Genomic Markers in Differentiated Thyroid Cancer

Samuel Wing Sang Chan *BM BSc MRCS(ENT)*

The Institute of Cancer Research

University of London

MD(Res)

(Word count = 46,602)

Contents

Detailed Contents

Declaration

I declare that the work presented herein is my own, unless otherwise stated.
It was carried out under the supervision of Dr Kate Newbold, Professor Dae Kim and Professor Kevin Harrington.

- Samuel Chan *BM BSc MRCS(ENT)*

Abstract

This thesis explores the potential use of genomic markers to assist the management of differentiated thyroid cancer (DTC) patients throughout the course of their treatment pathway. We explore the use of these markers to prognosticate and risk stratify patients, whilst also investigating their role as sensitive and specific markers of disease burden.

The first part of the thesis investigates a genomic signature capable of predicting the presence of high risk histopathological features. These features commonly upstage DTCs from low to higher risk tumours and potentially influence the decision for clinicians to recommend completion thyroidectomy after initial hemi-thyroidectomy.
We hypothesised that predictive genomic biomarkers obtained from tumour samples provide additional prognostic stratification and assist in surgical decision making.

We conducted a retrospective, observational cohort study in adult patients with stage I-IV DTC. The index tumours from these patients underwent multi-mutational analysis using next generation sequencing as well as microRNA (miRNA) differential expression analysis using Nanostring and quantitative polymerase chain reaction (qPCR). We identified mutational patterns that were significantly associated with aggressive pathological features such as disease recurrence, vascular invasion and distant metastases. By combining the dysregulation statuses of various miRNA species with the somatic mutation status of various oncogenes, we proposed a predictive genomic signature that correctly identified DTC samples with high risk histopathological features with high sensitivity and specificity. These findings could be translated clinically into adjunctive molecular testing that can complement our existing prognostic investigations for better risk stratification.

The second part of the thesis addresses the novel use of bespoke patient specific sequencing panels, to increase the limit of detection of circulating tumour DNA (ctDNA) in DTC.

We hypothesised that existing sensitive PCR based techniques were limited by the inability to perform large target multiplexing and that the use of Whole Exome Sequencing (WES) to inform the production of patient tumour specific sequencing panels combined with new sequencing techniques, could improve detection rates of ctDNA from peri-treatment blood samples.

In this prospective pilot study, we demonstrated that WES was more effective at identifying putative mutational targets with which ctDNA can be identified. We showed that traditional targeted high depth sequencing with singleplex droplet digital PCR (ddPCR) assays detected ctDNA in only a minority of our patient cohort, whilst the rate of detection significantly increased when using patient specific bespoke sequencing panels informed by WES. The latter approach improves the clinical viability of sequencing techniques for the detection of minimal residual disease. We found that ctDNA detection and tracking analysis is most effective in AJCC stage 2 disease and above, although clinical validity and utility will need to be confirmed in larger prospective trials.

Acknowledgements

The body of scientific work enclosed in this thesis, was achievable only with the support and guidance of many people who I have had the honour of working and collaborating with over the past few years.

Thank you to my supervisors Dr Kate Newbold, Dr Isaac Garcia-Murillas, Professor Kevin Harrington and Professor Dae Kim for your wisdom, intellect and moral support throughout this process. You have opened many doors for me, and have shared both academic and life lessons that have provided strong foundations for my early career.

To the Molecular Oncology team headed by Professor Nick Turner – thank you for providing me with vital work space and storage capacity for my samples as well as teaching me the techniques used in my experiments. Your patience and understanding through my seemingly endless questions was very much appreciated. In particular I would like to thank Ben O'Leary, Ros Cutts, Sarah Hrebien, Giselle Walsh, Alex Pearson, Maria Coakley, Javier Pascual and Nida Pasha for all the conversations, pointers and favours that were critical in helping me develop my understanding of concepts and techniques that were initially completely foreign to a clinician.

To the staff at St George's University Hospital and Royal Marsden hospitals, who assisted in the running of the clinical trial and kindly accepted additional logistical burden to help me and this project, thank you very much.

To the charities, Emmy Coates Royal Marsden Cancer Charity, Arcobaleno Trust and Royal College of Surgeons England who funded this work through generous charitable sponsorship – I am so grateful for your trust and investment in scientific research in the hope of making significant contributions to the management of thyroid cancer in the future.

To my friends and family, especially Anjali and Leo, who have all had to contend with my absence in body and mind, in order to complete this project over the years, I love you all and thank you for your endless support and encouragement.

But most of all, and to whom I dedicate this thesis, I thank all the brave patients who in your time of uncertainty and personal anxieties, had the grace and foresight to participate in our study, despite not benefiting directly.

List of abbreviations

AF	Allele frequency
AJCC	American Joint Committee on Cancer
AMP	Anchored Multiplex PCR
ATA	American thyroid association
ATC	Anaplastic thyroid cancer
BEAMing	Bead-based digital PCR in emulsions
cDNA	Complimentary DNA
cfDNA	Cell-free DNA
CI	Confidence intervals
cPTC	Classical PTC
CTC	Circulating tumour cell
ctDNA	Circulating tumour DNA
ddPCR	Droplet digital polymerase chain reaction
DNA	Deoxyribonucleic acid
DTC	Differentiated thyroid cancer
EDTA	Ethylenediaminetetraacetic acid
ETE	Extra-thyroidal extension
FFPE	Formalin fixed paraffin embedded
FTC	Follicular thyroid cancer
fvPTC	Follicular variant PTC
indel	Insertion/Deletion
LN	Lymph node
MAPK	Mitogen activated protein kinase
MCA	Multiple correspondence analysis
MDT	Multidisciplinary team
MI-FTC	Minimally invasive FTC
miR	MicroRNA
miRNA	MicroRNA
MOI	MicroRNA of interest
MTC	Medullary thyroid cancer
NFR	Nuclear Fast Red
NIS	Sodium-iodide symporter
NPV	Negative predictive value
OR	Odds ratio
PCR	Polymerase chain reaction
PDTC	Poorly differentiated thyroid cancer
PI3K	Phosphoinositide 3-kinase
PPV	Positive predictive value
PTC	Papillary thyroid cancer
qPCR	Quantitative polymerase chain reaction
RAI	Radioactive iodine
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
RTK	Receptor tyrosine kinase
RT-PCR	Reverse transcription polymerase chain reaction
SNP	Single nucleotide polymorphism

SNV	Single nucleotide variant
TCGA	The Cancer Genome Atlas
Tg	Thyroglobulin
TNM	TNM classification of malignant tumours
TSH	Thyroid stimulating hormone
UMI	Unique molecular identifier
VAF	Variant allele frequency
VI	Vascular invasion
WES	Whole exome sequencing
WI-FTC	Widely invasive FTC

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Thyroid Cancer

1.1.1 The thyroid gland

The thyroid gland is a bi-lobed butterfly shaped endocrine gland found in the midline of the neck. It weighs approximately 25-30 grams in a healthy adult and has a rich blood supply. Its primary function is the secretion of thyroid hormones. Two of these hormones, tri-iodothyronine (T3) and thyroxine (T4), have a number of physiological effects throughout all systems in the body, including basal metabolic rate and protein synthesis adjustment.

1.1.2 Thyroid histology

Microscopically the thyroid is composed of follicles filled with colloid. This colloid contains thyroid hormone precursor proteins, mainly thyroglobulin (Tg), and is enclosed within a surrounding single layer of epithelial follicular cells. The follicular cells synthesise thyroid hormones called triiodothyronine (T3) and thyroxine (T4) in response to a positive feedback loop regulated by thyroid stimulating hormone (TSH) released by the pituitary gland. Thyroid hormone synthesis requires a significant uptake of iodine by follicular cells. Dietary circulating iodine enters the follicular cells actively via a sodium-iodide symporter (NIS) and this physiological mechanism underpins oncological treatment with radioactive iodine (RAI). Adjacent and in-between follicles are satellite parafollicular “C” cells, which secrete the hormone calcitonin (Figure 1-1). The precise role of calcitonin remains elusive more than 50 years after its discovery, but it is thought to be involved in calcium homeostasis.

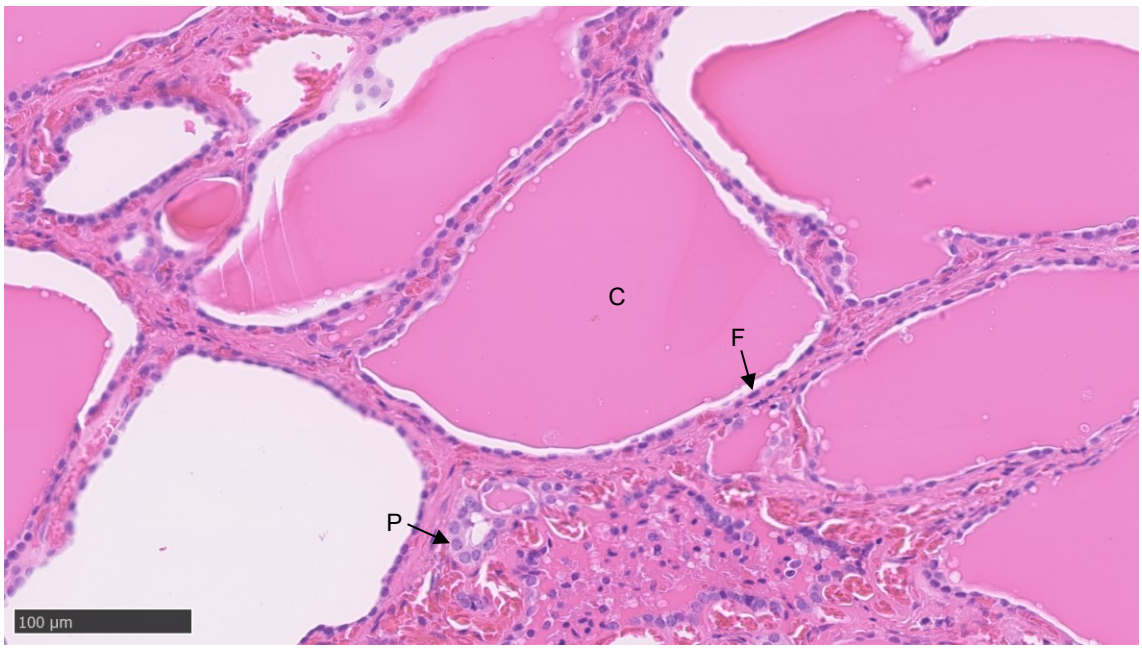


Figure 1-1. Light micrograph of normal thyroid tissue architecture.
C = Colloid, F = Follicular cell, P = Parafollicular "C" cell

1.1.3 Thyroid oncogenesis

The current paradigm to explain cancer formation is the progressive accumulation of critical genetic and epigenetic aberrations that drive a normal cell to a malignant derivative. These genomic alterations cause changes in gene expression, microRNA (miRNA) dysregulation and DNA structural abnormalities, resulting in malignant growth behaviours. Malignant features include uncontrolled clonal proliferation, breach of the basement membrane and spread to local or distant sites, neovascularisation as well as loss of apoptotic mechanisms and escape from innate immune cell control.

Somatic single nucleotide variants (SNV), small insertions and deletions (indels) and chromosomal structural changes, such as gene fusions, are all oncogenic drivers in thyroid cancer. These events occurring in somatic cells, lead to sporadic cases. Alternatively, genomic aberrations occurring in germline cells cause hereditary forms of thyroid cancer.

1.1.4 Thyroid cancer histological subtypes

Thyroid cancer is broadly divided into differentiated, poorly differentiated, undifferentiated or anaplastic, and medullary subtypes. Differentiated thyroid carcinomas (DTCs) account for approximately 85-90% of thyroid malignancies

and originate from follicular cells. DTCs are classified as papillary, follicular or Hürthle cell according to macro and microscopic disease architecture and behaviours of proliferation.

Papillary thyroid carcinoma (PTC) have characteristic microscopic features of enlarged nuclei, nuclear clearing and powdery chromatin giving a ground glass appearance. There is proliferation of malignant follicular cells into papillae shaped projections seen on microscopy (Figure 1-2). There are a further multitude of subtypes of PTC, which have different behaviours and varying degrees of pathological aggressiveness. Amongst these, tall cell variant and diffuse-sclerosing variant are the most aggressive subtypes with the highest propensity to recur or spread to distant sites (2, 3).

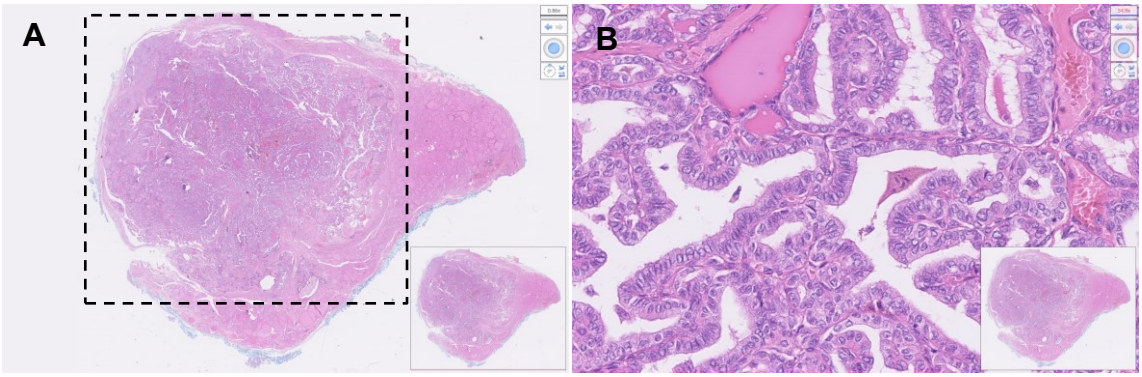


Figure 1-2. Light micrograph of a classical papillary thyroid cancer. A - H&E stain at 0.86x magnification, showing the margins between normal thyroid tissue and PTC tumour. B - H&E stain at 54x magnification showing characteristic “ground glass nuclei” or “Orphan-Annie eye nuclei”. Papillae architecture with multiple branches.

Follicular thyroid carcinoma (FTC) is defined as a malignant epithelial tumour showing follicular cell differentiation and lacking diagnostic nuclear features of PTC. This is in contrast to follicular adenoma (FA), which is a benign encapsulated tumour of the thyroid gland. The microscopic appearances of FTC are similar to FA and cannot be distinguished on cytological features alone. Carcinoma is distinguished through the identification of invasive phenotypes such as capsular invasion, vascular invasion, extrathyroidal tumour extension, lymph node metastasis or systemic metastases (Figure 1-3). FTC are more likely to metastasise to distant organs rather than regional lymph nodes due to a tendency to invade blood vessels causing haematogenous dissemination. World Health Organisation (WHO) classification divides FTC into minimally invasive (MI-FTC) and widely invasive (WI-FTC) depending on the extent of capsular or vascular invasion.

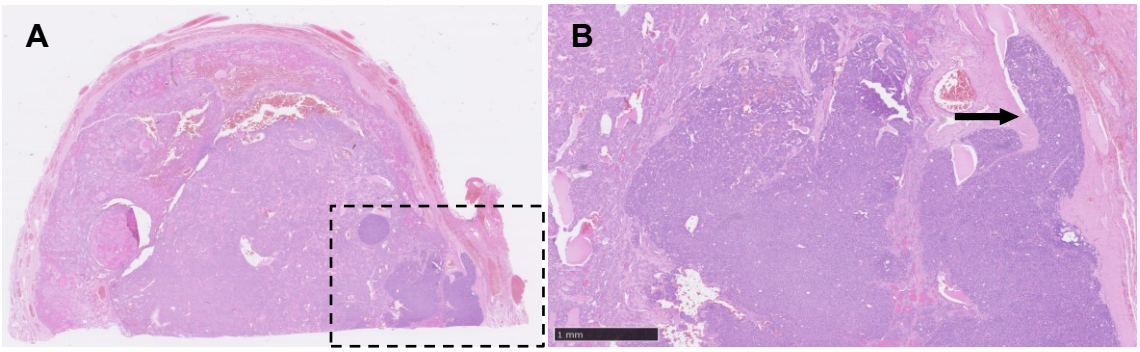


Figure 1-3. Light micrograph of a follicular thyroid cancer. A - H&E stain at 0.86x magnification, showing the margins between normal thyroid tissue and FTC tumour. B - H&E stain at 14.8x magnification showing evidence of capsular and vascular invasion (tumour breach of capsule and into a blood vessel) pointed by arrow.

Traditionally Hürthle cell thyroid carcinoma was considered a variant of FTC, as they are derived from follicular cells. Microscopically these cells are eosinophilic oxyphilic cells with round to oval nuclei and prominent nucleoli. They have small, densely packed mitochondria giving a granular appearance to the cytoplasm. Hürthle cell tumours represent around 5% of DTCs. They have a similar growth pattern to FTC, but the main clinical distinction is that only approximately 10% of Hürthle cell tumours take up radio-active iodine, hence response to treatment with RAI is much lower in these cases.

Poorly-differentiated (PDTC) and anaplastic thyroid carcinomas (ATC) are the result of further accumulation of genomic insults to existing DTC, causing de-differentiation and usually resulting in rapid proliferation, higher propensity to spread to distant sites and higher mortality (4, 5).

Medullary thyroid cancer (MTC) is a neuroendocrine tumour originating from the parafollicular “C” cells and account for 3-5% of thyroid cancers. As they are non-follicular in origin, they have different underlying pathological mechanisms and are not sensitive to RAI. MTC has poorer prognosis in terms of disease specific survival and disease recurrence compared with DTC (6-9).

1.1.5 Epidemiology of thyroid cancer

Thyroid cancer accounts for 2.1% of all newly diagnosed cancers worldwide, and is the most common endocrine cancer (10). It is more prevalent in women and has an age bimodal frequency distribution affecting the young and the elderly most commonly (11, 12). There has been an epidemic of thyroid cancer diagnoses in developed countries in the last two decades. Age standardised

three year average incidence rates have risen by at least 80% for both males and females from 2005 to 2015 in the United Kingdom (13). Similar trends have been reported in many developed countries across the world (14). There are a number of theories to explain this trend, but current consensus attributes the rise in incidence to the detection of subclinical thyroid nodules following investigation with ultrasound and cross-sectional imaging, leading to the diagnosis of small volume PTCs. This rapidly increasing incidence has resulted in a concomitant rise in the number of surgical thyroidectomies (15). However, despite this, mortality rates have remained relatively static over the same time period, suggesting possible over-diagnosis and treatment of potentially indolent disease. The increased incidence coupled with a 10-year survival rate of over 95% for stage 1-2 DTC, has resulted in a large survivor population (16). The treatment and monitoring of cancer survivors is a considerable financial burden to both health systems and individuals. Thyroid cancer costs $18-21 billion annually in the US healthcare system and ranks amongst the top five cancers to push patients into bankruptcy (17, 18). It is clear that better, more cost-effective methods to risk stratify and monitor thyroid cancer patients is needed.

1.2 Molecular pathology of thyroid cancer

Two cellular signalling pathways are affected by the most common genetic aberrations associated with thyroid malignancy. These are the Mitogen activated protein kinase (MAPK) and Phosphoinositide 3-kinase (PI3K) pathways (Figure 1-2). Both pathways allow the transmission of extracellular signals, via cell membrane bound receptor tyrosine kinases (RTKs), activating a cascade of protein interactions to ultimately influence DNA transcription within the nucleus.

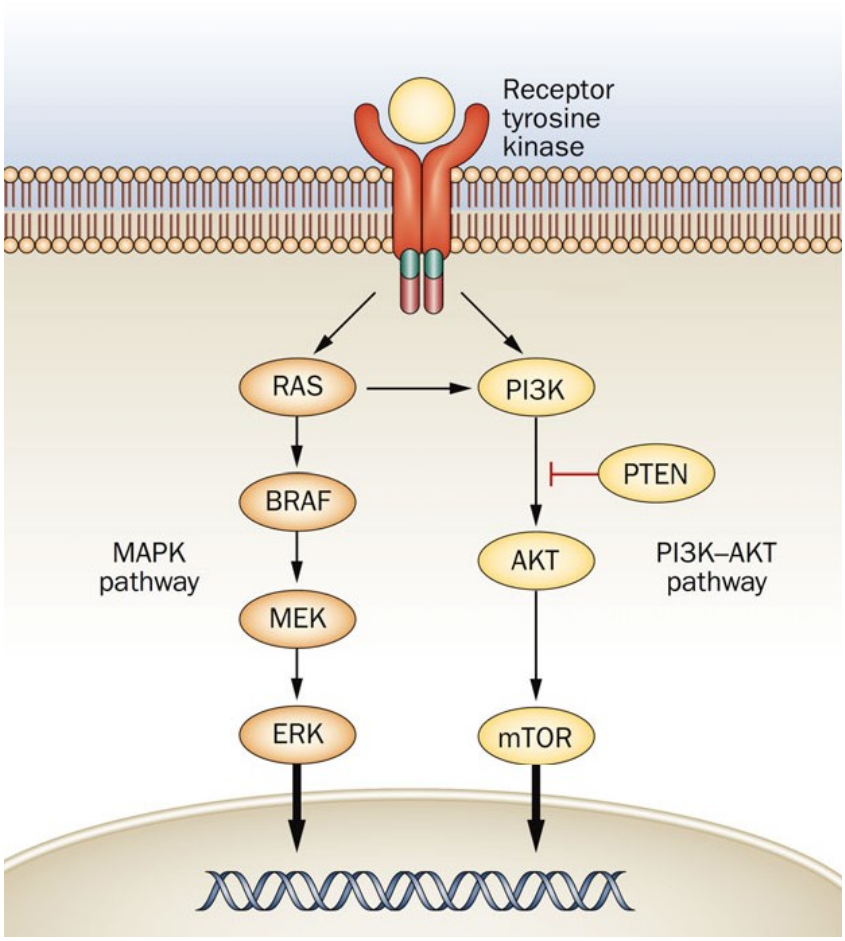


Figure 1-4. Simplified representation of the MAPK and PI3K pathways. Adapted from Campos et al 2014 (1)

1.2.1 Overview of MAPK & PI3K pathways

The MAPK pathway is a signal transduction pathway critical to the regulation of vital cellular functions involved in cell proliferation, differentiation, development and apoptosis. In normal physiology, the pathway activates when an extracellular ligand, such as epidermal growth factor, binds to RTKs in the cell membrane. This results in the activation of a number of downstream protein kinases, including RAS-RAF-MEK-ERK proteins, culminating in the production of transcription factors within the nucleus, thereby influencing protein transcription.

The PI3K pathway is another intracellular signal transduction pathway that promotes metabolism, proliferation, cell survival, growth and angiogenesis. It also similarly involves RTKs, but instead they activate Phosphoinositide 3-kinase (PI3K), which in turn phosphorylates Protein kinase B (AKT). AKT is a protein kinase with a number of downstream targets in the cytoplasm and

nucleus that influence cellular functions. Under normal physiologic conditions, the level of RTK activation is tightly controlled by a system of auto-inhibition and negative feedback loops. Alterations of genes encoding the proteins involved in these pathways can cause constitutive RTK activation, with the four principal alterations being gain-of-function mutations, genomic amplification, chromosomal rearrangements and/or autocrine activation (19). Dysregulation of RTK signalling leads to many human diseases, especially cancer.

1.2.2 Driver mutations and fusions in thyroid cancer

Multiple gene mutations have been implicated in the development of thyroid neoplasms, both benign and malignant. There are several integrated reviews that have outlined the genomic characterisation of DTC, implicating hotspot mutations and chromosomal rearrangements such as gene fusions, in the initiation and progression of the disease. Compared to other solid cancers, DTC has been shown to have one of the lowest tumour mutational burdens, usually harbouring only a single driver gene alteration (20). The Cancer Genome Atlas (TCGA) Research Network published a landmark paper in 2014 describing the genomic landscape of 496 PTC samples (21). This study excluded PDTC and ATC subtypes and the extracted DNA and total RNA samples were analysed with matched germline DNA from blood or normal thyroid. Analysis included single nucleotide polymorphism (SNP) arrays, whole exome sequencing (WES), RNA and microRNA sequencing and DNA methylation status. The top mutated genes in this study were *BRAF*, *RAS*, *EIF1AX*, *TSHR* and *PTEN* and 15.3% of samples harboured known or novel gene fusions with *RET* and *BRAF* fusions being the most frequent. A key finding from this large study is that PTC tumours can be subdivided by mutually exclusive mutations in either *BRAF* or *RAS* genes and this correlates strongly to histological subtype and therefore associated with differing oncogenic properties. *BRAF*-type PTC is associated with classical or tall-cell subtypes and are more likely to have recurrent disease or lymph node metastasis with a higher proportion that are refractory to RAI treatment. In contrast, *RAS*-type PTC is associated with follicular variant subtype, exhibiting more indolent behaviour. Functionally the *BRAFV600E* mutation results in a protein that is unresponsive to normal negative feedback from ERK, causing uncontrolled downstream activation of the MAPK pathway,

which explains the observed oncogenic effects compared with *RAS* mutated pathway activation which is still subject to ERK suppression.

Research on the genomic and transcriptomic characteristics of FTC is still quite sparse. Yoo et al employed RNA sequencing on 180 thyroid tumours, including 25 FA and 30 MI-FTCs. The most common somatic mutations found in FA and FTC were in *RAS*, *DICER1* and *EZH1*, however almost 30% of samples had unknown drivers. Additionally, the mutational profiles were unable to reliably distinguish between FA and FTC. Similarly, Jung et al analysed somatic mutations and copy number alterations using WES and a genomic hybridisation array in a cohort of FA and FTC tumours. They demonstrated similarities in mutational profiles between benign FA compared to FTC. However, malignant disease had higher variant allele fractions and significantly higher copy number alterations compared to benign disease (22). Follicular thyroid carcinoma therefore appears to be a more heterogeneous genetic entity when compared to PTC, with as yet unknown drivers that transform an initially benign FA to become malignant.

1.2.3 Tumour initiation mutations

RAS mutations

The rat sarcoma viral oncogenes homolog (*RAS*) genes encode GTPase proteins involved with cellular signal transduction. There are three *RAS* genes in humans: *HRAS*, *KRAS* and *NRAS*. RAS proteins are ubiquitously produced at varying levels in different tissue types and act as effector molecules in the MAPK and PI3K/AKT/mTOR signalling cascades. Activating mutations, such as point mutations in the GTP binding domain (codons 12 and 13) or the GTPase domain (codon 61) cause substitution of certain amino acid residues resulting in uncontrolled activation of downstream effectors. Approximately 99% of *RAS* mutations involve codons 12, 13 or 61 (23, 24). *RAS* mutations are found in both benign and malignant follicular thyroid disease. Approximately 20% of non-functioning follicular adenomas possess oncogene mutations such as in *NRAS* or *KRAS*, and have been implicated as playing a role in the evolution of follicular adenoma to follicular carcinoma (25). *RAS* mutations are found with highest frequencies in FTC (57%), follicular adenomas (30%), hyperplastic

nodules (5.6%), benign goitres (7-25%), and lesser frequency in follicular variant PTC (1.7-20%) (26-29).

BRAF mutations

Rapidly Accelerated Fibrosarcoma (RAF) kinases, of which there are three described (A-RAF, B-RAF and C-RAF), are activated by the RAS protein, which in turn activates the MAPK/ERK kinase (MEK) pathways by phosphorylation. Activating mutations allow continual MAPK pathway signal transduction independent of RAS activity. *BRAF* is an established oncogene and has been shown to be a key oncogenic driver in several malignancies including thyroid cancer and malignant melanoma. By far the most frequent mutation of the *BRAF* oncogene is a single base substitution at coding base 1799 where a thymine is replaced by an adenine (c.1799A>T). This causes a missense mutation, resulting in the transcription of glutamic acid instead of valine at amino acid position 600 (p.V600E) (30). *BRAF* is considered an initiator mutation, present in approximately 60% of PTCs but also found in papillary microcarcinomas (31). Large meta-analysis has demonstrated that *BRAFV600E* is significantly associated with increased tumour size, extra-thyroidal extension and clinical stage (32). Additionally, the long observed age-associated mortality risk in PTC appears to be dependent on *BRAF* variant status (33). *BRAF* mutated tumours exhibit reduced expression of genes involved with the production of thyroid hormone synthesis, including the sodium-iodide symporter, which explains their relative resistance to RAI therapy (34). Further, in the presence of concurrent *TERT* promoter mutation (described below) a negative synergistic effect is observed resulting in significantly higher rates of aggressive histopathological features such as extra-thyroidal extension, lymph node metastases, distant metastases and recurrent disease compared to *BRAF* wild type or *BRAFV600E* alone (35).

PTEN mutations

Phosphatase and tensin homolog (PTEN) is an inhibitor of the PI3K pathway through dephosphorylation. Inactivating variants prevent negative feedback in the signalling pathway, as is found in the germline mutations of *PTEN* that is intrinsic to Cowden’s syndrome - a disease that is characterised by

hamartomatous growths but also increased risk of benign and malignant thyroid disease (36, 37).

RET mutations and rearrangements

The Rearranged during Transfection (*RET*) gene is a proto-oncogene located on chromosome 10 and encodes a transmembrane tyrosine kinase receptor, which is normally expressed in neural crest derived cells, such as the parafollicular C cells. Mutations in the *RET* gene are rare in DTC but occur frequently in medullary thyroid cancer (MTC). Germline mutations in *RET* are found in familial MTC and Multiple Endocrine Neoplasia (MEN) type 2. *RET* is also involved in the creation of fusion oncoproteins in thyroid cancer, particularly in PTC. In these fusions the 3' RET receptor tyrosine kinase domain fuses with the 5' active promoter region of another gene, which leads to uncontrolled activation of the MAPK pathway (38, 39).

EIF1AX mutations

Eukaryotic translation initiation factor 1A X-linked (*EIF1AX*) gene encodes a critical component of the recruitment complex involved with translation of messenger RNAs at the 40S ribosomal subunit. Mutations of *EIF1AX* were first reported in uveal melanomas (40). They have since been reported in thyroid adenomas, FTCs and PTCs in a mutually exclusive manner with other drivers (21). Co-mutation with *RAS* mutations are more common in advanced thyroid cancers, and when present with *TERT* promoter or *TP53* mutations confer significantly worse survival in PDTCs and ATCs (41).

1.2.4 Tumour progression mutations

TERT promoter (TERTp) mutations

Telomeres protect chromosome ends from fusion and from being recognised as sites of DNA damage. Dysfunctional telomeres, which can occur by critical shortening in length in normal somatic cells during progressive cell division, elicit DNA damage responses that trigger cellular senescence (42). Telomerase maintains the telomere length at the end of chromosomes, which in turn maintains genomic integrity. Inappropriate activation of the telomerase reverse transcriptase (*TERT*) gene, responsible for coding the catalytic protein subunit of telomerase, is most commonly induced by mutations in the upstream

promoter region. This persistent activation promotes tumourigenesis by enabling replicative immortality in tumour cells (43, 44). There are a number of studies that have shown an association of *TERTp* mutations and thyroid cancers with aggressive phenotypes and reduced progression free survival (45-47). As mentioned previously, in PTC the negative effects on outcomes associated with *TERTp* variants are compounded in the presence of concurrent *BRAF* mutation (48-52). *TERTp* mutations are found in higher prevalence in PDTCs (~40%) and ATCs (~60-70%) suggesting an important role in tumour progression (5).

TP53 mutations

Tumour protein p53 (*TP53*) is a tumour suppressor gene if inactivated by point mutations, result in cancer formation. It is a late driver mutation, found with high prevalence in PDTCs and ATCs but less so in metastatic DTCs (53). In previous transgenic mice studies, those with mutant *BRAFV600E* and *TP53* developed PTCs that progressed into less differentiated PDTCs and ATCs (54, 55). This evidence implies *TP53* inactivation is a driver of de-differentiation in DTC.

1.3 Risk stratification and prognostication in thyroid cancer

Modern management of DTC requires individualised risk stratification and prognostication, as well as care plans that tailor the intensity of therapy and follow up. Clinical decision-making is influenced by estimating risk of mortality and recurrence, which in turn is based on various clinical, radiological and histological features of a patient and their individual cancer.

There are various staging systems designed to predict cancer-specific survival, including AGES, MACIS, AJCC and EORTC (56). The estimated 10-year disease specific survival rate in DTC is 98-100% for stage 1, 85-95% for stage 2, 60-70% for stage 3 and <50% for stage 4 disease (57). We have observed that the majority of new cases being diagnosed in the last decade are early, 1-4cm tumours without aggressive histopathological features, falling under stage 1-2 cancers.

The majority of patients therefore have a 10-year disease specific survival of over 85%. Historically, studies had shown that approximately 20% and 10% of DTC patients develop locoregional recurrences and distant metastases respectively within 10 years (58). Despite improvements in our surgical techniques; such as the advent of energy-based surgical tools, neuro-monitoring to facilitate closer dissection of tissue close to the recurrent laryngeal nerve and various endoscopic and robotic assisted approaches, as well as the addition of ablative RAI and TSH suppressive therapy, approximately 30% of DTC develop structurally identifiable persistent or recurrent disease within 10 years (16). Further, these recurrences can frequently be RAI-refractory, surgically challenging or unresectable with mortality reported as high as 38% to 69% (59, 60). Such patients can be treated with systemic therapy such as tyrosine kinase inhibitors including lenvatinib and sorafenib, as well as specific inhibitors for *BRAF*, *MEK* or *ALK* mutant tumours (61).

Many thyroid associations including the American, European and our own British Thyroid Association adopted a risk of recurrence classification (Figure 1-3). This recurrence classification scale uses a number of histological factors that confer varying increases in risk of recurrence. Clinical, radiological and cytological tumour features are readily available with minimal patient morbidity. However, definitive diagnosis and identification of high-risk histopathological features requires surgery, usually a hemi-thyroidectomy, to gain sufficient tissue and details of the cellular architecture. These histopathological features inform prognosis, and amongst them vascular invasion and extra-thyroidal extension are the most frequently identified features that upgrade a patient into higher risk categories. Identification of these features therefore has a direct impact on treatment extent.

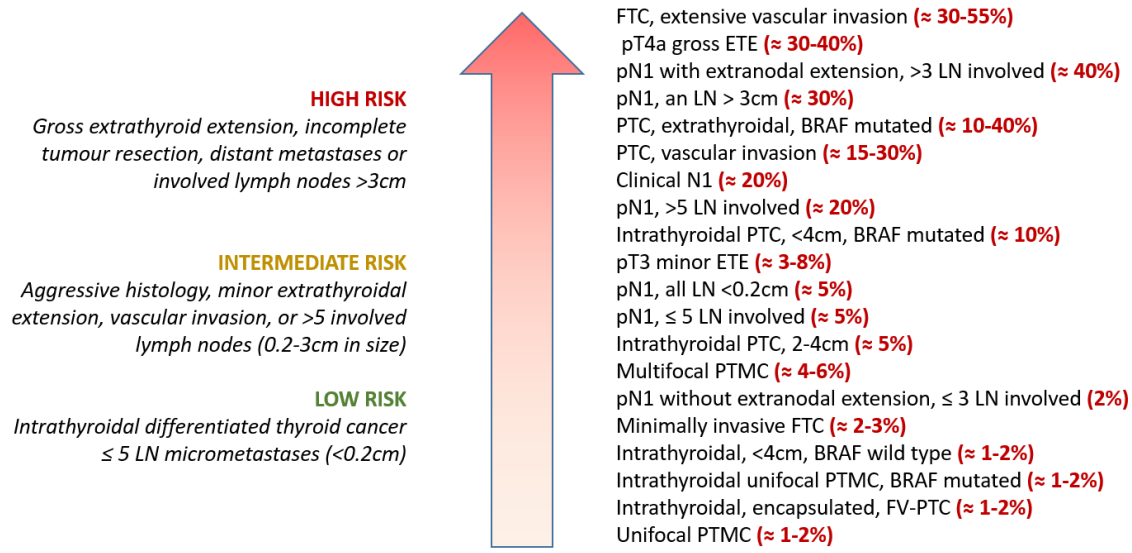


Figure 1-5. Scale for risk of recurrence, adapted from the American Thyroid Association 2015 guidelines

1.3.1 Vascular invasion as a high-risk histopathological feature in thyroid cancer

As demonstrated in other solid cancers, certain tumour phenotypes in thyroid cancer, have been associated with poorer outcomes. These are features such as large tumour size, aggressive cellular subtypes, extra-capsular extension, multi-focality, vascular invasion and peri-neural invasion (62). The presence of these features confers higher rates of local or regional nodal metastases as well distant spread.

Among these histopathological features, vascular invasion has been reported as an important and independent prognosticator in thyroid cancer. Systematic review and meta-analysis of vascular invasion in DTC showed higher risk for tumour persistence (OR = 2.75; 95% CI = 1.46–5.18), tumour recurrence (OR = 4.44; 95% CI = 2.94–6.71) and worse disease specific survival (HR = 2.47; 95% CI = 1.45–4.21) (63). Vascular invasion is defined as the presence of tumour cells in vascular spaces, underlying or invading through the endothelium of vascular channels or thrombus adherent to intravascular tumour. Immuno-histochemical markers anti-CD31 and D2-40 can be used to distinguish vascular invasion from lymphatic invasion or pseudo-invasion (64).

Vascular invasion typically occurs with other clinico-pathological features of aggressive behaviour and at a rate higher in FTC than PTC, which could

explain why FTC metastasises to distant organs more frequently (65). Whilst the independent prognostic ability of vascular invasion in PTC is controversial, in most major risk of recurrence classification systems for DTC the presence of vascular invasion confers intermediate to high risk, which would usually require total thyroidectomy and adjuvant RAI. Therefore, the pre-treatment identification of this histopathological feature via a direct or surrogate biochemical markers could guide extent of primary surgical treatment and potentially save two stage operations.

1.4 MicroRNAs in thyroid cancer

1.4.1 Structure and function of miRNAs

MicroRNAs (miRNAs) are defined as small non-coding RNAs of 19-25 nucleotides in length that control gene expression at the posttranscriptional level by hybridizing to target mRNAs at the 3' untranslated region (3'UTR) suppressing their translation or inducing their degradation (66).

MicroRNAs were discovered in the early 1990's during experiments involving the nematode model organism *Caenorhabditis elegans* (67, 68). They have since been detected in all animal model systems and some have been shown to be highly conserved across species (69). New miRNAs are still being discovered and are an exciting area of research. The majority of miRNAs are transcribed from DNA sequences into primary miRNAs and processed into precursor miRNAs, and finally into mature miRNAs. The interaction of miRNAs with their target genes is dynamic and dependent on many factors, such as the abundancy of miRNAs and target mRNAs, the degree of affinity of miRNA-mRNA interactions and the location within the cell where these interactions occur. If a miRNA molecule has the perfect complimentary base sequence to the 3' UTR region of a target mRNA, it induces the mRNA cleavage. In the absence of a perfect match, the miRNA will instead cause translational repression and inhibition of protein synthesis. This variability in action based on sequence affinity means that a single miRNA can target multiple genes.

MicroRNAs are studied using various techniques including arrays, Northern dot blot, RNA sequencing and RT-PCR. Experimental observations have found

miRNA species to be tissue and cell specific, even in adjacent cell types suggesting high evolutionary selection pressure favouring miRNA sequences that will yield biologically relevant miRNA-mRNA interactions (70). They are attractive as biomarkers in the study of human disease due to their relative stability and robustness in comparison to larger nucleic acid molecules; being more resistant to extreme temperatures, pH and the formalin-fixed-paraffin-embedding (FFPE) process (71, 72). More recently miRNAs have been shown to be relatively stable in a number of bodily fluids including saliva, cerebrospinal fluid, urine, lymph, tears and blood. The observed stability is attributed to complexing with proteins, lipoproteins and packaging into membrane-coated vesicles or exosomes (73).

1.4.2 miRNAs in differentiated thyroid cancer

Dysregulation of miRNAs is common in cancer cells and there is a large body of evidence implicating miRNA involvement in carcinogenesis. Altered expression of specific miRNAs related to either tumour suppressor or oncogenes have been shown to induce malignant effects on cell survival, proliferation and apoptosis (74, 75). Moreover, miRNA expression profiling has demonstrated specific miRNA signatures that can predict severity and clinical outcome in human cancers (76-80). The origin of this altered expression is attributed to genetic alterations to the miRNA-encoding genes, which are subject to the same changes that can affect protein coding genes.

A number of key miRNAs have been consistently shown to be dysregulated in DTC, some with high specificity and sensitivity to predict the presence of malignant disease versus benign or normal (81-83). Histological subtype specific miRNA dysregulation has been identified, which give an insight into the possible mechanisms of thyroid carcinogenesis.

1.4.3 Dysregulation of miRNA expression in PTC

Thirteen miRNAs (miR-31, miR-34b, miR-146, miR-155, miR-181a, miR-181b, miR-181c, miR-213, miR-220, miR-221, miR-222, miR-223 and miR-224) seem to be significantly and reproducibly upregulated in PTCs compared to normal thyroid tissue (84-87). In almost all studies using clinical samples, miR-146, miR-181b, miR-221 and miR-222 had large fold change increase in expression.

Interestingly, there are few studies describing many miRNAs that are significantly downregulated in PTC compared to normal. Most studies identify a miRNA signature that distinguishes PTC from normal thyroid tissue. There is evidence of miRNA dysregulation associated with various genetic aberrations common in PTC, for example Cahill et al used DNA microarrays and miRNA analysis to study normal thyroid cell lines compared to thyroid cell lines harbouring RET/PTC rearrangements and demonstrated strong correlation between upregulation of miR-128a, miR-128b, miR-139 and miR-200a and RET/PTC rearrangement (88). Upregulation of miR-221 and miR-222 has been demonstrated in mutated *BRAF* and *RAS* compared to wild type PTCs (89). Comparative upregulation of miR-31, miR-146b, miR-155, miR-221, miR-222 and downregulation of miR-1, miR-34b, miR-130b and miR-138 has been shown in aggressive versus non-aggressive PTC (87).

1.4.4 Dysregulation of miRNA expression in FTC

Finding high yield differential miRNA regulation that reliably discriminates between benign FA and malignant FTC is an ongoing but critically important venture. FTC cannot be distinguished from FA on the basis of cytology from fine needle aspirates, therefore hemi-thyroidectomy is required before definitive diagnosis is made. Nikiforova et al reported upregulation of miR-146b, miR-155, miR-187, miR-221, miR-222 and miR-224 in conventional FTC compared to normal thyroid tissue (89). This was corroborated by Rossing et al who went on to propose a 14 miRNA signature (miR-19a, miR-501-3p, miR-17, miR-335, miR-106b, miR-15a, miR-16, miR-374a, miR-542-5p, miR-503, miR-320a, miR-326, miR-330-5p and let7i) for the classification of FA and FTCs, with a reported negative predictive value of 92% and positive predictive value of 100% for FTC, although not prospectively validated (90). There is some evidence that downregulation of let-7a and miR-191 are early events in the development of FTC, and *in vitro* studies have shown that the restoration of miR-191 expression in FTC-derived cells inhibited cell growth by causing G1 phase arrest through the targeting of CDK6 (91).

1.5 Circulating DNA

1.5.1 Brief history of cell-free DNA

Pre-dating the publication of the double-helix structure of DNA by Watson and Crick, in 1948 two French scientists Mandel and Metais, published the surprising discovery of extracellular nucleic acids floating freely in plasma from patients with systemic lupus erythematosus (92). Several decades later a study by Leon et al described isolating serum derived “free DNA” using a radioimmunoassay capable of quantifying DNA to nanogram levels, and demonstrated higher levels in cancer patients compared to normal controls (93). Since those important discoveries, we now know that circulating DNA is highly fragmented genomic DNA that is thought to be released as a consequence of apoptosis, necrosis or active secretion (94, 95). This cell-free DNA (cfDNA) is detectable in the acellular compartments of blood, but has also been demonstrated in other bodily fluids including urine and saliva (96-99). It has a very short half-life of approximately 15 minutes to a few hours, thought to be rapidly degraded by nucleases or eliminated by the liver and kidneys (100). Cell-free DNA circulates in fragments ranging from 120-220bp with a maximum peak at 167bp. The peak size corresponds to the length of DNA wrapped around a single nucleosome (101). Haematopoietic cells are likely to be the source of the majority of circulating cfDNA, existing at a baseline level in healthy individuals (102). However fluctuations in cfDNA concentration has been shown in many other clinical scenarios including myocardial infarction, trauma, end-stage renal failure and even exercise (103-106). One of the most successful applications of cfDNA evaluation has been in the detection of foetal DNA in the circulation of expecting mothers, allowing for minimally invasive pre-natal screening for genetic abnormalities, such as point mutations and aneuploidies. The detection of cfDNA derived from tumour cells, known as circulating tumour DNA (ctDNA), is a rapidly expanding field within clinical oncology, with a number of important clinical applications. There are fundamental challenges with ctDNA detection, but slowly these are being circumvented through novel technologies and improved sequencing techniques.

1.5.2 Sources of circulating tumour DNA

Intact tumour derived epithelial cells have been shown to detach from primary or secondary tumours and disperse into the circulation as circulating tumour cells (CTCs) and the seeding and implantation of these cells is hypothesised to be the method of distant metastatic spread in some cancers (107). The detection, isolation and prognostic efficacy of CTCs has been demonstrated in a few cancer types, including colorectal, bladder and pancreatic cancers (108-110). Isolation of the extremely low concentration of CTCs from the background circulating normal cell population requires high volumes of blood and enrichment using various separation techniques. These techniques include using cell surface markers like immunomagnetic-assisted cell sorting or biophysical properties to separate cells based on size. These separated cells are then further analysed using cytometric or nucleic acid based techniques to verify their tumour origins. The isolation of CTCs often results in low yields of cells, which limits the clinical usefulness at the moment and therefore ctDNA continues to be favoured as a source of peripherally obtained tumour material.

Circulating DNA is found in both serum and plasma and can be isolated using simple centrifugation. Serum is the liquid portion of blood after it has clotted, whereas plasma is the liquid that remains when clotting is prevented with the addition of an anticoagulant such as Ethylenediaminetetraacetic acid (EDTA). Plasma has become the preferred sample fluid in most studies, due to the higher potential for contamination of genomic DNA resulting from lysis of a proportion of white cells that occurs in the clotting process.

1.5.3 Circulating tumour DNA

1.5.3.1 Tissue biopsy versus liquid biopsy

Tumour tissue is the gold standard for clinical and investigational genetic screening and is achieved via percutaneous or open biopsy. However, biopsies have a number of established disadvantages. They are invasive, have the potential for complications and can potentially delay treatment due to scheduling constraints or repeated attempts. Often metastatic deposits are found in anatomical locations that would present unacceptable risk or morbidity to patients if biopsy was attempted. The quantity and quality of tumour cells achieved is variable and often dependent on tumour cellularity and degree of

necrosis. From a processing perspective, most tumour tissue biopsies undergo fixing with formalin and embedding within paraffin, which crosslink DNA and can sometimes introduce excessive artefact and exclude samples from molecular analysis, particularly for whole genome and exome analyses.

A critical challenge, particularly in the context of genotyping a tumour, is tumour heterogeneity. This can be intra-tumoural heterogeneity as well as inter-metastatic heterogeneity. Advanced tumours are particularly characterised by these forms of heterogeneity, often consisting of varying clonal populations containing competing genotypes that confer survivability and resistance to treatment. Even with random sampling, cells from tissue biopsy may not provide a genetic picture representative of the whole of the tumour.

To overcome these limitations of tissue biopsies, less invasive techniques that capture tumour heterogeneity and the molecular changes inherent with tumour evolution in response to treatment is needed. Circulating tumour DNA in principle can address all of these. Sampling ctDNA through peripheral blood draws is minimally invasive, repeatable at minimal morbidity to the patient at varying time points throughout treatment and follow up, and provides dynamic monitoring of molecular changes from the tumour. The ctDNA fragments contain the genetic defects originating from all tumours; either primary, recurrent or metastatic and thus give a better representation of molecular heterogeneity as well as potentially highlighting the emergence of resistance conferring mutations.

1.5.3.2 Monitoring tumour burden

Tumour burden is typically measured both on clinical examination and radiological evaluation. Circulating biomarkers such as prostate specific antigen (PSA) in prostate cancer, carcinoembryonic antigen (CEA) in colorectal cancers and cancer antigen (CA) 125 in ovarian cancer are examples of protein biomarkers that are used to monitor response to treatment or disease progression. In thyroid cancer thyroglobulin (Tg) is used to monitor DTC subtypes, whereas calcitonin and carcinoembryonic antigen (CEA) are used in MTC. However, many of these protein markers lack specificity and may be elevated in clinical situations other than malignancy. Additionally some can

remain in the circulation for weeks, which reduces the accuracy and prolongs the assessment period. In the monitoring of thyroid cancer, thyroglobulin interpretation is limited by its non-specificity for cancer (as it is a surrogate marker for thyroid tissue only), its reduced accuracy in the presence of thyroglobulin autoantibodies and the difficulty in interpretation when residual normal thyroid tissue is present.

Circulating tumour DNA has by definition the same somatic mutations present in the original tumour but absent in matched normal DNA and therefore potentially has very high specificity, particularly when multiple tumour specific mutations are used as a signature. Several studies have shown that ctDNA levels can be a surrogate for tumour burden; decreasing following successful treatment and increasing with disease progression. This allows for more precise monitoring of tumour dynamics in advanced disease (111-113).

1.5.3.3 Minimal residual disease

Currently, predicting which patients are free of disease and those who have residual disease following definitive treatment, is largely based on clinical and pathological reporting. Margins are debated in multi-disciplinary cancer meetings across the country, and can sometimes be the difference between further treatment or clinical follow up. But accuracy of these predictions is hampered by various unavoidable factors including surgical resection artefact on the resected tumour margins from instrument handling or electrocautery devices, shrinking of tissue after resection and the fixing and embedding process. Circulating tumour DNA has been shown to be a potential marker of residual disease after resection. In a study by Diehl and colleagues, 18 subjects with stage II-IV colorectal cancer had blood sampling and ctDNA isolation and analysis before and after surgical treatment. For all cases that had complete surgical resection, ctDNA concentration rapidly declined. However in cases that had incomplete resection, ctDNA levels only reduced slightly or paradoxically increased. All patients with detectable postoperative levels of ctDNA experienced recurrent disease, whereas all patients with undetectable levels of ctDNA remained disease free (111). The detection of minimal residual disease using molecular markers like ctDNA have the potential to rationalise adjuvant treatment modalities, as well as inform post-treatment monitoring intensity.

Sensitivity of current techniques to detect minimal residual disease through ctDNA remains the most challenging aspect, particularly in early stage disease and will be discussed below in the limitations of ctDNA section.

1.5.3.4 Monitoring for treatment resistance

The acquisition of genetic aberrations that confer a survival benefit, and the partition of these cells through treatment selection pressures, are the hallmark of tumour treatment resistance. Molecular tools such as next generation sequencing, can be used to compare tissue genotypes before therapy and after recurrent disease to ascertain critical mutational differences. However these mutational differences or genotypic heterogeneity have also been demonstrated in temporal ctDNA collection and analysis. Patients with non-small cell lung cancer (NSCLC) who initially respond to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) therapy will eventually develop resistance. The most common cause of resistance is the development of the T790M EGFR variant, and this has been detected in ctDNA, with the development of RT-PCR based assays starting to enter routine clinical practice (114-116). The detection of these resistance conferring variants allows earlier termination of ineffective treatment with their respective side effects and modify treatment to available alternatives to further control disease. Similar resistance variants have not yet been definitively identified in recurrent or radio-iodine refractory DTC, but ctDNA could play a role in their early detection in the future.

1.5.4 Limitations of ctDNA

As in all clinical assays, sensitivity and specificity are critically important if meaningful clinical applications are intended. As mentioned above, ctDNA is highly fragmented, often at extremely low concentration and require identification from the background noise of wild-type DNA. When trying to detect allele fractions of under 0.01%, even highly sensitive techniques such as digital PCR can struggle with detection.

Further, the amount of ctDNA released by different tumours and disease stages have been shown to be variable depending on tissue type and extent of disease. Detection rates have been shown to be as high as 80-90% in advanced tumours known to have haematogenous spread, but can be less than

50% in primary brain, prostate and thyroid cancers depending on early or late disease states (117). The true amount of ctDNA released by tumour cells is unknown as studies are inherently limited by the number of variant targets being assayed and the sensitivity of the techniques being used to search for them. Recent innovation in unique molecular identifiers (UMIs) and ultra-high depth sequencing promise new limits of detection, which could potentially circumvent some of these historic limitations in ctDNA and make them more viable for routine clinical usage for the right patient cohort. Personalised sequencing panels that are generated from the patient's specific tumour genotype have shown huge potential in breast, bladder and colorectal cancers, particularly in early disease and minimal residual disease detection (118-122). This approach greatly increases the number of variant targets used to identify tumour specific ctDNA and in theory lowers the threshold of detection.

1.5.5 Circulating tumour DNA in thyroid cancer

There are limited peer-reviewed publications on ctDNA in DTC, with the vast majority of studies measuring *BRAFV600E* variants in prospectively collected blood samples of exclusively PTC tumours (123-128). This approach is understandable, given that PTC represents the majority of DTC diagnoses and that *BRAFV600E* is present in up to 87% of PTC tumours. The *BRAFV600E* mutation has been investigated in a number of cancers including melanoma and thyroid cancer, but has also been found to be detectable in lung cancers (129). A multi-variant approach is therefore clearly required to reduce false positives and improve sensitivity. Many ctDNA detection studies are limited to the use of a single mutational variant as techniques such as RT-PCR or digital PCR usually expends most of the extracted and purified cfDNA sample and multiplexing is limited. The only study that has demonstrated the assaying of multiple variant targets in cfDNA from plasma collected from patients with advanced thyroid cancers, is a study by Allin et al in 2018 (130). This study successfully demonstrated the detection and tracking of ctDNA in 73% and 60% of the PTC and FTC cohorts respectively. The study included patients with advanced thyroid cancer of all subtypes, including poorly differentiated and anaplastic thyroid carcinomas. They showed that ctDNA could provide detection of disease in the absence of traditional markers such as thyroglobulin, but also that detection of ctDNA preceded clinically detectable disease progression,

potentially shortening the lead time and providing an opportunity to change systemic therapy options. The rate of ctDNA detection in early (non-metastatic) DTC using a multi-variant approach has not yet been published. Additionally there are no published studies that have shown detection of ctDNA in DTC using direct sequencing of cfDNA with bespoke patient specific panels.

1.6 Theme of thesis

We know from published literature and our own clinical practice, the common debates and controversies within the multi-disciplinary team when discussing thyroid cancer patients. Central to these debates are prognostic markers to guide treatment extent, as well as sensitive biomarkers to detect minimal residual disease and guide follow up intensity. There is a need for prognostic markers that can assist clinicians and patients to decide on the optimum primary surgical extent; either total thyroidectomy up front or hemi-thyroidectomy followed by completion total thyroidectomy if high risk patient or histopathological features are present. Ideally these markers would be accessible through assessment of diagnostic sample material such as fine needle aspirate or through peripheral blood sampling, without the need for a surgical procedure. The first two chapters of this thesis focus on investigating potential somatic mutation and miRNA dysregulation signatures that can reliably predict the presence of vascular invasion and other high risk histopathological features in DTC. This is because they commonly influence the recommendation for two stage completion thyroidectomy, but are only discoverable after initial hemi-thyroidectomy and histological examination of the tumour. The final chapter examines the utility of ctDNA in early DTC as both a marker of tumour burden but also in the detection of minimal residual disease or early recurrence. Unique to this pilot study is the creation of bespoke patient specific sequencing panels based on somatic mutations observed through whole exome sequencing of primary tumours, and using these to detect ctDNA. This involves high depth sequencing of sequencing libraries created from the purified cfDNA extracted from the plasma time-points, which has to date never been published in thyroid cancer.

1.7 Hypotheses and aims of thesis

1.7.1 Hypotheses

1. There is a somatic mutational signature that significantly predicts the presence of vascular invasion or other high risk histopathological features in DTC.
2. There are dysregulated miRNAs that are significantly associated with the presence of vascular invasion or other high risk histopathological features in DTC.
3. Circulating tumour DNA is detectable in the plasma of patients with early DTC and that through the use of bespoke patient specific sequencing panels the sensitivity and therefore the rate of detection increases compared to single-plex digital PCR techniques.

1.7.2 Aims

A) To generate a biobank of FFPE DTC and extract total DNA and RNA.
B) To conduct targeted amplicon based sequencing of DNA extracted from DTC tumours and describe the somatic mutational landscape of these tumours.
C) To ascertain any mutational signature that significantly predicts the presence of vascular invasion or any other high risk histopathological feature.
D) To ascertain any differential expression signatures of miRNAs that significantly predicts the presence of vascular invasion or any other high risk histopathological feature.
E) To prospectively collect peri-operative blood samples from patients with DTC and assay for ctDNA based on variants detected in the tumour from their index operation.
F) To perform whole exome sequencing on a subset of tumours and create bespoke patient specific sequencing panels to sequence libraries made from cfDNA extracted from their corresponding time-points to detect ctDNA.

