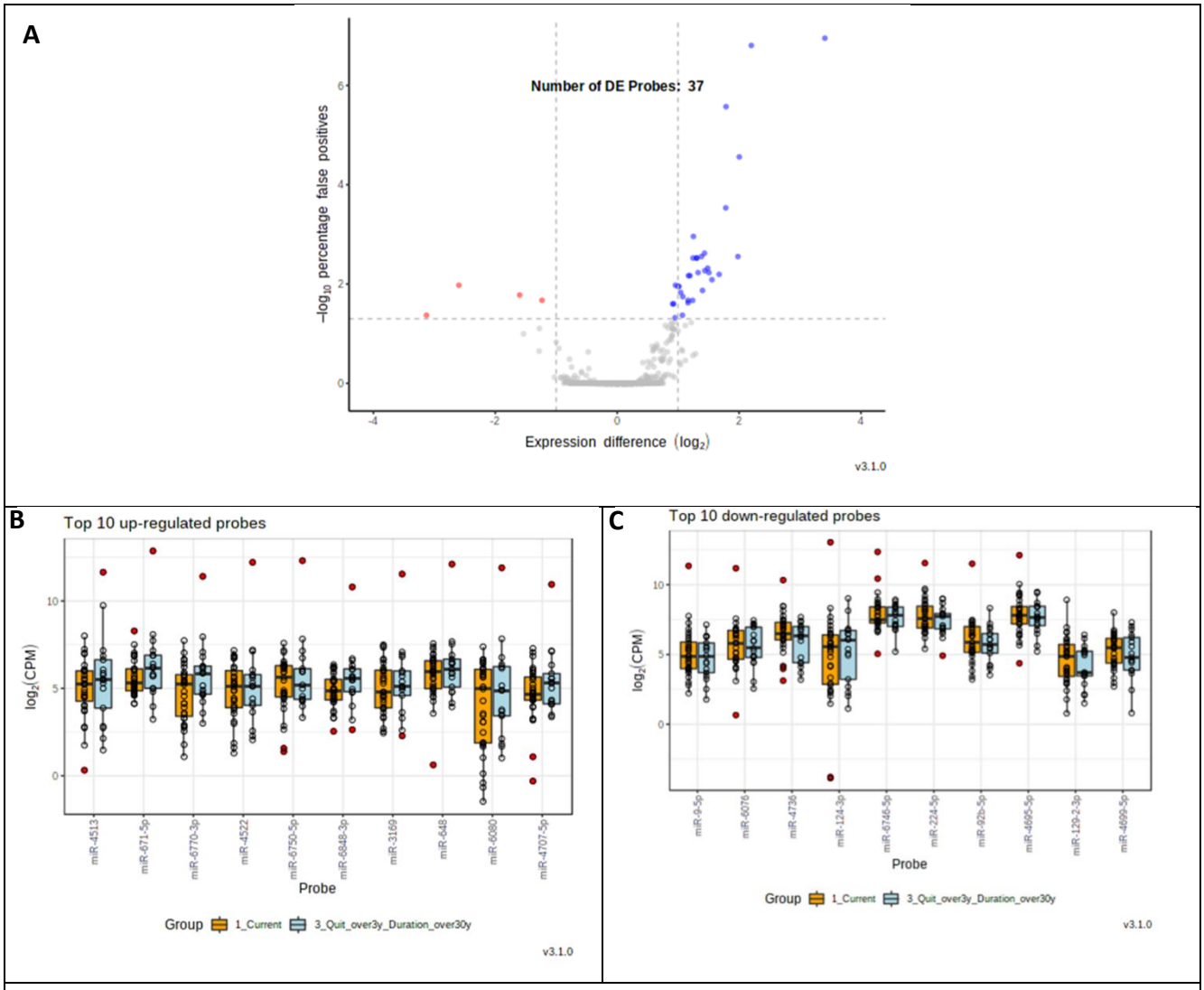


Figure 4.28. Volcano plot (A), top 10 up-regulated probes (B) and top 10 down-regulated probes (C) for long-term smokers who quit smoking more than 3 years ago compared to current smokers (3_quit_over3y_Duration_over30y *versus* 1_Current).

Shorter-term smokers who quit smoking more than 3 years ago compared to current smokers

We compared EV miRNA expression between current smokers and those who quit more than 3 years previously and had smoked for less than 30 years; we found 19 miRNAs significantly up-regulated for those who quit (Fig. 4.29).

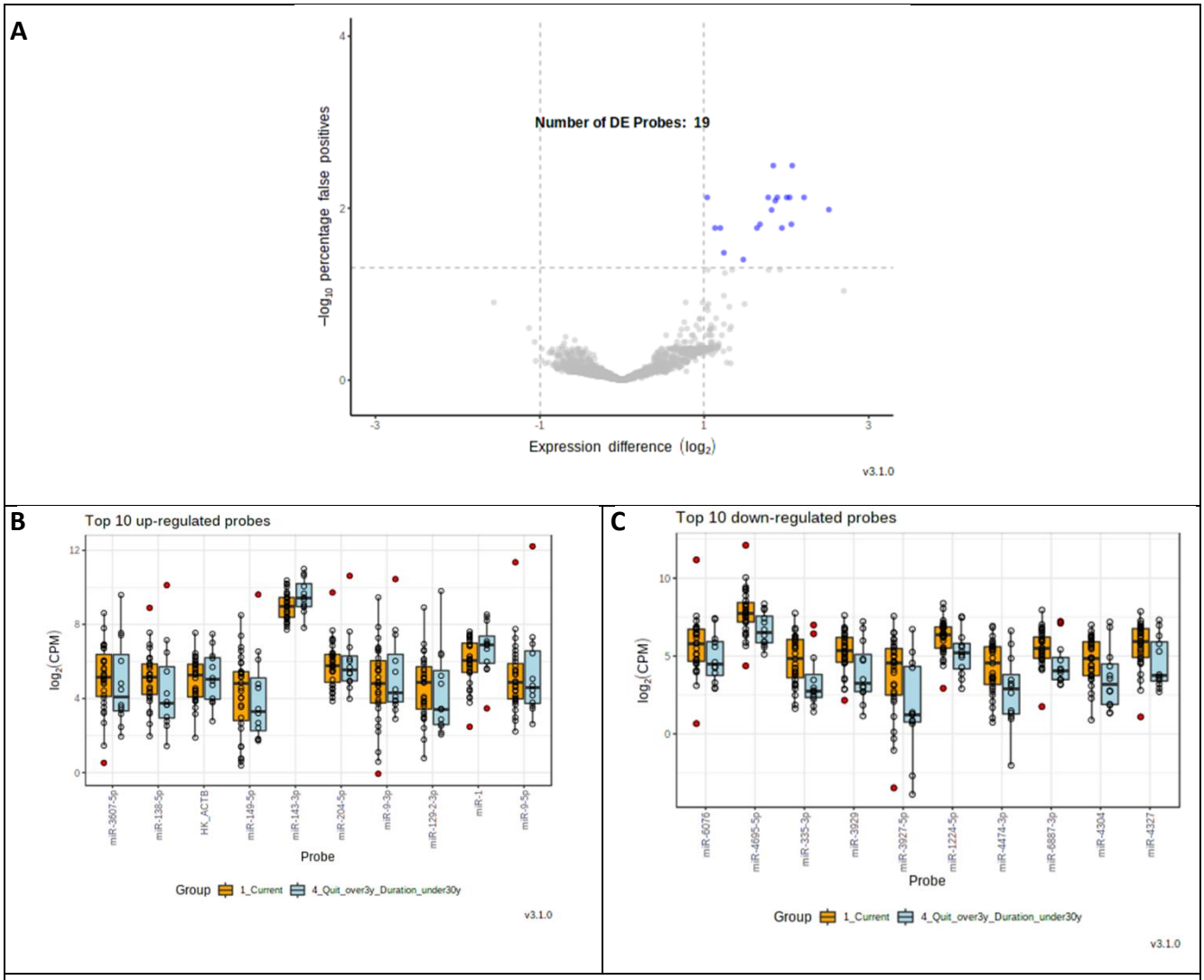


Figure 4.29. Volcano plot (A), top 10 up-regulated probes (B) and top 10 down-regulated probes (C) for shorter-term smokers who quit smoking more than 3 years ago compared to current smokers (4_quit_over3y_Duration_under30y *versus* 1_Current).

When looking at the differentially expressed miRNAs, a number of different patterns of expression can be seen in relation to when the subjects quit smoking and how long they had smoked for (Fig. 4.30).

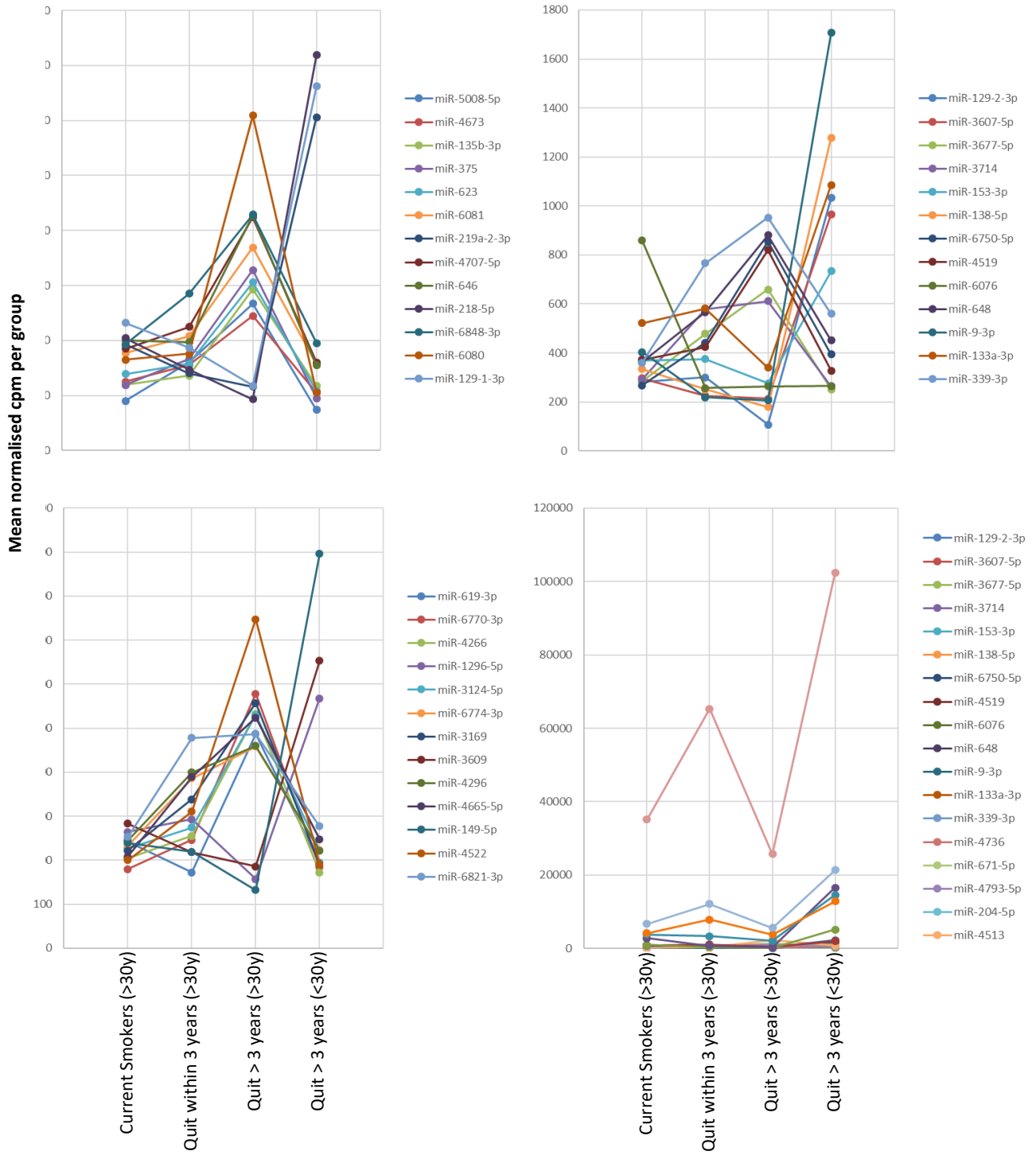


Figure 4.30. Top 52 differentially expressed miRNAs with respect to smoking history: sorted by average expression and grouped into 13 per graph (note Y axis different in each graph). All of the 52 miRNAs had a greater than 2-fold change in at least one comparison and at an FDR adjusted P < 0.05.

4.5.6 Histology

Different histological subtypes of lung cancer are characterised by their morphological appearance and also by differences in gene expression and to some extent mutation patterns. We might therefore expect some differences in EV miRNA released by the cancer cells, although not necessarily in EV miRNAs released as part of the host response. We therefore examined plasma EV miRNA differences between cases of adenocarcinoma (ADC) and squamous cell carcinoma (SqCCa). The proportion of individuals in each batch that were diagnosed with ADC or SqCCa are shown in Table 4.3. Overall, approximately half of the subjects had ADC (49%); the proportion was higher for batch 2 (52%) than batch 1 (44%), but not significantly so ($P = 0.61$).

For batch 1 (Fig. 4.31) despite no clear pattern of clustering on the PCA plot (Fig. 4.31A), 1 miRNA was significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.31B). There were 12 miRNAs where SqCCa > ADC at a raw $P < 0.01$, of which none had an FDR adjusted $P < 0.05$ (Table 4.9); bar graphs for the top 10 miRNAs higher in SqCCa are shown in Fig. 4.31C. There was 1 miRNA where SqCCa < ADC at a raw $P < 0.01$, it also had an FDR adjusted $P < 0.05$ (Table 4.9); bar graphs for the top 10 miRNAs higher in ADC are shown in Fig. 4.31D.

For batch 2 (Fig. 4.32), no clear pattern of clustering was seen on the PCA plot (Fig. 4.32A), 3 miRNAs were significantly differentially expressed at an FDR adjusted $P < 0.05$ (Fig. 4.32B). There were 85 miRNAs where SqCCa > ADC at a raw $P < 0.01$, 2 of which had an FDR adjusted $P < 0.05$ (Table 4.9); bar graphs for the top 10 miRNAs higher in SqCCa are shown in Fig. 4.32C. There were 6 miRNAs where SqCCa < ADC at a raw $P < 0.01$, of which 1 had an FDR adjusted $P < 0.05$ (Table 4.9); bar graphs for the top 10 miRNAs higher in ADC are shown in Fig. 4.32D.

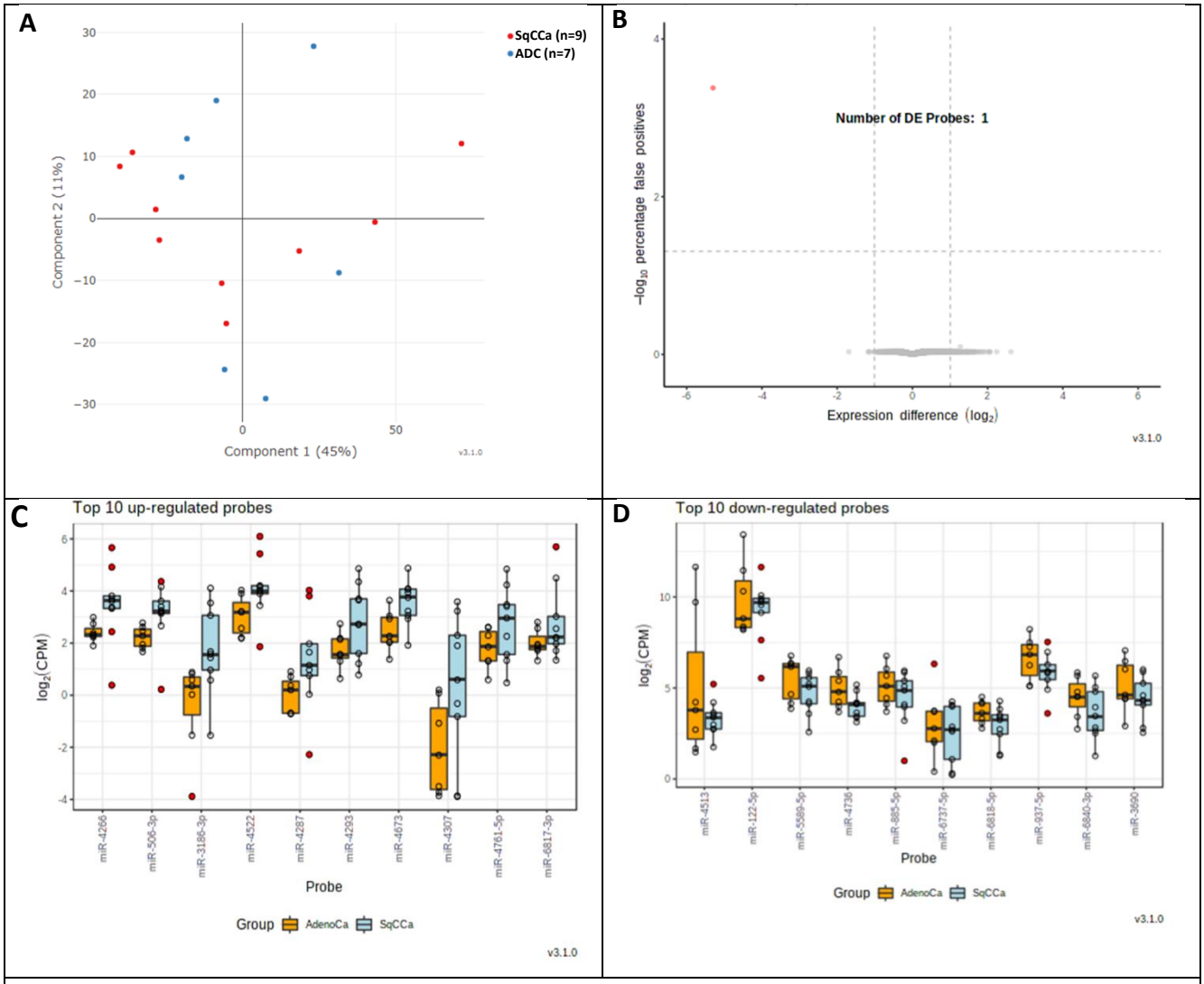


Figure 4.31. SqCCa *vs* ADC - PCA plot (A), volcano plot (B), top 10 probes higher in SqCCa (C) and top 10 probes higher in ADC (D) for batch 1. Excluding the QC failures, questionable controls, brain RNA controls and control samples.

Table 4.9. miRNAs differentially expressed by SqCCa *vs* ADC

	Batch 1					Batch 2					Both					
miRNA ID	**Mean normalized SqCCa**	**Mean normalized ADC**	**Fold Change SqCCa *vs* ADC**	**Raw P SqCCa *vs* ADC**	**adjP SqCCa *vs* ADC**	**Mean normalized SqCCa**	**Mean normalized ADC**	**Fold Change SqCCa *vs* ADC**	**Raw P SqCCa *vs* ADC**	**adjP SqCCa *vs* ADC**	**Mean normalized SqCCa**	**Mean normalized ADC**	**Fold Change SqCCa *vs* ADC**	**Raw P SqCCa *vs* ADC**	**adjP SqCCa *vs* ADC**	**Comparison Significant P values** SqCCa > ADC = Blue; ADC > SqCCa = Red; significant at an adjusted P < 0.05 in bold
miR-124-3p	46	24	1.75	2.05E-01	9.27E-01	**8596**	**527**	**21.21**	**9.29E-08**	**0.0001**	**4071**	**297**	**16.18**	**3.62E-07**	**0.0003**	**B2 & Both SqCCa > ADC**
miR-224-5p	4085	1229	3.09	1.66E-02	9.27E-01	7237	1788	5.26	0.0002	0.0582	**4882**	**1520**	**3.79**	**2.85E-05**	**0.012**	**Both SqCCa > ADC**
miR-4513	**102**	**3728**	**-39.42**	**1.99E-07**	**0.0004**	270	344	1.02	0.9299	0.9944	**183**	**2274**	**-10.50**	**2.34E-08**	**4.91E-05**	**B1 & Both ADC > SqCCa**
miR-6076	143	180	-1.35	3.05E-01	9.27E-01	**281**	**2012**	**-5.51**	**4.49E-05**	**0.0315**	**216**	**1253**	**-4.91**	**8.03E-07**	**0.0004**	**B2 & Both ADC > SqCCa**
miR-9-5p	135	98	1.28	4.46E-01	9.27E-01	**2899**	**297**	**12.68**	**1.27E-07**	**0.0001**	**1442**	**217**	**7.86**	**1.15E-07**	**0.0001**	**B2 & Both SqCCa > ADC**

adjP = FDR adjusted P, B1 = batch 1, B2 = batch 2

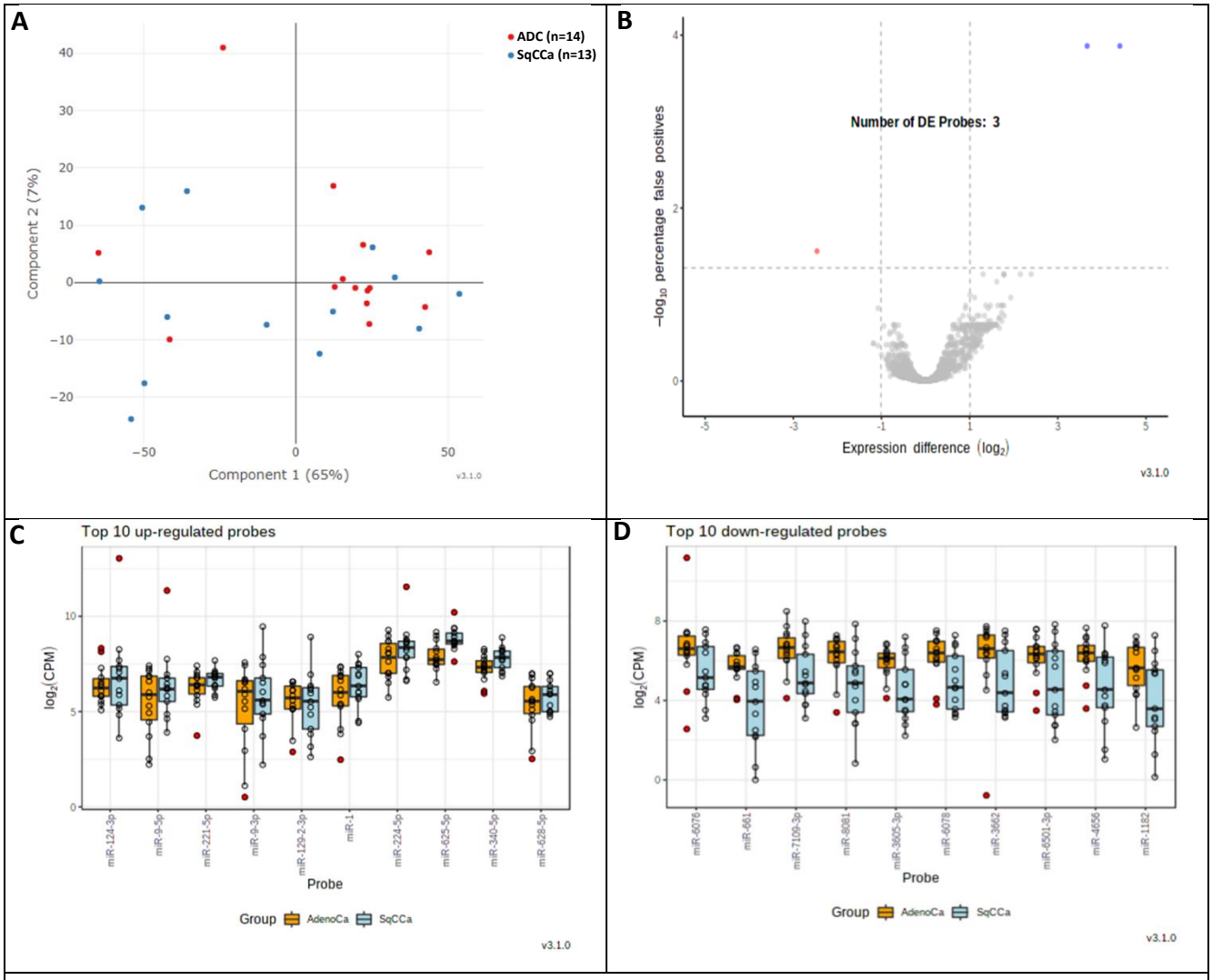


Figure 4.32. SqCCa *vs* ADC - PCA plot (A), volcano plot (B), top 10 probes higher in SqCCa (C) and top 10 probes higher in ADC (D) for batch 2. Excluding the QC failures, questionable controls, brain RNA controls and control samples.

For combined batch 1 and batch 2(Fig. 4.33), despite no clear pattern of clustering on the PCA plot (Fig. 4.33A), 5 miRNAs were significantly differentially expressed at an FDR adjusted P < 0.05 (Fig. 4.33B). There were 26 miRNAs where SqCCa > ADC at a raw P < 0.01, of which 3 had an FDR adjusted P < 0.05 (Table 4.9); bar graphs for the top 10 miRNAs higher in SqCCa are shown in Fig. 4.33C. There were 11 miRNAs where SqCCa < ADC at a raw P < 0.01, of which 2

had an FDR adjusted P < 0.05 (Table 4.9); bar graphs for the top 10 miRNAs higher in ADC are shown in Fig. 4.33D.

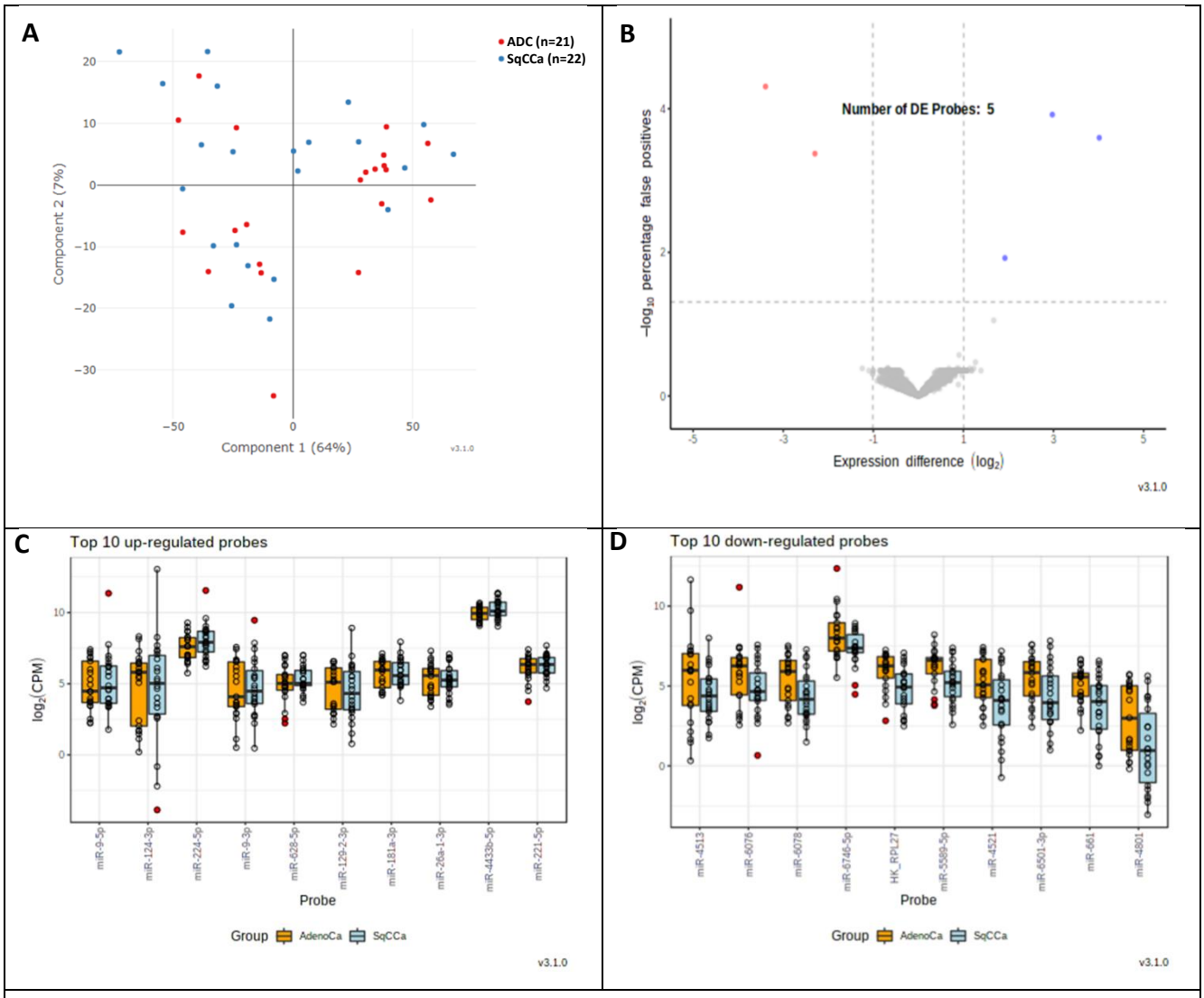


Figure 4.33. SqCCa *vs* ADC - PCA plot (A), volcano plot (B), top 10 probes higher in SqCCa (C) and top 10 probes higher in ADC (D) for the combination of batch 1 and batch 2. Excluding the QC failures, questionable controls, brain RNA controls and control samples.

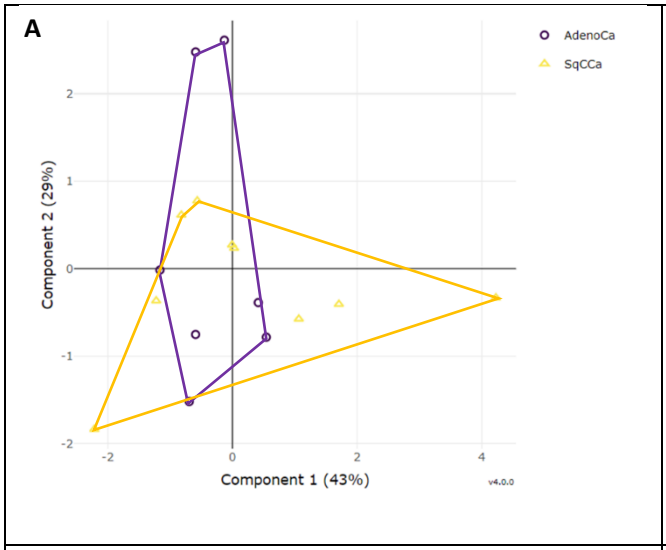
Comparison between differentially expressed miRNAs analysed by batch or combined are summarised in Table 4.9. No miRNA was consistently expressed at higher levels in SqCCa across all analyses at an FDR adjusted P < 0.05 and no miRNA consistently expressed at lower

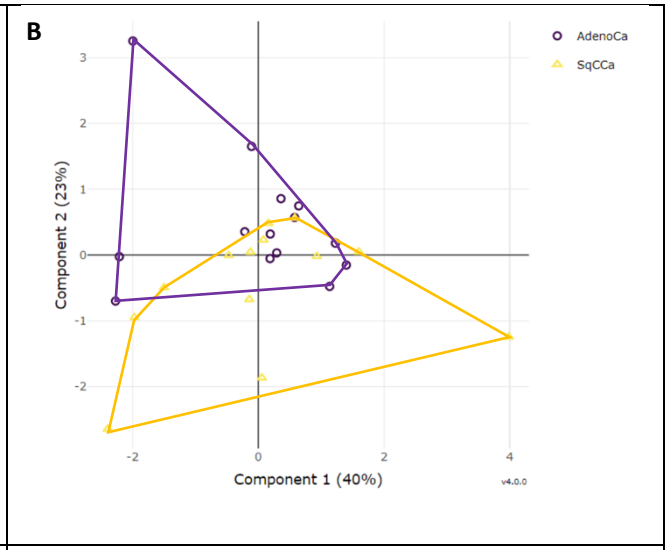
levels. Additionally, no miRNA was consistently differentially expressed by histological status across all analyses at a raw $P < 0.01$.

However, 1 miRNA (miR-4513) was higher in ADC (SqCCa < ADC) in both batch 1 and the combined at a raw $P < 0.01$ (raw $P = 1.99 \times 10^{-7}$ for batch 1, $P = 2.34 \times 10^{-8}$ for combined) and at an FDR adjusted $P < 0.05$ (FDR adjusted $P = 0.0004$ for batch 1, $P = 4.91 \times 10^{-5}$ for combined).

Additionally, 21 different miRNAs were higher in SqCCa (SqCCa > ADC) in both batch 2 and the combined at a raw $P < 0.01$, 2 of which were significantly higher in SqCCa at an FDR adjusted $P < 0.05$ in both batch 2 and the combined (miR-124-3p, FDR adjusted $P = 0.0001$ for batch 2, $P = 0.0003$ for combined; miR-9-5p, FDR adjusted $P = 0.0001$ for batch 2, $P = 0.0001$ for combined). Besides that, 3 miRNAs were higher in ADC (SqCCa < ADC) in both batch 2 and the combined at a raw $P < 0.01$, 1 of which (miR-6076) was significantly higher in ADC at an FDR adjusted $P < 0.05$ in both batch 2 and the combined ($P = 0.0315$ for batch 2, $P = 0.0004$ for combined).

When we limit the miRNAs to those with at least one FDR adjusted $P < 0.05$ (as per Table 4.9), there is only limited evidence of separation of ADC and SqCCa in batch 1 (Fig. 4.34A) or batch 2 (Fig. 4.34B). In the combined data, the clustering is dominated by the batches rather than by histology (Fig. 4.34C).





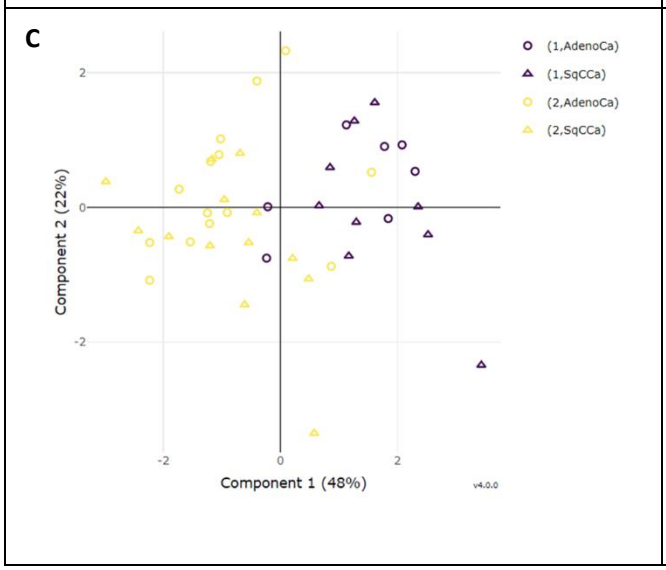


Figure 4.34. PCA plot based on 5 miRNAs with at least one FDR adjusted P < 0.05 when reanalysing HTG data for differential expression associated with histology, batch 1 (A), batch 2 (B) and the combined (C).

4.5.7 EV miRNA associated with multiple factors

Additionally, a few miRNAs were found to be associated with multiple factors (Table 4.10, Fig. 4.35). For example, miR-6076 is consistently up-regulated in both current smokers and those with COPD (Fig. 4.35A); whereas miR-375 is associated with lung cancer status and COPD, but in opposite directions (Fig. 4.35B), increased in COPD, but lower in cases.

Table 4.10. miRNAs associated with multiple factors

miRNA ID	Group	Batch 1					Batch 2					Both					Comparison Significant P values
COPD A = noCOPD Sex A = Female Status A = Control Smoking A = Former		Mean normalized A	Mean normalized B	Fold Change B *vs* A	Raw P B *vs* A	adjP B *vs* A	Mean normalized A	Mean normalized B	Fold Change B *vs* A	Raw P B *vs* A	adjP B *vs* A	Mean normalized A	Mean normalized B	Fold Change B *vs* A	Raw P B *vs* A	adjP B *vs* A	A < B = Blue; A > B = Red; significant at an adjusted P < 0.05 in bold
miR-215-5p	COPD	**47**	**172**	**3.38**	**0.0002**	**0.021**	135	211	1.55	0.0476	0.87	91	204	1.91	0.00063	0.063	**COPD > noCOPD B1 Both**
miR-215-5p	status	**160**	**44**	**-4.05**	**1.70E-06**	**0.0018**	161	189	-1.08	0.7036	0.999	165	123	-1.53	0.0263	0.4922	**Control > Case B1**
miR-375	COPD	**53**	**346**	**6.02**	**4.12E-05**	**0.012**	157	167	1.05	0.8243	0.9985	105	214	1.73	0.0141	0.259	**COPD > noCOPD B1**
miR-375	status	**308**	**52**	**-6.54**	**5.74E-06**	**0.004**	145	179	-1.03	0.895	0.999	195	122	-1.82	0.0059	0.2904	**Control > Case B1 Both**
miR-4449	COPD	**1001**	**3209**	**2.96**	**0.0007**	**0.039**	1699	1193	-1.44	0.1684	0.9985	1334	1657	1.06	0.7885	0.9186	**COPD > noCOPD B1**
miR-4449	status	**526**	**2118**	**3.64**	**0.0004**	**0.0474**	1597	1216	-1.67	0.046	0.829	1216	1650	1.19	0.415	0.878	**Case > Control B1**
miR-6076	COPD	164	132	-1.34	0.2325	0.6046	339	1067	3.12	0.0003	0.19	**257**	**832**	**2.76**	**1.86E-05**	**0.0089**	**COPD > noCOPD B2 Both**
miR-6076	Sex	114	168	1.62	0.031	0.236	302	947	2.70	0.0022	0.1309	**229**	**672**	**2.61**	**0.00029**	**0.0299**	**Male > Female B2 Both**
miR-6076	smoking	150	153	1.06	0.8094	0.9992	**352**	**1114**	**3.63**	**9.19E-06**	**0.0096**	**260**	**830**	**3.35**	**6.84E-08**	**0.00014**	**Current > Former B2 Both**
miR-3607-5p	Sex	92	105	1.26	0.235	0.582	**1173**	**337**	**-4.03**	**2.58E-07**	**3.60E-05**	**675**	**249**	**-3.05**	**6.01E-07**	**8.40E-05**	**Female > Male B2 Both**
miR-3607-5p	status	104	99	-1.17	0.447	0.8862	837	419	-2.54	0.001	0.176	**578**	**268**	**-2.45**	**0.00009**	**0.0273**	**Control > Case B2 Both**
miR-4513	Sex	2798	443	-5.76	0.006	0.123	259	357	1.19	0.4037	0.9999	**1334**	**424**	**-3.54**	**4.43E-05**	**0.0055**	**Female > Male B1 Both**
miR-4513	smoking	1705	137	-12.02	0.0024	0.9992	376	272	-1.21	0.3495	0.9986	**1185**	**236**	**-4.79**	**1.53E-06**	**0.0016**	**Former > Current B1 Both**
miR-4513	status	137	1733	11.41	0.0028	0.1921	329	309	-1.36	0.114	0.999	**266**	**1124**	**3.72**	**8.35E-05**	**0.0273**	**Case > Control B1 Both**
miR-671-5p	smoking	391	267	-1.41	0.2072	0.9992	**1303**	**319**	**-3.57**	**9.04E-06**	**0.0096**	**772**	**308**	**-2.39**	**2.55E-05**	**0.0134**	**Former > Current B2 Both**
miR-671-5p	status	234	409	1.58	0.0853	0.6439	**1078**	**421**	**-3.27**	**1.59E-05**	**0.0335**	732	434	-1.92	0.0014	0.1242	**Control > Case B2 Both**
miR-9-3p	COPD	**78**	**243**	**2.88**	**0.0004**	**0.0308**	1401	502	-2.82	0.008	0.48	717	445	-1.89	0.0506	0.4747	noCOPD > COPD B2 COPD > noCOPD B1
miR-9-3p	Sex	95	149	1.71	0.121	0.452	**2183**	**265**	**-9.54**	**1.48E-10**	**3.87E-08**	**1211**	**221**	**-6.16**	**1.82E-10**	**4.76E-08**	**Female > Male B2 Both**

adjP = FDR adjusted P, B1 = batch 1, B2 = batch 2

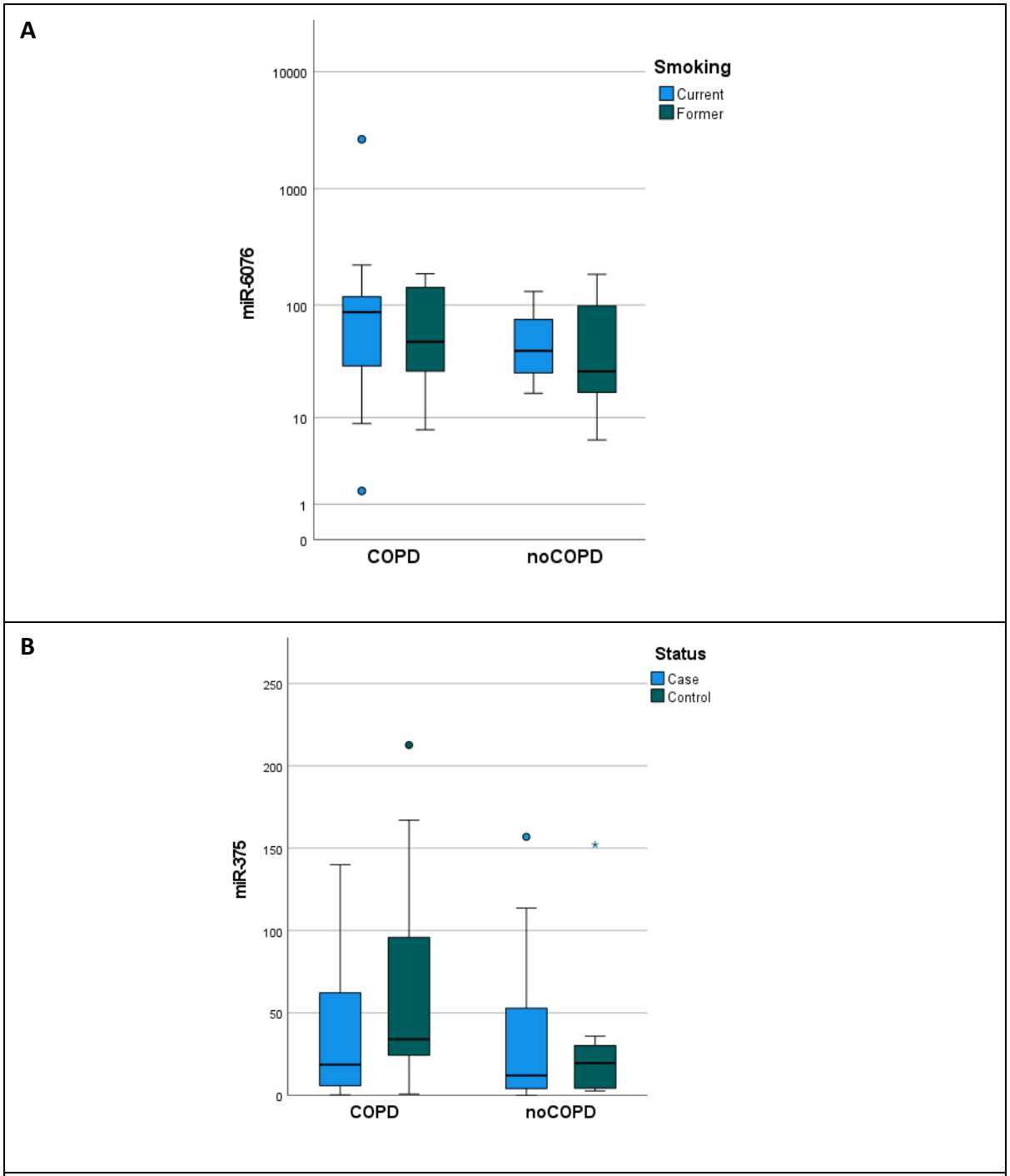


Figure 4.35. miRNAs associated with multiple factors. miR-6076 (A) associated with COPD and smoking status; miR-375 (B) associated with COPD and case status.

Chapter 5 Cross comparison between the four analysed datasets

Based on the preceding findings of HTG EV miRNA original analysis and reanalysis, as well as qPCR validation results of EV miRNA, we performed a comparison between the three analyses and additionally an HTG plasma miRNA dataset.

Comparisons between EV miRNA analyses aimed to address technical issues, such as differences in quality control and statistical approaches (HTG original analysis *vs* reanalysis) or methodologies, such as extraction techniques and assay method (HTG analysis *vs* qPCR).

In order to provide biological insight, in terms of which cancer-related biomarker miRNAs are enriched or depleted in the EV fraction, a comparison with plasma miRNA data (HTG) was performed. We also provide a more general overview of relative enrichment and depletion of miRNA in the EV fraction.

5.1 Comparison between the three EV miRNA analysed datasets

The qRT-PCR validation of EV miRNA was only partially successful, with 4 of the 18 miRNAs validated as being differentially expressed between lung cancer cases and controls, with a trend for 2 more (Chapter 3). However, additional EV miRNAs were differentially expressed when reanalysing the HTG data (Chapter 4), both between cases and controls and in relation to other factors, e.g. COPD, smoking. Here we examine these miRNAs in terms of fold-change and significance levels across the three EV miRNA datasets.

5.1.1 Comparison of qPCR validation to original HTG analysis

Given that the selection of EV miRNAs for validation was performed using lasso regression (which excludes highly correlated miRNAs and therefore not necessarily the most highly

differentially expressed miRNAs), we performed a comparison of fold-change and differential expression based on Mann Whitney test for the original HTG data and the qPCR data. Of note not all EV miRNAs selected were differentially expressed in the original data using this test (Table 5.1). In the original HTG data analysis, some EV miRNAs were down-regulated: miR-3149, miR-6872-5p, miR-1277-5p (Mann Whitney P < 0.01) and miR-1178-5p (Mann Whitney P < 0.05). Other EV miRNAs were up-regulated: miR-103a-3p, miR-107, miR-146b-5p and miR-301b (Mann Whitney P < 0.05). In contrast, all significantly differentially expressed EV miRNAs in the qPCR validation were up-regulated. This includes miR-3149 which was down-regulated in the original HTG analysis (Table 5.1).

Table 5.1. Selected miRNAs from original HTG analysis compared to qPCR quantitation

miRNA	Original HTG analysis		qPCR validation	
	Case (n) = 49 Control (n) = 48		Case (n) = 188 Control (n) = 187	
	FC	Mann-Whitney P values	FC	Mann-Whitney P values
let-7c-3p	-0.41	0.286	**1.86**	**0.007**
miR-26a-1-3p	0.59	0.128	1	0.981
miR-103a-3p	0.70	0.017	**5.1**	**0.001**
miR-107	0.64	0.030	**5.86**	**0.008**
miR-146b-5p	0.64	0.028	0.81	0.487
miR-301b	0.91	0.010	1.52	0.086 *
miR-342-5p	0.70	0.120	1.06	0.585
miR-514b-5p	-1.33	0.074	0.99	0.488
miR-1178-5p	-0.42	0.010	0.97	0.818
miR-1228-3p	0.18	0.223	1.29	0.057 *
miR-3149	**-1.84**	**0.002**	2.03	0.010
miR-6872-5p	**-0.92**	**< 0.001**	1	0.967
miR-1205	0.28	0.428	1.28	0.436
let-7a-3p	-0.41	0.377	1.1	0.539
miR-1185-5p	0.48	0.286	1.16	0.701
miR-1247-3p	-0.43	0.077	NA	0.687
miR-1277-5p	**-2.48**	**< 0.001**	NA	0.409
miR-5093	-0.23	0.247	NA	0.753

Note: FC= fold change; NA = not available; blue = up-regulated in cases; red = down-regulated in cases, coloured = P values < 0.05, bold = P values < 0.01; * trend in validation.

5.1.2 Comparison of qPCR validation to HTG reanalysis

In order to check if any of the EV miRNAs identified originally for validation by qPCR were differentially expressed in the HTG data reanalysis, we compared the fold-change values and FDR adjusted P values (Mann Whitney for the qPCR and EdgeR for HTG reanalysis, Table 5.2).

Only miR-3149 was significantly differentially expressed in both analyses, but was notably in a different direction (lower in cases in HTG reanalysis), consistent with the original HTG analysis (Table 5.1). One other miRNA (miR-1277-5p) was also significantly down-regulated in cases in HTG reanalysis, but only in batch 1 with FDR adjusted P < 0.05 and in combined analysis with raw P < 0.01.

Table 5.2. Selected miRNAs from original HTG analysis, comparing qPCR quantitation to HTG reanalysis

miRNA	qPCR validation		HTG reanalysis batch 1		HTG reanalysis batch 2		HTG reanalysis both	
	Case (n) = 188 Control (n) = 187		Case (n) = 16 Control (n) = 7		Case (n) = 27 Control (n) = 22		Case (n) = 43 Control (n) = 29	
	FC	Mann-Whitney P values	FC	FDR adjusted P values	FC	FDR adjusted P values	FC	FDR adjusted P values
let-7c-3p	**1.86**	**0.007**	1.12	0.9467	1.07	0.9991	1.02	0.9875
miR-26a-1-3p	1	0.981	1.42	0.8081	1.33	0.7075	1.34	0.4497
miR-103a-3p	**5.1**	**0.001**	1.42	0.8243	1.42	0.9991	1.5	0.748
miR-107	**5.86**	**0.008**	1.37	0.8523	1.39	0.9991	1.49	0.7698
miR-146b-5p	0.81	0.487	1.48	0.7948	1.59	0.9991	1.62	0.4737
miR-301b	1.52	0.086 *	1.41	0.8332	1.35	0.9991	1.45	0.6152
miR-342-5p	1.06	0.585	1.62	0.645	1.08	0.9991	1.34	0.6416
miR-514b-5p	0.99	0.488	-2.43	0.6169	-1.1	0.9991	-1.32	0.8714
miR-1178-5p	0.97	0.818	-1.16	0.8662	-1.39	0.9991	-1.37	0.6082
miR-1228-3p	1.29	0.057 *	1.29	0.7948	1.07	0.9991	1.09	0.8927
miR-3149	2.03	0.010	**-4.35**	**0.0004**	-3.47	0.0611	**-3.75**	**0.0002**
miR-6872-5p	1	0.967	-1.41	0.6896	-1.34	0.9991	-1.44	0.6152
miR-1205	1.28	0.436	1.16	0.9072	1.14	0.9991	1.04	0.9625
let-7a-3p	1.1	0.539	-1.03	0.98	-1.02	0.9991	-1.07	0.9361
miR-1185-5p	1.16	0.701	1.76	0.645	1.13	0.9991	1.23	0.8278
miR-1247-3p	NA	0.687	1.21	0.8831	-1.09	0.9991	1.59	0.813
miR-1277-5p	NA	0.409	**-3.71**	**0.0445**	-1.79	0.7633	-2.13	0.1242
miR-5093	NA	0.753	1.15	0.9081	1.07	0.9991	1.03	0.9714

Note: FC= fold change; NA = not available; blue = up-regulated in cases; red = down-regulated in cases. For qPCR: coloured = P values < 0.05, bold = P values < 0.01. For HTG reanalysis: coloured = raw P < 0.01; bold = FDR adjusted P < 0.05. * trend in validation.

5.1.3 Comparison of original HTG analysis to HTG reanalysis

In order to check if any of the EV miRNAs identified by either the original HTG analysis or the HTG reanalysis were also differentially expressed in both sets of HTG data, we compared the fold-change values and FDR adjusted P values (Mann Whitney for the original analysis and from EdgeR for HTG reanalysis, Table 5.3).

Amongst EV miRNAs selected for qPCR validation, both miR-3149 and miR-1277-5p were significantly differentially expressed in both analyses, miR-3149 more consistently across different batches.

Table 5.3. Selected miRNAs from original HTG analysis or from HTG reanalysis, comparing both sets of HTG data

	miRNA	Original HTG		HTG reanalysis batch 1		HTG reanalysis batch 2		HTG reanalysis both	
		Case (n) = 49 Control (n) = 48		Case (n) = 16 Control (n) = 7		Case (n) = 27 Control (n) = 22		Case (n) = 43 Control (n) = 29	
		FC	Mann-Whitney P values	FC	FDR adjusted P values	FC	FDR adjusted P values	FC	FDR adjusted P values
Selected for qPCR validation from original HTG analysis	let-7c-3p	-0.41	0.286	1.12	0.9467	1.07	0.9991	1.02	0.9875
	miR-26a-1-3p	0.59	0.128	1.42	0.8081	1.33	0.7075	1.34	0.4497
	miR-103a-3p	0.70	0.017	1.42	0.8243	1.42	0.9991	1.5	0.748
	miR-107	0.64	0.030	1.37	0.8523	1.39	0.9991	1.49	0.7698
	miR-146b-5p	0.64	0.028	1.48	0.7948	1.59	0.9991	1.62	0.4737
	miR-301b	0.91	0.010	1.41	0.8332	1.35	0.9991	1.45	0.6152
	miR-342-5p	0.70	0.120	1.62	0.645	1.08	0.9991	1.34	0.6416
	miR-514b-5p	-1.33	0.074	-2.43	0.6169	-1.1	0.9991	-1.32	0.8714
	miR-1178-5p	-0.42	0.010	-1.16	0.8662	-1.39	0.9991	-1.37	0.6082
	miR-1228-3p	0.18	0.223	1.29	0.7948	1.07	0.9991	1.09	0.8927
	miR-3149	**-1.84**	**0.002**	**-4.35**	**0.0004**	-3.47	0.0611	**-3.75**	**0.0002**
	miR-6872-5p	**-0.92**	**< 0.001**	-1.41	0.6896	-1.34	0.9991	-1.44	0.6152
	miR-1205	0.28	0.428	1.16	0.9072	1.14	0.9991	1.04	0.9625
	let-7a-3p	-0.41	0.377	-1.03	0.98	-1.02	0.9991	-1.07	0.9361
	miR-1185-5p	0.48	0.286	1.76	0.645	1.13	0.9991	1.23	0.8278
	miR-1247-3p	-0.43	0.077	1.21	0.8831	-1.09	0.9991	1.59	0.813
	miR-1277-5p	**-2.48**	**< 0.001**	**-3.71**	**0.0445**	-1.79	0.7633	-2.13	0.1242
	miR-5093	-0.23	0.247	1.15	0.9081	1.07	0.9991	1.03	0.9714
Identified by HTG data reanalysis	miR-6124	**1.45**	**< 0.001**	**4.5**	**0.0086**	-1.42	0.9991	-1.07	0.9509
	miR-1290	1.19	0.026	**4.16**	**0.0327**	-1.39	0.9991	-1.01	0.9972
	miR-4534	**1.02**	**0.001**	**3.08**	**0.0327**	-1.31	0.9991	-1.07	0.9431
	miR-4644	0.64	0.038	**2.92**	**0.0327**	-1.54	0.8288	-1.13	0.8927
	miR-5196-5p	**0.94**	**0.001**	**2.96**	**0.0327**	-1.28	0.9991	1.26	0.8278
	miR-6086	**0.88**	**0.002**	**2.7**	**0.0327**	-1.32	0.9991	1.02	0.9875
	miR-6870-5p	**0.92**	**0.004**	**3.47**	**0.0327**	-1.09	0.9991	1.25	0.8496
	miR-7111-5p	**0.79**	**0.001**	**2.68**	**0.0332**	-1.3	0.9991	1.08	0.9327
	miR-4484	1.37	0.021	**6.21**	**0.0334**	-1.29	0.9991	1.27	0.8726
	miR-3175	0.66	0.030	**2.72**	**0.0474**	-1.03	0.9991	1.39	0.4922
	miR-4449	0.85	0.033	**3.64**	**0.0474**	-1.67	0.8288	1.19	0.8784
	miR-215-5p	-0.70	0.010	**-4.05**	**0.0018**	-1.08	0.9991	-1.53	0.4922
	miR-183-5p	-0.75	0.048	**-2.9**	**0.0327**	1.34	0.9991	-1.32	0.8256
	miR-34a-5p	0.20	0.077	**-3.12**	**0.0327**	1.28	0.9991	-1.27	0.8403
	miR-375	-0.76	0.670	**-6.54**	**0.004**	-1.03	0.9991	-1.82	0.2904
	miR-671-5p	-0.19	0.300	1.58	0.6439	**-3.27**	**0.0335**	-1.92	0.1242
	miR-6877-5p	0.71	0.048	3.21	0.1273	1.15	0.9991	**2.2**	**0.0253**
	miR-4513	-0.36	0.259	11.41	0.1921	-1.36	0.9991	**3.72**	**0.0273**
	miR-6750-5p	-0.25	0.201	-1.35	0.7962	-2.55	0.1343	**-2.09**	**0.0448**
	miR-6080	-0.66	0.059	-1.3	0.8819	-3.18	0.0611	**-3**	**0.018**
	miR-4519	-0.17	0.201	1.08	0.9289	-2.31	0.0611	**-1.88**	**0.0253**
	miR-4522	-0.34	0.135	1.03	0.9848	-3.07	0.0611	**-2.42**	**0.018**
	miR-3607-5p	-0.06	0.903	-1.17	0.8862	-2.54	0.1761	**-2.45**	**0.0273**

Note: FC= fold change; blue = up-regulated in cases; red = down-regulated in cases. For HTG original analysis: coloured = P values < 0.05, bold = P values < 0.01. For HTG reanalysis: coloured = raw P < 0.01; bold = FDR adjusted P < 0.05.