1.8 References

1. Campos M, Kool MM, Daminet S, Ducatelle R, Rutteman G, Kooistra HS, et al. Upregulation of the PI3K/Akt pathway in the tumorigenesis of canine thyroid carcinoma. J Vet Intern Med. 2014;28(6):1814-23.
2. Nath MC, Erickson LA. Aggressive Variants of Papillary Thyroid Carcinoma: Hobnail, Tall Cell, Columnar, and Solid. Adv Anat Pathol. 2018;25(3):172-9.
3. Liu Z, Zeng W, Chen T, Guo Y, Zhang C, Liu C, et al. A comparison of the clinicopathological features and prognoses of the classical and the tall cell variant of papillary thyroid cancer: a meta-analysis. Oncotarget. 2017;8(4):6222-32.
4. Giannini R, Moretti S, Ugolini C, Macerola E, Menicali E, Nucci N, et al. Immune Profiling of Thyroid Carcinomas Suggests the Existence of Two Major Phenotypes: An ATC-Like and a PDTC-Like. J Clin Endocrinol Metab. 2019;104(8):3557-75.
5. Landa I, Ibrahimpasic T, Boucai L, Sinha R, Knauf JA, Shah RH, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. J Clin Invest. 2016;126(3):1052-66.
6. Maia AL, Wajner SM, Vargas CV. Advances and controversies in the management of medullary thyroid carcinoma. Curr Opin Oncol. 2017;29(1):25-32.
7. Jin LX, Moley JF. Surgery for lymph node metastases of medullary thyroid carcinoma: A review. Cancer. 2016;122(3):358-66.
8. Ernani V, Kumar M, Chen AY, Owonikoko TK. Systemic treatment and management approaches for medullary thyroid cancer. Cancer Treat Rev. 2016;50:89-98.
9. Frank-Raue K, Machens A, Leidig-Bruckner G, Rondot S, Haag C, Schulze E, et al. Prevalence and clinical spectrum of nonsecretory medullary thyroid carcinoma in a series of 839 patients with sporadic medullary thyroid carcinoma. Thyroid. 2013;23(3):294-300.
10. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. Nat Rev Endocrinol. 2016;12(11):646-53.
11. Roman BR, Morris LG, Davies L. The thyroid cancer epidemic, 2017 perspective. Curr Opin Endocrinol Diabetes Obes. 2017;24(5):332-6.
12. Chen AY, Jemal A, Ward EM. Increasing incidence of differentiated thyroid cancer in the United States, 1988-2005. Cancer. 2009;115(16):3801-7.
13. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. Br J Cancer. 2016;115(9):1147-55.
14. Deng Y, Li H, Wang M, Li N, Tian T, Wu Y, et al. Global Burden of Thyroid Cancer From 1990 to 2017. JAMA Netw Open. 2020;3(6):e208759.
15. Li M, Dal Maso L, Vaccarella S. Global trends in thyroid cancer incidence and the impact of overdiagnosis. Lancet Diabetes Endocrinol. 2020;8(6):468-70.
16. Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. Thyroid. 2010;20(12):1341-9.

17. Furuya-Kanamori L, Sedrakyan A, Onitilo AA, Bagheri N, Glasziou P, Doi SAR. Differentiated thyroid cancer: millions spent with no tangible gain? Endocr Relat Cancer. 2018;25(1):51-7.
18. Lubitz CC, Kong CY, McMahon PM, Daniels GH, Chen Y, Economopoulos KP, et al. Annual financial impact of well-differentiated thyroid cancer care in the United States. Cancer. 2014;120(9):1345-52.
19. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010;141(7):1117-34.
20. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;349(6255):1483-9.
21. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014;159(3):676-90.
22. Jung SH, Kim MS, Jung CK, Park HC, Kim SY, Liu J, et al. Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma. Oncotarget. 2016;7(43):69638-48.
23. Castellano E, Santos E. Functional specificity of ras isoforms: so similar but so different. Genes Cancer. 2011;2(3):216-31.
24. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 2012;72(10):2457-67.
25. Challeton C, Bounacer A, Du Villard JA, Caillou B, De Vathaire F, Monier R, et al. Pattern of ras and gsp oncogene mutations in radiation-associated human thyroid tumors. Oncogene. 1995;11(3):601-3.
26. Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LD, et al. Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors. JAMA Oncol. 2016;2(8):1023-9.
27. Seo JY, Park JH, Pyo JY, Cha YJ, Jung CK, Song DE, et al. A Multi-institutional Study of Prevalence and Clinicopathologic Features of Non-invasive Follicular Thyroid Neoplasm with Papillary-like Nuclear Features (NIFTP) in Korea. J Pathol Transl Med. 2019;53(6):378-85.
28. Rivera M, Ricarte-Filho J, Knauf J, Shaha A, Tuttle M, Fagin JA, et al. Molecular genotyping of papillary thyroid carcinoma follicular variant according to its histological subtypes (encapsulated vs infiltrative) reveals distinct BRAF and RAS mutation patterns. Mod Pathol. 2010;23(9):1191-200.
29. Nikiforova MN, Lynch RA, Biddinger PW, Alexander EK, Dorn GW, 2nd, Tallini G, et al. RAS point mutations and PAX8-PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. J Clin Endocrinol Metab. 2003;88(5):2318-26.
30. Xu X, Quiros RM, Gattuso P, Ain KB, Prinz RA. High prevalence of BRAF gene mutation in papillary thyroid carcinomas and thyroid tumor cell lines. Cancer Res. 2003;63(15):4561-7.
31. Zheng X, Wei S, Han Y, Li Y, Yu Y, Yun X, et al. Papillary microcarcinoma of the thyroid: clinical characteristics and BRAF(V600E) mutational status of 977 cases. Ann Surg Oncol. 2013;20(7):2266-73.
32. Kim TH, Park YJ, Lim JA, Ahn HY, Lee EK, Lee YJ, et al. The association of the BRAF(V600E) mutation with prognostic factors and poor clinical outcome in papillary thyroid cancer: a meta-analysis. Cancer. 2012;118(7):1764-73.
33. Shen X, Zhu G, Liu R, Viola D, Elisei R, Puxeddu E, et al. Patient Age-Associated Mortality Risk Is Differentiated by BRAF V600E Status in Papillary Thyroid Cancer. J Clin Oncol. 2018;36(5):438-45.

34. Romei C, Ciampi R, Faviana P, Agate L, Molinaro E, Bottici V, et al. BRAFV600E mutation, but not RET/PTC rearrangements, is correlated with a lower expression of both thyroperoxidase and sodium iodide symporter genes in papillary thyroid cancer. Endocr Relat Cancer. 2008;15(2):511-20.
35. Xing M, Liu R, Liu X, Murugan AK, Zhu G, Zeiger MA, et al. BRAF V600E and TERT promoter mutations cooperatively identify the most aggressive papillary thyroid cancer with highest recurrence. J Clin Oncol. 2014;32(25):2718-26.
36. Halachmi N, Halachmi S, Evron E, Cairns P, Okami K, Saji M, et al. Somatic mutations of the PTEN tumor suppressor gene in sporadic follicular thyroid tumors. Genes Chromosomes Cancer. 1998;23(3):239-43.
37. Yeager N, Klein-Szanto A, Kimura S, Di Cristofano A. Pten loss in the mouse thyroid causes goiter and follicular adenomas: insights into thyroid function and Cowden disease pathogenesis. Cancer Res. 2007;67(3):959-66.
38. Jhiang SM, Sagartz JE, Tong Q, Parker-Thornburg J, Capen CC, Cho JY, et al. Targeted expression of the ret/PTC1 oncogene induces papillary thyroid carcinomas. Endocrinology. 1996;137(1):375-8.
39. Powell DJ, Jr., Russell J, Nibu K, Li G, Rhee E, Liao M, et al. The RET/PTC3 oncogene: metastatic solid-type papillary carcinomas in murine thyroids. Cancer Res. 1998;58(23):5523-8.
40. Martin M, Masshofer L, Temming P, Rahmann S, Metz C, Bornfeld N, et al. Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. Nat Genet. 2013;45(8):933-6.
41. Krishnamoorthy GP, Davidson NR, Leach SD, Zhao Z, Lowe SW, Lee G, et al. EIF1AX and RAS Mutations Cooperate to Drive Thyroid Tumorigenesis through ATF4 and c-MYC. Cancer Discov. 2019;9(2):264-81.
42. Shay JW. Are short telomeres predictive of advanced cancer? Cancer Discov. 2013;3(10):1096-8.
43. Cong YS, Wen J, Bacchetti S. The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter. Hum Mol Genet. 1999;8(1):137-42.
44. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013;339(6122):959-61.
45. Liu X, Bishop J, Shan Y, Pai S, Liu D, Murugan AK, et al. Highly prevalent TERT promoter mutations in aggressive thyroid cancers. Endocr Relat Cancer. 2013;20(4):603-10.
46. Penna GC, Pestana A, Cameselle JM, Momesso D, de Andrade FA, Vidal APA, et al. TERTp mutation is associated with a shorter progression free survival in patients with aggressive histology subtypes of follicular-cell derived thyroid carcinoma. Endocrine. 2018;61(3):489-98.
47. Qasem E, Murugan AK, Al-Hindi H, Xing M, Almohanna M, Alswailem M, et al. TERT promoter mutations in thyroid cancer: a report from a Middle Eastern population. Endocr Relat Cancer. 2015;22(6):901-8.
48. Song YS, Lim JA, Choi H, Won JK, Moon JH, Cho SW, et al. Prognostic effects of TERT promoter mutations are enhanced by coexistence with BRAF or RAS mutations and strengthen the risk prediction by the ATA or TNM staging system in differentiated thyroid cancer patients. Cancer. 2016;122(9):1370-9.
49. Liu J, Liu R, Shen X, Zhu G, Li B, Xing M. The Genetic Duet of BRAF V600E and TERT Promoter Mutations Robustly Predicts Loss of Radioiodine Avidity in Recurrent Papillary Thyroid Cancer. J Nucl Med. 2020;61(2):177-82.

50. Melo M, da Rocha AG, Vinagre J, Batista R, Peixoto J, Tavares C, et al. TERT promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. J Clin Endocrinol Metab. 2014;99(5):E754-65.
51. Song YS, Yoo SK, Kim HH, Jung G, Oh AR, Cha JY, et al. Interaction of BRAF-induced ETS factors with mutant TERT promoter in papillary thyroid cancer. Endocr Relat Cancer. 2019;26(6):629-41.
52. Moon S, Song YS, Kim YA, Lim JA, Cho SW, Moon JH, et al. Effects of Coexistent BRAF(V600E) and TERT Promoter Mutations on Poor Clinical Outcomes in Papillary Thyroid Cancer: A Meta-Analysis. Thyroid. 2017;27(5):651-60.
53. Suzuki K, Matsubara H. Recent advances in p53 research and cancer treatment. J Biomed Biotechnol. 2011;2011:978312.
54. McFadden DG, Vernon A, Santiago PM, Martinez-McFaline R, Bhutkar A, Crowley DM, et al. p53 constrains progression to anaplastic thyroid carcinoma in a Braf-mutant mouse model of papillary thyroid cancer. Proc Natl Acad Sci U S A. 2014;111(16):E1600-9.
55. Zou M, Baitei EY, Al-Rijjal RA, Parhar RS, Al-Mohanna FA, Kimura S, et al. TSH overcomes Braf(V600E)-induced senescence to promote tumor progression via downregulation of p53 expression in papillary thyroid cancer. Oncogene. 2016;35(15):1909-18.
56. Lang BH, Lo CY, Chan WF, Lam KY, Wan KY. Staging systems for papillary thyroid carcinoma: a review and comparison. Ann Surg. 2007;245(3):366-78.
57. Ghaznavi SA, Ganly I, Shaha AR, English C, Wills J, Tuttle RM. Using the American Thyroid Association Risk-Stratification System to Refine and Individualize the American Joint Committee on Cancer Eighth Edition Disease-Specific Survival Estimates in Differentiated Thyroid Cancer. Thyroid. 2018;28(10):1293-300.
58. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. Am J Med. 1994;97(5):418-28.
59. Tubiana M, Schlumberger M, Rougier P, Laplanche A, Benhamou E, Gardet P, et al. Long-term results and prognostic factors in patients with differentiated thyroid carcinoma. Cancer. 1985;55(4):794-804.
60. Schlumberger M, Tubiana M, De Vathaire F, Hill C, Gardet P, Travagli JP, et al. Long-term results of treatment of 283 patients with lung and bone metastases from differentiated thyroid carcinoma. J Clin Endocrinol Metab. 1986;63(4):960-7.
61. Lamartina L, Grani G, Durante C, Filetti S. Recent advances in managing differentiated thyroid cancer. F1000Res. 2018;7:86.
62. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid. 2016;26(1):1-133.
63. Vuong HG, Kondo T, Duong UNP, Pham TQ, Oishi N, Mochizuki K, et al. Prognostic impact of vascular invasion in differentiated thyroid carcinoma: a systematic review and meta-analysis. Eur J Endocrinol. 2017;177(2):207-16.
64. Mete O, Asa SL. Pathological definition and clinical significance of vascular invasion in thyroid carcinomas of follicular epithelial derivation. Mod Pathol. 2011;24(12):1545-52.

65. Wreesmann VB, Nixon IJ, Rivera M, Katabi N, Palmer F, Ganly I, et al. Prognostic value of vascular invasion in well-differentiated papillary thyroid carcinoma. Thyroid. 2015;25(5):503-8.
66. Ramirez-Moya J, Santisteban P. miRNA-Directed Regulation of the Main Signaling Pathways in Thyroid Cancer. Front Endocrinol (Lausanne). 2019;10:430.
67. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. 1993;75(5):855-62.
68. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901-6.
69. Praher D, Zimmermann B, Dnyansagar R, Miller DJ, Moya A, Modepalli V, et al. Conservation and turnover of miRNAs and their highly complementary targets in early branching animals. Proc Biol Sci. 2021;288(1945):20203169.
70. Pogue AI, Clement C, Hill JM, Lukiw WJ. Evolution of microRNA (miRNA) Structure and Function in Plants and Animals: Relevance to Aging and Disease. J Aging Sci. 2014;2(2).
71. Kakimoto Y, Tanaka M, Kamiguchi H, Ochiai E, Osawa M. MicroRNA Stability in FFPE Tissue Samples: Dependence on GC Content. PLoS One. 2016;11(9):e0163125.
72. Balzano F, Deiana M, Dei Giudici S, Oggiano A, Baralla A, Pasella S, et al. miRNA Stability in Frozen Plasma Samples. Molecules. 2015;20(10):19030-40.
73. Cui C, Cui Q. The relationship of human tissue microRNAs with those from body fluids. Sci Rep. 2020;10(1):5644.
74. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005;65(16):7065-70.
75. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834-8.
76. Lovat F, Valeri N, Croce CM. MicroRNAs in the pathogenesis of cancer. Semin Oncol. 2011;38(6):724-33.
77. Baghaei F, Abdollahi A, Mohammadpour H, Jahanbin M, Naseri Taheri F, Aminishakib P, et al. PTEN and miR-26b: Promising prognostic biomarkers in initiation and progression of Oral Squamous Cell Carcinoma. J Oral Pathol Med. 2019;48(1):31-5.
78. Han B, Zheng Y, Wang L, Wang H, Du J, Ye F, et al. A novel microRNA signature predicts vascular invasion in hepatocellular carcinoma. J Cell Physiol. 2019;234(11):20859-68.
79. Kim Y, Sim J, Kim H, Bang SS, Jee S, Park S, et al. MicroRNA-374a Expression as a Prognostic Biomarker in Lung Adenocarcinoma. J Pathol Transl Med. 2019;53(6):354-60.
80. Sulaiman SA, Abu N, Ab-Mutalib NS, Low TY, Jamal R. Signatures of gene expression, DNA methylation and microRNAs of hepatocellular carcinoma with vascular invasion. Future Oncol. 2019;15(22):2603-17.
81. Santiago K, Chen Wongworawat Y, Khan S. Differential MicroRNA-Signatures in Thyroid Cancer Subtypes. J Oncol. 2020;2020:2052396.
82. Hosseinkhan N, Honardoost M, Blighe K, Moore CBT, Khamseh ME. Comprehensive transcriptomic analysis of papillary thyroid cancer: potential biomarkers associated with tumor progression. J Endocrinol Invest. 2020.

83. Yoo SK, Song YS, Lee EK, Hwang J, Kim HH, Jung G, et al. Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. Nat Commun. 2019;10(1):2764.
84. He H, Jazdzewski K, Li W, Liyanarachchi S, Nagy R, Volinia S, et al. The role of microRNA genes in papillary thyroid carcinoma. Proc Natl Acad Sci U S A. 2005;102(52):19075-80.
85. Pallante P, Visone R, Ferracin M, Ferraro A, Berlingieri MT, Troncone G, et al. MicroRNA deregulation in human thyroid papillary carcinomas. Endocr Relat Cancer. 2006;13(2):497-508.
86. Tetzlaff MT, Liu A, Xu X, Master SR, Baldwin DA, Tobias JW, et al. Differential expression of miRNAs in papillary thyroid carcinoma compared to multinodular goiter using formalin fixed paraffin embedded tissues. Endocr Pathol. 2007;18(3):163-73.
87. Yip L, Kelly L, Shuai Y, Armstrong MJ, Nikiforov YE, Carty SE, et al. MicroRNA signature distinguishes the degree of aggressiveness of papillary thyroid carcinoma. Ann Surg Oncol. 2011;18(7):2035-41.
88. Cahill S, Smyth P, Finn SP, Denning K, Flavin R, O'Regan EM, et al. Effect of ret/PTC 1 rearrangement on transcription and post-transcriptional regulation in a papillary thyroid carcinoma model. Mol Cancer. 2006;5:70.
89. Nikiforova MN, Tseng GC, Steward D, Diorio D, Nikiforov YE. MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility. J Clin Endocrinol Metab. 2008;93(5):1600-8.
90. Rossing M, Borup R, Henao R, Winther O, Vikesaa J, Niazi O, et al. Down-regulation of microRNAs controlling tumourigenic factors in follicular thyroid carcinoma. J Mol Endocrinol. 2012;48(1):11-23.
91. Lena AM, Mancini M, Rivetti di Val Cervo P, Saintigny G, Mahe C, Melino G, et al. MicroRNA-191 triggers keratinocytes senescence by SATB1 and CDK6 downregulation. Biochem Biophys Res Commun. 2012;423(3):509-14.
92. Mandel P, Metais P. Nuclear Acids In Human Blood Plasma. C R Seances Soc Biol Fil. 1948;142(3-4):241-3.
93. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977;37(3):646-50.
94. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. Clin Chim Acta. 2001;313(1-2):139-42.
95. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659-65.
96. Reckamp KL, Melnikova VO, Karlovich C, Sequist LV, Camidge DR, Wakelee H, et al. A Highly Sensitive and Quantitative Test Platform for Detection of NSCLC EGFR Mutations in Urine and Plasma. J Thorac Oncol. 2016;11(10):1690-700.
97. Li F, Huang J, Ji D, Meng Q, Wang C, Chen S, et al. Utility of urinary circulating tumor DNA for EGFR mutation detection in different stages of non-small cell lung cancer patients. Clin Transl Oncol. 2017;19(10):1283-91.
98. van Ginkel JH, Slieker FJB, de Bree R, van Es RJJ, Van Cann EM, Willems SM. Cell-free nucleic acids in body fluids as biomarkers for the prediction and early detection of recurrent head and neck cancer: A systematic review of the literature. Oral Oncol. 2017;75:8-15.

99. Arantes L, De Carvalho AC, Melendez ME, Lopes Carvalho A. Serum, plasma and saliva biomarkers for head and neck cancer. Expert Rev Mol Diagn. 2018;18(1):85-112.
100. Khier S, Lohan L. Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature. Future Sci OA. 2018;4(4):FSO295.
101. Henikoff S, Church GM. Simultaneous Discovery of Cell-Free DNA and the Nucleosome Ladder. Genetics. 2018;209(1):27-9.
102. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A. 2015;112(40):E5503-12.
103. Beiter T, Fragasso A, Hudemann J, Niess AM, Simon P. Short-term treadmill running as a model for studying cell-free DNA kinetics in vivo. Clin Chem. 2011;57(4):633-6.
104. Antonatos D, Patsilinakos S, Spanodimos S, Korkonikitas P, Tsigas D. Cell-free DNA levels as a prognostic marker in acute myocardial infarction. Ann N Y Acad Sci. 2006;1075:278-81.
105. Chang CP, Chia RH, Wu TL, Tsao KC, Sun CF, Wu JT. Elevated cell-free serum DNA detected in patients with myocardial infarction. Clin Chim Acta. 2003;327(1-2):95-101.
106. Tovbin D, Novack V, Wiessman MP, Abd Elkadir A, Zlotnik M, Douvdevani A. Circulating cell-free DNA in hemodialysis patients predicts mortality. Nephrol Dial Transplant. 2012;27(10):3929-35.
107. Lin D, Shen L, Luo M, Zhang K, Li J, Yang Q, et al. Circulating tumor cells: biology and clinical significance. Signal Transduct Target Ther. 2021;6(1):404.
108. Veyrune L, Naumann DN, Christou N. Circulating Tumour Cells as Prognostic Biomarkers in Colorectal Cancer: A Systematic Review. Int J Mol Sci. 2021;22(8).
109. Khetrapal P, Lee MWL, Tan WS, Dong L, de Winter P, Feber A, et al. The role of circulating tumour cells and nucleic acids in blood for the detection of bladder cancer: A systematic review. Cancer Treat Rev. 2018;66:56-63.
110. Pang TCY, Po JW, Becker TM, Goldstein D, Pirola RC, Wilson JS, et al. Circulating tumour cells in pancreatic cancer: A systematic review and meta-analysis of clinicopathological implications. Pancreatology. 2021;21(1):103-14.
111. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14(9):985-90.
112. Shinozaki M, O'Day SJ, Kitago M, Amersi F, Kuo C, Kim J, et al. Utility of circulating B-RAF DNA mutation in serum for monitoring melanoma patients receiving biochemotherapy. Clin Cancer Res. 2007;13(7):2068-74.
113. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4(136):136ra68.
114. Balak MN, Gong Y, Riely GJ, Somwar R, Li AR, Zakowski MF, et al. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. Clin Cancer Res. 2006;12(21):6494-501.
115. Thompson JC, Yee SS, Troxel AB, Savitch SL, Fan R, Balli D, et al. Detection of Therapeutically Targetable Driver and Resistance Mutations in

Lung Cancer Patients by Next-Generation Sequencing of Cell-Free Circulating Tumor DNA. Clin Cancer Res. 2016;22(23):5772-82.
116. Taniguchi K, Uchida J, Nishino K, Kumagai T, Okuyama T, Okami J, et al. Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. Clin Cancer Res. 2011;17(24):7808-15.
117. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6(224):224ra24.
118. McDonald BR, Contente-Cuomo T, Sammut SJ, Odenheimer-Bergman A, Ernst B, Perdigones N, et al. Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer. Sci Transl Med. 2019;11(504).
119. Coombes RC, Page K, Salari R, Hastings RK, Armstrong A, Ahmed S, et al. Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. Clin Cancer Res. 2019;25(14):4255-63.
120. Christensen E, Birkenkamp-Demtroder K, Sethi H, Shchegrova S, Salari R, Nordentoft I, et al. Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. J Clin Oncol. 2019;37(18):1547-57.
121. Christensen E, Nordentoft I, Vang S, Birkenkamp-Demtroder K, Jensen JB, Agerbaek M, et al. Optimized targeted sequencing of cell-free plasma DNA from bladder cancer patients. Sci Rep. 2018;8(1):1917.
122. Reinert T, Henriksen TV, Christensen E, Sharma S, Salari R, Sethi H, et al. Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. JAMA Oncol. 2019.
123. Pupilli C, Pinzani P, Salvianti F, Fibbi B, Rossi M, Petrone L, et al. Circulating BRAFV600E in the diagnosis and follow-up of differentiated papillary thyroid carcinoma. J Clin Endocrinol Metab. 2013;98(8):3359-65.
124. Chuang TC, Chuang AY, Poeta L, Koch WM, Califano JA, Tufano RP. Detectable BRAF mutation in serum DNA samples from patients with papillary thyroid carcinomas. Head Neck. 2010;32(2):229-34.
125. Cradic KW, Milosevic D, Rosenberg AM, Erickson LA, McIver B, Grebe SK. Mutant BRAF(T1799A) can be detected in the blood of papillary thyroid carcinoma patients and correlates with disease status. J Clin Endocrinol Metab. 2009;94(12):5001-9.
126. Kim BH, Kim IJ, Lee BJ, Lee JC, Kim IS, Kim SJ, et al. Detection of plasma BRAF(V600E) mutation is associated with lung metastasis in papillary thyroid carcinomas. Yonsei Med J. 2015;56(3):634-40.
127. Kwak JY, Jeong JJ, Kang SW, Park S, Choi JR, Park SJ, et al. Study of peripheral BRAF(V600E) mutation as a possible novel marker for papillary thyroid carcinomas. Head Neck. 2013;35(11):1630-3.
128. Salvianti F, Giuliani C, Petrone L, Mancini I, Vezzosi V, Pupilli C, et al. Integrity and Quantity of Total Cell-Free DNA in the Diagnosis of Thyroid Cancer: Correlation with Cytological Classification. Int J Mol Sci. 2017;18(7).
129. Yang Y, Shen X, Li R, Shen J, Zhang H, Yu L, et al. The detection and significance of EGFR and BRAF in cell-free DNA of peripheral blood in NSCLC. Oncotarget. 2017;8(30):49773-82.
130. Allin DM, Shaikh R, Carter P, Thway K, Sharabiani MTA, Gonzales-de-Castro D, et al. Circulating tumour DNA is a potential biomarker for disease progression and response to targeted therapy in advanced thyroid cancer. Eur J Cancer. 2018;103:165-75.

Chapter 2 Materials and Methods

2.1 Gene mutation signature for vascular invasion in thyroid cancer

2.1.1 Trial design and patient recruitment

"Gene Mutation Analysis as a biomarker in Thyroid Cancer (GEM-TC)" was a two part study (CCR 5150, JRES 2019.006, REC reference 19/LO/0342). The study was divided into two recruitment cohorts; a part 1 retrospective study (n = 69) and a part 2 prospective non-interventional study (n = 58). The sponsor was St George's University Hospitals NHS Foundation Trust (SGUH), with the Royal Marsden Hospital (RMH) as a recruitment sub-site. Part 1 patients were identified by searching histopathological databases and electronic patient records and part 2 patients were recruited prospectively following discussion at the regional Head & Neck multidisciplinary team meeting. All patients met the following criteria:

Inclusion criteria

- Adult (>18 years old at time of giving informed consent)
- Confirmed either/or:
  a) histopathological diagnosis of differentiated thyroid cancer
  b) cytological diagnosis of Thy4 or Thy5
- Treated at SGUH or RMH between 2008-2020 with archival tissue samples available

Exclusion criteria

- Age less than 18
- Diagnosis of medullary or anaplastic thyroid cancer
- Any concurrent thyroid disease other than malignancy
- Incomplete data or loss to follow up
- Significant medical co-morbidities or concurrent malignancy
- Micro-carcinoma (nodules <1cm in size)

Patients provided written informed consent to allow the research team access to surgical tissue surplus to diagnostic use. Part 2 patients were consented to four peri-treatment venepuncture time-points. Entry into the trial was opportunistic

and available at any point in their treatment pathway. The full trial protocol, patient information leaflet and consent form are shown in Appendix A.

2.1.2 Tumour DNA and RNA co-extraction

Archival formalin-fixed, paraffin embedded (FFPE) tissue blocks were used for extraction. Eight slides of 10μm thickness were cut from each tissue block using a microtome. The first slide was stained with haematoxylin and eosin (H&E) and mounted with a glass coverslip, and was examined by a consultant pathologist (GP) to confirm size and tumour content as well as histopathological diagnosis. The remaining sections were kept wax embedded and stored in independent slide boxes in cool, dry conditions until further processing.

Prior to DNA/RNA co-extraction, slides underwent paraffin de-waxing and Nuclear Fast Red (NFR) staining. Depending on the surface area of the tumour, between 1-3 slides underwent manual macro-dissection of the tumour tissue using a single-use sterile 25 gauge needle under light microscopy.
DNA and RNA was extracted using a spin column-based nucleic acid purification kit (Allprep DNA/RNA FFPE kit, Qiagen, Hilden, Germany). Quantification of extracted nucleic acids was performed using a dsDNA HS (high sensitivity) assay kit on a Qubit Fluorometer (ThermoFisher, Massachusetts, USA) for DNA and a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (ThermoFisher, Massachusetts, USA) for RNA. Extracted DNA and RNA samples were stored in barcode labelled tubes at -20°C and -80°C respectively until further processing.

2.1.3 Library preparation

Ion Torrent compatible library preparation was performed using the Ion AmpliSeq™ Library Kit 2.0 (Life Technologies, ThermoFisher, Massachusetts, USA) following the manufacturer's protocol. DNA input for each sample was 10ng, initially prepared in two separate primer pools with one-step enzymatic fragmentation, end repair, poly-A tailing and amplification using the Ion Ampliseq Hi-Fi Mix. Samples were processed in a thermocycler under the following conditions:

PCR program			
Stage	**Step**	**Temp**	**Time**
Hold	Activate enzyme	99°C	2 minutes
22x Cycles	Denature	99°C	15 seconds
	Anneal and extend	60°C	4 minutes
Hold	-	10°C	Up to 1 hour

The primer pools used were from a custom amplicon based target enrichment panel, designed using the ThermoFisher Scientific Ion Ampliseq Designer tool. This panel (ThyMa) was designed in-house and previously validated for use in DNA extracted from FFPE thyroid tumours as well as in cfDNA extracted from plasma. The panel targeted genes and exonic mutational hotspots known to be commonly associated with thyroid cancer, including one promoter region for *TERT* (see table 2-1).

ThyMa Panel Gene Hotspot List			
Gene	Exons	*Gene*	Exons
BRAF	15, 16	*AXIN1*	2-5, 7-11
RET	2, 10, 11, 15, 16	*GNAS*	8, 9
TSHR	8, 10, 11	*APC*	6-16
TERT	Promoter	*KRAS*	2-5
NRAS	2-4	*PTEN*	4-8
TP53	2-10	*AKT1*	2, 4
KMT2C	7-18	*EIF1AX*	1-7
HRAS	2, 3	*CHEK2*	4, 12, 16
PIK3CA	2, 5, 10, 21	*ALK*	7, 8, 12, 16, 19, 23, 25
CTNNB1	3, 11, 15	*IDH1*	6-4

Table 2-1 Summary of genes and exons covered by the ThyMa thyroid specific panel

The final panel had a total coverage of 20.47kb and an estimated 92.36% coverage of the submitted loci.

Post PCR primer digestion was performed by adding 1µl FuPa reagent to each sample and incubating at 50°C for 10 minutes, 55°C for 10 minutes and 60°C for 20 minutes in a thermocycler. Indexed adapters were ligated by adding 1µl Ion Ampliseq Adapters, 2µl Switch solution and 1µl DNA ligase to each sample and incubating at 22°C for 30 minutes and 72°C for 10 minutes in a thermocycler. Resultant libraries were purified using 1.5x volume AMPure XP beads (Beckman Coulter, Brea, CA, USA) and washed with 70% ethanol.

2.1.4 Library quantification

Purified libraries were quantified using the Ion Torrent Library Quantitation Kit (Life Technologies, ThermoFisher, Massachusetts, USA). Each library was diluted by 1:400 and was tested in duplicate. Reaction mixture was 10µl of 2x Ion Library TaqMan qPCR mix, 1µl of 20x Ion Library TaqMan Quantitation assay and 9µl of library dilution (total reaction volume 20µl). Standards of 6.8pM, 0.68pM and 0.068pM of *E.coli* DH10B Control Library were used to form a standard curve alongside no template control (NTC) wells and all run in triplicate. Samples and standards were processed in a QuantStudio 6 Flex from Thermofisher under the following conditions:

50°C	2 minutes	Hold
95°C	20 seconds	Hold
95°C	1 second	40 x cycles
60°C	20 seconds	

Libraries were normalised to 100pM concentration and pooled for sequencing.

2.1.5 Sequencing and bioinformatics

The pooled samples were loaded onto a Proton P1 chip and sequenced on an Ion Torrent platform (ThermoFisher Scientific, Waltham, MA, USA). Bioinformatics analysis for this study was run through the VariTAS pipeline (https://CRAN.R-project.org/package=varitas) v0.0.2 in R 3.6.0 (R Core Team, 2018). Reads were aligned to the GRCh37 assembly of the human genome using BWA mem v0.7.12 (1). The resulting BAM files were further processed with Picard v2.8.1 according to the Genome Analysis Toolkit (GATK) Best Practices with the exception of removing duplicate reads (2). Following these steps, mean and median coverage were calculated across all samples with BEDtools v2.25.0 to be 7,958x and 7,212x, respectively (3). Variant calling was performed with MuTect2 in GATK v4.0.5.1 and LoFreq v2.1.3.1 without matched normal samples (4, 5). Common variants found in gnomAD/dbSNP were provided to these two callers for filtering (6, 7). Variants were annotated using ANNOVAR (8). All variant calls were manually curated using the Integrated Genomics Viewer (IGV) software to ensure integrity of calls (9, 10).

2.1.6 Statistics

Statistical calculations and modelling were performed using a combination of Prism 9 software (Graphpad, San Diego, CA, USA) and various packages on R (R Core Team, 2014).

The Chi Squared and Fisher's Exact tests were used to quantify perceived associations between variables based on differing proportions.

Multiple Correspondence Analysis (MCA) was used to explore relationships between independent variables recorded for each tumour. It is a multivariate analysis method that is built upon Principal Component Analysis. The process was facilitated by the R package FactoMineR (11). The full R script can be found in Appendix B.

Multiple logistic regression was performed in the Prism 9 software to test multiple independent variables in their ability to predict the presence of vascular invasion. The full dataset with binary outputs (present or absent) was input into Prism. The analysis produced positive and negative predictive power percentages which were plotted on Receiver Operating Characteristic (ROC) curves.

2.2 Differential expression of miRNA associated with vascular invasion in thyroid cancer

2.2.1 Literature review

Using PubMed, Cochrane Library and Ovid MEDLINE a literature search was performed focusing on reviews of miRNA in association with thyroid disease of any kind, between 1997 to 2020. The following search terms were used: “thyroid” AND “microRNA” OR “miRNA”. Suitable papers were reviewed to canvas for known differentially expressed miRNAs in thyroid cancer as well as evidence for which reference genes were used in differential expression studies involving thyroid tissue (malignant or benign). A panel with specific miRNAs of interest (MOIs) was formed based on the findings from the literature search and this panel was used in the subsequent qPCR pilot experiment.

2.2.2 qPCR pilot experiment

Total RNA samples extracted from the FFPE tumours (mentioned in Chapter 2.1.2) were divided by histological subtype (PTC and FTC) and further divided into cohorts of samples with and without the presence of vascular invasion.

Four samples from each cohort (4 PTC with VI, 4 PTC without VI, 4 FTC with VI, 4 FTC without VI) as well as 8 total RNA samples extracted from normal thyroid tissue (as normal control) were processed for assessment of relative expression of a selected panel of MOIs based on literature review. The panel is shown in Table 4-2 in chapter 4. The miRCURY LNA qPCR kit (Qiagen) was used with individual assays for each miRNA and endogenous controls and UniSp6 as a technical control for the reverse transcription step.

Each total RNA sample was diluted to a concentration of 5ng/µl using nuclease-free water, and 2µl of this diluted template was added to 2µl of 5x miRCURY SYBR Green RT Reaction Buffer, 4.5µl RNase-free water, 1µl of 10x miRCURY RT Enzyme Mix and 0.5µl UniSp6 RNA spike-in to have a total reaction volume of 10µl. The reactions were incubated in a thermocycler for 60 minutes at 42°C and 5 minutes at 95°C and the completed cDNA samples stored at 4°C until

ready for qPCR. No enzyme control (NEC) reactions were performed in parallel for all samples.

Complimentary DNA samples were diluted to 1:60 using RNase-free water and were run on 96 well plates using the “all genes” layout (ie. All conditions run on the same plate, 1 plate per sample). For each condition and reference gene, qPCR reaction volumes were composed of 5µl of 2x miRCURY SYBR Green Mastermix, 0.05µl ROX reference dye, 0.95µl RNase-free water, 1µl of miRCURY LNA miRNA PCR Assay and 3µl of cDNA dilution. All samples were run in triplicate with NTC negative controls. Each plate was run on a QuantStudio 6 Flex from Thermofisher under the following conditions:

Step	**Temperature**	**Time**	**Comments**
PCR initial heat activation	95°C	2 minutes	Hold
Denaturation	95°C	10 seconds	2-step cycling for 40 cycles
Combined annealing/extension	56°C	60 seconds	

Raw Ct values were obtained and individual amplification curves reviewed with outlier replicates removed manually.

2.2.3 Data Analysis and Statistics

Average Ct values were calculated and Normfinder analysis (12) was performed to assess the most stable expressed reference gene. The delta-Ct method was used to calculate relative fold change with the geometric mean of the two most stably expressed gene per sample. Statistical calculations and modelling were performed using a combination of Prism 9 software (Graphpad, San Diego, CA, USA) and various packages on R (R Core Team, 2014).

2.2.4 Nanostring pilot experiment

To mirror the qPCR pilot experiment, the same sample cohorts were chosen from the samples outlined in Chapter 2.1. These samples were assessed for the expressional profile of 830 miRNA molecules using the nCounter Human v3

miRNA Expression Assay (Nanostring Technologies, Seattle, WA, USA) using the manufacturer protocol. Each RNA sample underwent purification with 2x volume RNA binding buffer and equal volume 100% ethanol with column based filtration using a Zymo-Spin™ IC Column (Zymo Research, California, US). Each column underwent three rounds of RNA wash buffer rinsing before elution with 15µl RNase free water. Post-clean up RNA was re-quantified and quality checked using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (ThermoFisher, Massachusetts, USA). All samples were diluted to approximately 100ng total RNA input in a total volume of 3µl. Samples underwent miRNA sample preparation using the manufacturers protocol. All steps were performed on ice. In short, an annealing master mix was prepared consisting of Annealing Buffer, nCounter Tag Reagent and 1:500 dilution of miRNA Assay Controls. This master mix was added to each 3µl RNA sample volume, mixed and underwent annealing in a heated lid thermocycler under the following conditions:

Temperature	**Time**
94°C	1 minute
65°C	2 minutes
45°C	10 minutes
48°C	Hold

Ligation Buffer and polyethylene glycol (PEG) were added and incubated for 5 minutes maintaining the temperature at 48°C throughout with subsequent Ligase enzyme added. Samples were then incubated under the following conditions:

Temperature	**Time**
48°C	3 minutes
47°C	3 minutes
46°C	3 minutes
45°C	5 minutes
65°C	10 minutes
4°C	Hold

Ligation Clean-Up enzyme was added and samples incubated under the following conditions:

Temperature	Time
37°C	1 hour
70°C	10 minutes
4°C	Hold

Following the clean-up process, 40µl of RNAse-free water was added to each sample and submitted for miRNA CodeSet Hybridization Setup Protocol. Samples were run on an nCounter® MAX Analysis System.

2.2.5 Data Analysis and Statistics

Raw count output files (RCC files) were normalised using the nSolver Analysis Software 4.0 consisting of a normalisation factor calculated from combining the geometric mean expression of the in-built positive controls and geometric mean of the most stably expressed miRNAs defined by co-efficient of variance <55%. Normalised data was analysed by ROSALIND® (https://rosalind.onramp.bio/), with a HyperScale architecture developed by ROSALIND, Inc. (San Diego, CA). Read Distribution violin plots, identity heatmaps, and sample MDS plots were generated as part of the QC step. The limma R library was used to calculate fold changes and p-values and perform optional covariate correction (13). Clustering of miRNA for the heatmaps of differentially expressed miRNAs were done using the PAM (Partitioning Around Medoids) method using the fpc R library. The top targeted gene predictions, validated genes, and related drugs and disease were analysed using the multiMiR R library (14). MicroRNA secondary structures were calculated and visualized using the ViennaRNA software (15).

2.2.6 Differential expression using qPCR

Based on the findings from the pilot experiments a panel consisting of 10 miRNAs of interest (MOIs), 4 endogenous reference genes and 2 technical controls was designed (Table 4-12). Custom prefilled 96-well plates were designed and manufactured by Qiagen. Reverse-transcription, qPCR methodology and data analysis pipeline were the same as in the qPCR pilot

experiment. In total 126 samples were included in this study, divided into 65 PTC, 44 FTC and 17 normal thyroid samples.

2.3 Detection of ctDNA in early thyroid cancer using a targeted sequencing panel approach

2.3.1 Samples

All prospectively recruited patients (n = 58) from the part 2 cohort of “Gene Mutation Analysis as a biomarker in Thyroid Cancer (GEM-TC)” study (described in section 2.1.1) were included in this experiment.

2.3.2 Sample processing

DNA and RNA co-extraction from FFPE tumour samples from each patient was performed as detailed in section 2.1.2. Library preparation, quantification, sequencing and bioinformatics pipelines were performed as detailed in sections 2.1.3-5.

2.3.3 Plasma cfDNA extraction and quantification

Where possible, 30mL of peripheral blood was collected from each patient at 4 separate peri-treatment time-points (pre-operative, post-operative, pre-radioactive iodine treatment and post-radioactive iodine treatment) alongside their routine blood tests. Streck Cell-free DNA Blood collection tubes (La Vista, NA, USA) were used to minimise cell lysis and maximise cfDNA concentration. All collection tube samples were maintained at room temperature and processed within 72 hours of collection. For each time-point, the peripheral blood samples were centrifuged at 1600g for 10 minutes and the upper plasma layer from each Streck tube replicate was aspirated into a single sterile 50mL conical centrifuge tube, leaving the white buffy coat layer intact. The plasma in the conical centrifuge tube was further centrifuged at 3100g for 10 minutes and the supernatant plasma was transferred into 2mL externally-threaded universal cryogenic tubes without disturbing the pellet. Plasma samples were stored at -80°C until further processing. The buffy coat layer was aspirated and transferred to a single 2mL externally-threaded universal cryogenic tube and also stored -80°C until further processing.

Cell-free DNA was extracted from 4mL of thawed plasma using a column-based vacuum manifold nucleic acid purification kit (QIAamp Circulating Nucleic Acid kit, Qiagen, Hilden, Germany). After extraction and purification, cfDNA was eluted from the column membrane with 50µL RNase-free water with 0.04% NaN_3 (Sodium azide). Quantification of extracted cfDNA was performed using digital droplet PCR with targeted assays against 3 individual endogenous genes (*HOGA1, IRAK4, OR4C12*). Extracted cfDNA was stored in barcode labelled tubes at -20°C until further processing.

2.3.4 Buffy coat DNA extraction and quantification

For each patient, genomic DNA was extracted from 50µL of thawed buffy coat sample using a spin column based extraction kit (QIAamp DNA Blood Mini Kit, Qiagen, Hilden, Germany) and quantified using a dsDNA HS (high sensitivity) assay kit on a Qubit Fluorometer (ThermoFisher, Massachusetts, USA). Extracted DNA was stored in barcode labelled tubes at -20°C until further processing.

2.3.5 Plasma genotyping

Custom TaqMan assays were designed using the Custom TaqMan® Assay Design Tool (Thermofisher) informed by the somatic variant calls detected in the targeted panel sequencing of each patients' tumour DNA. For each plasma time point, 4mL equivalent of DNA was tested with the custom assay and concurrently run with no template control (NTC), positive control (patient tumour DNA) and negative control (patient buffy coat DNA). Custom made primers and probes were diluted from 40X stock to 20X assay cocktail. Each PCR reaction was composed of 1.1µL of 20X gene specific Primer/Probe mix, 11µL of 2X ddPCR Supermix for Probes, 8.8µL of nuclease free water and 1.1µL of DNA template. PCR reactions were emulsified on a droplet generator and then incubated in a thermocycler under the following conditions:

PCR Program			
Step 1	95°C	10 minutes	
Step 2	94°C	30 seconds	40 cycles
Step 3	60°C	60 seconds	
Step 4	98°C	10 minutes	
Step 5	4°C	∞	

Samples were read on a QX-200 ddPCR droplet reader (Bio-Rad) and results were analysed using QuantaSoft software v1.7.4. A minimum of 10,000 droplets per well were required for valid downstream analysis. A minimum of 2 single FAM positive droplets was required to make a positive call.

2.3.6 Statistics

Prism 8 software (GraphPad, San Diego, CA, USA) was used for simple statistical testing. Statistical tests used were as follows:

Section 5.4.1 – Unpaired two-tailed t-test for continuous variables and Fisher's exact test for binary categorical variables.

Section 5.4.2 – One-way ANOVA for multiple comparisons of total cfDNA extracted between various time-points.

Section 5.4.3 – Testing of correlation between variant allele fractions detected by NGS and dPCR was performed using Pearson's rank correlation test.

2.4 Bespoke patient specific sequencing in early thyroid cancer

2.4.1 Samples

Fifteen patients were selected from the part 2 cohort of “Gene Mutation Analysis as a biomarker in Thyroid Cancer (GEM-TC)” study (described in section 2.1.1) and included in this experiment. This was a pilot study during the prospective recruitment time frame, and as such patients were selected on the basis of availability of plasma time-points (>3) and providing a range of histopathological characteristics and index operations.

2.4.2 Tumour whole exome sequencing and bioinformatics pipeline

Extracted and quantified DNA samples from the patients’ tumour and matched buffy coat samples were sent to an external company (GENEWIZ, Brooks Life Sciences, UK) to perform whole exome sequencing. Agilent SureSelect Human All Exon V6 chemistry was used for library preparation and sequenced on an Illumina platform with 2x150bp configuration.

Raw BCL files generated by the sequencer were converted to fastq files for each sample using bcl2fastq v.2.19. Sequence reads were trimmed to remove possible adapter sequences and nucleotides with poor quality using Trimmomatic v.0.38. The trimmed reads were mapped to the reference genome using the Illumina Dragen Bio-IT Platform. BAM files were generated as a result of this step. Somatic variants were called using the Illumina Dragen Bio-IT Platform in somatic mode. A panel of normal (PON) was also used to remove technical artefacts. Variants were further filtered and any variants in the follow categories were considered as false positives and removed: i) marked as common variants in dbSNP build 151 and ii) non_cancer_AC > 5 in gnomad exome database r2.1.1. The filtered VCF was then annotated with Ensembl Variant Effect Predictor (VEP) v95.

2.4.3 Bespoke patient specific panel creation

Based on the somatic variants detected by WES for each patients’ tumour, bespoke patient specific assays were designed using the Archer^{TM} Assay

Designer (ArcherDX, Boulder, Colorado, US). This is an online portal based designer with in-silico bench testing to predict primer coverage based on a number of metrics including base repeating sections and GC content. The assays are created using the VariantPlex™ assay in conjunction with the Archer Universal DNA reagent kit V2 for Illumina (AK0037-8) and Archer MBC Adapters for Illumina.

2.4.4 Library preparation and quantification

Illumina compatible library preparation was performed using the Archer™ Liquidplex for Illumina kit following the manufacturer's protocol. In summary, total purified cfDNA extracted from 4ml of plasma or matched fragmented buffy coat gDNA was used as input (in 50uL volume). The process involved end repair of DNA fragments, ligation of molecular barcodes, ligation of unique indexing sequences and two nested PCR steps using gene specific primers related to the bespoke panels created for each patient. Each stage was followed by AMPure XP bead (Beckman Coulter, Brea, CA, USA) clean-up and washing with 70% ethanol. All reactants for each stage were supplied in lyophilized beads and kept on ice throughout the library preparation process. The PCR cycling conditions were as follows:

	Step	Temperature (°C)	Time	Cycles
First PCR Reaction	1	95	3 min	1
	2	95	30 sec	11
	3	65	15 min (100% ramp rate)	
	4	72	3 min	1
	5	4	Hold	1
Second PCR Reaction	1	95	3 min	1
	2	95	30 sec	15*
	3	65	15 min (100% ramp rate)	
	4	72	3 min	1
	5	4	Hold	1

Completed libraries were quantified using the KAPA Universal Library Quantification Kit (Roche Sequencing, Switzerland) according to the manufacturer's protocol and using the supplied DNA standards (n = 6) with range from 20 to 0.0002 pM to generate the standard curve. Samples and standards were processed in a QuantStudio 6 Flex from Thermofisher under the following conditions:

Step	Temp.	Duration	Cycles
Initial denaturation	95°C	5 minutes	1
Denaturation	95°C	30 seconds	35 x cycles
Annealing/Extension/Data acquisition	60°C	45 seconds	
Melt curve analysis	65-95°C		

Libraries were normalised to 4 nM concentration and pooled for sequencing.

2.4.5 Sequencing and bioinformatics

Pooled libraries were sequenced on a NovaSeq SP 300 v1.5 cartridge on an Illumina platform (Illumina, San Diego, CA, USA) according to the manufacturer's instructions and 20% PhiX spike in. Raw BCL files generated by the sequencer underwent de-multiplexing and conversion to fastq files for each sample using bcl2fastq v.2.19. Fastq files underwent alignment, annotation and filtering using both in-house and Archer Analysis pipelines (Archer Analysis 6.2 site, ArcherDX, Boulder, Colorado, US). Positive variant calls were loci specific, and based upon intra-sample AF outlier p-value with significance set to $p < 0.001$. Additionally calls were manually curated on IGV. Positive ctDNA status was defined as any sample with 1 or more variants detected.

2.5 References

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.
2. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 0 1- 0 33.
3. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
4. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.
5. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40(22):11189-201.

6. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
7. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.
8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
9. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
10. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.
11. Lê S JJ, Husson F. FactoMineR: A Package for Multivariate Analysis. Journal of Statistical Software. 2008;25(1):1-18.
12. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245-50.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
14. Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. Nucleic Acids Res. 2014;42(17):e133.
15. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.

Chapter 3 Gene mutation signature for vascular invasion in thyroid cancer

3.1 Introduction

Vascular invasion is one of a few key histopathological features that can upstage DTCs from low to higher risk tumours when considering later risk of recurrence or distant metastatic spread. It is a key feature in the ATA risk of recurrence prognostication scale and for this reason, can influence the extent of surgical treatment and whether adjuvant RAI is recommended (1). Vascular invasion has been reported as an important and independent prognosticator in thyroid cancer and can only be identified on histological examination of a surgically resected specimen. The ability to predict the presence of higher risk histopathological tumour features, such as vascular invasion, through surrogate biomarkers found either circulating in the blood or from initial fine needle aspiration biopsy sample, has the potential to provide pre-treatment information prior to definitive surgical treatment. This information could reduce the rate of two-stage surgery (initial hemi-thyroidectomy and subsequent completion thyroidectomy), towards up-front single-stage total thyroidectomy.

Various surrogate biomarkers have been explored in the literature, but most promising are differential genomic aberrations, such as specific associated gene mutations, which could be identified on simple targeted sequencing of extracted DNA. Somatic mutational signatures associated with vascular invasion have been described in other solid cancers, such as breast, endometrial and hepatocellular cancers (2-4). The TCGA database describes the genomic landscape of 496 PTCs, but whilst publically available it is unfortunately deficient in the annotated histopathological vascular invasion status for each sample, and also does not contain any follicular thyroid carcinoma samples. There is therefore scope to focus specifically on vascular invasion in both PTC and FTC, to identify associated or predictive genomic biomarkers.

On reviewing current thyroid cancer literature, only two studies have been found that report a significant association between certain genetic mutations and the presence of vascular invasion (5, 6). These studies involved US and Saudi Arabian patient cohorts and are limited by selectively focusing on 1 or 2 gene mutation loci. This may limit the sensitivity and specificity of detection and strength of association. By using a panel of thyroid specific mutational hotspots, this study aims to identify a mutational signature that can reliably predict the presence of vascular invasion in DTC.

3.2 Hypothesis

A gene mutation signature is associated with the presence of vascular invasion in differentiated thyroid cancer.

3.3 Aims

1. To sequence thyroid tumours with and without vascular invasion using a thyroid cancer-specific targeted panel
2. To correlate the presence of single or combination of variants with vascular invasion and/or other histopathological phenotypes
3. To correlate the presence of single or combination of variants with other histopathological phenotypes such as extra-thyroidal extension (ETE), multifocality, lymph node involvement.

3.4 Results

3.4.1 Patient demographics and histopathological features

The total number of patients recruited from the "Gene Mutation Analysis as a biomarker in Thyroid Cancer (GEM-TC)" study (described in chapter 2.1) were 69 patients in the retrospective part 1 cohort and 58 patients in the prospective part 2 cohort giving a total cohort of 127 patients (Figure 3-1). The patient demographics are summarised in Table 3-1. After exclusions, the total cohort available for study was n = 118.

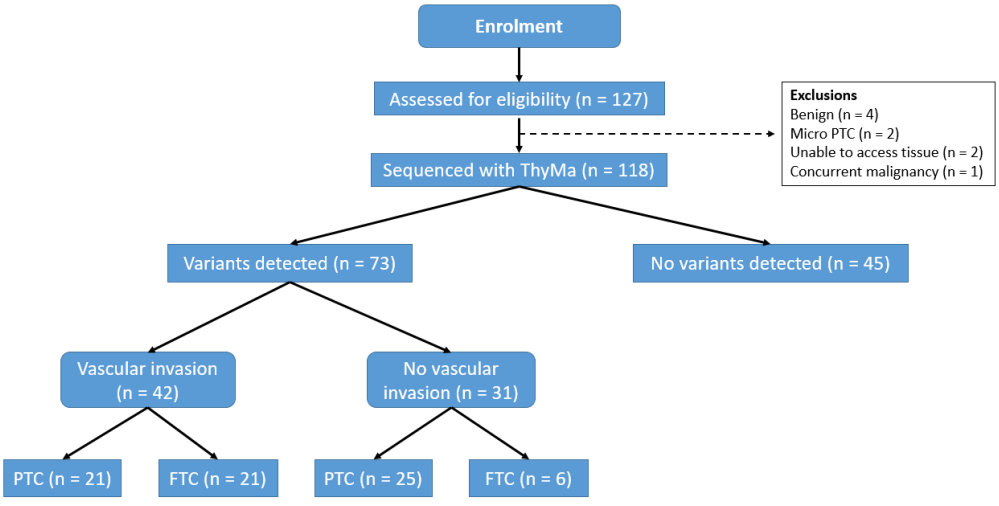


Figure 3-1. CONSORT diagram of study

As described in section 2.1.3 of the Materials and Methods chapter 2, the available cohort of patient tumours underwent targeted sequencing using a custom amplicon based target enrichment panel (called ThyMa). Of the 118 patient tumours sequenced with ThyMa, 73 (61.9%) had at least 1 variant detected and were analysed further.

3.4.2 Samples with variants

The histological subtypes of the cohort of tumours with variants were as follows: 42 classical PTC (cPTC), 3 follicular-variant PTC (fvPTC), 1 Tall Cell PTC, 19 minimally invasive FTC (MI-FTC), 8 widely invasive FTC (WI-FTC). 4 samples had poorly differentiated components that were <50% of the tumour.

3.4.3 Histopathological characteristics of tumours

Various histopathological features were documented for each tumour, including tumour size, presence or absence of vascular invasion, extra-thyroidal extension, multifocality, lymph node involvement and metastatic disease on presentation. Longer term clinical outcomes were limited to the presence or absence of recurrent disease (at study end-point) and if the patient had radioactive iodine treatment. Assessing survival outcomes were not feasible due to the limited long term patient follow up. The majority of tumours included in the study were less than 4cm in size (see Figure 3-2) and the average tumour size for the whole cohort was 36.19mm, with PTC tumours on average being smaller (29.85mm) compared to FTC tumours (47mm). There were a range of TNM stages, with the majority being T3 and below, N0 and M0. All histopathological characteristics according to histology are summarised in Table 3-1.