Amongst EV miRNAs identified by HTG reanalysis, there was a closer relationship between fold change in the original analysis and batch 1 reanalysis than for batch 2 or combined reanalyses. Of the 15 miRNAs differential expressed at an FDR adjusted $P < 0.05$ in batch 1, all but 2 were differentially expressed in the original HTG data (Mann Whitney $P < 0.05$), including 6 with $P < 0.01$; for all 13 differentially expressed miRNAs in both, the direction of the fold change was the same and there was a good correlation (Fig. 5.1). The direction of the fold change was also the same for 7 of the 8 miRNAs identified in HTG reanalysis in batch 2 or combined batches only.

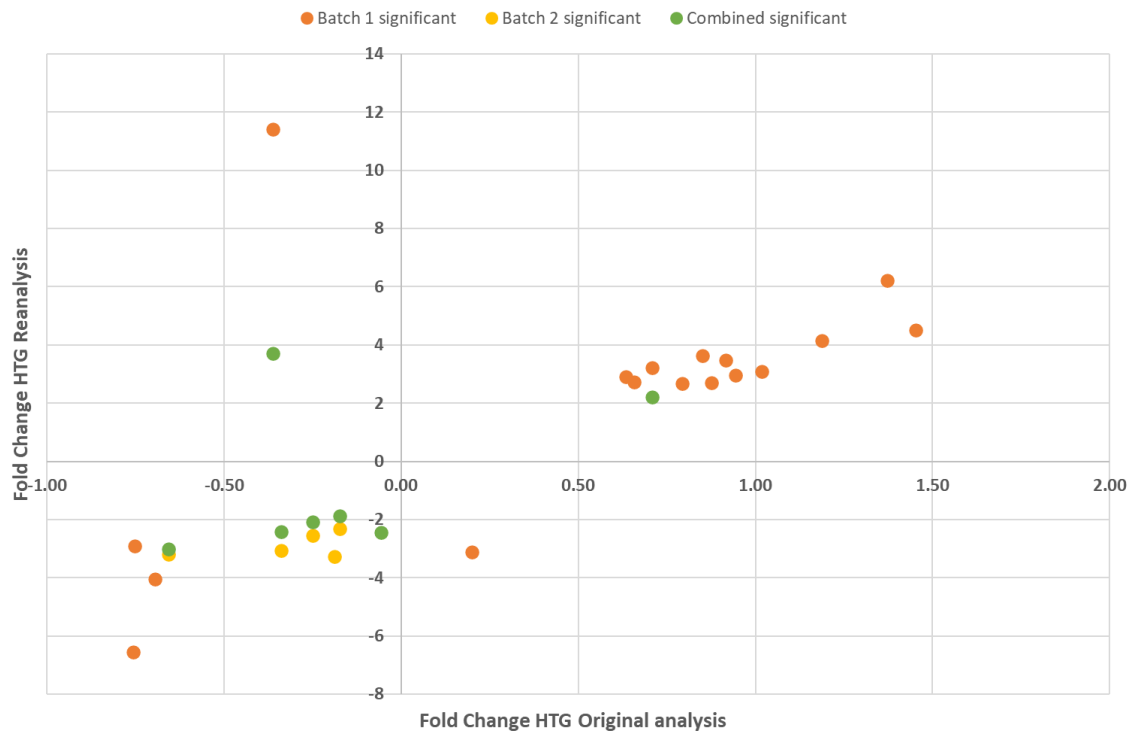


Figure 5.1. Scatterplot of significantly differentially expressed miRNAs in batch 1, batch 2 and the combination of batch 1 and 2 in HTG reanalysis, comparing fold change in the original HTG analysis to fold change in the HTG reanalysis.

5.2 Plasma miRNA HTG EdgeSeq analysis

The HTG EdgeSeq miRNA WTA was used to measure the expression of 2,083 human miRNA transcripts (plus 13 HK genes) from plasma samples of 26 NSCLC cases and 24 controls in three batches which passed the QC (using QC0, QC1 and QC2, as per the EV miRNA data reanalysis). Cases and controls were selected from the Liverpool Lung Project and matched on age, sex, and smoking status. The histologic types of the cases were ADC and SqCCa, and most of them were early stage (Table 5.4). Ages ranged from 51 to 81 (mean = 67) and 58% were male, all were smokers (mean smoking duration was 46 years for cases and 42 years for controls). The volcano plot (Fig. 5.2A) demonstrates the differential expression between the cases and the controls, and bar graphs for the top 10 up- and down-regulated miRNAs are shown in Fig. 5.2B&C. There were 169 miRNAs significantly over-expressed in cases compared to controls (Fig. 5.2A, adjusted $P < 0.05$), compared to 13 miRNAs for EV miRNA (HTG reanalysis, section 4.5.2), only two miRNAs (miR-4449 and miR-4484) were detected at significantly higher levels in both plasma and EVs in cases. There were 13 miRNAs significantly under-expressed in cases compared to controls in plasma (Fig. 5.2A, adjusted $P < 0.05$), compared to 12 miRNAs for EV miRNA (HTG reanalysis, Section 4.5.2), with only miR-3149 significantly under-expressed in cases in both.

Table 5.4. Post-QC NSCLC cases by histology and stage

	Stage							Total
Histology	**IA**	**IB**	**IIA**	**IIB**	**IIIA**	**IIIB**	**IV**	
ADC	5	1	2	1	2	1	2	14
SqCCa	1	0	6	3	1	1	0	12
Total	6	1	8	4	3	2	2	26

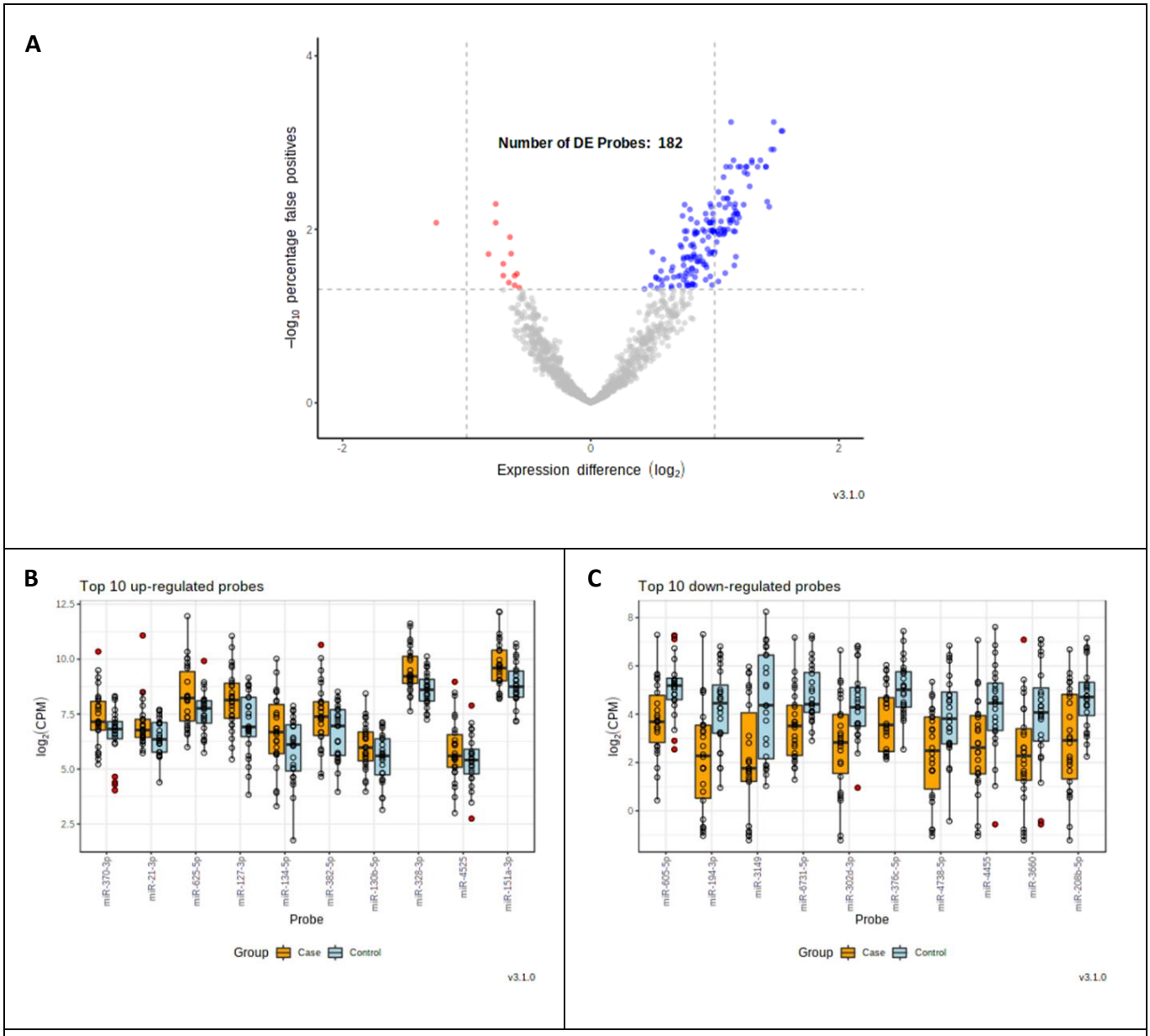


Figure 5.2. Significant differentially expressed plasma miRNAs by case or control status - volcano plot (A), top 10 up-regulated probes (B) and top 10 down-regulated probes (C).

5.3 Comparison of EV miRNA and plasma miRNA

When comparing miRNA expression between plasma and EV fractions, we must remember that some miRNAs exist only in the extracellular non-EV fraction and some are restricted to EVs, but that many will be present in both (we found 1895 miRNAs at 10 cpm or more in plasma and EV fractions). Furthermore, the plasma miRNA extraction using lysis reagent will include EV miRNAs. Nevertheless, by comparing levels of each miRNA in plasma and EV, we

can determine levels of enrichment/depletion, which may be related to biological function (e.g. whether the miRNA is actively involved in EV mediated signalling).

5.3.1 Differences between plasma and EV miRNA levels: dynamic range, enrichment and depletion

In comparing EV miRNA cpm data from the original HTG analysis and plasma, we can plot the counts per million (cpm) for each miRNA separately for controls only (Fig. 5.3), in order to investigate enrichment or depletion of the miRNAs in the EV fraction.

It should be remembered that given the lower absolute quantity of miRNA in the EV fraction and the smaller number of miRNAs found in EVs, compared to whole plasma, the cpm for any miRNA present in the EV fraction would be expected to be higher than seen in plasma (unless it is significantly depleted in the EV fraction). As shown in Fig. 5.3, for those miRNAs present in both plasma and the EV fraction, there was a good correlation overall between plasma and EV (Spearman R 0.816, $P < 10^{-6}$; $R^2 = 0.659$) but cpm were higher in EV, indicated by the fact that the line of best fit was offset upwards of a 1:1 relationship. A significant proportion of the miRNAs were above the upper 95% confidence interval (enriched in EV) or below the lower 95% confidence interval (depleted in EV). A similar pattern was seen for comparison of cases (Spearman R 0.849, $P < 10^{-6}$; $R^2 = 0.705$).

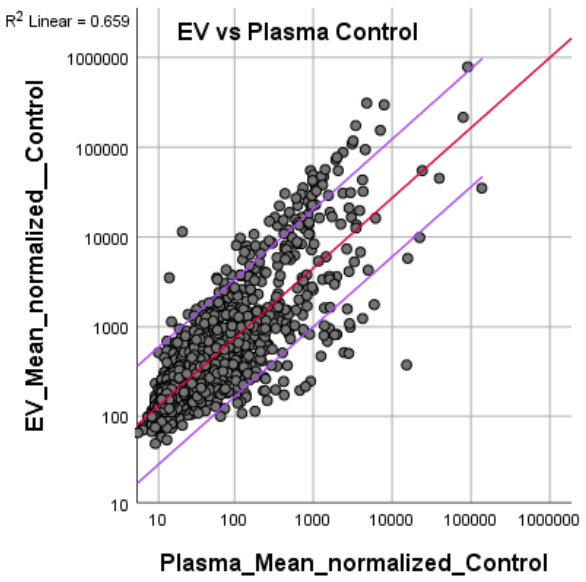


Figure 5.3. Comparison of mean plasma cpm *vs* mean EV cpm in control samples.

5.3.2 Differences between plasma and EV miRNA levels: case *vs* control

Plasma comparison between cases and controls (Fig 5.4A) indicated that apparently there was a greater proportion of differentially expressed miRNAs (i.e., outside of the 95% confidence interval for correlation between cases and controls, with a lower R^2 value) than seen for EV miRNA (Fig. 5.4B), especially considering up-regulation in cases compared to controls, in keeping with Fig. 5.2A. However, overall there was a very strong correlation between cases and controls in both EV (Spearman R 0.964, $P < 10^{-6}$; $R^2 = 0.978$) and plasma (Spearman R 0.951, $P < 10^{-6}$; $R^2 = 0.956$).

Of the 18 miRNAs originally selected for qRT-PCR validation, 5 were significantly differentially expressed in plasma at FDR adjusted $p < 0.05$. For all but 4 of these 18 miRNAs, the directionality of differential expression was the same in the original HTG analysis as in the plasma (3 up-regulated and 1 down-regulated).

Of the 17 miRNAs differentially expressed at FDR adjusted p < 0.05 in batch 1 of HTG reanalysis, 3 were significantly differentially expressed in plasma (2 up-regulated and 1 down-regulated). For all but 5 of these 17 miRNAs, the directionality of differential expression was the same in the batch 1 HTG reanalysis as in the plasma.

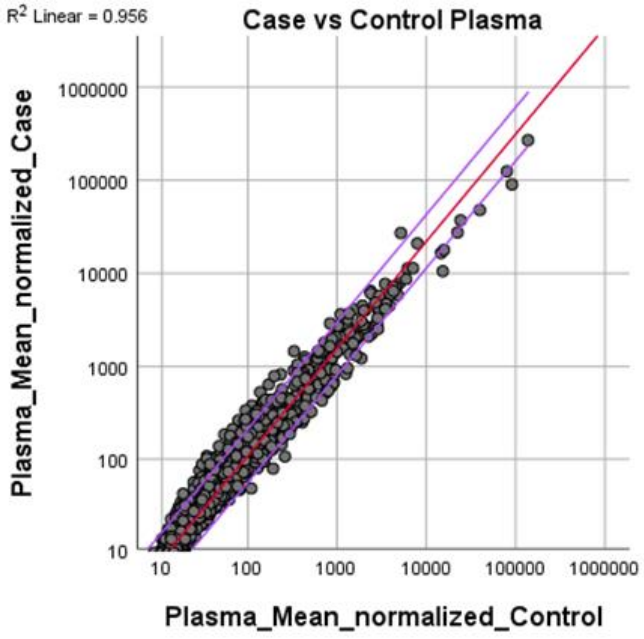
As shown in Table 5.5, of the 8 EV miRNAs differentially expressed when reanalysed both batch 1 and 2, only 1 miRNA was significantly differentially expressed in plasma, but all were expressed in the same direction in both EV and plasma (2 higher in cases and 6 higher in controls). Most were enriched in EV both in cases and controls (with > 2-fold enrichment for all but 2 for cases and all but 1 for controls). Enrichment was especially high for miR- 3149 (> 25-fold) and miR-4513 (45 fold for cases and 10.6 for controls).

Table 5.5. Enrichment and depletion of miRNA in the EV fraction compared to plasma, for the 8 most significantly differentially expressed EV miRNAs in HTG reanalysis of combined batch 1 and batch 2. Enrichment: EV/plasma > 1; depletion: EV/plasma < 1.

	HTG Reanalysis both		**Plasma**		**HTG Reanalysis both**		**Plasma**		**Enriched/depleted**	
	Case (n) = 43 Control (n) = 29		Case (n) = 26 Control (n) = 24		Mean cpm	Mean cpm	Mean cpm	Mean cpm	Ratio EV/plasma	
	FC	P	FC	P	Case	Control	Case	Control	Case	Control
miR-3149	**-3.75**	**0.0002**	**-2.37**	**0.0084**	**194**	**639**	**7**	**24**	27.7	26.6
miR-6877-5p	**2.2**	**0.0253**	1.4	0.3296	**939**	**375**	725	470	1.3	0.8
miR-4513	**3.72**	**0.0273**	1.03	NA	**1124**	**266**	25	25	45.0	10.6
miR-6750-5p	**-2.09**	**0.0448**	-1.31	0.4684	**307**	**565**	170	244	1.8	2.3
miR-6080	**-3**	**0.018**	-1.14	0.7524	**146**	**385**	72	86	2.0	4.5
miR-4519	**-1.88**	**0.0253**	-1.16	0.6931	**348**	**575**	84	101	4.1	5.7
miR-4522	**-2.42**	**0.018**	-1.16	0.7079	**212**	**450**	108	132	2.0	3.4
miR-3607-5p	**-2.45**	**0.0273**	1.09	0.833	**268**	**578**	25	23	10.7	25.1

Note: FC= fold change; NA = not available; blue = up-regulated in cases; red = down-regulated in cases, bold = FDR adjusted P values < 0.05.

A



B

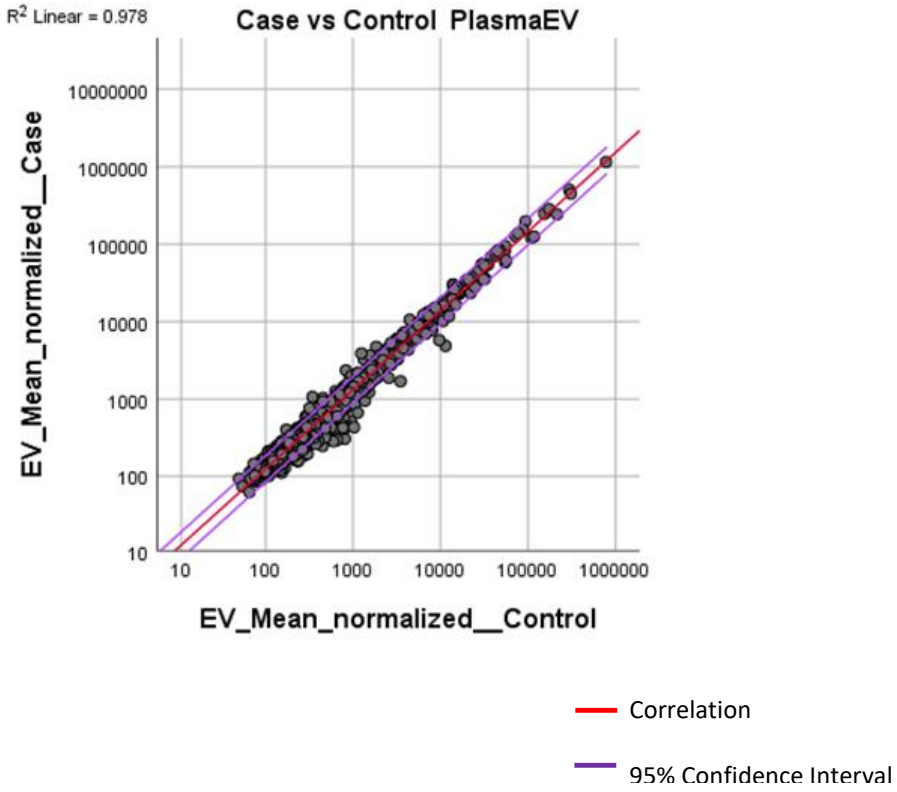


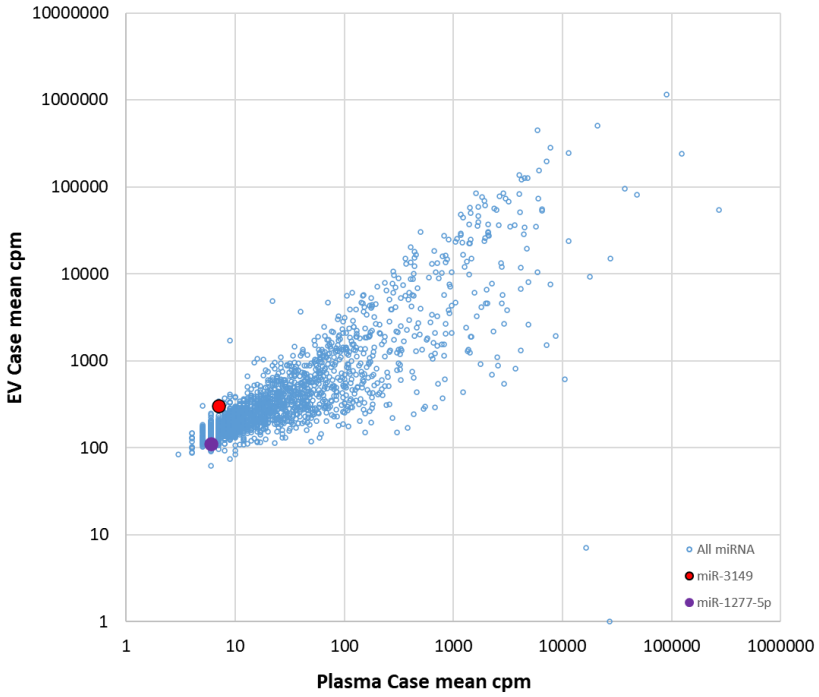
Figure 5.4. Comparison of mean miRNA levels between cases and controls in plasma (A) and the EV fraction (B).

If we look specifically at some examples of differentially expressed miRNAs:

- miR-3149 was down-regulated in cases in both EV HTG original analysis and plasma, but the mean cpm values were 306 for case and 807 for control in EV (FC = -3.42, FDR adjusted P = 0.07), compared to 7 for case and 24 for control in plasma (FC = -2.37, FDR adjusted P = 0.008). This indicates some enrichment in the EV fraction (Fig. 5.5).
- miR-1277-5p was down-regulated in cases in both EV HTG original analysis and plasma to approximately the same extent, but the mean cpm values were 111 for case and 153 for control in EV (FC = -1.78, raw P = 0.028), compared to 6 for case and 12 for control in plasma (FC = -1.74, raw P = 0.028). This indicates some enrichment in the EV fraction (Fig. 5.5).
- Apart from miR-3149, there were 3 miRNAs significantly differentially expressed in EV qPCR validation at P < 0.05 (let-7c-3p, P = 0.007; miR-103a-3p, p= 0.001; miR-107, p = 0.008). Among these, miR-103a-3p (FDR adjusted P = 0.018) and miR-107 (FDR adjusted P = 0.026) were also significantly differentially expressed in plasma in the same direction (up-regulated in cases), albeit with a lower fold change (5.1 & 5.9 respectively for EV by qPCR and 2.0 & 1.9 for plasma by HTG). As mentioned above, miRNA-3149 was significantly differentially expressed in EV qPCR validation (up-regulated) and plasma HTG analysis (down-regulated) with opposite directionality.

In order to investigate if the 18 EV miRNAs originally selected for qPCR validation, or as part of the 4-miRNA panel, were informative in plasma miRNA analysis, we produced PCA plots for plasma HTG data in REVEAL software (Fig. 5.6). There was some separation between cases and controls, indicating that the EV plasma signature is to some extent present within the total miRNA expression patterns.

A



B

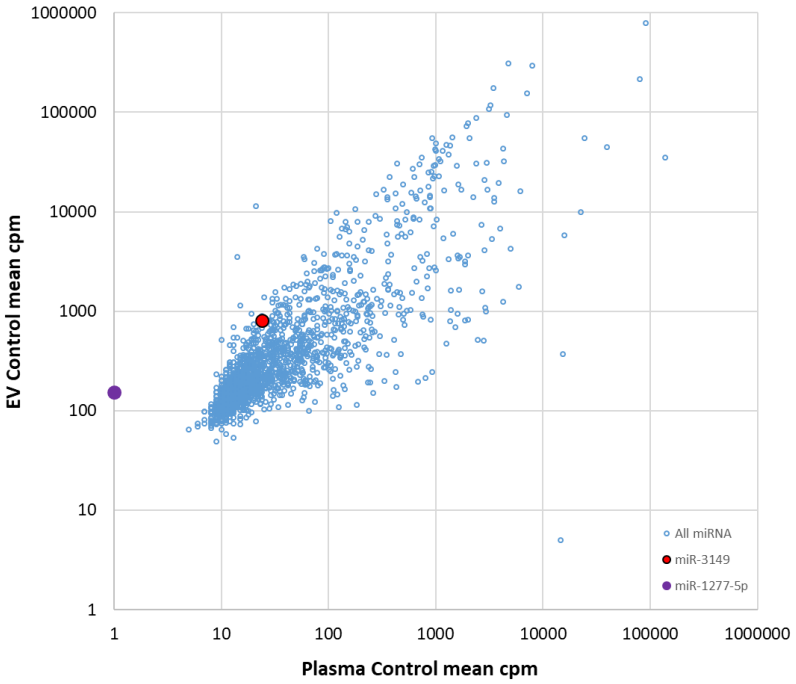
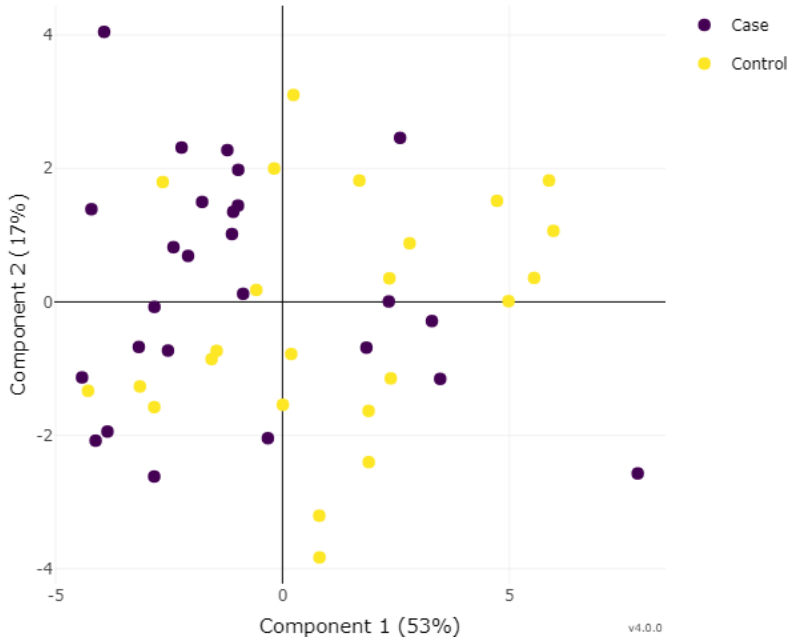


Figure 5.5. Mean cpm values of miR-3149 (red) and miR-1277-5p (purple) for cases (A) and controls (B) in plasma and the EV fraction.

A



B

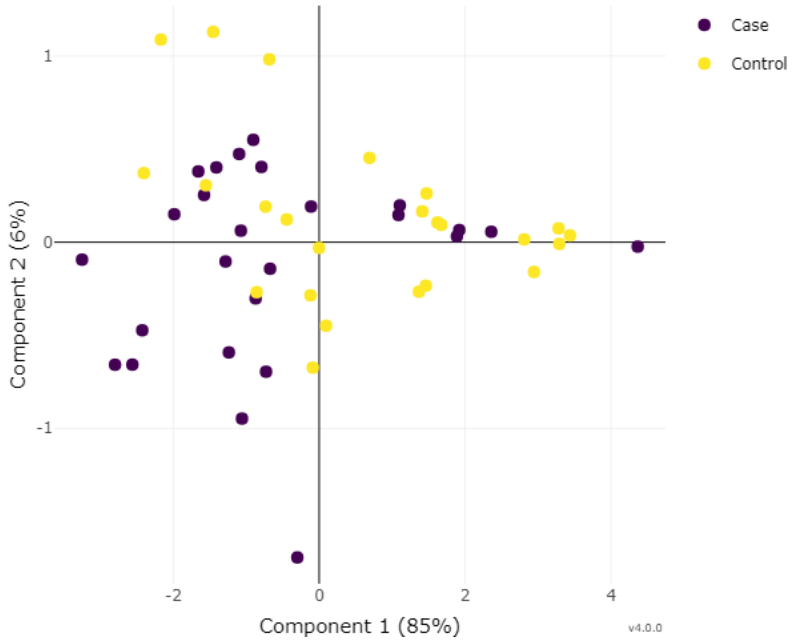


Figure 5.6. PCA plots of plasma HTG analysis, for 18 miRNAs selected for qPCR validation (A) and the 4-miRNA panel (B).

Chapter 6 Discussion

Extracellular vesicles (EVs) are implicated in cell-cell communication as shuttles for transferring a variety of cargo molecules (including miRNAs) from the donor cells to the recipient cells (Smith, Wajnberg et al. 2020). They are secreted by almost all cells and are widely present in diverse biofluids, such as blood and its components (Caby, Lankar et al. 2005). Tumour-derived extracellular vesicles (TEVs) have been reported to play a pivotal role in the onset, progression, invasion and metastasis of many cancers, including lung cancer (Frisch, Angenendt et al. 2019, Hasan, Sohal et al. 2022, Jahan, Mukherjee et al. 2022).

miRNAs represent an important mechanism of epigenetic control of gene expression. Due to their small size, encapsulation in EVs or binding to proteins, they can evade degradation by RNases and therefore can survive longer within various biological fluids. As a consequence, EV miRNAs are firmly associated with a wide variety of pathologies, such as Inflammation (Xie, Hou et al. 2019), kidney diseases (Zhang, Yi et al. 2020) and malignancies (Vu, Gong et al. 2020), and present as appealing biomarkers in medical research. A number of studies have shown that some EV miRNAs may have an early diagnostic value in distinguishing lung cancer from control subjects via various EV miRNA isolation and detection approaches (Cristiano, Leal et al. 2019).

Early lung cancer detection is a long-standing unmet healthcare need, that might benefit from integrating minimally invasive biomarker candidates to complement LDCT-based screening. However, so far, no miRNA (or other) biomarker has been clinically validated as a surrogate marker in early lung cancer diagnosis, (Sayyed, Gondaliya et al. 2022). One of the hypotheses in this PhD study is that EV miRNAs may provide sensitive and specific biomarkers for lung

cancer, due to their biological role as functional messengers as well as their abundance in biological fluids.

EV miRNAs as lung cancer biomarkers, validation by qRT-PCR

This project aimed to develop and validate a panel of EV miRNAs as diagnostic biomarkers for early detection of lung cancer. In the discovery phase, preceding my PhD, EVs were isolated from plasma using a standard ultracentrifugation protocol. Eighteen EV miRNAs were selected from the original analysis of the HTG EdgeSeq quantitation data of 2,083 miRNAs in 97 plasma samples (49 cases and 48 age/sex matched controls) based on the bioinformatic analysis using Mann-Whitney U test. Additionally, 4 miRNAs (miR-1277-5p, miR-509-3-5p, miR-514b-5p, miR-582-5p) were indicated with a high potential of discriminating lung cancer cases from individuals with no malignancy, based on model building for generalised linear models using Lasso penalized logistic regression and internal validation by ROC analysis with leave-one-out cross-validation. This comprehensive miRNA screening of EVs from lung cancer patients and controls provided an important profile analysis and suggested miRNA targets which may be used in early diagnosis. However, there was a need to validate these findings, using technology more easily applied in a clinical setting, avoiding ultracentrifugation in favour of column-based isolation of EV miRNA and using cheaper individual assays rather than a high throughput “omics” approach. The first objective in my PhD project was thus to validate the findings of the discovery phase in a large-scale cohort using a different methodological approach to test the hypothesis that plasma EV miRNA could be used for lung cancer diagnosis. In this validation phase, the exoEasy Midi Kit was used to isolate plasma EVs from 188 lung cancer cases and 187 age/sex matched controls; following RNA isolation and reverse transcription, qPCR, with TaqMan Advanced miRNA assays, was used to quantitate 18

preselected miRNAs. These miRNAs included only two of the miRNAs in the 4-miRNA panel, as a qPCR assay was not available for one and the wrong assay (as explained in Section 3.3) was used for another; hence it was not possible to validate this panel fully by qRT-PCR.

Mann-Whitney U test and ROC curve analysis were utilised to analyse the differentially expressed miRNAs in cases or controls. In this qPCR validation cohort, all 18 miRNAs were up-regulated in cases and 4 of them (miR-103a-3p, let-7c-3p, miR-107 and miR-3149) were significantly differentially expressed with Mann-Whitney P values ≤ 0.01. Furthermore, expression of many miRNAs was shown to correlate with each other, with Spearman correlation coefficient (rho) ranging from 0.91 (miR-107 vs miR-103a-3p) to 0.17 (miR-107 vs miR-3149). Some EV miRNAs were associated with COPD, but none with smoking or histology.

miR-103a-3p has been found to be up-regulated in many cancers. In both lung cancer (Li, Huhe et al. 2021) and gastric cancer (Hu, Miao et al. 2018), levels were higher in cancer tissue than normal tissue and a role in cancer progression was indicated by cell line gain- and loss-of-function studies. Conversely, a study by O'Farrell et al. indicated that plasma EV miRNA miR-103a-3p was significantly under-expressed in lung cancer participants compared to both healthy non-smokers and healthy smokers (O'Farrell, Bowman et al. 2021). Contradictions, dependent on the sample type examined, have also been seen in mesothelioma, where higher levels of miR-103a-3p in the cellular fraction of blood samples were suggested as a biomarker for malignant mesothelioma (Weber, Casjens et al. 2014). In contrast, a study by Cavalleri et al. measured plasma EV miRNAs in subjects with malignant pleural mesothelioma (MPM) and past asbestos exposure (PAE) using ultracentrifugation and miRNeasy purification kit for EV miRNA isolation from plasma, and QuantStudio 12K Flex OpenArray Real-Time PCR System for the assessment of miRNA profiling in the discovery phase and Custom TaqMan™ Low

Density Array was used in validation phase. miR-103a-3p was shown to be down-regulated in MPM patients compared to PAE subjects. This study indicated that EV miRNA miR-103a-3p could be used as biomarker to discriminate MPM patients from non-cancerous individuals exposed to asbestos fibres based on ROC curve analysis (Cavalleri, Angelici et al. 2017). Apparently, there are some similarities between this study and our qPCR cohort in terms of techniques been used, but ultracentrifugation for EV isolation was used instead of extraction approaches used in our qPCR validation cohort. miR-103a-3p has also been reported to be involved in various inflammatory responses, such as sepsis-induced acute kidney injury (Ding, An et al. 2022).

There were only two miRNAs that overlap between the selected 18 miRNAs for qPCR validation from the original EV miRNA HTG analysis and our EV miRNA HTG reanalysis. The reason for that is, as demonstrated in the HTG reanalysis, miRNAs differentially expressed between cases and controls are closely correlated, so were removed in Lasso penalized logistic regression applied for the original HTG analysis. That the Lasso-based selection is responsible for the minimal overlap is further demonstrated by the fact that miRNAs selected in the HTG reanalysis are differentially expressed in the original analysis (at least for batch 1, Table 5.3). Future work could include Lasso, as part of a new HTG analysis, taking into account improved QC metrics and batch effects.

In the original EV miRNA HTG EdgeSeq analysis, four miRNAs (miR-1277-5p, miR-509-3-5p, miR-514b-5p, miR-582-5p) were indicated as the best predicting model of lung cancer. As part of the HTG reanalysis, we were able to demonstrate that the 4-miRNA panel did provide a significant degree of discrimination between cases and controls (Fig. 4.5B), again more so for batch 1.

One of the 2 miRNAs (miR-1277-5p and miR-514b-5p) included in the qPCR validation that were significantly differentially expressed in HTG reanalysis was miR-1277-5p; this was significantly down-regulated in batch 1 at an FDR < 0.05, the combined batch 1 and 2 at a raw P = 0.0013 and batch 2 at raw P = 0.033). As discussed previously, this miRNA was also associated with COPD status.

The other qPCR-validated miRNA, miR-3149, is particularly noteworthy as it was significantly differentially expressed in all analyses of case vs. control. However, whereas it was down-regulated in cases in HTG original and re-analysis, it was up-regulated in the qPCR analysis. The main differences between the two types of analysis are the isolation method and the quantitation method. Different isolation methods could produce different profiles of EVs, for example, different proportions of exosomes and microvesicles, and it is known that some miRNAs are more abundant in one type of EV than another. For instance, miR-3149 is highly expressed in fibroblasts (MVs) (bioinfo.life.hust.edu.cn/EVmiRNA#!/ last accessed: 2022). The affinity-based isolation in the qPCR may have excluded EVs present in higher abundance in ultracentrifugation isolation used in HTG analysis. However, it seems unlikely that a similar effect would not have been seen for other miRNAs. Another possibility is that the qPCR assay has some cross-reactivity with a different miRNA that is abundant in cases (see technical issues with qPCR below). Nevertheless, EV miRNA isolation and detection approaches can have a remarkable influence on EV miRNA measurement results. Therefore, there is currently a need for detailed quality assurance on these methodologies. A study by Wang et al. suggested an alternative EV isolation approach, differential centrifugation and annexin A5 coated beads might be a promising modality to address the current method challenges (Wang, Gong et al. 2022). To address technical issues with isolation methods, it is probably

better to perform quantitation using the same method in order to produce comparable results. Furthermore, from a practical perspective, it is probably better to use whichever isolation method will be used clinically in the discovery phase.

A study by Zhang et al. indicated that plasma miR-3149 has promising diagnostic value for severe coronary artery disease (CAD) and myocardial infarction (Zhang, Cai et al. 2020), but ours is the first indication that it may be related to lung cancer diagnosis.

Apart from miR-3149, miR-103a-3p, let-7c-3p and miR-107 are significantly up-regulated in lung cancer cases. miR-103a-3p is up-regulated in the qPCR validation and in the original HTG analysis, as discussed previously. It has been reported that Ron receptor tyrosine kinase (RON) overexpression-associated miRNA let-7c-3p, a metastasis suppressor, suppresses cell migration and invasion by down-regulating K-RAS, MMP11, Bcl-2, CASP3, and PBX3 in lung tissue (Wang, Li et al. 2020, Ou, Chen et al. 2021). It is unclear how the suppression of cancer malignancy might be associated with apparent upregulation in plasma EV in lung cancer cases, but interesting to speculate that it may be a suppressive host response to the presence of lung cancer.

A study by Wu et al. revealed that miR-107 was down-regulated in NSCLC serum samples compared with normal serum samples, contradicting the results we found for plasma EVs. Furthermore, in vitro experiments showed that miR-107 could supress cell proliferation, migration and arrest cell cycle in both A549 and H1299 NSCLC cell lines (Wu, Yuan et al. 2020); like let-7c-3p, another example of a cancer suppressive miRNA found to be high in plasma EVs.

A number of miRNAs, although not included in the qPCR validation, were found to be significantly up-regulated in lung cancer cases in both the original and reanalysis of the HTG data (miR-6124, miR-1290, miR-4534, miR-4644, miR-5196-5p, miR-6086, miR-6870-5p and miR-7111-5p, Table 5.3). Recently, Wang and his colleagues identified miR-6124 as a potential new prognosis biomarker in lung SqCCa using tissue samples and the RNA-seq expression data of lung SqCCa were collected from the Cancer Genome Atlas (TCGA) (Qian Wang 2022); although this expression in lung cancer may explain our findings, they do not report on levels in plasma. A study by Li et al. suggested that miR-6870-5p plays suppressive roles in regulating malignant behaviours of lung SqCCa cell lines, based on the measurement of abundance of RNAs through qRT-PCR (Li, Zhao et al. 2021); another example of a cancer suppressive miRNA that we found to be up-regulated in plasma EV. A study by Saraei et al. showed that miR6870-5p and miR-7111-5p (another significantly expressed miRNA identified in our EV miRNA HTG EdgeSeq reanalysis) were significantly down-regulated in oral squamous cell carcinoma (OSCC) tissues compared with adjacent normal oral epithelial tissues (Mona SARAEI 2021); this indicates that the same miRNA might have different effects in different cancers, even closely related ones (two different respiratory tract squamous cancers). In a case report by Chinami et al., miR-5196-5p was shown to be suppressed by AC-ACT immunotherapy, and they speculate that this is related to its role in controlling the Fra2 gene, which is involved in cancer cell motility (Chinami, Iwabuchi et al. 2019). We found no evidence in the literature that miRNAs miR-1290, miR-4534, miR-4644 and miR-6086 have previously been associated with cancer.

When investigating COPD in our qPCR validation data, miR-1277-5p was higher in those with COPD and miR-1228-3p was higher in those without COPD. Both of these miRNAs showed

some predictive discrimination for COPD status (AUC P = 0.070 and 0.046, respectively), but their ratio (miR-1277-5p/miR-1228-3p) gave an even better prediction [AUC = 0.667 (95% CI 0.518 – 0.816), P = 0.029]. This demonstrates the power of combining miRNAs in this way, as has been shown previously for plasma miRNA (Boeri, Verri et al. 2011), which also overcomes the problem of identifying an endogenous loading control for normalisation. It has been revealed that miRNA miR-1277, which was predictive of COPD in a ratio to miR-1228-3p in our qPCR validation, has an association with inflammation (Guo, Wang et al. 2021, Yan, Li et al. 2022); this might fit with a role in COPD, which is an inflammatory disease. Meanwhile serum circulating miRNA miR-1228-3p, demonstrated some diagnostic efficacy for NSCLC when used in conjunction with-181a-5p and miR-1228-3p was an independent factor for the poor prognosis of NSCLC patients (Xue, Zhang et al. 2020). However, we found no significant association of miR-1228-3p with lung cancer status.

Reanalysis of HTG data identified technical challenges

Given the validation of the 18 miRNAs was only partly successful, we performed a reanalysis for the EV miRNA HTG data using a new software tool HTG REVEAL analysis with improved quality control and comprehensive integration of differential expression analysis (based on edgeR). Using a combination of quality control (QC) measures (ANT-QC and the new QC metrics, including QC0, QC1 and QC2), which were different from the QC measure applied for the original HTG analysis (ANT-QC only), we limited the number of samples used in the reanalysis, which would alter the differential expression in comparison to the original analysis, in that it removes miRNAs that were over-represented (or under-represented) in samples of poor quality (according to the QC metrics applied). Specifically, poor quality samples tended to have a smaller range of miRNA expression levels (as witnessed by QC2) or lower overall

levels of miRNA (as seen with QC0). The former was most closely associated with the previously used ANT-QC.

Although QC metrics were used to remove the poorest quality samples, it should be recognised that quality does not have an absolute threshold and the data will still contain samples of better quality (most representative, having wider variance in expression levels) and poorer quality (with narrower representation and variance due to lower overall counts). On the basis that there was a clear relationship between ANT-QC and sample distribution by PCA, investigating the placement of different QC flagged samples in PCA plots provides some indication of the likely quality of those remaining samples that passed all QC metrics.

Using PCA plots (Fig. 4.4A & 4.5A), it was seen that miRNA quality has a significant influence, e.g. ANT and QC2 failures cluster to the right (high PC1 values), with brain controls to the far left. It is reasonable to assume that poor quality samples that fall short of the threshold for exclusion will cluster close to the QC failures. When looking at the potential batch effects, based on initial sample processing, batch 2 was seen to cluster to the right and therefore might be of somewhat lower quality than batch 1 (Fig. 4.6). That quality issues might affect the EV miRNA relationship to lung cancer status, is demonstrated by the PCA plot based on the 4-miRNA panel (Fig 4.5B), which showed better separation of cases and controls on the left-hand side (low PC1 values, associated with better quality). Regardless of whether it is due to miRNA quality issues, or differences in the nature of the isolated EVs in the processing batches, there was good evidence of batch effects and for that reason in the reanalysis we performed differential expression within batches (or combined) and looked for cross-validation between batches.

The reanalysis was implemented using 72 samples (43 cases and 29 controls) which passed all QC metrics and investigated the differential expression of the miRNAs associated with lung cancer case/control status, COPD, sex, smoking status and histology. Some miRNAs were found to be associated with one or more factors in either one batch or the combined batch analyses. In keeping with the deduction that batch 1 quality was generally better, the differential expression in reanalysis of batch 1 was more in keeping with the original HTG analysis, i.e. a greater proportion of differentially expressed miRNAs were the same.

We also analysed the previous plasma miRNA HTG data to explore the miRNAs differentially expressed in plasma as well as the differences between EV miRNA and plasma miRNA in terms of dynamic range, enrichment and depletion. Lastly, a cross comparison was performed between the four analysed datasets discussed above. Some interesting findings will be demonstrated and discussed in detail in the following sections.

Technical issues related to the use of qPCR in the validation cohort and in clinical studies

Whilst using an assay that is more easily adopted in clinical testing laboratories has obvious advantages, it is not without issues, some of which are evident in our results. Although qPCR has such issues, there was a rational for utilising it in our validation cohort, which is tightly linked to the small absolute quantity of the extracted EV miRNA. In other words, we had very little template to begin with, given that EV miRNA is only a fraction of that present in plasma. As qRT-PCR utilises PCR amplification of a target DNA sequence to quantify the amount of DNA input in the reaction, it is inherently sensitive. However, to further aid with low starting amounts, we chose TaqMan Advanced miRNA cDNA Synthesis Kit for the RT step as the miR-Amp reaction boosts the quantity prior to qPCR, to allow multiple miRNAs to be measured from limited starting material.

Well-established limitations of qRT-PCR for miRNAs include that there is little flexibility over primer design (as the average miRNA is only 22 nucleotides long, which is the same length as a traditional PCR primer) and close similarities between some miRNAs (e.g. 1-2 bp differences, which leads to a lack of accuracy). Mis-priming, due to partial homology with closely related DNA sequences, especially when the true target sequence is absent or at very low levels, can result in false signal positive readings. These limitations are highly likely to contribute to failure to validate some miRNAs. In order to help overcome these limitations, the Exiqon platform might be suitable, as its locked nucleic acid (LNA) based technology currently appears to be the most specific [Mestdagh 2014].

Another disadvantage of qPCR is its potential for primer artefacts in the "no template controls" (NTC) and any samples with low amounts of template, where signals can arise due to dye molecules binding to primer dimers and related primer-extension artefacts. We saw some evidence of this when optimising the qPCR conditions (Fig. 3.9); one way it presented was in the lack of linear quantitation at the lower end of the dilutions tested (i.e. similar Ct values despite lower amounts of template added, Fig. 3.3A). It was shown that the artefacts were likely generated within the miR-Amp amplification reactions using random priming, but these were required in order to test multiple miRNAs from limited starting material (unlike the HTG assay and similar highly-multiplexed parallel detection methods). The HTG analysis indicates that the EV fraction contains the majority of miRNAs at some level, hence low target amounts, likely to have this type of false signal error, will mainly be an issue for the minority of samples which yield little total miRNA. Whilst these are easily identified by HTG QC metrics (based on multiple miRNA and control readings), samples with low quantity or quality of

miRNA are harder to identify based on qRT-PCR, without reliable low-level endogenous controls.

In order to further investigate the nature of the false signal (assumed artefact) seen in the no RNA controls, we could clone the PCR product (e.g. by T/A cloning) and sequence it using primers either side of the cloning site. This is especially useful if the size of the PCR artefact is the same as or similar to the expected product, which would mean we were unable to determine artefacts on sizing (e.g. agarose gel electrophoresis or melt-curve analysis). It should be noted that miRNA qRT-PCR relies on amplification of very short PCR products, where sizing is not very easy. As TaqMan probes were used, melt-curve analysis was not performed; it is often assumed that artefacts due to mis-priming will give no signal with TaqMan probes, so it should be borne in mind that the signal in no RNA controls could be contamination. Given that no such signal was seen in adjacent no-target PCR controls, it is unlikely to be contamination and the most obvious possibility is a product produced by random priming in the RT and/or miR-Amp reaction.

We further investigated the nature of the artefacts by performing a QIAquick column-based PCR purification after miR-Amp reaction and before qPCR. However, the outcome indicated that any artefact due to RT or miR-Amp reaction was not easily purified away from true RT product. At least 3 additional research group members (including Dr Liloglou) performed the same experiment and all attempts led to the same background signal. Therefore, it is unlikely to be an issue associated with operator skills. Another experiment showed that lower level undetermined signals (Ct of 39-41) occur for both RT and miR-Amp reaction products of 3 no RNA controls, i.e. these signals occurred with or without the miR-Amp reaction (Fig. 3.9), indicating that there is more than one source of such false signals. It is also noteworthy that

the artefacts are not amplified to the same extent as the true samples (Fig. 3.9), in keeping with them only having partial homology to the intended target.

One of the other disadvantages of using qPCR as a clinical method is that it is not easily scalable to the large numbers of miRNAs that might be needed for panels. We were able to measure 18 miRNAs, but that was challenging and may not be enough, especially if it includes endogenous controls for quality measures or loading corrections (which were not needed in the HTG analysis that revealed an accurate 4-miRNA panel).

In order to account for variations in input amounts, the standard approach is to use a ratio to an endogenous control (for miRNA studies these are often referred to as housekeeping genes, HK). However, despite many years and many efforts, no good candidates for universal control miRNAs have been widely recognised, hence, there remains lack of standardisation. In order to address quality issues, endogenous controls should be present at similar levels to the biomarker miRNAs, which means they can suffer from the same false-signal errors seen with poor quality or low quantity samples. The same limitations are not true for HTG analysis, or similar methods, such as RNA-Seq, in which the miRNA levels are reported in relation to bulk miRNA levels (of all measured) in terms of cpm and quality can be addressed using a number of QC metrics based on signal distribution.

We considered performing a separate reaction for a HK gene present in EVs, where there is more consensus, but this would require a different RT in parallel, so would be wasteful of the limited sample and not fully reflect the RT efficiency for the miRNA. Another alternative is using a spike-in control (non-human miRNA), this has the advantage of being a control for the whole process. However, it is not a true endogenous control and there is a danger, unless carefully calibrated amounts are used, that the spike control will swamp the RT reaction

(which would decrease the sensitivity for the other miRNAs). Unlike plasma, with significantly higher overall amounts of miRNA, the suitable quantity of spike control has not been determined for EV-based studies.

In order to help with plate-to-plate normalisation, we had included a cell line calibrator and a plasma calibrator on every RT plate. However, the Ct values of the calibrators in the qPCR reactions were found to vary significantly between the plates for different miRNAs. This is presumably related to the relative levels of the different miRNAs in the calibrators chosen. The plasma calibrator was not used for statistical analysis, as it was negative for some miRNAs. The cell line calibrator was more reproducible, but still introduced additional variation between plates (Fig.3.8), particularly for miRNAs that were expressed at low levels in the calibrator and therefor potentially suffered from some technical variability. We therefore settled on median normalisation, which is based on the principle that for large enough sets of samples the median expression will be similar. Where there were 92 samples on a plate, this seemed to work well (Fig. 3.8C), apart from plate 2 where there seemed to be a number of qPCR failures; however, it also worked well in plate 6, despite there only being 14 samples.

Despite being able to normalise between plates, without an endogenous control or another form of loading control, we were unable to make individual corrections for variation in EV miRNA starting amount or isolation losses between samples. However, it should be noted that there is no evidence that overall EV concentration in plasma, or miRNA concentration in EVs, are the same between individuals or vary with disease . We therefore rely on trying to ensure that the same amount of plasma is used and the isolation, RT and miR-Amp amplification steps are carried out with as little variation as possible. If a suitable endogenous control could be identified and qRT-PCR was performed in a fully quantitative manner (against

a standard curve), it might be possible to express EV miRNAs as an absolute value, albeit at considerable expense and complexity. Alternatively, digital droplet PCR would provide better quantitation. Even without a suitable endogenous control, it is possible to express EV miRNA as ratios between two targets that both have diagnostic value, e.g. (Boeri, Verri et al. 2011), as shown for COPD with miR-1277-5p and miR-1228-3p (Fig. 3.17C).

Technical advantages of the HTG EdgeSeq platform

HTG EdgeSeq miRNA Whole Transcriptome Assay (HTG Molecular Diagnostics) is an extraction-free, multiplexed nuclease protection assay with next-generation sequencing readout for difficult sample types. First, HTG EdgeSeq WTA allows measurement of the expression of 2,083 human miRNA transcripts in parallel, enabling hundreds of millions of miRNA molecules to be sequenced at a time, therefore a high-throughput probe-based technology. Moreover, probes that successfully hybridise to their cognate miRNAs in the sample are protected from nuclease digestion, amplified with the addition of barcodes, and then sequenced. In consequence, the output for EdgeSeq is read count, as in small RNA-Seq, but unlike small RNA-seq, the number of reads reflects the quantity of probes that were bound by miRNAs and protected from digestion. Godoy et al. compared small RNA sequencing to three targeted miRNA quantification platforms, including small RNA sequencing (RNA-seq), FirePlex, HTG EdgeSeq miRNA WTA and nCounter, to assess their reproducibility, bias, specificity, sensitivity and accuracy. The results of the comparison indicated that EdgeSeq is the most reproducible and has the least detection bias (Godoy, Barczak et al. 2019). We were also able to demonstrate the reproducibility by looking at in-house batch-to batch analysis by having the same sample run on different runs (e.g. brain control, Fig. 4.2). Furthermore, the HTG EdgeSeq platform has been shown to be robust enough for clinical use, and is already

approved in clinical labs for ALK testing. Additionally, there have been improvements in available tools for HTG data analysis, including quality control and bioinformatics (The HTG EdgeSeq Reveal software package), which provides more comprehensive and accurate analysis associated with a variety of factors. An important practical benefit of the HTG EdgeSeq platform, over RNA-Seq, is that it does not require alignment of sequence data to the genome to identify the miRNA (as this is not standardised and can be a source of variation between experiments). Instead, the sequence data is parsed against only to the pre-defined probe panel, which requires less bioinformatic experience; in fact, the only sequence data produced is that for the probes, so it is more efficient than RNA-Seq, which typically produces sequence data on all RNA present in the preparation. However, this needs to be weighed against the fact that RNA-Seq provides more comprehensive data, including other small RNA molecules.

The use of HTG EdgeSeq or RNA-Seq approaches, which utilise counts per million (cpm, normalisation in relation to the overall reads) provide a better indication of miRNA patterns which may be more informative and allow additional biomarker panels to be tested post-hoc, compared to qRT-PCR for which whole transcriptome approaches are impracticable. One significant barrier to clinical use is that they are expensive techniques. However, considering cost-effectiveness on a per-miRNA scale, HTG EdgeSeq is cheaper per target than the qPCR method, so might be cost-effective for larger panels. Other strengths of the HTG EdgeSeq method can be seen when exploring the pre-analytical steps of HTG and qPCR methods. For instance, for the HTG-based analysis, 2ml plasma sample was used for extracting 25µl EV miRNA, and half of it (equivalent to 1ml) was used in HTG EV analysis to produce results for nearly 2000 miRNAs. In contrast, using the 1ml column-based preparation of EV for qRT-PCR

did not provide full data, even for the relatively small number of miRNA tested. This may be in part because of losses during processing for the method used before qPCR, in which the bound vesicles were lysed and eluted with QIAzol (Qiagen)/Trizol (Invitrogen), then the Direct-zol RNA Isolation kit was used to isolate total RNA from plasma EV fractions. This extraction process can also introduce additional variation due to multiple steps, operator errors and process variability, and it is more time- and labour-consuming compared to HTG, which is extraction-free. EV miRNA HTG analysis was performed using a traditional ultracentrifugation process, which relies on sedimentation at high speed for separating EVs from other (extra) cellular components; this can also result in loses and variation, but we validated EV size using Nanosight, and composition using TEM and western blot. Ultracentrifugation is the most common EV isolation method to date and it is preferred for high-purity EV isolation. Our re-analysis of the HTG data did indicate there was some potential variation in quality between batches, but, even accounting for that, far more data was produced and the HTG QC metrics allowed us to remove a significant amount of the technical variation (albeit by excluding samples).

Confounding factors in EV miRNA analysis

Eight miRNAs were found to be associated with multiple factors in the HTG reanalysis. This highlights one of the difficulties of working with miRNA as biomarkers. If they are associated with multiple diseases, this can lead to false positive tests for a specific disease. This is compounded by the fact that many older people, most at risk of lung cancer, often have other conditions; the most obvious of these are COPD, other lung disease and heart disease, all associated with smoking. Inflammation associated miRNAs are a particular problem, as similar inflammatory processes are associated with multiple diseases. Whilst it may be possible to

select miRNA panels that are specific for a single disease, or that are sufficiently adjusted to accommodate some known confounding factors, there is always a risk of unknown confounding factors, if data about the confounding disease was not available at the time of collection. The Liverpool Lung Project sample set is very useful for such studies as a lot of data has been collected on participants: questionnaire data, a research nurse-led data collection (including medical history and medication), hospital case-note review, cancer registry (for cancer diagnosis both before and after the sample was taken), mortality data (including diseases that contributed to death) and Health Episode Statistics (providing data on any episodes in patient or outpatient diagnosis and treatment, linked to disease code). It is therefore possible to exclude subjects (e.g. with other cancers) or confirm if they had a disease of interest (e.g. COPD either known at the time of sample from the questionnaire, or diagnosed shortly after the sample was taken from Health Episode Statistics).

One of the difficulties of working with EVs is that they can vary due to stress which may be closely related to exercise or exposure to toxins (e.g. smoking, medications) and currently, there is often little or no control over these types of stress when collecting samples. Cancer is predominantly a disease of old age and older people are often on multiple drugs, many of which might alter EV or EV miRNA patterns (i.e. both confounding diseases and confounding treatments). A number of studies showed that a wide range of toxicants alter miRNA levels in target organs. For instance, miR-103a-3p was reported as an appropriate biomarker among the circulating miRNAs identified in rats with acetaminophen-induced hepatotoxicity (Yokoi and Nakajima 2013). Furthermore, a study in human and mouse models suggested that circulating miR-122-5p can potentially be used as a novel and reliable, early, predictive blood marker for viral-, alcohol- and chemical-induced liver injury (Zhang, Jia et al. 2010). Recently

it was shown, that miR-21-5p, miR-155-5p and miR-18a-5p were among the highest up-regulated miRNAs in the kidney after injury (Saikumar, Hoffmann et al. 2012). Cardiotoxicity is one of the major safety concerns in drug development; in doxorubicin induced cardiotoxicity, circulating levels of miR-34a-3p and miR-208a-3p were enhanced (Nishimura, Kondo et al. 2015, Piegari, Russo et al. 2016). Apart from those, a study by Frühbeis et al. has found that exosomes showed a nearly twofold increase immediately after exercise cessation (Fruhbeis, Helmig et al. 2015). It has also been reported that long-term fasting engages homeostatic mechanisms associated with specific miRNAs to improve metabolic signalling regardless of health status (Ravanidis, Grundler et al. 2021). Therefore, it is possible that these factors may have influence on EV miRNA expression of the subjects, however, we may not able to measure or address the impact to the findings of our research, as we do not have data on all these factors and it would be difficult to incorporate them into the analysis even if we did.