	Follicular (N=27)	*Papillary (N=46)*	*Total (N=73)*
Sex			
Male	14 (51.9%)	17 (37.0%)	31 (42.5%)
Female	13 (48.1%)	29 (63.0%)	42 (57.5%)
Age (years)			
Mean (SD)	55.7 (14.1)	51.5 (17.5)	53.1 (16.3)
Median [Min, Max]	56.0 [29.0, 78.0]	49.5 [23.0, 85.0]	52.0 [23.0, 85.0]
Tumour Size (mm)			
Mean (SD)	47.0 (22.8)	29.8 (15.3)	36.2 (20.1)
Median [Min, Max]	45.0 [12.0, 99.0]	29.0 [10.0, 80.0]	35.0 [10.0, 99.0]
T staging			
1b	2 (7.4%)	9 (19.6%)	11 (15.1%)
2	7 (25.9%)	14 (30.4%)	21 (28.8%)
3	16 (59.3%)	17 (37.0%)	33 (45.2%)
3a	1 (3.7%)	0 (0%)	1 (1.4%)
4a	1 (3.7%)	1 (2.2%)	2 (2.7%)
3b	0 (0%)	2 (4.3%)	2 (2.7%)
4	0 (0%)	2 (4.3%)	2 (2.7%)
4b	0 (0%)	1 (2.2%)	1 (1.4%)
N staging			
0	25 (92.6%)	22 (47.8%)	47 (64.4%)
1a	1 (3.7%)	11 (23.9%)	12 (16.4%)
1b	1 (3.7%)	12 (26.1%)	13 (17.8%)
2	0 (0%)	1 (2.2%)	1 (1.4%)
M staging			
1	1 (3.7%)	4 (8.7%)	5 (6.8%)
0	26 (96.3%)	42 (91.3%)	68 (93.2%)
Vascular Invasion			
Present	21 (77.8%)	21 (45.7%)	42 (57.5%)
Absent	6 (22.2%)	25 (54.3%)	31 (42.5%)
Extra-thyroidal extension			
Present	6 (22.2%)	20 (43.5%)	26 (35.6%)
Absent	21 (77.8%)	26 (56.5%)	47 (64.4%)
Lymph node invasion			
Present	3 (11.1%)	22 (47.8%)	25 (34.2%)
Absent	24 (88.9%)	24 (52.2%)	48 (65.8%)
Multi-focal disease			
Present	9 (33.3%)	17 (37.0%)	26 (35.6%)
Absent	18 (66.7%)	29 (63.0%)	47 (64.4%)
Recurrence			
Present	11 (40.7%)	9 (19.6%)	20 (27.4%)
Absent	16 (59.3%)	37 (80.4%)	53 (72.6%)
Mets on presentation			
Present	4 (14.8%)	3 (6.5%)	7 (9.6%)
Absent	23 (85.2%)	43 (93.5%)	66 (90.4%)
RAI treatment			
Yes	25 (92.6%)	42 (91.3%)	67 (91.8%)
No	2 (7.4%)	4 (8.7%)	6 (8.2%)

Table 3-1. Summary of patient clinical and histological characteristics

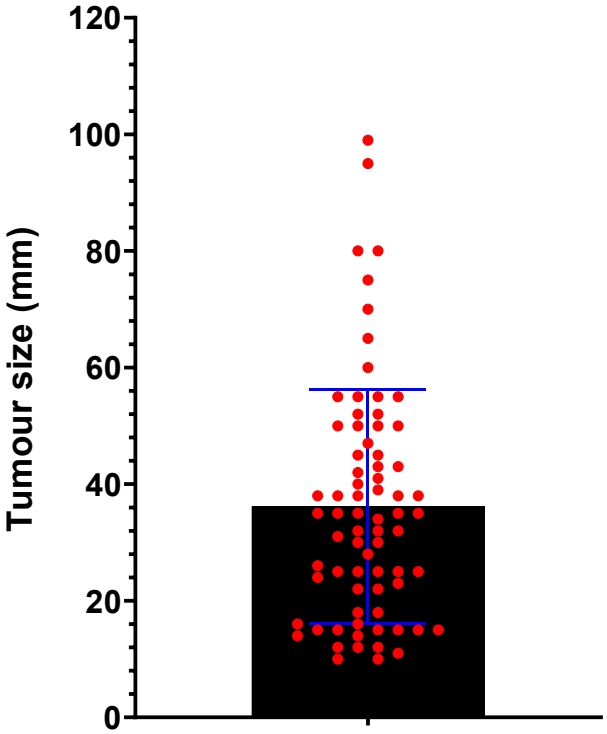


Figure 3-2. Boxplot showing distribution of tumour sizes in the whole patient cohort

3.4.4 Vascular invasion as a marker of poorer outcome

The total cohort was dichotomised by vascular invasion and disease relapse status and the proportion of relapse patients with vascular invasion was 80% (16/20). The Fisher's exact test for proportions suggested a significant association between vascular invasion and disease relapse ($p = 0.0192$). Sub-cohort analysis showed that in PTC samples, vascular invasion was significantly associated with the presence of extra-thyroidal extension ($p = 0.0360$) (see Figure 3-3).

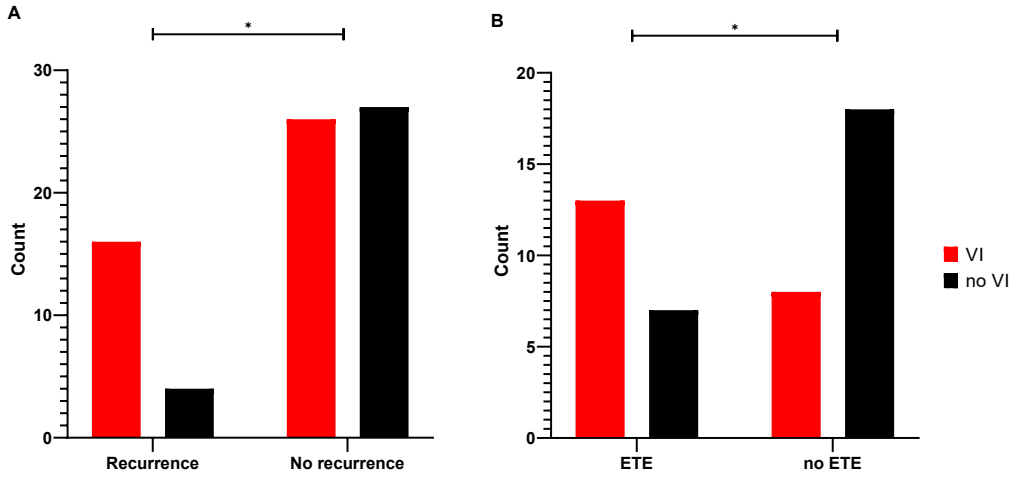


Figure 3-3. Vascular invasion as a marker of poorer outcome. A. Bar chart showing the number of samples with and without disease recurrence according to vascular invasion status. **B.** Bar chart showing the number of PTC samples with and without extra-thyroidal extension (ETE) according to vascular invasion (VI) status.
* represents p <0.05

3.4.5 Summary of variant discovery

All variant calls were filtered against the most up to date COSMIC database for thyroid cancer (accessed 26/03/2020) (7). A summary heatmap of variants detected alongside tumour phenotype for all analysed tumours is shown in Figure 3-4 below. The most common variant detected was *BRAF* p.V600E mutation, found exclusively within the PTC subgroup and at a frequency of 78.3% of PTC samples. This is higher than published figures in COSMIC (54.6%) and TCGA (58%) (Figure 3-5). The second and third most frequent variants were found in *TERT* promoter (*TERTp*) and *NRAS* regions respectively. In PTC samples, *TERTp* mutations were found in 19.6% of samples, with *NRAS* mutations present in 8.7% (Figure 3-5).

In FTC samples, *TERTp* mutations were found in 55.6% of samples with *NRAS* mutations present in 51.9%. The FTC tumours had a more heterogenous mutational profile compared to PTC (Figures 3-5 and 3-6). The relatively high frequency of *TERTp* variants called by the sequencing prompted validation studies. All *TERTp* calls were confirmed using an orthogonal technique (droplet digital PCR) where the tumour DNA was tested using bespoke assays to the *TERTp* variant loci.

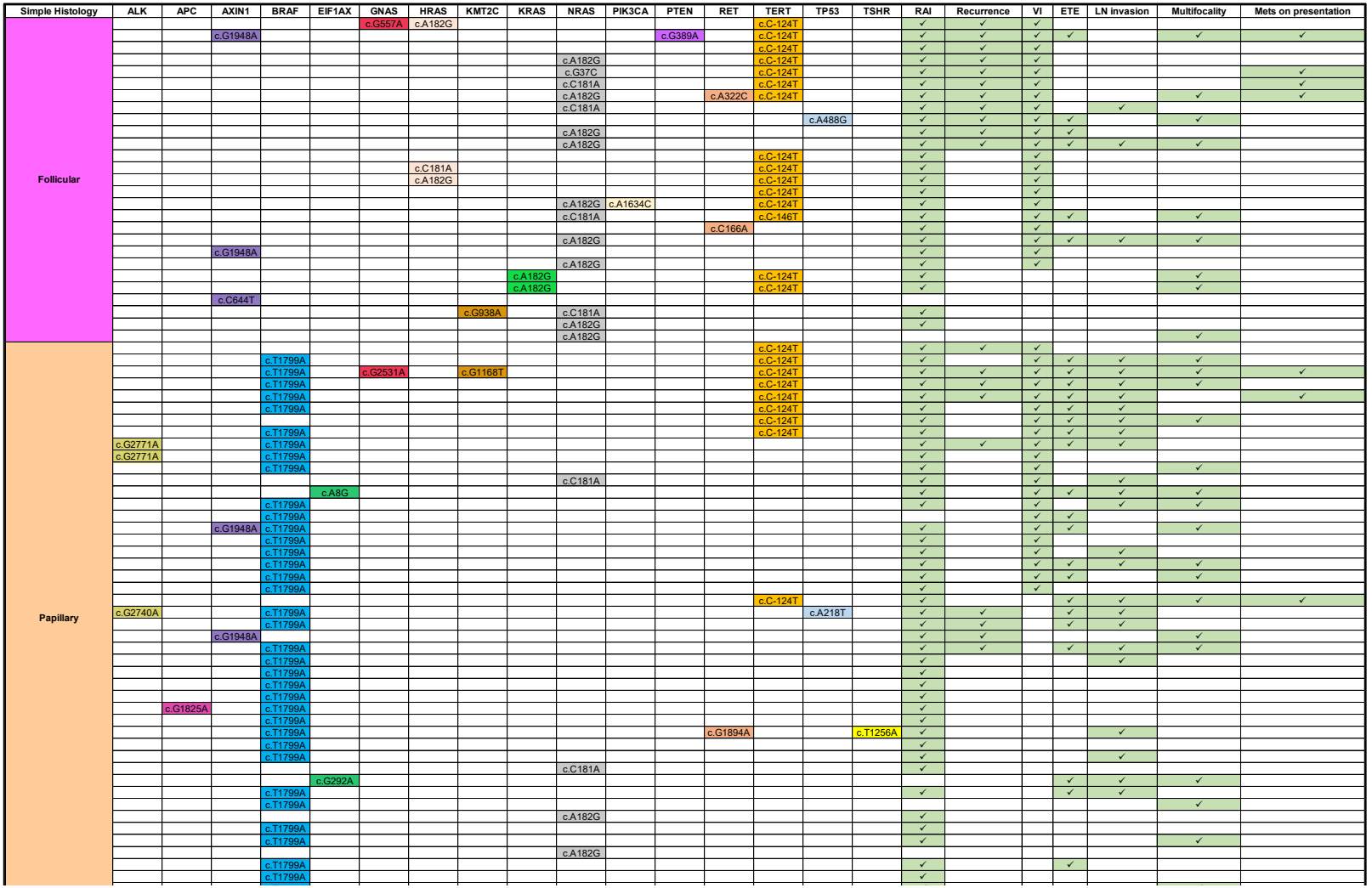


Figure 3-4. Mutation heat map from all tumours with a variant detected. Each row represents a single patient and the boxes are highlighted corresponding to the presence of the variant labelled at the top of the column. The HGVSc for each variant is detailed inside the box.

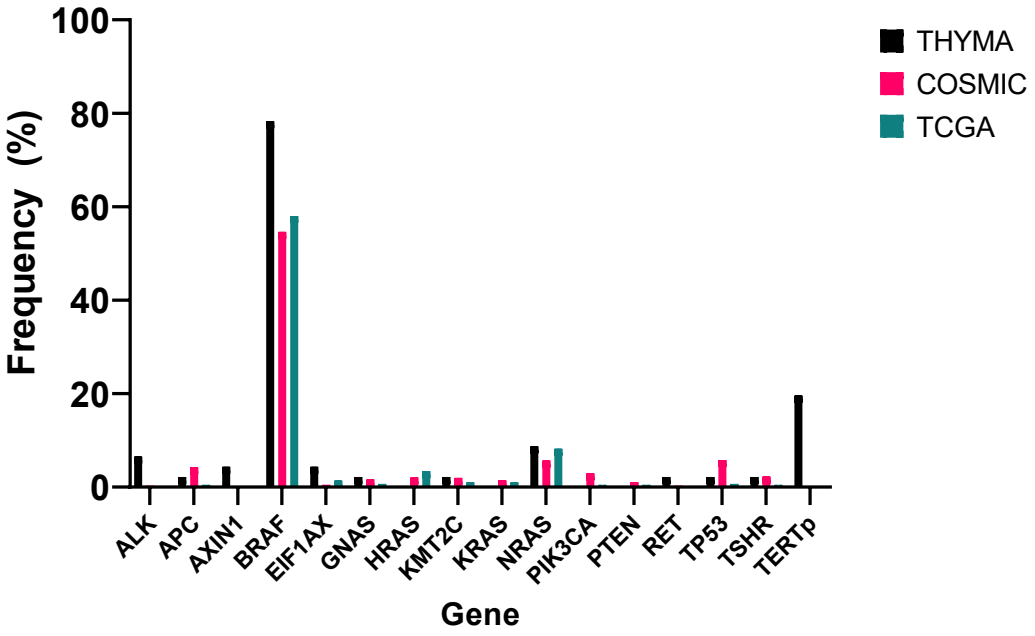


Figure 3-5. Frequency of gene variants detected in PTC tumours following targeted sequencing with the ThyMa thyroid specific panel compared to published data

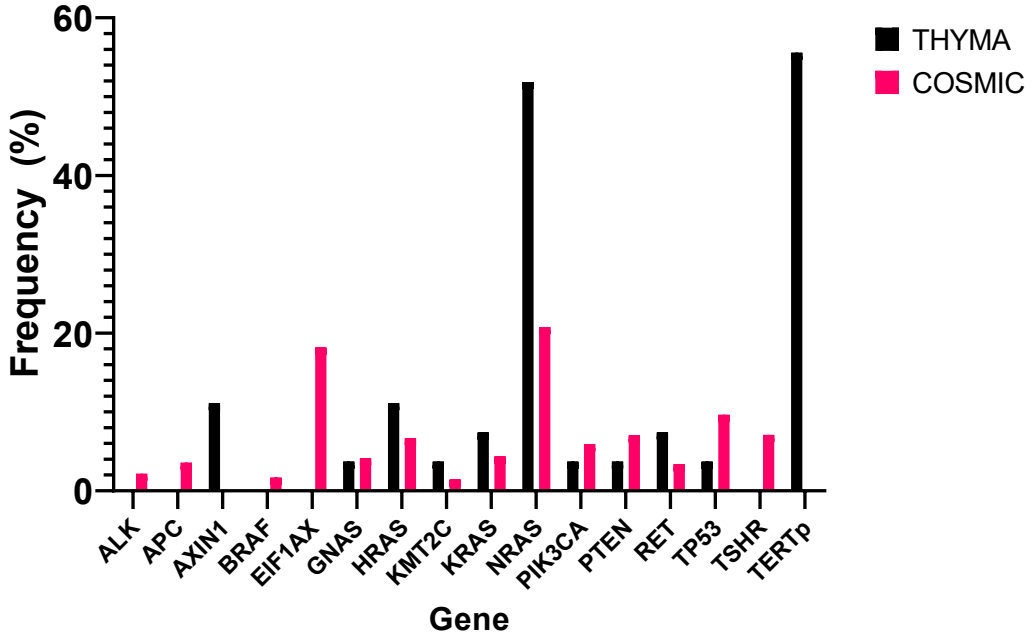


Figure 3-6. Frequency of gene variants detected in FTC tumours following targeted sequencing with the ThyMa thyroid specific panel compared to published data

The variants detected in the *NRAS* gene were all canonical p.Q61K or p.Q61R in exon 3. The *TERTp* variants were c.C1-123T except for one sample that had a c.C1-146T variant. In keeping with the findings from the TCGA study, PTC tumours exhibited mutually exclusive *BRAF*-like or *RAS*-like profiles (8). *TERTp* mutations were more common in follicular tumours (55.6%) compared to papillary tumours (19.6%).

On sub-analysis of the FTC cohort, we observed differences in variant frequency between MI-FTC and WI-FTC tumours (n = 19 and 8 respectively) (see Figure 3-7). There were a higher proportions of *GNAS*, *NRAS*, *PTEN*, *RET* and *TERTp* variants found in WI-FTC tumours, with 75% harbouring a *TERTp* mutation.

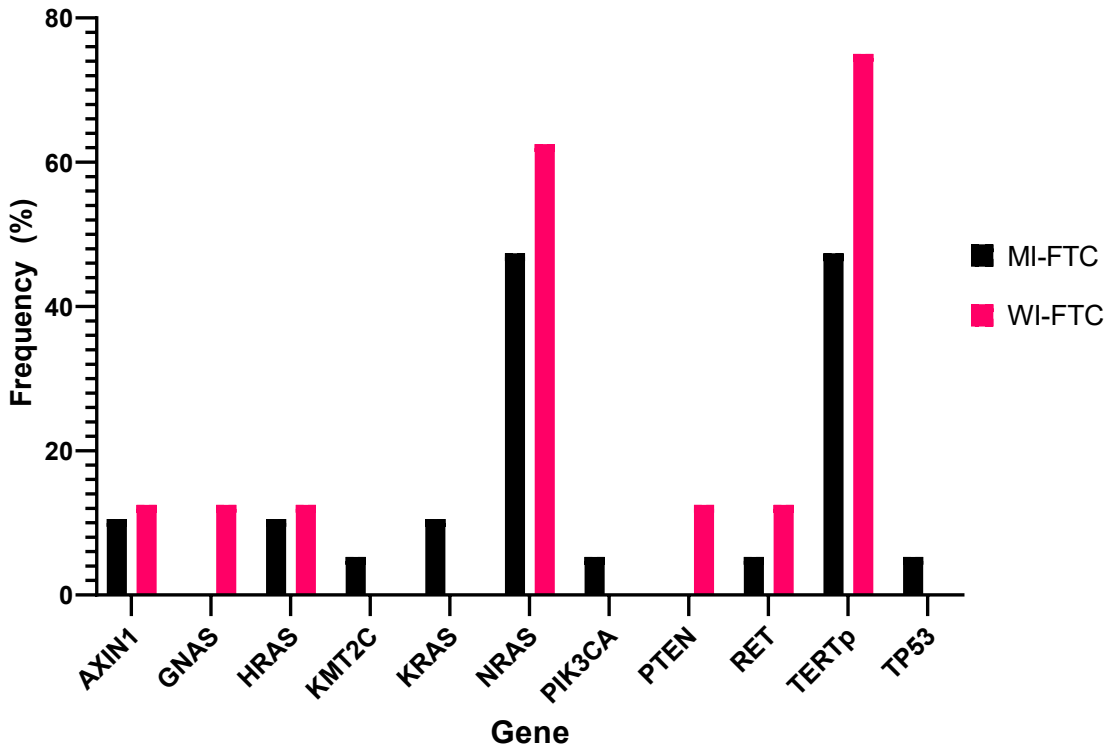


Figure 3-7. Frequency of gene variants detected in FTC tumour subdivided into minimally invasive and widely invasive histological subtypes. MI = minimally invasive, WI = widely invasive

To maximise tumour purity input into the nucleic acid extraction process, each FFPE slide underwent needle dissection using light microscopy to include only tumour tissue and leave adjacent normal thyroid tissue (as described in section 2.1.2). This maximises the chance of achieving an accurate representation of variant allele frequency of each tumour. Allele frequency for all detected variants ranged from 0.05 (found in *ALK*) to 0.75 (found in *NRAS*). Known driver mutations *BRAF*, *NRAS*, *KRAS*, *HRAS* had mean allele frequencies of 0.21 (±0.095), 0.38 (±0.13), 0.44 (±0.03), 0.35 (±0.25) respectively (see Figure 3-8).

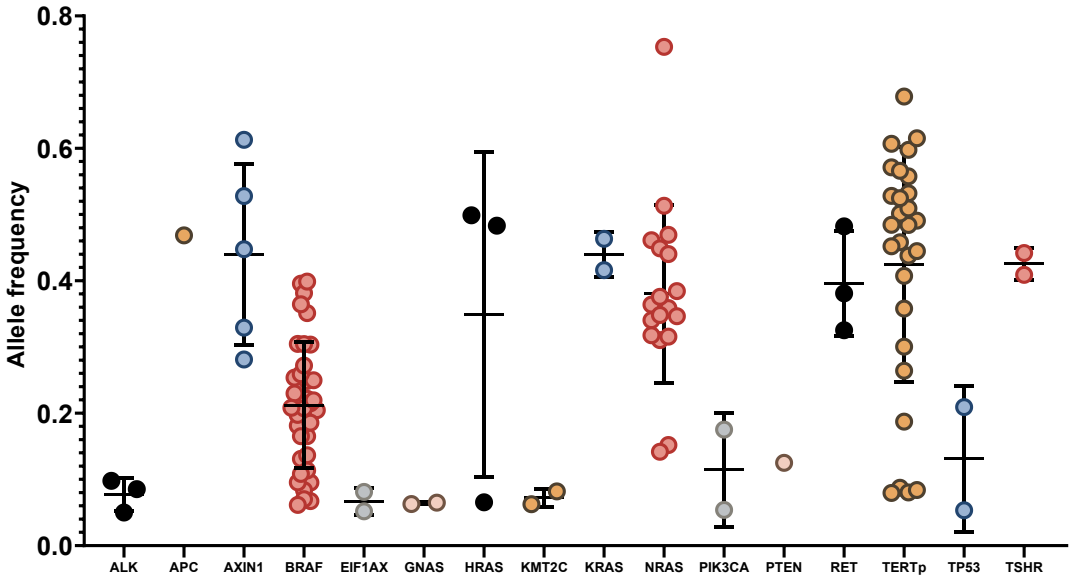


Figure 3-8. Strip chart showing allele frequency distribution of each detected gene variant. Higher allele frequency suggests clonal dominance of cells harbouring that particular variant.

3.4.6 Samples with multiple variants

Twenty-five samples had 2 or more variants detected, and these samples had higher proportions of aggressive pathological features (see Figure 3-9) with a statistically significant difference seen in disease recurrence, vascular invasion and distant metastases (p = 0.02, 0.02 and 0.0025 respectively). Seventy-two percent (18/25) of these samples with multiple variants had a *TERTp* mutation compared to 12.5% (6/48) of samples with only a single variant. There were no significant differences in tumour size or histological subtype when comparing samples with single versus multiple variants.

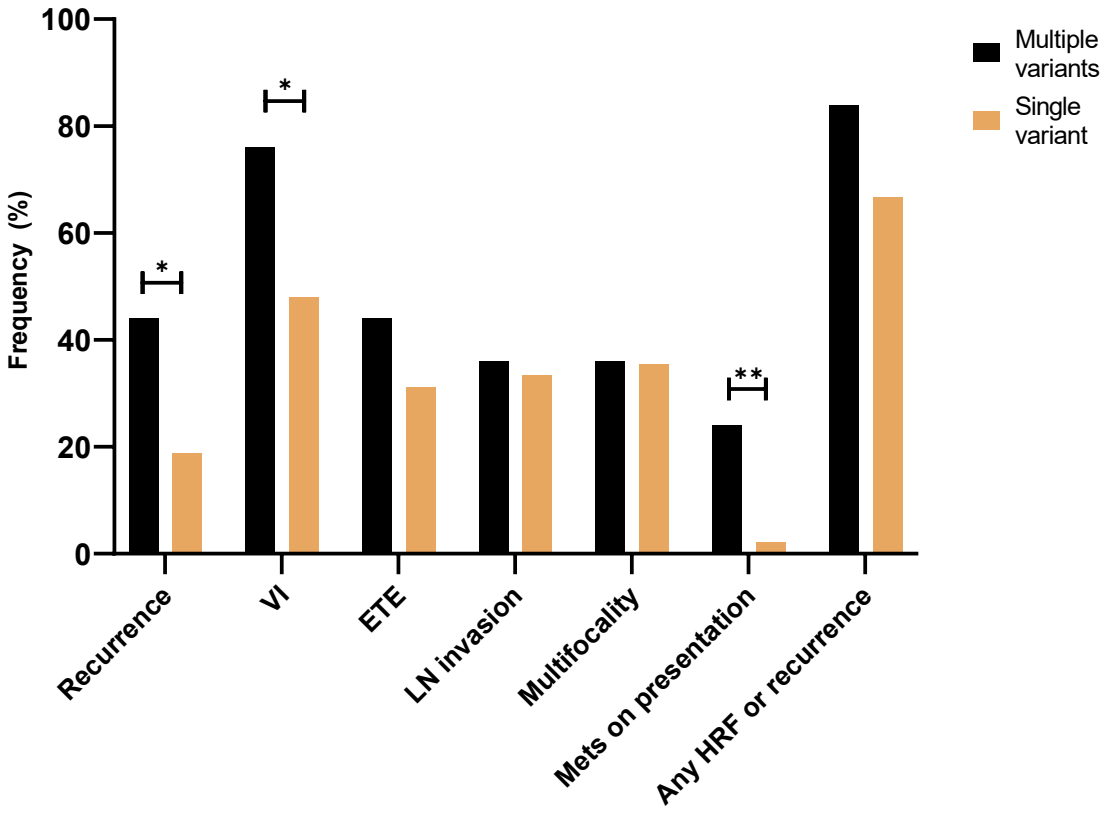


Figure 3-9. Bar chart showing different proportions of high risk pathological features between tumours with multiple variants compared to single variants. Chi squared test for proportions showed significant differences in recurrence (p = 0.02), vascular invasion (p = 0.02) and metastatic disease on presentation (p = 0.0025).

3.4.7 Multiple correspondence analysis (MCA)

To discern which single or combination of mutated genes were associated with the different histopathological features recorded for each tumour, a multiple correspondence analysis (MCA) was carried out using the FactoMineR package in R (9). Full methodology is described in section 2.1.6 in Chapter 2.

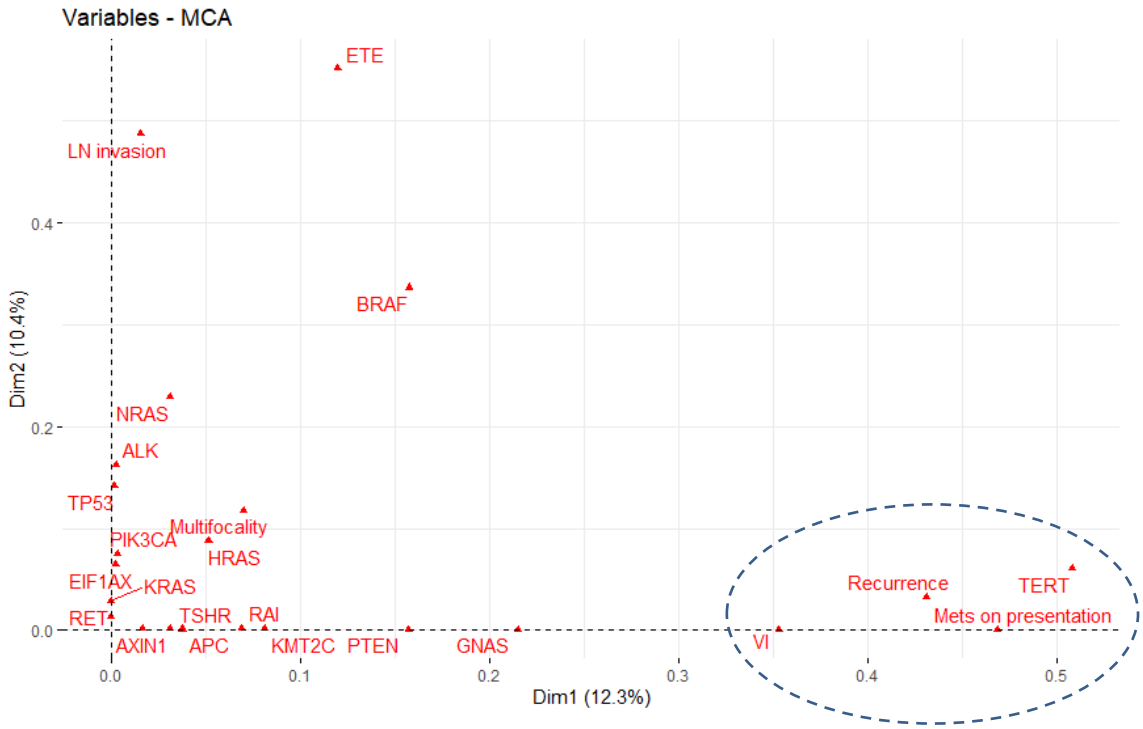


Figure 3-10. Multiple correspondence analysis plot to explore relationships between the independent variables recorded for each tumour in two dimensions. This plot helps to identify variables that are most correlated with each principal dimension. The two dimensions are sufficient to retain 22.7% of the total inertia (variability) in the dataset. Variable categories with similar profile are grouped together. Grouping between VI, disease recurrence and metastatic disease with TERTp mutation, highlighted in blue.

Multiple Correspondence Analysis (MCA) is a multivariate analysis method that is built upon Principal Component Analysis. The MCA technique was used as it allows analysis of nominal data and summarises the relationship between several categorical variables in a large dataset. These are then plotted on a two-dimensional plot according to the contribution of inertia attributed by the variance. The more a variable contributes to the variation of the dataset the further along each axis it will be plotted, and variables that associate tend to group together. This analysis suggested associative grouping of vascular invasion, recurrence and metastatic disease with the presence of a *TERTp* mutation (see Figure 3-10).

3.4.8 *TERTp* c.1-124C>T as a predictor of high risk histopathological features

The gene mutation heatmap (Figure 3-4) and multiple correspondence analysis (Figure 3-9) gave a visual indication of potential correlation between *TERTp* variants and the presence of vascular invasion, recurrent disease and distant metastases. Further associative analysis was performed using Fisher's Exact

statistical test. This showed that in the whole DTC cohort, presence of a *TERTp* variant was significantly associated with the presence of vascular invasion (p = 0.0003).

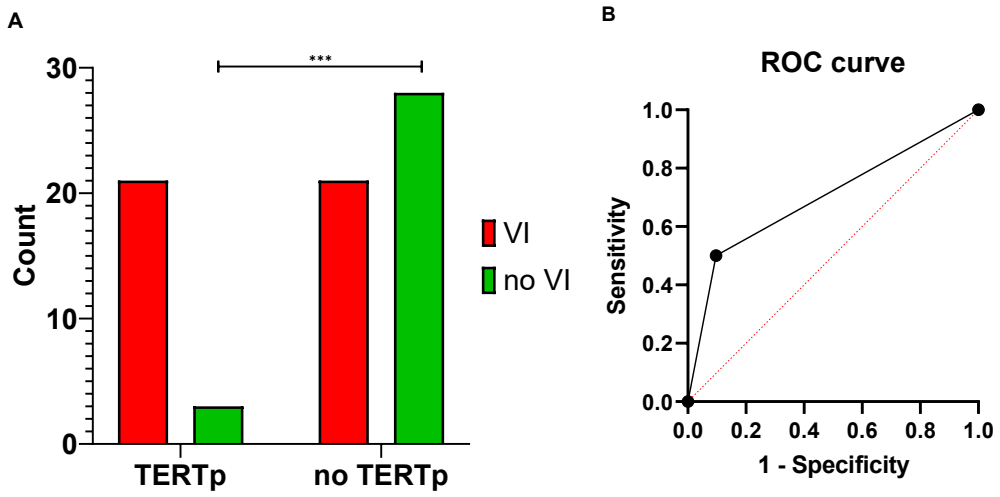


Figure 3-11. Results of Fisher's exact statistical test of *TERTp* status as a predictor of vascular invasion status. A. Bar chart showing the number of samples with and without *TERTp* variants according to vascular invasion status. **B.** Receiver operating characteristic (ROC) curve showing the diagnostic ability of *TERTp* status as a binary classifier of vascular invasion status. *** represents p = 0.0003

Used alone as an independent predictor of vascular invasion, presence of a *TERTp* variant would yield a positive predictive value (PPV) of 57.1%, negative predictive value (NPV) of 87.5%, sensitivity 50%, specificity 90%, area under the ROC curve 0.70 (see Figure 3-11).

Positive *TERTp* variant status was also significantly associated with the presence of disease recurrence and distant metastases (p = 0.02 and 0.0002 respectively). Of the samples with both vascular invasion and *TERTp* variant, 52.4% had disease recurrence and 28.6% had metastatic disease on presentation, compared with those samples with vascular invasion without *TERTp* variant where 23.8% had disease recurrence and none had metastatic disease on presentation.

On sub-analysis based on histology, the statistically significant association was only observed in PTC tumours and not FTC tumours (p = 0.0067, odds ratio 14.77 (95% CI 2.15 – 169.4).

3.4.9 Multiple logistic regression to improve predictive value

To improve the sensitivity of *TERTp* status as a predictive test, multiple binary independent variables were combined with *TERTp* status to perform multiple logistic regression on the total DTC dataset (see Section 2.1.6 in Chapter 2 for full methodology). A combination of *BRAF* variant positive, any *RAS* variant positive, presence of multiple variants and *TERTp* positive produced a predictive model with 59.5% sensitivity, 77.4% specificity, PPV 58.5%, NPV 78.1%, area under the ROC curve 0.74 and therefore 67% accuracy.

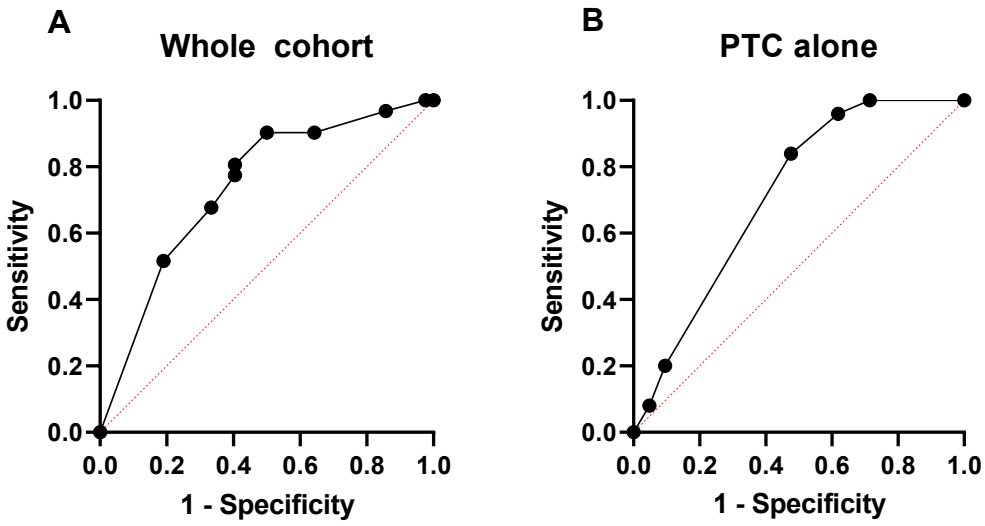


Figure 3-12. Results of multiple logistic regression combining independent variables into a predictive model. A. ROC curve showing the ability of the multiple logistic regression predictive model to predict the presence of vascular invasion in the whole cohort (AUC 0.74). **B.** ROC curve showing the ability of the multiple logistic regression predictive model to predict the presence of vascular invasion in the PTC cohort (AUC 0.71).

As the association was only statistically significant between PTC subtype and *TERTp* variants, this was repeated in the PTC cohort alone. The same combination of independent variables produced a predictive model with 52.4% sensitivity, 84.0% specificity, PPV 67.7%, NPV 73.3%, area under the ROC curve 0.71 and therefore 70% accuracy (see Figure 3-12).

3.5 Discussion

The presence of vascular invasion within DTC on definitive histological analysis has an impact on MDT recommendation for further treatment. There is clear evidence of the association between vascular invasion and poorer clinical outcomes in DTC, including increased rates of tumour recurrence and worse disease specific survival (10).

In our current thyroid cancer management paradigm of treatment de-escalation in low-risk 1-4cm DTC, pre-operative identification of vascular invasion using a molecular marker would be advantageous in guiding initial surgical extent. This could give clinicians and patients additional information to decide whether up-front total thyroidectomy should be performed instead of initial hemi-thyroidectomy and completion surgery. This may therefore reduce the number of two-stage surgical procedures with the associated costs and patient anxiety. The TCGA study in 2014 was the first integrative genomic analysis study in thyroid cancer and studied 496 PTCs. The clinical and genomic data for these tumours are publicly accessible on platforms like the NIH National Cancer Institute GDC data portal and cBioportal (11, 12). Unfortunately, the clinical data for these samples lacks accurate information about their vascular invasion status to be useful in association studies.

In this study, we have described the somatic mutational profile of 73 DTC tumours, 42 with vascular invasion and 31 without, from a UK adult patient population, using a thyroid specific targeted sequencing panel.

3.5.1 Biomarkers of high-risk histopathological features in DTC

Our study findings add to the limited body of evidence in the literature that have identified *TERTp* c.1-124C>T as not only a marker of aggressive disease, but significantly associated with the presence of vascular invasion in DTC.
The underlying oncogenesis of DTC appears to be driven by only a few somatic genetic aberrations. This has been confirmed in previous large cohort studies that have described DTC as a cancer of low tumour mutational burden. We hypothesised that despite this observation, tumours with more aggressive

phenotypes may be driven by multiple somatic mutations and that these observed mutations could potentially be combined into a mutational signature for reliable detection.

We have shown that in our cohort of patients, a DTC tumour with 2 or more mutational variants is significantly associated with aggressive pathological features such as disease recurrence (OR 3.40, 95% CI 1.21-10.47, p = 0.02), vascular invasion (OR 3.44, 95% CI 1.16-10.44, p = 0.02) and distant metastases (OR 14.84, 95% CI 2.05-173.1, p = 0.0025). This association appeared to be independent of histology or tumour size. This is consistent with findings previously by the TCGA study (8) but also more recently by Colombo et al who had a series of 208 PTCs that underwent genotyping and found that mutation density was significantly correlated with a higher risk for persistent disease (13).

The most frequently detected mutations in the whole DTC cohort were in *BRAF*, *TERTp* and *NRAS* genes, found in 49.3%, 32.9% and 24.7% of samples respectively. Interestingly, in those samples with two or more variants, *TERTp* variants were found in 18 out of 25 samples (72%). This could suggest that *TERTp* c.1-124C>T mutations occur more commonly in tandem with other somatic mutations as part of increasing mutational burden rather than in isolation.

3.5.2 Mutational associations in DTC

Overall, the somatic mutational landscape of our cohort was comparable to reported variant frequencies in publicly accessible databases (7, 8).
As has been previously reported in the literature, we observed that FTC tumours have a different mutational landscape compared to PTC (14-16). Variants found in *PIK3CA*, *PTEN* and *RAS* genes were more associated with FTC tumours whereas PTC tumours were predominantly driven by the single nucleotide variant *BRAF* p.V600E. In our PTC cohort, *BRAF* p.V600E mutation was detected in 36 out of 46 (78.3%), which is comparable to the frequency of 40-80% commonly quoted in literature review (8, 14, 15, 17, 18). There were no *BRAF* mutations detected in any FTC tumour. This single nucleotide variant is thought to cause constitutive activation of the MAPK pathway, refractory to

normal negative feedback control and uncontrolled downstream activation of cellular mechanisms of proliferation, cell survival and protein synthesis. Interestingly, we did not find a significant association between *BRAF* mutation status and any of the high-risk histopathological features. Whilst there are many studies that have demonstrated a role of *BRAF* p.V600E in thyroid tumour aggressiveness and mortality (19-21), there are several DTC studies that agree that this mutation may broadly have a role in the first step of tumorigenesis rather than a consistent predictor of worse outcome (22-25).

There are few studies in the literature that report the somatic mutational landscape of either widely or minimally invasive FTC tumours. In our cohort, we had 19 MI-FTC and 8 WI-FTC. We observed that WI-FTC had a higher frequency of *TERTp*, *PTEN*, *GNAS* and *NRAS* mutations compared to MI-FTC. The frequency of *RAS* related variants (*H/N/KRAS*) in our FTC cohort was comparable with a study by Yoo et al, who performed RNA sequencing on 180 thyroid tumours, including 30 MI-FTCs (16). They reported RAS variants in 15 out of 30 (50%) in their study cohort, whereas 70% of our FTC samples had a *RAS* variant. The comparative increase could be explained by our inclusion of more aggressive FTC tumours.

3.5.3 *TERT* promoter (*TERTp*)

Telomerase is a ribonucleoprotein complex that maintains telomere length at the end of chromosomes, which increases cellular immortality, proliferation and decreases apoptosis (26, 27). *TERT* is the catalytic subunit of telomerase, and overexpression has been shown to be associated with a number of human cancers including melanoma, gliomas and also thyroid cancer (28-33).
Two *TERTp* mutations, chr5:1,295,228C>T (C228T) and chr5:1,295,250C>T (C250T), represent changes at -124C>T and -146C>T (where -1 is the base upstream of the A base of the ATG translation start site) respectively in the *TERT* promoter region. These mutations confer increased transcriptional activity on the *TERT* promoter.

There is a distinct lack of *TERTp* mutations in benign thyroid tumours, like follicular adenomas, whereas other drivers like *RAS* mutations are readily detected. This suggests that *TERTp* mutations are malignancy specific and

likely related to later genetic events promoting more invasive aspects of tumorigenesis. *TERTp* mutations are have been shown to be more frequently detected in poorly differentiated and anaplastic thyroid cancers, further adding to evidence linking cancer progression with these variants. Our data supports the link between the *TERTp* c.1-124C>T variant and more aggressive pathology. In our PTC samples, the presence of concurrent *TERTp* variant with *BRAF* p.V600E conferred a relative risk of 3.75 for recurrent disease (p = 0.038, 95% CI 1.07-11.37) compared to *BRAF* p.V600E alone. We also specifically demonstrated a statistically significant association between *TERTp* and the presence of vascular invasion in PTC. As a predictor of vascular invasion in PTC, *TERTp* variant performs well as a highly specific screening test, but alone has poor sensitivity. In multiple logistic regression analysis our model (combining *BRAF* variant positive, any *RAS* variant positive, presence of multiple variants and *TERTp* variant positive) could correctly predict the presence of vascular invasion 70% of the time, with *TERTp* variants playing a key role.

The findings from this study suggest that the *TERT* promoter region should be routinely included in future somatic mutations screening panels. This is important when considering further diagnostic studies such as analysing DNA extracted from FNA needle biopsies within the thyroid nodule diagnostic phase. There are published studies that have demonstrated nucleic acid isolation of adequate quality and concentration suitable for PCR and DNA sequencing, from multiple tumour types including thyroid cancer (34, 35).

3.6 Conclusion

We have shown in our cohort of DTC tumours, *TERTp* c.1-124C>T variant can be used in combination with other thyroid cancer mutational hotspots, to predict the presence of vascular invasion with specificity and sensitivity approaching clinically applicable levels. A targeted sequencing approach is a relatively quick, cheap and sensitive method of identifying these somatic mutations from DNA extracted from degraded FFPE tissue. Using other genomic biomarkers such as microRNAs (investigated further in Chapter 4) and epigenetic changes such as methylation status could further increase the utility of a screening prognostic

test, which could guide clinicians and patients towards the most appropriate primary treatment.

3.7 References

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid. 2016;26(1):1-133.
2. Sulaiman SA, Abu N, Ab-Mutalib NS, Low TY, Jamal R. Signatures of gene expression, DNA methylation and microRNAs of hepatocellular carcinoma with vascular invasion. Future Oncol. 2019;15(22):2603-17.
3. Mannelqvist M, Wik E, Stefansson IM, Akslen LA. An 18-gene signature for vascular invasion is associated with aggressive features and reduced survival in breast cancer. PLoS One. 2014;9(6):e98787.
4. Mannelqvist M, Stefansson IM, Bredholt G, Hellem Bo T, Oyan AM, Jonassen I, et al. Gene expression patterns related to vascular invasion and aggressive features in endometrial cancer. Am J Pathol. 2011;178(2):861-71.
5. Xing M, Liu R, Liu X, Murugan AK, Zhu G, Zeiger MA, et al. BRAF V600E and TERT promoter mutations cooperatively identify the most aggressive papillary thyroid cancer with highest recurrence. J Clin Oncol. 2014;32(25):2718-26.
6. Qasem E, Murugan AK, Al-Hindi H, Xing M, Almohanna M, Alswailem M, et al. TERT promoter mutations in thyroid cancer: a report from a Middle Eastern population. Endocr Relat Cancer. 2015;22(6):901-8.
7. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-D7.
8. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014;159(3):676-90.
9. Lê S JJ, Husson F. FactoMineR: A Package for Multivariate Analysis. Journal of Statistical Software. 2008;25(1):1-18.
10. Vuong HG, Kondo T, Duong UNP, Pham TQ, Oishi N, Mochizuki K, et al. Prognostic impact of vascular invasion in differentiated thyroid carcinoma: a systematic review and meta-analysis. Eur J Endocrinol. 2017;177(2):207-16.
11. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-4.
12. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.
13. Colombo C, Muzza M, Proverbio MC, Tosi D, Soranna D, Pesenti C, et al. Impact of Mutation Density and Heterogeneity on Papillary Thyroid Cancer Clinical Features and Remission Probability. Thyroid. 2019;29(2):237-51.
14. Chai L, Li J, Lv Z. An integrated analysis of cancer genes in thyroid cancer. Oncol Rep. 2016;35(2):962-70.
15. Song YS, Park YJ. Genomic Characterization of Differentiated Thyroid Carcinoma. Endocrinol Metab (Seoul). 2019;34(1):1-10.

16. Yoo SK, Lee S, Kim SJ, Jee HG, Kim BA, Cho H, et al. Comprehensive Analysis of the Transcriptional and Mutational Landscape of Follicular and Papillary Thyroid Cancers. PLoS Genet. 2016;12(8):e1006239.
17. Song YS, Lim JA, Park YJ. Mutation Profile of Well-Differentiated Thyroid Cancer in Asians. Endocrinol Metab (Seoul). 2015;30(3):252-62.
18. Yoo SK, Song YS, Lee EK, Hwang J, Kim HH, Jung G, et al. Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. Nat Commun. 2019;10(1):2764.
19. Kim TH, Park YJ, Lim JA, Ahn HY, Lee EK, Lee YJ, et al. The association of the BRAF(V600E) mutation with prognostic factors and poor clinical outcome in papillary thyroid cancer: a meta-analysis. Cancer. 2012;118(7):1764-73.
20. Tufano RP, Teixeira GV, Bishop J, Carson KA, Xing M. BRAF mutation in papillary thyroid cancer and its value in tailoring initial treatment: a systematic review and meta-analysis. Medicine (Baltimore). 2012;91(5):274-86.
21. Xing M, Alzahrani AS, Carson KA, Viola D, Elisei R, Bendlova B, et al. Association between BRAF V600E mutation and mortality in patients with papillary thyroid cancer. JAMA. 2013;309(14):1493-501.
22. Trovisco V, Couto JP, Cameselle-Teijeiro J, de Castro IV, Fonseca E, Soares P, et al. Acquisition of BRAF gene mutations is not a requirement for nodal metastasis of papillary thyroid carcinoma. Clin Endocrinol (Oxf). 2008;69(4):683-5.
23. Eloy C, Santos J, Soares P, Sobrinho-Simoes M. The preeminence of growth pattern and invasiveness and the limited influence of BRAF and RAS mutations in the occurrence of papillary thyroid carcinoma lymph node metastases. Virchows Arch. 2011;459(3):265-76.
24. Sancisi V, Nicoli D, Ragazzi M, Piana S, Ciarrocchi A. BRAFV600E mutation does not mean distant metastasis in thyroid papillary carcinomas. J Clin Endocrinol Metab. 2012;97(9):E1745-9.
25. Gouveia C, Can NT, Bostrom A, Grenert JP, van Zante A, Orloff LA. Lack of association of BRAF mutation with negative prognostic indicators in papillary thyroid carcinoma: the University of California, San Francisco, experience. JAMA Otolaryngol Head Neck Surg. 2013;139(11):1164-70.
26. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet. 2005;6(8):611-22.
27. Mocellin S, Pooley KA, Nitti D. Telomerase and the search for the end of cancer. Trends Mol Med. 2013;19(2):125-33.
28. Vinagre J, Almeida A, Populo H, Batista R, Lyra J, Pinto V, et al. Frequency of TERT promoter mutations in human cancers. Nat Commun. 2013;4:2185.
29. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013;339(6122):959-61.
30. Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA, Jr., et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proc Natl Acad Sci U S A. 2013;110(15):6021-6.
31. Capezzone M, Marchisotta S, Cantara S, Pacini F. Telomeres and thyroid cancer. Curr Genomics. 2009;10(8):526-33.
32. Maggisano V, Celano M, Lepore SM, Sponziello M, Rosignolo F, Pecce V, et al. Human telomerase reverse transcriptase in papillary thyroid cancer:

gene expression, effects of silencing and regulation by BET inhibitors in thyroid cancer cells. Endocrine. 2019;63(3):545-53.
33. Donati B, Ciarrocchi A. Telomerase and Telomeres Biology in Thyroid Cancer. Int J Mol Sci. 2019;20(12).
34. Kizys MM, Cardoso MG, Lindsey SC, Harada MY, Soares FA, Melo MC, et al. Optimizing nucleic acid extraction from thyroid fine-needle aspiration cells in stained slides, formalin-fixed/paraffin-embedded tissues, and long-term stored blood samples. Arq Bras Endocrinol Metabol. 2012;56(9):618-26.
35. Treece AL, Montgomery ND, Patel NM, Civalier CJ, Dodd LG, Gulley ML, et al. FNA smears as a potential source of DNA for targeted next-generation sequencing of lung adenocarcinomas. Cancer Cytopathol. 2016;124(6):406-14.

Chapter 4 Differential expression of miRNA associated with vascular invasion in thyroid cancer

4.1 Introduction

In the previous chapter 3, we showed that *TERT* promoter (*TERTp*) mutational variants were significantly associated with the presence of vascular invasion in DTC. We proposed a mutational signature that predicted the presence of vascular invasion based on the tumour mutational burden and mutation status of *BRAF*, *RAS* and *TERTp* oncogenes. The ability of DTC tumour cells to breach endothelium and enter the vasculature clearly involves multiple biological processes that cannot be attributed to a somatic mutation alone. Sufficient DNA quality for downstream sequencing, is easily achievable from FFPE tissue blocks that have been stored appropriately, even after decades. Whilst there are inherent issues with processing artefact and sequencing error, these can be largely mitigated with a robust bioinformatics pipeline. RNA, however, is extremely sensitive to degradation and must be stored at -80°C prior to downstream processing. There is evidence however that smaller RNA molecules, like miRNAs, are relatively well preserved and resistant to ribonuclease digestion (1). There are a few hypotheses to explain this observation including their small size relative to larger RNA molecules assisting in evasion, the formation of secondary protective structures resisting degradation and protection within vesicular envelopes (2, 3). They are the subject of intense research within the circulating biomarkers field, due to their function in cell-to-cell communication whereby they are packaged into extracellular exosomes and released into circulating plasma (4-7). These observations promote miRNAs as a viable biomarker of disease in DTC. Studying post-transcriptional modifiers like miRNAs, could provide additional characterisation of DTC tumours and highlight which underlying pathways are being activated or suppressed to promote high risk histopathological features such as vascular invasion.

Dysregulation of miRNAs has been shown to promote oncogenesis. A number of review articles over the past decade have summarised the published literature base of miRNA differential expression in thyroid cancer (listed in Table 4-1 below).

Title	Year	Types of thyroid cancer	Reference
MicroRNA expression profiles in thyroid tumors	2009	All	Nikiforova et al (8)
Deregulation of microRNA expression in thyroid neoplasias	2014	All	Pallante et al (9)
MicroRNA expression profiles in the management of papillary thyroid cancer	2014	PTC	Lee et al (10)
Clinical pathological impacts of microRNAs in papillary thyroid carcinoma: A crucial review	2015	PTC	Chruscik et al (11)
Candidate microRNAs as biomarkers of thyroid carcinoma: a systematic review, meta-analysis, and experimental validation	2016	All	Hu et al (12)
The role of microRNAs in different types of thyroid carcinoma: a comprehensive analysis to find new miRNA supplementary therapies	2017	All	Pishkari et al (13)
MicroRNAs as biomarkers in thyroid carcinoma	2017	All	Celano et al (14)

Table 4-1. Summary of review articles of miRNA expression in thyroid cancer

Despite all these studies and the benefit of the TCGA RNASeq and miRNASeq DTC datasets being publicly available, there are few studies that have shown conclusive evidence describing any association between miRNA dysregulation and vascular invasion in DTC.

In this chapter, we used the same DTC cohort collected from the GEM-TC study (described in chapter 2.1) to study differential expression of miRNAs in order to observe if dysregulation of certain miRNAs could be associated with the presence of vascular invasion.

4.2 Hypothesis

There are dysregulated miRNAs that are significantly associated with the presence of vascular invasion in differentiated thyroid cancer.

4.3 Aims

1. To extract and quantify total RNA from FFPE thyroid tumour samples
2. To compare reference genes in published thyroid studies to observe the most stable endogenous reference
3. To observe differential expression of key miRNAs of interest (MOIs)
4. To correlate observed differential expression of MOIs with the presence or absence of vascular invasion

4.4 Results

4.4.1 Total RNA extraction and quantification

All patients with available FFPE tumour tissue that were recruited from the "Gene Mutation Analysis as a biomarker in Thyroid Cancer (GEM-TC)" study underwent total RNA extraction (described in section 2.2 of Materials and Methods Chapter 2). Average total RNA concentration extracted was 125.3ng/µL (SD = ±129.6, range 6.6 – 734.5). An average A260/280 ratio of 1.84 indicated highly pure RNA, however the extraction methodology clearly resulted in residual guanidine salts, as indicated by the on average low A260/230 ratios (see Figure 4-1). Due to the residual guanidine salts, we performed two pilot experiments to assess the usability of the RNA, opting to compare traditional qPCR analysis with a more recent Nanostring technique.

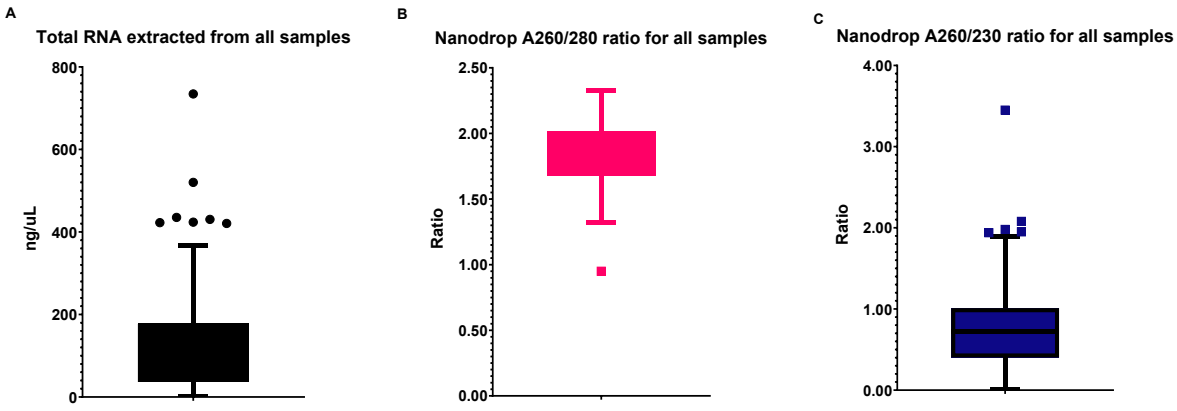


Figure 4-1. RNA extraction quality control summary. A. Boxplot showing total RNA extracted from all samples. **B.** Box plot showing A260/280 ratios for all samples measured using Nanodrop. **C.** Box plot showing A260/230 ratios for all samples measured using Nanodrop.

4.4.2 Pilot experiments

The RNA purity had potential implications for downstream analysis due to PCR inhibiting effects of retained guanidine salts. For this reason, we opted to perform pilot experiments to assess extracted RNA usability. Two pilot experiments were carried out as described in section 2.2 of Materials and Methods Chapter 2. Both pilot experiments aimed to demonstrate successful miRNA retention in the extraction process but also to provide initial differential expression data in a subset of samples prior to testing the full cohort.

The first pilot experiment utilised qPCR analysis of 11 miRNA species that have previously been shown to be dysregulated in DTC, with an additional 11 control candidates. These were chosen on the basis of literature review of miRNA dysregulation studies in DTC (see Table 4-2 below). The control genes were analysed for expression stability across samples, so that single or multiple controls could be used as expressional reference normalisers, to calculate fold change expression.

The second pilot experiment used a commercially validated panel of 830 mature human miRNAs using the Nanostring workflow (described in section 2.2.4 of Chapter 2). The rationale for using Nanostring over other techniques was based on institutional availability, financial cost and experimental time considerations. The nCounter® Human v3 miRNA Expression Assay (Nanostring Technologies, Seattle, WA, USA) allowed a broad panel of assayed miRNAs, providing an element of discovery to the second pilot experiment. This aimed to complement the first pilot experiment and identify further significantly dysregulated miRNAs of interest that may not have been highlighted previously in the literature.

4.4.2.1 qPCR pilot experiment

A panel of 22 assays were used for this pilot experiment, shown in Table 4-2. This pilot experiment used a subset of the total DTC cohort, with 24 samples comprising 8 PTC, 8 FTC and 8 normal controls. The PTC and FTC groups were divided into 4 samples with and without vascular invasion shown below in Table 4-3.

Assay	miRBase accession ID	Literature base	Dysregulation
hsa-miR-let7f-5p	MIMAT0000067	(9, 15)	Downregulated in PTC
hsa-miR-222-5p	MIMAT0004569	(14, 16-19)	Upregulated in PTC in a large number of studies and associated with aggressive pathology.
hsa-miR-221-5p	MIMAT0004568	(14, 16-19)	Upregulated in PTC in a large number of studies and associated with aggressive pathology.
hsa-miR-183-5p	MIMAT0000261	(20-23)	Upregulated in PTC and FTC
hsa-miR-182-5p	MIMAT0000259	(20-23)	Upregulated in PTC and FTC
hsa-miR-146b-5p	MIMAT0002809	(14, 16-19)	Upregulated in PTC in a large number of studies with large fold change.
hsa-miR-222-3p	MIMAT0000279	(14, 16-19)	Upregulated in PTC in a large number of studies and associated with aggressive pathology.
hsa-miR-221-3p	MIMAT0000278	(14, 16-19)	Upregulated in PTC in a large number of studies and associated with aggressive pathology.
hsa-miR-146b-3p	MIMAT0004766	(14, 16-19)	Upregulated in PTC in a large number of studies with large fold change.
hsa-miR-135b-5p	MIMAT0000758	(12, 14)	Upregulated in PTC
hsa-miR-32-5p	MIMAT0000090	(12, 14)	Upregulated in PTC
hsa-miR-191-5p	MIMAT0000440	(24)	Reference miRNA
hsa-miR-423-5p	MIMAT0004748	(24)	Reference miRNA
hsa-miR-425-5p	MIMAT0003393	(24)	Reference miRNA
hsa-miR-93-5p	MIMAT0000093	(24)	Reference miRNA
hsa-miR-151a-3p	MIMAT0004697	(25)	Reference miRNA
hsa-miR-197-3p	MIMAT0022691	(25)	Reference miRNA
hsa-miR-214-3p	MIMAT0004564	(25)	Reference miRNA
hsa-miR-103-3p	MIMAT0000101	(25)	Putative endogenous control
U6 snRNA	N/A	-	Commonly used miRNA surrogate normaliser
5S rRNA	N/A	-	Commonly used miRNA surrogate normaliser
UniSp6 spike in	N/A	-	Technical control

Table 4-2. Panel of miRNA assays used for the qPCR pilot experiment

Histological subtype	Staging	Tumour size (mm)	VI	ETE	LN invasion	Multifocality
cPTC	pT3N0Mx	26	No	No	No	No
cPTC	pT3N0Mx	23	No	No	No	No
cPTC	pT3N0Mx	41	No	No	No	No
cPTC	pT3N0Mx	80	No	No	No	No
cPTC	pT2N0M0	39	Yes	No	No	Yes
cPTC	pT1bN0Mx	11	Yes	No	No	No
cPTC	pT3N0M0	60	Yes	No	No	No
cPTC	pT3bN1aM0	52	Yes	Yes	Yes	No
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
MI FTC	pT2pN0M0	25	No	No	No	No
MI FTC	pT1bN0M0	20	No	No	No	No
MI FTC	pT2N0Mx	37	No	No	No	No
WI FTC	pT2N0M0	25	No	No	No	No
MI FTC	pT3N0Mx	12	Yes	No	No	No
WI FTC	pT3N0M1	70	Yes	No	No	No
WI FTC	pT3N0M0	45	Yes	No	No	No
MI FTC	pT2N0Mx	26	Yes	No	No	No
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-

Table 4-3. Histopathological phenotype of the samples included in the qPCR pilot experiment

Finding the most stable control miRNA

Normfinder (26) analysis (as described in section 2.2.3 of Chapter 2) showed that from the 24 samples analysed, the two miRNAs with the most stable expression from all control genes were hsa-miR-423-5p and hsa-miR-93-5p. Expression of these miRNAs were significantly more stable than U6 snRNA and 5S rRNA, which are widely used as endogenous reference RNAs in published miRNA studies (see Figure 4-2).

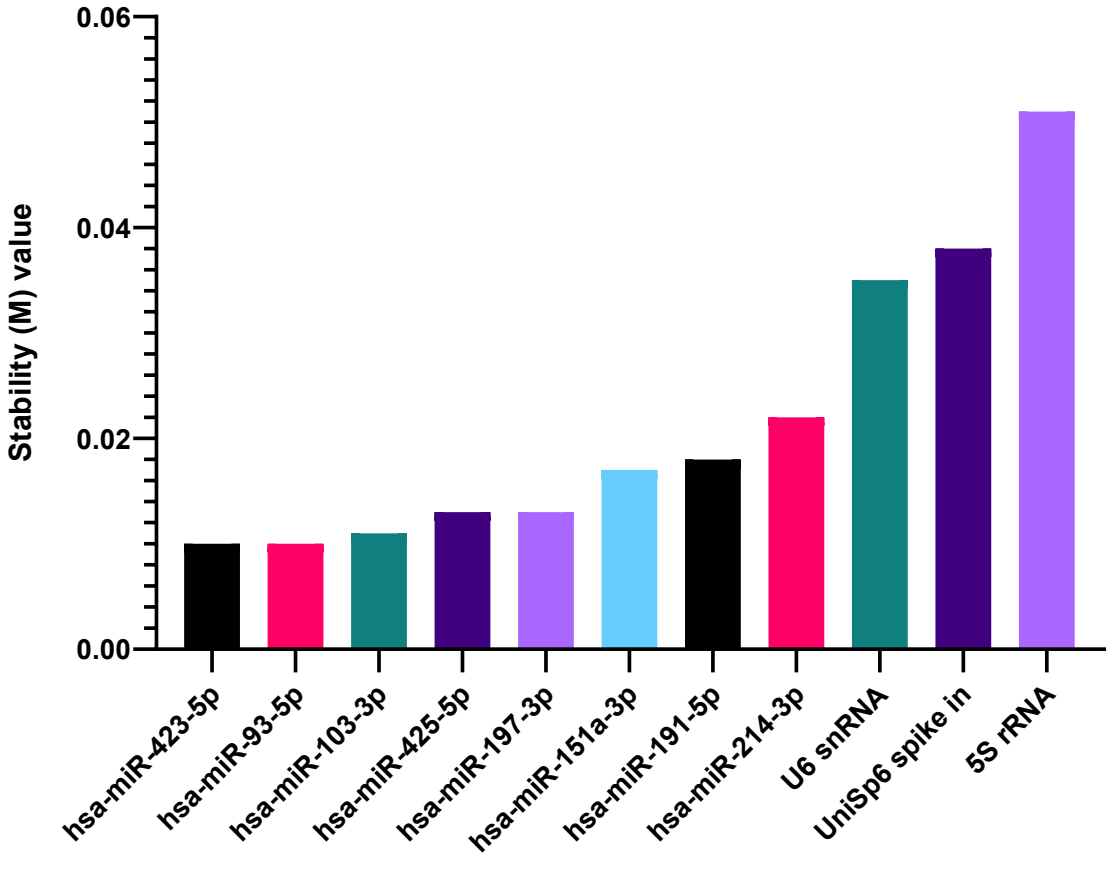


Figure 4-2. Bar plot showing stability values of the reference miRNA species used in the qPCR pilot experiment. The smaller the value the less expressional variance. Calculated using NormFinder.

Fold change expression

By averaging the expression of hsa-miR-423-5p and hsa-miR-93-5p as the comparative endogenous reference, fold change expression could be calculated for all miRNAs of interest in all samples. This is summarised in Table 4-4 below.

All DTC Samples				Sub-analysis PTC only				Sub-analysis FTC only			
miRNA	**Norm** Mean fold change	**nVI** Mean fold change	**VI** Mean fold change	miRNA	**Norm** Mean fold change	**nVI** Mean fold change	**VI** Mean fold change	miRNA	**Norm** Mean fold change	**nVI** Mean fold change	**VI** Mean fold change
hsa-miR-let7f-5p	1.1	2.0	1.4	hsa-miR-let7f-5p	1.0	2.7	1.6	hsa-miR-let7f-5p	1.1	1.2	1.1
hsa-miR-222-5p	1.5	3.7	5.3	hsa-miR-222-5p	1.5	5.2	7.9	hsa-miR-222-5p	1.4	2.1	2.7
hsa-miR-221-5p	1.1	4.0	7.7	hsa-miR-221-5p	0.7	5.8	10.9	hsa-miR-221-5p	1.5	2.2	4.6
hsa-miR-183-5p	1.2	2.9	3.1	hsa-miR-183-5p	1.0	1.9	2.6	hsa-miR-183-5p	1.3	3.9	3.6
hsa-miR-182-5p	1.1	3.1	3.9	hsa-miR-182-5p	1.1	2.5	4.1	hsa-miR-182-5p	1.1	3.7	3.7
hsa-miR-146b-5p	1.1	164.6	107.7	hsa-miR-146b-5p	1.1	325.8	213.2	hsa-miR-146b-5p	1.2	3.3	2.3
hsa-miR-222-3p	1.1	14.2	16.2	hsa-miR-222-3p	0.9	23.9	25.2	hsa-miR-222-3p	1.3	4.5	7.2
hsa-miR-221-3p	1.1	3.5	11.5	hsa-miR-221-3p	1.0	4.9	15.6	hsa-miR-221-3p	1.2	2.1	7.5
hsa-miR-146b-3p	1.3	73.5	60.2	hsa-miR-146b-3p	1.6	145.0	118.9	hsa-miR-146b-3p	1.0	2.1	1.6
hsa-miR-135b-5p	1.0	1.6	1.2	hsa-miR-135b-5p	1.0	1.9	1.6	hsa-miR-135b-5p	1.1	1.2	0.9
hsa-miR-32-5p	1.0	0.6	1.1	hsa-miR-32-5p	1.0	0.4	1.0	hsa-miR-32-5p	1.1	0.8	1.1

Table 4-4. Average fold change in each miRNA species according to vascular invasion status compared to normal control. Values calculated using the ΔΔCT method. Highlighted are significant fold change differences in each corresponding miRNA expression between nVI and VI categories; green is significant up-regulation and red is significant down-regulation

Cancer versus normal

In this cohort of 16 tumours and 8 normal controls, we observed dysregulation of all miRNAs of interest included in this limited qPCR assay panel, when comparing normal thyroid control samples with DTC samples. High fold change dysregulation was most apparent in PTC compared to FTC tumours. The highest fold change difference was seen in hsa-miR-146b-5p with mean expressional upregulation of 270x in PTC compared to normal controls. Both iso-miRs of hsa-miR-146b and hsa-miR-221 demonstrated large fold change upregulation in cancer versus normal (see Table 4-4). Hsa-miR-32-5p was the only significantly down-regulated species with 0.4x fold change (or 2.5x less expression) seen in PTC tumours without vascular invasion compared to normal.

Vascular invasion versus no vascular invasion

Differential expression between DTC with and without vascular invasion were observed most prominently in hsa-miR-146b and hsa-miR-221. Hsa-miR-146b-5p upregulation was 112.6x less in PTC with VI compared to nVI. Similarly, but to a lesser extent, hsa-miR-146b-3p upregulation was 26.1x less in PTC with VI compared to nVI (see Figure 4-3). These large differential expression differences in hsa-miR-146b iso-miRs were not seen in the FTC cohort. Both iso-miRs of hsa-miR-221 demonstrated over double the fold change

upregulation in samples with VI compared to nVI and this was independent of histological subtype (see Figure 4-3). This proof of principle pilot experiment demonstrated that miRNAs could be successfully isolated from our samples, and that dysregulation of hsa-miRs 146b, 221 and 32 were potentially good candidates for differentiating the presence of vascular invasion in DTC.

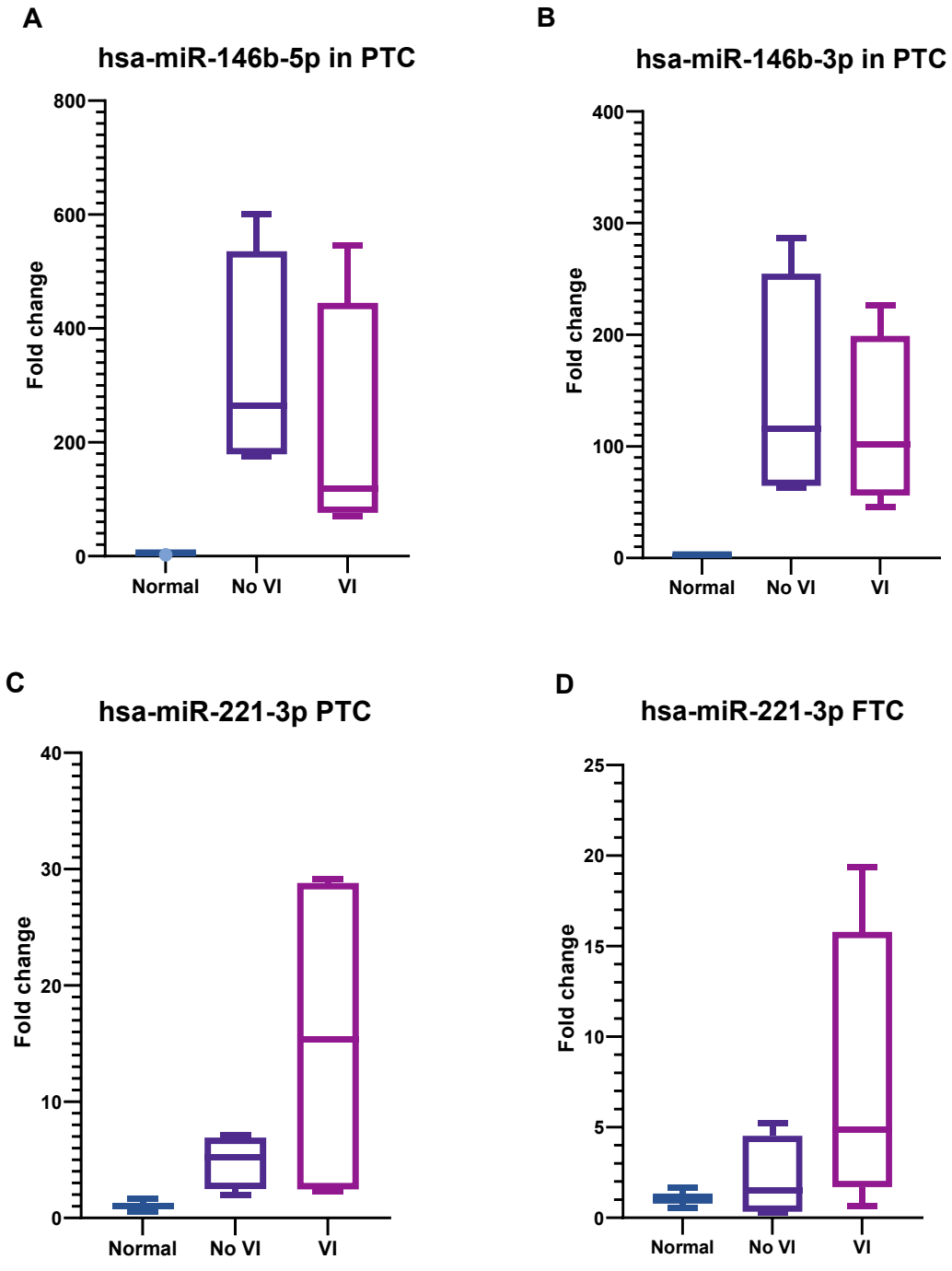


Figure 4-3. Boxplots showing averaged differential expression fold change according to vascular invasion status compared to normal control in key miRNA species from the qPCR pilot experiment. A. Boxplots of hsa-miR-146b-5p expression in PTC tumours according to vascular invasion status. **B.** Boxplots of hsa-miR-146b-3p expression in PTC tumours according to vascular invasion status. **C.** Boxplots of hsa-miR-221-3p expression in PTC tumours according to vascular invasion status. **D.** Boxplots of hsa-miR-221-3p expression in FTC tumours according to vascular invasion status.

4.4.2.2 Nanostring pilot experiment

The expression profile of 830 miRNAs in 24 samples comprising 8 PTC, 8 FTC and 8 normal controls were analysed using Nanostring (as described in section 2.2.4 in chapter 2) to identify differentially expressed miRNAs between VI and nVI samples. Due to purity constraints, not all of the same samples used in the qPCR pilot experiment were able to be used with the Nanostring technique. Similarly to the first pilot experiment, the PTC and FTC groups were divided into 4 samples with and without vascular invasion shown below in Table 4-5.

Histological subtype	Staging	Tumour size (mm)	VI	ETE	LN invasion	Multifocality
cPTC	pT2N0M0	32	No	No	No	No
cPTC	pT3N0Mx	23	No	No	No	No
cPTC	pT3N0Mx	26	No	No	No	No
cPTC	pT3N0Mx	80	No	No	No	No
cPTC	pT2N0M0	39	Yes	No	No	No
cPTC	pT1bN0Mx	11	Yes	No	No	Yes
cPTC	pT3N0M0	60	Yes	No	No	No
cPTC	pT3bN1aM0	52	Yes	Yes	Yes	No
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
MI FTC	pT2N0M0	25	No	No	No	No
MI FTC	pT2N0M0	37	No	No	No	No
MI FTC	pT1bN0M0	20	No	No	No	No
WI FTC	pT2N0Mx	25	No	No	No	No
MI FTC	pT3N0Mx	12	Yes	No	No	No
WI FTC	pT2N0Mx	38	Yes	No	No	Yes
WI FTC	pT3N0Mx	45	Yes	No	No	No
MI FTC	pT2N0Mx	26	Yes	No	No	No
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-
Normal ctrl	-	-	-	-	-	-

Table 4-5. Histopathological phenotype of the samples included in the Nanostring pilot experiment

Quality control of Nanostring run

The in-built positive and negative controls demonstrated good quality runs at the end of the Nanostring methodology and all quality metrics were passed (this included assay probe ligation, field of view, binding density and sample content). Visually this is represented in Figure 4-4, which shows consistency of control assay count percentage between all samples.

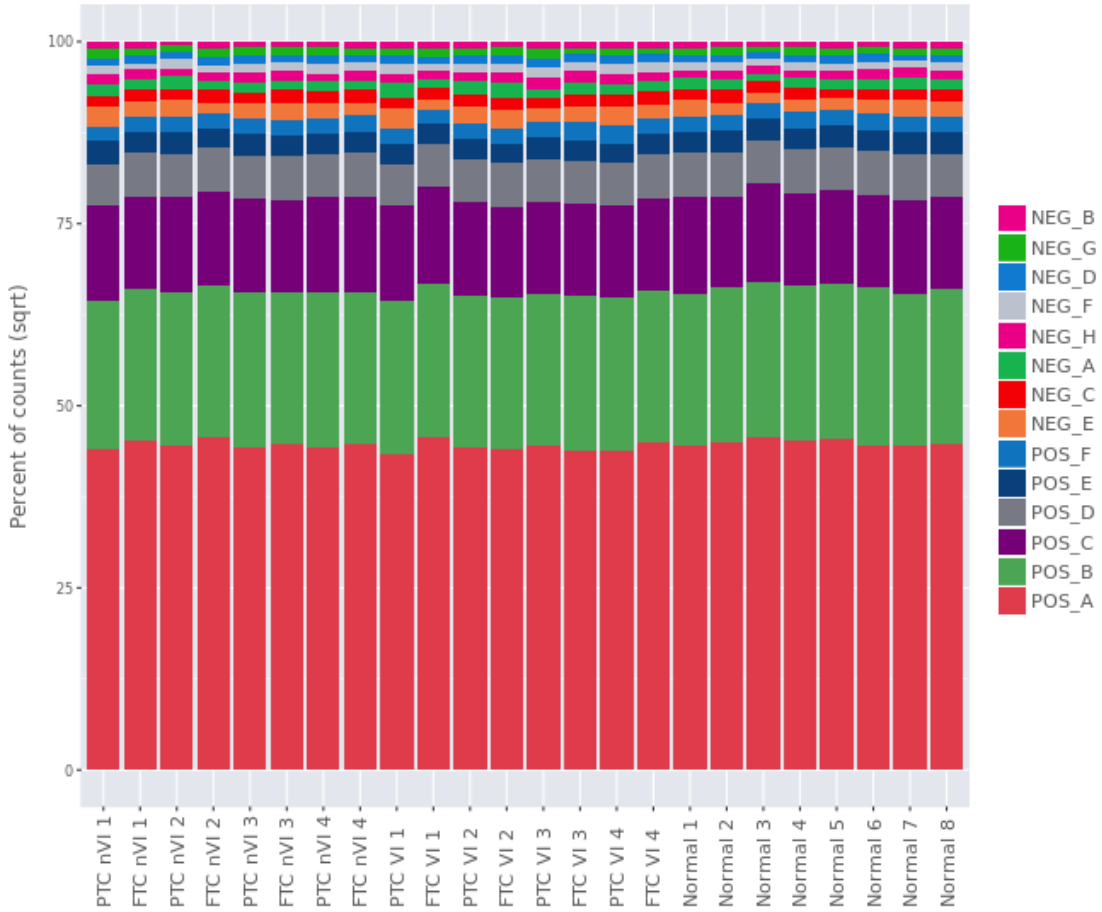


Figure 4-4. Stacked bar chart showing combined counts of spike-ins and negative controls. Combined counts are calculated using the square root of the expression level. Each column sums to 100%. Consistency in expression of controls suggests good quality runs and equity between samples.

Endogenous references and normalisation

The endogenous reference species (or Housekeeping genes) included as standard in the Nanostring panel (*RPL19*, *RPLP0*, *GAPDH*, *ACTB* and *B2M*) demonstrated variable expression stability as shown in Figure 4-5. The most stable of these was RPLP0 with 41.1% covariance. Due to the variation in expression amongst the samples, many of these pre-included endogenous references were not suitable for normalisation purposes. Instead, all detected miRNAs in the panel were ranked according to percentage covariance.

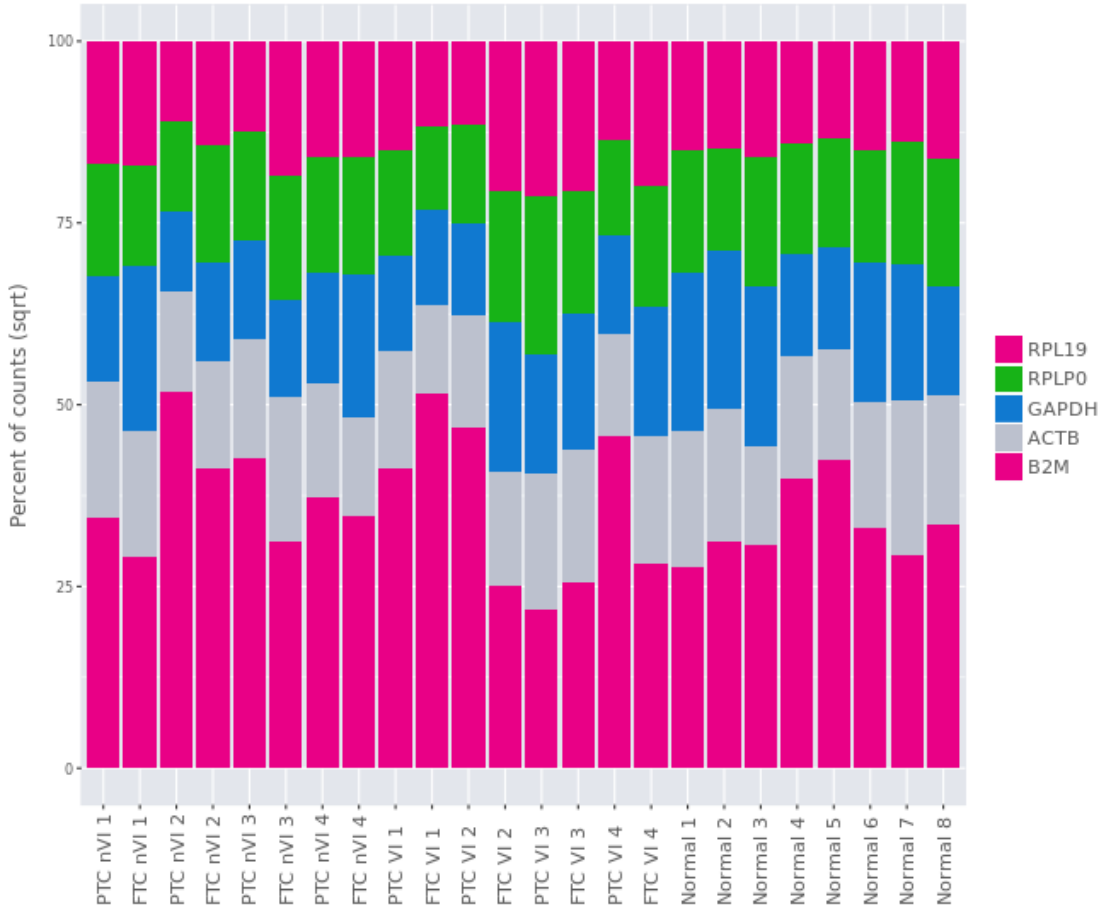


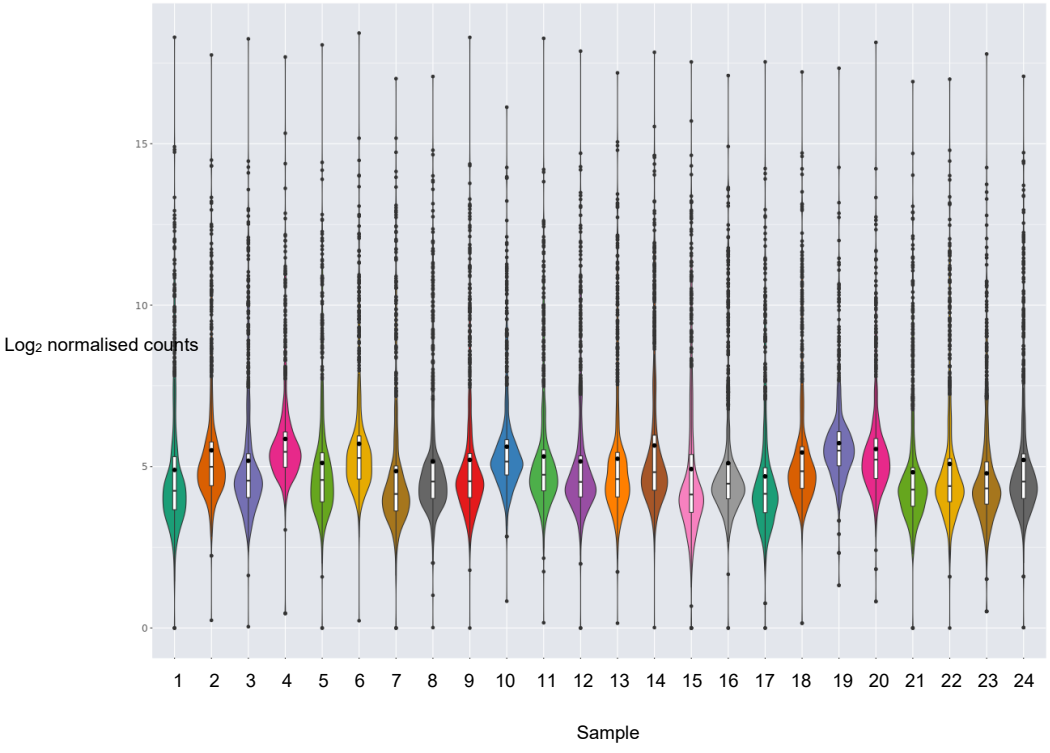
Figure 4-5. Stacked bar chart showing combined counts of endogenous references. Combined counts are calculated using the square root of the expression level. Each column sums to 100%. Useful endogenous references should demonstrate minimal variability in expression between samples.

Hsa-miR-423-5p was the most stable endogenous miRNA with an average molecule count of 380.04 and percentage covariance of 32.2%. This was the same miRNA species that was found to be most stable in the preceding qPCR pilot experiment. All miRNAs with less than 50% covariance were used for global mean normalisation to calculate differential expression. These are listed in Table 4-6 below.

Probe name	Class name	Avg Count	%CV
hsa-miR-423-5p	Endogenous	380.04	32.215
hsa-miR-361-3p	Endogenous	211.79	33.639
hsa-miR-1260a	Endogenous	966.42	36.214
hsa-miR-1180-3p	Endogenous	208.29	37.013
hsa-miR-664a-3p	Endogenous	177.46	37.72
hsa-let-7b-5p	Endogenous	6864.17	38.154
hsa-miR-378e	Endogenous	188.54	38.532
hsa-let-7d-5p	Endogenous	1874.17	39.212
hsa-let-7c-5p	Endogenous	2437.00	40.786
RPLP0	Housekeeping	590.08	41.133
hsa-miR-191-5p	Endogenous	1150.79	42.425
hsa-let-7a-5p	Endogenous	20304.96	42.923
hsa-miR-4454+hsa-miR-7975	Endogenous	200228.67	43.686
hsa-miR-331-3p	Endogenous	213.88	45.228
RPL19	Housekeeping	571.21	45.253
hsa-miR-100-5p	Endogenous	1556.08	45.59
hsa-miR-99a-5p	Endogenous	1790.79	45.815
hsa-miR-539-5p	Endogenous	175.96	45.835
ACTB	Housekeeping	670.83	46.805
hsa-miR-342-3p	Endogenous	883.46	47.699
hsa-let-7e-5p	Endogenous	603.96	47.828
hsa-miR-15b-5p	Endogenous	1006.13	47.862
hsa-miR-302d-3p	Endogenous	330.63	47.868

Table 4-6. miRNA species with less than 50% covariance used for global mean normalisation.

The effect of this normalisation is shown in the violin plot in Figure 4-6.



1	PTC nVI
2	FTC nVI
3	PTC nVI
4	FTC nVI
5	PTC nVI
6	FTC nVI
7	PTC nVI
8	FTC nVI
9	PTC VI
10	FTC VI
11	PTC VI
12	FTC VI
13	PTC VI
14	FTC VI
15	PTC VI
16	FTC VI
17	Normal control
18	Normal control
19	Normal control
20	Normal control
21	Normal control
22	Normal control
23	Normal control
24	Normal control

Figure 4-6. Violin plot of normalised expression levels of all samples included in the Nanostring pilot experiment. The objective of normalisation is to ensure equitable comparison between samples prior to calculating differential expression. This controls for minor differences in total RNA input, sample quality and technical variability.

Differential expression of cancer versus normal

14 miRNAs were significantly upregulated (>1.5x) and 1 significantly downregulated (<-1.5x) when comparing all cancer samples with normal controls (listed in Table 4-7). The most differential between cancer and normal was the expression of hsa-miR-34a-5p as shown in Figure 4-7 below.
A summary of miRNA dysregulation is demonstrated visually in the volcano plot and heatmap in Figure 4-8.

UPREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-146b-5p	13.1184	0.046951
hsa-miR-222-3p	9.66889	0.006151
hsa-miR-221-3p	6.36795	0.015179
hsa-miR-21-5p	5.24045	0.006151
hsa-miR-34a-5p	4.64707	1.63E-05
hsa-miR-221-5p	3.97292	0.006939
hsa-miR-1246	2.51349	0.046951
hsa-miR-181b-5p+hsa-miR-181d-5p	2.46289	0.025401
hsa-miR-15a-5p	2.43506	0.046951
hsa-miR-375	2.39089	0.037391
hsa-miR-181a-5p	2.36958	0.025401
hsa-miR-182-5p	2.10335	0.046951
hsa-miR-4284	1.97907	0.046951
hsa-miR-421	1.65466	0.046951
DOWNREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-199b-5p	-2.93787	0.025401

Table 4-7. Differential expression of miRNAs in cancer versus normal

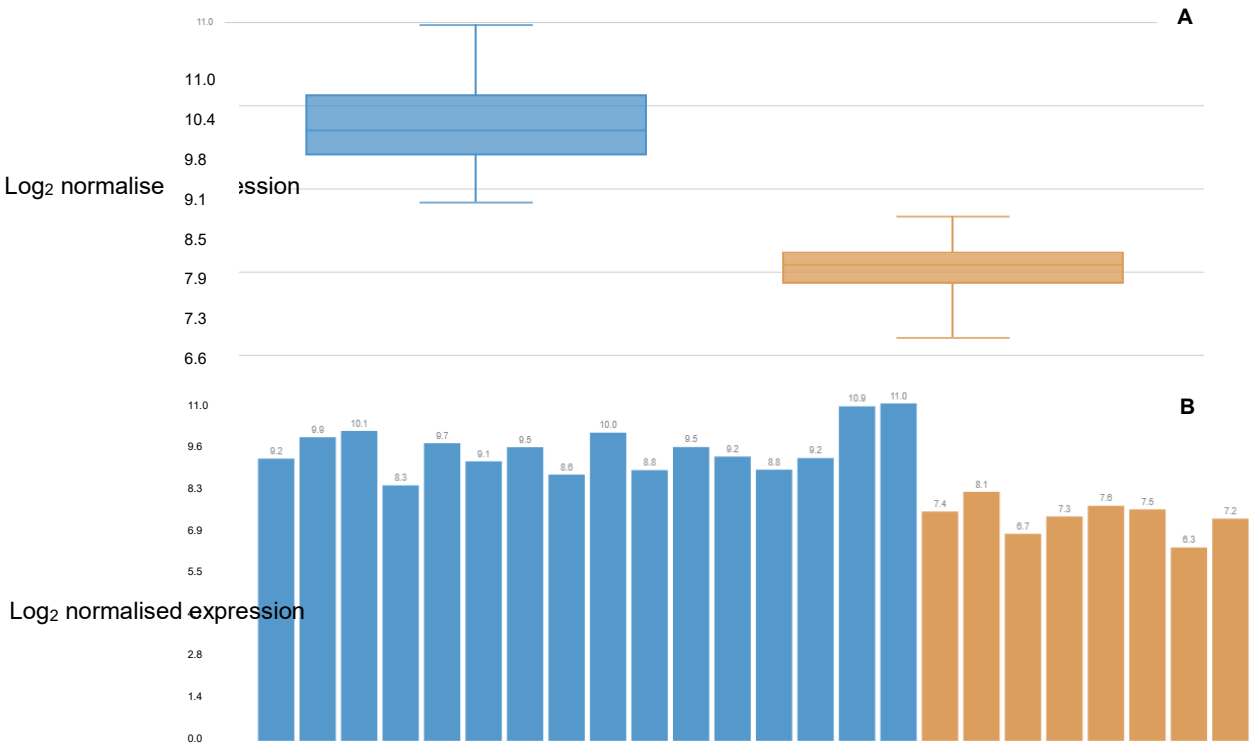


Figure 4-7. Differential expression of hsa-miR-34a-5p. Clear separation between cancer samples versus normal.
A – Boxplot showing median and range of Log_2 normalised expression between cancer (blue) and normal (orange).
B – Barplot showing individual Log_2 normalised expression in each sample of cancer (blue) and normal (orange).

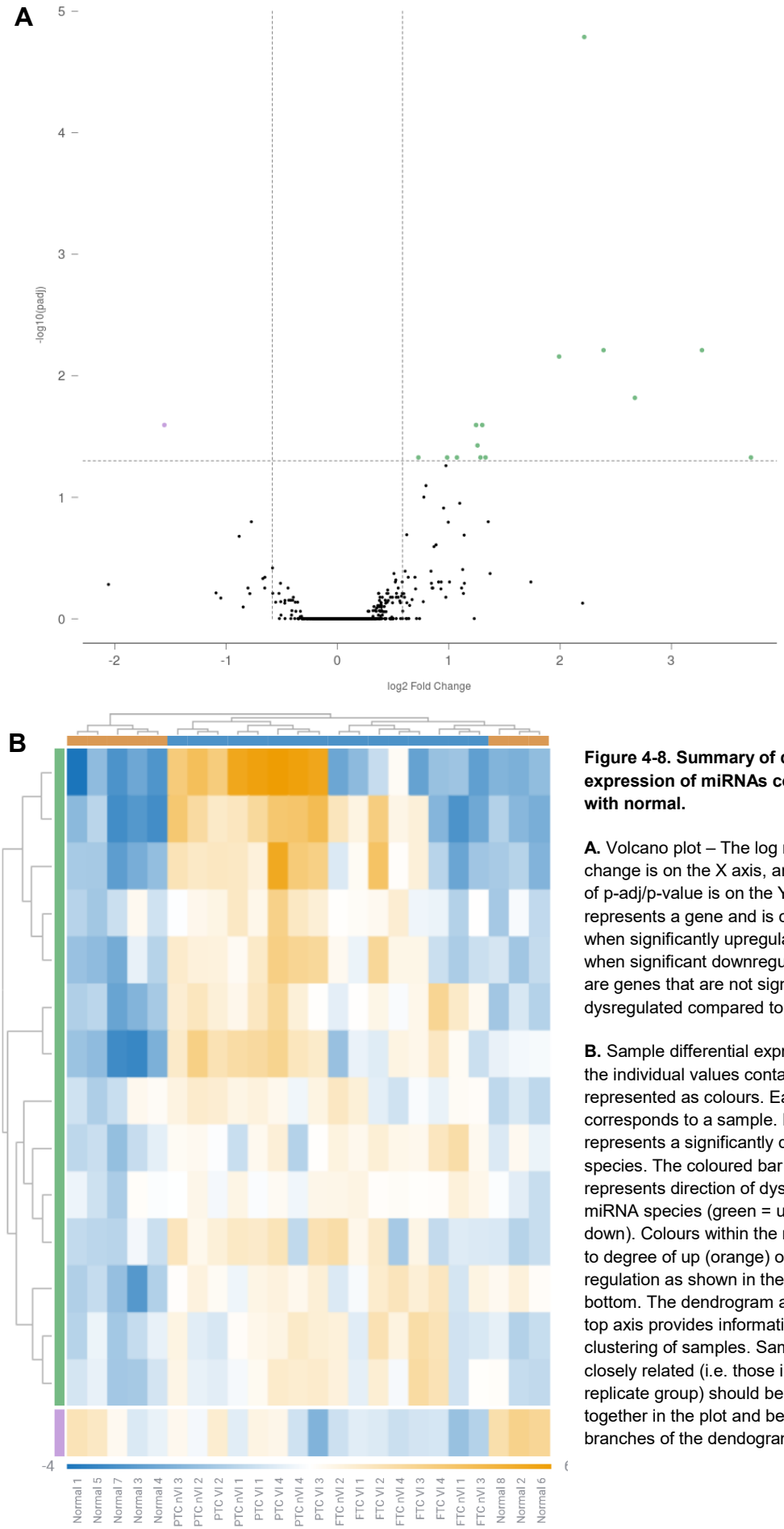


Figure 4-8. Summary of differential expression of miRNAs comparing cancer with normal.

A. Volcano plot – The log ratio of the fold change is on the X axis, and the negative log of p-adj/p-value is on the Y axis. Each dot represents a gene and is coloured green when significantly upregulated and light purple when significant downregulated. Black dots are genes that are not significantly dysregulated compared to normal control.

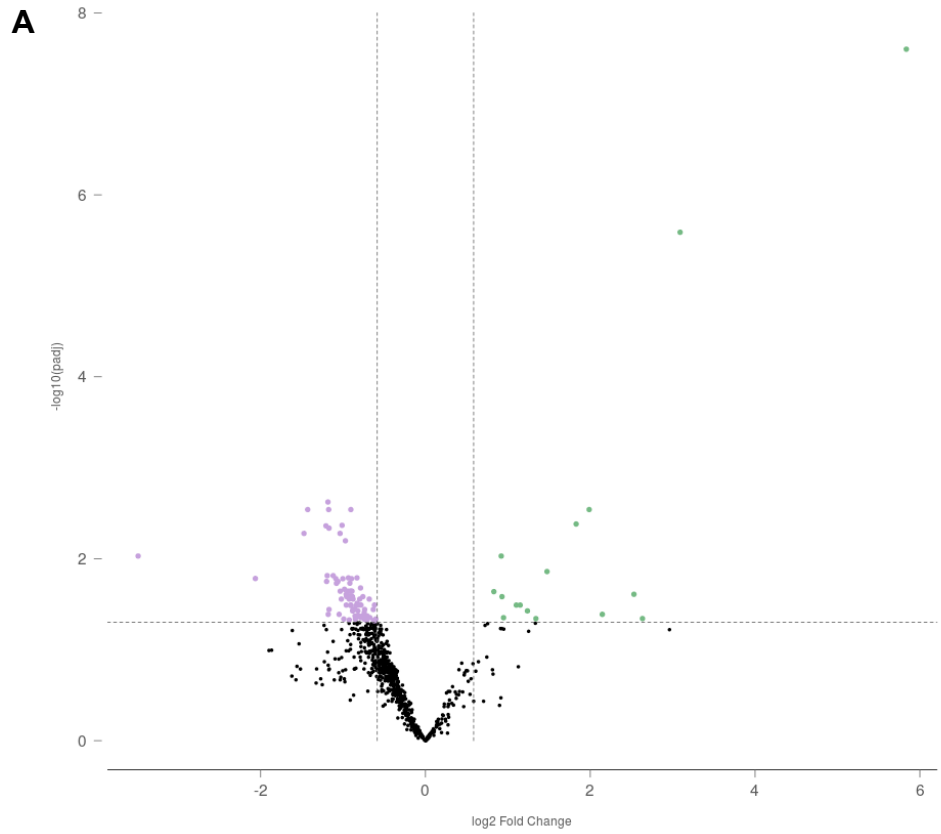
B. Sample differential expression heatmap - the individual values contained in a matrix are represented as colours. Each column corresponds to a sample. Each row represents a significantly dysregulated miRNA species. The coloured bar on the left represents direction of dysregulation of each miRNA species (green = up, light purple = down). Colours within the matrix correspond to degree of up (orange) or down (blue) regulation as shown in the scale at the bottom. The dendrogram annotation on the top axis provides information regarding the clustering of samples. Samples that are closely related (i.e. those in the same replicate group) should be strongly correlated together in the plot and be the closest branches of the dendogram.

Differential expression of PTC versus FTC

16 miRNAs were significantly upregulated (>1.5x) and 76 significantly downregulated (<-1.5x) when comparing PTC with FTC samples (listed in Table 4-8). This is demonstrated visually in the volcano plot and heatmap in Figure 4-9.

UPREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-146b-5p	57.0909	2.51E-08
hsa-miR-31-5p	8.51585	2.58E-06
hsa-miR-494-3p	6.20448	0.045462
hsa-miR-222-3p	5.77664	0.024597
hsa-miR-221-3p	4.42697	0.04087
hsa-miR-146b-3p	3.96316	0.00288
hsa-miR-21-5p	3.55207	0.004152
hsa-miR-630	2.7808	0.013831
hsa-miR-146a-5p	2.53153	0.045544
hsa-miR-575	2.35796	0.037447
hsa-miR-199a-5p	2.21825	0.032294
hsa-miR-223-3p	2.14761	0.032294
hsa-miR-551b-3p	1.92923	0.044459
hsa-miR-320e	1.90387	0.026114
hsa-miR-24-3p	1.89169	0.009334
hsa-miR-302d-3p	1.77766	0.022989
DOWNREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-7-5p	-11.2078	0.009334
hsa-miR-204-5p	-4.18153	0.016501
hsa-miR-148b-3p	-2.77822	0.005268
hsa-miR-183-5p	-2.69241	0.00288
hsa-miR-30a-3p	-2.3086	0.004351
hsa-miR-190a-5p	-2.29641	0.017769
hsa-miR-552-3p	-2.28501	0.015338
hsa-miR-510-5p	-2.27034	0.002379
hsa-miR-568	-2.26566	0.04087
hsa-miR-30e-3p	-2.2568	0.00288
hsa-miR-371b-5p	-2.25016	0.004613
hsa-miR-345-5p	-2.24988	0.036138
hsa-miR-590-3p	-2.1724	0.015338
hsa-miR-891b	-2.12494	0.016578
hsa-miR-564	-2.11412	0.018625
hsa-miR-95-3p	-2.08716	0.017769
hsa-miR-664b-5p	-2.06662	0.04087
hsa-miR-655-3p	-2.05065	0.005268
hsa-miR-760	-2.04811	0.022739
hsa-miR-362-3p	-2.02861	0.027775
hsa-miR-2278	-2.01636	0.004292
hsa-miR-548h-5p	-2.00286	0.016578

Table 4-8. Differential expression of miRNAs in PTC versus FTC tumours.



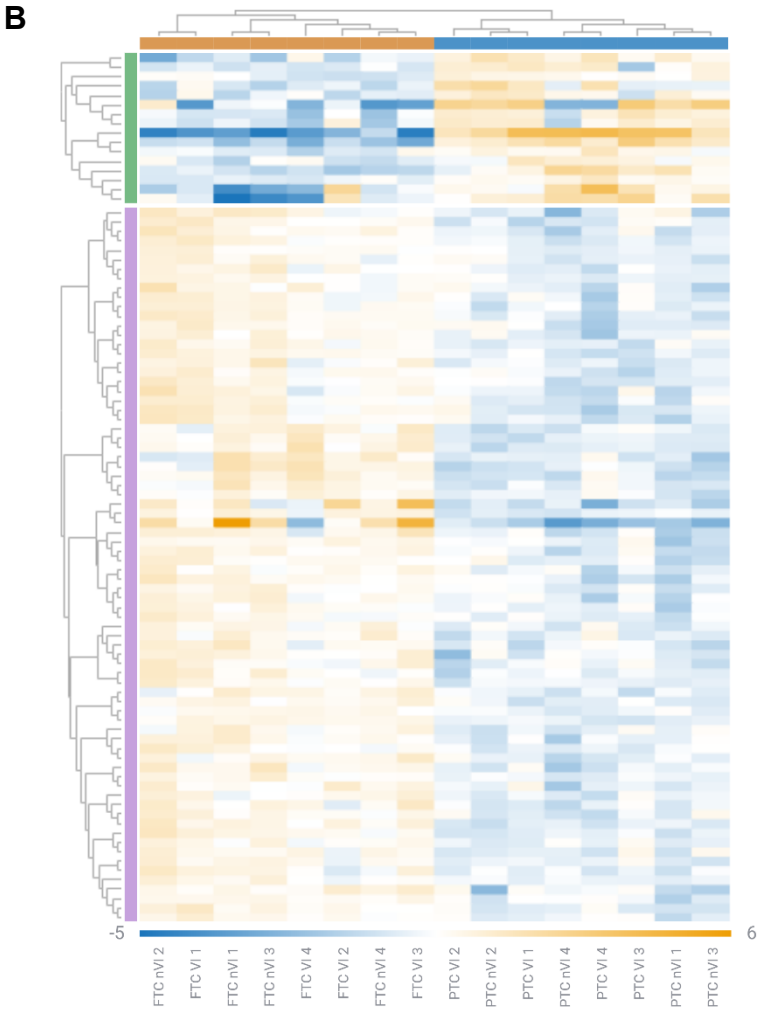


Figure 4-9. Summary of differential expression of miRNAs comparing PTC with FTC tumours.

A. Volcano plot – The log ratio of the fold change is on the X axis, and the negative log of p-adj/p-value is on the Y axis. Each dot represents a gene and is coloured green when significantly upregulated and light purple when significant downregulated. Black dots are genes that are not significantly dysregulated compared to normal control.

B. Sample differential expression heatmap - the individual values contained in a matrix are represented as colours. Each column corresponds to a sample. Each row represents a significantly dysregulated miRNA species. The coloured bar on the left represents direction of dysregulation of each miRNA species (green = up, light purple = down). Colours within the matrix correspond to degree of up (orange) or down (blue) regulation as shown in the scale at the bottom. The dendrogram annotation on the top axis provides information regarding the clustering of samples. Samples that are closely related (i.e. those in the same replicate group) should be strongly correlated together in the plot and be the closest branches of the dendogram.

Differential expression of VI versus nVI in all samples

8 miRNAs were upregulated and 4 downregulated when comparing VI samples with nVI samples (listed in Table 4-9). These were not statistically significant according to adjusted p-values.

UPREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-200b-3p	4.23379	0.996879
hsa-miR-200a-3p	3.82216	0.996879
hsa-miR-205-5p	2.45316	0.996879
hsa-miR-429	2.14795	0.996879
hsa-miR-1915-3p	2.05318	0.996879
hsa-miR-664b-5p	1.69219	0.996879
hsa-miR-1290	1.65124	0.996879
hsa-miR-25-3p	1.51002	0.996879
DOWNREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-193a-5p+hsa-miR-193b-5p	-1.75969	0.996879
hsa-miR-542-5p	-1.74918	0.996879
hsa-miR-145-5p	-1.73468	0.996879
hsa-miR-570-3p	-1.59901	0.996879

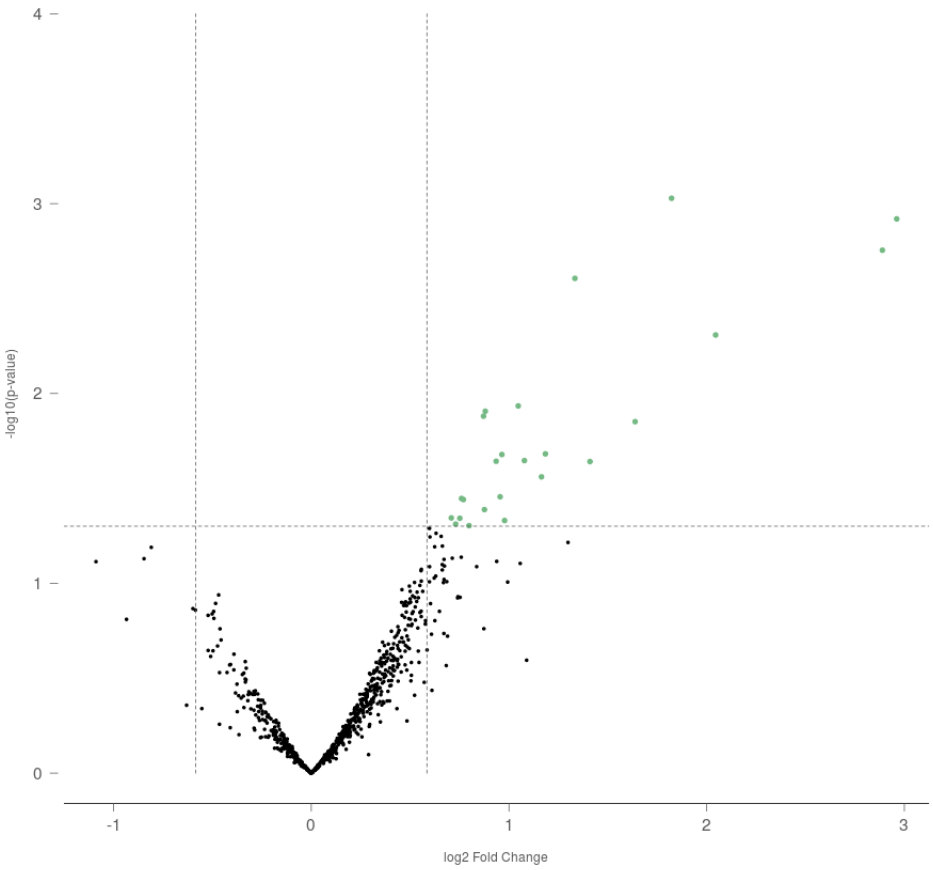
Table 4-9. Differential expression of miRNAs comparing samples with and without vascular invasion (VI).

Differential expression of VI versus nVI in PTC samples

24 miRNAs were upregulated in VI samples within the PTC cohort, with no obvious downregulated species (listed in Table 4-10). These were not statistically significant according to adjusted p-values. There was some visual separation of groups in the volcano plot and heatmap shown in Figure 4-10.

UPREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-200a-3p	7.78744	0.467801
hsa-miR-200b-3p	7.40675	0.467801
hsa-miR-525-3p	4.12756	0.784499
hsa-miR-429	3.53681	0.467801
hsa-miR-450a-5p	3.11266	0.999398
hsa-miR-424-5p	2.65715	0.999398
hsa-miR-664b-5p	2.52033	0.49363
hsa-miR-503-5p	2.27258	0.999398
hsa-miR-135a-5p	2.24131	0.999398
hsa-miR-648	2.11133	0.999398
hsa-miR-548n	2.06587	0.999398
hsa-miR-3605-3p	1.96944	0.999398
hsa-let-7i-5p	1.95103	0.999398
hsa-miR-1290	1.93884	0.999398
hsa-miR-592	1.91208	0.999398
hsa-miR-1285-5p	1.84079	0.999398
hsa-miR-517b-3p	1.83519	0.999398
hsa-miR-663a	1.82912	0.999398
hsa-miR-1268a	1.73875	0.999398
hsa-miR-190a-5p	1.70457	0.999398
hsa-miR-432-5p	1.6933	0.999398
hsa-miR-654-5p	1.68353	0.999398
hsa-miR-590-5p	1.65902	0.999398
hsa-miR-605-5p	1.63391	0.999398

Table 4-10. Differential expression of miRNAs comparing PTC samples with and without vascular invasion (VI).



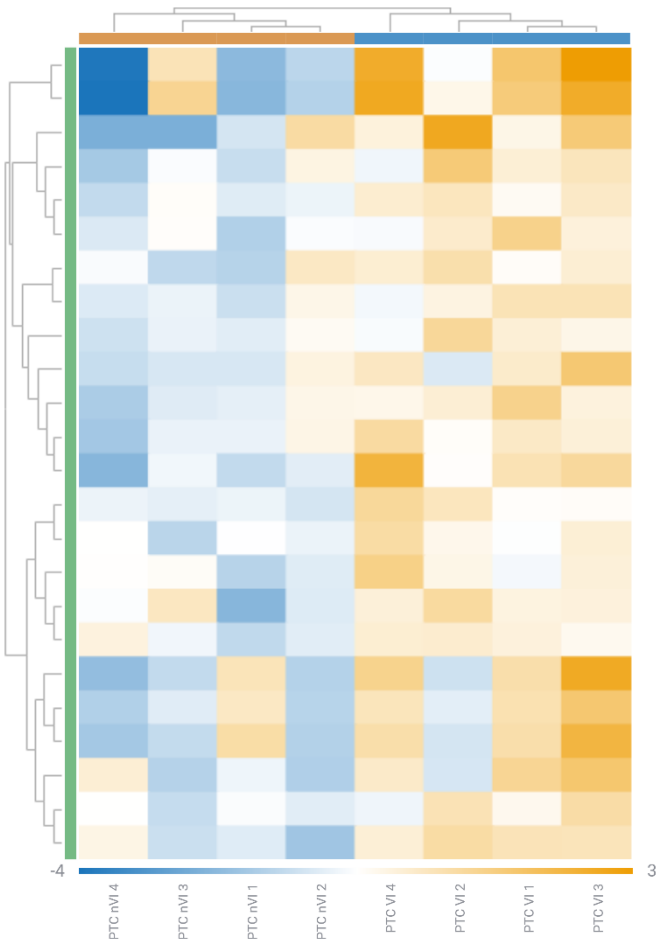


Figure 4-10. Summary of differential expression of miRNAs comparing PTC tumours with and without vascular invasion.

A. Volcano plot – The log ratio of the fold change is on the X axis, and the negative log of p-adj/p-value is on the Y axis. Each dot represents a gene and is coloured green when significantly upregulated and light purple when significant downregulated. Black dots are genes that are not significantly dysregulated compared to normal control.