Levels of EV miRNA are not only associated with pathological changes; a recently published pilot study indicated that short (median 12 days long) low-Earth orbit (LEO) spaceflight induced significant changes in plasma-derived EV miRNA expression as miR-4732-3p was found to be significantly up-regulated post-flight. It is also worth noticing that there were some substantial discrepancies (overexpressed in the opposite direction) between the expression of some miRNAs detected using two different approaches (sRNA-Seq and qPCR), which is similar to the findings of our research: miR-3149 was significantly down-regulated in lung cancer cases in the original analysis and reanalysis of EV miRNA HTG EdgeSeq data, as well as plasma miRNA HTG EdgeSeq data, but significantly up-regulated in qPCR analysis of EV miRNA validation cohort.

EV miRNA compared to plasma miRNA

Given that the EV isolation techniques apparently introduced some variation in EV miRNA (in terms of data quality between batches in HTG analysis) and potentially contributed to a failure to validate by qPCR (as different EV isolation techniques were used in discovery and validation), and that for most EVs there is a strong correlation between plasma miRNA and EV miRNA levels, it is worth speculating that it might be preferable to simply use plasma, avoiding the complications of EV isolation. It is noteworthy that the starting amount for ultracentrifugation was 1ml to obtain 15µl EV fraction for EV miRNA HTG analysis. However, only 15µl plasma sample is needed for plasma miRNA analysis and this requires less amplification during the library prep to provide the desired total miRNA read count (which can be seen by the fact that the lowest read count for plasma miRNA is only 5 cpm, compared to 50 cpm for EV miRNA). Hence, we get just as much, and potentially more, data using plasma than EVs using far less sample; this is only a practical advantage clinically if it is possible to collect the smaller amount (i.e. would we collect the same volume of blood to provide either 1ml or 15µl). For 15µl, might it be possible to collect less in a way that is less distressing or intrusive (e.g. finger prick)?

It has been reported that a significant proportion (estimated as 90%) of circulating miRNAs are protein bound (stabilised) nucleic acids, with only 10% protected within exosomes (Arroyo, Chevillet et al. 2011, Cortez, Bueso-Ramos et al. 2011, Vickers, Palmisano et al. 2011), but most whole plasma miRNA extraction techniques do not discriminate between these different forms. Nevertheless, when we compared plasma to EV miRNA, there were very strong correlations, both in controls (Fig. 5.3) and cases, although the correlation between plasma and EV was less than between case and control for each sample type (Fig. 5.4),

indicating that there was clearly enrichment and depletion in the EV fraction for a significant portion of the miRNAs.

Given the strong correlation between plasma and EV miRNA levels and the additional technical challenges of working with EVs (both in terms of larger starting amounts and variation in isolation techniques), the advantages of working with plasma need to be weighed against the potential additional biological relevance of working with EVs, especially exosomes (with their proven signalling roles).

Future work

In this project, the blood samples used for the three main parts (EV miRNA HTG EdgeSeq data, EV miRNA qPCR validation cohort and plasma miRNA HTG EdgeSeq data) were collected from independent groups of subjects to facilitate independent biological validation. The limitation of this is that it does not facilitate the technical comparison between the three approaches. Future work should include measuring EV miRNA expression in the same starting material, utilising the same isolation method and different detection methods, or different isolation methods with the same detection method. Since HTG EdgeSeq analysis is most informative (data on more miRNAs and on their relative levels) and has inherent, recognised QC metrics to judge sample quality, the best experiment would be to use that to compare the same sample isolated by different techniques. Any relationship to the differences in EV profiles (e.g. relative proportions of exosome and microvesicles) could be correlated with NanoSight results or investigated using miRNAs known to be found mainly in exosome or EV.

Similarly, it would be ideal to measure the expression of EV miRNA and plasma miRNA through direct comparison between EV and plasma in same blood samples to minimise the

variation and confounding factors. We might expect the correlation between miRNA levels to be even better than we found, as some of the variation in our experiment is likely to be biological, e.g. related to different disease profiles in the control samples used.

Future work could also include alternative ways of analysing the data. Rather than looking at single miRNA based on EV miRNA HTG data reanalysis, it is worth identifying multi-miRNA biomarker panel (prediction models), like Michael Marcus did with the original EV miRNA HTG data via Lasso penalized logistic regression, or similar statistical techniques that remove redundancy in biomarker panels.

We used the ratio between 2 miRNAs for COPD prediction and this was found to be beneficial in terms of accuracy. To avoid the need for an endogenous control, which have not been established for EV miRNA, it is worth creating ratios of all miRNAs and testing those combinations, an approach that was taken by Boeri et al. (Boeri, Verri et al. 2011). The problem is that this produces a lot of ratios, so the Bonferroni correction is massive; this was less of an issue in earlier studies, as the discovery platform used by Boeri et al. detected fewer miRNAs. However, we should be able to adapt the approach and apply it only to those that are significantly up- or down-regulated (i.e. each up ratioed with each down).

In our research, we focused on exploring differential expression of plasma EV miRNAs based on quantitative analysis. We can then further look into the biomarker pathways. miRNAs have been linked to expression of specific proteins, which fall onto certain signalling pathways or cell functions (e.g. proliferation, apoptosis, motility, invasion, etc.) especially for exosomal miRNAs as exosomes have been revealed to initiate or suppress various signalling pathways in the recipient cells via transmitting heterogeneous cargoes. Whereas a lot of gene-pathway interactions have been identified and validated, the linkage of specific miRNAs to the

expression of specific genes is not simple and relies on complex bioinformatic analysis, not just of expression datasets, but also of interactions between miRNA and functional gene motifs. It is therefore useful to deploy data such as HTG EdgeSeq miRNA WTA, which provides maximal numbers of miRNAs, allowing multiple hits to specific pathways, even if some are missing because the miRNA-gene interactions are not fully mapped.

Upon identification of a suitable validated biomarker panel, the next stage would be independent validation in a prospective cohort (collect new samples and test with a "locked-down" biomarker assay) or a different representative retrospective cohort (use biobanked samples, as with LLP, choosing which to include, so that specific patient groups have sufficient numbers for statistical analysis). Use of a non-selected prospective cohort, which means taking everyone who attends a diagnostic clinic, is the only way to get true positive predictive value (PPV) and negative predictive value (NPV), as they take account of the incidence of the disease. In contrast, selective cohorts (often retrospective such as ours) cannot provide NPV and PPV unless adjusted to reflect incidence, they therefore may also not give a true measure of accuracy in the clinical settings. Nevertheless, additional retrospective cohorts are a relatively cheap and easily accessed form of external validation, which have the advantage that there is follow-up data available. This is particularly useful for early diagnosis markers, as by implication, if there is a need for these, the current "gold standard" diagnosis techniques are not sufficient and a significant number of false negatives will be found to have a positive diagnosis shortly after the original negative diagnosis (Nikolaidis, Raji et al. 2012). For this reason, a preferred method is the prospectively collected, retrospectively analysed study; this has the benefit of standardised prospective sample collection (after which samples are all stored in the same way), but also allows for follow-up prior to sample analysis. One issue is

that some biomarkers may be affected by sample storage; although this does not seem to be a significant concern for miRNA, it might be a bigger issue for EVs.

Conclusions

This study has demonstrated that certain miRNAs are up- or down-regulated in plasma EV fraction in lung cancer cases. However, expression patterns were not very consistent between the discovery and validation data. The overexpression of the same miRNAs was sometimes in the opposite direction in EV miRNA HTG EdgeSeq analysis and EV miRNA qPCR validation cohort. It is assumed that this is partly due to different EV miRNA isolation and detection approaches, as well as statistical analysis methods, but might also reflect variation in the patient cohorts used. In both the qPCR validation cohort and the HTG EdgeSeq reanalysis, we could identify some miRNAs that were differentially expressed in regards to other factors, including COPD, smoking status and sex.

EV miRNA is of interest due to its wide range of biological functions in both normal and pathological physiology. However, challenges remain outstanding in terms of the absence of well-established and acknowledged EV miRNA endogenous control. The association of EV miRNAs with multiple diseases and physiological states produces confounding factors that have an impact on data interpretation, the drawing conclusions and ease of validation. The over-representation of some miRNAs might be related to the confounding factors that are associated with lung cancer rather than lung cancer itself, or even the factors that are not relevant to lung cancer, but only to the stress caused by medication or smoking, as well as diet or exercise. Undoubtedly, this kind of analysis still has predictive values, but unless we carefully select and validate EV miRNAs to avoid negative impacts of confounding factors, this could lead to some degree of waste in health care and unclear benefit for lung cancer patients.

The collective evidence in our research, and other studies in this research field, suggests that miRNAs in plasma and its EV fraction show promise as minimally invasive diagnostic, prognostic and therapeutic clinically valuable biomarkers. However, significant effort is still required to overcome the practical challenges of assay reproducibility and utility, data analysis and interpretation, validation of research findings and translation of research into clinical practice. More well-designed cohort studies should be conducted to investigate the clinical diagnostic value of plasma and plasma EV miRNAs. These should base their discovery phase isolation methods on those to be used in the clinical setting and include technical validation between the measurement methods used in discovery and validation (unless the discovery platform technology can be easily utilised in a clinical setting).

Bibliography

Adizie, J. B., A. Khakwani, P. Beckett, N. Navani, D. West, I. Woolhouse and S. V. Harden (2019). "Stage III Non-small Cell Lung Cancer Management in England." Clin Oncol (R Coll Radiol) **31**(10): 688-696.

Aga, M., G. L. Bentz, S. Raffa, M. R. Torrisi, S. Kondo, N. Wakisaka, T. Yoshizaki, J. S. Pagano and J. Shackelford (2014). "Exosomal HIF1alpha supports invasive potential of nasopharyngeal carcinoma-associated LMP1-positive exosomes." Oncogene **33**(37): 4613-4622.

Akers, J. C., V. Ramakrishnan, R. Kim, J. Skog, I. Nakano, S. Pingle, J. Kalinina, W. Hua, S. Kesari, Y. Mao, X. O. Breakefield, F. H. Hochberg, E. G. Van Meir, B. S. Carter and C. C. Chen (2013). "MiR-21 in the extracellular vesicles (EVs) of cerebrospinal fluid (CSF): a platform for glioblastoma biomarker development." PLoS One **8**(10): e78115.

Aleckovic, M. and Y. Kang (2015). "Regulation of cancer metastasis by cell-free miRNAs." Biochim Biophys Acta **1855**(1): 24-42.

Ambrogi, M. C., O. Fanucchi, P. Dini, F. Melfi, F. Davini, M. Lucchi, G. Massimetti and A. Mussi (2015). "Wedge resection and radiofrequency ablation for stage I nonsmall cell lung cancer." Eur Respir J **45**(4): 1089-1097.

Arita, T., D. Ichikawa, H. Konishi, S. Komatsu, A. Shiozaki, K. Shoda, T. Kawaguchi, S. Hirajima, H. Nagata, T. Kubota, H. Fujiwara, K. Okamoto and E. Otsuji (2013). "Circulating long non-coding RNAs in plasma of patients with gastric cancer." Anticancer Res **33**(8): 3185-3193.

Arroyo, J. D., J. R. Chevillet, E. M. Kroh, I. K. Ruf, C. C. Pritchard, D. F. Gibson, P. S. Mitchell, C. F. Bennett, E. L. Pogosova-Agadjanyan, D. L. Stirewalt, J. F. Tait and M. Tewari (2011). "Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma." Proc Natl Acad Sci U S A **108**(12): 5003-5008.

ascopost.com/news/november-2021/nsclc-in-the-united-states-update-on-incidence-prevalence-and-survival/. (last accessed: 2022). Retrieved 10 May 2022, from https://ascopost.com/news/november-2021/nsclc-in-the-united-states-update-on-incidence-prevalence-and-survival/.

atlasgeneticsoncology.org/solid-tumor/5141/lung-non-small-cell-carcinoma. (last accessed: 2022). Retrieved 18 Feb 2022, from https://atlasgeneticsoncology.org/solid-tumor/5141/lung-non-small-cell-carcinoma.

ATSDR (2009). Chemicals, Cancer, and You.

Auperin, A., R. Arriagada, J. P. Pignon, C. Le Pechoux, A. Gregor, R. J. Stephens, P. E. Kristjansen, B. E. Johnson, H. Ueoka, H. Wagner and J. Aisner (1999). "Prophylactic cranial irradiation for patients with small-cell lung cancer in complete remission. Prophylactic Cranial Irradiation Overview Collaborative Group." N Engl J Med **341**(7): 476-484.

Auperin, A., C. Le Pechoux, E. Rolland, W. J. Curran, K. Furuse, P. Fournel, J. Belderbos, G. Clamon, H. C. Ulutin, R. Paulus, T. Yamanaka, M. C. Bozonnat, A. Uitterhoeve, X. Wang, L. Stewart, R. Arriagada, S. Burdett and J. P. Pignon (2010). "Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer." J Clin Oncol **28**(13): 2181-2190.

Baj-Krzyworzeka, M., M. Majka, D. Pratico, J. Ratajczak, G. Vilaire, J. Kijowski, R. Reca, A. Janowska-Wieczorek and M. Z. Ratajczak (2002). "Platelet-derived microparticles stimulate proliferation, survival, adhesion, and chemotaxis of hematopoietic cells." Exp Hematol **30**(5): 450-459.

Battistelli, M. and E. Falcieri (2020). "Apoptotic Bodies: Particular Extracellular Vesicles Involved in Intercellular Communication." Biology (Basel) **9**(1).

Berghmans, T., M. Paesmans, A. P. Meert, C. Mascaux, P. Lothaire, J. J. Lafitte and J. P. Sculier (2005). "Survival improvement in resectable non-small cell lung cancer with (neo)adjuvant chemotherapy: results of a meta-analysis of the literature." Lung Cancer **49**(1): 13-23.

Bhan, A., M. Soleimani and S. S. Mandal (2017). "Long Noncoding RNA and Cancer: A New Paradigm." Cancer Res **77**(15): 3965-3981.

Bianchi, F., F. Nicassio, M. Marzi, E. Belloni, V. Dall'olio, L. Bernard, G. Pelosi, P. Maisonneuve, G. Veronesi and P. P. Di Fiore (2011). "A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer." EMBO Mol Med **3**(8): 495-503.

bioinfo.life.hust.edu.cn/EVmiRNA#!/. (last accessed: 2022). Retrieved 20 May 2022, from http://bioinfo.life.hust.edu.cn/EVmiRNA#!/.

Bleyer A, O. L. M., Barr R, Ries LAG (eds) (2006). Cancer Epidemiology in Older Adolescents and Young Adults 15 to 29 Years of Age, Including SEER Incidence and Survival: 1975-2000. N. C. Institute.

Blomquist, T., E. L. Crawford, D. Mullins, Y. Yoon, D. A. Hernandez, S. Khuder, P. L. Ruppel, E. Peters, D. J. Oldfield, B. Austermiller, J. C. Anders and J. C. Willey (2009). "Pattern of antioxidant and DNA repair gene expression in normal airway epithelium associated with lung cancer diagnosis." Cancer Res **69**(22): 8629-8635.

Boeri, M., C. Verri, D. Conte, L. Roz, P. Modena, F. Facchinetti, E. Calabro, C. M. Croce, U. Pastorino and G. Sozzi (2011). "MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer." Proc Natl Acad Sci U S A **108**(9): 3713-3718.

Boffetta, P. (2006). "Human cancer from environmental pollutants: the epidemiological evidence." Mutat Res **608**(2): 157-162.

Boffetta, P., N. Jourenkova and P. Gustavsson (1997). "Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons." Cancer Causes Control **8**(3): 444-472.

Bradley, S. H., N. L. F. Hatton, R. Aslam, B. Bhartia, M. E. Callister, M. P. Kennedy, L. T. Mounce, B. Shinkins, W. T. Hamilton and R. D. Neal (2021). "Estimating lung cancer risk from chest X-ray and symptoms: a prospective cohort study." Br J Gen Pract **71**(705): e280-e286.

Bray, F., J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal (2018). "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA Cancer J Clin **68**(6): 394-424.

Brothers, J. F., K. Hijazi, C. Mascaux, R. A. El-Zein, M. R. Spitz and A. Spira (2013). "Bridging the clinical gaps: genetic, epigenetic and transcriptomic biomarkers for the early detection of lung cancer in the post-National Lung Screening Trial era." BMC Med **11**: 168.

Brzezina, J., K. Kobolova and V. Adamec (2020). "Nanoparticle Number Concentration in the Air in Relation to the Time of the Year and Time of the Day." Atmosphere **11**(5).

Burgos-Ravanal, R., A. Campos, M. C. Diaz-Vesga, M. F. Gonzalez, D. Leon, L. Lobos-Gonzalez, L. Leyton, M. J. Kogan and A. F. G. Quest (2021). "Extracellular Vesicles as Mediators of Cancer Disease and as Nanosystems in Theranostic Applications." Cancers (Basel) **13**(13).

Bussolati, B. and G. Camussi (2017). "Renal injury: Early apoptotic extracellular vesicles in injury and repair." Nat Rev Nephrol **13**(9): 523-524.

Bydder, S., A. Nowak, K. Marion, M. Phillips and R. Atun (2009). "The impact of case discussion at a multidisciplinary team meeting on the treatment and survival of patients with inoperable non-small cell lung cancer." Intern Med J **39**(12): 838-841.

Byers, L. A. and C. M. Rudin (2015). "Small cell lung cancer: where do we go from here?" Cancer **121**(5): 664-672.

Caby, M. P., D. Lankar, C. Vincendeau-Scherrer, G. Raposo and C. Bonnerot (2005). "Exosomal-like vesicles are present in human blood plasma." Int Immunol **17**(7): 879-887.

Carter, B. W., B. S. Glisson, M. T. Truong and J. J. Erasmus (2014). "Small cell lung carcinoma: staging, imaging, and treatment considerations." Radiographics **34**(6): 1707-1721.

Carvalho, A. S., H. Baeta, B. C. Silva, M. C. S. Moraes, C. Bodo, H. C. Beck, M. S. Rodriguez, M. Saraswat, A. Pandey and R. Matthiesen (2020). "Extra-cellular vesicles carry proteome of cancer hallmarks." Front Biosci (Landmark Ed) **25**(3): 398-436.

Cassidy, A., J. P. Myles, S. W. Duffy, T. Liloglou and J. K. Field (2006). "Family history and risk of lung cancer: age-at-diagnosis in cases and first-degree relatives." Br J Cancer **95**(9): 1288-1290.

Cassidy, A., J. P. Myles, M. van Tongeren, R. D. Page, T. Liloglou, S. W. Duffy and J. K. Field (2008). "The LLP risk model: an individual risk prediction model for lung cancer." Br J Cancer **98**(2): 270-276.

Cavalleri, T., L. Angelici, C. Favero, L. Dioni, C. Mensi, C. Bareggi, A. Palleschi, A. Rimessi, D. Consonni, L. Bordini, A. Todaro, V. Bollati and A. C. Pesatori (2017). "Plasmatic extracellular vesicle microRNAs in malignant pleural mesothelioma and asbestos-exposed subjects suggest a 2-miRNA signature as potential biomarker of disease." PLoS One **12**(5): e0176680.

Cazzoli, R., F. Buttitta, M. Di Nicola, S. Malatesta, A. Marchetti, W. N. Rom and H. I. Pass (2013). "microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer." J Thorac Oncol **8**(9): 1156-1162.

CDC. "United States and Puerto Rico cancer statistics, 1999-2015 incidence request." Retrieved March 16, 2019, from https://wonder.cdc.gov/cancer-v2015.html.

Celik, M., S. Halezeroglu, C. Senol, M. Keles, Z. Yalcin, S. Urek, H. Kiral and B. Arman (1998). "Video-assisted thoracoscopic surgery: experience with 341 cases." Eur J Cardiothorac Surg **14**(2): 113-116.

Chan, B. A. and B. G. Hughes (2015). "Targeted therapy for non-small cell lung cancer: current standards and the promise of the future." Transl Lung Cancer Res **4**(1): 36-54.

Chen, D., Z. Jin, J. Zhang, C. Xu, K. Zhu, Y. Ruan, B. Zhang, B. Chen and J. Shen (2021). "Efficacy and Safety of Neoadjuvant Targeted Therapy vs. Neoadjuvant Chemotherapy for Stage IIIA EGFR-Mutant Non-small Cell Lung Cancer: A Systematic Review and Meta-Analysis." Front Surg **8**: 715318.

Chen, G., X. Wan, G. Yang and X. Zou (2015). "Traffic-related air pollution and lung cancer: A meta-analysis." Thorac Cancer **6**(3): 307-318.

Chen, L. T., S. D. Xu, H. Xu, J. F. Zhang, J. F. Ning and S. F. Wang (2012). "MicroRNA-378 is associated with non-small cell lung cancer brain metastasis by promoting cell migration, invasion and tumor angiogenesis." Med Oncol **29**(3): 1673-1680.

Chen, S. C. and C. M. Liao (2006). "Health risk assessment on human exposed to environmental polycyclic aromatic hydrocarbons pollution sources." Sci Total Environ **366**(1): 112-123.

Chen, Y., E. Zitello, R. Guo and Y. Deng (2021). "The function of LncRNAs and their role in the prediction, diagnosis, and prognosis of lung cancer." Clin Transl Med **11**(4): e367.

Chinami, M., K. Iwabuchi, Y. Muto, Y. Uchida, R. Arita, R. A. Shuraim and C. N. Adra (2019). "Assessment by miRNA microarray of an autologous cancer antigen-pulsed adoptive immune ensemble cell therapy (AC-ACT) approach; demonstrated induction of anti-oncogenic and anti-PD-L1 miRNAs." Clin Case Rep **7**(11): 2156-2164.

Choi, D. Y., S. You, J. H. Jung, J. C. Lee, J. K. Rho, K. Y. Lee, M. R. Freeman, K. P. Kim and J. Kim (2014). "Extracellular vesicles shed from gefitinib-resistant nonsmall cell lung cancer regulate the tumor microenvironment." Proteomics **14**(16): 1845-1856.

Christianson, H. C., K. J. Svensson, T. H. van Kuppevelt, J. P. Li and M. Belting (2013). "Cancer cell exosomes depend on cell-surface heparan sulfate proteoglycans for their internalization and functional activity." Proc Natl Acad Sci U S A **110**(43): 17380-17385.

Cohen, A. J. and C. A. Pope (1995). "Lung cancer and air pollution." Environmental Health Perspectives **103**: 219-224.

Committee, H. a. P. P. (1985). "Diagnostic thoracentesis and pleural biopsy in pleural effusions. Health and Public Policy Committee, American College of Physicians." Ann Intern Med **103**(5): 799-802.

commons.wikimedia.org/wiki/File:Pie_chart_of_lung_cancers.svg. (last accessed: 2022). Retrieved 14 Feb 2022, from https://commons.wikimedia.org/wiki/File:Pie_chart_of_lung_cancers.svg.

Conde-Vancells, J., E. Rodriguez-Suarez, N. Embade, D. Gil, R. Matthiesen, M. Valle, F. Elortza, S. C. Lu, J. M. Mato and J. M. Falcon-Perez (2008). "Characterization and comprehensive proteome profiling of exosomes secreted by hepatocytes." J Proteome Res **7**(12): 5157-5166.

Cortez, M. A., C. Bueso-Ramos, J. Ferdin, G. Lopez-Berestein, A. K. Sood and G. A. Calin (2011). "MicroRNAs in body fluids--the mix of hormones and biomarkers." Nat Rev Clin Oncol **8**(8): 467-477.

Craver, L. F. (1948). "The nitrogen mustards; clinical use." Radiology **50**(4): 486-493.

Crescitelli, R., C. Lasser, T. G. Szabo, A. Kittel, M. Eldh, I. Dianzani, E. I. Buzas and J. Lotvall (2013). "Distinct RNA profiles in subpopulations of extracellular vesicles: apoptotic bodies, microvesicles and exosomes." J Extracell Vesicles **2**.

Crino, L., W. Weder, J. van Meerbeeck, E. Felip and E. G. W. Group (2010). "Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Ann Oncol **21 Suppl 5**: v103-115.

Cristiano, S., A. Leal, J. Phallen, J. Fiksel, V. Adleff, D. C. Bruhm, S. O. Jensen, J. E. Medina, C. Hruban, J. R. White, D. N. Palsgrove, N. Niknafs, V. Anagnostou, P. Forde, J. Naidoo, K. Marrone, J. Brahmer, B. D. Woodward, H. Husain, K. L. van Rooijen, M. W. Orntoft, A. H. Madsen, C. J. H. van de Velde, M. Verheij, A. Cats, C. J. A. Punt, G. R. Vink, N. C. T. van Grieken, M. Koopman, R. J. A. Fijneman, J. S. Johansen, H. J. Nielsen, G. A. Meijer, C. L. Andersen, R. B. Scharpf and V. E. Velculescu (2019). "Genome-wide cell-free DNA fragmentation in patients with cancer." Nature **570**(7761): 385-389.

Daly, M. E., N. Singh, N. Ismaila, M. B. Antonoff, D. A. Arenberg, J. Bradley, E. David, F. Detterbeck, M. Fruh, M. A. Gubens, A. C. Moore, S. K. Padda, J. D. Patel, T. Phillips, A. Qin, C. Robinson and C. B. Simone, 2nd (2022). "Management of Stage III Non-Small-Cell Lung Cancer: ASCO Guideline." J Clin Oncol **40**(12): 1356-1384.

David, A. R. and M. R. Zimmerman (2010). "Cancer: an old disease, a new disease or something in between?" Nat Rev Cancer **10**(10): 728-733.

Davis, M. D., A. Montpetit and J. Hunt (2012). "Exhaled breath condensate: an overview." Immunol Allergy Clin North Am **32**(3): 363-375.

Dawe, D. E., D. Christiansen, A. Swaminath, P. M. Ellis, J. Rothney, R. Rabbani, A. M. Abou-Setta, R. Zarychanski and S. M. Mahmud (2016). "Chemoradiotherapy versus radiotherapy alone in elderly patients with stage III non-small cell lung cancer: A systematic review and meta-analysis." Lung Cancer **99**: 180-185.

De Marinis, F., V. Gebbia and L. De Petris (2005). "Neoadjuvant chemotherapy for stage IIIA-N2 non-small cell lung cancer." Ann Oncol **16 Suppl 4**: iv116-122.

Deng, X., N. Feng, M. Zheng, X. Ye, H. Lin, X. Yu, Z. Gan, Z. Fang, H. Zhang, M. Gao, Z. J. Zheng, H. Yu, W. Ding and B. Qian (2017). "PM2.5 exposure-induced autophagy is mediated by lncRNA loc146880 which also promotes the migration and invasion of lung cancer cells." Biochim Biophys Acta Gen Subj **1861**(2): 112-125.

Denissenko, M. F., A. Pao, M. S. Tang and G. P. Pfeifer (1996). "Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53." Science **274**(5286): 430-432.

Detterbeck, F. C., D. J. Boffa, A. W. Kim and L. T. Tanoue (2017). "The Eighth Edition Lung Cancer Stage Classification." Chest **151**(1): 193-203.

DeVita, V. T., Jr. and E. Chu (2008). "A history of cancer chemotherapy." Cancer Res **68**(21): 8643-8653.

Di Leva, G., M. Garofalo and C. M. Croce (2014). "MicroRNAs in cancer." Annu Rev Pathol **9**: 287-314.

Ding, G., J. An and L. Li (2022). "MicroRNA-103a-3p enhances sepsis-induced acute kidney injury via targeting CXCL12." Bioengineered **13**(4): 10288-10298.

Dixon, C. L., S. Sheller-Miller, G. R. Saade, S. J. Fortunato, A. Lai, C. Palma, D. Guanzon, C. Salomon and R. Menon (2018). "Amniotic Fluid Exosome Proteomic Profile Exhibits Unique Pathways of Term and Preterm Labor." Endocrinology **159**(5): 2229-2240.

Doll, R. and R. Peto (1981). "The Causes of Cancer - Quantitative Estimates of Avoidable Risks of Cancer in the United-States Today." Jnci-Journal of the National Cancer Institute **66**(6): 1191-+.

Doyle, L. M. and M. Z. Wang (2019). "Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis." Cells **8**(7).

Du Rand, I. A., P. V. Barber, J. Goldring, R. A. Lewis, S. Mandal, M. Munavvar, R. C. Rintoul, P. L. Shah, S. Singh, M. G. Slade, A. Woolley and G. British Thoracic Society Interventional Bronchoscopy Guideline (2011). "British Thoracic Society guideline for advanced diagnostic and therapeutic flexible bronchoscopy in adults." Thorax **66 Suppl 3**: iii1-21.

ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics_-_specific_cancers. (last accessed: 2022). Retrieved 17 Jan 2022, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics_-_specific_cancers.

Edelman, M. J. (2020). "Prophylactic Cranial Irradiation for Small-Cell Lung Cancer: Time for a Reassessment." Am Soc Clin Oncol Educ Book **40**: 24-28.

encyclopedia.pub/entry/683. (last accessed: 2022). Retrieved 19 Jan 2022, from https://encyclopedia.pub/entry/683.

Escrevente, C., S. Keller, P. Altevogt and J. Costa (2011). "Interaction and uptake of exosomes by ovarian cancer cells." BMC Cancer **11**: 108.

Ettinger, D. S. and J. Aisner (2006). "Changing face of small-cell lung cancer: real and artifact." J Clin Oncol **24**(28): 4526-4527.

Ettinger, D. S., W. Akerley, H. Borghaei, A. C. Chang, R. T. Cheney, L. R. Chirieac, T. A. D'Amico, T. L. Demmy, R. Govindan, F. W. Grannis, Jr., S. C. Grant, L. Horn, T. M. Jahan, R. Komaki, F. M. Kong, M. G. Kris, L. M. Krug, R. P. Lackner, I. T. Lennes, B. W. Loo, Jr., R. Martins, G. A. Otterson, J. D. Patel, M. C. Pinder-Schenck, K. M. Pisters, K. Reckamp, G. J. Riely, E. Rohren, T. A. Shapiro, S. J. Swanson, K. Tauer, D. E. Wood, S. C. Yang, K. Gregory, M. Hughes and n. National comprehensive cancer (2013). "Non-small cell lung cancer, version 2.2013." J Natl Compr Canc Netw **11**(6): 645-653; quiz 653.

Evison, M. and U. K. L. AstraZeneca (2020). "The current treatment landscape in the UK for stage III NSCLC." Br J Cancer **123**(Suppl 1): 3-9.

Faguet, G. B. (2015). "A brief history of cancer: age-old milestones underlying our current knowledge database." Int J Cancer **136**(9): 2022-2036.

Felip, E., N. Altorki, C. Zhou, T. Csoszi, I. Vynnychenko, O. Goloborodko, A. Luft, A. Akopov, A. Martinez-Marti, H. Kenmotsu, Y. M. Chen, A. Chella, S. Sugawara, D. Voong, F. Wu, J. Yi, Y. Deng, M. McCleland, E. Bennett, B. Gitlitz, H. Wakelee and I. M. Investigators (2021). "Adjuvant atezolizumab after adjuvant chemotherapy in

resected stage IB-IIIA non-small-cell lung cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial." Lancet **398**(10308): 1344-1357.

Field, J. K., S. W. Duffy, D. R. Baldwin, K. E. Brain, A. Devaraj, T. Eisen, B. A. Green, J. A. Holemans, T. Kavanagh, K. M. Kerr, M. Ledson, K. J. Lifford, F. E. McRonald, A. Nair, R. D. Page, M. K. Parmar, R. C. Rintoul, N. Screaton, N. J. Wald, D. Weller, D. K. Whynes, P. R. Williamson, G. Yadegarfar and D. M. Hansell (2016). "The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer." Health Technol Assess **20**(40): 1-146.

Field, J. K., S. W. Duffy, D. R. Baldwin, D. K. Whynes, A. Devaraj, K. E. Brain, T. Eisen, J. Gosney, B. A. Green, J. A. Holemans, T. Kavanagh, K. M. Kerr, M. Ledson, K. J. Lifford, F. E. McRonald, A. Nair, R. D. Page, M. K. Parmar, D. M. Rassl, R. C. Rintoul, N. J. Screaton, N. J. Wald, D. Weller, P. R. Williamson, G. Yadegarfar and D. M. Hansell (2016). "UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening." Thorax **71**(2): 161-170.

Field, J. K., D. L. Smith, S. Duffy and A. Cassidy (2005). "The Liverpool Lung Project research protocol." Int J Oncol **27**(6): 1633-1645.

Field, J. K., D. Vulkan, M. P. A. Davies, S. W. Duffy and R. Gabe (2021). "Liverpool Lung Project lung cancer risk stratification model: calibration and prospective validation." Thorax **76**(2): 161-168.

Force, L. M., I. Abdollahpour, S. M. Advani, D. Agius, E. Ahmadian, F. Alahdab, T. Alam, A. Alebel, V. Alipour, C. A. Allen, A. Almasi-Hashiani, E. M. Alvarez, S. Amini, Y. A. Amoako, N. H. Anber, J. Arabloo, A. Artaman, S. Atique, A. Awasthi, M. Bagherzadeh, H. Basaleem, E. T. Bekru, A. Bijani, K. A. Bogale, M. Car, F. Carvalho, C. Castro, F. Catalá-López, D.-T. Chu, V. M. Costa, A. H. Darwish, F. M. Demeke, A. B. Demis, G. T. Demoz, S. D. Dharmaratne, H. P. Do, L. P. Doan, M. Dubey, A. Eftekhari, Z. El-Khatib, M. H. Emamian, M. A. Farhangi, E. Fernandes, F. Fischer, R. Fouladi Fard, P. M. Friedrich, T. Fukumoto, G. A. Gedefaw, A. Ghashghaee, A. Gholamian, A. Haj-Mirzaian, A. Haj-Mirzaian, S. Hamidi, J. D. Harvey, H. Y. Hassen, S. I. Hay, C. L. Hoang, M. K. Hole, N. Horita, S. N. Hosseini, M. Hosseinzadeh, M. Hsairi, M. M. Hudson, K. Innos, F. Jalilian, S. L. James, A. Kasaeian, T. D. Kassa, N. J. Kassebaum, P. N. Keiyoro, Y. S. Khader, J. Khubchandani, N. Kianipour, J. Kirby, A. Kisa, S. Kisa, J. M. Kocarnik, P. Lauriola, A. D. Lopez, M. Mägi, M. A. Malik, A. Manafi, N. Manafi, M. A. Mansournia, B. B. Massenburg, V. Mehta, H. G. Meles, T. J. Meretoja, T. Mestrovic, S. M. Mir, M. Mirzaei-Alavijeh, D. K. Mohammad, A. Mohammad Darwesh, N. Mohammad Gholi Mezerji, R. Mohammadibakhsh, M. Mohammadoo-Khorasani, A. H. Mokdad, Y. Moodley, M. Moosazadeh, M. Moossavi, F. Moradpour, S. D. Morrison, K. F. Muchie, M. Naghavi, J. Nazari, J. W. Ngunjiri, C. T. Nguyen, L. H. Nguyen, S. H. Nguyen, T. H. Nguyen, M. R. Nixon, A. T. Olagunju, T. O. Olagunju, S. K. Ong, K. Paapsi, A. Pennini, D. M. Pereira, J. D. Pillay, M. Qorbani, M. Rabiee, N. Rabiee, S. Raoofi, D. L. Rawaf, S. Rawaf, R. C. Reiner, N. Rezaei, A. Rezapour, K. T. Roba, L. L. Robison, C. Rodriguez-Galindo, G. Roshandel, S. Safiri, M. Salahshoor, S. Salehi Zahabi, A. M. Samy, M. M. Santric Milicevic, M. Satpathy, S. M. Sawyer, S. Seyedmousavi, H. Shabaninejad, M. A. Shaikh, A. Shamshirian, M. Shamsizadeh, R. Shirkoohi, S. Siabani, J. A. Singh, K. SiRamlee, R. Tabarés-Seisdedos, M.-H. Temsah, B. X. Tran, I. Ullah, A. Vahedian-Azimi, S. E. Vollset, T. Vos, Y. Waheed, G. T. Weldesamuel, H. M. Workie, R. Xu, M. Yaseri, N. Yonemoto, C. Yu, V. Zadnik, T. Zahirian Moghadam, Z. Zaidi, A. Zangeneh, T. A. Zewale, A. Ziapour, S. Zodpey, C. J. L. Murray, N. Bhakta and C. Fitzmaurice (2019). "The global burden of childhood and adolescent cancer in 2017: an analysis of the Global Burden of Disease Study 2017." The Lancet Oncology **20**(9): 1211-1225.

Force, U. S. P. S. T., A. H. Krist, K. W. Davidson, C. M. Mangione, M. J. Barry, M. Cabana, A. B. Caughey, E. M. Davis, K. E. Donahue, C. A. Doubeni, M. Kubik, C. S. Landefeld, L. Li, G. Ogedegbe, D. K. Owens, L. Pbert, M. Silverstein, J. Stevermer, C. W. Tseng and J. B. Wong (2021). "Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement." JAMA **325**(10): 962-970.

Frisch, J., A. Angenendt, M. Hoth, L. Prates Roma and A. Lis (2019). "STIM-Orai Channels and Reactive Oxygen Species in the Tumor Microenvironment." Cancers (Basel) **11**(4).

Fruhbeis, C., S. Helmig, S. Tug, P. Simon and E. M. Kramer-Albers (2015). "Physical exercise induces rapid release of small extracellular vesicles into the circulation." J Extracell Vesicles **4**: 28239.

Garzon, R., G. A. Calin and C. M. Croce (2009). "MicroRNAs in Cancer." Annu Rev Med **60**: 167-179.

gco.iarc.fr/today/data/factsheets/populations/908-europe-fact-sheets.pdf. (last accessed: 2022). Retrieved 8 May 2022, from https://gco.iarc.fr/today/data/factsheets/populations/908-europe-fact-sheets.pdf.

George, J., J. S. Lim, S. J. Jang, Y. Cun, L. Ozretic, G. Kong, F. Leenders, X. Lu, L. Fernandez-Cuesta, G. Bosco, C. Muller, I. Dahmen, N. S. Jahchan, K. S. Park, D. Yang, A. N. Karnezis, D. Vaka, A. Torres, M. S. Wang, J. O. Korbel, R. Menon, S. M. Chun, D. Kim, M. Wilkerson, N. Hayes, D. Engelmann, B. Putzer, M. Bos, S. Michels, I. Vlasic, D. Seidel, B. Pinther, P. Schaub, C. Becker, J. Altmuller, J. Yokota, T. Kohno, R. Iwakawa, K. Tsuta, M.

Noguchi, T. Muley, H. Hoffmann, P. A. Schnabel, I. Petersen, Y. Chen, A. Soltermann, V. Tischler, C. M. Choi, Y. H. Kim, P. P. Massion, Y. Zou, D. Jovanovic, M. Kontic, G. M. Wright, P. A. Russell, B. Solomon, I. Koch, M. Lindner, L. A. Muscarella, A. la Torre, J. K. Field, M. Jakopovic, J. Knezevic, E. Castanos-Velez, L. Roz, U. Pastorino, O. T. Brustugun, M. Lund-Iversen, E. Thunnissen, J. Kohler, M. Schuler, J. Botling, M. Sandelin, M. Sanchez-Cespedes, H. B. Salvesen, V. Achter, U. Lang, M. Bogus, P. M. Schneider, T. Zander, S. Ansen, M. Hallek, J. Wolf, M. Vingron, Y. Yatabe, W. D. Travis, P. Nurnberg, C. Reinhardt, S. Perner, L. Heukamp, R. Buttner, S. A. Haas, E. Brambilla, M. Peifer, J. Sage and R. K. Thomas (2015). "Comprehensive genomic profiles of small cell lung cancer." Nature **524**(7563): 47-53.

Gibbons, D. L., L. A. Byers and J. M. Kurie (2014). "Smoking, p53 mutation, and lung cancer." Mol Cancer Res **12**(1): 3-13.

Gildea, J. J., J. E. Seaton, K. G. Victor, C. M. Reyes, D. Bigler Wang, A. C. Pettigrew, C. E. Courtner, N. Shah, H. T. Tran, R. E. Van Sciver, J. M. Carlson and R. A. Felder (2014). "Exosomal transfer from human renal proximal tubule cells to distal tubule and collecting duct cells." Clin Biochem **47**(15): 89-94.

Godoy, P. M., A. J. Barczak, P. DeHoff, S. Srinivasan, A. Etheridge, D. Galas, S. Das, D. J. Erle and L. C. Laurent (2019). "Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA Quantification Platforms." Cell Rep **29**(12): 4212-4222 e4215.

Goldman, R., L. Enewold, E. Pellizzari, J. B. Beach, E. D. Bowman, S. S. Krishnan and P. G. Shields (2001). "Smoking increases carcinogenic polycyclic aromatic hydrocarbons in human lung tissue." Cancer Research **61**(17): 6367-6371.

Gondi, V., J. Meyer and H. A. Shih (2021). "Advances in radiotherapy for brain metastases." Neurooncol Adv **3**(Suppl 5): v26-v34.

Goulart, B. H. and S. D. Ramsey (2013). "Moving beyond the national lung screening trial: discussing strategies for implementation of lung cancer screening programs." Oncologist **18**(8): 941-946.

Greco, V., M. Hannus and S. Eaton (2001). "Argosomes: a potential vehicle for the spread of morphogens through epithelia." Cell **106**(5): 633-645.

Greenblatt, M. S., W. P. Bennett, M. Hollstein and C. C. Harris (1994). "Mutations in the P53 Tumor-Suppressor Gene - Clues to Cancer Etiology and Molecular Pathogenesis." Cancer Research **54**(18): 4855-4878.

Gridelli, C., A. Rossi, D. P. Carbone, J. Guarize, N. Karachaliou, T. Mok, F. Petrella, L. Spaggiari and R. Rosell (2015). "Non-small-cell lung cancer." Nat Rev Dis Primers **1**: 15009.

Group, N. M.-a. C., R. Arriagada, A. Auperin, S. Burdett, J. P. Higgins, D. H. Johnson, T. Le Chevalier, C. Le Pechoux, M. K. Parmar, J. P. Pignon, R. L. Souhami, R. J. Stephens, L. A. Stewart, J. F. Tierney, H. Tribodet and J. van Meerbeeck (2010). "Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small-cell lung cancer: two meta-analyses of individual patient data." Lancet **375**(9722): 1267-1277.

Grunnet, M. and J. B. Sorensen (2012). "Carcinoembryonic antigen (CEA) as tumor marker in lung cancer." Lung Cancer **76**(2): 138-143.

Guescini, M., S. Genedani, V. Stocchi and L. F. Agnati (2010). "Astrocytes and Glioblastoma cells release exosomes carrying mtDNA." J Neural Transm (Vienna) **117**(1): 1-4.

Guo, F., F. Yu, J. Wang, Y. Li, Y. Li, Z. Li and Q. Zhou (2015). "Expression of MALAT1 in the peripheral whole blood of patients with lung cancer." Biomed Rep **3**(3): 309-312.

Guo, Z., H. Wang, F. Zhao, M. Liu, F. Wang, M. Kang, W. He and Z. Lv (2021). "Exosomal circ-BRWD1 contributes to osteoarthritis development through the modulation of miR-1277/TRAF6 axis." Arthritis Res Ther **23**(1): 159.

Gupta, K. H., J. W. Goldufsky, S. J. Wood, N. J. Tardi, G. S. Moorthy, D. Z. Gilbert, J. P. Zayas, E. Hahm, M. M. Altintas, J. Reiser and S. H. Shafikhani (2017). "Apoptosis and Compensatory Proliferation Signaling Are Coupled by CrkI-Containing Microvesicles." Dev Cell **41**(6): 674-684 e675.

Hainaut, P. and G. P. Pfeifer (2001). "Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke." Carcinogenesis **22**(3): 367-374.

Hanahan, D. (2022). "Hallmarks of Cancer: New Dimensions." Cancer Discov **12**(1): 31-46.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.

Hanna, N. H., A. G. Robinson, S. Temin, S. Baker, Jr., J. R. Brahmer, P. M. Ellis, L. E. Gaspar, R. Y. Haddad, P. J. Hesketh, D. Jain, I. Jaiyesimi, D. H. Johnson, N. B. Leighl, P. R. Moffitt, T. Phillips, G. J. Riely, R. Rosell, J. H. Schiller, B. J. Schneider, N. Singh, D. R. Spigel, J. Tashbar and G. Masters (2021). "Therapy for Stage IV Non-

Small-Cell Lung Cancer With Driver Alterations: ASCO and OH (CCO) Joint Guideline Update." J Clin Oncol **39**(9): 1040-1091.

Hannafon, B. N. and W. Q. Ding (2013). "Intercellular communication by exosome-derived microRNAs in cancer." Int J Mol Sci **14**(7): 14240-14269.

Hao, S., Z. Ye, F. Li, Q. Meng, M. Qureshi, J. Yang and J. Xiang (2006). "Epigenetic transfer of metastatic activity by uptake of highly metastatic B16 melanoma cell-released exosomes." Exp Oncol **28**(2): 126-131.

Hasan, H., I. S. Sohal, Z. Soto-Vargas, A. M. Byappanahalli, S. E. Humphrey, H. Kubo, S. Kitdumrongthum, S. Copeland, F. Tian, A. Chairoungdua and A. L. Kasinski (2022). "Extracellular vesicles released by non-small cell lung cancer cells drive invasion and permeability in non-tumorigenic lung epithelial cells." Sci Rep **12**(1): 972.

Hassanein, M., J. C. Callison, C. Callaway-Lane, M. C. Aldrich, E. L. Grogan and P. P. Massion (2012). "The state of molecular biomarkers for the early detection of lung cancer." Cancer Prev Res (Phila) **5**(8): 992-1006.

Hayes, J., P. P. Peruzzi and S. Lawler (2014). "MicroRNAs in cancer: biomarkers, functions and therapy." Trends Mol Med **20**(8): 460-469.

Health, N. C. f. C. D. P. a. H. P. U. O. o. S. a. (2014). "The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General."

Hennessey, P. T., T. Sanford, A. Choudhary, W. W. Mydlarz, D. Brown, A. T. Adai, M. F. Ochs, S. A. Ahrendt, E. Mambo and J. A. Califano (2012). "Serum microRNA biomarkers for detection of non-small cell lung cancer." PLoS One **7**(2): e32307.

Herbst, R. S., J. V. Heymach and S. M. Lippman (2008). "Lung cancer." N Engl J Med **359**(13): 1367-1380.

Hernandez-Boussard, T. M. and P. Hainaut (1998). "A specific spectrum of p53 mutations in lung cancer from smokers: Review of mutations compiled in the IARC p53 database." Environmental Health Perspectives **106**(7): 385-391.

Higgins, G., K. M. Roper, I. J. Watson, F. H. Blackhall, W. N. Rom, H. I. Pass, J. F. Ainscough and D. Coverley (2012). "Variant Ciz1 is a circulating biomarker for early-stage lung cancer." Proc Natl Acad Sci U S A **109**(45): E3128-3135.

Horeweg, N., C. M. van der Aalst, R. Vliegenthart, Y. Zhao, X. Xie, E. T. Scholten, W. Mali, E. Thunnissen, C. Weenink, H. J. Groen, J. W. Lammers, K. Nackaerts, J. van Rosmalen, M. Oudkerk and H. J. de Koning (2013). "Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial." Eur Respir J **42**(6): 1659-1667.

Hoshino, A., B. Costa-Silva, T. L. Shen, G. Rodrigues, A. Hashimoto, M. Tesic Mark, H. Molina, S. Kohsaka, A. Di Giannatale, S. Ceder, S. Singh, C. Williams, N. Soplop, K. Uryu, L. Pharmer, T. King, L. Bojmar, A. E. Davies, Y. Ararso, T. Zhang, H. Zhang, J. Hernandez, J. M. Weiss, V. D. Dumont-Cole, K. Kramer, L. H. Wexler, A. Narendran, G. K. Schwartz, J. H. Healey, P. Sandstrom, K. J. Labori, E. H. Kure, P. M. Grandgenett, M. A. Hollingsworth, M. de Sousa, S. Kaur, M. Jain, K. Mallya, S. K. Batra, W. R. Jarnagin, M. S. Brady, O. Fodstad, V. Muller, K. Pantel, A. J. Minn, M. J. Bissell, B. A. Garcia, Y. Kang, V. K. Rajasekhar, C. M. Ghajar, I. Matei, H. Peinado, J. Bromberg and D. Lyden (2015). "Tumour exosome integrins determine organotropic metastasis." Nature **527**(7578): 329-335.

HQIP, N. (2011). National Lung Cancer Audit Report 2011.

HTG. (2019). "A Research Use Only White Paper for HTG EdgeSeq miRNA Whole Transcriptome Assay QC Metrics." from https://www.htgmolecular.com/assets/htg/support/A_Research_Use_Only_HTG_EdgeSeq_miRNA_WTA_QC_Metrics_White_Paper.pdf.

Hu, X., J. Bao, Z. Wang, Z. Zhang, P. Gu, F. Tao, D. Cui and W. Jiang (2016). "The plasma lncRNA acting as fingerprint in non-small-cell lung cancer." Tumour Biol **37**(3): 3497-3504.

Hu, X., J. Miao, M. Zhang, X. Wang, Z. Wang, J. Han, D. Tong and C. Huang (2018). "miRNA-103a-3p Promotes Human Gastric Cancer Cell Proliferation by Targeting and Suppressing ATF7 in vitro." Mol Cells **41**(5): 390-400.

Hung, R. J., J. D. McKay, V. Gaborieau, P. Boffetta, M. Hashibe, D. Zaridze, A. Mukeria, N. Szeszenia-Dabrowska, J. Lissowska, P. Rudnai, E. Fabianova, D. Mates, V. Bencko, L. Foretova, V. Janout, C. Chen, G. Goodman, J. K. Field, T. Liloglou, G. Xinarianos, A. Cassidy, J. McLaughlin, G. Liu, S. Narod, H. E. Krokan, F. Skorpen, M. B. Elvestad, K. Hveem, L. Vatten, J. Linseisen, F. Clavel-Chapelon, P. Vineis, H. B. Bueno-de-Mesquita, E. Lund, C. Martinez, S. Bingham, T. Rasmuson, P. Hainaut, E. Riboli, W. Ahrens, S. Benhamou, P. Lagiou, D. Trichopoulos, I. Holcatova, F. Merletti, K. Kjaerheim, A. Agudo, G. Macfarlane, R. Talamini, L. Simonato, R. Lowry, D. I. Conway, A. Znaor, C. Healy, D. Zelenika, A. Boland, M. Delepine, M. Foglio, D. Lechner, F. Matsuda,

H. Blanche, I. Gut, S. Heath, M. Lathrop and P. Brennan (2008). "A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25." Nature **452**(7187): 633-637.

Hurter, T. and P. Hanrath (1992). "Endobronchial sonography: feasibility and preliminary results." Thorax **47**(7): 565-567.

IARC (2004). Tobacco Smoke and Involuntary Smoking.

Jacobaeus, H. C. (1910). "Uber die Moglichkeit, die Zystoskopie bei Untersuchung seroser Hohlungen anzuwenden." Munch Med Wschr. **40**: 2090-2092.

Jahan, S., S. Mukherjee, S. Ali, U. Bhardwaj, R. K. Choudhary, S. Balakrishnan, A. Naseem, S. A. Mir, S. Banawas, M. Alaidarous, H. Alyenbaawi, D. Iqbal and A. J. Siddiqui (2022). "Pioneer Role of Extracellular Vesicles as Modulators of Cancer Initiation in Progression, Drug Therapy, and Vaccine Prospects." Cells **11**(3).

Jemal, A., T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer and M. J. Thun (2005). "Cancer statistics, 2005." CA Cancer J Clin **55**(1): 10-30.

Jeon, J., T. R. Holford, D. T. Levy, E. J. Feuer, P. Cao, J. Tam, L. Clarke, J. Clarke, C. Y. Kong and R. Meza (2018). "Smoking and Lung Cancer Mortality in the United States From 2015 to 2065 A Comparative Modeling Approach." Annals of Internal Medicine **169**(10): 684-+.

Jett, J. R., S. E. Schild, K. A. Kesler and G. P. Kalemkerian (2013). "Treatment of small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines." Chest **143**(5 Suppl): e400S-e419S.

Jia, X. H., H. Xu, L. Y. Geng, M. Jiao, W. J. Wang, L. L. Jiang and H. Guo (2020). "Efficacy and safety of neoadjuvant immunotherapy in resectable nonsmall cell lung cancer: A meta-analysis." Lung Cancer **147**: 143-153.

Johnson, D. H., L. Fehrenbacher, W. F. Novotny, R. S. Herbst, J. J. Nemunaitis, D. M. Jablons, C. J. Langer, R. F. DeVore, 3rd, J. Gaudreault, L. A. Damico, E. Holmgren and F. Kabbinavar (2004). "Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer." J Clin Oncol **22**(11): 2184-2191.

Johny A. Verschakelen, W. D. W. (2018). Computed Tomography of the Lung, Springer.

Kalemkerian, G. P., W. Akerley, P. Bogner, H. Borghaei, L. Q. Chow, R. J. Downey, L. Gandhi, A. K. Ganti, R. Govindan, J. C. Grecula, J. Hayman, R. S. Heist, L. Horn, T. Jahan, M. Koczywas, B. W. Loo, Jr., R. E. Merritt, C. A. Moran, H. B. Niell, J. O'Malley, J. D. Patel, N. Ready, C. M. Rudin, C. C. Williams, Jr., K. Gregory, M. Hughes and N. National Comprehensive Cancer (2013). "Small cell lung cancer." J Natl Compr Canc Netw **11**(1): 78-98.

Kastenhuber, E. R. and S. W. Lowe (2017). "Putting p53 in Context." Cell **170**(6): 1062-1078.

Kenfield, S. A., E. K. Wei, M. J. Stampfer, B. A. Rosner and G. A. Colditz (2008). "Comparison of aspects of smoking among the four histological types of lung cancer." Tob Control **17**(3): 198-204.

Kerr, K. M., L. Bubendorf, M. J. Edelman, A. Marchetti, T. Mok, S. Novello, K. O'Byrne, R. Stahel, S. Peters, E. Felip and M. Panel (2014). "Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer." Ann Oncol **25**(9): 1681-1690.

Kuriyama, N., Y. Yoshioka, S. Kikuchi, N. Azuma and T. Ochiya (2020). "Extracellular Vesicles Are Key Regulators of Tumor Neovasculature." Front Cell Dev Biol **8**: 611039.

Lally, B. E., D. Zelterman, J. M. Colasanto, B. G. Haffty, F. C. Detterbeck and L. D. Wilson (2006). "Postoperative radiotherapy for stage II or III non-small-cell lung cancer using the surveillance, epidemiology, and end results database." J Clin Oncol **24**(19): 2998-3006.

Larke, F. J., R. L. Kruger, C. H. Cagnon, M. J. Flynn, M. M. McNitt-Gray, X. Wu, P. F. Judy and D. D. Cody (2011). "Estimated radiation dose associated with low-dose chest CT of average-size participants in the National Lung Screening Trial." AJR Am J Roentgenol **197**(5): 1165-1169.

Li, H., M. Huhe and J. Lou (2021). "MicroRNA-103a-3p Promotes Cell Proliferation and Invasion in Non-Small-Cell Lung Cancer Cells through Akt Pathway by Targeting PTEN." Biomed Res Int **2021**: 7590976.

Li, H., Q. Zhao and Z. Tang (2021). "LncRNA RP11-116G8.5 promotes the progression of lung squamous cell carcinoma through sponging miR-3150b-3p/miR-6870-5p to upregulate PHF12/FOXP4." Pathol Res Pract **226**: 153566.

Li, W., J. B. Liu, L. K. Hou, F. Yu, J. Zhang, W. Wu, X. M. Tang, F. Sun, H. M. Lu, J. Deng, J. Bai, J. Li, C. Y. Wu, Q. L. Lin, Z. W. Lv, G. R. Wang, G. X. Jiang, Y. S. Ma and D. Fu (2022). "Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring." Mol Cancer **21**(1): 25.

Li, X. and K. Hemminki (2004). "Inherited predisposition to early onset lung cancer according to histological type." Int J Cancer **112**(3): 451-457.

Liang, W., T. Lv, X. Shi, H. Liu, Q. Zhu, J. Zeng, W. Yang, J. Yin and Y. Song (2016). "Circulating long noncoding RNA GAS5 is a novel biomarker for the diagnosis of nonsmall cell lung cancer." Medicine (Baltimore) **95**(37): e4608.

Lin, P. Y., S. L. Yu and P. C. Yang (2010). "MicroRNA in lung cancer." Br J Cancer **103**(8): 1144-1148.

Lindberg, K., J. Nyman, V. Riesenfeld Kallskog, M. Hoyer, J. A. Lund, I. Lax, P. Wersall, K. Karlsson, S. Friesland and R. Lewensohn (2015). "Long-term results of a prospective phase II trial of medically inoperable stage I NSCLC treated with SBRT - the Nordic experience." Acta Oncol **54**(8): 1096-1104.