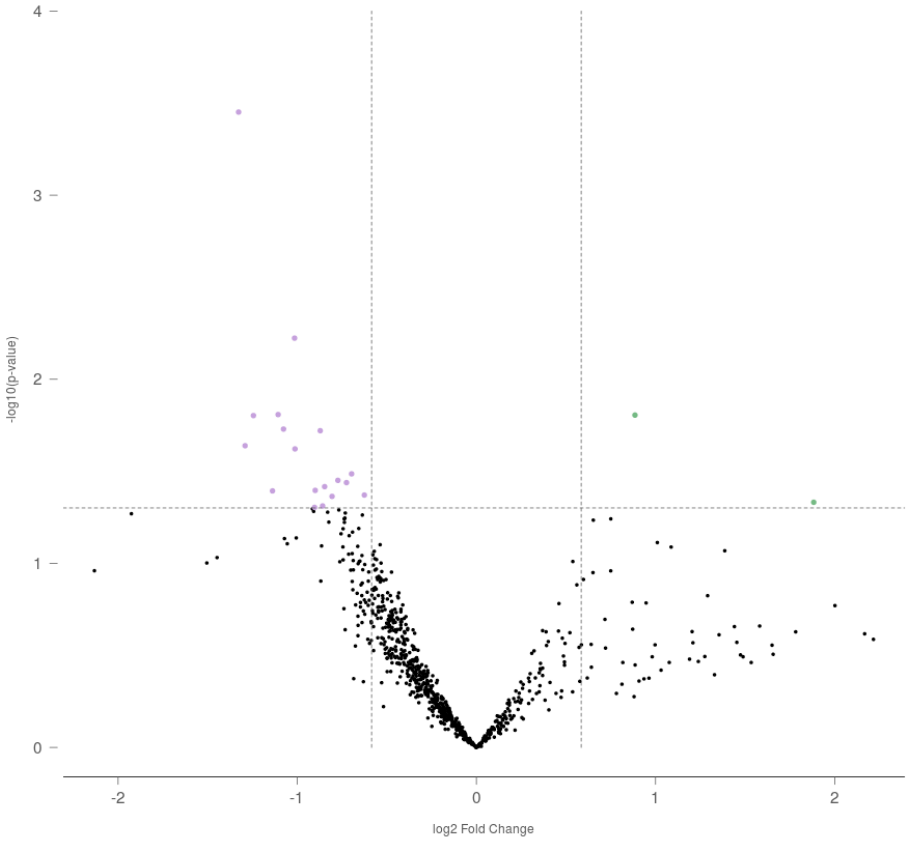
B. Sample differential expression heatmap - the individual values contained in a matrix are represented as colours. Each column corresponds to a sample. Each row represents a significantly dysregulated miRNA species. The coloured bar on the left represents direction of dysregulation of each miRNA species (green = up, light purple = down). Colours within the matrix correspond to degree of up (orange) or down (blue) regulation as shown in the scale at the bottom. The dendrogram annotation on the top axis provides information regarding the clustering of samples. Samples that are closely related (i.e. those in the same replicate group) should be strongly correlated together in the plot and be the closest branches of the dendogram.

Differential expression of VI versus nVI in FTC samples

Only 2 miRNAs were upregulated but a further 18 miRNA were downregulated in VI samples compared to nVI samples within the FTC cohort (listed in Table 4-11). These were not statistically significant according to adjusted p-values. Similar to PTC, there were visual separation of groups in the volcano plot and heatmap shown in Figure 4-11.

UPREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-221-3p	3.68831	0.854182
hsa-miR-146a-5p	1.84708	0.854182
DOWNREGULATED	***Fold Change***	***Adj p-value***
hsa-miR-542-5p	-2.50933	0.282124
hsa-miR-548aa+hsa-miR-548t-3p	-2.44694	0.854182
hsa-miR-365a-3p+hsa-miR-365b-3p	-2.36918	0.854182
hsa-miR-193a-5p+hsa-miR-193b-5p	-2.20109	0.854182
hsa-miR-570-3p	-2.15326	0.854182
hsa-miR-129-5p	-2.1089	0.854182
hsa-miR-1260b	-2.0209	0.854182
hsa-miR-548n	-2.01771	0.854182
hsa-miR-603	-1.87033	0.854182
hsa-miR-548j-5p	-1.86587	0.854182
hsa-miR-3144-5p	-1.82985	0.854182
hsa-miR-27a-3p	-1.8131	0.854182
hsa-miR-182-3p	-1.79903	0.854182
hsa-miR-152-5p	-1.74778	0.854182
hsa-miR-548q	-1.70899	0.854182
hsa-miR-1224-3p	-1.65299	0.854182
hsa-miR-331-3p	-1.62083	0.854182
hsa-miR-548d-5p	-1.54247	0.854182

Table 4-11. Differential expression of miRNAs comparing FTC samples with and without vascular invasion (VI).



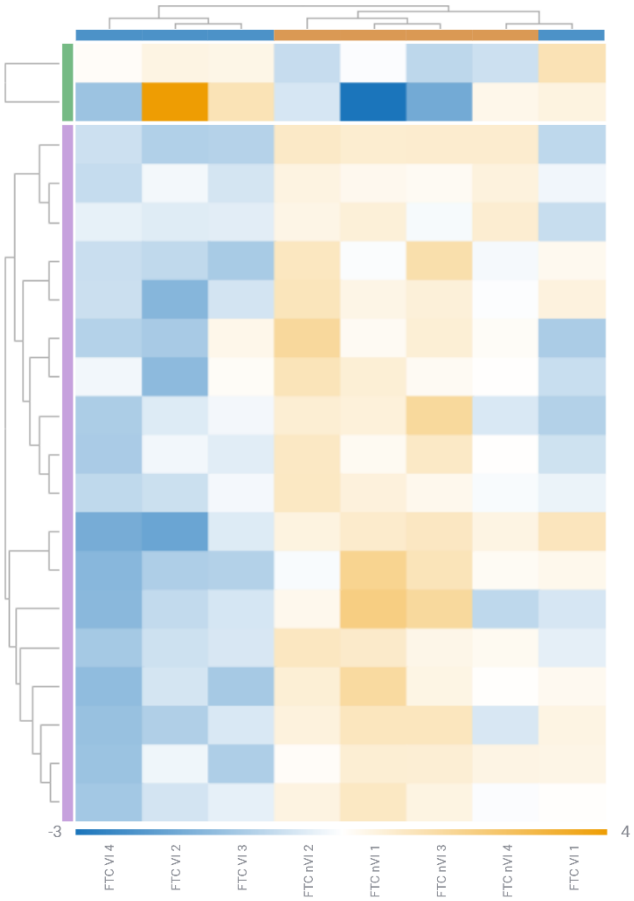


Figure 4-11. Summary of differential expression of miRNAs comparing FTC tumours with and without vascular invasion.

A. Volcano plot – The log ratio of the fold change is on the X axis, and the negative log of p-adj/p-value is on the Y axis. Each dot represents a gene and is coloured green when significantly upregulated and light purple when significant downregulated. Black dots are genes that are not significantly dysregulated compared to normal control.

B. Sample differential expression heatmap - the individual values contained in a matrix are represented as colours. Each column corresponds to a sample. Each row represents a significantly dysregulated miRNA species. The coloured bar on the left represents direction of dysregulation of each miRNA species (green = up, light purple = down). Colours within the matrix correspond to degree of up (orange) or down (blue) regulation as shown in the scale at the bottom. The dendrogram annotation on the top axis provides information regarding the clustering of samples. Samples that are closely related (i.e. those in the same replicate group) should be strongly correlated together in the plot and be the closest branches of the dendogram.

4.4.3 Differential expression using qPCR

The results from the pilot experiments demonstrated that our methodology produced assessable miRNA from both normal thyroid and DTC FFPE tissue blocks and that both qPCR and Nanostring could be performed successfully. The Nanostring pilot also confirmed the findings from the qPCR pilot, specifically regarding statistically significant upregulation of Hsa-miR-146b in cancer versus normal. Disappointingly, neither pilot clearly demonstrated a statistically significant association between vascular invasion and any particular or combination of dysregulated miRNAs. However, there remained the possibility that the low sample numbers used in the pilot experiments resulted in our analysis being underpowered to detect a difference. There was evidence to show large fold change in a number of miRNAs, raising the possibility that a statistically significant difference could be seen if an increased sample size was used. We therefore proceeded with a definitive differential expression experiment using a larger cohort of DTC tumours. 126 patient samples were included in this experiment, divided into 65 PTC, 44 FTC and 17 normal thyroid samples.

The design of the panel of miRNAs of interest (MOIs) was informed by the results of both pilot experiments and listed in Table 4-12.

ASSAY	Function
Hsa-miR-200a-3p	Condition
Hsa-miR-200b-3p	Condition
Hsa-miR-429	Condition
Hsa-miR-542-5p	Condition
Hsa-miR-1260b	Condition
Hsa-miR-204-5p	Condition
Hsa-miR-146a-5p	Condition
Hsa-miR-146b-5p	Condition
Hsa-miR-221-3p	Condition
Hsa-miR-525-3p	Condition
Hsa-miR-423-5p	Reference
Hsa-miR-93-5p	Reference
Hsa-miR-214-3p	Reference
RNU48	Reference

Table 4-12. Panel of miRNAs of interest (MOIs) for main qPCR differential expression experiment

4.4.3.1 Technical results

PCR replicates were generally of excellent quality with identical signal curves. Any obvious deviations were excluded from analysis and Ct values were averaged amongst remaining concordant curves.

The hsa-miR-525-3p assay consistently failed in all samples and plates. This suggested a prevalence of less than 10 molecules of this miRNA in the original sample prior to cDNA generation (low abundance). Results from this miRNA were therefore excluded from downstream analysis.

The included UniSp6 assay spike-in technical control demonstrated very good and consistent reverse-transcription efficiency. We observed distinct Ct number groupings, correlating with reverse transcription batching. The UniSp3 assay technical control was included as an inter-plate calibrator. This also showed consistent and tight Ct grouping, with only a few outliers due to expected variation from factors such as manufacturing anomalies (see Figure 4-12).

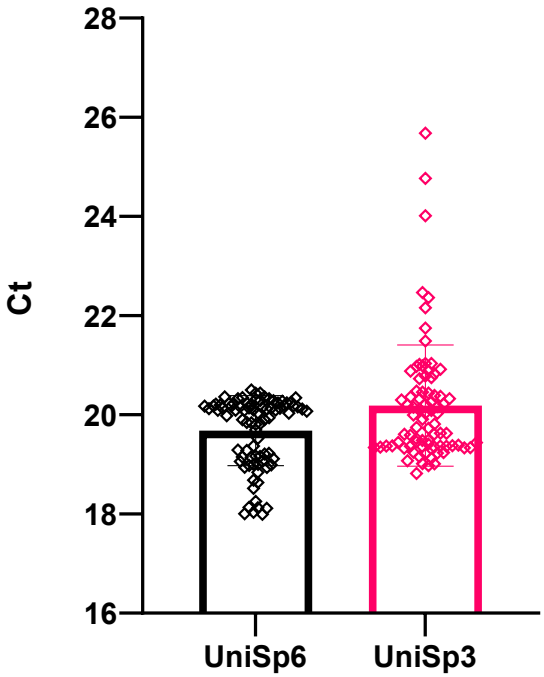


Figure 4-12. Raw Ct values of the technical controls in all samples. Bar graph shows mean Ct with standard deviation bars.

4.4.3.2 Stability of endogenous reference miRNAs

The endogenous references were selected on the basis of their expression stability from both pilot experiments. A summary of the Ct values for each endogenous reference included in the panel is shown in Table 4-13. Plotting the Ct values from these endogenous references, in Figure 4-13, showed visually that the most stable miRNA across all samples was hsa-miR-93-5p.

To calculate the fold change expression in each sample, the geometric mean of all endogenous reference gene expressions was used in the delta-delta-Ct calculation method.

	hsa-miR-423-5p	hsa-miR-93-5p	hsa-miR-214-3p	SNORD48
Number of values	84	84	84	84
Minimum	27.37	25.04	26.27	23.65
25% Percentile	30.08	27.22	28.06	26.02
Median	31.2	27.98	29.02	27.26
75% Percentile	32.82	28.91	30.49	29.06
Maximum	37.05	33.93	36.18	34.98
Range	9.681	8.891	9.907	11.33
Mean	31.44	28.21	29.41	27.64
Std. Deviation	1.917	1.542	1.916	2.383
Std. Error of Mean	0.2092	0.1682	0.2091	0.26
Lower 95% CI of mean	31.02	27.88	28.99	27.13
Upper 95% CI of mean	31.85	28.55	29.83	28.16
Coefficient of variation	6.10%	**5.47%**	6.52%	8.62%

Table 4-13. Summary of the Ct values for each endogenous reference. The most stable reference was hsa-miR-93-5p highlighted in green.

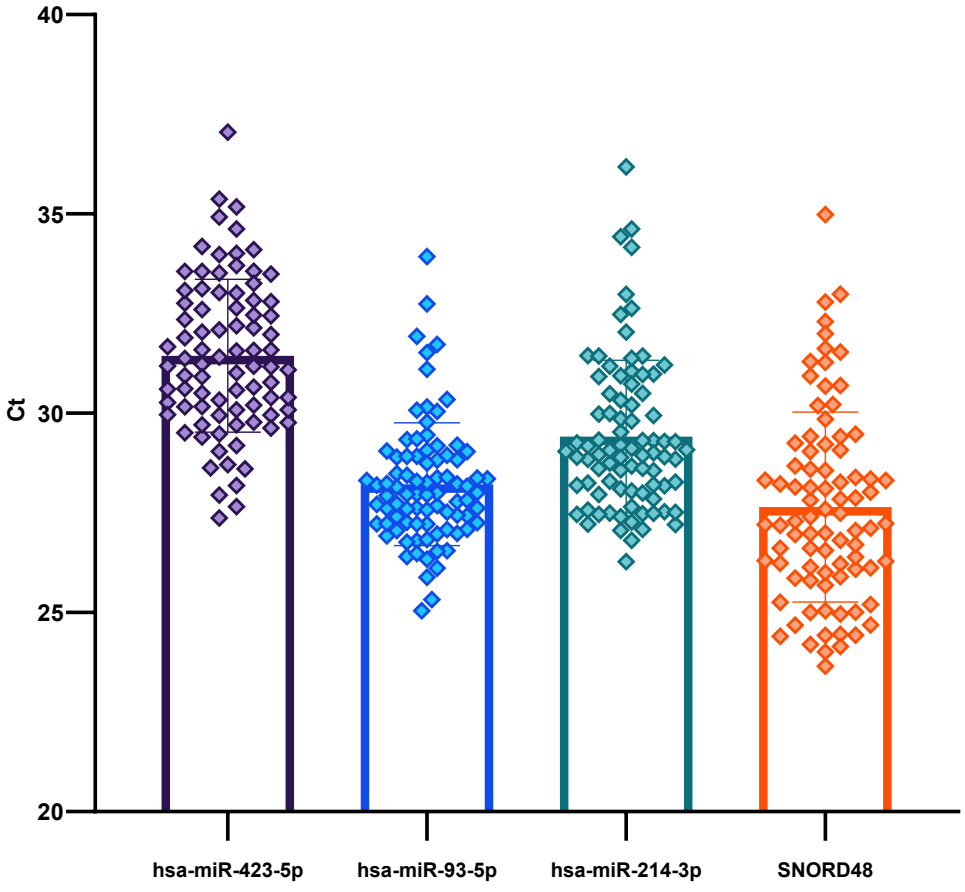


Figure 4-13. Bar chart plots of raw Ct values of the reference gene species in all samples. Bar graph shows mean Ct with standard deviation bars.

4.4.3.3 Upregulated expression of hsa-miR-146a-5p is associated with vascular invasion in DTC

A summary of the results of differential expression of all MOIs in the panel is shown in Table 4-14.

Our results showed that upregulation of hsa-miR-146a-5p was significantly associated with the presence of vascular invasion in DTC (see Figure 4-14). The association was strongest in PTC, with an averaged 2.35x increased expression in samples with vascular invasion ($p = 0.0081$).

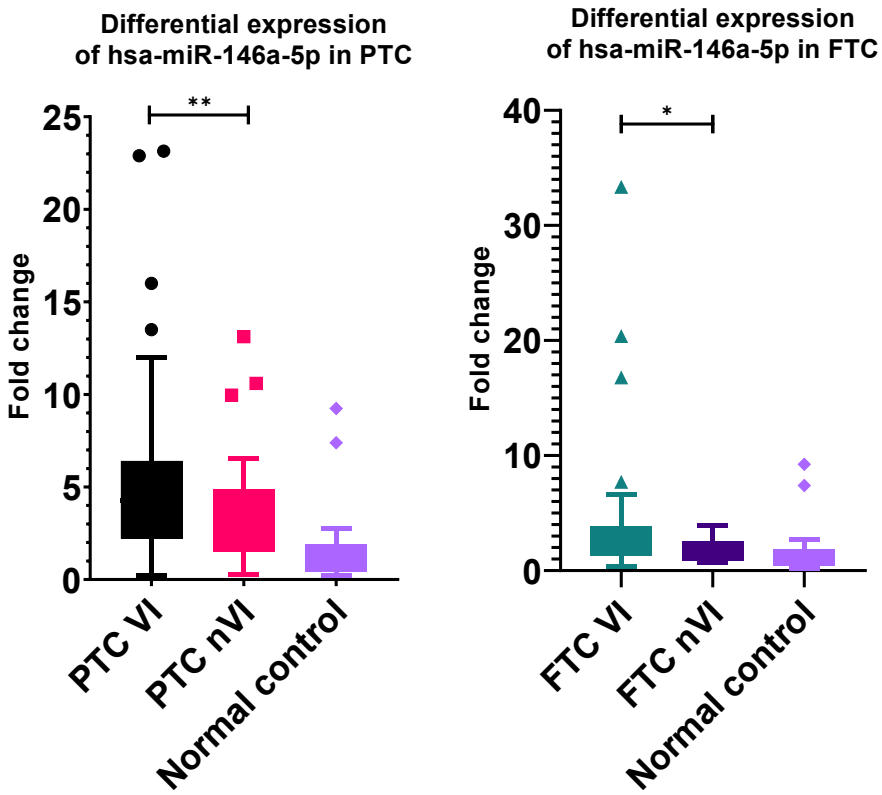


Figure 4-14. Box plots summarising the fold change difference in has-miR-146a-5p between samples with and without vascular invasion compared to normal control. * $p \leq 0.05$, ** $p \leq 0.01$

ASSAY	PTC VI	PTC nVI	FTC VI	FTC nVI	Control
Hsa-miR-200a-3p					
Mean fold change	4.54	4.22	6.85	10.9	2.04
SD	4.75	3.99	8.63	12.8	2.94
Lower 95% CI of mean	2.91	2.74	3.74	1.76	0.471
Upper 95% CI of mean	6.17	5.71	9.96	20.0	3.60
Hsa-miR-200b-3p					
Mean fold change	4.27	3.14	4.44	5.55	1.51
SD	5.34	2.84	6.04	6.24	1.85
Lower 95% CI of mean	2.44	2.08	2.26	1.36	0.527
Upper 95% CI of mean	6.10	4.20	6.61	9.75	2.50
Hsa-miR-429					
Mean fold change	7.16	6.32	14.6	16.2	1.76
SD	9.29	6.54	20.6	22.3	1.79
Lower 95% CI of mean	3.97	3.88	7.05	0.244	0.801
Upper 95% CI of mean	10.3	8.77	22.2	32.1	2.71
Hsa-miR-542-5p					
Mean fold change	2.34	1.94	4.63	3.88	1.18
SD	3.19	1.54	4.09	4.09	0.759
Lower 95% CI of mean	1.24	1.36	3.15	1.13	0.773
Upper 95% CI of mean	3.44	2.51	6.10	6.62	1.58
Hsa-miR-1260b					
Mean fold change	1.70	2.45	3.80	3.59	1.23
SD	2.32	2.09	3.44	3.57	0.872
Lower 95% CI of mean	0.902	1.67	2.56	1.19	0.765
Upper 95% CI of mean	2.50	3.23	5.04	5.99	1.69
Hsa-miR-204-5p					
Mean fold change	1.37	1.19	8.25	12.8	1.87
SD	1.78	1.31	11.2	25.6	2.96
Lower 95% CI of mean	0.757	0.704	4.21	-4.47	0.296
Upper 95% CI of mean	1.98	1.68	12.3	30.0	3.45
Hsa-miR-146a-5p					
Mean fold change	**5.79**	**3.30**	**4.60**	**1.90**	1.86
SD	5.63	3.10	6.78	1.18	2.63
Lower 95% CI of mean	3.85	2.14	2.16	1.11	0.460
Upper 95% CI of mean	7.72	4.46	7.05	2.69	3.26
Hsa-miR-146b-5p					
Mean fold change	229	213	14.0	16.9	2.48
SD	210	226	22.0	23.7	4.12
Lower 95% CI of mean	157	129	6.09	0.973	0.283
Upper 95% CI of mean	301	297	22.0	32.9	4.67
Hsa-miR-221-3p					
Mean fold change	26.8	28.3	42.5	35.9	1.37
SD	18.9	28.6	52.8	45.3	1.07
Lower 95% CI of mean	20.3	17.7	23.5	5.42	0.793
Upper 95% CI of mean	33.3	39.0	61.6	66.3	1.94

Table 4-14. Summary of the results of differential expression of all MOIs in the panel. The highlighted green values are statistically significant differences associated with vascular invasion.

4.4.3.4 Dysregulation of miRNAs is associated with high risk histopathological features

We defined fold change >5 or <0.5 to be considered up and down regulated respectively. A fold change of 0.5 designates 2 fold downregulation. A fold change of 0.1 is 10 fold downregulation. Based on calculated fold change for each miRNA in our panel, we annotated each sample with the binary variables of upregulation yes/no and downregulation yes/no.

By once again dichotomising the data according to VI status, we found that 26.9% (18/67) of samples with VI had upregulation of hsa-miR-204-5p compared to 4.9% (2/41) of samples with nVI and this was a statistically significant difference (Fisher's exact test, p = 0.044). Similarly we found that 32.8% (22/67) of samples with VI had upregulation of hsa-miR-146a-5p compared to 12.2% (5/41) of samples with nVI (Fisher's exact test, p = 0.02) (see Figure 4-15).

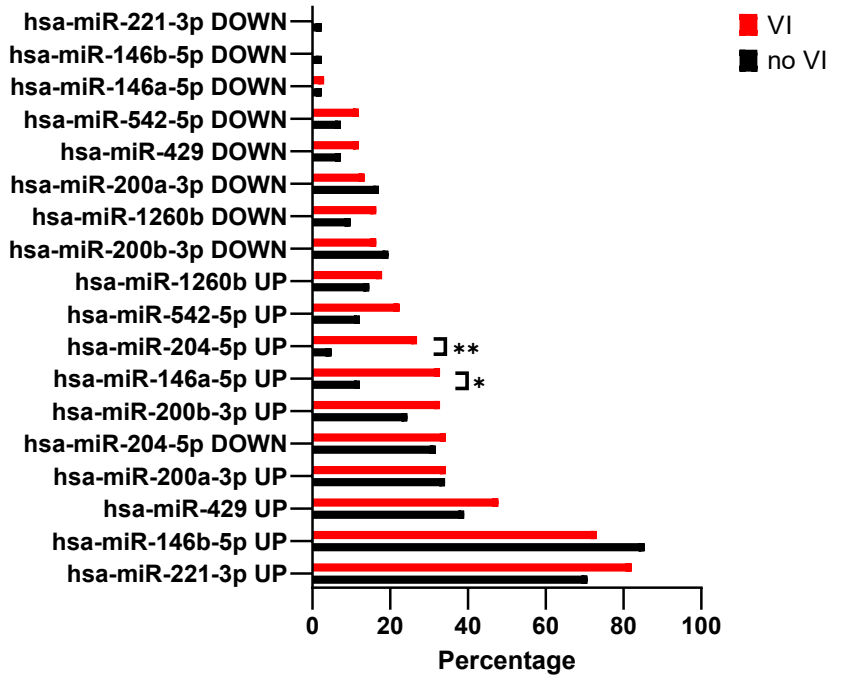


Figure 4-15. Bar plot showing proportion of DTC samples with up or down dysregulation of miRNAs of interest dichotomised by vascular invasion status * p ≤ 0.05, ** p ≤ 0.01

We then dichotomised the data according to ETE status and found that 39.5% (15/38) of samples with ETE had upregulation of hsa-miR-146a-5p compared to 17.1% (12/70) of samples with no ETE and this was a statistically significant difference (Fisher's exact test, p = 0.019). Additionally, we also observed that 50.0% (19/38) of samples with ETE had downregulation of hsa-miR-204-5p

compared to 24.3% (17/70) of samples with no ETE (Fisher's exact test, p = 0.01) (see Figure 4-16).

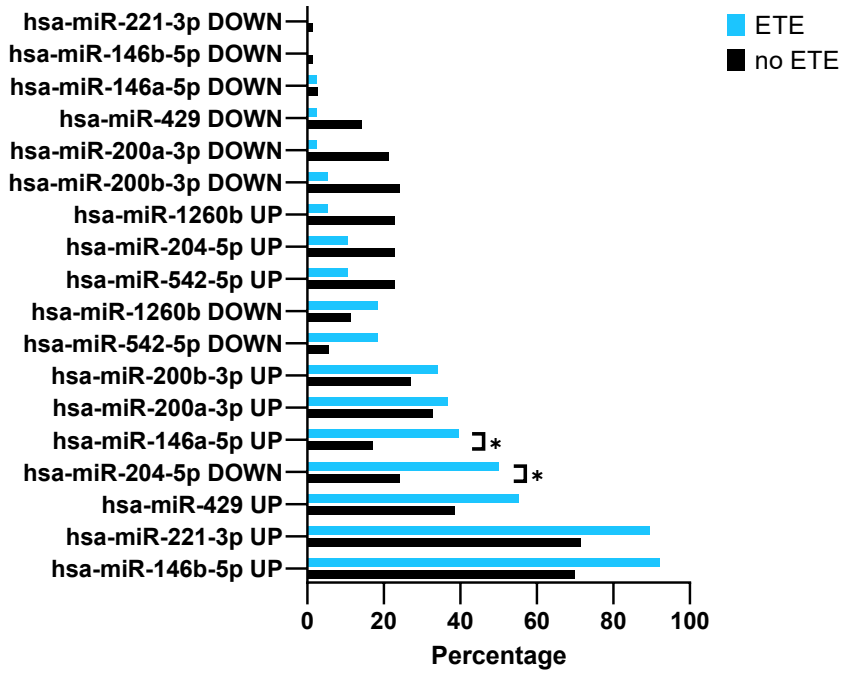


Figure 4-16. Bar plot showing proportion of DTC samples with up or down dysregulation of miRNAs of interest dichotomised by extra-thyroidal extension status * $p \leq 0.05$

Tumours with recurrent disease had upregulation of hsa-miR-429 in 60%, hsa-miR-146b-5p in 84% and hsa-miR-221-3p in 88%, although none showed statistically significant association (see Figure 4-17).

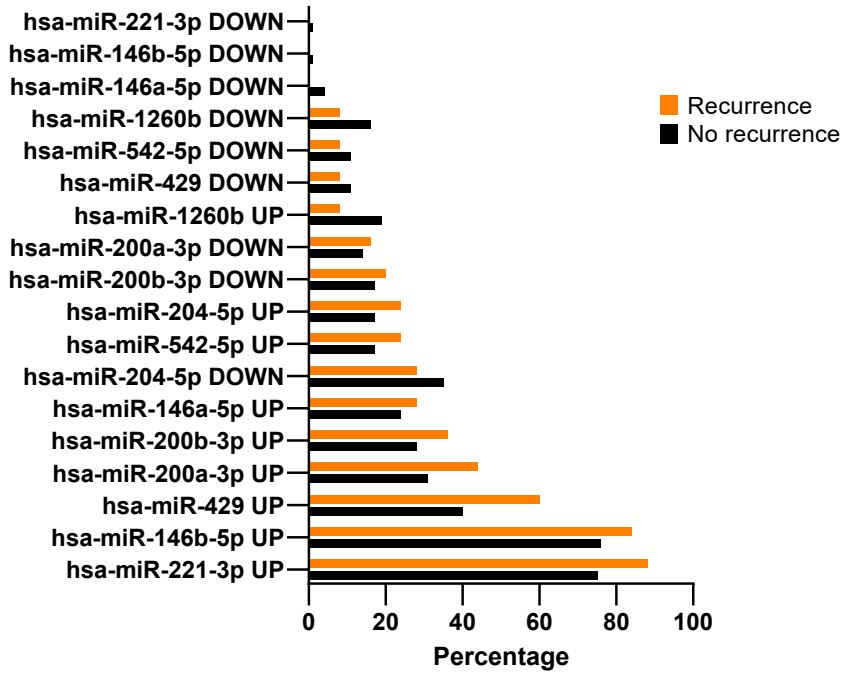


Figure 4-17. Bar plot showing proportion of DTC samples with up or down dysregulation of miRNAs of interest dichotomised by recurrent disease status

4.4.4 Improving multiple logistic regression model with miRNA dysregulation status

In the previous chapter, we characterised 73 DTC tumours from the "Gene Mutation Analysis as a biomarker in Thyroid Cancer (GEM-TC)" study. The characteristics included somatic mutation status from hotspot mutations in DTC as well as various histopathological phenotypes. Sixty nine (95%) of these DTC tumours also had miRNA differential expression data, analysed in this present study. The multiple logistic regression model previously described in section 3.4.9, demonstrated a somatic mutational signature that predicts the presence of vascular invasion in DTC with 59.5% sensitivity, 77.4% specificity, PPV 58.5%, NPV 78.1%, area under the ROC curve 0.74 and therefore 67% accuracy. With the additional miRNA differential expression for each sample, the model could be improved. By combining the dysregulation status of hsa-miRs 146a-5p, 204-5p and 221-3p with the somatic mutation status of *BRAF*, *RAS* and *TERTp*, the model improved dramatically with PPV of 70.9% and NPV of 84.2%, with area under the ROC curve of 0.84 ($p < 0.0001$, 95% CI 0.75-0.94). This model correctly classifies samples 78.3% of the time (see Figure 4-18).

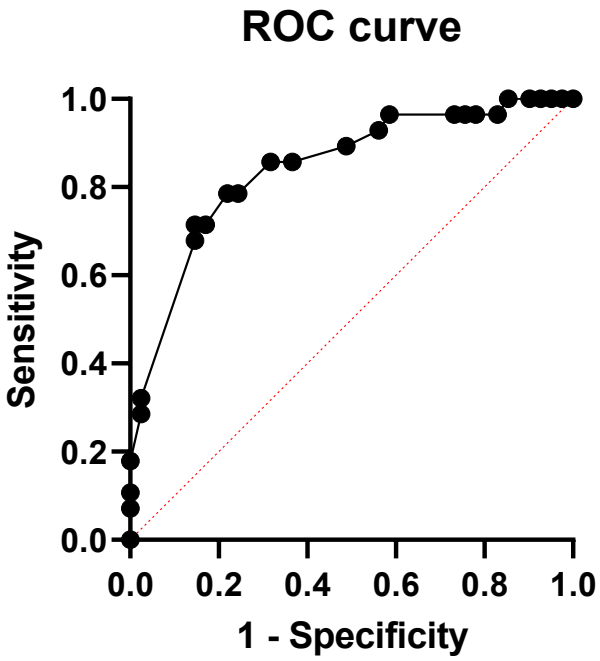


Figure 4-18. ROC curve of combined somatic mutation and miRNA dysregulation signature as a predictive model for the presence of VI. ROC curve showing a significant improvement in the ability of the predictive model to predict the presence of vascular invasion in the whole cohort (AUC 0.84).

We went further and applied the same model to predict the presence of any high risk feature (VI, ETE or lymph node invasion). This had a PPV of 87.5% and NPV of 82.0%, with area under the ROC curve of 0.86 ($p < 0.0001$, 95% CI 0.77-0.95). This model correctly classifies samples 82.6% of the time (see Figure 4-19).

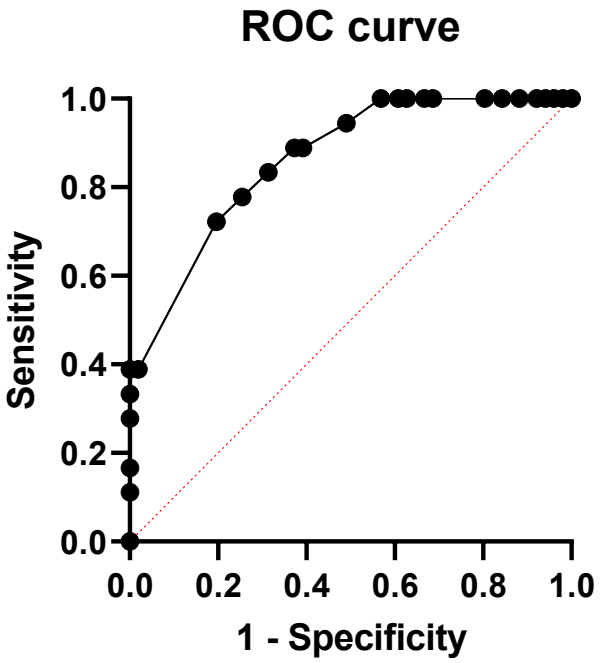


Figure 4-19. ROC curve of combined somatic mutation and miRNA dysregulation signature as a predictive model for the presence of any HRF. AUC 0.86 ($p < 0.0001$, 95% CI 0.77-0.95).

4.5 Discussion

4.5.1 Endogenous references are the foundation of accurate differential expression

A current literature search in PubMed returns over 1800 published studies involving miRNA expression in both benign and malignant thyroid disease. A fundamental problem with many of these studies, particularly when performing differential expression, is the lack of validated endogenous reference miRNAs used in the calculation of fold change difference. Historically, surrogate molecules like snoRNAs and GAPDH have been used as endogenous references (8, 21, 27, 28). Some published studies use reference miRNAs from other differential expression experiments involving completely different tissue types than is being investigated (29). This has resulted in the reporting of a heterogenous mix of dysregulated miRNAs in association with certain phenotypes or clinical outcomes. There is well documented evidence that the expression of miRNAs tends to be tissue specific, and therefore any endogenous references should be derived from both normal and diseased thyroid tissue (30, 31). Further, many studies also fail to have a normal thyroid tissue control cohort, which is absolutely necessary when deciding whether observed differential expression is significant or not. Many studies include a benign thyroid cohort, commonly described within a “normal” cohort, which can include lesions such as follicular adenoma. This could skew the degree of differential expression observed, as there is evidence that these benign lesions could have similar underlying miRNA dysregulation as is found in malignant disease (32, 33).

In our study, we first performed two pilot studies to demonstrate that our methodology of dual DNA and total RNA extraction from FFPE tissue blocks of varying age, could provide viable miRNAs. Extraction of miRNAs from FFPE tissue has been shown in previous studies, but was important to confirm in our own samples (34). MicroRNA expression levels can be studied by several methods: microarray analysis, real-time PCR, Northern blots, in situ hybridization, solution hybridization (such as Nanostring) and sequencing. Of these techniques RT-PCR is considered the gold standard method (35).

Considerations for which technique to employ included sample purity, financial cost and experimental time, availability of downstream analysis support and number of miRNA species being analysed. Using qPCR to analyse the small panel of miRNA species in the first pilot experiment was appropriate as there are readily available validated assays from large biotech companies that have demonstrated equivalent sensitivity and accuracy. However, when performing multiplex miRNA differential expression experiments aiming for novel discovery, broader techniques such as RNAseq or microarrays are more commonly used, covering many thousands of transcripts at once. These techniques are expensive and lack flexibility and reproducibility when evaluating low-quality RNA samples such as those from FFPE (36-38). The Nanostring Technologies nCounter platform is a relatively new technology that has been used within various clinical and research applications. The automated nCounter platform hybridizes fluorescent barcodes directly to specific nucleic acid sequences, allowing for the non-amplified measurement of over 800 targets within one sample (39). We chose to capitalise on the commercially available nCounter Human v3 miRNA Expression Assay that covered 830 mature human miRNA molecules, partly due to cost effectiveness in covering a large number of transcripts, but also in the freely available data analysis suite which allowed increased ease of downstream data analysis. We found this to be a very easy platform to use, with the main limitations being the higher cost (relative to qPCR) and the requirement for RNA samples to be relatively free from guanidine salt contamination. For these reasons, the subsequent large scale cohort analysis was performed using qPCR and custom designed plates instead of Nanostring.

An important secondary purpose of the pilot experiments were to determine suitable thyroid specific endogenous reference miRNAs for the purpose of accurate downstream differential expression calculation. Using two orthogonal techniques (both qPCR and Nanostring), we showed that for our DTC and normal thyroid cohorts, hsa-miR-423-5p and hsa-miR-93-5p are the most stably expressed miRNA with the least amount of expressional variation. These could potentially be reference miRNA candidates specific to thyroid.

4.5.2 Differential miRNA expression associated with high risk histopathological features in DTC

Our study provides further evidence that miRNA dysregulation can be used to differentiate between cancer and normal thyroid tissue, as well as between histological subtypes of PTC and FTC with a high degree of significance. There are a number of commercially available diagnostic kits in North America that incorporate combinations of genomic biomarkers into screening panels, some of which include miRNAs. They all claim high sensitivity and specificity in differentiating cancer from benign disease, particularly in the context of the indeterminate nodule. Many of these tests are not available in the UK and are prohibitively expensive, whilst also requiring the transport of patient samples abroad. A recent multicentre prospective blinded cohort study screened 286 cytologically indeterminate FNA samples from thyroid nodules with the ThyroSeq v3 multigene classifier (40). They showed that in Bethesda III and IV nodules combined, the test demonstrated 94% sensitivity and 82% specificity. False-negative rates are between 1-3% and therefore approaching that of benign cytology. It is conceivable that a molecular test will soon surpass cytology and radiology in cancer diagnostic ability in thyroid malignancy.

The objective of our study addressed a different clinical issue relating to prognostication and guiding initial management of a confirmed thyroid cancer. This involves defining surrogate markers of high risk histopathological features or poor outcomes such as disease recurrence or reduced disease specific survival. The latter outcomes are challenging to investigate given the frequently long period between initial treatment and disease recurrence and overall high disease-specific survival, particularly in low-risk differentiated thyroid cancer, which makes up the majority of all thyroid cancer.

High risk histopathological findings such as vascular invasion and extra-thyroidal extension are common factors that influence MDT recommendations for further surgical treatment after diagnostic hemi-thyroidectomy. In fulfilment of the primary aims of our study, we have shown that upregulation of hsa-miR-204-5p and hsa-miR-146a-5p are significantly associated with the presence of vascular invasion in DTC ($p = 0.044$ and 0.02 respectively) and that extra-thyroidal extension was significantly associated with upregulation of hsa-miR-

146a-5p and downregulation of hsa-miR-204-5p (p = 0.019 and 0.01 respectively). We also showed that a high proportion of tumours with recurrent disease had upregulation of hsa-miR-429, hsa-miR-146b-5p and hsa-miR-221-3p. The dysregulated miRNA species that have been highlighted in this study have been previously described in the literature as being involved in a number of signalling pathways that could explain an association with high risk pathological phenotypes in DTC.

Hsa-miR-146a is involved in Toll-like receptor, NF-kappa beta and ErbB signalling pathways, which play a role in the innate immune system and intracellular signalling respectively (41-43). Dysregulation of these pathways could lead to immune escape and facilitate both proliferation and migration of cancer cells. The target genes for hsa-miR-146a include *NRAS* (oncogene involved in cell proliferation, differentiation and adhesion regulation); *IRAK1* (the gene involved in the signalling pathway of interleukin-1 and apoptosis regulation); *LFNG* (gene that participates in the epithelial-to-mesenchymal transition (44). Further, a study by Czajka et al showed that inhibition of miR-146a and miR-146b resulted in a concomitant decrease in proliferation rates of thyroid cancer cell lines (45).

There is a growing body of evidence that describes hsa-miR-204 as having tumour suppressor activity in many different types of solid cancers, including triple negative breast cancer, laryngeal squamous cell carcinoma, renal cell carcinoma and endometrial cancer, with downregulation of hsa-miR-204 expression being associated with higher rates of aggressive disease (46-49). The exact mechanism is poorly understood, but increased expression of hsa-miR-204 is hypothesised to be involved in inducing apoptosis (50-52).

Dysregulation of hsa-miR-221 is usually paired with hsa-miR-222 as they act in a cluster. They have both been frequently identified as dysregulated across different cancer types. Multiple systematic reviews and meta-analyses agree on the association between hsa-miR-221/222 overexpression and poorer overall survival in cancer (53-55). Proposed mechanisms of action include the induction of cell proliferation through the activation of the cell cycle and the Akt pathway, through post-transcriptional expression modification of genes such as *PTEN*,

TIMP3, *p27Kip1* and *p57* (56, 57). Deeper investigation into the mechanisms and pathways involved in the dysregulation of these key miRNAs was outside the scope of this study.

Further to our findings of significant dysregulation of these key miRNAs in association with high risk histopathological features in our DTC cohort, we combined the dysregulation status of hsa-miRs 146a-5p, 204-5p and 221-3p with the somatic mutation status of *BRAF*, *RAS* and *TERTp* into an updated predictive model that correctly identified DTC samples with vascular invasion with 78% sensitivity and 79% specificity, with a positive likelihood ratio of 3.64. This model was also able to correctly identify DTC samples with at least one HRF (VI, ETE or LN invasion) with a PPV of 87.5% and NPV of 82.0%, with area under the ROC curve of 0.86 ($p < 0.0001$, 95% CI 0.77-0.95). Whilst some research groups are actively investigating molecular biomarkers to improve the diagnostic pathway of DTC, the genomic signature we have proposed could have important clinical utility in the pre-operative phase prior to the index operation. There is clear potential as an adjunct to existing prognostic factors (radiological and cytological findings) to further quantify risk at a molecular level of aggressive histopathological features and therefore help inform surgical extent at index operation.

4.5.3 Challenges of miRNA as a biomarker

The most challenging aspects of using miRNA differential expression as prognostic biomarkers, lies in the high variability of their expression and the ability of single miRNA species to have multiple post-transcriptional targets. Multiple characterising genomic targets need to be combined into a testing panel to increase the accuracy of phenotype prediction. This is clearly required if any useful clinical applications can be promoted.

The practicality of working with RNA samples is fraught with challenges including strict temperature control, unforgiving requirement to keep all physical contact nuclease free and propensity for RNA concentrations to fall dramatically with repeated freeze-thaw cycles. It was outside the scope of this study to compare and measure how these factors affect miRNA specifically.

4.5.4 Further work

This study used two different molecular techniques to determine miRNA differential expression in a cohort of DTC tumours. The Nanostring pilot experiment was limited by the relatively low sample number. If further sample numbers could be studied with the same 830 miRNA panel, it is possible that new significantly differentially expressed miRNA species might be associated with various high risk features. Nanostring, as a relatively new technology, is limited by higher cost per sample and a requirement for highly pure RNA input. Any further work with Nanostring would require an adjustment to the total RNA extraction protocol, to reduce any contamination with guanidine salts.
Based on the findings from this study, further work could be performed in defining a more financially economical panel of genomic markers that could be used to not only differentiate cancer from benign, but also prognosticate nodules in the pre-operative diagnostic phase. These potential studies are discussed in more detail in Chapter 6. Further large number retrospective or prospective clinical trials are recommended to validate the predictive model described in this study.

4.6 Conclusion

Differential expression of miRNAs both extracted directly from tumour cells or free floating in the circulation are exciting avenues of current research in thyroid cancer. Future studies can benefit from the findings in this study; of thyroid specific reference miRNAs as well as significantly dysregulated miRNAs that can differentiate between cancer and normal, PTC and FTC as well as the presence of high risk histopathological features like vascular invasion.

4.7 References

1. Zhang Z, Qin YW, Brewer G, Jing Q. MicroRNA degradation and turnover: regulating the regulators. Wiley Interdiscip Rev RNA. 2012;3(4):593-600.
2. Aryani A, Denecke B. In vitro application of ribonucleases: comparison of the effects on mRNA and miRNA stability. BMC Res Notes. 2015;8:164.
3. Belter A, Gudanis D, Rolle K, Piwecka M, Gdaniec Z, Naskret-Barciszewska MZ, et al. Mature miRNAs form secondary structure, which suggests their function beyond RISC. PLoS One. 2014;9(11):e113848.

4. Mathivanan S, Ji H, Simpson RJ. Exosomes: extracellular organelles important in intercellular communication. J Proteomics. 2010;73(10):1907-20.
5. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9(6):654-9.
6. Ramshani Z, Zhang C, Richards K, Chen L, Xu G, Stiles BL, et al. Extracellular vesicle microRNA quantification from plasma using an integrated microfluidic device. Commun Biol. 2019;2:189.
7. Dai D, Tan Y, Guo L, Tang A, Zhao Y. Identification of exosomal miRNA biomarkers for diagnosis of papillary thyroid cancer by small RNA sequencing. Eur J Endocrinol. 2020;182(1):111-21.
8. Nikiforova MN, Chiosea SI, Nikiforov YE. MicroRNA expression profiles in thyroid tumors. Endocr Pathol. 2009;20(2):85-91.
9. Pallante P, Battista S, Pierantoni GM, Fusco A. Deregulation of microRNA expression in thyroid neoplasias. Nat Rev Endocrinol. 2014;10(2):88-101.
10. Lee JC, Gundara JS, Glover A, Serpell J, Sidhu SB. MicroRNA expression profiles in the management of papillary thyroid cancer. Oncologist. 2014;19(11):1141-7.
11. Chruscik A, Lam AK. Clinical pathological impacts of microRNAs in papillary thyroid carcinoma: A crucial review. Exp Mol Pathol. 2015;99(3):393-8.
12. Hu Y, Wang H, Chen E, Xu Z, Chen B, Lu G. Candidate microRNAs as biomarkers of thyroid carcinoma: a systematic review, meta-analysis, and experimental validation. Cancer Med. 2016;5(9):2602-14.
13. Pishkari S, Paryan M, Hashemi M, Baldini E, Mohammadi-Yeganeh S. The role of microRNAs in different types of thyroid carcinoma: a comprehensive analysis to find new miRNA supplementary therapies. J Endocrinol Invest. 2018;41(3):269-83.
14. Celano M, Rosignolo F, Maggisano V, Pecce V, Iannone M, Russo D, et al. MicroRNAs as Biomarkers in Thyroid Carcinoma. Int J Genomics. 2017;2017:6496570.
15. Pallante P, Visone R, Ferracin M, Ferraro A, Berlingieri MT, Troncone G, et al. MicroRNA deregulation in human thyroid papillary carcinomas. Endocr Relat Cancer. 2006;13(2):497-508.
16. Acibucu F, Dokmetas HS, Tutar Y, Elagoz S, Kilicli F. Correlations between the expression levels of micro-RNA146b, 221, 222 and p27Kip1 protein mRNA and the clinicopathologic parameters in papillary thyroid cancers. Exp Clin Endocrinol Diabetes. 2014;122(3):137-43.
17. Geraldo MV, Kimura ET. Integrated Analysis of Thyroid Cancer Public Datasets Reveals Role of Post-Transcriptional Regulation on Tumor Progression by Targeting of Immune System Mediators. PLoS One. 2015;10(11):e0141726.
18. Sun M, Fang S, Li W, Li C, Wang L, Wang F, et al. Associations of miR-146a and miR-146b expression and clinical characteristics in papillary thyroid carcinoma. Cancer Biomark. 2015;15(1):33-40.
19. Wang Z, Zhang H, He L, Dong W, Li J, Shan Z, et al. Association between the expression of four upregulated miRNAs and extrathyroidal invasion in papillary thyroid carcinoma. Onco Targets Ther. 2013;6:281-7.
20. Dettmer M, Perren A, Moch H, Komminoth P, Nikiforov YE, Nikiforova MN. Comprehensive MicroRNA expression profiling identifies novel markers in follicular variant of papillary thyroid carcinoma. Thyroid. 2013;23(11):1383-9.

21. Jacques C, Guillotin D, Fontaine JF, Franc B, Mirebeau-Prunier D, Fleury A, et al. DNA microarray and miRNA analyses reinforce the classification of follicular thyroid tumors. J Clin Endocrinol Metab. 2013;98(5):E981-9.
22. Mancikova V, Castelblanco E, Pineiro-Yanez E, Perales-Paton J, de Cubas AA, Inglada-Perez L, et al. MicroRNA deep-sequencing reveals master regulators of follicular and papillary thyroid tumors. Mod Pathol. 2015;28(6):748-57.
23. Swierniak M, Wojcicka A, Czetwertynska M, Stachlewska E, Maciag M, Wiechno W, et al. In-depth characterization of the microRNA transcriptome in normal thyroid and papillary thyroid carcinoma. J Clin Endocrinol Metab. 2013;98(8):E1401-9.
24. Zhao L, Zhou X, Shan X, Qi LW, Wang T, Zhu J, et al. Differential expression levels of plasma microRNA in Hashimoto's disease. Gene. 2018;642:152-8.
25. Titov SE, Demenkov PS, Ivanov MK, Malakhina ES, Poloz TL, Tsivlikova EV, et al. Selection and validation of miRNAs as normalizers for profiling expression of microRNAs isolated from thyroid fine needle aspiration smears. Oncol Rep. 2016;36(5):2501-10.
26. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245-50.
27. Yip L, Kelly L, Shuai Y, Armstrong MJ, Nikiforov YE, Carty SE, et al. MicroRNA signature distinguishes the degree of aggressiveness of papillary thyroid carcinoma. Ann Surg Oncol. 2011;18(7):2035-41.
28. Nikiforova MN, Tseng GC, Steward D, Diorio D, Nikiforov YE. MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility. J Clin Endocrinol Metab. 2008;93(5):1600-8.
29. Han PA, Kim HS, Cho S, Fazeli R, Najafian A, Khawaja H, et al. Association of BRAF V600E Mutation and MicroRNA Expression with Central Lymph Node Metastases in Papillary Thyroid Cancer: A Prospective Study from Four Endocrine Surgery Centers. Thyroid. 2016;26(4):532-42.
30. Guo Z, Maki M, Ding R, Yang Y, Zhang B, Xiong L. Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. Sci Rep. 2014;4:5150.
31. Preusse M, Theis FJ, Mueller NS. miTALOS v2: Analyzing Tissue Specific microRNA Function. PLoS One. 2016;11(3):e0151771.
32. Nwadiugwu MC. Thyroid Tumor: Investigating MicroRNA-21 Gene Suppression in FTC and FTA. Cancer Inform. 2020;19:1176935120948474.
33. Stokowy T, Wojtas B, Krajewska J, Stobiecka E, Dralle H, Musholt T, et al. A two miRNA classifier differentiates follicular thyroid carcinomas from follicular thyroid adenomas. Mol Cell Endocrinol. 2015;399:43-9.
34. Boisen MK, Dehlendorff C, Linnemann D, Schultz NA, Jensen BV, Hogdall EV, et al. MicroRNA Expression in Formalin-fixed Paraffin-embedded Cancer Tissue: Identifying Reference MicroRNAs and Variability. BMC Cancer. 2015;15:1024.
35. Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, et al. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA. 2010;16(5):991-1006.

36. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordonez B, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. BMC Biotechnol. 2011;11:46.
37. Saba NF, Wilson M, Doho G, DaSilva J, Benjamin Isett R, Newman S, et al. Mutation and Transcriptional Profiling of Formalin-Fixed Paraffin Embedded Specimens as Companion Methods to Immunohistochemistry for Determining Therapeutic Targets in Oropharyngeal Squamous Cell Carcinoma (OPSCC): A Pilot of Proof of Principle. Head Neck Pathol. 2015;9(2):223-35.
38. Scott DW, Chan FC, Hong F, Rogic S, Tan KL, Meissner B, et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. J Clin Oncol. 2013;31(6):692-700.
39. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008;26(3):317-25.
40. Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, Sipos JA, et al. Performance of a Multigene Genomic Classifier in Thyroid Nodules With Indeterminate Cytology: A Prospective Blinded Multicenter Study. JAMA Oncol. 2019;5(2):204-12.
41. Lopez JP, Fiori LM, Cruceanu C, Lin R, Labonte B, Cates HM, et al. MicroRNAs 146a/b-5 and 425-3p and 24-3p are markers of antidepressant response and regulate MAPK/Wnt-system genes. Nat Commun. 2017;8:15497.
42. Bhaumik D, Scott GK, Schokrpur S, Patil CK, Campisi J, Benz CC. Expression of microRNA-146 suppresses NF-kappaB activity with reduction of metastatic potential in breast cancer cells. Oncogene. 2008;27(42):5643-7.
43. Quinn EM, Wang JH, O'Callaghan G, Redmond HP. MicroRNA-146a is upregulated by and negatively regulates TLR2 signaling. PLoS One. 2013;8(4):e62232.
44. Aksenenko M, Palkina N, Komina A, Tashireva L, Ruksha T. Differences in microRNA expression between melanoma and healthy adjacent skin. BMC Dermatol. 2019;19(1):1.
45. Czajka AA, Wojcicka A, Kubiak A, Kotlarek M, Bakula-Zalewska E, Koperski L, et al. Family of microRNA-146 Regulates RARbeta in Papillary Thyroid Carcinoma. PLoS One. 2016;11(3):e0151968.
46. Gao W, Wu Y, He X, Zhang C, Zhu M, Chen B, et al. MicroRNA-204-5p inhibits invasion and metastasis of laryngeal squamous cell carcinoma by suppressing forkhead box C1. J Cancer. 2017;8(12):2356-68.
47. Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, et al. Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Int J Cancer. 2012;130(5):1036-45.
48. Mikhaylova O, Stratton Y, Hall D, Kellner E, Ehmer B, Drew AF, et al. VHL-regulated MiR-204 suppresses tumor growth through inhibition of LC3B-mediated autophagy in renal clear cell carcinoma. Cancer Cell. 2012;21(4):532-46.
49. Toda H, Kurozumi S, Kijima Y, Idichi T, Shinden Y, Yamada Y, et al. Molecular pathogenesis of triple-negative breast cancer based on microRNA expression signatures: antitumor miR-204-5p targets AP1S3. J Hum Genet. 2018;63(12):1197-210.
50. Li T, Pan H, Li R. The dual regulatory role of miR-204 in cancer. Tumour Biol. 2016;37(9):11667-77.

51. Liu L, Wang J, Li X, Ma J, Shi C, Zhu H, et al. MiR-204-5p suppresses cell proliferation by inhibiting IGFBP5 in papillary thyroid carcinoma. Biochem Biophys Res Commun. 2015;457(4):621-6.
52. Vimalraj S, Miranda PJ, Ramyakrishna B, Selvamurugan N. Regulation of breast cancer and bone metastasis by microRNAs. Dis Markers. 2013;35(5):369-87.
53. Ravegnini G, Cargnin S, Sammarini G, Zanotti F, Bermejo JL, Hrelia P, et al. Prognostic Role of miR-221 and miR-222 Expression in Cancer Patients: A Systematic Review and Meta-Analysis. Cancers (Basel). 2019;11(7).
54. Zhang P, Zhang M, Han R, Zhang K, Ding H, Liang C, et al. The correlation between microRNA-221/222 cluster overexpression and malignancy: an updated meta-analysis including 2693 patients. Cancer Manag Res. 2018;10:3371-81.
55. Liang L, Zheng X, Hu M, Cui Y, Zhong Q, Wang S, et al. MiRNA-221/222 in thyroid cancer: A meta-analysis. Clin Chim Acta. 2018;484:284-92.
56. Di Martino MT, Rossi M, Caracciolo D, Gulla A, Tagliaferri P, Tassone P. Mir-221/222 are promising targets for innovative anticancer therapy. Expert Opin Ther Targets. 2016;20(9):1099-108.
57. Garofalo M, Quintavalle C, Romano G, Croce CM, Condorelli G. miR221/222 in cancer: their role in tumor progression and response to therapy. Curr Mol Med. 2012;12(1):27-33.

Chapter 5 Detection of ctDNA in early thyroid cancer

5.1 Introduction

In the context of oncology, liquid biopsy defined as a non-invasive method of collecting circulating tumour specific biomarkers for malignant disease, is an exciting prospect. The potential benefits are clear; serial blood sampling with minimal morbidity compared to percutaneous or open biopsy, collection as an outpatient and potential provision of tumour specific molecular characteristics that can inform prognosis and monitor disease extent.

Amongst the multiple candidate molecules that are being investigated as viable circulating biomarkers in cancer detection and monitoring, circulating tumour DNA (ctDNA) has been shown to be viable option in a number of solid cancers (1-5). These studies have shown the efficacy of ctDNA detection in monitoring minimal residual disease, but also as a strong independent predictor of subsequent recurrent disease. Whilst rates of detection vary between cancer types and extent of disease, detection is highest in advanced cancers with aggressive phenotypes (6-8).

In DTC specifically, there is a relative paucity of published research on ctDNA. Early studies have focused on the detection of ctDNA as a marker of benign versus malignant disease, mainly assaying single mutations such as *BRAFV600E* mutation in PTC. This is understandable given PTC accounts for the highest proportion of DTC and *BRAF* harbouring the most frequent somatic mutation. However, there appears to be huge variability in the rate of detection, with some researchers reporting very low or absent detection in plasma or serum using sensitive PCR based techniques. This variability in detection could be related to tumour pathology, where tumour necrosis or apoptosis does not feature highly within DTC and therefore ctDNA is not released into the circulation in concentration high enough to be detected. Variation could also be related to differing stage of disease, where burden of disease is clearly different when comparing early versus advanced stage and this could be reflected in the degree of ctDNA being released into the circulation. This variation in ctDNA

detection according to early or late stage disease is well known (6), but the limits to which it is detectable in early stage disease, particularly in DTC is currently poorly understood.