Linet, M. S., K. P. Kim and P. Rajaraman (2009). "Children's exposure to diagnostic medical radiation and cancer risk: epidemiologic and dosimetric considerations." Pediatr Radiol **39 Suppl 1**: S4-26.

Lopatina, T., C. Gai, M. C. Deregibus, S. Kholia and G. Camussi (2016). "Cross Talk between Cancer and Mesenchymal Stem Cells through Extracellular Vesicles Carrying Nucleic Acids." Front Oncol **6**: 125.

Lortet-Tieulent, J., D. Georges, F. Bray and S. Vaccarella (2020). "Profiling global cancer incidence and mortality by socioeconomic development." Int J Cancer **147**(11): 3029-3036.

Lu, T., X. Yang, Y. Huang, M. Zhao, M. Li, K. Ma, J. Yin, C. Zhan and Q. Wang (2019). "Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades." Cancer Manag Res **11**: 943-953.

Ma, F., J. Vayalil, G. Lee, Y. Wang and G. Peng (2021). "Emerging role of tumor-derived extracellular vesicles in T cell suppression and dysfunction in the tumor microenvironment." J Immunother Cancer **9**(10).

Ma, J., X. Li, S. Zhao, J. Wang, W. Zhang and G. Sun (2021). "Robot-assisted thoracic surgery versus video-assisted thoracic surgery for lung lobectomy or segmentectomy in patients with non-small cell lung cancer: a meta-analysis." BMC Cancer **21**(1): 498.

Ma, Q., H. Wu, Y. Xiao, Z. Liang and T. Liu (2020). "Upregulation of exosomal microRNA21 in pancreatic stellate cells promotes pancreatic cancer cell migration and enhances Ras/ERK pathway activity." Int J Oncol **56**(4): 1025-1033.

Marino, P., A. Preatoni and A. Cantoni (1995). "Randomized trials of radiotherapy alone versus combined chemotherapy and radiotherapy in stages IIIa and IIIb nonsmall cell lung cancer. A meta-analysis." Cancer **76**(4): 593-601.

Mashouri, L., H. Yousefi, A. R. Aref, A. M. Ahadi, F. Molaei and S. K. Alahari (2019). "Exosomes: composition, biogenesis, and mechanisms in cancer metastasis and drug resistance." Mol Cancer **18**(1): 75.

Masters, G. A., S. Temin, C. G. Azzoli, G. Giaccone, S. Baker, Jr., J. R. Brahmer, P. M. Ellis, A. Gajra, N. Rackear, J. H. Schiller, T. J. Smith, J. R. Strawn, D. Trent, D. H. Johnson and P. American Society of Clinical Oncology Clinical (2015). "Systemic Therapy for Stage IV Non-Small-Cell Lung Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update." J Clin Oncol **33**(30): 3488-3515.

Mauguen, A., C. Le Pechoux, M. I. Saunders, S. E. Schild, A. T. Turrisi, M. Baumann, W. T. Sause, D. Ball, C. P. Belani, J. A. Bonner, A. Zajusz, S. E. Dahlberg, M. Nankivell, S. J. Mandrekar, R. Paulus, K. Behrendt, R. Koch, J. F. Bishop, S. Dische, R. Arriagada, D. De Ruysscher and J. P. Pignon (2012). "Hyperfractionated or accelerated radiotherapy in lung cancer: an individual patient data meta-analysis." J Clin Oncol **30**(22): 2788-2797.

McCarthy, D. J., Y. Chen and G. K. Smyth (2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation." Nucleic Acids Res **40**(10): 4288-4297.

McCready, J., J. D. Sims, D. Chan and D. G. Jay (2010). "Secretion of extracellular hsp90alpha via exosomes increases cancer cell motility: a role for plasminogen activation." BMC Cancer **10**: 294.

McPhail, S., S. Johnson, D. Greenberg, M. Peake and B. Rous (2015). "Stage at diagnosis and early mortality from cancer in England." Br J Cancer **112 Suppl 1**: S108-115.

Mehic, B., L. Duranovic Rayan, N. Bilalovic, D. Dohranovic Tafro and I. Pilav (2016). "Lung adenocarcinoma mimicking pulmonary fibrosis-a case report." BMC Cancer **16**(1): 729.

Mettler, F. A., Jr., W. Huda, T. T. Yoshizumi and M. Mahesh (2008). "Effective doses in radiology and diagnostic nuclear medicine: a catalog." Radiology **248**(1): 254-263.

Milasan, A., N. Tessandier, S. Tan, A. Brisson, E. Boilard and C. Martel (2016). "Extracellular vesicles are present in mouse lymph and their level differs in atherosclerosis." J Extracell Vesicles **5**: 31427.

Miller, K. D., L. Nogueira, A. B. Mariotto, J. H. Rowland, K. R. Yabroff, C. M. Alfano, A. Jemal, J. L. Kramer and R. L. Siegel (2019). "Cancer treatment and survivorship statistics, 2019." CA Cancer J Clin **69**(5): 363-385.

Milman, N. (1995). "Percutaneous lung biopsy with a fine bore cutting needle (Vacu-Cut): improved results using drill technique." Thorax **50**(5): 560-562.

Mitchell, P. S., R. K. Parkin, E. M. Kroh, B. R. Fritz, S. K. Wyman, E. L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K. C. O'Briant, A. Allen, D. W. Lin, N. Urban, C. W. Drescher, B. S. Knudsen, D. L. Stirewalt, R. Gentleman, R. L. Vessella, P. S. Nelson, D. B. Martin and M. Tewari (2008). "Circulating microRNAs as stable blood-based markers for cancer detection." Proc Natl Acad Sci U S A **105**(30): 10513-10518.

Mittelbrunn, M., C. Gutierrez-Vazquez, C. Villarroya-Beltri, S. Gonzalez, F. Sanchez-Cabo, M. A. Gonzalez, A. Bernad and F. Sanchez-Madrid (2011). "Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells." Nat Commun **2**: 282.

Mona SARAEI, A. G., Faranak JAMSHIDIAN (2021). "miR-7111-5p and miR6870-5p may be Potential Biomarkers for Oral Squamous Cell Carcinoma." Turk J Oncol **36(3):252–58**.

Montecalvo, A., A. T. Larregina, W. J. Shufesky, D. B. Stolz, M. L. Sullivan, J. M. Karlsson, C. J. Baty, G. A. Gibson, G. Erdos, Z. Wang, J. Milosevic, O. A. Tkacheva, S. J. Divito, R. Jordan, J. Lyons-Weiler, S. C. Watkins and A. E. Morelli (2012). "Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes." Blood **119**(3): 756-766.

Morelli, A. E., A. T. Larregina, W. J. Shufesky, M. L. Sullivan, D. B. Stolz, G. D. Papworth, A. F. Zahorchak, A. J. Logar, Z. Wang, S. C. Watkins, L. D. Falo, Jr. and A. W. Thomson (2004). "Endocytosis, intracellular sorting, and processing of exosomes by dendritic cells." Blood **104**(10): 3257-3266.

Mukhopadhyay, S. (2012). "Utility of small biopsies for diagnosis of lung nodules: doing more with less." Mod Pathol **25 Suppl 1**: S43-57.

Muller, K. M. (1984). "Histological classification and histogenesis of lung cancer." Eur J Respir Dis **65**(1): 4-19.

Mulvenna, P., M. Nankivell, R. Barton, C. Faivre-Finn, P. Wilson, E. McColl, B. Moore, I. Brisbane, D. Ardron, T. Holt, S. Morgan, C. Lee, K. Waite, N. Bayman, C. Pugh, B. Sydes, R. Stephens, M. K. Parmar and R. E. Langley (2016). "Dexamethasone and supportive care with or without whole brain radiotherapy in treating patients with non-small cell lung cancer with brain metastases unsuitable for resection or stereotactic radiotherapy (QUARTZ): results from a phase 3, non-inferiority, randomised trial." Lancet **388**(10055): 2004-2014.

Murray, N. and A. T. Turrisi, 3rd (2006). "A review of first-line treatment for small-cell lung cancer." J Thorac Oncol **1**(3): 270-278.

Muscat, J. E. and E. L. Wynder (1995). "Lung cancer pathology in smokers, ex-smokers and never smokers." Cancer Letters **88**(1): 1-5.

Naderi-Meshkin, H., X. Lai, R. Amirkhah, J. Vera, J. E. J. Rasko and U. Schmitz (2019). "Exosomal lncRNAs and cancer: connecting the missing links." Bioinformatics **35**(2): 352-360.

Nafstad, P., L. L. Haheim, B. Oftedal, F. Gram, I. Holme, I. Hjermann and P. Leren (2003). "Lung cancer and air pollution: a 27 year follow up of 16 209 Norwegian men." Thorax **58**(12): 1071-1076.

National Lung Screening Trial Research, T., D. R. Aberle, A. M. Adams, C. D. Berg, W. C. Black, J. D. Clapp, R. M. Fagerstrom, I. F. Gareen, C. Gatsonis, P. M. Marcus and J. D. Sicks (2011). "Reduced lung-cancer mortality with low-dose computed tomographic screening." N Engl J Med **365**(5): 395-409.

Navani, N., M. Nankivell, D. R. Lawrence, S. Lock, H. Makker, D. R. Baldwin, R. J. Stephens, M. K. Parmar, S. G. Spiro, S. Morris and S. M. Janes (2015). "Lung cancer diagnosis and staging with endobronchial ultrasound-guided transbronchial needle aspiration compared with conventional approaches: an open-label, pragmatic, randomised controlled trial." The Lancet Respiratory Medicine **3**(4): 282-289.

Nerlich, A. G., H. Rohrbach, B. Bachmeier and A. Zink (2006). "Malignant tumors in two ancient populations: An approach to historical tumor epidemiology." Oncology Reports **16**(1): 197-202.

NHS (2019). Targeted Screening for Lung Cancer with Low Radiation Dose Computed Tomography. NHS.

Nicholson, A. G., K. Chansky, J. Crowley, R. Beyruti, K. Kubota, A. Turrisi, W. E. Eberhardt, J. van Meerbeeck, R. Rami-Porta, Staging, A. B. Prognostic Factors Committee, I. Participating, Staging, B. Prognostic Factors Committee Advisory and I. Participating (2016). "The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revision of the Clinical and Pathologic Staging of Small Cell Lung Cancer in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer." J Thorac Oncol **11**(3): 300-311.

Nikolaidis, G., O. Y. Raji, S. Markopoulou, J. R. Gosney, J. Bryan, C. Warburton, M. Walshaw, J. Sheard, J. K. Field and T. Liloglou (2012). "DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer." Cancer Res **72**(22): 5692-5701.

Nishimura, Y., C. Kondo, Y. Morikawa, Y. Tonomura, M. Torii, J. Yamate and T. Uehara (2015). "Plasma miR-208 as a useful biomarker for drug-induced cardiotoxicity in rats." J Appl Toxicol **35**(2): 173-180.

O'Farrell, H. E., R. V. Bowman, K. M. Fong and I. A. Yang (2021). "Plasma Extracellular Vesicle miRNAs Can Identify Lung Cancer, Current Smoking Status, and Stable COPD." Int J Mol Sci **22**(11).

Ocak, S., F. Duplaquet, J. Jamart, L. Pirard, B. Weynand, M. Delos, P. Eucher, B. Rondelet, M. Dupont, L. Delaunois, Y. Sibille and C. Dahlqvist (2016). "Diagnostic Accuracy and Safety of CT-Guided Percutaneous Transthoracic Needle Biopsies: 14-Gauge versus 22-Gauge Needles." J Vasc Interv Radiol **27**(5): 674-681.

Osterlind, K. (2001). "Chemotherapy in small cell lung cancer." Eur Respir J **18**(6): 1026-1043.

Oswald, N. C., K. F. Hinson, G. Canti and A. B. Miller (1971). "The diagnosis of primary lung cancer with special reference to sputum cytology." Thorax **26**(6): 623-627.

Ou, H., K. Chen and H. Wu (2021). "MicroRNA Expression Profiling of Lung Cancer with Differential Expression of the RON Receptor Tyrosine Kinase." J Oncol **2021**: 5670675.

Paez, J. G., P. A. Janne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson and M. Meyerson (2004). "EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy." Science **304**(5676): 1497-1500.

Papi, A., G. Casoni, G. Caramori, I. Guzzinati, P. Boschetto, F. Ravenna, N. Calia, S. Petruzzelli, L. Corbetta, G. Cavallesco, E. Forini, M. Saetta, A. Ciaccia and L. M. Fabbri (2004). "COPD increases the risk of squamous histological subtype in smokers who develop non-small cell lung carcinoma." Thorax **59**(8): 679-681.

Park, H. Y., D. Kang, S. H. Shin, K. H. Yoo, C. K. Rhee, G. Y. Suh, H. Kim, Y. M. Shim, E. Guallar, J. Cho and O. J. Kwon (2020). "Chronic obstructive pulmonary disease and lung cancer incidence in never smokers: a cohort study." Thorax **75**(6): 506-509.

Patkar, V., D. Acosta, T. Davidson, A. Jones, J. Fox and M. Keshtgar (2011). "Cancer multidisciplinary team meetings: evidence, challenges, and the role of clinical decision support technology." Int J Breast Cancer **2011**: 831605.

Patz, E. F., Jr., M. J. Campa, E. B. Gottlin, I. Kusmartseva, X. R. Guan and J. E. Herndon, 2nd (2007). "Panel of serum biomarkers for the diagnosis of lung cancer." J Clin Oncol **25**(35): 5578-5583.

Pegtel, D. M., K. Cosmopoulos, D. A. Thorley-Lawson, M. A. van Eijndhoven, E. S. Hopmans, J. L. Lindenberg, T. D. de Gruijl, T. Wurdinger and J. M. Middeldorp (2010). "Functional delivery of viral miRNAs via exosomes." Proc Natl Acad Sci U S A **107**(14): 6328-6333.

Peinado, H., M. Aleckovic, S. Lavotshkin, I. Matei, B. Costa-Silva, G. Moreno-Bueno, M. Hergueta-Redondo, C. Williams, G. Garcia-Santos, C. Ghajar, A. Nitadori-Hoshino, C. Hoffman, K. Badal, B. A. Garcia, M. K. Callahan, J. Yuan, V. R. Martins, J. Skog, R. N. Kaplan, M. S. Brady, J. D. Wolchok, P. B. Chapman, Y. Kang, J. Bromberg and D. Lyden (2012). "Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET." Nat Med **18**(6): 883-891.

Perez-Gonzalez, R., S. A. Gauthier, A. Kumar and E. Levy (2012). "The exosome secretory pathway transports amyloid precursor protein carboxyl-terminal fragments from the cell into the brain extracellular space." J Biol Chem **287**(51): 43108-43115.

Perez-Moreno, P., E. Brambilla, R. Thomas and J. C. Soria (2012). "Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities." Clin Cancer Res **18**(9): 2443-2451.

Perez-Sanchez, C., N. Barbarroja, L. C. Pantaleao, L. M. Lopez-Sanchez, S. E. Ozanne, B. Jurado-Gamez, E. Aranda, C. Lopez-Pedrera and A. Rodriguez-Ariza (2021). "Clinical Utility of microRNAs in Exhaled Breath Condensate as Biomarkers for Lung Cancer." J Pers Med **11**(2).

Perry, M. C., W. L. Eaton, K. J. Propert, J. H. Ware, B. Zimmer, A. P. Chahinian, A. Skarin, R. W. Carey, H. Kreisman, C. Faulkner and et al. (1987). "Chemotherapy with or without radiation therapy in limited small-cell carcinoma of the lung." N Engl J Med **316**(15): 912-918.

Piegari, E., R. Russo, D. Cappetta, G. Esposito, K. Urbanek, C. Dell'Aversana, L. Altucci, L. Berrino, F. Rossi and A. De Angelis (2016). "MicroRNA-34a regulates doxorubicin-induced cardiotoxicity in rat." Oncotarget **7**(38): 62312-62326.

Pignon, J. P., R. Arriagada, D. C. Ihde, D. H. Johnson, M. C. Perry, R. L. Souhami, O. Brodin, R. A. Joss, M. S. Kies, B. Lebeau and et al. (1992). "A meta-analysis of thoracic radiotherapy for small-cell lung cancer." N Engl J Med **327**(23): 1618-1624.

Planchard, D., S. Popat, K. Kerr, S. Novello, E. F. Smit, C. Faivre-Finn, T. S. Mok, M. Reck, P. E. Van Schil, M. D. Hellmann, S. Peters and E. G. Committee (2018). "Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Ann Oncol **29**(Suppl 4): iv192-iv237.

Pless, M., R. Stupp, H. B. Ris, R. A. Stahel, W. Weder, S. Thierstein, M. A. Gerard, A. Xyrafas, M. Fruh, R. Cathomas, A. Zippelius, A. Roth, M. Bijelovic, A. Ochsenbein, U. R. Meier, C. Mamot, D. Rauch, O. Gautschi, D. C. Betticher, R. O. Mirimanoff, S. Peters and S. L. C. P. Group (2015). "Induction chemoradiation in stage IIIA/N2 non-small-cell lung cancer: a phase 3 randomised trial." Lancet **386**(9998): 1049-1056.

Ponomaryova, A. A., E. Y. Rykova, N. V. Cherdyntseva, T. E. Skvortsova, A. Y. Dobrodeev, A. A. Zav'yalov, L. O. Bryzgalov, S. A. Tuzikov, V. V. Vlassov and P. P. Laktionov (2013). "Potentialities of aberrantly methylated circulating DNA for diagnostics and post-treatment follow-up of lung cancer patients." Lung Cancer **81**(3): 397-403.

Porcel, J. M., A. Esquerda, M. Vives and S. Bielsa (2014). "Etiology of Pleural Effusions: Analysis of More Than 3,000 Consecutive Thoracenteses." Archivos de Bronconeumología (English Edition) **50**(5): 161-165.

Postmus, P. E., K. M. Kerr, M. Oudkerk, S. Senan, D. A. Waller, J. Vansteenkiste, C. Escriu, S. Peters and E. G. Committee (2017). "Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Ann Oncol **28**(suppl_4): iv1-iv21.

Powell, H. A. and D. R. Baldwin (2014). "Multidisciplinary team management in thoracic oncology: more than just a concept?" Eur Respir J **43**(6): 1776-1786.

Powrozek, T., P. Krawczyk, D. M. Kowalski, B. Kuznar-Kaminska, K. Winiarczyk, M. Olszyna-Serementa, H. Batura-Gabryel and J. Milanowski (2016). "Application of plasma circulating microRNA-448, 506, 4316, and 4478 analysis for non-invasive diagnosis of lung cancer." Tumour Biol **37**(2): 2049-2055.

Powrozek, T., P. Krawczyk, D. M. Kowalski, K. Winiarczyk, M. Olszyna-Serementa and J. Milanowski (2015). "Plasma circulating microRNA-944 and microRNA-3662 as potential histologic type-specific early lung cancer biomarkers." Transl Res **166**(4): 315-323.

Price, A. (2003). "Lung cancer 5: state of the art radiotherapy for lung cancer." Thorax **58**(5): 447-452.

Pritchard, R. S. and S. P. Anthony (1996). "Chemotherapy plus radiotherapy compared with radiotherapy alone in the treatment of locally advanced, unresectable, non-small-cell lung cancer. A meta-analysis." Ann Intern Med **125**(9): 723-729.

Proctor, R. N. (2001). "Tobacco and the global lung cancer epidemic." Nature Reviews Cancer **1**(1): 82-86.

Purandare, N. C. and V. Rangarajan (2015). "Imaging of lung cancer: Implications on staging and management." Indian J Radiol Imaging **25**(2): 109-120.

Putila, J. J. and N. L. Guo (2011). "Association of arsenic exposure with lung cancer incidence rates in the United States." PLoS One **6**(10): e25886.

Qiagen (2019). exoRNeasy Midi/Maxi Handbook.

Quadrelli, S., G. Lyons, H. Colt, D. Chimondeguy and A. Buero (2015). "Clinical characteristics and prognosis of incidentally detected lung cancers." Int J Surg Oncol **2015**: 287604.

Raaschou-Nielsen, O., Z. J. Andersen, R. Beelen, E. Samoli, M. Stafoggia, G. Weinmayr, B. Hoffmann, P. Fischer, M. J. Nieuwenhuijsen, B. Brunekreef, W. W. Xun, K. Katsouyanni, K. Dimakopoulou, J. Sommar, B. Forsberg, L. Modig, A. Oudin, B. Oftedal, P. E. Schwarze, P. Nafstad, U. De Faire, N. L. Pedersen, C.-G. Östenson, L. Fratiglioni, J. Penell, M. Korek, G. Pershagen, K. T. Eriksen, M. Sørensen, A. Tjønneland, T. Ellermann, M. Eeftens, P. H. Peeters, K. Meliefste, M. Wang, B. Bueno-de-Mesquita, T. J. Key, K. de Hoogh, H. Concin, G. Nagel, A. Vilier, S. Grioni, V. Krogh, M.-Y. Tsai, F. Ricceri, C. Sacerdote, C. Galassi, E. Migliore, A. Ranzi, G. Cesaroni, C. Badaloni, F. Forastiere, I. Tamayo, P. Amiano, M. Dorronsoro, A. Trichopoulou, C. Bamia, P. Vineis and G. Hoek (2013). "Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)." The Lancet Oncology **14**(9): 813-822.

Rabinowits, G., C. Gercel-Taylor, J. M. Day, D. D. Taylor and G. H. Kloecker (2009). "Exosomal microRNA: a diagnostic marker for lung cancer." Clin Lung Cancer **10**(1): 42-46.

Rahmani, Z., M. Mojarrad and M. Moghbeli (2020). "Long non-coding RNAs as the critical factors during tumor progressions among Iranian population: an overview." Cell Biosci **10**: 6.

Rajpurkar, P., J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, B. N. Patel, K. W. Yeom, K. Shpanskaya, F. G. Blankenberg, J. Seekins, T. J. Amrhein, D. A. Mong, S. S. Halabi, E. J. Zucker, A. Y. Ng and M. P. Lungren (2018). "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists." PLoS Med **15**(11): e1002686.

Rami-Porta, R., D. Ball, J. Crowley, D. J. Giroux, J. Jett, W. D. Travis, M. Tsuboi, E. Vallieres, P. Goldstraw, C. International Staging, R. Cancer, Biostatistics, C. Observers to the and I. Participating (2007). "The IASLC Lung Cancer Staging Project: proposals for the revision of the T descriptors in the forthcoming (seventh) edition of the TNM classification for lung cancer." J Thorac Oncol **2**(7): 593-602.

Rasing, M. J. A., M. Peters, M. J. Aarts, G. J. M. Herder, A. S. R. van Lindert, F. Schramel, F. S. van der Meer, J. J. C. Verhoeff and P. S. N. van Rossum (2022). "Adjuvant Treatment Following Irradical Resection of Stage I-III Non-small Cell Lung Cancer: A Population-based Study." Curr Probl Cancer **46**(1): 100784.

Ravanidis, S., F. Grundler, F. Wilhelmi de Toledo, E. Dimitriou, F. Tekos, Z.-V. Skaperda, D. Kouretas and E. Doxakis (2021). "Fasting-mediated metabolic and toxicity reprogramming impacts circulating microRNA levels in humans." Food and Chemical Toxicology **152**: 112187.

Reitsma, M. B., L. S. Flor, E. C. Mullany, V. Gupta, S. I. Hay and E. Gakidou (2021). "Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990-2019." Lancet Public Health **6**(7): e472-e481.

Robinson, M. D. and A. Oshlack (2010). "A scaling normalization method for differential expression analysis of RNA-seq data." Genome Biol **11**(3): R25.

Rodriguez, M., D. Ajona, L. M. Seijo, J. Sanz, K. Valencia, J. Corral, M. Mesa-Guzman, R. Pio, A. Calvo, M. D. Lozano, J. J. Zulueta and L. M. Montuenga (2021). "Molecular biomarkers in early stage lung cancer." Transl Lung Cancer Res **10**(2): 1165-1185.

Rolfo, C. and A. Russo (2020). "Liquid biopsy for early stage lung cancer moves ever closer." Nat Rev Clin Oncol **17**(9): 523-524.

Ronai, Z. A., S. Gradia, L. A. Peterson and S. S. Hecht (1993). "G to a Transition and G to T Transversions in Codon-12 of the Ki-Ras Oncogene Isolated from Mouse Lung-Tumors Induced by 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanone (Nnk) and Related DNA Methylating and Pyridyloxobutylating Agents." Carcinogenesis **14**(11): 2419-2422.

Roscoe, R. J., K. Steenland, W. E. Halperin, J. J. Beaumont and R. J. Waxweiler (1989). "Lung Cancer Mortality Among Nonsmoking Uranium Miners Exposed to Radon Daughters." JAMA **262**(5): 629-633.

Saikumar, J., D. Hoffmann, T. M. Kim, V. R. Gonzalez, Q. Zhang, P. L. Goering, R. P. Brown, V. Bijol, P. J. Park, S. S. Waikar and V. S. Vaidya (2012). "Expression, circulation, and excretion profile of microRNA-21, -155, and -18a following acute kidney injury." Toxicol Sci **129**(2): 256-267.

Salido-Guadarrama, I., S. Romero-Cordoba, O. Peralta-Zaragoza, A. Hidalgo-Miranda and M. Rodriguez-Dorantes (2014). "MicroRNAs transported by exosomes in body fluids as mediators of intercellular communication in cancer." Onco Targets Ther **7**: 1327-1338.

Sandler, A., R. Gray, M. C. Perry, J. Brahmer, J. H. Schiller, A. Dowlati, R. Lilenbaum and D. H. Johnson (2006). "Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer." N Engl J Med **355**(24): 2542-2550.

Sanganalmath, P., J. E. Lester, A. G. Bradshaw, T. Das, C. Esler, A. E. F. Roy, E. Toy, J. F. Lester, M. Button, P. Wilson, C. Comins, P. Atherton, R. Pickles, K. Foweraker, G. A. Walker, M. Keni and M. Q. Hatton (2018). "Continuous Hyperfractionated Accelerated Radiotherapy (CHART) for Non-small Cell Lung Cancer (NSCLC): 7 Years' Experience From Nine UK Centres." Clin Oncol (R Coll Radiol) **30**(3): 144-150.

Sayyed, A. A., P. Gondaliya, P. Bhat, M. Mali, N. Arya, A. Khairnar and K. Kalia (2022). "Role of miRNAs in Cancer Diagnostics and Therapy: A Recent Update." Curr Pharm Des **28**(6): 471-487.

Scott, A. R., K. C. Stoltzfus, L. T. Tchelebi, D. M. Trifiletti, E. J. Lehrer, P. Rao, A. Bleyer and N. G. Zaorsky (2020). "Trends in Cancer Incidence in US Adolescents and Young Adults, 1973-2015." JAMA Netw Open **3**(12): e2027738.

seer.cancer.gov/data/. (last accessed: 2022). Retrieved 18 Feb 2022, from https://seer.cancer.gov/data/.

Semenova, E. A., R. Nagel and A. Berns (2015). "Origins, genetic landscape, and emerging therapies of small cell lung cancer." Genes Dev **29**(14): 1447-1462.

Senan, S. (2021). "The elusive pursuit of progress in limited stage small-cell lung cancer." Lancet Oncol **22**(3): 290-291.

Seung, S. J., M. Hurry, R. N. Walton and W. K. Evans (2020). "Real-world treatment patterns and survival in stage IV non-small-cell lung cancer in Canada." Curr Oncol **27**(4): e361-e367.

Shen, J., N. W. Todd, H. Zhang, L. Yu, X. Lingxiao, Y. Mei, M. Guarnera, J. Liao, A. Chou, C. L. Lu, Z. Jiang, H. Fang, R. L. Katz and F. Jiang (2011). "Plasma microRNAs as potential biomarkers for non-small-cell lung cancer." Lab Invest **91**(4): 579-587.

Sher, T., G. K. Dy and A. A. Adjei (2008). "Small cell lung cancer." Mayo Clin Proc **83**(3): 355-367.

Shi, T., G. Gao and Y. Cao (2016). "Long Noncoding RNAs as Novel Biomarkers Have a Promising Future in Cancer Diagnostics." Dis Markers **2016**: 9085195.

Shi, X., M. Sun, H. Liu, Y. Yao, R. Kong, F. Chen and Y. Song (2015). "A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer." Mol Carcinog **54 Suppl 1**: E1-E12.

Smith, N. C., G. Wajnberg, S. Chacko, N. T. Woldemariam, J. Lacroix, N. Crapoulet, D. C. Ayre, S. M. Lewis, M. L. Rise, R. Andreassen and S. L. Christian (2020). "Characterization of miRNAs in Extracellular Vesicles Released From Atlantic Salmon Monocyte-Like and Macrophage-Like Cells." Front Immunol **11**: 587931.

Socinski, M. A., T. Evans, S. Gettinger, T. A. Hensing, L. VanDam Sequist, B. Ireland and T. E. Stinchcombe (2013). "Treatment of stage IV non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines." Chest **143**(5 Suppl): e341S-e368S.