Further confounding considerations include which techniques are employed for ctDNA detection, the technical limit of detection sensitivity and also if single target assaying or multiplexing is utilised. Due to the low fractional abundance of ctDNA in proportion to total cell-free DNA, sampling effects become an important factor, particularly when attempting to detect low frequency events (9). Studies in breast cancer have shown how targeting multiple mutations can increase the rate of ctDNA detection. Garcia-Murillas et al performed targeted sequencing of DNA extracted from the baseline biopsy of early breast cancer (node negative, <T3 disease with no distant metastases) and used the subsequent somatic mutations to inform the design of personalised droplet digital PCR (ddPCR) probes to assay serial plasma cfDNA samples. They detected ctDNA in 69% of baseline plasma samples and were able to track the fluctuations in ctDNA based on tumour specific mutations allowing identification of patients at risk of disease recurrence (10). Further, Coombes et al demonstrated the use of whole exome sequencing (WES) on primary breast tumour DNA to inform the subsequent design of personalised panels of 16 variant targets to be tracked in serial plasma cfDNA samples using ultradeep sequencing. They showed that the detection of ctDNA preceded clinical or radiological evidence of disease relapse in 89% of relapse patients and that all patients that remained disease free also had no ctDNA detectable in any timepoint (11).

Our group published a study demonstrating the efficacy of targeted sequencing of patient's tumour DNA followed by personalised tracking of variants using ddPCR in the metastatic setting. In this study, involving a cohort of advanced thyroid cancers including poorly differentiated, anaplastic and medullary subtypes, we showed that tracking multiple variants equated to higher rates of ctDNA detection as well as the ability to show treatment resistant sub-clonal cell population evolution in serial blood sampling (12). These findings have yet to be replicated in a cohort of early thyroid cancer.

This present study aims to capitalise on the foundations of previous research detailed above, and exploit new technology in the form of anchored multiplex PCR with unique molecular indexes (UMIs) or molecular barcodes (13). These techniques hope to push the boundaries of ctDNA detection in early DTC, by using NGS that approaches sensitivity levels provided by partition PCR based methods but allowing for greater multiplexing abilities (14-16).

5.2 Hypothesis

That ctDNA is detectable in the plasma of early DTC and that using whole exome sequencing and patient specific bespoke sequencing panels significantly improves detection rate compared to traditional targeted sequencing and singleplex ddPCR assays.

5.3 Aims

1. To perform multi-mutational analysis and characterise the mutational landscape of DTC tumours with both targeted panel sequencing and WES.
2. To extract and quantify cfDNA from matched peri-treatment serial blood tests.
3. To determine the detection rate of ctDNA using bespoke singleplex ddPCR assays.
4. To design bespoke sequencing panels according to the mutational profile of each tumour.
5. To determine the detection rate of ctDNA using direct plasma cfDNA sequencing and patient specific sequencing panels.

5.4 Results

5.4.1 Detection of variants using a thyroid specific targeted panel

We investigated the potential utility of ctDNA analysis in differentiated thyroid cancer in a prospectively recruited cohort of 47 patients presenting with stage 1-4 disease (see Table 5-1).

We subjected primary tumour DNA, extracted from the tumour of the patient's diagnostic index operation, to high depth targeted sequencing using a thyroid specific panel (ThyMa) as detailed in section 2.1.3 of the Materials and Methods Chapter 2. We identified one or more somatic mutation(s) in 27 out of 47 (57.4%) cancers, with two or more mutations found in 8 cases. A summary of histopathological demographics from patients with and without variants is shown in Table 5-2 and a list of which variants and allele frequencies is shown in table Table 5-3.

	Patients with variants detected		Patients without variants detected		Sig. diff?
Mean age (years)	**51.9**		**47.9**		NS
Mean tumour size (mm)	**31.7**		**27.3**		NS
No. PTC (% of total)	**23**	(85.2)	**20**	(74.1)	-
No. FTC (% of total)	**4**	(14.8)	**7**	(25.9)	-
T stage					
1	**6**	(22.2)	**10**	(40.0)	-
2	**8**	(29.6)	**7**	(28.0)	-
3	**11**	(40.7)	**8**	(32.0)	-
4	**2**	(7.4)	**0**	0	-
AJCC stage					
1	**19**	(70.4)	**22**	(84.6)	-
2	**5**	(18.5)	**3**	(11.5)	-
3	**0**	0	**0**	0	-
4	**3**	(11.1)	**1**	(3.8)	-
Vascular invasion present (%)	**16**	(61.5)	**17**	(68.0)	NS
ETE present (%)	**12**	(46.2)	**7**	(28.0)	NS
Multifocal disease (%)	**13**	(52.0)	**12**	(48.0)	NS
Unilateral disease (%)	**14**	(58.3)	**17**	(68.0)	NS
Central LN involvement (%)	**9**	(37.5)	**4**	(16.0)	NS
Lateral LN involvement (%)	**5**	(22.7)	**3**	(12.5)	NS
LN extracapsular spread (%)	**5**	(20.0)	**1**	(4.0)	NS
Recurrent or persistent disease (%)	**9**	(33.3)	**3**	(12.5)	NS

Table 5-2. Summary of patient clinical and tumour histopathological features divided into those with and without variants detected on targeted panel sequencing. PTC – papillary thyroid cancer, FTC – follicular thyroid cancer, AJCC – American Joint Committee for Cancer, ETE – Extrathyroidal extension, LN – Lymph Node, NS – Not significant.

	Follicular (N=10)	Papillary (N=37)	Total (N=47)
Sex			
Male	4 (40.0%)	12 (32.4%)	16 (34.0%)
Female	6 (60.0%)	25 (67.6%)	31 (66.0%)
Age (years)			
Mean (SD)	53.2 (20.3)	49.3 (15.3)	50.1 (16.3)
Median [Min, Max]	56.5 [18.0, 81.0]	48.0 [23.0, 85.0]	49.0 [18.0, 85.0]
Tumour Size (mm)			
Mean (SD)	36.0 (13.0)	30.6 (20.5)	31.8 (19.1)
Median [Min, Max]	36.0 [17.0, 55.0]	22.0 [10.0, 82.0]	25.0 [10.0, 82.0]
T staging			
1b	1 (10.0%)	12 (32.4%)	13 (27.7%)
2	5 (50.0%)	7 (18.9%)	12 (25.5%)
3	4 (40.0%)	12 (32.4%)	16 (34.0%)
3a	0 (0%)	1 (2.7%)	1 (2.1%)
3b	0 (0%)	3 (8.1%)	3 (6.4%)
4a	0 (0%)	1 (2.7%)	1 (2.1%)
4b	0 (0%)	1 (2.7%)	1 (2.1%)
N staging			
0	9 (90.0%)	22 (59.5%)	31 (66.0%)
1a	1 (10.0%)	9 (24.3%)	10 (21.3%)
1b	0 (0%)	6 (16.2%)	6 (12.8%)
M staging			
1	2 (20.0%)	3 (8.1%)	5 (10.6%)
0	8 (80.0%)	34 (91.9%)	42 (89.4%)
VI			
Present	9 (90.0%)	24 (64.9%)	33 (70.2%)
Absent	1 (10.0%)	13 (35.1%)	14 (29.8%)
ETE			
Present	2 (20.0%)	19 (51.4%)	21 (44.7%)
Absent	8 (80.0%)	18 (48.6%)	26 (55.3%)
LN Invasion			
Present	1 (10.0%)	15 (40.5%)	16 (34.0%)
Absent	9 (90.0%)	22 (59.5%)	31 (66.0%)
Multi-focality			
Present	4 (40.0%)	20 (54.1%)	24 (51.1%)
Absent	6 (60.0%)	17 (45.9%)	23 (48.9%)
Recurrent/Residual Disease			
Present	3 (30.0%)	3 (8.1%)	6 (12.8%)
Absent	7 (70.0%)	34 (91.9%)	41 (87.2%)
Mets on presentation			
Present	2 (20.0%)	3 (8.1%)	5 (10.6%)
Absent	8 (80.0%)	34 (91.9%)	42 (89.4%)
RAI treatment			
Yes	10 (100%)	30 (81.1%)	40 (85.1%)
No	0 (0%)	7 (18.9%)	7 (14.9%)

Table 5-1. Summary of patient clinical and histological characteristics

Trial ID	Gene	Protein	HGVs	Allele Freq
GEMTC-P2-1	BRAF	c.1799T>A	p.V600E	0.10
GEMTC-P2-2	BRAF	c.1799T>A	p.V600E	0.25
GEMTC-P2-2	TERT	c.1-124C>T	N/A	0.53
GEMTC-P2-3	AXIN1	c.1948G>A	p.G650S	0.45
GEMTC-P2-3	BRAF	c.1799T>A	p.V600E	0.20
GEMTC-P2-6	BRAF	c.1799T>A	p.V600E	0.17
GEMTC-P2-8	BRAF	c.1799T>A	p.V600E	0.22
GEMTC-P2-12	BRAF	c.1799T>A	p.V600E	0.08
GEMTC-P2-12	GNAS	c.2531G>A	p.R844H	0.06
GEMTC-P2-12	KMT2C	c.1168G>T	p.V390L	0.08
GEMTC-P2-12	TERT	c.1-124C>T	N/A	0.19
GEMTC-P2-13	BRAF	c.1799T>A	p.V600E	0.13
GEMTC-P2-16	NRAS	c.182A>G	p.Q61R	0.36
GEMTC-P2-16	RET	c.322A>C	p.K108Q	0.33
GEMTC-P2-16	TERT	c.1-124C>T	N/A	0.08
GEMTC-P2-19	TERT	c.1-124C>T	N/A	0.46
GEMTC-P2-20	NRAS	c.182A>G	p.Q61R	0.34
GEMTC-P2-21	BRAF	c.1799T>A	p.V600E	0.07
GEMTC-P2-25	BRAF	c.1801A>G	p.K601E	0.30
GEMTC-P2-25	TERT	c.1-124C>T	N/A	0.51
GEMTC-P2-25	TSHR	c.727T>C	p.S243P	0.41
GEMTC-P2-27	TERT	c.1-124C>T	N/A	0.52
GEMTC-P2-30	BRAF	c.1799T>A	p.V600E	0.22
GEMTC-P2-31	TERT	c.1-124C>T	N/A	0.08
GEMTC-P2-36	BRAF	c.1799T>A	p.V600E	0.11
GEMTC-P2-38	BRAF	c.1799T>A	p.V600E	0.18
GEMTC-P2-40	BRAF	c.1799T>A	p.V600E	0.25
GEMTC-P2-42	BRAF	c.1799T>A	p.V600E	0.07
GEMTC-P2-44	BRAF	c.1799T>A	p.V600E	0.10
GEMTC-P2-45	TERT	c.1-124C>T	N/A	0.25
GEMTC-P2-45	BRAF	c.1799T>A	p.V600E	0.44
GEMTC-P2-48	BRAF	c.1799T>A	p.V600E	0.40
GEMTC-P2-48	TERT	c.1-124C>T	N/A	0.57
GEMTC-P2-49	BRAF	c.1799T>A	p.V600E	0.18
GEMTC-P2-51	NRAS	c.182A>G	p.Q61R	0.15
GEMTC-P2-52	BRAF	c.1799T>A	p.V600E	0.06
GEMTC-P2-53	NRAS	c.182A>G	p.Q61R	0.46
GEMTC-P2-56	BRAF	c.1799T>A	p.V600E	0.38
GEMTC-P2-56	TERT	c.1-124C>T	N/A	0.61

Table 5-3. List of variants detected in all tumours that underwent targeted panel sequencing with ThyMa

The most frequently identified variant was *BRAF c.T1799A (p.V600E)* found in 19 out of 27 (70.4%) DTCs. The other variants detected included *NRAS c.A182G (p.Q61R),TERT promoter c.C-124T (p.C228T), AXIN1 c.G1948A (p.G650S), GNAS c.G2531A (p.R844H), KMT2C c.G1168T (p.V390L), RET c.A322C (p.K108Q)* and *TSHR c.T727C (p.S243P)* (see Table 5-3).
The cohort of patients with somatic variants detected had a higher proportion of >T2 stage disease, higher AJCC stage and worse histopathological features including lymph node invasion, lymph node extracapsular spread and recurrent disease. These differences in proportions were not statistically significant on direct statistical comparison.

5.4.2 Plasma time points and cfDNA quantification

In total, 150 individual time-points were collected; 45 pre-op, 45 post-op, 26 pre-RAI and 34 post-RAI. The amount of cfDNA extracted from 4ml of plasma ranged from 4.07ng to 634.7ng (mean 43.31ng ± 68.34ng, 95% CI 32.28-54.33ng) (see Figure 5-1). There was on average slightly higher cfDNA concentration in the post-op time-point compared to pre-op, but this was not statistically significant. There was a significant difference between the post-op and post-RAI time-point (one-way ANOVA with Tukey's multiple comparisons, $p = 0.009$), with an overall trend of reducing cfDNA concentration with more treatment modalities.

The large range of cfDNA concentrations observed was due to a few outlier patients with very high levels of cfDNA. These outliers consisted of 14 individual patients, with 1 patient having high outlier cfDNA levels in all 4 of their plasma time-points. This patient had an index tumour of a 23mm Hürthle cell FTC with TNM staging pT2N0M1. Thirty-six percent of these individuals with outlier cfDNA concentrations had residual or recurrent disease after their index procedure.

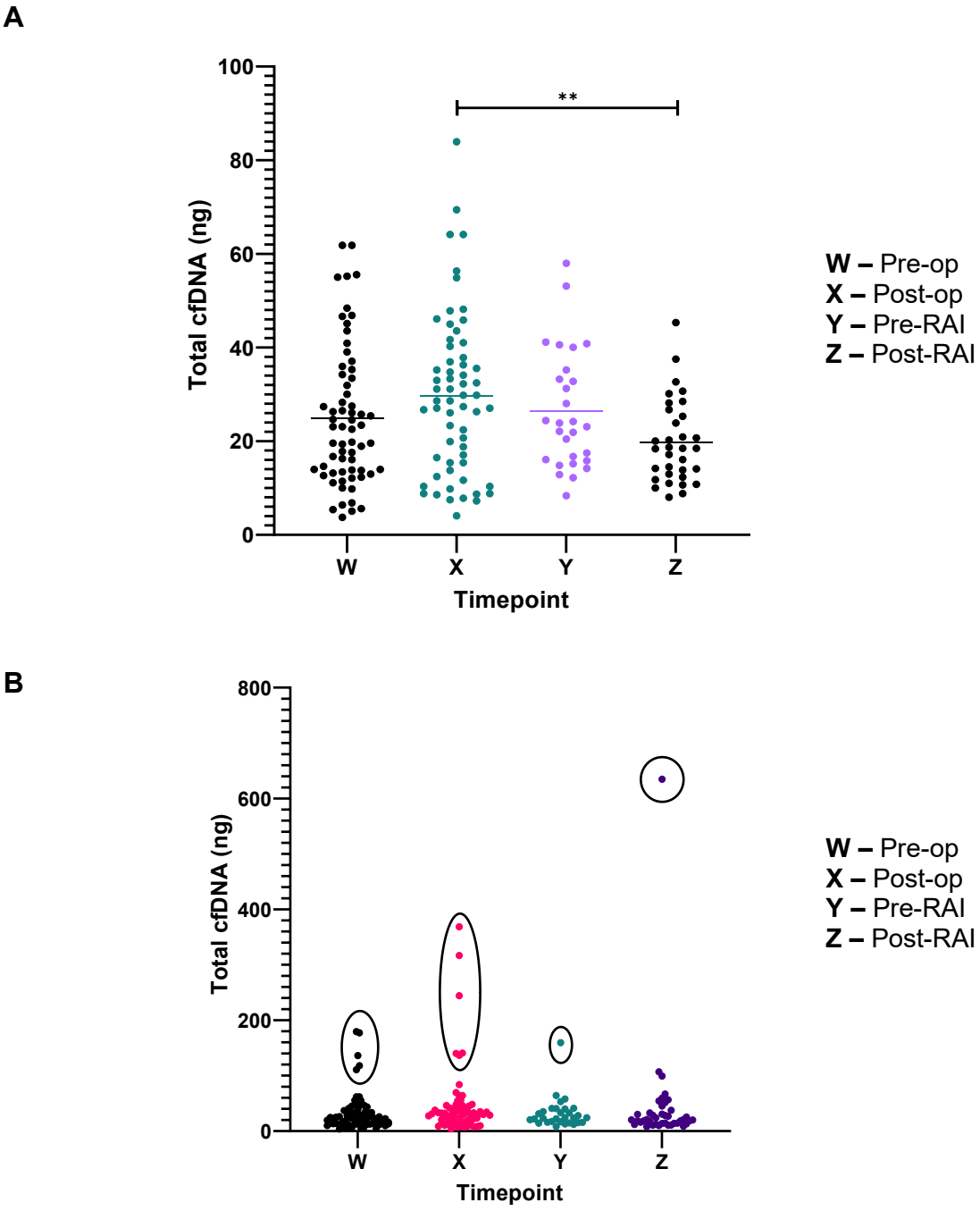


Figure 5-1. Total cfDNA extracted from 4ml plasma time points A - Outliers removed. Significant difference between time point X (post-op) and time point Z (post-RAI), p = 0.009. **B -** Highlighting high concentration outliers. ** p ≤ 0.01

5.4.3 Rate of ctDNA detection with ddPCR

To track the mutations detected from targeted sequencing of the index tumours in their corresponding cfDNA, we designed custom TaqMan ddPCR assays for each somatic mutation identified (full designs in Appendix C). As mentioned in the methods section, each plasma cfDNA analysis by ddPCR was concurrently run with a positive control (tumour DNA from the corresponding patient) and a negative control (buffycoat DNA from the corresponding patient). There was a strong positive correlation between targeted sequencing and ddPCR calculated mutant allele fraction for baseline tumour DNA, validating the presence of these mutations and efficacy of the custom assays (r = 0.9, 95% CI, 0.79-0.95, p = <0.001).

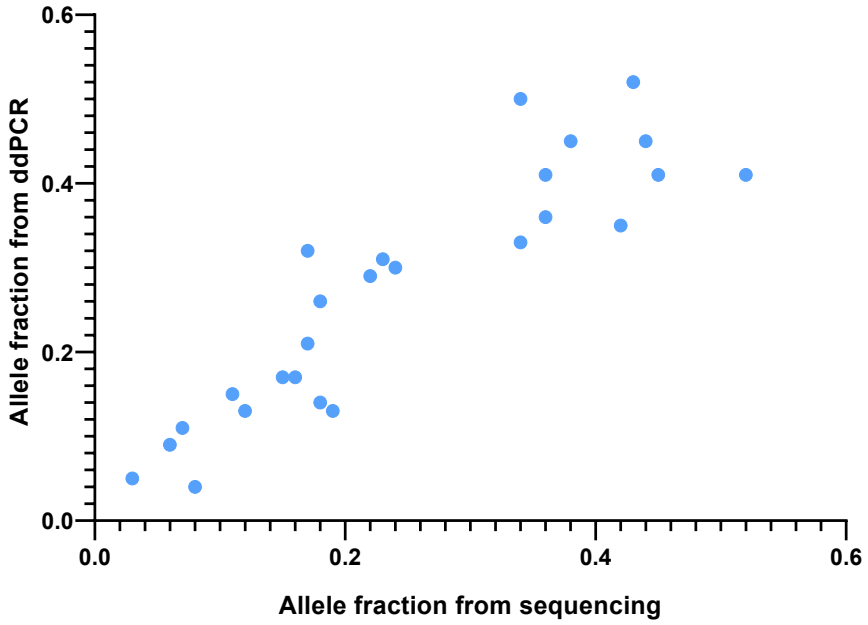


Figure 5-2. Graph showing correlation between mutant allele fractions detected with ddPCR and targeted sequencing. There was a strong positive correlation (r = 0.9, 95% CI, 0.79-0.95, p = <0.001)

Of the 27 patients with a variant detected by ThyMa targeted sequencing, only 3 patients (11.1%) had at least 1 time-point positive for ctDNA using singleplex ddPCR. These 3 patients have larger tumours (T3+) with either nodal or metastatic disease are listed in the Table 5-4 below.

Tumour stage	Histology	Variant	Timepoints	Mutant copies/ml	Wild type copies/ml	Approximate mutant allele fraction in cfDNA
pT4bN1aM1	cPTC	BRAF c.T1799A (p.V600E)	Pre-op Post-op	39 15	6332 14286	0.006 0.001
pT3N0M1	WI FTC	NRAS c.A182G (p.Q61R)	Pre-op Post-op	8 7	1273 1349	0.006 0.005
pT3N1aM0	WI FTC	NRAS c.A182G (p.Q61R)	Pre-op Post-op	1 4	540 377	0.002 0.009

Table 5-4. Patients with at least 1 timepoint positive for ctDNA using ddPCR

5.4.4 Detection of variants using whole exome sequencing (WES)

We next explored if using a broader sequencing technique would identify a higher proportion of samples with potential somatic mutation targets compared to targeted sequencing with a thyroid specific panel. From the original prospective cohort, a representative sub-cohort of 15 tumours, matched with their buffycoat DNA to provide germline variant filtering, underwent WES (described in section 2.4.2 of Chapter 2). The patients included in this cohort are in Table 5-5 below.

Trial ID	Pre-op blood sample	Post-op blood sample	Pre-RAI blood sample	Post-RAI blood sample	Histology	TNM stage	AJCC stage
GEMTC-P2-1	Yes	Yes	No	Yes	cPTC	pT3(m)N1a	1
GEMTC-P2-2	Yes	Yes	No	Yes	cPTC	pT3bN1aM0	2
GEMTC-P2-3	Yes	Yes	Yes	Yes	cPTC	pT2NxM0	1
GEMTC-P2-4	Yes	Yes	No	Yes	fvPTC	pT1b(m)N0M0	1
GEMTC-P2-6	Yes	Yes	No	Yes	cPTC	pT3N0Mx	1
GEMTC-P2-7	Yes	Yes	No	Yes	FTC	pT2N0Mx	1
GEMTC-P2-8	Yes	Yes	No	Yes	cPTC	pT2N0Mx	1
GEMTC-P2-9	Yes	Yes	Yes	Yes	cPTC	pT1b(m)N0M0	1
GEMTC-P2-10	Yes	Yes	No	Yes	FTC	pT3N0M0	1
GEMTC-P2-11	Yes	Yes	Yes	Yes	cPTC	pT3N1aMx	1
GEMTC-P2-12	Yes	Yes	No	Yes	cPTC	pT4bN1a	4b
GEMTC-P2-16	Yes	Yes	Yes	No	FTC	pT3N0M1	4b
GEMTC-P2-21	Yes	Yes	No	Yes	cPTC	pT3N0Mx	2
GEMTC-P2-22	Yes	Yes	Yes	Yes	cPTC	pT2N1a	2
GEMTC-P2-39	Yes	Yes	Yes	No	cPTC	pT1bN0	1

Table 5-5. Summary of patients that underwent WES

Overall 1,394,361,006 reads were achieved, yielding 418,308 Mbases of data with a mean quality score of 38.24 with 91.8% of samples achieving a score greater than 30. Tumour DNA samples were sequenced with median bait coverage of 211X (range 165-247), achieving 91.2% of target bases above 20X, 86.1% unique reads and 99.9% aligned unique reads. Buffycoat DNA samples

were sequenced with median bait coverage of 92X (range 58-101) and 80.7% of target bases above 20X, 89.9% unique reads and 99.9% aligned unique reads. Somatic variants were called using the Illumina Dragen Bio-IT platform in somatic mode. All calls from the tumour DNA were pair-filtered with the corresponding buffycoat DNA as well as a panel of normal to remove artefacts. The results from each sample is listed in the Table 5-6.

Trial ID	TNM stage	Total variants	SNV	Insertion	Deletion
GEMTC-P2-1	pT3(m)N1a	48	48	-	-
GEMTC-P2-2	pT3bN1aM0	239	220	6	13
GEMTC-P2-3	pT2NxM0	63	58	3	2
GEMTC-P2-4	pT1b(m)N0M0	94	81	4	9
GEMTC-P2-6	pT3N0Mx	72	66	3	3
GEMTC-P2-7	pT2N0Mx	215	188	13	14
GEMTC-P2-8	pT2N0Mx	261	210	21	30
GEMTC-P2-9	pT1b(m)N0M0	121	105	6	10
GEMTC-P2-10	pT3N0M0	66	63	2	1
GEMTC-P2-11	pT3N1aMx	90	83	2	5
GEMTC-P2-12	pT4bN1a	71	66	-	5
GEMTC-P2-16	pT3N0M1	128	109	9	10
GEMTC-P2-21	pT3N0Mx	73	66	4	3
GEMTC-P2-22	pT2N1a	67	61	4	2
GEMTC-P2-39	pT1bN0	335	270	31	34

Table 5-6. Summary of variants detected by WES

The Figure 5-3 below summarises the variants classification distribution. The most common type of mutation being a missense mutation.

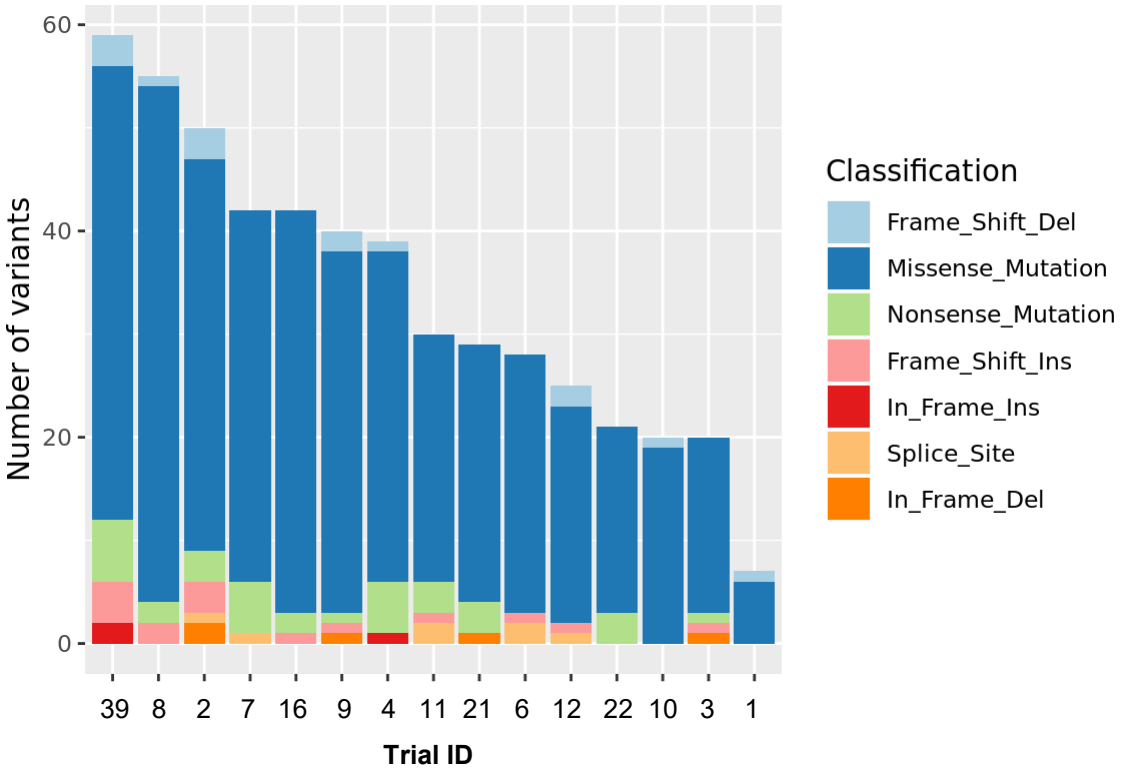


Figure 5-3. Summary stacked bar graph showing the frequency and distribution of variant types across all patients that underwent WES

The most common DNA substitution mutation observed was the transition (interchange of purines or pyrimidines) with C>T substitutions accounting for an approximate average of 50% of SNVs (Figure 5-4).

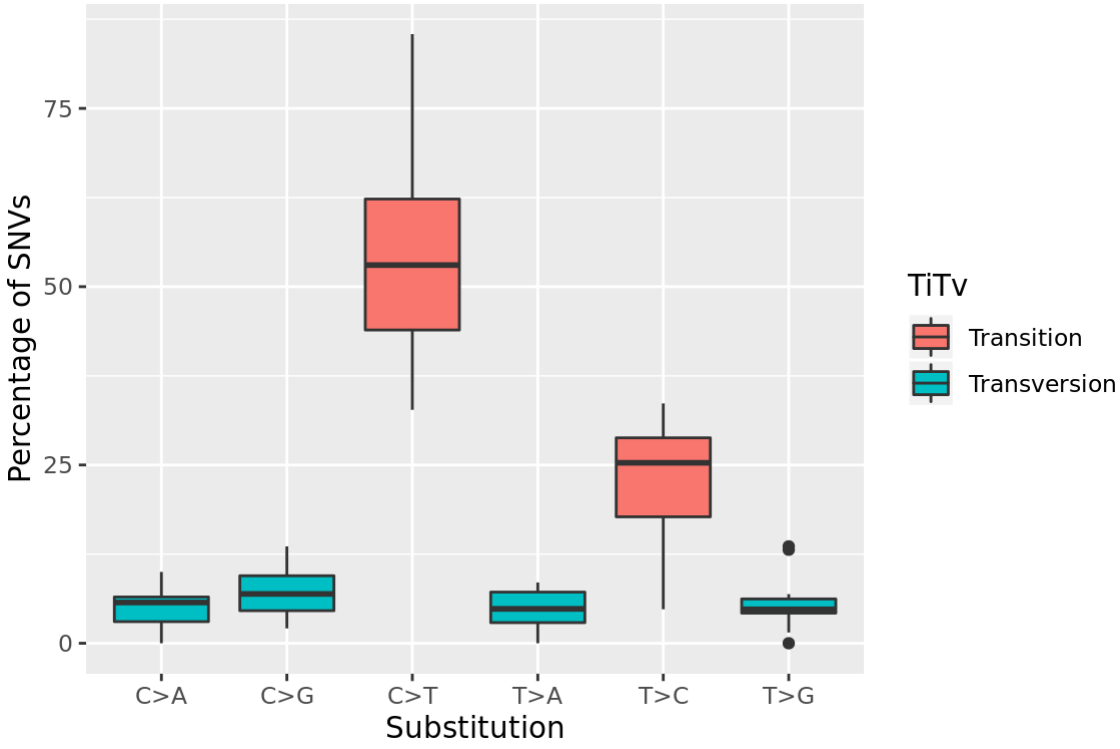


Figure 5-4. Boxplot showing proportion of different substitution types

After manual curation, an average of 16.2 SNVs per sample (range 4-34) were identified from WES. All 15 tumour DNA samples therefore had potential somatic variant targets to assay in their corresponding plasma cfDNA. Many of these variants were patient specific passenger mutations and were not canonical driver mutations associated with thyroid cancer.

All 15 samples underwent dual sequencing, initially with the targeted ThyMa panel and then again with WES. Eight of the fifteen samples had variants detected on ThyMa sequencing and we found complete concordance with the variants detected between the two techniques, adding further evidence to corroborate the presence of these mutations. There was a positive correlation between calculated allele fractions (Figure 5-5), with a trend towards higher allele fraction estimates from WES ($r = 0.79$, 95% CI, 0.18-0.95, $p = 0.02$).

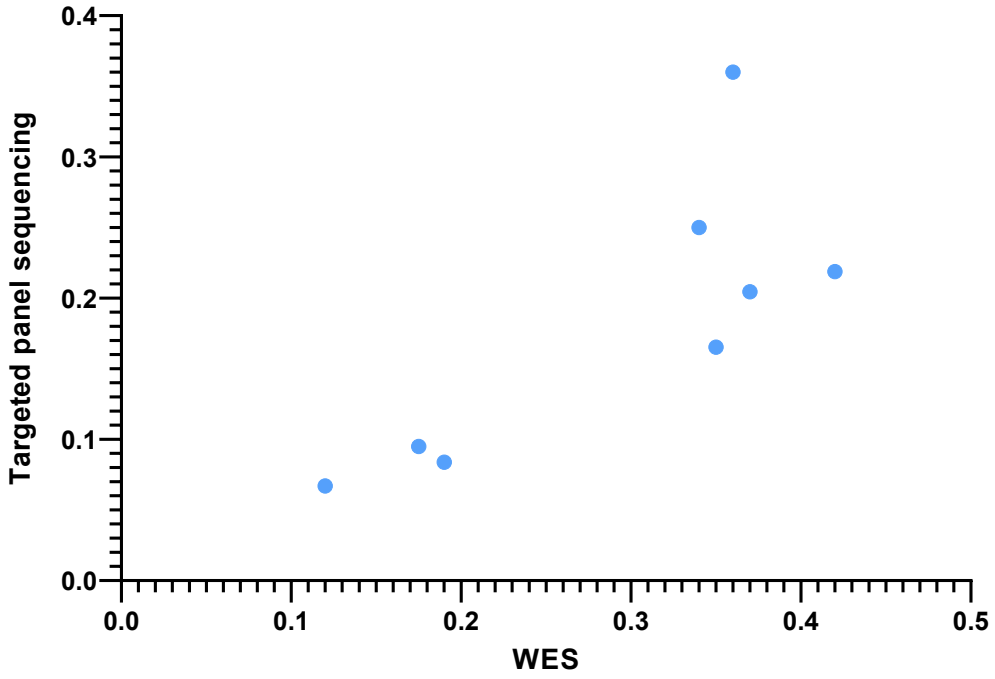


Figure 5-5. Graph showing correlation between mutant allele fractions called by WES and targeted panel sequencing. There was a strong positive correlation (r = 0.79, 95% CI, 0.18-0.95, p = 0.02)

5.4.5 Bespoke patient specific panel creation sequencing of plasma cfDNA

We created bespoke sequencing panels for each patient using the loci of each variant call to inform gene specific primer (GSP) design. In addition to those variants identified from the WES data, all panels included GSPs covering the TERT promoter region as well as common hotpot genes frequently mutated in cancer. The final designs for each patient with each target coordinates are available in Appendix D. Overall, there was an average of 74 targets per panel with 95% coverage of specified targets (see Table 5-7).

Trial ID	TNM stage	Total targets	Overall coverage
GEMTC-P2-1	pT3(m)N1a	76	95%
GEMTC-P2-2	pT3bN1aM0	73	95%
GEMTC-P2-3	pT2NxM0	75	95%
GEMTC-P2-4	pT1b(m)N0M0	84	95%
GEMTC-P2-6	pT3N0Mx	72	95%
GEMTC-P2-7	pT2N0Mx	65	95%
GEMTC-P2-8	pT2N0Mx	72	95%
GEMTC-P2-9	pT1b(m)N0M0	71	95%
GEMTC-P2-10	pT3N0M0	79	95%
GEMTC-P2-11	pT3N1aMx	82	95%
GEMTC-P2-12	pT4bN1a	65	95%
GEMTC-P2-16	pT3N0M1	72	95%
GEMTC-P2-21	pT3N0Mx	76	95%
GEMTC-P2-22	pT2N1a	71	95%
GEMTC-P2-39	pT1bN0	72	95%

Table 5-7. Summary of number of targets and coverage of each patient's bespoke sequencing panel

5.4.6 Buffy coat DNA fragmentation

Buffy coat derived DNA is a comparatively large molecule compared to cell-free DNA, which is highly fragmented. As a source of germ-line DNA it can be used as a negative control for somatic mutation calling. However, it was necessary to fragment the DNA to approximately the same size as cfDNA in order for the library preparation process to be equitable across samples. This was performed using a Covaris Ultrasonicator. Two rounds of optimisation tests were performed. The first test had the settings outlined in Table 5-8 below.

Setting	A	B	C	D
Predicted average size (bp)	100	150	200	300
Peak incident power	175	175	175	175
Duty factor	10%	10%	10%	10%
Cycles per burst	200	200	200	200
Treatment time	340 secs	280 secs	120 secs	50 secs

Table 5-8. Initial fragmentation settings on Covaris Ultrasonicator

None of these settings achieved the desired size of 120-220bp as shown in the Tapestation output in Figure 5-6.

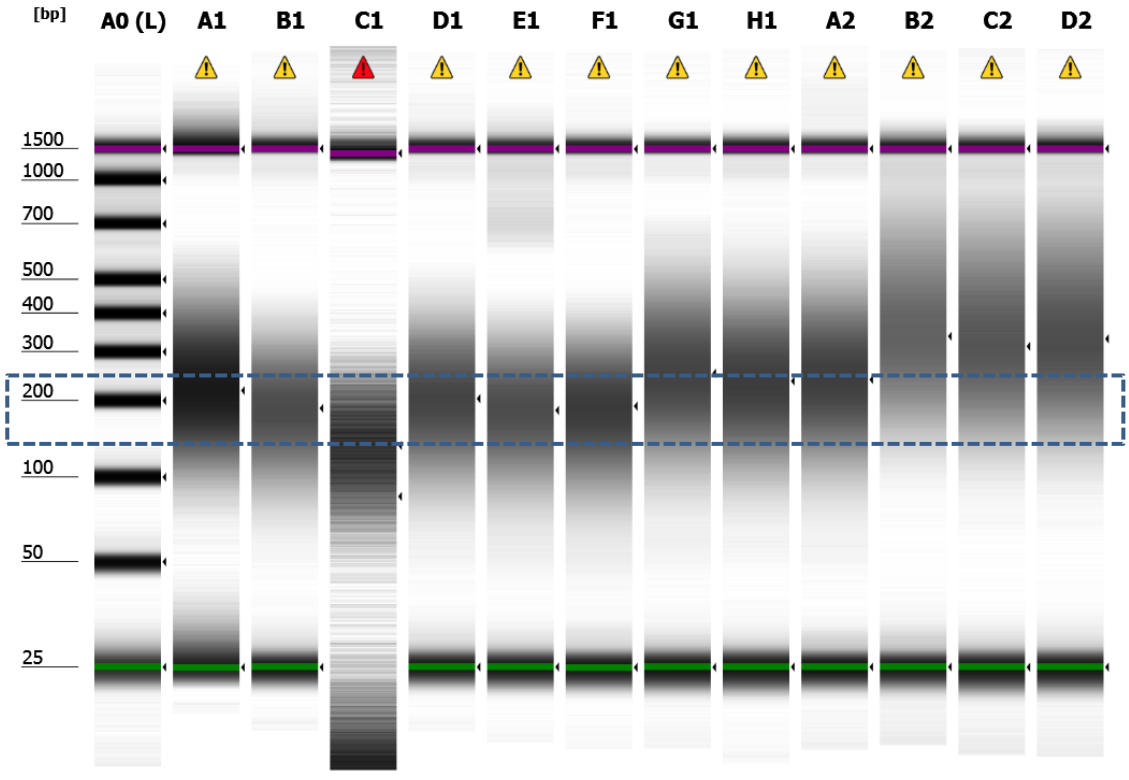


Figure 5-6. Tapestation gel output showing fragment sizes achieved from initial fragmentation settings. The desired fragment size was 120-220bp signified by the blue dashed box

The second optimisation test was run on modified settings as shown in Table 5-9 below.

Setting	A	B
Predicted average size (bp)	100	150
Peak incident power	75	75
Duty factor	15%	15%
Cycles per burst	1000	1000
Treatment time	500 secs	360 secs

Table 5-9. Second optimisation fragmentation settings on Covaris Ultrasonicator

The best average fragment size was achieved from setting A, and was therefore applied to all buffycoat samples. The resultant fragment sizes, tapestation gel bands and representative fragment size graph are shown in Figure 5-7 below.

5.4.7 Library preparation and quality control

Using the bespoke panels, sequencing libraries with ligated unique molecular identifiers (UMIs) were generated from the purified cfDNA extracted from plasma timepoints and matched fragmented buffycoat DNA for each patient. Where available this included baseline (pre-op), post-op, pre-RAI and post-RAI time-points. Mean DNA input was 45ng (range 9.4-190.1ng) and completed libraries were normalised to a final concentration of 4nM. The concentration of end sample library produced was proportional to the amount of input DNA. This is demonstrated in Figure 5-8 below, which compares the Bioanalyzer peak tracings from a completed library sample with 190.1ng of cfDNA input versus 25.8ng. The size of completed LiquidPlex library product was consistently around 180-200bp. All libraries (n = 64) were of sufficient quantity and quality to go on for high depth sequencing on an Illumina platform (NovaSeq SP 300 v1.5).

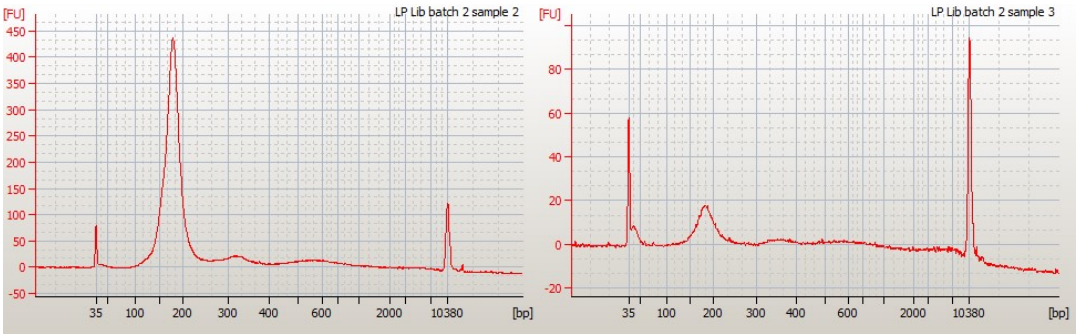
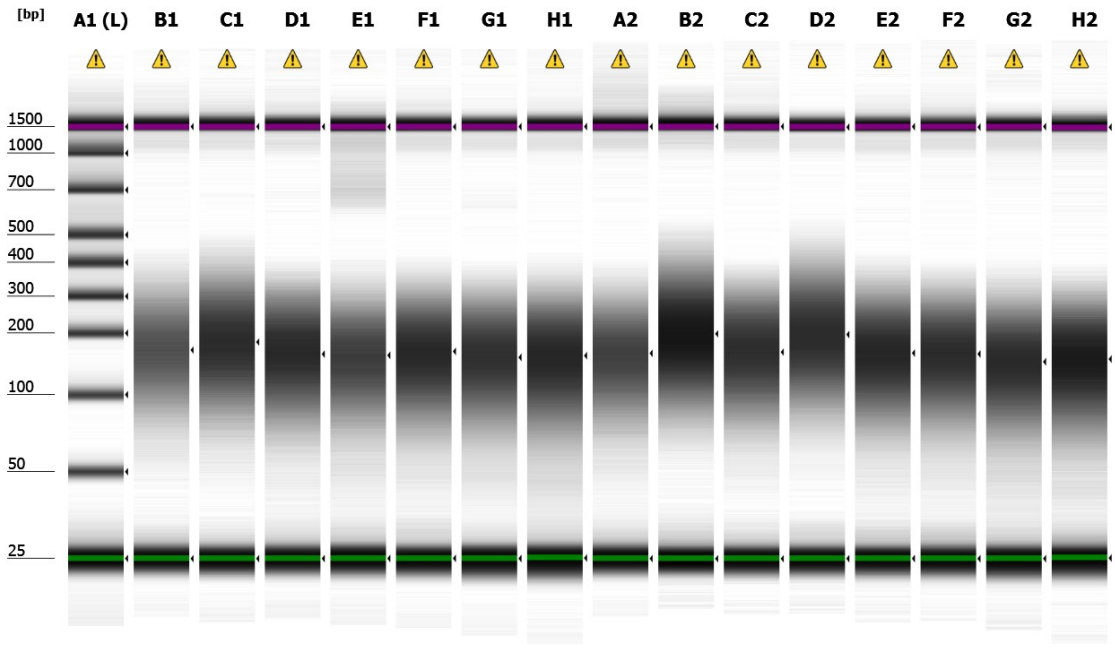
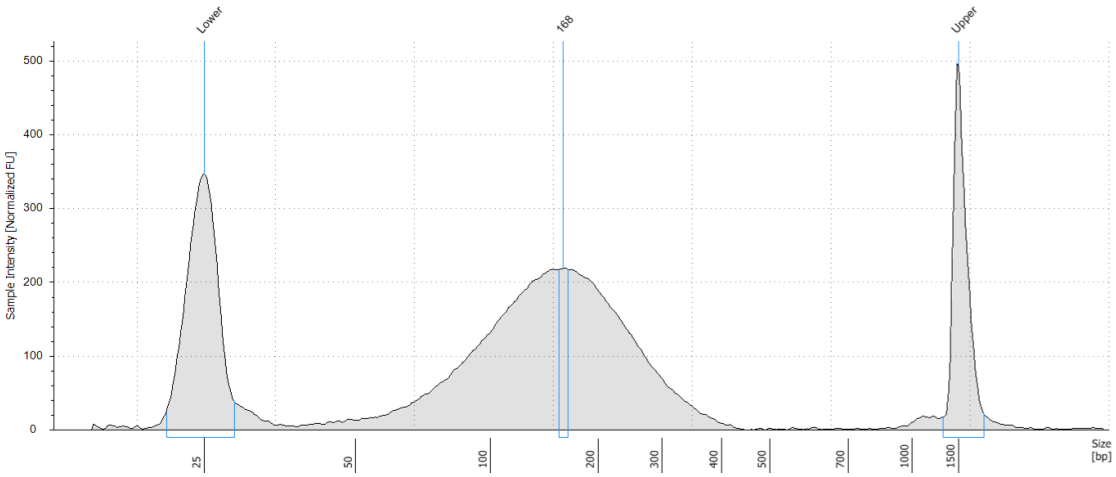


Figure 5-8. Examples of Agilent Bioanalyzer analysis of completed Liquidplex libraries. Completed library size is approximately 180-200bp. Amount of completed library (amplitude of peak tracing) was proportional to cfDNA input as shown in the tracing on the left with 190.1ng cfDNA input and right with 25.8ng cfDNA input.

A



B



C

Trial ID	Average size (bp)	Conc. (pg/ul)	Region molarity (pmol/L)	% of Total
GEMTC-P2-1	178	821	8560	92.6
GEMTC-P2-2	194	1220	11900	95.5
GEMTC-P2-3	171	1190	12900	94.0
GEMTC-P2-4	162	917	10500	86.0
GEMTC-P2-6	167	1160	13000	94.2
GEMTC-P2-7	162	1150	13200	94.3
GEMTC-P2-8	160	1240	14400	92.1
GEMTC-P2-9	164	862	9680	90.0
GEMTC-P2-10	210	1380	12300	94.2
GEMTC-P2-11	174	1090	11600	94.4
GEMTC-P2-12	210	1250	11300	95.8
GEMTC-P2-16	170	1220	13300	93.8
GEMTC-P2-21	166	1150	12700	92.8
GEMTC-P2-22	153	1150	13900	91.9
GEMTC-P2-39	156	1300	15400	93.1

Figure 5-7. Results of optimised fragmentation protocol on buffy coat gDNA. A – Tapestation gel output showing peak fragment concentration over the target range. **B** – Example of fragment size peaks showing normal distribution centred over desired fragment size. **C** – Table summarising the average fragment size,concentration and region molarity for all matched buffy coat gDNA samples.

5.4.8 Analysis of sequencing data and variant calling

The sequencing run generated 278.9 Gigabases of data with 88.9% Q30 score, 549,585,880 reads and 51.2% mean unique fragment on target percentage. As detailed in the methods section, the fastq files for each library underwent dual analysis and filtering through the ArcherDX analysis pipeline v6.2 as well as an in-house filtering pipeline. After manual curation of calls from both pipelines, 25 high confidence (outlier AF p value < 0.01) somatic variants calls were made in the 49 time points available for the 15 patients. The mean read depth for each of these calls was 7153 (SD ± 6721) and the median detected AF was 0.28% (range 0.02 – 12.13%).

5.4.9 Rate of ctDNA detection

From the whole cohort, ctDNA was detected in at least one time point in 6 out of 15 (40%) patients. This is summarised in Table 5-10 below.

Trial ID	Operation at entry to study	Pre-op blood sample	Post-op blood sample	Pre-RAI blood sample	Post-RAI blood sample	Histology	TNM stage	AJCC stage
GEMTC-P2-1	TT	UD	UD	N/A	UD	cPTC	pT3 N1a M0	1
GEMTC-P2-2	TT	POS	POS	N/A	POS	cPTC	pT3b N1a M0	2
GEMTC-P2-3	HT	UD	UD	UD	UD	cPTC	pT2 N0 M0	1
GEMTC-P2-4	TT	UD	UD	N/A	UD	fvPTC	pT1b N0 M0	1
GEMTC-P2-6	CT	UD	UD	N/A	UD	cPTC	pT3 N0 M0	1
GEMTC-P2-7	CT	UD	UD	N/A	UD	FTC	pT2 N0 M0	1
GEMTC-P2-8	CT	UD	UD	N/A	POS	cPTC	pT2 N0 M0	1
GEMTC-P2-9	CT	UD	UD	UD	UD	cPTC	pT1b N0 M0	1
GEMTC-P2-10	CT	UD	UD	N/A	UD	FTC	pT3 N0 M0	1
GEMTC-P2-11	TT	UD	UD	UD	UD	cPTC	pT3 N1a M0	1
GEMTC-P2-12	TT	POS	POS	N/A	POS	cPTC	pT4b N1a M0	4b
GEMTC-P2-16	TT	POS	UD	POS	N/A	FTC	pT3 N0 M1	4b
GEMTC-P2-21	HT	UD	POS	N/A	UD	cPTC	pT3 N0 M0	2
GEMTC-P2-22	HT	UD	UD	UD	POS	cPTC	pT2 N1a M0	2
GEMTC-P2-39	TT	UD	UD	UD	N/A	cPTC	pT1b N0 M0	1

Table 5-10. Summary of ctDNA detection in 15 DTC patients included in the study. TT – Total thyroidectomy, HT – Hemi-thyroidectomy, CT – Completion thyroidectomy, POS – positive detection, UD – undetectable, N/A – blood sample timepoint not available

Circulating tumour DNA was detected in the pre-op blood sample in 3 out of 15 (20%) and in at least one time point in 5 out of 10 (50%) surgically naïve patients (ie. not having completion surgery on entry to the study). Five patients had already underwent diagnostic hemi-thyroidectomy on entry into the study, and of these only 1 had disease after their second surgery in the completion thyroid lobe, which was a sub centimetre foci. Surprisingly, this patient (GEMTC-P2-8) had detectable mutant *BRAFV600E* (AF 0.7%, AF outlier p value < 0.0001) in the post-RAI blood samples. She remains clinically free from detectable residual or recurrent disease 1 year after treatment completion.

In sub-cohort analysis, if excluding the 4 patients with no clinically identifiable disease from recruitment to the end of the study (recruited after diagnostic hemi-thyroidectomy and had no disease in their completion thyroid lobe and no residual or recurrent disease), then the overall proportion of the remaining patients (n = 11) with at least one time point with ctDNA detected was 54.5% (6 out of 11).

Circulating tumour DNA was detected in tumours from T2 to T4 and N0 to N1 disease, but patients with AJCC stage 2 and above all had detectable ctDNA in at least 1 time point (see Figure 5-9). Four out of 15 patients had persistent disease at 12 month follow up and all but one had detectable ctDNA before and after surgical treatment.

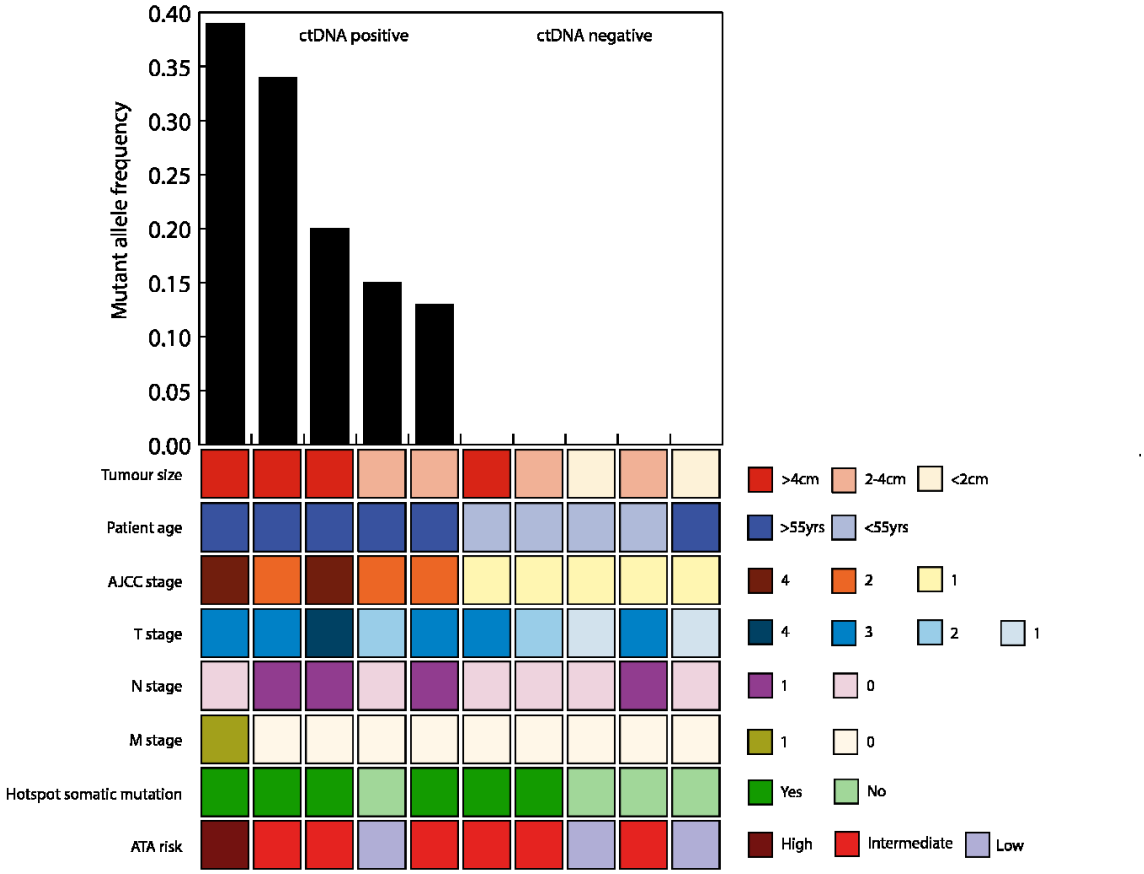


Figure 5-9. Association between ctDNA mutant allele frequency and histopathological characteristics. The 10 treatment naïve patients (index surgery was hemi-thyroidectomy or total thyroidectomy) were sorted based on the allele fraction of the somatic mutation detected in the primary tumour DNA that was also detected in plasma ctDNA. There appears to be a trend of higher allele frequency and more aggressive features. All tumours with detectable ctDNA were AJCC stage 2 or above. The status for each variable is given to the right of the graph.

5.4.10 Variant allele frequency and extent of treatment

The 3 patients with detectable ctDNA in the baseline pre-treatment blood sample demonstrated changes in mean variant allele frequency (the averaged allele frequency of all detected variants in each time-point) in relation to extent of treatment and residual disease burden. Patient GEMTC-P2-2 had a 52mm PTC with level 6 nodal involvement and extra-thyroidal extension into the adjacent strap muscles of the neck. We observed how surgical treatment (total thyroidectomy and level 6 neck dissection) resulted in a reduction in the majority of patient specific variants, but many remained detectable at a lower level (see Figure 5-10-A below). Further, whilst initially undetectable at baseline, *BRAFV600E* mutation became detectable in the post-operative period and the VAF continued to rise despite 3.7Gbq of radioactive iodine ablation. In contrast, thyroglobulin levels in this patient fell to <1 µg/L in response to surgical and adjuvant RAI treatment (data not shown). On clinical follow up, this patient had an abnormal level 4 lymph node discovered 12 months after completion of treatment, which was highly suspicious for residual disease but unfortunately was not amenable for confirmatory needle biopsy. This case further demonstrates the clinical utility of ctDNA as a biomarker of residual disease, potentially superseding the sensitivity of conventional markers such as thyroglobulin and also could confirm the presence of malignant disease when biopsy is unavailable.