Solis, L. M., C. Behrens, M. G. Raso, H. Y. Lin, H. Kadara, P. Yuan, H. Galindo, X. Tang, J. J. Lee, N. Kalhor, Wistuba, II and C. A. Moran (2012). "Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome." Cancer **118**(11): 2889-2899.

Sonett, J. R., M. Suntharalingam, M. J. Edelman, A. B. Patel, Z. Gamliel, A. Doyle, P. Hausner and M. Krasna (2004). "Pulmonary resection after curative intent radiotherapy (>59 Gy) and concurrent chemotherapy in non-small-cell lung cancer." Ann Thorac Surg **78**(4): 1200-1205; discussion 1206.

Song, Z., S. Wang and Y. Liu (2019). "The diagnostic accuracy of liquid exosomes for lung cancer detection: a meta-analysis." Onco Targets Ther **12**: 181-192.

Sozzi, G. and M. Boeri (2014). "Potential biomarkers for lung cancer screening." Transl Lung Cancer Res **3**(3): 139-148.

Stahl, A. L., K. Johansson, M. Mossberg, R. Kahn and D. Karpman (2019). "Exosomes and microvesicles in normal physiology, pathophysiology, and renal diseases." Pediatr Nephrol **34**(1): 11-30.

Steliarova-Foucher, E., M. Colombet, L. A. G. Ries, F. Moreno, A. Dolya, F. Bray, P. Hesseling, H. Y. Shin, C. A. Stiller and I.-. contributors (2017). "International incidence of childhood cancer, 2001-10: a population-based registry study." Lancet Oncol **18**(6): 719-731.

Stoica, S. C. and W. S. Walker (2000). "Video assisted thoracoscopic surgery." Postgrad Med J **76**(899): 547-550.

Stone, E., N. Rankin, S. Kerr, K. Fong, D. C. Currow, J. Phillips, T. Connon, L. Zhang and T. Shaw (2018). "Does presentation at multidisciplinary team meetings improve lung cancer survival? Findings from a consecutive cohort study." Lung Cancer **124**: 199-204.

Su, J., J. Li, Q. Yu, J. Wang, X. Li, J. Yang, J. Xu, Y. Liu, Z. Xu, L. Ji, F. Yin and X. Chen (2020). "Exosomal miRNAs as potential biomarkers for acute myocardial infarction." IUBMB Life **72**(3): 384-400.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray (2021). "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." CA Cancer J Clin **71**(3): 209-249.

Tammela, T. and J. Sage (2020). "Investigating Tumor Heterogeneity in Mouse Models." Annu Rev Cancer Biol **4**(1): 99-119.

Tang, Q., Z. Ni, Z. Cheng, J. Xu, H. Yu and P. Yin (2015). "Three circulating long non-coding RNAs act as biomarkers for predicting NSCLC." Cell Physiol Biochem **37**(3): 1002-1009.

Tantai, J., D. Hu, Y. Yang and J. Geng (2015). "Combined identification of long non-coding RNA XIST and HIF1A-AS1 in serum as an effective screening for non-small cell lung cancer." Int J Clin Exp Pathol **8**(7): 7887-7895.

Thai, P., S. Statt, C. H. Chen, E. Liang, C. Campbell and R. Wu (2013). "Characterization of a novel long noncoding RNA, SCAL1, induced by cigarette smoke and elevated in lung cancer cell lines." Am J Respir Cell Mol Biol **49**(2): 204-211.

Thomas, M., C. Rube, P. Hoffknecht, H. N. Macha, L. Freitag, A. Linder, N. Willich, M. Hamm, G. W. Sybrecht, D. Ukena, K. M. Deppermann, C. Droge, D. Riesenbeck, A. Heinecke, C. Sauerland, K. Junker, W. E. Berdel, M. Semik and G. German Lung Cancer Cooperative (2008). "Effect of preoperative chemoradiation in addition to preoperative chemotherapy: a randomised trial in stage III non-small-cell lung cancer." Lancet Oncol **9**(7): 636-648.

tobaccoatlas.org/. (last accessed: 2022). Retrieved 19 Feb 2022, from https://tobaccoatlas.org/.

Tokuhata, G. K. and A. M. Lilienfeld (1963). "Familial aggregation of lung cancer in humans." J Natl Cancer Inst **30**: 289-312.

Travis, W. D., E. Brambilla and K. R. Geisinger (2016). "Histological grading in lung cancer: one system for all or separate systems for each histological type?" Eur Respir J **47**(3): 720-723.

Travis, W. D., E. Brambilla, A. G. Nicholson, Y. Yatabe, J. H. M. Austin, M. B. Beasley, L. R. Chirieac, S. Dacic, E. Duhig, D. B. Flieder, K. Geisinger, F. R. Hirsch, Y. Ishikawa, K. M. Kerr, M. Noguchi, G. Pelosi, C. A. Powell, M. S. Tsao, I. Wistuba and W. H. O. Panel (2015). "The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification." J Thorac Oncol **10**(9): 1243-1260.

Travis, W. D., E. Brambilla, M. Noguchi, A. G. Nicholson, K. R. Geisinger, Y. Yatabe, D. G. Beer, C. A. Powell, G. J. Riely, P. E. Van Schil, K. Garg, J. H. Austin, H. Asamura, V. W. Rusch, F. R. Hirsch, G. Scagliotti, T. Mitsudomi, R. M. Huber, Y. Ishikawa, J. Jett, M. Sanchez-Cespedes, J. P. Sculier, T. Takahashi, M. Tsuboi, J. Vansteenkiste, I. Wistuba, P. C. Yang, D. Aberle, C. Brambilla, D. Flieder, W. Franklin, A. Gazdar, M. Gould, P. Hasleton, D. Henderson, B. Johnson, D. Johnson, K. Kerr, K. Kuriyama, J. S. Lee, V. A. Miller, I. Petersen, V. Roggli, R. Rosell, N. Saijo, E. Thunnissen, M. Tsao and D. Yankelewitz (2011). "International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma." J Thorac Oncol **6**(2): 244-285.

Travis, W. D., E. Brambilla and G. J. Riely (2013). "New pathologic classification of lung cancer: relevance for clinical practice and clinical trials." J Clin Oncol **31**(8): 992-1001.

Valadi, H., K. Ekstrom, A. Bossios, M. Sjostrand, J. J. Lee and J. O. Lotvall (2007). "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells." Nat Cell Biol **9**(6): 654-659.

Vallero, D. (2014). Fundamentals of Air Pollution.

van Meerbeeck, J. P. (2008). "The controversial role of surgery in stage III NSCLC." Lancet Oncol **9**(7): 607-608.

Verstegen, N. E., F. J. Lagerwaard, S. M. Hashemi, M. Dahele, B. J. Slotman and S. Senan (2015). "Patterns of Disease Recurrence after SABR for Early Stage Non-Small-Cell Lung Cancer: Optimizing Follow-Up Schedules for Salvage Therapy." J Thorac Oncol **10**(8): 1195-1200.

Vickers, K. C., B. T. Palmisano, B. M. Shoucri, R. D. Shamburek and A. T. Remaley (2011). "MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins." Nat Cell Biol **13**(4): 423-433.

Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr. and K. W. Kinzler (2013). "Cancer genome landscapes." Science **339**(6127): 1546-1558.

Vu, L. T., J. Gong, T. T. Pham, Y. Kim and M. T. N. Le (2020). "microRNA exchange via extracellular vesicles in cancer." Cell Prolif **53**(11): e12877.

Wales, T. E. A. G. o. C. t. t. C. M. O. o. E. a. (1995). A POLICY FRAMEWORK FOR COMMISSIONING CANCER SERVICES.

Wallace, M. J., S. Krishnamurthy, L. D. Broemeling, S. Gupta, K. Ahrar, F. A. Morello, Jr. and M. E. Hicks (2002). "CT-guided percutaneous fine-needle aspiration biopsy of small (< or =1-cm) pulmonary lesions." Radiology **225**(3): 823-828.

Wang, H. M., J. H. Lu, W. Y. Chen and A. Q. Gu (2015). "Upregulated lncRNA-UCA1 contributes to progression of lung cancer and is closely related to clinical diagnosis as a predictive biomarker in plasma." Int J Clin Exp Med **8**(7): 11824-11830.

Wang, L., J. Li, Y. Li and L. B. Pang (2020). "Hsa-let-7c exerts an anti-tumor function by negatively regulating ANP32E in lung adenocarcinoma." Tissue Cell **65**: 101372.

Wang, M. C., G. Y. Gong, C. L. Wang, H. W. Ko, R. X. Weng, P. Y. Chang and C. C. Chiou (2022). "Methods for Collection of Extracellular Vesicles and Their Content RNA as Liquid Biopsy for Lung Cancer Detection: Application of Differential Centrifugation and Annexin A5 Coated Beads." Curr Issues Mol Biol **44**(5): 2374-2386.

Wang, X., C. Ling, Y. Bai and J. Zhao (2011). "MicroRNA-206 is associated with invasion and metastasis of lung cancer." Anat Rec (Hoboken) **294**(1): 88-92.

Weber, D. G., S. Casjens, G. Johnen, O. Bryk, I. Raiko, B. Pesch, J. Kollmeier, T. T. Bauer and T. Bruning (2014). "Combination of MiR-103a-3p and mesothelin improves the biomarker performance of malignant mesothelioma diagnosis." PLoS One **9**(12): e114483.

Weber, D. G., G. Johnen, S. Casjens, O. Bryk, B. Pesch, K. H. Jockel, J. Kollmeier and T. Bruning (2013). "Evaluation of long noncoding RNA MALAT1 as a candidate blood-based biomarker for the diagnosis of non-small cell lung cancer." BMC Res Notes **6**: 518.

webpath.med.utah.edu/NEOHTML/NEOPL069.html. (last accessed: 2022). Retrieved 28 Feb 2022, from https://webpath.med.utah.edu/NEOHTML/NEOPL069.html.

Whitehouse, M. (1995). "A policy framework for commissioning cancer services." BMJ **310**(6992): 1425-1426.

WHO (2013). AIR POLLUTION AND CANCER

WHO (2018). World health statistics 2018: monitoring health for the SDGs, sustainable development goals. Geneva, World Health Organization.

Wieckowski, E. U., C. Visus, M. Szajnik, M. J. Szczepanski, W. J. Storkus and T. L. Whiteside (2009). "Tumor-derived microvesicles promote regulatory T cell expansion and induce apoptosis in tumor-reactive activated CD8+ T lymphocytes." J Immunol **183**(6): 3720-3730.

Wilczynska, A. and M. Bushell (2015). "The complexity of miRNA-mediated repression." Cell Death Differ **22**(1): 22-33.

Williams, A. B. and B. Schumacher (2016). "p53 in the DNA-Damage-Repair Process." Cold Spring Harb Perspect Med **6**(5).

Wistuba, II and A. F. Gazdar (2006). "Lung cancer preneoplasia." Annu Rev Pathol **1**: 331-348.

Wozniak, M. B., G. Scelo, D. C. Muller, A. Mukeria, D. Zaridze and P. Brennan (2015). "Circulating MicroRNAs as Non-Invasive Biomarkers for Early Detection of Non-Small-Cell Lung Cancer." PLoS One **10**(5): e0125026.

Wu, S. F., N. Noren Hooten, D. W. Freeman, N. A. Mode, A. B. Zonderman and M. K. Evans (2020). "Extracellular vesicles in diabetes mellitus induce alterations in endothelial cell morphology and migration." J Transl Med **18**(1): 230.

Wu, Z., Q. Yuan, C. Yang, X. Zhang, P. Qi, H. Huang and Z. Ma (2020). "Downregulation of oncogenic gene TGFbetaR2 by miRNA-107 suppresses non-small cell lung cancer." Pathol Res Pract **216**(1): 152690.

www.asco.org/research-guidelines/cancer-progress-timeline/lung-cancer. (last accessed: 2022). Retrieved 4 Feb 2022, from https://www.asco.org/research-guidelines/cancer-progress-timeline/lung-cancer.

www.bccancer.bc.ca/books/lung/management/non-small-cell-lung-cancer-nsclc/combined-modality-therapy-for-unresectable-stage-iii. (2022). Retrieved 9 Feb 2022, from http://www.bccancer.bc.ca/books/lung/management/non-small-cell-lung-cancer-nsclc/combined-modality-therapy-for-unresectable-stage-iii.

www.cancer.gov/about-cancer/diagnosis-staging/prognosis/tumor-grade-fact-sheet#r1. (last accessed: 2022). Retrieved 17 Feb 2022, from https://www.cancer.gov/about-cancer/diagnosis-staging/prognosis/tumor-grade-fact-sheet#r1.

www.cancer.gov/about-cancer/diagnosis-staging/staging. (last accessed: 2022). Retrieved 1 May 2022, from https://www.cancer.gov/about-cancer/diagnosis-staging/staging.

www.cancer.gov/about-cancer/treatment/types. (last accessed: 2022). Retrieved 1 May 2022, from https://www.cancer.gov/about-cancer/treatment/types.

www.cancer.gov/about-cancer/treatment/types/targeted-therapies. (last accessed: 2022). Retrieved 1 May 2022, from https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies.

www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq. (last accessed: 2022). Retrieved 19 May 2022, from https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq.

www.cancer.org/cancer/lung-cancer/about/what-is.html. (2022). Retrieved 20 Feb 2022, from https://www.cancer.org/cancer/lung-cancer/about/what-is.html.

www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html. (last accessed: 2022). Retrieved 17 Jan 2022, from https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html.

www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html. (last accessed: 2022). Retrieved 6 Feb 2022, from https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html.

www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html. (last accessed: 2022). Retrieved 11 Feb 2022, from https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html.

www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/staging-sclc.html. (last accessed: 2022). Retrieved 25 Feb 2022, from https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/staging-sclc.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html. (last accessed: 2022). Retrieved 1 May 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html. (last accessed: 2022). Retrieved 27 Feb 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html. (last accessed: 2022). Retrieved 7 Jan 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/by-stage.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/immunotherapy.html. (last accessed: 2022). Retrieved 1 May 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/immunotherapy.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html. (last accessed: 2022). Retrieved 10 Jan 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html. (last accessed: 2022). Retrieved 1 May 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/radiation-therapy.html.

www.cancer.org/cancer/lung-cancer/treating-non-small-cell/surgery.html. (last accessed: 2022). Retrieved 10 May 2022, from https://www.cancer.org/cancer/lung-cancer/treating-non-small-cell/surgery.html.

www.cancer.org/cancer/lung-cancer/treating-small-cell/by-stage.html. (last accessed: 2022). Retrieved 9 Feb 2022, from https://www.cancer.org/cancer/lung-cancer/treating-small-cell/by-stage.html.

www.cancer.org/content/dam/CRC/PDF/Public/6151.00.pdf. (last accessed: 2022). Retrieved 7 May 2022, from https://www.cancer.org/content/dam/CRC/PDF/Public/6151.00.pdf.

www.cancerresearchuk.org/about-cancer/lung-cancer/risks-causes#:~:text=We%20know%20that%20air%20pollution, c. c. i. t. U. (last accessed: 2022). Retrieved 8 March 2022, from https://www.cancerresearchuk.org/about-cancer/lung-cancer/risks-causes#:~:text=We%20know%20that%20air%20pollution,cancer%20cases%20in%20the%20UK.

www.cancerresearchuk.org/about-cancer/lung-cancer/treatment/immunotherapy-targeted. (last accessed: 2022). Retrieved 2 May 2022, from https://www.cancerresearchuk.org/about-cancer/lung-cancer/treatment/immunotherapy-targeted.

www.cancerresearchuk.org/about-cancer/lung-cancer/treatment/surgery/types. (last accessed: 2022). Retrieved 16 Jan 2022, from https://www.cancerresearchuk.org/about-cancer/lung-cancer/treatment/surgery/types.

www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers/incidence. (last accessed: 2022). Retrieved 18 Jan 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers/incidence.

www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-One. (last accessed: 2022). Retrieved 19 Feb 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-One.

www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-One. (last accessed: 2022). Retrieved 4 May 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-One.

www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Three. (last accessed: 2022). Retrieved 6 Feb 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Three.

www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero. (last accessed: 2022). Retrieved 3 Feb 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero.

www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero. (last accessed: 2022). Retrieved 5 Jan 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero.

www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/survival%22. (last accessed: 2022). Retrieved 18 Feb 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/survival%22.

www.cancerresearchuk.org/health-professional/cancer-statistics/young-people-cancers/incidence. (last accessed: 2022). Retrieved 1 Jan 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/young-people-cancers/incidence.

www.cdc.gov/cancer/lung/basic_info/screening.htm. (last accessed: 2022). Retrieved 19 Feb 2022, from https://www.cdc.gov/cancer/lung/basic_info/screening.htm.

www.cdc.gov/cancer/lung/nonsmokers/index.htm. (last accessed: 2022). Retrieved 14 Feb 2022, from https://www.cdc.gov/cancer/lung/nonsmokers/index.htm.

www.cedars-sinai.org/programs/imaging-center/exams/ct-scans/lung.html. (last accessed: 2022). Retrieved 23 Feb 2022, from https://www.cedars-sinai.org/programs/imaging-center/exams/ct-scans/lung.html.

www.fda.gov/. (last accessed: 2022). Retrieved 28 Feb 2022, from https://www.fda.gov/.

www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-atezolizumab-extensive-stage-small-cell-lung-cancer. (last accessed: 2022). Retrieved 1 Feb 2022, from https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-atezolizumab-extensive-stage-small-cell-lung-cancer.

www.fda.gov/drugs/drug-approvals-and-databases/fda-grants-accelerated-approval-lurbinectedin-metastatic-small-cell-lung-cancer. (last accessed: 2022). Retrieved 21 May 2022, from https://www.fda.gov/drugs/drug-approvals-and-databases/fda-grants-accelerated-approval-lurbinectedin-metastatic-small-cell-lung-cancer.

www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-durvalumab-extensive-stage-small-cell-lung-cancer. (last accessed: 2022). Retrieved 3 May 2022, from https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-durvalumab-extensive-stage-small-cell-lung-cancer.

www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-osimertinib-adjuvant-therapy-non-small-cell-lung-cancer-egfr-mutations. (last accessed: 2022). from https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-osimertinib-adjuvant-therapy-non-small-cell-lung-cancer-egfr-mutations.

www.fda.gov/drugs/resources-information-approved-drugs/withdrawn-cancer-accelerated-approvals. (last accessed: 2022). Retrieved 11 Jan 2022, from https://www.fda.gov/drugs/resources-information-approved-drugs/withdrawn-cancer-accelerated-approvals.

www.htgmolecular.com/systems/workflow. (last accessed: 2022). Retrieved 10 Feb 2022, from https://www.htgmolecular.com/systems/workflow.

www.lungcancerresearchfoundation.org/about/. (last accessed: 2022). Retrieved 24 Feb 2022, from https://www.lungcancerresearchfoundation.org/about/.

www.lungcancerresearchfoundation.org/research/why-research/treatment-advances/. (last accessed: 2022). Retrieved 9 May 2022, from https://www.lungcancerresearchfoundation.org/research/why-research/treatment-advances/.

www.lungevity.org/for-patients-caregivers/lung-cancer-101/types-of-lung-cancer/small-cell-lung-cancer-sclc. (last accessed: 2022). Retrieved 20 Feb 2022, from https://www.lungevity.org/for-patients-caregivers/lung-cancer-101/types-of-lung-cancer/small-cell-lung-cancer-sclc.

www.nccn.org/. (last accessed: 2022). Retrieved 28 Feb 2022, from https://www.nccn.org/.

www.thebestoncologist.com/Cancer_Diseases/Lung_Cancer/Staging_of_Lung_Cancer.html. (last accessed: 2022). Retrieved 27 Feb 2022, from http://www.thebestoncologist.com/Cancer_Diseases/Lung_Cancer/Staging_of_Lung_Cancer.html.

www.wcrf-uk.org/preventing-cancer/uk-cancer-statistics/. (last accessed: 2022). Retrieved 25 Feb 2022, from https://www.wcrf-uk.org/preventing-cancer/uk-cancer-statistics/.

www.who.int/news-room/fact-sheets/detail/cancer. (last accessed: 2022). Retrieved 7 May 2022, from https://www.who.int/news-room/fact-sheets/detail/cancer.

www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. (last accessed: 2022). Retrieved 3 Jan 2022, from https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

Xie, M. Y., L. J. Hou, J. J. Sun, B. Zeng, Q. Y. Xi, J. Y. Luo, T. Chen and Y. L. Zhang (2019). "Porcine Milk Exosome MiRNAs Attenuate LPS-Induced Apoptosis through Inhibiting TLR4/NF-kappaB and p53 Pathways in Intestinal Epithelial Cells." J Agric Food Chem **67**(34): 9477-9491.

Xie, Y., Y. Zhang, L. Du, X. Jiang, S. Yan, W. Duan, J. Li, Y. Zhan, L. Wang, S. Zhang, S. Li, L. Wang, S. Xu and C. Wang (2018). "Circulating long noncoding RNA act as potential novel biomarkers for diagnosis and prognosis of non-small cell lung cancer." Mol Oncol **12**(5): 648-658.

Xue, W. X., M. Y. Zhang, L. Rui, X. Liu, Y. H. Yin and Y. Q. Qu (2020). "Serum miR-1228-3p and miR-181a-5p as Noninvasive Biomarkers for Non-Small Cell Lung Cancer Diagnosis and Prognosis." Biomed Res Int **2020**: 9601876.

Yan, C., N. Li, T. Xiao, X. Ye, L. Fu, Y. Ye, T. Xu and J. Yu (2022). "Extracellular vesicles from the inflammatory microenvironment regulate the osteogenic and odontogenic differentiation of periodontal ligament stem cells by miR-758-5p/LMBR1/BMP2/4 axis." J Transl Med **20**(1): 208.

Yanez-Mo, M., P. R. Siljander, Z. Andreu, A. B. Zavec, F. E. Borras, E. I. Buzas, K. Buzas, E. Casal, F. Cappello, J. Carvalho, E. Colas, A. Cordeiro-da Silva, S. Fais, J. M. Falcon-Perez, I. M. Ghobrial, B. Giebel, M. Gimona, M. Graner, I. Gursel, M. Gursel, N. H. Heegaard, A. Hendrix, P. Kierulf, K. Kokubun, M. Kosanovic, V. Kralj-Iglic, E. M. Kramer-Albers, S. Laitinen, C. Lasser, T. Lener, E. Ligeti, A. Line, G. Lipps, A. Llorente, J. Lotvall, M. Mancek-Keber, A. Marcilla, M. Mittelbrunn, I. Nazarenko, E. N. Nolte-'t Hoen, T. A. Nyman, L. O'Driscoll, M. Olivan, C. Oliveira, E. Pallinger, H. A. Del Portillo, J. Reventos, M. Rigau, E. Rohde, M. Sammar, F. Sanchez-Madrid, N. Santarem, K. Schallmoser, M. S. Ostenfeld, W. Stoorvogel, R. Stukelj, S. G. Van der Grein, M. H. Vasconcelos, M. H. Wauben and O. De Wever (2015). "Biological properties of extracellular vesicles and their physiological functions." J Extracell Vesicles **4**: 27066.

Yang, P., J. Ma, B. Zhang, H. Duan, Z. He, J. Zeng, X. Zeng, D. Li, Q. Wang, Y. Xiao, C. Liu, Q. Xiao, L. Chen, X. Zhu, X. Xing, Z. Li, S. Zhang, Z. Zhang, L. Ma, E. Wang, Z. Zhuang, Y. Zheng and W. Chen (2012). "CpG site-specific hypermethylation of p16INK4alpha in peripheral blood lymphocytes of PAH-exposed workers." Cancer Epidemiol Biomarkers Prev **21**(1): 182-190.

Yang, S., Z. Zhang and Q. Wang (2019). "Emerging therapies for small cell lung cancer." J Hematol Oncol **12**(1): 47.

Yao, X., M. M. Gomes, M. S. Tsao, C. J. Allen, W. Geddie and H. Sekhon (2012). "Fine-needle aspiration biopsy versus core-needle biopsy in diagnosing lung cancer: a systematic review." Curr Oncol **19**(1): e16-27.

Yin, Y., G. V. Shelke, C. Lasser, H. Brismar and J. Lotvall (2020). "Extracellular vesicles from mast cells induce mesenchymal transition in airway epithelial cells." Respir Res **21**(1): 101.

Yokoi, T. and M. Nakajima (2013). "microRNAs as mediators of drug toxicity." Annu Rev Pharmacol Toxicol **53**: 377-400.

Yoon, S. B. and J. H. Chang (2017). "Extracellular vesicles in bile: a game changer in the diagnosis of indeterminate biliary stenoses?" Hepatobiliary Surg Nutr **6**(6): 408-410.

Young, R. P. and R. J. Hopkins (2011). "How the genetics of lung cancer may overlap with COPD." Respirology **16**(7): 1047-1055.

Yu, H., Z. Guan, K. Cuk, H. Brenner and Y. Zhang (2018). "Circulating microRNA biomarkers for lung cancer detection in Western populations." Cancer Med **7**(10): 4849-4862.

Yu, Y., Z. Chen, J. Dong, P. Wei, R. Hu, C. Zhou, N. Sun, M. Luo, W. Yang, R. Yao, Y. Gao, J. Li, G. Yang, W. He and J. He (2013). "Folate receptor-positive circulating tumor cells as a novel diagnostic biomarker in non-small cell lung cancer." Transl Oncol **6**(6): 697-702.

Zaborowski, M. P., L. Balaj, X. O. Breakefield and C. P. Lai (2015). "Extracellular Vesicles: Composition, Biological Relevance, and Methods of Study." Bioscience **65**(8): 783-797.

Zamay, T. N., G. S. Zamay, O. S. Kolovskaya, R. A. Zukov, M. M. Petrova, A. Gargaun, M. V. Berezovski and A. S. Kichkailo (2017). "Current and Prospective Protein Biomarkers of Lung Cancer." Cancers (Basel) **9**(11).

Zaporozhchenko, I. A., E. S. Morozkin, T. E. Skvortsova, A. A. Ponomaryova, E. Y. Rykova, N. V. Cherdyntseva, E. S. Polovnikov, O. A. Pashkovskaya, E. A. Pokushalov, V. V. Vlassov and P. P. Laktionov (2016). "Plasma miR-19b and miR-183 as Potential Biomarkers of Lung Cancer." PLoS One **11**(10): e0165261.

Zappa, C. and S. A. Mousa (2016). "Non-small cell lung cancer: current treatment and future advances." Transl Lung Cancer Res **5**(3): 288-300.

Zatloukal, P., L. Petruzelka, M. Zemanova, L. Havel, F. Janku, L. Judas, A. Kubik, E. Krepela, P. Fiala and L. Pecen (2004). "Concurrent versus sequential chemoradiotherapy with cisplatin and vinorelbine in locally advanced non-small cell lung cancer: a randomized study." Lung Cancer **46**(1): 87-98.

Zhang, J. T., H. Qin, F. K. Man Cheung, J. Su, D. D. Zhang, S. Y. Liu, X. F. Li, J. Qin, J. T. Lin, B. Y. Jiang, D. Song, R. Q. Liao, N. Qiang, X. N. Yang, H. Y. Tu, Q. Zhou, J. J. Yang, X. C. Zhang, Y. N. Zhang, Y. L. Wu and W. Z. Zhong (2019). "Plasma extracellular vesicle microRNAs for pulmonary ground-glass nodules." J Extracell Vesicles **8**(1): 1663666.

Zhang, W., B. Yi, S. K. Yang, A. Li, J. Wang and H. Zhang (2020). "Extracellular vesicles carrying miRNAs in kidney diseases: a systemic review." Clin Exp Nephrol **24**(12): 1103-1121.

Zhang, X., H. Cai, M. Zhu, Y. Qian, S. Lin and X. Li (2020). "Circulating microRNAs as biomarkers for severe coronary artery disease." Medicine (Baltimore) **99**(17): e19971.

Zhang, Y., Y. Jia, R. Zheng, Y. Guo, Y. Wang, H. Guo, M. Fei and S. Sun (2010). "Plasma microRNA-122 as a biomarker for viral-, alcohol-, and chemical-related hepatic diseases." Clin Chem **56**(12): 1830-1838.

Zhou, E., Y. Li, F. Wu, M. Guo, J. Xu, S. Wang, Q. Tan, P. Ma, S. Song and Y. Jin (2021). "Circulating extracellular vesicles are effective biomarkers for predicting response to cancer therapy." EBioMedicine **67**: 103365.

Zhu, H., L. Zhang, S. Yan, W. Li, J. Cui, M. Zhu, N. Xia, Y. Yang, J. Yuan, X. Chen, J. Luo, R. Chen, R. Xing, Y. Lu and N. Wu (2017). "LncRNA16 is a potential biomarker for diagnosis of early-stage lung cancer that promotes cell proliferation by regulating the cell cycle." Oncotarget **8**(5): 7867-7877.