Two patients had clinically detectable residual disease after primary surgical treatment. Patient GEMTC-P2-12 had a 55mm PTC with multifocal lympho-vascular invasion and localised spread into the thyroid cartilage and larynx. In this patient the post total thyroidectomy time-point showed a large reduction in ctDNA (shown in Figure 5-10-B), however the majority of variants remained detectable in keeping with the residual disease in the larynx. This patient went on to have further surgery and therapy dose RAI (5.5Gbq). Further long term follow up may reveal if the detectable ctDNA in her post-RAI blood sample was an early predictor of recurrent disease. Patient GEMTC-P2-16 had a 55mm FTC with capsular and vascular invasion, but also distant metastatic disease on entry to the study. Interestingly, despite such aggressive disease, VAF of detected variants were quite low at baseline blood testing and the commonly FTC associated *NRAS* Q61R mutation only became detectable after surgery

(Figure 5-10-C). Levels of ctDNA increased rapidly leading up to her treatment with therapy dose RAI and this was matched with concurrent rises in serum thyroglobulin. Unfortunately, no further time points were available for this patient after RAI treatment.

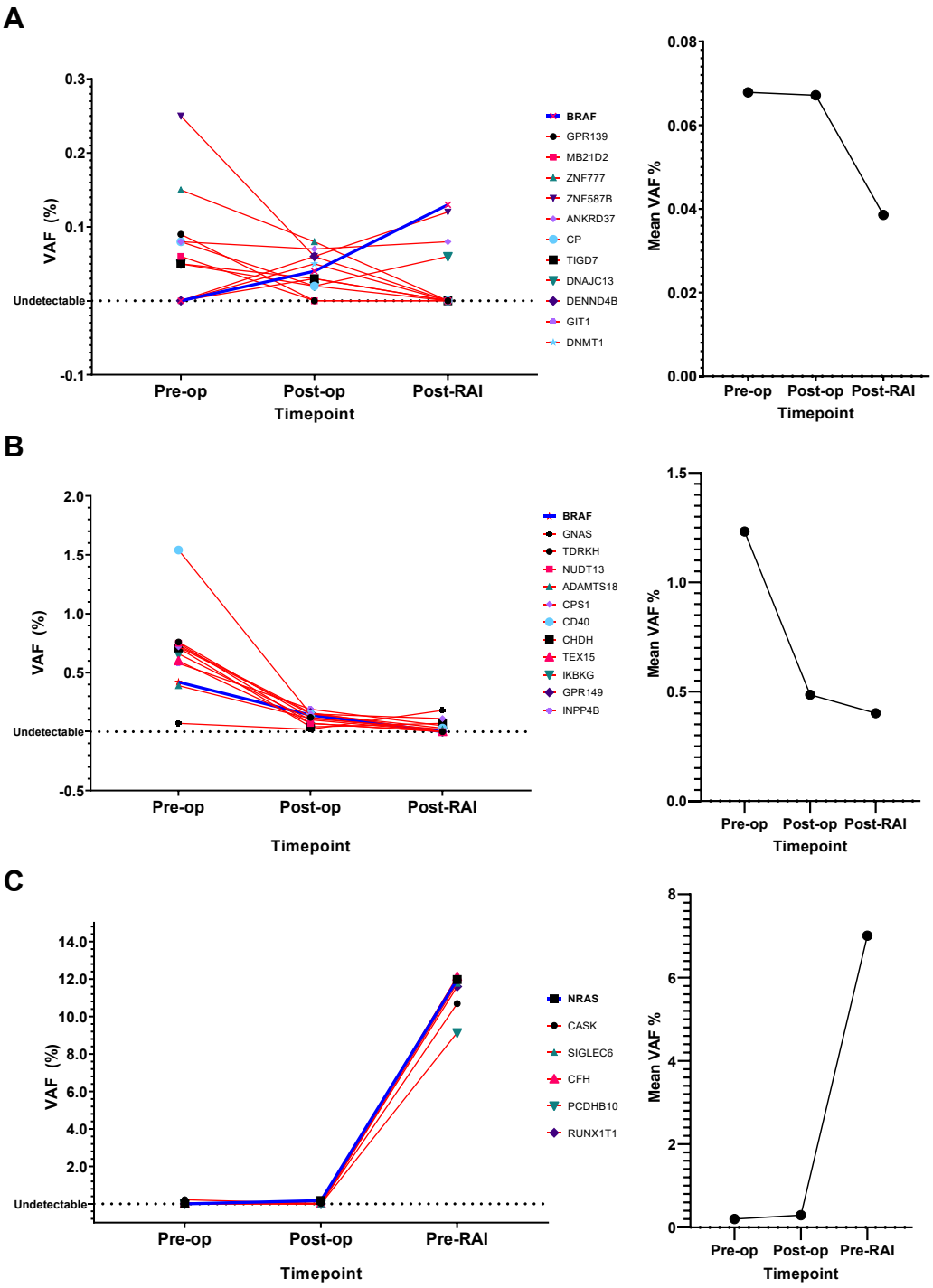


Figure 5-10. Mutation tracking in patients with detectable ctDNA in the pre-op timepoint. Line graph on the left show individual variant VAF trends over serial timepoints. Hotspot mutations are shown in dark blue line. Line graphs on the right show the overall trends of mean VAF of all detected variants. A - Patient GEMTC-P2-2, B - Patient GEMTC-P2-12, C - Patient GEMTC-P2-16.

5.5 Discussion

The search for sensitive and specific circulating biomarkers of disease in thyroid cancer is critical for all aspects of clinical management. The detection, prognostication and surveillance of DTC has become the focus of attention as the incidence of new DTC diagnoses and the number of post treatment survivors continues to rise at a rapid rate worldwide.

In this prospective, multi-centre study, we recruited 47 patients with a diagnosis of DTC and extracted total DNA from their respective index tumours as well as cfDNA from peri-treatment blood samples to study the detectability of ctDNA. We compared two different approaches and found that traditional targeted high depth sequencing with singleplex droplet digital PCR (ddPCR) assays detected ctDNA in only a minority of patients (10%), whilst the rate of detection increased when using patient specific bespoke sequencing panels informed by whole exome sequencing (54.5%). The latter approach improves the clinical viability of sequencing techniques for the detection of minimal residual disease in AJCC stage 2 disease and above, with all patients with detectable ctDNA in post-operative blood samples having T3+ or M1 disease. This study extends current paradigms of precision medicine that have been shown to be fruitful in a number of other solid cancers.

5.5.1 Detection of variants using a thyroid specific targeted panel versus WES

The rationale for utilising a targeted sequencing panel to identify somatic mutations in thyroid cancer originates from the observed low tumour mutational burden in most DTCs and the cost-effectiveness per sample of the sequencing technique. Coupled with a highly sensitive assaying technique such as ddPCR should, in theory, detect very small fractions of ctDNA from total cfDNA extracted from plasma.

For tumour DNA somatic variant detection, we used a previously validated thyroid specific panel called “ThyMa” which consisted of 336 individual amplicons covering exons in 28 genes previously shown to harbour somatic mutations in DTC. Previously this panel detected variants in 93% of advanced thyroid tumours, which included DTC, medullary, poorly differentiated and

anaplastic subtypes. In sub-analysis of this previous cohort including only DTC tumours, the detection rate of 1 or more variants was 46%.
Our inclusion/exclusion criteria were designed to achieve a cohort that reflects a broader representation of DTC tumours, including a range of disease stages and varying histopathological phenotypes. As is the typical natural presentation of DTC, the cohort has predominantly early stage disease. This is clearly reflected in 90% of the cohort with stage T3 disease or below and 66% were N0.

Whilst the 2014 TCGA study cohort of 496 PTC tumours described an oncogenic driver in 96.5%, this included the detection of gene fusions and somatic copy-number aberrations and did not include FTC tumours. In our cohort, we identified one or more somatic mutation(s) in 27 out of 47 (57.4%) DTCs, with two or more mutations found in 8 cases. The targeted sequencing panel was limited in its ability to only reliably detect SNVs and small indels in the loci covered by the included amplicons. We therefore deduce that the oncogenic drivers for the tumours with no somatic variants detected with ThyMa were either SNVs not covered in the panel or structural aberrations that require other techniques to detect. When comparing the cohort with variants versus without, those tumours with variants detected tended to be larger (higher proportion of T3+), more advanced tumours with higher proportions of aggressive histopathological features such as vascular invasion and extra-thyroidal extension. There was no obvious TNM cut off, as the full range of TNM staging was present in both cohorts with and without variants.
Aligned to other previous studies on DTC tumours, *BRAF p.V600E* was the most frequently detected variant; found in 78.3% of the PTC cohort. It was exclusive to PTC with none found in any FTC tumour. The most common variants detected in the FTC cohort were *NRAS p.Q61R* and *TERT* promoter *p.C228T*. A notable observation was that despite tumour purities being high (subjectively between 70-95% purity based on microscopy) the allele frequencies of these driver mutations were quite variable. The allele frequency range of the *BRAF* V600E variant for example, was between 0.06 – 0.44. This suggests heterogeneity in the clonal prevalence of tumour cells harbouring the mutation, despite homogeneity of microscopic appearance. Previous observations have suggested marked variation between pathologically

assessed and bioinformatically inferred tumour content (17-19). Intra-tumour heterogeneity presents an analytical conundrum and is hypothesised to enable tumours with adaptability to evade therapies and the immune system allowing disease progression and metastasis (20). Whilst intra-tumoural heterogeneity can be broadly divided into spatial and temporal, a number of techniques are being investigated to mitigate this problem, including multi-site sampling (21, 22), but also systemic sampling through liquid biopsy. The latter is the focus of this study and is dependent on sensitively identifying these circulating markers with sufficient specificity for the tumour being investigated.

Broader sequencing techniques such as WGS, WES or RNAseq are becoming more financially accessible as the technology improves and becomes less expensive due to scalability. We subjected DNA extracted from a subset of 15 tumours to dual sequencing, first with the ThyMa targeted panel and then again with WES. Seven of these tumours had no variant detected on ThyMa, but had multiple exonic variants detected on WES. In all 7 cases, the variants detected were non-canonical mutations that could be considered patient specific passenger mutations. The proportion of tumours with trackable somatic mutations improved from 54.7% using a targeted approach to 100% using WES. We have a high degree of confidence that these non-canonical mutations are not the result of sequencing error, as all samples were sequenced at sufficient depth, the variant calls were filtered against the gnomad exome and COSMIC databases to exclude non-cancer related variants and also the tumour DNA was co-sequenced with matched buffy coat DNA to exclude germline mutations. We also found very close concordance between targeted sequencing and WES in terms of canonical somatic variant calls and their respective allele fractions. We have not investigated what role these non-canonical mutations may have in the pathogenesis of DTC, if any, but if these somatic variants are only found from the DNA in the primary tumour, then they still remain viable as circulating biomarkers specific for the patient's individual tumour. We have therefore demonstrated the benefits of performing broader genotyping investigations like WES over targeted sequencing in DTC. Additionally, the cost per sample is steadily decreasing to the point where WES may become the preferred genotyping technique when attempting to identify

somatic mutations from tumour DNA for the purpose of identifying putative targets for ctDNA detection in future studies.

5.5.2 Improving sensitivity of ctDNA detection with patient specific sequencing panels

The inherent challenge with utilising ctDNA detection in monitoring minimal residual disease, is the capturing and identification of the diminutive proportion of tumour derived fragments from the total cfDNA in a given sample. Technologies for ctDNA analysis can be broadly divided into two categories; PCR based and NGS based methods. Enhanced detection down to 1 mutant allele in 10,000 or more wild-type alleles can be obtained using methods such as bead-based digital PCR in emulsions (BEAMing) or ddPCR. However, the identification of multiple rare or patient tumour specific mutations remains elusive with these methods due to the inability to test for many loci in the same assay. Whilst further technological advances are inevitable, currently ddPCR multiplexing is limited to combinations of fluorescent channels available and stand at approximately 10 targets (23). The advantage of NGS lies in the potential ability to canvas large sections of the genome in parallel, being able to detect somatic mutations, aneuploidy, copy number and methylation variations that all serve as markers to identify ctDNA. The size of the regions and the number of times each region is sequenced is limited only by sample input and financial constraints. Historically, the main disadvantage of NGS is that sensitivity was limited by sequencing artefacts and polymerase error, rendering it challenging to separate true variant calls from background noise.

The first part of this study used the traditional approach of designing personalised ddPCR assays to detect ctDNA in plasma, based on somatic mutational targets found in the primary tumour. This has been shown to be highly effective in minimal residual disease detection and tracking in breast cancer (10) as well as recently in advanced thyroid cancer (12). In our cohort of DTCs, 11% (3 of 27) of patients had detectable ctDNA in at least 1 plasma time point. These 3 patients all had tumours larger than 4cm, with 2 out of the 3 having distant metastases. Two of these patients had WI-FTCs with detectable circulating *NRASQ61R,* which has not previously been published.

With an 11% ctDNA detection rate, we could potentially conclude that early DTC tumours less than 4cm in size do not release ctDNA in sufficient quantity to be detectable, unless regionally or distantly metastatic. The published literature on ctDNA detection in thyroid cancer is mixed and confined to predominantly PTC and *BRAF* V600E, with some studies demonstrating ctDNA detection in patients with tumours of varying size, extent and disease stage (24-27), as has been shown previously in melanoma (28). However our results appear to parallel studies, such as Cradic et al who reported 19% of their *BRAFV600E* positive PTC cohort having the same variant detected in their blood using RT-PCR (29), or Kim et al who reported 6.1% detection, also using PCR methodology, with all ctDNA positive patients having lateral lymph node and lung metastases (30). Chuang et al reported a higher proportion (60%) of their *BRAFV600E* positive PTC cohort having detectable ctDNA from serum, but this study had low numbers of patients; 14 PTCs, with 5 having mutant BRAF found in the index tumours (31).

In order to improve sensitivity through multi-variant assaying and to subsequently determine if ctDNA detection rate could be increased, we utilised new library preparation technology in the form of Anchored Multiplex PCR (AMP) and unique molecular indexes (UMIs) to significantly reduce the error rate in calling low abundance variants and improve sensitivity. The technology incorporates the ligation of half functional adapters with unique 8-mer UMIs followed by 2 rounds of nested PCR with sequencer specific molecular barcodes and gene specific primers to allow for deduplication of reads and enrich for areas of interest dictated by personalised panel designs respectively. This method has been shown to be highly effective in low input or degraded DNA as is found from FFPE or cfDNA and increases the confidence in calling low abundance mutant alleles. A study by Namlos et al have shown the benefits of AMP in highly sensitive detection of single nucleotide variations and small indels in gastrointestinal stromal tumours (32). They reported a median detected AF of 4.6% (range 0.07 - 48.1%) for indels and 1.4% (range 1.1 - 1.5%) for SNVs.

In our study, we have shown that using bespoke patient specific panels, AMP library preparation and high depth sequencing increases the rate of ctDNA detection from 11% to 40% compared to targeted sequencing with a pan-thyroid

panel and ddPCR, in our cohort of DTC patients. If analysing only patients with clinically identifiable disease at study recruitment, then this proportion increases to 54.5%. Our median detected AF was 0.28% (range 0.02 – 12.13%) and all variant calls from cfDNA matched the loci and base changes found in their respective index tumours. Importantly, our findings suggest that ctDNA is present in the circulation in early thyroid cancer, but detection is more likely when using multiple variants rather than just a single hotspot mutation.

5.5.3 Viability of ctDNA in differentiated thyroid cancer

The mechanisms by which ctDNA is released into the circulation is hypothesised to be secondary to tumour cell death, apoptosis or direct secretion from tumour cells (33, 34). Therefore patients with large bulky disease that becomes centrally necrotic due to outstripping its blood supply or which spread locally and to distant sites would be most likely to have detectable ctDNA. In our study, 4 out of 6 patients (66.7%) with detectable ctDNA had either nodal or distant metastatic disease and all had AJCC stage 2 disease or above. We have shown that if detectable, ctDNA concentration (as quantified by mutant AF of detected variants) declines in response to increasing treatment modalities (surgery followed by RAI), but is still detectable when minimal residual disease is present. This adds evidence at a molecular level of the efficacy of adjuvant RAI treatment in treating minimal residual disease. In the era of treatment de-escalation, where the results of the IoN trial may soon result in fewer patients having adjuvant RAI, ctDNA detection could become an independent prognostic factor that would warrant adjuvant treatment. When focusing on the patients with detectable pre-op ctDNA, we showed evidence suggestive of sub-clonal treatment selection pressures with certain variant AFs falling after surgery, then subsequent undetectable variants becoming detectable with increasing AF in the presence of residual disease. This was despite concurrent reduction of thyroglobulin, the conventional surrogate marker of residual thyroid tissue. These observations add more evidence for the presence of initial tumour heterogeneity and subsequent sub-clonal evolution, which is exciting given the often quoted benefits of liquid biopsy. Our study adds further evidence for the role of ctDNA in monitoring disease evolution and directs further research into treatment resistance.

5.5.4 Limitations and further work

Building on the previous work in advanced thyroid cancer by Allin et al (12), this was a pilot study to investigate if ctDNA was detectable in early DTC. We chose to focus on DTC as it is the most prevalent subtype of thyroid cancer and most likely to prospectively recruit sufficient numbers within the study timeframe. We optimised recruitment by running a multi-centre study in two major London tertiary referral centres. Whilst we recruited 58 patients in total, the study was limited by the proportion of patients with a detectable somatic mutation that could be assayed in the plasma. As described above, this was improved with WES, but the number of patient tumours submitted for WES was limited due to time and cost restraints. This study is therefore hypothesis generating and further studies with DTC and ctDNA will have to take this into account and incorporate techniques that can look for other genomic aberrations such as fusions and methylation in order to capture a wider proportion of patients.

Additionally the number of time points in this study were limited to a maximum of 4 peri-treatment opportunities, in order to keep inconvenience to participating patients to a minimum and collect research bloods at routine standard of care blood draws. Previous studies in ctDNA tracking have demonstrated the benefit of regular and frequent blood draws, particularly in the follow up period, to provide granular data and precise tracking of variant fluctuations. It would be recommended that future studies incorporate multiple blood draws in the peri-treatment time frame, including multiple follow up blood draws after treatment completion, in order to demonstrate if ctDNA detection can sensitively predict recurrent disease.

5.6 Conclusion

This prospective multi-centre pilot study aimed to assess the efficacy of targeted sequencing with a thyroid specific panel coupled with singeplex ddPCR assays compared to whole exome sequencing with bespoke patient specific sequencing panels in the detection of ctDNA in DTC tumours. We found that detection in at least one peri-treatment timepoint increased from 10% to 54.5% in our cohort of patients. The findings in this study support the use of personalised sequencing panels for ctDNA detection in AJCC stage 2 DTC and

above. Further multicentre long term (> 5 year) clinical prospective trials are required to validate these findings and determine if detection of ctDNA predicts disease free survival or overall survival.

5.7 References

1. Christensen E, Nordentoft I, Vang S, Birkenkamp-Demtroder K, Jensen JB, Agerbaek M, et al. Optimized targeted sequencing of cell-free plasma DNA from bladder cancer patients. Sci Rep. 2018;8(1):1917.
2. Comino-Mendez I, Turner N. Predicting Relapse with Circulating Tumor DNA Analysis in Lung Cancer. Cancer Discov. 2017;7(12):1368-70.
3. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013;368(13):1199-209.
4. O'Leary B, Hrebien S, Morden JP, Beaney M, Fribbens C, Huang X, et al. Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. Nat Commun. 2018;9(1):896.
5. Reinert T, Scholer LV, Thomsen R, Tobiasen H, Vang S, Nordentoft I, et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. Gut. 2016;65(4):625-34.
6. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6(224):224ra24.
7. Merker JD, Oxnard GR, Compton C, Diehn M, Hurley P, Lazar AJ, et al. Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. J Clin Oncol. 2018;36(16):1631-41.
8. Zill OA, Banks KC, Fairclough SR, Mortimer SA, Vowles JV, Mokhtari R, et al. The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients. Clin Cancer Res. 2018;24(15):3528-38.
9. O'Leary B, Hrebien S, Beaney M, Fribbens C, Garcia-Murillas I, Jiang J, et al. Comparison of BEAMing and Droplet Digital PCR for Circulating Tumor DNA Analysis. Clin Chem. 2019;65(11):1405-13.
10. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. 2015;7(302):302ra133.
11. Coombes RC, Page K, Salari R, Hastings RK, Armstrong A, Ahmed S, et al. Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. Clin Cancer Res. 2019;25(14):4255-63.
12. Allin DM, Shaikh R, Carter P, Thway K, Sharabiani MTA, Gonzales-de-Castro D, et al. Circulating tumour DNA is a potential biomarker for disease progression and response to targeted therapy in advanced thyroid cancer. Eur J Cancer. 2018;103:165-75.
13. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex PCR for targeted next-generation sequencing. Nat Med. 2014;20(12):1479-84.

14. Gilani A, Donson A, Davies KD, Whiteway SL, Lake J, DeSisto J, et al. Targetable molecular alterations in congenital glioblastoma. J Neurooncol. 2020;146(2):247-52.
15. Ko TK, Lee E, Ng CC, Yang VS, Farid M, Teh BT, et al. Circulating Tumor DNA Mutations in Progressive Gastrointestinal Stromal Tumors Identify Biomarkers of Treatment Resistance and Uncover Potential Therapeutic Strategies. Front Oncol. 2022;12:840843.
16. Szurian K, Kashofer K, Liegl-Atzwanger B. Role of Next-Generation Sequencing as a Diagnostic Tool for the Evaluation of Bone and Soft-Tissue Tumors. Pathobiology. 2017;84(6):323-38.
17. Haider S, Tyekucheva S, Prandi D, Fox NS, Ahn J, Xu AW, et al. Systematic Assessment of Tumor Purity and Its Clinical Implications. JCO Precis Oncol. 2020;4.
18. Lou S, Zhang J, Yin X, Zhang Y, Fang T, Wang Y, et al. Comprehensive Characterization of Tumor Purity and Its Clinical Implications in Gastric Cancer. Front Cell Dev Biol. 2021;9:782529.
19. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015;6:8971.
20. Saunders NA, Simpson F, Thompson EW, Hill MM, Endo-Munoz L, Leggatt G, et al. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. EMBO Mol Med. 2012;4(8):675-84.
21. Erramuzpe A, Cortes JM, Lopez JI. Multisite tumor sampling enhances the detection of intratumor heterogeneity at all different temporal stages of tumor evolution. Virchows Arch. 2018;472(2):187-94.
22. Jie W, Bai J, Yan J, Chi Y, Li BB. Multi-Site Tumour Sampling Improves the Detection of Intra-Tumour Heterogeneity in Oral and Oropharyngeal Squamous Cell Carcinoma. Front Med (Lausanne). 2021;8:670305.
23. Zhong Q, Bhattacharya S, Kotsopoulos S, Olson J, Taly V, Griffiths AD, et al. Multiplex digital PCR: breaking the one target per color barrier of quantitative PCR. Lab Chip. 2011;11(13):2167-74.
24. Sato A, Tanabe M, Tsuboi Y, Niwa T, Shinozaki-Ushiku A, Seto Y, et al. Circulating Tumor DNA Harboring the BRAF(V600E) Mutation May Predict Poor Outcomes of Primary Papillary Thyroid Cancer Patients. Thyroid. 2021;31(12):1822-8.
25. Lubitz CC, Zhan T, Gunda V, Amin S, Gigliotti BJ, Fingeret AL, et al. Circulating BRAF(V600E) Levels Correlate with Treatment in Patients with Thyroid Carcinoma. Thyroid. 2018;28(3):328-39.
26. Lubitz CC, Parangi S, Holm TM, Bernasconi MJ, Schalck AP, Suh H, et al. Detection of Circulating BRAF(V600E) in Patients with Papillary Thyroid Carcinoma. J Mol Diagn. 2016;18(1):100-8.
27. Pupilli C, Pinzani P, Salvianti F, Fibbi B, Rossi M, Petrone L, et al. Circulating BRAFV600E in the diagnosis and follow-up of differentiated papillary thyroid carcinoma. J Clin Endocrinol Metab. 2013;98(8):3359-65.
28. Shinozaki M, O'Day SJ, Kitago M, Amersi F, Kuo C, Kim J, et al. Utility of circulating B-RAF DNA mutation in serum for monitoring melanoma patients receiving biochemotherapy. Clin Cancer Res. 2007;13(7):2068-74.
29. Cradic KW, Milosevic D, Rosenberg AM, Erickson LA, McIver B, Grebe SK. Mutant BRAF(T1799A) can be detected in the blood of papillary thyroid carcinoma patients and correlates with disease status. J Clin Endocrinol Metab. 2009;94(12):5001-9.

30. Kim BH, Kim IJ, Lee BJ, Lee JC, Kim IS, Kim SJ, et al. Detection of plasma BRAF(V600E) mutation is associated with lung metastasis in papillary thyroid carcinomas. Yonsei Med J. 2015;56(3):634-40.
31. Chuang TC, Chuang AY, Poeta L, Koch WM, Califano JA, Tufano RP. Detectable BRAF mutation in serum DNA samples from patients with papillary thyroid carcinomas. Head Neck. 2010;32(2):229-34.
32. Namlos HM, Boye K, Mishkin SJ, Baroy T, Lorenz S, Bjerkehagen B, et al. Noninvasive Detection of ctDNA Reveals Intratumor Heterogeneity and Is Associated with Tumor Burden in Gastrointestinal Stromal Tumor. Mol Cancer Ther. 2018;17(11):2473-80.
33. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. 2016;35(3):347-76.
34. Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. Cell Res. 2014;24(6):766-9.

Chapter 6 Conclusions and future work

Genomics based biomarkers have the potential to revolutionise the way we diagnose and treat patients with thyroid cancer. In parallel with existing clinical, radiological and histopathological techniques, these biomarkers can inform the entire patient pathway including diagnosis, prognosis, risk stratification and disease monitoring. With the progressively increasing literature base, current and planned prospective clinical trials and scientific laboratory collaborations, a paradigm shift is on the horizon whereby tumour-specific genomic signatures will be routinely measured as an invaluable tumour characterising factor.

This thesis aimed to further our understanding of key areas of differentiated thyroid cancer (DTC) management; both prognostication in the treatment phase of patient care and residual disease detection in the follow up phase. The observational clinical study that underpinned the sample and data collection for this thesis was divided into retrospective and prospective parts. The retrospective study aimed to describe a genomic signature that predicts the presence of vascular invasion and other high risk histopathological features. The prospective study aimed to investigate the limit of circulating tumour DNA (ctDNA) detection in the peri-treatment and short term follow-up period in early DTC using the most recent molecular techniques. The aims of the thesis were met as summarised below.

6.1 Gene mutation signature for vascular invasion in thyroid cancer

This study used a thyroid specific targeted sequencing panel to perform multi-mutational analysis on 118 DTC tumours. This mutational hotspot panel was previously validated on a cohort of advanced thyroid cancer tumour samples and covered chromosomal loci that have been shown to harbour somatic mutations associated with thyroid cancer in large population databases.

We showed that 62% of our tumour cohort contained at least one somatic mutational driver, confirming previous published findings that tumorigenesis in DTC is driven not only by single nucleotide variations or small insertions and

deletions, but also by other genomic aberrations. These could include fusions and copy number variations, which require other techniques to detect and would be an avenue for further studies, via modification of the existing thyroid specific amplicon based sequencing panel used in this study.

We found that in our cohort of patients, a DTC tumour with two or more somatic mutational variants is significantly associated with aggressive pathological features such as disease recurrence, vascular invasion or distant metastatic disease, independent of histological subtype or tumour size. Further, in those samples with multiple variants, they were most likely to harbour a *TERT* promoter (*TERTp*) mutation as a concurrent variant. A critical finding from this study was that *TERTp* mutations are important prognostic markers in DTC, demonstrating a link between the *TERTp* c.1-124C>T variant and more aggressive pathology, with statistically significant associations with the presence of vascular invasion, disease recurrence and distant metastases. By combining the mutational status of the *TERTp* region with key hotspot genes *BRAF* and *RAS* (H,N,K subtypes), a somatic mutational signature is achieved that can correctly predict the presence of vascular invasion 70% of the time.

The findings from this study suggest that a limited targeted sequencing panel could be an important adjunct in the characterisation of DTC and that *TERTp* mutations confer prognostic value that can assist in clinical decision making. This could be important when considering further diagnostic studies such as analysing DNA extracted from FNA needle biopsies within the thyroid nodule diagnostic phase. A targeted sequencing approach is a relatively quick, cheap and sensitive method of identifying these somatic mutations from DNA extracted from degraded FFPE tissue.
We would recommend that future studies utilising targeted sequencing techniques to characterise DTC tumours would benefit from the addition of assays that can cover structural genomic aberrations such as fusions and copy number variations. This would greatly increase the number of samples with identifiable mutation drivers and increase the proportion of samples available for associative analysis. Whilst it is possible to expand the coverage of the thyroid specific panel (ThyMa) used in this study with all genomic loci currently published in thyroid cancer, the frequency with which other somatic variants

occur is very low and would arguably provide marginal gains in terms of detection. Further experiments with larger sample sizes would be recommended to validate the predictive accuracy of the proposed somatic mutation signature from this study.

6.2 Differential expression of miRNA associated with vascular invasion in thyroid cancer

To further extend the genomic characterisation of our DTC cohort, we studied the differential expression of miRNAs using two different techniques, qPCR and Nanostring. We showed that differential expression of miRNAs are potentially a powerful tool to differentiate between cancer and normal, papillary versus follicular subtypes and importantly the presence or absence of high-risk histopathological features such as vascular invasion and extra-thyroidal extension. All of these metrics have a role in the diagnostic patient pathway and could assist in areas of diagnostic ambiguity such as in the indeterminate nodule.

We have highlighted in this study the importance of identifying a tissue specific endogenous reference species when studying differential miRNA expression. The traditional surrogate molecules like snoRNAs and GAPDH have been used as endogenous references, but in our DTC cohort they exhibited very high levels of expressional variability, which has a significant downstream effect on calculated fold change. Using two orthogonal techniques (both qPCR and Nanostring), we showed that for our DTC and normal thyroid cohorts, hsa-miR-423-5p and hsa-miR-93-5p are the most stably expressed miRNA with the least amount of expressional variation. These could potentially be reference miRNA candidates specific to thyroid. It is recommended in future studies involving miRNA differential expression, that reference species be identified and justified, particularly when a limited number of miRNAs are being measured. In the context of large canvassing experiments where many hundreds of miRNA species are being measured, then global mean normalisation is sufficient.

The key miRNA species that show significant dysregulation in our DTC cohort were hsa-miRs 146a-5p, 204-5p and 221-3p and that in combination with the

somatic mutation data from the prior experiment, a more comprehensive genomic signature containing the dysregulation status of hsa-miRs 146a-5p, 204-5p and 221-3p with the somatic mutation status of *BRAF*, *RAS* and *TERTp* was described. This signature formed a predictive model that correctly identified DTC samples with vascular invasion with 78% sensitivity and 79% specificity, with a positive likelihood ratio of 3.64. This model was also able to correctly identify DTC samples with at least one high risk histopathological feature (VI, ETE or LN invasion) with a positive predictive value of 87.5% and negative predictive value of 82.0%. The performance of this predictive model is comparable to the accuracy of many routinely used tests in current clinical practice. It has the potential to add a layer of prognostication prior to definitive surgical treatment, which could help clinicians and patients mutually decide on extent of treatment. Further studies with larger DTC sample sizes to validate these findings are recommended, with subsequent prospective interventional clinical trials to demonstrate clinical utility.

6.3 Detection of ctDNA in early thyroid cancer

This novel study aimed to identify the limits to which circulating tumour DNA (ctDNA) can be used as a marker of disease in DTC. By performing multi-mutational analysis to characterise the mutational landscape of our DTC tumours with both targeted panel sequencing and whole exome sequencing (WES) we were able to demonstrate that WES was more effective at identifying putative mutational targets with which ctDNA can be identified. Many of the identified somatic targets may not have been involved in the primary tumorigenic process, but still serve as patient tumour specific markers. This ultimately translates to a higher rate of ctDNA detection when using the latest sequencing techniques that incorporate Anchored Multiplex PCR (AMP) and unique molecular indexes (UMIs) into the library preparation process.

We showed in this study that traditional targeted high depth sequencing with singleplex droplet digital PCR (ddPCR) assays detected ctDNA in only a minority of patients (10%), whilst the rate of detection increased when using patient specific bespoke sequencing panels informed by WES (54.5%). The latter approach improves the clinical viability of sequencing techniques for the

detection of minimal residual disease. Based on detection rates analysed against clinico-pathological features, we would recommend ctDNA detection and tracking analysis is most beneficial in AJCC stage 2 disease and above, or with T3+ or M1 disease. This suggests that ctDNA detection and monitoring is most useful in advanced or recurrent disease, which is generally considered the patient cohort who would most benefit from longitudinal monitoring and earlier detection.

The methods detailed in this study have been shown to be highly effective in low input or degraded DNA as is found from FFPE or cfDNA and increases the confidence in calling low abundance mutant alleles in the post-sequencing bioinformatics process. The limits of detection are based on the confidence of calling low abundant mutant alleles at each genomic loci of interest. Confidence can be increased by optimising a number of variables, including high starting cfDNA input, high depth sequencing, increasing the number of tracked variants in each personalised sequencing panel and inclusion of a normal cohort to assist in background noise reduction at individual genomic loci. These aspects should be considered in further studies of ctDNA in thyroid cancer.

The study also highlighted areas in which ctDNA demonstrates added utility over conventional tumour markers. We showed how the identification and tracking of multiple variants in ctDNA from pre-treatment, through radio-active iodine (RAI) treatment and into follow-up periods showed fluctuations in allele frequency correlating with treatment extent. These findings add evidence at a molecular level of the efficacy of RAI, which is important in the age of treatment de-escalation. Fluctuating ctDNA levels (as measured by detected variant allele frequency) not only correlated with the presence of residual disease but was also able to demonstrate evidence of tumour heterogeneity and clonal evolution. This is invaluable in the recurrent or residual disease patient as it reduces the lead time to residual disease detection but also potentially informs the modification of adjuvant or salvage treatment according to underlying molecular mechanisms.

The findings from this pilot study assist in guiding future study design when investigating ctDNA in thyroid cancer. It is recommended that future studies that

utilise the bespoke personalised sequencing panel methodology should focus on cohorts of patients with large tumours (T3+) with and without metastatic or recurrent disease with longer term follow up periods. Blood sampling frequency should be increased to allow more granular analysis of ctDNA fluctuations. The process of WES, bespoke sequencing panel design, library preparation and sequencing of cfDNA is extremely time consuming and a large financial commitment. There is a considerable bioinformatics burden that will require a dedicated bioinformatician to assist in data interpretation.

6.4 Future work

The findings from this thesis stimulate further possible avenues of investigation to improve our management of DTC. Translation of these techniques into clinical applications that can benefit patient and surgeon alike, requires a collaborative effort between clinicians, industry and scientific laboratories.

The genomic signature incorporating both somatic mutation status and differential expression of miRNA is an exciting prospect that warrants further investigation. Further expansion with the inclusion of gene specific primers that can detect fusions or copy number variations is recommended. Whilst broad sequencing methods such as whole exome, whole genome and RNA sequencing are capable of identifying all of these genetic events, it is still not financially viable to perform such analyses on large numbers of samples. Furthermore, as DTC is a low tumour mutational burden cancer, these techniques could generate large amounts of redundant sequencing data as the tumorigenic drivers are confined to small and specific areas of the genome. Cost effective methods that can analyse the full spectrum of genomic abnormalities, but only the specific areas pertaining to thyroid cancer, would greatly increase the clinical applicability. Further large sample size cohort studies that could be either retrospective or prospective, would be required to validate the prognostic accuracy of this predictive model. This would provide the necessary data to justify prospective interventional clinical studies that would randomise patients between upfront total thyroidectomy versus standard of care on the basis of molecular profiling. Extracting nucleic acids directly from fresh frozen tissue rather than formalin fixed tissue may yield higher quality nucleic

acids and improve the bioinformatics process. This could be achieved at the point of fine needle aspiration biopsy of a thyroid nodule in the diagnostic phase. An additional benefit of recruiting patients at the point of diagnosis would be analysing the effectiveness of molecular profiling in identifying cancer from benign, and therefore further its role in improving diagnostic accuracy, particularly in the context of the indeterminate nodule.

Our findings of dysregulated miRNA species that are significantly associated with vascular invasion and are involved in pathways intrinsic to immune activation and modulation prompt the recommendation for further studies investigating how the immune microenvironment is altered in DTC. There are a number of theories regarding immune escape in oncology and disease progression and our considerable biobank of samples could be used in studies to investigate this. This could include spatial studies using specific immunohistochemistry techniques to visualise the concentrations of these miRNA molecules in the tumour and their relationship to infiltrating immune cells. Additionally, in vitro studies could be performed using thyroid cancer cell lines, to determine if miRNA inhibitors (or anti-miRs) can manipulate the concentrations of the miRNA species identified in this study, with any resultant reduction in cell proliferation or increase in apoptosis.

The use of bespoke patient specific sequencing panels in this thesis is novel in DTC, and we have identified their potential to increase the detection sensitivity of ctDNA above the gold standard of PCR based techniques. Many of the techniques performed manually in the methodology of these studies are becoming increasingly automated in the laboratory. If automation is utilised, throughput of patient samples would exponentially increase, allowing the design of studies with increasing sampling frequency and therefore more granular data.

Due to the financial cost of WES and bespoke panel design, the number of patients entered into this methodology was limited (n = 15), but future industry sponsored clinical trials may provide the support for study design incorporating larger sample size to demonstrate clinical efficacy. The low proportion of patients that develop recurrent or residual disease in DTC mandates larger multicentre prospective studies with follow up between 1-5 years (as most

recurrences occur in this time frame) in order to acquire sufficient numbers to demonstrate the sensitivity of ctDNA bespoke sequencing in residual or recurrent disease detection. The success of the methods in this thesis are planned for use in future studies on advanced thyroid cancer by our unit. It is evident that genomics offer the possibility of truly personalised medicine, and in the field of oncology, it will revolutionise how we diagnose, prognosticate and monitor patients after definitive treatment.

Appendices

Appendix A – Full trial protocol, PIS and consent form

St George's University Hospitals NHS
NHS Foundation Trust

St George's
University of London

FULL/LONG TITLE OF THE STUDY	**Gene mutation analysis as a biomarker in thyroid cancer**
SHORT STUDY TITLE / ACRONYM The short title should be: • Sufficiently detailed to make clear to participants what the research is about in simple English • If acronyms are used the full title should explain them. The proposed acronym should not drive the long title	**N/A**
PROTOCOL VERSION NUMBER AND DATE Version control: • All draft versions should be numbered 0.1, 0.2 etc. • The final version for submission should be numbered 1.0 • The changes made relative to the previous protocol version should be listed after submission	**Version 1.3** **25/02/2019**
IRAS Number:	**244553**
JRES Reference Number	**2019.006**
Funder Reference Number:	**TBA**
This protocol has regard for the HRA guidance and order of content	

STUDY PROTOCOL

Gene mutation analysis as a biomarker in thyroid cancer

1 BACKGROUND

The incidence of well differentiated thyroid cancer (WDTC) has increased dramatically in the past decade (1), more than any other type of malignancy. Increasing awareness and more sensitive diagnostic techniques has largely driven this rapid increase and as a result majority of cancers being diagnosed are small volume tumours with low-risk features. Patients with thyroid cancer have excellent overall survival resulting in an ever-increasing pool of cancer survivors. However, up to 30% develop recurrence or metastatic disease, sometimes 20 years after initial treatment. Therefore, patients are currently followed-up life-long, which adds considerable financial and logistical strain to a resource limited healthcare system.

Recognising the general indolent nature and the excellent survival rates for WDTC, there has been major shift towards treatment de-escalation. The HiLo trial demonstrated lower dosage of radio-iodide treatment was sufficient for low-medium risk cancers and the currently ongoing IoN trial is expected to reveal that RAI may not be required in such patients. As a result, the current debate is centred on the optimal extent of surgical resection in these patients with low-medium risk group that may not need adjuvant treatment. More extensive and traditional total-thyroidectomy is recognised to carry higher complication rates and should be avoided in favour of hemi-thyroidectomy (or lobectomy) if it adds little/no additional health benefit (2). In line with this, the most recent ATA and BTA guideline updates in 2014/16 have therefore suggested that conservative lobectomy may be suitable for low-risk WDTC. However, the guidelines are broad and acknowledging the deficiencies in the evidence basis for the suggested practice change, there has been inconsistent interpretation and uptake of recommendations leading to significantly varied practices both in the UK and worldwide (3,4).

Although decision-making by the tumour board/MDT is based upon multifactorial issues, the decision to recommend total-thyroidectomy or completion surgery after initial lobectomy is mainly dependent on clinical and histological features of advanced disease and 'high-risk' factors associated with poor outcome. For patients with small volume disease (<4cm) and without clear clinical/imaging features of local invasion or distant metastases, the main determining factor is histological features of vascular invasion and extra-capsular invasion. Our recent study (submitted and under review) is consistent with other reports that up to 50% of small WDTC may require further completion surgery after initial lobectomy due mainly to the presence of vascular invasion (5). This figure is higher if cancer-multifocality is deemed to increase risk of recurrence by the local tumour

board/MDT (BTA 2014 guidelines). However, if all patients underwent upfront total-thyroidectomy, up to 50% of patients may have not needed to undergo more extensive surgery.

Therefore, there are 2 major issues to address in the management of low-risk WDTC group (representing the majority of new cancer diagnosis). Firstly, to optimise decision-making to provide the most suitable and effective first operation (and minimise unnecessary more invasive/extensive surgery). Improvements in so called "pre-operative risk stratification" would provide better patient experience, QoL and cost-savings for the health service providers. Secondly, to address the rapidly increasing health burden of long-term survivors requiring life-long follow up.

Major strides in genetic technology, in particular massive parallel next-generation gene sequencing, potentially allows us to rapidly and accurately probe the cancer genome for mutations and changes in important signalling and metabolic pathways. The detection of such changes may provide important diagnostic and prognostic information to facilitate both better and more accurate surgical decision-making and also a reliable and cost-effective way to monitor disease recurrence during post-treatment surveillance.

It is now well-recognised that various driver mutations promote aggressive biology and progression in Thyroid cancer and to be associated with poor outcome and high-risk histological features. (6,7). Parallel multi-mutation profiling in thyroid cancer biopsies may help to identify those cancers that are likely to be more aggressive and have high-risk features such as vascular invasion. Recent studies have also highlighted the ability to detect free circulating-DNA (ctDNA) in the blood samples of patients with early cancers (including breast (8), colon (9) and lung 10)) and its potential utility as a biomarker. Further, these studies have demonstrated that patients with detectable ctDNA after surgery are associated with higher risk of clinical recurrence and poorer survival outcomes. Detection of ctDNA would allow identification of thyroid cancer patients with clinically unrecognised post-operative minimal residual disease (MRD) and who are at higher risk of relapse after first surgery; thus useful in the selection of patients likely to benefit from further surgery (total thyroidectomy if initial lobectomy undertaken) or adjuvant RAI (guiding the need for higher dosage for some).

Whilst there is very little data currently in thyroid cancer, recently, Cradic et al (11) reported feasibility in detecting circulating BrafV660E mutation in blood in patients with early stage low volume thyroid cancers (Stage 1, T1-2). Further, a study in MTC showed that presence of Ret mutations in blood after initial treatment was associated with poorer outcome and suggests its use as biomarker of prognosis and guide for consideration of

early additional adjuvant therapy (12). Although these studies have shown that ctDNA is detectable in thyroid cancers of different subtypes, because they examined only a single variant, the reported sensitivity rates were too low for use as a clinical biomarker. However, more recently, our group utilized a parallel multi-mutational analysis method and demonstrated the feasibility and potential value of ctDNA as a biomarker in advanced thyroid cancer (13 EJC).

Current methods of thyroid cancer surveillance include serum Thyroglobulin (Tg) and serial imaging (RECIST criteria). CT/MRI scans are known to be associated with significant lag-time effect and low-sensitivity resulting in late detection of disease recurrence. Detection of ctDNA during surveillance would allow earlier detection of occult recurrence. The use of simple non-invasive serial blood sampling would also avoid costly serial imaging and radiation exposure. Further, the detection of ctDNA during surveillance would be particularly useful as a new thyroid cancer biomarker in an ever-increasing cohort of patients where the only currently available biomarker, thyroglobulin (Tg), is not helpful; those patients being followed-up after lobectomy-only, those patients who have not received adjuvant RAI and in those with high anti-thyroglobulin antibody titres (up to 25%).

Study Aims:

i. We propose to an exciting 2-part ***'proof-of-concept'*** study that utilise our purpose-designed thyroid cancer-specific multi-gene mutation panel (ThyMa panel which allows detection of over 95% of known driver-mutations in thyroid cancer) to interrogate thyroid cancers with high-risk features (such as vascular invasion, extra-capsular extension, aggressive histology sub-type and multifocality), to identify key mutations that can be used to differentiate these from more favourable low-risk WDTC. Positive findings would then lead to feasibility of use in fine-needle cytology specimens to provide the critical pre-operative information to facilitate selection of the correct 1st-time surgery. Oncotype Dx ™ represents a highly successful application in breast cancer (14), used in almost 1 million cases worldwide, to stratify risk and help guide treatment.

ii. Further, the same gene-mutation panel would also be used to detect individual-specific driver-mutations to allow design of patient-specific cancer biomarkers (cancer specific circulating DNA) to be tested on blood samples, so called '**liquid biopsy**'. This is hoped to offer a more accurate and earlier detection of disease recurrence during post-treatment surveillance. Further, the specific mutation profile may provide prognostic information to stratify intensity of follow-up regimen and help reduce health service burden.

This study has been discussed at the National Thyroid NCRI CSG meeting and gained wide support. This pilot study would represent an early 'proof-of-concept' study, and if it demonstrates positive findings we would envisage moving forward with a larger, multi-centre trial for clinical evaluation/validation of its role as a biomarker of disease persistence after initial treatment or recurrence during surveillance.

References:

1. Cancer Research UK; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Three.
2. Lang BH, Wong CKH. Lobectomy is a more Cost-Effective Option than Total Thyroidectomy for 1 to 4 cm Papillary Thyroid Carcinoma that do not Possess Clinically Recognizable High-Risk Features. Ann Surg Oncol. 2016 Oct;23(11):3641-3652.
3. National Patient Surgeon Survey (unpublished): Presented at the Controversies in Thyroid Cancer Management Symposium, Royal College of Physicians, June 2018.
4. Hirshoren N, Kaganov K, Weinberger JM, et al. Thyroidectomy Practice after implementation of the 2015 American Thyroid Assoociation Guidelines on surgical options for patients with well-differentiated thyroid carcinoma. JAMA Otolaryngol Head Neck Surg 2018;144:427-432
5. Hauch A, Al-Qurayshi Z, Randolph G, et al. Total thyroidectomy is associated with increased risk of complications for low- and high-volume surgeons. Ann Surg Oncol. 2014 Nov;21(12):3844-52.
6. Zhang Q, Liu S, Zhang Q, et al. Meta-Analyses of Association Between BRAF(V600E) Mutation and Clinicopathological Features of Papillary Thyroid Carcinoma. Cell Physiol Biochem. 2016;38(2):763-76.
7. Su X, Jiang X, Wang W, et al. Association of telomerase reverse transcriptase promoter mutations with clinicopathological features and prognosis of thyroid cancer: a meta-analysis. Onco Targets Ther. 2016 11;9:6965-6976.
8. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Science translational medicine. 2015;7(302):302ra133.
9. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Science translational medicine. 2016;8(346):346ra92.
10. Chaudhuri AA, Chabon JJ, Lovejoy AF, Newman AM, Stehr H, Azad TD, et al. Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. Cancer discovery. 2017.
11. Cradic K, Milosevic D, Rosenberg A, McIver B, Grebe S. Mutant BRAFT1799A can be detected in the blood of papillary thyroid carcinoma patients and correlates with disease status. J Clin Endocrinol Metab. 2009, 94(12):5001-9
12. G.J. Cote, C. Evers, M.I. Hu, et al. Prognostic significance of circulating RET M918T mutated tumor DNA in patients with advanced medullary thyroid carcinoma. J Clin Endocrinol Metab 2017; 102,3591-3599
13. Allin DM1, Shaikh R2, Carter P2, et al. Circulating tumour DNA is a potential biomarker for disease progression and response to targeted therapy in advanced thyroid cancer. Eur J Cancer. 2018; 103:165-175.
14. J Sparano, R Gray, D Makower et al. Adjuvant Chemotherapy Guided by a 21-gene expression assay in Breast cancer. NEJM 2018; 379:111-121

2 RATIONALE

The rate of newly diagnosed thyroid cancers has increased dramatically in the past decade, more than any other type of malignancy. Increasing awareness and more sensitive diagnostic techniques has largely driven this rapid increase and as a result, most new thyroid cancers are small volume tumours that are generally treated very successfully with a low rate of recurrence. Subsequently there is an ever-increasing pool of cancer survivors requiring long term follow up, which represents an increasing burden on the health services.

If a cancer is found in one lobe of the thyroid, there is ongoing controversy as to the optimal primary surgery; hemithyroidectomy versus total thyroidectomy (half the gland versus all the gland). There are advantages and disadvantages of each option that create a dilemma in decision making.

There are two major issues to address in the management of low-risk thyroid cancer. Firstly, the rapidly increasing health burden of long-term survivors requiring robust, cost effective, life-long follow up. Secondly, optimisation of the decision-making process to provide the most suitable and effective first operation. Improvements in both aspects would provide better patient experience, quality of life and cost-savings for health service providers.

This is a two-part study utilising molecular gene testing to address these issues.

Part 1 – This part of the study aims to analyse thyroid cancer specimens (surplus to clinical need) after appropriate consent, to identify critical gene changes that predict more aggressive cancers. This could more accurately inform the decision-making process when deciding the extent of surgery, potentially saving many unnecessary procedures.

Part 2 – This part of the study aims to identify patient specific thyroid cancer markers in patient blood samples at different time points through the patient journey. Increased levels of these markers could signify the presence of remaining cancer and therefore alert the need for additional treatment.

3 THEORETICAL FRAMEWORK

The theory behind this study is an extension of already published work by colleagues within our group[1]. Much of the methodology is identical and has been proven to be successful in advanced thyroid cancers. We now want to extend that technology into well-differentiated thyroid cancer, which forms the majority of all thyroid cancer diagnoses and where the biggest potential benefit for patients and the NHS service lies.

We have the ability to use established genetic sequencing techniques to interrogate DNA extracted from solid tumours and use that information to inform the detection of circulating tumour DNA free-floating within patient plasma. If successful, this will provide "proof-of-concept" data that could have major future implications in the way we risk stratify patients as well as provide cost-effective long term follow up for cancer survivors.

1. Allin DM1, Shaikh R2, Carter P2, et al. Circulating tumour DNA is a potential biomarker for disease progression and response to targeted therapy in advanced thyroid cancer. Eur J Cancer. 2018; 103:165-175

4 RESEARCH QUESTION/AIM(S)

4.1 Objectives

In well-differentiated thyroid cancer tumours, are there gene mutational signatures that are strongly associated with high-risk histopathological features and can we reliably detect them using a custom designed DNA mutation detection panel?

Can we predict residual or recurrent disease based on the detection of circulating tumour DNA after definitive treatment?

4.2 Outcome

1. Positive detection of 1 or more gene mutational variant using the ThyMa gene panel in DNA extracted from well-differentiated thyroid cancer tumour samples.

2. Positive detection of circulating tumour DNA in plasma samples from patients with confirmed well-differentiated thyroid cancer in at least one time point in the peri-operative and follow up period.

5 STUDY DESIGN and METHODS of DATA COLLECTION AND DATA ANALYIS

Our study is divided into two parts, a retrospective first part (Part 1) and a prospective second part (Part 2).

Part 1

The retrospective part of the study will identify historic patient data that meets the study inclusion criteria using the electronic patient records. The thyroid MDT has pre-existing databases with all patients previously discussed at MDT meetings, including those with confirmed diagnoses of thyroid cancer. These databases will be used to identify patients. Patients have already consented for their surgical samples to be used for Research via existing local surgical consent forms. Only these patients will have their archived thyroid resection samples located and extracted. Representative histopathology slides will be processed and an independent consultant histopathologist will manually verify the presence of malignant cells and confirm the histopathological diagnosis. These confirmed well-differentiated thyroid cancer cells will be isolated using microdissection and processed to extract DNA using well recognised techniques. The extracted DNA will be analysed using various molecular techniques including our specialised thyroid cancer gene mutation panel and next generation sequencing.

Included patients will be divided into the following cohorts:

50 samples in 10 x 5 cohorts.

i) Follicular thyroid cancers less than 4 cm with High Risk Features. (n=10)

ii) Papillary thyroid cancers less than 4 cm with High Risk Features. (n=10)

iii) Follicular thyroid cancers less than 4 cm without High Risk Features. (n=10)

iv) Papillary thyroid cancers less than 4 cm without High Risk Features. (n=10)

v) Benign disease (negative control) (n=10)

Part 2

The prospective part of the study will identify appropriate patients through current standard referral pathways. Patients will be recruited to the study via specialised head and neck clinics and multidisciplinary team (MDT) cancer pathways. The referral pathway will usually be via other ENT surgeons, GPs or endocrinologists. Full informed consent will be taken at the point of first contact with one of the research team (PI, co-investigators and research nurse).

Research participants will have identical investigation and treatment pathways as non-participants, apart from additional blood samples being taken at normal clinical timepoints (no additional clinical visits will be required for patients).

There are two cohorts of patients that will be recruited;

i) Patients who have a diagnosis of well-differentiated thyroid cancer following diagnostic lobectomy and due completion surgery (Lobectomy cohort)
ii) Patients who have pre-treatment diagnosis of thyroid cancer after fine needle cytology (Thy 5) due for surgery (lobectomy or total thyroidectomy) (Thy 5 cohort)

Blood samples (of approximately 20mls each time) will be collected at the following time points:

Lobectomy cohort – 1. Pre-operative, 2. Post completion surgery, 3. Post adjuvant radio-iodine treatment or 6 month follow up (whichever comes first)

Thy 5 cohort – 1. Pre-operative, 2. Post primary surgery (lobectomy or total thyroidectomy), 3. Post secondary surgery (completion surgery if relevant) 4. Post adjuvant radio-iodine treatment or 6 month follow up (whichever comes first)

Surgical excision specimens (surplus to diagnosis) will be acquired from research participants that undergo thyroid lobectomy, total thyroidectomy or both as recommended by the MDT outcome. These specimens will be processed in the same way as the retrospective study described above.

Research participants will be followed up as per routine departmental protocol.

6 STUDY SETTING

Retrospective Part 1 patients will be identified via existing Thyroid MDT databases and electronic hospital records. Prospective part 2 patients will be identified through current standard referral pathways. Patients will be recruited to the study via specialised head and neck clinics and multidisciplinary team (MDT) cancer pathways. The referral pathway will usually be via other ENT surgeons, GPs or endocrinologists. Full informed consent will be taken at the point of first contact with one of the research team (PI, co-investigators and research nurse).

Research participants will have identical investigation and treatment pathways as non-participants, apart from additional blood samples being taken at normal clinical timepoints (no additional clinical visits will be required for patients).

It is a single-centre study however scientists and clinicians at the Institute of Cancer Research, London will be providing analysis of samples.

The entire Ear, Nose and Throat (ENT) department at St George's will be aware of the study. This is to ensure both clinician and support staff cooperation.

Access to a centrifuge and secure, lockable freezer facilities will be used to process and store patient plasma samples at -80°C in the histopathology department at St George's Hospital. Only the PI, co-investigators and research nurse will have access to these facilities.

7 SAMPLE AND RECRUITMENT

7.1 Eligibility Criteria

All adult patients (defined as >18 years of age at the time of giving informed consent) with a confirmed histopathological diagnosis of well-differentiated thyroid cancer will be eligible for inclusion in this study.

7.1.1 Inclusion criteria

- **Adult (>18 years old at time of giving informed consent)**
- **Confirmed histopathological diagnosis of well-differentiated thyroid cancer**

7.1.2 Exclusion criteria

- Age less than 18
- Any concurrent thyroid disease other than malignancy
- Incomplete data or lost to follow up
- Significant medical co-morbidities or concurrent malignancy
- Thyroid nodules less than 1cm in size.

7.2 Sampling

Will be opportunistic based on presentation in specialised Head & Neck clinics in the ENT department at St George's University Hospital.

7.2.1 Size of sample

We are aiming for a study population of n = 50 for each part of the study (n = 100 overall). This is an observational cohort study, aiming to provide pilot "proof-of-concept" data, and as such power calculations are limited. However, we have based our study sample size on recently published literature.

7.2.2 Sampling technique

Sampling will be chronological for all patients that fit the inclusion criteria and consent to participate.

7.3 Recruitment

Appropriate patients will be identified through current standard referral pathways. Patients will be recruited to the study via specialised head and neck clinics and multidisciplinary team (MDT) cancer pathways. The referral pathway will usually be via other ENT surgeons, GPs or endocrinologists. Patient will be approached by a member of their direct care team in collaboration with the research team, PIS will be given and discussed. Patient will be consented at the same visit unless they require more time to consider the trial.

The entire Ear, Nose and Throat (ENT) department at St George's will be aware of the study. This is to ensure both clinician and support staff cooperation.

7.3.1 Participant identification

The Thyroid MDT has prospectively populated databases of all patients discussed each week. This database is managed by the Head & Neck team and will be utilised to identify patients.

Part 1 study patients will be identified directly from the Thyroid MDT database and electronic records.

Part 2 study patients will be identified through current standard referral pathways. Patients will be recruited to the study via specialised head and neck clinics and multidisciplinary team (MDT) cancer pathways. The referral pathway will usually be via other ENT surgeons, GPs or endocrinologists. Patients will be approached by a member of their direct care team in collaboration with the Research team and a Patient Information Sheet will be given and discussed. Patients will be consented at the same visit unless they require more time to consider the trial.

Only members of the direct care team for the patients and the research team will have access to identifiable patient information.

7.3.2 Consent

Part 1 patients will have already consented for their surgical samples to be used in research via local surgical consent forms. No further consent will be sought as there are no extra requirements for these patients, and this may cause undue stress if for instance patient has passed away and their relatives are contacted.

Informed consent will be obtained prior to the participant undergoing any activities that are specifically for the purposes of the study.

For Part 2 as outlined in the protocol and IRAS ethics application, specific time additional to their routine appointment time will be required (approx. 10mins) to discuss the study and gain informed consent.

Full informed consent will be taken by the patient's direct care team in collaboration with Research staff after Patient information Sheet has been discussed unless patient needs more time to consider participation in the trial. This will involve discussing the nature and objectives of the study and possible risks associated with their participation. Emphasis will be made that this is a non-interventional observational study and as such the research treatment pathway will be identical to standard treatment pathways. Participants will be given adequate time to ask questions and an appropriate contact for any further queries. A copy of the written consent will be issued to each participating patient.

Consent provisions for collection and use of participant data and biological specimens

Participating patients will be consented for the use of their clinical data and biological specimens for research purposes. Biological specimens include surgically excised and archived thyroid tumours surplus to clinical diagnosis as well as four serial plasma samples collected in the peri-operative and post-operative period at routine clinical time points. No additional visits will be required for participation in the study.

Biological specimens will be stored in the Histopathology department at St George's University Hospital and transported in accordance to NHS and GCP guidelines to laboratories at the Institute of Cancer Research, London. Frozen blood plasma samples will be stored at -80°C in secure locked freezers, access to which will be limited to the PI and research fellow.

Participants are entitled to withdraw themselves or their biological samples from the study at any time, (would you be able to request blood samples and results are discarded after they've been sent off?) although data derived prior to request to withdraw may still be used in the final analysis. All published data will be anonymized.

7.3.3 Data collection tool

Patient data will be collected from pre-existing Thyroid MDT databases and electronic patient records. No additional data collection forms will be required. Total data collected will be collated on electronic spreadsheets with NHS encryption security. Each patient will be given a unique study number to allow for anonymisation. Corresponding cross-identification tables will only be available to PI and clinical research fellow.

7.3.4 Biological Sample Handling

Thyroid tumour resection samples surplus to clinical diagnosis will be accessed from the histopathology department. This will usually be as Formalin-Fixed, Paraffin-Embedded (FFPE) blocks. Samples will be cut with a microtome to produce 6 glass slide samples. The remaining FFPE blocked sample will be returned to routine NHS histopathology storage facilities as per standard practice.

1 glass slide samples will be used by our nominated Histopathology Consultant colleague to provide an accurate histopathological diagnosis. The remaining 5 glass slide samples will be processed for DNA and RNA extraction using recognised techniques. This DNA will then be analysed using various genetic techniques, including next generation sequencing.

Additionally, 20ml blood samples will be collected at 4 peri-operative time points for each participant. These will be at routine clinical timepoints therefore no additional appointments will be required. Blood samples will be processed within 2 hours of collection. The blood will be centrifuged and the plasma component will be retained and stored at -80°C in secure locked freezers, the remaining blood cells will be disposed of. Access to -80°C freezers will be limited to the PI and research fellow.

The plasma samples will be processed and circulating tumour DNA will be extracted to be processed with recognised genetic techniques.

All samples will be retained for this project and any future ethically approved studies within NHS biobanks.

8 ETHICAL AND REGULATORY CONSIDERATIONS

This study is observational and non-interventional in design, with minimal risk to participating patients. All patients will follow standard care pathways in accordance with the Thyroid MDT practice at St George's University Hospital. There is no direct benefit to participants. There are additional interventions for participating patients compared to non-participating patients. This is highlighted below:

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?
3. Average time taken per intervention/procedure (minutes, hours or days).
4. Details of who will conduct the intervention/procedure, and where it will take place.

Intervention or procedure	1	2	3	4
Given information sheet, initial discussion	1	0	10 mins	PI, co-investigators, trained research nurse – specialist head and neck clinics
Signed informed consent	1	0	10 mins	PI, co-investigators, trained research nurse – specialist head and neck clinics
Venesection (20mls blood)	4	3	5min	Co-investigators, trained research nurse or phlebotomy staff at St George's hospital

Our study protocol is robust and mindful of patients' concerns and expectations. Adequate time will be provided for patients to ask questions, as well as a contact pathway for any further queries regarding participation in the study. All sample collection, processing, storage and transportation will be within currently used NHS arrangements and contracts. The laboratory analysis will be conducted at the Institute of Cancer Research in London, which has a worldwide reputation of excellent scientific and ethical standards of practice.

8.1 Assessment and management of risk

This study is observational and non-interventional in design, with minimal risk to participating patients. All patients will follow standard care pathways in accordance with the Thyroid MDT practice at St George's University Hospital. There is no direct benefit to participants.

We do not anticipate any safeguarding issues.

8.2 Research Ethics Committee (REC) and other Regulatory review & reports

Before the start of the study, a favourable opinion will be sought from an appropriate REC for the study protocol, informed consent forms and other relevant documents e.g. advertisements.

For HRA- NHS REC reviewed research

- Substantial amendments that require review by NHS REC will not be implemented until that review is in place and other mechanisms are in place to implement at site.
- It is the Chief Investigator's responsibility to produce the annual reports and submit the REC within 30 days of the anniversary date on which the favourable opinion was given, and annually until the study is declared ended.
- The Chief Investigator will notify the REC of the end of the study within one year after the end of the study.
- If the study is ended prematurely, the Chief Investigator will notify the REC, including the reasons for the premature termination.

Regulatory Review & Compliance

Before any site can enrol patients into the study, the Chief Investigator/Principal Investigator or designee will ensure that appropriate approvals from participating organisations are in place. Specific arrangements on how to gain approval from participating organisations are in place and comply with the relevant guidance.

Amendments

For any amendment to the study, the Chief Investigator or designee, in agreement with the sponsor will submit information to the appropriate body in order for them to issue approval for the amendment. The Chief Investigator or designee will work with sites (R&D departments at NHS sites as well as the study delivery team) so they can put the necessary arrangements in place to implement the amendment to confirm their support for the study as amended.

8.3 Peer review

This study has been discussed at the National Thyroid NCRI CSG meeting and gained wide support.

The protocol has been reviewed by Professor Kevin Harrington and Dr Kate Newbold (Institute of Cancer Research and Royal Marsden Oncology department).

8.4 Patient & Public Involvement

N/A

8.5 Protocol compliance

Protocol deviations, non-compliances, or breaches are departures from the approved protocol.

All protocol deviations must be adequately documented on the relevant forms and reported to the Chief Investigator and Sponsor immediately.

Deviations from the protocol which are found to frequently recur are not acceptable, will require immediate action and could potentially be classified as a serious breach.

8.6 Data protection and patient confidentiality

All data will be handled in accordance with the Data Protection Act 2018 (UK implementation of the EU General Data Protection Regulation (GDPR)).

Any Case Report Forms (CRFs) will not bear the participant's name or other directly identifiable data. The participant's trial Identification Number (ID) only, will be used for identification. The Subject ID log can be used to cross reference participant's identifiable information.

Tissue and blood plasma samples with pseudonymised data to identify them will be transferred between research sites (between St Georges University Hospital and Institute of Cancer Research) but without patient identifiable information.

All research data will correspond to unique patient trial numbers, with conversion tables/codes stored separately on password protected drives and only available to the PI and Clinical Research Fellow. Confidentiality of data will be maintained when study findings are published and submitted for peer review and patients will only be referred to by trial number.

Data will be retained for 5 years.

8.7 Indemnity

St George's University Hospitals NHS Foundation Trust sponsored research:

St Georges University Hospitals NHS Foundation Trust is party to NHS Litigation Authority (NHSLA) / NHS Resolution. As an NHS body it is liable for clinical negligence and other negligent harm to individuals covered by their duty of care. NHS Institutions employing researchers are liable for negligent harm caused by the design of studies they initiate.

8.8 Access to the final study dataset

This is a single-centre study and access to the full dataset will be restricted to the PI and clinical research fellow. As mentioned above, tissue samples from fully consented participants will be transported to the research laboratories at the Institute of Cancer Research in London, but in all cases these samples will be anonymised with individual study identification numbers. Final analysis, interpretation and write up of results will be under the remit of the PI and clinical research fellow.

9 DISSEMINATION POLICY

9.1 Dissemination policy

Publication: "Any activity that discloses, outside of the circle of trial investigators, any final or interim data or results of the Trial, or any details of the Trial methodology that have not been made public by the Sponsor including, for example, presentations at symposia, national or regional professional meetings, publications in journals, theses or dissertations."

All scientific contributors to the Trial have a responsibility to ensure that results of scientific interest arising from Trial are appropriately published and disseminated. The Sponsor has a firm commitment to publish the results of the Trial in a transparent and unbiased manner without consideration for commercial objectives.

To maximise the impact and scientific validity of the Trial, data shall be consolidated over the duration of the trial, reviewed internally among all investigators and not be submitted for publication prematurely. Lead in any publications arising from the Trial shall lie with the Sponsor in the first instance.

Before the official completion of the Trial,

All publications during this period are subject to permission by the Sponsor. If an investigator wishes to publish a sub-set of data without permission by the Sponsor during this period, the **Steering Committee/the Funder** shall have the final say.

Exempt from this requirement are student theses that can be submitted for confidential evaluation but are subject to embargo for a period not shorter than the anticipated remaining duration of the trial.

Up to 180 days after the official completion of the Trial

During this period the Chief Investigator shall liaise with all investigators and strive to consolidate data and results and submit a manuscript for peer-review with a view to publication in a reputable academic journal or similar outlet as the Main Publication.

- The Chief Investigator shall be senior and corresponding author of the Main Publication.
- Insofar as compatible with the policies of the publication outlet and good academic practice, the other Investigators shall be listed in alphabetic order.

- Providers of analytical or technical services shall be acknowledged, but will only be listed as co-authors if their services were provided in a non-routine manner as part of a scientific collaboration.
- Members of the Steering Group shall only be acknowledged as co-authors if they contributed in other capacities as well.
- If there are disagreements about the substance, content, style, conclusions, or author list of the Main Publication, the Chief Investigator shall ask the Steering Group to arbitrate.

Beyond 180 days after the official completion of the Trial

After the Main Publication or after 180 days from Trial end date any Investigator or group of investigators may prepare further publications. In order to ensure that the Sponsor will be able to make comments and suggestions where pertinent, material for public dissemination will be submitted to the Sponsor for review at least sixty (60) days prior to submission for publication, public dissemination, or review by a publication committee. Sponsor's reasonable comments shall be reflected. All publications related to the Trial shall credit the Chief and Co-Investigators as co-authors where this would be in accordance with normal academic practice and shall acknowledge the Sponsor and the Funders.

9.2 Archiving Arrangements

Each site will be responsible for their onsite level study archiving. The trial essential TMF along with any central trial database will be archived in accordance with the sponsor SOP.

St George's University Hospitals NHS
NHS Foundation Trust

Study of genetic biomarkers in head and neck cancers

We invite you to take part in a research study

- Before you decide whether to take part, it is important for you to understand why the research is being done and what it will involve.
- Please take time to read the following information carefully. Discuss it with friends and relatives if you wish.
- You are free to decide whether or not to take part in this study. If you choose not to take part, this will not affect the care you get from your own doctors.
- Ask us if there is anything that is not clear or if you would like more information.

Important things that you need to know

- We want to find the best ways to treat patients with head and neck cancer.
- Our aim is to modernise and optimise the diagnosis and follow up approach for patients with head and neck cancer.
- We plan to use exciting new genetic technology to look for specific gene mutations that could predict more aggressive types of head and neck cancer and therefore help us treat them most appropriately.
- This study fits into your normal treatment, so there are no extra clinics visits or scans.
- You can stop taking part in the study at any time.

Contents

How to contact us

If you have any questions about this study, please talk to the doctors who organise it: Mr Dae Kim or Mr Samuel Chan on **07921 463788**.

Purpose of this study

Our aim is to discover genetic mutation markers within DNA taken from the cells of head and neck tumours that will enable us to predict more aggressive forms of cancer in the future. The results of this study would form the first steps towards producing a cost-effective minimally invasive test that could help both patients and doctors make decisions before and after cancer treatment.

Why am I being asked to take part?

You have been diagnosed with a type of cancer that we are investigating. After your surgery, your surgically removed tumour will be processed as normal in the histopathology department. With your permission, we would like access to a small portion of this tissue that is surplus to clinical diagnosis. We will then extract the DNA from the tumour cells and analyse the DNA for certain mutations. Our tests only look for mutations that are related to your cancer, so there will be no unexpected new genetic diagnoses from this testing. Additionally, we would like to collect blood samples around the time of your surgery and also in the follow up period. This is to look for biological markers in the blood that are related to your cancer, and to track how these change before and after surgery. These additional blood samples will be collected at the same time as routine blood samples normally taken before and after surgery as part of routine treatment pathways. This means there will be no additional inconvenience to you if you choose to participate in the study.

Description of the study

This is an observational study that will not make any changes to your routine treatment pathway (ie. If you choose to participate in the study, your treatment will be exactly the same as someone who isn't participating). Below is a flow diagram of your anticipated treatment pathway and when we may take samples, although your individual pathway is not confined to this model:

What will I need to do if I take part?

Looking at blood serum samples

In addition to your routine blood tests, we would like to take additional blood samples (approx. 10-30ml extra) from you during your regular clinical visits for additional analyses. We would like to monitor potential changes in protein levels in your blood during and after treatment to see if it can help us better predict treatment outcomes. There will be 4 individual timepoints for blood sampling at various points of your treatment and follow-up time periods.

Looking at tissue samples

As part of this clinical research study, we would like to be allowed to have access to samples of your tissues, which were taken as part of your routine surgical treatment. These samples will have been collected by your hospital team and stored in paraffin fixed wax blocks. With your consent, these samples will be collected and tested for the presence of a number of different chemicals known as biological markers by the Institute of Cancer Research, London. All your data will be anonymised so you cannot be identified through them.

Your rights regarding tissue/blood samples taken as part of this clinical research study

The results of the analysis of your individual samples will not routinely be given to you unless it is of clinical significance and of importance to your health. You will not benefit financially if this research leads to the development of a new treatment or medical test. By consenting to this study, you are gifting your tissue/blood samples to the research team to be used in ongoing research. The samples may be stored in secure NHS tissue bank facilities for future research but will not be sold for financial gain. Any unused samples will be disposed of securely and in accordance with Human Tissue Act 2004 laws.

New Information

We do not anticipate finding any new information that could directly impact your care or treatment pathway, however if this were to happen your study doctor will tell you about it and discuss with you whether or not you want to continue. If you decide to withdraw, your study doctor will ensure that your care will continue.

Voluntary participation and discontinuation

Your participation in this study is voluntary. If you agree to take part and then change your mind and wish to withdraw you may do so at any time without this decision affecting your future care.

Confidentiality and patient rights

If you agree to take part in the study you will be asked to sign and date the Informed Consent Form attached. Your medical notes will need to be accessed by authorised members of the research team and your direct care team only. Your unique study registration number will be used to make sure you cannot be identified outside of the research study. All information which is collected about you during the course of the research will be treated as strictly confidential. The confidentially of your medical records will be maintained at all times. For more information please see the Transparency Section document.

Collaboration with the Royal Marsden Hospital

If your routine visits and subsequent treatments occur at the Royal Marsden Hospital, then samples may be collected by the research team there. Your clinical data will be shared between care teams in each hospital.

Organisation and funding of this study

This research study is being sponsored by St George's University Hospital. The Royal Marsden Hospital is a research subsite. Samples will be processed and analysed at the Institute of Cancer Research in London. The research is funded by grants from several charities and health organisations. No commercial companies have stake in this research. During your involvement in this study unfortunately no travel costs incurred by you or your family will be paid. Your doctor or any other members of staff that are involved in your treatment and care have not been paid for entering you into the study.

Additional information

What are the potential disadvantages of taking part?

This study does not involve any additional invasive procedures compared to routine care. The main inconveniences are the short amount of extra time in clinic it will take to discuss the study and consent to taking part and also the extra blood bottle samples taken during routine blood testing. Taking part in the study will not change your diagnosis or alter your treatment pathway and therefore should have no effect on any insurance you have, but we recommend seeking advice if necessary.

What are the potential benefits of taking part?

There is not anticipated to be any clinical or financial benefit for you by taking part in this study. The information we will get from this study will help improve the treatment of people with head and neck cancer.

What if there is a problem?

Any complaint about the way you have been dealt with or any possible harm you might suffer will be addressed by your direct care team or Patient Liaison Services (PALS) but please do speak with the study doctor who will do their best to help. In the event that something goes wrong and you are harmed during the research and this is due to someone's negligence then you may have grounds for legal action for compensation against the sponsor (St George's University Hospitals NHS Foundation Trust), but you may have to pay your legal costs. The normal NHS complaints mechanism will still be available to you.

Local details to be added here

Will my taking part in the study be kept confidential?

Your confidential data will be safeguarded during and after the study. Access to data that could be identifiable to you, is restricted to your routine care team, members of the research team and SGUH and RMH Research & Development team (for purposes of monitoring the quality of research only). Upon entry into the study, you will be allocated a unique study number and all your subsequent samples and study data will be anonymised without any ability for them to be linked to you by anyone outside of the research team. Your data will be collected onto electronic forms and databases that will be stored within data encrypted NHS systems. We will not be informing your GP or any other health professional outside of your routine care team of your participation in this study. Your data may be stored in the same secure NHS systems for up to 5 years and any disposal of data or tissues will be undertaken securely maintaining your confidentiality.

What will happen to my tissue specimens after the study has ended?

Your tissue specimens may be retained and stored in secure NHS pathology and research facilities for future research pending further ethics approval if necessary.

Where can I find out the results of the study?

The results will be published in scientific peer-reviewed journals, but you will not be identified individually on any report or publication. Your study doctor can direct you towards these publications on request.

Who should you contact with questions

You will be given a copy of this information sheet and the signed consent form to keep. If you have any problems or questions about this study or your rights as a patient in clinical research you should contact:

Doctor..... Study Nurse.....

Tel No..... Tel No.....

We would like to thank you for reading this patient information sheet and for considering taking part in this clinical research study. If you have any further questions please talk to the study doctor or study nurse before considering entering into the study.

Transparency Statement for St George's University Hospitals NHS Foundation Trust

St George's University Hospitals NHS Foundation Trust is the sponsor for this study based in the United Kingdom. We will be using information from you and/or your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. We may keep information collected for the purpose of the study for up to 5 years after the study has finished. This is to ensure integrity of the results. All data will be stored in a secure manner.

We will use your name, contact details and other identifiers to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Individuals from St George's University Hospitals NHS Foundation Trust and regulatory organisations may look at your medical and research records to check the accuracy of the research study. The only people in St Georges University Hospitals NHS Foundation Trust who will have access to information that identifies you will be people who need to contact you as part of the research study you are involved in, or to audit the data collection process. The people who analyse the information will not be able to identify you and will not be able to find out your name or contact details.

When you agree to take part in a research study, the information about your health and care may be provided to researchers running other research studies in this organisation and in other organisations. These organisations may be universities, hospitals or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research.

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of healthcare research and cannot be used to contact you or to affect your care. It will not be used to make decisions about future services available to you, such as insurance.

Your rights to access, to change or move your information are limited, as we need to manage the data in specific ways to ensure the research we conduct is reliable and accurate. If you withdraw your consent to participate in a research project, this will not mean we will have to remove all data as well. We will keep the information about you that we have already obtained to ensure research integrity is maintained in the public's interest. To safeguard your rights, we will strive to use the minimum personally-identifiable information possible.

You can find out more about how we use your information here

https://www.stgeorges.nhs.uk/about/privacy-notice/

For general information on how the NHS uses research data please visit:

https://www.hra.nhs.uk/information-about-patients/

St George's University Hospitals NHS
NHS Foundation Trust

IRAS ID: **244553**

Centre Name:

Study Name:

Participant Identification Number for this trial:

CONSENT FORM

Title of Project: **Gene mutation analysis as a biomarker in head and neck cancer**

Name of Chief Investigator: **Mr Dae Kim (Consultant ENT Surgeon)**

Please initial box

1. I confirm that I have read the information sheet dated..... (version.....) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant sections of my medical notes and data collected during the study, may be looked at by responsible individuals and the Research Team from St. George's University Hospitals NHS Foundation Trust (SGHT) and The Royal Marsden NHS Foundation Trust, or regulatory authorities, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

4. I understand that the information collected about me will be used to support other research in the future and may be shared anonymously with other researchers. ☐

5. I consent to the access and use of surgically excised tumour samples surplus to clinical diagnosis and gift them to St George's University Hospitals NHS Foundation Trust (SGHT) and The Royal Marsden NHS Foundation Trust for future research. ☐

6. I consent to the taking of blood samples from me as part of the study, and gift them to the research team for future research. ☐

7. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature

When completed: 1 for participant; 1 for researcher site file; 1 to be kept in medical notes. Version 2.1, 12/06/19

Appendix B – Full ‘R’ script for MCA

```
install.packages("installr")
library(installr)
updateR()
setwd
install.packages("readxl")
library(readxl)
read_excel("ALL.xlsx")-> data_CMA
read_excel ("CA_table2.xlsx") -> data_CA
library (tidyverse)
data_CA %>% remove_rownames %>% column_to_rownames(var="Mutations") -> matrix_CA
dt <- as.table(as.matrix(matrix_CA))
dt
install.packages(c("FactoMineR", "factoextra"))
library("FactoMineR")
library("factoextra")
select (data_CMA, 2:24) -> all_data
res.mca <- MCA(all_data, ncp = 5, graph = FALSE)
print (res.mca)
eig.val <- get_eigenvalue(res.mca)
fviz_screeplot(res.mca, addlabels = TRUE, ylim = c(0, 45))
fviz_mca_biplot(res.mca,
               repel = FALSE,
               ggtheme = theme_minimal())
fviz_mca_biplot(res.mca,
               repel = TRUE, # Avoid text overlapping (slow if many point)
               ggtheme = theme_minimal())
var <- get_mca_var(res.mca)
var
fviz_mca_var(res.mca, choice = "mca.cor",
            repel = TRUE, # Avoid text overlapping (slow)
            ggtheme = theme_minimal())
fviz_mca_var(res.mca,
            repel = TRUE, # Avoid text overlapping (slow)
            ggtheme = theme_minimal())
head(var$cos2, 4)
fviz_mca_var(res.mca, col.var = "cos2",
            gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
            repel = TRUE, # Avoid text overlapping
            ggtheme = theme_minimal())
fviz_mca_var(res.mca, alpha.var="cos2",
            repel = TRUE,
            ggtheme = theme_minimal())
library("corrplot")
corrplot(var$cos2, is.corr=FALSE)
fviz_cos2(res.mca, choice = "var", axes = 1:2)
head(round(var$contrib,2), 4)
fviz_contrib(res.mca, choice = "var", axes = 1, top = 15)
fviz_contrib(res.mca, choice = "var", axes = 2, top = 15)
fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 30)
fviz_mca_var(res.mca, col.var = "contrib",
            gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
            repel = TRUE,
            ggtheme = theme_minimal()
)
ind <- get_mca_ind(res.mca)
ind
fviz_mca_ind(res.mca, col.ind = "cos2",
            gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
            repel = TRUE, # Avoid text overlapping (slow if many points)
            ggtheme = theme_minimal())
fviz_cos2(res.mca, choice = "ind", axes = 1:2, top = 20)
fviz_contrib(res.mca, choice = "ind", axes = 1:2, top = 20)
fviz_mca_ind(res.mca,
            label = "none", # hide individual labels
```

```
            habillage = "VI", # color by groups
            palette = c("#00AFBB", "#E7B800"),
            addEllipses = TRUE, ellipse.type = "confidence",
            ggtheme = theme_minimal())
fviz_ellipses(res.mca, c("VI", "TERT"),
              geom = "point")
fviz_ellipses(res.mca, c("VI", "BRAF", "TERT"),
              geom = "point")
res.desc <- dimdesc(res.mca, axes = c(1,2))
res.desc[[1]]
res.desc[[2]]
install.packages("gplots")
library("gplots")
balloonplot(t(dt), main ="Mutations and vascular Invasion", xlab ="", ylab="",
            label = FALSE, show.margins = FALSE)
chisq <- chisq.test(matrix_CA)
chisq
library(corrplot)
corrplot(chisq$residuals, is.cor = FALSE)
```

Appendix C – Custom TaqMan Assay designs

Gene	Nucleotide change	Amino acid change	Chr	Position	Number of bases	Sequence
BRAF	c.T1799A	p.V600E	7	140453136	118	TATATTTCTTCATGAAGACCTCACAGTAAAAATAGGTGATTTTGGTCTAGCTACAG[T/A]GAAA TCTCGATGGAGTGGGTCCCATCAGTTTGAACAGTTGTCTGGATCCATTTTGTGGATG
NRAS	c.A182G	p.Q61R	1	115256529	118	AAACAAGTGGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATACAGCTGGAC[A/G]A GAAGAGTACAGTGCCATGAGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTG
TERTp	c.C-124A	NA	5	1295228	124	ccccgtcccgacccctcccgggtccccggcccagcccc[c/a]tccgggccctcccagcccctcccctccttccgcggcccccgccCT CTCCTCGCGGCGCGAGTTTCAGGCAGCGCTGCGTC
APC	c.C3778T	p.Q1260*	5	112175069	124	CAGCCTCAAAAGGCTGCCACTTGCAAAGTTTCTTCTATTAACCAAGAAACAATA[C/T]AGACTT ATTGTGTAGAAGATACTCCAATATGTTTTTCAAGATGTAGTTCATTATCATCTTTGTCA
TSHR	c.T727C	p.S243P	14	81606057	125	TGCCTCCCAGGGACGTGTCTCAAACCAGTGTCACTGCCCTTCCA[T/C]CCAAAGGCCTGGAG CACCTGAAGGAACTGATAGCAAGAAACACCTGGACTCTTAAGAAACTTCCACTTTCCTTGA
PIK3CA	c.G56A	p.R19K	3	178916669	124	AGAATCAGAACAATGCCTCCACGACCATCATCAGGTGAACTGTGGGGCATCCACTTGATG CCCCCAA[G/A]AATCCTAGTAGAATGTTTACTACCAAATGGAATGATAGTGACTTTAGAATGC
GNAS	c.G2531A	p.R844H	20	57484421	119	ctgtttcggttggctttggtgagatccattgacctcaattttgtttcagGACCTGCTTCGCTGCC[G/A]TGTCCTGACTTC TGGAATCTTTGAGACCAAGTTCCAGGTGGACAAAGTCAACT

Appendix D – Final bespoke sequencing panel designs with amplicon coordinates for each patient

Patient study ID No 7

Chromosome	Start	Stop	Description
chr1	13716964	13716965	PRAMEF17_c.452T>C
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	160143961	160143962	ATP1A4_c.2053G>T
chr1	170688995	170688996	PRRX1_c.371G>A
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	52252216	52252217	LEO1_c.1039C>T
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	72648927	72648928	HEXA_c.284del
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	74695252	74695253	RFWD3_c.95G>T
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	7578261	7578262	TP53_c.587G>A
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr19	55798540	55798541	BRSK1_c.232-41C>T
chr2	24118803	24118804	ATAD2B_c.253G>A
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr20	1317082	1317083	AL136531.1_c.248G>C
chr22	32875189	32875190	rs11107_G_A_0.48742_1111

chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	111732464	111732466	TAGLN3_c.*68dup
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr4	48037042	48037043	NIPAL1_c.607A>G
chr4	55970796	55970797	KDR_c.1987+13T>A
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1
chr5	14509401	14509402	TRIO_c.*980G>C
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	72892783	72892784	RIMS1_c.1610G>T
chr6	75795139	75795141	COL12A1_c.*1122_*1123insGTG
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	38828069	38828070	PLEKHA2_c.*768G>C
chr8	38828080	38828081	PLEKHA2_c.*779T>C
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	86888832	86888833	KLHL4_c.1634C>A
chrX	153640335	153640336	TAZ_c.109+47G>A
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 39

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	149858900	149858902	HIST2H2AC_c.378_379insTGGC
chr1	224002025	224002026	TP53BP2_c.205T>C
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	15813476	15813477	EPS8_c.937+71A>G
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr14	31083510	31083511	AP4S1_c.367-1258T>G
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	5684300	5684301	HSD11B1L_c.-14-529C>A
chr19	6466367	6466368	CRB3_c.157-109A>G
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	39423316	39423318	MRPS12_c.396_397del
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr19	55761918	55761919	RFPL4A_c.119A>G
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	37441081	37441083	CEBPZ_c.2469_2470insTTAC
chr2	141459303	141459304	LRP1B_c.6413G>A
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	223065781	223065782	PAX3_c.*111C>T
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	52783825	52783826	NEK4_c.1388A>G

chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr5	1260643	1260644	TERT_c.2915G>A
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	57842954	57842955	CTD-2117L12.1_c.261+7A>G
chr5	57843011	57843012	CTD-2117L12.1_c.211C>T
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	65429329	65429330	GUSB_c.1769C>T
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	135291461	135291462	FHL1_c.749G>T
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 9

Chromosome	Start	Stop	Description
chr1	7792499	7792500	CAMTA1_c.2915-8T>G
chr1	44165665	44165666	KDM4A_c.2841+1982G>C
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	201981506	201981507	ELF3_c.421G>T
chr10	76732390	76732391	KAT6B_c.1055G>A
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	25398283	25398284	KRAS_c.35G>A
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	17752225	17752226	TOM1L2_c.1339-53C>T
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr18	62715971	62715972	PHLPP1_c.289G>A
chr19	10480461	10480462	TYK2_c.194-1368T>C
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	36979012	36979013	ZNF568_c.19T>C
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	27457541	27457542	CAD_c.3775G>A
chr2	178503895	178503896	PLEKHA3_c.*9A>T
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	210690729	210690730	UNC80_c.2431C>T
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111

chr3	105266008	105266009	ALCAM_c.1121G>A
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr5	179290803	179290804	TBC1D9B_c.3397G>T
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	86817479	86817480	DMTF1_c.1274G>A
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	101670030	101670031	SNX31_c.65T>C
chr8	101670042	101670043	SNX31_c.53T>C
chr8	103326141	103326142	UBR5_c.1897C>T
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	14748353	14748354	GLRA2_c.1106G>A
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 2

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	153914584	153914585	DENND4B_c.815A>G
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	111592546	111592547	SIK2_c.1945-7T>G
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	6128781	6128782	VWF_c.3802C>T
chr12	49297659	49297660	RP11-302B13.5_c.416C>T
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	3350043	3350044	TIGD7_c.571T>A
chr16	20043670	20043671	GPR139_c.448G>T
chr16	56884141	56884142	SLC12A3_c.1763C>T
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	27910538	27910539	GIT1_c.148C>A
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	2425231	2425232	TMPRSS9_c.2848G>A
chr19	10257182	10257183	DNMT1_c.2738C>A
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr19	58355619	58355620	ZNF587B_c.1123-4A>G
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	46524991	46524992	EPAS1_c.-59C>A
chr2	74674512	74674513	SEMA4F_c.838C>T
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	220355503	220355504	SPEG_c.9211C>T
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	9781071	9781072	BRPF1_c.989G>A

chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	12606198	12606200	RAF1_c.680+1_680+2delinsAA
chr3	31532984	31532985	STT3B_c.-14C>T
chr3	38538023	38538024	EXOG_c.163+4del
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	132175358	132175359	DNAJC13_c.1114G>A
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr3	148928124	148928125	CP_c.436G>A
chr3	192635518	192635519	MB21D2_c.111C>A
chr4	144621305	144621306	FREM3_c.523C>A
chr4	186320165	186320166	ANKRD37_c.256A>G
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140453135	140453136	BRAF_c.1799T>A
chr7	140481381	140481448	BRAF_Region_3
chr7	149133764	149133765	ZNF777_c.1240G>A
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	101670022	101670024	SNX31_c.72_73del
chr8	144662754	144662755	EEF1D_c.1631G>A
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	102463218	102463219	TCP11X2_c.551G>T
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 6

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	204234215	204234216	PLEKHA6_c.281-46_281-43del
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	19097827	19097828	DUXAP9_n.831+4822C>A
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr14	74393822	74393823	ZNF410_c.*16+2301A>C
chr14	75165464	75165465	AC007956.1_c.258+2T>C
chr14	75509046	75509047	MLH3_c.3379+35C>T
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	51704248	51704249	SPACA6_c.731-21del
chr19	52207501	52207502	SPACA6P_n.1649-21del
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	237369138	237369139	COL6A3_c.4324G>A
chr22	15820191	15820192	DUXAP8_n.293+4626G>T
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111

chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr4	9706088	9706089	ALG1L3P_n.449T>G
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	132150421	132150422	SOWAHA_c.1109G>A
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	11192530	11192531	NEDD9_c.663+47C>G
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	97855975	97855976	TECPR1_c.2513+1350A>T
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140453135	140453136	BRAF_c.1799T>A
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	100709704	100709705	PABPC1_c.999del
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chr9	130587255	130587256	ENG_c.817-3_818del
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 10

Chromosome	Start	Stop	Description
chr1	86952287	86952288	CLCA1_c.1034G>A
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	57436899	57436900	MYO1A_c.1054C>T
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr14	55518449	55518450	MAPK1IP1L_c.-76C>T
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr17	74732965	74732966	SRSF2_c.277G>C
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr19	54080790	54080791	ZNF331_c.976_977insGAGGCAG
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	80085246	80085247	CTNNA2_c.407G>A
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	189851869	189851870	COL3A1_c.528+5G>A
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1

chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	87021895	87021896	CROT_c.1524G>C
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chr9	138377955	138377956	PPP1R26_c.1600A>T
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 8

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr10	14884778	14884779	HSPA14_c.221+2623A>T
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	3845286	3845287	PGAP2_c.523C>T
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	64037685	64037686	BAD_c.502C>T
chr11	89443694	89443695	TRIM77_c.229A>G
chr11	118889018	118889019	RPS25_c.-25C>T
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	9310427	9310428	PZP_c.3304G>C
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	8965985	8965986	KIDINS220_c.108+1130A>G
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	37441081	37441083	CEBPZ_c.2469_2470insTAC
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	52412635	52412636	DNAH1_c.7217C>G
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	57108278	57108279	SPATA12_c.557T>C
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111

chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr3	155461071	155461072	AC104472.1_c.320T>C
chr4	106157484	106157485	TET2_c.2386G>A
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140453135	140453136	BRAF_c.1799T>A
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	36661711	36661713	KCNU1_c.377+106_377+107del
chr8	38828062	38828063	PLEKHA2_c.*761G>C
chr8	38828067	38828068	PLEKHA2_c.*766T>C
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 11

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	144852449	144852450	PDE4DIP_c.7000-7C>T
chr10	22498524	22498525	EBLN1_c.388G>A
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr13	52649892	52649893	NEK5_c.1797+1G>A
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	40595627	40595628	PLCB2_c.163-71T>G
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	56913665	56913679	SLC12A3_c.*266_*279del
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr17	80786534	80786535	TBCD_c.1318+13725T>C
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr19	55759075	55759076	RFPL4A_c.-109C>G
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	167328883	167328884	SCN7A_c.515C>A
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	36700083	36700084	MYH9_c.2347A>G
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111

chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr3	185906085	185906086	DGKG_c.2000C>A
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	42073443	42073444	C6orf132_c.2206T>A
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	38828067	38828068	PLEKHA2_c.*766T>C
chr8	38828069	38828070	PLEKHA2_c.*768G>C
chr8	67590067	67590068	C8orf44_c.125T>C
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	17775315	17775316	SH3GL2_c.188-11063C>T
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 12

Chromosome	Start	Stop	Description
chr1	1223902	1223903	SDF4_c.392C>T
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	151748678	151748679	TDRKH_c.1110G>T
chr1	248273282	248273283	OR2T33_c.532G>A
chr10	74884915	74884918	NUDT13_c.510_512del
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	14498044	14498045	ATF7IP_c.3809A>G
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	77334265	77334266	ADAMTS18_c.2568G>A
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	16012187	16012188	NCOR1_c.2094A>T
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr17	78820579	78820580	RPTOR_c.1314+206T>C
chr18	5891187	5891188	TMEM200C_c.876C>G
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	50423600	50423601	SPIB_c.340-4G>A
chr19	51714069	51714070	HAS1_c.1094C>T
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	211441073	211441074	CPS1_c.259C>T
chr2	220283688	220283689	DES_c.505C>T
chr20	44747015	44747016	CD40_c.34G>A
chr20	57484420	57484421	GNAS_c.2531G>A
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111

chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	53855784	53855785	CHDH_c.874G>A
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	98133213	98133214	OR5H1_c.517A>G
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr3	154145546	154145547	GPR149_c.982-50del
chr4	110662036	110662037	CFI_c.*12C>G
chr4	143029052	143029053	INPP4B_c.2374+193T>C
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140453135	140453136	BRAF_c.1799T>A
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	30700895	30700896	TEX15_c.5638T>A
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	153780221	153780222	IKBKG_c.209A>T
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 16

Chromosome	Start	Stop	Description
chr1	26345989	26345990	CRYBG2_c.1640G>A
chr1	115256521	115256540	NRAS_Region_1
chr1	115256528	115256529	NRAS_c.182A>G
chr1	115258728	115258793	NRAS_Region_2
chr1	196695992	196695993	CFH_c.2159C>G
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	61906227	61906228	INCENP_c.1159G>A
chr11	68367883	68367884	PPP6R3_c.2114A>C
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	44714739	44714740	NELL2_c.1146G>T
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr13	39425118	39425133	FREM2_c.6618_6632del
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	67645892	67645893	CTCF_c.821C>T
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr16	86602512	86602513	FOXC2_c.*66C>A
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	36666450	36666451	ARHGAP23_c.3720del
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	10467662	10467663	AC011529.1_c.*42A>T
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	36223675	36223676	KMT2B_c.6226C>T
chr19	52033147	52033148	SIGLEC6_c.842del
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr19	58355619	58355620	ZNF587B_c.1123-4A>G
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	242054566	242054567	PASK_c.3224G>A
chr20	54972733	54972734	CSTF1_c.482A>G
chr22	32875189	32875190	rs11107_G_A_0.48742_1111

chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	8671350	8671351	SSUH2_c.587C>T
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr3	170744501	170744502	SLC2A2_c.-43A>C
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	57842954	57842955	CTD-2117L12.1_c.261+7A>G
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr5	140573078	140573079	PCDHB10_c.954T>A
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	96034576	96034577	MANEA_c.262G>A
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	22471636	22471637	CCAR2_c.737G>A
chr8	93029588	93029589	RUNX1T1_c.124C>G
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	116259674	116259675	RGS3_c.832C>T
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	41379817	41379818	CASK_c.2621C>T
chrX	51486886	51486887	GSPT2_c.165C>A
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 1

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	27876995	27876996	MRPS35_c.399G>A
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr14	23416462	23416463	HAUS4_c.840-15T>G
chr14	74393810	74393811	ZNF410_c.*16+2289T>C
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	1268197	1268198	CACNA1H_c.5446-12C>T
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	130365181	130365182	COL6A6_c.5239+1336C>G
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1

chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr5	140841426	140841427	PCDHA8_c.106C>T
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	70452451	70452452	FAM135A_c.78-41_78-40inv
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140453135	140453136	BRAF_c.1799T>A
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	141003330	141003331	SPANXB1_c.91-100A>C
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 3

Chromosome	Start	Stop	Description
chr1	65144100	65144101	CACHD1_c.3091+108G>A
chr1	112303833	112303834	DDX20_c.963-14G>C
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr10	63184517	63184518	JMJD1C_c.6961+90C>T
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	20259893	20259894	AC126603.1_n.1047+1978_1047+1979inv
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	50888599	50888600	TRPM7_c.3164-22T>G
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	43098168	43098169	CEACAM8_c.65-117_65-112del
chr19	47913310	47913311	MEIS3_c.598-507C>G
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	203404037	203404038	ABI2_c.1192+1304T>C
chr20	35720066	35720067	RBM39_c.825+1673_825+1674inv
chr20	35720078	35720079	RBM39_c.825+1661_825+1662inv
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	41636821	41636822	XRCC6_c.589+52A>G
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	98133213	98133214	OR5H1_c.517A>G

chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	124912185	124912186	SLC12A8_c.52-2821C>G
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	124798315	124798316	FAM91A1_c.962+335T>C
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chr9	136590543	136590544	SARDH_c.915+4343A>G
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	54449273	54449274	FGD1_c.2149-6C>G
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 21

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	158063137	158063138	KIRREL_c.1481C>A
chr1	158093347	158093348	KIRREL1_c.1481C>A
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	26734188	26734189	SLC5A12_c.404C>G
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	6567295	6567296	TAPBPL_c.904+386A>G
chr12	50502046	50502047	GPD1_c.953+134G>C
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24595024	24595025	TPTE2P6_n.985+662A>C
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr13	63746742	63746743	AL445989.1_c.-58C>T
chr13	63746782	63746783	AL445989.1_c.-18T>G
chr14	18974855	18974856	POTEM_c.811-1193T>A
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	68846519	68846520	TANGO6_c.94+2808_94+2809delinsCA
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	10480461	10480462	TYK2_c.194-1368T>C
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr21	45960109	45960110	TSPEAR_c.304-6304_304-6295del
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111

chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	130365181	130365182	COL6A6_c.5239+1336C>G
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr3	146102913	146102914	PLOD2_c.680-62_680-55del
chr3	146102920	146102921	PLOD2_c.680-69G>C
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_I1
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	140753335	140753336	BRAF_c.1919T>A
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	32632933	32632934	TAF1L_c.2644G>A
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	77442852	77442853	TRPM6_c.682T>C
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chr9	136215318	136215319	QSOX2_c.1210-15T>A
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	50350972	50350973	SHROOM4_c.3169G>A
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 4

Chromosome	Start	Stop	Description
chr1	1413817	1413819	ATAD3B_c.283-203_283-202del
chr1	43919313	43919314	HYI_c.151G>A
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr1	196977711	196977712	CFHR5_c.1609G>C
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	57451937	57451938	INHBC_c.*1916G>A
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr12	81768526	81768530	PPFIA2_c.1149_1152del
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr14	50952854	50952855	MAP4K5_c.224G>A
chr14	74393822	74393823	ZNF410_c.*16+2301A>C
chr15	23363714	23363715	GOLGA8S_c.1293G>C
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr19	17026770	17026771	CPAMD8_c.13G>T
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr2	201736197	201736198	PPIL3_c.418G>A
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111

chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr5	179287225	179287227	CTC-241N9.1_c.-84_-83insTGTGTGTGT
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	106836766	106836767	HBP1_c.1385+171C>G
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	120886146	120886147	DEPTOR_c.61G>A
chr8	145167348	145167349	KIAA1875_c.3022-1G>C
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chr9	132671110	132671111	FNBP1_c.1295+58del
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Patient study ID No 22

Chromosome	Start	Stop	Description
chr1	115256521	115256540	NRAS_Region_1
chr1	115258728	115258793	NRAS_Region_2
chr10	76855325	76855326	rs3740318_A_G_0.426318_1111
chr10	84745255	84745256	rs2295933_C_T_0.469649_1111
chr10	99160151	99160152	rs1048442_A_G_0.400559_1111
chr10	115350596	115350597	rs1885434_G_A_0.360024_1111
chr11	55606817	55606818	rs11231253_C_T_0.453075_1111
chr11	121367625	121367626	rs12364988_T_C_0.44369_1111
chr12	2997396	2997397	rs2907608_G_A_0.525958_1111
chr12	12633122	12633123	DUSP16_c.815+34T>G
chr12	72372861	72372862	rs7305115_A_G_0.541733_1111
chr13	24876751	24876752	rs33990382_T_C_0.452077_1111
chr14	21109140	21109141	rs17277522_C_T_0.53155_1111
chr14	58785271	58785272	ARID4A_c.287G>T
chr15	27772675	27772676	rs140679_C_T_0.516374_1111
chr15	52757506	52757507	ONECUT1_c.*48A>G
chr15	57918068	57918069	rs2069133_A_G_0.534545_1111
chr15	100801697	100801698	rs4965613_G_A_0.692093_1111
chr15	101910549	101910550	rs20543_G_A_0.451677_1111
chr16	83065663	83065664	rs6565105_G_A_0.534145_1111
chr17	33433309	33433310	RAD51D_c.636+95G>C
chr17	33881630	33881631	rs321612_T_C_0.627196_1111
chr17	39081712	39081713	rs8037_A_G_0.570887_1111
chr17	46925475	46925476	CALCOCO2_c.303G>T
chr17	48940421	48940422	rs4626_T_C_0.63139_1111
chr17	71233129	71233130	rs11869253_A_G_0.519569_1111
chr18	8069867	8069868	rs2230601_C_T_0.474441_1111
chr18	21798216	21798217	MIB1_c.1226A>G
chr19	35998361	35998362	rs4254439_T_G_0.384784_1111
chr19	52496148	52496149	rs16983353_C_T_0.477436_1111
chr2	27163043	27163044	rs3739085_T_C_0.608227_1111
chr2	179427185	179427186	rs2366751_A_G_0.508786_1111
chr22	32875189	32875190	rs11107_G_A_0.48742_1111
chr22	44372068	44372069	rs2073084_G_A_0.488019_1111
chr3	11887990	11887991	rs408600_C_G_0.434305_1111
chr3	38798522	38798523	SCN10A_c.1078C>T
chr3	43073760	43073761	rs658958_G_A_0.438299_1111
chr3	49051470	49051471	WDR6_c.2594T>C
chr3	56771250	56771251	rs3772219_A_C_0.414736_1111
chr3	122288209	122288210	rs2332285_G_A_0.545927_1111
chr3	124739891	124739892	rs6438874_G_A_0.509585_1111
chr3	141327473	141327474	rs295323_G_A_0.450479_1111
chr3	143293011	143293012	rs6763202_A_G_0.340655_1111

chr4	77065425	77065426	NUP54_c.171G>C
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366_l1
chr5	1295119	1295366	bed_target_0_chr5:1295119-1295366
chr5	1295208	1295291	TERT_Region_1
chr5	42719238	42719239	rs6180_A_C_0.444489_1111
chr5	135287028	135287029	rs31517_T_C_0.636382_1111
chr6	56417544	56417545	rs4715631_T_C_0.662141_1111
chr6	101094553	101094554	rs239239_A_G_0.571286_1111
chr7	50680561	50680562	GRB10_c.1096-26A>G
chr7	99696929	99696930	MCM7_c.373C>T
chr7	108212352	108212353	rs1043615_G_A_0.478435_1111
chr7	123264803	123264804	rs11769381_C_T_0.469649_1111
chr7	135406175	135406176	rs4596594_A_G_0.45028_1111
chr7	140453131	140453164	BRAF_Region_2
chr7	140481381	140481448	BRAF_Region_3
chr7	150491083	150491084	rs3173833_T_G_0.544129_1111
chr8	101199337	101199338	SPAG1_c.702-10T>C
chr9	4576679	4576680	rs301430_T_C_0.458267_1111
chr9	33167130	33167154	B4GALT1_c.14_37del
chr9	34397544	34397545	rs11790577_A_G_0.404353_1111
chr9	37745557	37745558	FRMPD1_c.3529C>G
chr9	73461336	73461337	rs7862440_T_A_0.520567_1111
chr9	128002558	128002560	HSPA5_c.382_383dup
chr9	130485617	130485618	rs537881_G_A_0.540735_1111
chrX	4429257	4429315	chrX_sexID2_SEX_ID
chrX	11314432	11314485	chrX_sexID1_SEX_ID
chrX	129804177	129804178	ENOX2_c.548-6A>G
chrY	6738552	6738553	chrY_sexID3_SEX_ID
chrY	19490214	19490215	chrY_sexID4_SEX_ID

Appendix E – Abstracts of accepted poster and oral presentations

American Thyroid Association annual conference, Oct 2022, Montreal, Canada. Poster abstract 1 – ID 487

Title

Genomic Signature That Predicts the Presence of Vascular Invasion and Other High Risk Histopathological Features in Differentiated Thyroid Cancer.

Objectives & Methods

Vascular invasion (VI) is one of a few key histopathological features that can upstage differentiated thyroid carcinomas (DTCs) from low to higher risk tumours when considering risk of recurrence or distant metastatic spread. This can influence the decision for clinicians to recommend completion thyroidectomy after initial hemi-thyroidectomy. Predictive biomarkers obtained from biopsy samples could provide additional prognostic stratification and assist in surgical decision making.

We conducted a retrospective, observational cohort study in adult patients with stage I-IV differentiated thyroid cancer (DTC), with and without VI. Total DNA and RNA from Formalin Fixed Paraffin Embedded (FFPE) DTC samples underwent targeted high depth sequencing using a panel enriched for somatic mutations in thyroid cancer, as well as differential miRNA expression profiling using a combination of Nanostring and Reverse Transcription PCR (RT-PCR) techniques.

Results

We identified one or more somatic mutation(s) in 73 out of 118 (61.9%) DTCs. Twenty-five samples had 2 or more variants detected, and these samples were significantly associated with aggressive pathological features such as disease recurrence (OR 3.40, 95% CI 1.21-10.47, p = 0.02), vascular invasion (OR 3.44, 95% CI 1.16-10.44, p = 0.02) and distant metastases (OR 14.84, 95% CI 2.05-173.1, p = 0.0025). *TERTp* c.1-124C>T was significantly associated with the presence of VI (OR 14.77, 95% CI 2.15 – 169.4, p = 0.0067). Extra-thyroidal extension (ETE) was significantly associated with upregulation of hsa-miR-146a-5p and downregulation of hsa-miR-204-5p (p = 0.019 and 0.01 respectively). Upregulation of hsa-miR-146a-5p was significantly associated with the presence of VI (p = 0.02). By combining the dysregulation status of hsa-miRs 146a-5p, 204-5p and 221-3p with the somatic mutation status of *BRAF*, *RAS* and *TERTp*, we designed a predictive model that correctly identified DTC samples with vascular invasion with PPV of 70.9% and NPV of

84.2%, with area under the ROC curve of 0.84 (p < 0.0001, 95% CI 0.75-0.94). This model correctly classifies samples 78.3% of the time.

Conclusions

We propose a genomic signature that combines the presence of hotspot somatic mutations and key miRNA dysregulation statuses to predict the presence of vascular invasion with high sensitivity and specificity. Further multi-centre prospective clinical trials are recommended to validate our findings.

Keywords

DTC

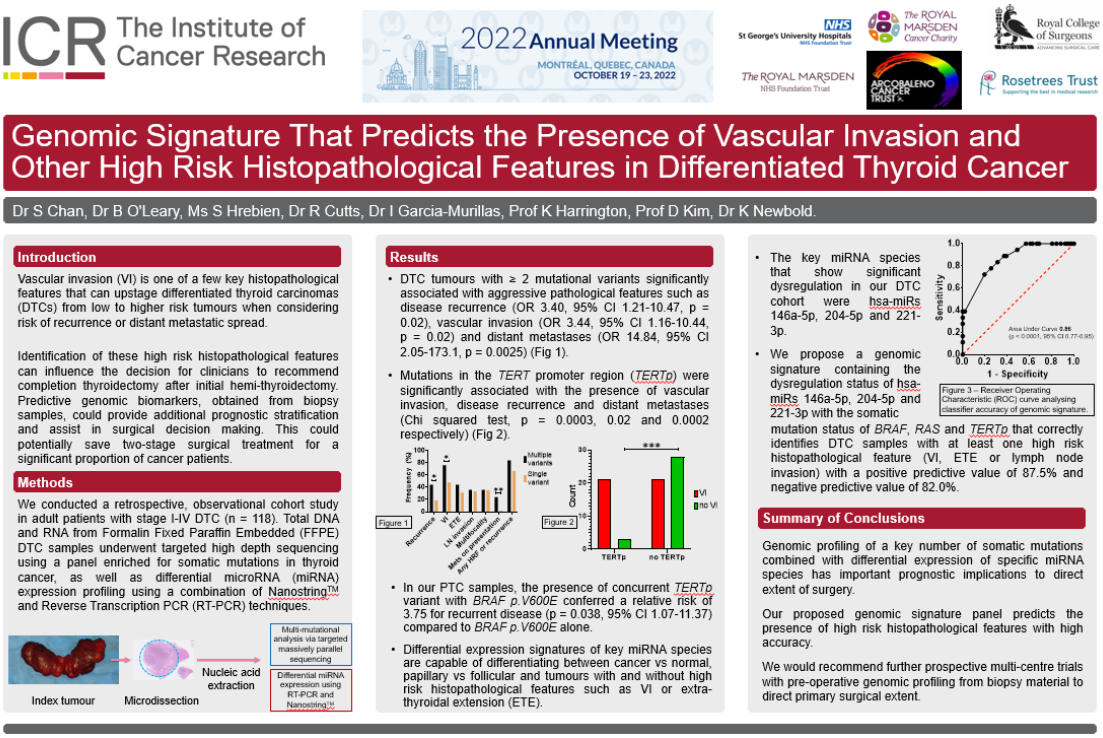
Vascular invasion

Genomic biomarkers

TERT promoter

microRNA

POSTER



American Thyroid Association annual conference, Oct 2022, Montreal, Canada. Oral and poster abstract 2 – ID 268

Title

Personalised Circulating Tumour DNA Analysis In Differentiated Thyroid Cancer Using Next Generation Sequencing Panels.

Objectives & Methods

Numerous studies have shown the clinical utility of circulating tumour DNA (ctDNA) as a non-invasive, cancer-specific biomarker of disease. Techniques such as droplet digital PCR (ddPCR) assays have shown exquisite sensitivity for ctDNA detection but are limited by multiplexing capabilities. Newer next generation sequencing techniques could improve the detection of ctDNA and therefore the clinical viability in routine oncological practice.

We conducted a multi-centre, prospective, observational cohort study to investigate ctDNA in adult patients with stage I-IV differentiated thyroid cancer (DTC). Total DNA from their respective index tumours underwent targeted high depth sequencing using a panel enriched for somatic mutations in thyroid cancer. Bespoke singleplex droplet digital PCR (ddPCR) assays were designed to track ctDNA in matched peri-treatment plasma samples. A sub-cohort of patient samples underwent concurrent whole exome sequencing (WES) to inform the design of highly sensitive personalised sequencing panels for tumour-specific variants using LiquidPlex Dx^{TM} technology (ArcherDX/Invitae), to analyse matched serial pre- and post-operative plasma samples for evidence of residual or recurrent disease.

Results

We identified one or more somatic mutation(s) in 27 out of 47 (57.4%) DTCs, with two or more mutations found in 8 cases. The most frequently identified variant was *BRAF c.T1799A (p.V600E)* found in 19 out of 26 (73.1%) DTCs. Of the 27 patients with a variant detected by targeted high depth sequencing, 11% had detectable ctDNA in at least one plasma timepoint using ddPCR. In the sub-cohort that underwent WES, trackable somatic mutations were identified in 15 of 15 patients. Bespoke patient specific sequencing panels comprising 71-84 targets per panel, increased the rate of ctDNA detection in at least 1 peri-operative plasma timepoint to 54.5%. The median detected mutation allele fraction was 0.28% (range 0.02 – 12.13%). Patients with AJCC stage 2 and above all had detectable ctDNA in at least 1 timepoint.

Conclusions

The findings of this exciting pilot study support the use of personalised sequencing panels for ctDNA detection in AJCC stage II DTC and above. Further multi-centre long term (> 5 year) clinical prospective trials are recommended to validate these findings and determine if detection of ctDNA predicts disease free survival or overall survival.

Keywords

Liquid biopsy

Bespoke sequencing panel

Droplet digital PCR

Circulating tumour DNA

Differentiated thyroid cancer

POSTER